Sawtooth Software

RESEARCH PAPER SERIES

Application of Covariates Within Sawtooth Software's CBC/HB Program: Theory and Practical Example

Bryan Orme and John Howell,
Sawtooth Software, Inc.

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

Application of Covariates within Sawtooth Software's CBC/HB Program: Theory and Practical Example

[1]Bryan Orme and John Howell, Sawtooth Software

Importance of CBC (Choice-Based Conjoint)

Over the last decade, hierarchical Bayes (HB) estimation of part-worths has had a significant and positive impact on the analysis of discrete choice (CBC) data. Certainly, HB has been key to the emergence of CBC (Choice-Based Conjoint) as the most popular conjoint-related method (Sawtooth Software, 2003). Sawtooth Software released its CBC/HB software for estimating part-worths from CBC questionnaires in 1999, based on earlier work published by Greg Allenby and Peter Lenk (Allenby *et al.* 1995, Lenk *et al.* 1996), as well as workshops given by Allenby & Lenk at the American Marketing Association's ART/Forum conferences. Although the focus of this paper is on the use of CBC/HB for analyzing CBC data, similar data such as MaxDiff will also benefit from the use of covariates.

The Basic Hierarchical Model

Hierachical Bayes models require repeated measures per respondent. In CBC questionnaires each respondent picks the best alternative from multiple choice tasks. Repeated measures are needed to measure within- and between-respondent variation. The hierarchical Bayes model is called *hierarchical* because it models respondents' preferences as a function of an upper-level (pooled across respondents) model and a lower level (within-respondents) individual-level model. At the individual-level, the respondent is assumed to choose product concepts according to the sums of part-worths as specified in the logit model. The lower-level model is also a standard feature of non-Bayesian models. The upper-level model describes the heterogeneity in the individual part-worths across the population of respondents. At the upper-level, the basic HB approach assumes that respondents are drawn from a multivariate normal distribution, with part-worths ($ß_i$) distributed with means α and covariance matrix D, $ß_i \sim \text{Normal}(\alpha, D)$, where the subscript i is the respondent. HB determines the optimal weight of the upper level and lower level models in estimating part-worth estimates for each individual, resulting in high posterior probability that the part-worths ($ß_i$) fit respondent choices, given that respondents are drawn from the population distribution. In application, the upper-level model plays the role of a prior when estimating each respondent's part-worths, and the lower-level model provides the likelihood. Because it leverages information from the population parameters α and D, HB is able to estimate stable and useful part-worths for each individual even in the case of relatively sparse data.

Versions 1 through 4 of Sawtooth Software's CBC/HB program used a simple assumption: that respondents were drawn from a single population of normally distributed part-worths.

[1] The authors gratefully acknowledge Rich Johnson, Peter Lenk, and Jeffrey Dotson who provided helpful comments and suggestions regarding this article.

Throughout this paper we will refer to this as *generic HB*. While this assumption may seem overly simple, it actually performs quite well in practice. The single-normal-population assumption is only an *influencing* factor on the final part-worth estimates, and does not constrain the final part-worths to reflect normality. Since HB represents a weighted combination of the upper- and lower-level models, if enough information (i.e. choice tasks) is available at the individual-level, the influence of the upper-level model is much less than the lower-level model, which seeks only to fit each individual's choices. For example, the authors have seen bimodal distributions for a brand part-worth result from HB estimation of CBC data, because the strong bimodal preference for the brand as exhibited by individuals in the sample was an overriding factor in the weighting. The influence of the upper-level model provides some degree of Bayesian shrinkage toward the global mean, which tends to smooth the distribution somewhat, with a *tendency* toward normality. But again, if a substantial number of choice tasks are available relative to the number of parameters to be estimated (as is typical with CBC applications in practice), the Bayesian shrinkage is usually modest, making the troughs less deep in multi-modal distributions of preference, and drawing the means for distinct populations of respondents closer together. The bottom line based on extensive simulation studies and experience is that HB estimation is fairly robust to the normal assumption of part-worth heterogeneity.

The Possibility of Covariates

Despite the general success of CBC/HB software with the assumption of single population normality, there are some instances where CBC/HB users have raised concerns:

- In situations in which a segment of respondents (with divergent preferences) is oversampled, this can bias the estimates for the segment means as well as the overall population means – which especially can bias the betas for the undersampled group of respondents (Howell, 2007).
- When conducting segmentation studies, some researchers have expressed concern that distances between segment means are diminished (because HB shrinks the individual estimates of the part-worths towards the population mean), whereas clients like to see large differences between segments.
- Advanced HB practitioners have recommended that in many cases, well-chosen covariates can provide additional information to improve parameter estimates and predictions.
- The notion that respondents are drawn from a single normal population has struck many researchers and their clients as unrealistic.

Some early applications of HB estimation of part-worths for conjoint experiments used a more flexible definition of the upper-level model in HB (Lenk *et al.* 1996). Rather than assuming that all respondents were drawn from a single multivariate-normal distribution, they allowed a more flexible definition of the population distribution based on respondent characteristics (covariates). *Covariates* is another term for additional independent variables that may be predictive of some outcome (the dependent variable). Often, we think of covariates such as common demographics like gender, age, income, company size, geographic location, etc. Unfortunately, these variables often have low correlation with preferences within choice contexts. The most useful covariates bring *exogenous* information (outside the information already available in the choice tasks) to the model to improve the estimates of part-worths and improve market predictions. In the example shown in this paper, the dollar amount respondents expected to pay for their next PC purchase was a useful covariate to improve part-worth estimates that included such attributes as brand preference, PC performance, and price sensitivity.

More formally, rather than assume respondents are drawn from a normal distribution with mean vector α and covariance matrix *D*, an HB model with upper-level covariates assumes that respondents' part-worths are related to the covariates through a multivariate regression model:

$$\beta_i = \Theta' z_i + \varepsilon_i \text{ where } \varepsilon_i \sim \text{Normal}(0,D)$$

where Θ is a q by b matrix of regression parameters, z_i is a q vector of covariates, and ε_i is a b vector of random error terms. The part-worths are drawn from a normal distribution with means $\Theta' z_i$. Instead of shrinking the individual estimates to the population mean α, the multivariate regression model shrinks them to the conditional mean Θz_i given the subject's covariates. For example, if gender is a covariate, and preferences differ between the genders, then a woman's part-worths are shrunk toward those of other women, while a man's part-worths are shrunk toward those of other men. In contrast, generic HB shrinks both men's and women's part-worths to a common population mean for all subjects. In this way, the multivariate regression, upper-level model can automatically use observed, segment basis variables (e.g. Gender, Security Seeking, Experience, etc.) to refine the estimation of the part-worths and increase the distinction between segments.

If there are n subjects in the study, the model for all of the $\{\beta_i\}$ can be written in matrix notation, which we give here for completeness. Let B be a n by b matrix where the i^{th} row consists of the part-worths β_i for subject i. Then

$$B = Z\Theta + \Delta$$

where Z is a n x q matrix with z_i in the i^{th} row; Θ is a q by b matrix of regression coefficients, and Δ is a matrix normal distribution with ε_i in the i^{th} row.

In the generic HB model with the single normal population assumptions, there are b + [b(b+1)]/2 parameters to be estimated in the upper-level model, where b is the number of part-worths for each individual. Breaking that equation down, assuming there are b parameters for the mean population parameters (α), and since the covariance matrix (D) is symmetric, there are [b(b+1)]/2 parameters to estimate on and above the diagonal. With covariates in the upper-level model, there are bq parameters to be estimated in the upper model, where b is the number of part-worths and q is the number of covarariate parameters (dummy-coded for categorical variables, and with the possibility of continuous variables). If Gender was the covariate, consisting of males and females, q would equal 2. There would be one set of parameters (for each part-worth) for the intercept and one additional set of parameters (again, one for each part-worth) for the dummy contrast between males and females. Including covariates in the upper-level model doesn't alter the estimation of covariance matrix D. This makes it a more parsimonious model than separating the sample by gender and running generic HB within the separate samples. In that case, the vector of sample means (16 parameters, in our example that follows)) plus the covariance matrix consisting of [16(16+1)]/2 = 136 parameters would have to be estimated for the upper model in each sample. Estimating as a single HB run with a dummy-coded covariate for gender saves 136 parameters in the upper-level model versus segmenting the data set and running generic HB within the separate samples. A few covariate columns (when dummy-coded) added to an HB run increases the length of the run time, but typically by only about 10 to 25%.

The covariates HB model can be estimated in much the same way as the generic HB model using MCMC and a Gibbs Sampler. The only difference is that instead of using a common α for all respondents, respondents are compared to their predicted mean based on the characteristics in their specific $Z\Theta$ vector. In addition, instead of estimating α from a multivariate normal distribution, Θ is drawn from a Matrix Normal distribution with means 0 and variances 100. For more details on MCMC and the Gibbs Sampler please see the CBC/HB Technical Paper from Sawtooth Software (Sawtooth Software 2005).

Sawtooth Software investigated extending the CBC/HB software to include covariates in the upper-level model as early as about 2001. Peter Lenk was helpful to us as we developed code for that investigation. After examining a few data sets, our feeling was that although covariates seemed to offer modest benefit in specific cases, we worried about burdening our CBC/HB users with additional decisions about the choice of covariates. We were more concerned that users could obtain good results quite automatically, without giving rise to user angst regarding whether more should be done by investigating a number of potential covariates. Because HB was a relatively new technique at the time, and because each HB run often took 4 hours or more given hardware technology in 2001, we decided to keep our CBC/HB software as straightforward and parsimonious as we could.

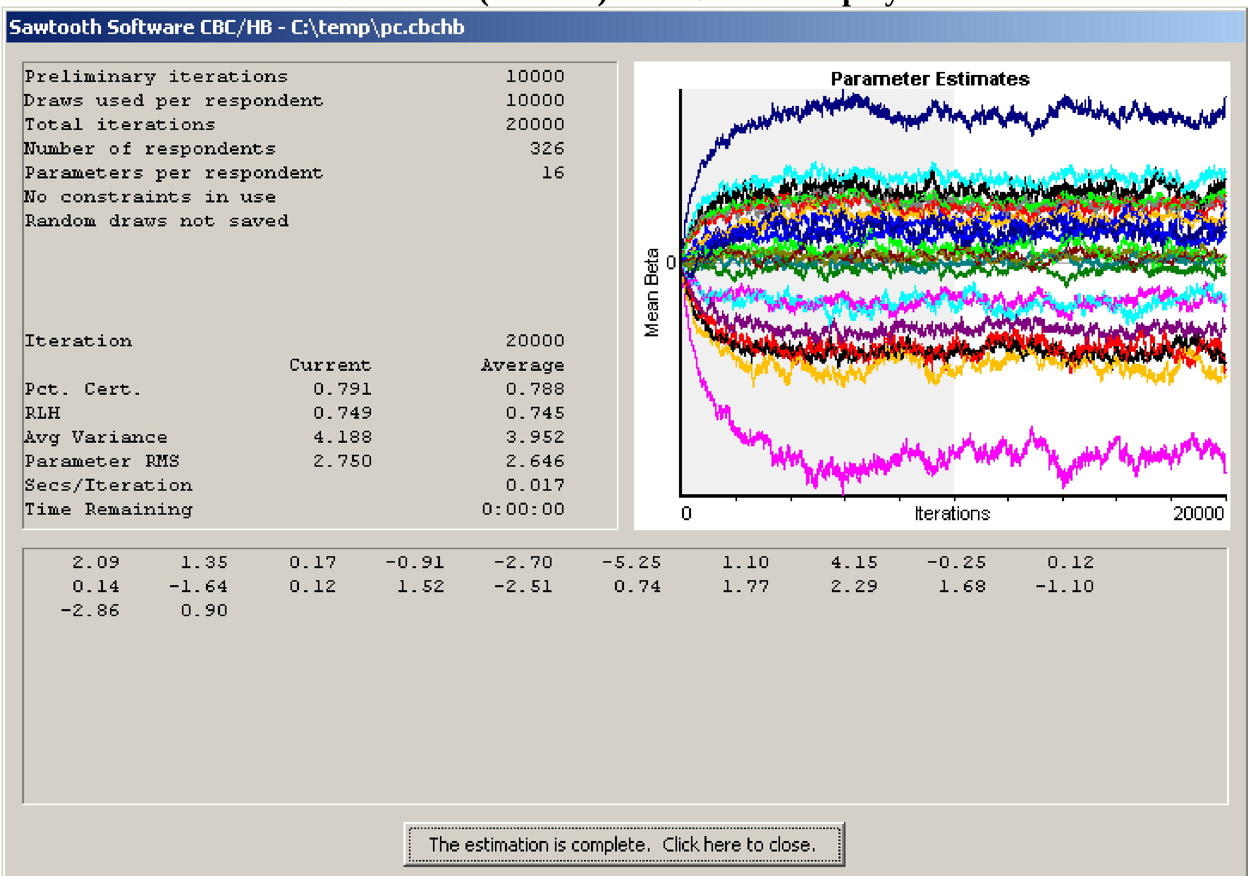
In 2001, Sentis and Li undertook a large investigation using our CBC/HB software to determine if improvements in hit rates could be seen by first segmenting the sample by meaningful covariates and running generic HB separately within each sample (Sentis and Li, 2001). The overall hit rates after recombining the samples were no better across 7 datasets (each with n=320 to n=800) than when respondents were simply combined. The additional parameters involved in estimating a new covariance matrix D for each subsample likely swamped the possible improvements that could have been seen by allowing respondents' part-worth estimates to be influenced (via Bayesian shrinkage) only by their peers.

Example Using the PC Dataset

In the early 1990s, IntelliQuest and McKinsey collected a CBC dataset regarding the purchase of PCs by purchase managers at organizations. They shared the data with Sawtooth Software, and we have shipped the data set as an example with installations of our SMRT platform. The data set includes 326 respondents who answered an 8-task CBC study. Each task included 3 product alternatives plus a "None" option. The products were defined on 6 attributes, as a $3^4 4^1 5^1$ design (16 parameters to be estimated, when including the "None" constant). Twelve percent of the choices were "None." In addition to the CBC tasks, respondents were asked a number of questions regarding their demographics/firmographics, including questions such as "how much do you plan to spend on the next PC you plan to purchase?"

Using CBC/HB v5 under its default settings (10,000 preliminary iterations, 10,000 used iterations), the history of the run looks as follows:

Exhibit 1
Standard (Generic) HB Runtime Display



The history of mean beta across iterations suggests convergence after about 5,000 iterations. The mean posterior estimates of betas are summarized for each respondent as point estimates (an average of all 10,000 draws for each respondent) within the pc_utilities.csv file. The average RLH (Root Likelihood) fit across respondents was 0.758; though in the data file, CBC/HB reports RLH as RLH x 1000, so the average fit you can compute by taking the mean of the RLH column is 758.

The mean point estimates of part-worths for the sample are shown in Table 1:

Table 1
Mean Part-Worths, Generic HB Run

1.78	Brand A
1.03	Brand B
0.32	Brand C
-1.01	Brand D
-2.12	Brand E
-4.78	Below average performance
1.11	Average performance
3.67	Above average performance
-0.19	Order over the telephone
0.06	Obtain from a retail store
0.14	Obtain from a sales person at your site
-1.65	90 day warranty
0.18	1 year warranty
1.46	5 year warranty
-2.21	Ship back to mfg for service
0.66	Service at local dealer
1.54	On-site service
2.10	Low Price
1.40	Med-Low Price
-0.95	Med-High Price
-2.55	High Price
0.78	NONE

One of the available variables asked outside the CBC questions was the amount the respondent planned to spend on the next PC purchased. It was coded as 5 categories:

Table 2
Distribution of Responses, ExpectToPay Covariate

Category	# Respondents	% of Respondents
1) $1,000 to $1,499	61	18.7%
2) $1,500 to $2,499	194	59.5%
3) $2,500 to $3,499	51	15.6%
4) $3,500 to $4,299	10	3.1%
5) $4,300 or more	7	2.1%
Missing	3	0.1%

If we collapse respondents into two groups (plan to spend <$2,500; plan to spend $2,500+), we can examine the differences between these two groups using simple Counting analysis of main effects (percent of times each level was chosen, when available):

Table 3
Counting Analysis by Segments based on ExpectToPay Covariate

Brand (customized per respondent) by ExpectToPay

	Total	<$2,500	$2,500+
Total Respondents	326	255	68
Brand A	0.41	0.39	0.48
Brand B	0.32	0.33	0.31
Brand C	0.29	0.30	0.27
Brand D	0.24	0.25	0.19
Brand E	0.20	0.21	0.16
Within Att. Chi-Square	140.93	80.61	70.94
D.F.	4	4	4
Significance	p < .01	p < .01	p < .01
Between Group Chi-Square	11.78		
D.F.	4		
Significance	p < .05		

Performance by ExpectToPay

	Total	<$2,500	$2,500+
Total Respondents	326	255	68
Below average performance	0.08	0.09	0.06
Average performance	0.30	0.31	0.26
Above average performance	0.50	0.50	0.52
Within Att. Chi-Square	792.97	583.69	207.74
D.F.	2	2	2
Significance	p < .01	p < .01	p < .01
Between Group Chi-Square	6.13		
D.F.	2		
Significance	p < .05		

Channel by ExpectToPay

	Total	<$2,500	$2,500+
Total Respondents	326	255	68
Order over the telephone	0.28	0.28	0.27
Obtain from a retail store	0.29	0.30	0.30
Obtain from a sales person at your site	0.31	0.31	0.28
Within Att. Chi-Square	3.21	3.50	0.64
D.F.	2	2	2
Significance	not sig	not sig	not sig
Between Group Chi-Square	0.77		
D.F.	2		
Significance	not sig		

Warranty by ExpectToPay

	Total	<$2,500	$2,500+
Total Respondents	326	255	68
90 day warranty	0.21	0.21	0.21
1 year warranty	0.30	0.30	0.28
5 year warranty	0.37	0.38	0.36
Within Att. Chi-Square	114.69	93.85	23.03
D.F.	2	2	2
Significance	p < .01	p < .01	p < .01
Between Group Chi-Square	0.25		
D.F.	2		
Significance	not sig		

Service by ExpctToPay

	Total	<$2,500	$2,500+
Total Respondents	326	255	68
Ship back to mfg for service	0.19	0.18	0.19
Service at local dealer	0.32	0.32	0.30
On-site service	0.38	0.39	0.36
Within Att. Chi-Square	172.02	144.22	27.99
D.F.	2	2	2
Significance	p < .01	p < .01	p < .01
Between Group Chi-Square	0.60		
D.F.	2		
Significance	not sig		

Price (customized per respondent) by ExpectToPay

	Total	<$2,500	$2,500+
Total Respondents	326	255	68
Low Price	0.40	0.42	0.33
Med-Low Price	0.36	0.37	0.31
Med-High Price	0.23	0.22	0.28
High Price	0.18	0.17	0.20

Within Att. Chi-Square	224.93	222.60	13.96
D.F.	3	3	3
Significance	p < .01	p < .01	p < .01
Between Group Chi-Square	16.12		
D.F.	3		
Significance	p < .01		

None by ExpectToPay

	Total	<$2,500	$2,500+
Total Respondents	326	255	68
None chosen:	0.12	0.11	0.15
Between Group Chi-Square	7.51		
D.F.	1		
Significance	p < .01		

We see that the percent of times different levels are chosen is often significantly different between the two segments of respondents. For example, the "low expect to pay group" chose Brand A 39% of the time, but the "high expect to pay group" chose Brand A 48% of the time. The Between Group Chi-Square test shows that Brand and Performance have p values <.05, and the Price attribute and None parameters have p values <.01. This indicates that the probability of observing these differences in the choice probabilities by chance for levels of these attributes is less than 5% for Brand and Performance, and less than 1% for Price and None.

We might summarize some of the differences observed in Table 3 with statements like the following:

> Compared to respondents who expect to pay less than $2,500, respondents who expect to pay $2,500 or more for their next PC…

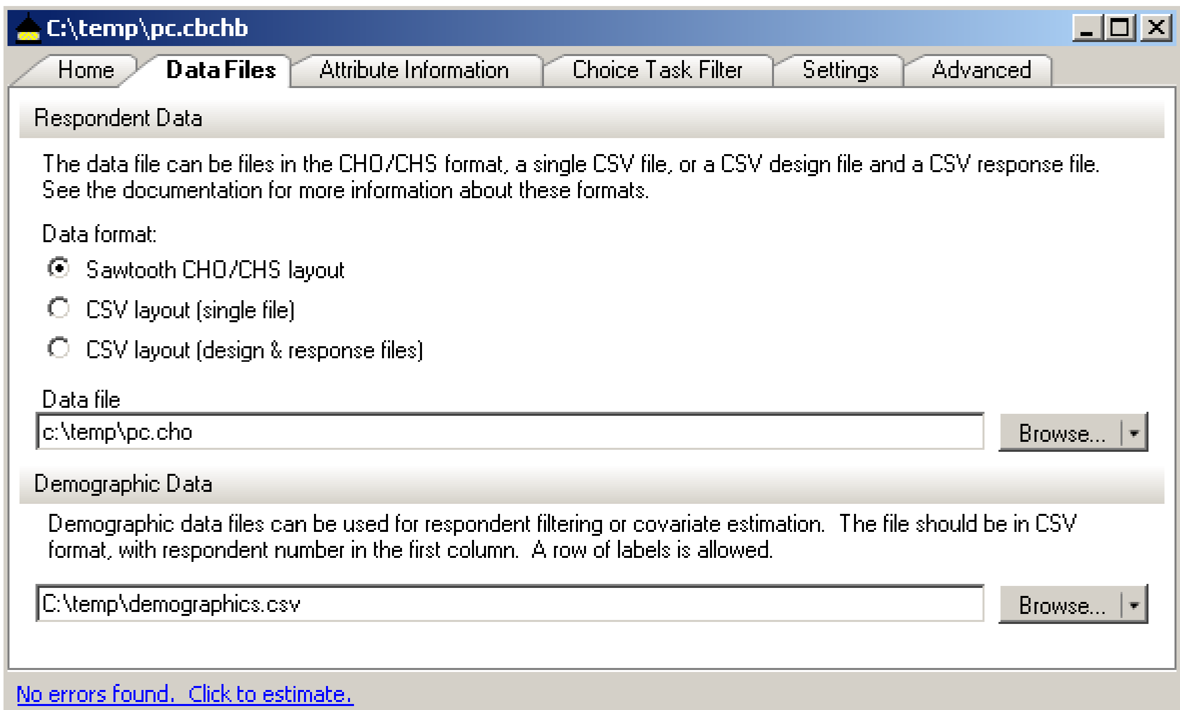
- Are more likely to favor Brand A (a premium brand), and less likely to prefer Brands D or E (discount brands)
- Are more likely to prefer better Performance
- Are less price sensitive
- Use the "None" option more often

Given some of the strong differences in preference between the groups (especially on price sensitivity), it might make sense to investigate using this variable (ExpectToPay) as a covariate in a new HB run.

Including a Covariate in CBC/HB v5 Software

To include a covariate in a CBC/HB run, we first need to associate a CSV formatted file with the project that includes the CaseID and demographic/segmentation information.

Exhibit 2
Specifying Data Files in CBC/HB v5



In the demographics.csv file, we did two things to pre-process the ExpectToPay variable. Recall that the ExpectToPay variable was coded as 5 categories. First, we resolved the 3 cases with missing data by simply imputing the modal value (2). Next (in Excel), we computed the mean of the ExpectToPay column across respondents, and then subtracted that mean value from each respondent's ExpectToPay value, thus zero-centering the values. This isn't necessary, but it makes it easier to interpret the Θ weights associated with the intercept and the ExpectToPay variable. These weights are written to the studyname_alpha.csv file (rather than the standard estimates of alpha, as in the generic HB run).

The demographics.csv file (when viewed in Excel) looks like:

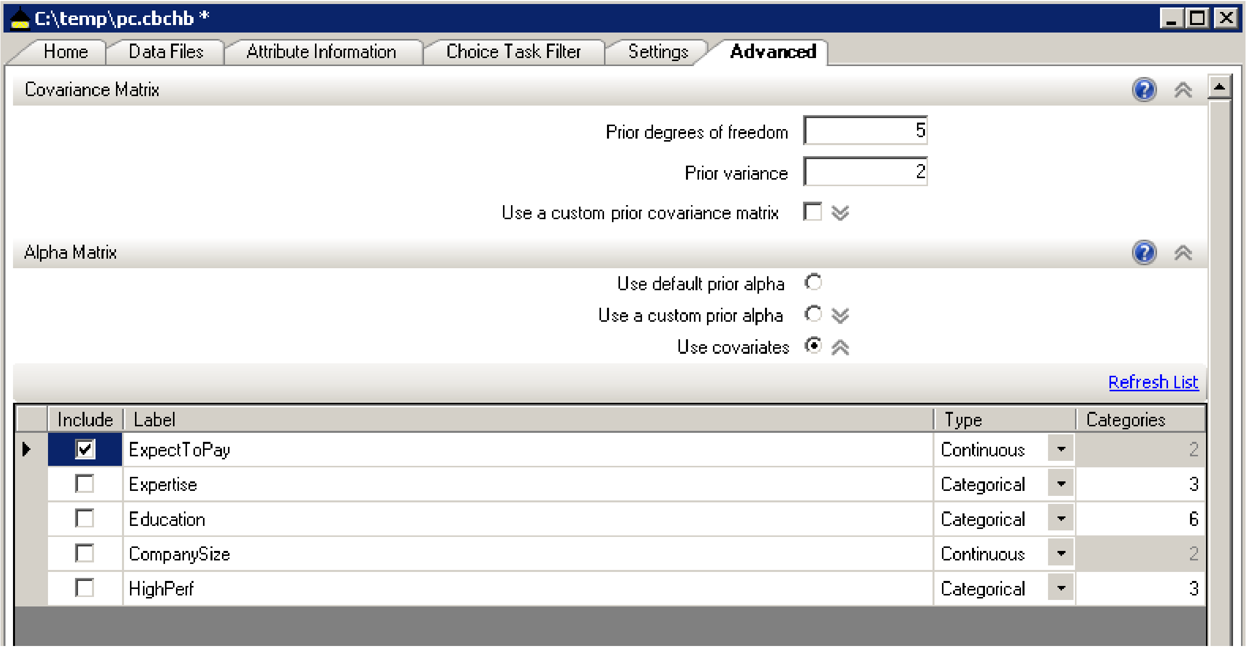
Exhibit 3
CSV-Formatted File Containing Demographic/Covariate Information

	A	B	C	D	E	F
1	CaseID	ExpectToPay	Expertise	Education	CompanySize	HighPerf
2	1001	0.896	3	0	0.113	1
3	1003	-0.104	3	5	-1.887	3
4	1005	-1.104	2	4	3.113	1
5	1012	-1.104	2	2	-1.887	2
6	1014	0.896	3	4	1.113	1
7	1018	0.896	3	0	2.113	3
8	1022	0.896	2	4	1.113	3
9	1024	-0.104	2	4	-1.887	1
10	1029	-0.104	2	3	1.113	3
11	1035	-1.104	3	3	-2.887	2
12	1052	-0.104	3	4	2.113	1
13	1058	0.896	3	5	1.113	3
14	1067	-0.104	2	5	1.113	1
15	1068	-0.104	2	3	-0.887	3
16	1073	-0.104	3	2	-1.887	3

CaseID is the first variable, the zero-centered ExpectToPay variable is the second column, and other segmentation variables that could be employed as covariates are shown in the other columns.

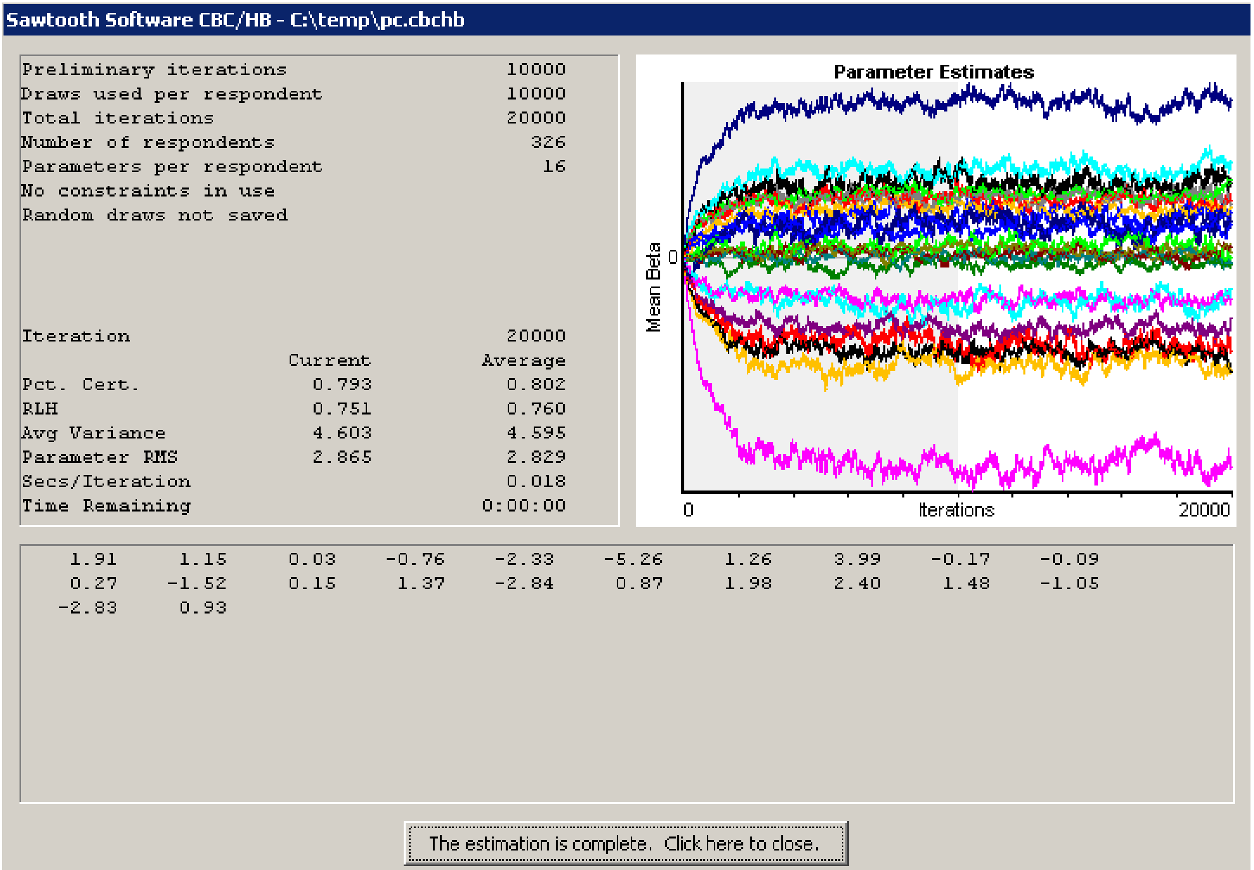
Then, on the *Advanced* tab in CBC/HB v5, we reveal the *Covariates* table, and select to include ExpectToPay. The software automatically detects that the variable is type *Continuous*.

Exhibit 4
Selecting Covariates in CBC/HB v5



When we estimate the model, the runtime display looks identical to the original default run:

Exhibit 5
Runtime Display of HB Run with Covariates



The speed of the iterations is one per every 0.018 seconds rather than one per every 0.017 second with the default run. Running HB with a single continuous covariate makes the iterations only a slight bit slower. To complete 20,000 iterations takes about 6 minutes, on just a mediocre to ordinary PC by today's standards. (Quite a far cry from the 6+ hour runs commonly experienced a decade ago!)

The summary statistics for this run look quite similar to the generic HB run shown earlier, except for the fit (RLH) and scale (Parameter RMS) of the parameters seem a bit higher. Examining the PC_utilities.csv file, we find that the average fit for the part-worths (RLH x 1000) is 770. This compares to 758 for the generic HB run. The fit has slightly improved, but an improvement in fit for the individual part-worths in predicting respondent choices does *not* necessarily mean a better set of part-worths and a better model. Overfitting has occurred if improved fit to choices used in estimating part-worths does not lead to improved fit to choices held out of part-worth estimation and used for validation. We'll speak more to that later.

Two important files that are produced are:

> **PC_utilities.csv** (contains the point estimates of part-worths for each respondent. The point estimates are the average of the 10,000 used draws of beta). The same information is saved to the PC.hbu file, for those who wish to take this information to Sawtooth Software's market simulator.
>
> **PC_alpha.csv** (contains the estimates of alpha for each of the 20,000 draws. In the generic HB run, these are the estimates of population means for each of the part-worths. But, with our covariates run that includes a single covariate coded as a continuous variable, we find two sets of weights associated with each of the levels in the study. The first set of weights represents the mean population estimates of part-worths (the intercept of the covariates regression) when the covariates are dummy-coded as a row of zeroes in the covariate design matrix (representing the reference/last levels if using categorical covariates). In our situation, we have just one covariate, it's a continuous variable, and it is zero-centered. So, the first set of weights represents the population estimates of part-worths when the ExpectToPay variable is at zero. Since a zero represents the average value for the population on ExpectToPay, the first set of part-worth population estimates will correlate nearly 1.0 with the mean point estimates of beta from the PC_utilities.csv file. The second set of weights associated with each part-worth level in the PC_alpha.csv file is the set of regression weights associated with each part-worth to represent the *adjustment in the population mean that is expected for each unit increase in ExpectToPay*.

Table 4 contains the average of the last 10,000 draws in the PC_alpha.csv file for both the mean population part-worth estimates when the covariate is in its zero state (the intercept), as well as the regression weights associated with the ExpectToPay variable. We have bolded any parameters where the draws are >95% positive or >95% negative.

Table 4
Theta Weights for Covariates Run

	Intercept	ExpectToPay
Brand A	**1.89**	**0.87**
Brand B	**1.08**	**0.44**
Brand C	0.34	-0.23
Brand D	**-1.08**	-0.26
Brand E	**-2.23**	**-0.83**
Below average performance	**-5.28**	**-0.63**
Average performance	**1.28**	-0.01
Above average performance	**3.99**	**0.64**
Order over the telephone	-0.20	-0.07
Obtain from a retail store	0.03	0.05
Obtain from a sales person at your site	0.17	0.02
90 day warranty	**-1.76**	0.02

1 year warranty	0.16	-0.13
5 year warranty	**1.60**	0.11
Ship back to mfg for service	**-2.40**	0.25
Service at local dealer	**0.75**	-0.11
On-site service	**1.65**	-0.14
Low Price	**2.32**	**-0.81**
Med-Low Price	**1.49**	**-0.37**
Med-High Price	**-1.01**	**0.45**
High Price	**-2.80**	**0.73**
NONE	**0.95**	0.40

This output means that the expected value or regression function for the part-worth for Brand A is:

$$\beta_{\text{Brand A}} = 1.89 \quad + 0.87\text{*ExpectToPay}$$

If a subject expects to pay a lot, then he or she most prefers Brand A (after accounting for the other attributes).

The mean of Theta draws for the intercept is almost perfectly correlated with the original mean part-worths shown in Table 1, but with a slightly larger scale (this is expected, given the slightly higher fit of the covariate run). Examining the magnitude of the Theta draws associated with the ExpectToPay covariate as well as the percent of draws that are positive helps us determine if the covariate was useful and has face validity. When the percent of draws that are positive is >95% or <5%, it suggests that these covariate weights are significantly different from zero, at or better than the 90% confidence level (two-sided test).

The draws of covariate weights suggest that as ExpectToPay increases, the following part-worths *increase*:

- Brand A
- Brand B
- Above Average Performance
- Med-High Price
- High Price

This makes perfect sense, because brands A and B are premium brands, people who plan to spend more will likely place greater emphasis on getting above-average performance, and such people will also have lower price sensitivity (leading to relatively higher utility for higher price points than respondents who expect to pay less).

And, the covariate weights suggest that as ExpectToPay increases, the following part-worths *decrease*:

- Brand E

- Below Average Performance
- Low Price
- Med-Low Price

It's clear that these relationships also make sense, since Brand E is a discount brand, below average performance should appeal less to folks planning to spend more, and the two low price points should be less attractive to those who plan to spend more.

This information very closely confirms the differences we saw from the counts analysis that compared respondents (by separating them into two groups) based on stated ExpectToPay (see Table 3). And, since the regression weights are based on a continuous representation of the ExpectToPay variable (instead of the simple truncation into two groups as used in the counting analysis), as well as a full Bayesian estimation of the choice model following the logit rule, we should place much greater confidence in the HB results and its estimates of how ExpectToPay affects preferences and choice.

Holdout Predictions

In a separate analysis, we re-ran the generic HB run as well as the covariate run described above. We also included a second covariate run where we indiscriminately threw all five covariate variables available to us into the run (without examining their potential usefulness as discriminators of preference or looking for multicolinearity problems). In this HB estimation, we held out one task for validation and used the other seven tasks for part-worth estimation (repeating this analysis eight times, holding out a different task in each case). We examined both the hit rate for predicting the holdout choice, as well as the likelihood according to the logit rule that the each respondent would pick the concept he/she actually chose. The hit rates and likelihood of the holdout choice (averaged across all eight replications) are given in Table 5 below. For each run, we bold the run that has the best performance.

Table 5
Hit Rates for HB Runs

Held-Out Task	Model	Hit Rate	Likelihood
Task 1	Generic HB Run	65.0%	62.8%
Task 1	1 Covariate Run	**65.3%**	**63.4%**
Task 1	5 Covariate Run	62.3%	61.4%
Task 2	Generic HB Run	64.7%	62.5%
Task 2	1 Covariate Run	**67.2%**	**64.3%**
Task 2	5 Covariate Run	65.0%	64.0%
Task 3	Generic HB Run	69.0%	66.7%
Task 3	1 Covariate Run	**69.9%**	66.2%
Task 3	5 Covariate Run	69.6%	**67.6%**

Task 4	Generic HB Run	63.8%	**62.8%**
Task 4	1 Covariate Run	**64.7%**	62.0%
Task 4	5 Covariate Run	62.6%	61.9%
Task 5	Generic HB Run	68.1%	64.8%
Task 5	1 Covariate Run	**69.3%**	**65.6%**
Task 5	5 Covariate Run	63.8%	63.9%
Task 6	Generic HB Run	**64.7%**	62.8%
Task 6	1 Covariate Run	64.1%	62.6%
Task 6	5 Covariate Run	63.8%	**64.0%**
Task 7	Generic HB Run	66.0%	63.6%
Task 7	1 Covariate Run	**66.3%**	64.3%
Task 7	5 Covariate Run	66.0%	**64.8%**
Task 8	Generic HB Run	**65.3%**	61.9%
Task 8	1 Covariate Run	64.4%	61.8%
Task 8	5 Covariate Run	62.9%	**62.8%**
Average	Generic HB Run	65.8%	63.5%
Average	1 Covariate Run	**66.4%**	**63.8%**
Average	5 Covariate Run	64.5%	**63.8%**

The average of all eight runs shows a slight edge for the covariates model for both measures. But, the covariate run with a single covariate (ExpectToPay) is the best across both the raw hit rate and probability of the holdout choice. Adding covariates indiscriminately to the run has not improved matters. But, the fact that holdout prediction stayed fairly consistent when fairly useless covariates were added to the run is evidence of general robustness to even poor specification of covariates.

Discrimination among Respondents and Segments

Holdouts, while common measures of quality of part-worths, are only one way to examine the usefulness of the data. An aim when using covariates is to promote Bayesian shrinkage toward respondents that share meaningful characteristics (such as, in our example, ExpectToPay). A common complaint about generic HB is that in the case of sparse data, it obscures differences between segments due to shrinkage to the global mean. To investigate this issue, we used one of the replicates from the holdout analysis above (HB runs using the first 7 choice tasks). We compare the importance scores (best minus worst levels for each attribute, percentaged across attributes to sum to 100) between three groups, segmented by the amount respondents expect to pay for their next PC (see Table 6). In the final column, we compute the absolute spread in importance scores across the three segments of respondents. For example, in the generic HB run, the absolute difference between Brand importance scores across the three segments is 23.39 minus 17.28, or 6.10 points. The larger the spread, the greater differentiation among segments on the part-worths. Clients typically like to see large and meaningful differences between groups, as it helps clarify and direct segmentation strategy.

Table 6
Importance Scores for Three Segments

Average Importances by ExpectToPay (Generic HB Run)

	Total (n=326)	<$1,500 (n=61)	$1,500 to $2,499 (n=194)	$2,500+ (n=68)	Spread
Brand	19.53	17.28	18.90	23.39	6.10
Performance	31.24	29.88	31.16	32.58	2.70
Channel	4.58	4.61	4.64	4.45	0.19
Warranty	12.35	11.84	12.96	11.31	1.65
Service	14.80	16.22	14.60	13.98	2.24
Price	17.49	20.17	17.74	14.30	5.87

Average Importances by ExpectToPay (Covariates HB Run)

	Total (n=326)	<$1,500 (n=61)	$1,500 to $2,499 (n=194)	$2,500+ (n=68)	Spread
Brand	20.22	15.96	19.58	25.90	9.94
Performance	31.59	28.55	31.43	34.62	6.07
Channel	4.30	4.65	4.30	4.00	0.65
Warranty	11.99	11.66	12.58	10.82	1.76
Service	14.48	15.97	14.38	13.30	2.67
Price	17.42	23.23	17.72	11.37	11.86

The second half of Table 6 shows that when the covariates run (the covariate run with a single covariate ExpectToPay) is used, the segments are separated much more. The spread is nearly double for Price and Performance importance (vs. the generic HB run), and 50% higher for Brand importance. The enhanced spread is not an artificial increase in differences between people, but a more true representation of their segment means because we have a more accurate representation of population means in the upper-level model. Enhanced discrimination between meaningful segments of the population not only makes segmentation analysis more robust and accurate, but it should also reduce IIA problems when conducting market simulations.

These results are based on a single data set, so we should be cautious in drawing general conclusions from this regarding the value of covariates in HB modeling. However, academics and leading practitioners have been applying covariates in HB modeling of conjoint data for over a decade now, and we at Sawtooth Software are relative novices regarding this subject. Leading academics tell us that they regularly achieve modest improvements in the predictive validity of models if the covariates are chosen wisely. But, they also make it clear that improving predictions isn't the main reason for using covariates; modeling heterogeneity and understanding segments is the primary benefit. Our results suggest that covariates slightly improve hit rates, but significantly enhance discrimination between segments for sparse CBC datasets. They also allow us to test more formally (by examining the percentage of draws of theta that are negative or positive) the differences between segments on the part-worths.

Suggestions for Practice

1. Avoid trying to use too many covariates in the model. It is generally unwise to throw lots of covariates into the model without first confirming their potential usefulness. Try to focus on just a few covariates that add relatively few columns to the covariate design matrix. If the variable can be treated as continuous, it is often helpful to do so rather than to categorize it as categorical dummies. One can save many parameters to be estimated without sacrificing much information by using a continuous variable as a covariate.

2. As with any multiple regression application, if using more than one covariate in the HB model, take care to examine whether the covariates are hindered by multicollinearity. If covariates exhibit high multicollinearity, the analyst could use factor scores in the upper-level model. For example, if the study includes batteries of behavioral items that are measured on a Likert scale, then first factor-analyzing the behavioral items and extracting their factor scores for the CBC/HB covariates would be a reasonable strategy. The extraction of the factor scores is done outside of CBC/HB in any general, statistical software package.

3. Covariates work best when the variables add *new* (exogenous) information to the CBC data, and when the covariate information is strongly predictive of respondent preferences (part-worths). For example, "amount expect to pay for a PC" was helpful for the data set described in this paper. Variables related to behavior and preferences will tend to be more valuable as covariates than descriptive information such as demographics. For example, segments developed from MaxDiff data on attribute preferences could be useful as covariates. Brand preference, past purchase choice, budget threshold, and similar variables are good candidates as valuable covariates. A segmentation solution based on a cluster analysis of dozens of variables including preferences, attitudes, and psychographics could also be valuable as a categorical covariate.

4. Covariates developed using only the choice data (for example using latent class segments developed from the same CBC data that will be used in the HB run with covariates) tend not to be helpful, and generally lead to overfitting. No new information from outside the CBC data is being used. Information already available within the CBC data is in essence being used twice.

5. The more sparse the dataset (relatively few choice tasks relative to number of parameters to estimate), the more the Bayesian shrinkage toward the pooled upper-level model. Thus, covariates are more likely to be most effective for sparse data sets. For datasets where there is a great deal of information at the individual level (relative to parameters to estimate), the Bayesian shrinkage is already relatively small, and covariates will play a lesser role in improving the quality of the estimates.

References

Allenby, G. M., Arora, N., and Ginter, J. L. (1995) “Incorporating Prior Knowledge into the Analysis of Conjoint Studies,” Journal of Marketing Research, 32 (May) 152-62.

Howell, John (2007), “Respondent Weighting in HB,” Sawtooth Software Conference Proceedings, Sequim WA, pp 365-377.

Lenk, P. J., DeSarbo, W. S., Green P. E. and Young, M. R. (1996) “Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs,” Marketing Science, 15, 173-191.

Sawtooth Software, "More Evidence CBC Is Most Popular Conjoint Method,” Sawtooth Solutions Newsletter, Summer 2003, available at www.sawtoothsoftware.com.

Sawtooth Software (2005), “CBC/HB Technical Paper,” available at www.sawtoothsoftware.com.

Sentis, Keith and Lihua Li, “One Size Fits All or Custom Tailored: Which HB Fits Better?” Sawtooth Software Conference Proceedings, Sequim, WA, pp167-177.