The Journal of Choice Modelling 16 (2015) 43–49

Contents lists available at ScienceDirect

The Journal of Choice Modelling

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

Manipulating a stated choice experiment

Mogens Fosgerau [a,b,*], Maria Börjesson [b]

[a] Technical University of Denmark, Denmark
[b] Royal Institute of Technology, Sweden

ARTICLE INFO

Article history:
Received 7 August 2014
Received in revised form
25 September 2015
Accepted 27 September 2015
Available online 3 November 2015

JEL:
C8
C9
C25

Keywords:
Stated choice
Willingness to pay
Misspecification
Experimental design

ABSTRACT

This paper considers the design of a stated choice experiment intended to measure the marginal rate of substitution (MRS) between cost and an attribute such as time using a conventional logit model. Focusing the experimental design on some target MRS will bias estimates towards that value. The paper shows why this happens. The resulting estimated MRS can then be manipulated by adapting the target MRS in the experimental design.

1. Introduction

This paper considers a stated choice experiment designed to measure the marginal rate of substitution (MRS) between cost and another attribute of the choice alternatives. We present a case where the design is centered on some target MRS and where the MRS is estimated by applying a conventional logit model. We demonstrate that if the random individual response heterogeneity is mainly driven by heterogeneity of the MRS rather than by response errors, then the estimated MRS will tend to be biased towards the target. This bias hinges on misspecification of the logit model, which has constant marginal utilities and additive random residuals. Thus, this bias will tend to confirm the target MRS and may thus not be informative about consumer preferences.

Seasoned choice experiment designers probably know this already. The objective of this paper is to make clear how the mechanism works and to make this insight more widely accessible.

Stated choice experiments are used within a range of applied fields of economics, including energy, transportation, health, tourism, agricultural and environmental economics. They are also used by consultants in a variety of contexts and in marketing. The mechanism explored in this paper may not only be present in stated choice data, but also when the MRS is estimated on revealed preference data.

The mechanism described in the present paper should not be confused with other sources of bias on the MRS inferred from stated choice data. The choice experiment can influence the output for a number of reasons. Many cognitive processes

* Corresponding author at: Technical University of Denmark, Denmark.
E-mail addresses: mf@transport.dtu.dk (M. Fosgerau), maria.borjesson@abe.kth.se (M. Börjesson).

http://dx.doi.org/10.1016/j.jocm.2015.09.005
1755-5345/

are at play when people make decisions, typically simplified by various decision rules, heuristics, implying that the experimental design can influence the result (Hensher, 2014; Leong and Hensher, 2012). There are empirical (Bliemer and Rose, 2011) and theoretical indications (Burgess and Street, 2005; Sandor and Wedel, 2005) that attribute levels and the order of them within the experiment influence output (see Rose and Bliemer, 2014, for a review). De Borger and Fosgerau (2008) and Hess et al. (2008) show that preference asymmetries such as different valuation of gains or losses, and reference effects influence respondent's stated choices, with De Borger and Fosgerau (2008) linking these phenomena to prospect theory (Tversky and Kahneman, 1991). In particular, the value of time from stated choice experiments has been found to depend on the size and sign of the attribute differences between the alternatives. A commonly found sign effect is that gains are valued less than losses; this is called loss aversion. Size effects refer to cases where the estimated MRS is found to depend on the size of the attribute difference between alternatives (e.g., Mackie et al., 2001; Hultkrantz and Mortazavi, 2001; Bates and Whelan, 2001; Fosgerau, 2006). A commonly found size effect is that small time savings are valued less per minute than large time savings.

Adaptive design techniques can introduce bias in the estimated MRS through yet another mechanism, namely that attribute levels and unobserved heterogeneity in the respondents preferences become correlated (Bradley and Daly, 1993). Adaptive designs then lead to a self-imposed endogeneity problem, violating the statistical assumptions underlying standard models, and making subsequent statistical inference invalid.

There is evidence that stated choices are sometimes subject to hypothetical bias (Harrison, 2014). In the context of risky choices, nonlinearities regarding attitudes and perceptions of risks imply that the design of the experiment in terms of risk levels and presentation of those influence the choices (Bates et al., 2001; Liu and Polak, 2007; Borjesson and Eliasson, 2011; Loomes and Blackburn, 2014). These other potential sources of bias are not in focus in this paper.

The mechanism we discuss in the present paper relates to the boundary value design approach (Fowkes and Wardman, 1988). The idea of this design approach is to choose boundary or trade-off values within a range where the analyst think the MRS distribution is located. Fowkes and Wardman point out that the boundary values should cover a reasonable range of potential variation in taste and uncertainty, but that it is often desirable to have them closer together in the range where the actual values are expected to be located. An implication of our results is that the boundary value approach should be avoided, since it runs the risk of biasing the estimated MRS toward the chosen boundary values. Moreover, it is not suited to models that account for preference heterogeneity.

A remedy to avoid the bias arising from the misspecification of the logit model is to explore the error structure of the data, using non-parametric techniques, before defining a parametric model. If such analysis indicates heterogeneity in the MRS, the parametric model should allow for heterogeneity in the MRS. The choice of parametric distribution for the MRS is crucial and should be tested (Fosgerau and Bierlaire, 2007; Fosgerau and Mabit, 2013).

To estimate the distribution of the MRS, a key condition is that the full distribution of the MRS is uncovered by the data. If this condition is not met, then the estimates of the mean and other moments of the MRS distribution have to rely on assumptions about the shape of the distribution in the range where it cannot be identified by the data. Such assumptions are hard to verify. Fosgerau (2006) shows that when the tail of the MRS distribution is not revealed, then the choice of parametric distributions can result in arbitrarily high estimates of the mean MRS. The resulting estimate of the mean MRS will depend on the parts of the distribution that are extrapolated outside the range of data.

The paper is organized as follows. Section 2 describes the experimental context and choice generating process assumed when applying the standard logit model. Section 3 describes a different choice generating process, governed by the random MRS model, and describes the mechanism that tends to bias the estimated MRS. Section 4 describes the experimental design used in the simulation exercise in Section 5. Section 6 uses part of the Danish value of time data to empirically validate the theoretical predictions of Section 3. Section 7 concludes.

2. Choice setup and the model to estimate

The context of a trip is used for concreteness. The alternatives can be different routes by some mode of transportation, for example; Section 6 uses data concerning rail trips. Each subject is asked to choose between two alternatives for a trip and each route is characterized by an associated monetary cost and a travel time.

This setting can be used to reveal the subjects' rate of substitution between travel time and cost. By design, one route will be faster but also more expensive than the other. With everything else being equal, subjects reveal through their choices whether their willingness to pay for the time saving associated with the faster route is greater or smaller than the cost difference between the two routes.

The canonical model for this situation is the logit model with the foundation in the theory of consumer demand developed by McFadden (1974). The conventional and widely used specification of the binary logit model in this setting uses indirect utilities for the two alternatives that are linear indices $v_{ij} = \alpha c_{ij} + \beta t_{ij} + \gamma 1_{\{j=2\}} + \varepsilon_{ij}$, where subscript i indexes individuals, $j = 1, 2$ indexes choice alternatives, c_{ij} is the travel cost for individual i in alternative j with corresponding marginal utility α, t_{ij} is the travel time with corresponding marginal utility β, γ is a constant specific to alternative 2 and ε_{ij} are i.i.d. extreme value type 1 random residuals. The marginal utilities α, β are expected to be negative. Each individual chooses the alternative yielding the highest indirect utility. Then the probability that alternative 1 is chosen is