Model-Based Estimation of CBC Attribute Impact

Christopher N. Chapman PhD, Microsoft

Summary

A long-standing problem in choice-based conjoint (CBC) studies is determining attribute importance. The idea is salient to practitioners and clients, but current methods are not completely satisfactory.

We propose a new method to estimate attribute impact by examining how a model changes as attribute data is systematically decoupled from observed choices.

This work rests on a simple idea: ***If an attribute matters in a CBC model, altering its data will predict observed choices with lower accuracy.***

The proposed method has several advantages over existing models of importance:

- The concept is clear and easy to explain
- It is theoretically grounded in model estimation
- It can detect "zero contribution" attributes

The Problem with Traditional "Importance"

Usual importance metric (IM) $\propto$ Attribute Range = (AttributeHighPW – AttributeLowPW) / sum(Ranges)

IM is a highly salient metric of great interest to clients – "which attribute is more important?" Yet IM is unsatisfactory for several reasons. It is:

1. Affected only by best & worst attribute levels
2. Inflated by unrealistically good or bad attributes
3. Not directly related to predictive accuracy
4. Susceptible to noise (individual, across attributes)
5. Claimed every attribute is "somewhat" important

In short, traditional "importance" is an indirect measure whose connection to actual respondent preferences is unclear … yet IM is almost certain to yield an outcome that *appears* useful (for discussion, cf. Orme, 2009, p. 81).

Attribute Impact Concept

Determine the contribution of an attribute using a procedure similar to "variable importance" in machine learning random forest models (Breiman)

Outline

A. Determine a base CBC model with all attributes and find its predictive power (correct choices)
B. For each attribute - one at a time - modify its data systematically and estimate a new model.
C. If the new model is worse than the base model, then the modified attribute is "important".

Definitions

- *CBC base model* = MNL model estimated by MLE (e.g., as in Chapman & Alford, 2010)
- *Predictive power* = % of correct predictions in observed choices, when a model is developed on a training sample and then tested against a holdout sample of respondents
- Systematic modification: shuffle, only, drop, randomize (see inset below)

Estimation Code

Code is available in the Rcbc package for the R statistics environment, available from the author.

Goals

The proposed Attribute Impact (AI) intends to address each of the IM problems:

1. Uses all attribute information
2. Lower sensitivity to attribute spread (TBD)
3. Tied to model's ability to predict choices
4. Less susceptible to noise (TBD)
5. Can propose – and perhaps detect – "zero-importance" attributes

Types of "Systematic Modification" of Observed Choice Data

Shuffle: values (rows) of an attribute are randomly mixed across cards, breaking the attribute-to-choice linkage (as tested: applied to holdout sample)

Only: the attribute levels in question are retained as the only predictive variables

Drop: the attribute in question is discarded while all others are retained

Randomize: replace the attribute's data with randomly generated values drawn from the attribute level range

Shuffle

	Original data			Modified data (example)		
	Attr 1	Attr 2	Attr 3	Attr 1	Attr 2	Attr 3
Card 1-1	1	2	3	1	1	3
Card 1-2	2	3	2	2	2	2
Card 1-3	3	1	2	3	3	2
Card 2-1	2	3	1	2	2	1
Card 2-2	3	2	1	3	3	1
Card 2-3	1	1	3	1	1	3

Only

	Original data			Modified data
	Attr 1	Attr 2	Attr 3	Attr 2
Card 1-1	1	2	3	2
Card 1-2	2	3	2	3
Card 1-3	3	1	2	1
Card 2-1	2	3	1	3
Card 2-2	3	2	1	2
Card 2-3	1	1	3	1

Drop

	Original data			Modified data	
	Attr 1	Attr 2	Attr 3	Attr 1	Attr 3
Card 1-1	1	2	3	1	3
Card 1-2	2	3	2	2	2
Card 1-3	3	1	2	3	2
Card 2-1	2	3	1	2	1
Card 2-2	3	2	1	3	1
Card 2-3	1	1	3	1	3

Randomize

	Original data			Modified data (example)		
	Attr 1	Attr 2	Attr 3	Attr 1	Attr 2	Attr 3
Card 1-1	1	2	3	1	2	3
Card 1-2	2	3	2	2	1	2
Card 1-3	3	1	2	3	2	2
Card 2-1	2	3	1	2	1	1
Card 2-2	3	2	1	3	3	1
Card 2-3	1	1	3	1	3	3

Case Studies

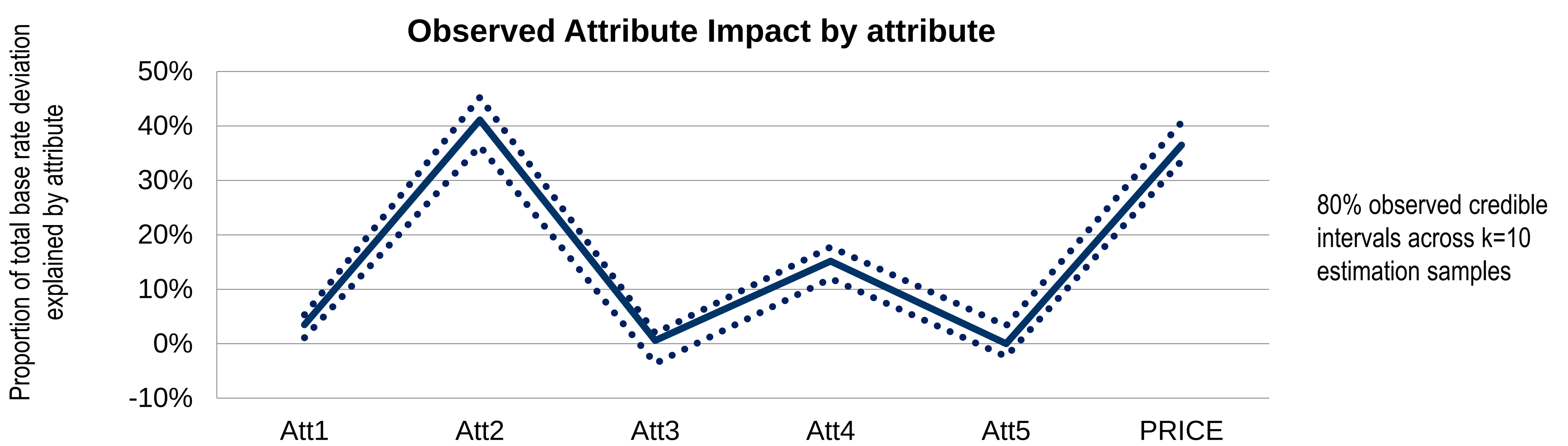
Study 1

Goal: Determine the attribute impact of 6 attributes in a CBC study of a consumer electronics item.

Data: online CBC study fielded with Sawtooth Software SSI/Web; N=202 respondents; 6 attributes including Price; 3-7 levels per attribute; K=12 choice sets per respondent.

Method: Use the "shuffle" procedure to determine 80% credible intervals for attribute impact

Result



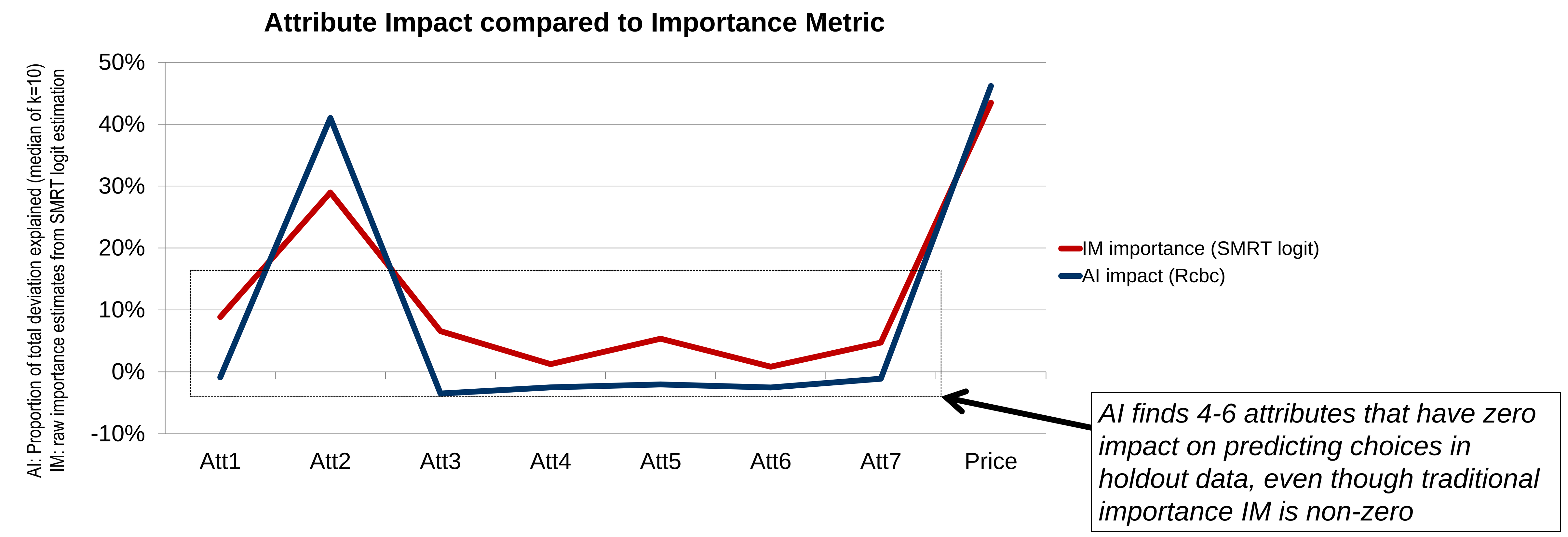
Study 2

Goal: Compare Attribute Impact (AI) to traditional Importance Metric (IM)

Data: online CBC study fielded with Sawtooth Software SSI/Web; N=792 respondents; 8 attributes including Price; 3-7 levels per attribute; K=12 choice sets per respondent.

Method: Use the "shuffle" procedure to determine mean estimate for AI; Sawtooth Software SMRT (logit model) attribute importance metric to determine IM values.

Result



Conclusion

The proposed AI measure yields results that are directionally similar to those of traditional importance IM, but are advantageous for several reasons:

1. AI is theoretically grounded in model accuracy (successful choice prediction)
2. AI can detect attributes that have "zero impact" on observed choices
3. The procedure allows bootstrapping and multiple methods of determining impact

Future work: Use HB estimation models in additional to standard MNL models
Explore suitability and differences among the "systematic modification" options

References

Breiman, L. (2001). Random forests. Machine Learning, 45:1, 5-32.

Chapman, C.N., and Alford, J.L. (2010). Rcbc: choice-based conjoint and multinomial logit estimation in R [poster]. Presented at the 21st annual Advanced Research Techniques Forum, (A/R/T Forum), San Francisco, June 2010. [Available by email from this author.]

Orme, B. K. (2009). *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*, 2nd ed. Research Publishers.

R Development Core Team (2011). R: A language and environment for statistical computing [Computer software]. Version 2.13.0. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org

Author information:

Chris Chapman
1 Microsoft Way (cchap)
Redmond, WA, 98052
chris.chapman@microsoft.com