Models of best-worst choice and ranking among multiattribute options (profiles)

A. A. J. Marley∗,1 and D. Pihlens2

Journal of Mathematical Psychology
(In Press, September 2011)

1A.A.J. Marley
*(corresponding author)
Department of Psychology
University of Victoria
Victoria BC V8W 3P5
Canada
email: ajmarley@uvic.ca

2School of Mathematical Sciences
Faculty of Science
University of Technology Sydney
PO Box 123 Broadway
Sydney, NSW 2007
Australia
email:david.pihlens@uts.edu.au
Abstract

We develop and characterize new representations for the maxdiff model (Marley & Louviere, 2005) for best-worst choice between multiattribute options; in particular, we state conditions under which the scale value of a multiattribute option is a product of independent ratio scales on each attribute. We show that there is a vector of simple “scores” that are sufficient for the model, with each score a closed-form function of the maximum likelihood estimates of the model’s parameters. Several related models are developed for rank orders obtained by repeated best and/or worst choices, and two of these models are tested on stated preferences between mobile phones.

keywords: best-worst choice; maxdiff model; multiattribute options; probabilistic models; profiles
1 Introduction

Louviere and Woodworth (1990) and Finn and Louviere (1992) developed a discrete choice task in which a person is asked to select both the best and the worst option in an available (sub)set of choice alternatives; they also presented and evaluated a probabilistic model of their data. There is growing interest in such best worst scaling (BWS), with three recent empirical applications receiving ‘best paper’ awards (Cohen, 2003; Cohen & Neira, 2004 (published version); Chrzan & Golovashkina, 2006 (published version)). Chrzan and Golovashkina (2006), Jaeger et al. (2008) and Louviere et al. (2008) present detailed material and references on the potential advantages of BWS over more traditional choice methods (such as those that ask a person to choose the best option; choose the worst option; rank the options; or rate the options) and Marley (2010) summarizes the main theoretical work on BWS to date.

Despite increasing use of BWS, until recently the underlying models had not been axiomatized, leaving practitioners without clear guidelines on appropriate experimental designs, data analyses, and interpretation of results. Marley and Louviere (2005) summarized earlier theoretical work and developed an integrative theoretical approach to three overlapping classes of probabilistic models for best, worst, and best-worst choices among ‘things’ (hereafter, Case 1), with the models in each class proposing specific ways in which such choices might be related; these things can be generic objects (such as brands of detergent), social issues (such as concern for the environment) or multiattribute options (such as computers with different levels on various attributes); we call the latter profiles. Marley, Flynn, and Louviere (2008) extended and applied these models to the study of best-worst choice of attributes in a profile (hereafter, Case 2). The present paper, the third in the series, presents additional theoretical results for best-worst choice among profiles (hereafter, Case 3), where it is desired to measure the utility of individual attribute levels, and/or their contribution to the overall utility of a multiattribute option.

The following study, which we analyze in detail in Section 6, illustrates Case 3. The study examined mobile phone feature trade-off among Australian pre-paid mobile phone users. Table 1 shows the attributes and levels studied. The data was collected in two waves, the first in
December 2007, with 465 respondents, the second in February 2008, with 399 respondents. Each participant was presented with 32 choice sets with four options per set, according to a Street and Burgess (2007) main effects design. Figure 1 shows a typical choice set. For each choice set, the participant chose the best option of the available four; then the worst option from the remaining three; and, finally, the best option of the remaining two.

Figure 1: An example choice set from the experiment.

The remainder of the paper is structured as follows. Section 2 presents the basic notation. Section 3 introduces the maxdiff model for Case 3, develops various results for the representation of the scale value of a profile in terms of the scale values on the attributes, and characterizes the resulting best-worst choice model. Section 4 presents the properties of various scores that are based on counts of how often each attribute level is chosen in the design. Section 5 presents two models of rankings derived from repeated best and/or worst choice and Section 6 applies those models to the mobile phone rank data. Section 7 presents our summary and conclusions and Section 8 contains proofs and derivations.

Before continuing, we note that there is an extensive, somewhat related, literature on mathematical and computational models of choice between multiattribute options, with experimental tests on individual and aggregate data (for instance, Busemeyer & Diederich, 2002; and the
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phone Style</td>
<td>Clam or flip phone&lt;br&gt;Candy Bar or straight phone&lt;br&gt;Slider phone&lt;br&gt;Swivel flip&lt;br&gt;Touch Screen phone&lt;br&gt;PDA phone with a HALF QWERTY keyboard&lt;br&gt;PDA phone with a FULL QWERTY keyboard&lt;br&gt;PDA phone with touch screen input</td>
</tr>
<tr>
<td>Brand</td>
<td>A&lt;br&gt;B&lt;br&gt;C&lt;br&gt;D</td>
</tr>
<tr>
<td>Price</td>
<td>$49.00&lt;br&gt;$129.00&lt;br&gt;$199.00&lt;br&gt;$249.00</td>
</tr>
<tr>
<td>Camera</td>
<td>No camera&lt;br&gt;2 megapixel camera&lt;br&gt;3 megapixel camera&lt;br&gt;5 megapixel camera</td>
</tr>
<tr>
<td>Wireless connectivity</td>
<td>No Bluetooth or WiFi connectivity&lt;br&gt;WiFi connectivity&lt;br&gt;Bluetooth connectivity&lt;br&gt;Bluetooth and WiFi connectivity</td>
</tr>
<tr>
<td>Video capability</td>
<td>No video recording&lt;br&gt;Video recording (up to 15 minutes)&lt;br&gt;Video recording (up to 1 hour)&lt;br&gt;Video recording (more than 1 hour)</td>
</tr>
<tr>
<td>Internet capability</td>
<td>Internet Access&lt;br&gt;No Internet Access</td>
</tr>
<tr>
<td>Music capability</td>
<td>No music capability&lt;br&gt;MP3 Music Player only&lt;br&gt;FM Radio only&lt;br&gt;MP3 Music Player and FM Radio</td>
</tr>
<tr>
<td>Handset memory</td>
<td>64 MB built-in memory&lt;br&gt;512 MB built-in memory&lt;br&gt;2 GB built-in memory&lt;br&gt;4 GB built-in memory</td>
</tr>
</tbody>
</table>

Table 1: Attributes and levels used in the experiment.
review by Rieskamp, Busemeyer & Mellers, 2006). We do not discuss that research for two principle reasons: first, that work is for best choice, whereas ours is for best-worst choice; second, it focuses on the detailed analysis of experimental data, including response times and various ‘paradoxes’, whereas our focus is survey stated preference data, with no response times, and designs that are not setup for the study of ‘paradoxes’.

2 Notation

We present notation for Case 1, then extend it for Case 3. Let \( X, |X| = n \geq 2 \), denote a typical available choice set. With \( x, y \in X \), \( BW_X(x, y) \) denotes the probability that \( x \) is chosen as best in \( X \) and \( y \neq x \) is chosen as worst in \( X \). At this point, we do not consider whether \( x \) is chosen as best in \( X \), then \( y \) is chosen as worst in \( X - \{x\} \); or \( y \) is chosen as worst in \( X \), then \( x \) is chosen as worst in \( X - \{y\} \); or \( x \) is chosen as best in \( X \) and \( y \neq x \) is chosen as worst in \( X \) “simultaneously”; we do consider such issues when we present specific models. In any such case, we have

\[
0 \leq BW_X(x, y) \leq 1
\]

and

\[
\sum_{x, y \in X \atop x \neq y} BW_X(x, y) = 1.
\]

We have a set of best-worst choice probabilities (on a master set \( P \)) when we have a set of best-worst choice probabilities on some of the \( X, X \subseteq P \); the set of choice probabilities is complete when we have choice probabilities on all \( |X| \geq 2 \). We assume that no choice probability equals 0 or 1.

The above notation is sufficient for Case 1, we now extend it for Case 3. We assume the following: there are \( m \) attributes, usually with \( m \geq 2 \), and we let \( M = \{1, ..., m\} \). Attribute \( i, i = 1, ..., m, \) has \( q(i) \) levels, and we let \( Q(i) = \{1, ..., q(i)\} \) and \( Q = \prod_{i=1}^{m} Q(i) \). A profile (traditionally called a multiattribute option) is an \( m \)-component vector with each component

\footnote{Best-worst designs usually have choice sets with at least 3 options; for completeness, we allow the possibility of choice sets with 2 options.}
i taking on one of the \(q(i)\) levels for that component. Thus we have a set of size \(\prod_{i=1}^{m} q(i)\) of possible profiles; we denote this set by \(Q\). Let \(D(P), P \subseteq Q\), denote the design, i.e., the set of (sub)sets of profiles that occur in the study; for notational convenience, we let \(D(P)\) include each profile \(r \in P\) that occurs in the design, i.e., in some subset of two or more profiles. We denote a typical profile by

\[ z = (z_1, \ldots, z_m), \quad (1) \]

where \(z_i, i = 1, \ldots, m,\) denotes the level of attribute \(i\) in profile \(z\). We call each \(z_i\) an attribute-level and sometimes use \(p, q,\) etc., to denote generic attribute-levels.

The basic theoretical results on BWS relevant to profiles in Marley & Louviere (2005) and Marley et al. (2008) assumed a complete set of best-worst choice probabilities on the relevant product set \(Q\), i.e., a full factorial design; further work on those results would be worthwhile to weaken those assumptions to more realistic designs. Here, we simply assume that various of the representations developed in those papers hold, and add additional properties that imply additional structure on the representation of the attribute space.

3 The Maxdiff Model of Best–Worst choice among Profiles (Case 3)

We now summarize relevant definitions and results from Marley and Louviere (2005) and Marley et al. (2008) on the maxdiff model of best-worst choice. We use boldface notation \(\mathbf{x}, \mathbf{y} \in X\) to remind the reader that we are interested is multiattribute profiles. However, the results from Marley and Louviere (2005) that we use do not depend on that interpretation, that is, on whether the options have no attribute structure specified; have a single attribute; or have multiple attributes.

We have the following definition, adapted slightly from Marley et al. (2008, Def. 4):

**Definition 1** A set of best-worst choice probabilities for a design \(D(P), P \subseteq Q, |P| \geq 2\), satisfies a multiattribute (MA) maxdiff model iff there exists a positive scale \(b\) on \(P\) such
that for every \( x, y \in X \subseteq D(P) \), \( x \neq y \), \(|X| \geq 2\),

\[
\text{BW}_X(x, y) = \frac{b(x)/b(y)}{\sum_{r, s \in X, r \neq s} b(r)/b(s)} \quad (x \neq y).
\] (2)

Marley et al. (2008, Theorem 8) state necessary and sufficient conditions for a complete set of best-worst choice probabilities, with no probability equal to 0 or 1, to satisfy a maxdiff model with \( b \) a ratio scale. They did not explicitly consider the attribute structure of the options, but such consideration is not required until one studies the relation of the scale \( b \) to scales \( b_i, i = 1, \ldots, m \), on the attributes; this is a major focus of the current work.

3.1 The scale of a profile as a function of scales on the attributes

Now, we study conditions under which the scale value \( b(z) \) for a typical \( z \in P \) can be written as a function of scale values \( b_i(z_i), i = 1, \ldots, m \), for each of its attribute values. The simplest, and classic (main effects), representation is: there exist independent ratio scales \( b_i \) such that

\[
b(z) = \prod_{i=1}^{m} b_i(z_i),
\]

We now define a maxdiff model for choice between profiles that has this property and then discuss mathematical conditions that make such a representation plausible.

**Definition 2** A set of best-worst choice probabilities for a design \( D(P), P \subseteq Q, |P| \geq 2 \), satisfies a **preference independent MA maxdiff model** iff there exists a separate positive scale \( b_i \) on each attribute \( i, i = 1, \ldots, m \), such that for every \( x, y \in X \subseteq D(P) \), \( x \neq y \), \(|X| \geq 2\),

\[
\text{BW}_X(x, y) = \frac{\prod_{i=1}^{m} [b_i(x_i)/b_i(y_i)]}{\sum_{r, s \in X} \prod_{i=1}^{m} [b_i(r_i)/b_i(s_i)]} \quad (x \neq y).
\] (3)

Note that the choice probabilities are unchanged if each scale \( b_i \) is replaced by a scale \( \alpha_i b_i \) with possibly different constants \( \alpha_i > 0 \), suggesting that the \( b_i \) should be assumed to be independent ratio scales. Theorem 6 of Section 3.3 presents a characterization of the model, with such properties. As discussed there, the assumptions are somewhat restrictive – in particular, they
require that no best-worst choice probability is equal to 0 or 1; however, the latter assumption is customary in application of this model to data.

We now present a theoretical motivation for the above representation, using an approach that has been applied to related issues in the measurement literature; the approach is closely related to dimensional analysis in physics and engineering. Specifically, we consider what relations are (theoretically) possible between the ratio scale $b$ and the independent ratio scales $b_i$, $i = 1, ..., m$, on the $m$ attributes. At this point, the latter ratio scales are hypothetical; Section 3.3, Theorem 6, presents conditions for the existence of such scales.

We assume that there is a function $F$ that maps a typical vector of scale values $(b_1(r_1), ..., b_m(r_m))$ on the attributes to the scale value $b(r) = b(r_1, ..., r_m)$ of the MA maxdiff model, Definition 1, i.e.,

$$F(b_1(r_1), ..., b_m(r_m)) = b(r_1, ..., r_m).$$

We also assume that the mapping is invariant under admissible transformations in the following sense: there is a function $G$ on the non-negative real numbers such that for each set of $\alpha_i > 0$, $i = 1, ..., m$,

$$F(\alpha_1 b_1(r_1), ..., \alpha_m b_m(r_m)) = G(\alpha_1, ..., \alpha_m) F(b_1(r_1), ..., b_m(r_m)).$$

The mathematical question then becomes: what are the possible solutions of the above equation for $F$ (and $G$)? Under quite weak regularity conditions\footnote{That certain functions are bounded on an arbitrarily small open $m$-dimensional interval.}, the solution has the form (Aczél, Roberts, & Rosenbaum, 1986): there are constants $\alpha > 0$ and $\beta_i$, $i = 1, ..., m$, such that

$$F(b_1(r_1), ..., b_m(r_m)) = \alpha \prod_{i=1}^m b_i(r_i)^{\beta_i},$$

i.e., with $\tilde{b}_i(r_i) = \alpha^{\frac{1}{\beta_i}} b_i(r_i)^{\beta_i}$

$$b(r_1, ..., r_m) = \prod_{i=1}^m \tilde{b}_i(r_i),$$

which gives a set of scale values as proposed in the preference independent MA maxdiff model, Definition 2, (3). In particular, the constant $\alpha$ and the exponents $\beta_i$ are not identifiable, in the
sense that the choice probabilities in (3) can be rewritten equally well with the scales $\tilde{b}_i$. We explore this fact further in the next section.

3.2 The relation of the above representations to those developed from a random utility model

In the above development of the preference independent MA maxdiff model, Definition 2, we worked with products of ratio scales. However, when the preference independent MA maxdiff model is developed in a random utility framework (see Marley & Louviere, 2005, for the maxdiff model in that framework), it is customary to work with sums (or weighted sums) of scales, and the scales are often (erroneously) considered to be interval scales. We now relate these two approaches, and show how the errors concerning the scales arise.

With the notation as in the previous section, define $u(z_i) = \log b(z_i)$ and $u_i(z_i) = \log b_i(z_i)$, $i = 1, ..., m$. Then we have

$$b(z) = \exp \left( \sum_{i=1}^{m} u_i(z_i) \right),$$

and so the maxdiff representation of (2) becomes

$$BW_X(x, y) = \frac{\exp \left( \sum_{i=1}^{m} [u_i(x_i) - u_i(y_i)] \right)}{\exp \left( \sum_{i=1}^{m} [u_i(r_i) - u_i(s_i)] \right)} \quad (x \neq y),$$

(4)

which is the obvious generalization to best-worst choice of the classic linear in-the-attributes form of the multinomial logit (MNL) for best choice. This representation of the best-worst choice probabilities is usually derived from an underlying random utility model (Marley & Louviere, 2005). Clearly, the choice probabilities are unchanged if each scale $u_i$ is replaced by a scale $u_i + a_i$ with possibly different constants $a_i$, suggesting that the $u_i$ should be assumed to be independent difference scales. However, there is an inclination for authors working from the random utility framework to act as if the scales $u_i$ in (4) are interval scales with a common unit and different origins, i.e., if $u_i$ are scales in (4), then to assume that one can equally well use scales $cu_i + a_i$ with arbitrary $c > 0$ and different constants $a_i$. This is incorrect because the best-worst choice probabilities given by the formula (4) are only invariant when $c = 1$. To the
extent this confusion arises, it is due to the fact that the researcher does not take account of the fact that the derivation of (4) from a random utility model assumes that the underlying Gumbel random variables are in standard form\(^3\). The elimination of this (potential) confusion between the scale properties of the utility values (the \(u_i\) values above) and the scale determined by the unstated variance of the assumed random variables is leading to important new theoretical and empirical results. For example, see Marley et al. (2008) for the set-dependent maxdiff model, which involves such scales; regarding the general issue of the scale parameter in random utility models, see Louviere & Swait (2010), Salisbury & Feinberg (2010a), and Swait & Louviere (1993); and for utility versus scale interpretations of temporal discounting data, see Hutchinson, Zauberman, & Meyer (2010) and Salisbury & Feinberg (2010b).

3.3 Preference independence and a characterization of the preference independent MA maxdiff model.

We now consider qualitative statements equivalent to certain properties of the best-worst models that we have introduced above, and use these properties in characterizing the preference independent MA maxdiff model (Def. 2).

As above, \(z = (z_1, ..., z_m)\) denotes an arbitrary vector of attribute-levels and so \(z_j\) is the level of \(z\) on the \(j\)th attribute. Then, for any attribute-level \(v_g\), let \(z\setminus v_g\) denote the vector that agrees with \(z\) except – possibly – on attribute \(g\), where its value is \(v_g\). The reason for the parenthetic ‘possibly’ is that if \(z_g \neq v_g\) then \(z\setminus v_g\) differs from \(z\), whereas if \(z_g = v_g\) then \(z\setminus v_g\) is identical to \(z\) – and each case occurs.

Remember that \(M = \{1, ..., m\}\) is the set of attributes. Throughout this section, we assume that \(BW_X(x, y) \neq 0, 1\) for any \(\{x, y\} \in X \in D(P), x \neq y\); this is the customary assumption for the various maxdiff models presented in this paper.

**Definition 3** A set of best-worst choice probabilities for a design \(D(P), P \subseteq Q, |P| \geq 2\), satisfies preference independence iff for every \(N \subset M\) and any \(r, s \in X \in D(P), r \neq s\),

\[^3\text{This means that } \Pr(\epsilon \leq t) = e^{-e^{-t}} \quad (-\infty < t < \infty), \text{ which has variance } \pi^2/6.\]**
with $r_i = s_i = a_i, i \in N$,

$$\frac{BW_X(r, s)}{BW_X(s, r)},$$

is unaffected by the attribute-levels $a_i, i \in N$ or by $X - \{r, s\}$.

This definition extends that of independence in the classic deterministic conjoint measurement literature (e.g., Krantz, Luce, Suppes, & Tversky, 1971, p. 301, Def. 11) to the probabilistic case. As far as we know, the deterministic equivalent of this property is the foundation of all axiomatizations of (deterministic) conjoint measurement. Paralleling the deterministic case, it is easily checked that, when no best-worst choice probability equals 0 or 1, preference independence is necessary for a MA maxdiff model, Definition 1, to be representable as a preference independent MA maxdiff model, Definition 2. For the results of this section, it is enough to assume the following condition, which is the special case of preference independence where $|N| = m - 1$, i.e., the available options agree on all dimensions but one.

**Definition 4** A set of best-worst choice probabilities for a design $D(P), P \subseteq Q, |P| \geq 2$, satisfies $(m - 1)$-preference independence iff for any $r, s \in D(P)$, and any attribute-levels $x_i, y_i, x_i \neq y_i, i \in \{1, ..., m\}, m \geq 2$, such that $\{r \setminus x_i, r \setminus y_i\} \in X \in D(P), r \setminus x_i \neq r \setminus y_i$,

$$\frac{BW_X(r \setminus x_i, r \setminus y_i)}{BW_X(r \setminus y_i, r \setminus x_i)}$$

depends only on $x_i$ and $y_i$, i.e., it is unaffected by the attribute-levels $r_j, j \in M - \{i\}$ or by $X - \{r \setminus x_i, r \setminus y_i\}$.

When (4) holds, (6) corresponds to

$$\frac{BW_X(r \setminus x_i, r \setminus y_i)}{BW_X(r \setminus y_i, r \setminus x_i)} = \exp 2[u_i(x_i) - u_i(y_i)],$$

i.e., the ratio of the best-worst choice probabilities in (6) depends only on the attribute-levels of attribute $i$.

The next theorem shows that, when the MA maxdiff model (Def. 1) holds with no choice
probability equal to 0 or 1, and a certain non-necessary condition holds (next), then \((m - 1)\)-preference independence is necessary and sufficient for the representation in (3).

**Definition 5** A design \(D(P)\) is **1-connected** iff for each \(u \in D(P)\) and each \(k_i, i \in M\), that appears in some profile in the design, the profile \(u \setminus k_i \in D(P)\) and the pair of profiles \(\{u, u \setminus k_i\}\) belongs to some set in \(D(P)\); we denote such a set by \(X(u, u \setminus k_i)\).

The assumption that a design is 1-connected is very important in that it allows us to work with choices sets where, effectively, only a single attribute-level is varying and hence we can apply \((m - 1)\)-preference independence (Def. 4). 1-connectedness is a strong assumption\(^4\); however, it is theoretically acceptable as it is satisfied by any design that includes all possible choices sets consisting of two (or, usually for best-worst choice sets, three) distinct profiles from the full set \(Q\) of possible profiles. Similar non-necessary existence (solvability) conditions are required to obtain the classic deterministic conjoint measurement representation (e.g., Krantz, Luce, Suppes, & Tversky, 1971, p. 301, Def. 12). Practical designs for estimating the scale values, when it is assumed that the preference independent maxdiff model (Def. 2) holds, are developed in Vermeulen, Goos, and Vandebroek (2010).

**Theorem 6** Assume that \(D(P), P \subseteq Q, |P| \geq 2\), is 1-connected and that a set of best-worst choice probabilities for \(D(P)\) satisfies a MA maxdiff model, Def. 1, with \(b\) a ratio scale with no best-worst choice probability equal to 0 or 1. Then the set of best-worst choice probabilities satisfies a preference independent maxdiff model, Def. 2, iff it satisfies \((m - 1)\)-preference independence, Def. 4, in which case each \(b_i, i = 1, \ldots, m\), is a separate ratio scale, and for \(r \in P\), \(b(r) = \prod_{i=1}^{m} b_i(r_i)\).

### 4 Properties of Scores for Maxdiff Models

We now consider estimation issues related to the MA maxdiff model, Definition 1, and the preference independent MA maxdiff model, Definition 2. We begin with a brief summary of why the following technical results are of interest.

---

\(^4\)In design theory, an attribute \(i\) with \(q(i)\) levels is connected if there are \(q(i) - 1\) independent estimable contrasts of the levels of the attribute (Eccleston & Russell, 1975).
As stated in the Introduction (Section 1), Louviere and Woodworth (1990) and Finn and Louviere (1992) developed and tested a discrete choice task in which a person is asked to select both the most important (“best”) and least important (“worst”) option in an available (sub)set of options; they also presented a probabilistic model of their data, which concerned food safety. Finn and Louviere also calculated, for each food safety issue, the number of times it was of most concern minus the number of times it was of least concern across the whole design. These difference “scores” had a linear relationship to the estimates of the importance of the food safety issues given by a multinomial logistic regression fitted to the observed most important/least important data. The model in Finn and Louviere has come to be known as the maxdiff model (our focus in this paper); and a linear relation has since been repeatedly found between best minus worst scores and estimates of the parameters of the maxdiff model given by, for instance, maximum likelihood (Bednarz, 2006). The purpose of the present section is to study the theoretical properties of such difference scores for the MA maxdiff model, Definition 1, and for the preference independent MA maxdiff model, Definition 2. To the extent that score differences have useful theoretical properties for such models, they can serve to make best-worst choice modeling more accessible to the practitioner, as demonstrated by their use in studies of national and international comparison of retail wine preferences (Goodman, 2009) and of consumer ethical beliefs across countries (Auger, Devinney & Louviere, 2006).

We use the representations that are normally stated when working from the random utility perspective because most estimation procedures are also developed at that level.

For the MA maxdiff model, Definition 1, we have

\[ BW_X(x, y) = \frac{\exp(u(x) - u(y))}{\sum_{r,s \in X} \exp(u(r) - u(s))} \quad (x \neq y), \]

where \( u \) is a difference scale; and for the preference independent MA maxdiff model, Definition 2, we have

\[ BW_X(x, y) = \frac{\exp(\sum_{i=1}^{m} |u_i(x_i) - u_i(y_i)|)}{\sum_{r,s \in X} \exp(\sum_{i=1}^{m} |u_i(r_i) - u_i(s_i)|)} \quad (x \neq y), \]

where each \( u_i \) is a difference scale with a different origin; when both models hold, the origins
can be chosen such that \( u(x) = \sum_{i=1}^{m} u_i(x_i) \).

Using notation from Marley & Louviere (2005), for choice among things (Case 1), \( \hat{b}(x) - \hat{w}(x) \) denotes the number of times thing \( x \) is chosen as best minus the number of times \( x \) is chosen as worst in the study. Similarly, for choice among profiles (Case 3), \( \hat{b}_i(p) - \hat{w}_i(p) \), \( p \in Q(i) \), \( i \in M \), denotes the number of times attribute-level \( p \) is ‘chosen’ as best minus the number of times \( p \) is ‘chosen’ as worst. Note that the empirical score difference for each thing \( x \) or attribute-level \( p \) is calculated in a similar manner for each case.

The following results, especially the parts a), depend on reasoning similar to that in Marley & Louviere (2005). Those in ii., the most complicated, are derived in Section 8.2; the derivations for i. are similar.

**i. MA maxdiff model** For \( x \in P \),

a) The set of values \( \hat{b}(x) - \hat{w}(x) \) is a sufficient statistic. b) \( \hat{b}(x) - \hat{w}(x) \) equals the sum over all \( X \in D(P) \) of the weighted difference between the (marginal) best and (marginal) worst probability of choosing thing \( x \) in set \( X \), with those probabilities evaluated at the values of the maximum likelihood parameter estimates; the weight for a set \( X \) is the number of times \( X \) occurs in the design \( D(P) \).

**ii. Preference independent MA maxdiff model** For \( p \in Q(i) \), \( i \in M \),

a) The set of values \( \hat{b}_i(p) - \hat{w}_i(p) \) is a sufficient statistic. b) \( \hat{b}_i(p) - \hat{w}_i(p) \) equals the sum over all \( X \in D(P) \) of the weighted difference between the (marginal) best and (marginal) worst probability of ‘choosing’ attribute-level \( p \) in set \( X \), with those probabilities evaluated at the values of the maximum likelihood parameter estimates; the weight for a set \( X \) is the number of times \( X \) occurs in the design \( D(P) \).

Looking at the derivation of ii.b) in Section 8.2, we see that each sufficient statistic (best minus worst score) is a closed-form function of the maximum likelihood estimates of the model’s parameters; a parallel result holds for i.b). Additionally, various empirical results suggest that in each case the best minus worst score differences (vis, score frequencies) are highly linearly

\[ \text{Of course, in this Case 3, attribute-level } p \text{ is (only) 'chosen' as a consequence of its being a component of the profile chosen on a given choice opportunity.} \]
related to the maximum likelihood estimates of the corresponding scale values (Bednarz, 2006; Finn & Louviere, 1992). To date, we have no detailed understanding of why the linear form should hold, other than locally.

5 Models of Ranking by Repeated Best and/or Worst Choice

Thus far, we have discussed models for the choice of the best and the worst option in a set of available options. To the extent that those choices are reliable, it is of interest to consider the use of repeated choices of the best and/or worst option to generate a rank order of the available options. We now summarize models of such processes and in Section 6 apply them to the mobile phone data.

Consider a typical subset, $X$, of available options\(^6\) in the design, $D(P)$. For continuity with the earlier part of the paper, consistency with the notation in Marley and Louviere (2005), and notational convenience, we let $BW_X(x, y)$, $x \neq y$, denote the probability that a person selects $x$ as best and $y$ as worst when the available set is $X$. This notation is not to be interpreted, necessarily, as meaning that $x$ and $y$ are selected ‘simultaneously’, which is one natural interpretation of the maxdiff model (Section 4 of Marley & Louviere, 2005).

For rankings, we use the following notation: $X$ is a typical choice set; $R(X)$ denotes the set of rank orders of $X$; $\rho \in R(X)$ denotes a typical rank order (from best to worst) of $X$; and $p_X(\rho)$ denotes the probability of that rank order occurring. That is, with $|X| = n$, $\rho = \rho_1\rho_2...\rho_{n-1}\rho_n$ is the rank order in which option $\rho_1$ is the most preferred, $\rho_2$ is the second most preferred, ..., $\rho_n$ is the least preferred. In particular, $\rho_1$ is ‘best’ and $\rho_n$ is ‘worst’. For instance, if the rank orders are generated by a sequence of pairs of best and worst choices, we have the following

\(^6\)The ranking notation in this section applies to both Case 1 and Case 3. When we turn to profiles (Case 3), we use boldface for the options.
for $\rho \in R(X), |X| = n,$

$$p_X(\rho) = \begin{cases} BW_X(\rho_1, \rho_n) & \text{if } n = 2, 3, \\
BW_X(\rho_1, \rho_n)BW_{X-\{\rho_1, \rho_n\}}(\rho_2, \rho_{n-1}) & \text{if } n = 4, 5, \\
BW_X(\rho_1, \rho_n)\ldots BW_{(\rho_j, \rho_{j+1})}(\rho_j, \rho_{j+1}) & \text{if } n = 2j, j \geq 3, \\
BW_X(\rho_1, \rho_n)\ldots BW_{(\rho_j, \rho_{j+1}, \rho_{j+2})}(\rho_j, \rho_{j+2}) & \text{if } n = 2j + 1, j \geq 3. 
\end{cases}$$ \hspace{1cm} (8)

The distinction between the cases with an even versus an odd number of elements arises because in the latter case, with $2j + 1$ elements, the rank position of the final element is determined after $j$ best-worst choices. An alternate (recursive) notation is

$$p_X(\rho) = \begin{cases} BW_X(\rho_1, \rho_n) & \text{if } n = 2, 3, \\
BW_X(\rho_1, \rho_n)p_{X-\{\rho_1, \rho_n\}}(\rho_2\ldots \rho_{n-1}) & \text{if } n > 3. 
\end{cases}$$ \hspace{1cm} (9)

The following are two plausible models for Case 3 for the basic choice probabilities that enter into (8), equivalently (9).

The first model is the \textit{multiattribute maxdiff} model (Def. 1).

The second model is expressed in terms of repeated best, then worst, choices; therefore, for every $x \in X \in D(P), |X| \geq 2,$ we let $B_X(x)$ (respectively, $W_X(x)$) denote the probability that $x$ is chosen as best (respectively, worst) in $X.$

\textit{Best-worst multiattribute multinomial logit (MNL) model} There is a ratio scale $b$ on $P$ such that for every $x, y \in X \in D(P), x \neq y, |X| \geq 2,$

$$B_X(x) = \frac{b(x)}{\sum_{r \in X} b(r)}, \quad W_X(y) = \frac{1/b(y)}{\sum_{s \in X} 1/b(s)},$$

and then

$$BW_X(x, y) = B_X(x)W_{X-\{x\}}(y)$$

$$= \frac{b(x)}{\sum_{r \in X} b(r)} \frac{1/b(y)}{\sum_{s \in X-\{x\}} 1/b(s)}.$$ \hspace{1cm} (10)

Lancsar and Louviere (2009) call this the \textit{sequential best-worst MNL} model.
6 An Application

A Discrete Choice Experiment (DCE) was conducted to examine mobile phone feature trade-offs among Australian pre-paid mobile phone users. The study was conducted online, with a sample from the pureprofile consumer panel\footnote{http://www.pureprofile.com/}. The first wave of data was collected in December 2007 and used for model calibration; the second wave was collected in January 2008 and used for out-of-sample testing. The first wave comprised 465 respondents drawn randomly from a population of Australian pre-paid mobile phone users; the second wave comprised 399 respondents drawn randomly from the same population (excluding any respondents contained in Wave 1).

6.1 Experimental design

The experimental design, $D(P)$, was a Street and Burgess (2007) design for testing the main effects of the attributes. There were $m = 9$ attributes, with a $8 \times 4^7 \times 2$ factorial structure (in particular, using the notation of Section 2, $q(1) = 8; q(j) = 4$, $j = 2, \ldots, 6, 8, 9$; $q(7) = 2$). The attributes used to describe mobile phones, in the above order, were phone style, brand, price, camera, wireless connectivity, video capability, internet capability, music capability, and handset memory. A detailed list of these generic attributes and their associated levels is given in Table 1 of Section 1. The design comprised 32 choice sets, each with four profiles. All respondents completed all sets. For each respondent, for each choice set, we observed the choice of ‘best’ from the initial set of 4 profiles, ‘worst’ from the remaining 3 profiles, and ‘best’ from the final two profiles. All profiles in the current choice set remained on-screen throughout the repeated choices. A screen shot of an example choice set is given in Figure 1 of Section 1.

6.2 Models

We fit the best-worst ranking models of Section 5 to the Case 3 rank data via maximum likelihood, with each assuming an additive representation of the utility of each profile, i.e. for each $z \in D(P)$,

$$b(z) = \exp u(z) = \exp \left( \sum_{i=1}^{m} u_i(z_i) \right),$$

(11)
We can assume that each scale \( u \) and \( u_i \), \( i = 1, \ldots, m \), is a difference scale, so we impose the constraints \( \sum_{p \in Q(i)} u_i(p) = 0, \ i = 1, \ldots, m \), to remove the arbitrary origins.

The first model assumes the \textit{preference independent maxdiff model} (2), with (11) at each best-worst stage, with choice probabilities: for \( x, y \in D(P), x \neq y \),

\[
BW_Y (x, y) = \frac{\exp (u(x) - u(y))}{\sum_{r, s \in Y} \exp (u(r) - u(s))} = \frac{\exp (\sum_{i=1}^{m} u_i(x_i) - \sum_{i=1}^{m} u_i(y_i))}{\sum_{r, s \in Y} \exp (\sum_{i=1}^{m} u_i(r_i) - \sum_{i=1}^{m} u_i(s_i))},
\]

where \( u_i(p), p \in Q(i) \) is the utility value of attribute-level \( p \) on attribute \( i \).

The second model assumes the \textit{best-worst multinomial logit (MNL) model}, (10), with (11) at each best-worst stage, with choice probabilities: for \( x, y \in Y \subseteq T, x \neq y \),

\[
BW_Y (x, y) = \frac{\exp u(x)}{\sum_{r \in Y} \exp u(r)} \cdot \frac{\exp -u(y)}{\sum_{s \in Y \setminus \{x\}} \exp -u(s)} = \frac{\exp (\sum_{i=1}^{m} u_i(x_i))}{\sum_{r \in Y} \exp (\sum_{i=1}^{m} u_i(r_i))} \cdot \frac{\exp - (\sum_{i=1}^{m} u_i(y_i))}{\sum_{s \in Y \setminus \{x\}} \exp - (\sum_{i=1}^{m} u_i(s_i))}.
\]

As noted earlier, in each choice set \( X \) in the mobile phone ranking task, we observed a respondent’s choice of ‘best’ from the initial set of 4 profiles, ‘worst’ from the remaining 3 profiles, and ‘best’ from the final two profiles. These three observations can then be used to derive a rank-order of the profiles in the choice set \( X \). With \( \rho = \rho_1 \rho_2 \rho_3 \rho_4 \) denoting a typical rank order obtained in this manner, the choice probability for \( \rho \) is given by \( BW_X (\rho_1, \rho_4) BW_X \setminus \{\rho_1, \rho_4\} (\rho_2, \rho_3) \), where the best-worst choice probabilities either all satisfy (12), giving the \textit{repeated maxdiff model} (with additive utility) or they all satisfy (13) giving the \textit{repeated best-worst MNL model} (with additive utility).

### 6.3 Model estimation results

Tables 2 and 3 provide a summary of the estimation results for each model for the in-sample fit; we summarize the out of sample fits, below, but do not include the corresponding tables. The last level of each attribute \( i \) is omitted from the tables as its estimated value is the negative of
the sum of the other estimated attribute-levels of that attribute; this approach incorporates the constraint that \( \sum_{p \in Q(i)} u_i(p) = 0 \) for each attribute \( i \), enabling an unconstrained maximization of the log-likelihood function.\(^8\)

Both models were successfully fit to the discrete choice data, and provide similar parameter estimates. The sign and direction of the effects are all consistent with logic/theory (e.g., utility is a decreasing function of price), and most are significantly different from zero (at the 99% confidence level).

The repeated maxdiff model (Table 2) provides a slightly better in-sample fit compared to the repeated best-worst MNL model (Table 3), with loglikelihood values of, respectively, \(-35,714.27\) and \(-35,732.98\). Note that each model has the same set of parameters. For these in-sample data, the log-likelihood value under the null hypothesis that the parameter vector \( u = 0 \) for the repeated maxdiff model is given by \( 465 \times 32 \times \ln(1/12 \times 1/2) = -47289.44 \); the corresponding value for the repeated BW MNL model has the same value, given by \( 465 \times 32 \times \ln(1/4 \times 1/3 \times 1/2) = -47289.44 \).

We use these values to calculate McFadden’s \( \rho^2 \) goodness-of-fit statistic (McFadden, 1974), \( 1 - \frac{LL_{\hat{u}}}{LL_{u=0}} \), where \( LL_{\hat{u}} \) and \( LL_{u=0} \) denote the value of the log-likelihood function at, respectively, the maximum likelihood estimates and the null hypothesis values of the parameters, with \( 0 \leq \rho^2 < 1 \). For the repeated maxdiff model, \( \rho^2 = 1 - \frac{-35,714.27}{-47289.44} = 0.245 \), and for the repeated BW MNL model \( \rho^2 = 1 - \frac{-35,732.98}{-47289.44} = 0.244 \). Note that both models assume homogenous preferences across the respondents. While we strongly reject the null hypothesis of random choice, much improved fits are likely to be obtained by considering the heterogeneity of preferences in the population. One approach would be to specify a latent class sequential best-worst model, in a manner similar to Kamakura and Russell (1989), who studied best choices, and Flynn et al. (2010), who studied best-worst choices. A visual representation of the estimated marginal utilities for the levels of four of the attributes is provided in Figure 2.

The out-of-sample results are similar to those of the in-sample data, with a log-likelihood value for the repeated maxdiff model of \(-30,592.18\) (\( \rho^2 = 1 - \frac{-30,592.18}{-40577.39} = 0.246 \)), compared with

\(^8\)In each table, the \( \hat{u}_i \) column provides the estimate of the relevant row attribute-level utility, the \( \text{std.err.} \) column the usual measure of the asymptotic variability of the estimate \( \hat{u}_i \), the \( z \)-value column the estimated value of \( u_i \) divided by the \( \text{std.err.} \), and the \( p \)-value column the probability of the null hypothesis \( \hat{u}_i = 0 \).
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Level</th>
<th>$u_i$</th>
<th>std. err.</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phone style</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clam or flip phone</td>
<td></td>
<td>0.082</td>
<td>0.015</td>
<td>5.487</td>
<td>4.10E-08</td>
</tr>
<tr>
<td>Candy Bar or straight phone</td>
<td></td>
<td>0.028</td>
<td>0.015</td>
<td>1.875</td>
<td>6.10E-02</td>
</tr>
<tr>
<td>Slider phone</td>
<td></td>
<td>0.061</td>
<td>0.015</td>
<td>4.134</td>
<td>3.60E-05</td>
</tr>
<tr>
<td>Swivel flip</td>
<td></td>
<td>-0.004</td>
<td>0.015</td>
<td>-0.248</td>
<td>8.00E-01</td>
</tr>
<tr>
<td>Touch Screen phone</td>
<td></td>
<td>0.004</td>
<td>0.015</td>
<td>0.249</td>
<td>8.00E-01</td>
</tr>
<tr>
<td>PDA phone with a HALF QWERTY keyboard</td>
<td></td>
<td>-0.043</td>
<td>0.015</td>
<td>-2.910</td>
<td>3.60E-03</td>
</tr>
<tr>
<td>PDA phone with a FULL QWERTY keyboard</td>
<td></td>
<td>-0.088</td>
<td>0.015</td>
<td>-5.943</td>
<td>2.80E-09</td>
</tr>
<tr>
<td><strong>Brand</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td>-0.019</td>
<td>0.009</td>
<td>-2.125</td>
<td>3.40E-02</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>0.085</td>
<td>0.009</td>
<td>9.519</td>
<td>1.70E-21</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>-0.023</td>
<td>0.009</td>
<td>-2.530</td>
<td>1.10E-02</td>
</tr>
<tr>
<td><strong>Price</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$49</td>
<td></td>
<td>0.161</td>
<td>0.009</td>
<td>17.932</td>
<td>6.60E-72</td>
</tr>
<tr>
<td>$129</td>
<td></td>
<td>0.036</td>
<td>0.009</td>
<td>4.032</td>
<td>5.50E-05</td>
</tr>
<tr>
<td>$199</td>
<td></td>
<td>-0.001</td>
<td>0.009</td>
<td>-0.072</td>
<td>9.40E-01</td>
</tr>
<tr>
<td><strong>Camera</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No camera</td>
<td></td>
<td>-0.265</td>
<td>0.009</td>
<td>-29.461</td>
<td>9.10E-19</td>
</tr>
<tr>
<td>2 megapixel camera</td>
<td></td>
<td>0.040</td>
<td>0.009</td>
<td>4.514</td>
<td>6.40E-06</td>
</tr>
<tr>
<td>3 megapixel camera</td>
<td></td>
<td>0.079</td>
<td>0.009</td>
<td>8.857</td>
<td>8.20E-19</td>
</tr>
<tr>
<td><strong>Wireless connectivity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Bluetooth or WiFi connectivity</td>
<td></td>
<td>-0.108</td>
<td>0.009</td>
<td>-12.005</td>
<td>3.30E-33</td>
</tr>
<tr>
<td>WiFi connectivity</td>
<td></td>
<td>-0.001</td>
<td>0.009</td>
<td>-0.092</td>
<td>9.30E-01</td>
</tr>
<tr>
<td>Bluetooth connectivity</td>
<td></td>
<td>0.036</td>
<td>0.009</td>
<td>4.048</td>
<td>5.20E-05</td>
</tr>
<tr>
<td><strong>Video capability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No video recording</td>
<td></td>
<td>-0.071</td>
<td>0.009</td>
<td>-7.967</td>
<td>1.80E-15</td>
</tr>
<tr>
<td>Video recording (up to 15 minutes)</td>
<td></td>
<td>0.008</td>
<td>0.009</td>
<td>0.937</td>
<td>3.50E-01</td>
</tr>
<tr>
<td>Video recording (up to 1 hour)</td>
<td></td>
<td>0.031</td>
<td>0.009</td>
<td>3.460</td>
<td>5.40E-04</td>
</tr>
<tr>
<td><strong>Internet capability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internet Access</td>
<td></td>
<td>0.062</td>
<td>0.005</td>
<td>12.042</td>
<td>2.10E-33</td>
</tr>
<tr>
<td><strong>Music capability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No music capability</td>
<td></td>
<td>-0.140</td>
<td>0.009</td>
<td>-15.606</td>
<td>6.70E-55</td>
</tr>
<tr>
<td>MP3 Music Player only</td>
<td></td>
<td>0.083</td>
<td>0.009</td>
<td>9.239</td>
<td>2.50E-20</td>
</tr>
<tr>
<td>FM Radio only</td>
<td></td>
<td>-0.045</td>
<td>0.009</td>
<td>-5.020</td>
<td>5.20E-07</td>
</tr>
<tr>
<td><strong>Handset memory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64 MB built-in memory</td>
<td></td>
<td>-0.073</td>
<td>0.009</td>
<td>-8.175</td>
<td>3.00E-16</td>
</tr>
<tr>
<td>512 MB built-in memory</td>
<td></td>
<td>0.013</td>
<td>0.009</td>
<td>1.453</td>
<td>1.50E-01</td>
</tr>
<tr>
<td>2 GB built-in memory</td>
<td></td>
<td>0.002</td>
<td>0.009</td>
<td>0.248</td>
<td>8.00E-01</td>
</tr>
</tbody>
</table>

Log-likelihood: -35,714.27

Table 2: Parameter estimates for the repeated maxdiff model (in-sample).
## Table 3: Parameter estimates for the repeated BW MNL model (in-sample).

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Level</th>
<th>( \mu_i )</th>
<th>std. err.</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phone style</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clam or flip phone</td>
<td>0.085</td>
<td>0.018</td>
<td>4.850</td>
<td>1.20E-06</td>
</tr>
<tr>
<td></td>
<td>Candy Bar or straight phone</td>
<td>0.034</td>
<td>0.018</td>
<td>1.953</td>
<td>5.10E-02</td>
</tr>
<tr>
<td></td>
<td>Slider phone</td>
<td>0.073</td>
<td>0.017</td>
<td>4.214</td>
<td>2.50E-05</td>
</tr>
<tr>
<td></td>
<td>Swivel flip</td>
<td>0.005</td>
<td>0.017</td>
<td>0.285</td>
<td>7.80E-01</td>
</tr>
<tr>
<td></td>
<td>Touch Screen phone</td>
<td>0.010</td>
<td>0.017</td>
<td>0.586</td>
<td>5.60E-01</td>
</tr>
<tr>
<td></td>
<td>PDA phone with a HALF QWERTY keyboard</td>
<td>-0.044</td>
<td>0.017</td>
<td>-2.516</td>
<td>1.20E-02</td>
</tr>
<tr>
<td></td>
<td>PDA phone with a FULL QWERTY keyboard</td>
<td>-0.103</td>
<td>0.017</td>
<td>-5.928</td>
<td>3.10E-09</td>
</tr>
<tr>
<td><strong>Brand</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>-0.025</td>
<td>0.011</td>
<td>-2.388</td>
<td>1.70E-02</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>0.097</td>
<td>0.011</td>
<td>9.155</td>
<td>5.40E-20</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>-0.020</td>
<td>0.011</td>
<td>-1.939</td>
<td>5.30E-02</td>
</tr>
<tr>
<td><strong>Price</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$49</td>
<td>0.172</td>
<td>0.011</td>
<td>16.151</td>
<td>1.10E-58</td>
</tr>
<tr>
<td></td>
<td>$129</td>
<td>0.054</td>
<td>0.011</td>
<td>5.188</td>
<td>2.10E-07</td>
</tr>
<tr>
<td></td>
<td>$199</td>
<td>0.016</td>
<td>0.010</td>
<td>1.520</td>
<td>1.30E-01</td>
</tr>
<tr>
<td><strong>Camera</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No camera</td>
<td>-0.319</td>
<td>0.011</td>
<td>-30.244</td>
<td>6.30E-201</td>
</tr>
<tr>
<td></td>
<td>2 megapixel camera</td>
<td>0.053</td>
<td>0.011</td>
<td>5.066</td>
<td>4.10E-07</td>
</tr>
<tr>
<td></td>
<td>3 megapixel camera</td>
<td>0.101</td>
<td>0.011</td>
<td>9.636</td>
<td>5.60E-22</td>
</tr>
<tr>
<td><strong>Wireless connectivity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No Bluetooth or WiFi connectivity</td>
<td>-0.127</td>
<td>0.011</td>
<td>-12.088</td>
<td>1.20E-33</td>
</tr>
<tr>
<td></td>
<td>WiFi connectivity</td>
<td>0.007</td>
<td>0.011</td>
<td>0.633</td>
<td>5.30E-01</td>
</tr>
<tr>
<td></td>
<td>Bluetooth connectivity</td>
<td>0.041</td>
<td>0.011</td>
<td>3.940</td>
<td>8.10E-05</td>
</tr>
<tr>
<td><strong>Video capability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No video recording</td>
<td>-0.085</td>
<td>0.011</td>
<td>-8.074</td>
<td>6.80E-16</td>
</tr>
<tr>
<td></td>
<td>Video recording (up to 15 minutes)</td>
<td>0.009</td>
<td>0.011</td>
<td>0.893</td>
<td>3.70E-01</td>
</tr>
<tr>
<td></td>
<td>Video recording (up to 1 hour)</td>
<td>0.037</td>
<td>0.011</td>
<td>3.506</td>
<td>4.60E-04</td>
</tr>
<tr>
<td><strong>Internet capability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Internet Access</td>
<td>0.069</td>
<td>0.006</td>
<td>11.377</td>
<td>5.40E-30</td>
</tr>
<tr>
<td><strong>Music capability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No music capability</td>
<td>-0.169</td>
<td>0.011</td>
<td>-16.065</td>
<td>4.50E-58</td>
</tr>
<tr>
<td></td>
<td>MP3 Music Player only</td>
<td>0.101</td>
<td>0.011</td>
<td>9.648</td>
<td>5.00E-22</td>
</tr>
<tr>
<td></td>
<td>FM Radio only</td>
<td>-0.049</td>
<td>0.011</td>
<td>-4.674</td>
<td>3.00E-06</td>
</tr>
<tr>
<td><strong>Handset memory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>64 MB built-in memory</td>
<td>-0.091</td>
<td>0.011</td>
<td>-8.664</td>
<td>4.60E-18</td>
</tr>
<tr>
<td></td>
<td>512 MB built-in memory</td>
<td>0.018</td>
<td>0.011</td>
<td>1.748</td>
<td>8.00E-02</td>
</tr>
<tr>
<td></td>
<td>2 GB built-in memory</td>
<td>0.006</td>
<td>0.011</td>
<td>0.601</td>
<td>5.50E-01</td>
</tr>
</tbody>
</table>

Log-likelihood: -35,732.98
Figure 2: Plot of the estimated marginal utilities for the phone style, camera, price and video capability attributes for each model.
for the repeated BW MNL model. This demonstrates the slight improvement in fit of the repeated maxdiff model being sustained out-of-sample.

6.4 Discussion of model estimation

We have provided an illustration of fitting best-worst models to choices observed in a DCE, and have compared the goodness-of-fit of the repeated maxdiff model to that of the repeated BW MNL model. The repeated maxdiff model provides a slightly better fit, although we do not suggest that this implies a general result. Both models give approximately the same fit to the data, which may not be surprising given the exponential form of the basic numerator ratios, i.e., \( \frac{b(x)}{b(y)} = \exp[u(x) - u(y)] \), in each.

The models estimated are aggregate models, assuming preference homogeneity in the population of pre-paid mobile phone consumers. While it is very likely that considerable preference heterogeneity exists in consumers’ preferences for mobile phone handsets, the preference homogeneity assumption was made to simplify this illustration of model estimation. In practice, one would want to specify a more general model that allows for a distribution of preferences in the population, and test the homogeneity restriction. For example, random coefficients could be specified, leading to mixed best-worst models, paralleling mixed logit models for best choices (e.g., Train, 2009, Chap. 6).

7 Summary and Conclusion

Marley et al. (2008) present a set of properties that characterize the maxdiff model (Case 1, their Th. 8) with a ratio scale, and a set of properties that characterize the attribute-level maxdiff model for choice in profiles (Case 2, their Th. 15) with a common ratio scale across all attributes. Here, we have developed a parallel characterization of the preference independent maxdiff model for choice among profiles (Case 3, Def. 2), with a separate ratio scale on each attribute.

We tested two models on rank data, namely the repeated maxdiff and the repeated sequential best-worst MNL, each with an additive utility representation of the profiles. Each model gave
reasonable, and slightly different, fits to the data. The similarity of the fits is quite possibly due
to the assumed additive representation (and the use of a main effects only experiment design),
plus the fact that both models fall within an ‘exponential family.’ Also, we did not consider
between-subject variance heterogeneity or within-subject variance heterogeneity over choice sets;
Flynn et al. (2010) apply the sequential best-worst MNL with the former type of heterogeneity
to stated preferences for quality of life, and Scarpa et al. (2011) apply a generalization of the
rank order logit (ROL) model with the latter type of heterogeneity to rankings obtained by
repeated best-worst choice, in a study of alpine visitation.

In conclusion, it is desirable to proceed in a way similar to Stott’s (2006) study of alternative
probabilistic models for risky and uncertain choice, in comparing best-worst models for
multiattribute choice with: representations other than additive for the utility; variance hetero-
genecity; and functional forms other than the ‘exponential family.’ We are currently exploring
such models on further stated preference data.

8 Proofs and Derivations

8.1 Proof of Theorem 6

In the following, the notation $\mathbf{z} \setminus s$ is as introduced in Section 3.3.

Proof of Theorem 6

Proof. It is routine to check that if a set of best-worst choice probabilities for a design
$D(P), P \subseteq Q, |P| \geq 2,$ satisfies a preference independent maxdiff model, Definition 2, then it
satisfies $(m - 1)$-preference independence, Definition 4. So it remains to prove that if a set of
best-worst choice probabilities for a design $D(P), P \subseteq Q, |P| \geq 2$ satisfies the conditions of the
theorem statement, then it satisfies a preference independent maxdiff model, Definition 2.

Let $\ast = (\ast_1, ..., \ast_m)$ be a fixed profile in $D(P)$.

If $m = 1$ there is nothing to prove, so assume $m \geq 2$.

For attribute $i, i = 1, ..., m,$ any attribute-level $v_i$ that appears in the design has to appear
as the $i^{th}$ attribute-level in at least one profile in $D(P);$ consistent with our notation, denote one
such profile by $s = s \setminus v_i \in D(P);$ then Definition 5 with $u = s \setminus v_i, k_i = \ast_i,$ i.e., $u \setminus k_i = s \setminus \ast_i$ gives
that there is a set \( X(s \setminus v_i, s \setminus s_i) \in D(P) \). Now define scale values \( b_i(v_i), i = 1, \ldots, m \), for each such \( v_i \) by:

\[
b_i(v_i) = \begin{cases} 
1 & \text{if } v_i = s_i \\
\left( \frac{BW_X(s_i, s_i)(s \setminus v_i, s \setminus s_i)}{BW_X(s_i, s_i)(s \setminus v_i, s \setminus s_i)} \right)^{ \frac{1}{2} } & \text{if } v_i \neq s_i \end{cases}
\]  

(14)

By \((m - 1)\)-preference independence, Definition 4, \( b_i(v_i) \) depends only on \( v_i \) (and the fixed \( s_i \)).

We now show that, for each \( r \in D(P) \),

\[
b(r) = \prod_{i=1}^{m} b_i(r_i) \quad b(*) = \prod_{i=1}^{m} b_i(*)
\]  

(15)

For any \( r \in D(P) \) and \( i = 0, \ldots, m \), define \( z^{(i)} \) by:

\[
\begin{align*}
z^{(0)} &= r = (r_1, r_2, \ldots, r_{m-1}, r_m) \\
z^{(1)} &= z^{(0)} \setminus s_1 = (*_1, r_2, \ldots, r_{m-1}, r_m) \\
z^{(2)} &= z^{(1)} \setminus s_2 = (*_1, *_2, \ldots, r_{m-1}, r_m) \\
& \quad \vdots \\
z^{(m-1)} &= z^{(m-2)} \setminus s_{m-1} = (*_1, *_2, \ldots, *_{m-1}, r_m) \\
z^{(m)} &= z^{(m-1)} \setminus s_m = (*_1, *_2, \ldots, *_{m-1}, *_m) = *.
\end{align*}
\]

By repeated application of Definition 5, starting with \( u = r \) and \( k_i = *_1 \), we have that each \( z^{(i)} \in D(P) \) and for \( i = 0, \ldots, m-1 \), \( \{z^{(i)}, z^{(i+1)}\} \in X(z^{(i)}, z^{(i+1)}) \in D(P) \). Note that the change from \( z^{(i)} \) to \( z^{(i+1)} \) corresponds to replacing \( r_i \) with \(*_i \). Also, for later, note that for \( i = 0, 1, \ldots, m-1 \),

\[
\begin{align*}
z^{(i+1)} \setminus r_{i+1} &= z^{(i)} \quad \text{and} \quad z^{(i+1)} \setminus *_{i+1} = z^{(i+1)},
\end{align*}
\]

(16)

Since we have not placed a restriction that prohibits \( r_i = *_i \) for some (or even all) \( i = 1, \ldots, m \), it will be possible in the following material to have a term of the form

\[
\frac{BW_X(v \setminus v_i, v \setminus v_i)(v \setminus v_i, v \setminus v_i)}{BW_X(v \setminus v_i, v \setminus v_i)(v \setminus v_i, v \setminus v_i)},
\]
which is undefined in the maxdiff model. However, no inconsistency arises if we bypass this notational issue by using the convention that

\[
BW_X(v_i, v_i) = \frac{BW_X(v_i, v_i)}{BW_X(v_i, v_i)} = 1,
\]

which is equivalent to carefully writing the relevant expressions without such undefined terms.

Now, we are assuming: the preference independent maxdiff model, (3); \((m - 1)\)-preference independence, (6); (14); and (16). Then simple algebra gives: for \(r \in D(P)\),

\[
\frac{b(r)}{b(*)} = \frac{b(z(0))}{b(z(m))} = \prod_{i=0}^{m-1} \frac{b(z(i))}{b(z(i+1))} = \prod_{i=0}^{m-1} \frac{b(z^t(r_{i+1}))}{b(z^t(*_{i+1}))} \frac{1}{2} = \prod_{i=0}^{m-1} b(r_{i+1}) = \prod_{i=1}^{m} b(i) = \prod_{i=1}^{m} b_i(r_i).
\]

However, by (14), \(b_i(*) = 1\) and, since, by assumption, \(b\) is a ratio scale, we can set \(b(*) = 1\), and the above becomes \(b(r) = \prod_{i=1}^{m} b_i(r_i)\).

Finally, we show that each \(b_i\) is a ratio scale. So suppose that \(b'_i\) is another such function for attribute \(i\). We first consider an attribute-level \(v_i\) with \(v_i \neq *_{i}\), and consider the special case \(v_i = *_{i}\) at the end of the argument. Since \(v_i\) appears in the design, we know that there is some \(s = s' \setminus v_i \in D(P)\) and then Definition 5 gives that there is a set \(X(s \setminus v_i, s' \setminus v_i) \in D(P)\). Next, by the assumption that each scale \(b'_i\) (respectively, \(b_i\)) is a representation of the preference independent maxdiff model, (3), we have

\[
\left( \frac{BW_X(s \setminus v_i, s' \setminus *)}{BW_X(s \setminus v_i, s' \setminus *)} \right)^{\frac{1}{2}} = \frac{b'_i(v_i)}{b'_i(*_i)} = \frac{b_i(v_i)}{b_i(*_i)}.
\]

Now let

\[
a' = b'_i(*_i).
\]
Then, remembering that \( b_i(\ast_i) = 1 \) and we have assumed that \( v_i \neq \ast_i \), (18) gives \( b_i'(v_i) = a'b_i(v_i) \) for \( v_i \neq \ast_i \). Also, the fact that \( b_i(\ast_i) = 1 \) and \( a' = b_i'(\ast_i) \) gives \( b_i'(\ast_i) = a'b_i(\ast_i) \). Therefore
\[
b_i' = a'b_i,
\]
i.e., \( b_i' \) is a ratio scale.

**An observation:** Looking at the product form (17), its relations to the ratio scale \( b \), and the constructed ratio scales \( b_i, i = 1, .., m \), it is clear that the following assumption can be used to derive ratio scales with the property that \( b = \prod_{i=1}^{m} b_i \):

For each \( r, s \in D(P) \), there exists a set \( X(r, s) \in D(P) \) and profiles \( w^{(i)} \in D(P), i = 1, ..., m \), such that for each \( i \) there exists a set \( X(w^{(i)} \setminus r_i, w^{(i)} \setminus s_i) \in D(P) \) such that
\[
\frac{BW_{X(r,s)}(r,s)}{BW_{X(r,s)}(s,r)} = \prod_{i=1}^{m} \frac{BW_{X(w^{(i)} \setminus r_i, w^{(i)} \setminus s_i)}(w^{(i)} \setminus r_i, w^{(i)} \setminus s_i)}{BW_{X(w^{(i)} \setminus r_i, w^{(i)} \setminus s_i)}(w^{(i)} \setminus s_i, w^{(i)} \setminus r_i)}. \tag{19}
\]

Conditional on the existence of the relevant choice sets, (19) is a necessary condition for the representation of the preference independent maxdiff model, (3). However, the proof we first provided requires only the assumption of a design property, namely 1-connected, whereas a proof based on the current reasoning requires the stated design properties plus (19).

### 8.2 Likelihood and score properties for the preference independent MA maxdiff model

For the samples \( n = 1, ..., N(X) \) and any \( x, y \in X \in D(P), \ x \neq y \), let
\[
\hat{bw}_{n,X}(x, y) = \begin{cases} 1 & \text{if } x \text{ is best, } y \text{ is worst for sample } n \text{ of } X \\ 0 & \text{otherwise} \end{cases}
\]
Then, for attribute-levels \( p \in Q(i) \), \( q \in Q(j) \), \( i, j \in M \), let\(^{9}\)

\[
\hat{b}_i(p) = \sum_{X \in D(P)} \sum_{x \in X} \sum_{y \in X - \{x\}} \sum_{n=1}^{N(X)} \hat{b}_{n,x}(x, y),
\]

\[\hat{w}_j(q) = \sum_{X \in D(P)} \sum_{x \in X} \sum_{y \in X} \sum_{n=1}^{N(X)} \hat{b}_{n,x}(x, y).
\]

Thus, \( \hat{b}_i(p) \) (respectively, \( \hat{w}_j(q) \)) is the number of times that attribute-level \( p \) (respectively, \( q \)) is ‘chosen’ as best (respectively, as worst). We have put ‘chosen’ as here the only way the attribute-level \( p \) can be ‘chosen’ is as a result of choosing a presented profile \( z \) for which \( z_i = p \).

For any attribute-level \( p \), \( p \in Q(i) \), \( i \in M \), let

\[
N_X(p) = \begin{cases} 
  N(X) & \text{if } x_i = p \text{ for some } x \in X \\
  0 & \text{otherwise}
\end{cases}
\]

with parallel notation for attribute-level \( q \). Also, for attribute-levels \( p, q \), with \( p \in Q(i), q \in Q(j) \), \( i, j \in M \), let

\[
B_X^{(M)}(p) = \sum_{(x \in X) \land (x_i = p)} \sum_{y \in X - \{x\}} BW_X(x, y),
\]

and

\[
W_X^{(M)}(q) = \sum_{(y \in X) \land (y_j = q)} \sum_{x \in X - \{y\}} BW_X(x, y).
\]

Thus, \( B_X^{(M)}(p) \) (respectively, \( W_X^{(M)}(q) \)) is the probability, across the design, of ‘choosing’ the attribute-level \( p \) as best (respectively, \( q \) as worst). We have put ‘choosing’ as \( B_X^{(M)}(p) \) and \( W_X^{(M)}(q) \) are not choice probabilities in the sense used in Case 1 as here the only way the attribute-level \( p \) can be ‘chosen’ is as a result of choosing a presented profile \( z \) for which \( z_i = p \). Nonetheless, as we now show, \( B_X^{(M)}(p) \) and \( W_X^{(M)}(q) \) act in ways similar to those earlier choice probabilities.

The likelihood of obtaining the data for the preference independent MA maxdiff model,

\(^{9}\)The subscript \( i \) (respectively, \( j \)) on the “best score” (respectively, “worst score”) is not technically necessary. However, we include it so that we can easily distinguish between Cases 2 and 3.
Definition 2, is

\[
\prod_{X \in D(P)} \prod_{n=1}^{N(X)} \prod_{x,y \in X, x \neq y} [BW_X(x, y)]^\widehat{bw}_{n,X}(x, y),
\]

\[(24)\]

\[
= \prod_{X \in D(P)} \prod_{x,y \in X, x \neq y}^{N(X)} \prod_{n=1} \left( \frac{\exp \left( \sum_{i=1}^m |u_i(x_i) - u_i(y_i)| \right)}{\sum_{r,s \in X \ r \neq s} \exp \left( \sum_{i=1}^m |u_i(r_i) - u_i(s_i)| \right)} \right) \widehat{bw}_{n,X}(x, y)
\]

\[
= \left( \prod_{X \in D(P)} \frac{1}{\sum_{r,s \in X \ r \neq s} \exp \left( \sum_{i=1}^m |u_i(r_i) - u_i(s_i)| \right)^{N(X)}} \right)
\]

\[
\times \prod_{X \in D(P)} \prod_{x,y \in X, x \neq y} \prod_{i=1}^m \exp \left( \frac{|u_i(x_i) - u_i(y_i)| \sum_{n=1}^{N(X)} \widehat{bw}_{n,X}(x, y)}{\sum_{n=1}^{N(X)} \widehat{bw}_{n,X}(x, y)} \right).\]

We now collect the terms in the product expression on the final line. Using the properties of the exponential function, we do this by replacing products with sums. Also, for notational
At the maximum of the log likelihood function, the derivative of the log likelihood with respect to
$\hat{b}_i$ and so the set of score differences $\sum_{r,s \in X} \exp \left( \sum_{i=1}^{m} [u_i(r_i) - u_i(s_i)] \right) N(X)^{-1}$

\[ \sum_{i=1}^{m} \sum_{X \in D(P)} \sum_{x,y \in X} [u_i(x_i) - u_i(y_i)] \sum_{n=1}^{N(X)} \hat{b}_{n,X} \{x,y\} \]

\[ \sum_{i=1}^{m} \sum_{X \in D(P)} \sum_{x,y \in X} u_i(x_i) \sum_{n=1}^{N(X)} \hat{b}_{n,X} \{x,y\} - \sum_{i=1}^{m} \sum_{X \in D(P)} \sum_{x,y \in X} u_i(y_i) \sum_{n=1}^{N(X)} \hat{b}_{n,X} \{x,y\} \]

\[ \sum_{i=1}^{m} \sum_{p \in Q(i)} u_i(p) \sum_{X \in D(P)} \sum_{x \in X\{x}\wedge(x_i=p)} \sum_{y \in X\{y\}} \sum_{n=1}^{N(X)} \hat{b}_{n,X} \{x,y\} \]

\[ \sum_{i=1}^{m} \sum_{p \in Q(i)} u_i(p) \sum_{X \in D(P)} \sum_{y \in X\{y\} \wedge(y_i=p)} \sum_{x \in X\{x\}} \sum_{n=1}^{N(X)} \hat{b}_{n,X} \{x,y\} \]

\[ \sum_{i=1}^{m} \sum_{p \in Q(i)} u_i(p) [\hat{b}_i(p) - \hat{w}_i(p)], \]

where the final equality follows from (20) and (21). Thus the likelihood becomes

\[ \left( \prod_{X \in D(P)} \left[ \frac{1}{\sum_{r,s \in X} \exp \left( \sum_{i=1}^{m} [u_i(r_i) - u_i(s_i)] \right) N(X)^{-1}} \right] \right)^{N(X)} \exp \left( \sum_{i=1}^{m} \sum_{p \in Q(i)} u_i(p) [\hat{b}_i(p) - \hat{w}_i(p)] \right), \]

and so the set of score differences $[\hat{b}_i(p) - \hat{w}_i(p)], p \in Q(i), i = 1, ..., m,$ is a sufficient statistic.

Also, the likelihood in log form becomes

\[ \sum_{i=1}^{m} \sum_{p \in Q(i)} u_i(p) [\hat{b}_i(p) - \hat{w}_j(p)] - \sum_{X \in D(P)} N(X) \log \sum_{r,s \in X} \exp \left( \sum_{i=1}^{m} [u_i(r_i) - u_i(s_i)] \right). \]

At the maximum of the log likelihood function, the derivative of the log likelihood with respect
to each $u_i(p)$ is zero, and a routine calculation of this derivative gives:

$$\left[ \hat{b}_i(p) - \hat{w}_i(p) \right] = \sum_{X \in D(p)} N(X)[\hat{B}_{X}^{(M)}(p) - \hat{W}_{X}^{(M)}(p)],$$

where $\hat{B}_{X}^{(M)}(p)$ (respectively, $\hat{W}_{X}^{(M)}(p)$) are the values of the (choice) probabilities (22) (respectively, 23) evaluated at the maximum likelihood values of the scale values $\hat{u}_i(p), p \in Q(i), i = 1,...,m$. In this sense, the maximum likelihood parameter estimates induce the model to reproduce the sample best minus worst scores.

**Acknowledgements**

This research has been supported by Natural Science and Engineering Research Council Discovery Grant 8124-98 to the University of Victoria for Marley. The work was carried out, in part, whilst Marley was a Distinguished Research Professor (part-time), and Pihlens was a Senior Researcher, at the Centre for the Study of Choice, University of Technology, Sydney. We thank Frances Terlich for her extensive work on estimation of the models, and Terry Flynn for helpful discussions on the use of data summaries (“scores”).

**References**


University of Technology, Sydney.


34


