Transportation Research Part A 46 (2012) 1097–1107

Contents lists available at SciVerse ScienceDirect

Transportation Research Part A

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

A behavioral housing search model: Two-stage hazard-based and multinomial logit approach to choice-set formation and location selection

Taha Hossein Rashidi [a,*], Joshua Auld [b,1], Abolfazl (Kouros) Mohammadian [b,2]

[a] Department of Civil Engineering, University of Toronto, Sandford Fleming Building, 10 King's College Road, Room 3001, Toronto, Ontario, Canada M5S 3G4
[b] Department of Civil and Materials Engineering, University of Illinois at Chicago, 842 W. Taylor St., Chicago, IL 60607, United States

ARTICLE INFO

Article history:
Received 22 December 2010
Received in revised form 27 January 2012
Accepted 31 January 2012

Keywords:
Housing search
Sampling correction
Choice set formation
Multinomial logit
Hazard models

ABSTRACT

Residential location search has become an important topic to both practitioners and researchers as more detailed and disaggregate land-use and transportation demand models are developed which require information on individual household location decisions. The housing search process starts with an alternative formation and screening stage. At this level households evaluate all potential alternatives based on their lifestyle, preferences, and utilities to form a manageable choice set with a limited number of plausible alternatives. Then the final residential location is selected among these alternatives. This two-stage decision making process can be used for both aggregate zone-level selection as well as searching disaggregate parcel or building-based housing markets for potential dwellings. In this paper a zonal level household housing search model is developed. Initially, a household specific choice set is drawn from the entire possible alternatives in the area based on the average household work distance to each alternative. Following the choice set formation step, a discrete choice model is utilized for modeling the final residential zone selection of the household. A hazard-based model is used for the choice set formation module while the final choice selection is modeled using a multinomial logit formulation with a deterministic sample correction factor. The approach presented in the paper provides a remedy for the large choice set problem typically faced in housing search models.

1. Introduction

Residential location search has been recently become an important research topic in many fields including transportation, urban planning, geography, economics, and other related disciplines. Metropolitan planning organizations, real estate companies, insurance companies and financial institutions are also among the non-academic organizations that are interested in having working housing search models. Since the early introduction of the discrete choice paradigm, such models of housing selection behavior have often been developed using this approach (McFadden, 1974). As has long been known, the predictive potential and accuracy of a discrete choice model itself are highly reliant on the choice set composition (Ben-Akiva and Lerman, 1985 ; Timmermans and Golledge, 1990). Therefore, in cases of spatial models, as in residential location choice, the handling of the choice set development becomes very important. Even though recent advances in computational power

* Corresponding author. Tel.: +1 647 638 5450.
E-mail addresses: taha.hosseinrashidi@utoronto.ca (T.H. Rashidi), auld@uic.edu (J. Auld), kouros@uic.edu (A. Mohammadian).
[1] Tel.: +1 312 996 0962; fax: +1 312 996 2426.
[2] Tel.: +1 312 996 9840; fax: +1 312 996 2426.

0965-8564/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tra.2012.01.007

allows researchers to work with large datasets, in practical applications, the difficulty of handling many alternatives makes it necessary to reduce the number of alternatives in the choice set into some manageable size. In the literature, there have been two extreme approaches for selecting the set of alternatives; first, randomly selecting a finite number of alternatives from the *universal choice set*, as it is defined by Ben-Akiva and Lerman (1985), second, considering all plausible alternatives (Salomon and Ben-Akiva, 1983 and Thill and Horowitz, 1991). It can be shown that both approaches can raise concerns. Although inclusion of all possible alternatives may seem to be a conservative approach, nonetheless, it can be unrealistic as it assumes decision makers have perfect knowledge about all alternatives. This approach can result in assigning non-zero selection probabilities to some alternatives that otherwise may not be known or be available to the decision maker. On the other hand, random selection of a few alternatives for the choice set by stratified sampling or other similar approaches can result in biased and possibly inaccurate parameter estimation.

In addition to the two aforementioned approaches, there are other methods to address the choice set formation issue. Disaggregate alternatives can be combined into more aggregated sets which consequently result in choice set size reduction. This alternative aggregation method is satisfactorily studied in the literature from different perspectives (Kitamura et al., 1979; Ben-Akiva and Lerman, 1985). Alternatively, a selected set of all possible alternatives can be chosen to form a smaller and more manageable choice set using a heuristic or non-heuristic approach, in which alternatives are evaluated by certain criteria for being included in the choice set. The later method has not been sufficiently studied and it is the main target of this paper (see for example Arnold et al., 1983).

This study aims to introduce a behavioral method for housing search choice set formation followed by an application of this behavioral choice set formation in a discrete choice model. The residential location choice process of this study starts with an alternative evaluation and screening step. The alternatives are filtered based on average household work distance using the individuals' priorities, lifestyle, preferences, and utilities. Note that the use of average work distance in choice set composition implies that residential choice is in this instance conditional on workplace choices of the individual's in the household. While there are several other factors that clearly affect the selection of housing alternatives (e.g., property value, commute distance, school quality, safety, tax rate, etc.), in order to show the practicality of the approach, only average work distance is considered in the choice set formation stage. The remaining variables will be accounted for in the final location choice model. The final residential location selection behavior is modeled using a multinomial logit formulation with the sampled choice sets, in which sampling correction factors are included to remove the sampling bias affecting the parameter estimations introduced from the choice set formation stage. Additionally, the systematic spatial dependencies among the alternatives are included in the model using an additional deterministic utility term added to the original utility function of the discrete choice model.

It is important to reiterate that this paper only discusses the household housing search behavior which is conditional on the household residential relocation timing decision. It has been previously discussed elsewhere (Rashidi et al., 2011) that the timing of the residential relocation is endogenously correlated with household employed members' job relocation timing decisions. In the current paper, therefore, the search behavior of the household for finding the most appropriate zone is modeled given that workplace choices are fixed. Note, however, that the converse situation of workplace choice being dependent on household location could similarly modeled in the manner presented in this work with little modification – although it would be necessarily conducted at the individual rather than household level.

The rest of the paper is organized as follows. First, a brief literature review is presented and the study approach is discussed. Then the choice set design algorithm along with the data used for its development are presented. Following this discussion, the discrete choice model, data, methodology and results, are given. Finally, conclusions and future research directions are discussed.

2. Background and study approach

Choice set formation has been the topic of many research studies following the introduction of the discrete choice model method. In this section, an overview of previous work regarding various methods to form an appropriate choice set is presented. This overview starts by discussing simple binary choice set formation methods, continues to discuss nested choice set models and finally more advanced probabilistic approaches used to construct a choice set are reviewed.

The choice set formation problem can be traced back to early applications of discrete choice models. Ben-Akiva and Lerman (1985) proposed the stratified sampling procedure to generate the alternative set and showed the efficiency of that approach. Spiggle and Sewall (1987) introduced three levels in screening the alternatives and finding the final choice set: awareness set, evoked set and choice set. He borrowed the term evoked set from another study by Howard who originally introduced it in 1963. According to his model, the awareness set consists of all alternatives the consumer is aware of. This set is then filtered to the evoked set which is a subset of the awareness set and consists of those alternatives that meet certain criteria for further consideration. Finally, the choice set is a subset of the evoked set in which there are very few alternatives including the final choice which is the immediate group of alternatives before making a decision. Shocker et al. (1991) employed the term consideration set for evoked set which was originally introduced in a study by Wright and Barbour in 1977.

Other than the different definitions for the choice set, various solutions have been introduced to deal with the choice set problem. Willumsen and Ortuzar, 2001 listed three ways for tackling the choice set problem available in the literature. These include using deterministic rule-based frameworks, asking individuals about the alternatives considered during the survey