Sawtooth Software

RESEARCH PAPER SERIES

What Are the Optimal HB Priors Settings for CBC and MaxDiff Studies?

Bryan Orme and Walter Williams,

Sawtooth Software, Inc.

1457 E 840 N Orem, Utah
+1 801 477 4700
www.sawtoothsoftware.com

What Are the Optimal HB Priors Settings for CBC and MaxDiff Studies?

Bryan Orme and Walter Williams
Sawtooth Software, Inc.

Introduction

Hierarchical Bayes (HB) estimation has become a staple in the market research industry for obtaining individual-level estimates from choice data such as CBC (Choice-Based Conjoint) and MaxDiff (Maximum Difference Scaling, also known as Best-Worst Measurement). HB involves *priors* settings: specifically, *prior variance* and *prior degrees of freedom*. The prior variance typically is set to around 1.0 or 2.0 and expresses a prior belief about how diffuse or peaked the distribution of tastes is across respondents. The prior degrees of freedom are expressed as integer values from 2 and up[1], typically scaling with sample size, and influence how strongly the prior variance assumption is enforced within the model. These prior settings influence the degree of Bayesian smoothing (shrinkage) of individuals to the population means.

Most analysts use the defaults provided in our CBC/HB software (or in other software) and hope that's good enough. However, the default prior variance and degrees of freedom are not necessarily appropriate and optimal for each data set. In fact, our findings based on a meta analysis of around 50 commercial CBC and MaxDiff data sets show that the default priors in Sawtooth Software's CBC/HB software are not optimal for *any* of these data sets.

The optimal HB priors depend on the characteristics of the data set: the true heterogeneity across respondents (differences in tastes), the characteristics of the experimental design, the sample size, and the response error. The priors affect how much respondent part-worth utilities (betas) are shrunk[2] to the population means and covariances. Too little shrinkage and the betas overfit and tend to explain too much noise. Too much shrinkage and the betas underfit and the many benefits of capturing heterogeneity of tastes begin to be lost. With many choice tasks per individual and few parameters to estimate, the priors have relatively little effect on the posterior estimates of beta. However, many CBC and MaxDiff datasets are sparse and the priors can have a fairly large impact on the posteriors.

With HB, obtaining the highest fit possible (RLH—root likelihood, which is the geometric mean of the likelihoods across the respondent's tasks) to each respondent's data is *not* the goal. HB works to maximize a joint fit criterion: finding parameters that can explain individuals' choices well while at the same time having a high likelihood of pertaining to the population distribution. Proper priors in HB are key to obtaining the right fit to the lower level (the individual betas) of the hierarchy. Overfitting occurs

[1] A 2 means prior degrees of freedom equal to the number of beta parameters in the model + 2.
[2] Even though increasing the prior variance in HB leads to an increase in the magnitude of the part-worth utilities, it is not the same effect as simply multiplying the utilities by a constant (such as with adjusting the scale factor, also known in Sawtooth Software literature as the *exponent*).

when the betas explain too much variance at the individual level, tending to fit random noise rather than explaining true relationships that will hold for new choice tasks, new respondents, or new scenarios such as share of preference predictions for market scenarios.

Some Background on HB Priors

We at Sawtooth Software are grateful for the help and guidance of Greg Allenby and Peter Lenk regarding HB algorithms. Greg taught courses at the ART/Forum in the mid-1990s on HB and shared code with our founder Rich Johnson, leading Johnson to code (with Allenby's blessing) the first version of our CBC/HB software around 1997. Greg and Peter suggested the priors that have to this point been implemented in the CBC/HB software: Prior Variance = 2; Prior Degrees of Freedom = 5. Bryan Orme later chose a slightly different prior variance (1.0) for HB estimation for Sawtooth Software's MaxDiff and ACBC utility software components. Orme suggested this for MaxDiff and ACBC because of his belief that MaxDiff could be quite sparse and that ACBC could often be used with large attribute lists (he worried about overfitting and wanted to err on the side of caution).

Pinnell and Fridley (2001) showed at the Sawtooth Software conference that the default CBC/HB[3] did a poor job estimating quality individual-level utilities for four out of nine partial-profile CBC data sets. The aggregate logit model in these cases predicted holdout choices for *individuals* better than HB! In follow-up research, Orme (2003) examined two of these problematic partial profile data sets and showed that the individual-level hit rates could be increased (to be superior to the aggregate logit solution) by *decreasing* the prior variance and *increasing* the strength of that assumption via increasing the prior degrees of freedom. Starting in CBC/HB v3, the software allowed researchers to modify the priors.

McCullough (2009) showed at the Sawtooth Software conference that for a particularly sparse CBC data set the default priors did not do as well fitting holdouts as properly tuned priors. He used a grid-search approach to estimate HB utilities under different settings for prior variance and degrees of freedom, then used each set of utilities to predict holdout choices at the individual level. The results of his grid search are displayed in Exhibit 1 (where prior degrees of freedom are the rows and prior variance is the columns):

[3] The early CBC/HB software version they had access to did not support modifying the priors.

Exhibit 1

Holdout Hit Rate by different values of Prior Variance and Prior Degrees of Freedom

	0.1	0.25	0.5	0.75		2
5						0.618
15		0.628	0.631	0.618		
30	0.635	0.651	0.641	0.615		
60	0.635	0.631	0.615	0.628		
90	0.631					
150	0.625					
300	0.575					
375	0.545					
450	0.548					

Default CBC/HB settings

The default CBC/HB settings (prior degrees of freedom=5, prior variance=2) resulted in a raw hit rate to holdout choice tasks of 61.8%. Optimal priors (prior degrees of freedom = 30, prior variance = 0.25) resulted in a hit rate of 65.1%. One should note that the hit rate measure is notoriously stubborn to move much based on changes in model specification, so an increase in over five hit rate points (*relative* hit rate points, on a percentage basis) seems to us to be significant and valuable. McCullough recommended that researchers adjust the priors to best fit holdouts and to include more holdout choice tasks in their studies to enable themselves to do so. Of course, clients and researchers are typically loath to include very many holdouts since they are viewed as costly and taking up precious respondent time. This is especially the case as the trend is for shorter CBC and MaxDiff surveys, given the realities of today's data collection environment and the proliferation in the use of mobile devices for taking surveys.

At the 2013 SKIM/Sawtooth Software European Conference, Wirth and Moore examined five CBC data sets and performed a similar grid search to McCullough. After examining the fit to holdouts, they found that the optimal priors depended on the data set. Even though the gains in hit rates they found were not as large as McCullough's, they concluded, "…it is still crucial to investigate the best settings based on your own data. Always include holdout tasks and try out different settings." Again, not a very pleasant message for clients and researchers; for clients due to the costs and for researchers due to the time involved to conduct extensive grid searches with HB modeling!

At the 2016 Turbo Choice event, Kevin Lattery of SKIM demonstrated the results of his optimal priors search for a particular CBC data set. He showed that the hit rate was best (offering a minor improvement for his data set) with prior variance = 0.5 rather than the default 2. In the Q&A period, Kevin commented that he had done this multiple times with different CBC data sets and *never* had found prior variance = 2 to be optimal. The optimal prior variance in every case had been less than 2. Kevin

used a multi-pronged approach to assess optimal priors based not only on individual-level hit rates but on across-respondent share of choice predictions for held out blocks of choice tasks.

A Meta Study of 50 Commercial Data Sets

In the latter half of 2015, we assembled about 50 CBC and MaxDiff datasets. Some of these were from our own internal methodological studies over the years and many others were donated to us by friends and colleagues in the industry (many thanks!). Only the raw CBC or MaxDiff data (without labels) were shared with us, so we know nothing about the product categories, attribute labels, or personally identifying information of the respondents.

Walter Williams at Sawtooth Software wrote an automated program (called the *Model Explorer*) that interacts with CBC/HB software's command interpreter to perform an automated grid search across the priors[4]. Furthermore, Walter implemented a jack-knife and bootstrap[5] resampling procedure for raw hit rate validation. For each sample replicate, the approach drew a new sample (with replacement) and partitioned the data into a training (calibration) data set and a validation (holdout) data set. We typically held out just one or two of the choice tasks for the validation data set and estimated the part-worth utilities using the remaining tasks. We repeated this resampling procedure enough times such that the number of respondents x holdout tasks x resampling replicates was equal to about 30,000.

The reader may wonder how long it takes to do so many HB estimations for each sample replicate. Well, we pulled some tricks to speed up the procedure dramatically. First, we took advantage of a finding from Sentis and Li (2000) that the hit rates for holdout tasks become fairly stable and near optimal with relatively few HB iterations (such as just 1K burn-in iterations and 5K used iterations), even though convergence hasn't yet occurred. Second, Walter used multiple instances of CBC/HB's command interpreter to take advantage of multi-core processing. Nearly every modern laptop or desktop computer has multiple cores in its processor. For example, if you have an 8-core processor, 8 HB runs may be done simultaneously at almost the same speed as a single HB run. Given these two tricks, we could perform a 5x5 grid search (across five levels of prior variance and five levels of degrees of freedom) with enough resampling replications to stabilize the hit rate estimates for each treatment in about 1 to 6 hours for most any CBC or MaxDiff data set. Because the response surface characterizing the raw hit rates over the two-dimensional grid search was quite smooth, with just one peak, we fit the hit rate results using a polynomial regression equation to estimate the hit rate maximizing prior variance and degrees of freedom.

The optimal estimated priors for different CBC data sets are reported below. First, in Exhibit 2, we report the results for CBC data sets with relatively few attributes and few concepts per task.

[4] We recognize that searching for priors using the data is not formerly acceptable among statistical purists. However, practitioners often are more pragmatic and the jack-knifing holdout procedure as describe herein seems a reasonable approach for obtaining better priors to improve modeling performance under HB.

[5] Some of our colleagues have wondered if the bootstrap resampling layer in our procedure is unnecessary, so the *Model Explorer* by default does not do bootstrap resampling (though it is an option in the software interface).

Exhibit 2

Optimal Priors for

CBC Data Sets with Fewer Attributes and Few Concepts

Dataset	Attributes	MaxLevels in Any Attribute	#Params	# Tasks	# Concepts	Sample Size	Optimal Settings: PriorVar	DF	Optimal Hit Rate	Hit Rate Default	Lift over Default
1	2	10	13	10	3	1000	1.55	6	85.93%	85.93%	0.00%
2	2	7	13	12	4	2469	1.45	47	79.95%	79.79%	0.20%
3	3	5	12	9	3	988	0.95	2	80.38%	80.17%	0.26%
4	3	3	6	10	3	895	0.85	2	65.51%	65.34%	0.26%
*5	3	4	7	9	4	350	0.70	2	74.01%	73.33%	0.93%
6	3	4	7	9	4	350	0.65	2	71.51%	71.10%	0.58%
7	4	5	12	10	3	1245	0.25	2	80.15%	79.58%	0.72%
8	5	5	18	10	3	1192	0.25	102	56.79%	55.41%	2.49%
*9	5	5	13	15	3	2005	0.10	2	74.73%	74.28%	0.61%
10	5	15	27	15	3	2005	0.25	105	74.20%	73.85%	0.47%
*11	5	15	27	15	3	2005	0.30	15	74.66%	74.37%	0.39%
12	5	3	8	10	3	895	0.25	2	62.01%	61.77%	0.39%
13	5	3	8	8	3	251	1.15	2	66.87%	66.87%	0.00%
14	5	7	24	24	5	302	0.10	2	46.09%	43.38%	6.25%
15	6	3	13	9	3	184	1.20	3	68.70%	68.70%	0.00%

Notes: Data set #5 includes covariates (copy of #6). #9 linear price; #11 covariates (copies of #10)

Exhibit 2 reports hit rate results for CBC data sets from 2 to 6 attributes. The hit rate at the default priors is given in the second-to-last column, along with the hit rate at the optimal priors settings (third-to-last column). The relative lift over the default hit rate is given in the last column. The largest gain (6.25% relative lift in hit rate) occurs for data set number 14. Note how much difference there is in the optimal prior variance, with two data sets benefitting from prior variance = 0.1 and other data sets benefitting from a prior variance greater than 1.0. Also note that a few data sets don't change much at all in terms of hit rates due to modifying the priors.

Exhibit 3 reports hit rates for CBC sets continuing in order from 6 to 14 attributes:

Exhibit 3
Optimal Priors for
CBC Data Sets with More Attributes, but Few Concepts

Dataset	Attributes	MaxLevels in Any Attribute	#Params	# Tasks	# Concepts	Sample Size	Optimal Settings: PriorVar	DF	Optimal Hit Rate	Hit Rate Default	Lift over Default
16	6	5	13	17	4	78	0.10	2	67.93%	66.20%	2.61%
17	6	6	21	15	4	600	0.30	2	67.89%	67.11%	1.16%
18	6	5	13	8	4	78	0.25	2	54.68%	54.54%	0.26%
19	6	6	18	14	4	309	1.25	2	60.00%	59.95%	0.08%
20	7	3	9	12	2	264	0.20	5	73.64%	71.94%	2.36%
21	7	4	13	16	2	700	0.20	2	77.16%	76.51%	0.85%
22	7	7	28	10	5	1339	0.95	21	62.80%	62.54%	0.42%
*23	8	5	16	6	3	289	0.45	2	66.17%	65.02%	1.77%
24	8	5	16	14	3	289	0.20	10	74.86%	74.01%	1.15%
25	8	5	16	14	3	289	1.10	2	76.04%	75.67%	0.49%
*26	9	9	28	6	3	239	0.25	55	50.31%	49.03%	2.61%
27	9	9	28	12	3	239	0.25	11	64.13%	63.06%	1.70%
28	10	6	18	12	3	571	0.15	10	53.52%	52.03%	2.86%
29	10	14	41	12	3	603	0.25	97	64.15%	63.60%	0.86%
30	12	11	28	24	3	515	0.25	22	69.89%	68.83%	1.54%
*31	12	11	28	10	3	515	0.15	30	69.91%	69.25%	0.95%
32	13	5	19	12	3	420	0.15	2	74.71%	73.95%	1.03%
33	14	8	27	12	4	2143	0.15	50	70.39%	69.92%	0.67%

*Notes:
Data set #23 created by throwing away 8 tasks from data set #24
Data set #26 created by throwing away 6 tasks from data set #27
Data set #31 created by throwing away 14 tasks from data set #30

Although the optimal prior variance again varies quite a bit across data sets, you can see a general trend that as the number of attributes in a full-profile CBC increases, the optimal prior variance tends to decrease (especially for data sets in which changing the priors makes a larger difference in hit rates). Also note that from three data sets we created new datasets by throwing away half or more of the choice tasks. We wondered if making the datasets more sparse would lead to much change in the optimal priors. It did not.

Exhibit 4 presents the results for CBC data sets with a lot of concepts but few attributes (such as would be common with shelf-facing displays and FMCG studies). Note that some of the attributes (presumably brands or SKUs) have lots of levels (frequently 50+ levels).

Exhibit 4
Optimal Priors for
CBC Data Sets with Few Attributes, but Many Concepts

Dataset	Attributes	MaxLevels in Any Attribute	#Params	# Tasks	# Concepts	Sample Size	Optimal PriorVar	Settings: DF	Optimal Hit Rate	Hit Rate Default	Lift over Default
34	2	51	54	22	10	780	1.2	38	60.54%	60.49%	0.08%
35	2	51	54	22	10	788	1.2	94	62.60%	62.60%	0.00%
36	2	41	45	12	21	2107	0.9	22	15.50%	15.30%	1.31%
37	2	40	44	12	30	255	1.2	6	40.44%	40.03%	1.02%
38	2	42	46	15	42	426	1.3	2	84.51%	84.51%	0.00%
39	3	48	52	22	10	800	1.3	75	61.27%	61.27%	0.00%
40	3	50	64	15	24	601	1.2	20	41.12%	40.53%	1.46%
41	3	50	64	15	24	626	1.1	22	53.04%	52.49%	1.05%
42	4	27	36	12	29	606	1.6	19	50.63%	50.62%	0.02%
43	4	55	64	12	36	1261	1.2	22	29.21%	28.88%	1.14%
44	4	57	66	12	36	1119	1.2	2	31.32%	31.16%	0.51%

The optimal prior variance ranges from 0.9 to 1.6 for these data sets, with 1.2 as the modal value. What's intriguing about these results is that it confirms the pattern that few attributes leads to larger optimal prior variance. There is likely something about how much response error is involved in considering more attributes in CBC that demands greater Bayesian shrinkage to the upper-level model to maximize individual-level hit rates.

A few of our data sets (data sets #45-49) involved alternative-specific and partial-profile data sets (2 data sets and 3 data sets respectively). In these cases, the optimal prior variance ranged from 0.2 to 1.2. However, because it is hard to draw conclusions for alternative-specific and partial-profile choice studies from so few datasets, we don't show the results here.

Last, we looked at MaxDiff data sets (which mathematically are similar to 1-attribute CBC studies with lots of levels). Given what we've seen already in the past few tables, we might have predicted the results: the optimal prior variance ranged from 1.2 to 1.65 for these MaxDiff data sets. Interestingly enough, this points to using a slightly *higher* prior variance (higher degree of capturing differences in respondent tastes) than is currently the default in Sawtooth Software's HB settings for MaxDiff datasets. Using the software defaults may actually be leading to lost opportunities to capture useful heterogeneity for MaxDiff data sets.

Exhibit 5

Optimal Priors for MaxDiff Data Sets

Dataset	Attributes	# Items	# Tasks	# Concepts	Sample Size	Optimal Settings: PriorVar	DF	Optimal Hit Rate	Hit Rate Default	Lift over Default
50	1	18	14	4	153	1.40	2	72.84%	72.43%	0.57%
51	1	31	19	5	257	1.30	2	65.93%	65.63%	0.46%
*52	1	31	7	5	257	1.40	2	57.09%	57.02%	0.12%
*53	1	20	5	4	251	1.20	2	58.37%	58.32%	0.09%
54	1	20	15	4	251	1.20	2	68.43%	68.40%	0.04%
55	1	28	12	4	800	1.45	2	48.95%	48.95%	0.00%
56	1	36	15	4	1637	1.65	12	55.53%	55.35%	0.18%

*Notes:
Data set #52 is created by throwing away 12 tasks from #51
Data set #53 is created by throwing away 10 tasks from #54

Note as before that making the data sets considerably more sparse (by throwing away more than half the tasks per respondent) did not change the optimal priors much.

ACBC and HB Priors

The attentive reader will notice we have not included any ACBC (Adaptive CBC) datasets within this meta analysis. Even though we could export .CHO files that could be used within the *Model Explorer* tool, we question whether randomly drawn tasks should be used as holdouts for ACBC data. Recall that ACBC's choice tasks can include data from at least four different choice contexts: 1) BYO, or the choice of best level among multiple levels of each attribute, 2) Screener Tasks, or the choice of concepts versus a "None" threshold, 3) Implied Screener Tasks, or the choice of the "None" threshold over concepts that include unacceptable levels, and 4) Choice Tournament Tasks, or (typically) the choice of one concept from among three concepts. If we allow the *Model Explorer* to randomly select among all these formatted choice tasks which involve different degrees of response error (e.g. the BYO tasks involve much less response error than the Choice Tournament tasks), the analysis would tend to favor much higher prior variance settings (less Bayesian smoothing to the hyperparameters) than would be useful for predicting new CBC-looking choice scenarios.

We recently examined just one ACBC dataset involving home purchase choices on 10 attributes that also included some CBC-looking holdout choice tasks asked after the ACBC section. This data set leads to around 50 separate choice tasks (depending on the respondents' answers) formatted in the .CHO file covering the three ACBC sections (BYO, Screeners, and Choice Tournament Tasks). We found that modifying the prior variance and degrees of freedom had very little change on the holdout hit rates unless very extreme settings were selected. This is not surprising given the rich amount of data available at the individual level for computing part-worth utilities for this data set. The more information available at the individual level, the less the priors affect the posterior part-worth estimates.

Investigation of Interaction Effects

Another valuable use of the *Model Explorer* is for searching for useful interaction effects. One cannot simply add an interaction term to a CBC model and use the increase in model fit to the training data (the calibration choice tasks) as an indicator of whether the interaction effect improved the model. Adding even useless terms to a CBC/HB model usually will improve the model fit (e.g. the RLH). HB estimation does provide an easy and logical test of significance for interaction terms if the analyst simply counts the population mean draws after convergence (available in the alpha.csv ouput file) and observes if 95% or more of the draws for a given interaction term are either *all* positive or negative[6] (e.g. assuming 95% confidence level). However, such tests are often too sensitive in our opinion and point to statistically significant interaction terms that sometimes don't practically improve the model much at all in terms of ability to predict holdout choices or new market scenarios.

The *Model Explorer* uses the same jack-knifing procedure as we've previously described to hold out randomly selected choice tasks and estimate the models on the remaining tasks. First, the base model is run using main effects only. Then, each two-way interaction effect (first-order interaction) is added to the main effects specification (one interaction at a time[7]), HB estimation is performed, and the hit rate is computed. As before, the process is repeated often dozens of times to stabilize the hit rate estimates. Prior to exploring all possible first-order interaction effects, one should first use the *Model Explorer* to find the optimal priors for main effects, then apply those settings for the interactions runs.

We examined 22 different full-profile CBC data sets with the *Model Explorer* to try to find interaction effects that could improve the hit rate. Although it is quite subjective regarding how much improvement in hit rate would be considered practically significant, we decided to use the threshold of 1.5 points of absolute hit rate improvement. For example, improving the hit rate from 50.0% to 51.5% would be, in our opinion, a practical and significant improvement. Among the 22 data sets we examined, four provided a lift of at least 1.5 in hit rate. The home purchase CBC data set involved a conditional pricing table and including the interaction between home size and price led to a 4 point increase in hit rate (the largest improvement we found). Three other data sets led to improvements equal to or greater than 1.5 points. So, four out of 22 data sets might benefit from interaction terms beyond main effects.

Exhibit 6 reports hit rates for a CBC data set with five attributes where attribute 1 is a style attribute and attribute 5 is a conditional price attribute based on brands specified in attribute 2. We used a forward step-wise procedure to detect significant interaction effects and compound the improvements across more than one significant interaction effect.

[6] We're assuming effects-coded main effects and interaction terms such that the parameters are zero-centered and that the interaction terms main be interpreted orthogonally of the main effects.
[7] For example, with 6 attributes 6*5/2 = 15 two way interactions must be considered.

Exhibit 6
Hit Rates Using a Forward
Stepwise Search Procedure for Significant Interactions

	Hit Rate
Base Model (main effects):	47.4%
Main effects + 1x2	49.7%
Main effects + 1x2 + 2x5	50.7%

First, we used the *Model Explorer* with its jack-knife resampling procedure to examine main effects plus all potential first-order effects and found that adding the interaction between attributes 1x2 to the main effects model increased the hit rate the most, from 47.4% to 49.7%. Then, we ran the *Model Explorer* a second time with the interaction between attributes 1x2 included in the base model. The routine searched all additional first-order interactions added to the *Main effects + 1x2* model and found that adding the interaction between attributes 2x5 increased the hit rate the most, from 49.7% to 50.7%. We ran the procedure a third time to look for any additional interaction effects that could improve the hit rate and found no additional interactions that seemed valuable to include in the model.

Although there are occasional opportunities for including interaction effects in CBC/HB models to improve results (4 out of 22 data sets), we'd conclude that HB does a fine job of fitting the data using main effects for most CBC data sets. The interactions often detected using pooled analysis (counts and aggregate logit) are mostly due to unrecognized heterogeneity (due to correlations in preference for levels across respondents). But, for the occasional data set where interactions under HB could improve the results, the *Model Explorer* is a useful tool to find and quantify their practical benefit. An interactions search for a data set with six or seven attributes should take about 2 to 8 hours. Of course, a stepwise procedure (such as we showed directly above) to find combinations of significant interaction terms would take more manual steps, each involving 2 to 8 hours.

Opportunities for Covariates

The evidence we've presented here generally encourages us to set priors that strengthen the influence of the upper-level model in our HB estimations for CBC data. This of course implies more Bayesian smoothing toward the hyperparameters, which shrinks the differences among respondents and subgroups on the lower-level part-worth utilities—usually not the direction researchers want to move toward for segmentation analysis and targeting. This may lead to greater opportunities for leveraging useful covariates in the upper-level model to tease out more true heterogeneity in the data. In our opinion, the best priors will typically be directly related to choice, such as responses to BYO (build-your-own) questions, past purchase behavior, brand preferences, performance needs, and questions regarding budgets.

Future of Priors Settings for Sawtooth Software Tools

Based on our findings, we have begun to update the defaults in our HB implementations for choice data. We have already updated Lighthouse Studio HB estimation for CBC to use a default prior variance of 1.0. CBC/HB v5.5.4 also implements the change. We will probably integrate the *Model Explorer* within Lighthouse Studio and CBC/HB software in the future, but for now it is a separate .EXE program file.

Conclusions

We shouldn't oversell what we've done here: for most data sets the *Model Explorer* offers a method of fine-tuning to achieve slightly better results. Yet, given the amount of money typically invested in choice data and the modest predictive gains associated with often big investments in terms of time and human capital to build better and better models, achieving the gains seen here by using an automated tool to search for better HB priors (and potentially useful interaction effects) seems like a win-win proposition for researchers and clients alike.

The *Model Explorer* extension to CBC/HB software allows the researcher to perform holdout hit rate validation using the existing experimentally designed tasks (often called random tasks) rather than requiring fixed holdout tasks. We still strongly recommend holdout tasks (especially out-of-sample holdout tasks) for academic publications. Yet, the ability to leverage existing random tasks as holdouts for the purpose of finding proper priors and valuable interaction effects in commercial data sets is a boon for researchers and should make clients much happier.

If you prefer not to use the *Model Explorer* to search for optimal priors for your dataset (statistical purists would not use the data to fit the priors), you might draw some general conclusions from our meta analysis to update your prior beliefs:

- Set your Degrees of Freedom depending on the sample size:

Sample Size	D.F.
0-200	2
200-400	5
400-700	10
700-1200	30
1200-2400	50

- For Full-Profile CBC (showing 5 or fewer concepts per task):

	Prior Variance
>=10 attributes	0.2
7-9 attributes	0.3
5-6 attributes	0.5
3-4 attributes	0.8
2 attributes	1.5

- For shelf-display CBC with very few attributes but many levels of a single attribute (e.g.>=20) and showing 10 or more concepts per task, set Prior Variance = 1.2

- For MaxDiff, set Prior Variance = 1.3

References

McCullough, Paul Richard (2009), "Comparing Hierarchical Bayes and Latent Class Choice: Practical Issues for Sparse Data Sets," Sawtooth Software Conference Proceedings, Sequim, WA, pp 273-284.

Orme, Bryan (2003), "New Advances Shed Light on HB Anomalies" Sawtooth Software Research Paper available at: https://www.sawtoothsoftware.com/download/techpap/hbadvanc.pdf

Pinnell, Jon and Lisa Fridley (2001), "The Effects of Disaggregation with Partial Profile Choice Experiments," Sawtooth Software Conference Proceedings, Sequim, WA, pp 151-166.

Wirth, Ralph and Chris Moore (2013), "HB Priors Do Matter – But When and How Exactly?," SKIM/Sawtooth Software European Conference.