Transport Policy 69 (2018) 106–121

Contents lists available at ScienceDirect

Transport Policy

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

Modeling departure time choice of metro passengers with a smart corrected mixed logit model - A case study in Beijing

Haiying Li[a], Xian Li[a], Xinyue Xu[a,*], Jun Liu[b], Bin Ran[c]

[a] State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, 100044, China
[b] School of Traffic and Transportation, Beijing Jiaotong University, Beijing, 100044, China
[c] Department of Civil and Environmental Engineering, University of Wisconsin-Madison, WI 53706, USA

ARTICLE INFO

Keywords:
Metro
Departure time choice
Mixed logit
Passenger flow control
Price endogeneity

ABSTRACT

It is critical to improve the effectiveness of demand management in metro systems with passenger departure time choice exactly learned during peak hours. In this study, a practical framework is developed to model departure time choice of metro passengers during peak hours. First, various attributes that influence departure time choice of metro passengers are investigated and the technique for order preference by similarity to ideal solutions (TOPSIS) is used to identify these main attributes. Then, a mixed logit (ML) model of departure time choice that accounts for price endogeneity is developed. To calibrate the model, a stated preference (SP) survey based on D-efficient design is conducted in the Beijing metro system. It is proved that the corrected ML model outperforms the uncorrected ML model according to the collected 1152 sample data. An elasticity analysis of these main attributes is further conducted, which indicates that metro fare and departure time change influence passenger departure time choice more than crowdedness in Beijing metro. Knowledge of these preferences assists traffic managers in balancing passenger departure time to mitigate overcrowding during peak hours. Heterogeneity of passenger socioeconomic and trip characteristics is also concerned taking advantage of ML model. Finally, a ML-based fare discount strategy to ease the crowdedness in Batong Line of Beijing metro is presented and evaluated via an existing simulation tool.

1. Introduction

Overcrowding in metros is a potential safety hazard, and the introduction of passenger flow control measures during peak hours has become common in China (Xu et al., 2016), as well as other countries (Sadhukhan et al., 2016; Tirachini et al., 2013). In 2016, average daily passenger volume of Beijing metro reached 8.2 million and train loading factor of 10 lines during peak hours were over 100%. There are many methods to manage demand on metros such as diverting passengers to other alternative travel modes (Kamargianni et al., 2015; Kou et al., 2017; Zou et al., 2015), rescheduling departure time during peak hours, and coordinating capacity with demand (Jiang et al., 2017; Xu et al., 2016, 2018). Particularly, the demand characteristic is the fundament of these methods given that demand significantly exceeds capacity during peak hours in Chinese metros. Therefore, learning the features of travel demand, especially passenger behavior of departure time choice is urgently required for addressing the problem of metro crowding.

Some studies have investigated the factors affecting departure time choice in related transportation problems (Jou, 2001; Kim et al., 2015; Nurul Habib et al., 2009; Saleh and Farrell, 2005; Thorhauge et al., 2016; Vovsha et al., 2014; Wang et al., 2017; Zou et al., 2015). However, scant research is on the factors affecting departure time choice in metros, especially in China. Furthermore, the aforementioned foundings for other travel modes may not apply to metros because of the significant differences in the demand features, network topology, and operational organization. Therefore, a comprehensive survey to identify the relevant attributes is necessary.

Departure time choice models, which decide an individual's choice of timing or an interval to take a trip, have been the subject of considerable interest. Many studies of departure time choice have employed stochastic models including multinomial models (Nurul Habib et al., 2009; Small, 2001), nested models (Hess et al., 2007), cross-nested models (Lemp et al., 2010), mixed logit models, probit models (Jou, 2001), and variations of the ordered generalized extreme value model (Chu, 2009). Hereinto, mixed logit (ML) model has significant advantages (Train, 2009) in the following aspects: (1) it is easier to simulate straightforward derivation and choice probabilities with an ML model than with a probit model because the former does not require a constrained normal distribution; (2) ML models can include various

* Corresponding author.
E-mail addresses: hyli@bjtu.edu.cn (H. Li), 10114213@bjtu.edu.cn (X. Li), xxy_1983@163.com (X. Xu), jliu@bjtu.edu.cn (J. Liu), bran@engr.wisc.edu (B. Ran).

https://doi.org/10.1016/j.tranpol.2018.06.005
Received 15 June 2017; Received in revised form 1 June 2018; Accepted 7 June 2018
Available online 18 June 2018

types of correlations; (3) ML models are better at accounting for random preference heterogeneity in estimating model parameters. Therefore, the ML approach is preferred for modeling departure time choice (de Jong et al., 2003; Thorhauge et al., 2015). However, the ML approach has not been used to model departure time choice for metro systems despite of the heterogeneity of passengers.

In typical departure time choice models, price is usually assumed to be exogenous. While in fact, price endogeneity exists in demand models considering that price and demand interact with each other. The omission of price endogeneity will lead to inconsistent estimates and biased results (Lurkin et al., 2017; Mumbower et al., 2014). Fortunately, there are multiple methods to correct for price endogeneity (Guevara, 2015). A fare discount strategy, where passengers enjoy 50% fare discount if they enter a station before 7:00am, has been employed both in Batong Line and Changping Line of Beijing metro to relieve the crowded stations since 2015. Thus, this study will address the price endogeneity in passenger departure time model for crowded metros.

How to collect the data to calibrate a departure time choice model is another topic in the field of metro. Two specific approaches, which are revealed preference (RP) and stated preference (SP), are commonly used to collect data (He, 2013; Sadhukhan et al., 2016; Saleh and Farrell, 2005; Zou et al., 2015). SP survey provides data on multiple options and subject desires while RP survey provides information about single options (Hensher et al., 2005). Therefore, SP surveys is adopted in this study.

Further, designing stated choice (SC) experiments has been an increasingly important area of SP surveys. Orthogonal designs is widely used to generate the choice situations presented to respondents. However, discrete choice models, whose nonlinearity is a significant feature in departure time choice, have been called into question in certain studies (Bliemer and Rose, 2011; Rose and Bliemer, 2009; Skirton et al., 2012). D-efficient design attracts increasing interest as a result of the Fisher information maximization on the respondent preferences obtained from observations (Bliemer and Rose, 2011; Rose et al., 2008). Rose et al. (2008) suggested that D-efficient designs had greater statistical efficiency than orthogonal designs. In an extensive empirical study of air travel choice behavior, Bliemer and Rose (2011) found that D-efficient designs had lower standard errors in estimating parameters than a conventional orthogonal design while requiring smaller sample sizes. However, few studies focus on efficient designs in experiments of departure time choice for metro systems.

To our knowledge, few studies attempt to model and calibrate departure time choice behavior of metro passengers. Specifically, the effect of attribute levels on departure time choice is not quantitatively examined, and the price endogeneity in departure time choice of metro passengers is not explored. In addition, efficient designs for departure time choice behavior have not been investigated.

The primary goal of this paper is to provide an integrated framework for modeling departure time choice where TOPSIS, D-efficient design for an SP survey, and correct ML models are integrated. To verify our framework, the effects of travel time saving, crowdedness, and metro fare are examined using TOPSIS. A ML-based departure time choice model is subsequently constructed that employs a two-stage control function (2SCF) method to address price endogeneity. To calibrate the proposed ML model, a D-efficient design survey collecting data is conducted in the Beijing metro system. Then, the elasticity and heterogeneity of each attribute for various groups of passengers are further discussed. Finally, this research proposes a model application in passenger flow control, which contributes to effective strategies for passenger flow management.

The remainder of the paper is structured as follows: A ML model that corrects for price endogeneity is presented in Section 2. The SP survey design is detailed in Section 3, the model development and results are discussed in Section 4. A model application for passenger flow control is proposed in Section 5. And Section 6 summarizes the significant findings and suggests possibilities for future research.

2. Theoretical background

A brief description of the corrected ML model which accounts for price endogeneity is given in this section.

The two-stage control function (2SCF) method, a typical CF method efficiently dealing with price endogeneity (Lurkin et al., 2017), is employed to correct for price endogeneity. This method consists of two key stages as follows:

Stage 1 Estimate price by ordinary least square (OLS)

$$p_{nj} = \widehat{\boldsymbol{\beta}_n}\widehat{\mathbf{X}}_{nj} + \alpha_1 IN_{nj}^1 + \cdots + \alpha_k IN_{nj}^k + ASC \tag{1}$$

where p_{nj} is the average price associated with alternative j for individual n, $\widehat{\boldsymbol{\beta}_n}$ is the vector of coefficients associated with all exogenous regressors, $\widehat{\mathbf{X}}_{nj}$ is the vector of variables except for price variables, α_k is the parameter associated with the kth instrument variable, IN_{nj}^k is the kth instrumental variable included in the price equation for alternative j for individual n, ASC is a constant term. Note that the instrument variables are selected according to two rules: (1) instrument variables are correlated with the endogenous variable; (2) instrument variables are independent of the error term (Lurkin et al., 2017; Vij and Walker, 2014; Walker et al., 2011).

Stage 2 Estimate the choice model using the residuals from **Stage 1**.

If an individual n faces an alternative j ($j = 1,2, .., J_t$) from a choice set t ($t = 1,2, .., T$), the corresponding utility, according to random utility theory, may be expressed as:

$$U_{njt} = \boldsymbol{\beta}_n \mathbf{x}_{njt} + \beta_p p_{nj} + \beta_{\widehat{\delta}} \widehat{\delta} + \varepsilon_{njt} \tag{2}$$

where $\mathbf{x}_{njt}$ is the vector of observed variables in a given choice set that relates the alternative and a respondent, $\boldsymbol{\beta}_n$ is a vector of the coefficients for preferences of individual n, β_p is the coefficient associated with price from **Stage 2**, $\widehat{\delta}$ is the difference between actual and predicted price from **Stage 1** ($\widehat{\delta} = p_{nj} - \widehat{p_{nj}}$), $\beta_{\widehat{\delta}}$ is the coefficient associated with the difference between actual and predicted prices from **Stage 1**, and ε_{njt} is a random term with an independently and identically distributed (IID) extreme value distribution (Train, 2009).

To obtain the error components of various correlated alternatives, the ML model incorporates additional stochastic elements in β_n ($\forall \beta_n \in \boldsymbol{\beta}_n$) as follows:

$$\beta_n = (\beta + \Delta z_n + \eta_n) \tag{3}$$

where β is the mean estimate of the parameter, η_n is a random term whose distribution over a sample population depends on underlying parameters, and z_n gives the observed variables of individuals that affect β_n. In this study, $\mathbf{x}_{njt}$ is the vector of attributes that influence the decision of departure time, z_n is the vector of socioeconomic and trip characteristics of metro commuters, and twelve choice sets containing three alternatives of departure time choice will be given in Section 3.

The ML model assumes a general distribution for each β_n and an IID extreme value type I distribution for ε_{njt}. If $\boldsymbol{\beta}_n$ is known, then the probability of a particular choice is a standard logit model. That is, the probability of choosing alternative i in a choice set t as a function of β_n is given by (Train, 1999):

$$P_{nit}(\boldsymbol{\beta}_n) = \frac{\exp(\boldsymbol{\beta}_n \mathbf{x}_{nit})}{\sum_{j=1}^{J_t} \exp(\boldsymbol{\beta}_n \mathbf{x}_{njt})}, \forall\, i \in J_t \tag{4}$$

If $\boldsymbol{\beta}_n$ is unknown, the ML probability is the expected value of the logit probability over all the possible values of $\boldsymbol{\beta}_n$.

$$P_{nit} = \int_{\boldsymbol{\beta}_n} P_{nit}(\boldsymbol{\beta}_n) f(\boldsymbol{\beta}_n|\Omega, z_n) d\boldsymbol{\beta}_n = \int_{\boldsymbol{\beta}_n} \frac{\exp(\boldsymbol{\beta}_n x_{nit})}{\sum_{j=1}^{J_t} \exp(\boldsymbol{\beta}_n x_{njt})} f(\boldsymbol{\beta}_n|\Omega, z_n) d\boldsymbol{\beta}_n \tag{5}$$

where $f(\boldsymbol{\beta}_n|\Omega, z_n)$ denotes the marginal joint density of the random coefficients $\boldsymbol{\beta}_n$ with Ω, which are the fixed parameters of the distribution.

A normal distribution and a maximum likelihood estimator are