Bayesian truth serum fused conjoint

Master thesis

Grigor Dimitrov

A thesis presented for the degree of
Master of science

Department of Business and Economics
Erasmus School of Economics
Rotterdam, The Netherlands
29 August, 2017"

Supervisor: Chen Li
First-reader: Aurelien Baillon

Abstract

"Conjoint methodology has been widely used in marketing research to answer a variety of business questions. However, it is usually performed in hypothetical settings, which raised questions by both practitioners as well as academicians on its validity. Recent studies investigated whether implementing incentive-compatible schemes would improve conjoint analysis predictive power. The results have been positive so far, but there are certain disadvantages associated with each approach. In this paper I will explore a relatively new incentive-compatible mechanism, the Bayesian truth serum. The method is easy to be implemented and it does not change the conjoint methodology. The method assigns high scores to respondents when they have surprisingly common answers, e.g when the frequency of their answer is higher than commonly predicted by the rest of the respondents. The Bayesian truth serum model outperformed the control group on several key metrics providing first plausible results for the adoption of the method into the preference measurement methodologies."

Contents

Introduction

In economics, often the decision-making process relies both on ‘soft factors’ and ‘hard numbers’ (Baillon 2015). The hard numbers are objective quantitative measurements, while the soft factors refer to beliefs, opinions, expectations, etc. that are subjective. Subjective data is heavily used to understand individuals’ behavior, attitude, intentions and choices in many different domains, including governmental institutions and businesses Prelec (2004). One specific domain is opinion-taking for market research purposes, which again relies heavily on subjective data gathered through surveys from panels of respondents. However, subjective data reliability depends on “its quality at the source - the thought process of an individual respondent” (Prelec 2004). To provide solid and credible recommendations to businesses, market researchers, who use surveys as a proxy of their decisions, crucially depend on respondents’ trustworthiness.

One of the widely used research methodologies for the past four decades is the conjoin methodology. It has been employed in many domains such as market research, policy making, exploring optimal pricing of products, new product development, segmentation, product positioning, etc. (Ding, Grewal, and Liechty 2005). It is a broad methodology for presenting and analyzing products, services or policies. Using conjoint analysis researchers are able to elicit respondents’ utilities of different characteristics of a product, service or policy. Those utilities are estimated based on data from respondents’ simultaneous evaluation of components configurations across different attributes levels of a product, service or policy, also referred to as stimuli (Green, Krieger, and Wind 2001; Green and Srinivasan 1978). Essentially, respondents are making trade-offs within a given set of product characteristics and state their preferences for the different stimuli. In the estimation process researchers are able to elicit respondents’ utilities based on those trade-offs on an individual level as well as aggregating the utilities in the sample.

Data collection is typically done in hypothetical settings, however, there has been a growing concern amongst market research practitioners regarding the usage of hypothetical approaches due to insufficient confidence in respondents’ truthful answers’ (Gittelman and Trimarchi 2012; Puleston and Eggers 2012; Toubia et al. 2012). Such settings do not provide respondents sufficient incentives to carefully and honestly indicate their preferences (Ding, Grewal, and Liechty 2005), respondents are not given neither incentive nor penalty to truthfully state their preferences. Hence, there could be a discrepancy between respondents’ stated preferences and their real-life behavior Ding, Grewal, and Liechty (2005). In an attempt to account for those concerns, the academia has been exploring the application of incentive-compatible mechanisms within the preference measurement techniques, including conjoint analysis (Ding, Grewal, and Liechty 2005; Ding 2007; Dong, Ding, and Huber 2010; Toubia et al. 2012). Ding, Grewal, and Liechty (2005) rewarded respondents with their actual choice at the end of the conjoint exercise, showing that this incentive-compatible mechanism increases the external validity of the data. Furthermore, Toubia et al. (2012) successfully implemented incentive-compatible scheme within a choice-based conjoint exercise that showed promising results on respondents’ engagement . However, there are three main disadvantages related to using those methods. First, when it comes to new product development, as in Ding, Grewal, and Liechty (2005), rewarding respondents with their actual choice is infeasible. Second, from an implementation standpoint, using Toubia et al. (2012) method requires a complex altering of the conjoint methodology in order to apply in the incentive compatibility. Third, from respondent standpoint again in Toubia et al. (2012) method, requires respondents understanding of the "conjoint-poker" rules in order to maximize their pay-off rather than just making choices between alternatives. I introduce an application of an incentive-compatible methodology which can account for those disadvantages - Bayesian truth serum (Prelec 2004). Bayesian truth serum can be applied in a hypothetical setting, it doesn't alter the conjoint methodology and it is relatively easier to be understood by the respondents.

In 2004, Drazen Prelec introduces a new incentive-compatible mechanism called the Bayesian truth serum (hereafter BTS) Prelec (2004). Essentially BTS is a method for eliciting respondents' subjective answers and meta-knowledge (i.e. knowledge of what other people know). Prelec argues that under the BTS condition respondents are acting as if they are being evaluated by an "omni-center in possession of the truth". The method works in an absence of verifiable truth. The method is based on a scoring system that takes into account respondents' answers as well as their probabilistic predictions about the distribution of answers within a given sample. BTS rewards truthful answers by assigning them high scores which will accordingly result in higher payoffs. The high scores are computed using both respondent's personal answer and their prediction about the distribution of others' answers. More precisely, high scores are given to an answer whose actual frequency is higher than its predicted frequency. Such an answer is referred to as "surprisingly common". For instance, if an answer is endorsed by 25% of the sample and its predicted frequency obtained from the same sample is 20%, it will be regarded as "surprisingly common" and given a high score. The method awards truthful responses for both rare and widely shared answers e.g. it does not depend on consensus as a norm for truth-telling. The method doesn't require respondents to understand any details of the scoring rule apart from the fact that it is being applied, making it easier to be applied in practice. As BTS is a relatively new methodology, it has not been employed widely, however, the results so far have been positive in terms of its validity (Weaver and Prelec 2013; John, Loewenstein, and Prelec 2012)

In this paper, I will adapt BTS to provide incentives for respondents to reveal their true preferences in a conjoint analysis of a new service development. Golbeck and Mauriello (2016) found that on average users of social media applications, such as Facebook, are concerned about their privacy and are not well informed regarding the access of data that are given to third party applications. At the same time, users consciously neglect the terms and conditions related to data privacy on different websites. Given the fact that data privacy is a sensitive topic and at the same time users tend to ignore the general terms and conditions of a privacy policy, respondents may give socially desirable answers and misreport their true preferences Krumpal (2013). In this study, I will employ BTS to evaluate respondents' preferences towards a new hypothetical service that aims to buy out individuals' online private data. I will jointly evaluate the effect of several factors on the probability of individuals to enroll in that service. The factors, also the main components of this service, are the following: first, devices to which individuals would be willing to grant access to; second, the extent of the collected data; third, the monetary amount individuals are willing to accept in order to share their data. Conjoint analysis is dependent on the truthfulness of the subjective data, thus creating incentives for "truthful" responses with BTS could potentially reduce problems in regards to truthful revelation and ultimately improve the data quality.

The motivation of this paper lays in the improvement of conjoint analysis reliability with an application of an incentive-compatible methodology, the BTS. The main objective of the paper is to explore whether the BTS can prove itself to be a superior truthful elicitation technique in comparison to the regular hypothetical conjoint setting. Furthermore, I will explore whether the BTS can prove itself to be a reliable method, applicable to conjoint analysis when truthful answers are not observed during the time of the research.

The main research question is as follows:

> Can the BTS improve the choice based conjoint data quality?

To my best knowledge, no research has explored improving the reliability of choice based conjoint utilizing the BTS. One of the research's contributions research will be to complement the emerging stream of literature exploring how the external validity of conjoint analysis can be improved using an incentive-compatible method. Moreover, the paper will contribute to BTS's application in preference measurement methodologies.

The paper is structured as follows. Firstly, the literature review section introduces in details conjoin analysis and BTS. I will additionally explore current limitations concerning online data collection. Based on the literature review section, I will develop my hypotheses. Secondly, in the methodology section, I will describe the data collection and analysis procedures regarding conjoint analysis and the BTS scoring system. Next, the result section describes a detailed analysis of the main results and findings. Finally, I will discuss the results, the main contributions and implications for both practitioners and academia and the main limitations of the study.

Literature review

Conjoint analysis

Origins of conjoint methodology and analysis

Conjoint analysis has been used for circa 44 years and it is still evolving. It was introduced to the academia by Green and Rao (1971) in their paper "Conjoint Measurement for Quantifying Judgmental Data". Shortly afterward, it was adopted by practitioners in diverse areas such as new product development, pricing, segmentation and positioning (Ding, Grewal, and Liechty 2005). Initially, the theory was referred to as conjoint measurement, but once it was appreciated by practitioners and transformed to evolving research stream by scientists, the term conjoint analysis (hereafter "CA") was adopted (Hauser and Rao 2004). Nowadays, the term CA is used both as a reference to the theory behind measuring judgmental data as well as for different methods used for this purpose Louviere (1988). In the next sections, I will introduce a brief description of the theory behind CA followed by the introduction in CA design. As this paper will employ choice based conjoint (hereafter “CBC”) the CA design section will be focused on CBC.

Multidimensional scaling, judgmental data and choice data

CA originates from mathematical psychology (Green and Srinivasan 1978) and it faces the mapping challenge of multidimensional scaling (MDS) to “translate a point from perceptual space into a corresponding point(s) of product feature space” (Hauser and Rao 2004). More precisely, the mapping challenge emerged from psychometrics, explored by Krantz et al. (1971) in their influential book “Foundations of measurement”, where the authors investigated behavioral axioms that would make possible decomposition of an overall judgment.

Hauser and Rao (2004) explain that essentially CA is a method for MDS and clustering applied to marketing questions. Paul Green uses MDS to decompose overall consumers’ preferences and perceptions toward a certain product into a partial contribution of product’s features (also called “part-worths”). Quoting Green and Srinivasan (1978) seems the most appropriate to explain the name of this elegant theory " . . . applications by psychometricians and consumer researchers have emphasized the scaling aspects - finding specific numerical scale values, assuming that a particular composition rule applies, possibly with some error. Accordingly, it now seems useful to adopt the name, conjoint analysis, to cover models and techniques that emphasize the transformation of subjective responses into estimated parameters“. Louviere (1988) defines judgmental data measurement as respondents’ evaluation of a product profiles represented by a set of alternatives of products’ attributes. The data is usually in at least ordinal scale resulted from rating or ranking the profiles. In the same paper, the author defines choice data as respondent’s preference of a particular profile among a given set of profiles. Profiles are again represented by a composition of product attributes. According to Louviere (1988), the terms judgmental and choice data have been used as mutually replaceable. However, there is an important distinction between judgmental and choice data, namely, judgment data may not satisfy important assumptions that are needed to forecast choice behavior.

Choice based conjoint was proposed by Louviere and Woodworth (1983) as an extension of the traditional conjoint analysis. The authors introduced CBC following by the recent development in the choice behavior modeling, namely, the introduction of multinomial and conditional logit models by McFadden and others (1973). CBC constitutes of a choice evaluation between profiles. It allows for interaction and cross effects between different attributes (Chrzan and Orme 2000). Moreover, with CBC it is possible to measure preferences at an individual level and it represents a realistic choice task in which respondents are required to make a certain trade-off.

Designing a choice-based conjoint

There are several critical steps involved in designing conjoint experiments that are common for all CA (Green and Srinivasan 1978; Green and Srinivasan 1990; Hauser and Rao 2004).

1. Decomposition of product on product attributes
2. Stimuli representation
3. Stimuli configuration
4. Data collection
5. Estimation

Due to the broadness of the term CA, it is difficult to assign a meaningful explanation of each step without focusing on a particular CA approach. Although the first 2 steps are common for all types of CA, stimuli configuration, data collection, and estimation tasks are interlinked. Louviere (1988) refers to the first three common steps as experimental design techniques and explains that researchers rely on them to create profile configuration. However, the data collection and estimation vary per different "paradigms" (or approaches used above). For example, collection and employing judgmental data implies evaluation of one profile at a time by rating it on an ordinal scale, which is referred to as "Rating-based conjoint", while choice data implies a choice of one profile among several at a time mimicking real purchase behaviour which is referred to as "Choice-based conjoint". Depending on the different data collection techniques, the method of analysis and estimation of the "part-worth" utilities differs. The focus of data collection and estimation will be on CBC while briefly describing the other possible approaches.

Decomposition of product on product attributes

There are 2 critical issues that should be taken into account in this first step (Hauser and Rao 2004). First, defining product's attributes and respectively the levels in which each attribute is divided. Second, defining the function of each attribute, respectively each attribute level, in the overall model of respondents' preferences. Functions can be: vector monotone preference (approximately linear), ideal point preference (convex or concave) and part-worth preference (discrete or categorical). In contemporary literature on the topic, all preference functions are referred to as "part-worths", despite the fact that they may have different functions in the overall preference structure according to earlier articles.

In order to illustrate the steps better, I will use an example. A company X is planning to release a new smartphone on the market. After a preliminary research, company X decides that the most important features of their smartphone are: Product name, Screen size, Memory and Price. After consultation with engineering and marketing departments, the following product attribute levels can be distinguished:

Table 1

If we consider independently price and memory, they would have approximately linear utility for the respondents (i.e. the higher the price, the lower utility). However, considering the screen size, we can expect a concave shape of the preference (i.e. most people would probably prefer neither too small nor too big phone). As brand name is a discrete preference, it belongs to the so-called part worth preferences.

Stimuli representation

There are many ways that stimuli can be presented to subjects such as text, pictures, video, physical objects and even verbal description. However, as in the recent years, the established way of conducting marketing research interviews is in online web-based forms, researchers are taking advantage of various kinds of available multimedia (Dahan and Srinivasan 2000). Naturally, the more realistic the stimuli are presented to the subjects the better quality of the data is to be expected.

Stimuli configuration

Once the product is decomposed on its attributes and attributes' levels, a large number of product profiles can be generated per each combination. However, ever since the emergence of CA, researchers recognized that showing all possible products profiles is not feasible as it is tiresome for the respondents and therefore, it is not cost effective and can potentially lead to biased estimates (Carmone, Green, and Jain 1978).

To limit respondent's burden and provide more robust estimates, several approaches to reduce product profiles can be applied. However, product profile configuration is closely related to the data collection method and estimations, thus, collection and estimation cannot be examined independently. For example, one way to reduce product profiles is orthogonal fractional factorial design or just "orthogonal design". In an orthogonal design, only the extreme combinations of products' features are being evaluated, "the levels of the features are chosen such that, for each pair of features, say a and b, the high level a appear equally often in profiles that have a high-level b as in profiles that have a low level of b, and vice versa" (Hauser 2007). In a demonstration paper, Green (1974) applied orthogonal design to an extreme example of 4x3x27 (1536) product profiles, was able to reduce the product profiles to only 16 in order to estimate the main effects. As expected, such reduction comes with certain trade-offs, that is, preference independence assumption, which does not allow any interactions among the different attributes to be estimated.

Table 2 Orthogonal design example

Card ID	Brand Name	Screen Size	Memory	Price
1	ProductA	4'	32	300
2	ProductB	3'	32	400
3	ProductB	5'	32	200
4	ProductB	6'	64	300
5	ProductA	5'	16	300
6	ProductB	5'	16	200
7	ProductB	6'	16	400
8	ProductA	5'	64	400
9	ProductA	3'	16	200
10	ProductA	6'	32	200
11	ProductB	3'	16	300

Card ID	Brand Name	Screen Size	Memory	Price
12	ProductA	6’	16	200
13	ProductA	3’	64	200
14	ProductB	4’	64	200
15	ProductB	4’	16	200
16	ProductA	4’	16	400

For instance, our smartphone example would generate a total of 72 (2x4x3x3) possible product profiles. As seen in Table 2, using orthogonal design we can reduce this number to 16. Orthogonal design, full factorial desigh and fractional factorial design are the starting point for more advanced stimuli configurations that are used in practice nowadays.

Data collection

The methods described below are proven to be reliable in large scale commercial applications (see Hauser and Rao (2004) for discussion).

- Partial profile CA

Also known as Two-attribute-at-a-time trade-off analysis, in this setting respondents are presented to two attribute trade-off tables where they rank their preferences for different attribute levels. The number of stimuli can vary from two to many, nonetheless, two attributes at a time are the most popular (Hauser and Rao 2004).

- Full-Profile CA

Products are described by the levels of the attributes they contain. Respondents are then asked to rank order all stimuli or to provide metric rating of each stimulus. By controlling the attribute combinations, based on correlations of the attributes with respondents’ preferences, the utility for each level of each attribute tested can be estimated.

- Choice-Based (CBC)/Discrete-Choice CA

Respondents are presented to a given set of product profiles, once they make their choice a new set is presented. The utilities are estimated based on respondents’ choices. Also known as stated preference models, they are based on random utility models (RUM). Respondent’s utility is presented as combination of product’s features and an error term. Based on assumption about the distribution of the error term, the probability of purchase of a product assembled out of a given set of features can be estimated.

- Adaptive CA

As currently the most preferred way of data collection is computer based interviews, this enables researcher to make computations during the interview itself and adapt consecutive questions based on utilities estimated during the initial stage of the exercise. This method allows for reducing the survey length without losing valuable information or diminishing the power of the analysis.

Estimation

- Regression based models

If the data obtained from the respondents for their preferences is in interval scale, then the partworths can be represented by dummy variables and ordinary least square (OLS) estimation can be adopted.

- Random utility models (RUM)

If the data obtained is from CBC, RUM can be adopted. The assumption of utility maximization combined with distributional assumptions of the unobserved term states that there is "a known function that maps the partworth levels onto the probabilities that each profile is chosen from a given choice set" Hauser and Rao (2004).

- Hierarchical Bayes Estimation (HB)

In order to improve the predictive power of CA, researchers approached the problem using HB estimation. However, using HB slightly changes the idea of the analysis i.e. researchers do not try to estimate the pathworths, but rather attempt to characterize the uncertainty behind them (Hauser and Rao 2004); e.g. the HB estimation allows for more accurate results while showing respondents fewer choice tasks.

Conjoint analysis has been widely used in many different domains in market research, allowing businesses to optimize their product development, pricing strategy, find the optimal product or portfolio configuration, etc. However, in a marker research workplace, typically the conjoint exercise in administered in hypothetical environment. In such an environment, respondents could potentially give 'hypothetical answers' (Scott 1965) as respondents are not offer a strong incentive to carefully and thoughtfully think about their answers (Ding, Grewal, and Liechty 2005). Recently, researchers have examined different incentive-compatible methods that could measure better respondents' true preferences in conjoint analysis (Ding, Grewal, and Liechty 2005; Ding 2007; Toubia et al. 2012). In the next section, I will focus on specific problems of online opinion taking and how incentive-compatible methodologies can improve model performance while at the same time increasing respondents' engagement and motivation.

Problems of online opinion taking

Practitioners have defined several systematic behaviors which potentially lead to low quality data in online data collection (Puleston and Eggers 2012, Gittelman and Trimarchi (2012)). Namely, cross-cultural factors, untruthfulness, speeding, question format and sample quality have been identified to have a negative impact on data quality, thus, causing noise in the data rather than variance. One of the major limitations of online opinion taking, and in particular preference measurement is the lack of motivation by respondents (Toubia et al. 2012), which can result in reduced respondents engagement, hence untruthful answers. For the purpose of this paper I will further elaborate on respondents' untruthfulness, engagement and motivation.

The hypothetical nature in which typically conjoint studies are executed questions the extent to which they are able to reveal the true consumer preferences (Ding, Grewal, and Liechty 2005). When respondents are not incentivized to reveal their true preferences and the data is collected in hypothetical settings, weaker external validity is observed (Ding, Grewal, and Liechty 2005). Respondents have little or no incentive to reveal their true preference or to be mindful about their answers in hypothetical settings (Noussair, Trautmann, and Kuilen 2014). Economic theory suggests that if economic agents believe their answers will have an impact on decisions made by businesses or governments for outcomes that the agents care about, they should respond in such a way as to maximize their payoffs and welfare (Carson and Groves 2007). However, due to the hypothetical nature of the surveys, in most conjoint studies respondents may misinterpret their preferences and the study may not yield truthful answers. Though hypothetical techniques are found to perform fine in many domains (Guiso, Jappelli, and Terlizzese 1992; Carman and Kooreman 2010; Guiso and Parigi 1999; Hurd 2009), recent conjoint studies reveal that incentive-aligned conjoint is superior than hypothetical conjoint in out-of-sample prediction, and it is more engaging and motivating (Ding, Grewal, and Liechty 2005; Ding 2007; Toubia et al. 2012). Typically, practitioners apply different rules of identifying "bad quality" respondents and eliminating them when analyzing the data, such as using misdirect questions (or traps). For example, Gittelman and Trimarchi (2012) recognize untruthfulness by asking respondents several questions with an extremely low incidence rate, and ask questions that have known distribution across the population. The authors argue that based on the score estimated from these questions, it is possible for researchers to screen out respondents and thus, remove respondents that do not have carefully considered answers. Speeding is also a quality measure often applied to achieve the necessary data quality. Puleston and Eggers (2012) claim that 85% of the respondents are speeding at some point of the questionnaires, which make it impossible to exclude these responses from the survey. The authors argue that speeding is an "epidemic" in online survey world and it is a generic issue that holds the potential to cause the biggest amount of noise in the data. In their paper, the authors divided the sample in two categories, above and below the average length of the interview. The analysis of comparison of both groups showed that respondents in the group below average times account for +/-6.2% higher variance on average. As expected, with progression of the survey the variance increases from 4% at the beginning to 8% at the end. The authors do not bind themselves to the claim that this variance is a source of noise in the data, however, one could speculate that the consequence of speeding is low quality data.

The recent technological advancements allowed for increased accessibility to respondents (Netzer et al. 2008). But this comes with a trade-off, as there has been decreased attentiveness and patience as respondents are easily distracted. With the aid of technology, there have been efforts to engage respondents with the task using improved interfaces, survey designs, question formats and even game like environments (see Netzer et al. (2008) for discussion). Hence, practitioners have attempted to tackle the disengagement by incorporating intuitive mechanisms in their survey designs. These mechanisms indicate only "suggestive evidence" of underlying issues and determination that respondents are “behaving badly”. Despite those attempts, the biggest source of bad data is still the disengaged respondent.

The academia also recognized the lack of motivation as a key limitation concerning preference measurement methods (Toubia et al. 2012; Ding, Grewal, and Liechty 2005; Netzer et al. 2008; Liechty, Fong, and DeSarbo 2005). However, the academia has investigated the source of these problems in order to provide solid solutions. Respondents are putting significantly less cognitive efforts when making survey choices compared to real life decisions, which leads to poor out of sample predictions (Toubia et al. 2012). To increase the level of involvement, based on insights from experimental economics, some scholars approached preference measurement issues applying incentive-compatible mechanisms. Experimental economists claim that using incentive compatible methods should induce truth telling as monetary incentives dominate over other factors. Hence, respondents are encouraged to reveal their true preferences as to maximize their final payoff (Smith 1982). In their Induced value theory, Smith and Parker (1976) define 5 percepts, 3 of which are sufficient conditions for incentive compatible behavior. The first principle is referred to as saliency, in other words participants’ final monetary payoff should depend on their performance, where better performance leads to better outcome. The saliency principle is also related to the nonsatiation principle, claiming that “utility is a monotone increasing function of monetary reward”, or simply put “the more the better”. The third principle is dominance. To assure control over the preference, the final payoff for the exerted cognitive efforts should dominate the subjective costs of participation. When those conditions are met, incentive-compatible methods should yield truthful answers.

Recent studies on adopting incentive-compatible mechanism to conjoint have shown that incentive-compatible mechanisms can result in boosting respondents’ motivation leading to respondents providing truthful answers.

Incentive aligned conjoint

A general concern amongst practitioners when using conjoint methodologies is that respondents are not motivated during the survey, and that choices made during the conjoint exercise differ from real-life purchasing choices (Toubia et al. 2012). Ding, Grewal, and Liechty (2005) compared the out-of-sample predictions for two conjoin designs; the first adopted the traditional hypothetical setting, while the second used Chinese dinner special as a context in which respondents had a positive chance of receiving a preferred alternative from each choice set in the end of the experiment. Results showed that the external validity is increased compared to a setting where respondents hypothetically provide answers for each of the choice tasks.

The authors find the most relevant principle to conjoint analysis to be saliency. In practice respondents are given a flat participation fee, which doesn’t induce honest responses. Hence, saliency hasn’t been satisfied as respondents are neither rewarded nor penalized based on their performance. Providing respondents with a flat participation fee cannot motivate them enough to carefully evaluate every given choice in the conjoint surveys, resulting in noise in the data. Hence, there is no reason for researchers to expect that predictions based on the survey data from non-salient responses will make valid out-of-sample predictions. In other words, respondents' stated preference will not be consistent with their revealed preference (Ding, Grewal, and Liechty 2005). The main disadvantage of this experiment, however, is that the researcher should have available all alternatives from each choice set.

Ding (2007) created a mechanism that does not require changes in the existing conjoint methodologies and allows for situations in which the researcher has available a few alternatives of the product. Respondents are offered the available alternatives as potential rewards. The method incorporates BDM (Becker, DeGroot, and Marschak 1964) procedure to compare subjects' valuations to a randomly generated price, the optimal strategy for a participant being to state his or hers true WTP. If the bid is higher, the subject pays the randomly generated price and gets the item. If the bid is lower than the elicited price, the subject doesn't pay and receives nothing. The method works in four steps. First, participants complete a standard conjoint exercise. Afterwards, the experimenter shows them a real product that they could potentially purchase. By not revealing the product in hand prior the conjoint exercise (as it is of limited availability), respondents' preferences are not biased. After the product if being shown, the interred WTP is calculated. Finally, using BDM it is determined whether a participant can purchase the real product or not. The drawback of this method, however, is that by using the BDM mechanism, respondents have the change to purchase the real product with the majority of the amount of money coming from the difference between the inferred price and the randomly drawn price. Hence, there might be a difference when in real life respondents need to use their own money. Moreover, the design cannot be applied to new goods as it requires an alternative to be present when performing the study and one of the main applications of conjoint analysis is for new product development.

Park, Ding, and Rao (2008), proposed an incentive-aligned conjoint method that allows respondents to update each attribute and configure their own product. When upgrading attributes, respondents state their willingness to pay per upgrade. The BDM mechanism is adopted in here as well, as to ensure that the best option for respondents is to truthfully state their willingness to pay. At the end of the experiment, respondents receive their upgraded product. The authors found that the method can significantly outperform the standard conjoint study. The drawback of this study is that again, researchers need to have the product available.

Toubia et al. (2012) argued that those methods "may not increase involvement to a level of real-life purchasing decisions". The reasoning is that respondents may behave differently when obtaining the product requires them to use own money instead of receiving a product as a reward at the end of the experiment. In such a case, respondents will pay more attention to certain attributes relevant to them in real life, but not to such an extent during the conjoint exercise, where there is a probability of obtaining the tested product. Hence, Toubia et al. (2012) proposed the "conjoin poker", which is an incentive-compatible conjoint study that collects preference data in a game setting in an attempt to increase respondents' involvement and attention. The study design results in increased respondent engagement and motivation, which was measured by the time they spend in the survey and by a follow-up questionnaire. Generally, the study design was found engaging and entertaining, but also complex and time-consuming.

The issue that still stands in order to comply with the saliency principle is how to evaluate respondents' performance in an absence of objective truth. In 2004 Drazen Prelec proposed a solution which he called Bayesian truth serum. This paper main objective will be to test if the BTS can help motivate respondents to truthfully reveal their preferences and improve the quality of the data. The method works in hypothetical settings i.e. in new product development context. It doesn't alter the experimental design to an extent such as the conjoint poker. It is relatively easy for understanding by researchers and respondents. The following section described BTS intuition, assumptions, and applications.

Revealing truthful preference with BTS

In his work on truth-telling incentives, (Prelec 2004) recognizes another source of bias that can result in low quality data. He suggests that in the absence of "external criteria" people can become subjects of "self-deception and false confidence even among the well-intentioned". In order to express the notion that some experts are never subject of reality checks, Prelec compares subjective judgment of business investors with art critic. Both experts are making subjective judgments but the end result of the investor can be used for evaluation of his judgments (i.e. reality check) while for the art critic there are no external criteria for proper assessment.

Bayesian truth serum

Introduction

Bayesian truth serum is a methodology for eliciting subjective judgments and "unverifiable truths", developed by an MIT professor Prelec (2004). Prelec recognizes the necessity of assessing subjective beliefs in absence of objective truth as opinions, attitudes, and intentions are often used in science and policy making.

For example, in marketing research, companies have been using respondents' subjective beliefs in order to create policies about products, prices, packages, features etc. Prelec recognizes that the quality of the subjective data is limited by "its quality of the source" and argues that if respondents act as if they are being evaluated by an "onmicenter scorer in possession of the truth" the data quality will be enhanced. The mechanisms of BTS methodology grants maximum incentives for truth telling based on Bayesian reasoning. Scoring system

BTS is a scoring system that assigns different scores to opinions based on their truthfulness where high scores imply high incentives. BTS consists of two components: information score and prediction score.

Information score intuition

Information score is estimated on question level and assigns high values to common answers and low values to the opposite. Surprisingly common answers are considered those whose actual frequency is higher than collectively predicted by the same population of interest (Prelec 2004). Once considered an irrational bias, afterward proven as a purely rational effect, the idea behind common/uncommon answers is that opinion holders would overestimate the frequency of their own opinion amongst the population. Ross, Greene, and House (1977) describe this effect as: one "tend to perceive a"false consensus"-to see their own behavioral choices and judgments as relatively common and appropriate to existing circumstances while viewing alternative responses as uncommon, deviant, or inappropriate." More precisely, it is a systematic deviation of assigning higher estimations or probabilities with regards to population frequency in directions of one's own opinion in a group of which the subject is a member.

In his paper, Dawes (1989) referred to the "false consensus bias" as "an egocentric bias to overestimate the degree in which others are like us". Ross, Greene, and House (1977) executed four studies in which they tested the "false consensus bias" in various settings such as behavioral choices, situations, and judgments as well as personal problems, expectations preferences and characteristics. For example, in a study among Stanford undergraduates Ross, Greene, and House (1977) showed that students who think about dying expected that 44% of students, in general, would share their problem against students who don't think about dying expected only 25.6% of the students to think about dying. Within the same experiment subjects who believed to die before turning age of 70 indicated that 57.6% of the students share their expectations while this number was only 43.9% for students with the opposite expectation.

This pattern was considered to be an irrational systematic deviation before Dawes (1989) has proven it to be a rational effect. Using Bayesian reasoning Dawes argues that subjects are considering their own opinion to be "informative sample of one" which shifts the estimation of endorsers of certain opinion in a direction toward their own opinion.

Using Dawes' interpretation Prelec (2004) constructed a critical assumption; the information score component of BTS which indicates that ones' opinion frequency will be underestimated by the others and vice versa, one will overestimate the frequency of his own opinion. Consequently, this implies that surprisingly common answers are those whose actual frequency overweighs the collectively predicted frequency. For example, let's consider we have data about respondents being asked to state their preferences at binary choice (e.g. Yes/No) question. The respondents are asked to provide their own opinion as well as their estimates about the fraction of people endorsing their answer. If the actual frequency is 10% per "Yes" and the predicted frequency is 5%, the respondents who indicated "Yes" will receive a positive score as their answer will be considered "surprisingly common". However, given the same actual frequency, if the collectively predicted frequency is 25% respondents will be penalized with a negative score as their answers will be considered uncommon.

Information score formula

Prelec (2004) defines the following formula for estimation of information score component for answer k:

$$\text{information score} = log(\frac{\overline{x}_k}{\overline{y}_k})$$

Each respondent r information score per k number of answers:

$$\text{information score} = \sum_k x_k^r log(\frac{\overline{x}_k}{\overline{y}_k})$$

Where:

- $\overline{x}_k$ is the actual frequency of answer k or $\overline{x}_k = \frac{1}{n}\sum_{r=1}^{n} x_k^r$
- $\overline{y}_k$ is the geometric average of predicted frequencies for answer k or $\overline{y}_k = (\prod_{i=1}^{n} y_k^r)^{\frac{1}{n}}$

Prediction score intuition

The second component of BTS scoring is the prediction score, which is "a penalty proportional to the relative entropy (Kullback-Leibler divergence) between the empirical distribution and respondent's prediction of that distribution" Prelec (2004). In other words, this score assigns zero value to accurate prediction (i.e. best prediction score) when one's prediction matches the actual frequency and penalizes for inaccurate one. It is an asymmetric estimate of the difference between two probability distributions which produces number smaller than zero.

Prediction score formula

$$\text{prediction score} = \alpha \cdot \sum_k \overline{x}_k \cdot log\frac{y_k^r}{\overline{x}_k}$$

α is a constant used for assigning different weight to the prediction score relative to the information score in the BTS final score. Setting α equal to 1 will assign equal weights to score of the equation which will result in zero sum game. In the zero sum game, the average of all BTS scores is equal to zero. When α approaches to zero BTS game results in "Pareto-dominate expected scores" in which at least one individual has a positive score.

BTS formula

The overall score for each respondent is:

$$\text{BTS score} = \text{Information score} + \text{Prediction score}$$

BTS Assumptions

Prelec (2004) relies on several assumptions for constructing BTS algorithm:

Bayesian Nash equilibrium

"Truth-telling is individually and collectively optimal" Prelec (2004). Taking into account the BTS assumptions, agents stating their true preferences is BTS Nash equilibrium for any Î±>0. Thus, it is individually optimal as BTS Nash equilibrium results in a maximum expected payoff. It is collectively optimal because there is no other equilibrium which results in higher information score for any agent. The equilibrium relyes on two assumptions:

1. There a large enough sample so that a single asnwer can not affect the overall distribution;
2. All agents are Bayesians. They have a common prior and they update their beliefs based on their opinions;

BTS validity

Although the BTS is a relatively easy method that can be implemented within a study design, little has been done to assess its validity. So far, the results are positive regarding the usage of the approach.

In an attempt to eliminate hypothetical bias in contingent valuation (CV) (Barrage and Lee 2010) compared three mechanisms to elicit subjects' true preferences along with real and hypothetical CV in two settings. The results showed that BTS was able to eliminate the hypothetical bias in one of the settings but only reduce it in the other. The authors conclude that BTS eliminates the hypothetical bias inconsistently. Interestingly enough, the authors also found an interaction between BTS and subjects' characteristics.

Howie, Wang, and Tsai (2011) emphasized that accurate prediction of performance on the market when products are not yet available is crucial to the businesses success. In a study that aims to predict a medical product adoption on the market, the authors made the first empirical validation of BTS. In order to improve the predictive performance, the approach is adapted, based on the notion that "bad" respondents "hurt" the data, hence the authors exclude respondents with a low score (i.e. the untruthful ones). Applying BTS on a data from 1763 physicians, BTS improved the overall performance up to 21% compared to a simple average from the full sample. An interesting particularity, contradictory to the original BTS article by Prelec (2004), is that the authors did not explain the BTS scoring concept to the participants, arguing that it is impractical and it hasn't been empirically validated to improve the data.

In the study of Weaver and Prelec (2013), the authors conducted 5 experiments testing different aspects of BTS. They tested the validity of BTS upon respondents recognizing real and fake items from different categories such as brands, movies, journals etc. The authors tested BTS in the following settings: Recognition questionnaires, in 2x2 between subject design which consists of truth telling (Control vs. BTS) and deception incentives (incentives for recognizing items vs. none). For the first experiment, the authors noted that it is not clear whether the performance of BTS is because of the truth telling incentives or because of the claim made about them. To eliminate this doubt, keeping the same experimental settings, respondents were presented with their information score after each question. BTS was also compared to another "truth serum", the solemn oath. It requires participants to sign a declaration of their honesty, and again, respondents were given incentives to exaggerate their knowledge. BTS was also applied in the evaluation of a public good, e.g. it was applied on CV, in an attempt to eliminate the hypothetical bias. The experiments results showed good performance of BTS in recognizing real and fake items. The method performed well even when respondents were given incentives to overclaim item recognition. The results indicated there is strong evidence that instructing respondents about BTS has a positive impact over truth telling. BTS results outperformed the "solemn oath" and successfully eliminated the hypothetical bias in CV. Important to notice in this research is that the authors provided an empirical evidence that proper BTS instructions will result in higher data quality, a claim that was fairly questioned by Howie, Wang, and Tsai (2011). Furthermore, the results showed that BTS eliminated the hypothetical bias in CV and the authors speculated that the mixed evidence about BTS performance in Barrage and Lee (2010), possibly result from a difference in the instructions. Weaver and Prelec (2013) also address interesting pragmatic issues that will be discussed in another part of the paper.

In an interesting article about Questionable Research Practices (QRPs) (John, Loewenstein, and Prelec 2012) used by the scientific community, the authors applied BTS to a big sample (2155) of research psychologists. The BTS condition showed a significant difference in the results compared to the control group in 4 out of 10 measures of QRPs. The authors argue that the answers obtained with BTS incentives are more truthful. Following the fashion of investigating shameful activities from the QRPs article, Kukla-Gryz et al. (2015) apply BTS to investigate downloading intellectual property from unauthorized sources. BTS elicited around 60 percent higher self-admission rates on average compared to the control group. The authors interpreted the results as plausible in favor of BTS.

BTS has not been applied in the preference measurement domain yet. As there might be a discrepancy between one's preference in real life and in a conjoint context, BTS can play a significant role when measuring respondents' preferences when constructing new product, service or policy. The successful introduction of an optimal combination of attributes for new product, service or policy will crucially depend on respondents' truthfulness. Given the importance of accurately predicting the optimal product, service or policy product features, truthful elicitation is key when obtaining respondents' preferences.

Hypotheses

Based on the research question, research objectives, and the literature review I hereby formulate the following hypotheses:

H1. BTS will outperform the control group on choices where on of the alternatives dominate the other.

H2. The model estimated from the BTS sample will provide higher prediction score comparing the actual choice and the fitted model.

H2.1 Comparing in-sample absolute difference between the actual and the predicted choice

H2.2 Comparing in-sample prediction-realization table

H2.3 Comparing out-of-sample absolute difference between the actual and the predicted choice

H2.4 Comparing out-of-sample prediction-realization table

Methodology

The methodology chapter of the paper explains in details all of the procedures done in the analysis. The methodology part of this paper describes in details the following five steps of the experiment.

1. Questionnaire design
2. Designing a discrete choice experiment
3. Analysis of the choice experiment
4. Empirical comparison methodology

Questionnaire design

The context chosen for the conjoint experiment is data privacy. More precisely, the questions assess respondents' choices related to giving up their private online behavioral data on different electronic devices they are using for various reimbursment schemes.

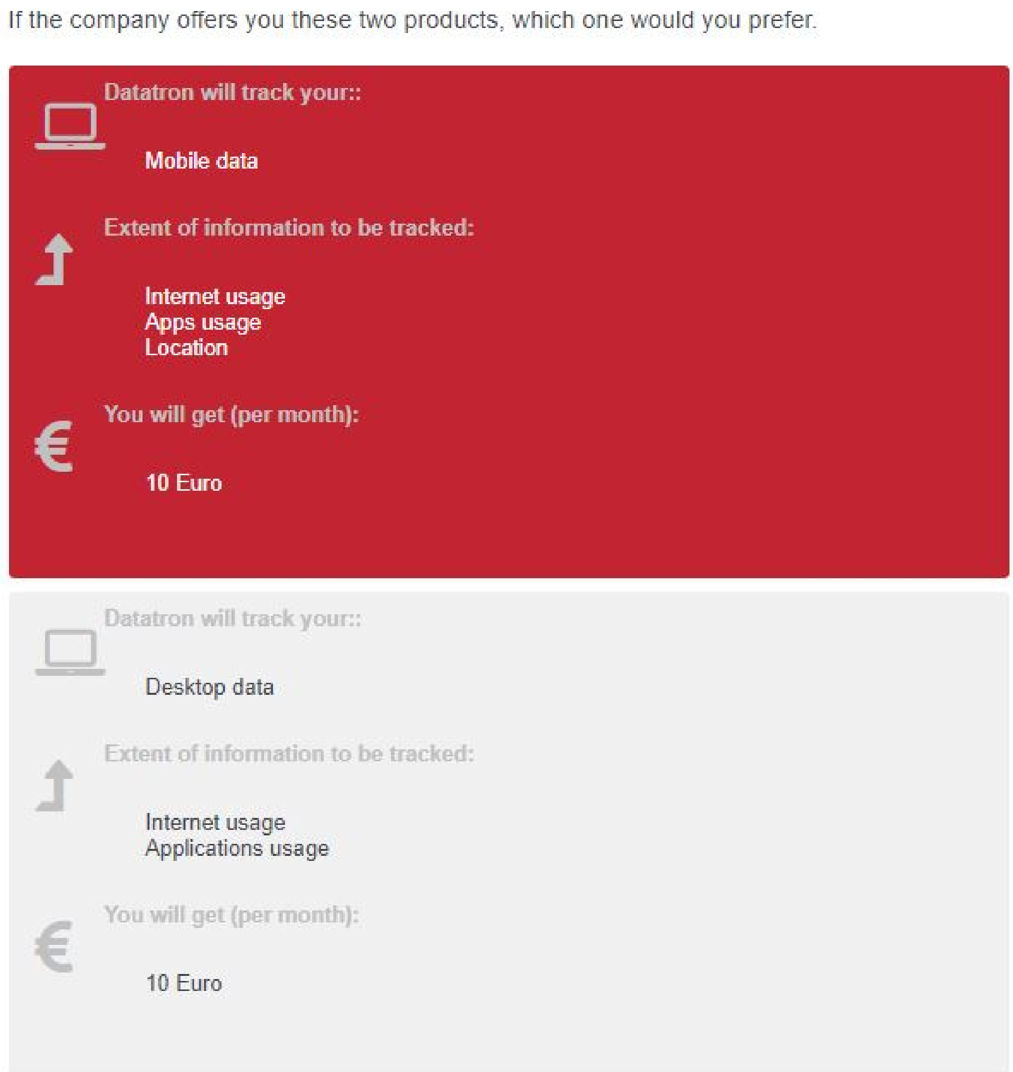
A computer-administered survey was executed. The survey included the following sections:

1. Instructions: regarding the incentives respondents will obtain;
2. A definition of what is behavioral data in order to understand the choice task;
3. A short description of a fictional company with a follow-up question, whether respondents understood the experimental context;
4. Main section, a conjoint exercise with 12 trade-off tasks, where each task consists of a choice between two profiles with different combinations of attribute levels;
5. Follow up questions demographic questions and contact information (if respondents wish to participate in the draw);

A between subject design was created and respondents were randomly allocated either to the BTS condition or to the control group. The instructions of the experiment were different. In the control group, respondents were instructed that at random one participant will receive an award of 20 euros. In the BTS condition, respondents were instructed that truthful scores are calculated and the person with the highest truth score will receive 20 euros. The main section of the conjoint experiment also differed between the two groups. Respondents were given a total of 12 conjoint tasks and in each task, they were required to make a choice between two options. In the control condition, respondents were asked to provide only their choice. In the BTS condition, additionally to each choice task, respondents were asked to predict the average distribution of people who would make the same choice.

Each task results in a variable indicating respondents' choice. This variable has been used to provide an input for the dependent variable of the logit model. The data is structured as a binary dependent variable indicating a choice and zero otherwise. The independent variables are represented by the attribute levels of each task. Each choice results in one observation, dependent variable indicating whether one task has been chosen over the other and the independent variables are indicating the level difference between the two alternatives that were shown to the respondents. In total there 12 observations per respondent.

Each task looks as follows:



In addition to each conjoint task, the respondents in the BTS condition had to answer an addional question:

If the company offers you these two products, which one would you prefer.

Datatron will track your::

Mobile data

Extent of information to be tracked:

Internet usage
Apps usage
Location

You will get (per month):

10 Euro

Datatron will track your::

Desktop data

Extent of information to be tracked:

Internet usage
Applications usage

You will get (per month):

10 Euro

Think about the rest of the respondents participating in the survey. What percentage of them do you feel are going to select the same option as you just did?

1 - 10%	51 - 60%
11 - 20%	61 - 70%
21 - 30%	71 - 80%
31 - 40%	81 - 90%
41 - 50%	91 - 100%

After each conjoint task respondents were presented a question asking them to estimate what proportion of the population they think will make the same choice as they did. To reduce respondents' burden and optimize the survey for mobile devices instead of typing in their guess in an open field, respondents were presented a scale with intervals. That way they did not give an exact estimate of the proportion of the population that would supply with the same answer, but rather an estimate within a given range.

BTS has been analyzed according to the scoring formula, explained in the BTS section of the literature review. As the survey uses interval scale instead of open answer input, for the prediction score estimation, the function uses the midpoint of the given range. To calculate respondents' truth scores from the BTS condition two inputs were required. First, their actual choice and second their midpoint of the given range indicated in the follow-up question. Based on the truth score calculation, the respondent with the highest truth score was rewarded 20 Euros.

Designing a discrete choice experiment

For the design of the choice experiment, I use Horne (2014) and Wheeler (2014). The method utilizes fractional factorial design for designing optimal choice experiments. Essentially the function works in the following way: as input values, it takes the number of attributes and levels associated with each attribute, the number of versions of the exercise, the number of desired choice tasks per each exercise and the number of alternatives per each task. Then it runs full factorial design for the given number of factors and levels. In accordance with the specified inputted number of tasks the function reduces the number of alternatives and generates optimal fractional factorial design. For a full discussion see Fedorov (1972).

The conjoint design includes three attributes and each attribute has three levels, as follows:

	Device	Extent	Reimbursement
Level 1	Desktop computer	Internet usage	8 Euro
Level 2	Mobile device	Internet usage and Apps usage	10 Euro
Level 3	Desktop and Mobile	Internet usage and Apps usage and, Location	12 Euro

The attributes and levels were determined based on consultation with a company provider of such data collections technology. The motivation for selecting this particular context is the increasing demand for behavioral data in the area of market research. Since this is an emerging trend, there is still no sufficient information about how much companies should pay to the participants for their data, nor do participants have experience selling it. The focus of this research, however, is not in the particular context of the survey.

In order to fulfill the design requirements of Bayesian truth serum, the respondents are presented with forced choice among two alternatives (i.e. "None" alternative has not been presented). Furthermore, all of the respondents are shown the same version of the questionnaire. This design poses certain limitations on the conjoint strength, which are discussed in details the "Limitations" chapter. The full conjoint design can be found in the appendix: *Appendix #3 Design and Survey.*

Analysis of the choice experiment

Logit model

To derive the estimated parameters logit model is used. To estimate the preferences for the different levels per attribute, utilities of each attribute levels are computed, the utility of the attribute is simply the sum of all of its levels. The "utilities are latent variables and are assumed to be a function of a set of explanatory variables X" (Walker and Ben-Akiva 2002):

The decision maker will choose an alternative with the highest utility:

$$\mathrm{Pr(Choice = 1 \mid D, E, P) = F(}\beta_0 + \beta_1\mathrm{Device} + \beta_2\mathrm{Extent} + \beta_3\mathrm{Price))}$$

Hence, the probability of choosing alternative I is:

$$\mathrm{Pr(Choice = 1 \mid D, E, P)} = \frac{\exp(\beta_0 + \beta_1\mathrm{Device} + \beta_2\mathrm{Extent} + \beta_3\mathrm{Price})}{1 + \exp(\beta_0 + \beta_1\mathrm{Device} + \beta_2\mathrm{Extent} + \beta_3\mathrm{Price})}$$

The final model includes the BTS treatment as an independent variable in the model and the effects of the treatment interacting with the independent variables i.e. device, extent and price. Therefore the model is taking the following form:

$$\mathrm{Pr(Choice = 1 \mid D, E, P, BTS) = F(}\beta_0+\beta_1\mathrm{Device}+\beta_2\mathrm{Extent}+\beta_3\mathrm{Price}+\beta_4\mathrm{BTS}+\beta_5\mathrm{(BTS^*Device)}+\beta_6\mathrm{(BTS^*Extent)}+\beta_7\mathrm{(B}$$

Model performance

In order to compare the models I will use the following performance indicators:

- Log likelihood
- Akaike Information Criteria
- McFadden R2

Predictions comparison

There are several ways to compare choice-model model performance based on a prediction-realization table (hit-rate table). If there is higher hit rate, that would imply a higher external validity.

		Predicted		
		Choice 1	Choice 2	
Observed	Choice 1	p_{11}	p_{21}	$p_{1.}$
	Choice 2	p_{21}	p_{22}	$p_{1.}$
		$p_{.1}$	$p_{.2}$	1

In-Sample prediction refers to producing a hit rate table for each observed choice in the data and comparing it to the predicted choice. The ratio between the correctly predicted choices and the incorrect ones is referred to as hit-rate. The higher the hit-rate the better the prediction realization of the model. Out-of-sample prediction refers to the same setting, however, a certain number of random observations are excluded from the sample and the model is estimated over the rest of the observations. The prediction realization table is produced over the observations that are not included in the sample. In this case, I left out 30% of the sample when producing the out-of-sample prediction realization table.

Residuals comparison

I will use sum of squared residuals (SSR) to compare the two conditions. SSR is the absolute difference between the actual choices and the fit model i.e. the deviations from the actual empirical values. Since the number of observations slightly differs per condition I will scale down to the number of observations. Smaller residual difference indicates less devation between the actual empirical values and the predicted values, thus, the smaller the SSR value the better.

$$\text{SSR} = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

Results

General results

A total of 79 respondents were collected for this study. Primarily in the age range of 25-34: 42 respondents and 18-24: 21. 34 are males, the other 44 females. Most of them are employed 63. Proportions are similar across the two conditions, 39 in the control group and 40 in BTS. More elaborate descriptive statistics can be found in the appendix.

Analysis

Model and coefficients

The outcome of the conjoint exercise results in 4 variables: * 3 categorical independent variables each representing an attribute with 3 levels. Each level of the first two attributes, “Device” and “extent” was converted to a binary variable taking value 1 if the level was present and zero otherwise. The third attribute “Reimbursement” was converted to a continuous variable with 3 degrees of freedom, taking values 8, 10 and 12. * A dependent variable indicating the choice of the respondent per task

The variables are described in the table below, the third level serves as a base level in the logit model, therefore it is omitted in the coefficients’ table. The last row indicates the dependent variable:

Table 0.3:

Attribute	Level	Variable name	Variable type
Device	Level 1	Desktop	Binary
Device	Level 2	Mobile	Binary
Device	Level 3	Both(Desktop and Mobile)	Binary
Extend	Level 1	Browsing history	Binary
Extend	Level 2	Browsing history and Applications	Binary
Extend	Level 3	Browsing history, Applications, Location	Binary
Reimbursement	-	Reimbursement	Continuous
-	-	Choice	Binary

Table 5.1 represents the logit parameters estimates for the overall population (All), overall population with the BTS condition as a covariate (All*BTS) and a model estimated on the isolated datasets from the two conditions. This table provides me with information regarding the direction of the relationship and also the significance. All variables are significant and have a positive relationship with the probability of making a choice in comparison to the base level. In the model that presents BTS treatment as a covariate I observe where BTS treatment alters the stated preference of the respondents' choices.

In general, keeping other factors constant:

1. Featuring desktop data in the product increases the probability of subjects making a choice compared with the base level i.e. featuring both desktop and mobile data;
2. Featuring mobile data in the product increases the probability of subjects making a choice compared with the base level i.e. featuring both desktop and mobile data;
3. Featuring browsing history data in the product increases the probability of subjects making a choice compared with the base level i.e. featuring browsing data, applications data, and location;
4. Featuring browsing history data and applications data in the product increases the probability of subjects making a choice compared with the base level i.e. featuring browsing data, applications data and location;
5. Increasing the reimbursement featured in the product increases the probability of subjects making a choice.

However based on the model with the BTS as a covariate (All*BTS) I observe:

1. The base level probability of subjects making a choice between (i.e. BTS variable value) BTS and the control group is not significantly different;
2. Featuring the different device levels does not significantly impact the choice of the two groups;
3. Featuring the different levels of information sharing i.e. Extent, also, does not significantly impact the choice of the two groups;
4. The Price has a significantly different impact on the choice between the two groups.

Looking at the model fit, I can see that the Akaike criterion is smaller for the BTS model compared to the control group. The Akaike criterion is used as a relative measure to compare the goodness of a

Table 0.4: Binary logit coefficients

	Dependent variable:			
	Choice = 1			
	All	All*BTS	Control	BTS
	(1)	(2)	(3)	(4)
Desktop (D)	1.721***	1.639***	1.719***	1.798***
	(0.174)	(0.551)	(0.240)	(0.269)
Mobile (M)	1.426***	2.178***	1.693***	1.207***
	(0.207)	(0.679)	(0.308)	(0.285)
Base: D + M				
Browsing History (BH)	2.321***	1.699***	2.172***	2.645***
	(0.186)	(0.584)	(0.251)	(0.299)
BH + Apps Usage (AU)	1.238***	1.468***	1.319***	1.169***
	(0.175)	(0.564)	(0.251)	(0.257)
Base: BH + AU + Location				
Price (Euro)	0.573***	0.085	0.426***	0.766***
	(0.063)	(0.196)	(0.082)	(0.107)
BTS		−0.022		
		(0.208)		
BTS:Desktop (D)		0.079		
		(0.361)		
BTS:Mobile (M)		−0.486		
		(0.420)		
BTS:Base: D + M				
BTS:Browsing History (BH)		0.473		
		(0.390)		
BTS:BH + Apps Usage (AU)		−0.149		
		(0.359)		
BTS:Base: BH + AU + Location				
BTS:Price (Euro)		0.340**		
		(0.135)		
Constant	0.075	0.127	0.105	0.082
	(0.101)	(0.327)	(0.145)	(0.149)
Observations	948	948	468	480
Log Likelihood	−374.740	−369.926	−196.656	−173.270
Akaike Inf. Crit.	761.480	763.851	405.312	358.540

model between several models; the lower the Akaike criterion, the better the fit. However the sample size between the two groups differs and ths this measure is not informative when it comes to comparison.

Exploring the odds ratios also gives a clear way of comparison of the attribute importance of a hypothetical product. One euro increase of the reimbursement gives Ã round 25% higher odds of choosing a product in the BTS comparing to the control group. The impact of featuring Browsing history and applications (Level 2 of the *Extent* attribute) in the product is almost as high as double in BTS compared to the control group. A detailed plot of the values from the table below can be found in *Appendix 1. Figures and Graphs*, *Section: Odds Ratios*.

Table 0.5: Odds ratios

	All	All∗BTS	Control	BTS
Desktop (B: Both)	5.592	5.151	5.576	6.037
Mobile (B: Both)	4.162	8.830	5.433	3.343
Browsing History (B: BH + AU + LI)	10.181	5.466	8.776	14.090
Browsing History + Apps (B: BH + AU + LI)	3.449	4.340	3.738	3.219
Price (Euro)	1.774	1.089	1.530	2.151
group		0.978		
group:Desktop (B: Both)		1.083		
group:Mobile (B: Both)		0.615		
group:Browsing History (B: BH + AU + LI)		1.605		
group:Browsing History + Apps (B: BH + AU + LI)		0.861		
group:Price (Euro)		1.405		

*Both: Desktop + Mobile; BH: Browsing History, AU: Application Usage, LI: Location Information

Furthermore, to see if there is a difference between the magnitude of each variable over the choice, I look at the marginal effects of each variable (table 5.3). There are certain differences between the two models. First, featuring desktop data seems to have a similar effect across the two models, while featuring mobile data is associated with much lower effect. However, a product featuring browsing history increases the probability of making a choice by 0.42 percentage points relative to a product featuring browsing history, application usage and location in the BTS condition, whereas this probability increases by 0.57 percentage points for the control group. The probability of making a choice increases by 0.02 percentage points when the reimbursement increases by one euro in the BTS conditions, while for the control group, it increases by 0.14 percentage points. A detailed plot of the values from the table below can be found in *Appendix 1. Figures and Graphs*, *Section: Marginal Effects*. Furthermore, an isolated effect of the treatment on the product characteristics is presented in *Section:Marginal Effects - BTS Treatment impact.*

Prediction realization tables

In-sample predictions

In sample prediction realization table has been obtained by crossing the number of prediction

Table 0.6: Marginal effects

	All	All∗BTS	Control	BTS
Desktop (B: Both)	0.424	0.405	0.423	0.445
Mobile (B: Both)	0.351	0.538	0.416	0.299
Browsing History (B: BH + AU + LI)	0.572	0.419	0.534	0.655
Browsing History + Apps (B: BH + AU + LI)	0.305	0.362	0.324	0.290
Price (Euro)	0.141	0.021	0.105	0.190
group		−0.006		
group:Desktop (B: Both)		0.020		
group:Mobile (B: Both)		−0.120		
group:Browsing History (B: BH + AU + LI)		0.117		
group:Browsing History + Apps (B: BH + AU + LI)		−0.037		
group:Price (Euro)		0.084		

*Both: Desktop + Mobile; BH: Browsing History, AU: Application Usage, LI: Location Information

estimation versus the actual numbers. As the estimates are probabilities, 0.5 cut-off point has been used i.e. all predictions higher that this threshold are considered as choice and all below otherwise. Graphical representation of the table below can be found in *Appendix 1. Figures and Graphs*, *Section: Receiver operating characteristic curve > ROC In-Sample Predictions.* The difference between the confusion matrix below and the ROC curve is that in the ROC curbe the cut-off point of 0.5 has been varied.

Table 0.7: Prediction realization table In-Sample Predictions

Condition	Predicted	Actual	Result
Control	0	0	0.444
Control	1	0	0.105
Control	0	1	0.056
Control	1	1	0.395
BTS	0	0	0.454
BTS	1	0	0.088
BTS	0	1	0.046
BTS	1	1	0.412

Control Accuracy 0.84
BTS Accuracy 0.867

In the tables above the BTS model performs slightly better producing a higher percentage of accurate prediction than the control group. Furthermore, the area below the ROC curve is larger for the BTS condition compared to the control group, which indicates better performance (See Appendix #1 for reference).

Out-of-sample predictions

Out-of-sample prediction realization table is similar to the in-sample one, however, when estimating the model I am using only random 70% of the observations and then fitting the model in the rest of the observations. The results in these tables are obtained over the random 30% of the observations not included in the model. Again similar to the in-sample table a 0.5% cut-off threshold has been applied. Graphical representation of the table below can be found in *Appendix 1. Figures and Graphs*, *Section: Receiver operating characteristic curve > ROC Out-Of-Sample Predictions.*

Table 0.8: Prediction realization table Out-of-Sample Predictions

Condition	Predicted	Actual	Result
Control	0	0	0.429
Control	1	0	0.164
Control	0	1	0.05
Control	1	1	0.357
BTS	0	0	0.396
BTS	1	0	0.09
BTS	0	1	0.056
BTS	1	1	0.458

Control Accuracy 0.786
BTS Accuracy 0.854

In the tables above the BTS model performs better producing a higher percentage of accurate prediction than the control group. Furthermore, the area below the ROC curve is larger for the BTS condition compared to the control group, which indicates better performance (See Appendix #1 for reference).

Residuals

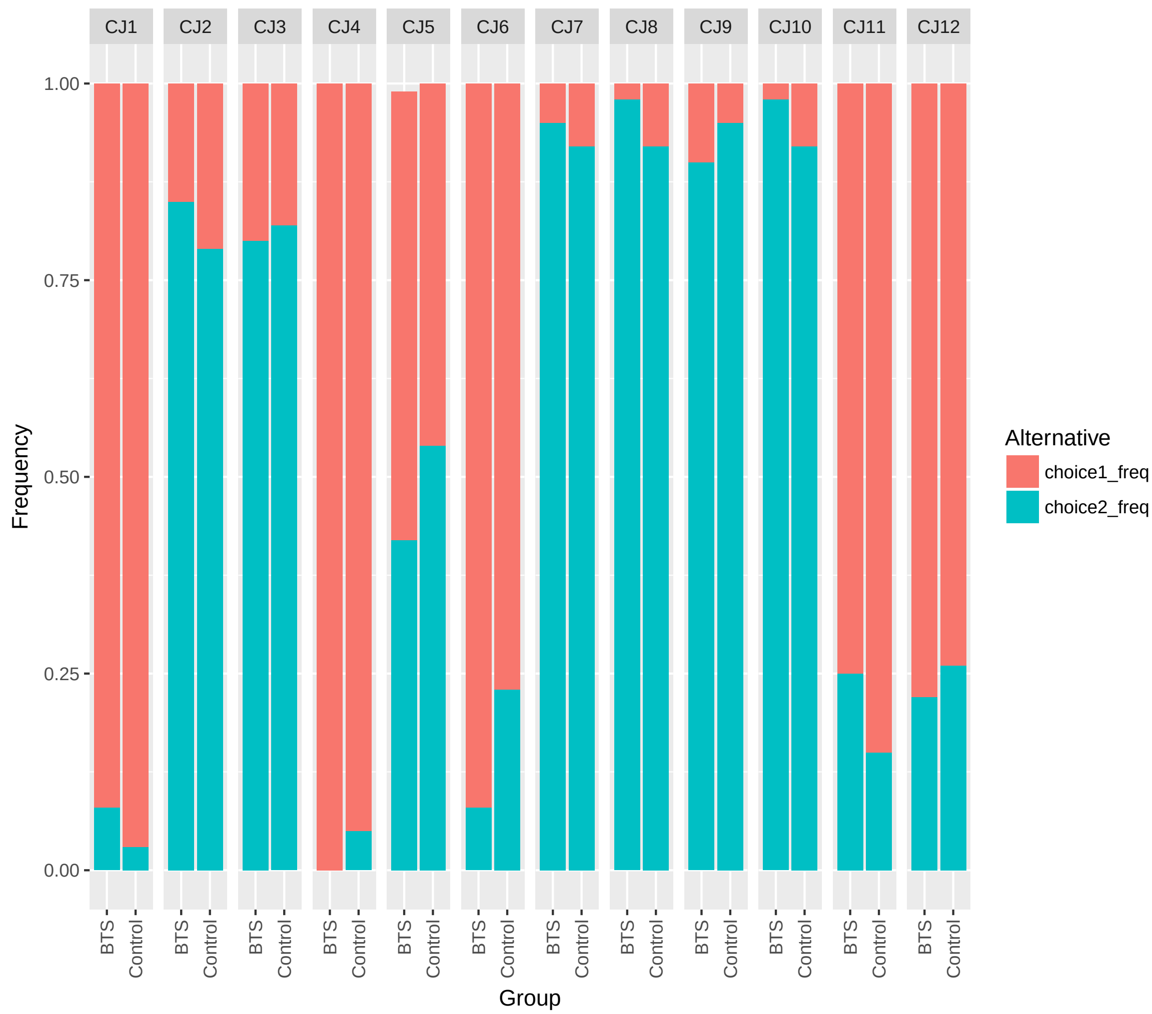
Table 0.9: (Absolute sum of residuals)/n

Data	Model	Value	Lower
In-Sample	Control	0.258	
In-Sample	BTS	0.217	$*$
Out-of-Sample	Control	0.278	
Out-of-Sample	BTS	0.219	$*$

The absolute sum of residuals of BTS model using the full data set i.e. In-Sample, as well as BTS model used to predict the Out-Of-Sample observation is lower. This indicates lower deviance from the actually observed empirical values when using the BTS treatment compared to the control group.

Comparing dominated alternatives

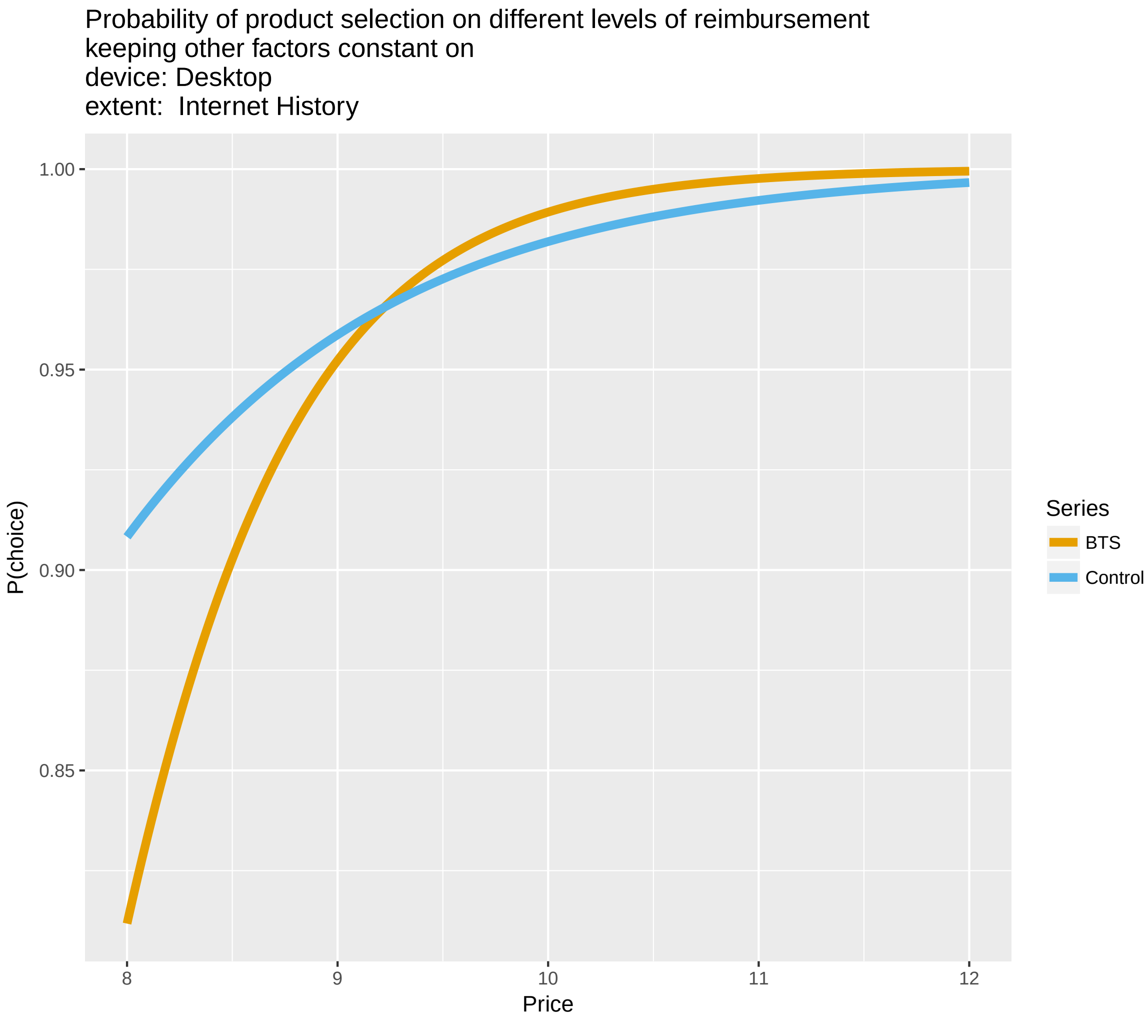
Assuming respondents are making rational choices I would expect higher price will always dominate respondents choices comparing to lower one. In the conjoint desing there were two choice tasks, CJ1 and CJ10 where one options was dominant. In CJ1 the price of option 1 dominated the price of option 2 while all of the other characteristics are the same. In CJ10 the price of option 2 dominated the price of option 1. From the graph below I can see that in CJ1 option 1 i.e. the dominant alternative was slightly more preferred in the Control group coparing to the BTS. When it comes to CJ10, option 2 i.e. the dominant alternative was slightly more preferred in the BTS compared to the control group.



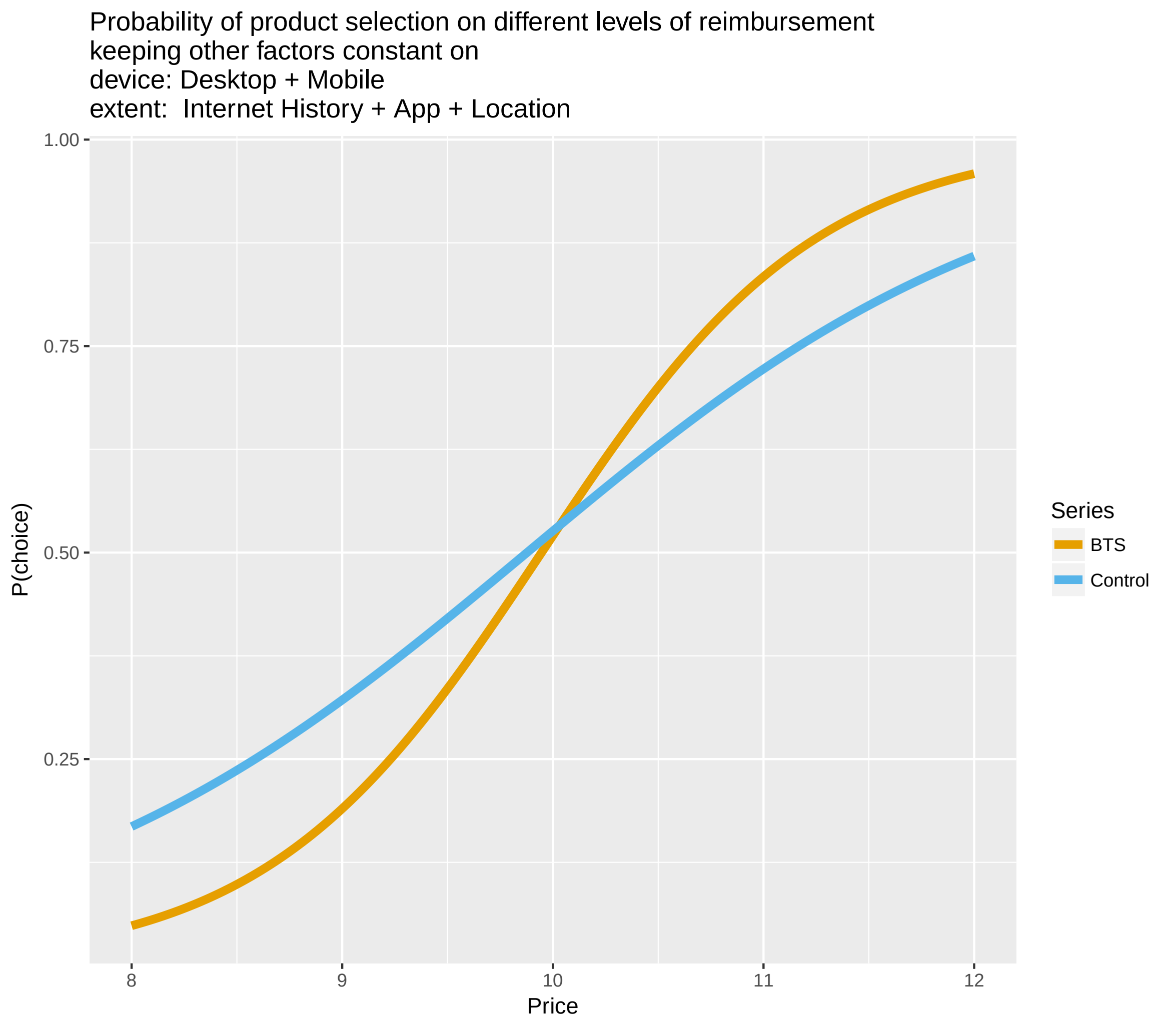
Product comparison

From a practical standpoint, the estimated contrasting coefficients pose a serious question since the decisions based upon them would be different. In order to assess these differences, I look at different product compositions.

Graph 1 displays the probability of a product selection with different reimbursement levels while keeping others factors constant: desktop as device and browsing history as an extent of sharing information. It is observed that the probability of making a choice is different between the two conditions when the reimbursement is the lowest and the highest. More specifically, for a reimbursement of 8 euros, respondents from the BTS condition indicate they are less likely to make a choice for a product that features sharing their browsing history on a desktop, whereas the probability of making a choice for the same product is higher in the control group. Furthermore, for a reimbursement of 12 euros, respondents from the BTS condition indicate they are slightly more likely to make a choice for a product that features sharing their browsing history on a desktop, whereas the probability of making a choice for the same product is lower in the control group.



The same tendency is observed in Graph 2, which displays the probability of a choice of a product of different reimbursement levels with devices: desktop + mobile and extent: browsing history, application, and location. It can be noted from the two graphs, that respondents are less willing to make a choice involving sharing their data on multiple devices.



The different output from the two models results in different outcomes which could lead to different choice by marketeers. This is perhaps the most important result of this paper. Keeping all other factors constant, would result in different predictions from the two models and ultimately lead to different decisions made by the practitioners

Conclusion

General discussion

The main goal of this paper was to explore whether a new incentive-compatible method, the Bayesian truth serum, could improve the validity of a conjoint methodology compared to a traditional hypothetical conjoint setting. The main hypotheses were that the BTS model will provide better goodness of fit and provide better prediction realization rate compared to the control group. The context of the research was a new service development regarding collecting private data. The research examined whether the BTS could result in providing more truthful respondents in a context where objective truth is absent. In this research, I applied the BTS in a context that is hadn't been tested before â€“ preference measurement methodologies. Specifically, I used a choice based conjoint to test whether the BTS could improve the quality of the data.

The results showed that the conjoint model based on the BTS and the control group differ in few key aspects. First, different effects of the variables are observed between the two models. Respondents from the BTS condition are more likely to choose an alternative with a higher reimbursement compared to respondents from the control condition. That would imply that the reimbursement is the 'deal-breaker' in the BTS condition in choosing to sell one's data, ceteris paribus. The comparison of the models reveals differences between the probabilities of model selection based on the different reimbursement levels. That would imply that respondents in the BTS condition are less willing to make a choice if the reimbursement is lower, but they are more willing to make a choice if the reimbursement is higher. From a practitioner standpoint, this implies different choices with the usage of the different methodologies.

With this study, I confirmed all of the outlined hypothesis:

> H1. BTS will outperform the control group on choices where on of the alternatives dominate the other: *Partially confirmed*
>
> H2. The model estimated from the BTS sample will provide higher prediction score comparing the actual choice and the fitted model.
>
> H2.1 Comparing in-sample absolute difference between the actual and the predicted choice: *confirmed*
>
> H2.2 Comparing in-sample prediction-realization table: *confirmed*
>
> H2.3 Comparing out-of-sample absolute difference between the actual and the predicted choice: *confirmed*
>
> H2.4 Comparing out-of-sample prediction-realization table: *confirmed*

Contributions

This research contributed methodologically to the stream of research testing the validity of the Bayesian truth serum. More precisely, it complements the stream of research that is in favor of using the approach for truth telling. Overall, the research found that the BTS is performing well in a setting where truthful answers are unknowable. The paper confirms that using BTS in a choice based conjoint experiment leads to improvement in data reliability. Furthermore, the research adds up to the emerging stream of literature exploring how incentive-compatible methodologies could be applied to preference measurement experiments. The research further confirms the notion that incentive-compatible methodologies could be utilized within preference measurement experiments.

This research also offers several managerial insights for market researchers. Comparing the probability of making a choice across the two conditions, showed that different results are obtained. A market researcher could infer different conclusions depending on whether they are using an incentive-compatible method, BTS or not.

Limitations

One of the limitations of this research was ensuring an efficient conjoint design. Conjoint design refers to "process of generating specific combinations of attributes and levels that respondents evaluate in choice questions" (F. R. Johnson et al. (2013)). An efficient conjoint design allows for better estimates of the parameters or the part-worths for a given number of choice tasks. To generate a design with 4 levels and 4 attributes per level, a total of 256 product profiles are generated, in order to reduce the number of profiles shown to respondents and have a robust statistical information, the choice tasks are varied for respondents i.e. each respondent receives a different set of alternatives to evaluate. Different design types can account for different effects that could be captured depending on the research objectives, e.g. main effects only, interaction, cross effects, etc. In order to apply BTS with conjoint, to calculate the truth score per choice task respondents were given one design version. Hence, a trade-off had to be made between ensuring efficient conjoint design and applying the BTS.

Furthermore, in order to make the BTS exercise easier for the respondent, the conjoint exercise is held up to only two alternatives with no "None" option. Depending on the experimental design conjoint tasks maybe composed by more than two choices and also include a "None" option, as it is realistic that some respondents wouldn't select any of the alternatives presented to them within a given task. BTS design setting would increase the respondent burden arguably too much if they have to fill in their predictions about more than two alternatives. A possible way to overcome this limitation would be to ask only for the percentage of the population that would take the same option and then assume a certain distribution across the other two (or more) alternatives.

Moreover, incentives were not distributed immediately and to everyone according to their truth scores. In the BTS condition, participants were told they could receive an award of 20 euro if they have the highest truth score. Therefore, not all respondents were provided incentives for their time and effort. Furthermore, the conjoint exercise is already a lengthy survey to be filled in and adding more questions to it could increase respondents' burden. Weaver and Prelec, 2012 argue that "it would be sufficient to elicit predictions from a small number of randomly selected respondents and use their predictions to calculate initial scores". In that way, estimations of an answer's average distribution could be done and in the survey itself, respondents can provide answers only to the conjoint experiment and only provide their opinion, reducing their burden. Additionally, utilizing this approach it would be possible to calculate respondents' truth score "on-the-go", hence after they finish the survey they would know their score and their reward, which could induce their motivation to participate.

The survey was distributed through social media, hence it did not account for a controlled experimental environment. A controlled experiment would allow reducing the noise in the data, as it reduces the effect of other variants different than the independent, which results in a positive impact on the reliability of the data and the results. Yet, this two limitations can be relaxed since the main objective of the paper is to access the reliability of BTS in market research application where no control over the environment has been posed.

The experiment can be replicated with higher sample size. Due to the low sample size, each data point heavily affects the outcomes of the research.

Due to a programming mistake, the product attributes were presented in different order for the control and BTS groups, in the control group the price was presented first while in the BTS condition the price was presented as the last attribute in each profile. This results in losing control for the effect of the order. I would expect that characteristics presented earlier will have higher effects, therefore this issue diminished the difference between the control and the treatment group. However, this is only a speculation. Even presented as last attribute price has a higher importance in the BTS condition compared to the control group.

Future research

It is common practice for researchers to include the so-called hold out cases that are used for model evaluations. Those conjoint tasks are held constant across the whole sample. In order to improve both BTS and the conjoint design efficiency I propose an experimental design where these hold out conjoint tasks are also applied BTS which would make BTS scores estimation possible, while the rest of the conjoint tasks are varied across the sample in order make the analysis more robust and estimation of interaction effects across parameters possible.

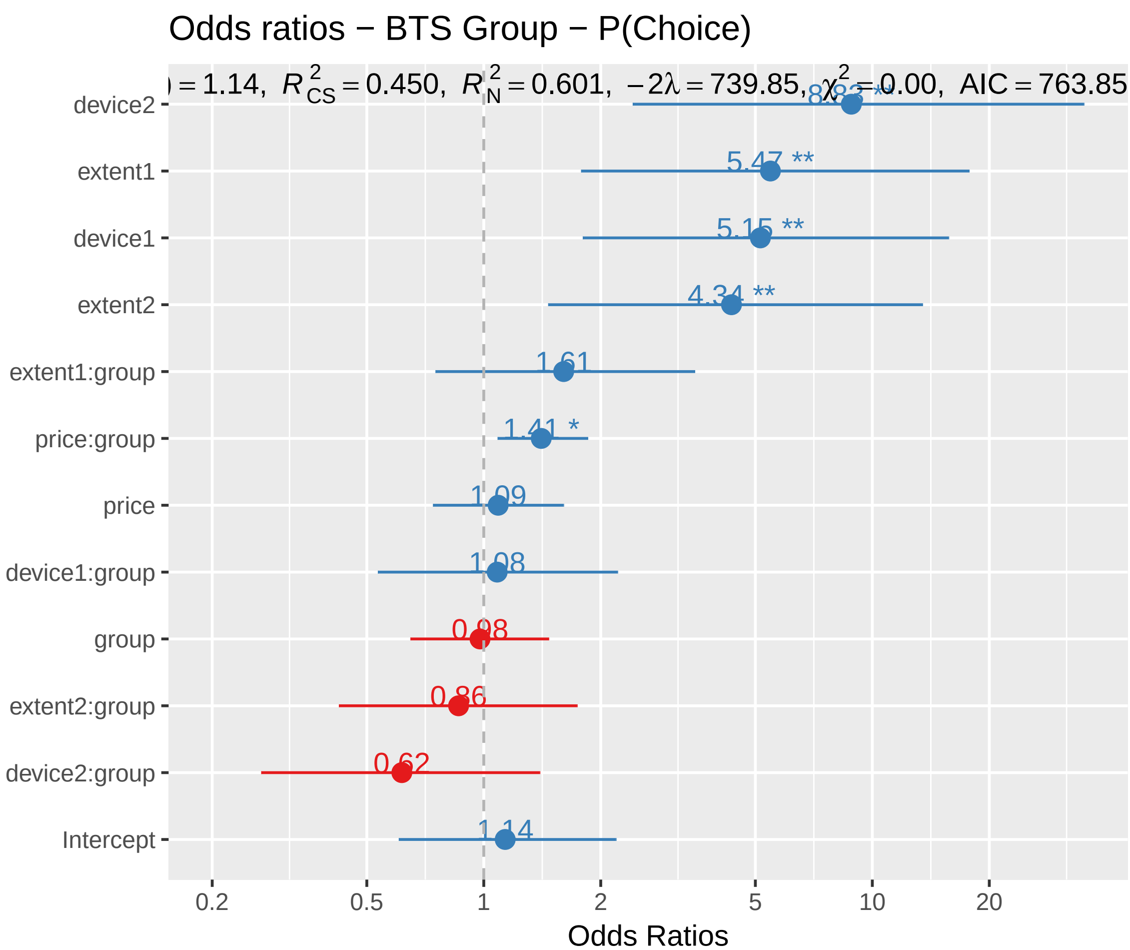
Appendix #1 Figures and Graphs

Appendix 1. Figures and Graphs

Odds rations

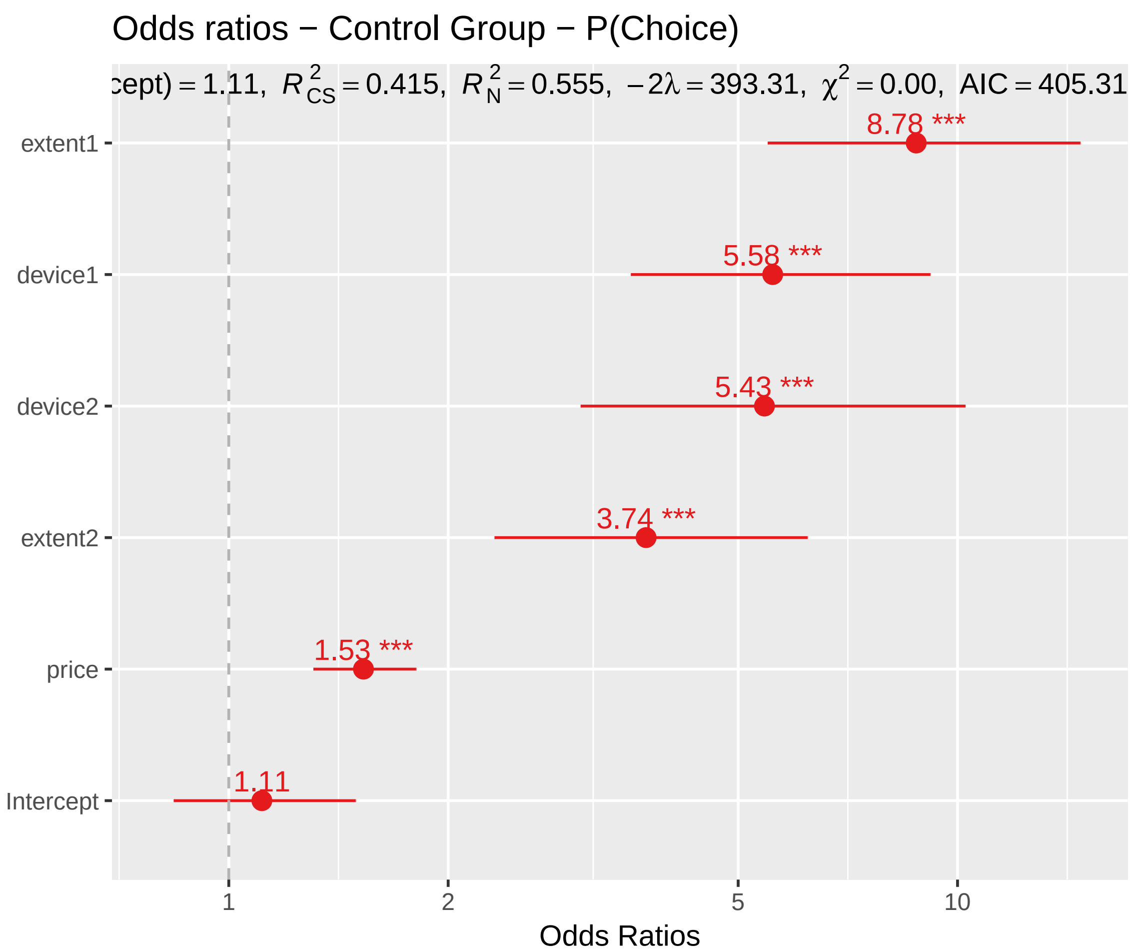
```
## Waiting for profiling to be done...
```

```
## Intercept = 1.14
## R2[cs] = 0.450
## R2[n] = 0.601
## Lambda = 739.85
## Chi2 = 0.00
## AIC = 763.85
```



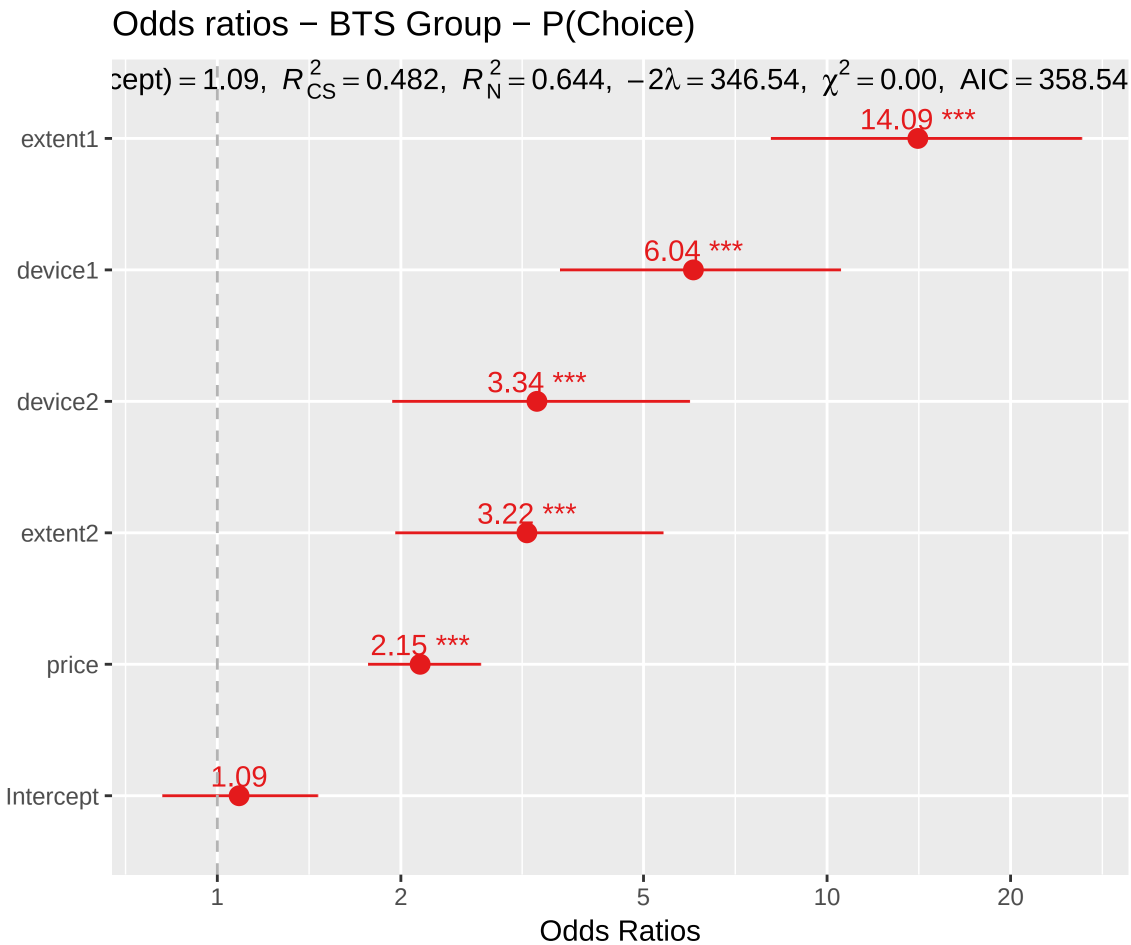

```
## Waiting for profiling to be done...
```

```
## Intercept = 1.11
## R2[cs] = 0.415
## R2[n] = 0.555
## Lambda = 393.31
## Chi2 = 0.00
## AIC = 405.31
```

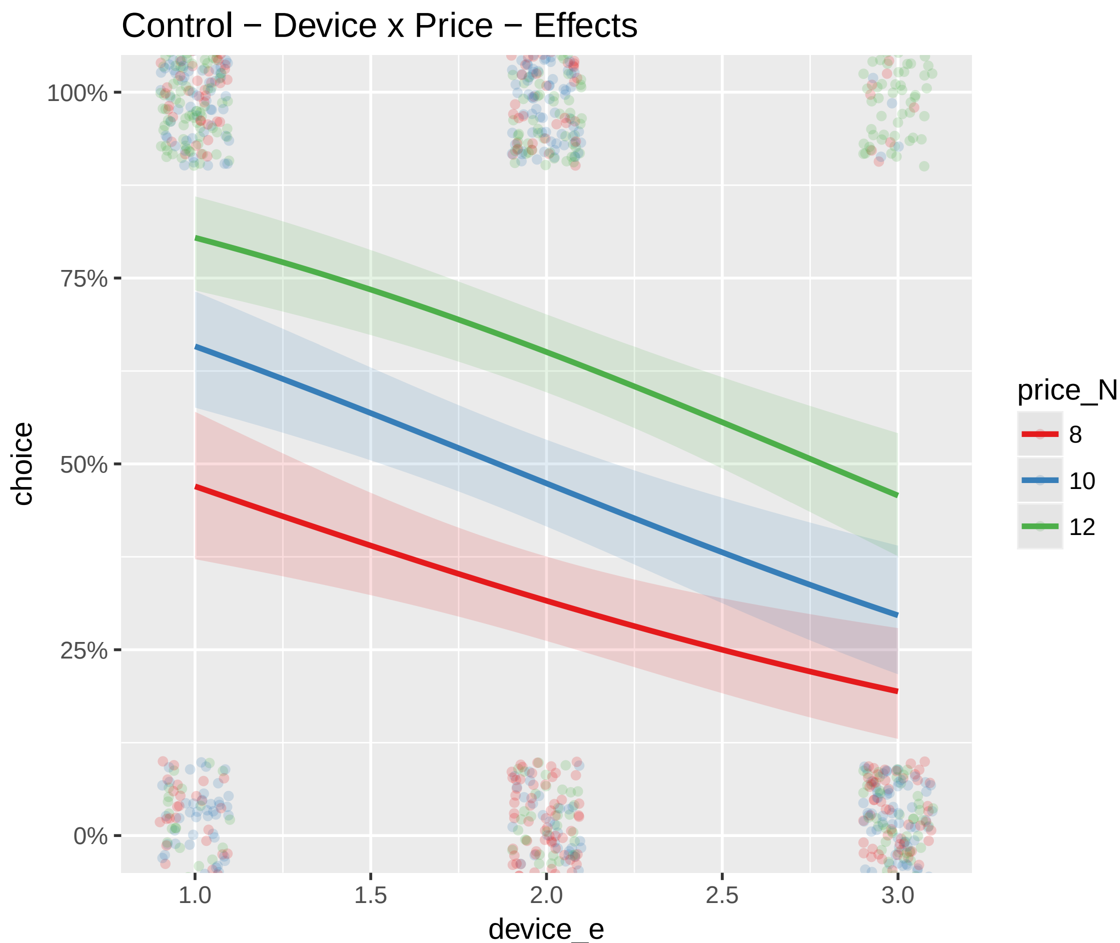


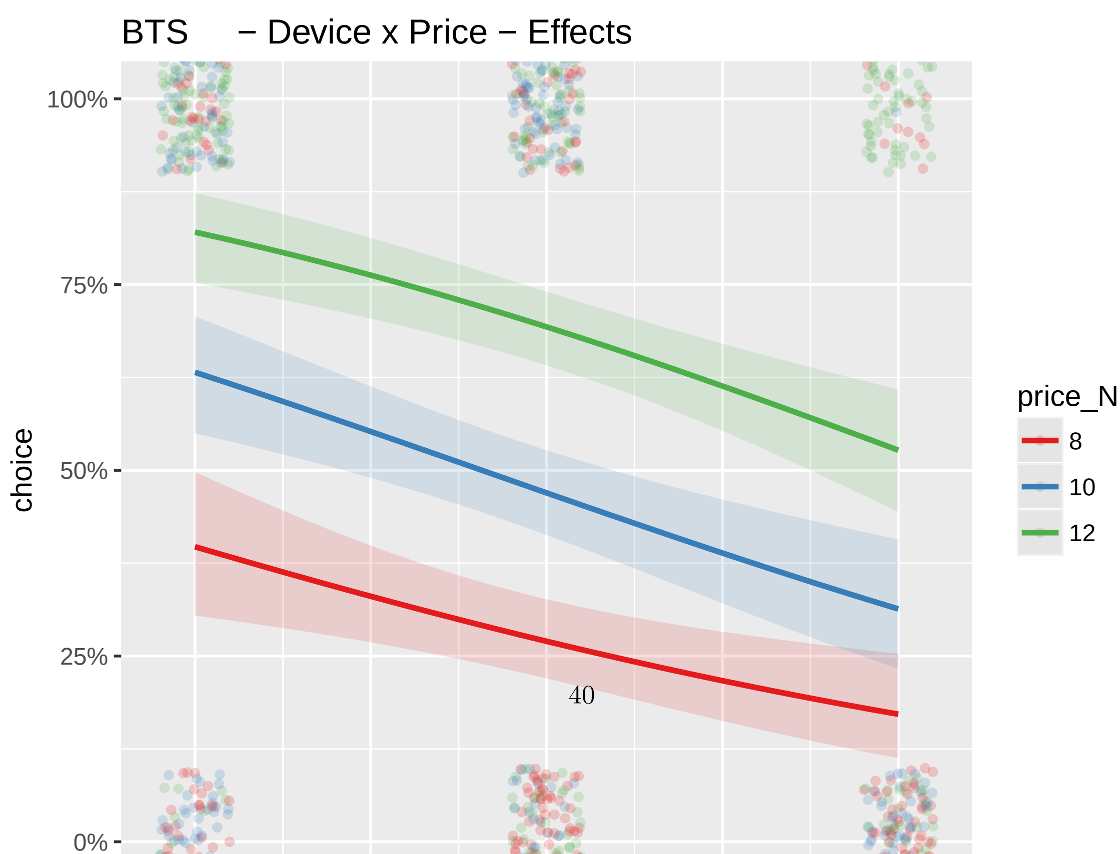

```
## Waiting for profiling to be done...

## Intercept = 1.09
## R2[cs] = 0.482
## R2[n] = 0.644
## Lambda = 346.54
## Chi2 = 0.00
## AIC = 358.54
```

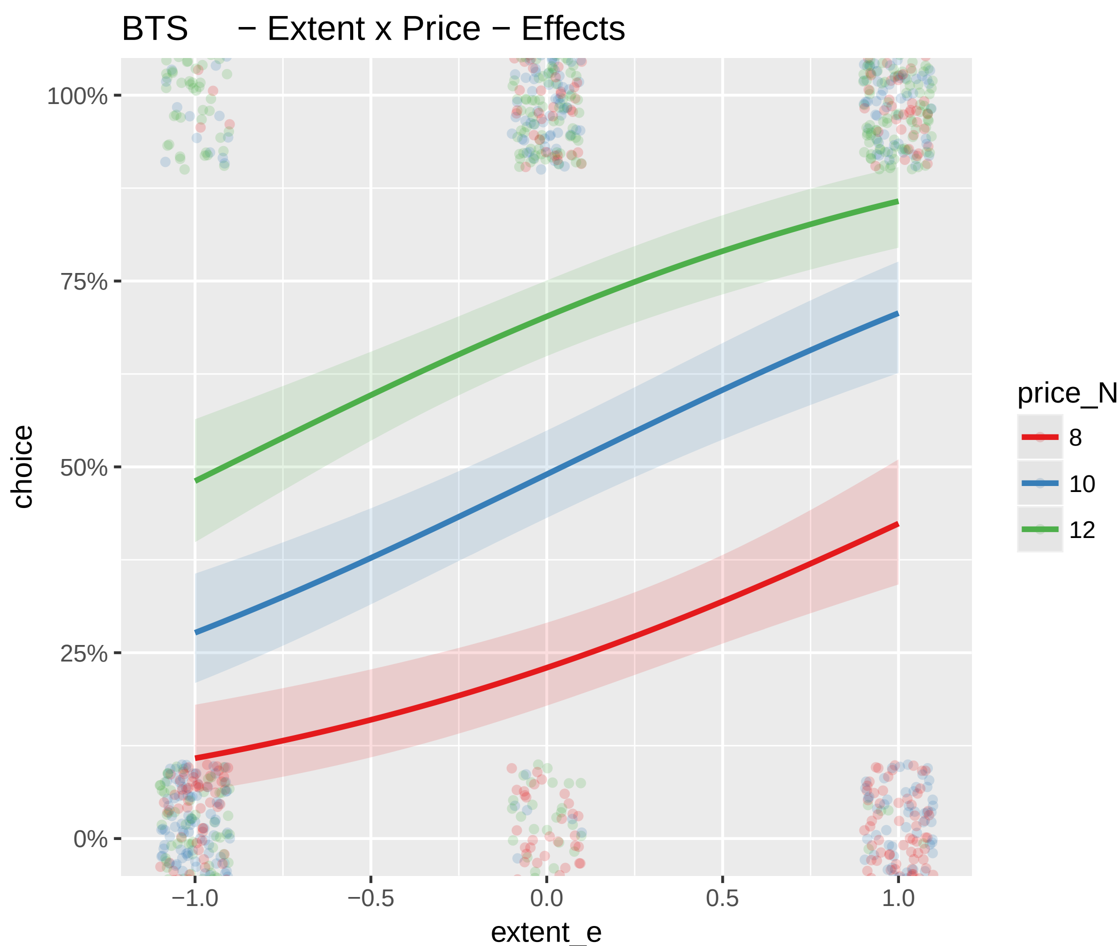

Odds ratios − BTS Group − P(Choice)
cept) = 1.09, R^2_{CS} = 0.482, R^2_N = 0.644, −2λ = 346.54, χ^2 = 0.00, AIC = 358.54
extent1
device1
device2
extent2
price
Intercept
14.09 ***
6.04 ***
3.34 ***
3.22 ***
2.15 ***
1.09
1
2
5
10
20
Odds Ratios

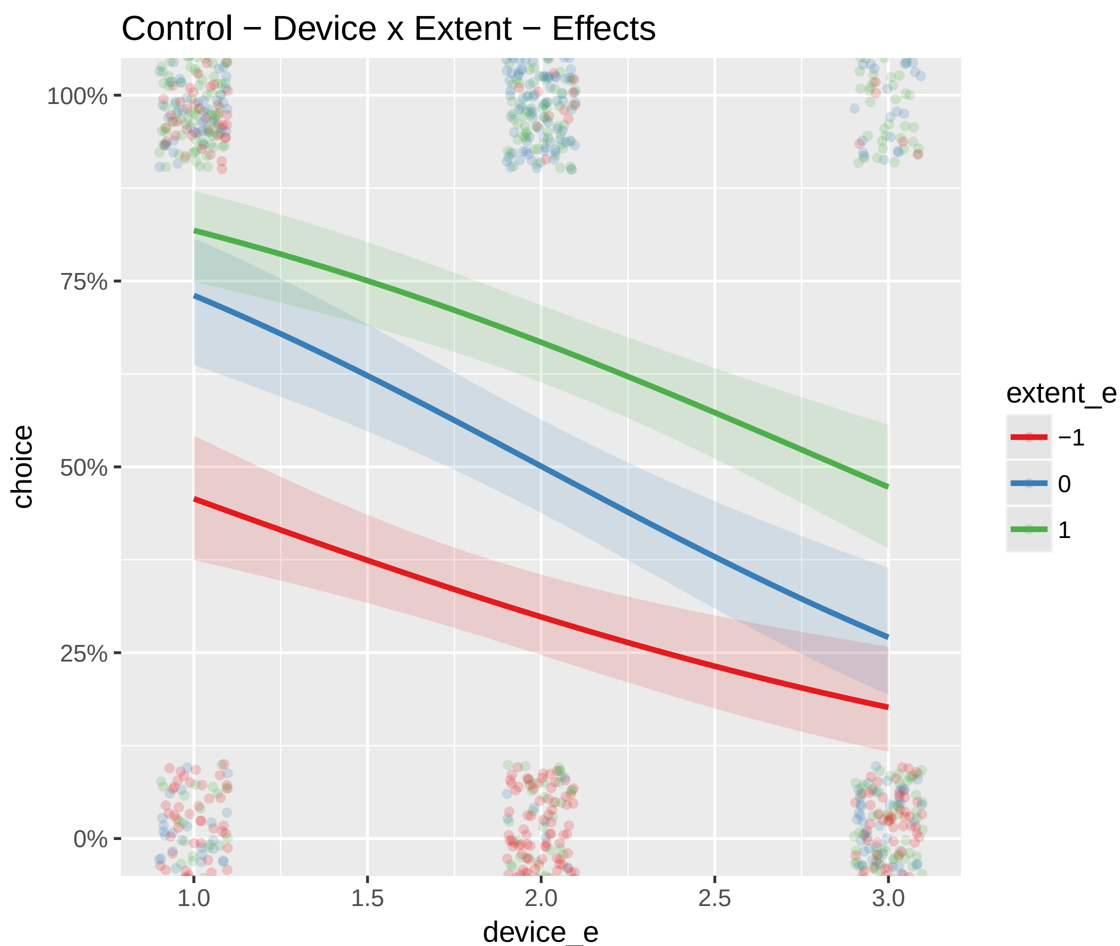
Marginal Effects





Control − Extent x Price − Effects
100%
75%
50%
25%
0%
choice
−1.0
−0.5
0.0
0.5
1.0
extent_e
price_N
8
10
12

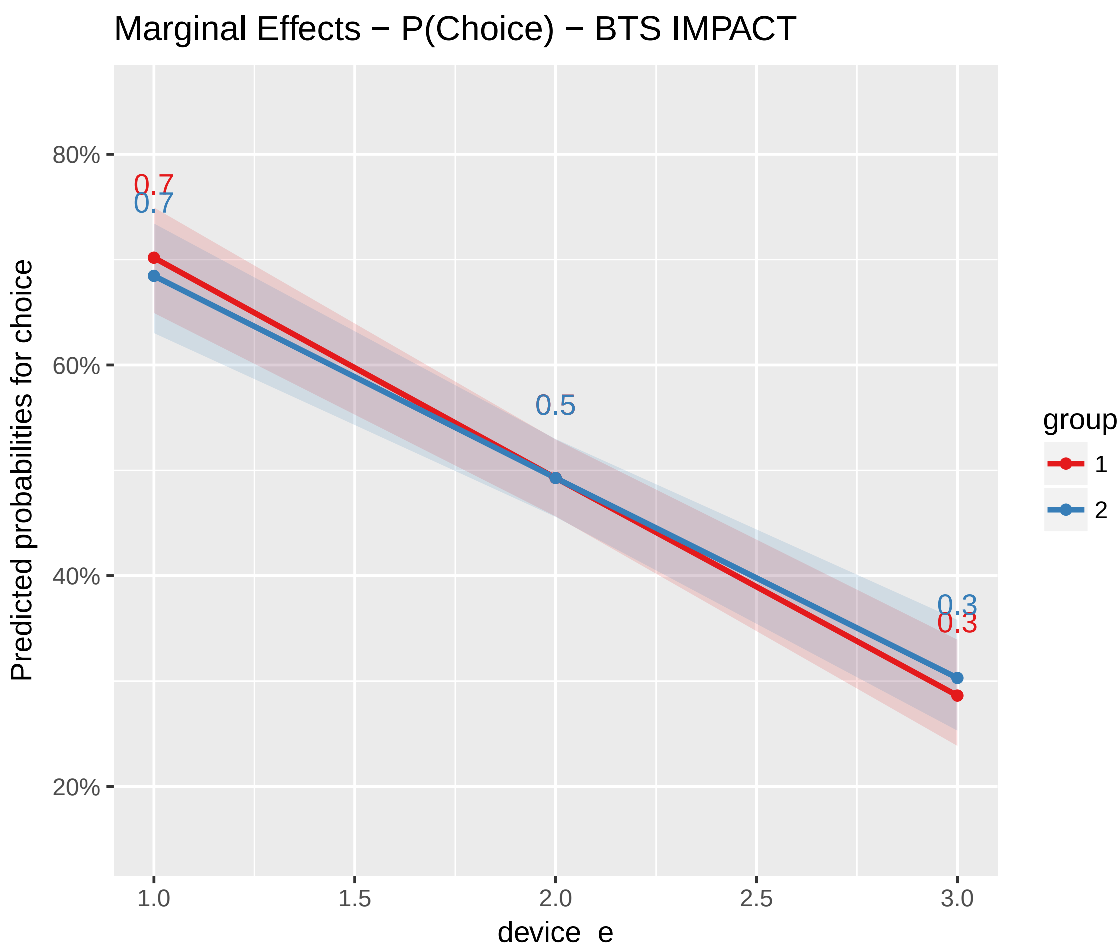

BTS − Extent x Price − Effects
100%
75%
50%
25%
0%
choice
−1.0
−0.5
0.0
0.5
1.0
extent_e
price_N
8
10
12

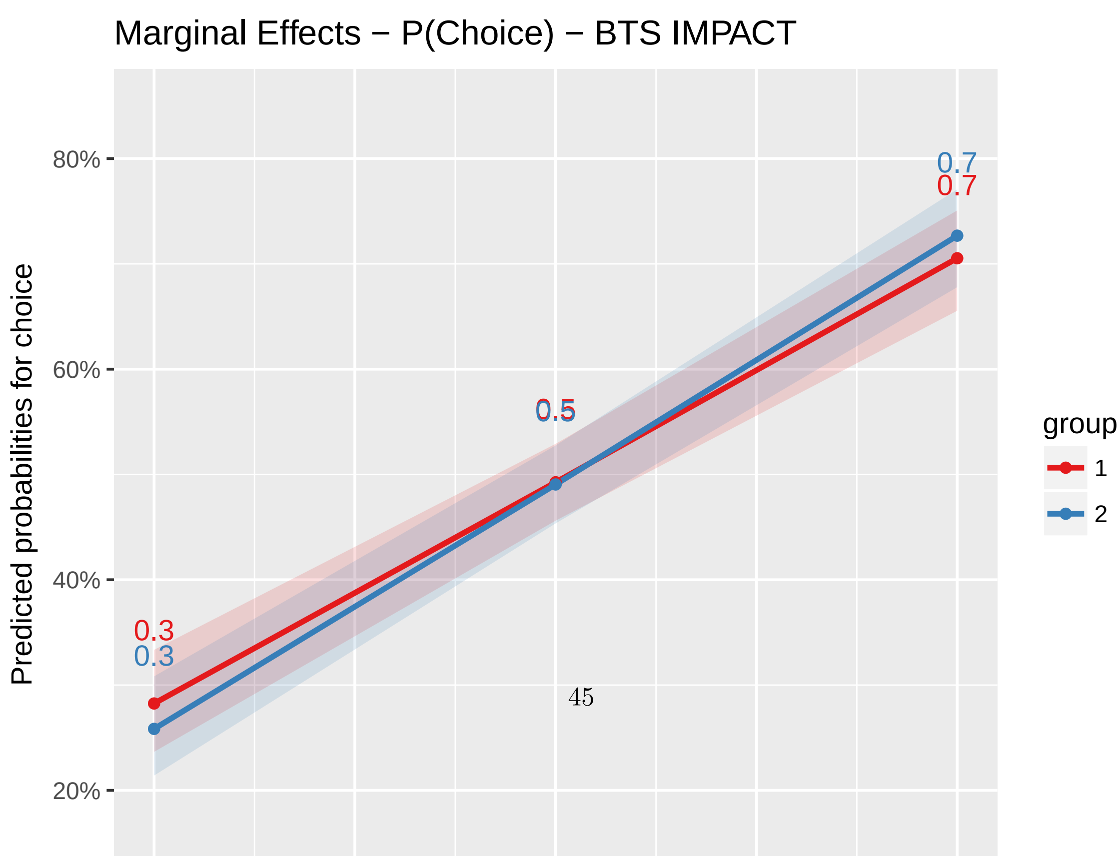


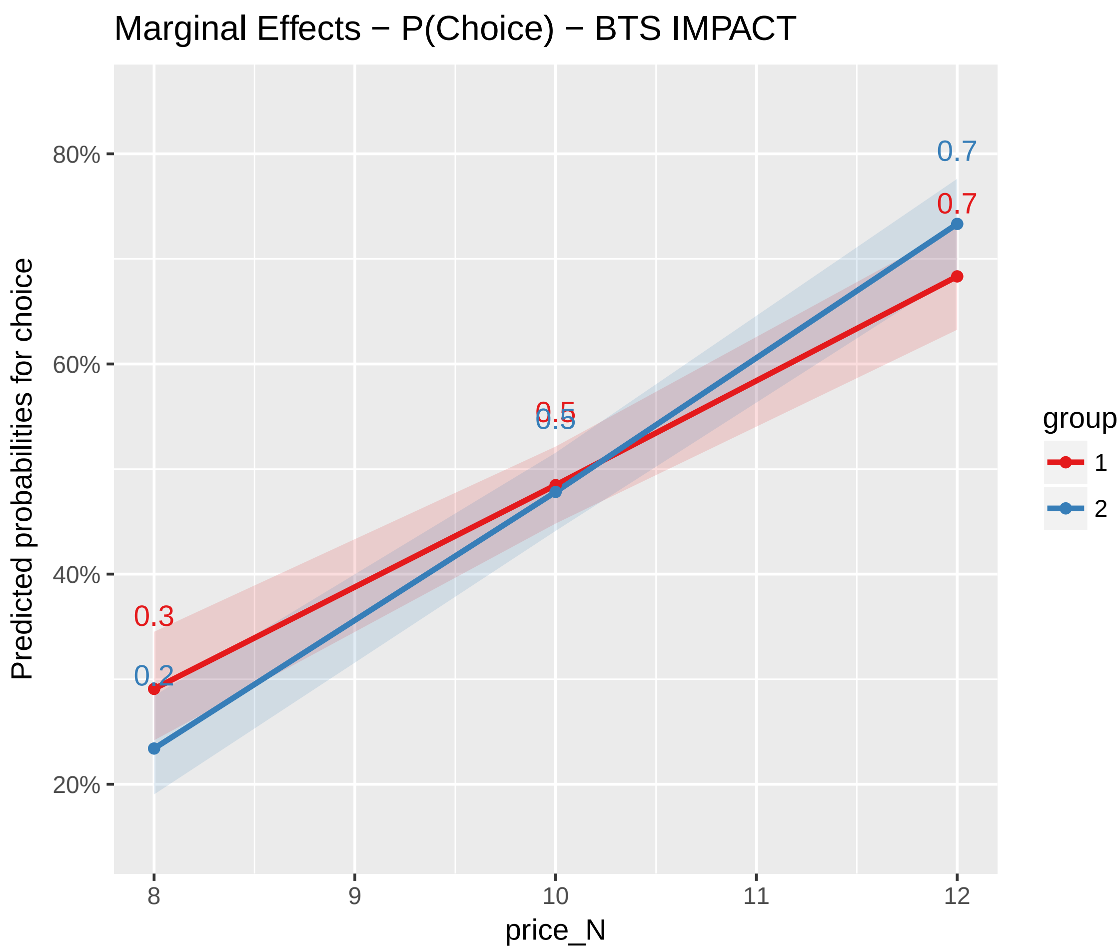
Control − Device x Extent − Effects
100%
75%
50%
25%
0%
choice
1.0
1.5
2.0
2.5
3.0
device_e
extent_e
−1
0
1

BTS − Device x Extent − Effects
choice
100%
75%
50%
25%
0%
1.0
1.5
2.0
2.5
3.0
device_e
extent_e
−1
0
1

Marginal Effects - BTS Treatment impact

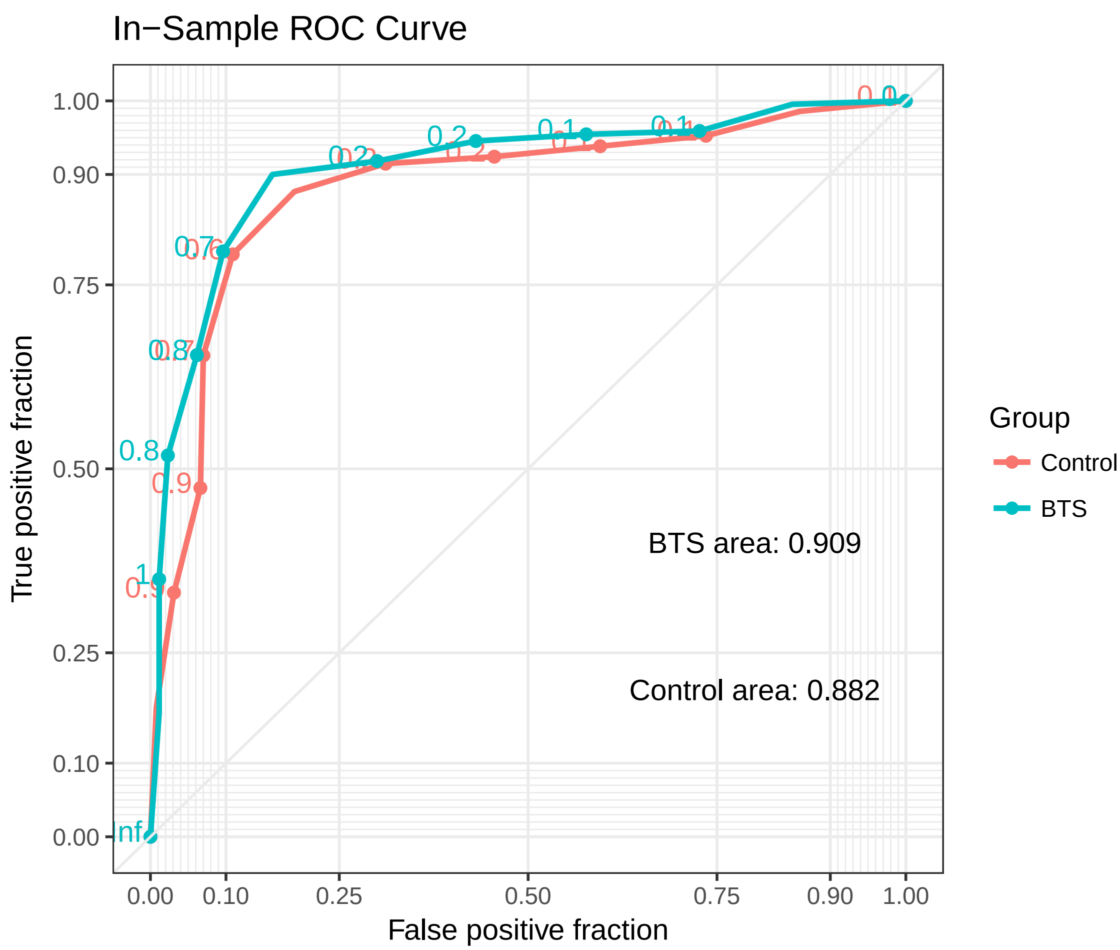




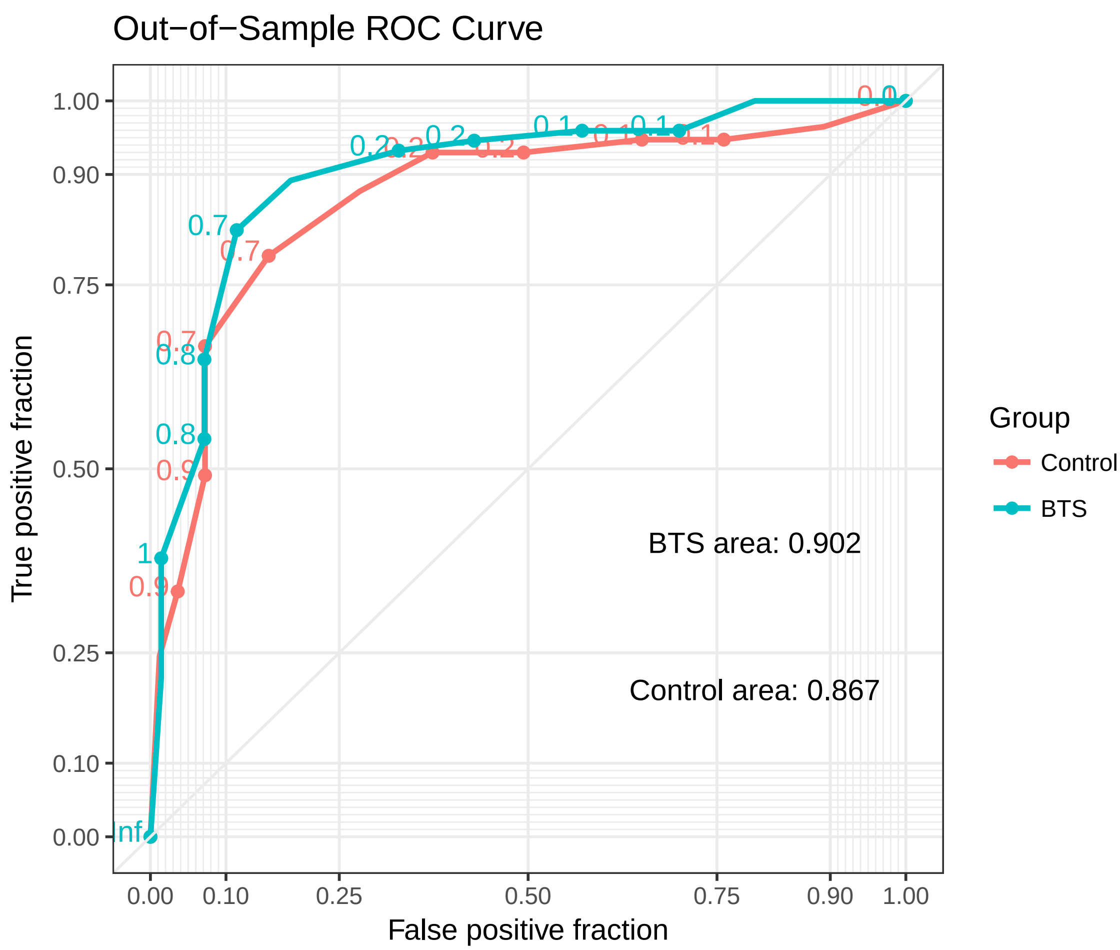


Receiver operating characteristic curve

ROC In-Sample Predictions



ROC Out-Of-Sample Predictions



Appendix #2 Code

Appendix 2. Code

```
library(reshape2)
library(mlogit)
library(effects)
library(stargazer)
library(mfx)
library(ggplot2)
library(tidyr)
library(ROCR)
library(pscl)
library(caret)
```

```
library(sjPlot)
library(ggplot2)
library(plotROC)
library(pROC)



# Reading in the survey data ---------------------------------------------


setwd(getwd())

s_data <- read.csv("data/20170816_data.csv", header = T, stringsAsFactors = F, na.strings = "")
s_data <- s_data[c(3:nrow(s_data)),]
s_data <- s_data[-which(is.na(s_data$Q46)),]

# Reading in the desing of the choice experiment -------------------------
design <- readRDS(file =  "data/design.rds")

full_design <- design$levels

names(full_design)[names(full_design)=="X1"] <- "device"
names(full_design)[names(full_design)=="X2"] <- "extent"
names(full_design)[names(full_design)=="X3"] <- "price"

full_design$vers <- NULL
full_design$task <- rep(x = 1:2,6)
full_design <- cbind(setNames( full_design[full_design$task==1,c("card","device","extent","price")],c
                     setNames( full_design[full_design$task==2,c("device","extent","price")], paste0(
full_design$card <- paste0("CJ",full_design$card)

#table(full_design$device1, full_design$device2)
#table(full_design$extent1, full_design$extent2)
#table(full_design$price1, full_design$price2)
#table(full_design$price1, full_design$extent1)
#table(full_design$price1, full_design$extent2)
#table(full_design$price2, full_design$extent1)
#table(full_design[,c("price1","price2")], full_design[,c("extent1","extent2")])
#table(design$levels$X1, design$levels$X2)
#table(design$levels$X1, design$levels$X3)
#table(design$levels$X2, design$levels$X3)
```

```r
# Data processing ----------------------------------------------------------

s_data$group <- NA
#s_data$group <- apply(s_data[,c("CJNT1","BCJNT1")], 1, FUN = function(x) ifelse(is.na(x),1,2))
s_data$group[which(s_data$CJNT12==1 | s_data$CJNT12==2)] <- 1
#s_data[which(s_data$CJNT1==1),"Q46"]
s_data$group[which(s_data$BCJNT12==1 | s_data$BCJNT12==2)] <- 2
#s_data[which(s_data$BCJNT1==1),"Q46"]

t_data <- as.data.frame(matrix(data = NA, nrow = 0, ncol = 14))
names(t_data) <- c("ResponseId","group", paste0("CJ",1:12))

demo <- s_data[, c("ResponseId","Q5","Q44","Q45","Q46")]

temp_c <- s_data[ s_data$group==1, c("ResponseId", "group",paste0("CJNT",1:12))]
temp_b <- s_data[ s_data$group==2, c("ResponseId", "group",paste0("BCJNT",1:12))]

t_data <- rbind(t_data, setNames(object = temp_c, nm = names(t_data)))
t_data <- rbind(t_data, setNames(object = temp_b, nm = names(t_data)))


prop.table(table(t_data$CJ1, t_data$group), 2)

prop.table(table(t_data$CJ4, t_data$group), 2)

prop.table(table(t_data$CJ7, t_data$group), 2)

prop.table(table(t_data$CJ9, t_data$group), 2)

t.test()
#head(t_data)


m_data <- melt(t_data, id.vars = c("ResponseId","group"))
names(m_data)[names(m_data)=="value"] <- "choice"
names(m_data)[names(m_data)=="variable"] <- "card"
m_data$choice <- paste0("choice",m_data$choice)

#remove NAs
m_data <- m_data[m_data$choice!="choiceNA",]
```

```
freq_result <- dcast(data = m_data, formula = card + group ~ choice, fun.aggregate = length)

freq_result <- cbind(freq_result, setNames(object = data.frame(round(t(apply(freq_result[,c("choice1"

plot_freq <- melt(freq_result[,c(1,2,5,6)], id.vars = c("card","group"))
plot_freq[plot_freq$group==1,"group"] <- "Control"
plot_freq[plot_freq$group==2,"group"] <- "BTS"


ggplot(plot_freq, aes(x = group, y = value, fill = variable)) +
  geom_bar(stat = 'identity', position = 'stack') + facet_grid(~ card)

final_all <- merge(m_data, full_design, by = "card")

names(final_all)[names(final_all)=="card"] <- "choiceid"
names(final_all)[names(final_all)=="ResponseId"] <- "id"

# Data preparation -----------------------------------------------------------

final_all[apply(expand.grid(c("device","extent","price"), 1:2), 1, paste, collapse="")] <-
  lapply(final_all[apply(expand.grid(c("device","extent","price"), 1:2), 1, paste, collapse="")], fun


final_all <- merge(final_all, demo, by.x = "id", by.y = "ResponseId")
final_all[, c("Q5", "Q44", "Q45", "Q46")] <- apply(final_all[, c("Q5", "Q44", "Q45", "Q46")], 2, as.n

final_all <- final_all[final_all$choice!="choiceNA",]


# Chen implementation -------------------------------------------------------
altvar <- list()

altvar$final_chen <- setNames(as.data.frame(matrix(data = NA, nrow = nrow(final_all), ncol = 4)), c('

#altvar$final_chen$device <- final_all$device1 - final_all$device2
#altvar$final_chen$extent <- final_all$extent1 - final_all$extent2
#altvar$final_chen$price <- final_all$price1 - final_all$price2

altvar$final_chen$device_1 <- final_all$device1
altvar$final_chen$extent_1 <- final_all$extent1
```

```r
altvar$final_chen$price_1  <- final_all$price1
altvar$final_chen$price_1N <- 6 + altvar$final_chen$price_1*2

altvar$final_chen$device_2 <- final_all$device2
altvar$final_chen$extent_2 <- final_all$extent2
altvar$final_chen$price_2  <- final_all$price2
altvar$final_chen$price_2N <- 6 + altvar$final_chen$price_2*2

#altvar$final_chen <- cbind(altvar$final_chen,
#setNames( data.frame(sapply(1:3, function(x) as.integer(x == final_all$device1))), paste0("device_1_
#setNames( data.frame(sapply(1:3, function(x) as.integer(x == final_all$device2))), paste0("device_2_
#setNames( data.frame(sapply(1:3, function(x) as.integer(x == final_all$extent1))), paste0("extent_1_
#setNames( data.frame(sapply(1:3, function(x) as.integer(x == final_all$extent2))), paste0("extent_2_
#)

altvar$final_chen <- cbind(altvar$final_chen,
                           setNames( data.frame(sapply(1:3, function(x) as.integer(x == final_all$dev
                           setNames( data.frame(sapply(1:3, function(x) as.integer(x == final_all$ext


altvar$final_chen$price <- altvar$final_chen$price_1N - altvar$final_chen$price_2N



altvar$final_chen$device_plot <-  final_all$device1 - final_all$device2
altvar$final_chen$extent_plot <-  final_all$extent1 - final_all$extent2
altvar$final_chen$price_plot <- 10+(altvar$final_chen$price_1N - altvar$final_chen$price_2N)/2

View(altvar$final_chen)


altvar$final_chen$choice <- ifelse(final_all$choice=="choice1", 1, 0)
altvar$final_chen$group <- final_all$group

View(altvar$final_chen)

altvar$final_chen_control <- altvar$final_chen[altvar$final_chen$group==1,]
altvar$final_chen_bts     <- altvar$final_chen[altvar$final_chen$group==2,]


altvar$model_full <- formula( choice ~  device1 + device2 + device3 + extent1 + extent2 + extent3 + p
altvar$model_bts  <- formula( choice ~  device1 + device2 + device3 + extent1 + extent2 + extent3 + p
```

```
altvar$results_all     <- glm( altvar$model_full, data = altvar$final_chen, family = binomial(link =
altvar$results_allbts  <- glm( altvar$model_bts, data = altvar$final_chen,  family = binomial(link =
altvar$results_control <- glm( altvar$model_full, data = altvar$final_chen,  family = binomial(link =
altvar$results_bts     <- glm( altvar$model_full, data = altvar$final_chen,  family = binomial(link =

altvar$nonas$model_full     <- formula( choice ~  device1 + device2 + extent1 + extent2 + price)
altvar$nonas$model_bts      <- formula( choice ~  device1 + device2 + extent1 + extent2 + price + gro
altvar$nonas$model_plot_bts <- formula( choice ~  device1 + device2 + extent_plot + price_plot + grou

altvar$nonas$results_all          <- glm( altvar$nonas$model_full,     data = altvar$final_chen,  fam
altvar$nonas$results_allbts       <- glm( altvar$nonas$model_bts,      data = altvar$final_chen,  fam
altvar$nonas$results_plot_allbts  <- glm( altvar$nonas$model_plot_bts, data = altvar$final_chen,  fam
altvar$nonas$results_control      <- glm( altvar$nonas$model_full,     data = altvar$final_chen,  fam
altvar$nonas$results_bts          <- glm( altvar$nonas$model_full,     data = altvar$final_chen,  fam


sum(abs(altvar$results_control$residuals))/length(altvar$results_control$residuals)
sum(abs(altvar$results_bts$residuals))/length(altvar$results_bts$residuals)

sum(sqrt(altvar$results_control$residuals^2))/length(altvar$results_control$residuals)
sum(sqrt(altvar$results_bts$residuals^2))/length(altvar$results_bts$residuals)


stargazer(
  altvar$results_all,
  altvar$results_allbts,
  altvar$results_control,
  altvar$results_bts,
  type = "text", column.labels = c("All","All*BTS", "Control", "BTS"))

# Alt implementation -------------------------------------------------------


final_all_mlogit <- mlogit.data(final_all,
                                choice = "choice",
                                shape = "wide",
                                varying = 5:10,
                                alt.levels = c("choice1", "choice2"),
                                #opposite = c("device","extent","price"),
                                sep = "")
```

```r
#final_all_mlogit$device <- factor(final_all_mlogit$device, labels = c("d","m","d+m"))
#final_all_mlogit$extent <- factor(final_all_mlogit$extent, labels = c("b","b+a","b+a+l"))
#final_all_mlogit$price <- factor(final_all_mlogit$price, labels = c("8","10","12"))

final_all_mlogit$price_8  <- as.logical(ifelse(final_all_mlogit$price==1,1,0))
final_all_mlogit$price_10 <- as.logical(ifelse(final_all_mlogit$price==2,1,0))
final_all_mlogit$price_12 <- as.logical(ifelse(final_all_mlogit$price==3,1,0))



final_all_mlogit$price_N <- ifelse(final_all_mlogit$price == 1,8,
                                   ifelse(final_all_mlogit$price == 2,10,
                                          ifelse(final_all_mlogit$price == 3,12,NA)))




final_all_mlogit$extent_1 <- as.logical(ifelse(final_all_mlogit$extent==1,1,0))
final_all_mlogit$extent_2 <- as.logical(ifelse(final_all_mlogit$extent==2,1,0))
final_all_mlogit$extent_3 <- as.logical(ifelse(final_all_mlogit$extent==3,1,0))

final_all_mlogit$device_1 <- as.logical(ifelse(final_all_mlogit$device==1,1,0))
final_all_mlogit$device_2 <- as.logical(ifelse(final_all_mlogit$device==2,1,0))
final_all_mlogit$device_3 <- as.logical(ifelse(final_all_mlogit$device==3,1,0))


#Effect coding
final_all_mlogit$device_e <- ifelse(final_all_mlogit$device == 1,1,
                                    ifelse(final_all_mlogit$device == 2,2,
                                           ifelse(final_all_mlogit$device == 3,3,NA)))
final_all_mlogit$extent_e <- ifelse(final_all_mlogit$extent == 1,1,
                                    ifelse(final_all_mlogit$extent == 2,0,
                                           ifelse(final_all_mlogit$extent == 3,-1,NA)))
final_all_mlogit$price_e <- ifelse(final_all_mlogit$price == 1,-1,
                                   ifelse(final_all_mlogit$price == 2,0,
                                          ifelse(final_all_mlogit$price == 3,1,NA)))


#final_all_mlogit_g1 <- final_all_mlogit[final_all_mlogit$group==1,]
#final_all_mlogit_g2 <- final_all_mlogit[final_all_mlogit$group==2,]
```

```r
#use this one to compare with alt ver
#model_formula <- formula( choice ~ device + extent + price )

#model_formula <- formula( choice ~  device_1 + device_2 + device_3 + extent_1 + extent_2 + extent_3
#model_formula <- formula( choice ~  device_1 + device_2 + device_3 + extent_1 + extent_2 + extent_3
#model_formula <- formula( choice ~  device_1 + device_2 + extent_1 + extent_2 + price_N )
#model_formula <- formula( choice ~  device_1 + device_2 + device_3 + extent_1 + extent_2 + extent_3
#model_formula <- formula( choice ~  device_1 + device_2 + device_1*device_2 + extent_1 + extent_2 +
#model_formula <- formula( choice ~  device_e + extent_e + price_e )

altvar$full_data$full_data_model_formula         <- formula( choice ~  device_e + extent_e + price_N
altvar$full_data$full_data_model_formula_bts_odr <- formula( choice ~  device_1 + device_2 + extent_e
altvar$full_data$full_data_model_formula_bts_mfx <- formula( choice ~  device_e + extent_e + price_N


altvar$full_data$results_fdata$all     <- glm( altvar$nonas$full_data_model_formula,         data = f
altvar$full_data$results_fdata$bts_odr <- glm( altvar$nonas$full_data_model_formula_bts_odr, data = f
altvar$full_data$results_fdata$bts_mfx <- glm( altvar$nonas$full_data_model_formula_bts_mfx, data = f
altvar$full_data$results_fdata$g1      <- glm( altvar$nonas$full_data_model_formula,         data = f
altvar$full_data$results_fdata$g2      <- glm( altvar$nonas$full_data_model_formula,         data = f

# Tests implementation ----------------------------------------------------

ref_notes <- "*Both: Desktop + Mobile; BH: Browsing History, AU: Application Usage, LI: Location Info
ref_labels_ormx <- c("Desktop (B: Both)", "Mobile (B: Both)", "Browsing History (B: BH + AU + LI)", "


# ODDs ratios
altvar$results_or$all_or    <- logitor( altvar$model_full, data = altvar$final_chen         )
altvar$results_or$allbts_or <- logitor( altvar$model_bts,  data = altvar$final_chen         )
altvar$results_or$g1_or     <- logitor( altvar$model_full, data = altvar$final_chen_control)
altvar$results_or$g2_or     <- logitor( altvar$model_full, data = altvar$final_chen_bts     )


altvar$results_or$table_or <- cbind(setNames(as.data.frame(c(altvar$results_or$all_or$oddsratio[1:5],
                                    setNames(as.data.frame(  altvar$results_or$allbts_or$oddsratio[1:
                                    setNames(as.data.frame(c(altvar$results_or$g1_or$oddsratio[1:5],
                                    setNames(as.data.frame(c(altvar$results_or$g2_or$oddsratio[1:5],

rownames(altvar$results_or$table_or) <- c(ref_labels_ormx,"group",paste0("group:",ref_labels_ormx))
```

```r
stargazer(altvar$results_or$table_or, type = "text", summary = FALSE, notes = ref_notes)


# Marginal effects
altvar$results_mfx$all_mfx    <- logitmfx( altvar$model_full, data = altvar$final_chen        )
altvar$results_mfx$allbts_mfx <- logitmfx( altvar$model_bts,  data = altvar$final_chen        )
altvar$results_mfx$g1_mfx     <- logitmfx( altvar$model_full, data = altvar$final_chen_control)
altvar$results_mfx$g2_mfx     <- logitmfx( altvar$model_full, data = altvar$final_chen_bts    )


altvar$results_mfx$table_mfx <- cbind(setNames(as.data.frame(c(altvar$results_mfx$all_mfx$mfxest[1:5]
                                      setNames(as.data.frame(  altvar$results_mfx$allbts_mfx$mfxest[1:
                                      setNames(as.data.frame(c(altvar$results_mfx$g1_mfx$mfxest[1:5],
                                      setNames(as.data.frame(c(altvar$results_mfx$g2_mfx$mfxest[1:5],

rownames(altvar$results_mfx$table_mfx ) <- c(ref_labels_ormx,"group",paste0("group:",ref_labels_ormx)
stargazer(altvar$results_mfx$table_mfx , type = "text", summary = FALSE, notes = ref_notes)



# in sample predictions ---------------------------------------------------

altvar$results_is <- list()

altvar$results_is$g1   <- predict(altvar$results_control , type='response')
altvar$results_is$g1_p <- altvar$results_is$g1
altvar$results_is$g1   <- ifelse(altvar$results_is$g1 > 0.5,1,0)
altvar$results_is$g1_results <- mean(altvar$results_is$g1 != altvar$final_chen_control$choice)

print(paste('Accuracy',1-altvar$results_is$g1_results))


altvar$results_is$g2   <- predict(altvar$results_bts , type='response')
altvar$results_is$g2_p <- altvar$results_is$g2
altvar$results_is$g2   <- ifelse(altvar$results_is$g2 > 0.5,1,0)
altvar$results_is$g2_results <- mean(altvar$results_is$g2 != altvar$final_chen_bts$choice)

print(paste('Accuracy',1-altvar$results_is$g2_results))


altvar$results_is$final_table <-setNames(data.frame(
```

```r
cbind( c(rep("Control",4),rep("BTS",4)),
       c(rep(c(0,1),4)),
       c(rep(c(0,0,1,1),2)),
rbind(
matrix(round(prop.table(table(altvar$results_is$g1, altvar$final_chen_control$choice)),3)),
matrix(round(prop.table(table(altvar$results_is$g2, altvar$final_chen_bts$choice    )),3))
))),
nm = c("Condition", "Predicted", "Actual", "Result"))

# out of sample predictions ----------------------------------------------

#altvar$results_os <- list()

#altvar$results_os$g1_obs <- sample(1:nrow(altvar$final_chen_control), nrow(altvar$final_chen_contro
#altvar$results_os$g2_obs <- sample(1:nrow(altvar$final_chen_bts),     nrow(altvar$final_chen_bts

altvar$results_os$g1_train <- altvar$final_chen_control[-altvar$results_os$g1_obs,]
altvar$results_os$g1_test  <- altvar$final_chen_control[ altvar$results_os$g1_obs,]
altvar$results_os$g2_train <- altvar$final_chen_bts[-altvar$results_os$g2_obs,]
altvar$results_os$g2_test  <- altvar$final_chen_bts[ altvar$results_os$g2_obs,]

altvar$results_os$g1_glm  <- glm( altvar$model_bts, data = altvar$results_os$g1_train, family = binom
altvar$results_os$g2_glm  <- glm( altvar$model_bts, data = altvar$results_os$g2_train, family = binom

altvar$results_os$g1_os        <- predict(altvar$results_os$g1_glm, newdata=altvar$results_os$g1_test
altvar$results_os$g1_os_p      <- altvar$results_os$g1_os
altvar$results_os$g1_os        <- ifelse(altvar$results_os$g1_os > 0.5,1,0)
altvar$results_os$g1_os_result <- mean(altvar$results_os$g1_os != altvar$results_os$g1_test$choice)

altvar$results_os$g2_os        <- predict(altvar$results_os$g2_glm, newdata=altvar$results_os$g2_test
altvar$results_os$g2_os_p      <- altvar$results_os$g2_os
altvar$results_os$g2_os        <- ifelse(altvar$results_os$g2_os > 0.5,1,0)
altvar$results_os$g2_os_result <- mean(altvar$results_os$g2_os != altvar$results_os$g2_test$choice)


print(paste('Accuracy',1-altvar$results_os$g1_os_result))
print(paste('Accuracy',1-altvar$results_os$g2_os_result))

round(prop.table(table(as.logical(altvar$results_os$g1_os), altvar$results_os$g1_test$choice)),2)
round(prop.table(table(as.logical(altvar$results_os$g2_os), altvar$results_os$g2_test$choice)),2)
```

```r
altvar$results_os$final_table <-setNames(data.frame(
  cbind( c(rep("Control",4),rep("BTS",4)),
         c(rep(c(0,1),4)),
         c(rep(c(0,0,1,1),2)),
         rbind(
           matrix(round(prop.table(table(altvar$results_os$g1_os, altvar$results_os$g1_test$choice)),
           matrix(round(prop.table(table(altvar$results_os$g2_os, altvar$results_os$g2_test$choice)),
         ))),
  nm = c("Condition", "Predicted", "Actual", "Result"))

# tests ------------------------------------------------------------------


pR2(results_models[["g1"]])
pR2(results_models[["g2"]])

# Residuals
altvar$residual_table <- setNames(
data.frame(
cbind(
  c("In-Sample","In-Sample","Out-of-Sample","Out-of-Sample"),
  c("Control","BTS","Control","BTS"),
round(c(
sum(abs( resid(altvar$results_control, type = "response")))/length(altvar$final_chen_control$choice),
sum(abs( resid(altvar$results_bts, type = "response")))/length(altvar$final_chen_bts$choice),
sum(abs( altvar$results_os$g1_test$choice - altvar$results_os$g1_os_p ))/length(altvar$results_os$g1_
sum(abs( altvar$results_os$g2_test$choice - altvar$results_os$g2_os_p ))/length(altvar$results_os$g2_
),3)),
c("","*","","*")),
c("Data","Model","Value","Lower")
)


# ROC Curve --------------------------------------------------------------
#
##TO DO
#
#ROCRpred <- prediction(fitresults$g1_os_p, as.numeric(final_os$g1_test$choice))
#ROCRperf <- performance(ROCRpred, 'tpr','fpr')
#plot(ROCRperf, colorize = TRUE, text.adj = c(-0.2,1.7))
```

```r
#
#
#ROCRpred <- prediction(fitresults$g2_os_p, as.numeric(final_os$g2_test$choice))
#ROCRperf <- performance(ROCRpred, 'tpr','fpr')
#plot(ROCRperf, colorize = TRUE, text.adj = c(-0.2,1.7))
#
#
#g <- roc(admit ~ prob, data = mydata)
#plot(g)
#
#
#
#library(ggplot2)
#library(plotROC)
#
#set.seed(2529)
#D.ex <- rbinom(200, size = 1, prob = .5)
#M1 <- rnorm(200, mean = D.ex, sd = .65)
#M2 <- rnorm(200, mean = D.ex, sd = 1.5)
#
#test <- data.frame(D = D.ex, D.str = c("Healthy", "Ill")[D.ex + 1],
#                   M1 = M1, M2 = M2, stringsAsFactors = FALSE)
#
#ggplot(melt_roc(test, "D", c("M1", "M2")),
#       aes(d = D, m = M, color = name)) +
#  geom_roc() +
#  style_roc()
#
#ggplot(melt_roc(setNames(data.frame(cbind(as.numeric(final_os$g2_test$choice),fitresults$g2_os_p,fi
#       aes(d = D, m = M, color = name)) +
#  geom_roc() +
#  style_roc()
#
# Plot time -----------------------------------------------------------------

bts_c <- list()

bts_c[["c_b0"]] <- altvar$results_control$coefficients[1]
bts_c[["c_d1"]] <- altvar$results_control$coefficients[2]
bts_c[["c_d2"]] <- altvar$results_control$coefficients[3]
bts_c[["c_e1"]] <- altvar$results_control$coefficients[5]
```

```
bts_c[["c_e2"]] <- altvar$results_control$coefficients[6]
bts_c[["c_p"]]  <- altvar$results_control$coefficients[8]

bts_c[["t_b0"]] <- altvar$results_bts$coefficients[1]
bts_c[["t_d1"]] <- altvar$results_bts$coefficients[2]
bts_c[["t_d2"]] <- altvar$results_bts$coefficients[3]
bts_c[["t_e1"]] <- altvar$results_bts$coefficients[5]
bts_c[["t_e2"]] <- altvar$results_bts$coefficients[6]
bts_c[["t_p"]]  <- altvar$results_bts$coefficients[8]

p_range <- seq(from=min(altvar$final_chen$price), to=max(altvar$final_chen$price), by=.01)

a_logits <- bts_c[["c_b0"]] +
  bts_c[["c_d1"]]*1 +
  bts_c[["c_d2"]]*0 +
  bts_c[["c_e1"]]*1 +
  bts_c[["c_e2"]]*0 +
  bts_c[["c_p"]] *p_range


b_logits <- bts_c[["t_b0"]] +
  bts_c[["t_d1"]]*1 +
  bts_c[["t_d2"]]*0 +
  bts_c[["t_e1"]]*1 +
  bts_c[["t_e2"]]*0 +
  bts_c[["t_p"]] *p_range

a_probs <- exp(a_logits)/(1 + exp(a_logits))
b_probs <- exp(b_logits)/(1 + exp(b_logits))


test <- confint(altvar$results_control, level = 0.95)

a_logits_lb <- test[1,1] + test[2,1]*1 + test[3,1]*0 + test[5,1]*1 + test[6,1]*0  + test[8,1]*p_range
a_logits_ub <- test[1,2] + test[2,2]*1 + test[3,2]*0 + test[5,2]*1 + test[6,2]*0  + test[8,2]*p_range
a_probs_lb <- exp(a_logits_lb)/(1 + exp(a_logits_lb))
a_probs_ub <- exp(a_logits_ub)/(1 + exp(a_logits_ub))
a_conf_int <- data.frame(lb=a_probs_lb, ub=a_probs_ub, price=p_range)
a_conf_int <- melt(a_conf_int)

bts_c[["plot.data_p1"]] <- data.frame(control=a_probs, bts=b_probs, price=p_range)
```

```r
bts_c[["plot.data_p1"]] <- gather(bts_c[["plot.data_p1"]], key=group, value=prob, c("control","bts"))


ggplot(data =bts_c[["plot.data_p1"]], aes(x=price, y=prob, color=group)) +
  geom_line(lwd=2) +
  labs(x="Price", y="P(choice)", title="Probability of product selection on different levels of price
  theme_gray()+
  scale_color_manual(labels = c("BTS", "Control"), values = c("#E69F00", "#56B4E9")) + scale_x_contin

#  geom_ribbon(a_conf_int, aes(ymin=lb, ymax=ub), alpha=0.3)

sjp.glm(fit =  altvar$results_control)
sjp.glm(fit =  altvar$results_bts)

sjp.glm(fit =  altvar$results_control, type = "slope")
sjp.glm(fit =  altvar$results_bts, type = "slope")

sjp.glm(fit =  altvar$results_control, type = "eff", show.ci = TRUE)
sjp.glm(fit =  altvar$results_bts, type = "eff",show.ci = TRUE)

sjp.glm(fit = altvar$results_control, type = "pred", vars = c("price","device1"), show.ci = TRUE)
sjp.glm(fit = altvar$results_bts, type = "pred", vars = c("price","device1"), show.ci = TRUE)
sjp.glm(fit = altvar$results_control, type = "pred", vars = c("price","device2"), show.ci = TRUE)
sjp.glm(fit = altvar$results_bts, type = "pred", vars = c("price","device2"), show.ci = TRUE)

sjp.glm(altvar$results_control, type = "pred", vars = c("price","device1"), facet.grid = FALSE, show.
sjp.glm(altvar$results_bts, type = "pred", vars = c("price","device1"), facet.grid = FALSE, show.ci =

sjp.glm(altvar$results_control, type = "pred", vars = c("extent1","device1"), facet.grid = FALSE, sho
sjp.glm(altvar$results_bts, type = "pred", vars = c("extent1","device1"), facet.grid = FALSE, show.ci

sjp.glm(altvar$results_control, type = "pred", vars = c("price","extent1"), facet.grid = FALSE, show.
sjp.glm(altvar$results_bts, type = "pred", vars = c("price","extent1"), facet.grid = FALSE, show.ci =



c_logits <- bts_c[["c_b0"]] +
  bts_c[["c_d1"]]*0 +
  bts_c[["c_d2"]]*0 +
  bts_c[["c_e1"]]*0 +
  bts_c[["c_e2"]]*0 +
```

```
  bts_c[["c_p"]] *p_range


d_logits <- bts_c[["t_b0"]] +
  bts_c[["t_d1"]]*0 +
  bts_c[["t_d2"]]*0 +
  bts_c[["t_e1"]]*0 +
  bts_c[["t_e2"]]*0 +
  bts_c[["t_p"]] *p_range

c_probs <- exp(c_logits)/(1 + exp(c_logits))
d_probs <- exp(d_logits)/(1 + exp(d_logits))


bts_c[["plot.data_p2"]] <- data.frame(control=c_probs, bts=d_probs, price=p_range)
bts_c[["plot.data_p2"]] <- gather(bts_c[["plot.data_p2"]], key=group, value=prob, c("control","bts"))


ggplot(bts_c[["plot.data_p2"]], aes(x=price, y=prob, color=group)) +
  geom_line(lwd=2) +
  labs(x="Price", y="P(choice)", title="Probability of product selection on different levels of price
  theme_gray()+
  scale_color_manual(labels = c("BTS", "Control"), values = c("#E69F00", "#56B4E9"))+ scale_x_continu


# ROC Curve ---------------------------------------------------------------


#altvar$results_is$g1_p != altvar$final_chen_control$choice
#altvar$results_is$g2_p != altvar$final_chen_bts$choice
#
#altvar$results_os$g1_os_p != altvar$results_os$g1_test$choice
#altvar$results_os$g2_os_p != altvar$results_os$g2_test$choice

altvar$roc$is_area_c   <- paste("Control area:", round(auc(altvar$final_chen_control$choice, altvar$r
altvar$roc$is_area_bts <- paste("BTS area:",     round(auc(altvar$final_chen_bts$choice, altvar$resul

altvar$roc$is_contol <- data.frame(D = altvar$final_chen_control$choice, M = altvar$results_is$g1_p,
altvar$roc$is_bts    <- data.frame(D = altvar$final_chen_bts$choice, M = altvar$results_is$g2_p, grou
altvar$roc$is_data <- rbind(altvar$roc$is_contol, altvar$roc$is_bts)
```

```
ggplot(altvar$roc$is_data, aes(d = D, m = M, color = group)) + geom_roc() + style_roc() +  labs(title
  annotate(geom = "text", x = 0.8, y = 0.2, label = altvar$roc$is_area_c) +
  annotate(geom = "text", x = 0.8, y = 0.4, label = altvar$roc$is_area_bts)

altvar$roc$os_area_c   <- paste("Control area:", round(auc(altvar$results_os$g1_test$choice, altvar$r
altvar$roc$os_area_bts <- paste("BTS area:",     round(auc(altvar$results_os$g2_test$choice, altvar$r

altvar$roc$os_contol <- data.frame(D = altvar$results_os$g1_test$choice, M = altvar$results_os$g1_os_
altvar$roc$os_bts    <- data.frame(D = altvar$results_os$g2_test$choice, M = altvar$results_os$g2_os_
altvar$roc$os_data <- rbind(altvar$roc$os_contol, altvar$roc$os_bts)

ggplot(altvar$roc$os_data, aes(d = D, m = M, color = group)) + geom_roc() + style_roc() +  labs(title
annotate(geom = "text", x = 0.8, y = 0.2, label = altvar$roc$os_area_c) +
annotate(geom = "text", x = 0.8, y = 0.4, label = altvar$roc$os_area_bts)
```

Appendix #3 Design and Survey

Appendix 3. Design and Survey

Conjoint Design

Table 0.10: Conjoint Design

card	device1	extent1	price1	device2	extent2	price2
CJ1	2	1	2	2	1	1
CJ2	2	3	2	1	2	2
CJ3	3	1	1	2	2	3
CJ4	1	3	3	3	3	2
CJ5	1	2	3	1	1	2
CJ6	3	1	3	1	3	1
CJ7	2	3	3	2	1	3
CJ8	2	3	1	2	2	2
CJ9	3	2	1	2	2	1
CJ10	3	1	2	1	1	3
CJ11	1	1	1	3	3	3
CJ12	3	2	3	1	3	2

Full Survey

The full survey appears at the end of the document

References

Baillon, Aurélien. 2015. "Subjective Truths." E. http://hdl.handle.net/1765/78157.

Barrage, Lint, and Min Sok Lee. 2010. "A Penny for Your Thoughts: Inducing Truth-Telling in Stated Preference Elicitation." *Economics Letters* 106 (2). Elsevier: 140–42.

Becker, Gordon M, Morris H DeGroot, and Jacob Marschak. 1964. "Measuring Utility by a Single Response Sequential Method." *Behavioral Science* 9 (3). Wiley Online Library: 226–32.

Carman, Katherine Grace, and Peter Kooreman. 2010. "Flu Shots, Mammogram, and the Perception of Probabilities."

Carmone, Frank J, Paul E Green, and Arun K Jain. 1978. "Robustness of Conjoint Analysis: Some Monte Carlo Results." *Journal of Marketing Research*. JSTOR, 300–303.

Carson, Richard T, and Theodore Groves. 2007. "Incentive and Informational Properties of Preference Questions." *Environmental and Resource Economics* 37 (1). Springer: 181–210.

Chrzan, Keith, and Bryan Orme. 2000. "An Overview and Comparison of Design Strategies for Choice-Based Conjoint Analysis." *Sawtooth Software Research Paper Series*.

Dahan, Ely, and V Srinivasan. 2000. "The Predictive Power of Internet-Based Product Concept Testing Using Visual Depiction and Animation." *Journal of Product Innovation Management* 17 (2). Wiley Online Library: 99–109.

Dawes, Robyn M. 1989. "Statistical Criteria for Establishing a Truly False Consensus Effect." *Journal of Experimental Social Psychology* 25 (1). Elsevier: 1–17.

Ding, Min. 2007. "An Incentive-Aligned Mechanism for Conjoint Analysis." *Journal of Marketing Research* 44 (2). American Marketing Association: 214–23.

Ding, Min, Rajdeep Grewal, and John Liechty. 2005. "Incentive-Aligned Conjoint Analysis." *Journal of Marketing Research* 42 (1). American Marketing Association: 67–82.

Dong, Songting, Min Ding, and Joel Huber. 2010. "A Simple Mechanism to Incentive-Align Conjoint Experiments." *International Journal of Research in Marketing* 27 (1). Elsevier: 25–32.

Fedorov, Valerii Vadimovich. 1972. *Theory of Optimal Experiments*. Elsevier.

Gittelman, Steven H, and Elaine Trimarchi. 2012. "Rules of Engagement the War Against Poorly Engaged Respondents Guidelines for Elimination." Esomar.

Golbeck, Jennifer, and Matthew Louis Mauriello. 2016. "User Perception of Facebook App Data Access: A Comparison of Methods and Privacy Concerns." *Future Internet* 8 (2). Multidisciplinary Digital Publishing Institute: 9.

Green, Paul E. 1974. "On the Design of Choice Experiments Involving Multifactor Alternatives." *Journal of Consumer Research* 1 (2). The Oxford University Press: 61–68.

Green, Paul E, and Vithala R Rao. 1971. "Conjoint Measurement for Quantifying Judgmental Data."

Journal of Marketing Research. JSTOR, 355–63.

Green, Paul E, and Venkat Srinivasan. 1990. "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice." *The Journal of Marketing*. JSTOR, 3–19.

Green, Paul E, and Venkatachary Srinivasan. 1978. "Conjoint Analysis in Consumer Research: Issues and Outlook." *Journal of Consumer Research* 5 (2). The Oxford University Press: 103–23.

Green, Paul E, Abba M Krieger, and Yoram Wind. 2001. "Thirty Years of Conjoint Analysis: Reflections and Prospects." *Interfaces* 31 (3 supplement). INFORMS: S56–S73.

Guiso, Luigi, and Giuseppe Parigi. 1999. "Investment and Demand Uncertainty." *The Quarterly Journal of Economics* 114 (1). Oxford University Press: 185–227.

Guiso, Luigi, Tullio Jappelli, and Daniele Terlizzese. 1992. "Earnings Uncertainty and Precautionary Saving." *Journal of Monetary Economics* 30 (2). Elsevier: 307–37.

Hauser, John R. 2007. "A Note on Conjoint Analysis." *Massachusetts Institute of Technology, MIT Sloan Courseware*.

Hauser, John R, and Vithala R Rao. 2004. "Conjoint Analysis, Related Modeling, and Applications." In *Marketing Research and Modeling: Progress and Prospects*, 141–68. Springer.

Horne, Jack. 2014. "Package 'ChoiceDes'."

Howie, Patrick J, Ying Wang, and Joanne Tsai. 2011. "Predicting New Product Adoption Using Bayesian Truth Serum." *Journal of Medical Marketing: Device, Diagnostic and Pharmaceutical Marketing* 11 (1). Sage Publications: 6–16.

Hurd, Michael D. 2009. "Subjective Probabilities in Household Surveys." *Annu. Rev. Econ.* 1 (1). Annual Reviews: 543–62.

John, Leslie K, George Loewenstein, and Drazen Prelec. 2012. "Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling." *Psychological Science*. Sage Publications, 0956797611430953.

Johnson, F Reed, Emily Lancsar, Deborah Marshall, Vikram Kilambi, Axel Mühlbacher, Dean A Regier, Brian W Bresnahan, Barbara Kanninen, and John FP Bridges. 2013. "Constructing Experimental Designs for Discrete-Choice Experiments: Report of the Ispor Conjoint Analysis Experimental Design Good Research Practices Task Force." *Value in Health* 16 (1). Elsevier: 3–13.

Krantz, David, Duncan Luce, Patrick Suppes, and Amos Tversky. 1971. *Foundations of Measurement*. academic press.

Krumpal, Ivar. 2013. "Determinants of Social Desirability Bias in Sensitive Surveys: A Literature Review." *Quality & Quantity*. Springer, 1–23.

Kukla-Gryz, Anna, Joanna Tyrowicz, Michal Krawczyk, and Konrad Siwinski. 2015. "We All Do It, but Are We Willing to Admit? Incentivizing Digital Pirates' Confessions." *Applied Economics Letters*

22 (3). Taylor; Francis: 184–88.

Liechty, John C, Duncan KH Fong, and Wayne S DeSarbo. 2005. “Dynamic Models Incorporating Individual Heterogeneity: Utility Evolution in Conjoint Analysis.” *Marketing Science* 24 (2). INFORMS: 285–93.

Louviere, Jordan J. 1988. “Conjoint Analysis Modelling of Stated Preferences: A Review of Theory, Methods, Recent Developments and External Validity.” *Journal of Transport Economics and Policy*. JSTOR, 93–119.

Louviere, Jordan J, and George Woodworth. 1983. “Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data.” *Journal of Marketing Research*. JSTOR, 350–67.

McFadden, Daniel, and others. 1973. “Conditional Logit Analysis of Qualitative Choice Behavior.” Institute of Urban; Regional Development, University of California.

Netzer, Oded, Olivier Toubia, Eric T Bradlow, Ely Dahan, Theodoros Evgeniou, Fred M Feinberg, Eleanor M Feit, et al. 2008. “Beyond Conjoint Analysis: Advances in Preference Measurement.” *Marketing Letters* 19 (3-4). Springer: 337–54.

Noussair, Charles N, Stefan T Trautmann, and Gijs van de Kuilen. 2014. “THigher Order Risk Attitudes.” *Demographics, and Financial Decisions, UReview of Eco*.

Park, Young-Hoon, Min Ding, and Vithala R Rao. 2008. “Eliciting Preference for Complex Products a Web-Based Upgrading Method.” *Journal of Marketing Research* 45 (5). American Marketing Association: 562–74.

Prelec, Drazen. 2004. “A Bayesian Truth Serum for Subjective Data.” *Science* 306 (5695). American Association for the Advancement of Science: 462–66.

Puleston, J, and MITCH Eggers. 2012. “Dimensions of Online Survey Data Quality. What Really Matters?” In *ESOMAR, Congress Atlanta*.

Ross, Lee, David Greene, and Pamela House. 1977. “The ‘False Consensus Effect’: An Egocentric Bias in Social Perception and Attribution Processes.” *Journal of Experimental Social Psychology* 13 (3). Elsevier: 279–301.

Scott, Anthony. 1965. “The Valuation of Game Resources: Some Theoretical Aspects.” *Canadian Fisheries Report* 4: 27–47.

Smith, John Maynard. 1982. *Evolution and the Theory of Games*. Cambridge university press.

Smith, John Maynard, and Geoffrey A Parker. 1976. “The Logic of Asymmetric Contests.” *Animal Behaviour* 24 (1). Elsevier: 159–75.

Toubia, Olivier, Martijn G De Jong, Daniel Stieger, and Johann Fuller. 2012. “Measuring Consumer Preferences Using Conjoint Poker.” *Marketing Science* 31 (1). INFORMS: 138–56.

Walker, Joan, and Moshe Ben-Akiva. 2002. “Generalized Random Utility Model.” *Mathematical*

Social Sciences 43 (3). Elsevier: 303–43.

Weaver, Ray, and Drazen Prelec. 2013. "Creating Truth-Telling Incentives with the Bayesian Truth Serum." *Journal of Marketing Research* 50 (3). American Marketing Association: 289–302.

Wheeler, Bob. 2014. "AlgDesign: Algorithmic Experimental Design." *R Package Version*, 1–1.

BTS Welcome

Welcome!

Nowadays the world is all about **data**. This survey contains 14 questions about your attitude towards your **willingness to sell your digital behavioral data.** The digital behavioral data capture individual actions, interactions, and movements in a given website.

In order to assure your **truthfulness**, your responses will be evaluated against a **scoring formula** where **truthful answers receive higher score** comparing to non-truthful ones.

If **your answers are the most truthful ones** according to the formula you will receive the amount of **20 euro**. You can leave your email at the end of the survey.

If you have any questions please forward them to:

Grigor Dimitrov
grdimitrov@gmail.com

Non BTS Welcome

Welcome!

Nowadays the world is all about **data**. This survey contains 14 questions about your attitude towards your **willingness to sell your digital behavioral data.** The digital behavioral data capture individual actions, interactions, and movements in a given website.

In order to **compensate** you for your time and effort a random respondent will be selected to receive the amount of **20 euro**.

In case you want to participate in the draw you can fill in your email at the end of the survey.

If you have any questions please forward them to:

Grigor Dimitrov
grdimitrov@gmail.com

WelcomeBlock

Datatron is a new company entering on the market of **internet data collection and trading**.

Datatron ***tracks*** the **internet usage of individuals**, collects the data and sells it to its partners. **Individuals are paid for sharing their digital behavioral data.**

How does it works?

After giving a **consent** individuals **install** a **tracking application on their devices**. The tracking application sends all of the visited websites and applications usage to the **Datatron** servers.

What is **Datatron**?

Internet provider company	Behavioral data collection company	Survey company
○	○	○

Given that:

1. Different social media are selling your data to third party companies, but you are not reimbursed with any monetary incentives for your data.
2. Your data only costs as much as someone is ready to pay for it

Will you consider to sell your data for **any** amount?

- Yes
- Maybe
- No

Now **Datatron** wants to buy **your** data!

In the next exercises you will make a choice between **two** offerings from Datatron.

The offerings vary across:

1. Amount of **Euro** you are offered
2. **Devices**:
   - Your home computer
   - Your mobile device
   - Both Home computer + mobile device
3. **Extent of data collection**:
   - Your internet usage (Visited websites)
   - Your Internet and online application usage (Facebook, Instagram, Maps etc.)
   - Internet usage, applications + Your location

Keep in mind, **individuals'** data will:

- Never be reported outside of a group
- Always be stripped down of sensitive information which can lead to revealing the personality of the subject

CJNT Tasks - Jul 12, 2017

If the Datatron offers you these two products, which one would you prefer.

- 10 Euro
  Mobile data
  Internet usage

- 8 Euro
  Mobile data
  Internet usage

If the Datatron offers you these two products, which one would you prefer.

- 10 Euro
  Mobile data
  Internet usage
  Apps usage
  Location
- 10 Euro
  Desktop data
  Internet usage
  Applications usage

If the Datatron offers you these two products, which one would you prefer.

- 8 Euro
  Desktop data
  Mobile data
  Internet usage
- 12 Euro
  Mobile data
  Internet usage
  Applications usage

If the Datatron offers you these two products, which one would you prefer.

- 12 Euro
  Desktop data
  Internet usage
  Apps usage
  Location
- 10 Euro
  Desktop data
  Mobile data
  Internet usage
  Apps usage
  Location

If the Datatron offers you these two products, which one would you prefer.

- 12 Euro
  Desktop data
  Internet usage
  Applications usage
- 10 Euro
  Desktop data
  Internet usage

If the Datatron offers you these two products, which one would you prefer.

- 12 Euro
  Desktop data
  Mobile data
  Internet usage
- 8 Euro
  Desktop data
  Internet usage
  Apps usage
  Location

If the Datatron offers you these two products, which one would you prefer.

- 12 Euro
  Mobile data
  Internet usage
  Apps usage
  Location
- 12 Euro
  Mobile data
  Internet usage

If the Datatron offers you these two products, which one would you prefer.

- 8 Euro
  Mobile data
  Internet usage
  Apps usage
  Location

- 10 Euro
  Mobile data
  Internet usage
  Applications usage

If the Datatron offers you these two products, which one would you prefer.

- 8 Euro
  Desktop data
  Mobile data
  Internet usage
  Applications usage
- 8 Euro
  Mobile data
  Internet usage
  Applications usage

If the Datatron offers you these two products, which one would you prefer.

- 10 Euro
  Desktop data
  Mobile data
  Internet usage
- 12 Euro
  Desktop data
  Internet usage

If the Datatron offers you these two products, which one would you prefer.

- 8 Euro
  Desktop data
  Internet usage
- 12 Euro
  Desktop data
  Mobile data
  Internet usage
  Apps usage
  Location

If the Datatron offers you these two products, which one would you prefer.

- 12 Euro
  Desktop data
  Mobile data
  Internet usage
  Applications usage
- 10 Euro
  Desktop data
  Internet usage
  Apps usage
  Location

BTS CJNT Tasks - Jul 12, 2017

If the company offers you these two products, which one would you prefer.

- Mobile data
  Internet usage
  10 Euro
- Mobile data
  Internet usage
  8 Euro

Think about the rest of the respondents participating in the survey. What percentage of them do you feel are going to select the same option as you just did?

- 1 - 10%
- 11 - 20%
- 21 - 30%
- 31 - 40%
- 41 - 50%
- 51 - 60%
- 61 - 70%
- 71 - 80%
- 81 - 90%
- 91 - 100%

If the company offers you these two products, which one would you prefer.

- Mobile data
  Internet usage
  Apps usage
  Location
  10 Euro

- Desktop data
  Internet usage
  Applications usage
  10 Euro

Think about the rest of the respondents participating in the survey. What percentage of them do you feel are going to select the same option as you just did?

- 1 - 10%
- 11 - 20%
- 21 - 30%
- 31 - 40%
- 41 - 50%
- 51 - 60%
- 61 - 70%
- 71 - 80%
- 81 - 90%
- 91 - 100%

If the company offers you these two products, which one would you prefer.

- Desktop data
  Mobile data
  Internet usage
  8 Euro
- Mobile data
  Internet usage
  Applications usage
  12 Euro

Think about the rest of the respondents participating in the survey. What percentage of them do you feel are going to select the same option as you just did?

- 1 - 10%
- 11 - 20%
- 21 - 30%
- 31 - 40%
- 41 - 50%
- 51 - 60%
- 61 - 70%
- 71 - 80%
- 81 - 90%
- 91 - 100%

If the company offers you these two products, which one would you prefer.

- ○ Desktop data
  Internet usage
  Apps usage
  Location
  12 Euro
- ○ Desktop data
  Mobile data
  Internet usage
  Apps usage
  Location
  10 Euro

Think about the rest of the respondents participating in the survey. What percentage of them do you feel are going to select the same option as you just did?

- ○ 1 - 10%
- ○ 11 - 20%
- ○ 21 - 30%
- ○ 31 - 40%
- ○ 41 - 50%
- ○ 51 - 60%
- ○ 61 - 70%
- ○ 71 - 80%
- ○ 81 - 90%
- ○ 91 - 100%

If the company offers you these two products, which one would you prefer.

- ○ Desktop data
  Internet usage
  Applications usage
  12 Euro
- ○ Desktop data
  Internet usage
  10 Euro

Think about the rest of the respondents participating in the survey. What percentage of them do you feel are going to select the same option as you just did?

- ○ 1 - 10%
- ○ 11 - 20%
- ○ 51 - 60%
- ○ 61 - 70%

- 21 - 30%
- 31 - 40%
- 41 - 50%
- 71 - 80%
- 81 - 90%
- 91 - 100%

If the company offers you these two products, which one would you prefer.

- Desktop data
  Mobile data
  Internet usage
  12 Euro
- Desktop data
  Internet usage
  Apps usage
  Location
  8 Euro

Think about the rest of the respondents participating in the survey. What percentage of them do you feel are going to select the same option as you just did?

- 1 - 10%
- 11 - 20%
- 21 - 30%
- 31 - 40%
- 41 - 50%
- 51 - 60%
- 61 - 70%
- 71 - 80%
- 81 - 90%
- 91 - 100%

If the company offers you these two products, which one would you prefer.

- Mobile data
  Internet usage
  Apps usage
  Location
  12 Euro
- Mobile data
  Internet usage
  12 Euro

Think about the rest of the respondents participating in the survey. What percentage of them do you feel are going to select the same option as you just did?

- 1 - 10%
- 11 - 20%
- 21 - 30%
- 31 - 40%
- 41 - 50%
- 51 - 60%
- 61 - 70%
- 71 - 80%
- 81 - 90%
- 91 - 100%

If the company offers you these two products, which one would you prefer.

- Mobile data
  Internet usage
  Apps usage
  Location
  8 Euro
- Mobile data
  Internet usage
  Applications usage
  10 Euro

Think about the rest of the respondents participating in the survey. What percentage of them do you feel are going to select the same option as you just did?

- 1 - 10%
- 11 - 20%
- 21 - 30%
- 31 - 40%
- 41 - 50%
- 51 - 60%
- 61 - 70%
- 71 - 80%
- 81 - 90%
- 91 - 100%

If the company offers you these two products, which one would you prefer.

- Desktop data
  Mobile data
  Internet usage
  Applications usage
  8 Euro
- Mobile data
  Internet usage
  Applications usage
  8 Euro

Think about the rest of the respondents participating in the survey. What percentage of them do you feel are going to select the same option as you just did?

- 1 - 10%
- 11 - 20%
- 21 - 30%
- 31 - 40%
- 41 - 50%
- 51 - 60%
- 61 - 70%
- 71 - 80%
- 81 - 90%
- 91 - 100%

If the company offers you these two products, which one would you prefer.

- Desktop data
  Mobile data
  Internet usage
  10 Euro
- Desktop data
  Internet usage
  12 Euro

Think about the rest of the respondents participating in the survey. What percentage of them do you feel are going to select the same option as you just did?

- 1 - 10%
- 11 - 20%
- 21 - 30%
- 31 - 40%
- 41 - 50%
- 51 - 60%
- 61 - 70%
- 71 - 80%
- 81 - 90%
- 91 - 100%

If the company offers you these two products, which one would you prefer.

- Desktop data
  Internet usage
  8 Euro
- Desktop data
  Mobile data
  Internet usage
  Apps usage
  Location
  12 Euro

Think about the rest of the respondents participating in the survey. What percentage of them do you feel are going to select the same option as you just did?

- 1 - 10%
- 11 - 20%
- 21 - 30%
- 31 - 40%
- 41 - 50%
- 51 - 60%
- 61 - 70%
- 71 - 80%
- 81 - 90%
- 91 - 100%

If the company offers you these two products, which one would you prefer.

- Desktop data
  Mobile data
  Internet usage
  Applications usage
  12 Euro
- Desktop data
  Internet usage
  Apps usage
  Location
  10 Euro

Think about the rest of the respondents participating in the survey. What percentage of them do you feel are going to select the same option as you just did?

- 1 - 10%
- 11 - 20%
- 21 - 30%
- 31 - 40%
- 41 - 50%
- 51 - 60%
- 61 - 70%
- 71 - 80%
- 81 - 90%
- 91 - 100%

END

Of course, Datatron is a fictional company.

You are almost done, only some demographic questions left.

Please enter your e-mail in order to participate for the draw.

- [ ] I'm curious and I want to receive the results

Gender

- Male
- Female

Age

- Under 18
- 18 - 24
- 25 - 34
- 35 - 44
- 45 - 54
- 55 - 64
- 65 - 74
- 75 - 84
- 85 or older

Employment

- Student
- Employed