Journal of Choice Modelling 28 (2018) 28–40

Contents lists available at ScienceDirect

Journal of Choice Modelling

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

Identifying the presence of heterogeneous discrete choice heuristics at an individual level

Felipe Gonzalez-Valdes *, Sebastián Raveau

Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile, Av. Vicuna Mackenna 4860, Macul, Santiago, Chile

ARTICLE INFO

Keywords:
Heterogeneous heuristics
Discrete choice
Random utility maximization
Elimination by aspects
Stochastic satisficing

ABSTRACT

Discrete choice models that allow heterogeneous choice heuristics have been recently proposed and applied in different contexts. These models are traditionally based on latent classes, where each class represents a choice heuristic. Different specification challenges arise from applying the traditional approaches and identifiability issues are not unusual. We propose a Mixed Heuristic Model (MHM) to identify the presence of different choice heuristics through a latent class approach, giving higher flexibility to the class membership function through a random variable and analyse it at an individual level. The MHM identifies individuals who are likely to have followed the choice heuristic without explicitly specifying the class membership function. The MHM also avoids specifying the class membership and choice levels simultaneously. The MHM is tested on simulated data and applied to model an air travel survey. Results show that the MHM is able to identify the presence and structure of the heuristics with high accuracy within the simulated data. On the real data, the MHM identifies Random Utility Maximization and Stochastic Satisficing behaviour within the individuals.

1. Introduction

Individuals may follow different decision rules, which can be modelled through specific choice heuristics, to choose an alternative out of a potentially extensive choice set (Wansink and Sobal, 2007; Denstadli et al., 2012). Several choice heuristics have been proposed for modelling behaviour, such as Random Utility Maximization (RUM) (McFadden, 1974), Elimination by Aspects (EBA) (Tversky, 1972a; b), Random Regret Minimization (RRM) model (Chorus et al., 2008), and Satisficing (Simon, 1955).

Although there are several alternative choice heuristics, traditional modelling approaches assume that only one of them is present within a population. In this sense, heterogeneity in preferences is usually considered, but no heterogeneity in decision rules is considered. However, if different choice heuristics coexist within a population, traditional models could lead to serious problems when understanding people's behaviour and forecasting their decisions (Williams and Ortuzar, 1982).

To avoid these potential problems, models that simultaneously accommodate different choice heuristics are desirable. Nevertheless, formulating such models is a complex task, since different decision rules must be incorporated into a single model. Indeed, models that explicitly account for different choice heuristics have been used only recently (Araña et al., 2008; Hess et al., 2012; Leong and Hensher, 2012; McNair et al., 2012; Adamowicz and Swait, 2013). The most popular approach to model multiple choice heuristics is based on latent class (LC) models, which have two decision levels. In the first level, the class membership level, the choice heuristic followed by the individuals is modelled. In the second level, the choice level, the preferences for alternatives are modelled conditional on the choice

* Corresponding author.
E-mail address: fagonzalezv@uc.cl (F. Gonzalez-Valdes).

https://doi.org/10.1016/j.jocm.2018.05.001
Received 29 May 2017; Received in revised form 8 May 2018; Accepted 11 May 2018

heuristic followed.

There are three main challenges that must be addressed when using a LC framework to model multiple choice heuristics: i) class membership is hard to model, ii) both LC levels must be modelled simultaneously, and iii) there is no certainty that the considered choice heuristics are present within the population.

Class membership is usually modelled by using the inverse Logit function. The most frequent formulations only have a population-wide choice heuristic constant (i.e. all individuals in the population have the same probability of following a specific choice heuristic). This simple formulation is unfortunately necessary, since more general and informative formulations have been found to be unidentifiable in practice. To the best of the authors' knowledge, only two studies has been reported where a non-constant population-wide class membership function has been estimated (Leong and Hensher, 2012; Hess and Stathopoulos, 2013). Leong and Hensher (2012) work with a class of RUM and another of RRM individuals. In their experiment, some parameters of the RUM and the RRM class had to be fixed (as one equals the minus of the other) in order for the model to be identifiable. Hess and Stathopoulos (2013) also worked with RUM and RRM; in their experiment, they had to use latent variables to explain the class membership function and choices to estimate the heuristics' parameters, otherwise, no identifiability could be found just from the choices of the individuals. Therefore, identifying heuristics and class membership function simultaneously is not straightforward.

Specifying the class membership level and the choice level simultaneously has further complications than identifiability. If modelling individuals' choices following a single choice heuristic is challenging (as it has been a research topic for decades), modelling their choices through multiple choice heuristics is even harder, since each of them must be modelled. As additionally a class membership level needs to be modelled, then the complexity of the model increases further more. This issue is enhanced by the interaction between the two levels, since modifying one of them affects the other.

Selecting the choice heuristics to be considered is a significant challenge. Even though in practice RUM models are the most popular alternative, there are numerous alternative choice heuristics to model individuals' behaviour. Ideally, a large number of choice heuristics could be considered, and a LC model would indicate if a given heuristic is present or not in the population by assigning a low probability to it. However, in practice, LC models have identifiability problems even when simply modelling two classes. Therefore a limited number of choice heuristics must be properly selected beforehand, or different combinations of heuristics can be tested.

Is important to consider that these three challenges interact with each other. The LC model could be used to test the presence or absence of a particular choice heuristic within a population. However, if the model's results indicate the absence of a heuristic, this could be due to an erroneous class membership function, due to a misspecification of the choice heuristic itself in the choice level, or due to its real absence.

We propose a model that addresses the three challenges of the LC approach. Our model, the Mixed Heuristic Model (MHM), can help identify the presence of the choice heuristics by giving higher flexibility to the class membership function. This flexibility is given by a random variable that allows to capture heterogeneity in decision rules. This heterogeneity is confirmed by analysing the class membership probability at an individual level.

The remainder of this paper is organized as it follows. In Section 2, we formalise the traditional and the proposed LC models for identifying the presence of heterogeneous heuristics. In Section 3, we test the accuracy of the MHM with synthetic data, exploring the effect of different sample sizes and responses per individual. In Section 4, we apply the MHM to an air travel survey, contrasting the classical RUM against five other choice heuristics. Finally, in Section 5 we present the main conclusions of the study, showing how the MHM tackles the three traditional modelling challenges.

2. The Mixed Heuristic Model

When applying LC models to the context of multiple choice heuristics, each class represents an independent heuristic. This way, the probability of individual $q \in Q$ choosing alternative *i*, P_{qi} has two components. The first component is the probability of individual q following choice heuristic $m \in M$, defined by the class membership function $\pi_{q,m}$. The second component is the probability of individual q choosing alternative i conditional to following choice heuristic m, $P_{q,i}(m)$. Then, the probability of individual q choosing alternative i ($P_{q,i}$) is given by the *total probability* presented in Equation (1).

$$P_{q,i} = \sum_{m \in M} P_{q,i}(m) \cdot \pi_{q,m} \tag{1}$$

In Equation (1), the functional form of $P_{q,i}(m)$ depends on the specific choice heuristic *m*. For example, if the classic RUM model is one of the considered heuristics, then its probability $P_{q,i}(m)$ is given by the well-known ratio of the exponential of the utility functions.

The probability of individual q following choice heuristic m, $\pi_{q,m}$, is typically given by the inverse Logit function (Equation (2)). To explain the fact that different individuals may follow different choice heuristics, this probability takes as an input a $\gamma_{q,m}$ function that varies across heuristics and individuals.

$$\pi_{q,m} = \frac{exp(\gamma_{q,m})}{\sum_{n \in M} exp(\gamma_{q,n})} \tag{2}$$

The main difference between the current approach to model heterogeneous discrete choice heuristics and the proposed MHM lies on the specification of the $\gamma_{q,m}$ function. This new functional form handles the three challenges exposed in section 1 as presented in section 2.2.