The Journal of Choice Modelling 9 (2013) 3–13

Contents lists available at ScienceDirect

The Journal of Choice Modelling

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

Simplified probabilistic choice set formation models in a residential location choice context

Alireza Zolfaghari *, Aruna Sivakumar [1], John Polak [2]

Centre for Transport Studies, Imperial College London, London, United Kingdom

ARTICLE INFO

Available online 10 January 2014

Keywords:
Choice set formation
Semi-compensatory models
Ordered probit

ABSTRACT

The implementation of a theoretically sound, two-stage discrete-choice modelling paradigm incorporating probabilistic choice sets is impractical when the number of alternatives is large, which is a typical case in most spatial choice contexts. In the context of residential location choice, Kaplan et al. (2009), (2011), (2012) (KBS) developed a semi-compensatory choice model incorporating data of individuals searching for dwellings observed using a customised real estate agency website. This secondary data is used to compute the probability of considering a choice set that takes the form of an ordered probit model. In this paper, we illustrate that the simplicity of the KBS model arises because of an unrealistic assumption that individuals′ choice sets only contain alternatives that derive from their observed combination of thresholds. Relaxing this assumption, we introduce a new probabilistic choice set formation model that allows the power set to include all potential choice sets derived from variations in thresholds′ combinations. In addition to extending the KBS model, our proposed model asymptotically approaches the classical Manski model, if a suitable structure is used to categorise alternatives. In order to illustrate the biases inherent in the original KBS approach, we compare it with our proposed model and the MNL model using a Monte Carlo experiment. The results of this experiment show that the KBS model causes biases in predicted market share if individuals are free to choose from any potential choice sets derived from combinations of thresholds.

1. Introduction

Standard discrete choice models are powerful in choice situations where individuals′ choice sets are known to the analyst with a degree of confidence. However, there are many choice contexts where this is not the case. Examples are typically found in spatial choices such as residential location choice where the number of alternatives evaluated is constrained by individuals′ limited capacity for gathering and processing information (Fotheringham et al., 2000). The inappropriate use of the universal choice set in modelling can lead to significant misspecification errors where the actual choice sets considered by decision makers are different from the universal choice set (see Stopher, 1980).

Experimental research suggests that in a complex choice situation (e.g., choice among a large number of alternatives) decision makers adopt non-compensatory screening strategies (e.g., elimination-by-aspects, see Tversky, 1972) to reduce the

* Corresponding author. Tel.: +44 2075946086.
E-mail addresses: a.zolfaghari07@imperial.ac.uk (A. Zolfaghari), a.sivakumar@imperial.ac.uk (A. Sivakumar), j.polak@imperial.ac.uk (J. Polak).
[1] Tel.: +44 2075946036.
[2] Tel.: +44 2075946089.

1755-5345/$ - see front matter
http://dx.doi.org/10.1016/j.jocm.2013.12.004

number of alternatives to a smaller number before using a compensatory decision rule to make a final decision (see Manrai and Andrews, 1998). This has led to the view that decisions may be made in two stages: (a) non-compensatory stage and (b) compensatory stage.

Manski (1977) proposed a two-stage discrete-choice modelling framework incorporating probabilistic choice sets. The general Manski model considers all potential choice sets and requires summation over the power set G which is a set of all non-empty subsets of the master set. The size of the power set G, however, increases exponentially with the increase in the number of alternatives. Therefore, it is practically impossible to estimate the general Manski model with a large number of alternatives in the master choice set.

Some authors have tried to overcome the computation burden of the Manski approach by imposing *a priori* restrictions on the composition of choice sets based on some exogenous evidence from the choice set formation process (see Swait and Ben-Akiva, 1987; Siddarth et al., 1995; Andrews and Manrai, 1998; Zheng and Guo, 2008; Kaplan et al., 2009; Hicks and Schnier, 2010). In the context of residential location choice, Kaplan et al. (2009), (2011), (2012) (KBS, hereafter) developed a probabilistic choice set model incorporating data of individuals searching for dwellings observed using a customised real estate agency website. This secondary data was used to compute the probability of considering a choice set that takes the form of an Ordered Probit model. In this paper, we illustrate that the simplicity of the KBS model arises because of an unrealistic assumption that individuals′ choice sets only contain alternatives that derive from their observed combination of thresholds. By relaxing this assumption we introduce a new probabilistic choice set formation model that allows the choice set space to include all potential choice sets derived from variations in the combinations of thresholds.

In addition to extending the KBS model, our proposed model asymptotically approaches the classical Manski model, provided a suitable structure is used to categorise alternatives. In order to illustrate the biases inherent in the original KBS approach, we compare it with our proposed model and the MNL model using a Monte Carlo experiment. The results of this experiment show that the KBS model causes biases in predicted market share if individuals are free to choose from any potential choice sets derived from combinations of thresholds.

The rest of this paper is organised as follows. Section 2 reviews the literature in choice set formation problem discussing different approaches and methodologies applied to reduce the computation burden of the Manski model. Section 3 elaborates more on the model developed by Kaplan et al. (2009), (2011), (2012)and discusses the shortcomings of this model. Section 4 presents the proposed modelling framework of this research. Section 5 presents the details of the Monte Carlo experiment, including data generation, different model estimation and evaluation criteria as well as a discussion of the findings. Section 6 concludes this paper.

2. Literature review

An important issue in models with a massive universal choice set concerns the fact that decision makers do not choose from the universal choice set when they are facing a large number of alternatives. This issue is known in the literature as the choice set formation problem (Thill, 1992, in the context of destination choice). Since the observed choice data do not reveal any information about the actual choice sets, researchers have proposed different methodologies to deal with the issue of choice set formation based on decision theory.[3] Thill (1992) and more recently Pagliara and Timmermans (2009) reviewed different choice set formation approaches applied in the spatial choice context. Focusing on residential choice, different choice set formation approaches are reviewed in the following sub-sections.

2.1. Deterministic choice set formation

In the deterministic (or exogenous) choice set formation approach, it is assumed that decision makers′ choice sets can be determined deterministically and exogenous to the choice process. The choice set for each decision maker, in the deterministic approach, is typically generated by restricting the universal choice set using deterministic constraints on one or more attributes of alternatives. Many applications of this approach exist in the literature for different choice contexts. In the mode choice context, for example, knowledge that an individual has no driving licence may lead the analyst to an *a priori* removal of the auto-mode from the individual′s choice set (see Ben-Akiva and Lerman, 1974; Train, 1980). In the destination/activity location choice context, a distance threshold, for example, is determined deterministically and used to generate the choice sets based on the hypothesis that individuals do not consider alternatives that would imply travelling more than a specified maximum distance (see Parsons and Hauber, 1998; Termansen et al., 2004; Scott, 2006).

Some authors have also proposed applying the importance sampling approach (without considering the alternative-specific correction terms) to incorporate more behavioural realism in the choice set formation stage for spatial models. For example, in the residential location choice component of ILUTE, households′ choice sets are constructed by taking 75% of a random sample from dwellings that were within 15 km of their previous locations and the remaining 25% from dwellings that were not within the 15 km threshold (Farooq and Miller, 2012). Elgar et al. (2009) also applied the importance sampling approach (without alternative-specific correction terms) in a firm location choice context by oversampling of alternatives

[3] There have been some attempts to model choices as an outcome of a search process based on the search theory rather than the decision theory and using the “choice process” data rather than the choice data (see Lerman and Mahmassani, 1985; Caplin and Dean, 2011).