Sawtooth Software

RESEARCH PAPER SERIES

Special Features of CBC Software for Packaged Goods and Beverage Research

Bryan Orme,
Sawtooth Software, Inc.

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

Special Features of CBC Software for Packaged Goods and Beverage Research

August, 2003

Bryan Orme, Sawtooth Software

We assume the reader is proficient with Sawtooth Software's CBC or CBC/Web software, including knowledge of the capabilities of the Advanced Design Module. The interested reader can obtain more information on conditional pricing, alternative-specific designs, and design strategies (Complete Enumeration, Balanced Overlap, Shortcut) by reading the CBC and CBC Advanced Design Module technical paper and manuals.

Introduction

Choice-Based Conjoint (CBC) software has been used extensively over the last decade for a variety of conjoint analysis problems. Among Sawtooth Software customers, its use has now eclipsed the use of ACA (Sawtooth Software, 2003). CBC-type questionnaires are the most widely used conjoint method among all market researchers. CBC particularly has found widespread use in packaged goods and beverage research. CBC's popularity for this type of research is due to a number of benefits:

- Asking respondents to make choices among sets of products is more realistic than rating them individually.
- CBC can measure interaction effects more effectively than traditional conjoint (when using aggregate estimation routines, or HB which leverages estimates of population parameters), which often occur among brand, packaging, and price.
- The resulting market simulation tool can estimate shares of choice, demand curves, and substitution (including cannibalization) effects.
- The CBC questionnaire can be made to look quite a bit like choosing products from actual store shelves.

Under favorable conditions, CBC can produce quite accurate predictions of market share (Orme and Heft, 1999; Feurstein and Natter, 1999, Rogers and Renken, 2003). Of course, predictions are never perfect. For example, as Rogers and Renken point out, CBC may consistently over or underestimate price sensitivity in certain cases. We believe the bulk of evidence suggests that CBC is one of the most useful research tools available for testing pricing, packaging, repositioning, and new product introductions for packaged goods. It cannot *perfectly* predict market shares or estimate elasticity curves, but we shouldn't expect nor require it to do so. There are many other elements that influence market share, such as product life cycle, prominence of shelf facing, promotions, and distribution, to name a few. Price elasticity is also greatly influenced by temporary stock-up behavior, whereas CBC is more predictive of long-range, equilibrium elasticity. Under the assumptions of equal availability and information, CBC is able to predict useful shares of choice which, when applied in the context of what-if market simulations, can significantly increase the likelihood of making profitable marketing decisions.

The purpose of this document is to discuss some of the common approaches (and mistakes made) with past versions of our CBC software, and to point out some new capabilities available with the latest version of the Advanced Design Module for CBC/Web.

Shelf-Facing Presentation

The newest version of our CBC/Web Advanced Design Module supports "shelf-facing" presentation, as shown below. (This graphic has been sized to fit within this document—it is bigger and clearer on a PC monitor.)

To achieve this look, the user supplies graphics in which each product is situated on a shelf "segment," such as:

When the graphics are placed side-by-side (with no space between them), and a black border is shown directly underneath the graphic, the resulting display looks like a continuous shelf.

We needed to implement three new features in the software to support the shelf-facing look.

1. In the shelf display shown on the previous page, there are 29 different products. Previous versions of CBC/Web only supported up to 15 levels per attribute and a maximum of 16 concepts per task, so it was impossible to show so many unique products on the screen at once. The new version of the CBC/Web Advanced Design Module can include up to 100 levels for an attribute and up to 100 concepts within a task, which should offer great flexibility for showing quite complex packaged goods displays.

2. Some package sizes are larger than others (or the researcher may want to include more units in a graphic to represent more linear shelf space), so the software needed to support differing widths of product concepts. Also, we needed to make the software flexible so that if multiple rows of products were displayed, the number of products shown per shelf did not need to be constant. The table below more explicitly demonstrates these properties.

<table>
<tr><td colspan="2">Concept 1</td><td colspan="2">Concept 2</td><td colspan="2">Concept 3</td><td colspan="2">Concept 4</td></tr>
<tr><td>Concept 5</td><td colspan="4">Concept 6</td><td>Concept 7</td><td>Concept 8</td></tr>
<tr><td>Concept 9</td><td colspan="2">Concept 10</td><td>Concept 11</td><td>Concept 12</td><td>Concept 13</td><td>Concept 14</td></tr>
</table>

Notice that some concepts (such as concept 6) are much wider than others. The total row widths and the number of products per row may differ.

3. We needed to allow the researcher to specify that the brands should not change positions on the screen across choice tasks (suppress the randomization of level order for brand). We expect most researchers choosing shelf-type display will prefer fixed positions for brands. But, if desired, brands can have randomized position, either across tasks, or held constant within a respondent interview but randomized across respondents.

CBC/Web leverages web browser technology to display the product concepts. This leads to a great deal of flexibility in the programming and in fielding the questionnaire. HTML and JavaScripting elements can be added by the author, to add elements such as pop-ups showing more detail about a package (detailed image and text) if the graphic is clicked. With JavaScripting, pop-ups can automatically appear upon "mouse-over" (without the need for the click). The surveys can be either fielded over the Web or in CAPI mode, from a PC or laptop not connected to the Internet, or even via paper-and-pencil. These benefits lead to greater realism in the interview, and greater flexibility for interviewing respondents.

The remainder of this article will focus on design and analysis issues related to CBC and packaged goods research, paying special attention to common mistakes made in the past, and how the newest version of the CBC/Web Advanced Design Module can overcome those problems.

The Generic "Conjoint-Style" Case

In the 1970s, researchers began to use conjoint analysis for business problems. With traditional conjoint analysis, each attribute had multiple levels, and the levels for each attribute could generically combine with all other levels of the other attributes. For example, a small conjoint study might have three brands, three sizes, and three prices:

Brand:
Brand A
Brand B
Brand C

Package Size:
Small
Medium
Large

Price:
$1.00
$1.25
$1.50

With a generic, balanced conjoint design, each level of price occurs an equal number of times with each brand and size level. This immediately posed limitations. The prices for different brands, or especially package sizes, in reality might be quite different, but the conjoint interview couldn't reflect that.

The Prohibitions Trap

One of the most common mistakes CBC users make is to try to solve the problem introduced in the previous section by using level prohibitions. For example, some users have expanded the number of prices, and created prohibitions between, say, package size and price. The table below represents such a "prohibitions table," where the "Xs" indicate combinations that are *prohibited*.

	$0.75	**$1.00**	**$1.25**	**$1.50**	**$1.75**
Small				**X**	**X**
Medium	**X**				**X**
Large	**X**	**X**			

Indeed, the prohibitions table above leads to a CBC survey in which the Small package is generally shown at lower prices, and the Large package at higher prices. The combinations presented to respondents seem more realistic. But, the resulting data are often poor, or entirely unusable. That is because the researcher is still asking the CBC software to estimate the part worth utilities for package and price *as if they were independent* (main effects). But, because the levels within the attributes are quite correlated, one often cannot estimate the separate effects of the levels with good precision. Sometimes, the prohibitions specified may be so extreme such that the part worths cannot be estimated at all (given the current main-effects model specification). (Though, this problem can sometimes be remedied by re-coding the data in the .CHO and attribute labels .ATT files as a single collapsed attribute, with all not-prohibited cells as the levels.)

CBC software includes a *Test Design* function that shows the relative efficiency values for the main effects, and can thus point to specific problem areas in the design. If you include any prohibitions between attributes, you must use the *Test Design* function prior to fielding the study.

In addition to the efficiency problem for main effects, using prohibitions between package and price doesn't permit CBC software to estimate the interaction effects between these same two attributes. This is very problematic if the price function (the demand curve) differs depending on the package size. In general, CBC software cannot estimate an interaction effect between two attributes involved in a prohibition. (When CBC codes the design matrix, columns representing prohibited combinations of levels have no independent variation.)

Conditional Pricing

Due to the problems discussed in the previous section, researchers sought ways to customize the price ranges for the different brands and package sizes, but that would still lead to efficient estimates for main effects and interactions. Conditional pricing offers a solution. (CBC for Windows has always offered conditional pricing, but conditional pricing wasn't available within CBC/Web until the most recent release.)

With conditional pricing, a "look-up" table is provided. We can use the example from the previous section to create a conditional pricing table.

	Low Price	Medium Price	High Price
Small	**$0.75**	**$1.00**	**$1.25**
Medium	**$1.00**	**$1.25**	**$1.50**
Large	**$1.25**	**$1.50**	**$1.75**

Notice that we still have specified just two attributes for our design, each with three levels. But, we have created a series of conditional prices that are shown in the questionnaire, depending on the package size and price level. The interview looks correct, because the right combinations of prices are shown with the package sizes. But, more importantly, the data no longer are hindered by any prohibitions. We've solved these initial problems, but left ourselves with slightly more challenging issues for back-end analysis.

By default, the software still estimates the main effects for the brands and prices. However, these main effects are no longer interpreted as the preference for each level, holding all else constant. All else was *not* held constant. For example, the main effect for the Large package size captures the preference for Large package sizes given the *average* prices shown for Large packages. Thus, included in the parameter estimate for Large package size is a negative utility intercept to compensate for the average increased prices shown with Large packages.

When one builds conditional pricing tables, this often leads to the need to estimate additionally the interaction between the attributes involved in the conditional pricing grid. This is especially the case if the price differences from level to level, for each package size, were not so uniform as portrayed in the grid above. With aggregate logit and latent class, additional interactions are often required. That is because many interactions observed at the aggregate level are just due to unrecognized heterogeneity (i.e. the same people who prefer premium brands are also less price sensitive). With individual-level estimation (HB), if the interactions are principally due to unrecognized heterogeneity, one can often obtain excellent models with main effects estimation only (Orme and Heft, 1999).

In the example above, if you estimate main effects plus the first-order interaction effects between Package Size and Price, you are free to use any pattern of prices in the table. In that case, CBC estimates the effect of each cell in the table independently of the others. But, we recommend you leave open the possibility of being able to model the data with main effects only (parsimonious models, if they fit well, are always preferred). When estimating main effects plus all part worth interactions, many "reversals" can result (due to random noise), wherein increases in price are associated with increases in share of preference. Thus, we encourage you to specify pricing increments that are proportional (constant percentage changes from level to level across the package sizes) or reflect constant price increments for adjacent price levels from brand to brand, as this increases the likelihood of fitting the data well with main effects.

We've discussed conditional pricing with respect to one attribute: Package Size. However, in our CBC software it is possible to make prices conditional on the combination of up to three attributes other than price. Researchers studying packaged goods categories often need to create conditional prices, depending on the package size *and* the brand. But this often leads to the "brand/package size prohibitions trap," described in the next section.

The Brand/Package Size Prohibitions Trap

Clients often approach CBC users with a list of, say, 18 brand and package size combinations, where the prices also need to be unique for these SKUs. The CBC user recognizes that the past CBC software only permitted up to 15 levels per attribute. So, the only way to represent all 18 brand and package size combinations was to specify brand and package size as separate attributes, associate a conditional graphic with those combinations, and prohibit any of the brand and package size combinations that didn't apply. For example, perhaps there were 6 unique brands and 5 unique package sizes. That leads to 30 possible combinations, of which the researcher needs to prohibit 12 of them. This is clearly the same problem as the "prohibitions trap" described earlier, and such prohibitions may lead to an inefficient (or even deficient) design.

Even if the prohibitions between brand and package were very modest, leading to reasonable efficiencies for main effects, this procedure would not permit the estimation of interactions between brand and package. If the preference for a package really depends on the brand attached to it, main effects estimation will lead to improper conclusions. Many CBC researchers have used such hindered designs over the years, with two results: reduced overall design efficiency and the inability to assess whether modeling the main effects for brand independent of package size provided an adequate fit to represent peoples' preferences for joint brand/package combinations. We suspect this has been one of the most common mistakes committed with our CBC software over the years.

An adequate solution to this problem was not available in CBC software until most recently, with the most recent release of CBC/Web Advanced Design Module. In this situation, the researcher needs to be able to specify the brand/package size combinations as a *single* attribute, with all 18 levels. This explicitly accounts for the independent preference for each brand/package combination (the interaction between brands, packages and price). Once the 18-level attribute is specified, a conditional pricing table can be specified for these 18 brand/package levels.

This ability to collapse multiple attributes into a single attribute doesn't come without a statistical cost. It leads to larger numbers of parameters to be estimated. Given the same sample size, the precision of these parameters is much less than for an attribute with the typical five levels. Thus, researchers should strive for much larger sample sizes when dealing with many separate brands with unique price ranges for each. In some cases, this may require sample sizes well in excess of 1000 respondents for good predictive models.

Number of Levels Effect

There are also interesting issues for packaged goods research regarding the Number of Levels (NOL) effect. The NOL effect causes an attribute defined on more levels (given the same total range) to receive an artificial relative increase in importance. For example, a price attribute defined with 6 levels ($10, $12, $14, $16, $18, $20) receives significantly more importance than if defined using, say, three levels ($10, $15, $20), even though the same total difference in price is measured from highest to lowest level. The NOL effect occurs for other attributes, including packages and brands. However, how NOL might potentially affect pricing experiments for packaged goods research isn't altogether clear.

For example, assume 4 brands and 4 package types, along with price. We can represent products in text form:

Coke 6-pack of cans $1.29

If we formulate the problem with brand and package as two separate attributes (each with just 4 levels) or if we formulate the problem with brand and package combined as a 16-level attribute, the respondent still cannot detect any difference in how the products are shown. Moreover, coding the model as main effects plus interaction effects between brand and package size provides an equivalent model to the one coded as "main effects" for the combined 16-levels of brand and package size. Thus, simply collapsing two attributes into a single attribute with more levels shouldn't lead to a number of levels effect relative to those same attributes defined separately in the software.

Still, with packaged goods research, the potential for a number of levels effect exists. If 10 brands are included, but just 2 package sizes, the classic number of levels argument suggests that the importance of brand may be artificially biased upward relative to package size. However, there is a counterargument that if buyers in the real world actually see these 10 brands on the shelf each with two package sizes, the same NOL effect may influence real-world decisions.

How Many Levels for Price?

One of the most difficult decisions is the number of levels to include for price. If using conditional pricing tables, the question focuses on how many levels of price to include per brand (the number of levels associated with each brand is held constant across the table). There is a trade-off when considering this issue: with more price points, it provides more flexibility to detect areas of nonlinearity in the price function (sometimes referred to as "kinks" or "sweet spots"); but, with more data points (holding sample size constant), the precision of the estimates for each price point is reduced, increasing the likelihood of seeing "reversals" in the data. Less sophisticated clients often push for six or more levels for price, without recognizing this trade-off with precision (or budgeting enough money to achieve the sample size needed to adequately model so many effects).

Generally, for part worth models, we recommend three to five price points per brand at most. If you have determined that a linear (or log-linear function) is appropriate for price, then it doesn't matter how many levels of price are included (ignoring the potential bias from the number of levels effect) as you just measure a single coefficient for price. The problem is that you usually don't know ahead of time whether a linear or log-linear function fits the data well.

Alternative-Specific Prices, or Conditional Pricing Table?

CBC's Advanced Design Module lets you deal with conditional prices in two ways: either with a conditional pricing table, or with alternative-specific prices. (With alternative-specific prices, the researcher defines a separate price attribute for each brand, or brand/package size joint attribute.) Either way, the interview looks the same to the respondent. But, with alternative-specific coding, the coding of the design matrix is a bit different.

With conditional pricing tables, the default coding in the design matrix is main effects only. If there is an interaction between brand and price (which is often the case when using conditional pricing), the researcher can additionally ask the software to include the first-order interaction between brand and price (and the software internally cross-multiplies the main effects columns to generate new columns representing interaction terms).

With alternative-specific pricing, the software codes the design matrix to account for the main effect of brand, and the "main effect" of the brand-specific price attributes. This model includes the same number of terms (and provides the same model fit) as the main effects plus first-order interaction specification for conditional pricing. But, one gives up the ability to quickly toggle in the software between a main effects only model versus an interaction model if using alternative-specific pricing (though, with some data processing using the ASCII-formatted .CHO file, you can collapse the codes for the separate alternative-specific price attributes into a single price attribute, such as would result from conditional pricing tables).

With alternative-specific price attributes, one gains the ability to specify monotonicity constraints using our Latent Class or HB software. This can be an important feature, as it allows you to eliminate reversals in the separate price functions for brands (which is not possible with our software if you use the conditional pricing approach, because the net effect of price for a brand stems from two components: the main effect for price plus an interaction effect). This suggests another interesting possibility. If you use conditional pricing and determine that you want to estimate separate price functions for different brands, but find that the resulting interaction model leads to reversals in derived demand curves, you can perform data processing to reformat the .CHO file to reflect alternative-specific price effect coding (see the Advanced Design Module documentation for details). With alternative-specific price effects in place, you could use monotonicity constraints within our HB or Latent Class software. However, alternative-specific effects plans can lead to many parameters to be estimated, with the usual cautions regarding sample size.

Note: if you choose to use alternative-specific effects, please recognize that SMRT software is currently limited to 30 total attributes in the design. Thus, you'll be limited to no more than 29 brand-specific price attributes (after defining the brand attribute).

Level Overlap within Task

In the previous section, we discussed the differences between coding conditional pricing designs and alternative-specific designs. We only discussed the differences with respect to estimation. But there are also differences with respect to the characteristic of level overlap in the design.

Level overlap refers to the number of times a level may be present within the same task. Let's consider the simple case of two attributes: three brand levels and three price levels. If we show three product concepts per task, then we could design tasks such that each brand only occurs once, and each price only occurs once. This is called "minimal overlap" (each level occurs as few times as possible per task), and is the strategy employed by our Complete Enumeration and Shortcut design strategies. Minimal overlap leads to highly efficient estimation of main effects. However, if the researcher wants to increase the precision of the interaction terms (the brand-specific price effects), then some degree of overlap is desirable (e.g. permitting the same price level to occur more than once within a task). But, adding some overlap usually leads to a loss in precision for the main effects. Our software includes a compromise overlap strategy called "Balanced Overlap" which tends to lead to increased precision for the interaction terms while sacrificing relatively little in terms of precision for main effects. Please see Chrzan, Orme 2000 for more information on design strategies for CBC.

Consider what occurs with respect to overlap in price if using conditional pricing versus alternative-specific pricing. With conditional pricing, the software "thinks" there is just one price attribute, and the customized prices for each brand are inserted when the survey is executed by referring to a look-up table. Thus, if using a minimal overlap strategy, once price level 1 had been used in a task, the software would try using all other price levels before repeating price level 1. This can lead to lower efficiency if the researcher during analysis specifies interaction terms with price. With alternative-specific prices, a separate attribute is typically defined for each brand. Thus, if price level 1 is used for Brand A, it doesn't constrain the design from also using price level 1 for Brand B. This is a more efficient design if the researcher intends to capture the brand-specific price effects.

One might conclude, therefore, that if using conditional pricing, one might want to use the "Balanced Overlap" strategy; and that one should use a minimal overlap strategy if using alternative specific plans (only Shortcut strategy is available for alternative-specific plans). However, using Balanced Overlap with conditional pricing plans can lead to bizarre situations in which the same brand is offered twice, but at different prices, within the same choice task. With Balanced Overlap, overlap is permitted for *all* attributes. But, with alternative-specific pricing (and if using the Shortcut method), the brands have minimal overlap. And, the brand-specific price attributes also have minimal overlap with respect to their brand, but overlap with respect to other price attributes for other brands (e.g. price level 1 for Brand A can appear with price level 1 for Brand B).

However, there is yet another way to achieve overlap in the price attribute for better measurement of interaction terms. Consider a design in which there are 12 brands and a four-level price attribute (conditional pricing design). If we show 12 product concepts per task and use the Complete Enumeration design method, each brand is represented once. But, each price will be repeated three times (12 concepts divided by 4 prices). Thus, by virtue of the number of concepts required per task, we have enforced minimal overlap for brand (which is realistic) and allowed overlap with respect to the price attribute.

Packaged Goods Research and Market Simulations

Sawtooth Software provides different market simulation approaches for competitive contexts: first choice, share of preference, share of preference with correction for product similarity (not suggested, but available for historical purposes only), and Randomized First Choice (RFC). RFC's approach provides correction (share reduction) for pairs of products that are defined similarly in terms of their attribute levels. In general, we have found RFC to perform well in methodological studies involving, say, five or more attributes, especially when using aggregate logit or latent class results. There is evidence that RFC provides some benefit over standard share of preference simulations when using individual-level part worths (such as from HB), but the benefits are not as dramatic as with group-based models.

Recently, we've recognized that RFC may be less useful or not useful at all for two-attribute studies involving brand/package and conditional prices. In conditional pricing tables, level "1" for one brand/package may mean $2.00, but level "1" for another brand/package might mean "$5.25." However, RFC is blind to this difference, assuming only that since they both share level 1 for price, they must be identically defined on price. The resulting correction may not be desirable.

We generally recommend using HB estimation for CBC studies, and especially if using conditional pricing tables we recommend that you test whether RFC or share of preference best fit holdout observations. Make sure to tune the exponent (scale factor) for the different models to best predict holdouts prior to comparing results. Recent evidence presented by Arenoe (Arenoe, 2003) involving packaged goods research and real-world sales data suggests that RFC may offer little benefit over share of preference in these cases. If this finding holds, you can save a great deal of computing effort using the faster share of preference method, which will especially pay off if using the computationally-intensive Advanced Simulation Module for optimization searches.

HB Estimation for Many-Leveled Attributes

We strongly recommend using HB estimation for pricing research. However, unless one uses the most recent version of CBC/HB software (version 3), when using very many levels for an attribute, the HB estimation may be poor.

When we changed the CBC/Web Advanced Design Module to accommodate up to 100 levels per attribute, we also upgraded our CBC/HB software to provide better estimation for the kinds of data sets that might result. When data are sparse (which is usually the case for CBC studies and pricing research) and one is estimating parameters for many levels of an attribute under effects-coding, the final ("omitted") parameter can be severely underestimated, and its variance can become too large.

Rich Johnson identified the overstatement of variance issue a few years ago (Johnson 1999), but for typical CBC data sets, no problems for estimating the mean were observed, and the finding was viewed as a curiosity rather than a real problem that could substantially affect results. Recently, we've seen data sets where estimating 20 or more levels for a "brand" attribute can result in very poor estimation of the last brand level. We didn't have a solution for the problem until the release of CBC/HB version 3, based on input from Peter Lenk from the University of Michigan.

The solution involves introducing appropriate negative covariances in the off-diagonal elements of the prior covariance matrix for levels within attributes, and using zero as the starting points for elements of alpha and beta (rather than the "pseudo-OLS" starting points). The reader interested in more details should download the CBC/HB v3 technical paper at www.sawtoothsoftware.com.

Summary

Choice-Based Conjoint (CBC) is widely used throughout the world for a variety of conjoint analysis problems. It is particularly popular for consumer packaged goods research. The new capabilities of the CBC/Web Advanced Design Module provide even more power and flexibility for CBC users to conduct realistic-appearing, useful studies.

There are a number of potentially complex design and analysis decisions, such as how to deal with detailed conditional pricing tables. Conditional pricing tables that are designed with uniform or proportional increments between prices are the safest approach, as they leave the door open for main effects models, should the interaction terms not provide significant additional fit. When unique price functions are estimated for many brands or SKUs, sample sizes often need to be increased beyond typical CBC standards to permit stable estimation. If using HB to estimate parameters, recent advances in version 3 of CBC/HB will lead to better estimation for many-leveled attributes.

References

Arenoe, Bjorn (2003), “Determinants of External Validity in CBC,” Sawtooth Software Conference Proceedings, pp 217-232.

Chrzan, Keith and Bryan Orme (2000), “An Overview and Comparison of Design Strategies for Choice-Based Conjoint Analysis,” Sawtooth Software Research Paper, available at www.sawtoothsoftware.com.

Feurstein, Markus and Martin Natter (1999), “Forecasting Scanner Data by Choice-Based Conjoint Models,” Sawtooth Software Conference Proceedings, pp 169-181.

Johnson, Richard M. (1999), “The Joys and Sorrows of Implementing HB Methods for Conjoint Analysis,” Sawtooth Software Research Paper, available at www.sawtoothsoftware.com.

Orme, Bryan K. and Michael Heft (1999), “Predicting Actual Sales with CBC: How Capturing Heterogeneity Improves Results,” Sawtooth Software Conference Proceedings, pp 183-199.

Rogers, Greg and Tim Renken (2003), “Validation and Calibration of Choice-Based Conjoint for Pricing Research,” Sawtooth Software Conference Proceedings, pp 209-215.

Sawtooth Software (2003), “More Evidence CBC Is Most Popular Conjoint Method,” Sawtooth Solutions, Summer 2003.