Best Worst Scaling workshop

Michael Burton
School of Agriculture and Environment
University of Western Australia
May 2021

University of
South Australia

Institutions to support intensification, integrated decision making and inclusiveness in agriculture in the East Gangetic Plain

(ACIAR Project # LWR-2018-104)

Funding Organisation & Project Partners

Australian Centre for International Agricultural Research

Australian Aid

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI

IIDS
Institute for Integrated Development Studies 1979

THE UNIVERSITY OF WESTERN AUSTRALIA

CI's

Prof Lin Crase (Project Leader)
Professor Mohammad Jahangir Alam
Dr Avinash Kishore
Dr Bishnu Dev Pant

The structure: 5 sessions

1. Conceptual issues and design
2. Models of choice and ways to analysis
3. Interpretation/presentation of results
4. Anchors and absolute rankings, and other issues
5. Resources and software to implement

Conceptual issues

- Best Worst Scaling: a method to generate a ranking of things
  - If you want to quantify the extent of differences between items
  - If you have a large number of items
- More efficient and reliable than other methods (i.e. ratings: Chrzan et al 2006, Burton et al 2021))

- NB: there are three different types of BWS
- Case 1: object
- Case 2: profile
- Case 3: multi-profile

Case 1: object

- If you want to rate a number of discrete objects
- Create subsets of items and select “best” and “worst”

1. Please choose your most preferred and least preferred service from your water provider

Least Preferred		Most Preferred
○	Dealing with customers enquiries and complaints, resolving satisfaction of customers (e.g. swifter response with friendly personnel)	○
○	Delivering uninterrupted and high-quality water and sewerage service (e.g. ensure delivery of the core service despite population growth and climate change)	○
○	Focusing on small and decentralized solutions (e.g. more regular maintenance to prevent burst and leaks, harvesting storm water)	○
○	Be proactive and not reactive with large scale projects (e.g. having long-term perspective with minimum political intervention on large assets such as desalination and recycled water)	○

Khosroshahi, S., Crase, L., Cooper, B., Burton, M. (2021) Matching customers’ preferences for tariff reform with managers’ appetite for change: The case of volumetric-only tariffs in Australia *Australian Journal of Agricultural and Resource Economics*.

Case 2: profile

- Rating elements of a product/policy etc
- Create versions of the product in a profile and select best and worst elements

Table 2 Example of a B-W choice set presented to respondents

Imagine you are at your usual grocery store, which of these attributes do you think are the most important and least important when purchasing a craft beer (choose only one as most important and one as least important)?		
Most important		**Least important**
☐	Taste	☐
☐	Country of origin	☐
☐	Alcohol content	☐
☐	Certification (*organic*)	☐

Lerro, M., Marotta, G., Nazzaro, C. (2020) Agricultural and Food Economics (2020) 8:1

Case 3: multi-profile

- A profile with attributes, but choosing across profiles

Features		Option A Maintain Current Situation	Option B Use water to:	Option C Use water to:
Culturally important waterholes		No natural waterholes remain	Preserve 5 natural waterholes	Preserve 3 natural waterholes
Water supply for towns		Groundwater supply falling	Supply 60 additional years of water	Supply 20 additional years of water
Grazing land		120,000 hectares degraded	Restore 15,000 hectares	Restore 75,000 hectares
Household cost *Per year for 5 years*	$	$0	$50	$100

	Option A	Option B	Option C
I like this option MOST: *Click on one box only*	☐	☐	☐
I like this option the LEAST: *Click on one box only*	☐	☐	☐

Hatton-McDonald et al (2019) Australian Journal of Agricultural and Resource Economics, 63, pp. 897–921

Focus now on case 1: object based

- Why not use Likert scales instead? i.e. rate each object on a scale, and then compare across objects?
  - Need to maintain consistency in calibration of the scale across objects
  - People may use different interpretation of end points
  - People tend to 'cluster' responses at end points
  - And at the limit can give all objects the same rating i.e. "very important"

- BWS avoids those issues
  - And has been found to be more consistent and reliable, even with young children

Burton,N., Burton,M., Fisher,C., González Peña,P., Rhodes,G. & Ewing, L Beyond Likert ratings: Improving the robustness of developmental research measurement using best-worst scaling *Behavior Research Methods* forthcoming

Case 1: design

- Repeated choice of best and worst
  - Need subsets of items
- Paired comparisons- not often used
  - Pair every object with every other: pick ‘best’ in each pair
  - Large number of pairs: J objects -> J(J-1)/2
    - J=10-> 45 pairs
  - Not efficient

Case 1: design

- Balanced Incomplete Block Designs (BIBD)
  - Designs with more than 2 items per ‘block’ or question
  - Balanced in that each item appears the same number of times
  - And co-occurs with other items the same number of times
  - *But its not a complete factorial*
- BIBD do not exist for all J
- Catalogues of BIBD exist

Table 2.3 *Illustrative list of potential BIBDs*

18 The BWS object case

Table 2.3 *Illustrative list of potential BIBDs*

Design no.	Objects (v)	No. sets (b)	Occurs (r)	Set size (k)	Co-occurs (λ)
1	4	4	3	3	2
2	5	5	4	4	3
3	5	10	6	3	3
4	6	10	5	3	2
5	7	7	3	3	1
6	7	7	4	4	2
7	7	21	15	5	10
8	8	14	7	4	3
9	9	12	4	3	1
10	9	18	8	4	3
11	9	12	8	6	5
12	9	18	10	5	5
13	10	15	6	4	2
14	10	30	9	3	2
15	10	18	9	5	4
16	10	15	9	6	5
17	11	11	5	5	2
18	11	11	6	6	3
19	11	55	15	3	3
20	12	44	11	3	2
21	12	33	11	4	3
22	12	22	11	6	5
23	13	13	4	4	1
24	13	26	6	3	1
25	13	26	12	6	5
26	13	39	15	5	5
27	14	26	13	7	6
28	15	35	7	3	1
29	15	35	14	6	5
30	16	20	5	4	1
31	16	16	6	6	2
32	16	24	9	6	3
33	16	80	15	3	2
34	16	48	15	5	4
35	19	57	9	3	1
36	19	57	12	4	2
37	21	21	5	5	1
38	21	70	10	3	1
39	21	42	12	6	3
40	25	30	6	5	1
41	25	50	8	4	1
42	25	100	12	3	1

(Louviere et al 2015 p18)

Case 1: design

- Balanced Incomplete Block Designs (BIBD)
  - Designs with more than 2 items per ‘block’ or question
  - Balanced in that each item appears the same number of times
  - And co-occurs with other items the same number of times
  - *But its not a complete factorial*
- BIBD do not exist for all J
- Catalogues of BIBD exist
- Can be generated in R (also see later)

Case 1: design

Table 2.2 *A BIBD for nine objects*

Subset	Objects in each subset			Issues in each subset		
1	2	4	8	K-12 education	Parks and recreation	Broadband access/speed
2	1	4	5	Streets and roads	Parks and recreation	Sports facilities
3	4	7	9	Parks and recreation	Job creation	Tourism facilities
4	3	4	6	Tertiary education	Parks and recreation	Housing developments
5	1	2	3	Streets and roads	K-12 education	Tertiary education
6	2	5	7	K-12 education	Sports facilities	Job creation
7	2	6	9	K-12 education	Housing developments	Tourism facilities
8	1	8	9	Streets and roads	Broadband access/speed	Tourism facilities
9	5	6	8	Sports facilities	Housing developments	Broadband access/speed
10	3	7	8	Tertiary education	Job creation	Broadband access/speed
11	1	6	7	Streets and roads	Housing developments	Job creation
12	3	5	9	Tertiary education	Sports facilities	Tourism facilities

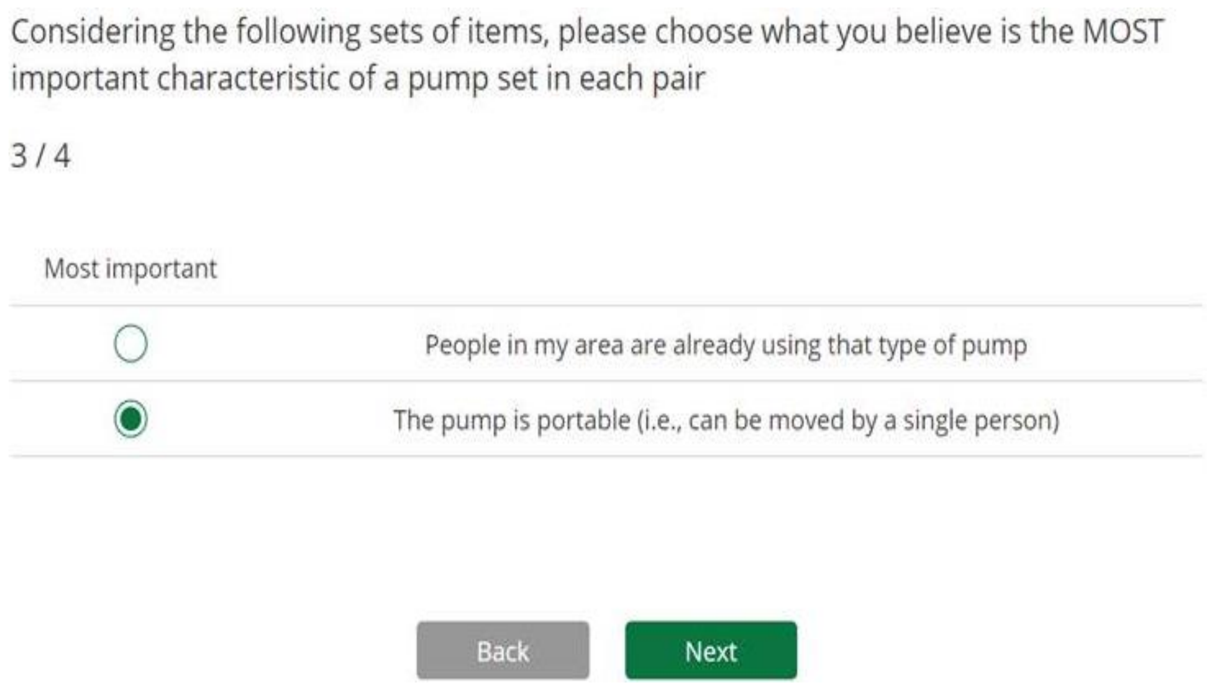
(Louviere et al 2015 p19)

Case 1: design

- What if no BIBD for your J?
  - Mix and match existing designs (see Louviere et al 2015 p19)
  - Approximations of BIBD designs (e.g. sawtooth)

- Which is best design to use?
  - Do you want to estimate individual ranks?
    - Need objects to occur at least 4 times
  - How many choice sets can your respondents cope with?

Case 1: design



Sophie Lountain, Bethany Cooper, Lin Crase and Michael Burton
Technology, gender and sustainable livelihoods: Insights into preferences for irrigation pumps in West Bengal

Paper prepared *for Institutions to support intensification, integrated decision making and inclusiveness in agriculture in the East Gangetic Plain*

Part 2

Case 1: Models of choice and ways to analysis

Case 1: Models of Choice

- Respondents are being asked to select best and worst from a subset of objects
  - NB we will use Best/Worst here but framing depends on context
- Assume that there is some latent measure for each object u(i)
  - For the set of objects in the set they will pick the one with the highest/lowest utility as best/worst
  - With an appropriate assumption about the nature of the random elements of choice this can be represented as a multinomial logit model:

Case 1: Probability of picking best and/or worst

If you assume there are random elements to choice of a particular form, then it's a multinomial logit model

$$P_B(i \mid X) = \frac{\exp(\beta_i)}{\sum_{j \notin X} \exp(\beta_j)}$$

$$P_W(i' \mid X) = \frac{\exp(-\beta_{i'})}{\sum_{j \notin X} \exp(-\beta_j)}$$

Case 1: Probability of picking best-worst pair

Assume that respondent compared all possible combinations and picked the combination with largest difference (MAXDIF)

$$P_B(ii' \mid X) = \frac{\exp(\beta_i - \beta_{i'})}{\sum_{\substack{j,j' \notin X \\ j \neq j'}} \exp(\beta_j - \beta_{j'})}$$

Case 1: Estimation of preference ratings

- Sequential best/worst using aggregate data
- If we just use 'best' choices:
- Estimate a multinomial logit model using data from all individuals

$$P_B(i \mid X) = \frac{\exp(\beta_i D_i)}{\sum_{j \notin X} \exp(\beta_j D_j)}$$

Case 1: Estimation of preference ratings

- If we just use ‘worst’ choices:
- Estimate a multinomial logit model using data from all individuals, but multiply all dummy variables by -1

$$P_w(i' | X) = \frac{\exp(\beta_{i'}(-D_{i'}))}{\sum_{j \notin X} \exp(\beta_j(-D_j))}$$

Case 1: Estimation of preference ratings

- If we use ‘best’ and ‘worst’ choices:
- ‘stack’ the BW data into a single data frame

- Issue: should the item selected as best be included in the set for worst?
  - Drop, if you are sure about order of choice
  - Some use complete sets for both (e.g. sawtooth)

Case 1: Estimation of preference ratings

- An example of how the data is prepared

Analysis 37

Table 2.15 *Selection of raw data for conditional logit model of best and worst*

							Presence (1 or –1)/absence (0) of objects								
ID	Total sets	Set	New set	BW	Option	Object #	O2	O3	O4	O5	O6	O7	O8	O9	Choice
1	1	1	1	1	1	2	1	0	0	0	0	0	0	0	1
1	1	1	1	1	2	4	0	0	1	0	0	0	0	0	0
1	1	1	1	1	3	8	0	0	0	0	0	0	1	0	0
1	2	1	2	–1	2	4	0	0	–1	0	0	0	0	0	0
1	2	1	2	–1	3	8	0	0	0	0	0	0	–1	0	1
1	3	2	3	1	1	1	0	0	0	0	0	0	0	0	0
1	3	2	3	1	2	4	0	0	1	0	0	0	0	0	1
1	3	2	3	1	3	5	0	0	0	1	0	0	0	0	0
1	4	2	4	–1	1	1	0	0	0	0	0	0	0	0	0
1	4	2	4	–1	3	5	0	0	0	–1	0	0	0	0	1

(Louviere et al 2015 p37)

Case 1: Estimation of preference ratings

- Because of singularity, need to drop one item from model: it becomes the ‘base’ with zero weight
- Doesn’t matter which one is dropped
- Parameter estimates now give preference ratings for objects

Case 1: A note on coding

- **Dummy coding**: dummy variable takes a value of 1 if present in set, zero if not
  - Parameters are estimated relative to base
- **Effects coding**: dummy variable takes a value of 1 if present in set, zero if not, and -1 if not and the base case is present
  - Parameters are estimated relative to the mean of all parameters
- Has no impact on the explanatory power of the model, just interpretation

(see Daly et al 2016)

An example: dummy coding

Conditional (fixed-effects) logistic regression

Number of obs	=	1,224
LR chi2(9)	=	55.62
Prob > chi2	=	0.0000
Pseudo R2	=	0.0656

Log likelihood = -396.39789

choi	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
d1	.569771	.2738044	2.08	0.037	.0331242	1.106418
d2	.0060952	.260852	0.02	0.981	-.5051653	.5173557
d3	-.1812733	.2453639	-0.74	0.460	-.6621778	.2996312
d4	-1.109367	.2766381	-4.01	0.000	-1.651567	-.5671662
d5	-.0949606	.2532575	-0.37	0.708	-.5913362	.4014151
d6	-.1759256	.2666333	-0.66	0.509	-.6985172	.3466661
d7	.4912596	.2992907	1.64	0.101	-.0953394	1.077859
d8	.1055038	.2751919	0.38	0.701	-.4338624	.64487
d9	-.4483684	.2744141	-1.63	0.102	-.9862102	.0894734

An example: effects coding

```
Conditional (fixed-effects) logistic regression

                                                Number of obs     =      1,224
                                                LR chi2(9)        =      55.62
                                                Prob > chi2       =     0.0000
Log likelihood = -396.39789                     Pseudo R2         =     0.0656
```

choi	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
d1e	.6534975	.1779575	3.67	0.000	.3047071	1.002288
d2e	.0898217	.1575392	0.57	0.569	-.2189495	.3985929
d3e	-.0975468	.1582635	-0.62	0.538	-.4077376	.2126441
d4e	-1.02564	.1862645	-5.51	0.000	-1.390712	-.6605685
d5e	-.0112341	.1675333	-0.07	0.947	-.3395933	.3171252
d6e	-.0921991	.1749942	-0.53	0.598	-.4351814	.2507832
d7e	.5749861	.1969912	2.92	0.004	.1888905	.9610817
d8e	.1892303	.1747858	1.08	0.279	-.1533436	.5318041
d9e	-.3646419	.1737879	-2.10	0.036	-.7052599	-.0240238

Compared:

Log likelihood = -396.39789

choi	Coef.
d1	.569771
d2	.0060952
d3	-.1812733
d4	-1.109367
d5	-.0949606
d6	-.1759256
d7	.4912596
d8	.1055038
d9	-.4483684

0.5636758

Log likelihood = -396.39789

choi	Coef.
d1e	.6534975
d2e	.0898217
d3e	-.0975468
d4e	-1.02564
d5e	-.0112341
d6e	-.0921991
d7e	.5749861
d8e	.1892303
d9e	-.3646419

0.5636758

The value of the base:

choi	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
d1e	.6534975	.1779575	3.67	0.000	.3047071	1.002288
d2e	.0898217	.1575392	0.57	0.569	-.2189495	.3985929
d3e	-.0975468	.1582635	-0.62	0.538	-.4077376	.2126441
d4e	-1.02564	.1862645	-5.51	0.000	-1.390712	-.6605685
d5e	-.0112341	.1675333	-0.07	0.947	-.3395933	.3171252
d6e	-.0921991	.1749942	-0.53	0.598	-.4351814	.2507832
d7e	.5749861	.1969912	2.92	0.004	.1888905	.9610817
d8e	.1892303	.1747858	1.08	0.279	-.1533436	.5318041
d9e	-.3646419	.1737879	-2.10	0.036	-.7052599	-.0240238

```
. nlcom -(_b[d1]+_b[d2]+_b[d3]+ _b[d4]+_b[d5]+_b[d6]+ _b[d7]+_b[d8]+_b[d9])

       _nl_1:  -(_b[d1]+_b[d2]+_b[d3]+ _b[d4]+_b[d5]+_b[d6]+ _b[d7]+_b[d8]+_b[d9])
```

choi	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_nl_1	.0837265	.186135	0.45	0.653	-.2810913	.4485443

Case 1: Estimation of preference ratings

- But one can just use counts!

$$\mathrm{N}ormalised\mathrm{BW} \qquad \frac{N_x^b - N_x^w}{N_x}$$

- Shown to be linear with parameter estimates

Case 1: Estimation of preference ratings

- More complex versions:

$$Analytical\mathrm{BW} \quad \ln\left(\frac{1+\dfrac{N_x^b - N_x^w}{N_x}}{1-\dfrac{N_x^b - N_x^w}{N_x}}\right)$$

(Marley et al 2016 Journal of Choice Modelling 21: 15-24)

An example:

Institutions and policies for enhancing farm household livelihoods: An analysis of the coherence of expert opinion in the East Gangetic Plain.

Bethany Cooper, Lin Crase, Michael Burton, Dan Rigby, Mohamad Jahangir Alam, Avinash Kishore

Item description in BWS
Cheaper farm inputs
Easier access to farm inputs
Higher farm output prices
More stable farm output prices
More income from non-farm sources
Farmers adopting different types of crops
Farmers increasing non-crop farming
Easier access to modern technology

Most Effective			Least Effective
○		**More variety in the crops grown** (e.g. subsidies/credit to grow different crops such as vegetables, oil, pulses etc.)	○
●		**Easier access to farm inputs** (e.g. quality seeds, in-time irrigation water, electricity; credit; good roads)	○
○		**Increasing non-crop farming** (e.g. credit/subsidies to support livestock/fishing or non-crop farm activities)	○
○		**Higher farm output prices** (e.g. more competition among buyers; easier access to markets with more buyers)	●

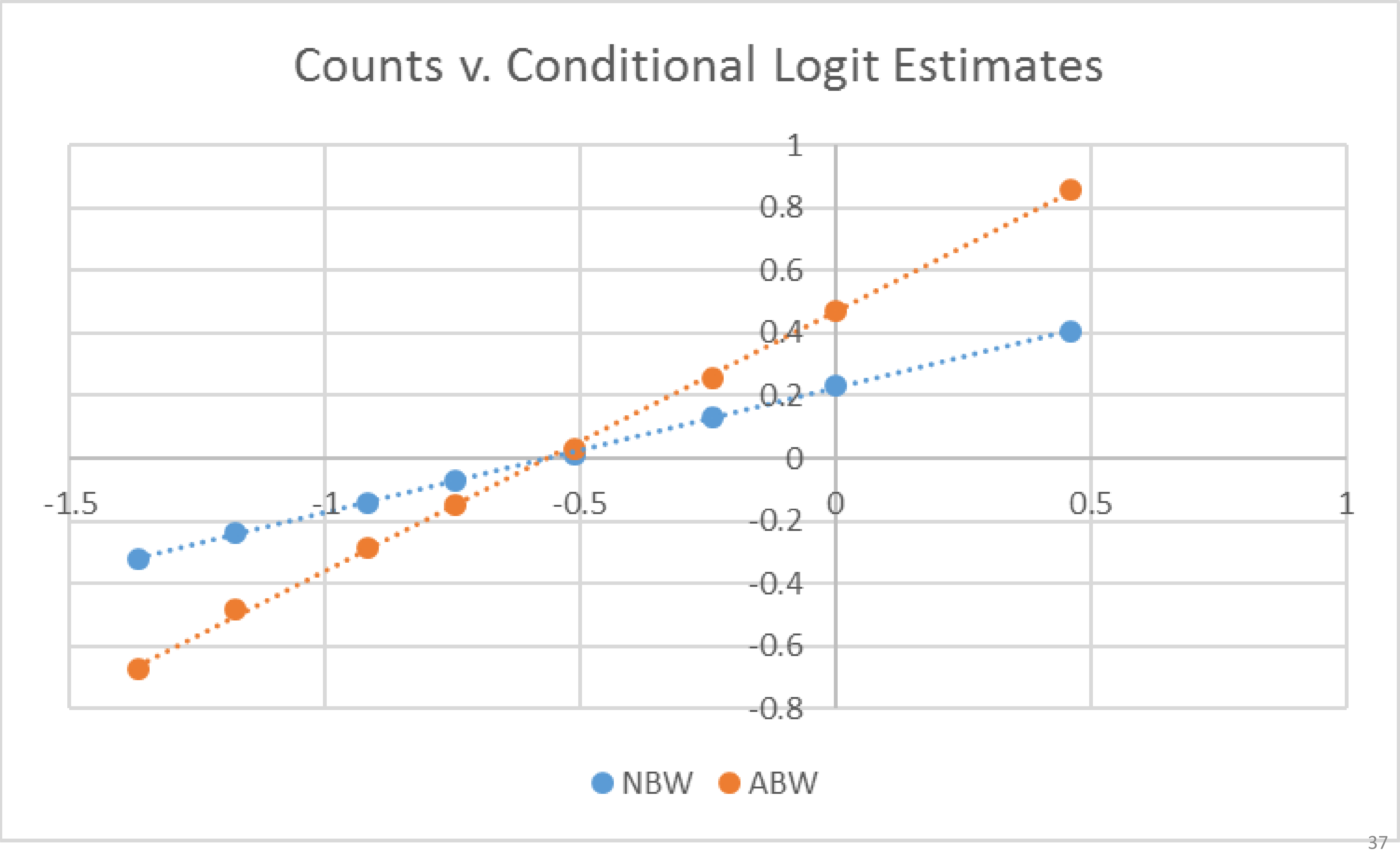
Conditional logit results: Nepal

Item	Nepal	
Cheaper farm inputs	-0.916***	(0.187)
Easier access to farm inputs	0.459**	(0.185)
Higher farm output prices	-0.511***	(0.185)
More stable farm output prices	-0.242	(0.185)
More income from non-farm sources	-0.745***	(0.184)
More variety in the crops grown	-1.367***	(0.188)
Increasing non-crop farming	-1.176***	(0.188)
Choices	296	
Individuals	37	
LL value	-742.41	

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Reference item: Easier access to modern technology

Counts analysis

Item	N	Best	Worst		NBW	ABW
1	148	28	49		-0.142	-0.286
2	148	70	10		0.405	0.860
3	148	37	35		0.014	0.027
4	148	52	33		0.128	0.258
5	148	34	45		-0.074	-0.149
6	148	16	64		-0.324	-0.673
7	148	15	50		-0.236	-0.482
8	148	44	10		0.230	0.468



Counts v. Conditional Logit Estimates
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1.5
-1
-0.5
0
0.5
1
NBW
ABW

Which approach to use?

- Counts
  - Simple
  - Can be used with individual data (to compare across individuals)
- Econometrics
  - Can be used for formal testing of differences across samples
  - Can potentially allow for interactions with sociodemographics (e.g. does age systematically change preferences?)
  - Will have issues with individual data if choices are deterministic

Part 3

- Interpretation/presentation of results

What do the estimates mean?

- NB They can only inform you about relative weights of objects, not absolute values
- Comparison of estimates gives relative weights on a line, but cannot be used as ratio scale (it has no absolute zero)
- Proposed transformation of Conditional logit estimates:

  Scaled Probability Scores (SPS)

Scaled probability scores

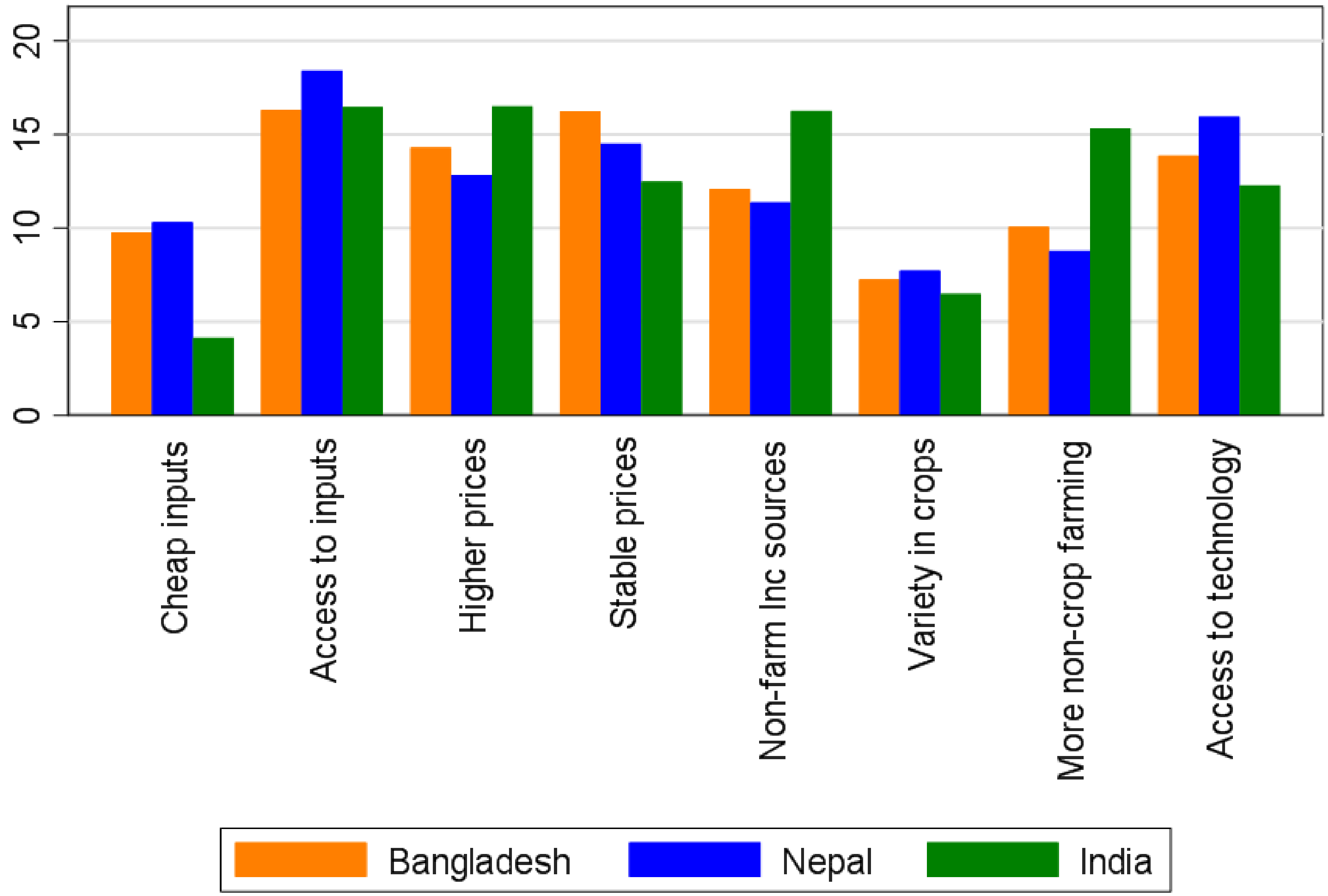
$$P_i = \frac{\exp\left(\bar{\beta}_i\right)}{1+\exp\left(\bar{\beta}_i\right)}$$

$\bar{\beta}_i$ = Parameter estimates defined as mean deviations

P is the probability of picking item *i* as best from a set of two items, where the ‘other’ item is average.

Then rescale so all J probabilities sum to 100

Sawtooth Software, Inc (2020) *The MaxDiff System Technical Paper.*



0
5
10
15
20
Cheap inputs
Access to inputs
Higher prices
Stable prices
Non-farm Inc sources
Variety in crops
More non-crop farming
Access to technology
Bangladesh
Nepal
India

Scaled probability scores

Can now be interpreted as a ratio scale: an SPS twice as large means the item is twice as likely to be picked as best

NB: these scores are influenced by the 'scale' parameter i.e. how much random 'noise' there is in choices.

Higher noise leads to SPS of all items being pushed towards the mean (i.e. 100/8=12.5 for a set with 8 items)

Alternative framings

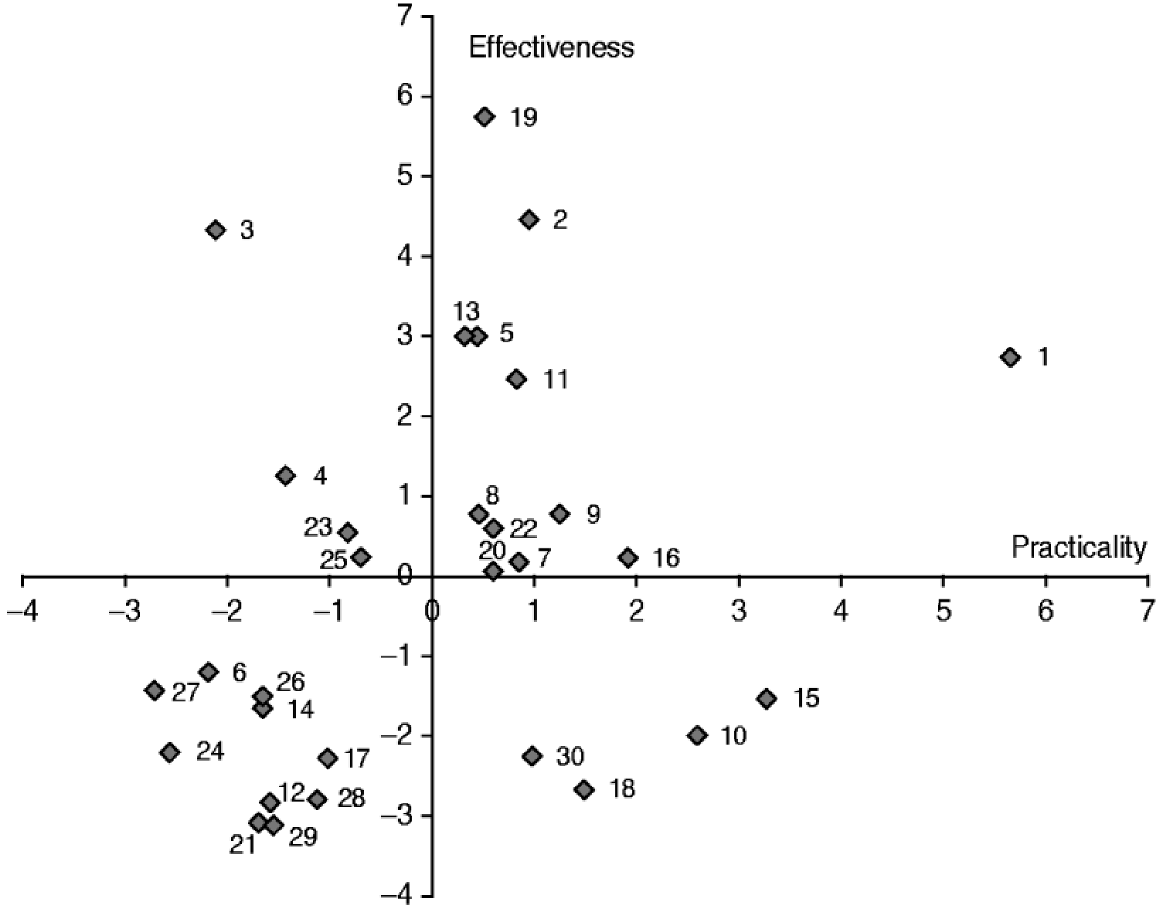
- One can ask the same people about the same item using different framings

CROSS, P., RIGBY, D., & EDWARDS-JONES, G. (2012). Eliciting expert opinion on the ***effectiveness*** and ***practicality*** of interventions in the farm and rural environment to reduce human exposure to Escherichia coli O157.

30 interventions

31 experts interviewed

Separate BWS questions for effectiveness and practicality



Zero-centred scatterplot of mean effectiveness and practicality scores for the 30 control measures

Part 4

- Anchors and absolute rankings and other issues

Are best and worst choices the same?

- Suggestion that people may use a different rating/utility system when picking worst compared to best

- Statistically that can be checked for:
  - Estimate separate models for best and worst, and test if parameters can be restricted to be the same (i.e. stacked data)

Bangladesh data: checking for B=W

Conditional (fixed-effects) logistic regression

Log likelihood = -332.43825

choi	Coef.
d1	-.2474472
d2	.6359794
d3	.4123668
d4	.6186124
d5	-.1177144
d6	-.7010632
d7	-.5182924

Log likelihood = -328.83119

choi	Coef.
d1	-.9437614
d2	.0212516
d3	-.426686
d4	.2653816
d5	-.706115
d6	-1.276726
d7	-.9067428

Log likelihood = -663.98849

choi	Coef.	Std. Err.	z
d1	-.5718185	.1971666	-2.90
d2	.3880302	.1963993	1.98
d3	.0600605	.1970667	0.30
d4	.4514399	.1957909	2.31
d5	-.373773	.1955897	-1.91
d6	-.9624714	.1973258	-4.88
d7	-.6500818	.1984134	-3.28

Likelihood-ratio test LR chi2(7) = 5.44
Prob > chi2 = 0.6067

Nepalese data

```
Conditional (fixed-effects) logistic

Log likelihood = -369.18792
```

choi	Coef.	Std. Err.
d1	-.5667543	.2575368
d2	.6264585	.2213878
d3	-.3400261	.2456212
d4	.0904804	.2305155
d5	-.4757806	.2496633
d6	-1.139319	.2976889
d7	-1.357397	.3192527

```
Conditional (fixed-effects) logistic

Log likelihood = -364.40365
```

choi	Coef.	Std. Err.
d1	-1.776511	.3599969
d2	.0103175	.4541314
d3	-1.302457	.3704032
d4	-1.341913	.3693341
d5	-1.612275	.3602653
d6	-2.091312	.3540555
d7	-1.743905	.3613868

```
Conditional (fixed-effects) logistic regression

                                                Number of obs
                                                LR chi2(7)
                                                Prob > chi2
Log likelihood = -742.54909                     Pseudo R2
```

choi	Coef.	Std. Err.	z	P>|z|	[95%
d1	-.9815239	.1873956	-5.24	0.000	-1.348
d2	.4511289	.185107	2.44	0.015	.0883
d3	-.5980513	.1859084	-3.22	0.001	-.9624
d4	-.3650875	.1850661	-1.97	0.049	-.7278
d5	-.8407162	.1850333	-4.54	0.000	-1.203
d6	-1.409452	.1883704	-7.48	0.000	-1.778
d7	-1.217338	.1888133	-6.45	0.000	-1.582

```
. lrtest m12 (m1 m2)

Likelihood-ratio test                        LR chi2(7)  =     17.92
                                             Prob > chi2 =    0.0124
```

Allowing for difference in variance

```
Heteroscedastic logistic regression               Number of obs     =       2368
                                                  Number of groups  =        592
                                                  LR chi2(1)        =       1.37
Log likelihood = -741.86274                       Prob > chi2       =     0.2413
```

choi	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
variables						
d1	-1.178	.2704169	-4.36	0.000	-1.708008	-.6479927
d2	.5269883	.2352974	2.24	0.025	.0658138	.9881628
d3	-.7494853	.2537402	-2.95	0.003	-1.246807	-.2521637
d4	-.5211527	.2612089	-2.00	0.046	-1.033113	-.0091926
d5	-1.023802	.2635334	-3.88	0.000	-1.540318	-.5072864
d6	-1.602753	.2649781	-6.05	0.000	-2.122101	-1.083406
d7	-1.383366	.2542077	-5.44	0.000	-1.881604	-.8851282
het						
vern	-.2419079	.2062068	-1.17	0.241	-.6460658	.1622499

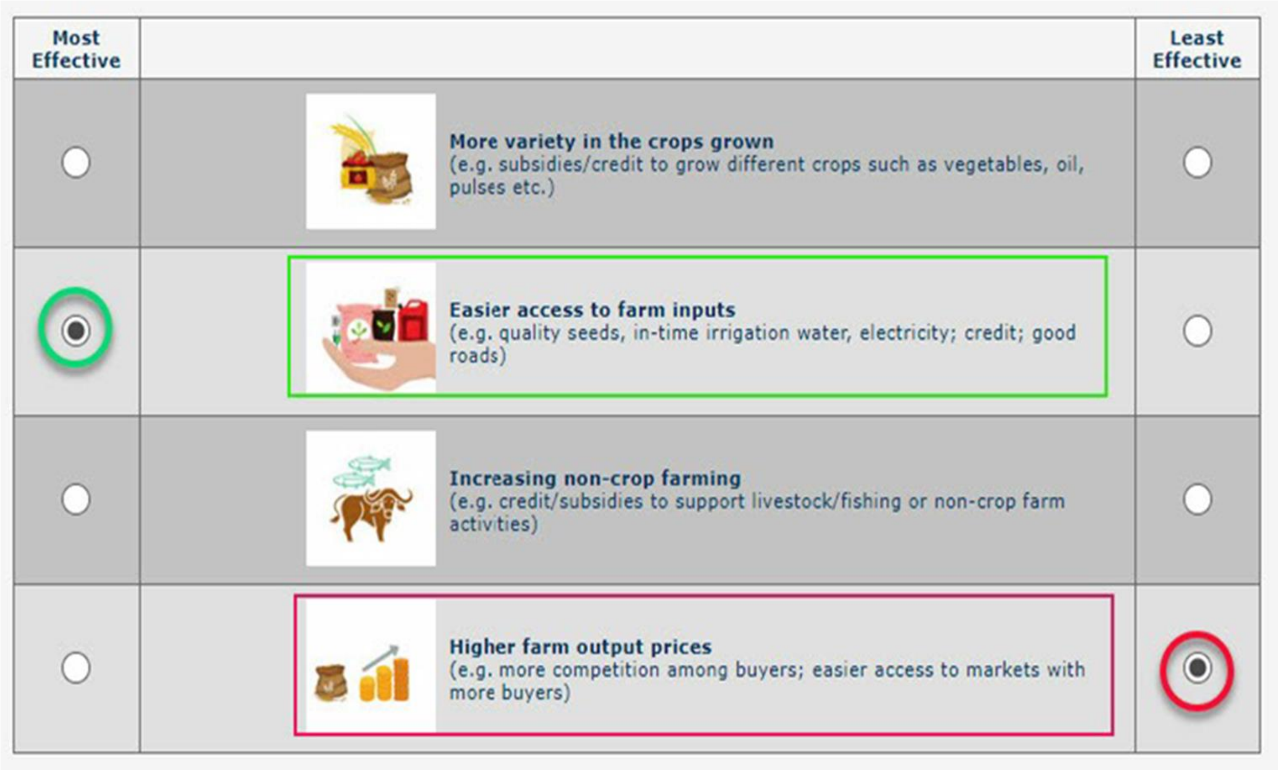
```
. lrtest m12het (m1 m2)

Likelihood-ratio test                                 LR chi2(6)  =     16.54
                                                      Prob > chi2 =    0.0111

Assumption: (m12het) nested in (m1, m2)
```

Absolute scale

- Respondents have to rate as best and worst in set:
  - But that doesn't mean objects would be acceptable



Anchored best-worst

- You can include an additional question after each BWS question

Considering only these four features, which is the Most Important and which is the Least Important?

Least Important	Feature	Most Important
	The price paid for the product offers 'value for money'	
	The product bought is washed and sprinkled with running tap water at the market	
	The product is transported to the market in a hygienic way (covered and in a clean container)	
	Pesticides have been applied using the recommended dosage for a given symptom	

Considering just these four features…

- **None of these four is important**
- **Some are important, some are not**
- **All four are important**

How do you use anchored information?

- If they select "None of these four is important" then all lie below a point of indifference, or zero
- If they select " All four are important" then all lie above some point of indifference, or zero
- If they select "Some are important, some are not" then the "best" object lies above, and the "worst" lies below.

Considering just these four features…

- None of these four is important
- Some are important, some are not
- All four are important

Adding this to estimation…

- Introduces an anchor with a value of zero, and those that are positive are deemed worth having, and those below are not.

- Details are in : - Sawtooth Software, Inc (2020) *The MaxDiff System Technical Paper.*

BWS-acceptability

Please consider the 4 control measures below.

Which is the most acceptable to you, and which is the least acceptable?

Assume no change in the cost of your chicken, and that all the measures are equally effective.

Most Acceptable		**Least Acceptable**
○	**Neck Skins** *Neck Skins removed after slaughter*	○
○	**"Do Not Wash"** *All whole, non-frozen, chickens to be sold with prominent "do not wash" labels on the packaging*	○
○	**Vaccination** *Vaccination of chickens at the farm against Campylobacter*	○
○	**Chilling** *Chilling the surface of chicken carcass after slaughter*	○

Below we list all the control measures to control Campylobacter

In the previous questions you were asked to identify the most and least acceptable control measures from the 4 shown on a page.

Now we would like you to identify which you find Acceptable and which you find Unacceptable.

	Acceptable To Me	Unacceptable To Me
Chlorine wash ***Dipping chicken carcass into chlorine wash after slaughter***	○	○
Heat ***Dipping chicken carcass into hot water bath after slaughter***	○	○
Feed Additives ***Chickens receive food additives to reduce how many of them get Campylobacter***	○	○
Irradiation ***Exposure of chicken carcass to irradiation after slaughter***	○	○
Ozone Gas ***Exposure of chicken carcass to ozone gas after slaughter***	○	○
Chilling ***Chilling the surface of chicken carcass after slaughter***	○	○
Farmers Paid ***Farmers being paid more for Campylobacter free chickens***	○	○
Frozen ***All fresh chicken sold to have been previously frozen***	○	○

Threshold Question

Anchored BWS Logit Results

Label	Item Number	Coeff	Std Error
Farmers Paid Farmers being paid more for Campylobacter free chickens	1	1.301	0.040
Chilling Chilling the surface of chicken carcass after slaughter	3	0.836	0.039
Roast-in-the-Bag All whole, non-frozen, chickens to be sold as pre-packed, roast-in-the-bag chickens.	9	0.790	0.039
Neck Skins Neck Skins removed after slaughter	8	0.538	0.039
"Do Not Wash" All whole, non-frozen, chickens to be sold with prominent "do not wash" labels on the packaging	10	0.515	0.038
Vaccination Vaccination of chickens at the farm against Campylobacter	11	0.431	0.038
Heat Dipping chicken carcass into hot water bath after slaughter	4	0.399	0.038
Feed Additives Chickens receive food additives to reduce how many of them get Campylobacter	2	0.390	0.038
Anchor		**0.000**	**N/A**
Ozone Gas Exposure of chicken carcass to ozone gas after slaughter	6	-0.305	0.038
Frozen All fresh chicken sold to have been previously frozen	12	-0.470	0.039
Irradiation Exposure of chicken carcass to irradiation after slaughter	7	-0.592	0.039
Chlorine wash Dipping chicken carcass into chlorine wash after slaughter	5	-0.769	0.039

Part 5

- Resources, and software to implement

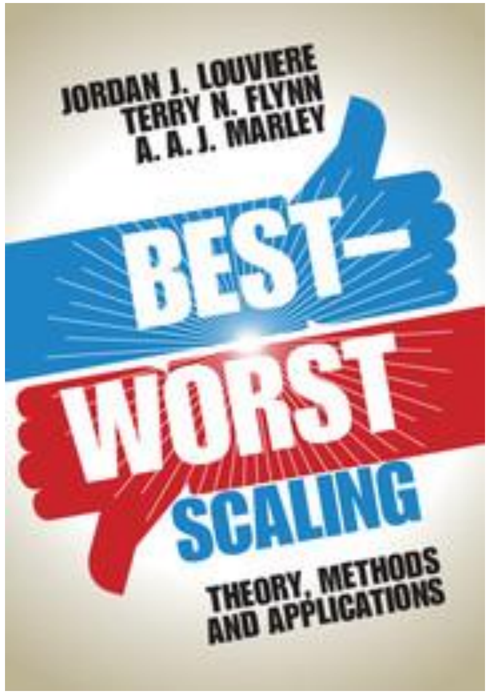
Texts

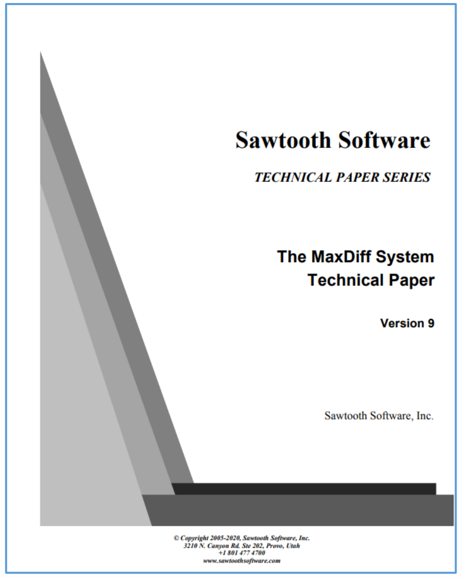
Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). *Best-worst scaling: Theory, methods and applications*: Cambridge University Press.

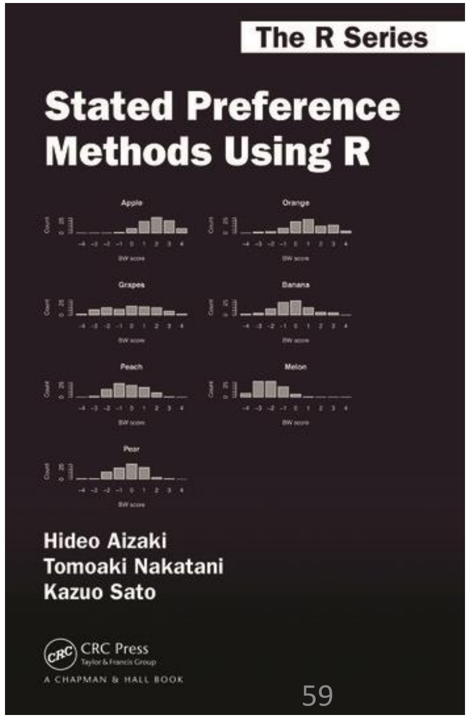
Aizaki, H., Nakatani, T., & Sato, K. (2014). *Stated preference methods using R*: Chapman and Hall/CRC

Sawtooth Software, Inc (2020) *The MaxDiff System Technical Paper.* Available at https://sawtoothsoftware.com/resources/technical-papers/maxdiff-technical-paper

Burton,N., Burton,M., Rigby,D. Sutherland, C.A.M., Rhodes, G. (2019) Best-worst scaling improves measurement of first impressions *Cognitive Research: Principles and Implications* 4(1),36 https://doi.org/10.1186/s41235-019-0183-2







Software

- Low minimal requirements
  - Could be paper based and analysed using counts e.g. in excel
- Any online survey software that can present a table of items, and allow you to select 2 from a set (e.g. Qualtrics)
- Any statistical software that can estimate a conditional logit model (e.g. R or Stata)
- Sawtooth software: can design choice sets from object list, format and present in online mode, and has advanced analysis capability

References

- Aizaki, H., Nakatani, T., & Sato, K. (2014). *Stated preference methods using R*: Chapman and Hall/CRC
- Burton, N., Burton, M., Rigby, D., Sutherland, C.A.M., & Rhodes, G. (2019). Best-worst scaling improves measurement of first impressions. *Cognitive Research: Principles and Implications, 4 (1)*, 1 – 10. https://doi.org/10.1186/s41235-019-0183-2
- Burton,N., Burton,M., Fisher,C., González Peña,P., Rhodes,G. & Ewing, L (2021) Beyond Likert ratings: Improving the robustness of developmental research measurement using best-worst scaling *Behavior Research Methods* forthcoming https://doi.org/10.3758/s13428-021-01566-w
- CROSS, P., RIGBY, D., & EDWARDS-JONES, G. (2012). Eliciting expert opinion on the effectiveness and practicality of interventions in the farm and rural environment to reduce human exposure to Escherichia coli O157. *Epidemiology and Infection, 140*(4), 643-654. doi:10.1017/S0950268811001257
- Chrzan K., Golovashkina N. (2006) An empirical test of six stated importance measures International Journal of Market Research, 48 (6) , pp. 717-740.
- Daly,A., Dekker,T. and Hess,S. (2016) Dummy coding vs effects coding for categorical variables: Clarifications and extensions, Journal of Choice Modelling,21:36-41
- Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). *Best-worst scaling: Theory, methods and applications*: Cambridge University Press.
- Sawtooth Software, Inc (2020) *The MaxDiff System Technical Paper.* Available at https://sawtoothsoftware.com/resources/technical-papers/maxdiff-technical-paper