Conceptual Relations Between Expanded Rank Data and Models of the Unexpanded Rank Data

A. A. J. Marley*1,2

June 14, 2011

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

2 Centre for the Study of Choice (CenSoC)  
Faculty of Business  
University of Technology Sydney  
PO Box 123 Broadway  
Sydney, NSW 2007  
Australia
Abstract

Louviere et al. (2008. *J. of Choice Modelling, 1, 126-163*) present two main empirical examples in which a respondent rank orders the options in various choice sets by repeated best, then worst, choice. They expand the ranking data to various “implied” choices in subsets and fit the expanded data in various ways; they do not present models of the original rank data, except in one case (that of the rank ordered logit). We build on their work by constructing models of the original rank data that are consistent with the “weights” implied by the data expansions. This results in two classes of models: the first includes the reversible ranking model and has useful “score” properties; the second includes the rank ordered logit model and has natural “process” interpretations. Finally, we summarize known and new results on relations between the two classes of models.
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. Recent research related to this topic by Louviere and his collaborators has begun to make a clear distinction between best-worst choice as a method, called Best Worst Scaling (BWS), and models that are applied to data obtained using BWS. For instance, Scarpa et al. (2011) obtained rank orders of options by asking respondents to make a sequence of best, then worst, choices - each choice set had 5 alternatives, and a rank order was obtained by a respondent making the following sequence of choices: the most preferred alternative out of 5; the least preferred of the remaining 4; the most preferred of the remaining 3; the least preferred of the final 2. Scarpa et al. fit the data by a model based on repeated best choices (namely, an extension of the rank-ordered logit) that includes three variance terms, each of which depends on one aspect of the design: the placement in the survey of the current choice set; the size of the current choice set; and the order in which the subsets of the current choice set are considered in producing the rank order. The important thing to note is that the choices were obtained by BWS, i.e., a sequence of best, then worst, choices, yet the model was formulated and tested as a sequence of 4 best choices. The model did have variance terms to account for the difference between the order in which the options were selected and the order in which they appear in the model. It remains for future research to determine whether this approach is reasonable, or whether it is preferable to apply models that include representations of best and worst choices and the order in which they are applied - as in Lancsar and Louviere (2009) and Collins and Rose (2011).

A related issue, if one treats BWS as a data collection method, rather than a model, is: what model should be fit to the resulting (ranking) data? This question is explored in Louviere et al. (2008), with emphasis on fitting models to the best (first) choices, only, of individuals, even when partial or full rank order data were collected by, say, repeated best, then worst. Specifically, the two main empirical examples in that paper had respondents rank order sets (of size 3, 4, or 5) by repeated best, then worst, choice. These rank data were then expanded to “implied” best choices on various subsets, and the expanded data used, in various ways, to fit the original best choice data with a multinomial logit (MNL) model; except in one case (that of the rank ordered logit), rank order models were neither presented, or fit, by Louviere et al. We build on their work by constructing models of the unexpanded rank data that are consistent with the “weights” proposed by the Louviere et al. expansions; those rank models belong to a general class, which we call weighted utility ranking (WUR) models, the general properties of which can be stated independently of the Louviere et al. expansions.

Full details on the use of expansions in fitting best choice data are given in Louviere et al. (2008); Islam, Louviere and Pihlens (2009); and Ebbling, Frischnecht and Louviere (2010). In this paper, we only provide sufficient detail
of those methods to illustrate particular weighted utility ranking models.

The remainder of the paper is as follows: Section 1 illustrates the data expansion methods in Louviere et al. (2008); Section 2 introduces the notation; Section 3 introduces the class of weighted utility ranking models, some of which, as shown in Section 4, can be partially motivated by the Louviere et al. expansions. Section 5 introduces ranking models based on repeated best and/or worst choice; the rank ordered logit is the best known model in this class; this section also presents known and new results relating weighted utility and repeated best and/or worst ranking models. Finally, Section 6 presents a discussion and conclusion. Appendices A and B include some technical material.

1 Data Expansion Methods

The following example, taken from Louviere et al. (2008), illustrates their methods and results; all such methods and results generalize to larger examples. After introducing this material, we develop related probabilistic ranking models.

Consider drink combinations described by type of drink (coffee, tea, coke, water), price ($1.00, $1.50, $2.00, $2.50) and container (bottle, can). Construct 16 choice options using an orthogonal main effects plan (OMEP), where the 1st (respectively, 2nd) attribute is type of drink (respectively, price), each with 4 levels 0, 1, 2, 3, and the 3rd is container with 2 levels 0, 1 (Table 1).

<table>
<thead>
<tr>
<th>Option</th>
<th>1st Level</th>
<th>2nd Level</th>
<th>3rd Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

We assume that a balanced incomplete block design (BIBD) is used to put the 16 drink combinations of Table 1 into the 20 choice sets of Table 2, with
each choice set of size 4; each drink combination occurs in 5 choice sets; and
each pair of drink combinations occurs in one choice set.

Table 2. BIBD of 20 choice sets of size 4.

<table>
<thead>
<tr>
<th>Choice Set:</th>
<th>Options in choice set:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 5 8 14</td>
</tr>
<tr>
<td>2</td>
<td>1 5 6 7</td>
</tr>
<tr>
<td>3</td>
<td>5 9 12 16</td>
</tr>
<tr>
<td>4</td>
<td>4 5 11 15</td>
</tr>
<tr>
<td>5</td>
<td>3 5 10 13</td>
</tr>
<tr>
<td>6</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>7</td>
<td>2 6 9 11</td>
</tr>
<tr>
<td>8</td>
<td>2 7 13 16</td>
</tr>
<tr>
<td>9</td>
<td>2 10 12 15</td>
</tr>
<tr>
<td>10</td>
<td>1 8 9 10</td>
</tr>
<tr>
<td>11</td>
<td>6 8 13 15</td>
</tr>
<tr>
<td>12</td>
<td>4 7 8 12</td>
</tr>
<tr>
<td>13</td>
<td>3 8 11 16</td>
</tr>
<tr>
<td>14</td>
<td>1 14 15 16</td>
</tr>
<tr>
<td>15</td>
<td>3 6 12 14</td>
</tr>
<tr>
<td>16</td>
<td>7 10 11 14</td>
</tr>
<tr>
<td>17</td>
<td>4 9 13 14</td>
</tr>
<tr>
<td>18</td>
<td>1 11 12 13</td>
</tr>
<tr>
<td>19</td>
<td>4 6 10 16</td>
</tr>
<tr>
<td>20</td>
<td>3 7 9 15</td>
</tr>
</tbody>
</table>

Suppose a participant has, separately, rank ordered the four options in each of the 20 choice sets of Table 2 from best to worst, and “...consider what happens if we “pretend” that we have more data than only the ranks, but do this in a very systematic and structured way.” (Louviere et al., 2008, p. 138). We now rephrase the details of this “pretence” (given on p. 138-9 of Louviere et al., 2008) and then, in Section 4, use the results as a motivation for probabilistic models of the unexpanded rank data. There, we also describe the similarities and differences between our approach (ranking models applied to rank data) and Louviere et al.’s approach (MNL models for best choices in rank data, with that rank data expanded to “implied” best choices on subsets).

Let $X = \{w, x, y, z\}$ denote a typical one of the 20 choice sets in Table 2, and assume that the participant has given the rank order $w \succ x \succ y \succ z$, where $\succ$ means “preferred to” - i.e., the ranking is from best (first) to worst (last). Now consider all two- and three-element subsets of $X$ - Table 3 shows them, where, for each column, 1 (respectively, 0) means that the option at the head of that column is (is not) in the choice set in the current row. Following Louviere et al., we now “pretend” that because we know the preference order

\footnote{For the remainder of the paper, best (respectively, worst) refers to the first (respectively, last) option in such a (partial or full) rank order.}
$w \succ x \succ y \succ z$, we can apply that order to predict the best choice for each 2- and 3-element subset of $X = \{w, x, y, z\}$. For example, choice set 10 in Table 3 is the 3-element subset $\{x, y, z\}$ and the “implied” best choice is $x$ as it precedes both $y$ and $z$ in the rank order $w \succ x \succ y \succ z$. The final row of Table 3, Total, gives the “implied” number of best choices of each element of $X$ for the given (data) rank order.

Comment: Louviere et al. include subsets with one element (singletons) in their expansion of Table 3, as they do in all their expansions; this gives Totals of 8, 4, 2, 1 in Table 3. The ranking probabilities of each of the ranking models that we motivate by such expansions are the same, with or without the inclusion of such singleton sets; so we do not mention this point again.

Table 3. “Implied” best choices for each 2- and 3-element subset of $\{w, x, y, z\}$ given the rank order $w \succ x \succ y \succ z$.

<table>
<thead>
<tr>
<th>Rank Position</th>
<th>Preference Order:</th>
<th>Choice Set:</th>
<th>Implied Choice:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1$^{st}$</td>
<td>2$^{nd}$</td>
<td>3$^{rd}$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

A parallel expansion is applied to the participant’s data rank order for each of the 20 choice sets in Table 2. Louviere et al. then use the (natural log of the) Total (score, weight) for each option in each of the 20 choice sets as the dependent variable in fitting an MNL model to the best (first) choices in the rank order data for those sets. They use various estimation methods, such as weighted least squares (WLS), weighted conditional logit (WCL), and maximum likelihood (ML), and obtain similar estimates (up to a scale factor) with each method. The most recent related work (Ebling et al., 2010) obtains excellent in- and out-of-sample fits of an MNL model to the best choices in partial rank data using a WLS method based on an expansion of that partial rank data - see Section 4.2 for related expansions.

As noted above, our main purpose is to further explore the use of such “expanded” best choices, and the resultant “implied” choice totals, in the motivation of probabilistic models of the unexpanded rank data. The details of the arguments appear in Section 4, with a brief summary here. Specifically, using
Table 3 as illustration, we agree with Louviere et al. that, when choices are deterministic, the rank order \( w \succ x \succ y \succ z \), where \( \succ \) means “preferred to”, is equivalent to (contains the same information as) the (implied) best choice for each of the choice sets in Table 3. However, we reason further that, when choices are probabilistic, the assumption of this equivalence (or other related ones discussed below) can be interpreted as specifying a particular (deterministic) functional relation between the probability of a ranking of a set and the best choice probabilities on subsets of that set; the details of such functional relations are given in Section 4. Said another way, in the probabilistic case, the (sample) rank order of a set and the (sample) best choices in the non-empty subsets of that set are only “equivalent” for a model satisfying a particular functional relation between the probabilities of ranking and of best choice. Importantly, that functional relation is not, in general, compatible with the multinomial logit (MNL) model holding exactly for the best choices in the rank orders.

Thus, our perspective is rather different than that of Louviere et al. (2008) and Ebling et al. (2010). Generally, their work involves using the expanded data in fitting the best choices, only, in the rank data, with those best choices assumed to satisfy an MNL model; the exception is their fitting the rank data by the rank ordered logit - see Section 5 for that model. Hence, it seems that they are implicitly assuming that the expanded data (choice totals) provides estimates of best choices that are improved relative to those obtained using the unexpanded rank data. We, on the other hand, focus on modeling the original rank data, but with ranking models motivated, to some extent, by the Louviere et al. expansions (choice totals). In fact, the likelihoods given by the two approaches are such that they yield the same sufficient statistics (Section 4.1.1).

The remainder of the paper is as follows: Section 2 introduces the notation; Section 3 introduces the class of weighted utility ranking models, some of which, as shown in Section 4, can be partially motivated by the Louviere et al. expansions. Section 5 introduces ranking models based on repeated best and/or worst choices; the rank ordered logit is the best known model in this class; this section also presents preliminary known and new results on relations between the two model classes. Finally, Section 6 presents a discussion and conclusion. Appendices A and B include some technical material.

## 2 Notation

### 2.1 General notation

Let \( T \) with \( |T| \geq 2 \) denote the finite set of potentially available choice options, and let \( D(T) \) denote the design, i.e., the set of (sub)sets of choice alternatives that occur in the study. For any subset \( X \subseteq T \), with \( |X| \geq 2 \), \( B_X(x) \) denotes the probability that the alternative \( x \) is chosen as best\(^2\) in \( X \), \( W_X(y) \) the probability

---

\(^2\)Most previous work using similar mathematical notation has used \( P_X(x) \) or \( P(x|X) \) where we use \( B_X(x) \). We use the latter for best, and \( W_X(x) \) for worst, to distinguish clearly between
that the alternative \( y \) is chosen as worst in \( X \), and \( BW_X(x, y) \) the probability that the alternative \( x \) is chosen as best in \( X \) and the alternative \( y \neq x \) 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 \( X \); or \( y \) is chosen as worst in \( X \), then \( x \) is chosen as best in \( X \); 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 of repeated best and/or worst choice. Thus

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

and

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

Much of the following material uses the multinomial logit (MNL) model for best choices: there is a difference scale \( u \) such that for all \( y \in Y \subseteq T \),

\[
B_Y(y) = \frac{e^{u(y)}}{\sum_{z \in Y} e^{u(z)}}.
\]

The representation restricted to \( Y \subseteq T, |Y| = 2 \), is the binary MNL model. The natural parallel for worst choices is the multinomial logit (MNL) model for worst choices: there is a difference scale \( v \) such that for all \( y \in Y \subseteq T \),

\[
W_Y(y) = \frac{e^{v(y)}}{\sum_{z \in Y} e^{v(z)}}.
\]

Marley and Louviere (2005) present a theoretical argument for the case where \( v = -u \), i.e., we have

\[
W_Y(y) = \frac{e^{-u(y)}}{\sum_{z \in Y} e^{-u(z)}}.
\]

Thus, the worst scale value is the negative of the best scale value; this is not a necessary restriction. One interpretation of this constraint is that, for the MNL model, it yields the result that the probability that \( y \in Y \) is selected as best in a set \( Y \) with scale values \( u(z) \), \( z \in Y \), is equal to the probability that \( y \in Y \) is selected as worst in a set \( Y \) with scale values \(-u(z) \), \( z \in Y \). In particular, when (1) and (2) both hold, we have that for all \( x, y \in X, x \neq y \), \( B_{(x, y)}(x) = W_{(x, y)}(y) \), and write the common value as \( p(x, y) \); empirically, this relation may not always hold (Shafr, 1993). Collins and Rose (2011) present good fits of generalizations of (1) and (2) to ranking data obtained by repeated best, then worst, choices.

Next we need the idea of a ranking of a set \( X \) from its best (most preferred) to its worst (least preferred) element, and similarly of a ranking of \( X \) from its worst (least preferred) to its best (most preferred) element. We refer to such Best and Worst choice probabilities.
these as best to worst, and worst to best, rankings, respectively. Such rankings may be empirical, i.e., resulting from a person’s judgments, but of at least equal importance in this paper is the theoretical role such rankings play in the development of probabilistic choice models, including those for best (worst) choices. For a set $X$, $X \subseteq T$, with $|X| = n$, $\rho = \rho_1\rho_2\ldots\rho_{n-1}\rho_n$ denotes a typical best to worst rank order of $X$, and $R(X)$ denotes the set of rank orders of $X$; sometimes, we write such a rank order in the form $\rho_1 \succ \rho_2 \succ \ldots \succ \rho_n$, where $\succ$ means “is preferred to”. $B_{R(X)}(\rho)$ denotes the probability that $\rho$ occurs as a best to worst rank order, and $W_{R(X)}(\rho)$ denotes the probability that $\rho$ occurs as a worst to best rank order - thus, in the former case $\rho_1$ is the best element in the rank order, in the latter case it is the worst element. The assumption that $B_{R(X)}(\rho)$ and $W_{R(X)}(\rho)$ are probabilities and that a ranking occurs at each choice opportunity is summarized by: for each $\rho \in R(X)$,

$$0 \leq B_{R(X)}(\rho), W_{R(X)}(\rho) \leq 1$$

and

$$\sum_{\rho \in R(X)} B_{R(X)}(\rho) = 1 = \sum_{\rho \in R(X)} W_{R(X)}(\rho).$$

The reverse of a rank order $\rho = \rho_1\rho_2\ldots\rho_{n-1}\rho_n$ is the rank order $\rho^* = \rho_n\rho_{n-1}\ldots\rho_2\rho_1$. There is no necessary theoretical or empirical relation between $B_{R(X)}(\rho)$ and $W_{R(X)}(\rho^*)$, though it is often of interest to consider the constraint $B_{R(X)}(\rho) = W_{R(X)}(\rho^*)$.

The above notation does not explicitly indicate whether a ranking is obtained by, say, a sequence of best and/or worst choices; repeated best choices; or repeated worst choices. We introduce such notation in Section 5; otherwise we write $B_{R(X)}(\rho)$ for the probability of the rank order $\rho$, where $\rho_1$ is best, $\rho_2$ second best, etc., without reference to how it was obtained.

The above general notation has to be extended in various ways to handle the three main cases of BWS, defined in Louviere, Flynn, and Marley (2011): Case 1: (repeated) best and worst choice among ‘things’ - these things are usually generic objects (such as brands of detergent) or social issues (such as concern for the environment); Case 2: (repeated) best and worst choice of attributes in a profile of attributes - such as the best and worst attribute-level in a health state; Case 3: (repeated) best and worst choice among profiles - such as computers with different levels on various attributes. In Case 3, it is desired to measure the utility of individual attribute-levels, and their combination in giving the overall utility of a profile; and to develop, and test, possible theoretical relations between the representations in Cases 2 and 3 (Marley & Pihlens, 2010). Louviere et al.’s application of data expansion has been mainly to Case 3, possibly because the attribute structure of the profiles, plus efficient designs, make it possible to fit the data of individual respondents in this case. However, the basic expansion methods and ranking models can be described independent of whether Case 1, 2 or 3 is being studied, and that is what we do here.

For this paper, we present the models as if they are for individual participants.
3 Weighted utility ranking models

We now develop a class of ranking models that involve weights similar to those that arise in the Louviere et al. expansions. Certain of the models in this class can be indirectly motivated by those expansions - as is done Section 4. However, the results of the current section are in no way dependent on those expansions.

Assume that, for every choice set \( X \), there is a “local” importance weight \( r_X(i) \) for position \( i \), \( i = 1, ..., |X| \), of a rank order of the options in \( X \), and that the weights satisfy the ordinal constraints

\[
r_X(1) \geq r_X(2) \geq ... \geq r_X(|X| - 1) \geq r_X(|X|),
\]

with \( r_X(1) > r_X(|X|) \). Now assume that the ranking probabilities are of the form

\[
B_{R(X)}(\rho) = \frac{e^{\sum_{i=1}^{|X|} r_X(i)u(\rho_i)}}{\sum_{\eta \in R(X)} e^{\sum_{i=1}^{|X|} r_X(i)u(\eta_i)}}.
\]

We call this representation a weighted utility ranking (WUR) model.

Comment The ranking probabilities given by (4) are identical to those of a model with the weights: for each choice set \( X \),

\[
r_X'(i) = r_X(i) - \frac{1}{|X|} \left( \sum_{i=1}^{|X|} r_X(i) \right),
\]

which satisfy the condition: for each choice set \( X \),

\[
\sum_{i=1}^{|X|} r_X'(i) = 0.
\]

We generally do not impose the latter constraint as the representation of most of the (ranking) models are simpler without it.

Next, for \( x \in X \) and \( \rho \in R(X) \), define \( \rho(x) = i \) if \( x \) is in rank position \( i \) in \( \rho \). For convenience, put \( r_X[\rho(x)] = 0 \) whenever \( x \notin X \). Then \( r_X[\rho(x)] \) is well-defined and (4) becomes

\[
B_{R(X)}(\rho) = \frac{e^{\sum_{x \in X} r_X[\rho(x)]u(x)}}{\sum_{\eta \in R(X)} e^{\sum_{x \in X} r_X[\eta(x)]u(x)}}.
\]

For each set \( X \subseteq T \) in the design\(^3\) \( D(T) \), and rank order \( \rho \in R(X) \), let

\[
s_X(\rho) = \begin{cases} 
1 & \text{if } \rho \text{ is selected from } R(X) \\
0 & \text{otherwise}
\end{cases}.
\]

\(^3\)We assume each set \( X \in D(T) \) is presented once. The results generalize easily to the case where each set \( X \) is presented \( N_X \) times.
Then, given the representation (6), equivalently (4), the likelihood of the data is
\[
\prod_{X \in D(T)} \prod_{\rho \in R(X)} B_{R(X)}^\rho (\rho)^{s_X(\rho)}
\]
\[= \prod_{X \in D(T)} \prod_{\rho \in R(X)} \left( \frac{e^{\sum_{x \in T} r_X[\rho(x)] u(x)}}{\sum_{\eta \in R(X)} e^{\sum_{x \in T} r_X[\eta(x)] u(x)}} \right)^{s_X(\rho)}
\]
\[= \prod_{X \in D(T)} \left( \frac{1}{\sum_{\eta \in R(X)} e^{\sum_{x \in T} r_X[\eta(x)] u(x)}} \right) \times \prod_{X \in D(T)} e^{\sum_{\rho \in R(X)} \sum_{x \in T} s_X(\rho) r_X[\rho(x)] u(x)}
\]
\[= \left( \prod_{X \in D(T)} \frac{1}{\sum_{\eta \in R(X)} e^{\sum_{x \in T} r_X[\eta(x)] u(x)}} \right) \times e^{\sum_{x \in T} \left( \sum_{X \in D(T)} \sum_{\rho \in R(X)} s_X(\rho) r_X[\rho(x)] \right) u(x)}.
\]
So the set of scores defined by: for \( x \in T \),
\[
s(x) = \left( \sum_{X \in D(T)} \sum_{\rho \in R(X)} s_X(\rho) r_X[\rho(x)] \right),
\]
is a sufficient statistic.

**Comments**

1. Using general language (with undefined terms in quotation marks), the following states a result due to Huber (1963); Appendix A states the terms and result exactly. Assume that one is interested in the rank order, only, of the scale values in the WUR model (6). An acceptable loss function is a “penalty” function with a value that remains constant under a common permutation of the scores and the scale values, and that increases if the ranking is made worse by misordering a pair of scale values. Then, given the model in (6), ranking the scale values in descending order of the scores, breaking ties at random, has “minimal average loss” amongst all (“permutation invariant”) ranking procedures that depend on the data only through the set of scores.

2. Given the properties of the scores stated in Comment 1., they are likely useful starting values in estimating the maximum likelihood values of the utilities \( u(x), x \in T \). In fact, various empirical work on the maximum difference (maxdiff) model for best-worst choice (a special case of the representation in (6) - see below) gives linear relations between the scores and (maximum likelihood) estimates of the utilities (Louviere et al., 2008, 2011).

Thus, WUR models have relatively simple representations for ranking probabilities, with very nice score properties. However, as with the reversible ranking
model (Section 5.1), the best and worst choice probabilities are complex. In particular, let \( B_X(x) \) for \( x \in X \) be the sum of the rank order probabilities, given by (6), in which \( x \) is ranked first, and let \( W_X(x) \) for \( x \in X \) be the sum of the rank order probabilities, given by (6), in which \( x \) is ranked last. Then the representations of the best and worst choice probabilities are relatively complex, and have no “simple” process interpretation (other than as a sum of rank orders, with each rank order generated as above); in this sense, the rank orders can be considered basic, and the best and worst choice probabilities derived.

The representation (6) includes various standard models of best, worst, and best-worst choice as special cases:

**The MNL model for best choices**: This is the special case of (6) where: for \( z \in T \)

\[
r_X[\rho(z)] = \begin{cases} 
1 & \text{if } \rho(z) = 1 \\
0 & \text{otherwise} 
\end{cases}
\]

Thus, we only have discriminating information (from the participant) on the first option in the ranking (“best”). So replacing the rank notation \( B_{R(X)}(\rho) \) by the best choice notation \( B_X(x) \) gives

\[
B_X(x) = \frac{e^{u(x)}}{\sum_{r \in X} e^{u(r)}}.
\]

**The MNL model for worst choices**: This is the special case of (6) where: for \( z \in T \)

\[
r_X[\rho(z)] = \begin{cases} 
-1 & \text{if } \rho(z) = |X| \\
0 & \text{otherwise} 
\end{cases}
\]

Thus, we only have discriminating information (from the participant) on the last option in the ranking (“worst”). So replacing the rank notation \( B_{R(X)}(\rho) \) by the worst choice notation \( W_X(x) \), gives

\[
W_X(x) = \frac{e^{-u(x)}}{\sum_{r \in X} e^{-u(r)}}.
\]

**The maxdiff model for best-worst choice**: This is the special case of (6) where: for \( z \in T \)

\[
r_X[\rho(z)] = \begin{cases} 
1 & \text{if } \rho(z) = 1 \\
-1 & \text{if } \rho(z) = |X| \\
0 & \text{otherwise} 
\end{cases}
\]

Thus, we only have discriminating information (from the participant) on the first option in the ranking (“best”) and the last option in the ranking (“worst”). So replacing the rank notation \( B_{R(X)}(\rho) \) by the best-worst notation \( BW_X(x,y) \), \( x \neq y \), gives

\[
B_X(x,y) = \frac{e^{[u(x)-u(y)]}}{\sum_{r,s \in X} e^{[u(r)-u(s)]}}.
\]
i.e., the *maxdiff model* of best-worst choice (Marley & Louviere, 2005). For each set $X \subseteq T$ in the design $D(T)$, and $z \in X$, let

$$
bw_X(z) = \begin{cases} 
1 & \text{if } z \text{ is selected as best in } X \\
-1 & \text{if } z \text{ is selected as worst in } X \\
0 & \text{otherwise}
\end{cases},
$$

For convenience, also set $bw_X(z) = 0$ if $z \notin X$, and define: for $z \in T$,

$$
bw(z) = \sum_{X \in D(T)} s_X(z),
$$

i.e., $bw(z)$ is the number of times $z$ is chosen as best minus the number of times $z$ is chosen as worst. Then, as a special case of the development of (10) and the results that follow from it, the set of scores $bw(z)$, $z \in T$, is a sufficient statistic and ranking in descending of these scores is optimal in the sense of minimizing the expected loss.

The *reversible ranking model* (Section 5.1): Consider the special case where:

for $z \in T$,

$$
r_X[\rho(z)] = |X| - i \quad \text{if } \rho(z) = i.
$$

In this case, (10) is the *Borda score*, which has important desirable properties in voting (e.g., see Saari, 2008).

4 Rank Data Expansion Methods and Associated WUR Models

We now describe how the weights are derived in various of Louviere et al.’s expansions in the deterministic case for full and partial rank data; extend those motivations to the nondeterministic (probabilistic) case; and present specific WUR models motivated by that framework.

4.1 Full rank data

4.1.1 Expansion to all binary subsets

We continue using the example of Section 1. Louviere et al. (2008) do not present the following binary expansion, but it is presented in Louviere et al. (2011). As do Louviere et al., we begin by supposing that all choices are deterministic. Let $\rho_1 \succ \rho_2 \succ \rho_3 \succ \rho_4$ be a typical (data) rank order for each of the 4-element sets in Table 2. Louviere’s expansion of each such rank order to binary choices is as follows: i) replace the rank order by the “implied” binary choices $\rho_1 \succ \rho_2 , \rho_1 \succ \rho_3 , \rho_1 \succ \rho_4 , \rho_2 \succ \rho_3 , \rho_2 \succ \rho_4 , \rho_3 \succ \rho_4$; ii) “pretend” (Louviere et al, 2008, p. 138) that those binary choices are independent; iii) “predict” the total number of implied choices for each option in the set \{\rho_1, \rho_2, \rho_3, \rho_4\} from all possible binary comparisons of options in that set, given the rank order.
\( \rho_1 \succ \rho_2 \succ \rho_3 \succ \rho_4 \); this gives the Total in the final row of Table 4; iv) apply the same procedure to each data rank order obtained for each of the 20 sets in Table 2; this gives the “expanded” data that is then used in fitting an MNL model to the best choices in the original rank data. For instance, as already mentioned, Louviere et al. use the following estimation methods: weighted least squares (WLS); weighted conditional logit (WCL); and maximum likelihood (ML).

Table 4. “Implied” best choices for each 2-element subset of \( \{\rho_1; \rho_2; \rho_3; \rho_4\} \)
given the rank order \( \rho_1 \succ \rho_2 \succ \rho_3 \succ \rho_4 \).

<table>
<thead>
<tr>
<th>Preference</th>
<th>( \text{1st} )</th>
<th>( \text{2nd} )</th>
<th>( \text{3rd} )</th>
<th>( \text{4th} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order ( \rightarrow )</td>
<td>( \rho_1 )</td>
<td>( \rho_2 )</td>
<td>( \rho_3 )</td>
<td>( \rho_4 )</td>
</tr>
<tr>
<td>Choice Set:</td>
<td>Implied Choice:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Now we extend the above approach to the relation between probabilistic binary choices and probabilistic rankings (Table 5). We assume that the “expanded” binary choices are controlled by the binary choice probabilities in the final column of Table 5 and that those probabilities satisfy the MNL; for instance, in row 4, \( \rho_2 \) is “chosen” over \( \rho_3 \) with probability \( p(\rho_2; \rho_3) \). The MNL assumption is crucial in extending Louviere’s weight results (though not the method) from the deterministic to the probabilistic case\(^4\). We also assume that there is a set of best choice probabilities \( P_X(x), x \in X, X \in D(T), |X| \geq 3 \), that are generated by the same scale values as the binary choice probabilities. For now we leave it unspecified whether or not these are the theoretical best choice probabilities for sets with 3 or more options. Ultimately, Louviere et al. (2008) and Ebling et al. (2010) do assume that those best choice probabilities on larger sets (3 or more options) satisfy an MNL model with those scale values. Specifically, they use the data expansion methods in fitting an MNL model to the best choices in the rank data, not as a model of that (full) rank data. In contrast, we use the expansions as a way to motivate models of the rank orders. In particular, our models force the (marginal) best choice probabilities on larger sets (3 or more options) to not satisfy an MNL model.

Comment The general statements in the previous paragraph apply to all our uses in this paper of the Louviere et al. expansions and weights. We do not repeat those statements in later sections.

---

\(^4\) The method “works” for models other than the MNL, but the results are complex, and do not, in general, produce “weights” (total scores).
Table 5. “Implied” best choices for each 2-element subset of \( \{ \rho_1, \rho_2, \rho_3, \rho_4 \} \) given the rank order \( \rho_1 \succ \rho_2 \succ \rho_3 \succ \rho_4 \).

<table>
<thead>
<tr>
<th>Rank Position</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preference</td>
<td>( \rho_1 )</td>
<td>( \rho_2 )</td>
<td>( \rho_3 )</td>
<td>( \rho_4 )</td>
</tr>
<tr>
<td>Choice Set</td>
<td>( p(\rho_1, \rho_2) )</td>
<td>( p(\rho_1, \rho_3) )</td>
<td>( p(\rho_1, \rho_4) )</td>
<td>( p(\rho_2, \rho_3) )</td>
</tr>
<tr>
<td>Choice Probability</td>
<td>( p_1 )</td>
<td>( p_2 )</td>
<td>( p_3 )</td>
<td>( p_4 )</td>
</tr>
</tbody>
</table>

The probability of the pattern of “implied” choices in Table 5 (which includes the deterministic case as a special case) is

\[ p(\rho_1, \rho_2)p(\rho_1, \rho_3)p(\rho_1, \rho_4)p(\rho_2, \rho_3)p(\rho_2, \rho_4)p(\rho_3, \rho_4), \]

which is

\[ \prod_{1 \leq i < j \leq 4} p(\rho_i, \rho_j). \] (11)

Note that this probability equals 1 (i.e., the deterministic case) only if each of the stated binary choice probabilities in the final column of Table 5 equals 1 - i.e., \( p(\rho_i, \rho_j) = 1 \) for \( 1 \leq i < j \leq 4 \); these probabilities can approach 1 for the MNL model, (1), if each of \( u(\rho_i) - u(\rho_{i+1}) \), \( i = 1, 2, 3 \), is large and positive. However, except in the deterministic case (which is not exactly achievable by an MNL model), the sum of the probabilities (11) over all the choice patterns obtained by expanding all possible rank orders of a choice set \( X \) can be less than 1 (see next), and hence (11) does not produce a ranking model.

To show that the sum of (11) over all possible rank orders of a current choice set \( X = \{ w, x, y, z \} \) can be less than 1, assume every binary choice probability is in the open interval \((0, 1)\) and consider a participant making independent choices of the best element in each binary subset of \( X \), of which there are 6. Then the probability over all possible patterns of such independent 6 choices is 1; includes the terms in (11) for all possible rank orders of \( X \); but also includes, for example, the term \( p(w, x)p(y, w)p(z, w)p(x, y)p(z, y)p(x, z) \), which has its value in the open interval \((0, 1)\). However, that pattern includes the “cycle” of choices \( w \succ x, x \succ y, y \succ z, z \succ w \), which is not compatible with the Table 5 expansion of any rank order of \( X \). Hence this pattern of binary choices has positive probability, but is never produced by Table 5. Therefore, the sum of the terms in (11) over all rank orders is less than 1, and hence those terms do not correspond to a model of the data rank orders. So we next turn to such a ranking model, suggested by the representation in (11).
Comment With the assumption of a common MNL model across the binary choice probabilities, (11) becomes
\[
e^{[3u(\rho_1) + 2u(\rho_2) + u(\rho_3)]}
\]
\[
C[(u(\rho_i), i \in \{1, 2, 3, 4\}]
\]
where
\[
C[(u(\rho_i), i \in \{1, 2, 3, 4\}] = \left[ e^{u(\rho_1)} + e^{u(\rho_2)}\right] \left[ e^{u(\rho_1)} + e^{u(\rho_3)}\right] \left[ e^{u(\rho_1)} + e^{u(\rho_4)}\right] \cdot
\]
\[
\times \left[ e^{u(\rho_2)} + e^{u(\rho_3)}\right] \left[ e^{u(\rho_2)} + e^{u(\rho_4)}\right] \left[ e^{u(\rho_3)} + e^{u(\rho_4)}\right] \cdot
\]
has the same value for every rank order on the given set.

Starting with (11), one “natural” way to generate a probability distribution over the rank orders is given by the following process (hinted at above): Given a presented set \(X\), the respondent makes all binary comparisons of (distinct) pairs in that set; in our example, there are 6 such comparisons. If the resulting comparisons are compatible with a rank order, that rank order is selected; otherwise, the process starts over. This gives the probability of rank order \(R\) as
\[
B_R(X)(\rho) = \frac{\prod_{1 \leq i < j \leq 4} p(\rho_i, \rho_j)}{\sum_{\eta \in R(X)} \prod_{1 \leq i < j \leq 4} p(\eta_i, \eta_j)}. \quad (13)
\]
In particular, when the binary choice probabilities satisfy a common MNL model, (13) can be rewritten in the form
\[
B_R(X)(\rho) = \frac{e^{[3u(\rho_1) + 2u(\rho_2) + u(\rho_3)]}}{\sum_{\eta \in R(X)} e^{[3u(\eta_1) + 2u(\eta_2) + u(\eta_3)]}}. \quad (14)
\]
Equation (13) is the reversible ranking model (Marley, 1968) for choice sets of size 4.

Results and questions parallel to those in the remainder of this subsection apply to all the expansions and related WUR models in this paper. We include them only for the present case.

Comments
1. Let \(X\) be a typical set with \(|X| = 4\), then the ranking model (14) is a WUR model, (6), with weights \(r_X(i) = |X| - i\) and sufficient statistic given by the scores in (10).
2. Note that (12) only differs from (14) in the form of the denominator which, in both forms, is independent of the data rank order. A routine check shows that, with (12) in the likelihood (8), the argument from that likelihood to the scores in (10) being a sufficient statistic still holds. Remember, though, that the representation in (12) is a result of a Louviere et al. expansion, and is not a model of the original rank order data.
3. Define a set of best choice probabilities \(P_X\) for choice set \(X\) in the design by: for \(x \in X\),
\[
P_X(x) = \frac{e^{u(x)}}{\sum_{z \in X} e^{u(z)}}, \quad (15)
\]
Then the representation (14) is equivalent to the form

$$B_{R(X)}(\rho) = \frac{P_X(\rho_1)^3 P_X(\rho_2)^2 P_X(\rho_3)}{\sum_{\eta \in R(X)} P_X(\eta_1)^3 P_X(\eta_2)^2 P_X(\eta_3)},$$

which can be thought of as a normalized weighted conditional logit (WCL) model. Related (non-normalized) WCL models are used by Louviere et al. (2008) in fitting the best choices in the rank order data. However, Louviere et al. assume that those best choices satisfy an MNL model, which is not the case for the above normalized WCL model (see the next comment).

4. The following results are from Marley (1968). The representation in (14) across subsets $X \subseteq T$ is not compatible with a random utility model. The probability $B_X(w)$ of choosing a particular option $w$ from $X = \{w, x, y, z\}$ as best in the rank orders of $X$ is given by the sum of the relevant ranking probabilities given by (14) - that is,

$$B_X(w) = \sum_{\eta \in R(X \setminus \{w\})} BW_X(\eta).$$

Importantly, the marginal best choice probabilities $B_X$ do not satisfy an MNL model for best choices. This leads to the question of what is the correct interpretation of the probabilities $B_X$ - for instance: are they the probabilities of best choices only when a rank order is asked of a participant; or are they also the probabilities of best choice when a single best choice is asked of a participant. Finally, note that the model in (14) can be stated, and motivated, without the involvement of the (“dummy”) binary MNL model actually used in its development above - though its development from binary choice probabilities underscores its interpretation as a process model.

5. Assume that the design and Louviere expansion of this section are used to fit an MNL model to the best choices in the rank order data, using maximum likelihood applied to the likelihood expressions (12) - that is, for a typical set $X = \{w, x, y, z\}$, if the participant responds with the rank order $\rho = wxyz$, then the term in the likelihood function for that choice is be given by (12) with $\rho_1 = w$, $\rho_2 = x$, $\rho_3 = y$, $\rho_4 = z$. Based on the results in various papers on the use of expansion methods (e.g., Louviere et al., 2008; Islam et al., 2009; Ebling et al., 2010), we expect a good fit of the estimated MNL model to the best choices in the data rank orders. However, we have pointed out in 4., above, that the best choices given by the related reversible ranking model in (14) do not satisfy an MNL model. Thus, it is of interest for future research to study: i. the extent to which the probability of best choices given by the reversible ranking model can deviate from, or be approximated by, an MNL model, and ii. whether the parameter estimates of the best-fitting MNL model for the best choices using the expansion methods yield a good a fit of the related reversible ranking model to the rank order data.

4.1.2 Expansion to all non-empty, non-singleton subsets

The second expansion of full rank order data that we, and Louviere et al. (2008), consider is based on the assumption that a single sample rank order
\[ \rho = \rho_1 \rho_2 \rho_3 \rho_4 \] for a set \( X \) can be expanded, in a consistent way, to all non-empty and non-singleton subsets of \( X \) (Table 6). We assume that the “expanded” choices are controlled by the choice probabilities in the final column of Table 5 and that those probabilities satisfy the MNL model; for instance, in row 7, \( \rho_1 \) is “chosen” over \( \rho_2 \) and \( \rho_3 \) with probability \( P_{(\rho_1, \rho_2, \rho_3)}(\rho_1) \). Remember, we are using the expansion as a motivation for a ranking model; as pointed out in previous sections, this is not its use in Louviere et al.

Table 6. “Implied” best choices for each 2- and 3-element subset of \( \{\rho_1, \rho_2, \rho_3, \rho_4\} \) given the rank order \( \rho_1 \succ \rho_2 \succ \rho_3 \succ \rho_4 \).

<table>
<thead>
<tr>
<th>Rank Position</th>
<th>Preference</th>
<th>Choice</th>
<th>Implied Choice</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>( \rho_1 )</td>
<td>1</td>
<td>1</td>
<td>( p(\rho_1, \rho_2) )</td>
</tr>
<tr>
<td>2nd</td>
<td>( \rho_2 )</td>
<td>1</td>
<td>0</td>
<td>( p(\rho_1, \rho_3) )</td>
</tr>
<tr>
<td>3rd</td>
<td>( \rho_3 )</td>
<td>0</td>
<td>0</td>
<td>( p(\rho_1, \rho_4) )</td>
</tr>
<tr>
<td>4th</td>
<td>( \rho_4 )</td>
<td>0</td>
<td>1</td>
<td>( p(\rho_2, \rho_3) )</td>
</tr>
<tr>
<td>5th</td>
<td></td>
<td>0</td>
<td>1</td>
<td>( p(\rho_2, \rho_4) )</td>
</tr>
<tr>
<td>6th</td>
<td></td>
<td>0</td>
<td>1</td>
<td>( p(\rho_3, \rho_4) )</td>
</tr>
<tr>
<td>7th</td>
<td></td>
<td>1</td>
<td>1</td>
<td>( P_{(\rho_1, \rho_2, \rho_3)}(\rho_1) )</td>
</tr>
<tr>
<td>8th</td>
<td></td>
<td>1</td>
<td>0</td>
<td>( P_{(\rho_1, \rho_2, \rho_4)}(\rho_1) )</td>
</tr>
<tr>
<td>9th</td>
<td></td>
<td>1</td>
<td>0</td>
<td>( P_{(\rho_1, \rho_3, \rho_4)}(\rho_1) )</td>
</tr>
<tr>
<td>10th</td>
<td></td>
<td>0</td>
<td>1</td>
<td>( P_{(\rho_2, \rho_3, \rho_4)}(\rho_2) )</td>
</tr>
<tr>
<td>11th</td>
<td></td>
<td>1</td>
<td>1</td>
<td>( P_{(\rho_2, \rho_3, \rho_4)}(\rho_3) )</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>7</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

As before, the probability of the pattern of choices in Table 6 can approach 1 (i.e., be near deterministic) if each of the stated choice probabilities in the final column of Table 6 approaches 1, which they can for the MNL model, \( 1 \), if each of \( u(\rho_i) - u(\rho_{i+1}) \), \( i = 1, 2, 3, 4 \), is large and positive. However, as for the expansion of Table 5, except in the deterministic case (which is not exactly achievable with the assumed MNL representations), the sum of the probabilities of the product of the expressions in Table 6 over all (partial or full) rank orders can be less than 1.

Thus, assuming the choice probabilities in Table 6 satisfy an MNL model, and proceeding in a manner paralleling that in the previous section, we obtain the probability of rank order \( \rho \in R(X) \) as

\[
B_R(X)(\rho) = \frac{e^{[7u(\rho_1) + 3u(\rho_2) + u(\rho_3)]}}{\sum_{\eta \in R(X)} e^{[7u(\eta_1) + 3u(\eta_2) + 3u(\eta_3)]}}.
\]

Results parallel to the Comments of Section 4.1.1 also hold for the above WUR model. In particular, the “dummy” MNL best probabilities in the final
column of Table 6, used in motivating the ranking model, are not the (marginal) best choice probabilities of that model.

4.2 Partial rank data

Let \( \{w, x, y, z\} \) be the set of available options and suppose that \( w \) is selected as best and \( z \) as worst. If we now check each (sub)set of size 2, 3, 4 of \( \{w, x, y, z\} \) in turn, then we see that the subsets \( \{x, y\} \) and \( \{x, y, z\} \) are the only ones where the best element is not determined either by the information that \( w \) is best or by the information that \( z \) is worst. We next present the (Louviere) expansions for this case in a way that parallels the approach in Ebling et al (2010), though the weights we use differ slightly from theirs; at the end of the section, we present, and discuss, a WUR model with their weights.

4.2.1 Expansion to implied binary choice sets

As before, we assume that the binary choice probabilities satisfy the MNL model; the related data expansion to all binary choice sets (vis, Table 7) is mentioned in Louviere et al. (2008), but is not tested there. Denote a typical (data) partial rank order of \( X = \{w, x, y, z\} \) by \( \rho_1 \succ \{\rho_2, \rho_3\} \succ \rho_4 \), with the expansion to all binary choice sets in Table 7; the dashes (–) in row 4 indicate that we have no relevant data for those rows. Remember, we are using that expansion as a motivation for a ranking model; as pointed out in previous sections, this is not its use in Louviere et al.

Table 7. “Implied” best choices for each 2-element subset of \( \{\rho_1, \rho_2, \rho_3, \rho_4\} \)
given the partial rank order \( \rho_1 \succ \{\rho_2, \rho_3\} \succ \rho_4 \).

<table>
<thead>
<tr>
<th>Choice Set →</th>
<th>Rank Position</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subset:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>( \rho_1 )</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>( \rho_2 )</td>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>( \rho_3 )</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>( \rho_4 )</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Proceeding as in previous sections, the probability of the pattern of “implied” choices in Table 7 is

\[
p(\rho_1, \rho_2) p(\rho_1, \rho_3) p(\rho_1, \rho_4) p(\rho_2, \rho_4) p(\rho_3, \rho_4)
= p(\rho_1, \rho_2) p(\rho_1, \rho_3) p(\rho_1, \rho_4) [p(\rho_2, \rho_3) + p(\rho_3, \rho_2)] p(\rho_2, \rho_4) p(\rho_3, \rho_4).
\]

19
Adding the assumption of a common MNL model across the probabilities, this becomes
\[
\frac{e^{(3u(r_i) + 2u(r_j) + u(r_k))} + e^{(3u(r_i) + 2u(r_j) + u(r_l))}}{C((u(r_i), i \in \{1, 2, 3, 4\})}
\]
where
\[
C((u(r_i), i \in \{1, 2, 3, 4\}) = [e^{u(r_1)} + e^{u(r_2)}][e^{u(r_1)} + e^{u(r_2)}][e^{u(r_1)} + e^{u(r_2)}] \\
\times [e^{u(r_2)} + e^{u(r_3)}][e^{u(r_2)} + e^{u(r_3)}][e^{u(r_2)} + e^{u(r_3)}]
\]
has the same value for each of the rank orders on the given set.

As before, the probability of this pattern of choices can approach 1 (i.e., be near deterministic) if each of the stated binary choice probabilities in the final column of Table 7 approaches 1, which they can for the MNL model, (1), if each of \(u(r_i) - u(r_j), i = 2, 3, 4\), and \(u(r_i) - u(r_j), i = 2, 3\), is large and positive. However, except in the deterministic case (which is not exactly achievable), the sum of the probabilities over all (partial or full) rank orders given by the above expression is less than 1; this can be seen by noting, for example, that the pattern of choices \(r_1 \succ r_2, r_3 \succ r_4, r_2 \succ r_1, r_4 \succ r_3\) occurs with probability \(p(r_1, r_2)p(r_3, r_1)p(r_2, r_3)p(r_4, r_4)\). However, that pattern includes the “cycle” of choices \(r_1 \succ r_2, r_2 \succ r_3, r_3 \succ r_4, r_4 \succ r_1\), which is not compatible with the constraint \(r_r \succ \{r_s, r_t\} \succ r_u\) for any selection of distinct \(r, s, t, u\) with each of them belonging to the set of integers \(\{1, 2, 3, 4\}\).

The following process generates a probability distribution over the observed best-worst choices (partial rank orders). Assume that the participant makes all 6 binary comparisons of distinct options in the 4-element set; if those choices give a rank order, the participant reports the first (respectively, last) option in that rank order as best (respectively, worst); otherwise, the participant resamples until a rank order is obtained. As we know only the best and worst choice, we replace the notation \(B_{R(X)}(r)\) with \(BW_X(r_1, r_4)\), and the above process gives the best-worst choice probability
\[
BW_X(r_1, r_4) = e^{(3u(r_i) + 2u(r_j) + u(r_k))} + e^{(3u(r_i) + 2u(r_j) + u(r_l))} \\
\sum_{\eta \in R(X)} e^{3u(\eta_1) + 2u(\eta_2) + u(\eta_3)}
\]
where the terms in the second expression are those of the reversible ranking model, (14).

### 4.2.2 Expansion to implied non-empty, non-singleton subsets

As usual, we assume that all the choice probabilities in the following Table 8 satisfy the MNL model. Now we consider the expansion of the (data) partial rank order \(r_1 \succ \{r_2, r_3\} \succ r_4\) to best choices on all non-empty, non-singleton subsets; this parallels, but is not identical to, the case presented in Ebling et al. (2010); we present their representation later in this section. Remember, we are using the expansion as a motivation for a ranking model; as pointed out in previous sections, this is not its use in Louviere et al.
The observed best-worst pair \( (\rho_1 \text{ best}, \rho_4 \text{ worst}) \) gives no information of the relative rank of \( \rho_2 \) and \( \rho_3 \) in the set \( \{\rho_2, \rho_3\} \) (row 4) or in the set \( \{\rho_2, \rho_3, \rho_4\} \) (row 10) - though, in the latter set, we do know that one of them has to be best as \( \rho_4 \) is worst in the choice set \( \{\rho_1, \rho_2, \rho_3, \rho_4\} \). Also, if we assume that there is an underlying rank order\(^5\), then either \( \rho_2 \succ \rho_3 \), which means we place \( p(\rho_2, \rho_3) \) for choice set 4 and \( P_{\{\rho_2, \rho_3, \rho_4\}}(\rho_2) \) for choice set 10, or \( \rho_3 \succ \rho_2 \), which means we place \( p(\rho_3, \rho_2) \) for choice set 4 and \( P_{\{\rho_2, \rho_3, \rho_4\}}(\rho_3) \) for choice set 10.

As before, the probability of the pattern of choices in Table 8 can approach 1 (i.e., be near deterministic) if each of the stated choice probabilities in the final column of Table 8 approaches 1, which they can for the MNL model, (1), if each of \( u(\rho_1) - u(\rho_i), \ i = 2, 3, 4, \) and \( u(\rho_i) - u(\rho_4), \ i = 2, 3, \) is large and positive. However, as for previous expansions, except in the deterministic case (which is not exactly achievable with the assumed MNL representations), the sum of the probabilities over all (partial or full) rank orders given of the product of the expressions in Table 8 is less than 1.

Turning to the related WUR representation, we know only that \( \rho_1 \) is best and \( \rho_4 \) is worst, so we replace \( B_{R(X)}(\rho) \) with \( BW_X(\rho_1, \rho_4) \), and reasoning as in the previous section, we obtain the representation

\(^5\)We could develop a model that uses only the partial order information of Table 3. However, the resulting representation does not simplify in the way that the model based on full rank orders does. In particular, it does not produce a model with the weights \((8,3,3,1)\) that appear in Ebling et al. (2010) WLS expansion.
\[ BW_X(\rho_1, \rho_4) = \frac{e^{[7u(\rho_1) + 3u(\rho_2) + u(\rho_3)]} + e^{[7u(\rho_1) + 3u(\rho_2) + u(\rho_3)]}}{\sum_{\eta \in R(X)} e^{[7u(\eta_1) + 3u(\eta_2) + 3u(\eta_3)]}} \]

where the terms in the second expression are those of the reversible ranking model, (16).

The above representation can be rewritten as

\[ BW_X(\rho_1, \rho_4) = \frac{e^{[8u(\rho_1) + 4u(\rho_2) + 2u(\rho_3) + u(\rho_4)]} + e^{[8u(\rho_1) + 4u(\rho_2) + 2u(\rho_3) + u(\rho_4)]}}{\sum_{\eta \in R(X)} e^{[8u(\eta_1) + 4u(\eta_2) + 2u(\eta_3) + u(\rho_4)]}}. \]

A weighted utility representation similar to those above, but based on Ebling et al.'s (2010) weights (8, 3, 3, 1), is

\[ BW_X(\rho_1, \rho_4) = \frac{e^{[8u(\rho_1) + 3u(\rho_2) + 3u(\rho_3) + u(\rho_4)]}}{\sum_{\eta \in R(X)} e^{[8u(\eta_1) + 3u(\eta_2) + 3u(\eta_3) + u(\rho_4)]}}. \]

Appendix B shows that this representation can have \( BW_X(\rho_1, \rho_4) \) approach 1, i.e., deterministic choice, by making each of the differences \( u(\rho_i) - u(\rho_i), i = 2, 3, 4, \) and \( u(\rho_1) - u(\rho_1), i = 2, 3, \) large and positive.

Note that the pattern of weights (8, 3, 3, 1) is the average of the (8, 4, 2, 1) weights for the rankings \( \rho_1 \rho_2 \rho_3 \rho_4 \) and \( \rho_1 \rho_3 \rho_2 \rho_4 \) in the representation (17). However, I do not think arguments paralleling those we have given yield this “average” model, even if one introduces singleton choice sets to Table 8.

5 Models of Ranking by Repeated Best and/or Worst Choices, and Their Relations to WUR Models

The classic expression for a ranking probability in terms of best choice probabilities, usually developed with the best choice probabilities satisfying the MNL model, (1), has the form: for \( \rho = \rho_1 \rho_2 \ldots \rho_{|X|} \),

\[ B_{R(X)}(\rho) = B_X(\rho_1)B_X(\rho_2) \ldots B_{X_{|X|}}(\rho_{|X|-1}) \] (18)

However, the above expansion does not require the choice probabilities to satisfy the MNL model. We now summarize the classic argument for using the MNL in this context, then develop various ranking models in terms of repeated best and/or worst choices.

Assume that the (best) choice probabilities and the ranking probabilities on all subsets of a set \( T \) satisfy a random utility model in the sense that there are random variables \( U_z, z \in T \), such that for all \( x \in X \subseteq T \) with \( |X| \geq 2, \rho = \rho_1 \rho_2 \ldots \rho_{|X|} \),

\[ B_{R(X)}(\rho) = \Pr(U_{\rho_1} > U_{\rho_2} > \ldots > U_{\rho_{|X|-1}}). \]
and

\[ B_X(x) = \Pr(U_x > U_y, \forall y \in X - \{x\}). \]

Then a classic result is that a random utility model satisfies (18) if and only if the preference probabilities satisfy the MNL model, (1), (see Luce & Suppes, 1965, Th. 50).

As we have already noted, the relation in (18) between ranking and choice probabilities does not imply that the choice probabilities satisfy the MNL model. This is obvious in the sense that one can have a set of choice probabilities on the subsets \( X \subseteq T \) that do not satisfy the MNL model and then define the ranking probabilities on those subsets by (18). We now take such an approach to developing ranking models based on repeated best and/or worst choice.

For 4 options, there are 8 distinct patterns of best and/or worst questions, any one of which an experimenter can ask a participant to use for the generation of a rank order on those options; and, also, at least 8 possible (MNL-based) models of best and/or worst choice for the final rank order. Thus, if one truly believes that BWS is a method, not a model, then there are 64 possible combinations of method and model for the ranking of 4 options. Given such combinatorial possibilities, it is likely highly desirable to constrain the order in which a participant can produce a rank order - for instance, by instructing the participant in the order of best and/or worst choices to be used; removing each selected option from view once it has been selected; and modeling the data with the same sequence of best and/or worst choices - this is the approach in Scarpa and Marley (2011). Here, we introduce notation for 4 such models, illustrated with 4-element sets. First, we emphasize that we are using the notation \( \rho = \rho_1\rho_2\rho_3\rho_4 \) for the rank order in which \( \rho_i, i = 1, 2, 3, 4 \), is the option in rank position \( i \), and the ranking is: best, \( i = 1 \); 2\(^{nd}\) best, \( i = 2 \); 3\(^{rd}\) best, \( i = 3 \); 4\(^{th}\) best, \( i = 4 \) - thus, this notation does not carry information on the choices leading to this rank order; the latter information is introduced next.

Let \( r \) and \( s \) be two indicators that can, independently, take on the values \( b \) (for best) and \( w \) (for worst). Let \( P_{R(X)}^{r,s}(\rho_1\rho_2\rho_3\rho_4) \) denote the probability of the rank order \( \rho = \rho_1\rho_2\rho_3\rho_4 \) (from best to worst), when the rank order is obtained by: the first choices is of type \( r \); the second of type \( s \); the third, again, of type \( r \). Then we can have:

i. repeated best:

\[ P_{R(X)}^{b,b}(\rho) = B_X(\rho_1)B_{X-(\rho_1)}(\rho_1)B_{(\rho_3,\rho_4)}(\rho_3). \]  \hfill (19)

ii. repeated worst:

\[ P_{R(X)}^{w,w}(\rho) = W_X(\rho_4)W_{X-(\rho_4)}(\rho_3)W_{(\rho_1,\rho_2)}(\rho_2). \]  \hfill (20)

iii. repeated best-worst:

\[ P_{R(X)}^{b,w}(\rho) = B_X(\rho_1)W_{X-(\rho_1)}(\rho_4)B_{(\rho_2,\rho_3)}(\rho_2). \]  \hfill (21)
iv. repeated worst-best:

$$P_{R(X)}^{w;b}(\rho) = W_X(\rho_4)B_X-(\rho_4)(\rho_1)W_{(\rho_2,\rho_3)}(\rho_3).$$ (22)

For each of (19) and (21), it is easily checked that the sum of the rank orders in which \( \rho_1 \) is first equals \( B_X(\rho_1) \); and for each of (20) and (22), the sum of the rank orders in which \( \rho_4 \) is last equals \( W_X(\rho_4) \). With slightly more work - assuming, say MNLs of the form given in (1) and (2) - one can show that, in general, for each of (19) and (21), the sum of the rank orders in which \( \rho_4 \) is last does not equal \( W_X(\rho_4) \); and for each of (20) and (22), the sum of the rank orders in which \( \rho_1 \) is first does not equal \( B_X(\rho_1) \).

The natural first assumption in testing these models is to assume that the best (respectively, worst) choice probabilities satisfy the MNL model (1) (respectively, (2)), and, as needed by data, generalizations of those models that include a scale factor that depends on the location of the current choice set in the rank order; this scale factor relates to the possibly changing variability (“consistency”) of the choices across sets. For example, Scarpa et al. (2011) collected ranking data by repeated best, then worst, choices and fit that data quite successfully with a model based on repeated best choices, with those choices satisfying a generalization of the MNL that took account of the difference between the data collection method (repeated best, then worst) and the model (repeated best); it would be interesting to see if their data could be better fit by a model that matched their data collection method - i.e., the model of case iii., above. Collins and Rose (2011) fit related models to stated preference data on dating choices. Also, as required by data, the MNL can be replaced by other models, such as the generalized multinomial logit model (GMNL) for best choices, also adapted to worst choices (Fiebig et al., 2010). And one can consider latent class extensions of these models.

5.1 Known relations between models of repeated best and/or worst choice and WUR models

As throughout the paper: \( T \) denotes a finite set of options; we use the convention that, for each \( x \in T \), \( B_{\{x\}}(x) = W_{\{x\}}(x) = 1 \); and we write \( p(x, y) \) for \( B_{\{x,y\}}(x) \) and, if needed, \( p^*(x, y) \) for \( W_{\{x,y\}}(x) \). We also assume that the binary choice probabilities are transitive in the sense that, for any \( r, s, t \in T \), \( p(r, s) = p(s, t) = 1 \) implies that \( p(r, t) = 1 \), and \( p^*(r, s) = p^*(s, t) = 1 \) implies that \( p^*(r, t) = 1 \); transitivity ensures that the various choice probabilities defined below are, in fact, probabilities.

1. The rank order logit

The rank order logit (ROL) is the ranking model (19) with the best choice probabilities satisfying the MNL model. Lam, Konig & Franses (2010) have developed a fascinating WUR model approximation to the ranking probabilities given by the ROL model, using a Taylor expansion around the scale value \( u(x_o) \) of a referent option \( x_o \). The authors present applications of the resulting WUR
model to several data sets, and show that their interpretations of the data from the WUR model approximation to the ROL are similar to those obtained by fitting the ROL. An interesting theoretical question is whether, given a ROL model, there is a WUR model that gives exactly the same choice probabilities as that ROL model on one set or multiple sets. Answering this question will likely involve the fact that the ROL is a random utility model, and thus satisfies the constraints of such a model, whereas, in general, a WUR model is not a random utility model.

2. The reversible ranking model

The following results from Marley (1968) are needed for a result that we develop for WUR models that are also models based on repeated best and/or worst choices.

Definition 1 (Marley, 1968, Def. 6) A reversible ranking model is a set of choice and ranking probabilities on all the non-empty subsets of a finite set \( T \) that satisfies (19), (20), and, for each \( \rho \in R(X) \), \( X \subseteq T \), \( P_{R(X)}^{b,b}(\rho) = P_{R(X)}^{w,w}(\rho) \).

Definition 2 (Marley, 1968, Def. 2) A concordant choice model is a set of choice probabilities on all the non-empty subsets of a finite set \( T \) such that for each \( x, y \in X \subseteq T \), \( x \neq y \),

\[
B_{X\setminus\{x\}}(y) = W_{X\setminus\{y\}}(x).
\]

Theorem 3 (Marley, 1968, Th. 8). Assume a set of choice and ranking probabilities on all the non-empty subsets of a finite set \( T \) for which the binary choice probabilities are transitive. Then the following two conditions are equivalent:

(i) the probabilities form a reversible ranking model;
(ii) the probabilities form a concordant choice model and for each \( X \), \( X \subseteq T \),

\[
P_{R(X)}^{b,b}(\rho) = P_{R(X)}^{w,w}(\rho) = \frac{\prod_{1 \leq i < j \leq |X|} p(\rho_i, \rho_j)}{\sum_{\eta \in R(X)} \prod_{1 \leq i < j \leq |X|} p(\eta_i, \eta_j)}.
\]

Marley and Pihlens (2010) show that a reversible ranking model satisfies repeated best-worst, (21), and repeated worst-best, (22), in addition to repeated best and repeated worst (the latter two given by (23)).

Now we change our focus to WUR models. Suppose there is a WUR model that is also a reversible ranking model - that is, the model satisfies (6) and (23), and therefore (19) and (20). In particular, using the binary case of (6), this gives that, for \( x, y \in T \),

\[
p(x, y) = \frac{e^{r_{(x,y)}(1)u(x) + r_{(x,y)}(2)u(y)}}{e^{r_{(x,y)}(1)u(x) + r_{(x,y)}(2)u(y)} + e^{r_{(x,y)}(1)u(y) + r_{(x,y)}(2)u(x)}}.
\]

Multiplying each term in the numerator and denominator by \( e^{-r_{(x,y)}(2)u(x) + u(y)} \) and letting, for any \( r \in T \), \( v_{(x,y)} = [r_{(x,y)}(1) - r_{(x,y)}(2)] \), the above expression
becomes
\[ p(x, y) = \frac{e^{v(x,y)u(x)}}{e^{v(x,y)u(x)} + e^{v(x,y)u(y)}}. \] (24)

However, we have assumed that the WUR model is also a reversible ranking model, and so by Theorem 3 it has the representation (23). Substituting (24) in that representation, and using the notation of (6) for the WUR model, we have that, for \( i = 1, ..., |X| \),
\[ r_X(i) = \sum_{i < j \leq |X|} v_{\{\rho_i, \rho_j\}}. \]

In particular, in the special case where \( v_{\{\rho_i, \rho_j\}} = 1 \) for all \( \rho_i, \rho_j \), i.e., the binary MNL holds, we have \( r_X(i) = |X| - i \) - that is, the Borda score.

Summarizing, there is essentially only one WUR model (class) that is also a reversible ranking model.

6 Discussion and Conclusion

We have used the Louviere et al. (2008) expansion methods to motivate models of full and partial rank order data, several of which have been studied previously, others being new. The weights suggested by the expansion methods play an important role in motivating specific weighted utility ranking (WUR) models - in particular, each expansion (when the MNL model holds), and the corresponding ranking model, have a common sufficient statistic that depends on the weights in each approach. If one is interested in the rank order, only, of the scale values in the model, then assuming those scale values are ranked in the same order as the scores in the sufficient statistic is “optimal” (with optimal as defined in Appendix A). To the extent that WUR models with different weights, but the same scale values, give the same rank order of the scores, they will give the same estimated rank order for the scale values. This fact, and the close relation between Louviere et al. expansions and specific WUR models, might partially explain why various expansions give scale estimates related up to scale (as in Louviere et al., 2008).

In general, there is a clear difference between the properties of WUR models and of ranking models based on repeated best and/or worst choices. The main advantage of the WUR models is that they have useful score properties for estimating the rank order of the utility values, and, empirically, of estimating the utility values up to scale; the main advantage of the repeated best and/or worst models of ranking is that they provide natural process interpretations of the rankings and can be based on relatively standard (e.g., MNL) representations of the best and/or worst choice probabilities.

The reversible ranking model, with set-dependent binary MNLs, belongs to both classes of representation. The advantage of this model lies in the fact that it has a simple sufficient statistic, based on scores, and that ranking in descending order of these scores is optimal in the sense of minimizing the expected loss;
the disadvantage is its relatively complex forms for best, and worst, choice probabilities.

An open empirical and theoretical issue is the extent to which a WUR model can 'mimic' a repeated best and/or worst choice model, and vice-versa.

Given the diversity of ranking models presented in this paper, which are only a small subset of current approaches (e.g., Agarwal et al., 2009; Doignon et al., 2004), it is up to further empirical research to decide which ranking models are the most useful, in terms of, say, ease of use and in- and out-of-sample predictions. Models based on repeated best and/or worst choices can often be given (psychological) process interpretations, and those based on weights may be appropriate for policy making contexts, given their close links to scoring rules in voting theory (Saari, 2008).

Appendices

Appendix A Score Properties (Based on Huber, 1963)

Preliminary comments

In order to apply Huber’s (1963) results under exactly the conditions he assumed, we need the scores in (10) to satisfy the constraints\footnote{Note that for every $X$ and every rank order $\rho \in R(X)$, $\sum_{x \in X} r_X[\rho(x)] = \sum_{i = 1}^{|X|} r_X(i)$.}: for each choice set $X$,

$$\sum_{i = 1}^{|X|} r_X(i) = 0.$$  

(25)

In this case, it is easily checked that the scores in (10) satisfy the property:

$$\sum_{x \in T} s(x) = 0.$$  

(26)

When (25), and hence (26), hold, the conditions of our application agree with the assumptions in Huber (1963), especially the example in his Section 7. Therefore, we now show that these constraints can be derived from our general assumptions, with no impact on our other assumptions and conclusions.

So assume we have the scores as in (10) with the weights satisfying the ordinal constraint (3), i.e., for each $x \in T$ and choice set $X$,

$$s(x) = \left(\sum_{X \in D(T)} \sum_{\rho \in R(X)} s_X(\rho) r_X[\rho(x)]\right),$$

and

$$r_X(1) \geq r_X(2) \geq \ldots \geq r_X(|X| - 1) \geq r_X(|X|),$$

with $r_X(1) > r_X(|X|)$.

Now, for $i = 1, \ldots, |X|$, define

$$r'_X(i) = r_X(i) - \frac{1}{|X|} \left(\sum_{i = 1}^{|X|} r_X(i)\right),$$
and for $x \in T$, define
\[ s'(x) = s(x) - \sum_{X \in D(T)} \frac{1}{|X|} \left( \sum_{i=1}^{|X|} r_X(i) \right). \]  
(27)

Then it is routine to show that for each choice set $X$,
\[ \sum_{i=1}^{|X|} r'_X(i) = 0, \]
and for each $x \in T$,
\[ s'(x) = \left( \sum_{X \in D(T)} \sum_{\rho \in R(X)} s_X(\rho) r'_X(\rho(x)) \right), \]
which in turn yield
\[ \sum_{x \in T} s'(x) = 0. \]

Therefore, as required, the weights $r'_X$ satisfy (25) and, consequently, the scores $s'$ satisfy (26). Also, as already noted in the body of the paper, the ranking probabilities for the WUR with weights $r'_X(i), i = 1, ..., |X|$, are identical to those of the WUR with the weights $r_X(i)$; and, by (27), the rank order of the scores $s'(x), x \in T$, for this common model is the same as that of the scores $s(x), x \in T$. These are the conditions on the weights and scores that are required for our application of Huber’s results.

In conclusion, we apply Huber’s results assuming a set of weights $r_X$ satisfying (25), and hence scores $s$ satisfying (26), even though in the main text we have not assumed that (25) holds.

**The application of Huber’s (1963) results**

Let $T$ be a set with $|T| = n$. Consider the group of permutations of the numbers 1 to $n$. For a vector of (utility) values $u = (u(t_1), u(t_2), ..., u(t_n))$ and a permutation $\sigma$, let $u^\sigma = (u(t_{\sigma(1)}), u(t_{\sigma(2)}), ..., u(t_{\sigma(n)}))$. The decision space $D$ is the set of all possible rank orders; thus, a typical $d \in D$ has the form
\[ d(1, 2, ..., n) = (d(1), d(2), ..., d(n)), \]
where $d(i), i = 1, ..., n$, is a permutation of the numbers 1 to $n$, and, thus, $d(i)$ is the rank of $i$ in the ranking $d$. For $d \in D$ and permutation $\sigma$, let
\[ d^\sigma(1, 2, ..., n) = (d(\sigma(1)), d(\sigma(2)), ..., d(\sigma(n))). \]

We assume that the “true” ranking of the utility scale values (in the relevant models in the main text) is in descending order of their values $u(z), z \in T$, and departures from the true ranking are punished by some real-valued loss; denote by $L(u, d)$ the loss incurred when $u$ is the true value of the parameters and decision (rank order) $d \in D$ is taken.
Definition 4 (Adapted from Huber, 1963, p. 513) A loss function $L$ is called \textbf{acceptable} if it satisfies the following two condition:

(i) $L$ is permutation invariant: for all $u, d$ and $\sigma$,
$$L(u^\sigma, d^\sigma) = L(u, d).$$

(ii) $L$ does not decrease if the ranking, $d$, is made worse by interchanging two items. More precisely, assume that $u(t_i) \geq u(t_j)$ and let $d^{(i,j)}$ be the ranking in which $i$ and $j$ are interchanged relative to their position in $d$. If $d$ ranks item $i$ before item $j$, $d(i) < d(j)$, then
$$L(u, d) \leq L(u, d^{(i,j)}).$$

Property (ii) implies that ranking in descending order of the $u(t_i)$ minimizes the loss, as desired.

We now define the \textbf{risk} of a decision procedure. Let $S$ be the (discrete) space of a statistic $s$; define $s^\sigma(t_i) = s(t_{\sigma(i)})$, and let $p(s|u)$ be the probability (for the assumed model) of the statistic $s$ given the parameter vector $u$. Let $\varphi_s$ be a randomized decision procedure that depends on the data only through $s$, that is, a probability density $\varphi_s(d)$ on $D$ such that $\sum_{d \in D} \varphi_s(d) = 1$. The risk is then:
$$R(u, \varphi) = \sum_{d \in D} \sum_{s \in S} p(s|u)L(u, d)\varphi_s(d).$$

We have written the statement of the following theorem in terms of a discrete probability mass. The analogous theorem in Huber (1963) is proved for general (discrete or continuous) probability distributions.

In the theorem statement, $T = \{t_1, t_2, \ldots, t_n\}$ is a set, $s = (s(t_1), s(t_2), \ldots, s(t_n))$ with $\sum_{i=1}^n s(t_i) = 0$ and $u = (u(t_1), u(t_2), \ldots, u(t_n))$ are vectors of length $n$, and for any vector $r$ of length $n$, $r^{(i,j)}$, $i \neq j$, is the vector in which items $r_i$ and $r_j$ are interchanged.

Theorem 5 (Adapted from Huber, 1963, p. 513) Assume that the joint distribution $p(s|u)$ of $s$ conditional on $u$ can be written as
$$p(s|u) = c(u)f(u, s),$$
where $f$ satisfies

i. $f(u, s) = f(u^{(i,j)}, s^{(i,j)})$,

ii. $f(u, s) \geq f(u^{(i,j)}, s)$ whenever $u(t_i) \geq u(t_j)$ and $s(t_i) \geq s(t_j)$.

If $L$ is an acceptable loss function, Def. 4, then ranking in descending order of the $s(t_i)$, breaking ties at random, has minimal risk among all permutation invariant ranking procedures that depend on the data through $s$. 

29
Property i. says that $f$ is invariant under permutations and property ii. says that the size of $f$ decreases if the order of the elements $t_i$ and $t_j$ is different in $u$ than in $s$.

We now rewrite the likelihood, (9), for a WUR, in a form that allows us to use the result of the above theorem.

Let $|T| = n, T = \{t_1, t_2, \ldots, t_n\}$ (in no particular order), $u = (u(t_1), u(t_2), \ldots, u(t_n))$, and $s = (s(t_1), s(t_2), \ldots, s(t_n))$. Then the likelihood, (9), of the score vector $s$ of (10) - that is, the probability of $s$ conditional on $u$ - can be written as

\[ p(s|u) = c(u)f(u, s), \]

where

\[ c(u) = \left( \prod_{X \in D(T)} \frac{1}{\sum_{\eta \in R(X)} e^{\sum_{x \in T} r_x[p(x)] u(x)}} \right), \]

and

\[ f(u, s) = \exp \sum_{i=1}^n s(t_i) u(t_i). \]  \hspace{1cm} (28)

It is routine to check that $f$ in (28) satisfies i. and ii. of Theorem 5, hence the result of that theorem holds and ranking in descending order of the $s(t_j)$, breaking ties at random, has minimal risk among all permutation invariant ranking procedures that depend on the data through $s$.

**Appendix B Calculation of (Near) Deterministic Limit for the (8,3,3,1) WUR Model**

Assume the utility values for the four options $w, x, y, z$ have the values $a, b, c, d$, respectively, with $a > b \geq c > d$ - that is, $u(w) = a, u(x) = b, u(y) = c, u(z) = d$. Then, (4) becomes

\[ BW_X(w, z) = \frac{e^{[8a+3(b+c)+d]}}{\sum_{\eta \in R(X)} e^{[8u(\eta_1)+3u(\eta_2)+3u(\eta_3)+u(\rho_4)]}} \]

\[ = \frac{1}{1 + \sum_{\eta \in R(X)} e^{[8u(\eta_1)+3u(\eta_2)+3u(\eta_3)+u(\rho_4)]} - [8a + 3(b+c) + d]}. \]

Then, for $BW_X(w, z)$ to approach 1, we need to show that, for any permutation (rank order) $(g, h, i, k)$ of the options in the set \{w, x, y, z\}, other than \{w, x, y, z\} or \{w, y, x, z\}, the quantity

\[ [8u(g) + 3(u(h) + u(i)) + u(k)] - [8a + 3(b+c) + d] \]

can be made as negative as desired by making the differences $a - b, a - c, b - d, c - d$, and, therefore, $a - d$, large and positive. That is, we need to show that, for any permutation (rank order) $(g, h, i, k)$ of the options, the quantity

\[ [8a + 3(b+c) + d] - [8u(g) + 3(u(h) + u(i)) + u(k)] \]

\[ = 8(a - u(g)) + 3(b + c) - (u(h) + u(i)) + (d - u(k)) \]

30
can be made as positive as desired by making the differences \(a - b, b - c, c - d\) large and positive.

The following Table demonstrates this by full enumeration, where it can be checked that each term on the right in the following table can be made as positive as desired under the constraints on \(a, b, c,\) and \(d\) given above.

<table>
<thead>
<tr>
<th>(u(g), u(h), u(i), u(k))</th>
<th>(8(a - u(g)) + 3((b + c) - (u(h) + u(i))) + (d - u(k)))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a, b, c, d)</td>
<td>0</td>
</tr>
<tr>
<td>(a, c, b, d)</td>
<td>0</td>
</tr>
<tr>
<td>(a, b, d, c)</td>
<td>2(c - d)</td>
</tr>
<tr>
<td>(a, d, b, c)</td>
<td>2(c - d)</td>
</tr>
<tr>
<td>(a, c, d, b)</td>
<td>2(b - d)</td>
</tr>
<tr>
<td>(a, d, c, b)</td>
<td>2(b - d)</td>
</tr>
<tr>
<td>(b, a, c, d)</td>
<td>5(a - b)</td>
</tr>
<tr>
<td>(b, c, a, d)</td>
<td>5(a - b)</td>
</tr>
<tr>
<td>(b, a, d, c)</td>
<td>5(a - b) + 2(c - d)</td>
</tr>
<tr>
<td>(b, d, a, c)</td>
<td>5(a - b) + 2(c - d)</td>
</tr>
<tr>
<td>(b, d, c, a)</td>
<td>7(a - b) + 2(b - d)</td>
</tr>
<tr>
<td>(b, c, d, a)</td>
<td>7(a - b) + 2(b - d)</td>
</tr>
<tr>
<td>(c, a, b, d)</td>
<td>5(a - c)</td>
</tr>
<tr>
<td>(c, b, a, d)</td>
<td>5(a - c)</td>
</tr>
<tr>
<td>(c, a, d, b)</td>
<td>5(a - c) + 2(b - d)</td>
</tr>
<tr>
<td>(c, d, a, b)</td>
<td>5(a - c) + 2(b - d)</td>
</tr>
<tr>
<td>(c, b, d, a)</td>
<td>7(a - c) + 2(c - d)</td>
</tr>
<tr>
<td>(c, d, b, a)</td>
<td>7(a - c) + 2(c - d)</td>
</tr>
<tr>
<td>(d, a, b, c)</td>
<td>5(a - d) + 2(c - d)</td>
</tr>
<tr>
<td>(d, b, a, c)</td>
<td>5(a - d) + 2(c - d)</td>
</tr>
<tr>
<td>(d, a, c, b)</td>
<td>5(a - d) + 2(b - d)</td>
</tr>
<tr>
<td>(d, b, a, c)</td>
<td>5(a - d) + 2(b - d)</td>
</tr>
<tr>
<td>(d, c, a, b)</td>
<td>7(a - d)</td>
</tr>
<tr>
<td>(d, c, b, a)</td>
<td>7(a - d)</td>
</tr>
<tr>
<td>(d, b, c, a)</td>
<td>7(a - d)</td>
</tr>
</tbody>
</table>

Acknowledgments

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 in the Centre for the Study of Choice, University of Technology, Sydney. I thank Bart Frischknecht, Christine Eckert, and especially Towhidul Islam, for their patient discussions of what I call the “Louviere expansions.” Any misinterpretations of those expansions are my own.

References


ranking among, and in, multiattribute options (profiles). Working Paper 10-
003, Centre for the Study of Choice, University of Technology Sydney.

Cambridge University Press, Cambridge, UK.

best and/or worst ranking procedures. Presentation, Second International 
Choice Conference, Leeds, July 4-6, 2011.

Exploring scale effects of best/worst rank ordered choice data to estimate 
visitors’s benefits from alpine transhumance. American Journal of Agricultural 
Economics. In Press.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both 
better and worse. Memory and Cognition, 21, 546-556.