Sawtooth Software

RESEARCH PAPER SERIES

Menu-Based Choice Modeling Using Traditional Tools

Bryan K. Orme,
Sawtooth Software, Inc.

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

Menu-Based Choice Modeling Using Traditional Tools

Bryan Orme
Sawtooth Software, Inc.

Executive Summary

Menu-Based Choice (MBC) research studies are becoming more popular as businesses implement *mass customization* sales models. The good news for conjoint researchers is that the existing tools for designing and analyzing CBC studies may also be used for MBC. The not-so-good news is that until software is developed to automate many of the steps (and we're working on that), the process is typically more challenging and time-consuming than for CBC studies.

Both traditional "fixed" and randomized design strategies are quite effective for designing MBC questionnaires. If randomized designs are used, the intuitive counting analysis approach can convey top-line results, including detailed price sensitivity curves. Multinomial logit (MNL) may be used in analysis, including aggregate logit, latent class, and hierarchical Bayes (HB) variants. Simulators may be built with Excel.

Background

Increasingly, stated preference choice projects involve Menu-Based Choice scenarios (MBC) where respondents can select from one to multiple options from a menu. This is not surprising, given the fact that buyers are commonly allowed to customize products and services (mass customization). Examples include choosing options to put on an automobile, designing employee benefits packages, combination drug therapy choices for pharma, selections from a restaurant menu, banking options, configuring an insurance policy, or purchasing bundled vs. *a la carte* services including mobile phones, internet, and cable.

Here is a very simple menu, where the respondent chooses options and a total price is shown:

Which of the following options would you buy? Select as many as you wish, or none of the items.

- ☑ Option A $12
- ❑ Option B $24
- ☑ Option C $7
- ❑ Option D $55
- ☑ Option E $3

Total Price of Selected Options: $22

Figure 1

Respondents can select between zero to five choices on the menu in Figure 1. This particular respondent has selected three items, for a total price of $22. There are 2^5=32 possible ways respondents can complete this menu. The prices shown might always stay the same, or perhaps the questionnaire is designed so that some or all of the prices vary between respondents, or even across repeated menus given to the same respondent.

If prices are varied across menu questions, we can observe whether changing the prices influences what respondents pick. Economic theory suggests that as the price of a menu item *increases* its likelihood of choice will *decrease*. How strong is that relationship? Is it fairly linear? Does reducing the price for an item cause a different item on the menu to be more likely (or even *less* likely) to be chosen? In other words, are items on the menus *substitutes* or *complements*? Menu-Based Choice (MBC) experiments can investigate such issues.

MBC questionnaires are often used to investigate bundling vs. *a la carte* strategies. We've been showing a case study involving fast-food menu choices at Sawtooth Software's advanced CBC trainings for over five years now:

<table>
<tr><td colspan="3">Menu Scenario #1: Please imagine you pulled into a fast-food restaurant to order dinner for <u>just yourself</u>.
If this were the menu, what (if anything) would you purchase?</td></tr>
<tr><td>❑ Deluxe Hamburger Value Meal
-Deluxe Hamburger
-Medium fries
-Medium drink

$3.99</td><td>❑ Chicken Sandwich Value Meal
-Chicken Sandwich
-Medium fries
-Medium drink

$5.59</td><td>❑ Fish Sandwich Value Meal
-Fish Sandwich
-Medium fries
-Medium drink

$3.99</td></tr>
<tr><td colspan="2">(Only order sandwiches, fries or drinks from this area if you did not pick a value meal above.)

Sandwiches:
❑ Deluxe Hamburger $1.99
❑ Chicken Sandwich $3.59
❑ Fish Sandwich $1.99

Fries:
❑ Small $0.79
❑ Medium $1.49
❑ Large $1.69

Drinks:
❑ Small $0.99
❑ Medium $1.69
❑ Large $2.19</td><td>Salads:
❑ Cobb dinner salad $4.79
❑ Grilled chicken salad $4.39

Healthy Sides:
❑ Carrots/Celery with Ranch dressing $1.19
❑ Apple slices/Grapes with dipping sauce $0.99

Desserts:
❑ Apple/Cherry/Berry pie $0.99
❑ Cookies $1.19</td></tr>
<tr><td colspan="3">❑ I wouldn't buy anything from this menu.
I'd drive to a different restaurant, or do something else for dinner.</td></tr>
</table>

Figure 2

Past Literature and Sawtooth Software Presentations

A number of articles have been published on menu-based choice, including:

- Ben-Akiva, M. and S. Gershenfeld (1998), "Multi-featured Products and Services: Analysing Pricing and Bundling Strategies," Journal of Forecasting, 17.

- Liechty, J., Ramaswamy, V., and S. Cohen (2001), "Choice-Menus for Mass Customization: An Experimental Approach for Analyzing Customer Demand with an Application to a Web-based Information Service," JMR, 39 (2).

- Cohen, S. and J. Liechty, (2007), "Have it Your Way: Menu-based conjoint analysis helps marketers understand mass customization," Marketing Research, 19:3.

The following was presented at the ART/Forum (American Marketing Association):

- Conklin. M., B. Paris, T. Boehnlien-Kearby, C. Johnson, K. Juhl, A. Zanetti-Polzi, K. Gustafson, B. Palmer, (2007) "Menu Based Choice Models," ART Forum.

And, the Sawtooth Software Conference has also seen some useful papers on MBC:

- Bakken, David and Len Bayer (2001), "Increasing the Value of Choice-Based Conjoint with 'Build Your Own' Configuration Questions," Sawtooth Software Conference Proceedings, pp 99-110.

- Bakken, David and Megan Kaiser Bond (2004), "Estimating Preferences for Product Bundles vs. *a la carte* Choices," Sawtooth Software Conference Proceedings, pp 123-134.

- Johnson, Richard, Bryan Orme and Jon Pinnell (2006), "Simulating Market Preference with 'Build Your Own' Data," Sawtooth Software Conference Proceedings, pp 239-253.

- Rice, Jennifer and David Bakken (2006), "Estimating Attribute Level Utilities from 'Design Your Own Product' Data—Chapter 3," Sawtooth Software Conference Proceedings, pp 229-238.

David Bakken's papers have been especially useful, and some of the ideas we present here are drawn directly from his work.

Designing MBC Studies

For traditional conjoint and CBC, the focus has been on designing a set of product concepts that respondents rate or choose. With MBC studies, the focus is on asking respondents to *configure their preferred choice* by making from zero to multiple selections from a menu of possible selections. Respondents are permitted to take a more *proactive* approach in designing appropriate products in MBC, whereas they tend to be placed in a more *reactive* stance with CBC questionnaires.

In conjoint analysis, we consider multiple factors (attributes), where each attribute has at least two levels. Menu-based choice problems also involve multiple factors, each having multiple levels. Whereas we often think of a CBC question as being composed of multiple product concepts (cards), we should think of the entire MBC menu question being controlled by a *single* card. This allows researchers to use the familiar tools for conjoint design with MBC experiments (e.g. CBC or CVA software, Warren Kuhfeld's SAS routines), except that the number of factors for MBC experiments will often be larger than for traditional conjoint or CBC.

In the MBC studies we've designed at Sawtooth Software, we've done web-based surveys and have used CBC's Complete Enumeration, Balanced Overlap, or Shortcut design strategies. These are randomized design routines that explicitly control for level balance, and in the case of Complete Enumeration and Balanced overlap, also control for orthogonality. Each respondent is randomly selected to receive one of many versions of the design, where each version has been carefully constructed. However, even purely random design strategies, such as would be available using randomized list functions available in many web interviewing systems (including SSI Web), will produce robust designs that work quite well for MBC studies.

Sample Study

Early in 2010, we conducted a methodological research study among approximately 1600 respondents (pre-screened for intention to purchase a new car in the next few years). 800 respondents were used for building the MBC models (calibration respondents), and 800 respondents were used as holdout sample. We used Western Wats' panel for respondent recruitment and invitations, and fielded the study using our SSI Web platform. (We thank our colleagues at Western Wats for their support of our R&D efforts and for their excellent service.)

Our MBC study actually consisted of two separate MBC exercises (shown below), which we analyzed independently.

At the beginning of the survey, we asked respondents how much they expected to pay for their next new vehicle, and to rate their preferences for automobile options (to be used as covariates in HB modeling). Next, we asked respondents to select the three new vehicles they were most likely to consider purchasing, and to indicate for each how much they expected to pay. The vehicle choices were provided in drop-down lists, developed using information from: http://www.automotive.com/new-cars/index.html. All respondents provided three vehicle choices in their consideration set.

In MBC Exercise 1, we showed the vehicle that respondents picked as their top considered vehicle, at a base price $2,000 *less than* the amount they said they were expecting to pay. The exercise consisted of eight choice tasks like the one directly below:

Let's assume you were going to purchase the **Honda Accord** and it didn't have any of the options below as standard features. If the prices for the options were as shown below, which options would you add to your vehicle?

(If you would add no options, just click the "Next" button)

Base Price: $23,000

☐	$1,500	Alloy Wheels
☐	$900	Moonroof/Sunroof
☐	$300	XM Radio (+ $13/month)
☐	$800	Leather Seats
☐	$350	Security System
☐	$1,000	Backup/parking assist sensor with rearview camera
☐	$600	Hands-Free Phone System
☐	$1,300	Navigation system (in dash)

Total: $23,000

Figure 3

The prices for the different options were varied (within respondent) using an experimental design (CBC's Complete Enumeration design method). Four price levels were varied per option, as shown below:

	Price 1	Price 2	Price 3	Price 4
Alloy Wheels	$1,500	$1,750	$2,000	$2,500
Sunroof	$500	$700	$900	$1,200
XM Radio	$300	$400	$500	$600
Leather Seats	$600	$800	$1,000	$1,200
Security System	$150	$200	$250	$350
Parking Assist	$600	$700	$800	$1,000
Hands-Free Phone	$400	$500	$600	$800
Navigation System	$1,000	$1,300	$1,600	$2,000

Figure 4

Exercise 2 was a bit more complex. We showed *all three* top considered vehicles. Respondents were asked to choose one of the vehicles, and then to add any options to that chosen vehicle. Respondents completed eight tasks as shown in Figure 5:

If you were deciding between the following three cars, and the prices were as shown, which car would you select, and which options would you add to it?

Honda Accord	BMW 3-Series	Hyundai Elantra
Base Price: $28,000	Base Price: $27,000	Base Price: $20,000
$2,000 Alloy Wheels	$1,750 Alloy Wheels	$1,500 Alloy Wheels
$1,200 Moonroof/Sunroof	$900 Moonroof/Sunroof	$700 Moonroof/Sunroof
$300 XM Radio (+ $13/month)	$600 XM Radio (+ $13/month)	$500 XM Radio (+ $13/month)
$1,600 Navigation system (in dash)	$1,000 Navigation system (in dash)	$2,000 Navigation system (in dash)
Total: $28,000	Total: $27,000	Total: $20,000

Figure 5

The four options (Alloy Wheels, Moonroof, XM Radio, and Navigation System) were each varied over four prices, as was shown in Figure 4. This time, we used CBC's Balanced Overlap design strategy to generate the experimental design (as some level overlap in the design would support estimation of cross-effects better than Complete Enumeration's minimal overlap strategy). Additionally, the base price of the vehicle was varied over three price points: $3,000 less than expected price, expected price, and $3,000 more than expected price.

Counting Analysis

Randomized designs are not only a robust and straightforward way to design complex MBC tasks, but they permit a simple form of top-line analysis: counts. With counting analysis, we simply compute the percent of times that an option was chosen. We can count that likelihood of choice overall, or split out by various prices shown on the menu.

Exercise 1

For Exercise 1, where respondents configured their top-most considered vehicle, we've summarized the percent of times options were chosen (across all price variations) in the table below. The most commonly selected item was security system (selected for 67% of the configured vehicles).

Exercise 1
Counting Results: Choice of Menu Options

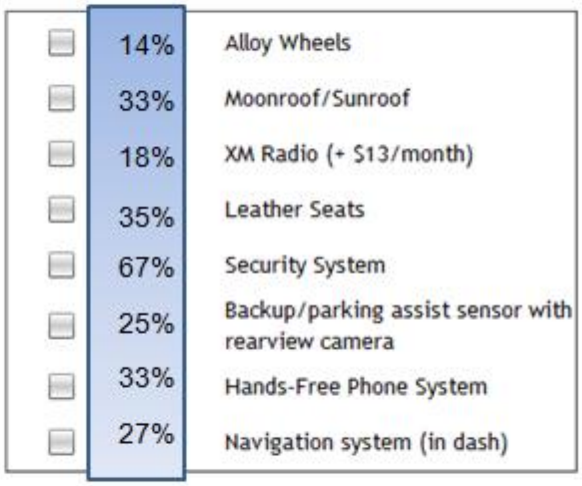

Figure 6

We can also count the percent of times each item was chosen when offered at each of its four price points, and plot the results in a chart. We've also performed a simple log-log regression to compute the price elasticity of demand for each price curve (shown in parentheses after each item's label).

Counting Results: Options by Price

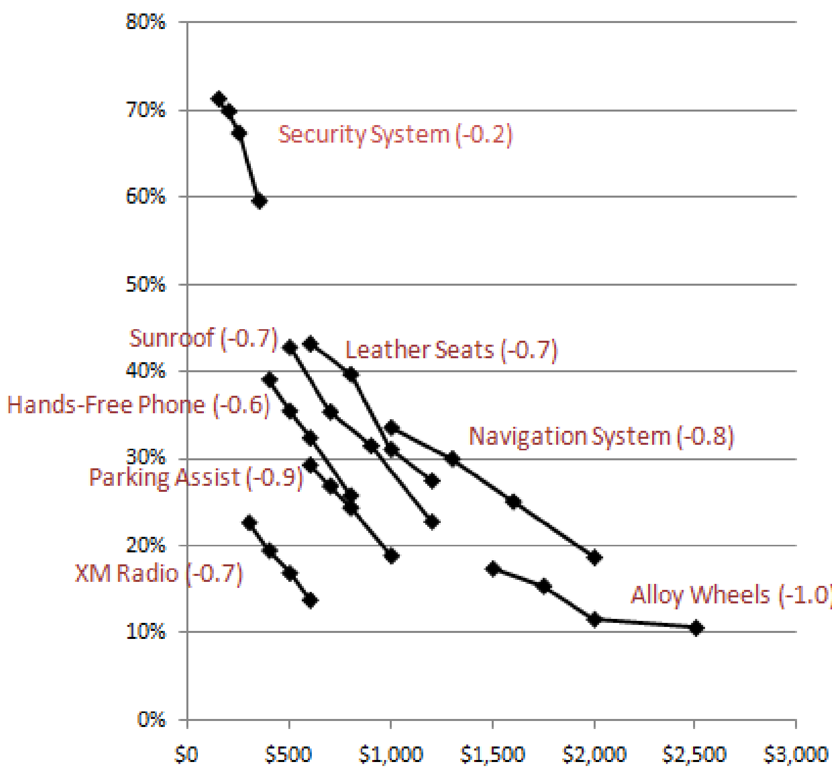

Figure 7

Each of the charted proportions in Figure 7 has a base size of 800 respondents x 8 tasks / 4 price points = 1,600. In other words, each item was available to be chosen at each price point 1,600 times. Because each price point occurred about an equal number of times with every other item's price in the design, we can summarize the independent price effect for each item using counting analysis. Assuming the worst-case scenario 50% proportion, the margin of error (95% confidence) is +/- 2.4% for each estimate. Figure 7 contains a lot of information, and the data are quite precise (+/- 2.4%), given the 1,600 data points supporting each proportion. As you can see, simple counting analysis can relay quite intuitive and useful information.

We may also count the *combinations* of items that were configured in the menu. With 800 respondents x 8 choice tasks, there were 6,400 total vehicles that were configured. We can tally the combinations selected, and report the top 10 most configured options (Figure 8).

Counting Results: 10 Most Common Combinations Selected

Alloy Wheels	Sunroof	XM Radio	Leather Seats	Security System	Parking Assist	Hands-Free Phone	Navi-gation System	%
				✓				12.0%
								8.9%
	✓			✓				3.8%
			✓	✓				3.5%
				✓		✓		3.1%
	✓		✓	✓				2.3%
				✓	✓			2.2%
			✓					2.2%
	✓							2.0%
	✓			✓		✓		1.8%
							Total:	41.7%

Figure 8

The most common configuration across the 6,400 tasks was to add only Security System (12.0% of the menus completed resulted in this choice). If two options were chosen, the most common selection was Sunroof and Security system, with 3.8% of the choices. The top 10 most common configurations account for 41.7% of the total choices made.

We also employed counting analysis to examine the cross-effects among menu items. For example, what was the effect of the price of the Security System on the choice for Sunroof? It turns out that *no* cross effects were significant! This surprised us, and makes Exercise 1 a very easy data set (but somewhat boring) to model.

Exercise 2

Exercise 2 is more complicated, since there were three vehicles shown, with the base price of the vehicle as well as option prices varying (within respondent). The base prices of the top three considered vehicles were varied by -$3,000 to +$3,000 of expected price. The summary counts for items selected on the menu (across all price manipulations) are:

Exercise 2
Counting Results: Choice of Menu Options

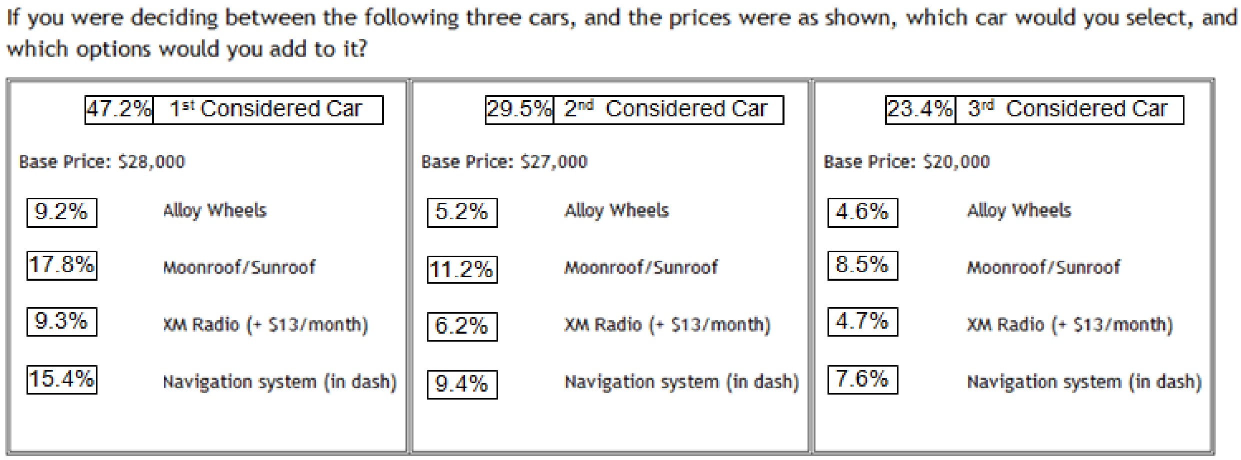

Figure 9

The price changes on the base price of the vehicles (-$3,000 to +$3,000) had a very large effect on the choice likelihood for the vehicle (Figure 10). If the price of the top considered vehicle was increased by $3,000, respondents were only about 20% likely to select that vehicle, and nearly 80% likely to switch to one of the other two vehicles.

Counting Analysis: Likelihood Choosing Vehicle Due to Changes in Base Price

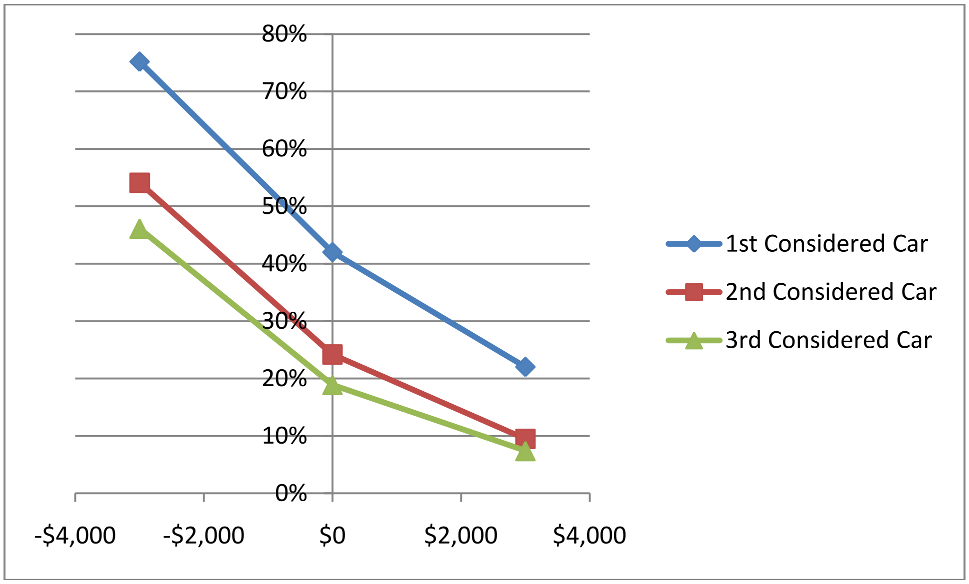

Figure 10

Given that respondents picked a particular vehicle, we can count the likelihood of selecting the four options (Alloy Wheels, Sunroof, XM Radio, Navigation System) at each of their prices.

Counting Analysis: Choice of Option x Prices, Given the Choice of Vehicle

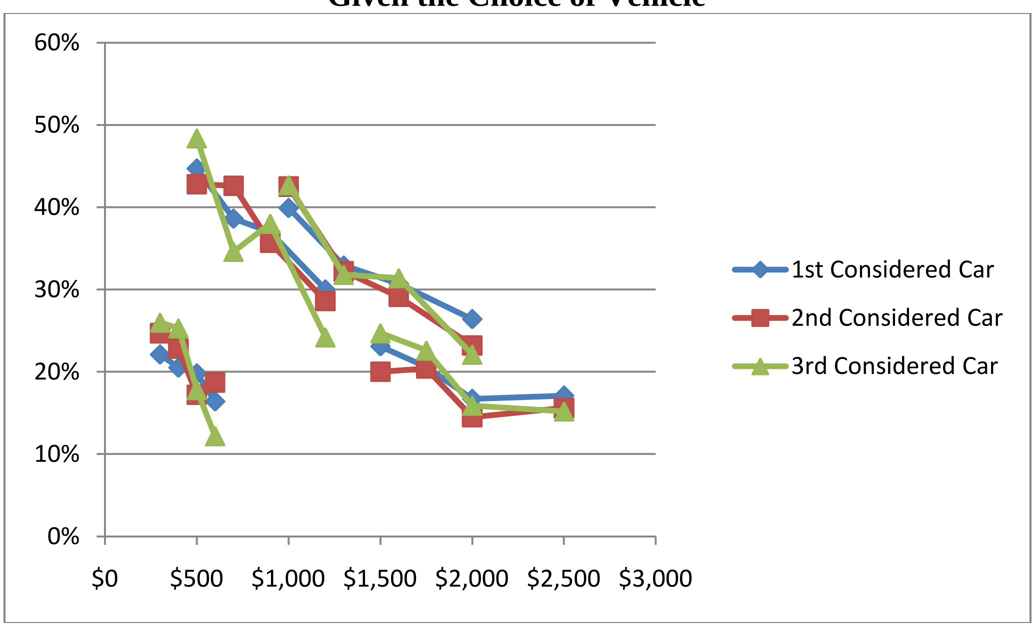

Figure 11

We also used counting analysis to examine cross effects for Exercise 2, finding many of them strongly significant.

Modeling via Multinomial Logit (MNL)

The Sawtooth Software community is quite familiar with three different utility estimation routines that all employ the logit rule: aggregate logit, latent class, and hierarchical Bayes (HB). We employed all three approaches for developing utility weights and building market simulators for these menu choice data.

Other approaches have been proposed in the literature, including multivariate probit (Liechty *et al.* 2001). We have not investigated multivariate probit. The purpose of this paper is to investigate how well the standard tools available to researchers can work in dealing with menu-based choice problems. In the future, we may find out that other methodologies also work well, or are even superior. Our hope is that the simpler models we propose here perform well enough to deliver accurate and robust results for the typical kinds of simulators managers demand. The excellent fit to holdouts we report below for this data set gives us hope that the MNL models may accomplish those aims.

We have experimented with three main approaches to specifying and coding the models for MNL estimation:

- Volumetric CBC Model
- Exhaustive Alternatives Model
- Serial Cross-Effects Model

You may think of these as tools you can mix-and-match to solve a variety of MBC problems. Again, these approaches just reflect different ways to code the choice sets. In all three cases, we are using MNL estimation.

Volumetric CBC Model

This model borrows a trick that we've been describing in our advanced CBC training workshops for some years now (and that is also described in a paper within these proceedings by Tom Eagle). The classic volumetric CBC example involves purchase of breakfast cereal. Imagine that eight product alternatives are available on the shelf, and the respondent is asked to state how many of each product she will purchase. This isn't a constant-sum task, as the respondent can allocate from 0 purchases to as many boxes of cereal as she thinks she can cart out of the store. Let's imagine the respondent completes 12 such tasks.

To model the data, we can simply scan the 12 tasks this respondent completed to identify the largest quantity of items ever purchased in a *single* choice task. Imagine that this maximum is found in task #4, where this respondent "purchased" the following quantities:

Purchase Volume: Task #4

Alternative	Quantity
Alternative 1	0
Alternative 2	3
Alternative 3	0
Alternative 4	2
Alternative 5	0
Alternative 6	0
Alternative 7	4
Alternative 8	0
Total:	9

The maximum volume for this respondent, for any *one* task, is 9.

To analyze the data, we'll generate a .CHS (or .CSV) file (the file for constant-sum allocation data supported by our CBC/HB and Latent Class software). A key thing to remember is that our CBC/HB and Latent Class systems automatically normalize the allocations within each choice task to sum to 100%. The trick, therefore, is to add a "None" alternative to the .CHS/.CSV file, so that the software believes that the respondent also faced a None alternative in each task. We reformat Task #4 as follows:

Purchase Volume: Task #4

Alternative	Quantity
Alternative 1	0
Alternative 2	3
Alternative 3	0
Alternative 4	2
Alternative 5	0
Alternative 6	0
Alternative 7	4
Alternative 8	0
None	0
Total:	9

Next, we also add a None alternative to the other 11 choice tasks. But, for these remaining tasks, the amount "purchased" of the None alternative is equal to 9 minus the volume purchased in that task. So, if in Task #1, the respondent "purchased" five boxes of cereal, a quantity of four will be given to the None. During market simulations, respondents are weighted by their maximum quantity value (9, for this respondent).

Now that we have introduced the coding of the model, it's very easy to see how to apply this to Exercise 1 for our automobile options study. We'll assume that the maximum number of items "purchased" for all respondents is eight (there are eight possible menu options).

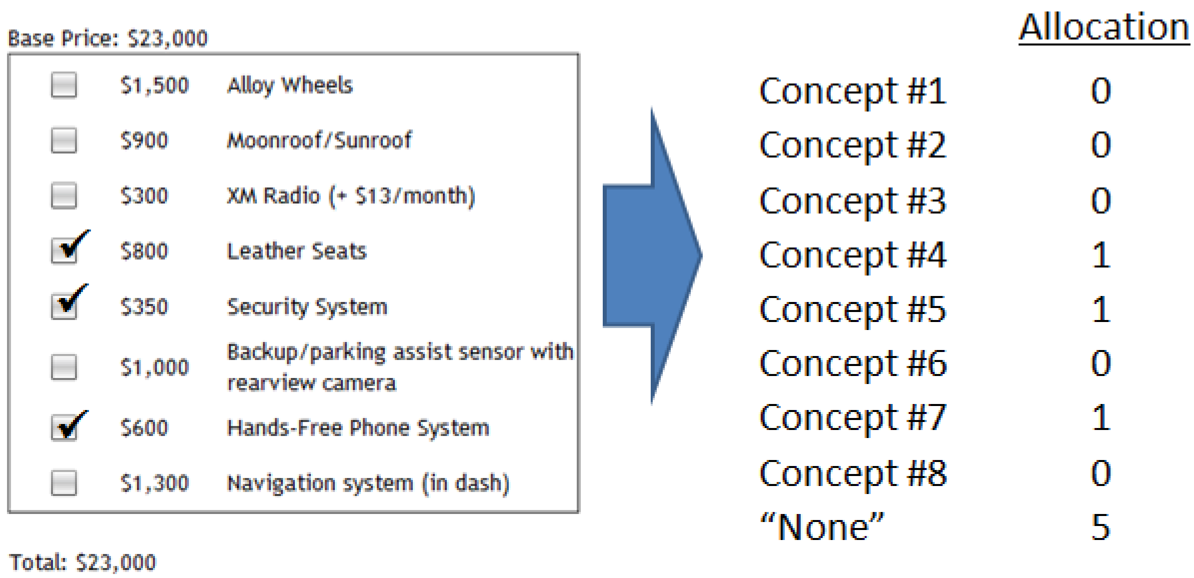

Figure 12

The next challenge is to code the independent variable matrix. To ensure reasonably fast convergence in our CBC/HB software (given its default settings for prior variance), we have found that it works best if the independent variables are coded to have absolute magnitudes in the single digits. It also tends to work out better if the independent variables are zero-centered. Our preference[1] is to code the eight alternative-specific price variables, representing the eight options on the menu, as zero-centered with a range of 2. We also need to capture the inherent desirability of the eight options on the menu, plus the None alternative. We do this with K-1, or 7 columns coded either as effects-coding or dummy-coding, plus a separate column for the None parameter. Thus, the total number of columns in the independent variable matrix is:

7 dummy or effects codes (desirability of 8 menu options)
8 alternative-specific price coefficients (zero-centered)
1 dummy code (None weight)
= 16 total parameters

Once the CBC/HB model is run, in allocation mode, a set of utilities is made available in a .CSV file, and a simulator may be built in Excel. When using the logit rule to simulate choices for the sample for Exercise 1, these logit "shares of preference" need to be multiplied by 8 for each

[1] We have described capturing the price effects as simple linear terms in this paper. Researchers should be on the lookout for non-linear price functions that are captured better using non-linear specifications, such as log-linear, quadratic, piecewise, or part-worth functions.

respondent (the weight representing the maximum number of items that can by "purchased" on the menu).

The key benefit of using the Volumetric CBC Model for MBC experiments is that it can estimate the likelihood of choosing multiple binary items on the menu using a single model. The main drawback is that it is ***not a theoretically sound model***, since the predictions are volumes rather than probabilities of choice, and those volumes are not bounded by 0 and 1.0.

Despite the obvious flaw in this model, it actually seems to work well for the data set employed in this paper. The aggregate predictions match the choice likelihoods for the holdout respondents very well. Aggregate predictions are one thing, but one may question what is happening at the individual level. For Exercise #1, even if we set all items on the menu to their lowest prices, 95% of the predicted "volumes" for alternatives (after multiplying the share of preference by a volume of 8) fell within the 0 to 1.0 range at the individual level. These, of course, are supposed to be likelihoods of choice bounded by 0 and 1.0, but the Volumetric CBC Model does not formally recognize these as likelihoods. It treats them as volumes of items purchased (without constraining the volume per person to be limited to 1). Even with all items set to lowest prices, the largest predicted volume for any one respondent (out of n=800) for any one item on the menu was 1.22. So, despite the theoretical shortcomings, the model seems reasonably well behaved.

Exhaustive Alternatives Model

This model assumes that respondents approached the menu task by considering *all* possible ways that the menu could be completed, and choosing the *one* most preferred way. For example, if the menu included just four binary items, there are 2^4=16 possible combinations of checks that can be done on the menu. To code the data for this example using the Exhaustive Alternatives Model, we treat each menu task as a discrete choice among 16 alternatives. Each alternative has a total price associated with it (or, the prices can be separated as item-specific price coefficients). The desirability of the 16 possible combinations is coded as K-1 or 15 dummy-coded columns in the independent variable matrix.

This approach only works well in practice when the total number of possible ways that respondents can complete the questionnaire is a reasonably small number, rather than in the multiple hundreds or thousands of possibilities.

For Exercise 1, there were 2^8 = 256 possible ways that respondents could complete the menu. This would have resulted in each choice task being coded with at least 256 total columns in the independent variable matrix and 256 rows per task. With 800 respondents and 8 tasks each, such a problem becomes too large to run in reasonable time with HB estimation. We can reduce the size of the problem if we recognize that only about 150 of the possible 256 combinations were ever chosen by respondents (and assume the other combinations have zero likelihood of choice), but this still results in a very large problem.

But, for Exercise 2 (see Figure 5), there were only 48 unique possible ways to complete the menu ($3(2^4)$). This is quite manageable with CBC/HB software. Thus, each task may be coded as 48 alternatives, and the choice is coded as a single vote for one of those rows. The data matrix for one choice task is as follows:

48 rows Per task, 8 tasks per resp.

47 effects-coded columns	PriceBase	PriceOpt1	PriceOpt2	PriceOpt3	PriceOpt4	Choice
-1 -1 -1............-1	-0.5	0	0	0	0	0
1 0 0............ 0	-0.5	0	0	0	-0.18	0
0 1 0............ 0	-0.5	0	0	0.17	0	1
...						
...						
...						
...						
...						
...						
...						
...						
...						
0 0 0............ 1	0.5	0.06	0.54	-.50	0.53	0

Figure 13

The inherent desirability of each of the 48 possible ways to complete the menu task is captured as a categorical variable with 48-1 = 47 effects-coded columns (row 1 in Figure 13 is the reference level). Next, we have captured separate price effects for the base price of the vehicle (PriceBase), and separate effects (PriceOpt1...4) for the prices for the options included in each way to complete the menu (note the zero prices when options are not a part of the alternative). Again, this reflects the idea that the respondent actually considered all 48 ways the menu could be completed, together with the price implications for each, and chose the *one* way with the perceived highest overall utility.

Once one estimates the model using MNL, there is some work on the back end to convert the predicted likelihoods for each of the 48 alternatives in the coded tasks into predictions of choices of items from the original menu. For example, option 1 on the menu might be coded in the "on" state only in alternatives 25 through 48. Therefore, one accumulates the total likelihood of respondents picking option 1 from the menu by summing the predicted likelihoods across coded alternatives 25 through 48.

The benefits of the exhaustive alternatives model is that it formally recognizes and predicts the *combinatorial outcomes* of menu choices, rather than just the marginal choices of each individual item on the menu. It is thus a more complete model of consumer choice. The drawback is that the model can become quite sparse at the individual level, with large numbers of independent variables and relatively few choice sets. This can lead to overfitting. Researchers may find that aggregate logit or latent class models do quite well with these exhaustive alternatives models, given the sparse nature of the data. Furthermore, because of the size of the models, runtime speed can become a real issue for HB.

Serial Cross Effects Model

This approach breaks the menu down into a series of separate choice models. For example, with Exercise 1 (Figure 3), we could treat this as eight separately run binary logit models. In each model, we are predicting the likelihood of selecting that menu item or not, given the inherent desirability of the item, its price, and the prices of other items on the menu.

Consider the choice of Alloy Wheels in Exercise 1. We may build a separate logit model to predict whether Alloy Wheels is selected or not (we treat "or not" the same as selecting the "None" within standard CBC model coding). The predictor variables include: the desirability of Alloy Wheels, the price of Alloy Wheels, and each of the prices of the other items in the menu. Conceptually, the model looks like Figure 14, where the arrows represent the effect of prices of different menu options on the choice of Alloy Wheels:

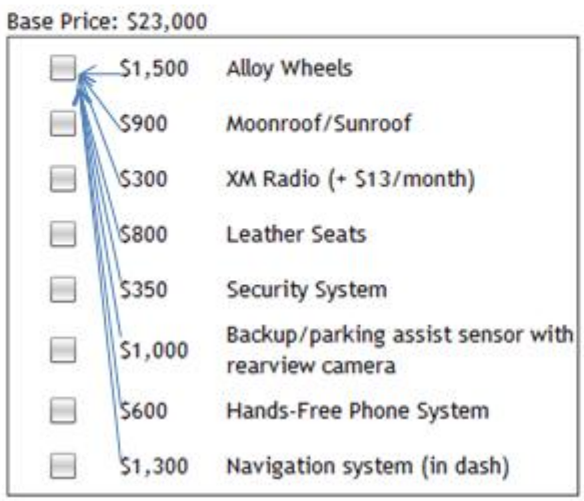

Figure 14

We build eight such separate logit models, where each model predicts the likelihood of picking a different item on the menu. The models are interconnected via the cross-effects terms (for example, the price of Moonroof affecting the purchase likelihood of Alloy Wheels, etc.).

While most menu studies in practice may use binary items (select or do not select), there are many menu situations that involve more than two mutually-exclusive choice outcomes, such as: 1) Standard cloth seats, 2) Black leather seats, 3) Cream leather seats. This is very simple to manage with the Serial Cross-Effects Models and MNL. Rather than coding each choice set with two alternatives, we expand to incorporate three alternatives.

There is an old saying: The best way to eat an elephant is one bite at a time. The key advantage of the Serial Cross-Effect Model approach is that very complex menus may be broken up into smaller, digestible pieces. For each checkbox on the menu, we can develop a separate quite manageable model—especially if we only include the cross-effects that seem to have a significant effect on choice.

One of the main disadvantages of the Serial Cross-Effect Model approach is that it can be a hassle (and error-prone) to build so many separate models. Another problem is that if all possible cross-effects are allowed into the model, then the resulting what-if simulator may produce some strange results that are just due to random error. For example, decreasing the price of carrots on the menu may lead to a tiny (non-significant) *increase* in the likelihood for purchasing the fishburger. In reality, this effect may be non-significant, and if the relationship lacks face validity, it may only cause the client some consternation. Pruning the model of non-significant effects is one way to reduce this problem. Imposing utility constraints is another.

As with the Volumetric CBC Model, a big weakness with the Serial Cross Effects Model is that it doesn't formally recognize combinatorial outcomes (multiple items being selected together), but instead focuses on being able to predict the marginal choices of each separate item on the menu. While this may not be especially detrimental for predicting the average choice likelihood for items across the menu for the sample, it leads to less accurate individual-level predictions of the actual combinations of menu items selected. That said, most managers are interested in the average predictions of choice likelihood for the menu (given price changes) rather than predictions of the *actual combinations* that will be purchased. If the latter is the goal, then Cohen and Liechty recommend multivariate probit (Cohen and Liechty, 2007).

Some Menu Choices Depend on Other Menu Choices

A common hurdle to overcome in MBC questionnaires is when some choices on the menu may only be made if another choice has first been selected. For example, in Exercise 2 (Figure 5), the respondent cannot pick Alloy Wheels for vehicle #1 unless he has first chosen vehicle #1. For the fast food example in Figure 2, we restricted respondents from picking medium French Fries from the menu if they also picked a value meal (which included the Medium French Fries).

There is a straightforward way to handle dependent choices. Consider the choice of French Fries from the *a la carte* section of the menu in Figure 2. Respondents can only pick French Fries if they first rejected all value meals. Therefore, to predict the likelihood of respondents picking among the *a la carte* French Fry options, we only include in the model estimation the tasks where respondents rejected the Value Meals. Next, at the individual-level, we use the logit rule to predict the likelihood of choices on the menu. We multiply the likelihood that the respondent rejected all Value Meals by the likelihood from the second model (that involved task filtering) of picking each *a la carte* French Fry option. This formally recognizes via the choice simulator that respondents cannot pick both value meals and *a la carte* French fries.

For Exercise 2's automobile configuration task (Figure 5), we might similarly assume that respondents follow a 2-stage decision process: first, choose the vehicle; second, configure the chosen vehicle. To accomplish this, we build an MNL model that predicts the likelihood of selecting vehicle 1, vehicle 2, or vehicle 3 (given their base prices and the prices of their options). This very much resembles a standard CBC MNL formulation, since it is a forced mutually-exclusive choice. Next, we can use either the Volumetric CBC Model, the Exhaustive Alternatives Model, or the Serial Cross-Effects Model, to predict the likelihood of selecting among the four options (Alloy Wheels, Moonroof, XM Radio, Navigation System), *given* that this vehicle (column) was selected. (Again, this is done by using only the choice tasks where the respondents picked the vehicle to develop the models that predict the selection of items within that vehicle.) When we simulate the likelihood of configuring options within each vehicle at the individual level, we multiply the likelihood of selecting that vehicle by the likelihood of configuring items on that vehicle (given the choice of that vehicle).

As we report the results below, we'll refer to this as the "2-Stage Model."

Results

Earlier, we described that the data collection employed 1600 total respondents. Each respondent completed eight choice tasks, where the prices varied across the tasks. 800 respondents were randomly selected to be *calibration respondents*, used for estimating models and building a What-If simulator in Excel. These respondents were given one of 300 versions of the questionnaire, generated using CBC's design methodology. The other 800 respondents received an identical-looking questionnaire, except that these respondents completed one of just 3 versions of the questionnaire, again generated using CBC's design methodologies. Thus, on average, each of the three holdout questionnaire versions was answered by 800/3 = 267 respondents.

We simply tallied the percent of holdout respondents who chose each of the items, for each of the 8 choice tasks x 3 questionnaire versions = 24 menu tasks. With Exercise 1, there were 8 items that could be checked on the menu. Thus, there were 24 tasks x 8 menu items = 192 separate holdout probabilities to be predicted using the market simulator. This reflects a great deal of holdout information for validating the models.

Exercise #1 Results

The R-Squared fit for holdout predictions and the Mean Absolute Error (MAE) of prediction, under the Volumetric CBC Model, were as follows:

	R-Squared	MAE
Volumetric CBC/HB model	0.925	0.0370
Volumetric CBC/HB model (with covariates)	0.928	0.0358

Scatter Plot: Predictions vs. Actual Holdout Probabilities Using Best Model (with Covariates)

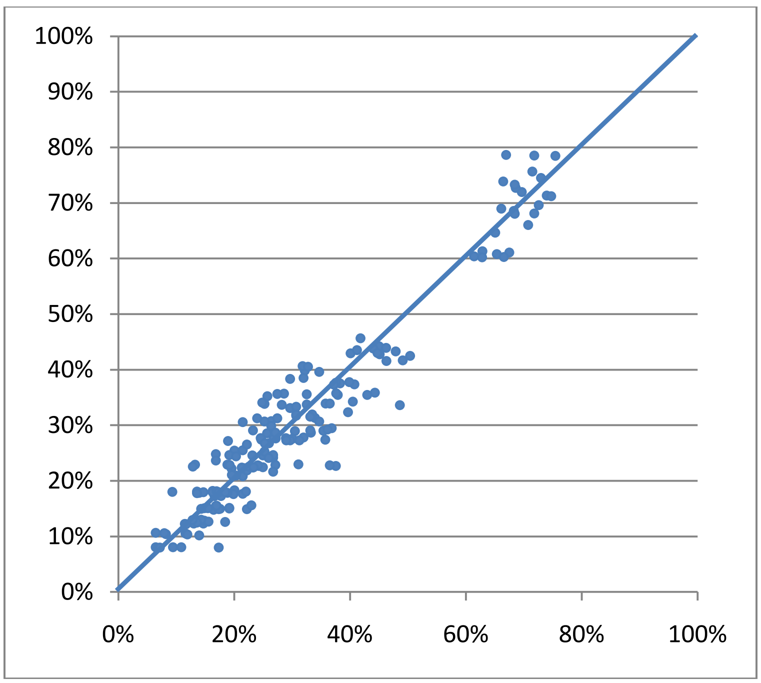

Figure 15

These predictions look very good, in the aggregate, especially considering that we cannot expect to achieve perfect predictions due to sampling error. Predictions based on 800 respondents were compared to actual choices for *different choice scenarios* completed by a *different group* of 267 respondents. Even if respondents answered without error, and our model specifications were perfect, there would still be unexplained error due to sampling error (a group of respondents predicting a separate group of respondents).

Because we found no evidence of significant cross-effects, we did not attempt the Serial Cross-Effects Model. Because there are 256 possible combinations of selections from this menu (resulting in huge data matrices and glacial run times), we did not attempt the Exhaustive Alternatives Approach. We focus much more attention on Exercise 2, which was a more complex and interesting data set.

Exercise #2 Results

The predictive fit to holdout choices for Exercise 2 was as follows, for the different approaches:

	R-Squared	**MAE**
2-Stage model (Aggregate Logit)	0.965	0.0201
2-Stage model (Latent Class)	0.960	0.0223
2-Stage model (CBC/HB)	0.960	0.0225
2-Stage model (CBC/HB, with covariates)	0.961	0.0224
Exhaustive alternatives model (Aggregate Logit)	0.954	0.0231
Exhaustive alternatives model (Latent Class)	0.956	0.0229
Exhaustive alternatives model (CBC/HB)	0.956	0.0234
Exhaustive alternatives (CBC/HB, with covariates)	0.957	0.0229
Serial cross-effects model (Aggregate Logit)	0.954	0.0226
Serial cross-effects model (Latent Class)	0.952	0.0236
Serial cross-effects model (CBC/HB)	0.942	0.0265
Serial cross-effects model (CBC/HB, with covariates)	0.951	0.0249

Scatter Plot: Predictions vs. Actual Holdout Probabilities
Using Best Model (2-Stage Aggregate Logit)

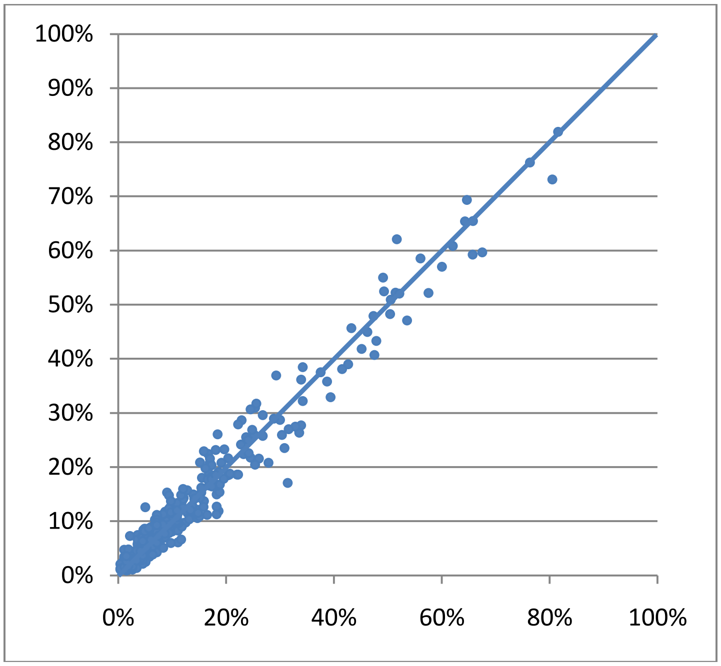

Figure 16

In general, all attempted models worked well, with R-squared values better than 0.94. Given that we cannot achieve perfect fit to holdout data given sampling error, we are doing about as well as possible.

For Exercise 2, we also tried aggregate logit and latent class. It is very interesting to note, and somewhat surprising, that the aggregate logit approach achieved the highest predictive validity (the 2-Stage model, with R-squared of 0.965) for this data set. This is surprising, since our experience with traditional CBC is that HB generally shows higher predictive validity for holdouts than aggregate logit (at least generic aggregate logit models *without* cross-effects). Aggregate logit is notoriously prone to IIA difficulties, and one of the benefits of HB is the ability to reduce IIA troubles for standard CBC studies. But, with MBC models, IIA should be less of a concern. IIA is concerned with maintaining the ratio of choice likelihoods for competing alternatives, when a given alternative is added or modified within a choice scenario involving at least three alternatives. If using a series of binary logits to estimate choice likelihoods of each item on the menu, there are only two alternatives in each model. Furthermore, differential substitution effects among items can be accounted for using cross-effect terms (which we've done here with the aggregate models). As a final point, IIA is less of a concern if the full context of all available products was shown in each choice task, which is often the case with MBC studies[2].

[2] MBC studies may also vary the presence of items on the menu (availability designs). When this is the case, the availability of an item may be used as a cross-effect for predicting the likelihood of choosing other items on the menu. With availability designs, HB models may have an additional advantage over aggregate models, but this remains to be proven with additional datasets.

Given the amount of evidence across the industry in favor of HB methods, it would be unwise to dismiss HB modeling as unnecessary for MBC studies based on our findings with this one dataset. HB provides the liberating convenience of easy on-the-fly filtering by respondent segments in the market simulator, without having to go back to square one and re-estimate the model (as with aggregate logit) for each segment. This benefit alone will make it worth the effort for most practitioners to use HB rather than aggregate logit. Also, the only standard for success reported here is aggregate predictive validity. We haven't worried about individual-level prediction, especially prediction of combinations of choices (such as reported in Figure 8). Individual-level models would seem to have an upper hand if this were the goal.

The approaches that formally recognize that there were logical exclusions in the way that respondents could complete the menu (prohibited choice combinations) tended to perform better than the method (Serial Cross-Effects Model) that ignored these exclusions. This seems quite logical and reasonable, and researchers should take care to use models that formally recognize any logical exclusions within MBC questionnaires. If we were modeling a questionnaire that didn't include any logical exclusions, the Serial Cross Effects model may have been the preferred model. For example, at this conference, Chris Moore of GfK reported solid results for a restaurant menu study when using the Serial Cross Effects approach.

As a final note, using covariates within CBC/HB makes very little difference in predictive accuracy over standard HB, though the results suggest perhaps a tiny directional improvement. This is in line with other research we've conducted on covariates, and the results of two other papers delivered at this conference (Sentis; Kurz and Binner). Based on our experiences with covariates in HB, their real value would be seen in enhanced discrimination among market segments if reporting MBC what-if simulations by segment.