

DESIGNING INFORMATION-SERVICE PRODUCTS: A HIERARCHICAL BAYESIAN APPROACH

Nan-Ting Chou, University of Louisville, Louisville, KY
David Steenhard, LexisNexis, Dayton, OH

ABSTRACT

Most methods used to analyze choice-based conjoint data combine data for all participants. One of the weaknesses in analyzing data this way is that it could obscure important individual aspects of the data. Hierarchical Bayes Estimation is one analysis of estimating individual part-worths (how each individual values various attributes of a product). This method can reasonably estimate individual part-worths even with relatively limited information from each respondent. This paper introduces the Hierarchical Bayes Estimation, codes this algorithm in SAS IML®, and conducts a choice-based conjoint analysis. Using a proprietary data from a marketing survey, we find customers value various attributes of an “information service” product differently. The results of the analysis help the firm design optimal packaging and pricing strategies.

INTRODUCTION

Disaggregate or individual discrete choice modeling is fast becoming a favorite research tool among market research professionals due to the technique's ability to answer a wide range of marketing questions.) In recent years, the hierarchical Bayes (HB) choice model has generated widespread interest and acceptance in marketing research (see Wedel et al. 1999 for a review) because of its ability to provide individual-level estimates. In a discrete choice analysis, the participants/consumers are asked to choose among two or more hypothetical products which are described by a list of attributes. This allows the respondent to easily compare among alternative products. The participant chooses the product that maximizes his/her utility (value) (Baltas and Doyle, 2001). The choice depends on product attributes and consumer preferences. The standard discrete choice model typically assumes that all consumers have identical preferences since one set of coefficients are estimated for all consumers in the sample. In other words, the standard model ignores possible interpersonal differences in consumer's evaluation of product attributes. The explicit treatment of individual part-worths is important not only because this valuable information helps in designing and marketing product but also because untreated consumer heterogeneity can potentially compromise the model's accuracy (Hsiao, 1986). HB analysis explicitly accounts for the differences in consumers' preferences by estimating individual partworths.

Landmark articles by Allenby and Ginter (1995) and Lenk, DeSarbo, Green, and Young (1996) describe the estimation of individual part-worths using HB models. This approach significantly enhances marketers' abilities to understand the inter-personal differences in preferences, since it could estimate reasonable individual part-worths even with relatively little amount of data from each respondent. However, this method is computationally intensive, and usually requires many thousands of iterations before it converges.

Our paper codes a Hierarchical Bayes discrete choice estimation in SAS IML® and uses this HB estimation to analyze a proprietary survey data. The results can be used in designing and marketing an information-service product.

The paper proceeds as follows. Section 2 explains Hierarchical Bayes model and presents a portion of SAS codes of the HB discrete choice model used in this paper. Section 3 describes the data. Results of HB discrete choice model are presented in section 4. Conclusions are provided in section 5.

HIERARCHICAL BAYES MODEL

HB analysis significantly enhances marketers' abilities to understand the heterogeneous attitudes and behaviors of customers thus refine more effective market segmentations. Aggregate and disaggregate models differ in that aggregate models provide one set of parameter estimates characterizing the behavior of a representative, or average respondent in the sample, whereas disaggregate models provide parameter estimates for each respondent in the sample. HB model is a disaggregate model.

The HB model used here is called “hierarchical” because it has two levels.

- At the upper level, we assume that individuals' part-worths are described by a multivariate normal

distribution with the following notation:

$$\beta_i \sim N(\alpha, D)$$

where:

β_i = a vector of part-worths for the i th individual.

α = a vector of means of the distribution of individuals' part-worths.

D = a matrix of variances and covariances of the distribution of part-worths across individuals.

- At the lower level we assume that, given an individual's part-worths, his/her probabilities of choosing particular alternatives are governed by a multinomial logit model. The probability of the i th individual choosing the k th alternative in a particular task is:

$$p_k = \exp(x_k' \beta_i) / \sum_j \exp(x_j' \beta_i) \quad (1)$$

where:

p_k = the probability of an individual choosing the k th concept in a particular choice task.

x_j = a vector of values describing the j th alternative in that choice task.

The parameters to be estimated are the vectors β_i part-worths for each individual, the vector α means of the distribution of part-worths, and the matrix D the variances and covariances of that distribution.

ESTIMATION OF THE PARAMETERS

The parameters β , α , and D are estimated by an iterative process. This process is quite robust, and its results do not depend on starting values. However; to make the process converge as quickly as possible, one should start with estimates of the parameters that are reasonably close to final values.

- Initial estimates of the β_i were set equal to the parameters of the aggregate multinomial logit model.
 - Initial estimates of α is the average of the β_i .
 - Our initial estimates of D consists of variance and covariances of the aggregate multinomial logit model.
- Given these initial values, each iteration consists of the following three steps.
- Using the estimates of the β_i and D , generate a new estimate of α , assuming that α is distributed normally with mean equal to the average.
 - Using the estimates of the α and β_i , generate a new estimate of D , from the inverse Wishart distribution.
 - Using the estimates of the α and D , generate a new estimate of β_i from a procedure known as "Metropolis Hasting Algorithm" Which will be discuss in detail in the next section.

For each of these three steps we re-estimate one set of the parameters (α , β_i or D) based on current values of the other two sets. This technique is know as "Gibbs sampling", and converges to the correct distribution for each of the three sets of parameters. Another name for this procedure is a "Monte Carlo Markov Chain", because the fact that the estimates in each iteration are determined from those of the previous iteration by a constant set of transition rules.

This process is carried out for a large number of iterations. The first few thousand are used to achieve convergence, with successive iterations fitting the data better and better. These iterations are called "burn-in" or "transitory" iterations. After the transitory iterations are completed we start to save the estimates of the β_i , α , and D for each iteration. To get a point estimates of the part-worths for each respondent, we take the average of the β_i from these iterations.

METROPOLIS HASTINGS ALGORITHM

The Metropolis Hasting Algorithm (Chib and Greenberg, 1995) is used to draw each new set of betas for each individual. We use the symbol β_{OLD} to indicate the previous estimate of β_i . We then generate a trial value for the new estimate of β_i , which we call β_{NEW} , and then test whether it represents an improvement. If so we accept it as our next estimate, if not we accept or reject it with probability depending on how much worse it is than the previous estimate.

To get β_{NEW} we draw a random vector d of "differences" from a distribution with mean of zero and covariance matrix proportional to D , and let $\beta_{NEW} = \beta_{OLD} + d$

We then calculate the probability of the data given each set of part-worths, β_{OLD} and β_{NEW} , using the formula for the multinomial logit model (1). That is done by calculating the probability of each choice that individual made, using the multinomial logit formula for p_k and then multiplying all these probabilities together and call these resulting values p_{OLD} and p_{NEW} , respectively.

Next we calculate the relative density of the distribution of the betas corresponding to β_{OLD} and β_{NEW} , given current estimates of parameters α and D (these serve as priors in the Bayesian updating). Call these values d_{OLD} and d_{NEW} . The relative density of the distribution at the location of a point β is given by the following formula:

$$\text{Relative Density} = \exp \left[-\frac{1}{2} (\beta - \alpha)' D^{-1} (\beta - \alpha) \right]$$

Finally calculate the ratio:

$$r = p_{NEW} d_{NEW} / p_{OLD} d_{OLD}$$

From Bayesian updating the posterior probabilities are proportional to the product of the likelihood times the priors. The probabilities p_{NEW} and p_{OLD} are the likelihood's of the data given the parameter estimates β_{OLD} and β_{NEW} . The densities d_{OLD} and d_{NEW} are proportional to the probabilities of drawing those values of β_{OLD} and β_{NEW} , respectively, from the distribution of part-worths, and play the role of priors. Therefore, r is the ratio of posterior probabilities of β_{OLD} and β_{NEW} .

If r is greater than unity, the new estimate has a higher posterior probability than the previous one, and we accept β_{NEW} . If r is less than unity we accept β_{NEW} with probability equal to r .

Two influences are at work in deciding whether to accept the new estimate of beta. First, if a respondent's choices fit well, their estimated β_i depends mostly on his own data and is influenced less by the population distribution (relative density). But if their choices fit poorly then their estimated β_i depends more on the population distribution and is influenced less by their data. In this way HB makes use of every respondent's data in producing estimates for each individual. This sharing of information is what gives HB the ability to produce reasonable estimates for each respondent even when there may be inadequate information for each individual.

The following SAS IML® code performs the Metropolis Hasting Algorithm.

```
start beta(nind,subj,set,x,beta,alpha,d, jd,umean,arate);
```

```
/* Matrix ucov is a covariance matrix that is proportional to D. The proportionality factor is jd "jumping
distribution". It determines the size of the random jump from the old estimates of the individual betas to the
new estimates*/
```

```
ucov=jd*d;
accept=0;
decline=0;
invd=inv(D);
seed=int(ranuni(0)*10000);
```

```
* Break all the information by individual respondents;
```

```
do i=1 to nind;
  xi=x[loc(subj=i),];
  yi=y[loc(subj=i),];
  seti=set[loc(subj=i),];
```

```
/* To get the new estimate for beta draw a random vector delb from a multivariate normal distribution with
mean of zero and covariance matrix proportional to D, and let the new betan = beta + delb, where
beta is the previous estimate of betan */
```

```
call vnormal(delb,umean,ucov,1,seed);
delb=delb;
seed=seed+i;
```

```

    betao=beta[,i];
    betan=betao+delb;

* Find the exponential of the utilities for the new and old estimates;

    eutilo=exp(xi*betao);
    eutiln=exp(xi*betan);

/* Find the probability of each choice and for each choice task that the individual made then multiply all the
probabilities together */

    maxseti=max(seti);
    po=1;
    pn=1;

    do j=1 to maxseti;
        yset=yi[loc(seti=j),];
        tutilo=eutilo[loc(seti=j),];
        tutiln=eutiln[loc(seti=j),];

/* Find the sum of all of the exponential utilities for each choice task */

        sutilo=sum(tutilo);
        sutiln=sum(tutiln);

/* Find the probability of each choice that the individual made then calculate the product of the probabilities
for each individual */

        ptempn= tutiln[loc(yset=1),]/sutiln;
        ptempo= tutilo[loc(yset=1),]/sutilo;

/* Calculate the product of the probabilities for each individual*/

        po=po*ptempo;
        pn=pn*ptempn;
    end;

/* Calculate the relative density of the distribution of the betas corresponding to betao and betan given
current estimates of parameters alpha and D. Call these values etmpo and etmpn. Finally calculate the
ratio (pn*etmpn)/(po*etmpo) */

    diffo=betao-alpha;
    diffn=betan-alpha;
    tmpo=diffo`*invd*diffo;
    tmpn=diffn`*invd*diffn;
    etmpo=exp(-0.5*tmpo);
    etmpn=exp(-0.5*tmpn);

/* Select either betao or betan for the new estimate of beta based on the ratio. If this ratio is greater than or
equal to unity accept betan as the new estimate for beta for that individual. If the ratio less than unity, then
use a random process to decide whether to accept betan or retain betao. Accept betan with probability equal
to the ratio. */

    ratio=(pn*etmpn)/(po*etmpo);
    minr=min(ratio,1);
    rand=uniform(0);

/* Determine if you want to save the new estimate of beta or not */

    if rand <= minr then do;
        beta[,i]=betan;
        accept=accept+1;
    end;
    else do;
        beta[,i]=betao;

```

```

        decline=decline + 1;
    end;
end;

/* Find the acceptance rate*/

arate=accept/(accept+decline);

free xi seti delb eutiln eutiln maxseti po pn tutilo tutiln sutilo sutiln ptempn
ptempo invd diffo diffn tmpn tmpn etmpo etmpn ratio minr rand accept decline;
finish;

```

DATA

The data utilized in this study are obtained from a firm wanting to offer web-based information products to potential and existing customers. This web-based information product is a summary report that provides financial and non-financial information of a specific subject which is compiled from various sources. As part of a larger marketing study, a discrete-choice survey was conducted over the internet.

The survey was designed to present potential customers with different trade-offs of attributes of an online information product. Some of the product attributes are: features and functions, content, price plans, and brands. Each attribute had at least two quality- levels. The attributes of the online products and the levels of each attribute are presented in Table 1. The brand name and some attribute levels are disguised to protect the proprietary interests of the cooperating firm.

The survey consisted of a conjoint choice task. Each respondent/customer was asked to evaluate fifteen independent "buying scenario" or choice sets. Each "buying scenario" contains a set of three product packages that were described by the ten attributes. The respondent was then asked to indicate which one they would most likely buy. The respondent could pick one of those three packages or none by choosing "none of these packages appeal to me". The attribute levels for each alternative package were systematically varied in the choice sets. Combinations of attribute levels that are not feasible were omitted. For example, if attribute "price plan" is fee per report then the "subscription pricing" attribute levels were omitted for this product choice.

There are 530 respondents completed the survey. These respondents have either used a web-based information product within the past six months, or would find such a product of value. In order to qualify for this study, respondents had to be at least somewhat involved in making purchasing decision of web-based information product for their respective organizations. The participants were randomly assigned to one of the six groups. Each group was presented with 15 sets of "buying scenario". Two choice sets were used as "holdout" group to test the goodness of fit of HB estimation.

Two scenarios were simulated for this analysis. The first simulation compares "Basic Brand A" with its three major competitors: Brands B, C, and D. This simulation was designed to represent the online information product offering available in the current marketplace. The second simulation compares "Deluxe Brand A" and its three major competitors. "Deluxe Brand A" includes upgraded downloading capabilities; increased coverage of geographic area; and higher subscription prices. The competitors product offerings remained unchanged for both simulations.

RESULTS

Using the Hierarchical Bayes Estimation in SAS IML ® to analyze this proprietor survey data, we are able to derive the following results:

- Estimates of the preference shares for the different brands in the current market. This provides information about customers' preferences of the existing products.
- Estimates of the preference shares for the "Deluxe Brand A" product. It allows us to predict the change of customers' preference towards Brand A if the upgrade version were offered.
- Simulate different "What If" scenarios by varying the different attributes and measuring the preference share changes of different brands.
- Estimates of individual and overall average (across all individuals) utility values for each level of attribute. This allows us to predict the potential improvement of preference share for each attribute of the existing product.
- Estimates of average (across all individuals) importance of each attribute. This provides information about which attributes are more important in the product's composition in general.

Table 1
Description of Product Attributes

1. Price Plan:	- Subscription Price: Flat rate by # of users - Transaction Price: Transactional per report - Report Price: Flat rate by # of committed reports
2. Brands:	- Brand A - Brand B - Brand C - Brand D
3. Screening Capabilities:	- Sophisticated - Limited
4. Supplemental Content:	- Limited - Moderate - Extensive - Extensive plus legal, patent & trademark
5. Download/Editing Capabilities:	- Print only, no download - Download, no edit - Download and edit
6. Company Coverage:	- U.S. Public companies only - All U.S. (public & private companies) - All U.S. plus U.K. and Europe - All U.S. plus U.K., Europe, Asia, & Latin America
7. Timeliness:	- Real time - Recent time
8. Linking to Full-Text Source Documents:	- Cannot link - Can link
9. Content Selection:	- One part at a time - Parts/entire report, in one step
10. Pricing	- Subscription Price: - Subscription pricing 1 - Subscription pricing 2 - Subscription pricing 3 - Subscription pricing 4 - Transaction Price: - Transaction pricing 1 - Transaction pricing 2 - Transaction pricing 3 - Report Price: - Report pricing 1 - Report pricing 2 - Report pricing 3 - Report pricing 4

Table 2 displays the average importance of each attribute. The importance is defined as its weight, or maximum influence it can have on a product choice, given the range of attribute levels defined in the study. The HB discrete model provides unbiased estimates of attribute importance because the technique takes into account of individual utility information. From the importance table, the price plan, brand, and company coverage are the top three attributes in influencing the purchase of this specific information-base product.

Table 2
Importance of Each Attribute

Attribute	Average Importance
Price Plan	12.6%
Brand	13.1%
Screening Capabilities	4.6%
Supplemental Content	9.0%
Download/Editing Capabilities	9.7%
Company Coverage	14.6%
Timeliness	4.1%
Linking to Full-Text Source Documents	5.0%
Content Selection	3.1%
Subscription Pricing	9.7%
Transaction Pricing	5.8%
Report Pricing	8.7%

100.0%

To test the goodness of fit of HB discrete choice model, two out of the fifteen sets of "buying scenario" were set aside as holdout sample in our analysis. The HB approach provided the part worths of each customer's evaluation of attributes. We took account of this information in predicting the likely choice decision in the holdout sample. The HB discrete choice model forecasts are compared to the forecast generated by standard discrete choice model. Table 3 and Table 4 compare the forecasting results of these two approaches. In general, the predictions generated by HB model were closer to the actual selections.

Table 3
Percentage of Customers Choosing the Product
Holdout Choice Set #1

Choices	Actual Selection	HB Estimates	Standard Model Estimates
Product 1	45.1%	47.2%	44.6%
Product 2	25.5%	23.7%	22.2%
Product 3	15.5%	15.4%	18.5%
None of the above	13.9%	13.7%	14.6%

Table 4
Percentage of Customers Choosing the Product
Holdout Choice Set #2

Choices	Actual Selection	HB Estimates	Standard Model Estimates
Product 1	25.8%	27.7%	27.8%
Product 2	33.2%	29.9%	28.4%
Product 3	33.2%	33.9%	34.9%
None of the above	7.8%	8.5%	8.9%

CONCLUSIONS

Discrete choice models are widely used by researchers and marketing practitioners who need to understand how consumers choose among multi-attribute alternatives. These models help predict the market shares as product attributes change. However, a drawback of the standard discrete choice models is that it only produces aggregate level statistics thus it provides the choice behavior of a representative, or an average, consumer. It ignores the interpersonal differences in consumers' evaluation of products.

Hierarchical Bayes(HB) approach facilitate the in-depth study of interpersonal differences by estimating reasonable individual-level part worths even when the individual information is limited. By explicitly recognizing the individual differences, the HB approach enables researchers and managers to gain a better understanding of complex consumer decision-making and to conduct insightful simulation of potential impacts of changes in product attributes.

This paper introduces the Hierarchical Bayes Estimation, codes a HB discrete choice procedure in SAS IML®, and uses this SAS program to analyze a proprietary survey data of an online information product. We show the HB estimates outperform standard discrete choice model in predicting customers' choices in a holdout sample. By assessing the effects of attributes at individual level, our results from the HB estimation help the firm in custom-designing products, forming market segments, targeting market actions, and improving pricing strategy.

REFERENCES

- Allenby, G.M. and Ginter, J.L. (1995). "Using Extremes to Design Products and Segment Markets," *Journal of Marketing Research*, 32, (November) 392-403.
- Baltas, G., and DOYLE, P. "Random Utility Models in Marketing Research: A Survey." *Journal of Business Research*, 51, 2 (2001): 115-25.
- Chib, S. and Greenberg, E. (1995) "Understanding the Metropolis-Hasting Algorithm," *American Statistician*, 49, (November) 327-335.
- Hsiao, C. *Analysis of Panel Data*. Cambridge, UK: Cambridge University Press, 1986.
- Lenk, P.J., DeSarbo, W.S., Green P.E. and Young, M.R. (1996) "Hierarchical Bayes Conjoint Analysis: Recovery of Part-Worth Heterogeneity fro Reduced Experimental Designs," *Marketing Science*, 15, 173-191.
- SAS Institute Inc. (1999). *SAS/IML User's Guide*, Version 8, Cary NC: SAS Institute Inc.
- Sawtooth Software (2000). *The CBC/HB Module, For Hierarchical Bayes Estimation*, Version 1.5, Sequim, WA: Sawtooth Software.
- Sawtooth Software (1999). *The Client Conjoint Simulator*, version 1.0, Sequim, WA: Sawtooth Software.
- Wedel, M., Arora, N., Bemmaor, A., Chiang, J., Elrod, T., Johnson, R., Lenk, P., Neslin, S., and Poulsen, C. S. (1999) "Discrete and Continuous Representations of Unobserved Heterogeneity in Choice Modeling," *Marketing Letters*, 10(3), 219-232.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Nan-Ting Chou
Economics Department, College of Business
University of Louisville
Louisville, KY 40292
502-852-4840
502-852-7672
ntchou01@louisville.edu

David Steenhard
Email: david.steenhard@lexisnexis.com