Sawtooth Software

RESEARCH PAPER SERIES

Including Holdout Choice Tasks in Conjoint Studies

Rich Johnson and Bryan Orme,
Sawtooth Software, Inc.

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

Including Holdout Choice Tasks in Conjoint Studies

Rich Johnson and Bryan Orme, Sawtooth Software

Note: this work represents an update to our 1997 technical paper of this same title.

We think it is wise to include holdout choice tasks in conjoint interviews, even though they may not appear to be needed for the main purpose of the study. Holdout choice tasks typically look just like regular CBC choice tasks, but they are not used (are "held out") during utility estimation. They almost always turn out to be useful, for these reasons:

- They provide a proximal indication of validity, measured by the utilities' ability to predict choices not used in their estimation. This can help expose errors in survey programming, data processing, or fielding, as well as help clients gain additional confidence that the conjoint utilities are useful for predicting complex choices.
- They provide a check on the scaling of the utilities. If the most popular concepts are over-predicted, then the scale parameter (Exponent, in Sawtooth Software's simulator) should be reduced. If the predictions are too flat, then the scale parameter should be increased. CBC utilities should already have appropriate scale to predict CBC-looking holdout tasks, but ratings-based conjoint and ACBC typically require scale parameter adjustment to predict holdouts well.
- They permit identification and removal of inconsistent respondents.
- They can be used for testing specific product configurations under consideration. Much value can be added by direct measurement of these concepts.

It's hard to design good holdout concepts without some prior idea of respondent preferences. There's no point in asking people to choose among concepts where one dominates in the sense that everyone agrees which is best. And, similarly, it's good to avoid presenting concepts that are equally attractive, since equal shares of preference would be predicted by a completely random simulator. If you present triples of concepts, it's probably best if their shares of choices are somewhere in the neighborhood of 50/30/20.

Over the last 25 years, most CBC studies and holdout questions have featured *minimal overlap,* which means that an attribute level is not repeated within a choice task (unless there are more products to show in the choice task than levels in the attribute). Over the last few years, the minimal overlap questionnaire has been shown to be less effective at extracting preferences for respondents who do not trade off attribute levels, but use heuristic decision rules such as "must haves," "unacceptables", or other non-compensatory decision strategies. For example, if a respondent requires a certain brand, and there is only one such brand offered per choice task *and* holdout task, it makes it trivial for this respondent to answer the CBC questionnaire, and trivial for the estimated utilities to predict this person's holdout choice. High fit to the data in this case

does not necessarily indicate that we've been successful at estimating this respondent's complex preferences beyond brand requirement, and we may face real difficulty predicting this respondent's real world choices (where the required brand typically has multiple variants on the other dimensions).

Since probably most respondents use non-compensatory rules when facing complex conjoint questionnaires, there is a strong movement toward using CBC questionnaires featuring level overlap (such as provided using CBC software's *Balanced Overlap* design approach). Real world market decisions also tend to involve considering multiple product offerings featuring level overlap. Thus, we strongly recommend using holdouts that feature level overlap. And, it is important to use a healthy amount of level overlap on attributes that respondents are known to screen on (apply as non-compensatory rules).

When conducting CBC studies, if you plan to do segmentation with latent class analysis, it's wise to consider the kinds of groups you expect to get, and to design products in holdout choice sets so that one alternative will be much more preferred by each group. This maximizes your ability to confirm the validity of the multi-group Latent Class simulator.

It isn't necessary to have many holdout sets to check the validity of your utilities, or their scaling. However, if you want to use those choices to identify and eliminate inconsistent respondents, you need several choice sets.

For ACBC, CVA, and ACA studies, holdout concepts can be included in the computer-administered interview using the CiW System. For CBC studies, there is an automatic capability to include "fixed" CBC tasks in the survey.

The position of the holdout tasks is important. It is well known that the first CBC tasks contain the most noise and the lowest scale. Respondents tend to learn through the process of completing multiple CBC tasks. They tend to rely more on brand in the first few tasks and more on price in later tasks, and the use of the None choice increases in later tasks. We generally suggest that holdout tasks be spaced evenly throughout a CBC questionnaire (e.g. in the 4^{th}, 8^{th}, and 12^{th} positions if using a 16-task survey).

We've shown an example of a holdout choice task below:

If you were shopping for a credit card and these were your only options, which would you choose?

Visa	Mastercard	Discover	Visa
No Annual Fee	$30 Annual Fee	$60 Annual Fee	$30 Annual Fee
15% Interest Rate	12% Interest Rate	9% Interest Rate	9% Interest Rate
Frequent Flier Program	No Frequent Flier Program	Frequent Flier Program	No Frequent Flier Program
$4,000 Credit Line	$6,000 Credit Line	$2,000 Credit Line	$4,000 Credit Line

It is not very useful to include a "None" option in holdout choice tasks when these are paired with traditional conjoint exercises which don't have a "None" option.

Finally, if you do have several choice sets, it's useful to repeat at least one of them so you can obtain a measure of the reliability of the holdout choices. Suppose your conjoint utilities are able to predict only 50% of the respondents' holdout choices. Lacking data about reliability, you might conclude that the conjoint exercise had been a failure. But if you were to learn that repeat holdout tasks had reliability of only 50%, you might conclude that the conjoint utilities were doing about as well as they possibly could, and that the problem lies in the reliability of the holdout judgments themselves.