

Getting the Most Bang From The Fewest Questions:

A New Approach to Designing and Analyzing Conjoint Studies

Conjoint Studies are renowned for yielding a great deal of strategic insight. In a single conjoint study, one can address price optimized relative to profit, revenues, unit sales or market share, optimal product feature set, cannibalization patterns, market response to competitor actions or new product introductions, dollar value of brand equity and many other issues.

Author: Paul Richard "Dick" McCullough

Published in the
PROFESSIONAL MARKETING RESEARCH SOCIETY
CONFERENCE PROCEEDINGS,
April 2003,
Vancouver, B.C., CANADA

Getting the Most Bang From The Fewest Questions:

A New Approach to Designing and Analyzing Conjoint Studies

Introduction

Conjoint Studies are renowned for yielding a great deal of strategic insight. In a single conjoint study, one can address price optimized relative to profit, revenues, unit sales or market share, optimal product feature set, cannibalization patterns, market response to competitor actions or new product introductions, dollar value of brand equity and many other issues.

However, conjoint studies capable of providing such rich findings tend to be somewhat lengthy and often time-consuming to field. A typical choice-based conjoint study, for example, will have 12-20 choice tasks in addition to any other questions included in the survey. If the number of attributes is large, even more choice tasks may be desired. Often, conjoint exercises can be confusing and fatiguing experiences for respondents.

The development of Hierarchical Bayes techniques in the late 90s not only has allowed the estimation of individual level utilities for choice-based conjoint but also for the more accurate individual level utility estimation of ratings-based conjoint. What has been generally ignored by the commercial research community is the fact that the efficiency of HB also allows for the reduction of the number of choice tasks required to support individual level utility estimation. Current practice is to design choice-based conjoint studies as if HB did not exist and then to apply HB to the resulting data. This is safe but inefficient.

The introduction of web-based surveys into common practice has given the practitioner a relatively inexpensive method for creating large sample studies. The combination of large samples with the efficiency of HB may create an opportunity to dramatically reduce the number of choice tasks per respondent necessary for estimating acceptably accurate disaggregate models.

There are several advantages to reducing the number of choice tasks/conjoint ratings shown per respondent:

- 1) time in field may be shortened,
- 2) data collection costs can be less,
- 3) additional questions can be included in the survey to address other issues,
- 4) data can be collected in more modes,
- 5) Number of level effect, order bias, learning bias, framing bias and respondent fatigue may all be minimized with an abbreviated task set.

There are also several potential disadvantages:

- 1) early responses and late responses appear to differ: early responses tend to emphasize brand while later responses may emphasize price,
- 2) later responses may be better predictors than early responses,
- 3) respondent propensity to select “none” may increase in the later tasks.

Potential applications for the abbreviated task set approach could include: Any study benefiting from a very brief interview but capable of generating a large sample, e.g., Trade Show and Conference floor intercepts, Web surveys or Telephone surveys, Studies combining conjoint with other issues such as segmentation, brand positioning or attitude and usage, resulting in an excessively long interview, Realistic environment studies, e.g., Laboratory simulations or Control store tests.

The purpose of this paper is to, in an empirical way, assess the net effect of reduced task set to model error, rather than address each potential factor separately.

Method

For three commercial data sets, each of large sample size ($n > 1,700$), models were estimated using the total sample and all available tasks, excluding holdouts. Additional models were estimated for various reduced numbers of tasks and various smaller sample sizes. The subsamples were generated by drawing independent random samples from the original data set. Each resulting choice model was evaluated using Mean Absolute Error (MAE) and hit rates, where appropriate. All three data sets included at least one hold out card.

Mean Absolute Error is the average of the absolute difference between each predicted and actual preference share for each holdout card (fixed task). When more than one holdout card is available, the reported MAE will be the average of the MAEs of all holdout cards. MAEs are calculated at the aggregate level. In practice, MAEs of 4 or 5 are typical and acceptable.

Hit rates are the percentage of individuals for whom their preferred alternative in a holdout task is correctly predicted by their individual-level model. Hit rates are calculated at the individual level. In practice, hit rates of 60% or higher are typical and acceptable.

Each of the three data sets analyzed for this paper is described below in Table 1.

Table 1. Data Set Profiles

| Data Set | Beverages | Games | Books |
|--|---------------------|-----------------|-----------------|
| Sample Size | 2,367 | 3,276 | 1,794 |
| Attributes | 15 | 18 | 5 |
| Levels | 57 | 42 | 16 |
| Parameters | 43 | 25 | 11 |
| Random Tasks | 4 | 6 | 4 |
| Fixed Tasks | 1 | 2 | 2 |
| Alternatives per Task (excluding no-buy) | 8 | 3 | 3 |
| No buy alternative | no | yes | no |
| Data Collection In-person | Online | Online | |
| Survey Versions | 3 | 999 | 999 |
| Analytic Method | Constant Sum Choice | Partial Profile | Discrete Choice |

Note that all HB runs were made with 100,000 iterations burned and every 10th of the next 10,000 saved. This number of burned iterations is much larger than typical. It was discovered early on in the analysis that HB does not converge as quickly when the number of tasks is reduced and that 100,000 burned iterations was a safe number to use for all data sets. It should also be noted that computer run time was substantially lengthened by the increase in number of iterations. Run times for this analysis varied from two hours to 30 hours depending on study parameters and computer capabilities. Computers used in this analysis had clock speeds that ranged from 333 MHz to 1.2 GHz. In all cases, the experimental designs were tested prior to field by estimating an aggregate multinomial logit (MNL) model using random data. Model convergence was confirmed and the standard errors of all partworths were examined for uniformity and magnitude.

Results

In general, acceptable MAEs and hit rates were consistently obtained with 4 or fewer tasks per respondent, even when there were a large number of parameters to be estimated. In two cases, MAEs were acceptable with models estimated using only one task per respondent.

Sample size was less of a factor than anticipated. Although model error does increase as sample size decreases, adequate models were consistently obtained using relatively few tasks (4 or fewer) and sample sizes as small as 500 to 1,000.

Beverages

The Beverages Study was conducted among grocery shoppers in a South American country. Respondents were shown a series of 5 boards depicting 8 different beverage products they might buy in a grocery store. Four of the boards were used to estimate individual level choice utilities. The fifth board was used as a holdout task.

The interviews were personal, one-on-one interviews conducted in six regions within the South American country. Sample size was approximately 400 per region. Respondents were shown a board of 8 alternative beverages and asked how many of each they would buy if these were the beverages available to them in the grocery store they typically frequented. These numeric data were converted to constant sum for the purpose of utility estimation.

In practice, MAEs of 4 or 5 are typical and acceptable. For the Beverage study in Table 2 below, MAEs of under 3 were obtained using just 1 choice task per person with a sample size of 2,367 or 2 tasks per person with a sample size of 500. This is particularly remarkable considering the large number of parameters to be estimated (43) and the small number of questionnaire versions available (3).

Table 2.
Beverages MAEs (43 parameters)

| n= | 2,367 | 1,000 | 500 | 200 |
|----------------|-------|-------|------|-------|
| <u>tasks =</u> | | | | |
| 4 | 1.05 | 2.11 | 2.42 | 3.94 |
| 2 | 1.86 | 2.2 | 2.9 | 3.75 |
| 1 | 2.89 | 4.33 | 6.28 | 10.29 |

Hit rates were not calculated for the Beverages data set because hit rates are not appropriate for constant sum data.

Games

The Games Study was an online conjoint study among registered users of a particular online gaming site. Registered users were sent an email inviting them to participate in an online study. For respondent convenience, a hyperlink to the online survey was embedded in the email invitation.

The Games study was designed using partial profile choice. Approximately one-third of the total number of attributes was represented at any one time. Thus, a sample of 1,000 in the Games data set is roughly equivalent to a sample size of 330 using a full-profile data set, in terms of attribute level exposure. The robustness of HB is evident in its ability to estimate good models with sample sizes as low as 200 and as few as 3 tasks (Table 3) for this partial profile design.

Note that where larger sample size or greater number of tasks yields MAEs of 4 or greater, further MAEs are not calculated. Also note there is some slight instability in MAE estimates due to sampling error in the subsample draws.

Table 3.
Games MAEs (25 parameters)

| n= | 3,276 | 2,000 | 1,000 | 500 | 200 |
|----------------|-------|-------|-------|-------|------|
| tasks = | | | | | |
| 6 | 1.79 | 1.93 | 1.87 | 2.41 | 2.56 |
| 4 | 1.85 | 2.51 | 2.83 | 3.33 | 3.59 |
| 3 | 2.63 | 2.64 | 2.75 | 3.4 | 3.85 |
| 2 | 3.52 | 3.36 | 3.34 | 3.61 | 5.37 |
| 1 | 4.47 | 5.89 | 4.94 | 14.19 | |

Hit rates are unusually high (see Table 4). This is most likely due to a large no-buy share. What is noteworthy, however, is the modest decline in hit rate as task number decreases.

Table 4.
Games Hit Rates

| | n=3,276 |
|----------------|---------|
| tasks = | |
| 6 | 81.4% |
| 4 | 79.9% |
| 3 | 78.4% |
| 2 | 78.3% |
| 1 | 76.3 |

Books

The Books Study was also an online conjoint study. Respondents were shoppers of a particular bookstore. Shoppers were sent an email inviting them to participate in an online study. For respondent convenience, a hyperlink to the online survey was embedded in the email invitation.

The Books model, as shown in Table 5, has the poorest MAEs of any data set examined. However, with only 4 tasks per person and given a fairly large sample size of 1,794, the MAE of 4.32 is marginally acceptable. Note that the MAE estimates at smaller sample sizes n=500 and n=200 were extremely volatile and not reported. Asterisks were inserted to denote instability. This is most probably due to a combination of sampling error and relatively poor model performance.

Table 5.
Books MAEs (11 parameters)

| n= | 1,794 | 1,000 | 500 | 200 |
|----------------|-------|-------|-----|-----|
| tasks = | | | | |
| 4 | 4.32 | 5.03 | ** | ** |
| 3 | 5.66 | 6.45 | ** | ** |
| 2 | 7.96 | ** | ** | |

Hit rates were again extremely high, most likely due to the dominance of one brand in the marketplace. However, notice the very modest declines in hit rates as number of tasks decreases.

Table 6. Books Hit Rates

| n=3,276 | |
|----------------|--------|
| <u>tasks =</u> | |
| 4 | 87.7% |
| 3 | 87.04% |
| 2 | 86.73% |

The relatively large MAE at 4 tasks per respondent may be due to the small number of attributes in the study (5) failing to model respondents' choice behavior and/or the failure to include the most relevant attributes to respondent choice behavior in the study. A qualitative examination of the attributes would suggest the latter alternative as the likely explanation for the relatively large MAE value.

Discussion

Results of this study may offer additional hypotheses concerning two findings recently published:

- Sentis and Li (2000) reported HB convergence for several commercial data sets after as few as one thousand iterations
- Sentis and Li (2001) reported that, again for numerous commercial data sets, HB alone performed as well as Latent Class followed by HB within LC segment

In both cases, these finding may be the result of practitioners using more choice tasks than necessary. HB may converge more quickly when there is an abundance of individual-level data. It appears clear that the reverse is true, namely, that when fewer tasks are used, a larger number of iterations is required to reach convergence.

Similarly, Latent Class segmentation may not offer much assistance in those cases where the individual-level model is information rich, that is, where the upper level HB model does not contribute much to the lower level model. Further work must be done to verify or deny these hypotheses.

If this second hypothesis is true, then abbreviated task set models using extremely few tasks per respondent, such as one or two, should benefit from Latent Class segmentation preceding HB estimation. This hypothesis could be explored by extending the analysis presented here to include Latent Class segmentation with HB estimation within segment. A comparison of MAEs and hit rates should confirm or deny the hypothesis.

The Beverages data set performed particularly well. The Beverages study differed from the other two in several ways: constant sum choice, large number of alternatives per task, in-person interview, visual representation of products (rather than written descriptions). It would be useful to know the degree to which, if any, each of these factors contributed to the excellent performance of the Beverages model.

Several biases thought to be inherent in conjoint studies, namely number of level effect, order bias, learning bias, framing bias and respondent fatigue, may all be diminished with an abbreviated task set. Further study needs to be undertaken to determine whether or not and if so, to what degree, any of these biases might be affected by the use of abbreviated task sets.

Summary

It appears that adequate individual-level choice models can be constructed with as few as two or three choice tasks per respondent when using HB. This approach requires substantial sample size, typically 1,000 respondents or more and a large number of burned iterations within HB, perhaps as many as 100,000. Computer run times can be significantly and adversely affected by the increase in sample size and burned iterations.

Care must be taken to thoroughly test the experimental design before fielding to ensure a convergable model with adequate coefficient standard errors will result. Aggregate attribute coefficient standard errors, using randomly generated test data, of under approximately 0.05 appear to generate good models.

There are numerous practical situations where sample size is more easily attainable than a large number of choice tasks. In those situations, the reduced task set approach may prove valuable and useful.

References

Johnson, Richard M. (2000), "Understanding HB: An Intuitive Approach," 2000 Sawtooth Software Conference Proceedings, Sawtooth Software, Inc., Sequim, WA.

Johnson, Richard M. and Bryan K. Orme (1996), "How Many Questions Should You Ask In Choice-Based Conjoint Studies?" 1996 Advanced Research Techniques Forum Proceedings, American Marketing Association, Chicago, IL.

Sentis, Keith and Lihua Li (2000), "HB Plugging and Chugging: How Much Is Enough?" 2000 Sawtooth Software Conference Proceedings, Sawtooth Software, Sequim, WA.

Sentis, Keith and Lihua Li (2001), "One Size Fits All or Custom Tailored: Which HB Fits Better?" 2001 Sawtooth Software Conference Proceedings, Sawtooth Software, Sequim, WA.

© 2003 / MACRO Consulting, Inc.

Published in the Professional Marketing Research Society Conference Proceedings, April 2003, Vancouver, B.C., CANADA



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.

312 SOUTHWOOD DRIVE
SCOTT'S VALLEY
CALIFORNIA 95066
TEL 831.454.8927
FAX 831.480.8419
www.macroinc.com