European Journal of Operational Research 239 (2014) 556–564

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Interfaces with Other Disciplines

Likelihood estimation of consumer preferences in choice-based conjoint analysis

Merja Halme *, Markku Kallio

Department of Information and Service Economy, Aalto University School of Business, POB 21210, FI-00076 Aalto, Finland

ARTICLE INFO

Article history:
Received 14 October 2013
Accepted 30 May 2014
Available online 12 June 2014

Keywords:
Conjoint analysis
Utility function
Preference estimation
Maximum likelihood
Marketing research

ABSTRACT

In marketing research the measurement of individual preferences and assessment of utility functions have long traditions. Conjoint analysis, and particularly choice-based conjoint analysis (*CBC*), is frequently employed for such measurement. The world today appears increasingly customer or user oriented wherefore research intensity in conjoint analysis is rapidly increasing in various fields, OR/MS being no exception. Although several optimization based approaches have been suggested since the introduction of the Hierarchical Bayes (*HB*) method for estimating *CBC* utility functions, recent comparisons indicate that challenging *HB* is hard. Based on likelihood maximization we propose a method called *LM* and compare its performance with *HB* using twelve field data sets. Performance comparisons are based on holdout validation, i.e. predictive performance. Average performance of *LM* indicates an improvement over *HB* and the difference is statistically significant. We also use simulation based data sets to compare the performance for parameter recovery. In terms of both predictive performance and *RMSE* a smaller number of questions in *CBC* appears to favor *LM* over *HB*.

1. Introduction

In marketing research the measurement of preferences and assessment of utility functions have long traditions. Often conjoint analysis (*CA*) is employed as the utility measurement instrument and the estimation takes place on segment or individual level. The world today seems increasingly customer or user oriented. In October 2013 ISI Web of Knowledge found 2758 hits for "conjoint analysis" (within title, abstract or keywords) with a significant share devoted to Operations Research/Management Science. References from the past five years indicate that research intensity in conjoint analysis is rapidly increasing in various fields, management science being no exception.

This paper concerns choice based conjoint analysis (*CBC*) which appears the most popular type of *CA* today. For revealing individual preferences, estimation approaches and software for *CBC* analysis have developed significantly during past decades. The Hierarchical Bayes *HB* method for estimating individual utility functions (e.g., Allenby & Ginter, 1995) remains most popular, it is found to perform well and there is commercial easy-to-use software.[1] Estimation results are frequently used to predict, for instance, market shares; see e.g., Natter and Feurstein (2002). The fine qualities of *CBC* analysis combined with *HB* estimation are acknowledged, for instance, by Karniouchina, Moore, van der Rhee, and Verma (2008) who compare *CBC* analysis with the ratings-based conjoint analyses, another commonly used method.

The performance of alternative estimation approaches for *CBC* analysis (without adaptive question design) is compared in Halme and Kallio (2011) with *HB* as a benchmark. The study indicates that challenging *HB* is hard. However, based on likelihood maximization we propose in this article a new and highly promising method called *LM* (for Likelihood Maximization).

The log-likelihood function in *LM* is based on three sources of uncertainty as follows. First, considering the choices of a respondent in individual questions of *CBC* we adopt the likelihood for each choice from the multinomial logit model (McFadden, 1974). Second, in terms of valuation errors the respondents are heterogeneous and we assume that the individual standard deviations of such errors are independent random draws from an inverse-gamma distribution. Third, as in *HB*, considering the interdependence of preferences among respondents we assume that parameter vectors defining individual utility functions are random draws from a multivariate normal distribution (e.g. Allenby & Ginter, 1995).

Using twelve sets of field data *LM* achieves a statistically significant average improvement over *HB* in terms of predictive power.

* Corresponding author. Tel.: +358 40 353 8081.
E-mail addresses: merja.halme@aalto.fi (M. Halme), markku.kallio@aalto.fi (M. Kallio).

[1] If segment-wise utility functions are adequate then latent class approaches also have turned out promising (DeSarbo, Ramaswamy, & Cohen, 1995).

http://dx.doi.org/10.1016/j.ejor.2014.05.044
0377-2217/

In these tests, some of the respondent's questions are left for holdout. Having estimated the individual utility functions we attempt to validate the choices in the holdout question, i.e. test if the utility functions produced confirm the choices made by respondents. Predictive power is then measured by the share of confirmed choices among all holdout questions.

Furthermore, we employ simulated data sets to compare utility function parameter recovery as well. With a small number of questions per respondent, *LM* outperforms *HB* in terms of *RMSE* (the root mean square error).

Halme and Kallio (2011) also propose for *CBC* estimation an optimization based method *CP* with some similarities with *LM*: (i) both assume that individual part-worth vectors may be interpreted as random draws from a multivariate normal distribution; and (ii) in terms of valuation errors the respondents are heterogeneous. However, *CP* and *LM* differ in the methods of determining individual error levels (cross validation vs. likelihood maximization) as well as in choice models (projective penalty vs. multinomial logit). Numerical tests in Halme and Kallio (2011) indicate that average performance of *CP* is neither superior nor inferior to *HB*.

Next, Section 2 introduces the *CBC* model and the likelihood method *LM*, Section 3 presents performance comparisons of *LM* and *HB*, and Section 4 concludes.

2. Likelihood estimation for *CBC* analysis

We begin this section by introducing the choice model employed in preference estimation. Then the log-likelihood problem for estimation of part-worth vectors is defined. Finally, the approach is operationalized in a computer implementation.

2.1. The choice model for CBC

CBC analysis is a multi-attribute method where the value of a product or service stems from a given set of attributes which may only attain a small number of possible *levels*. Product *concepts* or *profiles* are defined by levels for each attribute. The *utility function* for a concept is assumed additive separable in attributes.

In *CBC* analysis each respondent is presented a questionnaire including a set of Q questions (tasks). In each question, a set of P concepts (product/service profile specifications) is presented for evaluation and the respondent indicates the best one. Typically Q is in the range from 5 to 20 and P from 2 to 6. The questions often are different for each respondent or group of respondents, and they are defined employing, for example, fractional factorial, random experimental, or orthogonal question design (see Chrzan & Orme, 2000; Nair, 2013). In this article, we exclude adaptive question designs assuming that each task in the questionnaire is independent of responses to preceding tasks.

We adopt additional notation from Halme and Kallio (2011) as follows:

- i = respondent, $i = 1, 2, \ldots, N$
- j = question, $j = 1, 2, \ldots, Q$
- k = profile alternative in a question, $k = 1, 2, \ldots, P$
- l = profile index to attribute levels, $l = 1, 2, \ldots, L$ (see example below)
- x_{ijk} = profile of alternative k in question j (row vector in R^L) for person i
- x_{ij1} = preferred profile to person i in question j
- $\Delta_{ijk} = x_{ij1} - x_{ijk}$ = *preferred direction* for all i, j and k.

For notational convenience and without loss of generality, given data concerning the preferred choices, the alternatives of each question are reordered so that $k = 1$ refers to the preferred one.

For example, a profile $x \in R^L$ with three attributes and $L = 5 + 5 + 4 = 14$ is depicted as follows:

$$x = (\underbrace{0\ 1\ 0\ 0\ 0}_{\text{attribute 1}}\quad \underbrace{0\ 0\ 1\ 0\ 0}_{\text{attribute 2}}\quad \underbrace{0\ 1\ 0\ 0}_{\text{attribute 3}})$$

In this example, there are five possible levels of attribute 1, and the second level is chosen for this profile.

For each respondent i and profile x, assume a utility function that is additive in attributes. Let *part-worth* vector $\beta_i \in R^L$ be a column vector of weights such that the respondent's utility function of profile x is $v_i(x) = x\beta_i$. Given preferred directions Δ_{ijk} of question j, the value margin of the preferred profile x_{ij1} with respect to a non-preferred profile x_{ijk} is $v_i(x_{ij1}) - v_i(x_{ijk}) = \Delta_{ijk}\beta_i$. For person i, let ϵ_{ijk} be a random *profile valuation error*. Then profile x_{ij1} is conceived the most preferred profile by person i in question j, if $v_i(x_{ij1}) + \epsilon_{ij1} \geqslant v_i(x_{ijk}) + \epsilon_{ijk}$ for all $k > 1$ or equivalently, if $\Delta_{ijk}\beta_i \geqslant \epsilon_{ijk} - \epsilon_{ij1}$ for all $k > 1$. In the multinomial logit model (McFadden, 1974) we assume that the profile valuation errors ϵ_{ijk} are independent and Gumbel distributed with location parameter zero, scale parameter γ_i. Then the standard deviation of ϵ_{ijk} is given by

$$\sigma_i = \frac{\pi}{\sqrt{6}\gamma_i} \tag{1}$$

and the probability of person i choosing profile x_{ij1} is

$$p_{ij1} = \exp(\gamma_i x_{ij1}\beta_i) / \sum_k \exp(\gamma_i x_{ijk}\beta_i) = 1 / \sum_k \exp(-\gamma_i \Delta_{ijk}\beta_i). \tag{2}$$

2.2. The log-likelihood problem

In this section we formulate the log-likelihood function of *LM* to be used for estimation of the part-worth vectors. We employ a three-level hierarchy: (i) as in *HB*, we assume that the vectors β_i are independent random realizations from a multivariate normal distribution; (ii) the standard deviations of the profile valuation errors are independent draws from an inverse-gamma distribution; and (iii) the likelihood of choices made by respondents follows from (2). Given suitable independence assumptions below, the total log-likelihood is the sum of the log-likelihoods of the realizations of vectors β_i, the standard deviations of profile valuation errors, and individual choices. Next, we discuss the log-likelihood functions for each three levels.

(i) Given that individual data tends to be scarce in the estimation of relatively many part-worth components in β_i, we aim to borrow data from other respondents similarly as in *HB*. We take into account that the part-worth vectors among the respondents are correlated. For example, people generally prefer a low price to high and a high quality to low. In order to account for the interdependence of the respondents' part-worths, we proceed as follows. Let $\tilde{\beta}$ be an L dimensional random vector with multivariate normal distribution $N(\alpha, V)$ where $\alpha \in R^L$ is the expected value of $\tilde{\beta}$ and $V \in R^{L \times L}$ is the covariance matrix with $|V| = 1$.[2] We interpret part-worth vectors β_i, as independent realizations of the random vector $\tilde{\beta}$. The log-likelihood $c_i(\beta_i)$ for a realization β_i (omitting a constant term) is

$$c_i(\alpha, \beta_i, V) = -\frac{1}{2}(\beta_i - \alpha)^T V^{-1}(\beta_i - \alpha) - \frac{1}{2}\log|V|. \tag{3}$$

[2] Normalizations $|V| = \kappa \neq 1$ could be used equally well. This would just lead to rescaling the part-worth vectors as well as the standard deviations of the profile valuation errors.