Journal of Transport Geography 72 (2018) 151–165

Contents lists available at ScienceDirect

Journal of Transport Geography

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

Modeling discretionary activity location choice using detour factors and sampling of alternatives for mixed logit models

Gabriel Leite Mariante[a], Tai-Yu Ma[b,*], Véronique Van Acker[b]

[a] Katholieke Universiteit Leuven, Oude Markt 13, Leuven 3000, Belgium
[b] Department of Urban Development and Mobility, Luxembourg Institute of Socio-Economic Research, 11 Porte des Sciences, Esch-sur-Alzette 4366, Luxembourg

ARTICLE INFO

Keywords:
Destination choice modeling
Mixed logit model
Constrained choice
Sampling of alternatives
Detour factor
Trip chaining

ABSTRACT

Excessively large choice sets have been identified as an important issue influencing the prediction accuracy of individuals' activity location models. In this work, a constrained choice modeling approach with sampling of alternatives is applied to analyze an individual's location choice for discretionary activities. The issue of large choice sets is tackled through two constraining methods: (i) adequacy of destination, in which only locations which are suited to a certain type of activity are selected and (ii) use of a delimitation rule depending on the type of trip chain to which the discretionary activity belongs. A mixed logit model with sampling of alternatives is specified to estimate individuals' location choice for different types of discretionary activities. The estimation results show sampling alternatives using an individual's constrained choice set based on both adequacy destination and detour provides significantly better prediction accuracy compared to that using only adequacy of destination. We conducted several experiments with respect to constraining methods, number of sampled alternatives and bias-correcting methods for sampling of alternatives in a mixed logit model. The results show that the Naïve method for sampling of alternatives in a mixed logit model provides better goodness-of-fit than estimation with correction terms.

1. Introduction

Modeling of individuals' location choices for daily activities has been widely studied in recent decades (Kitamura and Kermanshah, 1984; Thill and Horowitz, 1991; Timmermans, 1996; Arentze and Timmermans, 2005, 2007; Scott and He, 2012). Because someone's travel decisions depend on his or her activities, location choice modeling obtains an important role in activity-based travel demand forecasting. When choosing the location to undertake a given activity, individuals are facing a discrete choice problem. So far, a wide variety of discrete choice models have been applied to model the process through which individuals choose one of the available alternatives of activity locations (McFadden, 1978; Ben-Akiva and Bowman, 1998; Lee and Waddell, 2010). However, one major problem in this process is the definition of an activity location choice set. The set of possible activity locations can be very large. Therefore, past research has focused on how to generate a constrained activity location choice set.

The most common approach to deal with large choice sets is random sampling of the alternatives. A pre-specified smaller number of alternatives is drawn randomly from a larger universal choice set. This approach not only improves feasibility and practical estimation of discrete choice models, but also yields consistent parameter estimates (McFadden, 1978; Tseng and Mcconnell, 2000; Nerella and Bhat, 2004). Various studies have investigated how different sample sizes influence parameter estimates (Azaiez, 2010; Guevara and Ben-Akiva, 2013a, 2013b). Using simulated data, significant biases were found when comparing estimated coefficients and simulated target values with small sample sizes. Moreover, very small sample sizes are associated with poor model fit and predictive performance (Nerella and Bhat, 2004; Azaiez, 2010; Lemp and Kockelman, 2012; Guevara and Ben-Akiva, 2013b). Nerella and Bhat (2004) showed that one-eighth of the full choice set is the minimum number of alternatives and one-fourth is the recommended number in order to obtain small biases in the coefficient estimates. In theory, discrete choice models can be corrected for this type of biases by adding a correction term to the utility function. McFadden (1978) first addressed this issue when applying a logit model for modeling residential location choice. Guevara and Ben-Akiva (2013a, 2013b) extended McFadden's work for Multivariate Extreme Value (MEV) models, and further developed sampling of alternatives for mixed logit models. The authors propose theoretical support for deriving several approximations to correct model estimates when sampling alternatives from a universal choice set. They suggest

* Corresponding author.
E-mail address: tai-yu.ma@liser.lu (T.-Y. Ma).

https://doi.org/10.1016/j.jtrangeo.2018.09.003
Received 8 August 2017; Received in revised form 23 May 2018; Accepted 5 September 2018
0966-6923/

the Naïve method (i.e. using sampled alternatives for model estimation disregarding the issue of potential bias) is preferred when applying sampling of alternatives in mixed logit models.

In addition to random sampling, several options exist to reduce choice sets, which are especially relevant when it comes to reducing a location choice set. The first option is based on individual's socio-economic situation to generate 'considered choice set' for decision making (Meyer, 1979; Swait and Ben-Akiva, 1987; Swait, 2001). The rationale is based on individual rational behavior that only a set of feasible choices is considered by a decision maker (Simon, 1955). Different factors to identify feasible/infeasible alternatives and the process of search might influence feasible choice set generation. Meyer (1979) showed limited information and past experiences influence individual's destination choice set formation. Richardson (1982) suggested individual's choice set generation is the outcome of a search under various attributes (information, budget limits, individual or household attributes etc.) of the search process. The author proposed a search model to decide whether to stop his search by comparing the additional utility gain and the cost for extra search. Swait and Ben-Akiva (1987) developed a probabilistic choice set generation model to incorporate different attributes under different levels of behavior realism. As considering factors to a decision maker is generally unobservable, another option is based on time geography theory (Hägerstrand, 1970) and considers the individual's spatial and temporal constraints to reduce the location choice set (Thill and Horowitz, 1991; Thill, 1992; Kwan and Hong, 1998; Kitamura and Kermanshah, 1984). Using GIS tools, network-based potential path areas (PPA) are designed taking into account all the available activity locations and considering someone's travel time budget and realistic road network situations (Kwan and Hong, 1998; Miller, 2004; Scott and He, 2012). However, in order to delimit PPA, detailed transportation network data is needed which are not always easily available. A third option is based on the calculation of detour factors, defined as the ratio between any given route between two points and the optimal (shortest-path) route. It ranges from one (i.e., when the given route equals the optimal) to infinity (Witlox, 2007). In the context of location choices for discretionary activities, the detour factor is often calculated to measure the degree of deviation between the individual's fixed home-to-work trajectory and candidate locations where someone can undertake discretionary activities (Arentze and Timmermans, 2007; Justen et al., 2013). The idea is that individuals tend to participate in non-work related activities at locations around their way from home to work, and not far away from it. By selecting only those locations which have a detour factor lower than a specific upper limit, activity location choice sets can be reduced considerably while still obtaining high rates of inclusion of the chosen alternative (Justen et al., 2013; Ma et al., 2017b). Using a subset of alternatives, McFadden (1978) illustrated that statistically consistent parameter estimates can be obtained for a logit model by adding a term to the utility function that is based on the expansion of alternatives in his subset. Guevara and Ben-Akiva (2013a, 2013b) applied this approach to multivariate extreme value models and mixed logit models. As there is no standard way to generate constrained choice sets, different important sampling techniques have been developed and applied for choice behavior modeling (Ben-Akiva and Bowman, 1998; Frejinger et al., 2009). Moreover, it is also important to account for trip chaining behavior. Empirical studies have illustrated how neglecting trip chaining behavior in activity location prediction leads to an overestimation of travel distance (Arentze and Timmermans, 2007). When people make location choice decision for a discretionary activity located in a work-based tour, i.e. a trip chain of work- (one or several) discretionary activities- work, this anchor point of the workplace could be considered as a reference point to measure individual's travel time/cost to reach that activity location. Consequently, when discretionary activities are located on the tour-based activity chains with home or work activity as anchor points, using home or work place as a center point to generate location choice candidates would be relevant to delimit potential reachable areas under individuals' activity schedule constraints (Arentze and Timmermans, 2007; Primerano et al., 2008; Currie and Delbosc, 2011).

This paper will contribute to the literature on choice set reduction by combining random sampling with empirical detour factors while accounting for the influence of trip chaining behavior. We investigate how different sample sizes and constrained choice set generated by detour factors and trip chaining constraints affect model estimation in terms of model fit and predictive performance. For comparative purposes, results are compared with random sampling from unconstrained choice sets. The latter random sampling approach was used in past studies to model constrained destination choice for shopping (e.g. Scott and He, 2012 among others). To the best of our knowledge, there are still few studies considering the impact of sampling protocols and trip chaining constraints on choice model estimation. This work contributes to this gap by applying importance sampling concept and sampling of alternatives in a mixed logit model framework for discretionary activity location choice modeling. Different experiments with the addition of bias-correcting factors to the utility functions in a mixed logit choice model are tested to evaluate its effect on model fitness statistics. The performance of all our estimated models is computed using an 80%–20% validation scheme, in which randomly selected 80% of the data is used for training and the remaining 20% for testing. The reported results of correct prediction rates always refer to performance of the test group.

The remainder of this paper is organized as follows. Section 2 provides some methodological background. In this section, we recall the methodology for constrained choice modeling. It starts with a description of how a constrained choice set can be generated based on trip chain specific detour factors. Then, it presents the mixed logit model and describes the experiments with sampling of alternatives. Section 3 summarizes the data used in this study. We first analyze the observed trip chaining behavior and define different types of activity chains. Then, we describe the determinants that are used to model location choices for different types (categories) of discretionary activities (i.e. eating out, picking up or dropping off someone, services and others). Section 4 presents the results of a set of mixed logit models. Various experiments are conducted with respect to constraining methods, sample sizes, and expansion factors when using random sampling of alternatives. Conclusions are drawn in Section 5.

2. Constrained choice modeling with sampling of alternatives

We will apply a constrained choice modeling approach to model individuals' location choice for discretionary activities as individuals consider only a limited number of choice alternatives over all possible activity locations under his/her scheduling constraints and preferences (Ben-Akiva and Boccara, 1995; Swait and Ben-Akiva, 1987). The theoretical foundation is based on Manski (1977), where the probability of choosing an alternative j in a universal choice set C_n of individual n is formulated as the multiplication of probability of choosing an alternative in a constrained choice set as

$$P_n(j) = \sum_{D \in C_n} P_n(j| \; D \Big) P_n(D) \tag{1}$$

where $P_n(j|D)$ is the conditional probability of choosing alternative j on the subset D for individual n. $P_n(D)$ is the probability of subset D on individual n's choice set C_n.

The constrained choice modeling approach provides a two-stage framework by a constrained choice set generation in a first stage and then a choice model estimation in a second stage (Ben-Akiva and Boccara, 1995; Thill and Horowitz, 1997; Scott and He, 2012). We detail the two stages next.