



Sawtooth Software

RESEARCH PAPER SERIES

Perspectives on the Recent Debate over Conjoint Analysis and Modeling Preferences with ACA

Bryan Orme,
Sawtooth Software, Inc.

Perspectives on Recent Debate over Conjoint Analysis and Modeling Preferences with ACA

Bryan Orme, Sawtooth Software
Copyright 2002, Sawtooth Software

A Controversial Article Inspires Debate

In a recent article in *Marketing Research*, "What's Wrong with Conjoint Analysis," Larry Gibson argued that conjoint analysis is not needed given the simplicity and predictive accuracy of self-explicated methods. He wrote that conjoint analysis cannot handle the number of factors that self-explicated methods can and that clients usually need to study. This argument may be valid for full-profile conjoint, but it largely ignores the strengths of the more flexible hybrid methods, like ACA. Regarding hybrid methods, he expressed that self-explicated methods produce valid enough data that subsequent conjoint tasks are not justified.

It is important to note that Gibson's firm (Eric Marder & Associates) has developed a unique method of self-explicated scaling called "SUMM," (see a description of the technique in the useful book The Laws of Choice by Eric Marder). In SUMM, respondents record their preference for each level of each attribute using an "unbounded scale." Rather than using a predefined numeric scale, respondents write "L" letters to indicate that they like a level (as many "L" letters as they wish—the more used, the greater the preference). Respondents write "D" symbols next to levels they dislike (as many "D" letters as they wish—the more used the more disliked the feature). A "N" letter is used to signify neutral. "Utilities" are developed by scoring +1 for every "L" and -1 for every "D." The technique, as described, employs paper-and-pencil surveys.

Gibson's position swims against the prevailing tide of practice and opinion in our industry—which by itself doesn't mean he is wrong. His article clearly struck a nerve, as a number of researchers published responses in defense of conjoint analysis in the subsequent issue of *Marketing Research* (Paul Green & Abba Krieger, Rich Johnson, and Dick McCullough). These authors provided evidence regarding the value of full-profile methods for small conjoint problems, and asking conjoint questions in addition to collecting self-explicated ratings data for larger problems.

I suspect Gibson knew his article would spark controversy, and tailored his prose for healthy effect. I don't fault him for stirring the pot, but I am concerned about some serious errors of fact and logic in his article. In his defense, his self-explicated technique is different (and perhaps superior among some populations, for certain product categories, and in paper-and-pencil mode) from the one the responding authors used for comparisons. Until more comparative work is done with his SUMM method, the argument cannot be resolved by referring to current research. But future research cannot validate or disprove the use of SUMM for all situations. There are many useful techniques for modeling preferences, and no one technique dominates the others.

We should not dismiss his claim that SUMM seems to work better for him than the conjoint techniques he has tried. But it is misleading to imply that a single method, such as SUMM, generally dominates conjoint and choice analysis.

I wonder how the SUMM technique might be used to model interaction effects, alternative-specific effects (designs where some attributes only apply to certain brands), conditional pricing, or cross-effects, which are important for many choice studies. I also wonder whether SUMM can work well for computerized interviewing.

Despite Gibson's recent condemnation of conjoint analysis, when smaller problems involving, say, brand, package and price are being studied, Gibson's firm uses a technique they call STEP. STEP is really just a single full-profile CBC (discrete choice) task, featuring an allocation-based response. Respondents allocate a fixed number (say 10) of stickers across alternative products within a set. Gibson misleads his readers by seeming to dismiss the use of conjoint analysis when his firm often uses this CBC-like approach.

But Gibson and his colleagues are quick to distinguish STEP from CBC. They argue that only one choice task should ever be asked of a respondent (per product category being studied) as subsequent tasks result in biased information. Indeed, if money were no object, one could achieve good results using very large sample sizes, with each respondent receiving one task. However, Rich Johnson and I showed in a 1996 paper entitled "How Many Tasks Should You Ask in CBC Studies" (available at www.sawtoothsoftware.com in the Technical Papers library) that the responses to subsequent tasks provide similar information, with a moderate but predictable bias. Brand becomes less important in later tasks, and price becomes more important. I would argue that the responses to the first CBC task might reflect what happens the first time a price change occurs (buyers may not notice and might stay loyal to their favored brand). But, later tasks may, in some cases, better reflect long-range equilibrium share. But what makes asking respondents additional tasks so valuable is that it enables us, under HB, to estimate individual-level utilities. Individual-level utilities resolve many IIA problems that plague aggregate-level modeling.

I haven't used the self-explicated SUMM technique, but it is likely to lead to greater discrimination among most and least important attributes than the basic rating-scale method used in ACA. Green correctly noted in his article "What's Right with Conjoint Analysis?" that respondents do not discriminate enough among the most and least important attributes when using self-explicated importance rating scales. In practice, I often observe importance ratings from the self-explicated section of ACA that show on average a 2:1 ratio between two attributes' importances. But when the conjoint information is analyzed separately, derived importances often suggest a 3:1 ratio or even greater.

Personal Perspectives on ACA

In the nine years of fairly intensive experience that I have with ACA, I admit to having had a hot and cold relationship with the technique. Upon first seeing ACA, I was enthralled. Later, I became disenchanted when I observed how unimportant attributes seemed to be ascribed too

much importance, and the importance of critical attributes seemed to be damped. ACA's pricing information was also disappointing, with price receiving too little importance. ACA simulators would often over-predict preference for expensive, feature-rich products. Another concern revolved around the assumption of linear utility increments for *a priori* or ranked attributes within the priors.

Jon Pinnell helped me overcome my dislike for ACA in pricing problems, by documenting both the bias and a solution (that he and prior researchers had been using) called "dual conjoint" (see "Multistage Methods for Measuring Price Sensitivity"). Dual conjoint requires that respondents complete some full-profile (or near-full profile) conjoint tasks in addition to ACA. The price weight from ACA can then be adjusted to fit the additional data. For conjoint problems involving more than six or so attributes, this solution often works reasonably well. Peter Williams later published a paper at our conference entitled "Calibrating Price in ACA: The ACA Price Effect and How to Manage It" that uses fixed holdout choice tasks rather than a separate full-profile experiment. Williams didn't invent this approach, but he was the first to formally present it at our conference. The holdout approach is easier to implement, and I believe it is more commonly used than the dual conjoint approach. Both of these papers are available at www.sawtoothsoftware.com in the Technical Papers library.

Enter Hierarchical Bayes

Even after overcoming the pricing hurdle, I remained concerned that ACA "flattened" attribute importances. About three years ago, I pondered whether something could be done so that the importances from the self-explicated section in ACA might provide greater discrimination. Reading about the "unbounded scale" approach in SUMM further stimulated my thinking in this direction. Within the next few months, Rich Johnson completed the ACA/HB product for hierarchical Bayes estimation of ACA. I marveled at how well ACA/HB improved ACA modeling. After this breakthrough, I realized that HB provided an excellent solution and discarded ambitions to search for a modified priors question. Let me discuss how ACA/HB largely dispensed with most of my lingering concerns.

The traditional ACA estimation combines the metric information from the priors and pairs section using OLS. If the priors importance weights are too flat, then the resulting final utilities are also biased toward flatness. Moreover, if attributes are ranked or specified as *a priori*, ACA assumes equidistant utilities within the priors, again biasing the final utilities toward this assumption. In contrast, the default estimation technique for ACA/HB uses the priors information to enforce ordinal constraints only. The metric information, which determines the importance of attributes and scaling of the levels within each attribute, is estimated solely from the conjoint pairs.

ACA/HB usually leads to greater discrimination in the importance of attributes than when using OLS estimation. It leads to better prediction of holdout choices for all but two data sets I know of. Even after using ACA/HB, you will usually see that ACA importances still show less discrimination than those derived from full-profile methods—especially CBC. Why does this happen? Respondents use simplification strategies to deal with full-profile tasks that include

many attributes. When faced with so much information, they focus on just the top few features. I suspect that ACA's flatter importances may be quite predictive of actual purchases for many high involvement product categories wherein which buyers make careful decisions.

I'm surprised that so many ACA users have not yet embraced ACA/HB. Yes, it is a complex statistical procedure—but it is also a very simple product to use. By using the defaults, it is as simple as pressing a few keystrokes and then waiting for the run to finish. With today's fast 1 GHz+ speed computers, it usually takes 30 minutes to an hour per utility run for a typical data set with 500 respondents. Let me discuss a recent project we were involved in to illustrate HB's value.

Why I Prefer HB Estimation: An Example

Our research wing, Sawtooth Analytics, was contracted by a recent ACA buyer to assist with a pricing project. The study involved 11 attributes, most of them being binary (on/off) features. We suggested ACA, with follow-up full-profile CBC-like holdout questions. Each respondent received three holdout tasks, and each holdout featured a selection from three possible products: a cheap one (with few features), a mid-priced product, and an expensive feature-rich product.

We used the regular OLS utilities to predict choices to the holdouts and saw the typical ACA pricing problem: too much share (by a significant margin) for the most expensive products in each set. We coached our ACA client regarding the adjustment for price (multiply the price part worths by a factor greater than unity, and import the new results back into the market simulator). After adjusting price by about a factor of 4, the fit was much improved for the least and most expensive products, but predictions were dismal for the medium-priced product. The errors in prediction weren't due to just price bias. The average MAE (mean absolute error) for prediction of all products across all three holdout sets was over 12. Those familiar with this type of analysis will recognize that this is a poor fit. An MAE of 12 means, for example, that the target share is 30, but we predict 42.

Next, we used ACA/HB to reanalyze the ACA data. Many of the features within this product were quite new to respondents, and we suspected that they may not have been able to settle on a reliable scaling of preferences and importance early on in the priors section. To test this assertion, we used ACA/HB to estimate two separate sets of utilities. In the first set, we fit the pairs information after applying the priors information as ordinal constraints. For the second set, we estimated utilities only using the information from the conjoint pairs. Comparing the two sets of utilities, we noted a large discrepancy in the importance of many of the features—particularly price. We used both sets of ACA/HB utilities to predict the holdouts, again adjusting price for best fit. The ACA/HB run that fit only the pairs information provided the best fit (not a typical finding, given our experience with many data sets). These utilities required a price weight of less than 2, and the resulting MAE was around 4 (significantly better than the MAE of 12 when using the default ACA OLS utilities).

We realized that we were discarding information by ignoring the priors data, leading to lower within-respondent precision of estimates. But we believed the priors information in this case

damaged predictive accuracy. The sample size was a healthy 1000 (which meant that the added error would have minimal effect on the precision of aggregated share predictions). Since the client was most concerned about achieving an accurate market simulator, it seemed clear that using HB with the conjoint pairs information only was the right choice.

In short, having ACA/HB makes one a much more confident and competent ACA researcher. The research is much more defensible, because it applies a leading-edge statistical technique with a more correct method for combining priors and pairs information. The researcher has greater options for building a model that is more predictive of real-world behavior. Under OLS, there is one default approach, and the typical user can do little but accept it.

In this most recent experience, by using ACA/HB we were able to identify a disconnect between the priors and pairs information. We ended up discarding the priors. Through HB's ability to stabilize estimates for individuals by borrowing information from the population parameters, we estimated more realistic utilities than under traditional ACA estimation and delivered a more predictive market simulator.

Suggestions and Conclusion

Despite my improved outlook for ACA, I wouldn't use it exclusively. I am suspicious of claims that there is one best technique for all (or even most) preference modeling situations. For studies involving about six or fewer attributes and sample sizes greater than 100, I generally first consider CBC. When sample sizes become quite small, I often favor traditional full-profile conjoint (with HB estimation under HB-Reg). For problems involving more attributes than is reasonable for full-profile methods, I tend to favor ACA—paying special attention to adjust for the price bias, and using ACA/HB to try different ways of combining the priors and pairs data.

Even after choosing the approach, there are numerous issues to be resolved regarding the formulation of attribute levels, choice of response scales, allocation versus discrete choice, full-versus partial-profile, number of alternatives per task, number of pairs/tasks. These decisions are often vexing, but it is the challenge that makes the process all the more enjoyable. If there was just a single easy and best approach for all situations, our profession wouldn't be nearly as fulfilling and managerially useful.