A User's Guide to Conjoint Analysis

Before starting out, you need to know where the land mines are.

Conjoint analysis is the most powerful and important family of analytic techniques in marketing research. But it's only the best method if you do it right. Conjoint techniques tend to be complex, which means there are more ways than ever to make mistakes. Learning about the possible pitfalls of conjoint analysis—and how to avoid them—can help researchers use this technique more successfully..

Author: Paul Richard "Dick" McCullough

A version of this article was published by *the AMA* in
MARKETING RESEARCH MAGAZINE
Summer 2002

A User's Guide to Conjoint Analysis

Before starting out, you need to know where the land mines are.

Executive Summary

Conjoint analysis is the most powerful and important family of analytic techniques in marketing research. But it's only the best method if you do it right. Conjoint techniques tend to be complex, which means there are more ways than ever to make mistakes. Learning about the possible pitfalls of conjoint analysis - and how to avoid them - can help researchers use this technique more successfully.

The term "conjoint" is said to be derived from two words: "considered jointly." Whether or not this is actually true, it illustrates the fundamental idea behind this technique. In conjoint analysis, researchers describe products or services by sets of attribute values or levels and then measure respondents' purchase interest. Thus, a respondent might be shown a red Ford pickup with a V-8 engine priced at $20,000. He or she must "consider jointly" all the attributes describing that pickup when deciding whether or not to purchase the vehicle.

The primary purpose of conjoint analysis is to model human behavior, usually purchase behavior. By measuring purchase interest in a "complete" product or service, conjoint analysis captures the essential dilemma of market choice: The perfect product is seldom available, but lesser alternatives are.

By forcing respondents to trade off competing values and needs, conjoint analysis uncovers purchase motivations the respondent may be unwilling to admit to and may even be unaware of.

Conjoint analysis addresses big issues with specific answers. As a result, when it fails, it often fails spectacularly. Nonsense conclusions such as "doubling price will double sales" don't sit well with experienced marketers. Study disasters contribute not only to the poor reputation of conjoint analysis within some organizations, but also to the reputation of the marketing research department in general.

Conjoint failures are generally the result of researchers who fail to properly design their conjoint studies or correctly interpret the output. Powerful, user-friendly software gives us opportunities to make mistakes we may not even be aware of.

What's Your Technique?

Conjoint analysis is a growing family of techniques broken into three branches: ratings-based conjoint, choice-based conjoint, and hybrid techniques. I do not include self-explicated scaling as a stand-alone conjoint technique since it does not force

The first step in doing conjoint analysis right is to pick the most appropriate method for your particular objectives and circumstances. In principle, the right technique will be the one that most closely mimics your marketplace dynamics. In practice, that will most often be choice-based conjoint, which offers respondents a series of choice sets—generally two to five alternative products. Respondents can pick any of the available alternatives or even elect not to buy if none of the alternatives in that choice set are sufficiently attractive. This format closely mimics buying environments in markets with competition.

Ratings-based conjoint involves monadically rating individual product alternatives or pairwise rating two product alternatives simultaneously. No-buy options aren't easily accommodated in ratings-based conjoint. This technique may be more appropriate for non-competitive markets, such as oligopolies, monopolies, or emerging categories.

Hybrid techniques, approaches that combine self-explicated scaling with either ratings-based conjoint or choice-based conjoint, are generally appropriate when a large number of attributes must be included. Sawtooth Software's Adaptive Conjoint Analysis (ACA) is a widely used example of a hybrid technique.

Hybrid techniques, approaches that combine self-explicated scaling with either ratings-based conjoint or choice-based conjoint, are generally appropriate when a large number of attributes must be included. Sawtooth Software's Adaptive Conjoint Analysis (ACA) is a widely used example of a hybrid technique.

Both ratings-based conjoint and choice-based conjoint can be conducted as full-profile or partial-profile studies. Full-profile tasks involve one level from every attribute in the study. If there are six attributes in your full-profile study, then each product alternative will have six attribute levels defining it. Partial-profile tasks involve a subset of the total set of attributes. If there are thirty attributes in your partial-profile study, then each product alternative may have six attribute levels defining it.

Full-profile studies should ideally contain no more than six attributes. The critical issue is to define products simply enough to be understood by respondents. If your attributes are extremely complex and unfamiliar, perhaps six is too many. If they're extremely simple and familiar, you may be able to include more than six.

Partial-profile designs can include up to 50 or more attributes. Partial-profile designs, a relatively recent development in conjoint analysis, typically compete with hybrid designs when a large number of attributes must be included.

Full-profile designs are generally preferred over partial profile designs if the number of attributes is sufficiently small because full-profile designs can accommodate interaction terms more easily, require less sample, and are more familiar to most market researchers. Full-profile designs are generally preferred over hybrid designs if the number of attributes is sufficiently small because hybrid designs usually can't accommodate interaction terms and are considered to employ a less natural question format.

A potential concern for any approach that accommodates a large number of attributes is attribute additivity (AA). Seldom mentioned in the literature, AA is the phenomenon where a large number of less important attributes may overwhelm one or two extremely important ones. For example, a feature-rich product may have more total utility than a low-priced one simply because all the small utility weights of the various product features, when summed, exceed the utility weight of the price attribute. There's currently no consensual "right" way to address this problem. One possible approach is to, on an individual level, limit the number of attributes included in model simulations to the six most important. This is consistent with the rationale for limiting the number of attributes in a conjoint task to six.

Attributes and Levels

If one research objective is to understand the impact of introducing a new brand into your category, for example, it's essential that brand be an attribute in your study and the new brand be a level within the brand attribute. There are two problematic attribute-related issues that merit consideration: number of levels effect (NOL) and attribute range (AR). With NOL, attribute importance is affected by the number of levels specified in the design. For example, if price has two levels ($6 and $12) in one study and price has four levels ($6, $8, $10, and $12) in another study that's exactly the same as the first (except for the price levels), price in the second study will be more important than in the first. Other than attempting to keep the number of levels of all attributes as close to one another as is practical, there's no known solution to this problem. ACA, however, does suffer substantially less from NOL than other techniques.

Similarly, attribute range also affects attribute importance. If, in the second study above, price only had two levels, but those levels were $6 and $24, price would again show more importance in the second study. The best we can do here is to define the minimum range of attribute levels necessary to realistically address the research objectives for each attribute in the study.

Breaking It Down

Conjoint studies, with the notable exception of ACA, require an experimental design to determine the appropriate set of product combinations for testing. Commercial software today offers powerful flexibility in study design and can be surprisingly easy to use. Often, design software provides diagnostic information allowing the researcher to evaluate the design. However, designs of any complexity should be tested with synthetic (or other) data prior to field to ensure their viability.

One design issue to note involves attribute specification. Numerical attributes, such as price, can be defined as part-worth attributes or vector attributes. If defined as a part-worth attribute, each level within price would receive its own utility weight. If defined as a vector attribute, one utility weight would be calculated for the attribute as a whole and would then be multiplied by each level value to determine the utility weight by level. Part-worth attributes require more information to estimate, but vector attributes assume linearity. The best approach is to define all attributes as part-worth attributes so that you are free to model non-linear relationships. Price, for example, is often non-linear.

Experimental design is an important part of conjoint analysis, and commercially available design software is extremely powerful. Computer-assisted interviews and Web-based interviews both allow each respondent to receive a set of conjoint tasks unique to him or her, a feature generally impractical with paper and pencil studies. This facility greatly enhances your study's design efficiency. Thus, using individualized interviews may allow you to use fewer tasks, have smaller sample size, or perhaps simply complete a difficult and ambitious study successfully.

Tasks

There are three types of conjoint questions that should be included in any conjoint exercise: warm-up, conjoint, and holdout tasks. Sometimes respondents need time to warm up because they take awhile to "get it." Their responses don't stabilize until they've done a few tasks. Two to four warm-up tasks should be included at the beginning of the conjoint exercise to educate and familiarize the respondent to the exercise at hand. As an added safeguard, task order should be randomized whenever possible.

Holdout tasks are tasks that won't be included in the utility estimation process. They are "held out" of the analysis and used to validate the model after utility weights have been estimated. Even if your study is a ratings-based conjoint study, your holdout tasks should be choice-based to make model validation more meaningful.

As a practical matter, clients often have specific scenarios they're interested in testing. These scenarios can be specified in the holdout tasks, with no compromise to the study design. The holdout tasks can then serve the dual purposes of validating the model and providing "hard" data that some clients will find more credible than model simulations. Another practical suggestion is that holdout tasks should be designed so that responses aren't flat across alternatives. This will make validating the model easier.

For choice-based conjoint, studies show 20 or more tasks can be given to respondents without degradation of data quality. That number is largely dependent on the number of attributes displayed, the familiarity of respondents with the category and terms, the level of involvement the respondent has with the category, the length of the questionnaire prior to the conjoint section, and numerous other factors.

If you want a conjoint study that works, be brief. This is a surprisingly difficult standard to meet. Most choice-based studies I've designed have worked well with as few as 10 tasks. Add in two warm-up tasks and two holdout tasks and you're already up to 14, at a minimum.

Sample size

Another important question with no clear answer is sample size. Little literature exists examining the impact of sample size on conjoint model error, but current evidence suggests that models can be reliably estimated with samples as low as 75, regardless of type of conjoint technique employed. However, keep in mind that 75 is the minimum size of any analytic cell you might want to examine. Thus, if you had a market with five regions and you wished to model each region separately, you'd need a sample of 375 (5 times 75). If you wanted to model males and females separately within each region, your minimum sample size would be twice that, or 750.

Although numerous technical pitfalls exist, the most common error in commercial conjoint studies is probably asking respondents questions they can't answer accurately. If respondents don't understand terms and concepts, if they're confused by product descriptions that are too complex and lengthy, or if they become disinterested or tired due to questionnaire length, your analysis will suffer.

To make sure respondents are capable of answering questions, be sure all attributes and levels are clearly defined prior to the conjoint exercise. Making a glossary of terms available for the respondent to review prior to the conjoint exercise and as a reference throughout the exercise can help. Visually organize the conjoint tasks to help the respondent quickly understand the choices. Don't include so many attributes in each product alternative that only a chess champion could keep them straight. Always pretest conjoint studies to confirm you can implement them. Statistical diagnostics won't tell you if humans can or cannot comprehend the questions you're about to ask.

There is an essential size problem that all designers of conjoint studies face. If the estimated model is fairly complex, it will require a great deal of information to estimate it, particularly at the disaggregate (individual) level. Experienced researchers know this information can be extracted through (1) number of conjoint tasks, (2) complexity of conjoint tasks, (3) sample size, (4) experimental design, and (5) utility estimation technique.

Utility Estimation and Models

Once data have been collected, the researcher faces another set of options. Historically, ratings-based conjoint utilities have been estimated using OLS (Ordinary Least Squares) regression at the individual respondent level, and choice-based conjoint utilities have been estimated using logit regression at the aggregate (total sample) level. Hierarchical Bayes (HB) modeling has changed all that.

In general, disaggregate models are preferred over aggregate models. There are several reasons for this, but the primary reason is that aggregate models don't capture heterogeneity. For example, consider a sample given the choice of Coke or Pepsi. If half the sample loves Coke and hates Pepsi and the other half loves Pepsi and hates Coke, an aggregate model will show the total sample indifferent to brand. The Coke lovers and the Pepsi lovers cancel each other out. In a disaggregate model, brand will appear to be extremely important because all the Coke lovers will exhibit large utilities for Coke and all the Pepsi lovers will exhibit large utilities for Pepsi.

Choice-based conjoint has historically been preferred over ratings-based conjoint because of its more natural question format, its ability to handle interaction terms, and its ability to easily model the no-buy option. Its biggest drawback has been its inability to generate disaggregate models. HB allows for individual utilities estimation of choice-based conjoint data. It has also been shown that HB estimates are superior to OLS regression estimates for ratings-based conjoint.

The primary drawback to HB estimation is that it is computationally intensive. Computation time can run from 30 minutes to 30 hours, depending on the sample size, the number of parameters being estimated, and the power of the computer running the calculations. In general, the advantages of HB far outweigh this one disadvantage.

Current research suggests that finite mixture models may estimate individual level choice utilities as well as HB. However, HB models have proven to be extremely robust, and recently introduced user-friendly HB software eliminates any excuse for not using this technique.

In some software packages, constraints can be included in the estimation routine that force certain attribute levels to always be the same or higher than other levels. For example, you may feel strongly that consumers truly would prefer to buy your product at a lower price. Therefore, you know a priori that the utility of the lowest price level should be greater than or equal to every higher price level. You can constrain your utility estimates to conform to this relationship. It's been shown that constraints tend to improve holdout prediction accuracy. However, a word of caution about constraints. The goal of most research is to learn how the market works, not to confirm what we already know about how the market works. Sometimes surprises aren't bad research - they're insight. I prefer letting the data run free as often as possible. If necessary, the data can always be rerun using constraints.

After utilities are estimated, preferably at the individual level using HB, simulations can be run. The five methods of simulation include (1) first-choice, (2) share of preference, (3) share of preference with correction, (4) purchase probability, and (5) randomized first-choice (RFC).

First-choice models are only available for disaggregate data and follow the maximum utility rule. That is, if three products are included in a scenario, each individual is assumed to pick the product for which his or her total utility is highest. This approach often suffers from volatility (i.e., minor changes in product configurations can result in unrealistically large shifts in preference shares).

Share of preference models can be run against either aggregate or disaggregate data. These models distribute preference proportional to each product's total utility. For example, if, in an aggregate model of two products, product A had total utility of 10 and product B had total utility of 20, product A would have 33% share of preference (10/(10+20)) and product B would have 67% share of preference (20/(10+20)).

Share of preference models are less volatile than first-choice models, but are subject to a bias called irrelevance of independent alternatives (IIA). If two products are similar, such as a red bus and a blue bus in a transportation alternatives study, their net share is overestimated. In effect, there is double counting. Share of preference models with correction are an attempt to adjust for the IIA bias. First-choice models are not subject to IIA bias.

In my opinion, the best approach is a recently developed technique called randomized first-choice (RFC), which exhibits much less IIA bias than share of preference models and is less volatile than first-choice models. It also offers several ways to tune the model for increased accuracy.

Regardless of the simulation technique selected, the model should be validated and tuned. Market scenarios should be defined and simulated that replicate the choices available in each holdout task. The model predictions of choices should be compared to the actual choices made by respondents. For disaggregate models, there are two measures of model accuracy, hit rates, and mean absolute error (MAE). For aggregate models, only MAE is appropriate.

Hit rates are calculated by comparing the choice predicted for an individual respondent by the model (using the maximum utility rule) to the actual choice made by the respondent. When the model correctly predicts the respondent's choice, it's counted as a hit. The total number of hits divided by total sample size equals the hit rate.

MAE is defined to be the sum of the differences between predicted share of preference and actual share of preference for all products in a holdout task divided by the number of products in the holdout task. Initial hit rates and MAE (prior to model tuning) can be compared to hit rates and MAE from a random model to give the researcher a feel for how successfully the model has been able to capture and model respondent choices.

For example, if there are four choices available in a holdout task—say three products and no-buy—a random model could be expected to have a hit rate of 25% (1/4). If your initial model has a hit rate of 65%, you can feel somewhat assured that your model performs better than random.

Similarly, MAEs for a random model can be calculated by subtracting 25% from the percent of respondents who picked each of the four options, summing the absolute value of the differences and dividing by four. If your random model has an MAE of 12 and your model has an MAE of four, again you can feel somewhat reassured.

For this analysis, you want to construct holdout tasks likely to have unequal preference across alternatives. In general, hit rates above 60% and MAEs below five points will reflect a reasonably good fitting model.

Once initial hit rate and MAE calculations have been examined, model tuning may be appropriate. Share of preference and RFC models can be tuned to maximize hit rates and minimize MAE. Tuning the model will increase its accuracy and, therefore, managerial utility. In some rare and fortuitous instances, actual market data can be used to tune the model, rather than holdout tasks.

Doing It "Right"

Although there are so many exceptions that the word "right" loses much of its meaning, I would generalize the "right" method of doing conjoint analysis as follows:

- Choice-based conjoint
- Including warm-up and holdout tasks
- Hierarchical Bayes for utility estimation
- RFC for market simulations
- Tuning the final simulator

In 1990, Batsell and Elmer wrote "The introduction in 1971 by Green and Rao of conjoint analysis marked a significant step in the evolution of marketing research from art to science." I agree. With a heritage in both psychometrics and econometrics, no marketing research technique comes close to offering either the managerial power or the economic efficiency of conjoint analysis.

But conjoint analysis is a complex family of techniques. Many difficult decisions await the conscientious researcher, often with no clear cut, "right" answer. Conjoint analysis has pushed marketing research much closer to a science, but it remains an art. The diligent researcher will be aware of both the possible pitfalls and the available antidotes. In the end, the reward far outweighs the effort.

A version of this article was published in *MARKETING RESEARCH MAGAZINE*, Summer 2002

We are an independent ***marketing research consulting firm*** *dedicated to helping you make the most informed, insightful marketing decisions possible. We specialize in technology, consumer, and new product research, and are well recognized for our* ***State-of-the-Art Research*** *techniques.*

Ultimately, we provide more than just technical expertise. We focus on developing ***pragmatic solutions*** *that will have a positive impact on the profitability of our clients.*

CONTACT US:

Telephone: 650-823-3042

General Inquiries:
info@macroinc.com

Advanced Analysis Inquiries:
analysis@macroinc.com

richard@macroinc.com

www.macroinc.com