Journal of Statistical Planning and Inference 141 (2011) 2276–2286

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Effect of choice complexity on design efficiency in conjoint choice experiments

Vishva Manohara Danthurebandara [a,*], Jie Yu [a,1], Martina Vandebroek [b,2]

[a] Faculty of Business and Economics, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven, Belgium
[b] Faculty of Business and Economics & Leuven Statistics Research Centre, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven, Belgium

ARTICLE INFO

Article history:
Received 30 October 2009
Received in revised form
6 January 2011
Accepted 13 January 2011
Available online 20 January 2011

Keywords:
Choice complexity
Optimal experimental design
Entropy
Heteroscedastic conditional logit model
Between respondent variability

ABSTRACT

Conjoint choice experiments have become a powerful tool to explore individual preferences. The consistency of respondents' choices depends on the choice complexity. For example, it is easier to make a choice between two alternatives with few attributes than between five alternatives with several attributes. In the latter case it will be much harder to choose the preferred alternative which is reflected in a higher response error. Several authors have dealt with this choice complexity in the estimation stage but very little attention has been paid to set up designs that take this complexity into account. The core issue of this paper is to find out whether it is worthwhile to take this complexity into account in the design stage. We construct efficient semi-Bayesian *D*-optimal designs for the heteroscedastic conditional logit model which is used to model the across respondent variability that occurs due to the choice complexity. The degree of complexity is measured by the entropy, as suggested by Swait and Adamowicz (2001). The proposed designs are compared with a semi-Bayesian *D*-optimal design constructed without taking the complexity into account. The simulation study shows that it is much better to take the choice complexity into account when constructing conjoint choice experiments.

1. Introduction

Conjoint choice experiments have become increasingly popular for collecting and studying preferences of individuals. Discrete choice models are usually derived under the assumption of utility-maximizing behavior of the decision makers. In a random utility model, the utility that a decision maker obtains from an alternative is described by a structural part, with information about the alternative, and an error term which represents all other influences. In most applications, this error term is assumed to have the same variance in all choice sets. However, according to Swait and Adamowicz (2001), people use different strategies to cope with complex situations. Therefore one can expect that the error variance will vary with the complexity of the choice set. In this paper we will use the heteroscedastic model that was proposed to model this between respondent variability and develop optimal designs to estimate this model efficiently.

* Corresponding author. Tel.: +32 16 32 69 63; fax: +32 16 32 66 24.
E-mail addresses: Vishva.Danthurebandara@econ.kuleuven.be (V.M. Danthurebandara), Jie.Yu@econ.kuleuven.be (J. Yu), Martina.Vandebroek@econ.kuleuven.be (M. Vandebroek).
[1] Tel.: +32 16 32 69 62.
[2] Tel.: +32 16 32 69 75.

0378-3758/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2011.01.008

In the literature of conjoint choice experiments, there is only limited research on how changes in the structure of the choice set changes choice outcomes (see for example DeShazo and Fermo, 2002). The complexity of a choice set is however crucial, because it directly affects the choice consistency. That it is easier to make a choice between two alternatives with few attributes than between five alternatives with a lot of attributes will be reflected in a higher response error in the latter case. In the literature, several measurements have been introduced to quantify the choice complexity. DeShazo and Fermo (2002) used five measurements, which describe the structure of the choice set. Sándor and Franses (2009) used two of the complexity measures of DeShazo and Fermo (2002) and one price related measure, which is the key factor of their empirical study. Severin (2000) used one major complexity measurement which is the number of trade-offs that respondents have to make in their decision process. This can also be referred to as the similarity of alternatives in terms of attribute levels. Mazzotta and Opaluch (1995) and Dellaert et al. (1999) use the number of attributes that vary across the alternatives to measure the complexity in their research on choice consistency and complexity. In all the studies mentioned above, statistics that describe the structure of the choice set are used to measure the complexity. Swait and Adamowicz (2001) argued that each of these measurements is a component of complexity rather than an overall measure. Therefore, they introduced entropy as an overall complexity measure, which summarizes the impact of the number of alternatives, the number of attributes, the correlation among attributes and the similarity among utilities of alternatives.

The statistical design is one of the key challenges in implementing a conjoint choice experiment, since the efficiency of the parameter estimates depends on the design. Most of the authors referred to above do not assess the effect of choice complexity on the statistical design of the experiment. By ignoring choice complexity when designing the experiment, the choice data obtained will be inconsistent with the estimation model. Hence, the experimental design obtained cannot be optimal.

The core issue we address in our paper is whether it is worthwhile to take the choice complexity into account when constructing the design. Standard models assume that respondents have unlimited information processing capacity, which allows them to make their choice in a strictly optimal way irrespective to the complexity of the choice situation (de Palma et al., 1994). The heteroscedastic conditional logit model proposed by Swait and Adamowicz (2001) however uses the scale factor to bring the complexity into the model. We use their parameterization to model the between respondent variability that occurs due to the choice complexity. In our research, we propose efficient semi-Bayesian *D*-optimal designs, constructed by considering the choice complexity. The proposed designs are compared with two semi-Bayesian *D*-optimal designs which are constructed ignoring the choice complexity but for the rest uses the same design setting as the proposed design.

We organize this paper as follows. In Section 2, we present the theoretical model, discuss the complexity measure, the design efficiency criterion, the design construction algorithm and the benchmark designs we used. This is followed in Section 3 by a relative design efficiency study. In Section 4, we present the simulation study setup, the proposed design settings and the estimation results. We discuss the impact of the misspecification of the complexity function in Section 5 and, finally, in Section 6 we evaluate and summarize our key findings.

2. Methodology

2.1. Heteroscedastic conditional logit model

First consider the homoscedastic conditional logit model (McFadden, 1974), which is popular for analyzing the data from conjoint choice experiments. The random utility a given respondent n attaches to an alternative k in choice set s is given as

$$U_{ksn} = \mathbf{x}_{ks}'\boldsymbol{\beta} + \varepsilon_{ksn}, \tag{1}$$

where $\mathbf{x}_{ks}$ is a m-dimensional vector containing the attribute values of alternative k in choice set s, $\boldsymbol{\beta}$ is a m-dimensional vector of parameters and U_{ksn} is the utility that the decision maker n actually obtained from alternative k in choice set s. The error term ε_{ksn} is assumed to have an extreme value distribution. Assuming there are K alternatives in a choice set, the probability that alternative k is chosen from choice set s is

$$q_{ks} = \frac{\exp\,(\mu\mathbf{x}_{ks}'\boldsymbol{\beta})}{\sum_{i=1}^{K}\exp\,(\mu\mathbf{x}_{is}'\boldsymbol{\beta})}, \quad k = 1,\ldots,K, \tag{2}$$

where μ is the scale factor. The scale factor is defined as $\pi/\sqrt{6}\sigma$, where σ is the standard error of ε. This model is called the homoscedastic conditional logit model since the scale factor is assumed constant.

An increase in choice set complexity will add noise to the error term of the random utility function (DeShazo and Fermo, 2002) which is reflected in a higher response error σ. Thus, the error variance σ^2 and, hence, also the scale μ depends on the choice complexity and we will denote this dependency explicitly in

$$p_{ks} = \frac{\exp(\mu(C_s)\mathbf{x}_{ks}'\boldsymbol{\beta})}{\sum_{i=1}^{K}\exp(\mu(C_s)\mathbf{x}_{is}'\boldsymbol{\beta})}, \tag{3}$$

where C_s measures the complexity of choice set s. The resulting model is called the heteroscedastic conditional logit model (HCLM) by DeShazo and Fermo (2002) and Swait and Adamowicz (2001).