Journal of Choice Modelling 32 (2019) 100171

Contents lists available at ScienceDirect

Journal of Choice Modelling

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

Software Paper

An R package and tutorial for case 2 best–worst scaling

Hideo Aizaki[a,*], James Fogarty[b]

[a] Research Faculty of Agriculture, Hokkaido University, Kita-9, Nishi-9, Kita-ku, Sapporo, Hokkaido, 060-8589, Japan
[b] Agricultural and Resource Economics, School of Agriculture and Environment, The University of Western Australia, 35 Stirling Highway, Perth, 6009, Western Australia, Australia

ARTICLE INFO

Keywords:
Best worst scaling
Package
Profile
R
Tutorial

ABSTRACT

Case 2 (profile case) best–worst scaling (BWS) is a question-based survey method for measuring preferences for attribute levels. Several existing **R** packages help to implement the construction of Case 2 BWS questions (profiles) and the discrete choice analysis of the responses to the questions. Structuring the dataset for Case 2 BWS analysis is, however, complicated: there are several model variants for the analysis, and independent variables are set according to the variants. This complexity makes it difficult for non-expert users to prepare datasets for Case 2 BWS analysis. To improve the capability of **R** with respect to Case 2 BWS and facilitate easier data analysis, the package support.BWS2 has been developed. The package provides a function to map raw survey data into a format suitable for analysis, and also includes other useful functions, such as a function to calculate count-based BWS scores. A free online tutorial for Case 2 BWS in **R** has also been made available. These works make it easier for those new to Case 2 BWS to complete research using **R**, and facilitate the use of Case 2 BWS in various research fields.

1. Introduction

Stated preference (SP) methods, such as contingent valuation (CV), discrete choice experiments (DCEs), and best–worst scaling (BWS), can be used to measure people's preferences for market and non-market goods/services and their characteristics. The strengths and weakness of SP methods have been broadly discussed, but the methods are now in wide use (Carson, 2012).

BWS, which is a new and promising SP method, requires survey respondents to select their most preferred and least preferred options from a choice set, and the method generates more precise preference information relative to approaches that involve selecting only the best options (Louviere et al., 2015). BWS is categorized into three types: Case 1 (object case), Case 2 (profile case), and Case 3 (multi-profile case). In Case 2 BWS, which is the focus of this paper, respondents are presented with what are called profiles. Each profile contains information about several attributes, and for each attribute there are several levels. From each profile, respondents select the best and worst attribute levels. Analyzing the responses reveals information about preferences for attribute levels. Case 2 BWS has strengths relative to Case 1 BWS: for example, the unconditional demand for a good/service can be estimated by adding a question that asks respondents to accept/reject an entire profile after a question on selecting the best and worst levels from the profile (e.g., Coast et al., 2006; Louviere et al., 2015).[1] Despite the advantages of Case 2 BWS, the approach is used much less frequently than

* Corresponding author.
E-mail address: azk-r@spa.nifty.com (H. Aizaki).

[1] Case 2 BWS also has a strength relative to Case 3 BWS: respondents generally find Case 2 BWS questions easier to understand than Case 3 questions (Louviere et al., 2015). However, each of the three BWS approaches have strengths and weaknesses, and so it is not the case that any one version of BWS has an absolute advantage over the other two formats. It is only the case that some formats are better suited to some specific scenarios than others. See Louviere et al. (2015) for a review of the three types of BWS.

https://doi.org/10.1016/j.jocm.2019.100171
Received 5 May 2019; Accepted 10 June 2019
Available online 24 June 2019

Case 1 BWS (Louviere et al., 2015). At the moment Case 2 BWS has gained some traction in health/medical economics (Cheung et al., 2016; Mühlbacher et al., 2016), but use of the approach in other areas of economics is limited (see appendix A for details).

One of the reasons for the limited use of Case 2 BWS identified in Louviere et al. (2015) is the lack of software packages to analyze Case 2 BWS data. For example, **R** (R Core Team, 2019) has packages to assist with the experimental design aspect of constructing profiles, and some of the econometric methods needed for analysis, but an **R** function to assist with preparing the dataset for analysis was unavailable, until recently.

Information on the structure of a Case 2 BWS dataset, and how to create a dataset from the profiles and responses to Case 2 BWS questions can be found in a number of places (e.g., Flynn et al., 2007; Louviere et al., 2015; Hensher et al., 2015), but the process is complicated. The complications arise because the user must select a model variant according to the assumptions made regarding the respondents' choices, and the dataset must be created according to the selected model. There are three standard models in Case 2 BWS: paired, marginal, and marginal sequential models; and each model can also be divided into sub-variants according to the way the independent variables used in each model are created. A further complication is that the number of alternatives in a choice set differs across models, for the same question. For example, for a Case 2 BWS question consisting of four levels, the number of alternatives is 12, 8, and 7 for paired, marginal, and marginal sequential models (see section 2.3 for detail). This means that mapping data to the required format for Case 2 BWS is more difficult than for DCEs. It is reasonable to think that the added complexity of creating the dataset for analysis is one of the reasons for the lack of use of Case 2 BWS, relative to Case 1 BWS. The non-expert user, in particular, is likely to find the dataset creation step daunting.

To improve the capability of **R** with respect to Case 2 BWS, and facilitate data analysis, the package support.BWS2 (Aizaki, 2019) was developed. The package provides a function to map raw survey data into a format suitable for analysis, and also includes other useful functions, such as a function to calculate count-based BWS scores. A free online tutorial for Case 2 BWS in **R** (Aizaki and Fogarty, 2019) is also available. The package support.BWS2 fills an existing gap in the tools available for the non-expert user to conduct Case 2 BWS analysis in **R**.

This paper explains the package support.BWS2 and details how to create datasets for Case 2 BWS analysis using the package. The structure of this paper is as follows: Section 2 outlines the steps involved in Case 2 BWS. Section 3 provides details on model variants and the structure of the different dataset formats needed for each model variant. Section 4 identifies the relevant packages and functions required to conduct Case 2 BWS in **R** and then details how to use the support.BWS2 package to create a dataset for analysis. Section 5 provides an example of Case 2 BWS in **R** with a focus on the modeling approach. Concluding remarks are presented in Section 6. Note that part of the paper is a revised version of material included as part of the help for the package support.BWS2 (Aizaki, 2019) and also draws on the content in the online tutorial (Aizaki and Fogarty, 2019), but the substantive parts of the paper, including the details on the structure of the dataset formats for the model variants and how to construct the dataset for each variant in **R** are entirely new.

2. Outline of case 2 best–worst scaling

2.1. Steps of implementing case 2 best–worst scaling study

Each Case 2 BWS question has a profile that has three or more attributes and each attribute has two or more levels. The profile is expressed as a combination of levels. The set of profiles used are constructed using experimental designs. Each profile's attributes are fixed, and the combination of levels in each profile is varied according to the profile. A profile selected from all the constructed profiles is presented to respondents, who are then asked to choose the best and worst levels. This style of question is repeated until all the profiles are evaluated. Preference information regarding attribute levels is measured by analyzing the responses to the questions. The work flow can be divided into seven steps (Flynn et al., 2007, 2008; Louviere et al., 2015):

- Step 1 is to characterize the situation in which respondents need to make a decision, and to specify the attributes and their levels.
- Step 2 is to construct the profiles using an orthogonal array (OA).
- Step 3 is to map the profiles into survey questions.
- Step 4 is to actually implement the survey, where for each profile presented to a respondent they select both the best and worst level.
- Step 5 is to compile a raw dataset, which contains two responses for per question (profile): a response indicating the best level and a response indicating the worst level.
- Step 6 is to put the raw dataset into the format required for analysis, based on the specific model variant assumed.
- Step 7 is to analyze the data on preferences using either a counting approach and/or a modeling approach.

Steps 2 and 7 are the core statistical tasks for a Case 2 BWS study and these steps are explained in greater detail below.

2.2. The construction of profiles

Assume that the profiles have K attributes and each attribute has L_k levels ($k = 1,2, ...,K$). If all the attributes have the same number of levels, L, a L^K OA is used to construct the profiles. The columns of the OA correspond to attributes, while the rows correspond to profiles.

Fig. 1 shows an example of the process of creating Case 2 BWS questions from the OA. The first panel of Fig. 1 shows the basic information for each profile. In this example there are four attributes and each attribute has three levels: attribute A with levels A1,

1) Define attributes and their levels

Attribute	Level
A	A1, A2, A3
B	B1, B2, B3
C	C1, C2, C3
D	D1, D2, D3

2) Search an orthogonal array (OA) corresponding to the attributes and their levels

1	3	2	3
3	1	2	2
3	3	3	1
2	3	1	2
2	2	2	1
1	1	1	1
1	2	3	2
3	2	1	3
2	1	3	3

3) Transform the OA according to the attributes and their levels

A1	B3	C2	D3
A3	B1	C2	D2
A3	B3	C3	D1
A2	B3	C1	D2
A2	B2	C2	D1
A1	B1	C1	D1
A1	B2	C3	D2
A3	B2	C1	D3
A2	B1	C3	D3

4) Convert the OA into a series of questions

Q1 Please select your best and worst levels from the following four:

Best	Attribute	Worst
[ ]	A1	[ ]
[ ]	B3	[ ]
[ ]	C2	[ ]
[ ]	D3	[ ]

Fig. 1. Example of creating a Case 2 BWS question.

A2, and A3; …; and attribute D with levels D1, D2, and D3. The second panel of Fig. 1 shows the shortest possible 3^4 OA that corresponds to the assumptions. Now, let the attributes A, B, C, and D be assigned to the first, second, third, and fourth column of the OA, and let the values 1, 2, and 3 correspond to the level values. For this mapping of attributes to columns, and levels to the integer values, the third panel of Fig. 1 shows the nine question set combinations. The bottom panel of Fig. 1 then shows the profile corresponding to the first row of the OA, and the respondents who face this question are asked to select their best and worst levels from the four levels: A1, B3, C2, and D3.

2.3. Analysis of the responses

2.3.1. Counting approach

The counting approach calculates scores based on the number of times level i is selected as the best (B_{in}: B score for level i) and the worst (W_{in}: W score for level i) among all the questions for respondent n. A best-minus-worst (BW) score is defined as:

$$BW_{in} = B_{in} - W_{in}, \tag{1}$$

and the standardized variant is defined as:

$$std.\ BW_{in} = \frac{BW_{in}}{f_i}, \tag{2}$$

where f_i is the frequency with which level i appears across all questions. These scores reveal respondents' preferences for levels.

2.3.2. Modeling approach: three standard models

The modeling approach uses discrete choice models to analyze the responses. When using the modeling approach, a model type must be selected according to the assumption made about the way respondents make choices. For Case 2 BWS questions there are three standard models: paired, marginal, and marginal sequential. Although all three models assume respondents derive utility for each level in a profile, the assumption regarding how they select level i as the best, and level j ($i \neq j$) as the worst from the set of K

levels differs among the three models.

The paired model assumes that the difference in utility between the two levels represents the greatest utility difference among all $K \times (K-1)$ utility differences. Consider the example profile shown at the bottom of Fig. 1. This profile contains four levels: A1, B3, C2, and D3. The number of possible pairs is 12 ($= 4 \times (4-1)$): there are 12 possible pairs of the best and worst levels. If, for each pair, the first value denotes the best and the second value the worst, we have: (A1, B3), (A1, C2), (A1, D3), (B3, A1), (B3, C2), (B3, D3), (C2, A1), (C2, B3), (C2, D3), (D3, A1), (D3, B3), and (D3, C2) as the possible combinations. If a respondent selects A1 as the best level and C2 as the worst, the paired model assumes that the respondent calculates 12 utility differences as per the 12 possible pairs, and that the difference in utility between A1 and C2 is the maximum difference across all 12 utility differences.

The marginal model assumes that the utility for level i is the maximum among the utilities for K levels, and that level j is the minimum. In the paired model, we assumed that the respondent selected A1 as the best and C2 as the worst from the four levels (A1, B3, C2, and D3). From the marginal model perspective, for this same example, there are four possible best levels and four possible worst levels, and the respondent's choice behavior is interpreted as follows. The utility for A1 is the maximum among the four utilities for A1, B3, C2, and D3, and the utility for C2 is the minimum among the four.

As the best level in a profile must differ from the worst, in the same profile, it has been suggested that the marginal model assumption that the worst level is selected from K levels is not appropriate. The resolution to this issue leads to the marginal sequential model. The marginal sequential model assumes that the utility for level i is the maximum utility among the K levels, and that level j is the minimum among the remaining $K-1$ levels. So, for our example, the marginal sequential model assumes, there are four possible best levels and three possible worst levels in the profile. For the specific realization we have shown in Fig. 1, the model assumes that the respondent selects A1 as the best from the four possible best levels because the utility from A1 is the highest among A1, B3, C2, and D3; but selects C2 as the worst from the three possible worst levels (B3, C2, and D3) because the utility for C2 is the least of the three.

The three models generally assume that the utility for the level selected as the worst is the negative of the one selected as the best. With this assumption, and given the assumption for the stochastic component of the utility, the probability of selecting level i as the best and level j ($i \neq j$) as the worst from a choice set C, for each model, can be expressed using a conditional logit model with the systematic component of utility v, as follows:

- The paired model

$$\Pr(\text{best} = i, \text{worst} = j) = \frac{\exp(v_i - v_j)}{\sum_{p,q \subset C, p \neq q} \exp(v_p - v_q)}. \tag{3}$$

- The marginal model

$$\Pr(\text{best} = i, \text{worst} = j) = \frac{\exp(v_i)}{\sum_{p \subset C} \exp(v_p)} \frac{\exp(-v_j)}{\sum_{p \subset C} \exp(-v_p)}. \tag{4}$$

- The marginal sequential model

$$\Pr(\text{best} = i, \text{worst} = j) = \frac{\exp(v_i)}{\sum_{p \subset C} \exp(v_p)} \frac{\exp(-v_j)}{\sum_{q \subset C-i} \exp(-v_q)}. \tag{5}$$

For each model, the coefficients of variables show respondents' preferences regarding attributes and/or levels.

3. Model variants used in previous works

In addition to the three standard models, it is possible to identify a number of specific model variants. To illustrate, Table 1 lists nine model variants identified following a review of: a fundamental Case 2 BWS paper (Flynn et al., 2007); a major textbook of SP methods (Hensher et al., 2015); and published empirical papers (see appendix A for a list of Case 2 BWS empirical study papers). Across the examples identified the types of independent variables (i.e., attribute variables and/or level variables) considered vary widely.

Table 1
Classification of model variants found in the literature.

No.	Model	Attribute variable	Level variable
1	Paired	Dummy coding with reversed signs for the worst	Effect coding with a base level for each attribute
2	Paired	None	Dummy coding with a base level for all attributes
3	Marginal	Dummy coding with reversed signs for the worst	Effect coding with a base level for each attribute
4	Marginal	Dummy coding without reversed signs for the worst	Dummy coding with a base level for each attribute
5	Marginal	None	Dummy coding with a base level for each attribute
6	Marginal	None	Dummy coding with a base level for all attributes
7	Marginal sequential	Dummy coding without reversed signs for the worst	Dummy coding with a base level for each attribute
8	Marginal sequential	Effect coding with reversed signs for the worst	Effect coding with a base level for each attribute
9	Marginal sequential	None	Effect coding with a base level for each attribute

3.1. Paired model with attribute and level variables

Model 1 (e.g., Flynn et al., 2007) is a paired model with both attribute and level variables. With this approach the attribute variables are created as dummy-coded variables (attribute-specific constants), with reversed signs when the attributes are treated as the worst. So, attribute variable D_k takes a value of 1 if any level in attribute k is treated as the best in the possible pair; -1 if any level in attribute k is treated as the worst in the possible pair; and 0 otherwise. When estimating Model 1, an arbitrary attribute variable is omitted: the coefficient of the omitted attribute variable is normalized to be zero. This means that the coefficients of the remaining attribute variables are estimated relative to the omitted attribute. The level variables are created as effect-coded variables. The signs of the effect-coded level variables are also reversed when levels are treated as the worst. For the possible best cases, the level variable D_{ki} takes a value of 1 if level i in attribute k is treated as the best in a possible pair; -1 if the base (reference) level in attribute k is treated as the best; and 0 otherwise. For the worst case selections, the level variable D_{ki} takes a value of -1 if level i in attribute k is treated as the worst in a possible pair; 1 if the base level in attribute k is treated as the worst; and 0 otherwise. Unlike dummy coding, effect coding allows the calculation of a coefficient of the base level in each attribute: it is the negative of the sum of the remaining coefficients in each attribute. Consider the example shown in Fig. 1. Assume that attribute D is omitted, and that the third level in each attribute (i.e., levels A3, B3, C3, and D3) is the base level. The systematic component of the utility function is then:

$$v = \beta_A D_A + \beta_B D_B + \beta_C D_C + \beta_{A1} D_{A1} + \beta_{A2} D_{A2} + \beta_{B1} D_{B1} + \beta_{B2} D_{B2} + \beta_{C1} D_{C1} + \beta_{C2} D_{C2} + \beta_{D1} D_{D1} + \beta_{D2} D_{D2}, \tag{6}$$

where the βs are the coefficients to be estimated. As explained above, the coefficient of the attribute variable corresponding to D is zero, and the coefficients of the base levels A3, B3, C3, and D3 are calculated using $-(\beta_{A1} + \beta_{A2})$, $-(\beta_{B1} + \beta_{B2})$, $-(\beta_{C1} + \beta_{C2})$, and $-(\beta_{D1} + \beta_{D2})$, respectively. Strictly speaking, the coefficients of attribute variables are called attribute impacts, whereas coefficients of level variables are called level scale values (Flynn et al., 2007).

Table 2 shows a design matrix corresponding to the first Case 2 BWS question (profile) shown in the bottom of Fig. 1, on the basis of Model 1. The first two columns show pairs of levels, and the remaining columns are the design matrix. The profiles in this example have four attributes, so there are 12 (= 4 × 3) possible pairs in each question. This means that a set of 12 rows corresponds to all possible pairs for a Case 2 BWS question. The first row shows a possible pair in which A1 is treated as the best level (the value in the first column is A1) and B3 is treated as the worst (the value in the second column is B3): the values of attribute variables D_A, D_B, and D_C are 1, -1, and 0, respectively. The value of the level variable D_{A1} is 1; the values of level variables D_{B1} and D_{B2} are 1 (level variables are effect-coded and B3 is the base level in attribute B and is treated as the worst in the first row). The values of the remaining level variables are 0. The second row shows a pair in which level A1 is treated as the best and level C2 is treated as the worst: the values of attribute variables D_A, D_B, and D_C are 1, 0, and -1, respectively; the values of level variables D_{A1} and D_{C2} are 1 and -1, respectively, and all other values are 0.

3.2. Paired model with level variables

Model 2 (e.g., Al-Janabi et al., 2011) is a paired model with only level variables. The level variables for this specification are dummy-coded variables. The sign of the level variables are reversed when the levels are treated as the worst: the level variable D_{ki} takes a value of 1 if level i in attribute k is treated as the best in the possible pair, -1 if level i in attribute k is treated as the worst in the possible pair, and 0 otherwise. When estimating this model variant, an arbitrary level variable is omitted: the coefficient of the omitted level variable is normalized to be zero. This means that the coefficients of the remaining level variables are interpreted as relative to the omitted one. For the example shown in Fig. 1, if level D3 is assumed to be omitted level the systematic component of the utility function is:

$$v = \beta_{A1} D_{A1} + \beta_{A2} D_{A2} + \beta_{A3} D_{A3} + \beta_{B1} D_{B1} + \beta_{B2} D_{B2} + \beta_{B3} D_{B3} + \beta_{C1} D_{C1} + \beta_{C2} D_{C2} + \beta_{C3} D_{C3} + \beta_{D1} D_{D1} + \beta_{D2} D_{D2}. \tag{7}$$

Table 3 shows a design matrix corresponding to the profile shown in the bottom of Fig. 1 on the basis of Model 2. The number of rows corresponding to the profile in this case is the same as that shown in Table 2. The first row shows a possible pair in which level

Table 2
Design matrix corresponding to the profile shown at the bottom of Fig. 1, on the basis of Model 1. In this case, attribute D is omitted and the third level in each attribute (i.e., A3, B3, C3, and D3) is the base level.

Best	*Worst*	D_A	D_B	D_C	D_{A1}	D_{A2}	D_{B1}	D_{B2}	D_{C1}	D_{C2}	D_{D1}	D_{D2}
A1	B3	1	− 1	0	1	0	1	1	0	0	0	0
A1	C2	1	0	− 1	1	0	0	0	0	− 1	0	0
A1	D3	1	0	0	1	0	0	0	0	0	1	1
B3	A1	− 1	1	0	− 1	0	− 1	− 1	0	0	0	0
B3	C2	0	1	− 1	0	0	− 1	− 1	0	− 1	0	0
B3	D3	0	1	0	0	0	− 1	− 1	0	0	1	1
C2	A1	− 1	0	1	− 1	0	0	0	0	1	0	0
C2	B3	0	− 1	1	0	0	1	1	0	1	0	0
C2	D3	0	0	1	0	0	0	0	0	1	1	1
D3	A1	− 1	0	0	− 1	0	0	0	0	0	− 1	− 1
D3	B3	0	− 1	0	0	0	1	1	0	0	− 1	− 1
D3	C2	0	0	− 1	0	0	0	0	0	− 1	− 1	− 1

Table 3
Design matrix corresponding to the profile shown at the bottom of Fig. 1, on the basis of Model 2. In this case, level D3 is the omitted level.

Best	*Worst*	D_{A1}	D_{A2}	D_{A3}	D_{B1}	D_{B2}	D_{B3}	D_{C1}	D_{C2}	D_{C3}	D_{D1}	D_{D2}
A1	B3	1	0	0	0	0	− 1	0	0	0	0	0
A1	C2	1	0	0	0	0	0	0	− 1	0	0	0
A1	D3	1	0	0	0	0	0	0	0	0	0	0
B3	A1	− 1	0	0	0	0	1	0	0	0	0	0
B3	C2	0	0	0	0	0	1	0	− 1	0	0	0
B3	D3	0	0	0	0	0	1	0	0	0	0	0
C2	A1	− 1	0	0	0	0	0	0	1	0	0	0
C2	B3	0	0	0	0	0	− 1	0	1	0	0	0
C2	D3	0	0	0	0	0	0	0	1	0	0	0
D3	A1	− 1	0	0	0	0	0	0	0	0	0	0
D3	B3	0	0	0	0	0	− 1	0	0	0	0	0
D3	C2	0	0	0	0	0	0	0	− 1	0	0	0

A1 is treated as the best and level B3 is treated as the worst: the value of the level variable D_{A1} is 1, the value of level variable D_{B3} is − 1, and the values of the remaining level variables are 0. The third row shows a possible pair in which level A1 is treated as the best and level D3 is treated as the worst: the value of the level variable D_{A1} is 1, and the values of the remaining levels are 0. Level D3 is the base level, and the level variable corresponding to D3 is omitted in Table 3.

3.3. Marginal model with attribute and level variables (1)

Model 3 (e.g., Flynn et al., 2007) is one of two sub-variants belonging to the set of marginal models with both attribute and level variables. In this approach the attribute variables are dummy-coded variables, with reversed signs when the attributes are treated as the possible worst. This means that attribute variable D_k takes a value of 1 if any level in attribute k is treated as the possible best, − 1 if any level in attribute k is treated as the possible worst, and 0 otherwise. An arbitrary attribute variable is omitted: coefficients of the remaining attribute variables are estimated relative to the omitted one. The level variables are effect-coded. The sign of effect-coded level variables are also reversed when levels are treated as the possible worst. For the possible best cases, the level variable D_{ki} takes a value of 1 if level i in attribute k is treated as the possible best, − 1 if the base level in attribute k is treated as the possible best, 0 otherwise. For the possible worst, the level variable D_{ki} takes a value of − 1 if level i in attribute k is treated as the possible worst, 1 if the base level in attribute k is treated as the possible worst, and 0 otherwise. For the example shown in Fig. 1, assume that attribute D is omitted and the third level in each attribute (i.e., levels A3, B3, C3, and D3) is the base level. The systematic component of the utility function is then:

$$v = \beta_A D_A + \beta_B D_B + \beta_C D_C + \beta_{A1} D_{A1} + \beta_{A2} D_{A2} + \beta_{B1} D_{B1} + \beta_{B2} D_{B2} + \beta_{C1} D_{C1} + \beta_{C2} D_{C2} + \beta_{D1} D_{D1} + \beta_{D2} D_{D2}. \tag{8}$$

This is the same structure as that for Model 1, however, the number of rows of the design matrix for each profile is different to that for Model 1.

Table 4 shows a design matrix corresponding to the profile shown in the bottom of Fig. 1, for Model 3. As explained above, the marginal model assumes that respondents select a level as the best or the worst from all the levels in a profile. For the example this means that respondents select the best and the worst from four levels. A set of eight rows in the design matrix therefore corresponds to the possible best and worst levels in a question: the first four rows of the design matrix show the levels treated as the possible best, and the last four rows show levels treated as the possible worst. The first column (*BW*) is an indicator that is 1 for rows corresponding to the possible best, and − 1 for rows corresponding to the possible worst. The second column (*Level*) denotes levels treated as the possible best or the possible worst in each row. The first row shows that level A1 is treated as the possible best and thus the values of attribute variable D_A and level variable D_{A1} are 1, while the values of the remaining variables are 0. The second row shows that level B3 is treated as the possible best and thus the value of attribute variable D_B is 1, the values of level variables D_{B1} and D_{B2} are − 1, and the values of the remaining variables are 0. The fifth row shows that level A1 is treated as the possible worst, and thus the values of D_A and D_{A1} are − 1, while the values of the remaining variables are zero. The last row shows that level D3 is treated as the possible worst and thus the values of all the attribute variables are 0 while the values of

Table 4
Design matrix corresponding to the profile shown at the bottom of Fig. 1, on the basis of Model 3. In this case, attribute D is omitted and the third level in each attribute (i.e., A3, B3, C3, and D3) is the base level.

BW	*Level*	D_A	D_B	D_C	D_{A1}	D_{A2}	D_{B1}	D_{B2}	D_{C1}	D_{C2}	D_{D1}	D_{D2}
1	A1	1	0	0	1	0	0	0	0	0	0	0
1	B3	0	1	0	0	0	− 1	− 1	0	0	0	0
1	C2	0	0	1	0	0	0	0	0	1	0	0
1	D3	0	0	0	0	0	0	0	0	0	− 1	− 1
−1	A1	− 1	0	0	− 1	0	0	0	0	0	0	0
−1	B3	0	− 1	0	0	0	1	1	0	0	0	0
−1	C2	0	0	− 1	0	0	0	0	0	− 1	0	0
−1	D3	0	0	0	0	0	0	0	0	0	1	1

D_{D1} and D_{D2} are 1, and the values of the remaining level variables are 0.

3.4. Marginal model with attribute and level variables (2)

Model 4, which is derived from Hensher et al. (2015 Table 6B.5), is another sub-variant belonging to the marginal model with both attribute and level variables. Although the attribute variables for this approach are created as dummy-coded variables, the sign of the value is not reversed when attributes are treated as the possible worst. The attribute variable D_k takes the value of 1 if any level in attribute k is treated as the possible best or the possible worst, and 0 otherwise. An arbitrary attribute variable is omitted from the model, and the coefficients of the remaining attribute variables are interpreted as relative to the omitted attribute. The level variables are created as dummy-coded variables in which base level is set for each attribute. The level variable D_{ki} takes a value of 1 if level i in attribute k is treated as the possible best, − 1 if treated as the possible worst, and 0 otherwise. An arbitrary level variable for each attribute is omitted and the coefficients of the remaining level variables for each attribute are interpreted relative to the omitted level. For the example shown in Fig. 1, and assuming that attribute D and levels A3, B3, C3, and D3 are omitted from the model, the systematic component of the utility function is:

$$v = \beta_A D_A + \beta_B D_B + \beta_C D_C + \beta_{A1} D_{A1} + \beta_{A2} D_{A2} + \beta_{B1} D_{B1} + \beta_{B2} D_{B2} + \beta_{C1} D_{C1} + \beta_{C2} D_{C2} + \beta_{D1} D_{D1} + \beta_{D2} D_{D2}. \tag{9}$$

Table 5 shows the design matrix corresponding to the profile shown in the bottom of Fig. 1 on the basis of Model 4. Similar to Table 4, the first four rows of the design matrix show levels treated as the possible best, while the last four rows show levels treated as the possible worst. Note that unlike Table 4, the value of the attribute variables D_A, D_B, and D_C are 0 or 1 (not − 1) even if those are in rows corresponding to the possible worst. Additionally, the level variables are dummy-coded variables: − 1 does not appear in the row corresponding to the possible best.

3.5. Marginal model with level variables (1)

Model 5, which is derived from Hensher et al. (2015 Table 6B.4), is the marginal model with only level variables. The level variables are dummy-coded and so for our example that part of the design matrix is the same as for Model 4. The attribute variables (D_A, D_B, and D_C) in Table 5, however, are not used in Model 5, and the systematic component of the utility function is:

$$v = \beta_{A1} D_{A1} + \beta_{A2} D_{A2} + \beta_{B1} D_{B1} + \beta_{B2} D_{B2} + \beta_{C1} D_{C1} + \beta_{C2} D_{C2} + \beta_{D1} D_{D1} + \beta_{D2} D_{D2}. \tag{10}$$

3.6. Marginal model with level variables (2)

Model 6 (e.g., Pollitt et al., 2016) is another sub-variant of the marginal model with only level variables. For Model 6 the level variables are dummy-coded, just as they are in Models 4 and 5, but for Model 6 the base level is different and an arbitrary level from all the levels is omitted. That is, Models 4 and 5 omit K level variables from the utility function, whereas Model 6 omits only one level variable from the utility function. For the example shown in Fig. 1, and with the assumption that level D3 is the base level, the systematic component of the utility function is:

$$v = \beta_{A1} D_{A1} + \beta_{A2} D_{A2} + \beta_{A3} D_{A3} + \beta_{B1} D_{B1} + \beta_{B2} D_{B2} + \beta_{B3} D_{B3} + \beta_{C1} D_{C1} + \beta_{C2} D_{C2} + \beta_{C3} D_{C3} + \beta_{D1} D_{D1} + \beta_{D2} D_{D2}. \tag{11}$$

Table 6 shows the design matrix corresponding to the profile shown in the bottom of Fig. 1, on the basis of Model 6 with the level variable D3 omitted from the design matrix.

3.7. Marginal sequential model with attribute and level variables (1)

Model 7, which is derived from Hensher et al. (2015 Table 6B.6), is one of two sub-variants belonging to the marginal sequential model with both attribute and level variables. For Model 7 the attribute and level variables are created as dummy-coded variables that are the same as for Model 4. For the example shown in Fig. 1 and the assumption regarding the attribute and level variables used in Model 4, the systematic component of the utility function is the same as that for Model 4.

Table 5
Design matrix corresponding to the profile shown at the bottom of Fig. 1, on the basis of Model 4. In this case, attribute D and levels A3, B3, C3, and D3 are omitted.

BW	*Level*	D_A	D_B	D_C	D_{A1}	D_{A2}	D_{B1}	D_{B2}	D_{C1}	D_{C2}	D_{D1}	D_{D2}
1	A1	1	0	0	1	0	0	0	0	0	0	0
1	B3	0	1	0	0	0	0	0	0	0	0	0
1	C2	0	0	1	0	0	0	0	0	1	0	0
1	D3	0	0	0	0	0	0	0	0	0	0	0
− 1	A1	1	0	0	− 1	0	0	0	0	0	0	0
− 1	B3	0	1	0	0	0	0	0	0	0	0	0
− 1	C2	0	0	1	0	0	0	0	0	− 1	0	0
− 1	D3	0	0	0	0	0	0	0	0	0	0	0

Table 6
Design matrix corresponding to the profile shown at the bottom of Fig. 1, on the basis of Model 6. In this case, level D3 is the omitted level.

BW	*Level*	D_{A1}	D_{A2}	D_{A3}	D_{B1}	D_{B2}	D_{B3}	D_{C1}	D_{C2}	D_{C3}	D_{D1}	D_{D2}
1	A1	1	0	0	0	0	0	0	0	0	0	0
1	B3	0	0	0	0	0	1	0	0	0	0	0
1	C2	0	0	0	0	0	0	0	1	0	0	0
1	D3	0	0	0	0	0	0	0	0	0	0	0
−1	A1	−1	0	0	0	0	0	0	0	0	0	0
−1	B3	0	0	0	0	0	−1	0	0	0	0	0
−1	C2	0	0	0	0	0	0	0	−1	0	0	0
−1	D3	0	0	0	0	0	0	0	0	0	0	0

The row structure of the design matrix for Model 7 is however different to that of Model 4. For the example shown in Fig. 1, assume that a respondent selects A1 as the best level in the profile. Table 7 shows the design matrix corresponding to the profile for the respondent on the basis of Model 7. The first four rows correspond to the four possible best levels, from which the respondent selects the best level. The last three rows correspond to the three possible worst, from which the respondent selects the worst level. The number of rows differs among the best choice situation and the worst choice situation. This is because the marginal sequential model assumes that the respondents select level i as the best from K levels in a profile, and then select level j as the worst among from the remaining $K - 1$ levels. In this instance, because the respondent is assumed to select A1 as the best level, the row corresponding to A1 is deleted from the rows corresponding to the worst choice situation.

3.8. Marginal sequential model with attribute and level variables (2)

Model 8 (e.g., Huynh et al., 2017) is another sub-variant of the marginal sequential model, with both attribute and level variables. The attribute variables are effect-coded, and an arbitrary attribute is the base attribute. The level variables are also effect-coded, and an arbitrary level in each attribute is the base level. The signs of effect-coded attribute and level variables are reversed when attributes/levels are treated as the possible worst.

Table 8 shows the design matrix on the basis of Model 8, under the same conditions as in Table 7. The attribute variable corresponding to D and level variables corresponding to A3, B3, C3, and D3 are omitted from the design matrix. The systematic component of the utility for Model 8 is therefore:

$$v = \beta_A D_A + \beta_B D_B + \beta_C D_C + \beta_{A1} D_{A1} + \beta_{A2} D_{A2} + \beta_{B1} D_{B1} + \beta_{B2} D_{B2} + \beta_{C1} D_{C1} + \beta_{C2} D_{C2} + \beta_{D1} D_{D1} + \beta_{D2} D_{D2}. \tag{12}$$

3.9. Marginal sequential model with level variables

Model 9 (e.g., dosReis et al., 2017) is the marginal sequential model with only level variables. The level variables are effect-coded, and an arbitrary level in each attribute is the base level. The signs of effect-coded level variables are reversed when levels are treated as the possible worst. If we assume that the third level in each attribute is the base level, and that the respondent selects A1 as the best in the example profile shown in the bottom of Fig. 1, then the design matrix for Model 9 is the same as for Model 8. However, the attribute variables D_A, D_B, and D_C in Table 8 are not used in Model 9. The systematic component of the utility function is:

$$v = \beta_{A1} D_{A1} + \beta_{A2} D_{A2} + \beta_{B1} D_{B1} + \beta_{B2} D_{B2} + \beta_{C1} D_{C1} + \beta_{C2} D_{C2} + \beta_{D1} D_{D1} + \beta_{D2} D_{D2}. \tag{13}$$

4. Software design

4.1. Roles of the package support.BWS2 and other packages for case 2 BWS

Fig. 2 shows the flow of **R** functions used to implement Case 2 BWS. Step 1 is to characterize the choice situation and thus is a project

Table 7
Design matrix corresponding to the profile shown at the bottom of Fig. 1, on the basis of Model 7. In this case, attribute D and levels A3, B3, C3, and D3 are omitted and level A1 is selected as the best.

BW	*Level*	D_A	D_B	D_C	D_{A1}	D_{A2}	D_{B1}	D_{B2}	D_{C1}	D_{C2}	D_{D1}	D_{D2}
1	A1	1	0	0	1	0	0	0	0	0	0	0
1	B3	0	1	0	0	0	0	0	0	0	0	0
1	C2	0	0	1	0	0	0	0	0	1	0	0
1	D3	0	0	0	0	0	0	0	0	0	0	0
−1	B3	0	1	0	0	0	0	0	0	0	0	0
−1	C2	0	0	1	0	0	0	0	0	−1	0	0
−1	D3	0	0	0	0	0	0	0	0	0	0	0

Table 8
Design matrix corresponding to the profile shown at the bottom of Fig. 1, on the basis of Model 8. In this case, attribute D and levels A3, B3, C3, and D3 are omitted and level A1 is selected as the best.

BW	*Level*	D_A	D_B	D_C	D_{A1}	D_{A2}	D_{B1}	D_{B2}	D_{C1}	D_{C2}	D_{D1}	D_{D2}
1	A1	1	0	0	1	0	0	0	0	0	0	0
1	B3	0	1	0	0	0	−1	−1	0	0	0	0
1	C2	0	0	1	0	0	0	0	0	1	0	0
1	D3	−1	−1	−1	0	0	0	0	0	0	−1	−1
−1	B3	0	−1	0	0	0	1	1	0	0	0	0
−1	C2	0	0	−1	0	0	0	0	0	−1	0	0
−1	D3	1	1	1	0	0	0	0	0	0	1	1

specific qualitative process that does not require software. Step 2 is to construct the profiles using an OA. From the package DoE.base (Grömping, 2018) the function `oa.design()` can be used to construct OAs. The columns of the OA generated using `oa.design()` correspond to attributes and the rows correspond to profiles. Step 3 is to map the profiles into survey questions. The `bws2.questionnaire()` function in the package support.BWS2 can be used for this step. Step 4, the survey function step, is separate to any analysis or data management undertaken in **R**. After the survey step, a raw dataset is compiled (Step 5). Step 6 involves taking the raw survey response data and putting this information in the format required for analysis in **R**. The function `bws2.dataset()` from support.BWS2 allows the raw survey data and the OA to be merged and converted to the format required for each of the model variants. Step 7 is the data analysis step. For the counting approach, the support.BWS2 package also provides the function `bws2.count()` which can be used to calculate BWS scores. There are several existing **R** packages that include functions that can be used for the modeling approach. For example, the survival package (Therneau, 2018; Therneau and Grambsch, 2000); the mlogit package (Croissant, 2019); the gmnl package (Sarrias and Daziano, 2017); and the apollo package (Hess and Palma, 2019) all have functions that can be used to estimate conditional logit models (the latter three packages also enable estimation of advanced discrete choice models).

In summary, the package support.BWS2 provides functions to convert an OA into Case 2 BWS questions (Step 3), create a dataset for the analysis from the OA and the responses to the questions (Step 6), and calculate BWS scores (Step 7). Other packages are needed to complete the implementation of Case 2 BWS with **R**: a package to construct the OA (Step 2) and to analyze the responses using the modeling approach (Step 7). How to use the support.BWS2 package to create a dataset for analysis is set out in detail in the next section. The other functions required to implement Case 2 BWS are explained in the example section.

4.2. R function for creating the dataset for the analysis

The `bws2.dataset()` function of the support.BWS2 package creates a dataset suitable for Case 2 BWS analysis. The generic call to the function is (see Table 9 for details on the arguments):

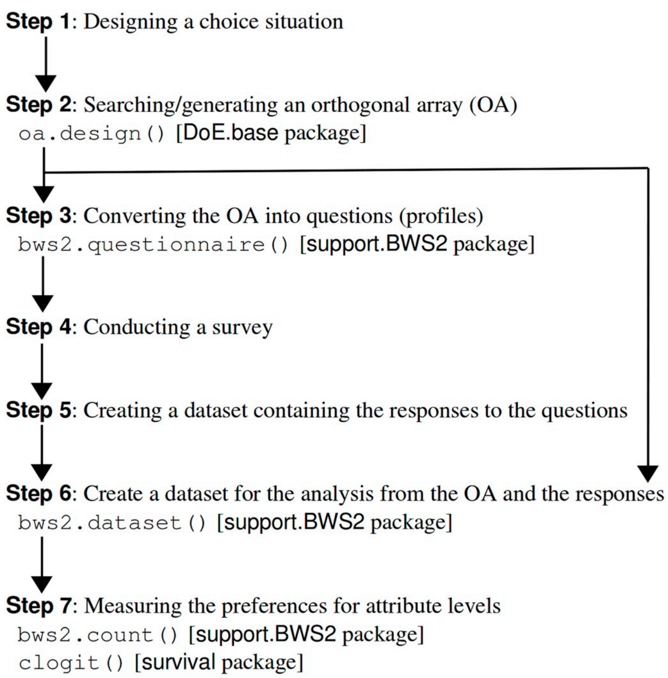


Fig. 2. R functions used in each step of the Case 2 BWS implementation.

Table 9
Main arguments of the function `bws2.dataset()`.

Argument	Definition
`data`	A data frame containing a respondent dataset.
`id`	A character showing the name of the respondent identification number variable used in the respondent dataset.
`responses`	A vector containing the names of response variables in the respondent dataset, showing the best and worst levels selected for each Case 2 BWS question.
`choice.sets`	A data frame or matrix containing an orthogonal array.
`attribute.levels`	A list containing the names of the attributes and their levels.
`base.attribute`	A character showing the base attribute; the argument is used when attribute variables are created as effect-coded ones and `NULL` is assigned to the argument when attribute variables are created as dummy-coded ones.
`base.level`	A list containing the base level in each attribute; the argument is used when level variables are created as effect-coded ones and `NULL` is assigned to the argument when level variables are created as dummy-coded ones.
`reverse`	A logical value denoted by `TRUE` when the signs of the attribute variables are reversed for the possible worst, or `FALSE` when not doing so.
`model`	A character showing a type of dataset created by this function: "`paired`" for a paired model, "`marginal`" for a marginal model, and "`sequential`" for a marginal sequential model.

```
bws2.dataset(data, id, responses, choice.sets, attribute.levels,
  base.attribute = NULL, base.level = NULL, reverse = TRUE,
  model = "paired")
```

The respondent dataset, in which each row corresponds to a respondent and each column corresponds to variables, must be organized by users and then assigned to the `data` argument. The dataset must include the respondent's identification number (id) variable in the first column and the response variables in the subsequent columns, each indicating which levels are selected as the best and worst for each question. Other variables in the respondent dataset are treated as the respondents' characteristics, such as gender and age. Respondents' characteristic variables are also stored in the dataset created by the `bws2.dataset()` function. Although the names of the id and response variables are left to the discretion of the user, those of the id and response variables are assigned to the `id` and `responses` arguments.

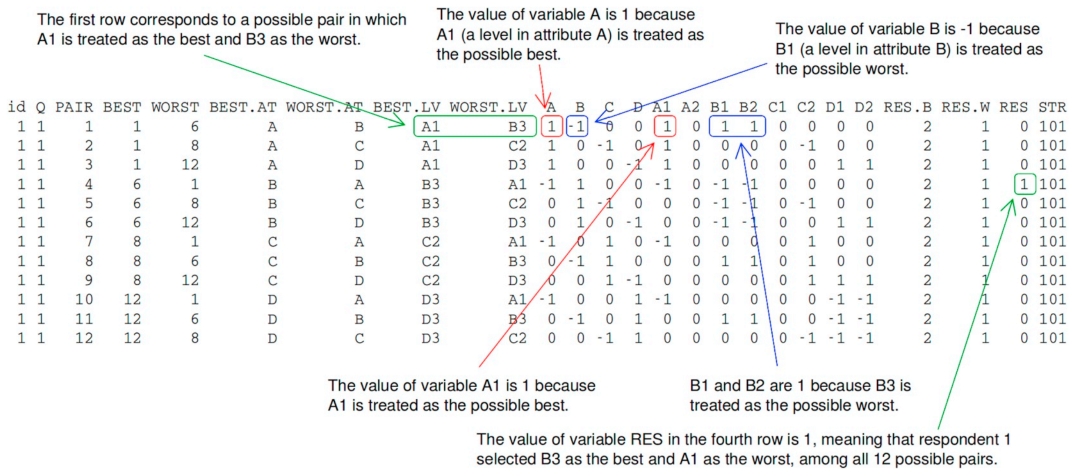
The response variables must be constructed such that the best levels alternate with the worst by question. For example, when there are nine BWS questions, the variables are $B1, W1, B2, W2, \ldots, B9$, and $W9$. Here, B_i and W_i are the levels selected as the best and worst for the *i*-th question. The row numbers of the levels selected as the best and worst are stored in the response variables. For example, suppose a respondent was asked to answer the BWS question shown in the bottom panel of Fig. 1, and that the respondent selected A1 (level in the *first* row) as the best and C2 (level in the *third* row) as the worst. The response variables B1 and W1, corresponding to the respondent's answer to this question, take values of 1 (= the level in the *first* row) and 3 (= the level in the *third* row).

An OA used to construct the BWS questions is assigned to the argument `choice.sets`, while the attributes and their levels are assigned to the argument `attribute.levels` in list format. The order of questions in `choice.sets` has to be the same as that in the respondent dataset.

Combinations of the `base.attribute`, `base.level`, `reverse`, and `model` arguments set the structure of the resultant dataset, which differs according to the model variants. These arguments are described in detail later.

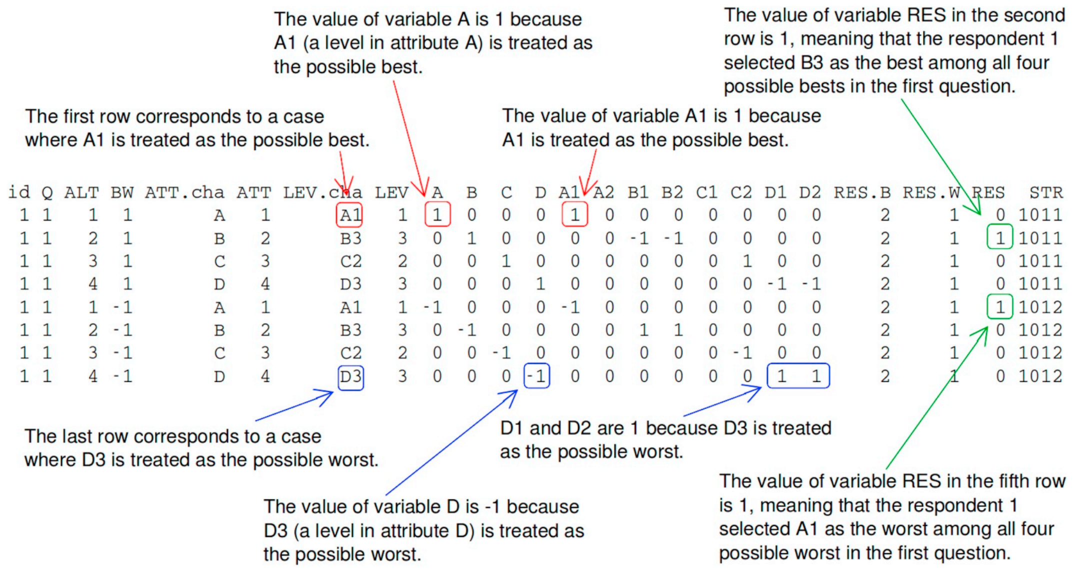
The function `bws2.dataset()` returns a dataset in data frame format for the paired model variant or one for the marginal (sequential) model variant. Figs. 3 and 4 show part of an example dataset for the paired model and for the marginal model corresponding to the profile shown in the bottom of Fig. 1, respectively. The profile shown in the bottom of Fig. 1, which corresponds to the first among nine Case 2 BWS questions, has four levels A1, B3, C2, and D3. Therefore, the number of all possible pairs of the four levels is 12. The 12 rows shown in Fig. 3 correspond to part of the dataset, and represent the 12 possible pairs in the first profile for respondent 1: the value of the variable `id` (the respondent's identification number) is `1` and the value of `Q` (the serial number of the BWS questions) is `1`. From the 8th and 9th columns (the variables `BEST.LV` and `WORST.LV`, which show the best and worst levels in a possible pair), we can see that the first row in the dataset indicates a possible pair in which A1 is treated as the best and B3 is treated as the worst. Accordingly, the value of the attribute variable `A` (10th column) and that of `B` (11th column) are `1` and `-1`, respectively, and the values of the remaining attribute variables such as `C` and `D` (12th and 13th columns) are `0`. Similarly, the value of the level variable `A1` is `1`, the values of level variables `B1` and `B2` are `1` (level variables are effect-coded variables and the sign of the values are reversed for the worst levels), and the values of the remaining level variables are `0`. The variable `RES` is a response variable, taking a value of `1` if a possible pair is actually selected by respondents from all the possible pairs, and `0` otherwise. In this example, the value of `RES` is `1` in the 4th row, and the remaining 11 rows are `0`: so respondent 1 selected a pair in which B3 is the best and A1 is the worst.

Fig. 4 shows part of the example dataset for the marginal model. The number of rows corresponding to a Case 2 BWS question for a respondent is $2 \times K$: the 8 rows of the dataset correspond to the first profile (`Q = 1`) for respondent 1 (`id = 1`). The first four rows show a set of four possible bests for the first question, for respondent 1, and the last four rows show a set of four possible worsts. The variable `BW` (4th column) takes a value of `1` for rows showing the possible best, and `−1` for rows showing the possible worst. The 7th column `LEV.cha` shows which levels are treated as the possible best or worst. The first row indicates that A1 is treated as the possible best, and thus, the value of the attribute variable `A` (9th column) is `1`, the value of the level variable `A1` (13th column) is `1`, and the values of the remaining attribute and level variables are `0`. From the value of the response variable `RES` (23rd column), respondent 1



id	Q	PAIR	BEST	WORST	BEST.AT	WORST.AT	BEST.LV	WORST.LV	A	B	C	D	A1	A2	B1	B2	C1	C2	D1	D2	RES.B	RES.W	RES	STR
1	1	1	1	6	A	B	A1	B3	1	-1	0	0	1	0	1	1	0	0	0	0	2	1	0	101
1	1	2	1	8	A	C	A1	C2	1	0	-1	0	1	0	0	0	0	-1	0	0	2	1	0	101
1	1	3	1	12	A	D	A1	D3	1	0	0	-1	1	0	0	0	0	0	1	1	2	1	0	101
1	1	4	6	1	B	A	B3	A1	-1	1	0	0	-1	0	-1	-1	0	0	0	0	2	1	1	101
1	1	5	6	8	B	C	B3	C2	0	1	-1	0	0	0	-1	-1	0	-1	0	0	2	1	0	101
1	1	6	6	12	B	D	B3	D3	0	1	0	-1	0	0	-1	-1	0	0	1	1	2	1	0	101
1	1	7	8	1	C	A	C2	A1	-1	0	1	0	-1	0	0	0	0	1	0	0	2	1	0	101
1	1	8	8	6	C	B	C2	B3	0	-1	1	0	0	0	1	1	0	1	0	0	2	1	0	101
1	1	9	8	12	C	D	C2	D3	0	0	1	-1	0	0	0	0	0	1	1	1	2	1	0	101
1	1	10	12	1	D	A	D3	A1	-1	0	0	1	-1	0	0	0	0	0	-1	-1	2	1	0	101
1	1	11	12	6	D	B	D3	B3	0	-1	0	1	0	0	1	1	0	0	-1	-1	2	1	0	101
1	1	12	12	8	D	C	D3	C2	0	0	-1	1	0	0	0	0	0	-1	-1	-1	2	1	0	101

Fig. 3. Example dataset for the paired model with both attribute and level variables (Model 1).



id	Q	ALT	BW	ATT.cha	ATT	LEV.cha	LEV	A	B	C	D	A1	A2	B1	B2	C1	C2	D1	D2	RES.B	RES.W	RES	STR
1	1	1	1	A	1	A1	1	1	0	0	0	1	0	0	0	0	0	0	0	2	1	0	1011
1	1	2	1	B	2	B3	3	0	1	0	0	0	0	-1	-1	0	0	0	0	2	1	1	1011
1	1	3	1	C	3	C2	2	0	0	1	0	0	0	0	0	0	1	0	0	2	1	0	1011
1	1	4	1	D	4	D3	3	0	0	0	1	0	0	0	0	0	0	-1	-1	2	1	0	1011
1	1	1	-1	A	1	A1	1	-1	0	0	0	-1	0	0	0	0	0	0	0	2	1	1	1012
1	1	2	-1	B	2	B3	3	0	-1	0	0	0	0	1	1	0	0	0	0	2	1	0	1012
1	1	3	-1	C	3	C2	2	0	0	-1	0	0	0	0	0	0	-1	0	0	2	1	0	1012
1	1	4	-1	D	4	D3	3	0	0	0	-1	0	0	0	0	0	0	1	1	2	1	0	1012

Fig. 4. Example dataset for the marginal model with both attribute and level variables (Model 3).

selected B3 as the best (`RES` = `1` in the second row and `0` in the first, third, and fourth rows) and A1 as the worst (`RES` = `1` in the 5th row and `0` in the 6th to 8th rows). For detailed descriptions of the variables in the resultant dataset, see the help file for the function `bws2.dataset()`, and Aizaki and Fogarty (2019).

The argument settings to create the dataset, the row structure of the resultant dataset, and the types of response, attribute, and level variables in the resultant dataset, including notes on setting independent variables for a discrete choice analysis, for each model are as follows.

- Model 1
  - Arguments are set as `model` = "`paired`" and `reverse` = `TRUE`, while a list of a base level in each attribute is assigned to `base.level`.
  - One row in the dataset shows a possible pair of level i selected as the best and level j selected as the worst ($i \neq j$). The number of possible pairs of the best and worst levels for each question is $K \times (K-1)$. Thus, the total number of rows for the resultant dataset is $N \times Q \times K \times (K-1)$, where N is the number of respondents and Q is the number of questions per respondent.
  - The response variable takes a value of 1 if a possible pair of the best and worst levels is actually selected by the respondents and 0 otherwise.

  - Dummy-coded attribute variables (attribute-specific constants) with reversed signs when the attributes are treated as the possible worst are created for all the attributes. An arbitrary attribute variable is omitted from the discrete choice analysis.
  - Effect-coded level variables are created according to base levels defined by the `base.level` argument.
- Model 2
  - Arguments are set as `model = "paired"` and `reverse = TRUE` or `FALSE`.
  - The row structure of dataset is the same as that for Model 1.
  - The response variable is the same as that for Model 1.
  - Although the attribute variables are created in the dataset, these are not used in this model.
  - Dummy-coded level variables are created for all levels. An arbitrary level variable is omitted from the discrete choice analysis.
- Model 3
  - Arguments are set as `model = "marginal"` and `reverse = TRUE`, while a list of a base level in each attribute is assigned to `base.level`.
  - One row in the dataset shows one possible level i selected as the best or worst. The number of possible levels selected as the best for each question is K and the number of possible levels selected as the worst for each question is also K. The total number of rows for the resultant dataset is therefore $N \times Q \times (K + K)$.
  - The response variable takes a value of 1 if a possible level is actually selected as the best or worst by the respondents and 0 otherwise.
  - Dummy-coded attribute variables with reversed signs when the attributes are treated as the possible worst are created for all the attributes. An arbitrary attribute variable is omitted from the discrete choice analysis.
  - Effect-coded level variables are created according to base levels defined by the `base.level` argument.
- Model 4
  - Arguments are set as `model = "marginal"` and `reverse = FALSE`.
  - The row structure of dataset is the same as that for Model 3.
  - The response variable is the same as that for Model 3.
  - Dummy-coded attribute variables are created for all the attributes. An arbitrary attribute variable is omitted from the discrete choice analysis.
  - Dummy-coded level variables are created for all levels. An arbitrary level variable for each attribute is omitted from the discrete choice analysis.
- Model 5
  - Arguments are set as `model = "marginal"` and `reverse = TRUE` or `FALSE`.
  - The row structure of dataset is the same as that for Model 3.
  - The response variable is the same as for Model 3.
  - Although the attribute variables are created in the dataset, these are not used in this model.
  - Dummy-coded level variables are created for all levels. An arbitrary level variable for each attribute is omitted from the discrete choice analysis.
- Model 6
  - Arguments are set as `model = "marginal"` and `reverse = TRUE` or `FALSE`.
  - The row structure of dataset is the same as that for Model 3.
  - The response variable is the same as for Model 3.
  - Although the attribute variables are created in the dataset, these are not used in this model.
  - Dummy-coded level variables are created for all levels. An arbitrary level variable is omitted from the discrete choice analysis.
- Model 7
  - Arguments are set as `model = "sequential"` and `reverse = FALSE`.
  - One row in dataset shows a possible level i selected as the best or worst. The number of possible levels selected as the best for each question is K, and the number of possible levels selected as the worst for each question is $K - 1$ because the level selected as the best is deleted from the worst choice set. The total number of rows for the resultant dataset is $N \times Q \times (K + K - 1)$.
  - The response variable is the same as for Model 3.
  - Dummy-coded attribute variables are created for all attributes. An arbitrary attribute variable is omitted form the discrete choice analysis.
  - Dummy-coded level variables are created for all levels. An arbitrary level variable for each attribute is omitted from the discrete choice analysis.
- Model 8
  - Arguments are set as `model = "sequential"` and `reverse = TRUE`, while a character showing a base attribute is assigned to `base.attribute` and a list of a base level in each attribute is assigned to `base.level`.
  - The row structure of dataset is the same as that for Model 7.
  - The response variable is the same as for Model 3.
  - Effect-coded attribute variables with reversed signs when the attributes are treated as the possible worst are created according to the base attribute defined by the `base.attribute` argument.
  - Effect-coded level variables are created according to base levels defined by the `base.level` argument.
- Model 9
  - Arguments are set as `model = "sequential"` and `reverse = TRUE` or `FALSE`, while a list of a base level in each attribute is

assigned to `base.level`.
– The row structure of dataset is the same as that for Model 7.
– The response variable is the same as for Model 3.
– Although the attribute variables are created in the dataset, these are not used in this model.
– Effect-coded level variables are created according to base levels defined by the `base.level` argument.

It should be noted that the users must select a model according to their assumptions regarding the respondents' behavior in Case 2 BWS questions.

5. Example application

This section is a shorter, modified version of the online tutorial for Case 2 BWS (Aizaki and Fogarty, 2019), and provides an overview of how to use support.BWS2 in conjunction with other packages to implement Case 2 BWS in **R**. The example highlights the creation of datasets for each of the nine models, and the subsequent analysis of these datasets. The online tutorial provides complete details for an original and full-length example for Models 1 and 3. Complete **R** code for the example is available from the online supplementary file.

The following packages are needed for the example:

```
library("support.BWS2"); library("survival"); library("DoE.base")
```

Note that we printed out results to four digits in the example, and to set the same appearance format in **R**, execute:

```
options(digits = 4)
```

Step 1 involves defining the choice situation. The example uses a survey of preferences for agritourism activities. Agritourism refers to the activities offered by farms to visitors, such as hands-on farm work or outdoor recreation activities. In the BWS questions, respondents were asked to evaluate agritourism packages provided by dairy farms (ranches). The agritourism package consists of the following four types of activities, each with three activity items (variable names are in parentheses):

- Hands-on ranch chores (`chore`): milking a cow (`milking`); feeding a cow (`feeding`); and nursing a calf (`nursing`).
- Hands-on food processing (`food`): butter making (`butter`); ice-cream making (`ice`); and creamy caramel making (`caramel`).
- Hands-on craft making (`craft`): making a product from wool (`wool`); making a product from wood (`wood`); and making a product from pressed flowers (`flower`).
- Outdoor activities (`outdoor`): horse riding (`horse`); tractor riding (`tractor`); and walking with cows (`cow`).

Step 2 is the creation of the profiles for the survey. The agritourism package has four attributes (activities), each with three levels (activity items), so a 3^4 OA is needed. This is created using the function `oa.design()`. Although the function has many arguments, only `nlevels` and `randomize` are used in this example. The argument `nlevels` sets the number of levels corresponding to the OA that you wish to create. The `randomize` argument specifies whether the order of the rows in the resultant design is randomized. The following lines of code were used to create the 3^4 OA and then store it in the `des` object:

```
(des <- data.matrix(oa.design(nl = c(3,3,3,3), randomize = FALSE)))

##   A B C D
## 1 1 1 1 1
## 2 1 2 3 2
## 3 1 3 2 3
## 4 2 1 3 3
## 5 2 2 2 1
## 6 2 3 1 2
## 7 3 1 2 2
## 8 3 2 1 3
## 9 3 3 3 1
```

The resultant OA is a matrix A where the element a_{ij} denotes, for profile i, the level for attribute j. In the example, $i = 1$ to 9 and $j = 1$ (A) to 4 (D).

Step 3 involves using the `bws2.questionnaire()` function to convert the `oa.design()` output into a format that matches the question format for a Case 2 BWS survey. In preparation for this step, the names of the attributes (activities) and levels (activity items) are stored in the list `attr.lev`:

```
attr.lev <- list(
  chore   = c("milking", "feeding", "nursing"),
  food    = c("butter",  "ice",     "caramel"),
  craft   = c("wool",    "wood",    "flower"),
  outdoor = c("horse",   "tractor", "cow"))
```

The series of Case 2 BWS questions is converted from the OA using the **R** code below. An OA is assigned to the argument `choice.sets`. Attributes and their levels are assigned to the argument `attribute.levels` in list format. The argument `position` is used to change the position of the attribute column in the resultant questions.

```
bws2.questionnaire(choice.sets = des, attribute.levels = attr.lev,
  position = "left")


Q1
 Attribute Best Worst
 milking   [ ]  [ ]
 butter    [ ]  [ ]
 wool      [ ]  [ ]
 horse     [ ]  [ ]


 ...


Q9
 Attribute Best Worst
 nursing   [ ]  [ ]
 caramel   [ ]  [ ]
 flower    [ ]  [ ]
 horse     [ ]  [ ]
```

The responses to the Case 2 BWS questions were collected (Step 4) and stored in the dataset `agritourism` in the support.BWS2 package (Step 5). The raw dataset is loaded into the current session as follows:

```
data("agritourism", package = "support.BWS2")
dim(agritourism)

## [1] 240  21

names(agritourism)

##  [1] "id"     "b1"     "w1"     "b2"     "w2"     "b3"     "w3"
##  [8] "b4"     "w4"     "b5"     "w5"     "b6"     "w6"     "b7"
## [15] "w7"     "b8"     "w8"     "b9"     "w9"     "gender" "age"
```

The dataset contains 240 observations with 21 variables. The `id` is the respondents' id variable. Response variables `b1–w9` correspond to the Case 2 BWS questions: `b` denotes a response as the best, w denotes a response as the worst, and the numbers `1` to `9` indicate the BWS question number. The `gender` is the respondents' gender variable: `1` denotes male, `2` denotes female. The `age` is the respondents' age category variable: `2, 3, 4`, and `5` correspond to 20s, 30s, 40s, and 50s, respectively.

Step 6 is to create a dataset for the analysis. In preparation, the names of the response variables used in the respondent dataset agritourism are stored in the vector `response.vars`:

```
response.vars <- colnames(agritourism)[2:19]
```

Furthermore, the base level in each attribute for Models 1, 3, 8, and 9, which use effect-coded level variables, is assumed to be `nursing` for attribute `chore`, `caramel` for attribute `food`, `flower` for attribute `craft`, and `cow` for attribute `outdoor`. These base levels are stored in the object `base.lev` in list format:

```
base.lev <- list(
  chore = c("nursing"), food = c("caramel"), craft = c("flower"),
  outdoor = c("cow"))
```

Information regarding the base levels for models with effect-coded level variables is used to create the dataset for analysis: therefore, the list object mentioned above is set in advance. On the other hand, information regarding the base levels for models with dummy-coded level variables will be set when estimating models.

The datasets for the nine models are created using the function `bws2.dataset()` and then stored in the nine objects `model1.dat` to `model9.dat`, respectively:

```
model1.dat <- bws2.dataset(
  data = agritourism, id = "id", response = response.vars,
  choice.sets = des, attribute.levels = attr.lev,
  reverse = TRUE, base.level = base.lev, model = "paired")
model2.dat <- bws2.dataset(
  data = agritourism, id = "id", response = response.vars,
  choice.sets = des, attribute.levels = attr.lev,
  reverse = TRUE, model = "paired")
model3.dat <- bws2.dataset(
  data = agritourism, id = "id", response = response.vars,
  choice.sets = des, attribute.levels = attr.lev,
  reverse = TRUE, base.level = base.lev, model = "marginal")
model4.dat <- model5.dat <- model6.dat <- bws2.dataset(
  data = agritourism, id = "id", response = response.vars,
  choice.sets = des, attribute.levels = attr.lev,
  reverse = FALSE, model = "marginal")
model7.dat <- bws2.dataset(
  data = agritourism, id = "id", response = response.vars,
  choice.sets = des, attribute.levels = attr.lev,
  reverse = FALSE, model = "sequential")
model8.dat <- bws2.dataset(
  data = agritourism, id = "id", response = response.vars,
  choice.sets = des, attribute.levels = attr.lev,
  reverse = TRUE, base.level = base.lev,
  base.attribute = "craft", model = "sequential")
model9.dat <- bws2.dataset(
  data = agritourism, id = "id", response = response.vars,
  choice.sets = des, attribute.levels = attr.lev,
  reverse = FALSE, base.level = base.lev, model = "sequential")
```

As explained above, the structure of the resultant datasets differs across model variants. For example, the dataset `model1.dat` consists of 25,920 rows (= 240 respondents × 9 questions × 12 possible pairs per question) and 27 columns (variables), while the dataset `model7.dat` consists of 15,120 rows (= 240 respondents × 9 questions × (4 possible best levels per question + 3 possible worst levels per question)) and 30 columns:

```
dim(model1.dat)

## [1] 25920    27

dim(model7.dat)

## [1] 15120    30
```

The final step is to analyze the responses using the counting and modeling approaches. We calculate the BWS scores using the `bws2.count()` function with any dataset created above (Here, we use the dataset `model1.dat`):

```
scores <- bws2.count(model1.dat)
```

The resultant scores, which are stored in the object `scores`, contain B, W, BW, standardized BW scores, and respondent characteristics variables for each respondent. Therefore, the `scores` object consists of 240 rows (respondents) and 51 columns (the id variable; 12 B, 12 W, 12 BW, and 12 standardized BW score variables; and 2 respondent characteristic variables):

```
dim(scores)

## [1] 240  51
```

Methods `sum()` and `barplot()` are available for the output from the function `bws2.count()`, which is an object of the S3 class `bws2.count`, inheriting from the S3 class `data.frame`. As such, we can aggregate the scores for each level among all respondents using the function `sum()` (see Aizaki and Fogarty (2019) for the resultant bar plots from the `barplot()`):

```
sum(scores, "level")

##            B   W   BW     stdBW
## milking  131 129    2  0.002778
## feeding  106 156  -50 -0.069444
## nursing  157 109   48  0.066667
## butter   358  51  307  0.426389
## ice      354  58  296  0.411111
## caramel  349  65  284  0.394444
## wool      79 297 -218 -0.302778
## wood      63 362 -299 -0.415278
## flower    34 428 -394 -0.547222
## horse    272  75  197  0.273611
## tractor  135 235 -100 -0.138889
## cow      122 195  -73 -0.101389
```

Finally, we fit the nine models to the Case 2 BWS responses on the basis of the conditional logit model specification. The function `clogit()` in the survival package is used to fit the models. Although this function has many arguments, only two are used in this example. The call to the function is:

```
clogit(formula, data)
```

where `formula` is a model formula and `data` is a dataframe containing the variables used in the model formula. The structure of the model formula is similar to that for a linear regression function, `lm()`: the dependent variable is set on the left-hand side of $\sim$, whereas the independent variables are set on the right-hand side according to a systematic component of the utility function to be fitted. Unlike `lm()`, however, a stratification variable must be added to the end of the right-hand side using the `strata()` function. The stratification variable used in `clogit()` is automatically generated as a variable `STR` and stored in the dataset created using the function `bws2.dataset()`. The systematic component of the utility function for the example under the nine models is as follows:

- Models 1, 3, 4, 7, and 8

$$v = \beta_1 chore + \beta_2 food + \beta_3 outdoor + \beta_4 milking + \beta_5 feeding + \beta_6 butter + \beta_7 ice + \beta_8 wool + \beta_9 wood + \beta_{10} horse + \beta_{11} tractor, \tag{14}$$

- Models 2 and 6

$$v = \beta_1 milking + \beta_2 feeding + \beta_3 nursing + \beta_4 butter + \beta_5 ice + \beta_6 caramel + \beta_7 wool + \beta_8 wood + \beta_9 horse + \beta_{10} tractor + \beta_{11} cow, \tag{15}$$

- Models 5 and 9

$$v = \beta_1 milking + \beta_2 feeding + \beta_3 butter + \beta_4 ice + \beta_5 wool + \beta_6 wood + \beta_7 horse + \beta_8 tractor, \tag{16}$$

where *chore*, *food*, and *outdoor* are attribute variables; and *milking*, *feeding*, *nursing*, *butter*, *ice*, *caramel*, *wool*, *wood*, *horse*, *tractor*, and *cow* are level variables.

The model formulas for `clogit()`, corresponding to the systematic component mentioned above, are described as:

```
model1.mf <- model3.mf <- model4.mf <- model7.mf <- model8.mf <-
  RES ~ chore + food + outdoor + milking + feeding + butter + ice +
        wool + wood + horse + tractor + strata(STR)
model2.mf <- model6.mf <-
  RES ~ milking + feeding + nursing + butter + ice + caramel +
        wool + wood + horse + tractor + cow + strata(STR)
model5.mf <- model9.mf <-
  RES ~ milking + feeding + butter + ice +
        wool + wood + horse + tractor + strata(STR)
```

We fit the nine models using `clogit()` with the datasets and model formulas set above (the detailed outputs are omitted below except for Model 1):

```
model1.out <- clogit(formula = model1.mf, data = model1.dat)
model2.out <- clogit(formula = model2.mf, data = model2.dat)
model3.out <- clogit(formula = model3.mf, data = model3.dat)
model4.out <- clogit(formula = model4.mf, data = model4.dat)
model5.out <- clogit(formula = model5.mf, data = model5.dat)
model6.out <- clogit(formula = model6.mf, data = model6.dat)
model7.out <- clogit(formula = model7.mf, data = model7.dat)
model8.out <- clogit(formula = model8.mf, data = model8.dat)
model9.out <- clogit(formula = model9.mf, data = model9.dat)
model1.out
```

```
## Call:
## clogit(formula = model1.mf, data = model1.dat)
##
##            coef exp(coef) se(coef)    z      p
## chore     0.735     2.085    0.042 17.6 <2e-16
## food      1.444     4.237    0.045 32.1 <2e-16
## outdoor   0.758     2.135    0.042 18.1 <2e-16
## milking   0.016     1.016    0.047  0.3    0.7
## feeding  -0.169     0.844    0.047 -3.6  4e-04
## butter    0.041     1.042    0.049  0.8    0.4
## ice       0.009     1.009    0.049  0.2    0.9
## wool      0.309     1.362    0.048  6.4  2e-10
## wood      0.024     1.024    0.049  0.5    0.6
## horse     0.629     1.875    0.048 13.1 <2e-16
## tractor  -0.363     0.695    0.047 -7.6  2e-14
##
## Likelihood ratio test=1432  on 11 df, p=<2e-16
## n= 25920, number of events= 2160
```

Table 10 shows estimates for the nine models, including estimates of the base levels. Estimates of dummy-coded base level variables are zero, while effect-coded ones are the negative of sum of estimates of the relevant level variables. The detailed outputs of the nine models are available by executing the **R** code in the online supplementary file.

6. Discussion and conclusion

As a general rule, the dataset structure required by SP methods is different to that used in traditional cross-section data analysis. Mapping the dataset to the required format is therefore complicated, especially for non-expert users. For Case 2 BWS the task is especially complicated as there are many model variants, and sub-variants, and the dataset structure differs across the model variants. As the use of SP methods in general increases, the number of non-expert users who will face research questions where Case 2 BWS is the most appropriate method will also increase. For this group of people it may be difficult for them to understand how the dataset for Case 2 BWS is constructed, and how a discrete choice model is applied for the analysis, without support. The difficulty of dataset creation for Case 2 BWS may be one reason limiting the application of Case 2 BWS in research fields other than health economics.

This paper has shown how **R** functions can be used to implement Case 2 BWS. The package support.BWS2 provides two main functions: one is for constructing the dataset for Case 2 BWS analysis from the choice sets and the responses to the questions; and the other is for calculating BW scores from the dataset. The former function converts the choice sets and the responses into the required data format, according to the users' requirement. The latter function calculates best, worst, and best-minus-worst scores for each level by respondent from the dataset generated using the former function, and then returns a dataset containing variables regarding these

Table 10
Estimates in the nine models. The value of NA in the table means a model has no independent variables corresponding to NA.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
chore	0.7349	NA	0.9280	−0.3290	NA	NA	−0.2759	0.0050	NA
food	1.4439	NA	1.8281	−0.0802	NA	NA	0.1890	0.8284	NA
craft	0.0000	NA	0.0000	0.0000	NA	NA	0.0000	−0.8361	NA
outdoor	0.7583	NA	0.9542	−0.1154	NA	NA	−0.0233	0.0028	NA
milking	0.0158	1.0832	0.0215	0.1447	0.1481	1.3439	0.2040	0.0229	0.0092
feeding	−0.1695	0.8979	−0.2221	−0.1080	−0.0739	1.1002	−0.0361	−0.1961	−0.1596
nursing	0.1536	1.2210	0.2006	0.0000	0.0000	1.5230	0.0000	0.1732	0.1504
butter	0.0414	1.8179	0.0482	1.1037	1.1411	2.2706	0.9141	0.0508	0.0494
ice	0.0087	1.7851	0.0118	1.0672	1.1036	2.2343	0.8788	0.0068	−0.0022
caramel	−0.0501	1.7263	−0.0600	0.0000	0.0000	2.1625	0.0000	−0.0576	−0.0472
wool	0.3089	0.6414	0.3730	−0.4611	−0.5331	0.7674	−0.4024	0.3419	0.2859
wood	0.0236	0.3561	0.0213	−0.7857	−0.8778	0.4157	−0.7173	0.0049	0.0070
flower	−0.3325	0.0000	−0.3944	0.0000	0.0000	0.0000	0.0000	−0.3468	−0.2929
horse	0.6288	1.7196	0.8101	0.9053	0.9135	2.1587	0.8656	0.7513	0.6361
tractor	−0.3633	0.7276	−0.4690	−0.2970	−0.2996	0.8796	−0.2642	−0.4587	−0.3824
cow	−0.2655	0.8253	−0.3411	0.0000	0.0000	1.0075	0.0000	−0.2926	−0.2537

scores and respondents' characteristics. The resultant dataset enables additional, valuable analysis: for example, respondents could be classified into groups via a cluster analysis of the scores; and these scores could be used as independent variables in a regression analysis of a respondent' behavior.

The package support.BWS2 and other packages related to experimental designs and discrete choice models can be used to construct an **R**-based platform to implement Case 2 BWS. This is helpful for the non-expert users because they only need to learn how to use one software platform to implement Case 2 BWS. An online tutorial explaining how **R** can be used to implement a Case 2 BWS study is also available.

This paper has also explained the model variants and the associated dataset structure in detail. These explanations also help non-expert users gain an understanding of the model formulation for Case 2 BWS analysis.

These works are based on **R** and are distributed/published under free licenses: a GNU General Public License is applied to the package support.BWS2; and a Creative Commons license is applied to the online tutorial. **R** is, of course, free software and runs on various operating systems including Windows, Mac, and Unix/Linux. Further, the **R** functions for creating datasets for Case 2 BWS analysis could also be of value to those who use software applications other than **R** for their analysis. This is because they can map their raw dataset to the appropriate format using an **R** function with the other tasks implemented using the software application they prefer.

The main limitation of the contribution is that users still have to write **R** code to use functions in the package support.BWS2. Some users who are not familiar with writing **R** code may still face some difficulties using the functions.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP25450341 and JP16K07886.

Appendix A

Papers that used Case 2 BWS in empirical studies were searched through the Web of Science on 16th August 2018. The search terms were as follows: field tag is set as TS = ("best-worst scaling" OR "best worst scaling" OR "best worst" OR "maxdiff" OR "max diff" OR "maxdiff scaling"); language is set as English; document type is set as article; and timespan is set as all years. Among 557 papers identified through the search, a total of 53 papers were identified as applications of Case 2 BWS in empirical studies. In the screening process, a Case 2 BWS paper is defined as a paper satisfying all the three conditions: 1) each profile is expressed using all attributes; 2) each profile has a level from each attribute; and 3) analysis of the responses to Case 2 BWS questions is not integrated with an analysis of responses to other SP questions, such as DCE questions. The identified papers were then divided into two groups: 44 papers belong to health or health-related areas; and eight papers belong to non-health areas. The following is a list of the two groups:

- Papers in health or health-related areas
  - Coast et al. (2006); Flynn et al. (2007); Coast et al. (2008); Flynn et al. (2008); Guenther et al. (2010); Flynn et al. (2010); Potoglou et al. (2011); Al-Janabi et al. (2011); Ratcliffe et al. (2012); Knox et al. (2012); Molassiotis et al. (2012); Flynn et al. (2013); Weisberg et al. (2013); Severin et al. (2013); Yoo and Doiron (2013); Peay et al. (2014); Whitty et al. (2014); Damery et al. (2014); Franco et al. (2015); Hollin et al. (2015); Jones et al. (2015); Flynn et al. (2015); dosReis et al. (2015); Ungar et al. (2015); Howell et al. (2016); Pollitt et al. (2016); Weernink et al. (2016); Ratcliffe et al. (2016a); Coast et al. (2016); Ratcliffe et al. (2016b); van Dijk et al. (2016); Lynd et al. (2016); Pooripussarakul et al. (2016); Weernink et al. (2016); Janssen et al. (2016); dosReis et al. (2017); Ng et al. (2017); Hollin et al. (2017); Meregaglia et al. (2017); Huynh et al. (2017); Weernink et al.

(2017); Krucien et al. (2017); Kok et al. (2018); and Ozawa et al. (2018).
- Papers in non-health areas
  - Dorow et al. (2009); Balbontin, D. Ortúzar, and Swait (2015); Soto et al. (2016); Waintrub et al. (2016); Kreye et al. (2016); Kreye et al. (2017); Soto et al. (2018); and Dawes et al. (2018).

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jocm.2019.100171.

References

Aizaki, Hideo, 2019. support.BWS2: basic functions for supporting an implementation of case 2 best-worst scaling. https://CRAN.R_project.org/package=support.BWS2 R Package version 0.2-2.
Aizaki, Hideo, Fogarty, James, 2019. An illustrative example of case 2 best–worst scaling. In: Non-Market Valuation with R. NMVR Project Team. http://lab.agr.hokudai.ac.jp/nmvr/.
Al-Janabi, Hareth, Flynn, Terry N., Coast, Joanna, 2011. Estimation of a preference-based carer experience scale. Med. Decis. Mak. 31 (3), 458–468. https://doi.org/10.1177/0272989X10381280.
Balbontin, C., Ortúzar, J. de D., Swait, J.D., 2015. A joint best–worst scaling and stated choice model considering observed and unobserved heterogeneity: an application to residential location choice. J. Choice Model. 16, 1–14. https://doi.org/10.1016/j.jocm.2015.09.002.
Carson, Richard T., 2012. Contingent Valuation: A Comprehensive Bibliography and History. Edward Elgar, Massachusetts, USA.
Cheung, Kei Long, Wijnen, Ben F.M., Hollin, Ilene L., Janssen, Ellen M., Bridges, John F., Silvia, M., Evers, A.A., Hiligsmann, Mickael, 2016. Using best–worst scaling to investigate preferences in health care. Pharmacoeconomics 34 (12), 1195–1209. https://doi.org/10.1007/s40273-016-0429-5.
Coast, Joanna, Flynn, Terry N., Natarajan, Lucy, Sproston, Kerry, Lewis, Jane, Louviere, Jordan J., Peters, Tim J., 2008. Valuing the ICECAP capability index for older people. Soc. Sci. Med. 67 (5), 874–882. https://doi.org/10.1016/j.socscimed.2008.05.015.
Coast, Joanna, Huynh, Elisabeth, Kinghorn, Philip, Flynn, Terry, 2016. Complex valuation: applying ideas from the complex intervention framework to valuation of a new measure for end-of-life care. Pharmacoeconomics 34 (5), 499–508. https://doi.org/10.1007/s40273-015-0365-9.
Coast, J., Salisbury, C., de Berker, D., Noble, A., Horrocks, S., Peters, T.J., Flynn, T.N., 2006. Preferences for aspects of a dermatology consultation. Br. J. Dermatol. 155 (2), 387–392. https://doi.org/10.1111/j.1365-2133.2006.07328.x.
Croissant, Yves, 2019. Mlogit: multinomial logit model. R Package version 0.4-1. https://CRAN.R_project.org/package=mlogit.
Damery, S., Biswas, M., Billingham, L., Barton, P., Al-Janabi, H., Grimer, R., 2014. Patient preferences for clinical follow-up after primary treatment for soft tissue sarcoma: a cross-sectional survey and discrete choice experiment. EJSO 40 (12), 1655–1661. https://doi.org/10.1016/j.ejso.2014.04.020.
Dawes, Leaundre C., Adams, Alison E., Escobedo, Francisco J., Soto, José R., 2018. Socioeconomic and ecological perceptions and barriers to urban tree distribution and reforestation programs. Urban Ecosystems. Springer, pp. 1–15. https://doi.org/10.1007/s11252-018-0760-z.
Dorow, Malte, Ben, Beardmore, Haider, Wolfgang, Arlinghaus, Robert, 2009. Using a novel survey technique to predict fisheries stakeholders' support for European Eel (Anguilla Anguilla L.) conservation programs. Biol. Conserv. 142 (12), 2973–2982. https://doi.org/10.1016/j.biocon.2009.07.029.
dosReis, Susan, Ng, Xinyi, Frosch, Emily, Reeves, Gloria, Cunningham, Charles, Bridges John, F.P., 2015. Using best-worst scaling to measure caregiver preferences for managing their child's adhd: a pilot study. Patient-Patient Cent. Outcomes Res. 8 (5), 423–431. https://doi.org/10.1007/s40271-014-0098-4.
dosReis, Susan, Park, Alex, Ng, Xinyi, Frosch, Emily, Reeves, Gloria, Cunningham, Charles, Janssen, Ellen M., Bridges, John F.P., 2017. Caregiver treatment preferences for children with a new versus existing attention-deficit/hyperactivity disorder diagnosis. J. Child Adolesc. Psychopharmacol. 27 (3), 234–242. https://doi.org/10.1089/cap.2016.0157.
Flynn, Terry N., Huynh, Elisabeth, Peters, Tim J., Al-Janabi, Hareth, Clemens, Sam, Moody, Alison, Coast, Joanna, 2015. Scoring the icecap-a capability instrument. Estimation of a UK general population tariff. Health Econ. 24 (3), 258–269. https://doi.org/10.1002/hec.3014.
Flynn, Terry Nicholas, Louviere, Jordan J., Peters, Tim J., Coast, Joanna, 2010. Using discrete choice experiments to understand preferences for quality of life. Variance-scale heterogeneity matters. Soc. Sci. Med. 70 (12), 1957–1965. https://doi.org/10.1016/j.socscimed.2010.03.008.
Flynn, Terry N., Louviere, Jordan J., Peters, Tim J., Coast, Joanna, 2007. Best–worst scaling: what it can do for health care research and how to do it. J. Health Econ. 26 (1), 171–189. https://doi.org/10.1016/j.jhealeco.2006.04.002.
Flynn, Terry N., Louviere, Jordan J., Peters, Tim J., Coast, Joanna, 2008. Estimating preferences for a dermatology consultation using best–worst scaling: comparison of various methods of analysis. BMC Med. Res. Methodol. 8 (1), 76. https://doi.org/10.1186/1471-2288-8-76.
Flynn, Terry N., Peters, Tim J., Coast, Joanna, 2013. Quantifying response shift or adaptation effects in quality of life by synthesising best-worst scaling and discrete choice data. J. Choice Model. 6 (March), 34–43. https://doi.org/10.1016/j.jocm.2013.04.004.
Franco, Marcia R., Howard, Kirsten, Sherrington, Catherine, Ferreira, Paulo H., Rose, John, Gomes, Juliana L., Ferreira, Manuela L., 2015. Eliciting older people's preferences for exercise programs: a best-worst scaling choice experiment. J. Physiother. 61 (1), 34–41. https://doi.org/10.1016/j.jphys.2014.11.001.
Grömping, Ulrike, 2018. R package DoE.base for factorial experiments. J. Stat. Softw. 85 (5), 1–41. https://doi.org/10.18637/jss.v085.i05.
Guenther, Oliver H., Kürstein, Beate, Riedel-Heller, Steffi G., König, Hans-Helmut, 2010. The role of monetary and nonmonetary incentives on the choice of practice establishment: a stated preference study of young physicians in Germany. Health Serv. Res. 45 (1), 212–229. https://doi.org/10.1111/j.1475-6773.2009.01045.x.
Hensher, David A., Rose, John M., Greene, William H., 2015. Applied Choice Analysis, second ed. Cambridge University Press, Cambridge, UK.
Hess, Stephane, Palma, David, 2019. Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application. J. Choice Model. https://doi.org/10.1016/j.jocm.2019.100170.
Hollin, Ilene L., Peay, Holly L., Apkon, Susan D., Bridges, John F.P., 2017. Patient-centered benefit-risk assessment in duchenne muscular dystrophy. Muscle Nerve 55 (5), 626–634. https://doi.org/10.1002/mus.25411.
Hollin, Ilene L., Peay, Holly L., Bridges, John F.P., 2015. Caregiver preferences for emerging duchenne muscular dystrophy treatments: a comparison of best-worst scaling and conjoint analysis. Patient-Patient Cent. Outcomes Res. 8 (1), 19–27. https://doi.org/10.1007/s40271-014-0104-x.
Howell, Martin, Wong, Germaine, Rose, John, Tong, Allison, Craig, Jonathan C., Howard, Kirsten, 2016. Eliciting patient preferences, priorities and trade-offs for outcomes following kidney transplantation: a pilot best–worst scaling survey. BMJ Open 6 (1). https://doi.org/10.1136/bmjopen-2015-008163.
Huynh, Elisabeth, Coast, Joanna, Rose, John, Kinghorn, Philip, Flynn, Terry, 2017. Values for the ICECAP-supportive care measure (ICECAP-SCM) for use in economic evaluation at end of life. Soc. Sci. Med. 189 (September), 114–128. https://doi.org/10.1016/j.socscimed.2017.07.012.
Janssen, Ellen M., Segal, Jodi B., Bridges, John F.P., 2016. A framework for instrument development of a choice experiment: an application to type 2 diabetes. Patient-Patient Cent. Outcomes Res. 9 (5), 465–479. https://doi.org/10.1007/s40271-016-0170-3.
Jones, Leanne G., Hawkins, Guy E., Brown, Scott D., 2015. Using best-worst scaling to improve psychological service delivery: an innovative tool for psychologists in organized care settings. Psychol. Serv. 12 (1), 20–27. https://doi.org/10.1037/ser0000011.
Knox, Stephanie A., Viney, Rosalie C., Street, Deborah J., Haas, Marion R., Fiebig, Denzil G., Weisberg, Edith, Bateson, Deborah, 2012. What's good and bad about contraceptive products? A best-worst attribute experiment comparing the values of women consumers and GPs. Pharmacoeconomics 30 (12), 1187–1202. https://doi.org/10.2165/11598040-000000000-00000.
Kok, Marlies M., Weernink, Marieke G.M., von Birgelen, Clemens, Fens, Anneloes, van der Heijden, Liefke C., van Til, Janine A., 2018. Patient preference for radial versus femoral vascular access for elective coronary procedures: the PREVAS study. Cathet. Cardiovasc. Interv. 91 (1), 17–24. https://doi.org/10.1002/ccd.27039.
Kreye, Melissa M., Adams, Damian C., Escobedo, Francisco J., Soto, José R., 2016. Does policy process influence public values for forest-water resource protection in

Florida? Ecol. Econ. 129, 122–131. https://doi.org/10.1016/j.ecolecon.2016.06.007.
Kreye, Melissa M., Pienaar, Elizabeth F., Soto, José R., Adams, Damian C., 2017. Creating voluntary payment programs: effective program design and ranchers' willingness to conserve Florida panther habitat. Land Econ. 93 (3), 459–480. https://doi.org/10.3368/le.93.3.459.
Krucien, Nicolas, Watson, Verity, Ryan, Mandy, 2017. Is best-worst scaling suitable for health state valuation? A comparison with discrete choice experiments. Health Econ. 26 (12), E1–E16. https://doi.org/10.1002/hec.3459.
Louviere, Jordan J., Flynn, Terry N., Marley, A.A.J., 2015. Best–Worst Scaling: Theory, Methods and Applications. Cambridge University Press, Cambridge, UK.
Lynd, Larry D., Anthony, Traboulsee, Marra, Carlo A., Mittmann, Nicole, Evans, Charity, Li, Kathy H., Carter, Melanie, Hategekimana, Celestin, 2016. Quantitative analysis of multiple sclerosis patients' preferences for drug treatment: a best-worst scaling study. Ther. Adv. Neurol. Disord. 9 (4), 287–296. https://doi.org/10.1177/1756285616648060.
Meregaglia, Michela, Cairns, John, Alfieri, Salvatore, Favales, Federica, Mazzitelli, Daniela, Orlandi, Ester, Licitra, Lisa, Bossi, Paolo, 2017. Eliciting preferences for clinical follow-up in patients with head and neck cancer using best-worst scaling. Value Health 20 (6), 799–808. https://doi.org/10.1016/j.jval.2017.01.012.
Molassiotis, Alex, Emsley, Richard, Ashcroft, Darren, Caress, Ann, Ellis, Jackie, Wagland, Richard, Bailey, Chris D., et al., 2012. Applying best-worst scaling methodology to establish delivery preferences of a symptom supportive care intervention in patients with lung cancer. Lung Cancer 77 (1), 199–204. https://doi.org/10.1016/j.lungcan.2012.02.001.
Mühlbacher, Axel C., Kaczynski, Anika, Zweifel, Peter, Reed Johnson, F., 2016. Experimental measurement of preferences in health and healthcare using best–worst scaling: an overview. Health Econ. Rev. 6 (1), 2. https://doi.org/10.1186/s13561-015-0079-x.
Ng, Xinyi, Bridges, John F.P., Ross, Melissa M., Frosch, Emily, Reeves, Gloria, Cunningham, Charles E., dosReis, Susan, 2017. A latent class Analysis to identify variation in caregivers' preferences for their child's attention-deficit/hyperactivity disorder treatment: do stated preferences match current treatment? Patient-Patient Cent. Outcomes Res. 10 (2), 251–262. https://doi.org/10.1007/s40271-016-0202-z.
Ozawa, Sachiko, Zhou, Mo, Wonodi, Chizoba, Chen, Hui-Han, John, F., Bridges, P., 2018. Parents' preferences for interventions to improve childhood immunization uptake in northern Nigeria. Vaccine 36 (20), 2833–2841. https://doi.org/10.1016/j.vaccine.2018.03.073.
Peay, Holly L., Hollin, Ilene, Fischer, Ryan, Bridges, John F.P., 2014. A community-engaged approach to quantifying caregiver preferences for the benefits and risks of emerging therapies for duchenne muscular dystrophy. Clin. Ther. 36 (5), 624–637. https://doi.org/10.1016/j.clinthera.2014.04.011.
Pollitt, Alexandra, Potoglou, Dimitris, Patil, Sunil, Peter, Burge, Guthrie, Susan, King, Suzanne, Wooding, Steven, Grant, Jonathan, 2016. Understanding the relative valuation of research impact: a best-worst scaling experiment of the general public and biomedical and health researchers. BMJ Open 6 (8). https://doi.org/10.1136/bmjopen-2015-010916.
Pooripussarakul, Siriporn, Riewpaiboon, Arthorn, Bishai, David, Muangchana, Charung, Tantivess, Sripen, 2016. What criteria do decision makers in Thailand use to set priorities for vaccine introduction? BMC Public Health 16. https://doi.org/10.1186/s12889-016-3382-5.
Potoglou, Dimitris, Peter, Burge, Flynn, Terry, Netten, Ann, Malley, Juliette, Forder, Julien, John, E., Brazier, 2011. Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. Soc. Sci. Med. 72 (10), 1717–1727. https://doi.org/10.1016/j.socscimed.2011.03.027.
Ratcliffe, Julie, Flynn, Terry, Terlich, Frances, Stevens, Katherine, Brazier, John, Sawyer, Michael, 2012. Developing adolescent-specific health state values for economic evaluation. Pharmacoeconomics 30 (8), 713–727. https://doi.org/10.2165/11597900-000000000-00000.
Ratcliffe, Julie, Huynh, Elisabeth, Chen, Gang, Stevens, Katherine, Swait, Joffre, Brazier, John, Sawyer, Michael, Roberts, Rachel, Flynn, Terry, 2016a. Valuing the child health utility 9D: using profile case best worst scaling methods to develop a new adolescent specific scoring algorithm. Soc. Sci. Med. 157 (May), 48–59. https://doi.org/10.1016/j.socscimed.2016.03.042.
Ratcliffe, Julie, Huynh, Elisabeth, Stevens, Katherine, Brazier, John, Sawyer, Michael, Flynn, Terry, 2016b. Nothing about us without us? A comparison of adolescent and adult health-state values for the child health utility-9d using profile case best-worst scaling. Health Econ. 25 (4), 486–496. https://doi.org/10.1002/hec.3165.
R Core Team, 2019. R: A Language and Environment for Statistical Computing. Vienna, Austria. R Foundation for Statistical Computing. https://www.R-project.org/.
Sarrias, Mauricio, Daziano, Ricardo, 2017. Multinomial logit models with continuous and discrete individual heterogeneity in R: the gmnl package. J. Stat. Softw. 79 (2), 1–46. https://doi.org/10.18637/jss.v079.i02.
Severin, Franziska, Schmidtke, Jörg, Mühlbacher, Axel, Rogowski, Wolf H., 2013. Eliciting preferences for priority setting in genetic testing: a pilot study comparing best-worst scaling and discrete-choice experiments. Eur. J. Hum. Genet. 21 (11), 1202–1208. https://doi.org/10.1038/ejhg.2013.36.
Soto, José R., Adams, Damian C., Escobedo, Francisco J., 2016. Landowner attitudes and willingness to accept compensation from forest carbon offsets: application of best–worst choice modeling in Florida USA. For. Policy Econ. 63, 35–42. https://doi.org/10.1016/j.forpol.2015.12.004.
Soto, José R., Escobedo, Francisco J., Khachatryan, Hayk, Adams, Damian C., 2018. Consumer demand for urban forest ecosystem services and disservices: examining trade-offs using choice experiments and best-worst scaling. Ecosyst. Serv. 29, 31–39. https://doi.org/10.1016/j.ecoser.2017.11.009.
Therneau, Terry, 2018. Survival: A Package for Survival Analysis in S. R package verion 2.43-3. https://CRAN.R_project.org/package=survival.
Therneau, Terry M., Grambsch, Patricia M., 2000. Modeling Survival Data: Extending the Cox Model. Springer, New York.
Ungar, Wendy J., Hadioonzadeh, Anahita, Najafzadeh, Mehdi, Tsao, Nicole W., Sharon, Dell, Lynd, Larry D., 2015. Parents and adolescents preferences for asthma control: a best-worst scaling choice experiment using an orthogonal main effects design. BMC Pulm. Med. 15. https://doi.org/10.1186/s12890-015-0141-9.
van Dijk, Joris, D., Groothuis-Oudshoorn, Catharina G.M., Marshall, Deborah A., IJzerman, Maarten J., 2016. An empirical comparison of discrete choice experiment and best-worst scaling to estimate stakeholders' risk tolerance for hip replacement surgery. Value Health 19 (4), 316–322. https://doi.org/10.1016/j.jval.2015.12.020.
Waintrub, Natan, Greene, Margarita, Dios Ortúzar, Juan de, 2016. Designing incentive packages for increased density and social inclusion in the neighbourhood of mass transit stations. Habitat Int. 55 (July), 133–147. https://doi.org/10.1016/j.habitatint.2016.03.006.
Weernink, Marieke G.M., Groothuis-Oudshoorn, Catharina G.M., IJzerman, Maarten J., van Til, Janine A., 2016. Valuing treatments for Parkinson disease incorporating process utility: performance of best-worst scaling, time trade-off, and visual analogue scales. Value Health 19 (2), 226–232. https://doi.org/10.1016/j.jval.2015.11.011.
Weernink, Marieke G.M., van Til, Janine A., Groothuis-Oudshoorn, Catharina G.M., IJzerman, Maarten J., 2017. Patient and public preferences for treatment attributes in Parkinson's disease. Patient-Patient Cent. Outcomes Res. 10 (6), 763–772. https://doi.org/10.1007/s40271-017-0247-7.
Weisberg, Edith, Bateson, Deborah, Knox, Stephanie, Haas, Marion, Viney, Rosalie, Street, Deborah, Fiebig, Denzil, 2013. Do women and providers value the same features of contraceptive products? Results of a best-worst stated preference experiment. Eur. J. Contracept. Reprod. Health Care 18 (3), 181–190. https://doi.org/10.3109/13625187.2013.777830.
Whitty, Jennifer A., Ratcliffe, Julie, Chen, Gang, Scuffham, Paul A., 2014. Australian public preferences for the funding of new health technologies: a comparison of discrete choice and profile case best-worst scaling methods. Med. Decis. Mak. 34 (5), 638–654. https://doi.org/10.1177/0272989X14526640.
Yoo, Hong Il, Doiron, Denise, 2013. The use of alternative preference elicitation methods in complex discrete choice experiments. J. Health Econ. 32 (6), 1166–1179. https://doi.org/10.1016/j.jhealeco.2013.09.009.