Sawtooth Software

RESEARCH PAPER SERIES

The Options Pricing Model: An Application of Best-Worst Measurement

Keith Chrzan,
Maritz Research

530 W. Fir St.
Sequim, WA 98382
(360) 681-2300
www.sawtoothsoftware.com

The Options Pricing Model:
An Application of Best-Worst Measurement

Keith Chrzan
Maritz Research

I. Introduction

Marketing researchers often employ one or another type of conjoint analysis to model price sensitivity: typically price is one attribute among several and calculated utilities for its levels allow simulation and price sensitivity analyses. The model described below has this special property: it can be used to estimate separate price sensitivity curves for individual attributes.

II. Background

Some history on maximum difference scaling and best-worst "conjoint" analysis will set the stage for the research problem and the proposed solution described below.

Maximum Difference Scaling
Developed by Finn and Louviere (1992), maximum difference scaling (maxdiff) has several benefits over traditional perceptual scaling:

- It is a general attitude scaling method – Sa Lucas, in this volume (Proceedings of the 2004 Sawtooth Software Conference), illustrates its use in measuring the perceived severity of criminal offenses and as an alternative to a Likert agreement scale (Sa Lucas 2004)
- It is a more discriminating way to measure attribute importance than either rating scales or the method of paired comparisons (Cohen 2003)
- It has greater predictive validity as an importance measurement than either ratings scales or the method of paired comparisons (Cohen 2003)
- Like other choice-based methods, it prevents scale-use heterogeneity, making it ideal for cross-cultural studies, or as the basis for needs-based segmentation (Cohen 2004).

Maxdiff is the multinomial extension of the traditional method of paired comparisons (Thurstone 1927, David 1988). Whereas a paired comparison question asks a respondent to make a binary choice ("Would you rather have a TV with a flat screen or with a built-in DVD player?") maxdiff has the respondent specify "best" and "worst" choices from sets of three or more objects:

"Which of the following features would MOST make you want to buy the TV described below and which would LEAST make you want to buy it?

	Most	Least
Built-in DVD player	[ ]	[ ]
Flat screen	[ ]	[ ]
Cable-ready	[ ]	[ ]
5 year warranty	[ ]	[ ]
Made by Sony	[ ]	[ ]

Design strategies for maxdiff question sets range from using traditional orthogonal fractions of 2^n factorial designs to the Sawtooth Software B/W Designer (Sawtooth Software 2003) that creates balanced designs with fixed numbers of alternatives per question.

Analysis options include:

- Simple counting and arithmetic transforms when using perfectly balanced designs (Finn and Louviere 1992; Swait, Louviere and Anderson 1995)
- Aggregate MNL, including SAS and SPSS programs
- Latent class MNL with Latent Gold or Sawtooth Software's CBC Latent Class Module
- HB MNL using Sawtooth Software's CBC/HB product

An important twist in coding maxdiff questions is that one reverse-codes the design matrix for the "worst" choices (i.e. multiplies all design codes by -1) and then concatenates the best and worst data matrices into a single data set. Though this method produces good results in practice, it is theoretically unappealing, because it frequently happens that the utilities would be different if you ran separate "best" and "worst" models. Most practitioners just ignore this slight contradiction and combine the "best" and "worst" choices anyway.

Optionally, one could further simplify the design coding task by modeling the "best" choice as a multinomial choice among the several alternatives, and dividing the "worst" choice into a series of pairwise choices wherein each non-best and non-worst alternative is, in turn, shown to be preferred over the "worst" alternative. For CBC users, this means you can use standard .cho file coding rather than creating a user specified design coding matrix. A more complete description of coding options in CBC appears in the Appendix.

Best-Worst "Conjoint" Analysis
Swait, Louviere and Anderson (1995) describe how to extend maxdiff scaling to a conjoint-like application they call Best-Worst Conjoint Analysis or B-W. One could argue, since the technique produces no decomposition of the utility of multiattribute stimuli, that it is not a type of conjoint analysis at all.

Taxonomy aside, The respondent's task in B-W is the same as in maxdiff, but the experimental design differs. Using an orthogonal main-effects design or an efficient design generated by a SAS design macro, one creates a set of profiles. Each profile becomes a separate choice question, where respondents choose the "best" and "worst" attribute/level combination in each question:

> "Which of the following features would MOST make you want to buy the TV described below and which would LEAST make you want to buy it?

	Most	Least
Built-in DVD player: No	[ ]	[ ]
Flat screen: No	[ ]	[ ]
Cable-ready: Yes	[ ]	[ ]
Warranty: 5 year	[ ]	[ ]
Manufacturer: Sony	[ ]	[ ]
Price: $299	[ ]	[ ]

In direct comparisons with choice-based conjoint models containing the same attributes and levels,

- B-W experiments have been found to contain less respondent error (Swait, Louviere and Anderson 1995, Chrzan and Skrapits 1996)
- Mixed results occur as to whether or not B-W and choice-based conjoint produce equivalent part-worth utilities, after adjusting for the difference in respondent error, with one finding that they do (Swait, Louviere and Anderson 1995) and one that they do not (Chrzan and Skrapits 1996)
- B-W may have a slight edge in predictive validity, though the test that showed this was a weak one (Chrzan and Skrapits 1996)

A further advantage of B-W is that it puts all attribute levels on the *same* interval scale; this is different than choice-based conjoint, which puts all attributes' levels on different scales with different arbitrary origins. This property uniquely allows cross-attribute level comparisons, an advantage illustrated below.

III. The Options Pricing Model

Some products are sold with optional features that are available at additional cost. For example, many personal computer manufacturers sell their products online, with base models that can be customized with optional features and upgrades for incremental cost. Likewise, automobile manufacturers offer optional features, at separate additional costs.

In these cases, each optional feature has a separate price and, conceivably, a separate sensitivity to differences in price. A client wanting to understand the price sensitivity attributable to each feature may be ill-served by using a typical choice-based conjoint model:

Which of these cars would you rather buy?

Car A	Car B
Honda	Ford
Minivan	Sedan
$26,000	$34,000
Sunroof @ $400	Sunroof @ $350

In this case, the cost difference in the base prices swamps that attributable to the sunroof, making it difficult to measure the sunroof price adequately. Even leaving the base price and the other attributes out of the experiment, however, price sensitivity of individual options may be poorly measured:

Which of these cars would you rather buy?

Car A	Car B
Antilock brakes @ $200	Antilock brakes at $400
Sunroof @ $600	Sunroof @ $300
CD player @ $200	CD Player @ $400
Heated Seats @ $300	Heated seats @ $200
Total: $1,300	Total: $1,300

Respondents would understandably be indifferent between the above packages. Moreover, the reality of the choice process to be measured is that respondents can pick and choose individual options, not packages as in this question. Perhaps better would be to do paired comparisons, like

Would you rather buy

[ ] A sunroof for $400, or
[ ] A CD changer for $350

The multiple choice extension of this is, of course, a type of B-W model we can call the Options Pricing Model (OPM).

In OPM, attributes' levels are their price points. Using an efficient experimental design, one can create a set of profiles to be evaluated with maxdiff scaling. If the number of attributes is less than about eight or ten, these may be full profile questions. Partial profile questions may be used when the number of attributes is larger. Analysis can be done to produce aggregate (MNL), segment (latent class MNL) or respondent level models (hierarchical Bayesian MNL).

IV. Case Study

Research Design and Analysis
The client needed to test price sensitivity for 11 options available on automobiles, each at 4 price points. Attributes and price points (partially disguised) were:

- **6 Disk In-Dash CD Changer** at $(600, 675, 750, 900)
- **Sunroof** at $(800, 900, 1,000, 1,200)
- **Anti-Lock Brakes** at $(400, 500, 600, 800)
- **MP3 Player** at $(410, 470, 530, 650)
- **Rear Side Airbags** at $(300, 350, 400, 500)
- **Heated Front Seats** at $(300, 350, 400, 500)
- **Fog Lights** at $(250, 300, 350, 450)
- **Cassette Player** at $(150, 200, 250, 350)
- **Keyless Entry** at $(210, 260, 310, 410)
- **Cruise Control** at $(135, 175, 215, 295)
- **100,000 Mile Powertrain Warranty** at $(900, 1,200, 1,500, 2,100)

202 household automobile purchase decision makers were qualified and surveyed using an internet panel sample source and a web-based interview.

Given the large number of attributes, we employed a partial profile design. Using an orthogonal fraction of a 2^{11} design, we settled on a 15 run design to determine which attributes would be present in each of 15 choice questions (the 16th was the null set with no attributes present). Order of questions was randomized, and levels were randomly assigned to each attribute, in each question, for each respondent. Across respondents this creates a near-orthogonal design. This would have been a poor design had the client wanted respondent-level utilities, because the design would be unbalanced and inefficient for individual respondents. An aggregate model fit the client's need (and budget) so this potential limitation was not injurious to the study's objectives.

Results

<u>Utility Model</u>
Two possibilities are that price could be generic (have the same utility for single dollar differences across all attributes) or attribute-specific. The utility of a single dollar differed across attributes by as much as a factor of 10, however, so the simple generic model described the data poorly and was abandoned. The resulting utility model appears in Table 1.

Table 1

Utilities

Attribute	Lowest Price	Mid-Low	Mid-High	High Price
CD(600, 675, 750, 900)	1.289	1.197	0.973	0.837
Sunroof(800, 900, 1,000, 1,200)	1.493	1.066	0.830	0.907
Brakes(400, 500, 600, 800)	1.351	1.242	1.055	0.648
MP3(410, 470, 530, 650)	1.266	0.994	1.002	1.034
Airbags(300, 350, 400, 500)	1.214	1.197	1.030	0.855
Seats(300, 350, 400, 500)	1.184	1.149	1.152	0.811
Lights(250, 300, 350, 450)	1.265	1.106	1.114	0.811
Cassette(150, 200, 250, 350)	1.173	1.173	1.107	0.843
Keyless (210, 260, 310, 410)	1.218	1.218	1.013	1.203
Cruise (135, 175, 215, 295)	1.203	1.297	1.039	0.757
Warranty (900, 1,200, 1,500, 2,100)	2.418	1.134	0.744	0.000

Note that the last level of the last attribute is the reference level – set to the arbitrary zero of the interval scale. All other levels of all attributes are measured on this same scale, with this same origin. This is different than all other forms of conjoint analysis, in which each attribute's levels are interval scales with separate origins. McFadden's ρ^2 was .217, suggesting a reasonably good model fit; put another way, the root likelihood of the aggregate model was .253, versus an expected likelihood of .175 from respondents choosing randomly.

Simulation

Because all attributes' levels are on the same interval scale, it makes sense to measure the relative appeal of two different options at two different price levels. For example, take the Anti-Lock Brakes at $400 (utility 1.35) and the MP3 Player at $650 (utility 1.034). From these we can use the standard MNL choice rule to calculate that demand for the MP3 Player at this price will be 42% that of the Anti-Lock Brakes at $400:

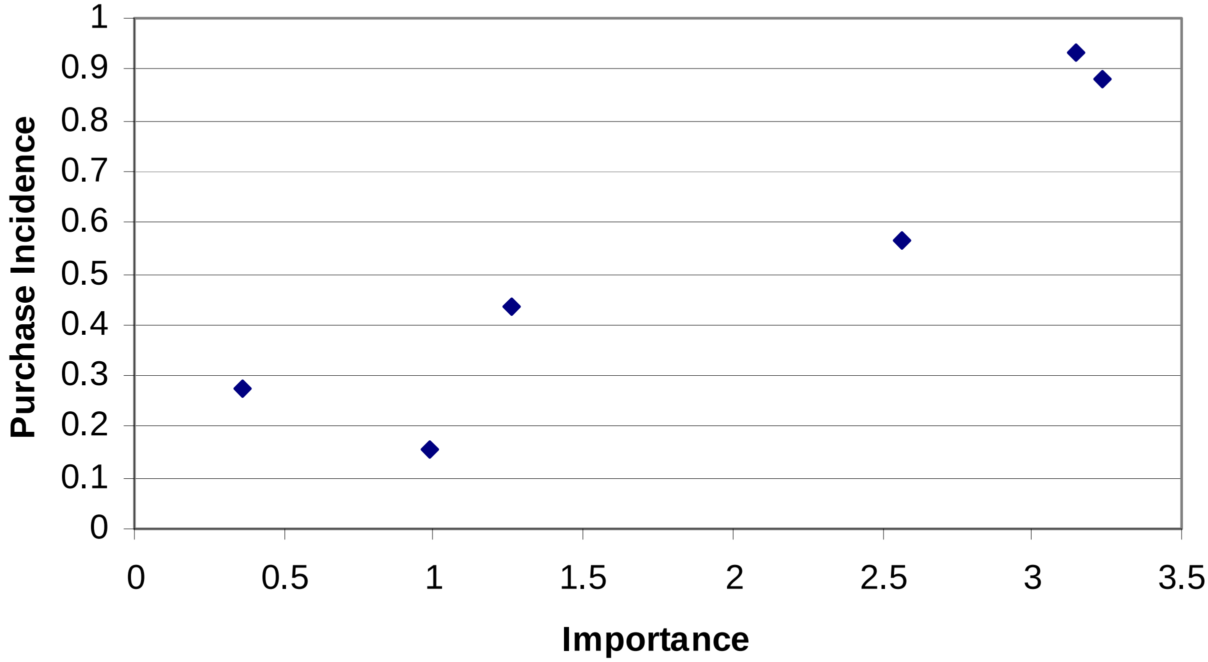
- Exponentiating the utility of the Anti-Lock Brakes yields $e^{1.35} = 3.857$
- Exponentiating the utility of the MP3 Player yields $e^{1.034} = 2.812$
- The logit choice rule would put the relative share for the MP3 player at 2.812/(2.812+3.857) = 42%

This allows the client to forecast demand for new features.

Validation

The utilities of existing options at existing price point should be related to the actual demand for those options. Prior to the B-W exercise, respondents reported which options they had added to their most recent vehicle purchases. The correlation between self-reported purchases and OPM utilities was 0.92. See the scatterplot in Figure 1

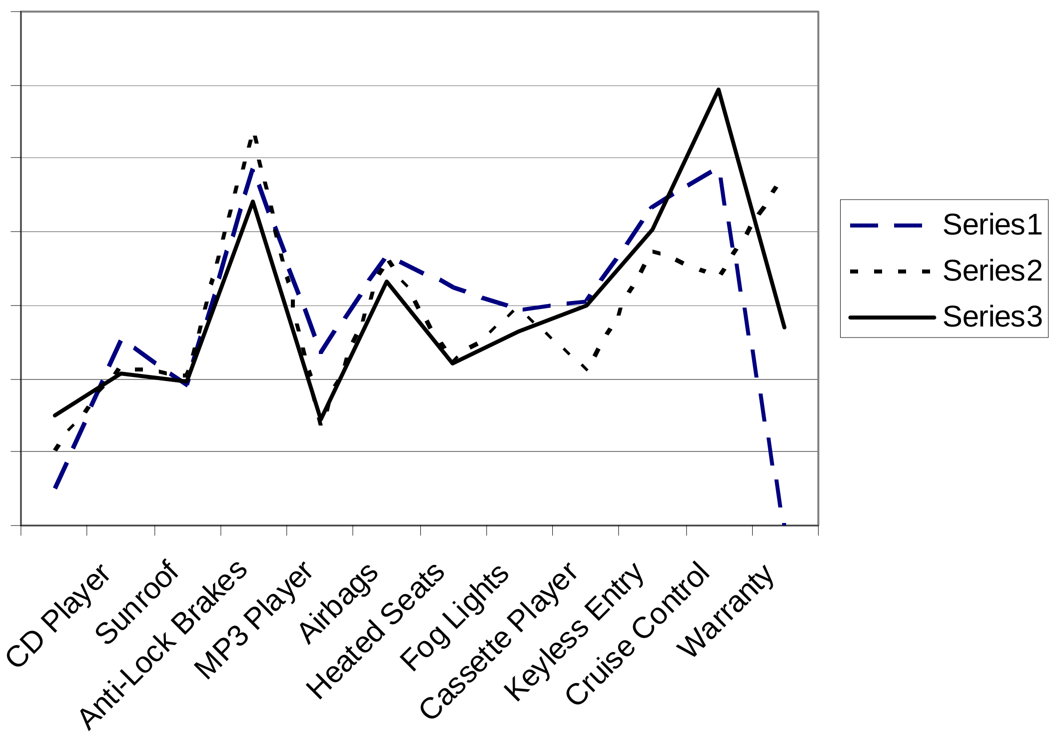
Figure 1



Segmentation

Using latent class MNL, or HB with cluster analysis, respondents could be grouped into segments. Though this study was not designed for respondent-level estimation, the client requested an exploratory segmentation, using cluster analysis of attribute utilities (rather than part-worth utilities). Since utilities were estimated using the second method described in the Appendix, attribute utilities were computed as the mean of the levels' utilities for each attribute. One solution that resulted appears in Figure 2, though it is not the solution the client actually used.

Figure 2



Segment 1 is most interested in heated seats, cruise control and airbags, and least interested in the warranty. Segment 2 seems more safety conscious: they are the least interested in entertainment attributes and most interested in anti-lock brakes and the warranty. Segment 3 showed the least interest in anti-lock brakes and the highest interest in the CD player and in cruise control.

V. Conclusion

Based on maxdiff scaling and more specifically on BW conjoint analysis, the Options Pricing Model is a designed choice experiment for a specific pricing situation. OPM looks like a reasonable solution to this very specific price sensitivity modeling need.

Likely OPM could be extended to handle more general menu price problems like those that restaurants face, perhaps with separate OPM exercises for appetizers, soup and salad, entrees and desserts.

Appendix:
Making a .cho File

Some ways of organizing OPM data for analysis via Sawtooth Software's CBC utility estimation tools appear below.

One could use the Swait, Louviere and Anderson (1995) recommendation for coding to produce attribute utilities in addition to part worth utilities. For the case of the 4^{11} experiment, this would produce 10 = (11-1) attribute utilities and 33 = (11[4-1]) part worth utilities, for a total of 43 parameters. The ability of both CBC/HB and Sawtooth Software's Latent Class Module to handle user-specified design coding is vital for the effects coding and for the reverse coding of the "worst" levels used in this strategy.

An easier way is to consider each of the 44 levels to be separate objects and to code them as you would the objects in a standard maxdiff experiment: 43 dummy codes for 44 objects. Each attribute utility is just the mean of its levels' part worth utilities. This still requires user-specified design coding to reverse code the design matrix for the "worst" choices.

As the CBC/HB software is currently programmed, neither of these methods allows you to impose monotonicity constraints, a handy feature indeed for a pricing study. To do this you have to model all the instances of attribute/level combinations being better than others, and this involves breaking down each B-W question into several inequalities. For example, in a set of objects A – E, say a respondent chooses A to be the best and D to be the worst. In this case we recast these best and worst choices into four choice observations:

- Choose A from A – E
- Choose B from B & D
- Choose C from C & D
- Choose E from D & E

One can now use standard CBC .cho file attribute coding and a standard constraint (.con) file.

Another way to impose constraints is to save the HB draws and use the manual method of tying the draws and collapsing into a point estimate per individual, as described by Rich Johnson in his "Monotonicity Constraints" paper for HB (Johnson 2000).

References

Chrzan, Keith and Mike Skrapits (1996) "Best-Worst Conjoint Analysis: An Empirical Comparison with a Full Profile Choice-Based Conjoint Experiment," paper presented at the INFORMS Marketing Science Conference, Berkeley.

Cohen, Steve (2003) "Maximum Difference Scaling: Improved Measures of Importance and Preference for Segmentation," 2003 Sawtooth Software Conference Proceedings, 61-74.

David, H. A. (1988) The Method of Paired Comparisons. New York: Oxford University Press.

Finn, Adam and Jordan J. Louviere (1992) "Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety," Journal of Public Policy and Marketing, 11: 12-25.

Johnson, Richard M. (2000) "Monotonicity Constraints in Choice-Based Conjoint with Hierarchical Bayes," Technical Paper available at www.sawtoothsoftware.com.

Sa Lucas, Luis (2004) "Scale Development with MaxDiffs : A Case Study," 2004 Sawtooth Software Conference Proceedings, in press.

Sawtooth Software (2003) "Best/Worst Designer," white paper available at www.sawtoothsoftware.com.

Swait, Joffre, Jordan J. Louviere and Don Anderson (1995) Best Worst Conjoint: A New Preference Elicitation Method to Simultaneously Identify Overall Attribute Importance and Attribute Level Partworths," working paper.

Thurstone, L. L. (1927) "A Law of Comparative Judgment," Psychological Review, 4: 273-286.