Humboldt University Berlin

Institute of Marketing

Prof. Dr. Daniel Klapper & Dr. Narine Yegoryan

Customer Analytics and Customer Insights

WS 2020/21

Special Work Performance 4: Choice-Based Conjoint analysis

Your answers including all tables and graphs must not exceed 10 pages (no appendix). You do not need to start a new page when providing your report to a new subtask. Please use typeface Times Roman in 12pt with 1.15 line spacing (in tables and graphs you may use 10pt and 1.0 line spacing) and 1 inch space on all sides. Do not forget to report your name and student number and a page number on each page, starting with number one on the first answering page. Do not include a title page or content page.
Send your report as pdf to my email address daniel.klapper@hu-berlin.de and narine.yegoryan@hu-berlin.de not later than Mar 31, 2021, 4:00pm.

Download the following files from Moodle:

- Version_StudentID.csv
- idList.csv
- Indivdata.csv
- cbc_data.csv
- readme.txt
- mxl_betaibluetooth.csv
- marketsimulationbluetooth_swp4.csv

Read/Load all data to R and carefully read the readme.txt file.

Each students will work with a random sample of 400 respondents from a population of almost 600 respondents. The structure of the data across different student samples is identical but results might differ to some extent so that each student must carefully analyze her or his data and estimation results carefully.

In order to get your individual data run the following code line in R.

```
Library(gmnl)
idList <- read.csv("idListBlueTooth.csv")
idList<-as.data.frame(idList)VersionStudentId <-
read.csv("Version_StudentID.csv")
subset(VersionStudentId,VersionStudentId$StudentId=="???????")
```

```
id<-idList$V2?????
id <- as.data.frame(id)mxl_betai <- read.csv("mxl_betaibluetooth.csv")
indivData <- read.csv("indivData.csv")
MarketSimulation <- read.csv("marketsimulationbluetooth_swp4.csv")
data.cbc  <- read.csv("cbc_data.csv")
mxl_betai <- merge(mxl_betai,id, by="id")
indivData <- merge(indivData, id, by="id")
data.cbc  <- merge(data.cbc, id, by="id")
data.cbc  <- data.cbc[order(data.cbc[,1],data.cbc[,2],data.cbc[,3]),]
# check the dimensions
nrow(indivData) == 400 # the answer must be "TRUE"
nrow(mxl_betai) == 400 # the answer must be "TRUE"
nrow(data.cbc)/12/4 == 400 # the answer must be "TRUE"
```

The file "mxl_betai" contains individual preference estimates from the CBC experiments where parameters have been estimates by a random coefficient logit model. Use this information (!!!, you do not need to re-estimate the individual level preference parameters nor you should do it) to explain the preferences of your 400 respondents. Then for example, you may want to document summary statistics or identify preference clusters. Relate the preference estimates to information from the dataset "indivData" and also explain individual preferences or cluster preferences based on the information in the dataset "indivData".

Use the individual preference data to document the willingness-to-pay (WTP) of your 400 respondents in a meaningful way. Interpret your results carefully. Make use of the information that describes the individuals and carefully describe everything in detail.

Assume the following market structure (`"marketsimulationbluetooth_swp4.csv"`):
Product 1: Sound 5.0, Weight 600gr, Battery life 12 hours
Product 2: Sound 4.0, Weight 400gr, Battery life 16 hours
Product 3: None

Compute for these products the profit maximizing price in your sample of 400 respondents if the cost of a Product 1 is 75 Euro and that of Product 2 is 70 Euro.

Document your approach always carefully and clearly. Explain everything in detail. If you have to make assumptions (e.g. for the market size) report those assumptions clearly and motivate them. You must provide your R-code as R-script in a separate document.
Do not use the "I did that"-format. You can use color in graphs and tables.

SWP 4 – Choice-Based Conjoint Analysis

In the following report, the results and implications of a choice based conjoint analysis that was done to understand the things that consumers value in a Bluetooth speaker and how consumers' preferences regarding them (mxl data) change given different factors, product attributes and product attribute levels will be examined. The examination will be done with the individual level estimates of 400 respondents (later reduced to 342), that were chosen randomly from the full sample of 600 respondents and were obtained using a random coefficient/mixed logit model. Also, the examination will be carried out in combination with the priorly done market research questionnaire conducted with the same individuals (indiv data). As will be seen below, respondents' demographic information as well as their opinions about Bluetooth speakers in indiv data will be used to define preference clusters.

Prior to examination of the mxl data, the individuals that chose 8 for their income information, 3 for gender and 5 for education are excluded from both of the data sets. These values correspond to "rather not say", which, except for the gender, doesn't convey any information while still highly affecting the calculations such as mean or the variance. The gender does convey information with the value 3 but they are excluded only to simplify the interpretability of the model. One solution could be the imputation of the "8" incomes, "3" genders and "5" educations by the mean of the rest of the individuals. This could take away the bias potentially to some extent but since there is no way of estimating these missing values 100% correctly, thus also an imputation which solves the missing values problem perfectly, this method was not chosen. Another step in the data preparation process is that, since only the proportion of people who chose residence as Germany was high enough, namely %58,19 of the data, to get meaningful results when compared with other places, people who live in Germany are coded as 1 and the others as 0. Lastly, "Subjective Knowledge, Product Involvement and Brand Awareness" columns in indiv data are reduced to one by taking their mean within each other.

The following section will provide a preliminary look at the raw results from the study. Additionally, in the section after, an examination of the mxl data that is more detailed and that also confirms some of the insights concluded from the raw results, using cbc table, indiv data and preference clusters will be done.

Examination of the results of mxl data

The utility that the respondents would get from not buying a Bluetooth speaker (none estimates, which are all negative) has a mean of -12.40 and its distribution resembles a normal-like distribution shape. Also, the distribution of the utility that individuals would get from a unit increase in price (price estimates) is right-skewed with the mean value of -7.50 and max value of -2.08. There are 8 extreme price sensitive individuals according to the IQR method. The relative price importance values of these individuals from indiv data and price importance from the cbc table being 74.23% and 84.03% higher from the population mean respectively points out in the direction that both of the studies' findings are in line with each other.

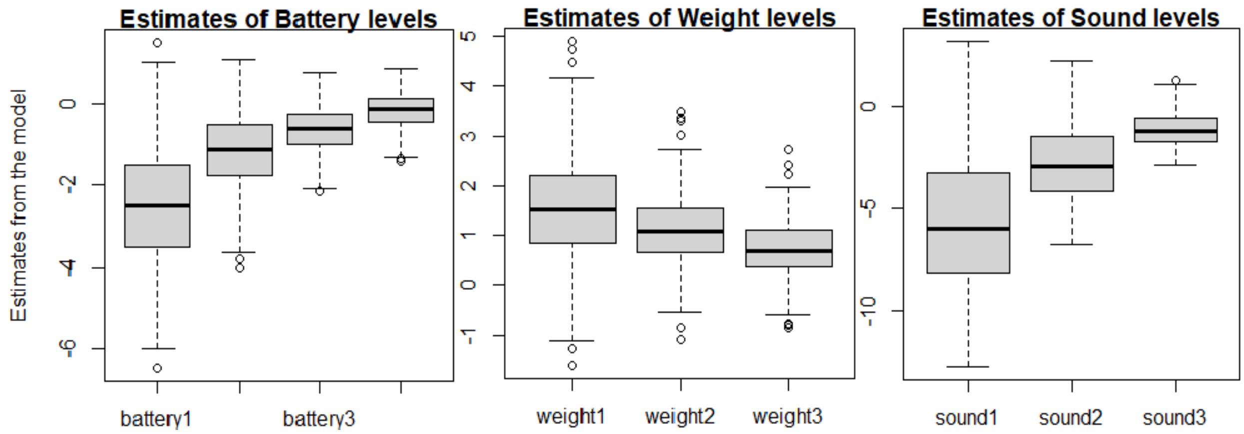

Figure 1. Raw estimates from the CBC study

With the product attribute levels, we observe a general trend that makes sense, namely that on average, people like the speaker that is lighter, that has more battery and that has a better sound quality. However, there are some unexpected results too. For example, while most of the individuals prefer a lighter speaker, there are around 15-20 same people that wouldn't like a decrease in the weight of the speaker. More specifically, the weight importance percent of all of the respondents is respectively 129.40%, 132.52% and 112.51% lower than the individuals with values lower than 0 for weight 3, weight 2 and weight 1. Regarding sound values, most people prefer a higher quality of sound but there are some people that would be satisfied with the lowest quality of sound. While this could be because even the least preferred product attribute levels might be enough for some minority, it could also be because they might have considered the price that they would have to pay for higher sound quality speakers. The reason for that is the price importance percent of all of the respondents being respectively 29.18% 44.51% and 45.54% lower than the price importance of the people with values bigger than 0 for sound 3, sound 2 and sound 1. Regarding battery level, most people also prefer a long-lasting speaker but some are still slightly satisfied with a shorter lasting speaker. The reason is likely to be the same reason as in the sound case above because the price importance percent of all of the respondents is respectively 15.41%, 18.78%, 21.38%, 2.03% lower than the price importance of the people with values bigger than 0 for battery 4, 3, 2 and 1. The last difference being so low compared to other battery levels and attribute levels in general could be due to calculation errors. One other insight that could be formed with the help of figure 1 is that as the change in the product attribute level increases, the variances in individuals' Bluetooth speaker preferences increase as well. This insight could possible be used in the following way. If a Bluetooth speaker company is designing a new product targeting a specific cluster, it can combine the information that the percent change in the variances of individuals' preferences, means of the importances and willingness to pay values of the cluster to decide on whether the combination of the product attributes that the new product has will satisfy the targeted cluster of consumers' needs and be preferred by them. This approach could be one way of making new product design decisions considering the trade-off effect between the product attributes using the results of the study. Another and possibly a simpler approach could be to use only importance values. In the end, after having different product ideas, which one will be introduced to the market could be decided by a ranking based conjoint analysis depending on the budget and time constraints of the firm.

CBC table examination

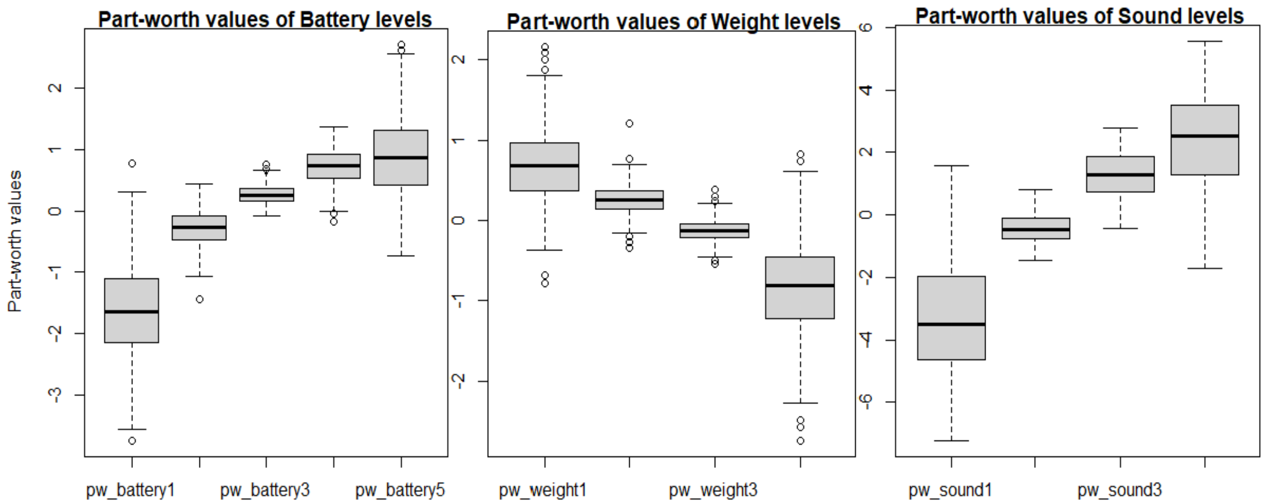

Figure 2. Part-worth values of product attributes

Figure 2 supports the idea of some potential unexpected results such as positive part worth values for the lowest battery and sound levels. Moreover, it also shows the fact that since part worth values include the left out product attribute levels too, the variances in individuals' preferences are not only high in the lowest levels but also in the highest ones. One other interpretation possible is that if a certain brand wants to reach the whole market it should have roughly levels of battery 4 or 5, weight 1 or 2 and sound 3 or 4. Also, according to same interpretation, if another brand wants to position itself as a mediocre quality brand, then the speaker they produce should have levels of battery 2 or 3, weight 3 and sound 2 as these levels' part worths are approximately half negative and half positive. In this case, for the low-end brands the lowest product quality levels are left and they might try to target individuals with positive part-worths or near 0 values. However, for such conclusions, firms would also need to consider the attribute trade-off effects and the price factor.

WTPs & Importances in combination with demographics from indiv data

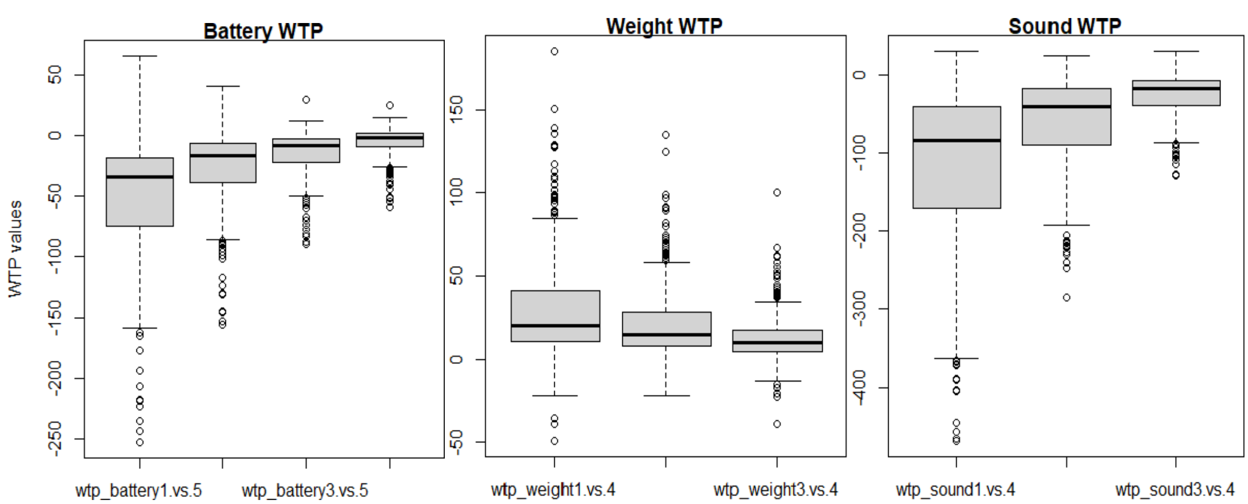

Figure 3. Willingness to pay values of product attributes

In figure 3, we see that people are less willing to pay for products that have lower sound quality, battery level and more weight in general. Also, one can see that there are some extreme willingness to pay values. The mean importances of price, sound, battery and weight are 0.36, 0.36, 0.18, 0.11 respectively. However, there are some individuals that put high importance on different attributes and, in most cases, these individuals are the same individuals that have an extreme willingness to pay values as willingness to pays and importances highly correlate with each other. Thus, what brands can do is that they can try to make more profit by exploiting this relationship and the extreme willingness to pay values because of the importance individuals give to certain attributes.

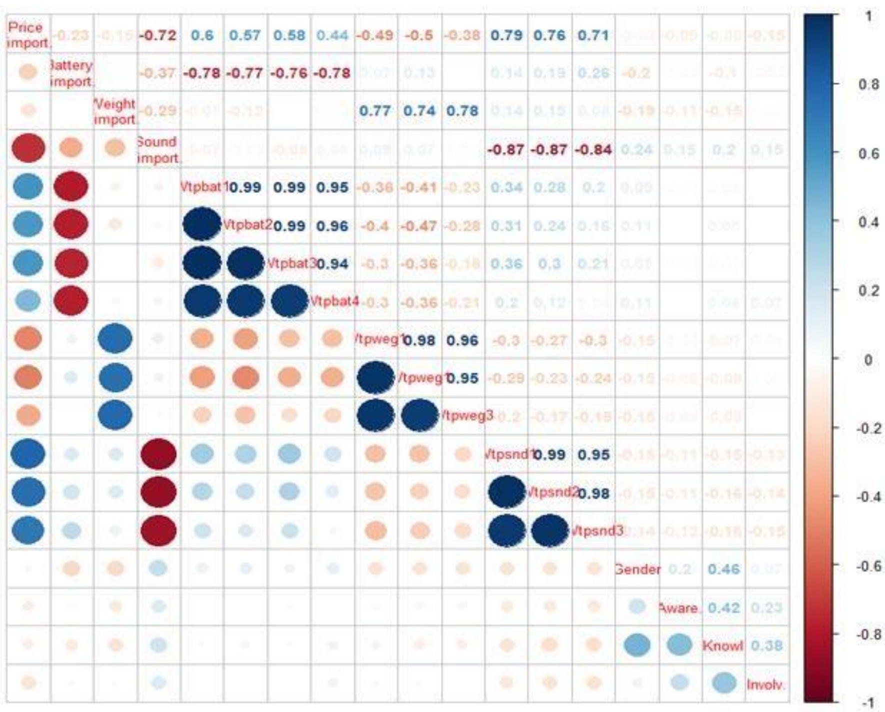

Figure 4. Correlation plot of a combination of cbc table and indiv data

In figure 4, the correlation plot combines importances, willingness to pay values as well as gender, brand awareness, subjective knowledge and product involvement information values. From the plot, there are couple of insights that can be derived. Firstly, one can confirm how significantly high the correlation between willingness to pay values and importances is, that was mentioned above. Secondly, the trade-off between different product attributes, with the highest one being in between price and sound importance, are reported as it was with the relative importances, which proves the consistency of the two studies. One other potentially interesting finding is that the correlation between the sound importance and brand awareness, knowledge and involvement columns are 0.15, 0.20 and 0.15 respectively. Brands can possibly try to make use of this insight by targeting the individuals that value sound more than the others by showing them online ads that includes slightly more technical information about the product. Last insight that can be derived from figure 4 is that males value sound more whereas females value weight and battery more. Differences in individuals' preferences based on various demographics. will be examined in more detail below, which will then be followed by the formation and the analysis of the preference clusters.

```
> gender_importances
                female      male
imp_price   0.3633004 0.3496480
imp_battery 0.2001311 0.1589775
imp_weight  0.1233446 0.0952393
imp_sound   0.3132238 0.3961351
```

```
> gender_wtps
                      female        male
wtp_battery1.vs.5 -55.577561  -46.943061
wtp_battery2.vs.5 -29.367619  -23.016771
wtp_battery3.vs.5 -16.061207  -13.269923
wtp_battery4.vs.5  -6.772927   -4.205875
wtp_weight1.vs.4   34.817474   25.328422
wtp_weight2.vs.4   24.550055   18.164768
wtp_weight3.vs.4   15.422114   11.061886
wtp_sound1.vs.4   -99.672546 -131.548821
wtp_sound2.vs.4   -51.275319  -69.856124
wtp_sound3.vs.4   -21.155976  -28.795101
```

Figure 5. Importances and willingness to pay values based on gender

From figure 5, we see that there is no significant difference in price importance based on genders. Females value weight and battery more. Males value sound more. Same effect can be observed in willingness to pay values. One potentially interesting finding regarding education level is that on average undergraduates have approximately 11.45% higher mean of price importance compared to high school and graduate segments separately. When the willingness to pay values of undergraduates are examined, it is found that they require less compensation for a reduction in the level of products especially for battery and sound, which is something that the brands could benefit from. Another possibly interesting finding regarding the income level of the individuals is that in price importance mean there is a 10.40% decrease from income below 500 to between 500 to 1000. However, the change in between next 3 income levels don't exceed 3.40% which indicates that in the given data, in between the second and the third income levels there might be an "income breaking point". Lastly, one other potentially interesting finding is that the speaker owners value price 8.82% more and value sound 11.02% less compared to the non-owners. This creates an opportunity for low segment brands if they can convey this information to the individuals who haven't bought a Bluetooth speaker yet intend to do so in the future.

Clustering

To see whether there exist distinct clusters of individual's preferences that could be targeted by the brands, clustering is done. In figure 6, on the right, the best result are acquired with hierarchical clustering (ward) of 4 clusters with gower metric and using importances as the variables only.

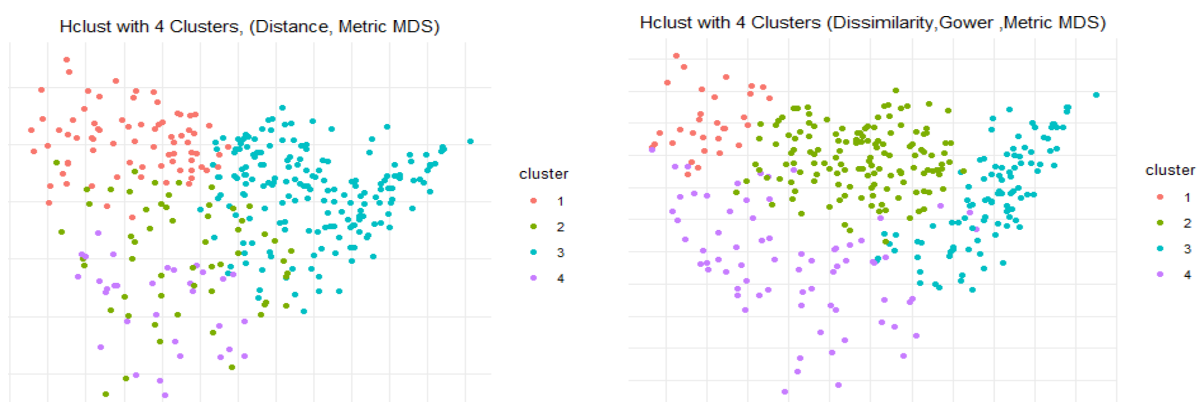

Figure 6. Multidimensional Scaling with Hierarchical Clustering

To decide on that the combinations of the following clustering decisions were tried: which distance will be used (Euclidean, Gower), which variables will be used (combination of the two studies, importances, part worths, willingness to pays and part worths + price estimates) and which clustering method will be used (hierarchical with single complete, average, centroid, ward, k-means, model-based with 3 to 5 clusters). Lastly, importances were chosen as the variables they captured all the information, were low in number and were the most important from the brands' point of view.

	imp_price	imp_battery	imp_weight	imp_sound	Gender	Age	Income	brand_awareness	knowledge	involvement
1	0.6693253	0.1218717	0.07309032	0.1357126	1.483871	3.064516	2.645161	0.2822581	3.348387	3.696774
2	0.4064646	0.1335218	0.09014527	0.3698683	1.582734	2.848921	2.316547	0.4136691	3.962590	4.303597
3	0.1842870	0.1780597	0.09608867	0.5415647	1.590000	3.020000	3.330000	0.4250000	4.028000	4.620000
4	0.3619492	0.2879037	0.17510581	0.1750413	1.402778	3.138889	2.777778	0.3663194	3.450000	4.294444

Figure 7. Clustering results

The 1st group consist of 31 people. They are very price sensitive, they have average income and low knowledge and brand awareness. One way to target this group is by informing the individuals about the specific brand and market the benefits of having this specific product, which would hopefully decrease their price importance to some extent. The 2nd group consists of 139 people. This group values price/performance ratio of the product as they value both price and sound. Additionally, there are slightly more males than females, their income is low and they are knowledgeable about Bluetooth speakers in general. One targeting approach could be to target them with ads that focus on youth related themes such as being energetic, being popular, having fun etc. as they are the youngest of the 4 clusters. The 3rd group consists of 100 people. One feature of this group that make them stick out from the others is that they have the highest sound importance. In this group there are also slightly more males than females, which is what is expected as it is known that males tend to value sound more than females. They have above average knowledge of Bluetooth speakers. Also, this group can be targeted by higher end brands with good sound quality as they value sound, are not price sensitive and have the highest income. The 4th group consists of 72 people. This is the group that values battery and weight of the product the most as well as to some extent price but not so much to sound. The mean levels of importances of this group matches with the finding that females value weight and battery more as we can see there are slightly more females than males. Brands could target this group considering the importance they give to weight and battery and its average age as they are slightly older than the other groups.

Market Simulation

In this section, a market simulation consisting of 3 options is assumed in order to see what possible insights can be derived from the results of the study and how these results can be used for decisions that Bluetooth speaker brands can take in real-world situations. First option is product 1 which has a battery level of 3 out of 5, a weight level of 3 out of 4 and a sound level of 4 out of 4. The second product in the market simulation has a battery level of 5, weight level of 1 and a sound level of 2. The third option is the option that where individuals find neither of the products and their attribute combinations attractive, thus decide to opt out of the market by choosing the none option. To summarize, product 1's advantage is having the highest quality sound while having slightly less desirable qualities than product 2 in terms of weight and battery. On the other hand, product 2's advantage is having the most desirable battery and weight levels; however, it has a sound level 2.

First of all, how much value do the individuals put to the given attribute combinations while both products have a price of 70, 90, 110, 130 and 150 euros respectively is examined.

	shares_p1_p2_0.7	shares_p1_p2_0.9	shares_p1_p2_1.1	shares_p1_p2_1.3	shares_p1_p2_1.5	shares_p1_1.5	shares_p2_1.5
[1,]	0.682707649	0.66352754	0.6181180	0.5555285	0.4956831	0.2632599	0.83643407
[2,]	0.309878770	0.29679483	0.2680355	0.2387820	0.2123069	0.6002828	0.06525398
[3,]	0.007413581	0.03967764	0.1138465	0.2056896	0.2920100	0.1364573	0.09831195

Figure 8. Market shares with different pricing scenarios

From figure 8, we see that the attributes of product 1 lose their value in the eyes of the customer faster than the 2nd product after a price of 90 euros as, after this price, the loss of share of product 1 increases while 2nd product's keeps decreasing steadily and the none option's gain of share increases. This finding could possibly indicate that if brands wants to secure their share of markets more, then they may want to produce products with more desirable weights and batteries as products with only a good sound are more vulnerable to share losses in case of a price over 90 euros. The same finding can be confirmed from the negative correlation of 0.38 between price importance and shares of product 1. Additionally, last two columns show when only one product's price is changed to 150 and the other is kept 100 euros. One can see that product 1 has 26.33% shares when its price was changed to 150 euros while in the same scenario product 2 has only 06.53% shares, which shows the importance of sound level and how product 1's attributes outperform product 2's. Also, the correlation between the raw model estimates (mxl data) and purchasing probabilities of individuals if both of the products were priced as 100 is examined. It is found that the correlation between the change of utility resulting from a one unit increase in the product 1's price, and the change in product 1's shares is 0.28. The same correlation with product 2 is 0.11. In other words, as the price of the products increase, the share of product 1 increases more than product 2's shares, which points out how important the sound quality level in consumers' purchasing decision. Another is that when the price of product 1 is 120 and the price of product 2 is 100 the market shares are 45.23% and 43.68% respectively. This indicates that for the market shares of the two products to be equal the price of product 1 needs to be approximately 20 euros higher than the 2nd one. As the price levels get very high or low this value can change but it remains more or less the same throughout also different prices.

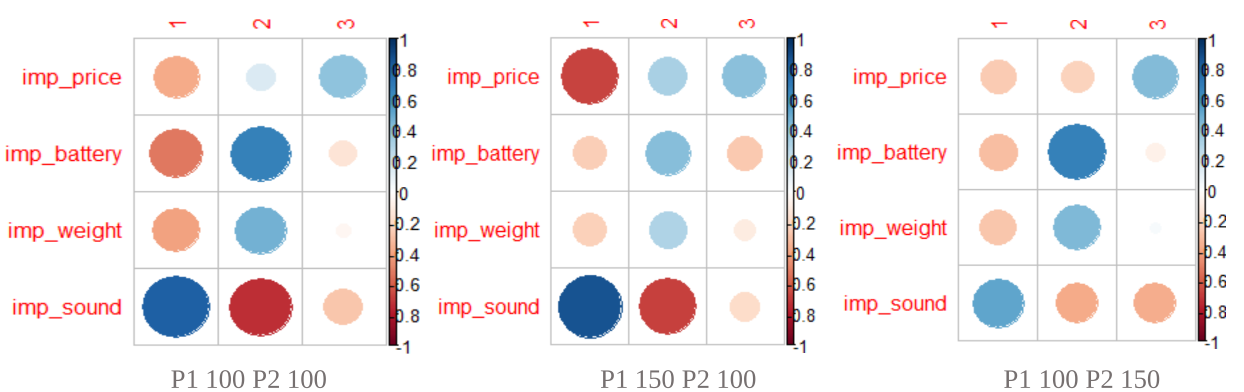

Figure 9. Importances in different price scenarios

Figure 9 proves two points. The first is that in the case where both of the products are 100 hundred euros, the fact that price importance and share of the first product correlating negatively shows that when individuals have enough money (thus also a lower price importance) they would care about sound more and they get more likely to buy the product with a better sound quality rather than the one that is lighter and lasts longer. Second point that can be made using figure 9 is that when the price of a product is much higher than the other one, people that value that the more expensive product's desirable attributes are likely to still buy the product: however, from the current study it can be observed that the ones that decide that the more expensive product is not worth it anymore then switches to the other product or the none option.

Another insight that could be derived when the relationship between the indiv data and the shares of the products is examined given that both of them has a price of 100, it can be seen that significantly more males prefer the 1st product likely because it has the best sound quality. In this examination, one can also conclude people with higher brand awareness and knowledge are more likely to purchase the 1st product as these categories correlate positively with sound importance. Similarly, because the 2nd product has a more desirable weight and battery level, females tends to purchase it more as they value these attributes more. Last insight one could come up with is that, the correlation between brand awareness and knowledge and product 2 shares are -0.15 and -0.21, which are the same numbers as in the first product's case but just the negative versions. Last insight can also be confirmed by the 0.25 positive correlation between gender and knowledge, which suggests female knowledge level of Bluetooth speakers are lower than males on average.

Changing shares depending on price levels

In this section, prices of each products is separately increased from 0 to 200 by a unit of 1, keeping the other product's price constant and the results are elaborated on below.

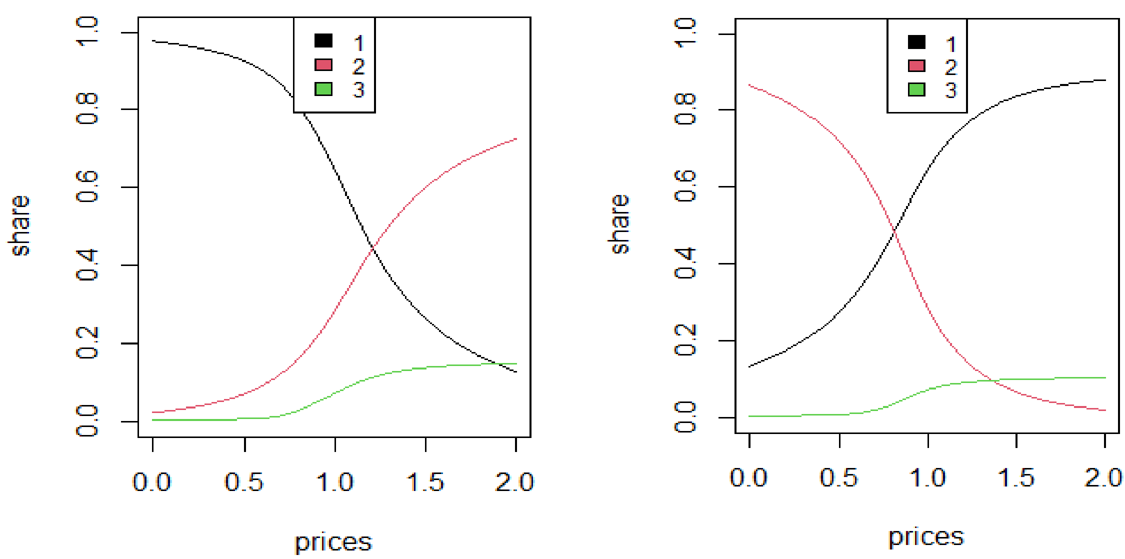

Figure 10. Change of market shares in different price scenarios

In figure 10, the color black shows the 1st product's shares, color pink shows the 2nd product's shares and color green shows the none option's shares in the market simulation. The plot on the left shows the change in shares of options as the 1st product's price goes from 0 to 200 and plot on the right shows the change in shares of options as the 2nd product's price goes from 0 to 200.

There are couple of insights that can be derived from this figure. First is that the first product takes over the market way earlier than the second product does, more specifically, about thirty euros before. This one more time proves that the first product is more desirable than the second product on average. Second insight to be derived from these plots is that share of product 1 when its price is 200 is higher than the product 2's, which points out that individuals tend to stick with the product 1 at higher prices more than they do with product 2. Lastly, because of the same reason, the share of the none option exceeds the share of the product with constant price about 50 euros earlier in the case where the second product's price changes and none reaches a higher share in the other case.

Price for Maximum Profit Calculation

One can also predict the optimum price at the same time for them to generate maximum profit. However, for this case to be real in the real world there are some assumptions that needs to be fulfilled. This case can be realistic only if there is a cartel formation in the market meaning that couple of brands decide the product prices together so that every brand gets "the most" profit while other brands are still in the market. Another way for this to be usable in a real life situation is when there is a monopoly that has two competing products with each other, which needs a tuning in their price strategies. Another assumption that needs to be fulfilled so that this price calculation can be usable in real life is that the utility functions of both of the products are known exactly, which includes some problems. First is that there could be omitted variable bias meaning that some consumers can get utility from an attribute that has not been thought of. Second is that, although impossible, even if the utility function was perfectly defined there is also some unexpected variation in the utility that cannot be calculated or predicted. If these assumptions are met, then the calculation can be done.

The calculation is done assuming a market size of 1. The reason for this is that because the exact market size is hard to estimate and assuming a market size of 1 makes it easier to scale the results for bigger number of individuals. The idea behind this calculation is that since the optimal price for one product changes given the other product's price change, one needs to iteratively try or calculate the optimal price in which both of the products make the most profit.

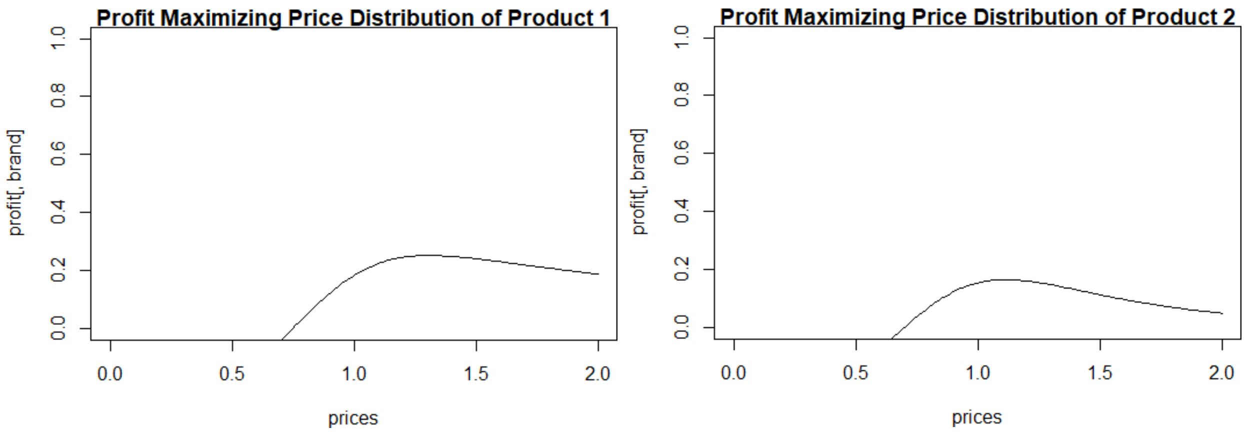

Figure 11. Changing profits in different pricing scenarios

In this case, the maximum profits are calculated assuming that cost of the first product is 75 euros and the cost of the second product is 70. After couple of tries, it is concluded that the profit maximizing price for both of the products in the market are 131 euros for the first product and 113 euros for the second product. In figure 11, one can see how both of the products' prices can be determined based on the effect they have on each other, observe the profit curve of each of them and confirm whether the stated prices are indeed profit maximizing prices for both of the products. In this study, the market size consisted of 342 individuals. Taking into account this number of market size, the first product's optimal price corresponds to a profit of 8563.56 euros and the second product's price, a profit of 5584.09 euros. Also, these prices reflect a share of 44.71% for the first product, 37.94% for the second one and 17.34% for the none option. However, to be able to apply these findings to a real-life situation, one needs to estimate the true market size and in actual markets the number of competing products are much higher, which makes the calculations and the interpretation much more complicated.

Individual Optimal Pricing

One can also compute the maximum prices that each individual is willing to pay in a given market setting. This could be done rather easily if the market simulation consisted of 1 product and a none option only by calculating the last price willing to be paid and utility an individual would get before he or she goes on with the none option but when there is more than 2 choices, in our case 3, calculations get more complicated as one needs to compute the maximum utility and price for the decision in between product 1 and 2. One potential solution that to calculate optimal prices according to the preference clusters' willingness to pay values that could be done as a further analysis. In real life, these calculations could be much harder to do. Also, even if it is calculated, it wouldn't be trustable for the values that are close to some other utility or price estimates due to the fact that we assume that there is no idiosyncratic shock in the model.