



# Sawtooth Software

*RESEARCH PAPER SERIES*

## One Size Fits All or Custom Tailored: Which HB Fits Better?

Keith Sentis  
and  
Lihua Li,  
Pathfinder Strategies

Originally published in the 2001  
Sawtooth Software Conference Proceedings

# ONE SIZE FITS ALL OR CUSTOM TAILORED: WHICH HB FITS BETTER?

*Keith Sentis & Lihua Li*  
*Pathfinder Strategies<sup>1</sup>*

## INTRODUCTION

As most of you know from either your own experience with Hierarchical Bayes Analysis (HB) or from reports by colleagues, this relatively new analytic method offers better solutions to a variety of marketing problems. For example, HB analyses yield equivalent predictive accuracy with shorter questionnaires when estimating conjoint part-worths (Huber, Arora & Johnson, 1998). HB also gives us estimates of individual utilities that perform well where before we would have had to settle for aggregate analyses (Allenby and Ginter, 1995; Lenk, DeSarbo, Green & Young, 1996).

HB methods achieve an “analytical alchemy” by producing information where there is very little data – the research equivalent of turning lead into gold. This is accomplished by taking advantage of recently developed analytical tools (the Gibbs sampler) and advances in computing speed to estimate a complex two-level model of individual choice behavior. In the upper level of the model, HB makes assumptions about the distributions of respondents’ vectors of part-worths. At the lower level of the model, HB assumes a logit model for each individual. The analytical alchemy results from using information from the upper level to assist with the fitting of the lower level model. If a given respondent’s choices are well estimated from his own data, the estimates of his part-worths are derived primarily from his own data in the lower level and depend very little on the population distribution in the upper level. In contrast, if the respondent’s choices are poorly estimated from his own data, then his part-worths are derived more from the distributions in the upper level and less from his individual data in the lower level. Essentially, HB “borrows” information from the entire sample to produce reasonable estimates for a given respondent, even when the number of choices made by the respondent is insufficient for individual analysis.

This process of “borrowing” information from the entire sample to assist in fitting individual level data requires considerable computational “grunt” and is a potential barrier to widespread use of HB methods. However, the rapid increase in computer speed and some of our own work (Sentis & Li, 2000) that identified economies achievable in the analysis, have made HB analysis a viable tool for the practitioner. The focus of our paper today is a particular aspect of how HB “borrows” information to fit individual level data. I mentioned a moment ago that the upper level model in HB makes some assumptions about the distribution of vectors of part-worths in the population. In the simplest case, the upper level model assumes that all respondents come from the same population distribution. More complex upper level models make further assumptions about the nature of the population. For example, the upper level model may allow for gender differences in choice behavior. Of course, these more complex upper level models require more parameters and additional computational grunt.

---

<sup>1</sup> The authors thank Rich Johnson for helpful comments.

In popular implementations of HB for estimating conjoint part-worths such as the Sawtooth HB modules, the upper level model is simple. All respondents' vectors of part-worths are assumed to be normally distributed. That is, all respondents' choices are assumed to come from a single population of choice behavior. On the face of it, this assumption runs counter to much work done in market segmentation. Indeed, the fundamental premise of market segmentation is that different segments of respondents have different requirements which are manifest as different patterns of choice behavior. Ostensibly, this demand heterogeneity enables differentiated product offerings, niche strategies and effective target marketing efforts. This view was first posited by Smith (1956) who defined market segmentation as making product decisions by studying and characterizing the diversity of wants that individuals bring to a market.

Our paper examines what happens when HB analyses are allowed to “borrow” information from more relevant subpopulations. The idea was to attempt to improve our predictions by “borrowing” information from a more appropriate segment of respondents rather than borrowing from the entire sample.

Consider this analogy. If you want to buy a new suit, there are two ways to proceed. You can shop for an off-the-rack suit and hope or assume that your particular body shape fits within the distribution of body shapes in the population. Alternatively, you can have a suit custom-tailored to your exact shape. Custom tailoring will almost always yield a better look than the “one-size-fits-all” alternative. This custom tailoring yields a better fit but is more costly.

Similarly, in our current project, we explored whether custom tailoring HB utilities within segments yields a better fit. That is, we explored whether better fitting models can be achieved by having the analysis “borrow” information from a more appropriate base than the entire sample — namely segments of the sample.

We do not have access to HB software that allows complex upper level models. Instead, we used the Sawtooth HB CBC module to explore the impact on predictive accuracy from first dividing respondents into groups with presumably similar choice patterns and then estimating the utilities separately for each group.

We compared the predictive accuracy of HB utilities derived from the entire sample to those derived from within *a priori* segments and also from within latent segments. To customize the HB utilities in this way requires more effort and therefore increases the cost of the analysis. Keeping with our sartorial analogy, our paper poses the following question:

***These custom-tailored HB analyses have a higher price tag but do they yield a nicer fit?***

## APPROACH

We took a simple-minded approach to this question. First, we computed HB utilities using the entire sample. Actually, we computed three separate sets of utilities to reduce any random jitters in the results. Then we calculated the hit rates for hold out tasks using the three sets of utilities and we averaged the three hit rates.

Next, we divided the sample into segments –either *a priori* segments or latent segments – and computed HB utilities within each segment. Then we calculated the hit rates for the same holdouts using the within-segment utilities. We computed three sets of HB utilities within each segment and averaged the hit rates as we did for the entire sample analyses. Then we compared the hit rates based on the total sample utilities with the hit rates based on the within-segment utilities.

Here is a summary of our approach:

Step 1: Compute HB utilities using entire sample

- 3 separate sets

Step 2: Calculate hit rates for hold outs using three separate sets of utilities

- average the three hit rates

Step 3: Divide sample into segments (*a priori* or latent) and compute HB utilities within each segment

- 3 separate sets in each segment

Step 4: Calculate hit rates for hold outs within each segment using three separate sets of utilities

- average the three hit rates within each segment

Step 5: Compare the mean hit rates

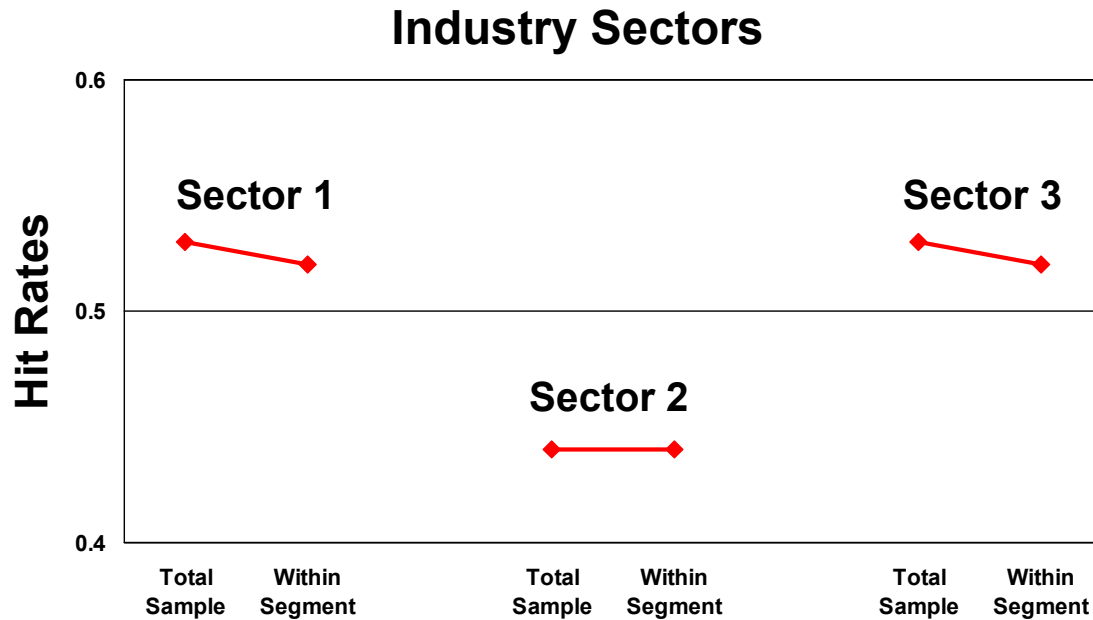
## Results

The first dataset we looked at had these characteristics:

- business to business study
- 280 respondents
- tasks = 16
  - 4 concepts plus NONE
- holdouts = 2
- attributes = 10
- partial profile design

This study focused on a range of farm enterprises that were engaged in quite different farm activities. Some of these farms produced fine Merino wool and some produced fine chardonnay grapes. The range of farm enterprises broke into three broad industry sectors and we used these industry sectors as *a priori* segments.

The results of our analyses are shown on this graph. Each of the points is the mean of the hit rates from three separate sets of utilities. It would be safe to summarize this slide as “Custom tailoring does not yield dramatically better fits.”

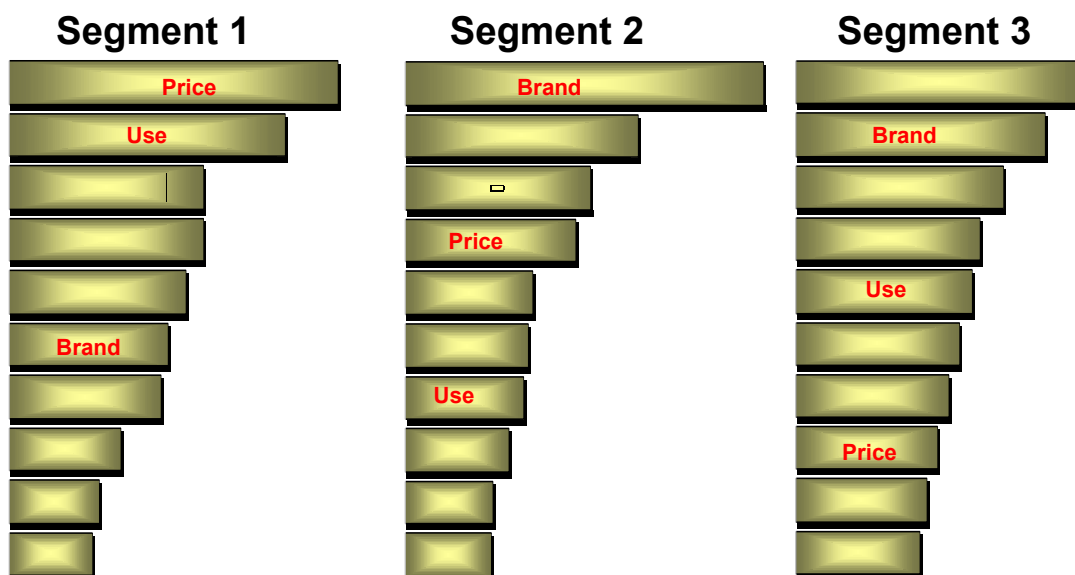


We were somewhat surprised by this and decided to explore what happens to the fit when latent segments are defined on the basis of different choice patterns. We examined three latent segments that we had identified within this same dataset.

Three segments were defined using KMEANS clustering of the HB utilities from the total sample. These three segments comprised 40%, 34% and 26% of the farming enterprises and had all of the characteristics that we like to see when we conduct segmentation projects. They looked different, they made sense, they were statistically different and most importantly, the client gave us a big head nod.

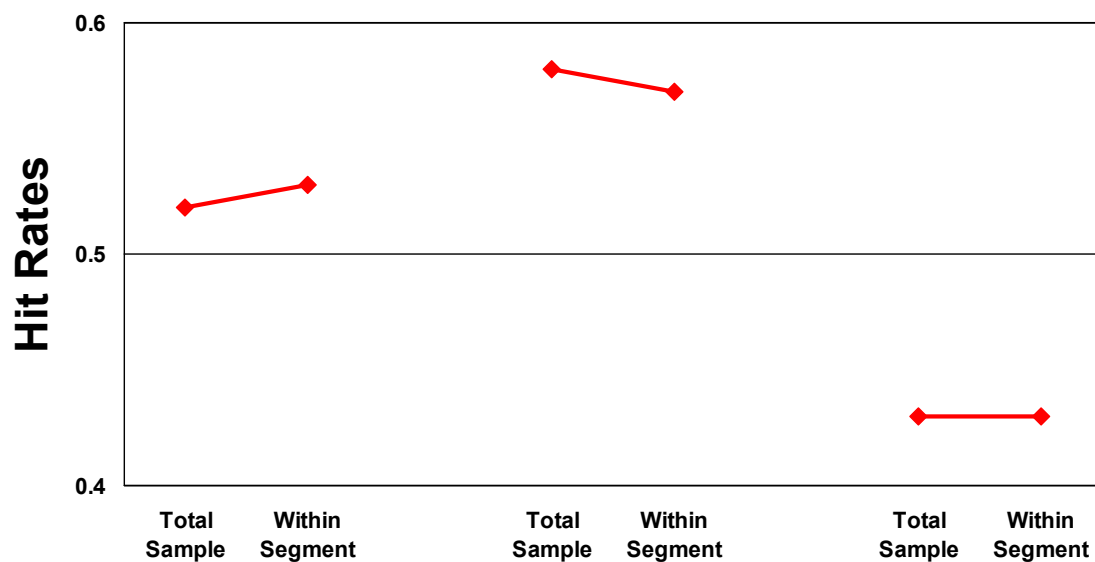
This graph shows the relative importance of the attributes for each of the segments. We have highlighted three of the attributes to demonstrate the differences in the pattern across the segments. These differences meet the usual significance thresholds for both univariate and multivariate tests.

## Feature Importance by Segment



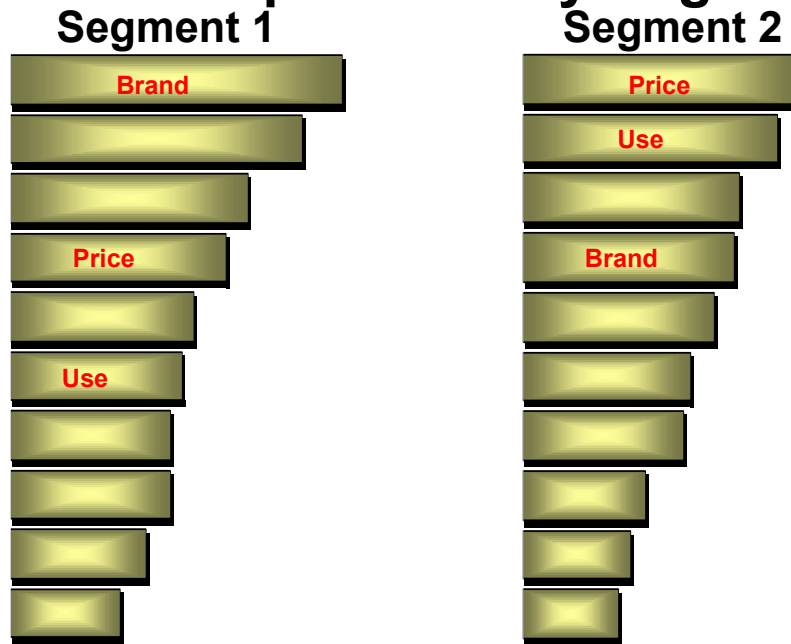
The next graph shows how much better the fit is when we customize the HB runs to borrow information from only the most relevant segment. Once again, custom-tailoring the HB utilities do not yield better fits.

## Clustering Segments – HB Utilities

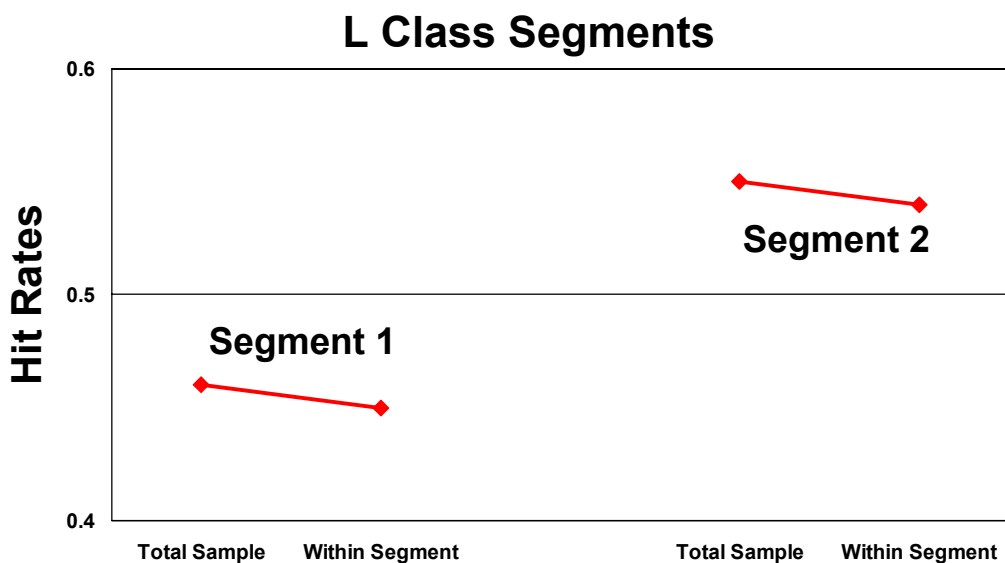


We thought that perhaps an alternative segmentation method would yield more expected results. So we ran a Latent Class segmentation using the Sawtooth LClass module to define two segments that comprised 53% and 47% of the sample. These segments are similar to the ones we found using the KMEANS method and they do exhibit a different pattern of attribute importance scores.

## Feature Importance by Segment



The results are shown here. Again, the within-segment hit rates were not any better than those from the total sample.

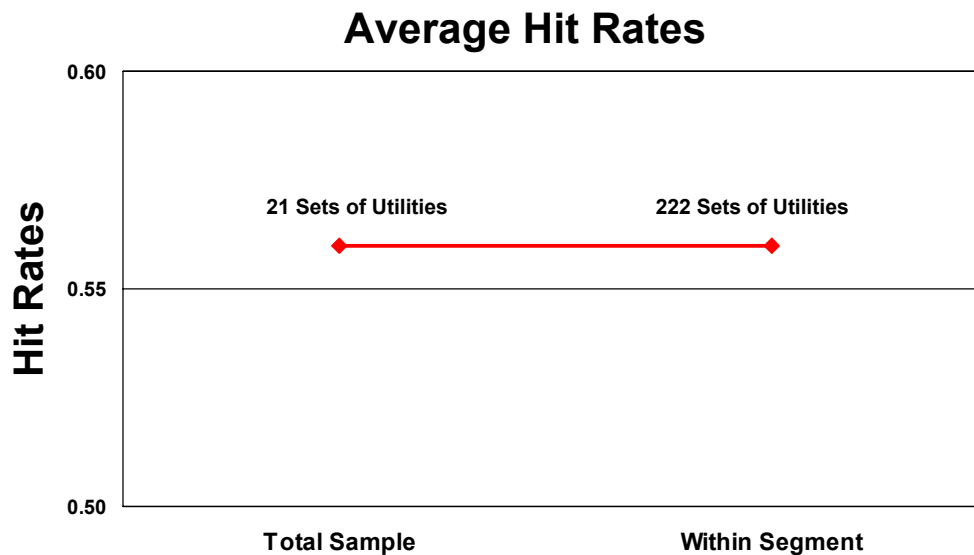


Undeterred by these unexpected results, we continued using this approach to examine six other datasets. These additional datasets were from business to business studies as well as FMCG studies. The sample sizes ranged from 320 to 800, the number of tasks ranges from 11 to 20, the number of attributes ranged from 4 to 7 with both full profile and partial profile designs. On these datasets, we examined *a priori* segments as well as latent segments derived using KMEANS and L Class methods. Across the six datasets, we examined latent segment solutions with between two and seven segments.

In some instances, there were slight improvements in the within-segment hit rates and in some instances the obverse result obtained. The graph below shows the results across the seven datasets. On the left is the mean hit rate from the 21 sets of HB utilities based on the total samples in our seven datasets. On the right is the mean hit rate from the 222 sets of HB utilities that were customised for the various segments. Even blind Freddy can see that the null hypothesis does not get much nuller than this.

This graph illustrates that the effort and expense of custom-tailoring 222 sets of utilities yields a fit that is no better than the 21 “off-the-rack” utilities. The finding across the seven datasets can be stated quite simply:

- there is no consistent improvement in predictive accuracy when going to the trouble to compute HB utilities within segments



So after all of this computation, Lihua and I were faced with a good news – bad news scenario. The good news is that the time and effort associated with customising HB to produce within-segment utilities does not appear to yield anything worthwhile. Therefore, we can run only the total sample analyses and then head for the surf with our sense of client commitment fully intact.

The bad news is that our worldview on market segments had been severely challenged. In discussing our findings with colleagues, we encountered a continuum of reaction. This continuum was anchored at one end by responses like “that’s so implausible you must have an error in your calculations”. At the other end of the spectrum, we heard reactions like “just what I expected, segmentation is actually like slicing a watermelon” or “social science data is usually one ‘big smear’ that we cut up in ways that suit our needs”.

Returning to our analogy about buying a suit for a moment, suppose we were to attempt to segment the buyers of suits using a few key measurements like length of sleeve, length of inseam, waist size and so forth. In this hypothetical segmentation exercise, we would expect to identify at least two segments of suit buyers. One segment would cluster around a centroid of measurements that is known in the trade as “42 Long”. The members of this segment are more than 6 feet tall and reasonably slim. Another segment likely to emerge from our segmentation is known as “38 Short”. Members of this segment tend to be vertically challenged but horizontally robust. Despite the fact that members of the 42 Long segment look very different from the members of the 38 Short segment, they all buy their suits off the rack from a common distribution of sleeve lengths, inseams and waist measurements.

In examining the literature more broadly, we came across other findings that are similar to ours. For example, Allenby, Arora and Ginter (1998) examined three quite different datasets looking for homogenous segments. They did not find convincing evidence of homogeneity of demand:

- “For all parameter estimates in the three datasets, the extent of within-component heterogeneity is typically estimated to be larger than the extent of across-component heterogeneity, resulting in distributions of heterogeneity for which well defined and separated modes do not exist. In other words, across the three data sets investigated by us, a discrete approximation did not appear to characterize the market place completely or accurately.”

In the aftermath of this project, Lihua and I have come to revise our worldview on market segments by embracing the “watermelon theory”. And as is often the case when one’s fundamentals are challenged, our revisionist view of market segments is a more comfortable one. So while we set out to find nicer fitting HB utilities, we ended up with a better fitting view of market segments.

## REFERENCES

- Allenby, G. M.; Arora, N; and Ginter J. L (1998) "On the heterogeneity of demand." *Journal of Marketing Research* 35, 384-389.
- Allenby, G. M. and Ginter J. L. (1995) "Using Extremes to Design Products and Segment Markets." *Journal of Marketing Research*, 32, 392-403.
- Huber, J., Arora, N. and Johnson, R. (1998) "Capturing Heterogeneity in Consumer Choices." *ART Forum*, American Marketing Association.
- 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.
- Sentis, K. and Li, L. (2000) "HB Plugging and Chugging: How Much Is Enough." *Sawtooth Software Conference Proceedings*, Sawtooth Software, Sequim.
- Smith, W. (1956) "Product Differentiation and Market Segmentation as Alternative Marketing Strategies." *Journal of Marketing*, 21, 3-8.