Designing for Attribute-Level Best-Worst Choice Experiments

Deborah J. Street* and Stephanie Knox,

University of Technology, Sydney, Sydney, Australia deborah.street@uts.edu.au

*Corresponding Author: Deborah J. Street,
email: deborah.street@uts.du.au

Abstract

In a traditional discrete choice experiment, respondents indicate which one of a set of products(services or profiles) they think is best, and they usually do this for a number of sets. Howeverthis does not allow the experimenter to gain any information about the relative contributionto the utility of each of the levels in a profile. To rectify this deficiency, researchers can showrespondents a profile, described by the levels of a number of attributes, and ask each respondentto choose the best and the worst feature (attribute-level) of the profile. Such a task is called anattribute-level best-worst choice task. Using the D-optimality criterion, resolution 3fractional factorial designs perform as well as the complete factorial design in attribute-level best-worst choice experiments, assuming both that all attribute levels are equally attractive and that only main effects of attribute levels are to be used to explain the results. We consider the performance of some small designs for various different prior assumptions about attribute levels.

Key Words: Multinomial logit model, stated preference experiments, profile case best-worst scaling

1   Introduction

Best-worst scaling is a preference elicitation method in which respondents are asked to indicate which of the items currently being shown to them they think is best and which worst. Recently there have been a number of studies in which the items to be compared are levels of a set of attributes that describe a policy or product, that is, a profile in a traditional discrete choice experiment (DCE). These attribute-levels best-worst scaling tasks are thus an alternative way to DCEs to elicit preferences. Potoglou et al. (2011), for example, give an empirical comparison of the two methods and conclude, "preference weights from best-worst scaling and discrete choice experiments do reveal similar patterns in preference and in the majority of cases preference weights - when normalised/rescaled - are not significantly different." The appeal of attribute-level best-worst (BW) (sometimes called profile-based BW or Case 2 BW) is that the task is cognitively less demanding for respondents, particularly if the area of application is unfamiliar.

In this paper we focus on design considerations. If a profile presented to a respondent has one attribute at its most attractive level and the other attributes at lower levels then it is likely that determination of the best is easy and hence choices are made with more consistency than is likely to be observed in a choice set in which the attributes are at intermediate levels. Thus it is of interest to determine a set of profiles (which here are choice sets) in which all choices are of consistent difficulty but which has an information matrix that is the same as

that of the complete factorial, which is D-optimal under the assumption of no diffrences between levels.

In the next section we give the form of the MNL model used, and the form of the information matrix in terms of both the parameters of the fractional factorial design and the unknown parameters for the attribute levels. In the absence of any prior estimates of the attribute-level parameters, these are assumed to be constant when designing a study (Potoglou et al. (2011)) and in that case resolution 3 fractional factorial designs and the complete factorial have the same information matrix. By using a smaller set of profiles it is more likely that a set of profiles of consistent choice difficulty can be found. We then investigate the performance of several designs for three binary attributes, and for four ternary attributes, for a number of different parameters for the attribute levels. We close with some suggestions for future research.

2   The Model

Suppose that k attributes, F_q, $1 \leq q \leq k$, are used to describe each profile and that F_q has ℓ_q levels. Then given any profile $(x_1, x_2, \ldots, x_k) = \mathbf{x}$ the implicit choice set of pairs from which a respondent is making a choice, $C_{\mathbf{x}}$, is given by

$$\{(x_1, x_2), (x_1, x_3), \ldots, (x_1, x_k), (x_2, x_3), \ldots, (x_{k-1}, x_k), (x_2, x_1), \ldots, (x_k, x_{k-1})\}.$$

We let P be the set of profiles that form the choice sets for the attribute-level BW choice task.

Marley et al. (2008) introduced the attribute-level maxdiff model. They defined $BW_{\mathbf{x}}(x_i, x_j)$ to be the probability that jointly (level x_i of factor F_i is chosen as best, level x_j of factor F_j is chosen as worst) from profile $\mathbf{x}$. Then $BW_{\mathbf{x}}$ is called the *best-worst choice probability for profile* $\mathbf{x}$. We define $BW_{\mathbf{x}} \, \forall \, \mathbf{x} \in P$.

$BW_{\mathbf{x}}$ satisfies the *attribute-level maxdiff model* if and only if there exists a positive scale b on the attributes such that for every $\mathbf{x} \in P$ and for any two distinct factors,

$$BW_{\mathbf{x}}(x_i, x_j) = \frac{b(x_i)/b(x_j)}{\sum_{q=1}^{k} \sum_{r=1, r \neq q}^{k} (b(x_q)/b(x_r))}.$$

To incorporate the attribute level information let $b(x_i) = \exp[\beta_{F_i} + \beta_{F_i, x_i}]$. Then the set of best-worst choice probabilities, $BW_{\mathbf{x}}$, satisfies 2-invertibility, 3-reversibility and 4-reversibility. Hence we can use Theorem 15 of Marley et al. (2008) to show that the set of choice probabilities satisfies an attribute-level max-diff model with b a ratio scale. Thus we can estimate the impact $b(x_i)$ of the attribute level x_i and we can calculate the average of the impacts of the levels of factor F_i to determine the average impact of factor F_i. This representation of $b(x_i)$ is similar to that in equation (1) of Flynn et al. (2007), with the exception that they include a constant which is inconsistent with the formulation required by Marley et al. (2008).

As in Street and Knox (2012) we define

$$\pi(F_i x_i, F_j x_j) = \frac{b(x_i)}{b(x_j)} = \exp[\beta_{F_i} + \beta_{F_i, x_i} - (\beta_{F_j} + \beta_{F_j, x_j})].$$

The goal of an attribute-level BW choice experiment is to estimate the $\pi(F_i x_i, F_j x_j)$ and hence the β_{F_q}, $1 \leq q \leq k$ and the β_{F_q, x_q}, $1 \leq q \leq k$, $0 \leq x_q \leq \ell_q - 1$. We will let $\boldsymbol{\pi}$ be a vector containing the distinct $\pi(F_i x_i, F_j x_j)$.

Let $a = \sum_q \ell_q$ be the total number of levels over all of the attributes. Then the total number of pairs of levels, and hence the total number of entries in $\boldsymbol{\pi}$, is $p = \sum_{q=1}^{k} \ell_q (a - \ell_q)$. For the first $p/2$ pairs of levels, we will order the entries of $\boldsymbol{\pi}$ by first ordering the factor pairs in lexicographic order, $F_1 F_2$, $F_1 F_3$ to $F_{k-1} F_k$, for some fixed but arbitrary order of the factors. Within each pair of factors we will order the levels lexicographically. This gives the pairs for π_1 to $\pi_{p/2}$. For $\pi_{i+p/2}$ reverse the order of the levels in π_i. Thus if there are 3 attributes each with two levels, and writing 0_2 for the first level of the second attribute, for example, the first few pairs are $(0_1, 0_2), (0_1, 1_2), (1_1, 0_2), (1_1, 1_2), (0_1, 0_3), (0_1, 1_3)$ and so on.

Suppose that in the attribute-level BW choice experiment each respondent evaluates the N profiles in P. Suppose that $n_{\mathbf{x}}$ is the number of times that profile $\mathbf{x}$ is evaluated by each respondent. (Thus usually $n_{\mathbf{x}}$ is 0 or 1.) Let $\lambda_{\mathbf{x}} = n_{\mathbf{x}}/N$.

Let Λ be a $p \times p$ matrix with rows and columns labelled by pairs of attribute levels, ordered in the same way as the entries in $\boldsymbol{\pi}$. Let $t_1 = (F_i x_i, F_j x_j)$, $t_2 = (F_q x_q, F_s x_s)$, $t_1 \neq t_2$, say. Then define

$$\Lambda_{t_1, t_2} = -\pi_{t_1} \pi_{t_2} \sum_{\{\mathbf{x} \in P | t_1, t_2 \in C_{\mathbf{x}}\}} \frac{\lambda_{\mathbf{x}}}{(\sum_{t \in C_{\mathbf{x}}} \pi_t)^2}$$

and

$$\Lambda_{t_1, t_1} = \sum_{\{t_2 | t_2 \neq t_1\}} \Lambda_{t_1, t_2}.$$

so the row and column sums of Λ are 0.

As defined, Λ is the information matrix for $\boldsymbol{\gamma} = \ln(\boldsymbol{\pi})$ for the MNL model with selection probabilities given by $BW_{\mathbf{x}}$ for the set of profiles P with corresponding choice sets $C_{\mathbf{x}}$ (Bradley (1955); Pendergrass and Bradley (1960); Street and Burgess (2007), Section 3.3). But we have assumed that

$$\gamma(F_i x, F_j y) = \beta_{F_i} + \beta_{F_i x} - (\beta_{F_j} + \beta_{F_j y})$$

in establishing the appropriateness of the attribute-level maxdiff model and so we need to transform Λ to the information matrix for the β_{F_i} and the $\beta_{F_i x}$.

To do this let

$$\boldsymbol{\beta}' = (\beta_{F_1}, \beta_{F_2}, \ldots, \beta_{F_k}, \beta_{F_1 0}, \ldots, \beta_{F_1 \ell_1 - 1}, \beta_{F_2 0}, \ldots, \beta_{F_2 \ell_2 - 1}, \ldots, \beta_{F_k 0}, \ldots, \beta_{F_k \ell_k - 1}),$$

and define X to be the matrix such that $\boldsymbol{\gamma} = X\boldsymbol{\beta}$. Then the information matrix for $\boldsymbol{\beta}$ is given by $X'\Lambda X$. But the rows of X are not linearly independent. We let $C_{\boldsymbol{\beta}_r}$ denote the full-rank information matrix of a reduced $\boldsymbol{\beta}$ with entries

$$\boldsymbol{\beta}'_r = (\beta_{F_1}, \ldots, \beta_{F_{k-1}}, \beta_{F_1 0} - \beta_{F_1 \ell_1 - 1}, \ldots, \beta_{F_1 \ell_1 - 2} - \beta_{F_1 \ell_1 - 1}, \ldots, \beta_{F_k \ell_k - 2} - \beta_{F_k \ell_k - 1}).$$

We let $r_{F_i x}$ be the number of times that level x of attribute F_i appears in the treatment combinations in P. The next result gives the entries in $C_{\boldsymbol{\beta}_r}$ under the assumption that all the entries in $\boldsymbol{\beta}$ are equal. We regard $C_{\boldsymbol{\beta}_r}$ as a block matrix. The first $k-1$ rows and columns of $C_{\boldsymbol{\beta}_r}$ are labelled by the first $k-1$ attributes. The remaining rows and columns are grouped into sets of $\ell_q - 1$ rows and columns and labelled by the corresponding attribute. So rows, and columns, k to $k + \ell_1 - 1$ are labelled F_1, rows, and columns, $k + \ell_1$ to $k + \ell_1 + \ell_2 - 1$ are labelled by F_2 and so on.

■ **THEOREM 2.1**

  Street and Knox (2012) *If all the entries in $\boldsymbol{\beta}$ are equal then the entries in $C_{\boldsymbol{\beta}_r}$ are given by*

$$
C_{\beta,i,j} = \begin{cases} \frac{2}{k}, & i = j, \\ \frac{-2}{k(k-1)}, & i \neq j, \end{cases}
$$

$$
C_{\beta,q,F_h} = \begin{cases} \frac{-2}{Nk(k-1)}\left[\, r_{F_h0} - r_{F_h\ell_h-1}, \quad \cdots, \quad r_{F_h\ell_h-2} - r_{F_h\ell_h-1} \,\right], & h \neq q, \\ \frac{2}{Nk}\left[\, r_{F_q0} - r_{F_q\ell_q-1}, \quad \cdots, \quad r_{F_q\ell_q-2} - r_{F_q\ell_q-1} \,\right], & h = q. \end{cases}
$$

$$
C_{\beta,F_g,F_h} = \begin{cases} (\frac{2}{Nk(k-1)}(-r_{F_gaF_hb} + r_{F_gaF_h\ell_h-1} + r_{F_g\ell_g-1F_hb} - r_{F_g\ell_g-1F_h\ell_h-1}))_{a,b}, & g \neq h \\ \frac{2}{Nk}\begin{bmatrix} r_{F_g0} + r_{F_g\ell_h-1} & r_{F_g\ell_h-1} & \cdots & r_{F_g\ell_h-1} \\ r_{F_g\ell_h-1} & r_{F_g1} + r_{F_g\ell_h-1} & \cdots & r_{F_g\ell_h-1} \\ \vdots & \vdots & \vdots & \vdots \\ r_{F_g\ell_h-1} & r_{F_g\ell_h-1} & \cdots & r_{F_g\ell_h-2} + r_{F_g\ell_h-1} \end{bmatrix}, & g = h \end{cases}
$$

This matrix is the same for a resolution 3 fraction as for the complete factorial.

3   Design Performance and Assumed Priors

In this section we consider the performance of various designs for a number of assumed prior values. For ease we will specify the values of $\ln(b(x_q))$ which we will denote by x_q.

3.1   Three binary attributes

In this case there are 8 profiles, 000, 001, 010, 011, 100, 101, 110, 111 and each profile can be included or not in the BW task. Thus there are $2^8 - 1 = 255$ different designs to compare using the determinant of the information matrix. A graph of the results for $(0_1, 0_2, 0_3, 1_1, 1_2, 1_3)=(1,1,1,1,1,1)$ is given below and confirms the results of Theorem 2.1, that is, the best designs are the the two resolution 3 designs and the complete factorial.
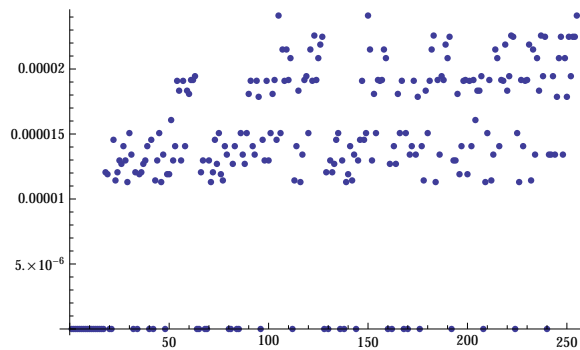

If all the 0_qs are equal and all the 1_qs are equal then the best BW design is the one with the two profiles $(0_1, 0_2, 0_3)$ and $(1_1, 1_2, 1_3)$. This design is 87% efficient in the case of equal parameter values. If the levels within each attribute are equal, but the values for the attributes are different, then for small differences the resolution 3 fractions and the complete factorial are still the best designs to use but as the difference between the attributes increase the complete factorial is the unique best design.

3.2   Four ternary attributes

In this case we consider the performance of the unique 64 designs that come from the three non-isomorphic 18 run 7 attribute orthogonal arrays by omitting columns. These designs can be divided into 34 geometric equivalence classes and 10 combinatorial equivalance classes but we include all of the designs in our discussion as neither of these subdivisions of the designs appears to coincide with their performance for different assumed prior values of $\ln(b(x_q))$.

If all of the levels are assumed to be equally attractive, or even if just all the levels within each attribute are equally attractive, then the 64 designs perform equally well. But this is not necessarily the case for other values of $\ln(b(x_q))$ as illustrated below. Here the levels all have $\ln(b(x_0)) = 2$, $\ln(b(x_1)) = 1$ and $\ln(b(x_2)) = 0.5$.
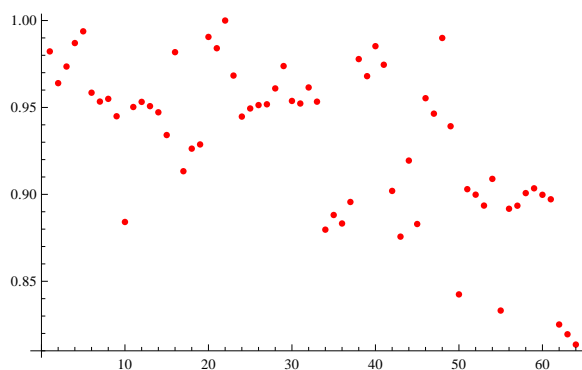
Further work on the performance of orthogonal arrays in the context of attribute-level BW designs is clearly needed.

4   Concluding Remarks

Although Theorem 2.1 shows that if all the $\beta_{F_q,x}$ are equal the equal replication of levels within all attributes or all but one attribute gives sub-matrices with 0 entries, other designs can perform well. For example, Knox et al. (2012) describe a study which uses an orthogonal main effects plan with three 3-level attributes, two 4-level attributes and one 8-level attribute in 32 runs. This OMEP has pairs of attributes in which neither attribute has all attribute levels appearing equally often and yet the simulation study that they undertake show that all assumed (non-zero) parameter values are recovered correctly. Thus OMEPs would appear to be an appropriate choice for attribute-level best-worst choice experiments even if unequal level replication occurs for more than one attribute. Future work might investigate the limitations, if any, of OMEPs in the context of attribute-level BW choice experiments.

We thank Dr Leonie Burgess for making her *Mathematica* code for the calculation of both the Λ matrix and the contrast matrix for DCEs available to us.

References

Bradley, R. A. (1955). Rank analysis of incomplete block designs: III. Some large-sample results on estimation and power for a method of paired comparisons. *Biometrika*, 42:450–470.

Flynn, T. N., Louviere, J. J., Peters, T. J., and Coast, J. (2007). Best-worst scaling: what it can do for health care research and how to do it. *Journal of Health Economics*, 26:171–189.

Knox, S. A., Viney, R. C., Street, D. J., Haas, M. R., Fiebig, D. G., Weisberg, E., and Bateson, D. (2012). What's good and bad about contraceptive products? A best-worst attribute experiment comparing the values of women consumers and GP providers. *Pharmacoeconomics*, 30:1187–202.

Marley, A. A. J., Flynn, T. N., and Louviere, J. J. (2008). Probabilistic models of set-dependent and attribute-level best-worst choice. *Journal of Mathematical Psychology*, 52:281–296.

Pendergrass, R. N. and Bradley, R. A. (1960). Ranking in triple comparisons. In *Contributions to probability and statistics*, pages 331–51. Stanford University Press, Stanford, CA.

Potoglou, D., Burge, P., Flynn, T., Netten, A., Malley, J., Forder, J., and Brazier, J. E. (2011). Best-worst scaling vs. discrete choice experiments: An empirical comparison using social care data. *Social Science and Medicine*, 72:1717–1727.

Street, D. J. and Burgess, L. (2007). *The Construction of Optimal Stated Choice Experiments: Theory and Methods*. Wiley, New York.

Street, D. J. and Knox, S. A. (2012). Designing for attribute-level best-worst choice experiments. *Journal of Statistical Theory and Practice*, 6:363–375.