Laboratory of Customer and Business Analytics

Alberto De Benedittis

December 2021

1 Abstract

The aim of this paper is to present the results of a **Choice-based Conjoint Analysis**. More precisely, the analysis wants to assess the determinants of consumers preferences with respect to a specific product : trading-investing apps. To do so, an ad hoc **choice-based survey** with a complete random design has been created . Then once, a reasonable number of answers were collected the data has been used to fit different models: **Multinomial Logit Models** and **Mixed Multinomial Logit Models**.

2 Introduction

In recent times a new socio-economic phenomenon has taken place: the wide spread of trading-investing apps. These apps allow almost everyone to invest their money in the financial market and nowadays it is difficult to know exactly how many people turned themselves into traders. While this phenomenon has been defined as an attempt to 'democratizing finance' by the creators of these apps, expert from Wall Street and other stock exchange warn that an high number of non expert users could threaten the stability of financial markets.

The potential of these apps has become clear to everyone during the month of January 2021. For the first time in history small traders and investors have beaten hedge funds. More precisely, hedge founds were investing against GameStop a retailer of videogames that due to Covid-19 pandemic was facing a serious crise. At the same time, a group of small private investors organized themselves through Reddit, a social news website, and thanks to Robinhood, a mobile app for trading, decided to invest in mass on GameStop shares. The strategy of small investors had the effect to drastically change the price of GameStop shares which went from 18$ to more than 350$. This caused loss of billions of dollars for professional investors and on the other hand, big profits for small investors.
Moreover, in this particular historical period the attractiveness of trading apps is even higher due to the high inflation that is going to persist for the next year or two according to the European Central Bank.
The rise of inflation is mainly caused by three factors :

- our economy is reopening fast and so companies are facing problems in rebuilding the supply chains that were badly hit by the pandemic,
- higher energy prices are pushing up inflation,
- during the pandemics prices were exceptionally low and so comparing today's higher prices to those very low levels means differences will seem large (this is called base effect by statisticians).

Hence, it is interesting to understand which are customers preferences with respect to trading apps. This is particularly relevant also because this is a relatively new product and thus there is no saturation in the market and new players can still enter into it. However, this could also be a problem because people may not have clear and well-defined preferences on trading apps.

Nevertheless, in order to properly understand customers preferences there is the need to accurately recreate a selection context. This can be accomplished through a **conjoint analysis**.

3 Survey Description

The first step of the conjoint analysis is the creation of the survey. This is a key moment because from the survey design derives the kind of model and the quality of the data used to fit it.
With respect to conjoint analysis there are two main approaches:

- the **Traditional conjoint survey** where respondents are asked to rank different product profiles. This is a context of multilevel linear regression.
- the **Choice-based conjoint survey** where customers are asked to make choices among alternative products with differing characteristics. In this scenario the response variable is no more quantitative and so it is the context of generalized linear model.

The approach used for the following analysis is the second because it is considered more realistic. Indeed, during the purchasing process customers do not rank all the options but typically pick only the preferred product.
Once the kind of survey is decided, it is important to define the products attributes and their respective levels. For this survey, six attributes have been defined with different levels each. They are the following:

- Platform type: Web App, Desktop App, Mobile App
- Deposit min and max: 0€-1000€, 10€-10.000€ , 100€-infinite
- Fees: 0%, 0.015%, 0.025%
- Financial instrument: Stocks, Crypto, ETFs, Commodities
- Leverage: YES, NO

- Social Trading: YES, NO

From a statistical point of view, two fundamental aspects that have to be taken into consideration when planning a conjoint survey: the number of respondents that should participate, that is the *sample size* and the choice of the *experimental design*.

3.1 Experimental design

Firstly, it is important to remember that the higher is the number of attributes and levels in a conjoint survey, given a fixed number of respondents, the lesser precise the part worth estimates will be. Secondly, finding the proper questions to ask can lead to more precise coefficients estimates and predictions of preference shares. However, it is not convenient to show all the possible product profiles in each survey. As a matter of fact, when respondents are asked to answer to too many questions there is the risk that they quit the questionnaire or that they answer randomly. A particularly good approach is assigning different conjoint questions to each survey participant and generating the product profiles randomly. However, many survey platforms, such as Google Forms, allow only to give the same questions to each respondent.

For this reason, an ad-hoc survey has been created through Shiny, an R package. This survey built from scratch allows to pick a random sample of product profiles, from the set of all possible product profiles, for each question every time the survey is opened. In other words, each respondent does a different survey. Lastly, in each questionnaire there are twelve choice-based questions.

3.2 Sample Size

The sample size plays an important role as well because it determines the accuracy of the results. The bigger is the sample size the more precise are the estimates. Namely, the standard error is smaller. A good way to assess the reduction of standard errors due to the increase of sample size consists on simulating data from known (assumed) parameters, estimate the model from the simulated data, and examine the standard errors or the resulting intervals for the predicted preference shares.
Such an approach can help in determining how many respondents are necessary with respect to a given number of attributes and levels.
However, there is no control on the number of respondents because it depends on the response rate. Nevertheless, as a rule of thumb, we may consider thirty respondents as the lowest number of respondents in order to start making hypothesis on customer preferences.

3.3 Demographic information and Appendix

In addition to the choice-based survey, there is a preliminary section in the survey where are asked demographic questions. They are the following:

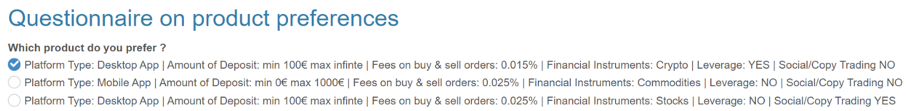

Figure 1: Example of a Choice-based question

- Indicate your gender: Male, Female, Other
- Indicate your age with respect to different age classes: 0-18, 19-25, 25-35, 35-45, 55-65, > 65
- Indicate your education level: Middle School, High School, Bachelor's, Master's, PhD
- Indicate your Job Sector: Agriculture, Services, Industry, Other
- Indicate your marital status: Married, Not married

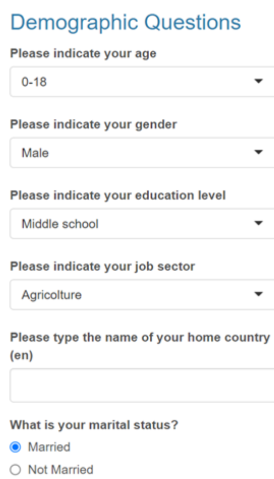

Figure 2: Demographic question

The answers to the above questions can be useful in the final stages of the analysis in order to detect whether some eventual niche in the market is based on respondent level variables such as the gender, the age or the education level. Lastly, since the topic can be complex and requires some previous knowledge, a glossary has been created. It can be accessed by clicking on the appendix tab in the blue panel at the top of the survey's page. In this glossary a brief definition for key concepts such as ETFs, crypto etc. is provided. These definition are not visible in the main page because there could have been the risk to annoy the respondents. Actually, a small catchy intro is provided in order to involve respondents more to the topic.

4 Description of data

In the conjoint survey each respondent answers twelve questions. Since each question has three alternatives, each respondent evaluates a total number of thirty-six product profiles. In addition, each respondent answers to the demographic question listed above.
Data is collected automatically into a google spreadsheet divided in three sheets:

- One where are collected demographic data
- One where are collected respondents selected choices
- One where are collected all the alternatives in each survey

The sheet containing the respondents' preferences is organized in the wide format, one column for each of the twelve answer. The sheet containing all the alternatives in the survey is organized in the long format, each row correspond to an alternative. This last sheet is the one that will be used for the choice-based conjoint analysis. But, there is the need to create unique indexes. To do so, three columns need to be added:

- a column with the respondent's ID
- a column that represents for each alternative the associated question number (from 1 to 12)
- a column that represents the position of the alternative namely, upper, middle and lower

The participants to the survey were mainly recruited through messages in university chats and by fixing posts, with the QR code of the link of the survey, on the bulletin boards of Trento main libraries. However, the response rate has been quite low and just thirty-six respondents participated to the survey. Moreover, the answers of few participants are affected by a bug in google sheets which did not correctly save their preferences. For this reason some observations cannot be included in the analysis. At the end the of the pre-processing the dataset used in the analysis was made by 1007 rows and 11 columns.
Moreover, due to the above mentioned recruitment process, the sample of respondents is quite homogeneous. As a matter of fact, the average respondent is between 19 and 25 years old, came from Italy, has a quite high education level (Bachelor's or Master's) and it is not married.

This will probably cause a reduction in the significance of the analysis' results. Indeed, even before a large number of respondents, the sample should be random in order to make the results reliable.

resp.id	ques	position	Platform	Deposit	Fees	Financial_Instrument	Leverage	Social_Trading	Gender	choice
1	1	top	Web App	0	0.015%	Crypto	NO	YES	Female	0
1	1	middle	Desktop App	0	0%	Crypto	YES	NO	Female	0
1	1	low	Web App	10	0.025%	ETFs	YES	YES	Female	1
1	2	top	Desktop App	100	0.025%	Crypto	YES	NO	Female	0
1	2	middle	Desktop App	10	0%	Stocks	YES	YES	Female	1
...	...	...	...	...	...	...	...	...	...	...
36	11	middle	Web App	100	0.025%	Crypto	NO	NO	Male	0
36	11	low	Mobile App	0	0%	Commodities	YES	NO	Male	0
36	12	top	Web App	10	0%	Stocks	NO	YES	Male	0
36	12	middle	Desktop App	10	0%	Crypto	YES	YES	Male	1
36	12	low	Mobile App	10	0.015%	Commodities	YES	YES	Male	0

Figure 3: Overview of the data

5 Explanation of the methodological choices and empirical strategies

To understand customers' preference both on individual and population level, there is the need to have multiple observations for each respondent.

Hence, applying a simple linear model to multilevel data will produce inaccurate results with the risk that the importance of same predictors is underestimated or overestimated. This occurs due to the fact that the assumption of independence among the observations is violated. Moreover, when the multilevel structure of the data is ignored it is likely to miss important relationships involving each level in the data.
For this reason, using a more complex model such as the Multinomial Logit Model leads to a better understanding of the data. As a matter of fact, by performing this kind of analysis, it is possible to estimate the values in the model for each respondent, while exploiting the contextual information about the entire sample of respondents. This is a common practice in the field of customer analytics because it is very useful to determine individual-level effects, such as which customers are more interested in a product attribute. Thus, to estimate both a population-level effect and an individual level effect jointly, it is better to use a Multilevel Linear Model
Furthermore, the choice of the type of survey is another important determinant for the analysis. As already said, the kind of methodology adopted is the so called **Choice-based conjoint survey**. This choice has been made because it is based on more realistic assumptions compared to the traditional conjoint survey. Indeed, choosing among alternative products is a more natural activity as it mimic better what customers actually do when they buy.

It is important to remember that with intangible products empirical results suggest that the traditional conjoint model is more appropriate.

Nevertheless, a choice-based conjoint model has been chosen because there is the risk that customers get confused or bored by a ranking task since the simple task of selecting the preferred product profile may be already complicated due to the complexity of the product itself.

For the same reason, not too many attributes and levels were included in order to make the the task of the respondents more straightforward.

6 Presentation and interpretation of the results

The analysis has been done proceeding from the simplest model to the more complex ones.

Although a simple logistic regression is not the most appropriate choice to analyse the data, the analysis begins from there because it could be useful to start understanding which are the important attributes according to the respondents.

```
Coefficients:
                             Estimate Std. Error z value Pr(>|z|)    
(Intercept)                  -1.47919    0.25254  -5.857 4.71e-09 ***
PlatformMobile App            0.50026    0.17335   2.886  0.00390 ** 
PlatformWeb App               0.40460    0.17395   2.326  0.02002 *  
Deposit10                     0.05150    0.16584   0.311  0.75613    
Deposit100                   -0.37086    0.17316  -2.142  0.03221 *  
Fees0.015%                   -0.06715    0.16631  -0.404  0.68638    
Fees0.025%                   -0.25690    0.17374  -1.479  0.13923    
Financial_InstrumentCrypto    0.04836    0.20708   0.234  0.81536    
Financial_InstrumentETFs      1.02680    0.19517   5.261 1.43e-07 ***
Financial_InstrumentStocks    0.59890    0.19932   3.005  0.00266 ** 
Leverage YES                  0.17280    0.13888   1.244  0.21342    
Social_TradingYES             0.27872    0.13877   2.008  0.04460 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1282.4  on 1006  degrees of freedom
Residual deviance: 1220.5  on  995  degrees of freedom
AIC: 1244.5
```

Figure 4: Logistic Regression Coefficients

From the above summary, it is possible to start having a partial idea of what could be the key attributes for understanding consumers' preferences. The above results although, partial and not statistically appropriate, give a useful hint. Indeed, it is reasonable to think that the financial instrument available on the trading app plays a central role in customers preferences. The choice of ETFs could be justified by the fact that users prefer a less volatile product and so less risky product when they chose to do trading or investments. This is also confirmed by the fact that also stocks are significant for respondents and have a positive coefficient.

Other attributes that seem to be relevant are relative to the kind of platform used and the deposit.

However, it is important to remember that these results may be overestimation and therefore some of the variable could actually be not relevant for predicting customer's preferences. Further analysis are needed to assess customers preferences. Thus, to better understand customers preference a more complex model that takes into account both a population-level effect and an individual level effect jointly is needed : the MNL model.

```
Coefficients :
                              Estimate Std. Error z-value  Pr(>|z|)
(Intercept):middle           0.0011163  0.1465574  0.0076  0.993923
(Intercept):top              0.1822337  0.1399917  1.3017  0.193003
PlatformMobile App           0.5193716  0.1768289  2.9371  0.003313 **
PlatformWeb App              0.4248038  0.1761781  2.4112  0.015899 *
Deposit10                    0.0544378  0.1659834  0.3280  0.742933
Deposit100                  -0.4418045  0.1795120 -2.4611  0.013850 *
Fees0.015%                  -0.0547807  0.1718433 -0.3188  0.749891
Fees0.025%                  -0.2457132  0.1813147 -1.3552  0.175362
Financial_InstrumentCrypto   0.1196414  0.2107310  0.5677  0.570209
Financial_InstrumentETFs     1.1093949  0.2063593  5.3760 7.614e-08 ***
Financial_InstrumentStocks   0.6603908  0.2119955  3.1151  0.001839 **
Leverage YES                 0.1725634  0.1398804  1.2336  0.217333
Social_TradingYES            0.3288047  0.1448400  2.2701  0.023200 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -335.31
McFadden R^2:  0.088887
Likelihood ratio test : chisq = 65.425 (p.value = 8.9657e-10)
```

Figure 5: MNL model Coefficients

The MNL model output looks similar to that of the logit model. The *estimates* column provides the estimated average part worths for each level. They have to be interpreted with respect to the reference level of each attribute:

- Platform : Desktop App
- Deposit : 0€
- Financial instrument: Commodities
- Leverage: No
- Social Trading: No

From the summary of the model, it is possible to see coherence with the results of the previous simple logistic model. Indeed, the important variables are still the same as before.
It is also important to remember that the order of magnitude of the estimates

provides how strong the preferences are. Moreover, the MNL model coefficients are on the logit scale and so they range mainly between -2 and 2.

In addition, when analysing the estimates it is important to look at the sign of the coefficients together with the values to better understand the results. For example, the highest value is related to ETFs as financial instruments, this strength the initial hypothesis that this is the preferred option for customers. It is more than twice the value associated to stocks. Furthermore, it is interesting to notice that the type of platform plays an important role as well. Indeed, customers prefer the mobile app compared to the web app and desktop app. Also this results seems reasonable. Indeed, it is possible to hypothesize that having a mobile app allows to easily check investments.

Deposit 100€ seems to be a relatively important predictor either, with a negative coefficient. This results could be explained by hypothesizing that customers do not trust this kind of product that much or they are newbies and so they are not willing to invest big amounts of money.

Lastly, before considering the standard error, it is nice to notice that all the variables associated to higher fees have negative coefficients but they are not significant. This suggests that customers do not like higher fees but they are not determinant for their choices. This can be justified by the fact that these fees are relatively low e.g. 0.015% or 0.025%. However, from the business point of view this can be precious information because the firm could introduce small fees that boost their revenues without losing customers.

Moving on to the standard error, it is possible to notice that these values are relatively quite high. This is definitely not a good sign. This is caused by the low number of observations and the high number of attributes in the conjoint survey. Hence, there is the need to be careful in drawing conclusions from this analysis because its precision is quite low. This can be noticed also by looking at the McFadden R^2.

The intercepts, that represent the so-called *alternative specific constants* represent the effect of the position of the options. Namely, how much customers prefer the middle/bottom alternative instead of the top one. However, from the summary it is possible to see that customers are not affected by the position of the alternatives. This is good because it means that respondents have replied rationally. In other words, respondents do not chose an alternative due to its position. For this reasons, it is useful to drop the intercepts in order to gain in parsimony and precision. This hypothesis can be checked through a likelihood ratio test. Firstly, there is no substantial difference among the two models in terms of magnitude of coefficients, standard error and significance.

Moreover, the comparison between the full model and the model with no intercept (Figure 6) leads to a p-value of 0.3184. According to this relatively big value, it is possible to conclude that the two models are not significantly different in terms of goodness of fit and hence they explain the data equally well. This indicates that the alternative specific constants are not necessary to adequately model the data.

```
Likelihood ratio test

Model 1: choice ~ Platform + Deposit + Fees + Financial_Instrument + Leverage +
    Social_Trading | -1
Model 2: choice ~ Platform + Deposit + Fees + Financial_Instrument + Leverage +
    Social_Trading
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  11 -336.45
2  13 -335.31  2 2.2889     0.3184
```

Figure 6: Likelihood ratio test Model with and without intercept parameters

6.1 Willingness to Pay

In order to improve the interpretability of the results, it could be useful to consider the deposit as a quantitative variable, and to recompute the model and perform a Likelihood ration test. This kind of procedure typically offers another kind of advantage: it allows to compute the *willingness to pay*.

Even though this kind of product typically does not have prices, due to the fact that companies base their revenues on fees and commissions, it could be interesting to consider as a kind of price the amount of money needed to start investing with the app.

This can be useful to understand how customers value the various trading app characteristics. In addition, this is helpful because there is no profit related to the initial deposit for the companies who issue the trading apps.

```
Coefficients :
                                  Estimate Std. Error z-value Pr(>|z|)
PlatformMobile App               0.5222516  0.1763907  2.9608 0.003069 **
PlatformWeb App                  0.4238644  0.1752358  2.4188 0.015571 *
as.numeric(as.character(Deposit)) -0.0047928  0.0016759 -2.8599 0.004238 **
Fees0.015%                      -0.0536971  0.1715107 -0.3131 0.754218
Fees0.025%                      -0.2503217  0.1811425 -1.3819 0.167001
Financial_InstrumentCrypto       0.1156640  0.2100006  0.5508 0.581785
Financial_InstrumentETFs         1.1051717  0.2060692  5.3631 8.18e-08 ***
Financial_InstrumentStocks       0.6779814  0.2112362  3.2096 0.001329 **
Leverage YES                     0.1669098  0.1391029  1.1999 0.230178
Social_TradingYES                0.3185634  0.1440707  2.2112 0.027025 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -336.65
```

Figure 7: MNL model with Deposit as numerical

The results of this model are similar to the ones of the previous models and according to the likelihood ration test there is no substantial difference. Hence, this last model is preferred both in terms of interpretation and performance since a degree of freedom is gained. In addition, given as reference level a trading app which has as platform the desktop, zero fees, commodities as financial instrument and without leverage and social trading, we get that customers will respectively accept an increase in the deposit of:

- 108 euros for having the app in mobile version;
- 88 euros for having the app in web version;
- -11 euros for having fees at 0.015%;
- -52 euros for having fees at 0.025%;
- 24 euros for having criptos as financial instrument;
- 230 euros for having ETFs as financial instrument;
- 130 euros for having stocks as financial instrument
- 34 euros for the leverage
- 66 euros for having the possibility to do social trading

Thanks to this analysis it is possible to assess that the most valuable factors for the respondents are having ETFs as financial instrument, followed by stocks, and by mobile app.

6.2 Simulating Preference Shares

Besides the willingness-to-pay measure, another useful approach to assess the role of product attributes consists on using the model to obtain preference share predictions. This can be useful for firms when they have to launch new products on the market in order to have an **idea** of what could be the market share. To compute them have been used the product profiles defined in Figure 7. The

Platform <fctr>	Deposit <fctr>	Fees <fctr>	Financial_Instrument <fctr>	Leverage <fctr>	Social_Trading <fctr>
Mobile App	0	0%	ETFs	YES	YES
Desktop App	0	0.015%	Commodities	YES	NO
Desktop App	100	0.025%	Stocks	YES	NO
Mobile App	10	0%	Crypto	NO	NO
Web App	0	0%	Stocks	NO	YES
Mobile App	10	0%	Crypto	YES	YES

Figure 8: Product profiles used for computing preference shares

preference shares has been compute with respect to both the model with no intercept and the model with no intercept and deposit as quantitative variable. They present very similar results. The 1^{st} product has slightly more than 40% of the preference shares, and it is followed by the 5^{th} and the 6^{th} product respectively with 20% and 16% of the preference shares, the other alternatives have all values below the 10%.
However, it is always important to not over-read the results. Indeed, the previous preference shares do not reflect the real world and the real market ones.
Firstly, it is important to be aware that these predicted shares are relative to a specific given set of potential competitors.

In addition, it is important to not treat the obtained preference share predictions as actual market share forecasts. Indeed, these predictions represent the respondents' behavior in a survey context but they do not necessarily translate to actual sales in real marketplace. For example, it could be difficult for a customer to find in the market the trading app with all the characteristics that he/she likes the most. Moreover, it is always important to remember that customers may behave differently when they have to make a real-life decision that involve money. Nevertheless, these preference shares can be useful to companies in order to get at least partially understanding of the potential customers' preferences.

6.3 Sensitivity Chart

From a business perspective it could be useful to predict how the preference share for the planned product design would change if variations on the levels of the attributes were considered. The above **Sensitivity Chart** considers

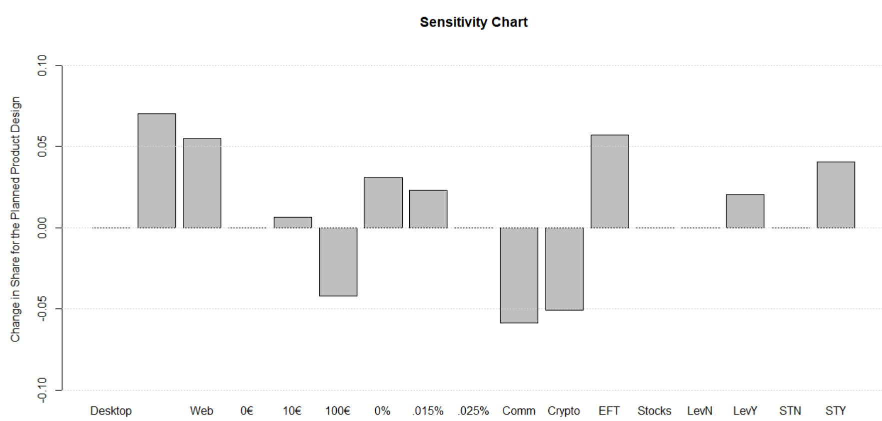

Figure 9: Sensitivity Chart

as reference level a trading app design with Desktop as platform, 0€ for the deposit, fees at 0.025%, Stocks as financial instrument and without leverage and social trading.
When a company analysis this data, there is always a trade off between costs and potential revenues. For example, realizing a product with specific attributes could be more expensive but there could be an increase in the preference share that counterbalance that costs.

In this context, a decrease in fees produces a small increase in preference shares, but loosing some customers could be accepted if it is payed off by the revenues related to the fees.

6.4 95% Confidence Interval

As a good practice, it is always better to have a *confidence interval* for the predictions instead of a unique values.

The same holds when companies compute market share. This is especially true for preference shares because they are an approximation of the real world and so are less reliable. Computing confidence intervals is better than having a

share <dbl>	2.5% <dbl>	97.5% <dbl>	Platform <fctr>	Deposit <fctr>	Fees <fctr>
0.12926122	0.06804879	0.1951184	Desktop App	0	0.025%
0.09487133	0.04874198	0.1617273	Desktop App	0	0.015%
0.09867644	0.05248849	0.1561170	Desktop App	100	0.025%
0.16880721	0.12181901	0.2121766	Mobile App	10	0%
0.23259777	0.17244152	0.3001983	Mobile App	10	0%
0.27578603	0.20264534	0.3618655	Mobile App	10	0%

Figure 10: 95% Confidence Interval

unique value because they give a less rigid idea of what could be the preference shares of customers.

6.5 Dealing with Customer Heterogeneity

Until now the analysis was based on the average part worths. However, it is important to consider that could exist some niches in the market and that customers can have heterogeneous preferences. Hence, it is reasonable to use the **mixed MNL** model in order to assess a unique coefficient to each respondent. This should increase the goodness of fit and provide more accurate preference share predictions that models with fixed effects. From the summary of the

random coefficients

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
PlatformMobile App	-Inf	-1.568507	16.3213796	16.3213796	34.2112657	Inf
PlatformWeb App	-Inf	-14.811451	13.8445840	13.8445840	42.5006194	Inf
Deposit10	-Inf	-32.881158	0.6286406	0.6286406	34.1384388	Inf
Deposit100	-Inf	-49.688927	-32.7087452	-32.7087452	-15.7285632	Inf
Fees0.015%	-Inf	-23.496831	-17.6536219	-17.6536219	-11.8104132	Inf
Fees0.025%	-Inf	-56.376243	-28.0750820	-28.0750820	0.2260789	Inf
Financial_InstrumentCrypto	-Inf	-185.934168	-33.6873334	-33.6873334	118.5595007	Inf
Financial_InstrumentETFs	-Inf	22.985297	89.5556515	89.5556515	156.1260064	Inf
Financial_InstrumentStocks	-Inf	-28.059641	49.1179004	49.1179004	126.2954417	Inf
Leverage YES	-Inf	-36.591643	6.3679914	6.3679914	49.3276261	Inf
Social_TradingYES	-Inf	-30.635155	24.4352045	24.4352045	79.5055636	Inf

Figure 11: Mixed MNL model results

mixed MNL model, it is possible to notice that the standard deviation for many variables is high meaning that there is *high heterogeneity in customer preferences.*

Moreover, by looking at the table 'random coefficients', which provides summary measures for each distribution of the individual-level coefficients it clearly visible that a change of sign occurs frequently. This is another sign of heterogeneity in customers preferences.

The variable with the largest distribution is definitely ‘Financial-InstrumentCrypto‘ followed by ‘Social-TradingYES‘ and ‘Leverage YES‘. Then it is possible to hypothesize that, among the respondents, there is a subset of them that feels more confident with finance and investments and thus one is more prone to use more complex financial instruments such as cryptos and in using the financial leverage in order to try to boost investments. With respect to ‘Social-TradingYES‘ it is possible to assume that the majority of people prefer to have it but other may be just indifferent and do not look for this kind of feature. Nevertheless, from a business point of view it is better to include this feature in the app because it is not associated to extra-costs and it is not mandatory to use. Thus, who likes this feature can use it. Another possible business strategy could be to include more financial products in order to satisfy a larger set customer.
However, before considering this results there is the need to assess if the output of this new model is reliable.

```
Likelihood ratio test

Model 1: choice ~ Platform + Deposit + Fees + Financial_Instrument + Leverage +
    Social_Trading | -1
Model 2: choice ~ Platform + Deposit + Fees + Financial_Instrument + Leverage +
    Social_Trading | -1
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  11 -336.45
2  22 -324.86 11 23.184    0.01665 *
```

Figure 12: Likelihood ration test MNL vs Mixed MNL

According to the likelihood ratio test the mixed MNL specification significantly reduce the goodness of fit.
One last improvement to the model could be obtained by allowing random coefficients to be correlated. This permits to assess whether customers who favor one attribute also tend to favor other attributes.

Unfortunately, there is no improvement in the model which is actually really complex. In addition, the only correlation that is significant according to the model is difficult to interpret. For this reasons, it is better to no to draw hazardous conclusion also because the t-test shows that there is a reduction in the goodness of fit.

6.6 Respondent level variable and customer heterogeneity

During the survey, demographic data has also been collected. Unfortunately, the sample group seems to be quite homogeneous with respect to education level, nationality, job sector and age, there are just few outliers. However, one discriminant factor could be the gender. Hence, it could be interesting to notice if the differences in the preferences could be related to the gender. For example, according to the literature on risk perception men seem to be less risk averse

compared to women. This could partially explain some of the results related to the preference of the financial product.
However, the above assumption has been rejected because there is no substantial difference among males and females. Hence, this model does not provide useful information.

7 Results' practical implications

To conclude, the results obtained through the MNL model without intercept and with deposit as quantitative variable are the most reliable to detect respondents' preferences.

The results showed that the financial instrument is definitely the most important attribute. Moreover, ETFs seem to be the preferred option, followed by Stocks.

This results is expected since the financial instrument is the main driver. It is interesting tough that people prefer ETFs over the others. It can be hypothesized that this is the preferred option because with ETFs it possible to build a diversified portfolio with relatively low investment amounts.

The platform seems to play a quite relevant role as well. Indeed, respodents chose the mobile app as the preferred option on average.

The deposit is another important attribute. The results show that respondent prefer to have lower deposit. This is quite reasonable. Indeed, it is possible to assume that people prefer to start investing with low amounts of money probably because they do not completely trust this kind of tool for investing.
From a business point of view, it is interesting stating again that customers seem to be indifferent to fees, probably due to their low levels. This means that a company that is creating a trading app can set higher fees without significantly loosing market shares.
Nevertheless, it is important to state that the analysis has serious limits. First of all, the sample of respondents is not completely random and it is definitely not big enough. Indeed, it would be interesting to extend the analysis to a broader class of respondents who are interested in this kind of products to get more accurate and reliable results. In addition, although the presence of the glossary, the respondents could be newbies in the field and so the answers that they give do not properly represent the ones that actual potential customers could give.
In addition, some of the assumptions that have been made on the trading app' attributes do not reflect the real world. For example, trading apps could offer more than one financial instrument. The same holds for the kind of platform, because companies can create multi-platform apps without many costs.
However, this analysis still gives a first understanding of which could be customers preferences relatively to this new products.

8 Appendix

For further references about the analysis see Github Repository