ARTICLE IN PRESS

Journal of Choice Modelling xxx (2017) 1–18

Contents lists available at ScienceDirect

Journal of Choice Modelling

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

Measuring respondent uncertainty in discrete choice experiments via utility suppression☆

Yu-Cheng Ku[a,*], John Wu[b]

[a] Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA
[b] The Federal Reserve Bank of Richmond, Charlotte, NC 28230, USA

ARTICLE INFO

Keywords:
Discrete choice experiment
Inverse Mills ratio
Multinomial probit
Respondent uncertainty
Truncated normal distribution

ABSTRACT

Discrete choice experiments (DCEs) are an important methodology in survey research. Although DCEs assume that respondents know their true preferences and can make choices precisely, in reality, respondents may be uncertain. Respondent uncertainty has long been recognized as an issue in the DCE context, but unfortunately, its explicit modelling has received only limited attention. This paper proposes using hierarchical Bayes multinomial probit models to construct respondent uncertainty measures based on the utility difference and a concept called utility suppression, which is quantified by the derivative of the inverse Mills ratio. Two empirical studies with different DCE design formats are conducted to assess and compare the performance. The first includes a no-choice option, and the second is forced-choice formatted with a simpler design. The hold-out validation shows that the utility difference and our proposed measure, which are consistent with the random utility maximization assumption, have better overall performance, particularly in identifying the high-uncertainty respondents. We further show how the information about respondent uncertainty can be utilized in practice.

1. Introduction

Discrete choice experiments (DCEs), also known as choice-based conjoint analysis, are an important methodology applied to a broad range of survey research areas, including resource and environment [Viscusi et al. (2008) and Hoyos (2010)], health [Ryan and Gerard (2003) and Blinman et al. (2012)], marketing [Zwerina (2013)], public policy [Vossler et al. (2012)] and transportation [Hensher and Rose (2005)]. As a stated preference (SP) method based on the random utility maximization assumption (RUM, Marschak (1960)), DCEs assume that respondents know their true preferences and can make choices precisely, given the utilities derived from the products presented to them. However, due to several factors, such as respondents' socio-demographic characteristics, design complexity, cognitive burden, and lack of attention, the respondents may be uncertain about which products they actually prefer. Thus, their responses could be unreliable and lead to biased inference.

Long recognized as an issue for SP studies, respondents' preference uncertainty was formalized by Li and Mattsson (1995) in the context of Contingent Valuation (CV), which is another type of SP method usually aimed at gauging people's willingness to pay (WTP). Wang (1997) hypothesized that the preference uncertainty (hereafter referred to as "respondent uncertainty") can be characterized by

☆ The opinions expressed in this article belong to the authors and do not represent official positions of the Federal Reserve Bank of Richmond or the Federal Reserve System.

* Corresponding author.
E-mail address: yku2@ncsu.edu (Y.-C. Ku).

http://dx.doi.org/10.1016/j.jocm.2017.08.002
Received 1 November 2016; Received in revised form 15 August 2017; Accepted 15 August 2017
Available online xxxx

the difference between the true WTP and indicated payments. Specifically, a small difference implies that an individual's preference is vague and thus indicates high uncertainty. In CV studies, an important application of respondent uncertainty is the mitigation of hypothetical bias,[1] in which the certainty degrees obtained from follow-up questions are utilized to recode the data for *ex post* calibration of the stated WTP to the actual level [see e.g., Champ et al. (1997), Champ and Bishop (2001), Akter and Bennett (2013) and Loomis (2014)]. The existing answer-recoding methods from the CV literature was first systematically evaluated by Lundhede et al. (2009). These methods include elimination from the sample and recoding the answers to either status quo or to the second-best alternative. Lundhede et al. (2009) find that the effects of these recoding schemes on adjusting WTP measures are insignificant and argue that, because the recoding schemes either change respondents' stated preferences or leave out information, none of them seems to be a satisfactory way of handling uncertain answers. On the other hand, they also find that incorporating the stated uncertainty into the logit scaling parameter can reduce the model's unexplained variance considerably. To mitigate potential endogeneity bias and measurement error issues implied in most applied CV methods that utilize the stated certainty, Dekker et al. (2016) proposed an integrated choice and latent variable model [ICLV, see e.g., Ben-Akiva et al. (2002)], which treats choice uncertainty as a latent factor that simultaneously affecting respondents' choices and responses to the certainty questions.

For DCEs, several uncertainty metrics have been introduced and studied. A commonly used one is the root likelihood [RLH, see Orme (2009) and Hofstede et al. (2014)]. As a goodness-of-fit measure, RLH is often used to assess whether the information captured by the model is greater than random guess. Swait and Adamowicz (2001) propose using entropy [Shannon (2001)] to measure respondent uncertainty from the aspects of choice complexity and preference similarity. The entropy value is typically used to quantify the information explained by the choice model [see e.g., Hauser (1978)]. Both RLH and entropy are applied in multinomial logit (MNL) models and are derived from the respondents' representative utilities only, which is arguably not consistent with the spirit of RUM. Olsen et al. (2011) extend Wang (1997)'s hypothesis into DCE studies. They hypothesize and show that respondents' self-reported certainty in choices increases with the utility difference, defined as the difference between the utility of the chosen alternative and the utility of the second-best alternative. In a recent work, Uggeldahl et al. (2016) investigate whether eye movements during the completion of choice tasks in DCEs are related to respondents' choice certainty and can explain variations in the scale factor. They find that, although eye movements can serve the purposes, the response time, which is a simpler measure routinely recorded in web-based surveys, performs better as a proxy for the stated choice certainty. While the aforementioned works provide seminal contributions to the understanding of respondent uncertainty in DCEs, unfortunately the explicit modelling of respondent uncertainty that fully follows RUM remains partially explored in the DCE context. This serves as the motivation for the study in this paper.

Based on the theory of Olsen et al. (2011), we apply hierarchical Bayes[2] (HB) multinomial probit (MNP) methods to estimate the utility difference for measuring respondent uncertainty. A related idea is discussed in Uggeldahl et al. (2016), where they use the difference in the representative utilities between the chosen and second-best alternatives as the uncertainty measure. By its construction, this measure again is not in line with RUM. We will discuss this issue in the methodology section. Another related idea is the utility deviation proposed by Ku et al. (2015), which can be viewed as an extension of Bradlow et al. (1998)'s largest absolute realized deviation to DCEs. Since utility deviation is based upon the absolute deviation between a respondent's true and representative utilities about the choice made, it is conceptually different from the respondent uncertainty characterized by Olsen et al. (2011). Our proposed approach is illustrated and tested using two empirical DCEs of different design formats and complexities.[3] The first DCE has a more complex design and includes a "no-choice" option, while the second is forced-choice formatted with less complexity. For the performance comparative study based on hold-out evaluation, we employ the standard performance metric, hit rate [see e.g., Natter and Feurstein (2002)], to validate the results. In both cases, it was found that respondents' hold-out hit rates decrease when the respondent uncertainty level increases, and vice versa. On the other hand, for respondents with exceptionally low uncertainty, the hit rate can reach 90% or higher, showing that the identified low-uncertainty respondents are very consistent in their preferences. The results, while showing the effectiveness of the utility difference, also provide support to the hypothesis of Olsen et al. (2011).

Additionally, we explore different ways of modelling respondent uncertainty. In a typical implementation of the Markov chain Monte Carlo (MCMC) for HB-MNP, the utility of the chosen alternative (U_1) and the utility of the second-best alternative (U_2) are convoluted. Specifically, the posterior draw of U_1 is made from a left truncated normal distribution in which the cut-off point is U_2. Consequently, the suppression resulting from U_2 to the range of U_1 determines the utility difference, $U_1 - U_2$. This observation inspired us to utilize the derivative of the inverse Mills ratio, denoted δ, as a quantification of the "utility suppression" to account for respondent uncertainty. The utility difference and δ are different from other aforementioned uncertainty measures in that they are derived based on the estimation of respondents' true utilities and hence have a direct connection with RUM. We benchmark the utility difference and δ against RLH and entropy in differentiating respondents' level of uncertainty. The benchmarking results show that the utility difference and δ perform nearly identically and have the best overall performance. Specifically, for the simpler DCE design, the four measures have comparable performances; when the design is more complex, the utility difference and δ perform significantly better in identifying high-uncertainty respondents, and entropy performs better for certain low-uncertainty levels.

The contribution of this work is three-fold. First, it verifies the hypothesis of Olsen et al. (2011) by showing the effectiveness of the utility difference as a measure of respondent uncertainty. Second, it proposes a novel and effective approach that measures respondent

[1] Hypothetical bias is broadly known as the divergence between the hypothetical WTP and real values.

[2] According to Train (2009), compared to classical estimation methods, the Bayesian procedure has two advantages. First, it does not require any function maximization, which can be challenging in many circumstances; second, in general, estimates derived with Bayesian procedures can attain desirable properties, such as efficiency and consistency, under more relaxed conditions.

[3] There are multiple dimensions in defining the design complexity [see e.g., Swait and Adamowicz (2001), Caussade et al. (2005) and Campbell (2008)], in this article we look at the number of attributes, largest number of attribute levels and number of alternatives.