

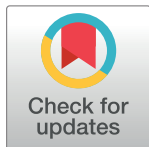
RESEARCH ARTICLE

# A constrained multinomial Probit route choice model in the metro network: Formulation, estimation and application

Yongsheng Zhang<sup>1</sup>, Enjian Yao<sup>2\*</sup>, Heng Wei<sup>3</sup>, Kangning Zheng<sup>4</sup>

**1** School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China, **2** MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing, China, **3** Advanced Research in Transportation Engineering and Systems (ART-EngineS) Laboratory, College of Engineering and Applied Science, The University of Cincinnati, Cincinnati, United States of America, **4** School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

\* [enjiyao@bjtu.edu.cn](mailto:enjiyao@bjtu.edu.cn)



## Abstract

Considering that metro network expansion brings us with more alternative routes, it is attractive to integrate the impacts of routes set and the interdependency among alternative routes on route choice probability into route choice modeling. Therefore, the formulation, estimation and application of a constrained multinomial probit (CMNP) route choice model in the metro network are carried out in this paper. The utility function is formulated as three components: the compensatory component is a function of influencing factors; the non-compensatory component measures the impacts of routes set on utility; following a multivariate normal distribution, the covariance of error component is structured into three parts, representing the correlation among routes, the transfer variance of route, and the unobserved variance respectively. Considering multidimensional integrals of the multivariate normal probability density function, the CMNP model is rewritten as Hierarchical Bayes formula and M-H sampling algorithm based Monte Carlo Markov Chain approach is constructed to estimate all parameters. Based on Guangzhou Metro data, reliable estimation results are gained. Furthermore, the proposed CMNP model also shows a good forecasting performance for the route choice probabilities calculation and a good application performance for transfer flow volume prediction.

## OPEN ACCESS

**Citation:** Zhang Y, Yao E, Wei H, Zheng K (2017) A constrained multinomial Probit route choice model in the metro network: Formulation, estimation and application. PLoS ONE 12(6): e0178789. <https://doi.org/10.1371/journal.pone.0178789>

**Editor:** Xiaolei Ma, Beihang University, CHINA

**Received:** December 16, 2016

**Accepted:** May 18, 2017

**Published:** June 7, 2017

**Copyright:** © 2017 Zhang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within its Supporting Information files.

**Funding:** This work is supported by Beijing Natural Science Foundation (grant number 8171003). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

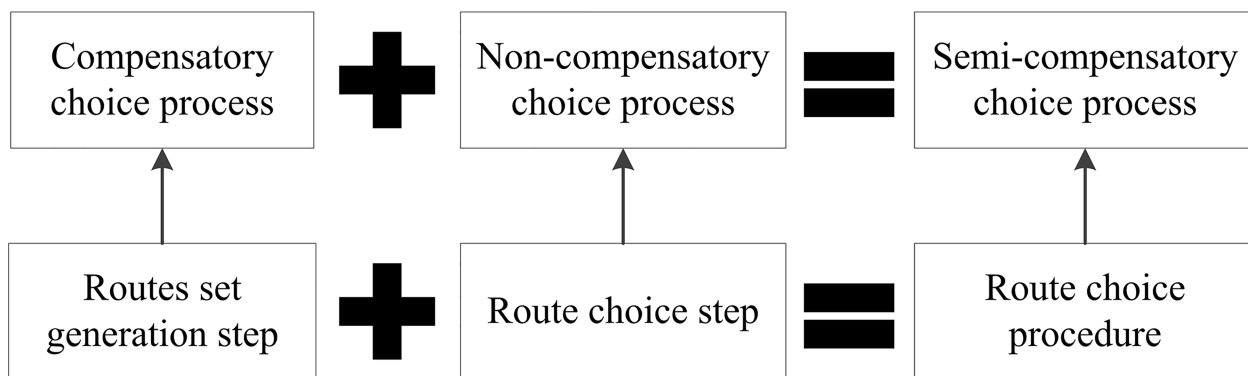
With new lines put into operation almost every year, the large scaled metro system has formed in some major cities in China, such as Beijing, Shanghai, Guangzhou and Shenzhen. Taking Guangzhou Metro for instance, up to the year of 2014, it is the sixth busiest metro system in the world and the third largest metro network in China with 9 lines, 164 stations including 21 transfer stations, and 260.5 km of tracks. In a large scaled metro network, the large number of transfer stations which brings plenty of routes for some origin-destination (OD) pairs increases the complexity of route choice modeling. Usually, according to a specific scheme, an

individual chooses the best route among many alternative routes with comprehensive consideration of multiple factors, including the variables denoting the level of service of metro system, such as in-vehicle travel time, number of transfers, transfer time, congestion level, etc. and the variables describing the influence of topological structure of the metro network and route direction on passengers' route choice preferences, such as angular cost [1–2].

The complex nature of route choice process responses to a large scaled metro system has brought challenges in establishing route choice model to reveal realistic behavioral decisions in the actual route choice process. Traditionally, with respect to route choice in a large scaled metro network, a route is chosen from a route set which is derived from attributes' limitations, such as travel time and number of transfers. For example, for an OD pair, if the shortest travel time of one route is 30min, it is a common sense that passengers will not consider the route with more than 60min travel time. In this case, the 60min travel time is the limitation and the routes with less than 60min travel time constitutes the routes set. However, route choice and routes set steps are usually carried out separately and independently in the metro system which leads to losing the consistence between the two steps. For the two steps, the route choice step is a compensatory choice process which focuses on calculating the trade-offs among multiple influencing factors and the routes set generation step is a non-compensatory process which pays attention to the cut-offs associated with the attributes' limitations. The non-compensatory behavior has been proved in the choice process [3–5]. And semi-compensatory route choice modelling which combines the routes set generation and route choice steps has attracted much more attention. The relationships between semi-compensatory, compensatory and non-compensatory choice processes are shown in Fig 1.

Meanwhile, route over-lapping problem in the large scaled metro network has already been figured out by Yai et al. [6]. Especially for some OD pairs with long direct distance, the fact that some alternative routes will share some links brings the correlation among the routes. Although most Logit-based models were satisfactory in representing route choice behavior associated with route over-lapping problem, they were still the approximate responses to the real behavior. In order to exactly express the interdependency among alternatives, Probit model [7] is more suitable though its estimation is a little harder than Logit models. Faced with elaborate operational requirements and services, the operational department is looking forward to a more advanced route choice model to reveal passengers' actual route choice behaviors so as to support personalized travel service and travel demand prediction.

Therefore, it is necessary to establish a semi-compensatory Probit route choice model and design an easier estimation approach for practice. In this paper, a constrained multinomial



**Fig 1. The relationships of different choice processes.** In the figure, 'A+B = C' means C is the combination of A and B. The arrow displays the one-to-one match.

<https://doi.org/10.1371/journal.pone.0178789.g001>

Probit route choice model is proposed to reveal the realistic route choice process along with the estimation approach, focusing on analyzing the semi-compensatory choice behavior and representing the interdependency among alternative routes.

## Literature review

Route choice model based on random utility maximization (RUM) theory [8] mainly consists of two types, including Logit and Probit models. Among various models, Multinomial Logit (MNL) model [9] is the most widely used due to its easy estimation and application. Ramming, Raveau et al., Zhang et al. and Liu et al. successfully analyzed route choice behavior with the consideration of level of service, social demographics, travel purpose and route direction based on MNL model [1, 2, 10–11]. But the assumption that the error component follows an identical and independent (IID) Gumbel distribution induces many weaknesses. In order to alleviate one of the weakness, known as route over-lapping problem which is caused by the interdependency among routes, many extended Logit models are developed, such as C-Logit [12], Path Sized Logit (PSL) [13–14], Paired Combinatorial Logit (PCL) [15–16], Cross Nested Logit (CNL) [17], Generalized Nested Logit (GNL) [18], Mixed Logit [19], etc. For the application in the metro network, Raveau et al. applied successfully C-Logit model to analyze passengers' route choice preferences [20]. However, Logit models cannot avoid IID distribution assumption to formulate probability equation with closed form, weakening the interdependency among alternative routes, while Probit model [7] can reflect deeply the interdependency by covariance, closer to passenger's actual route choice behavior. Yai et al. proposed a Probit model with structured variance to analyze route choice behavior in the railway network, saving the computational time to an extent [6].

Those models mainly focus on the process that an individual chooses the best route from a given routes set. The consistence between the routes set generation and route choice processes is usually neglected for metro passengers. Considering the interplay between the two sub-processes, Zhang et al. [21] successfully introduced constrained multinomial logit (CMNL) model [22–23] into route choice modeling in the metro network to analyze passengers' semi-compensatory choice behavior. Semi-compensatory models combining compensatory and non-compensatory behaviors have been paid more and more attention [24–26]. As one of two major approaches, the two-stage approach is widely used, consisting of two stages: generating all possible consideration routes sets and then choosing routes from the generated routes sets [27]. The consideration routes set is a subset of master routes set which is limited by some specific attributes. The two-stage approach is attractive that different models are allowed to explain each stage and many successful applications have already been found in the literature [28–30]. However, it leads to computational complex because too many consideration routes sets need to be constructed from master routes set [31]. And it also would have no sufficient robustness of choice prediction at the level of individual sets [32]. In order to avoid such a complex combinatorial number of choice sets, a kink called non-compensatory component is added to utility function, known as the second semi-compensatory choice modeling approach [33–34]. The non-compensatory component won't affect the utility when the attribute value lies in the domain, while it will negatively and significantly affect the utility when the attribute value exceeds the threshold. The non-compensatory component simplifies the structure of semi-compensatory choice model and saves the computational time by avoiding huge number of consideration choice sets compared with the two-stage approach. However, in those researches, kinks in the utility function make it non-differentiable at the cut-off, which is difficult to be applied in equilibrium and optimization processes.

To solve this problem, a constrained multinomial logit (CMNL) route choice model [21] is developed where the non-compensatory component is a continuous function, making utility function differentiable at the cutoff. However, its error component still follows (IID) Gumbel distribution and the route over-lapping problem in the route choice context still needs to be solved. Moreover, compared with the Gumbel distribution, the normal distribution is more approximate to the actual distribution of error component. To address the aforementioned problems, a study for developing a constrained multinomial probit (CMNP) route choice model for metro passengers is proposed in the paper. In this model, the error component follows the normal distribution instead of the IID Gumbel distribution to avoid the weaknesses. The correlations among alternative routes are measured by the covariance matrix.

The following sections of this paper are organized as follows: Section 3 introduces the CMNP route choice modeling methodology; Section 4 is about the estimation approach which is carried out based on MCMC method after transforming the CMNP model into Bayesian formulation; in Section 5, the CMNP model is estimated by the proposed estimation approach based on surveyed RP data in Guangzhou Metro and applied in forecasting the transfer passengers volumes; Section 6 is the conclusions.

## Modeling methodology

Based on random utility theory, the utility function with constrained characteristic attributes mainly consists of three parts: compensatory, non-compensatory and error components.

$$U_{n,k}^{rs} = V_{n,k}^{rs} + C_{n,k}^{rs} + \varepsilon_{n,k}^{rs} \quad (1)$$

where with respect to route  $k$  for OD pair  $rs$ ,  $U_{n,k}^{rs}$  is the generalized utility perceived by passenger  $n$ ;  $V_{n,k}^{rs}$  is the compensatory component;  $C_{n,k}^{rs}$  is non-compensatory component;  $\varepsilon_{n,k}^{rs}$  is the random error component.

The compensatory component  $V_{n,k}^{rs}$  is a trade-off function of characteristic attributes, including level of service variables, network topology, etc. This function represents the compensatory trade-offs among attributes. For simplicity, the compensatory component is defined as a linear function of attributes as shown below.

$$V_{n,k}^{rs} = \sum_{h=1}^H \theta_{n,h} X_{k,h} \quad (2)$$

where  $H$  is the number of characteristic attributes;  $X_{k,h}$  denotes the attributes;  $\theta_{n,h}$  is the corresponding parameters needed to be estimated.

The non-compensatory component  $C_{n,k}^{rs}$  is a cut-off function which should satisfy below conditions:

1. It should be a continuous function which guarantees the application in traffic equilibrium and optimization process.
2. If a constrained attribute value of one route exceeds threshold, non-compensatory component will let the route utility tend to be negative infinity. Otherwise, non-compensatory component tends to be zero.

Therefore, the non-compensatory component  $C_{n,k}^{rs}$  is formulated in this paper as follows:

$$C_{n,k}^{rs} = \ln \phi_{n,k}^{rs} \quad (3)$$

where  $\phi_{n,k}^{rs}$  is a continuous function limited in  $(0, 1)$ . The part  $\phi_{n,k}^{rs}$  can be formulated as a

probability function which represents the probability that route  $k$  is considered after the comparison between the constrained attributes of route  $k$  and the corresponding thresholds. Usually, more than one attribute is the constraint, according to conjunctive screen rule [35],  $\phi_{n,k}^{rs}$  can be defined as follows:

$$\phi_{n,k}^{rs} = \prod_{i=1}^{I_n} \varphi_{n,k}^{rs}(X_{k,i}^{rs}) \quad (4)$$

where for OD pair  $r-s$ ,  $X_{k,i}^{rs}$  is the constrained attribute  $i$  of route  $k$ ;  $I_n$  is the number of attributes constrained by individual  $n$ ; the function  $\varphi_{n,k}^{rs}(X_{k,i}^{rs})$  measures the considered probability influenced only by characteristic attribute  $X_{k,i}^{rs}$ . To specify this function, we can assume a scenario that a constrained attribute  $X_{k,i}^{rs}$  with perceived error  $\Psi_1$  has an upper bound  $b_{n,i}^{rs}$  with perceived error  $\Psi_2$  (if it is a lower bound, the sign will reverse), then the function  $\varphi_{n,k}^{rs}(X_{k,i}^{rs})$  can be calculated by

$$\begin{aligned} \varphi_{n,k}^{rs}(X_{k,i}^{rs}) &= P(X_{k,i}^{rs} + \psi_1 \leq b_{n,i}^{rs} + \psi_2) \\ &= P(\psi_1 - \psi_2 \leq b_{n,i}^{rs} - X_{k,i}^{rs}) \end{aligned} \quad (5)$$

where for OD pair  $r-s$ ,  $b_{n,i}^{rs}$  is the threshold of attribute  $i$  constrained for individual  $n$ .

In this paper, we assume that the perception errors of the constrained attribute and threshold both follow the normal distributions and are independent from each other. Further, the errors  $\Psi_1$  and  $\Psi_2$  respectively follow normal distributions  $\psi_1 \sim N(0, \sigma_1^2)$  and  $\psi_2 \sim N(\gamma_i, \sigma_2^2)$ , where  $\sigma_1^2$  and  $\sigma_2^2$  are the variances, and  $\gamma_i$  is the mean denoting the location parameter. According to the property of normal distribution,  $\Psi_1 - \Psi_2$  still follows normal distribution, that is  $\Psi_1 - \Psi_2 \sim N(-\gamma_i, \sigma^2)$ , where  $\sigma^2 = \sigma_1^2 + \sigma_2^2$ . And the function  $\varphi_{n,k}^{rs}(X_{k,i}^{rs})$  is equal to

$$\begin{aligned} \varphi_{n,k}^{rs}(X_{k,i}^{rs}) &= \Phi\left[\frac{b_{n,i}^{rs} - X_{k,i}^{rs} + \gamma_i}{\sigma}\right] \\ &= \Phi[\omega_i \cdot (b_{n,i}^{rs} - X_{k,i}^{rs} + \gamma_i)] \end{aligned} \quad (6)$$

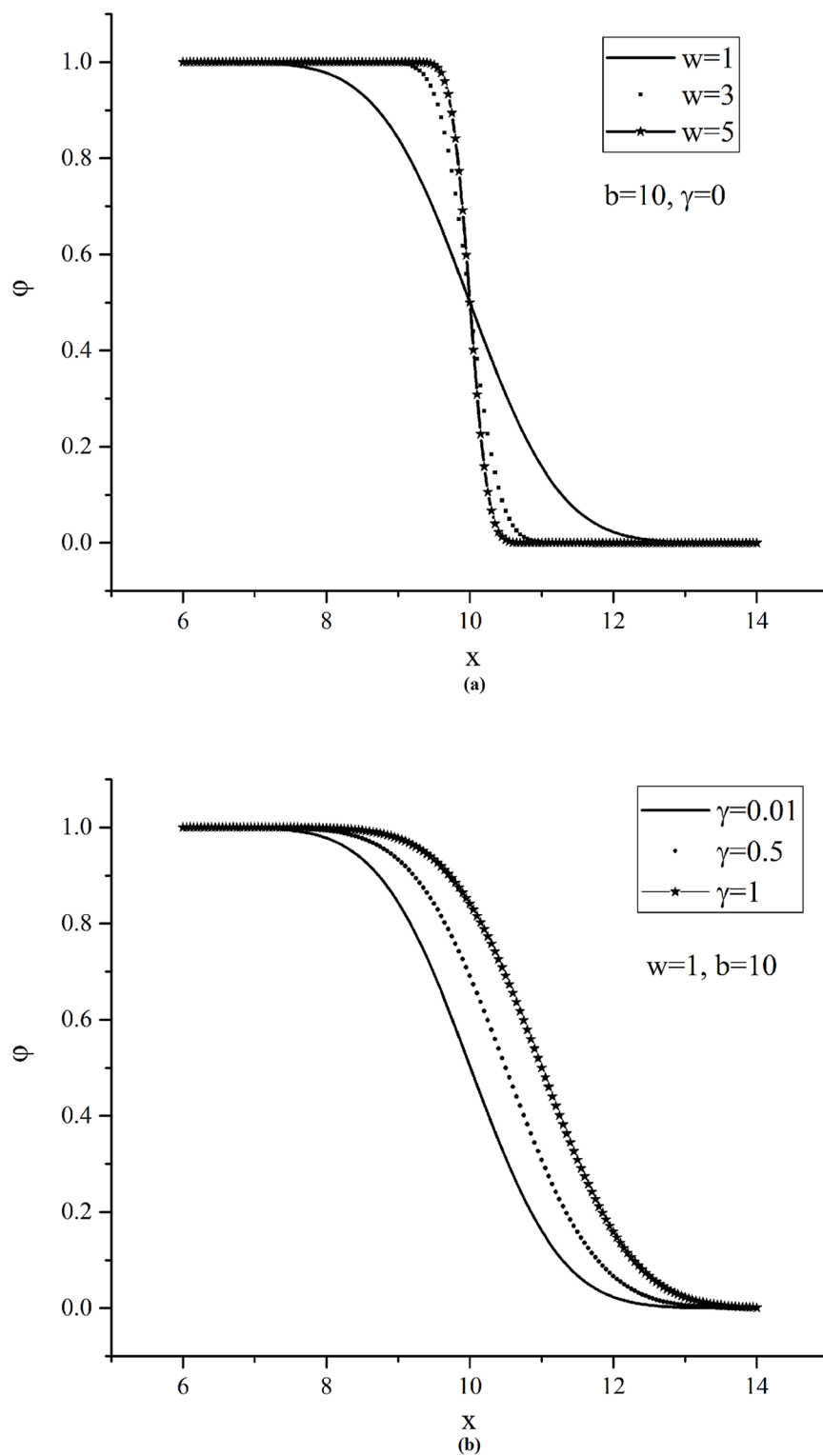
where  $\omega_i$  is the scale parameter related to the variance ( $\omega_i > 0$ , if  $b_{n,i}^{rs}$  is the upper bound of  $X_{k,i}^{rs}$ ; else,  $\omega_i < 0$ ) which affects the changing speed of the probability from 0 to 1;  $\Phi(\cdot)$  is the cumulative probability function of standard normal distribution. On account of the location parameter  $\gamma_i$ , even if the constrained characteristic attribute value is equal to the threshold, the considered probability may not be 0.5, depending on individual preference. The impacts of these parameters (e.g. the scale parameter  $\omega_i$  and location parameter  $\gamma_i$ ) on the function are shown in Fig 2.

Usually, in the metro system, the spatiotemporal constraints, referring to travel time and number of transfers, are taken as the constrained characteristic attributes. The fact that the passengers generally prefer to the route with smaller values of the two attributes leads to the result that both of the two constraints only have the upper bounds. In the route choice context for metro passengers, the thresholds to a specific attribute vary with OD pairs. The deterministic parts of the thresholds of constrained travel time and number of transfers are shown below respectively.

$$b_{n,t}^{rs} = \alpha_n \cdot \ln(T_{\min}^{rs} + 1) + T_{\min}^{rs} \quad (7)$$

$$b_{n,\tau}^{rs} = \beta_n + M_{\min}^{rs} \quad (8)$$

where  $b_{n,t}^{rs}$  is the bound of travel time (including in-vehicle time and transfer time) of



**Fig 2. The impacts of scale and location parameters.** The impacts of scale and location parameters on considered probability function are displayed separately by changing the values.

<https://doi.org/10.1371/journal.pone.0178789.g002>

individual  $n$  for OD pair  $rs$ ,  $h$ ;  $b_{n,\tau}^{rs}$  is the bound of number of transfers;  $T_{min}^{rs}$  is the shortest travel time for OD pair  $rs$ ,  $h$ ;  $M_{min}^{rs}$  is the minimum number of transfers;  $\alpha_n$  and  $\beta_n$  are the bound parameters needed to be calibrated.

Assuming that error component  $\epsilon_{n,k}^{rs}$  follows the multivariate normal distribution, that is  $\epsilon_{n,k}^{rs} \sim MVN(0, \Sigma_n^{rs})$ , where  $\Sigma_n^{rs}$  is the covariance matrix associated with the correlation among alternative routes, together with the constraints on route availability in the utility function, it is called CMNP (constrained multinomial probit) model. With respect to the route choice scenario, the route over-lapping problem in the railway network has been identified by Yai et al. [6] which is similar to metro network. This paper rewrites the covariance matrix into three parts, where the first part depends on the correlation among routes, the second one denotes the transfer variance of the route, and the last one denotes the unobserved variance. The latter two parts distribute independently by route.

$$\Sigma_n^{rs} = \sigma_L^2 \begin{pmatrix} \delta_{11}^{rs} & \delta_{12}^{rs} & \cdots & \delta_{1m}^{rs} \\ \delta_{21}^{rs} & \delta_{22}^{rs} & \cdots & \delta_{2m}^{rs} \\ \cdots & \cdots & \cdots & \cdots \\ \delta_{m1}^{rs} & \delta_{m2}^{rs} & \cdots & \delta_{mm}^{rs} \end{pmatrix} + \sigma_N^2 \begin{pmatrix} X_{12}^{rs} & 0 & \cdots & 0 \\ 0 & X_{22}^{rs} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & X_{m2}^{rs} \end{pmatrix} + \sigma_0^2 I \quad (9)$$

$$\delta_{kj}^{rs} = \sum_{i \in \Gamma_k} \kappa_{ikj} l_i \quad (10)$$

where  $m$  is the number of alternatives in the routes set;  $\sigma_L^2$  is the unit variance which is independent from each other;  $\sigma_N^2$  is the variance of transfer 1 time;  $\sigma_0^2$  is constant and identical to all routes;  $\delta_{kj}^{rs}$  is the over-lapping length between route  $k$  and  $j$  for OD pair  $rs$ ;  $X_{m2}^{rs}$  is the number of transfers of route  $m$ ;  $I$  is the identity matrix;  $l_i$  is the length of link  $i$ ;  $\Gamma_k$  is the links set of route  $k$ ; if link  $i$  is shared by route  $k$  and  $j$ ,  $\kappa_{ikj} = 1$ , otherwise,  $\kappa_{ikj} = 0$ . There are only three parameters in this covariance matrix, but we just need to estimate the ratio  $\lambda_1$  of  $\sigma_L^2$  to  $\sigma_0^2$  and the ratio  $\lambda_2$  of  $\sigma_N^2$  to  $\sigma_0^2$ .

Then based on the random utility maximization, given the values of all parameters, the chosen probability of route  $k$  is equal to

$$P_k^{rs} = P[V_{n,k}^{rs} + C_{n,k}^{rs} + \epsilon_{n,k}^{rs} \geq \max(V_{n,j}^{rs} + C_{n,j}^{rs} + \epsilon_{n,j}^{rs}; j \neq k, j \in \mathbf{A}_n^{rs})] \quad (11)$$

where  $\mathbf{A}_n^{rs}$  is the routes set between OD pair  $rs$  for passenger  $n$ . With respect to current scale of the metro network, the largest size of the routes set  $\mathbf{A}_n^{rs}$  can be set as 10.

## Model estimation

### The Bayesian formulation to the CMNP model

Faced with multidimensional integrals of the multivariate normal densities especially for large routes set, MNP model is usually estimated by Bayesian formulation and Monte Carlo Markov Chain (MCMC) approach [36]. This paper mainly wants to exhibit how to transform CMNP model with structured covariance into Bayesian formula and introduce Cholesky Decomposition to descend the dimension of integral so that the computational time can be saved. After the dimension reduction process and integral domain transformation, the calculation of multi-dimensional integrals of the multivariate normal densities given the parameters can be carried out based on quasi-Monte Carlo method.

Here, we construct a vector  $\zeta = \mu \cup \sigma \cup \Sigma = \theta \cup \omega \cup \gamma \cup z \cup \lambda$  including all unknown parameters, where  $\theta$  contains the parameters in the compensatory component;  $\omega$  is the vector with the

scale parameters;  $\gamma$  is the vector with the location parameters;  $\mathbf{z}$  contains the parameters in the threshold function, that is  $\alpha_n$  and  $\beta_n$  in this paper;  $\lambda$  covers the parameters in the covariance matrix. Meanwhile, the vector  $\mathbf{Y}$  denotes the indicators of the observations referring to the chosen routes. Compared with the traditional probit model and the proposed CMNP model, the non-compensatory component in the CMNP model leads to the difference between the two models. However, the value of the non-compensatory component can be easily calculated given the unknown parameters, benefiting from the independent bivariate normal distribution assumption. Based on Bayes' theorem, the posterior distribution  $\pi(\zeta|\mathbf{Y})$  is proportional to the priors on all unknown parameters, that is, the joint posterior distribution for the Hierarchical Bayes model is as follows.

$$\pi(\zeta|\mathbf{Y}) \propto P(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{\gamma}, \mathbf{z}, \lambda) \cdot \pi(\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{\gamma}, \mathbf{z}, \lambda|\zeta) \cdot \pi(\zeta) \quad (12)$$

where  $\pi(\cdot)$  is the probability density function;  $P(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{\gamma}, \mathbf{z}, \lambda)$  is the probability of observation  $\mathbf{Y}$  given all unknown parameters which is equal to Eq (11). Supposing that all parameters are independent from each other, we can get the below equations.

$$\pi(\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{\gamma}, \mathbf{z}, \lambda|\zeta) = \pi(\boldsymbol{\theta}|\boldsymbol{\mu}_\theta, \boldsymbol{\sigma}_\theta) \cdot \pi(\boldsymbol{\omega}|\boldsymbol{\mu}_\omega, \boldsymbol{\sigma}_\omega) \cdot \pi(\boldsymbol{\gamma}|\boldsymbol{\mu}_\gamma, \boldsymbol{\sigma}_\gamma) \cdot \pi(\mathbf{z}|\boldsymbol{\mu}_z, \boldsymbol{\sigma}_z) \cdot \pi(\lambda|\mu_\lambda, \sigma_\lambda) \quad (13)$$

$$\pi(\zeta) = \pi(\boldsymbol{\mu}_\theta) \cdot \pi(\boldsymbol{\sigma}_\theta) \cdot \pi(\boldsymbol{\mu}_\omega) \cdot \pi(\boldsymbol{\sigma}_\omega) \cdot \pi(\boldsymbol{\mu}_\gamma) \cdot \pi(\boldsymbol{\sigma}_\gamma) \cdot \pi(\boldsymbol{\mu}_z) \cdot \pi(\boldsymbol{\sigma}_z) \cdot \pi(\mu_\lambda) \cdot \pi(\sigma_\lambda) \quad (14)$$

Probit model is hard to be calculated even if all the unknown parameters are given because of the multivariate normal distribution. The Eq (11) can be rewritten into the D-dimensional integrals of the multivariate normal density as follows.

$$P_k^{rs} = \int_{\epsilon_1^{rs}=-\infty}^{\epsilon_1^{rs}+V_k^{rs}+C_k^{rs}-V_1^{rs}-C_1^{rs}} \cdots \int_{\epsilon_k^{rs}=-\infty}^{+\infty} \cdots \int_{\epsilon_m^{rs}=-\infty}^{\epsilon_k^{rs}+V_k^{rs}+C_k^{rs}-V_m^{rs}-C_m^{rs}} \Omega(\boldsymbol{\epsilon}^{rs}) d\epsilon_m^{rs} \cdots d\epsilon_1^{rs} \quad (15)$$

$$\Omega(\boldsymbol{\epsilon}^{rs}) = (2\pi)^{-\frac{m}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \boldsymbol{\epsilon}^{rs} \Sigma^{-1} (\boldsymbol{\epsilon}^{rs})^T\right] \quad (16)$$

where  $\boldsymbol{\epsilon}^{rs} = [\epsilon_1^{rs}, \epsilon_2^{rs}, \dots, \epsilon_m^{rs}]$  is the random error vector.

Considering that the covariance matrix is a Hermitian, positive-definite matrix, the integral of the general multivariate normal distribution can be transformed into that of standard normal distribution via Cholesky Decomposition to the covariance matrix and other substitutions [37–38]. By this means, the integral domain is referred to as transforming the  $m$ -variate integral into one over the  $(m-1)$ -dimensional hypercube.

Based on Cholesky Decomposition, the covariance matrix can be written as

$$\Sigma = D \times D^T \quad (17)$$

where  $D$  is a lower triangular matrix and  $D^T$  is its conjugate transpose. We set

$$\boldsymbol{\epsilon}^{rs} = D \times \mathbf{q}^{rs} \quad (18)$$

where  $\mathbf{q}^{rs} = [q_1^{rs}, q_2^{rs}, \dots, q_m^{rs}]$  is a vector substituting random error vector  $\boldsymbol{\epsilon}^{rs}$ . Then the Eq (11)

can be transformed as

$$P_k^{rs} = (2\pi)^{-\frac{m}{2}} \int_{-\infty}^{b'_1} \exp(-\frac{q_1^2}{2}) \cdots \int_{-\infty}^{b'_m} \exp(-\frac{q_m^2}{2}) d\mathbf{q}^{rs} \quad (19)$$

$$b'_i = \begin{cases} (\sum_{j=1}^k d_{kj} q_j + V_k^{rs} + C_k^{rs} - V_i^{rs} - C_i^{rs} - \sum_{j=1}^{i-1} d_{ij} q_j) / d_{i,i}, & i \neq k \\ +\infty, & i = k \end{cases} \quad (20)$$

where  $b'_i$  is the upper limit of the  $i$ -th layer's integration;  $d_{i,i}$  is the element in  $D$ . We assume

$$q_i = \Phi^{-1}(v_i) \quad (21)$$

where  $\mathbf{v}^{rs} = [v_1^{rs}, v_2^{rs}, \dots, v_m^{rs}]$  is a vector substituting vector  $\mathbf{q}^{rs}$ ;  $\Phi^{-1}(\cdot)$  is the inverse cumulative probability function of standard normal distribution, that is  $v_i = \Phi(q_i)$ . Meanwhile, we suppose  $v_i = w_i e_i$ . Thus, we get the equation

$$\begin{aligned} P_k^{rs} &= \int_0^{e_1} \cdots \int_0^{e_k} \cdots \int_0^{e_m} d\mathbf{v}^{rs} \\ &= \int_0^1 \int_0^1 \cdots \int_0^1 \prod_{i=1}^m e_i d\mathbf{w} \end{aligned} \quad (22)$$

$$e_i = \begin{cases} \Phi((\sum_{j=1}^k d_{kj} \Phi^{-1}(w_j \cdot e_j) + V_k^{rs} + C_k^{rs} - V_i^{rs} - C_i^{rs} - \sum_{j=1}^{i-1} d_{ij} \Phi^{-1}(w_j \cdot e_j)) / d_{i,i}), & i \neq k \\ 1, & i = k \end{cases} \quad (23)$$

where  $\mathbf{w} = (w_1, w_2, \dots, w_m)$  denotes the parameters vector.

When  $i = k$ ,  $e_i = 1$ , the  $m$ -dimensional integration declines into  $(m-1)$ -dimensional integration. And quasi-Monte Carlo method can be used to calculate the probability of multivariate normal distribution given all unknown parameters.

## The parameter identification problem

With respect to the constrained multinomial logit model, Castro et al. discussed the parameter identification problem derived from the fact that one attribute exists both in the compensatory and non-compensatory components [23]. This problem has been avoided in this paper via the process that the threshold is regarded as a function whose value varies with the change of the OD scale. By this means, the parameters associated with the same attribute both in the compensatory and non-compensatory components are identifiable. Another parameter identification problem arises when the threshold parameter plus directly location parameter. In this paper, the constraint of number of transfers suffers this problem as  $\beta + \gamma_r$ , where  $\beta$  is the threshold parameter and  $\gamma_r$  is the location parameter. The two parameters cannot be identified, but we can estimate the sum of them without influence on other parameters. For simplicity, we can assume the location parameter is equal to 0, and then we can get the value of the threshold parameter.

## Probability calculation based on Monte Carlo simulation

With respect to the solution to Eq (11) associated with the probit probability, the quasi-Monte Carlo method is carried out. Supposing that every element in  $\mathbf{w}$  follows the uniform distribution  $w_i \sim U(0, 1)$  and the elements in  $\mathbf{w}$  are independent from each other, we use Halton sequence to generate random data. The vector  $\mathbf{w}_j = (w_{1j}, \dots, w_{ij}, \dots, w_{mj})$ , contains the random values generated by the Halton sequence in the  $j$ -th iteration. We can get the approximate solution to Eq (11) given unknown parameters  $\theta, \omega, \gamma, \mathbf{z}, \lambda$ , that is,

$$\begin{aligned} P_k^{rs} &= E\left(\prod_{i=1}^m e_i(\mathbf{w})\right) \\ &\approx \frac{1}{J} \sum_{j=1}^J \left(\prod_{i=1}^m e_i(\mathbf{w}_j)\right) \end{aligned} \quad (24)$$

where  $E(\cdot)$  denotes the expected value.

## Estimation algorithm

The Metropolis-Hastings (M-H) algorithm [39–40] which is known as one of the MCMC approach is widely used to generate samples of the parameter from a prior distribution without the prior knowledge. In order to generate the final samples, candidates are drawn iteratively and they will be accepted as current samples with a certain probability in every iteration. With the increase of the number of iterations, Markov Chain ensures that we will gain a stationary posterior distribution of the parameter. In order to improve estimating efficiency, the variable-at-a-time Metropolis sampling scheme [39] is used to generate candidate for every parameter in turn in the parameters' set. The estimation process is organized as follows.

**Step 1:** Generate randomly initial values from pre-defined prior distribution, that is  $\zeta^{(0)} = (\zeta_1^{(0)}, \dots, \zeta_i^{(0)}, \dots, \zeta_{NP}^{(0)})$ , where  $NP$  denotes the number of parameters. And generate routes sets  $\mathbf{A}^{rs}$  for OD pairs based on physical length. Set iteration  $t = 1$  and  $i = 1$ .

**Step 2:** Draw a candidate  $\zeta^* = (\zeta_1^{(t)}, \dots, \zeta_{i-1}^{(t)}, \zeta_i^*, \zeta_{i+1}^{(t-1)}, \dots, \zeta_{NP}^{(t-1)})$  from a jumping distribution  $J_t(\zeta_i^* | \zeta_i^{(t-1)})$  based on the Gaussian random walk Metropolis sampling method. This method suggests that the jumping distribution is supposed to be a normal distribution which is a symmetric distribution satisfying the equation  $J_t(\zeta_i^* | \zeta_i^{(t-1)}) = J_t(\zeta_i^{(t-1)} | \zeta_i^*)$ . Here we set the jumping distribution as  $J_m(\zeta_i^* | \zeta_i^{(m-1)}) \sim N(\zeta_i^{(m-1)}, \xi^2)$ , where  $\xi^2$  is the proposal variance for the  $i$ -th parameter.

**Step 3:** Calculate the acceptance ratio  $\mathcal{G} = \min\left\{1, \frac{\pi'(\zeta_i^*) J_m(\zeta_i^{(m-1)} | \zeta_i^*)}{\pi'(\zeta_i^{(m-1)}) J_m(\zeta_i^* | \zeta_i^{(m-1)})}\right\} = \min\left\{1, \frac{\pi'(\zeta_i^*)}{\pi'(\zeta_i^{(m-1)})}\right\}$ ,

where

$$\begin{aligned} \pi'(\zeta_i^*) &= \pi(\zeta_1^{(t)}, \dots, \zeta_{i-1}^{(t)}, \zeta_i^*, \zeta_{i+1}^{(t-1)}, \dots, \zeta_{NP}^{(t-1)} | \mathbf{Y}) \\ &\propto \prod_{k \in \mathbf{Y}} [P(k | \zeta_1^{(t)}, \dots, \zeta_{i-1}^{(t)}, \zeta_i^*, \zeta_{i+1}^{(t-1)}, \dots, \zeta_{NP}^{(t-1)})] \cdot \prod_{j=1}^{i-1} \pi(\zeta_j^{(t)}) \cdot \prod_{j=i+1}^{NP} \pi(\zeta_j^{(t-1)}) \cdot \pi(\zeta_i^*) \end{aligned} \quad (25)$$

$$\begin{aligned} \pi(\zeta_i^{(t-1)}) &= \pi(\zeta_1^{(t)}, \dots, \zeta_{i-1}^{(t)}, \zeta_i^{(t-1)}, \zeta_{i+1}^{(t-1)}, \dots, \zeta_{NP}^{(t-1)} | \mathbf{Y}) \\ &\propto \prod_{k \in \mathbf{Y}} [P(k | \zeta_1^{(t)}, \dots, \zeta_{i-1}^{(t)}, \zeta_i^{(t-1)}, \zeta_{i+1}^{(t-1)}, \dots, \zeta_{NP}^{(t-1)})] \cdot \prod_{j=1}^{i-1} \pi(\zeta_j^{(t)}) \cdot \prod_{j=i+1}^{NP} \pi(\zeta_j^{(t-1)}) \cdot \pi(\zeta_i^{(t-1)}) \end{aligned} \quad (26)$$

where  $P(k | \zeta_1^{(t)}, \dots, \zeta_{NP}^{(t-1)}) = P_k^{rs}$ , referring to Eq (11). The same parts can be canceled out and

the ratio is

$$\vartheta = \min\left\{1, \frac{\prod_{k \in Y} [P(k|\zeta_1^{(t)}, \dots, \zeta_{i-1}^{(t)}, \zeta_i^*, \zeta_{i+1}^{(t-1)}, \dots, \zeta_{NP}^{(t-1)}) \cdot \pi(\zeta_i^*)]}{\prod_{k \in Y} [P(k|\zeta_1^{(t)}, \dots, \zeta_{i-1}^{(t)}, \zeta_i^{(t-1)}, \zeta_{i+1}^{(t-1)}, \dots, \zeta_{NP}^{(t-1)}) \cdot \pi(\zeta_i^{(m-1)})]}\right\} \quad (27)$$

*Step 4:* Draw a value  $u$  from the uniform distribution  $U(0, 1)$ . If  $u \leq \vartheta$ ,  $\zeta^{(m)} = \zeta^*$ ; otherwise,  $\zeta^{(m)} = \zeta^{(m-1)}$ .

*Step 5:* If  $i < NP$ ,  $i = i + 1$ , repeat *Step 2—Step 4*; otherwise, continue *Step 6*.

*Step 6:* If  $m < M$ ,  $m = m + 1$ ,  $i = 1$ , repeat *Step 2—Step 5*; otherwise, stop sampling.

## Results and discussions

### Data

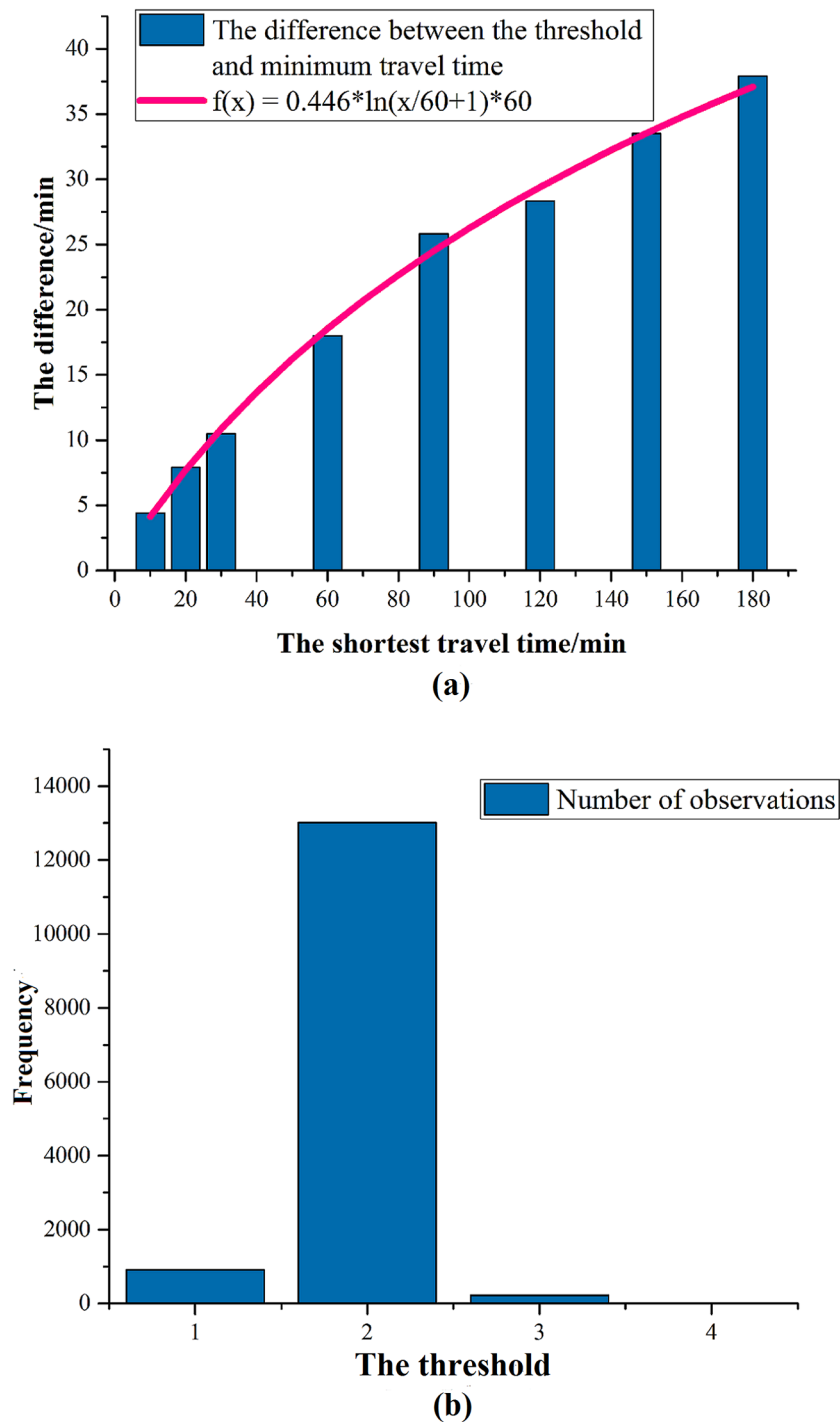
All data are available in the supporting information file ([S1\\_Dataset](#)).

With new lines put into operation ceaselessly, Guangzhou Metro becomes the sixth busiest metro system in the world and the third largest metro system in China. Up to July 2014, there are 8 lines and 136 stations (including 19 transfer stations) in operation, forming 256.6km operating length and carrying about 6.2 million daily ridership, except the APM Line. The APM Line isn't in our consideration, because it belongs to a unique system which needs to swipe through again although you were in other lines. The metro system covers the major urban areas of the city, reaches into some large suburban area and connects Guangzhou city and Foshan city. Through statistical analysis, there are 3721 OD pairs for which the routes with the shortest travel times are not the routes with the minimum number of transfers. It provides the possibility that more than one route will be considered by passengers with comprehensive consideration of multiple factors. The large scaled metro network increases the complexity of route choice analysis.

In July 2014, Guangzhou Metro Corporation organized a survey in the metro stations to collect passengers' travel characteristics, such as respondents' actual travel routes. Totally, the effective sample size is 14142. Based on the survey data, [Fig 3\(a\)](#) shows the relationship between the difference (namely the threshold minus the shortest travel time) and the shortest travel time. It can be seen that the difference increases logarithmically with the increase of the shortest travel time which demonstrates that the travel time threshold formula in [Eq \(7\)](#) is suitable. By data fitting, when  $\alpha_n = 0.446$ , the mean absolute percentage error (MAPE) is the minimum 3.657%. [Fig 3\(b\)](#) shows that when the minimum number of transfers is 0, the weighted mean value of the threshold of number of transfers is 1.951, that is,  $\beta_n$  can be assumed as 1.951.

### Estimations

In the compensatory component, the in-vehicle travel time ( $X_{k,1}^{rs}$ , h), number of transfers ( $X_{k,2}^{rs}$ , time), transfer time ( $X_{k,3}^{rs}$ , h), comfort degree ( $X_{k,4}^{rs}$ , 0–1 variable) and revised angular cost ( $X_{k,5}^{rs}$ , km) are considered with the corresponding parameters  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ ,  $\theta_4$  and  $\theta_5$ , where revised angular cost  $X_{k,5}^{rs}$  measures the deviation degree of a route by transforming  $\sin()$  into  $\tan()$  and comfort degree  $X_{k,4}^{rs}$  represents the congestion level in the train whose value is  $X_{k,4}^{rs} = 1$  when average load factor of one route is smaller than 20%, otherwise,  $X_{k,4}^{rs} = 0$ . In the non-compensatory component, the travel time ( $X_{k,0}^{rs} = X_{k,1}^{rs} + X_{k,3}^{rs}$ , h) and number of transfers  $X_{k,2}^{rs}$  are considered with the threshold parameters  $\alpha$  and  $\beta$ , scale parameters  $\omega_t$  and  $\omega_m$  as well as location parameters  $\gamma_t$  and  $\gamma_m$  respectively. In case of the parameter identification problem in the



**Fig 3. The thresholds of travel time and number of transfers.** The figure displays the survey data associated with thresholds. By fitting, the parameters in the thresholds can be estimated.

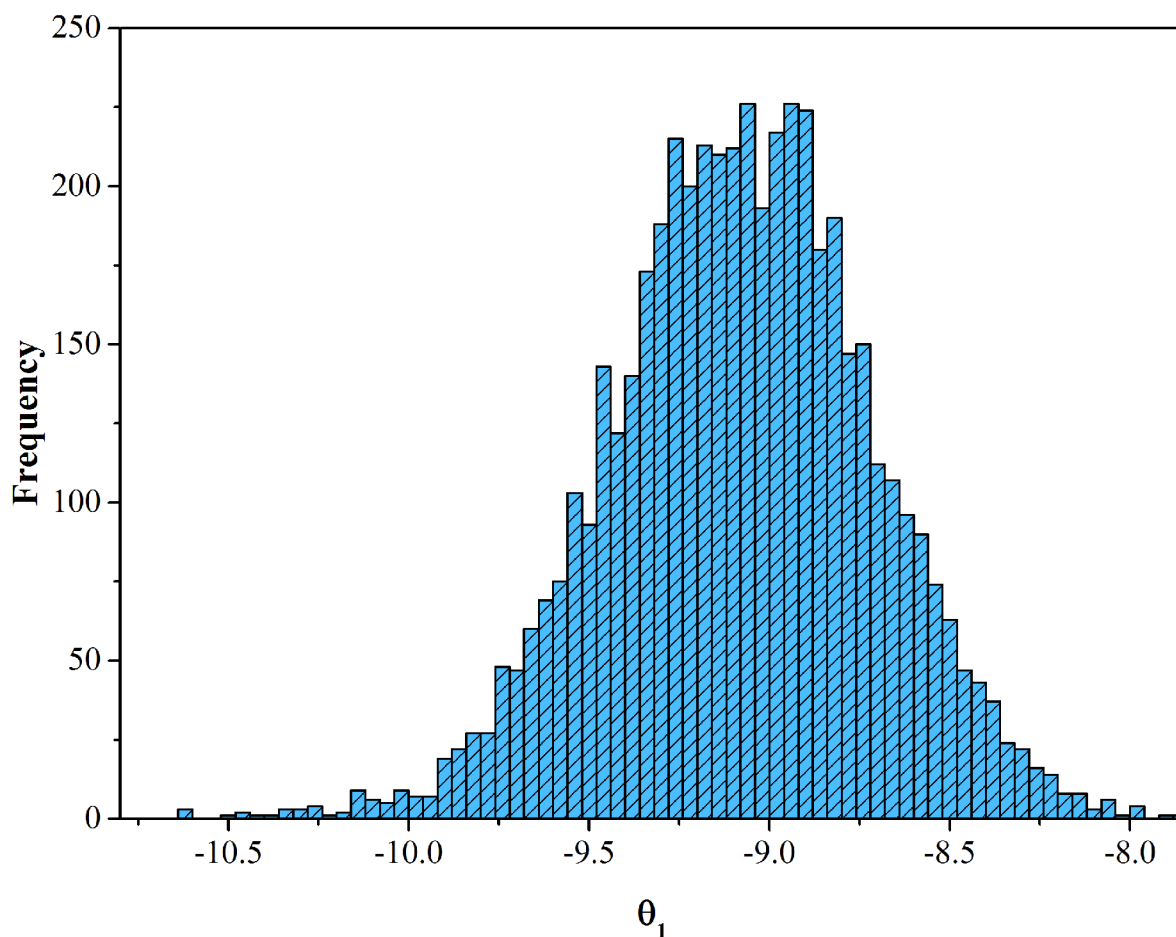
<https://doi.org/10.1371/journal.pone.0178789.g003>

threshold of number of transfers, the location parameter is assumed to be 0, that is  $\gamma_m = 0$ . Moreover, the parameter  $\lambda_1$  and  $\lambda_2$  in the covariance matrix needs to be estimated.

Based on the surveyed data, the proposed model, MNP model, MNL model and CMNL model are estimated respectively. The latter two models are estimated based on the

maximization likelihood estimation method, while the proposed model and MNP model are estimated based on the estimation approach proposed by this paper. Under the non-informative condition, the prior distributions for all parameters are assumed to follow uniform distribution. Totally, what parameters we need to estimate are  $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \alpha, \beta, \omega_b, \omega_m, \gamma_t, \lambda_1$  and  $\lambda_2$ . Considering the signs of the parameters, we assume that  $\theta_1, \theta_2, \theta_3$  and  $\theta_5$  follow  $U(-20, 0)$ ;  $\theta_4, \alpha, \beta, \omega_m, \lambda_1$  and  $\lambda_2$  follow  $U(0, 20)$ ;  $\omega_t$  follows  $U(60, 200)$ ;  $\gamma_t$  follows  $U(-1, 1)$ . The proposal variance  $\xi^2$  is 0.05. The surveyed RP data is divided into two parts, referring to the 12039 data for estimation and 2103 data for examination. The scheme to screen the data will be described later. Based on MCMC approach, we tried 10000 iterations to estimate all parameters in the CMNP model, where the fore 5000 samples for each parameter are abandoned as burn-in period and the left 5000 effective samples for each parameter are drawn. The distribution of the effective samples for  $\theta_1$  is taken as an example shown in Fig 4. Based on the samples, we can get the mean and 95% Bayesian conference interval (CI) shown in Table 1, as well as the estimations of MNL, MNP and CMNL models.

Kolmogorov-Smirnov (KS) Test is used to determine whether the samples follow normal distribution. Fig 4 shows the distribution of the effective samples for parameter  $\theta_1$  for an example. Descriptive statistics show that the average value is -9.074 and the standard deviation is 0.366. By KS test, the  $p$  value is 0.433 which is greater than 0.05, proving that the samples



**Fig 4. The distribution of effective samples for  $\theta_1$ .** The figure exhibits the frequency distribution of the samples to represent the convergence directly.

<https://doi.org/10.1371/journal.pone.0178789.g004>

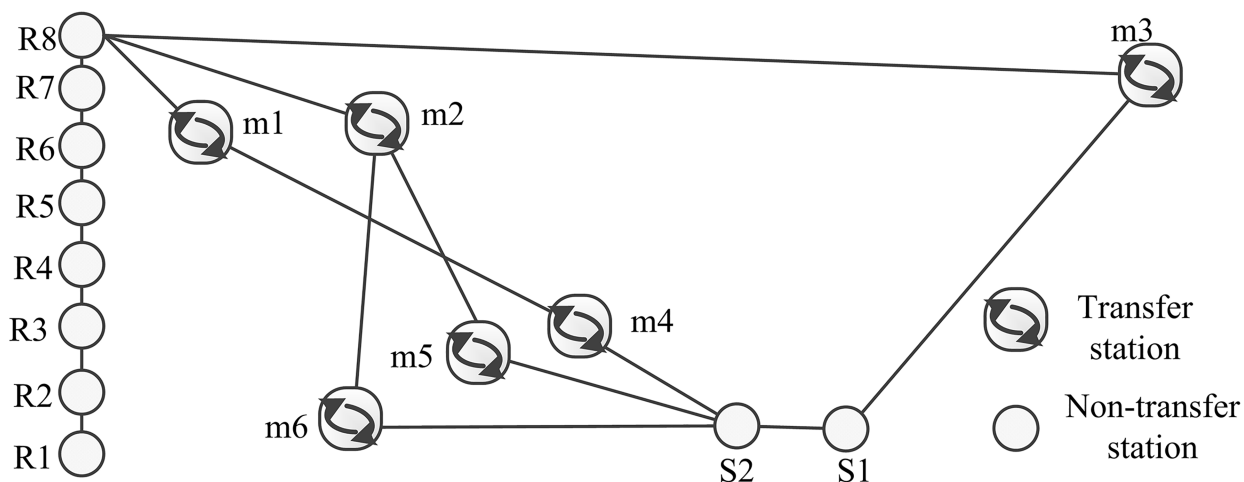
**Table 1. Estimations of MNL, CMNL and CMNP models.**

Parameter	MNL-value (t-value)	MNP-mean (95% CI)	CMNL-value (t-value)	CMNP-mean (95% CI)
$\theta_1$	-14.411 (-31.103)	-13.174 ([[-13.165, -13.184]])	-10.224 (-40.714)	-9.074 ([[-9.065, -9.084]])
$\theta_2$	-2.256 (-17.267)	-2.071 ([[-2.061, -2.081]])	-1.641 (-15.324)	-1.479 ([[-1.468, -1.489]])
$\theta_3$	-15.623 (-16.654)	-13.111 ([[-13.100, -13.121]])	-11.712 (-13.125)	-11.100 ([[-11.088, -11.111]])
$\theta_4$	0.005 (6.068)	0.008 ([0.00799, 0.00801])	0.004 (3.224)	0.008 ([0.00799, 0.00801])
$\theta_5$	-0.018 (-2.016)	-0.052 ([[-0.0520, -0.0523]])	-0.061 (-2.514)	-0.075 ([[-0.0751, -0.0753]])
$\alpha$	--	--	0.460 (3.412)	0.455 ([0.450, 0.460])
$\beta$	--	--	1.961 (2.121)	1.959([1.943, 1.975])
$\omega_t$	--	--	65.012 (21.403)	177.392 ([177.240, 177.548])
$\omega_m$	--	--	6.051 (5.711)	7.431 ([7.411, 7.452])
$\gamma_t$	--	--	0.001 (1.991)	0.001 ([0.0009, 0.0011])
$\lambda_1$	--	0.621([0.620, 0.622])	--	0.675([0.674, 0.676])
$\lambda_2$	--	0.544([0.542, 0.546])	--	0.552([0.550, 0.554])
$\rho^2$	0.442	0.571	0.533	0.723
Sample Size	10000	10000	10000	10000

<https://doi.org/10.1371/journal.pone.0178789.t001>

follow normal distribution at 5% significance level. Thus, the sampling process converges. Other parameters have the same characteristics. As shown in Table 1, the different results for the parameters in MNP and CMNP models indicate that the MCMC approach can successfully distinguished all parameters though some of them have the same initial values. And the CI for every parameter ensures us to accept the means of the drawn samples. Meanwhile, coefficients of in-vehicle time, number of transfers, transfer time and revised angular cost are negative, meaning that the chosen probability of one route decreases along with the increase of in-vehicle time, number of transfers, transfer time or revised angular cost. Coefficients of comfort degree in both models are all positive, meaning that their increase will improve individual preference to the route. It is consistent with the common sense. Furthermore,  $t$ -values of the coefficients for MNL and CMNL models exceed 1.96, indicating that the null hypothesis that the true values of the coefficients are zero can be rejected at the 0.05 significance level. And the  $\rho^2$  of all models are greater than 0.2, indicating that all models have a good goodness-of-fit. Compared with the  $\rho^2$ , CMNP model are the greatest, illustrating that the proposed CMNP model is the best among all models.

In addition to the estimation performance, the forecasting performances of all models are compared. In order to gain plenty of actual choice results which are drawn from the surveyed data, the route choices between some similar OD pairs are aggregated. As shown in Fig 5, we combine the origins as an identical origin **R** as well as the identical destination **S**, that is, **R** contains Guangzhou South Railway Station (R1), Shibi (R2), Huijiang (R3), Nanpu (R4), Luoxi (R5), Nanzhou (R6), Dongxiao South (R7) and Jiangtai Road (R8); **S** contains Jingxi Nanfang Hospital (S1) and Meihuayuan (S2). The transfer stations are Haizhu Square ( $m_1$ ), Gongyuan-qian ( $m_2$ ), Jiahewanggang ( $m_3$ ), Yantang ( $m_4$ ), Guangzhou East Railway Station ( $m_5$ ) and



**Fig 5. A diagram of deleted OD pairs and routes.** The origin stations are denoted as R; the destination stations are denoted as S; the transfer stations are denoted as m. Other stations are omitted in the figure. It means transferring if the route passes by a transfer station.

<https://doi.org/10.1371/journal.pone.0178789.g005>

Tiyu West Road ( $m_6$ ). Excluding the routes with chosen probabilities smaller than 0.0001, we have four routes left in Table 2 along with the chosen probabilities according to different models. The number of the actual choices between the specific OD pairs in the surveyed data is 2103 and the absolute error is calculated to compare the forecasting performance as shown in Table 2. We can see that CMNP model has the smallest MAE (Mean Absolute Error) which demonstrates that the proposed CMNP model has the best forecasting performance.

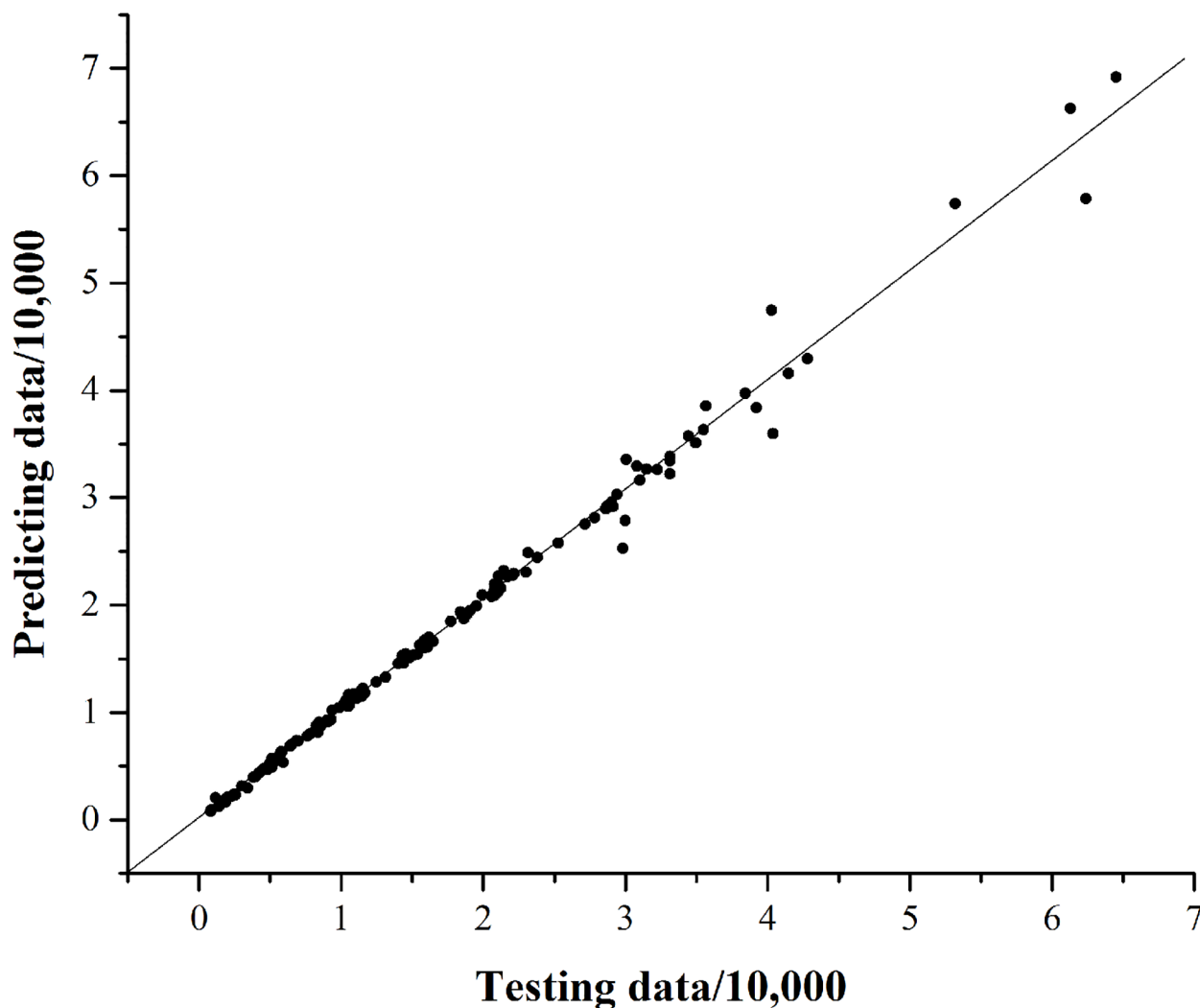
## Application

The route choice model can be used to predict the transfer flow volume, section flow volume, etc. which are the basis of scheduling the train plan, guiding individual travel route, etc. The proposed CMNP route choice model determines the route choice probability for every OD pair in the metro network. And then the flow volume on the route can be derived from the product of the probability and the OD volume. By counting the number of passengers transferring between two different lines based on the train timetable, the transfer flow volume can be calculated. All testing data are provided by Guangzhou Metro Corporation. The results are shown in Fig 6 where testing data is on the horizontal axis, predicting data is on the vertical axis, and the solid line is the basic line indicating that the predicting data is equal to the testing data if the data spot is on the line. Every spot represents the flow volume transferring from one running direction of one line to one running direction of another line. Usually, every line has

**Table 2. The comparisons of all models.**

Route	Actual Prob.	MNL Prob.(AE)	MNP Prob.(MAE)	CMNL Prob.(AE)	CMNP Prob.(AE)
K1: R -> m2 -> m6 -> S	51%	45% (6%)	46% (5%)	58% (7%)	53% (2%)
K2: R -> m2 -> m5 -> S	19%	38% (19%)	27% (8%)	23% (4%)	21% (2%)
K3: R -> m3 -> S	21%	13% (8%)	16% (5%)	12% (9%)	20% (1%)
K4: R -> m1 -> m4 -> S	9%	4% (5%)	11% (2%)	7% (2%)	6% (3%)
Sum	100%	100% (38%)	100% (20%)	100% (22%)	100% (8%)

<https://doi.org/10.1371/journal.pone.0178789.t002>



**Fig 6. The transfer flow volume forecasting performance.** For each spot, it has two values, including the testing data corresponding to horizontal axis and forecasting value corresponding to vertical axis.

<https://doi.org/10.1371/journal.pone.0178789.g006>

two running directions. The mean absolute percentage error (MAPE) is 4.91% which shows that the proposed CMNP model has a good application prospect.

## Conclusion

In a large scaled metro network, the complex nature of route choice process brings us a challenge to exactly figure out passengers' actual decision rules. This paper focuses on integrating the impacts of routes set and the interdependency among alternative routes on route choice probability into route choice modeling in the metro network. The impact of routes set on route choice probability expresses the semi-compensatory choice process which is a combination of routes set generation and route choice stages. Thereafter, a constrained multinomial probit (CMNP) model is proposed by this paper, in which, the utility function consists of compensatory, non-compensatory and error parts. The compensatory part is a linear function of in-vehicle travel time, number of transfers, transfer time, congestion level and revised angular cost. The non-compensatory part measures the impact of considered probability of one route

on the route's utility by a logarithm function, where considered probability is calculated by a binary probit equation denoting the relationship between the constrained attributes (e.g. travel time and number of transfers) and the corresponding thresholds proposed by this paper. The error part follows a multivariate normal distribution, whose variance is structured into three parts, including measuring the correlation among routes, representing the transfer variance of the route, and denoting the unobserved variance.

With respect to the estimation, considering multidimensional integrals of the multivariate normal probability density function, the CMNP model is rewritten as Bayesian formulation and MCMC approach is constructed to estimate all parameters. As a key point to calculate the acceptance rate, given the unknown parameters, the multidimensional integrals of the multivariate normal probability density function can be transformed into those of standard normal distribution via Cholesky Decomposition to the covariance matrix and other substitutions. Then the integrals can be easily simulated by quasi-Monte Carlo algorithm.

At last, the proposed model is estimated by the proposed estimation approach based on the surveyed RP data in Guangzhou Metro. The estimations show that every parameter can be distinguished though they have the same initial values. And the Bayesian CI indicates the reliability of the mean of the samples. Moreover, compared with MNL, MNP and CMNL models, the proposed CMNP model shows the best forecasting performance with respect to the prediction on the route choice probabilities and transfer flow volumes.

In the future, we will try to estimate the proposed model based on the smart card data and the travel time reliability will also be considered in the model.

## Supporting information

**S1 Table. Estimations of MNL, CMNL and CMNP models.**  
(DOCX)

**S2 Table. The comparisons of all models.**  
(DOCX)

**S1 Dataset. All data used in this paper.**  
(XLSX)

## Acknowledgments

Many thanks for helpful suggestions from reviewers and the editor.

## Author Contributions

**Conceptualization:** YZ EY.

**Funding acquisition:** EY.

**Methodology:** YZ EY HW KZ.

**Writing – original draft:** YZ EY HW KZ.

**Writing – review & editing:** YZ EY HW KZ.

## References

1. Raveau S, Muñoz J C, De Grange L. A topological route choice model for metro. *Transportation Research Part A: Policy and Practice*. 2011; 45(2): 138–147.

2. Zhang Y, Yao E, Dai H. Transfer volume forecasting method for the metro in networking conditions. *Journal of the China Railway Society*. 2013; 23(11): 1–6.
3. Simon H A. A behavioral model of rational choice. *The Quarterly Journal of Economics*. 1955; 69: 99–118.
4. Tversky A. Elimination by aspects: a theory of choice. *Psychological Review*. 1972; 79: 281–299.
5. Bovy P H L. On modelling route choice sets in transportation networks: a synthesis. *Transport Reviews*. 2009; 29: 43–68.
6. Yai T, Iwakura S, Morichi S. Multinomial Probit with structured covariance for route choice behavior. *Transportation Research Part B: Methodological*. 1997; 31(3): 195–207.
7. Daganzo C F, Sheffi Y. On stochastic models of traffic assignment. *Transportation Science*. 1977; 11(3): 253–274.
8. McFadden D. The revealed preferences of government bureaucracy. *Bell Journal of Economics Management Science*. 1968; 6: 401–416.
9. Dial R B. A probabilistic multipath traffic assignment model: which obviates path enumeration. *Transportation Research*. 1971; 5: 83–111.
10. Ramming M S. Network knowledge and route choice. Massachusetts: Massachusetts Institute of Technology. 2001: 111–212.
11. Liu S, Yao E, Zhang Y. Personalized route planning algorithm for urban rail transit passengers. *Journal of Transportation Systems Engineering and Information Technology*. 2014; 14(05): 100–104.
12. Cascetta E, Nuzzolo A, Russo F, Vitetta A. A modified Logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks, 13th International Symposium on Transportation and Traffic Theory. 1996; France.
13. Ben-Akiva M, Bierlaire M. Discrete choice methods and their applications to short term travel decisions. 1999; Springer, US.
14. Bekhor S, Ben-Akiva M E, Ramming M S. Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*. 2006; 14(1): 235–247.
15. Chu C. A paired combinatorial Logit model for travel demand analysis, 5th World Conference on Transport Research, 1989; Japan.
16. Koppelman F S, Wen C H. The paired combinatorial Logit model: properties, estimation and application. *Transportation Research Part B: Methodological*. 2000; 34(2): 75–89.
17. Vovsha P. The cross-nested Logit model: application to mode choice in the Tel-Aviv metropolitan area, 76th Transportation Research Board Annual Meeting. 1997; Washington D. C., USA.
18. Wen C H, Koppelman F S. The generalized nested Logit model. *Transportation Research Part B: Methodological*. 2001; 35(7): 627–641.
19. McFadden D, Train K. Mixed MNL models for discrete response. *Journal of Applied Econometrics*. 2000; 15(5): 447–470.
20. Raveau S, Guo Z, Muñoz J C, Wilson N H M. A behavioural comparison of route choice on metro networks: time, transfers, crowding, topology and socio-demographics. *Transportation Research Part A: Policy and Practice*. 2014; 66: 185–195.
21. Zhang Y, Yao E, Liu S. A constrained MNL route choice model for metro passengers, 94th Transportation Research Board Annual Meeting. 2015; Washington D. C., USA.
22. Martínez F, Aguila F, Hurtubia R. The constrained multinomial Logit: a semi-compensatory choice model. *Transportation Research Part B: Methodological*. 2009; 43(3): 365–377.
23. Castro M, Martínez F, Munizaga M A. Estimation of a constrained multinomial Logit model. *Transportation*. 2013; 40(3): 563–581.
24. Cantillo V, Ortuzar J D. A semi-compensatory discrete choice model with explicit attribute thresholds of perception. *Transportation Research Part B: Methodological*. 2005; 39(7): 641–657.
25. Kaplan S, Prato C G. Joint modeling of constrained path enumeration and path choice behaviour: a semi-compensatory approach, European Transport Conference. 2010; Glasgow.
26. Kaplan S, Prato C G. Closing the gap between behavior and models in route choice: the role of spatio-temporal constraints and latent traits in choice set formation. *Transportation Research Part F: Psychology and Behaviour*. 2012; 15(1): 9–24.
27. Manski C. The structure of random utility models. *Theory and Decision*. 1977; 8(3): 229–254.
28. Morikawa T. A hybrid probabilistic choice set model with compensatory and non-compensatory choice rules, 7th World Conference on Transport Research. 1995; Pergamon, Oxford.

29. Rob V N, Hoogendoorn-Lanser S, Koppelman F S. Using choice sets for estimation and prediction in route choice. *Transportmetrica*. 2008; 4(2), 83–96.
30. Kaplan S, Shiftan Y, Bekhor S. Development and estimation of a semi-compensatory model with a flexible error structure. *Transportation Research Part B: Methodological*. 2012; 46(2): 291–304.
31. Swait J D, Ben-Akiva M E. Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological*. 1987; 21(2): 91–102.
32. Bliemer M C J, Bovy P H L. Impact of route choice set on route choice probabilities, 87th Transportation Research Board Annual Meeting. 2008; Washington, D. C., USA.
33. Cascetta E, Papola A. Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. *Transportation Research Part C: Emerging Technologies*. 2001; 9(4): 249–263.
34. Swait J D. A non-compensatory choice model incorporating attribute cut-offs. *Transportation Research Part B: Methodological*. 2001; 35(7): 903–928.
35. Timothy J G, Greg M A. A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science*. 2004; 23(3): 391–406.
36. Nobile A. A Hybrid Markov Chain for the Bayesian Analysis of the Multinomial Probit Model. *Statistics and Computing*. 1995; 8: 229–242.
37. Genz A. Numerical computation of the multivariate normal probabilities. *Journal of Computational and Graphical Statistics*. 1992; 1: 141–150.
38. Genz A. Comparison of methods for the computation of multivariate normal probabilities. *Computing Science and Statistics*. 1993; 25: 400–405.
39. Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H, Teller E. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*. 1953; 21 (6): 1087–1092.
40. Hastings W K. Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*. 1970; 57: 97–109.