NORTHWESTERN UNIVERSITY
SCHOOL OF PROFESSIONAL STUDIES

Discrete Choice Experiment

Yining Feng

February 13, 2021

1. Introduction & Problem Statement

Star Technologies Company (STC) is planning on entering the computer tablet market with a new product that is most preferred by their targeted user segment. A choice-based conjoint (CBC) analysis is thus devised to estimate customer preference shares over choice scenarios for attributes of computer tablets. There are two different choice scenarios for preference share modeling which are based on a wider set of survey results administered by Neverending Marketing Insights. The survey results are then used to fit two Hierarchical Bayes (HB) Multinomial Logit (MNL) models for preference shares. The purpose of these models is to allow for the measurement of price sensitivity of respondent choices which are brand specific and also for examination of possible effects of prior STC product ownership and respondent gender on the attributes' contribution to customer preferences.

2. Research Design and Modeling Method

2.1 Data Description and Processing

Customer survey data are collected from 424 respondents. The survey is designed and administered by STC and Neverending Marketing Insights and includes questions focused on customer interest, demographics and choice sets. The full set of survey questions includes:

- STC Owner: Previously own STC remote or audio accessory
- 36 choice set questions
- Purchase interest questions:
  - Purchasing a new tablet?
  - Purchasing a new smart phone?
  - Using cloud storage for storing your personal digital content?
  - Taking an online course to improve relevant skills?
- Gender of respondent
- Age of respondent
- Income Level and Residential State

The 36 choice set questions are built from the following five attributes:

- Brand: 4 levels: STC, Somesong, Pear, Gaggle (level codes: 0,1,2,3)

- Price: 3 levels: \$199, \$299, \$399 (level codes: 0,1,2)
- Screen: 3 levels: 5-inch, 7-inch, 10-inch (level codes: 0,1,2)
- RAM: 3 levels: 8 Gb, 16 Gb, 32 Gb (Gb = “gigabytes”) (level codes: 0,1,2)
- Processor: 3 levels: 1.5 GHz, 2 GHz, 2.5 GHz (GHz = “gigahertz”) (level codes: 0,1,2)

Each of 36 choice set questions has three alternatives with each alternative representing a specific combination of the above attribute levels (108 possibilities). Each customer is asked to rank the three alternatives presented for each choice set with a score from one to three. This dataset of 108 possible combinations of alternatives for the 36 choice set questions will be employed to form the predictor variable set for model estimation. An additional dataset containing 6 alternatives based on the choice sets of two additional choice scenarios will be used for subsequent preference share modeling. A number of data pre-process routines are conducted to manipulate the data into a format which can easily allow fitting of the MNL regression models. These routines and their outputs are explained in their order of execution below:

1) Effects matrix: Generate a matrix which includes an effect coded version of the 108 possible combinations of attributes from the original dataset.
2) Brands matrix: Extract the coded levels associated with brand from the ‘effects matrix’.
3) Price vector: Create a price vector of the difference between the price associated with each attribute combination and the mean price of all possible attribute combinations.
4) Brand by price matrix: Multiply the ‘brands matrix’ by the ‘price vector’ in order to create a matrix of interactions between brand and price.
5) X-matrix: Merge the ‘effects matrix’ and the ‘brands by price matrix’ to create a ‘X-matrix’ of predictor variables for the modelling phase.
6) y-data: Extract each respondents selection to the 36 choice set questions from the original dataset.
7) List of y by X-matrix: Create a list of combinations of each respondent’s selection to the 36 choice questions (y-matrix) and the matrix of predictor variables (X-matrix).

2.2 Methodology

In this study, two HB MNL models are implemented for analyzing customer’s preference shares using the R function rhierMnlDP from the Bayesm R package. One model includes a covariate indicating previous ownership of a STC product and a covariate accounting for each respondent’s gender, while the other model excludes these covariates. Both models are estimated using Markov Chain Monte Carlo (MCMC) simulation, for which 30,000 iterations will be included and every 10th sample will be retained. The first HB MNL model (Model 1) includes the choice set responses for all 424 respondents (y-data) along with predictor variables (X-matrix). The functional form of the MNL likelihood model used for Model 1 can be shown as:

$$y_{ijk} = X_{ijk} \cdot \beta_i + e_{ijk} \text{ where } \beta_i \sim MVN\,(\bar{\beta}, V_\beta)$$

There are a number of statistical methods available to assess the overall model fit, including Mean Squared Error (MSE), Mean Absolute Error (MAE), likelihood measures (such as LL, RLH, and pseudo R-squared), Bayes factor, and posterior predictive checks. To calculate these statistical measures, one would need to first find the posterior means for each beta of each respondent. Next, a matrix would be derived from multiplying the original X-matrix by the posterior means, which would then be exponentiated and divided by the row sums of exponentiated betas. For this assessment however, a simple 'sniff test' is elected instead to be conducted, by first ensuring enough iterations are included for proper 'burn-in' of coefficients, and then the beta means are assessed directly to determine whether each seem logical. The betadraws recorded over the MCMC simulation for Model 1 are shown in Figure 1. Likewise, the Log Likelihood values recorded over the MCMC simulation for Model 1 are shown in Figure 2. Table 1 shows the beta means for each attribute. It is evident that the beta and likelihood value tend to level out at around 2,000 iterations, suggesting that more than enough iterations are included to achieve a proper burn-in.

	Attribute	Beta Mean
screen_1	7-inch Screen	-0.181
screen_2	10-inch Screen	0.486
RAM_1	16Gb RAM	0.080
RAM_2	32Gb RAM	0.616
processor_1	2GHz Processor	0.939
processor_2	2.5GHz Processor	1.262
price_1	$299	0.311
price_2	$399	-2.907
brand_1	Somesong Brand	-0.111
brand_2	Pear Brand	0.046
brand_3	Gaggle Brand	-0.393
brand_1_by_Price	Somesong Brand by Price	0.126
brand_2_by_Price	Pear Brand by Price	0.047
brand_3_by_Price	Gaggle Brand by Price	0.008

Table 1. *Beta means for each tablet attribute generated by Model 1.*

Since an effects coded version of the attribute data is generated as part of the pre-processing routine, beta mean values for the reference level of each attribute (level code: 0) will not be directly estimated. Instead, the beta mean for the reference level will be estimated by solving the sum of beta means for all levels of that attribute for zero. In this case, solving for zero is appropriate as our specification lacks an intercept term. An example would be for the RAM attribute, where a beta mean is estimated as 0.080 and 0.616 for the second and third levels of this attribute (level code 1: 16 Gb, level code 2: 32 Gb). In this case, the beta estimate for the first level of the RAM attribute (level code 0: 8 Gb) would be equal to 0.304 (1-0.616-0.080).

The second HB MNL model (Model 2) includes the choice set responses for all 424 respondents (y-data) along with predictor variables (X-matrix), but also includes additional covariates of responses to the survey questions "STCOwner" and "Gender". Then the beta's can be expressed as dependent variables of the predictors (covariates):

$$\beta_i = \varphi \cdot Z_i + u_i \text{ where } u_i \sim MVN\,(0, V_u)$$

As noted previously, there are a number of statistical methods available to assess the overall model fit. For this assessment however, a simple 'sniff test' is instead chosen to be conducted. The betadraws recorded over the MCMC simulation for Model 2 are shown in Figure 3. Likewise, the Log Likelihood values recorded over the MCMC simulation for Model 2 are shown in Figure 4. Table 2 shows the beta means and delta means for each attribute. It is evident that the beta and likelihood value tend to level out at around 1,000 iterations, suggesting that more than enough iterations are performed.

	Attribute	Beta Mean	STCOwner Delta Mean	Gender Delta Mean
screen_1	5-inch Screen	0.695	0.299	0.197
screen_2	7-inch Screen	-0.181	0.021	-0.112
screen_3	10-inch Screen	0.486	-0.320	-0.085
RAM_1	8Gb RAM	0.304	0.332	0.219
RAM_2	16Gb RAM	0.080	-0.141	-0.076
RAM_3	32Gb RAM	0.616	-0.191	-0.143
processor_1	1.5GHz Processor	-1.201	0.194	0.324
processor_2	2GHz Processor	0.939	0.055	-0.118
processor_3	2.5GHz Processor	1.262	0.751	-0.206
price_1	$199	2.596	0.189	-0.040
price_2	$299	0.311	0.026	-0.120
price_3	$399	-2.907	0.442	0.160
brand_1	STC Brand	1.458	0.974	0.506
brand_2	Somesong Brand	-0.111	-0.458	0.076

brand_3	Pear Brand	0.046	1.175	-0.116
brand_4	Gaggle Brand	-0.393	0.258	-0.466
brand_1_by_Price	STC Brand by Price	0.819	0.218	0.134
brand_2_by_Price	Somesong Brand by Price	0.126	0.258	0.060
brand_3_by_Price	Pear Brand by Price	0.047	-0.136	0.186
brand_4_by_Price	Gaggle Brand by Price	0.008	-0.340	0.023

Table 2. *Beta means and delta means for each tablet attribute generated by Model 2.*

Delta means of both the 'STCOwner' and 'Gender' covariates can be assessed to determine whether prior ownership of an STC product or gender has an impact on attribute preferences. For example, the STCOwner delta means above suggest that those who have previously owned an STC branded product can be seen to have a greater preference for tablets with a 5-inch screen, 8 Gb of RAM, and 2.5 GHz processor speed, at a $399 price point. This result is at odds to those suggested by the beta mean values. It is also surprising that those who previously owned an STC branded product had a much greater preference for Pear branded products. A similar bias toward performance attributes for males can be observed. That is, males can also be seen to have a greater preference for tablets with a 5-inch screen and 8 Gb of RAM. But in this case, males also have a greater preference for 1.5GHz processor speed and STC branded products.

3. Results

3.1 Analysis and Interpretation

From an assessment of beta means for each attribute, it is noteworthy that respondents seem to prefer tablets with a 10-inch screen, 32 Gb of RAM, 2.5 GHz processor speed, at a $199 price point, and of the STC brand. The attribute preferences for screen size, RAM and processor speed seem reasonable, noting that greater levels for these attributes coincide with values which can be seen to provide 'functional value'. i.e., a greater amount of RAM is seen to provide more value than less RAM. Likewise, the preference for a lower price point seems reasonable. There is little justification for the STC brand preference, despite the small magnitude of coefficient for each brand attribute. Finally, there does not seem to be a great deal of variation in mean betas for our brand by price interaction terms and likewise the beta means for each are close to zero. This would suggest little price sensitivity between brands. However, to further assess whether price sensitivity does in fact vary by brand, the next step is to derive the distribution of beta estimates for the final 300 draws of each of the brand by price attribute, and then determine the number of draws which fall outside of the 5th and 95th percentiles of each distribution. The results can be seen in the Table 3, which shows that each brand has only a small amount of draws which fall beyond the 5th and 95th percentiles of each distribution.

	% of draws outside 5th and 95th percentiles
brand_1_by_Price	2.3%
brand_2_by_Price	1.3%
brand_3_by_Price	0.1%

Table 3. *Distribution of beta estimates for the final 300 draws of each of the brand by price attribute, which fall outside of the 5th and 95th percentiles of each distribution.*

3.2 New Choice Scenario

The two HB MNL models would also be deployed to estimate preference shares for the alternatives in each of the two additional scenarios. First, an X-matrix which includes an effects coded version of the additional provided scenarios will be created. Next, posterior means for each beta for each subject would be calculated by using the betas from Model 1 and then multiply this matrix by the transpose of the mean betas matrix. The result is a matrix of the choice sets which can be converted to a matrix of choice probabilities by exponentiating each row and dividing it by its row sum. Choice probabilities can then be used to derive a set of preference shares for each attribute. To do this, the conjoint part-worth utilities for each respondent must be calculated. Part-worth utility can be obtained by dividing the sum of the number of times an attribute is selected by the number of times the attribute is available for selection. This ratio provides an indication of attribute preference and can be used to measure the relative importance of attributes. Table 4 shows the preference share for each attribute over the two additional scenarios. It is worth noting that there is a relatively higher preference share for 10-inch Screen, at the $199 price point and of the STC Brand.

	Attribute	Preference Share	5% percentile	95% percentile
screen_1	5-inch Screen	0.031	0.025	0.039
screen_2	7-inch Screen	0.044	0.039	0.049

screen_3	10-inch Screen	0.254	0.243	0.265
RAM_1	8Gb RAM	0.030	0.024	0.036
RAM_2	16Gb RAM	0.059	0.053	0.065
RAM_3	32Gb RAM	0.044	0.037	0.052
processor_1	1.5 GHz Processor	0.033	0.026	0.041
processor_2	2GHz Processor	0.021	0.016	0.027
price_1	$199	0.114	0.102	0.125
price_2	$399	0.044	0.039	0.049
brand_1	STC Brand	0.254	0.243	0.265
brand_2	Pear Brand	0.059	0.053	0.065
brand_3	Gaggle Brand	0.044	0.037	0.052

Table 4. *Preference shares for each attribute over two different choice scenarios that are not part of the choice task.*

4. Conclusion

In this study, two Hierarchical Bayes (HB) Multinomial Logit (MNL) models are applied to measure the price sensitivity of respondent choices which are brand specific and to also examine the possible effects of prior STC product ownership and respondent gender on the attributes' contributions to preferences. First, an evaluation of the mean betas for brand by price interaction terms of Model 1 implies that there is little variation between mean beta values and therefore little price sensitivity between brands. This is a positive indicator for STC's decision to join the market. The remaining mean betas of Model 1 are also evaluated, and the results suggest that survey respondents prefer tablets with a 10-inch screen, 32 Gb of RAM, 2.5 GHz processor speed, at a $199 price point. In terms of the effects of prior STC product ownership, the STC product owner's delta means of Model 2 suggest that those who had previously owned an STC product tend to prefer Pear branded products. Clearly, STC would do well by targeting any advertising campaign as part of their market entry against the Pear brand. The Gender delta means of Model 2 also show some interesting preferences of male respondents, with their tendency to favor a smaller screen size, less RAM and slower processor speed. It is suggestive that STC can benefit from providing two product lines, the first with a greater screen size, greater amount of RAM and faster processor speed at a higher price point. The second with a smaller screen size, less amount of RAM and slower processor speed at a lower price point. The first of these products should be targeted towards female consumers, while the second should be targeted towards male consumers. However, the marketing campaign to introduce both of these products should be targeted against the Pear brand.

Appendix

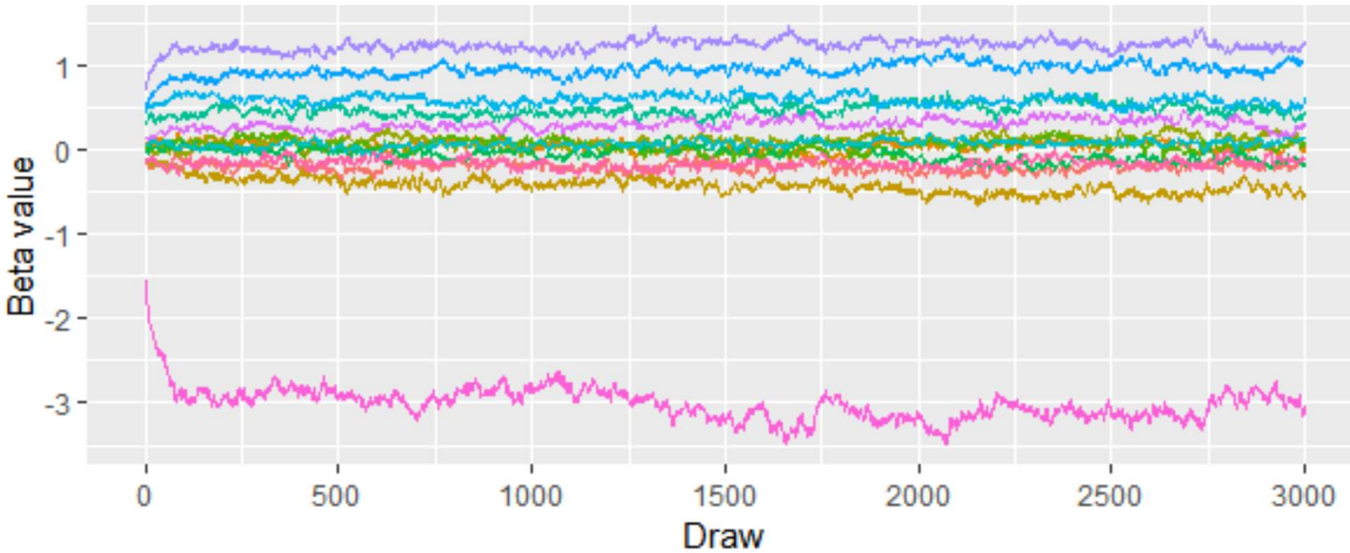

Figure 1. *Betadraws recorded over the MCMC simulation for Model 1.*

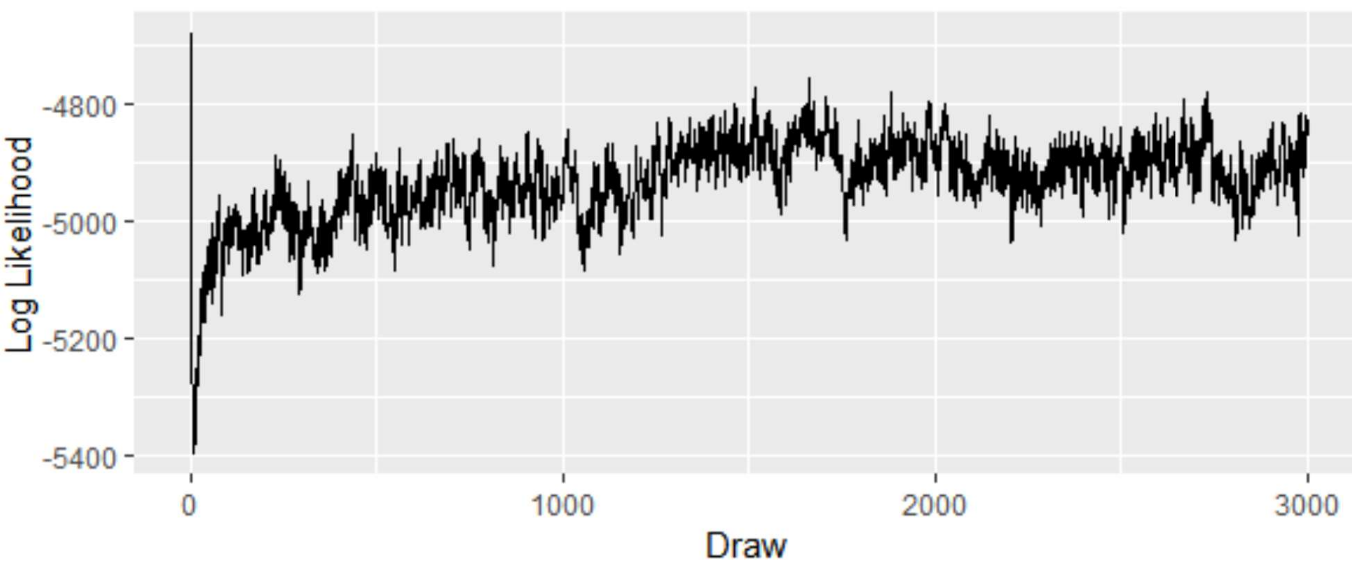

Figure 2. *The Log Likelihood values recorded over the MCMC simulation for Model 1.*

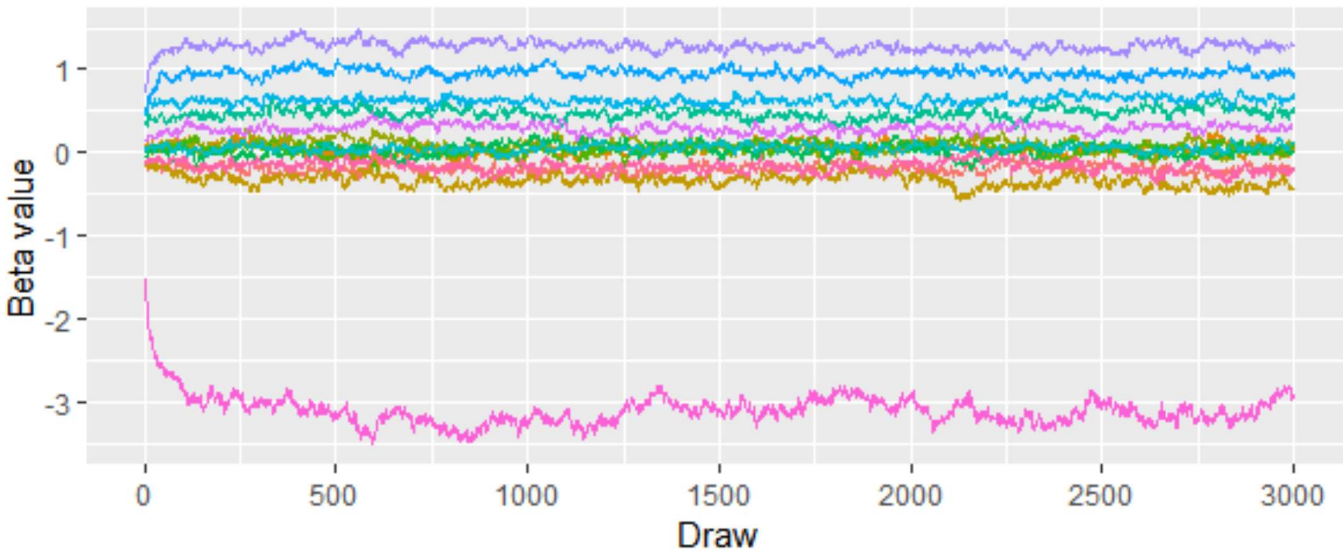

Figure 3. *Betadraws recorded over the MCMC simulation for Model 2.*

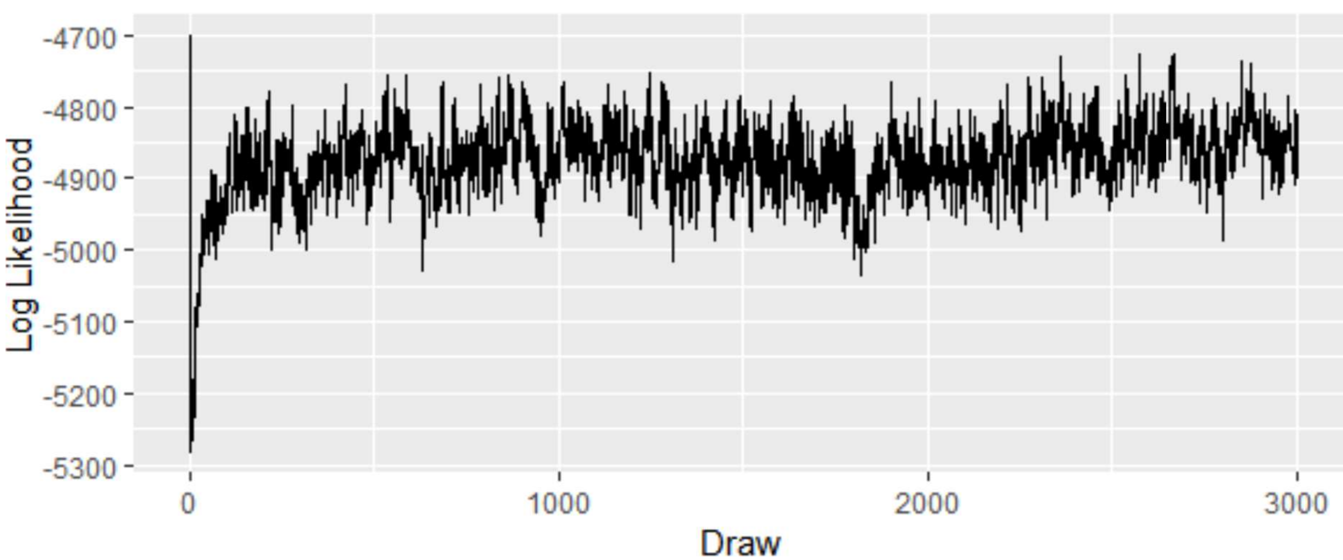

Figure 4. *The Log Likelihood values recorded over the MCMC simulation for Model 2.*