Master Thesis
Spring 2009

An Empirical Investigation of the Determinants of Attention to Attributes in Choice Experiments

Antonin Danalet
<antonin.danalet@a3.epfl.ch>

SUPERVISORS:

Prof. Michel Bierlaire
Laboratoire Transport et Mobilité
Ecole Polytechnique Fédérale de Lausanne

Prof. Riccardo Scarpa
Waikato Management School
The University of Waikato

"Popper's exile, in the late 1930s, in Christchurch, New Zealand, provided him with a nice personal experience of the power of falsification. While in Europe, he had taken it for granted that all swans were white, but in Christchurch he found to his surprise, as everyone who has visited Christchurch knows, that some swans were black! Against his own philosophical insight, he had fallen into the trap of inductive reasoning and had wrongly generalized his previous experience of white swans."

W. Ulrich[a]

[a] *Rethinking Critically Reflective Research Practice: Beyond Popper's Critical Rationalism*, Journal of Research Practice, Volume 2, Issue 2, Article P1, 2006

Acknowledgments

This project has been completed with the aid of a grant from the Schindler Foundation.

I would like to thank Riccardo Scarpa, who welcomed me in Waikato and helped me with this work, providing guidance and support.

I also would like to thank Bill Kaye-Blake for welcoming me in Lincoln University, for his data and for his careful reading of this thesis, Mara Thiene and Jurgen Meyerhoff for their data, and Michel Bierlaire who gave me the opportunity to do my Master Thesis in New Zealand.

Abstract

Recent results showed that taking into account attention to attributes in random utility models leads to better fitting models. In this paper, we study a proposition to model attention to attributes instead of monitoring it. We test this model considering attention as a choice. The results show that these factors have a low power of prediction of stated attention to single attributes.

Contents

Introduction

> "On the one hand there is your local English professor; your great-aunt Irma, who never married and liberally delivers sermons; your how-to-reach-happiness-in-twenty-steps and how-to-become-a-better-person-in-a-week book writer. It is called the Utopian Vision, associated with Rousseau, Godwin, Condorcet, Thomas Paine, and conventional normative economists (of the kind who ask you to make rational choices because that is what is deemed good for you), etc. They believe in reason and rationality – that we should overcome cultural impediments on our way to becoming a better human race – thinking we can control our nature at will and transform it by mere edict in order to attain, among other things, happiness and rationality. Basically this category would include those who think that the cure for obesity is to inform people that they should be healthy.
>
> On the other hand there is the Tragic Vision of humankind that believes in the existence of inherent limitations and flaws in the way we think and act and requires an acknowledgment of this fact as a basis for any individual and collective action. This category of people include Karl Popper (...), Friedrich Hayek and Milton Friedman (...), Herbert Simon (bounded rationality), Amos Tversky and Daniel Kahneman (...), the speculator George Soros, etc."

This quote is from Fooled by Randomness by Nassim Nicholas Taleb [29] that I read towards the beginning of my writing of this Master Thesis. Although I have a better opinion about the Utopioan Vision of humankind than Taleb – thinking we sometimes need Utopia for creativity, emotions and dreams – the analysis provided in by the paper is definitively tragic.

In 1955, Herbert Simon wrote in his foundation paper about bounded rationality [25]: "Recent developments in economics, and particularly in the theory of the business firm, have raised great doubts as to whether this schematized model of economic man [i.e., rational] provides a suitable foundation on which to erect a theory". Nowadays, recent developments tend to continue to raise great doubts about the rationality of the economic man.

Trying to avoid the obvious example of the current economic crisis, here is a quick example read in the Science magazine during the preparation for this Master Thesis. Antonakis [1] showed that people looking at pictures of unknown individuals choose the winner of a real election with a success rate of 70%. The importance of "looking good" is not a complete surprise since it was already a subject of discussion in Antic Greece, but it poses the question of the cost of processing information about different alternatives in a context of choice. The information about facial expression is much easier to gather than the political program or the different positions of candidates available in newspapers.

Choice making is one of the fundamental actions of human being and is important in politics, management, economics, or love. Most classical economics models consider that the economic man is rational and information is free. Lack of information for a decision maker has long been studied in game theory. For example with Aumann's theory of correlated equilibrium, in the framework of discrete choice models most analysts still live in a perfect world with rational choice-maker who have a perfect knowledge of their environment.

This paper tries to focus on theory of discrete choice models with some digressions about bounded rationality. The focus is on attention to attributes and the ability to model it in order to improve models of discrete choice.

The first chapter introduces bounded rationality and some recent applications of this concept. Then a number of articles on attention to attributes in random utility models are reviewed. In chapter two, I present a theoretical model developed by Cameron in a recent working paper. Then, in chapter three I describe a methodology where we consider attendance as a choice. Finally, in the two last chapters I present the different data sets I worked with and the results before concluding.

Chapter 1

Literature

It is impossible to begin this review of the literature with anybody else than Herbert Simon, whose name was in Taleb's quote in the introduction, and his "tragic vision of humankind", bounded rationality[1].

The seminal article about bounded rationality, even if the expression was not used in this paper, was published in 1955 by Simon, "A behavioral Model of Rational Choice". In this article Simon proposed some limits to the classical models of choice. He presented the "economic man" who is known to be rational and defined rationality as a "clear and voluminous" knowledge of the environment by this man, a stable system of preferences and the skill to reach the highest point in his preference scale, i.e., a capacity to maximize his utility. Simon showed doubts about this man and proposed to replace him with "a kind of rational behavior" taking into account access to information and computation capabilities in reality. He underlined the lack of empirical knowledge about this behavior and assumed a common experience to be able to discuss the subject. The goal of his paper is partially to "take into account the simplification the choosing organism may deliberately introduce into its model of the situation in order to bring the model within the range of its computing capacity". Simon presented common elements of his model with the more global models, the classical concepts of rationality: a set of alternatives A, a subset of perceived alternatives $\dot{A} \subset A$, the set of outcomes of choice S, a pay-off function representing the value or utility of each outcome of choice $V(s)\ \forall s \in S$, and information about which outcome would occur if a particular alternative $a \in A$ is chosen. Then he introduced simplifications, approximating procedures. These simplifications apply to pay-off functions, information gathering and ordering of pay-offs, assumed here to be

[1] Note however that concepts of limited cognition were available before 1955, with "limited intelligence" in 1840, "finite intelligence" in 1880, "incomplete rationality" in 1922, "limited rationality" in 1945, "administrative rationality" in 1945 or "approximate rationality" in 1948. For history of the emergence of bounded rationality, see [17]. In [9] the focus is on the period 1981-1996 but we can though read: "*It is evident that the rational thing to do is to be irrational, where deliberation and estimation cost more than they are worth.* Frank Knight (1921, p. 67, footnote)"

partial[2].

Simon dismissed neoclassical theory and proposed a behavioral approach. He used the name "bounded rationality" for the first time in 1957 in Models of Man [26] and used it to "designate rational choice that takes into account the cognitive limitations of the decision-maker-limitations of both knowledge and computational capacity" [27], p. 291. In other words, this theory basically admits suboptimal decisions that are indeed rational in a limited decision making environment. Simon defined two "kinds" of rationality. *Substantive* rationality means that the choice maker is never satisfied with anything less than the optimum, the best possible choice. It also means this economic man has a clear notion of success in a choice and knows when he has done the best choice. It is a result-oriented rationality. Conversely *procedural* rationality assumes that the choice maker use a process to deliberate and make a decision but with time and computation constraints that could vary. It is a form of process-oriented rationality.

Nowadays bounded rationality is broadly cited – and so is the term "satisficing", also one of Simon's neologisms. The concept is well known and recognized, and D. Kahneman when he received his Nobel prize in 2003 said a speech whose name was "Maps of Bounded Rationality" [15][3]. There have since been numerous attempts to create applications of bounded rationality in different fields. Most of them begin with classical models of optimal behavior and create new constraints on the choice maker in this framework. For a review about bounded rationality, see John Conlisk, 'Why Bounded Rationality?' [9]. It shows that models including bounded rationality have excellent results and endorse the notion that human cognition is a scarce resource. In 1998 Ariel Rubinstein published a book, *Modeling Bounded Rationality* [23], where he discussed models of bounded rationality, in particular modeling of choice, and their fundamental difficulties. The book includes a critique by Simon and an answer by Rubinstein. Rubinstein proposed to explicitly define decision procedures and model them.

After this brief introduction to bounded rationality, with the idea of resource constraints in our toolbox for decision making, we present now two particular and recent applications of this strategy by Xavier Gabaix more in details.

In 2000, Gabaix and Laibson [10] "extends the satisficing literature, which was pioneered by Herbert Simon" and develops an algorithm that simplifies decision trees by removing low probabilities branches. According to the authors, this algorithm is broadly applicable, allows to make behavioral predictions, is psychologically plausible and, last but not least, is empirically testable and actually tested in their article. This algorithm is inspired by Simon's bounded rationality and is parameterized in order to include as a special case perfect rationality and costless information gathering.

A decision tree is composed of (i) vertices that represent flow payoffs and of

[2]i.e., not only one scalar utility function but different utilities.

[3]And I secretly tried to put the words "bounded rationality" in my title, think that this would look "smart"...

(ii) edges, each representing the probabilities to choose a vertex starting from another one. The decision path always goes from the left to the right, and it is not a tree in the mathematical sense since there are cycles. In other words, each decision process is represented by choosing a vertex in the first column and then by following an edge going to the next column, until the last column is reached. Edges going in a vertex and edges going out both sum to one, since edges represent probabilities of following an edge from one vertex to another.

The number of possible paths is growing quickly in the number of columns. Thus, the expected value of a starting vertex is difficult to compute, since the decision maker should integrate over all possible paths. In practice, the most chosen starting vertex does not have the highest expectation in Gabaix's experiment.

In order to model the observed behavior, the authors proposed a "follow the leader" heuristic inspired by their intuition about how people analyze trees. A parameter p defines the cutoff probability. When $p = 0$, the heuristic is perfect rationality, not removing any edge. In this experiment, they exogenously set $p = 0.25$, intuitively. A *p-constrained path* $(V_0, p_1, V_1, p_2, V_2, ..., p_N, V_N)$ is a path starting in a vertex V_0 of the first column and with edges of probability $p_1, ..., p_{N-1} \geqslant p$, except possibly the last one, p_N. For any path, we define a cumulative probability $\pi = \prod_{i=1}^{N} p_i$ and cumulative payoff $U = \sum_{i=0}^{N} V_i$. Assuming there are K p-constrained path from a starting vertex with the associated set of cumulative probabilities $\{\pi_k\}_{k=1}^{K}$ and cumulative payoffs $\{U_k\}_{k=1}^{K}$, the constrained expected value is $\mathrm{V} = \sum_{k=1}^{K} \pi_k U_k$. It is an approximation of the true expectated value, with fewer paths to take into account.

This "follow the leader" heuristic (FTL) is the model preferred by the authors and is compared with three other models. The first one is a column-cutoff model, where the choice maker is perfectly rational but pays attention only to the Q first columns. The second one is a discounting model, where the choice maker follows all paths but exponentially discounts payoffs with a discount factor δ. Finally, the third model is the rational one, the particular case of FTL without cutoff probability, i.e., with $p = 0$, used in all economic models. It assumes rationality and zero cognition cost.

The payoff evaluations $V_1, V_2, ..., V_C$ of these models (with C possible choices) are translated in choice using the probability of choosing c:

$$P_c = \frac{e^{\mu V_c}}{\sum_{c'=1}^{C} e^{\mu V_{c'}}}.$$

with μ estimated in order to minimize the Euclidean distance between the empirical distribution and the distribution predicted by the model.

About the experimental design, 259 Harvard undergraduates were asked to choose a starting vertex on 12 decision trees in 40 minutes. All the trees were randomly generated and had 10 rows. Half of them had 5 columns and the other half had 10 columns. The author decided to use large trees to force students to use heuristics. The students were paid the expected value of one of the 12 trees picked randomly. A debriefing eliminated the students who did not understand the concept of expected value (8%).

If students behave randomly, they would gain an average amount of \$1.30. If they behave perfectly rationally, they would gain an average amount of \$9.74. In fact, they received \$6.72. Comparing rational, FTL, Q-column-cutoff and δ-discounting models with the full range of Q and δ values and with cutoff probability set to 0.25, FTL fits the data best. Other models outperform rationality in a statistically significant manner, FTL with the biggest margin using the squared Euclidian distance between the predictions of the model and the empirical data. All models are rejected compared to FTL, the closer with a t statistic of 6.06. However, the authors warn about the wrong prediction of FTL in special examples with outliers.

In a second article in 2006, Gabaix [11] proposed a directed cognition model that approximates option-value calculations. Decision makers select their next cognitive operation with partial myopia and act as if their next operation was the last opportunity, so they put the focus of their (direct) cognition on attributes with high value.

This strategy of modeling with an iterative structure breaks the infinite regress problem with costly cognition. Indeed, generally, in model of optimal cognition, if cognition is costly, the model needs to optimize cognition. And doing so is costly, so the model needs to optimize the optimization of cognition use, etc., thereby creating an infinite regress problem ([11], p. 1043)

Gabaix used this model to study information acquisition and to empirically evaluate it in two different experiments. One with financial cost and the other one with scarce time. In the first experiment, defined as "simple", an agent needs to choose one of three goods. These goods can be seen as projects that can either succeed or not. They have stochastic payoffs X_1, X_2, X_3: the projects could either give a payoff with probability p and $X_i = V_i$, or give nothing with probability $(1 - p)$ and $X_i = 0$. With a cost c the agent could "investigate" a project, i.e., know the payoff with certainty. There is an optimal sequence acquisition, using a Gittins-Weitzman index, but the directed cognition model, considering each acquisition as the last one, better describes the actual behavior of agents.

In the second experiment, the choice is more complex. There are N goods represented by rows with attributes in the columns. Each attribute is a payoff and the good is the sum of the payoffs of the chosen row. The payoffs are randomly drawn from a normal distribution with mean 0. The information about these payoffs are hidden by boxes and the choice maker has to use the mouse to open them. Only one box could be open at a time and order and duration are recorded using Mouselab[4]. Moreover, time is scarce (endo- and exogenous time pressure are tested). Variability of payoffs goes down from left to right. It

[4] `http://www.mouselabweb.org/`

is a abstraction of choice problems. Gabaix defines importance and variability of attribute straight in the experimental design. Here, attributes are directly commensurable. In reality, in a choice process, we need to evaluate importance of attributes as noted by Cameron [6], p. 6. In this second experiment, the directed cognition model evaluate each incremental search option as if it was the last. It computes the expected benefits and costs with a myopic horizon and chooses the search operation with highest benefit-cost ratio. It stops either when the time limit is reached or when the expected ratio falls below a threshold. There is no computationally rational solution and no cost (but still time constraints). Directed cognition successfully describes experimental information acquisition.

It seems very difficult to apply this kind of theory to random utility models. Basically, random utility models is a neoclassical economic theory assuming that the decision maker has a perfect discrimination capability. The random part of the model comes from the analyst, whose information is incomplete, but not from the choice maker. Sources of uncertainty are unobserved attributes, unobserved taste variation, measurements errors or instrumental variables [5]. Thus, the utility function is defined with a deterministic part and an error term, the stochastic part that models uncertainty of the analyst. The neoclassical theory says that the alternative with highest utility will be chosen. We generally define the deterministic part using attributes of alternatives with parameters and make some assumptions about the error terms. Then, using this utility to compute the probability of each chosen alternative, we can estimate the parameters of the deterministic part of the utility with highest likelihood.

This process does not mimic real choice processes. It focuses on most likely utility functions and not on the cognition processes. Gabaix's directed cognition and Simon's bounded rationality try to understand how decision makers think about real problems. It is a descriptive goal and this is not the case with random utility models. Moreover, when Gabaix compared his different models with the rational one, he used likelihood. Here we use likelihood to define our models. In the context of random utility function it is difficult to correct a potentially perfect rational behavior and it is difficult to compare them with each other.

There are mainly two strategies in the literature to mix information processing strategies with random utility models. In one hand, some authors do not modify the conceptual framework of random utility models but add information about cognition limitations in this framework. In particular, they focus on information acquisition and gather information about attention to attributes. On the other hand, Cameron and DeShazo [6] propose not to add information about attention to attributes in the model but to model directly the effect of attention allocation to attributes.

Previous results suggest that attention to attributes is of importance. In 2005, Hensher [14] modeled information processing strategies using a model with all attributes and a model without a few attributes and showed that this modifies estimates. In 2006, Hensher [13] also showed that the number of considered attributes goes down when the values of the attributes have a smaller variability (among other influences). Note that it is a global consideration and not a model based on attention allocation to each attribute. It shows that information about the choice set modifies individuals' strategy about attributes. In 2007 Campbell et al. [7] asked respondents if they attended all attributes when answering surveys about rural landscape. After the whole sequence of choice respondents were 64% to say they attend all attributes. The 36% of respondents who didn't attend all attributes had to declare which ones the didn't attend. Note that cost was the least attended attribute. Using this information in the model improves fit, and error variance was clearly higher for respondents who didn't pay attention to all attributes. Similarly, Gilbride and Allenby (2004, [12]), Rigby and Burton in 2006 [22] and Carlsson [8] in 2008 monitored attendance after the whole sequence of questions (what we will call "serial attendance"). In this last study, as an exception in the list, there is no significant difference between models with all attributes and models taking serial attendance into account.

In 2009, Puckett and Henscher [21] suggested that attention to attributes may vary between choice task for a same respondent. Scarpa et al. [24] asked respondents which attributes they did not attend (without an intermediate question asking if they attended all attributes) after *each* choice task. They built serial attendance from this information considering that an attribute was serially ignored when it was ignored in each choice task. Their results show that attendance substantially varies between choice tasks, i.e., a respondent does not systematically ignore the same attributes in all choice tasks s/he performs. As expected from this observation, models using serial attendance perform better than models assuming complete attention and models using choice-task attendance perform better than model using serial attendance. Meyerhoff [19] shows that a minority of respondents behave as if they ignored serially attributes. 50% of them attend all attributes while 42% ignore different attributes through different choice tasks and only 8% ignore always the same attributes (serial attendance). In this survey, price was the most attended attribute.

Kaye-Blake [16] used a different strategy to monitor respondents' attention to attributes. Respondents were not explicitly asked to declare which attributes they did or did not attended but attendance was directly implied by monitoring information access with a computerized survey. This is the same approach that Mouselab used in Gabaix [10] with an information display board on a web page. Here the authors developed their own software. The columns are the alternatives and the rows are the attributes. Each couple in the table has a value, like in all stated preference surveys, but this value is hidden behind a box that the user has to click to access the information. Here, once a box is opened, it stays opened until the end of the choice task (remember that in Gabaix [10] the box closed once the mouse moved out of the box). Another difference with Gabaix [10] is the absence of time limit or of cost to access information. The only cost involved is pointing the mouse and clicking. Even in this context, respondents didn't access all the information: only 78% of all the boxes were opened. In terms of a classical economic theory it would mean that the cost of moving the mouse and clicking is reputed higher than the benefits of the information accessible. This information display board has the great advantage of avoiding to ask respondents to consciously choose which attribute they attended in their choice task. In this way, the choice of attendance to attributes is more natural. Moreover, information was captured for each attribute and each alternative. In all other experiments, respondents could either attend or not an attribute. Here, they are able to gather information about an attribute only for one or two alternatives and not to gather attributes' values for all alternatives. Results show that a large number of respondents attend attributes only partially, accessing information about an attribute not for all alternatives. 55.6% of alternatives had all their boxes (i.e., attributes) opened (i.e., attended). But for 44.4% respondents left at least one card unopened. For the chosen alternatives, these proportions are significantly larger, as it may be expected. Three models were estimated, one with complete information, one without ignored attributes and one with the ignored attribute replaced by their average value since respondents were informed of the different levels of each attribute. The log-likelihood, the AIC and the BIC improve for each of these models compared to the previous one in the list. Another result of this experiment is that "use of information is highly correlated with the importance of attributes, given by the willingness to pay, and with the error associated with ignoring information use" (p. 17).

Finally, all these different studies concord on the fact that attention to attributes is incomplete and it differs from attribute to attribute, between respondents, and between choice-tasks performed by the same respondent. Kaye-Blake [16] even show that attendance to a given attribute may be incomplete because some of its levels may remain unknown event when the cost of access is very low. These studies suggest that accounting for attention to attributes improves model performance and quality of estimates. This effect is enforced if we use choice-task specific attention and not serial attention. It could be interesting to compare choice-task attention with attention to each attribute and each alternative (that we could name "box attention").

Assuming that attention is of importance in choice-modeling, it could be interesting to know if it is possible to predict attention by identifying its specific determinants. In this way we could use these results with data where the information about attention to attributes is not available. It would also allows us to understand how respondents pay attention and maybe modify policies about surveys and choice experiments to make them more effective and help their behavioral interpretation.

In a recent paper, Cameron and DeShazo [6] proposed a theoretic model for attention allocation. When ignoring an attribute, the choice maker does a sub-optimal decision. The expected value of the lost utility in this case is related to attention to attributes and depends of two components. The first one represent the distance of the top ranking alternatives when ignoring an attribute. The second one represents the relative importance of the attribute on the overall utility function. We develop this model in the next chapter in details.

Cameron and DeShazo [6] develop a methodology to implement this propensity to attend. They introduce a multiplicative parameter for the k^{th} attribute that represents propensity to attend. Thus each attribute is represented in the utility function as $\beta_k a_k x_k$, where β_k is the parameter for attribute x_k and a_k is the "propensity to attend" parameter. This parameter depends of different factors y_i, $i = 1, ..., n$, and could be incorporated with different strategies proposed by the authors: $a_k = 1 + \sum_i \gamma_i y_i$, $a_k = F(\sum_i \gamma_i y_i)$ or $exp(\sum_i \gamma_i y_i)$. Thus we need to estimate a model that is non-linear in the parameters. Finally, they present an empirical example of their "attention-corrected model" with data about mortality risk reduction programs.

Chapter 2

A Theory about Differential Attention to Attributes

In a choice experiment, assuming that acquiring information is costly and not complete, an attribute important to subjects could be ignored in the context of alternative evaluation and its marginal utility would hence be null in that choice-task. More generally, attention to attributes could be incomplete rather than completely ignored. For example, attention might not be paid to attributes when their value is satisfactory. This incomplete attention leads to underestimating its effect on choice compared to the real importance of this attribute to the choice maker. This is different from the case of a null marginal utility of an attribute as implied in the case of complete non-attendance.

If inattention to attributes is uniformly distributed across all attributes, utility parameters may be shifted in such a way that the overall effect does not create a bias in estimation from observed choice. However, it is likely that inattention will differ across attributes.

In order to model attention to attributes, the basic assumption proposed by Cameron and DeShazo is that the choice maker's attention to an additional attribute depends on both (i) the *expected loss* of utility of a suboptimal choice resulting from ignoring this attribute and (ii) the *marginal cost* of considering the attribute.

2.1 Marginal Cost of Considering the Attribute

This marginal cost depends on (i) the amount of cognitive resources available to the individual at a given time period and on (ii) the forgone benefits from using this capacity elsewhere. Attention to the first attribute is likely to be relatively cheap. But the marginal opportunity cost of attention to an additional attribute will probably increase with the number of attributes. To be able to estimate the effect of the number of attributes, we need choice sets with different number of attributes. Cameron and DeShazo also propose to consider other factors which could affect the marginal cost of attention to different attributes, such as information accessibility ("fine prints" or missing price attribute in advertising with contact to obtain it), order of a particular attribute in a long list or pre-occupation of the choice maker by other cognitive challenges (hours of work in a week, activities in the considered day). In their empirical example, they do not estimate any of these factors and thus the marginal cost of considering the attribute by lack of data.

2.2 Expected Loss of Utility of a Suboptimal Choice

About utility loss of a suboptimal choice resulting from ignoring the attribute, the authors propose two components to explain it : the "other-attribute utility dissimilarity" and the "own-attribute utility dissimilarity". For each of them, they first define it in a two-alternative case and then in a generalized form.

In a binary choice model, each alternative has an indirect utility function assumed to be linear: $U = \sum_k \beta_k X_k + \varepsilon$ with K different attributes X_k, each with their coefficient β_k and the random error ε representing unobserved utility by the researcher. Choice between alternatives A and B deals with the difference of utility of each alternative:

$$\begin{aligned} U^A - U^B &= \sum_k \beta_k (X_k^A - X_k^B) + (\varepsilon^A - \varepsilon^B) \\ &= \sum_k \beta_k x_k + \overline{\varepsilon} \end{aligned}$$

Let's call $X_k^A - X_k^B$ by x_k, the difference between alternative A and B and $\overline{\varepsilon}$ the difference of the error terms.

In the two-alternative case called A and B, a suboptimal choice means choosing alternative A when the optimal solution using *all* attributes is alternative B. Or conversely, choosing B when choice A is optimal. The expected utility loss of a suboptimal choice resulting of ignoring an attribute is thus:

$$\begin{aligned} E(\text{U Loss}) &= Pr(\text{A chosen}|\text{B optimal}) \cdot (U^B - U^A) \\ &+ Pr(\text{B chosen}|\text{A optimal}) \cdot (U^A - U^B) \end{aligned}$$

with

$$\begin{aligned} U^A - U^B &= \sum_k \beta_k x_k + \overline{\varepsilon} \\ &= x'\beta + \overline{\varepsilon} \\ &= x'_{-k}\beta_{-k} + x'_k\beta_k + \overline{\varepsilon} \end{aligned}$$

where β_k is the indirect utility-difference coefficient of the k-th attribute x_k in the first line. The third line of the equation is the decomposition of the inner product between two components: $x'_{-k}\beta_{-k}$ represents all attributes in the utility-difference function but the k-th one, and $x'_k\beta_k$ the k-th one only.

Probability for an alternative to be optimal is equal to the probability for its utility to be higher than the utility of the other alternative, and thus for the difference of utilities to have the right sign. In this case, we consider all attributes:

$$
\begin{aligned}
Pr(\text{A optimal}) &= Pr(x'\beta + \overline{\varepsilon} > 0) = Pr(\overline{\varepsilon} < x'\beta) \\
Pr(\text{B optimal}) &= Pr(\overline{\varepsilon} > x'\beta)
\end{aligned}
$$

When ignoring an attribute, we consider only all other attributes and the difference between the utility functions without this attribute. By the same reasoning, it gives:

$$
\begin{aligned}
Pr(\text{A chosen}) &= Pr(x'_{-k}\beta_{-k} + \overline{\varepsilon} > 0) = Pr(\overline{\varepsilon} < x'_{-k}\beta_{-k}) \\
Pr(\text{B chosen}) &= Pr(\overline{\varepsilon} > x'_{-k}\beta_{-k})
\end{aligned}
$$

Probability of making a mistake is then linked to either choosing alternative A when alternative B is optimal or conversely choosing B when A is optimal:

$$
\begin{aligned}
Pr(\text{A chosen} \cap \text{B optimal}) &= Pr\Big((\overline{\varepsilon} < x'_{-k}\beta_{-k}) \cap (\overline{\varepsilon} > x'\beta)\Big) \\
Pr(\text{B chosen} \cap \text{A optimal}) &= Pr\Big((\overline{\varepsilon} > x'_{-k}\beta_{-k}) \cap (\overline{\varepsilon} < x'\beta)\Big)
\end{aligned}
$$

Since $x'\beta = x'_{-k}\beta_{-k} + x'_k\beta_k$, the event $(\overline{\varepsilon} < x'_{-k}\beta_{-k}) \cap (\overline{\varepsilon} > x'\beta)$ is nonempty when $x_k\beta_k$ is positive, and thus $x'\beta < x'_{-k}\beta_{-k}$. Conversely, for the second event "B chosen and A optimal" to have a nonzero probability, $x_k\beta_k$ is negative and $x'_{-k}\beta_{-k} < x'\beta$.

Now we can write the expected utility loss as:

$$
\begin{aligned}
E(\text{U Loss}) &= \frac{Pr(\text{A chosen} \cap \text{B optimal})}{Pr(\text{B optimal})} \cdot (U^B - U^A) \\
&\quad + \frac{Pr(\text{B chosen} \cap \text{A optimal})}{Pr(\text{A optimal})} \cdot (U^A - U^B) \\
&= \frac{Pr(x'\beta < \overline{\varepsilon} < x'_{-k}\beta_{-k})}{Pr(x'\beta < \overline{\varepsilon})} \cdot (U^B - U^A) \\
&\quad + \frac{Pr(x'_{-k}\beta_{-k} < \overline{\varepsilon} < x'\beta)}{Pr(\overline{\varepsilon} < x'\beta)} \cdot (U^A - U^B) \\
&= \begin{cases} \frac{F(x'_{-k}\beta_{-k}) - F(x'\beta)}{1 - F(x'\beta)} \cdot (U^B - U^A) & \text{if } x'_{-k}\beta_{-k} > x'\beta \\ \frac{F(x'\beta) - F(x'_{-k}\beta_{-k})}{F(x'\beta)} \cdot (U^A - U^B) & \text{otherwise} \end{cases} \\
&= \begin{cases} \frac{F(x'_{-k}\beta_{-k}) - F(x'_{-k}\beta_{-k} + x'_k\beta_k)}{1 - F(x'_{-k}\beta_{-k} + x'_k\beta_k)} \cdot (U^B - U^A) & \text{if } x'_k\beta_k < 0 \\ \frac{F(x'_{-k}\beta_{-k} + x'_k\beta_k) - F(x'_{-k}\beta_{-k})}{F(x'_{-k}\beta_{-k} + x'_k\beta_k)} \cdot (U^A - U^B) & \text{otherwise} \end{cases}
\end{aligned}
$$

with F the cumulative distribution function of the difference of error term, usually a logistic distribution in logit models.

From this result, the authors call $x'_{-k}\beta_{-k}$ the "other-attribute utility difference" and $x_k\beta_k$ the "own-attribute utility difference". The link between the expression of expected utility loss, intuitively an explanatory variable of attention, and own- and other-attribute utility differences is not clear to me, even if the use of these two differences is attractive when reading the article.

2.2.1 Other-Attribute Utility Dissimilarity

In the case when there are three or more alternatives, Cameron and DeShazo present a few statistics by analogy with the two-alternative case.

First of all, the difference of utility between the two leading alternatives when ignoring the k-th attribute. They call this statistic $lead(x'_{-k}\beta_{-k})$. It means to calculate each of utility differences based on all attributes but the k-th one. It is also proposed to use the standard deviation in utility difference, called $sd(x'_{-k}\beta_{-k})$. The skewness could also be computed, $skew(x'_{-k}\beta_{-k})$. The authors present a fourth measure of dissimilarity, an entropy measure, employed by Swait and Adamowicz, for example.

All these measures of the dissimilarity based on other attributes are referred to using $dissim(x'_{-k}\beta_{-k})$ and represent a way to evaluate if there is a clear winner in the alternatives when not considering the k-th attribute. If there is a constant in the expression of the utility, it would be considered in the computation of $lead(x'_{-k}\beta_{-k})$.

2.2.2 Own-Attribute Utility Dissimilarity

In the general case, "own-attribute utility dissimilarity" is a measure of the potential for the k-th attribute to change the identity of the chosen alternative. By analogy to $x_k\beta_k$ in the two-alternative case, it depends of the importance of this attribute in the utility function, i.e., the marginal utility associated with the attribute, and of the level of this attribute.

As for the other-attribute utility differences, the authors propose a statistic of the difference of utility between the two leading alternatives for the k-th attribute, $lead(x_k\beta_k)$, skewness for this attribute, $skew(x_k\beta_k)$, and standard deviation, $sd(x_k\beta_k)$. Let's call them $dissim(x_k\beta_k)$.

Other-attribute dissimilarities could be tricky to represent in our minds. How is it possible to know before evaluating attributes what are their effect on the choice? If these factors allow to predict attention, it means that there is a necessity to evaluate all attributes before knowing which ones the respondent is going to attend. In a descriptive point of view, it is a loophole. We can consider this first estimation with all attributes as an *a priori* knowledge of the respondent but it is likely that this knowledge stems from a background of the choice maker and not from this particular choice process.

Chapter 3

Data Sets and Models

To implement Cameron's model, we use three different data sets. In the first one, different waves of surveys contain different numbers of attributes and respondents were asked to choose the attributes they had not attended to in their decision process at the end of each choice-task. In the second one, respondents had to first answer if they attended all attributes, and then – if their answer was "no" – they were asked to list the ones they did not attend. In the third one, respondents did not declare their attention to attributes. But a computerized system allowed the modeler to know when each respondent had accessed hidden information about attributes. For each of these data sets, we estimate different choice models that exclude information about attendance. Thus we always estimate models with full information.

3.1 Cortina d'Ampezzo and attendance by respondents

The first dataset was collected in a survey by Riccardo Scarpa and Mara Thiene [24]. The study concerns visitor services in the Alpine Park of Cortina d'Ampezzo in North of Italy, in the Dolomites. This data set is of particular interest to test Cameron's model, since choice makers had to identify the attributes they did not attend in their choice, for each single choice-task they undertook. This data set is a multi-attribute stated preferences survey reporting the choice of visitors of the park. Choices are about outdoor recreation activities.

3.1.1 Data collection

The visitors were divided in strata, depending on the main purpose of their visit to the park:

1. hikers,
2. climbers,
3. mountain bikers,

4. visitors using via-ferratas,
5. and visitors enjoying a short walk or picnicking

For each strata, 96 respondents were interviewed face-to-face and thus the data contains 480 surveys.

Some of the attributes are of particular interest to some strata of visitors and less relevant to others. There are ten attributes, and each one except cost has three levels:

1. building of additional thematic itineraries, focusing on flora, fauna and historical aspects,
2. increasing the network of trails and hiking paths within the Park,
3. improving the system of trail signs,
4. adding new itineraries, new trails with different technical challenge, length and effort,
5. adding climbing routes
6. improving quality and security of via-ferratas,
7. adding new shelters,
8. different levels of congestion describing the number of people met along the trails,
9. different levels of information material about the park,
10. and finally an entrance fee, with four levels

Obviously, the fifth attribute, about climbing routes, is of high interest for people whose main purpose of the visit in the park is climbing. People engaged in mountain-biking or other activities could of course also be interested, since these activities are not mutually exclusive. However, there may be some people not at all interested by new climbing routes who are likely to ignore this attribute.

3.1.2 Survey design

The survey contains four waves. In all waves, the price attribute is present. In the first wave, all nine non-monetary attributes are available. Then, in the following waves, there are seven, five and finally three non-monetary attributes. The attributes are discarded either because they are significant enough after estimating a basic multinomial logit or because they are of minor interest for the specific visitor being considered. This way, each two levels of each attribute were significantly estimated. Dropping the two attributes with highest significance in the multinomial logit allows choice-makers to give more attention to other attributes since there are fewer of them. A complete table of the discarded attributes depending of waves and groups is available in Scarpa [24].

Each respondent performed 12 choice-tasks. There are 24 visitors surveyed in each wave-group. The design was balanced and thus we have 120 surveys for each wave, i.e., 480 completed surveys. Choice tasks were in blocks to insure orthogonality and balancing of levels. The design at each stage after the first was a WTP-efficient Bayesian design and the priors were at each stage built on the previous pooled sample.

For each choice task, there are two experimental designed alternatives and a *statu quo* alternative with no cost. After each choice task, respondents have to report the attribute they feel they ignored in their decision process. It is a binary choice, either they pay attention to an attribute or not.

3.1.3 Models

We use four different models to estimate the parameters of each attribute. For each of them, we remove the factors that are not available to the respondent depending of the waves, but we don't consider the information about attendance. The first one is a classical multinomial logit, with linear utility functions and taking into account the availability of attributes since attributes are not always available depending of the wave. In the second one, we scale error terms by wave. In the third one, we scale error terms by wave and category of respondents. Scarpa [24] showed that the more precisely you scale the error term (no scaling, by wave, by wave and category) the better the model fits the data.

Finally, in the fourth model, we estimate a mixed logit allowing to have specific respondent betas. We used a error component mixed logit, i.e., we add a random error term to the two alternatives that are not status quo. It models the fact that it is easier for the respondent to experience the status quo alternative, so this alternative is not sharing this extra error term. It is normally distributed with mean zero and standard deviation one. An estimation of the model is given with a simulation using 100 halton draws and 10 iterations. The coefficients are assumed to be normally distributed.

3.2 Wind Power and attendance by respondents

This choice experiment tried to estimate the effect of different factors of onshore wind power generation in Westsachsen in the eastern part of Germany. Respondents had to choose between three alternatives about wind power generation in their area in 2020. Each alternative was described with four non-monetary attributes: the size of wind farms, the maximum height of the turbines, the effect on a population of birds, and the minimum distance to settlements. Each of them has 3 levels, presented in Table 3.1. All respondents were given a small description of each attribute and its effects. In addition, there is a price attribute with 5 levels.

Table 3.1: Attributes and levels for data about Wind Power, with levels of the first alternative in bold.

Attributes	Levels
Size of Wind Farms	**large (16 to 18 mills)**
	medium (10 to 12 mills)
	small (4 to 6 mills)
Maximum Height of Turbines	**200 meters**
	150 meters
	110 meters
Minimum distance to settlements	**750 meters**
	1.100 meters
	1.500 meters
Effect on red kite population	-5%
	-10%
	-15%
Monthly surcharge to power bill	**€ 0**
	€ 1
	€ 2.5
	€ 4
	€ 6

In the three alternatives, the first one has always the same levels for each attributes (in bold in Table 3.1). It is not exactly a status quo but a future situation. It is a benchmark level for 2020. Respondents were informed that for each restriction in the first four attributes, the cost would increase, since the attributes of the first alternative are the less strict (but for the effect on red kite population). The carbon dioxide emissions are assumed to be constant independently of the different attributes to put the focus on the effect of wind farms on landscape.

The survey was conducted by telephone on 708 respondents, each of one had to answer 5 choice tasks (and so 3540 observations). 40 choice sets were designed using a D-optimal fractional factorial design.

Attention to attributes was recorded after each choice-task, like in the data about Cortina, but with a slightly different strategy. While in the Cortina's survey [24] respondent were asked directly which attributes respondents had not considered – directly assuming that incomplete attribute attendance is likely. In the wind power survey by Meyerhoff – instead – respondents were first asked: "Did all attributes of wind power generation matter when you were choosing among the alternatives on the previous choice card?". Then, if the answer was a "no" they were also asked to describe which one they had not attended, as they were in the Cortina's survey.

3.2.1 Models

As previously mentioned, the first alternative is fixed and does not really represent a status quo, but the future situation depending of the existing decisions about regulation for 2020, a kind of “future status quo”, *with no* additional cost. The two other alternatives propose tighter regulation *with* additional costs. In order to model this, the author uses an error component logit (like the one used in the mixed logit model with the data about Cortina) estimated with Biogeme [4]. This extra error term for the two “non future status quo” alternatives is normally distributed with mean zero and standard deviation one. In this model, two groups are created: respondents who answered a complete attention to attributes and respondents attending only partially to attributes. These two groups have different scale parameters. One of these two parameters is fixed to one, while the other is estimated. Only the first alternative has a specific constant.

The authors estimated a second model using panel specifications allowing individuals to have specific error parameters. The extra error component of the two “non future status quo” alternatives does not thus vary across observations but across respondents and takes into account the information that every respondent answered five questions.

This forest wind power data were used to estimate a mixed logit model, in order to have individual specific estimates/betas. In this model, we also use an extra error component and do the same process than for the mixed logit model used with data about Cortina.

3.3 Potatoes and information display board

This data set was designed and conducted by Bill Kaye-Blake, Senior Research Officer in the Agribusiness and Economics Research Unit of Lincoln University in Christchurch, New Zealand, to test respondents’ attention to attributes with data about potatoes.

3.3.1 Survey Instrument

Unlike in the survey used in Cortina, respondents were not asked to state whether or not they attended each attribute, but they were expected to actively gather the information to make their choice. What they searched for provided a clear indication of what they knew and could consider in the context of process evaluation. The attribute information search was monitored via a computer based survey conducted using a purpose-built software. It software creates information display boards for each choice-task, with columns for each alternatives (potatoes) and rows for each attributes but with hidden attribute information in the intersection of the rows and the columns. To access this information, the respondents had to click on the box to reveal the hidden value. Then the box would open and stay so until the end of the choice-task. There were no time limit and no cost (apart of time and the effort to move the mouse) in opening the boxes and accessing information. We know when respondents opened the boxes and started paying attention to the value of the attribute. Thus we are also able to determine the order in which respondents accessed the information. Finally, the order of presentation of the attributes was randomized for each respondent. So each respondent received question in the same order, but different respondents had attributes presented in different orders.

This strategy of information display board is similar to MouseLab Web. Both use the same technology. Note, however, that the default display settings in MouseLab let the respondent access the information only when the mouse is on the top of the box. As soon as the respondent moves his attention to another box (with the mouse), the previous box is closed again. The duration of the exposure to the information by the modeler is thus different. In the potatoes' data, a respondent's accessing attribute information is no guarantee that this attribute is paid attention to in the evaluation of the alternatives.

3.3.2 Survey Design

Six attributes are available, as described in table 3.2 with their levels. The experimental design was optimized by using the D-efficiency criterion, using the chosen levels and a set of priors. For more information about the survey design and the creation of D-efficient choice sets using prior estimates of betas into the design, see Kaye-Blakes [16].

The survey was administered to 92 campus members of Lincoln University, resulting in 87 valid responses. They were asked to answer four different survey sections. First, respondents had to rank the six attributes in order of importance. Then, they made ten choice-tasks using information display boards described before. For each choice-task, they had the choice between three alternatives/potatoes. The two last sections of the survey were questions about their beliefs, attitudes and socioeconomic information. During the experiment respondents could refer to a reference sheet with the list of attributes and their levels. The order of presentation of attributes in the survey was randomly selected for each different respondent. It allows us to investigate whether the order of attributes presentation had an effect on whether attention was paid to them. The order of presentation differed across respondents but not within each respondent.

This data were used to estimate two models: (i) a basic multinomial logit and (ii) a panel mixed logit so as to derive individual-specific distribution parameter estimates (the same as the two mixed logit estimated with the two previous data).

Table 3.2: Attributes and levels for data about potatoes.

Attributes	Levels
Texture	waxy
	floury
Price	\$1.00
	\$1.50
	\$2.00
	\$2.50
Color	pink
	yellow
	white
Production	conventional
	GM
	organic
Nutrition	ordinary
	low-GI
	high omega3
Country of origin	Australia
	China
	New Zealand

Chapter 4

Attendance as a Choice

In stated preference survey we always consider the main choice as being the only choice. However, if we consider that the attention to attributes might be incomplete, then there is another level of choice concerning which attributes are attended to. We can model this as a binomial choice for each attribute: either the respondent attends the attribute k or not. And this process is repeated for each attribute k and at each choice-task. From this perspective one can explain the data about attendance and test whether some factors considered to be meaningful to explain attendance really are so. We focus on choice-task attribute non-attendance since different studies show that serial attribute non attendance, even if it is present for some respondents, is not a valid assumption (Meyerhoff [19] - in which only 8% of respondents serially attend attributes[1] against 42% for choice-task attendance and 50% of complete attendance, Scarpa [24]). In reality, non attended attributes vary across choice-tasks.

4.1 Dependent variables

Before describing the explanatory variables, a better definition of what we are trying to explain is needed. With Cortina Park management data, attendance is recorded after each choice task. The respondent, after choosing his favorite alternative, was asked to answer which attribute he did not attend to. So we have a binary variable, with value 1 denoting attribute attendance and 0 non-attendance.

The data about wind power were collected using the same strategy used in the Cortina Park management survey data. After each choice task, respondents were asked to list the attributes they had not attended. Thus, the dependent variable is exactly the same as described just before. However, in this experiment, respondents were first asked if they attended all attributes or not, and if not, were asked which ones they didn't attend. This difference could modify the answers of respondents. For example, some respondent might have grown

[1] i.e., respondents don't attend some attributes in their choice-tasks and these non-attended attributes are always the same across all the choice-tasks

reluctant – perhaps due to tiredness – to state any degree of attribute neglect towards the last part of the choice-task sequence.

In the potatoes data, respondents were not asked to answer the question, but their access to the level of each attribute was monitored. As in the Cortina survey, we have the information about attribute attendance: does the respondent access any value for any alternative for a given attribute? If yes, the variable is 1, otherwise it is 0. Note that sometimes respondents accessed information for a particular attribute only for a few alternatives and not for all of them. We consider here if they accessed any information about the attribute. Respondents could access the information about a particular attribute for all alternatives or only for a particular alternative. In both case, the dependent variable would be 1.

With these data, it could also be possible to try to explain not only *if* respondents access or not the information but also *when* they accessed the information. This information is not exactly the time respondents paid attention to an attribute, since they could open a box and then open another one and disregard the information from the first box in their choice. Instead, it gives us the information on when they first attend a particular value of an attribute. Since we have the information for each alternative and each attribute and we try here to explain only attendance to attributes in general, we could summarize the information for each attribute, taking into account all the alternatives. In other words, let's consider one specific attribute from the list of all attributes. Assume this attribute has 3 different values for each possible alternative. We know that the respondent accessed these information after respectively 2.3, 3.8 seconds and never for the third value. We could proxy these values to be able to use them as a dependent variable and we could consider the sum (or the mean, as it is equivalent). It would represent a summary of when the respondent accessed the information. We could also imagine to consider the minimum value different from zero, in order to model the order of attendance to attributes instead of the mean.

We won't do this analysis about the mean of the time respondents opened the boxes or the order of access to the boxes in this report. Here we just try first to explain attendance or not, and nothing more, since the first step is clearly to be able to predict attendance or not, and only then to predict the "amount of attention" in time or the order of attention to attributes.

4.2 Explanatory variables

Different authors have suggested different factors to explain attendance. Here we try to test the effect of some of these. For some of them, proposed by Cameron and DeShazo, there are two ways to compute them, either using the utility without a particular attribute or using the utility with only the linear factor of this utility, i.e., either $x'_{-k}\beta_{-k} := \sum_{k' \neq k} x'_{k'}\beta_{k'}$ or $x'_k\beta_k$. It is the case for the leadership (*lead*), standard deviation, skewness and entropy.

A example of computation for the values is available in the annex.

Leadership As described previously, leadership is the difference between the utilities of the leading alternative and the second leading alternative:

$$\begin{aligned} lead(x_k\beta_k) &:= (x_k'\beta_k)_1 - (x_k'\beta_k)_2 \\ lead(x_{-k}'\beta_{-k}) &:= (x_{-k}'\beta_{-k})_1 - (x_{-k}'\beta_{-k})_2 \end{aligned}$$

We can use either the differences or their absolute values.

It is important also to specify which utility is used to define which alternative is the leading one, since it could be different if we compute utilities using *only* the k-th attribute, all attributes *except* the k-th or all of the attributes *including* the k-th attribute.

In the $lead(x_1'\beta_1)$ example, we need to compute the difference of the utilities using only the first attribute of the two leading alternatives, i.e., $(x_1'\beta_1)_1 - (x_1'\beta_1)_2$.

However, there are two different ways to compute *lead*. The leading alternatives could be defined on the complete utility function or on the bounded utility function considered in the particular case. In our example – and assuming there are three alternatives – it means that we can consider the leading alternatives of the set:

$$\left\{(x_1'\beta_1)^A, (x_1'\beta_1)^B, (x_1'\beta_1)^C\right\}$$

of the bounded utilities or the set:

$$\left\{(x'\beta)^A, (x'\beta)^B, (x'\beta)^C\right\}$$

of the complete utilities. In the first case, we use the notation $lead_{local}$ for the *lead* computed with the leading alternatives using only the bounded utility. Thus we have:

$$(x_1'\beta_1)_1 = \left\{(x_1'\beta_1)^j \middle| (x_1'\beta_1)^j = \max_{j'=1,\ldots,J}\left((x_1'\beta_1)^{j'}\right)\right\}$$

with J the number of alternatives. $(x_1'\beta_1)_2$ is computed in the same way, using the second largest value instead of the maximum. Thus, we have $lead_{local}(x_1'\beta_1) = (x_1'\beta_1)_1 - (x_1'\beta_1)_2$.

In the other case, we have:

$$(x_1'\beta_1)_1 = \left\{(x_1'\beta_1)^j \middle| (x'\beta)^j = \max_{j'=1,\ldots,J}\left((x'\beta)^{j'}\right)\right\}$$

and here $lead(x_1'\beta_1) = (x_1'\beta_1)_1 - (x_1'\beta_1)_2$ using these values for $(x_1'\beta_1)_1$ and $(x_1'\beta_1)_2$.

The leadership computed with only the bounded number of attributes will be called "local" and we thus have four different values: $lead(x_k\beta_k)$, $lead(x'_{-k}\beta_{-k})$, $lead_{local}(x'_k\beta_k)$ and $lead_{local}(x'_{-k}\beta_{-k})$.

Standard deviation If we have three alternatives A, B and C, and considering only the k-th attribute, we have three utility functions: $(x'_k\beta_k)^A$, $(x'_k\beta_k)^B$ and $(x'_k\beta_k)^C$. Then:

$$SD(x'_k\beta_k) := sd\big((x'_k\beta_k)^A, (x'_k\beta_k)^B, (x'_k\beta_k)^C\big)$$

And identically for the utility without the k-th attribute:

$$SD(x'_{-k}\beta_{-k}) := sd\big((x'_{-k}\beta_{-k})^A, (x'_{-k}\beta_{-k})^B, (x'_{-k}\beta_{-k})^C\big)$$

Choice-tasks with small standard deviations are obviously made up of similar alternatives, and hence the relative advantage of each is hard to evaluate well because none alternative is a clear winner over the other.

Skewness The skewness of the utility function for each alternatives is:

$$\begin{aligned} Skew(x'_k\beta_k) &:= skewness\big((x'_k\beta_k)^A, (x'_k\beta_k)^B, (x'_k\beta_k)^C\big) \\ Skew(x'_{-k}\beta_{-k}) &:= skewness\big((x'_{-k}\beta_{-k})^A, (x'_{-k}\beta_{-k})^B, (x'_{-k}\beta_{-k})^C\big) \end{aligned}$$

Entropy measure Cameron and DeShazo also propose to use an entropy measure and give as an example Swait and Adamowicz [28]. In their article, Swait and Adamowicz present a measure of complexity of the decision process in order to study choice behavior. The number of alternatives and the number of attributes are part of the complexity of the choice-task. The closeness of alternatives could also increase the complexity since a clear dominant alternative would be easier to choose. A good measure of complexity should take these elements into account and should allow commensurability of these values between themselves and with each other. The metric proposed by Swait and Adamowicz is based on entropy. Entropy is a theoretic measure of information. Information has a very broad mathematical meaning. In the context of choice modeling, the outcomes are the alternatives, $\{a^j, j = 1, ..., J\}$, with distribution $\pi(a^j)$ and the entropy of the choice-task is defined as

$$H = -\sum_{j=1}^{J} \pi(a^j) \log \pi(a^j)$$

Entropy is a measure of certainty. It is greater than or equal to zero. It reaches its maximum when all the alternatives are equally likely. Conversely, it is small when a choice is dominant. In particular, it is equal to zero when one choice is sure. When the number of alternatives increases, the maximum entropy increases and more generally entropy is larger. For Swait and Adamovicz [28], p.138, "[t]he number of attributes and the degree of attribute correlation also impact entropy since these elements will affect the probabilities $\pi(x)$". It seems possible to use this measure in a global way, using the probability distribution as $\pi(a)$ to compute the entropy of the general choice experiment, or in a choice-task level, using the probability to choose each alternative as $\pi(a)$. Moreover, in this last case, it is possible to consider either the specific-attribute utility function (using Cameron's notation "own-attribute entropy"), or even - "other-attribute entropy". It would represent complexity of the choice task globally when using the complete utility function, complexity of the particular attribute for "own-attribute entropy" and finally complexity of the choice using all other attributes but the one considered. In order to compute entropy, we need a measure of the probability and we will use the classical logit probability using the utilities with only the k-th attribute :

$$\pi(a_k^j) \;:=\; \frac{e^{(x_k\beta_k)^j)}}{\sum_{j'=1}^{J} e^{(x_k\beta_k)^{j'}}}$$

And identically with the utility with all the attributes but the k-th:

$$\pi(a_{-k}^j) \;:=\; \frac{e^{(x'_{-k}\beta_{-k})^j)}}{\sum_{j'=1}^{J} e^{(x'_{-k}\beta_{-k})^{j'}}}$$

And finally we get the own- and other-attribute entropy:

$$\begin{aligned} H_k &:= -\sum_{j=1}^{J} \pi(a_k^j)\log\pi(a_k^j) \\ H_{-k} &:= -\sum_{j=1}^{J} \pi(a_{-k}^j)\log\pi(a_{-k}^j) \end{aligned}$$

Pivotality An attribute is pivotal when considering it or not is going to change the prediction of the model about the preferred choice, i.e., when the alternative with maximum utility is not the same. Using $x'\beta$ to describe the full utility with all attributes, we define being pivotal as a binomial variable:

$$Pivotal_k = \begin{cases} 1 & \text{if } \left\{j|(x'\beta)^j = \max_{j'}(x'\beta)^{j'}\right\} \neq \left\{i|(x'_{-k}\beta_{-k})^i = \max_{i'}(x'_{-k}\beta_{-k})^{i'}\right\} \\ 0 & \text{otherwise} \end{cases}$$

Number of attributes For Swait and Adamovicz, the number of attributes is correlated with entropy. For the Cortina data we are able to estimate the effect of the number of attributes on attention since the number of attribute varies in the different waves (10, 8, 6 and finally 4 attributes).

Order of attributes In the potatoes data, the order of the attributes is different and randomly defined for each respondent. It is thus possible to test if the order of the attributes affects the way respondents are attending attributes. We usually read from top to bottom and left to right, and maybe we give more attention to the attributes described on top or towards the left.

The effect of the number of alternatives cannot be tested, since all the data sets have the same number of alternatives. It is impossible to take into account time pressure since we have no data on it. Moreover there is no explicit time pressure in the three choice experiments we consider in this paper.

Since there are two different random utility models, one to estimate the different β's and one to explain attendance, and it could be confusing, we define here clearly the terminology we will use in the following part of this report. In the case of binomial models explaining attendance to attributes as a choice using the variables we have just defined, we will call these the "attribute attendance probability models". The dependent variables of these models were defined in Section 4.1; they are either 1 if attribute is attended, or 0 otherwise. We have just defined the explanatory variables of these models in Section 4.2.

We use other models in this study to estimate utility parameters β's, in order to compute the explanatory variables of the "attribute attendance probability models". These models explain the choice of the respondent considering a complete attention to attributes. They are the classic random utility models. We will refer to these models by "choice probability models".

Chapter 5

Results

5.1 Absolute value for *lead*

As explained before, *lead* is the difference of utilities using the two leading alternatives. To determine what are the two leading alternatives, we use the complete utility function of the choice probability model. It is, however, possible for this difference to be negative. Indeed the utility of the leading alternative is larger than the utility of the second leading alternative by definition. However, when we consider only a part of the utility function, as for $lead(x_k\beta_k)$ and $lead(x'_{-k}\beta_{-k})$, this need not be the case any more. So this difference $lead(x_k\beta_k)$ (respectively $lead(x'_{-k}\beta_{-k})$) could be negative in the case of a leading alternative having a smaller utility for a particular attribute (respectively for a particular subgroup of attributes).

Since *lead* could be negative and as it could be considered as a distance between the two leading alternatives, we could consider that it is more meaningful to consider its absolute value.

To test these two different specifications, we estimate two different attribute attendance probability models using only $lead(x_k\beta_k)$ and $lead(x'_{-k}\beta_{-k})$ as explanatory variables, once *with* absolute value and once *without* absolute value. The first attribute attendance probability model, without absolute value, is:

$$y = ASC + \beta_{other} * lead(x'_{-k}\beta_{-k}) + \beta_{own} * lead(x_k\beta_k)$$

And the second attribute attendance probability model, with absolute value, is:

$$y = ASC + \beta_{other} * |lead(x'_{-k}\beta_{-k})| + \beta_{own} * |lead(x_k\beta_k)|$$

for each attribute k and where y is the dependent variable with value 1 if attribute is attended and 0 otherwise and ASC is an constant.

Attribute attendance probability model results are reported in tables 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8 and 5.9 using estimates of each different choice probability models for different data sets. We test the null hypotheses that these values explain attendance for each attribute. We also consider the overall attention to attributes; the "Global" row is a model with all attributes. We estimate this model with all the dependent variables together for all attributes. It gives a general meaning of the explanatory variables for attention to attributes in the whole. Results for the first attribute attendance probability model, without absolute value, are in the left half of the tables and results for the second attribute attendance probability model, with absolute value, are in the right half of the tables. For each of the two models and for each attribute, we present (*i*) the sign of the estimates for other- and own-attribute utility *lead*, β_{other} and β_{own}, (*ii*) their p-values, (*iii*) the log-likelihood of the sample for the models ("Final log-L"), (*iv*) the likelihood ratio tests ("L ratio test") and (*v*) the adjusted likelihood ratio index $\overline{\rho}^2$, a goodness-of-fit measure. The largest value between the two attribute attendance probability models is in bold. For the different values of β's we use red color for p-values higher than 0.05. The "Composite" column gives the value of the log-likelihood of a composite attribute attendance probability model using both absolute values and signed values for *lead*, for comparison.

We use four different choice probability models for Cortina data, a basic MNL (Table 5.1), scaling by category (Table 5.2), scaling by category and wave (Table 5.3) and finally estimating a mixed logit with individual specific betas (Table 5.4). Each of these choice probability models fits the data better than the previous one, with the mixed logit giving the best fit of all. From this last model we compute the means of the individual specific distributions of β_n. We use these to compute utilities for alternatives of each respondent set of choice-tasks as well as all the various factors determining attendance. Indeed we observe different behaviours about attendance and using the means of individual specific distributions of β_n should improve the quality of the variables we use to predict attendance. Using a classical MNL (see Table 5.1) we have obtain disappointing results results. The coefficient for *lead* (without absolute value) is not significant. The majority of the coefficient estimates are not significantly different from zero. Moreover when it is globally significant, the sign of β_{own} has not got the expect sign. It should be positive since a large difference of the own-attribute utility of the two leading alternatives should give a higher probability to pay attention to the attribute. Looking at the values for $\overline{\rho}^2$ $|lead|$ clearly provides a better factor in the explanation of attribute attendance.

Using the choice probability model with scaling by category to estimate β's does not really change the results, as we can see in Table 5.2. It is globally insignificant for *lead* and again, the sign is wrong for β_{own}. However, this attribute attendance probability model is the best at explaining attribute attendance for the Cortina data. $\overline{\rho}^2$ is the highest amongst the various attribute attendance probability model specifications tried. Note that, if these estimates for β are the best ones to explain attendance, they are not the best ones to explain choices of respondents in the choice probability model. Scaling by category *and wave* is better to explain choices. We could expect a better explanation of attention to attributes in our attribute attendance probability models when estimates come from a better choice probability model; it is not the case here.

Table 5.1: Comparison of *lead* and $|lead|$ for Cortina data using basic MNL.

	Using *lead*							Using $\|lead\|$							Composite
	β_{other}		β_{own}					β_{other}		β_{own}					
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Final logL
Attribute 1	+	0.42	+	0.41	-1208.665	2772.957	0.533	−	0.14	+	0.01	-1204.862	2780.562	**0.535**	-2784.188
Attribute 2	+	0.12	+	0.36	-815.089	2362.350	0.590	+	0.41	−	0.02	-813.302	2365.924	**0.591**	-2119.245
Attribute 3	−	0.22	+	0.04	-1391.325	4004.648	0.589	−	0.07	+	0.64	-1393.056	4001.186	0.589	-1746.295
Attribute 4	−	0.02	−	0.21	-1396.227	1998.580	0.416	−	0.02	+	0.00	-1390.632	2009.769	**0.418**	-2049.751
Attribute 5	+	0.03	−	0.85	-662.188	1071.141	0.445	−	0.04	+	0.38	-662.138	1071.240	0.445	-1556.871
(Attribute 6)	−	0.03	+	0.00	-847.785	1498.452	**0.467**	−	0.04	−	1.00	-852.210	1489.602	0.464	-1863.139
Attribute 7	−	0.54	+	0.21	-1742.196	2903.652	0.454	−	0.16	−	0.00	-1715.249	2957.547	**0.462**	-2073.125
Attribute 8	+	0.00	+	0.00	-2779.654	2026.495	0.266	+	0.30	−	0.00	-2705.536	2174.730	**0.286**	-2082.043
(Attribute 9)	−	0.00	+	0.10	-2285.866	3014.070	**0.397**	−	0.00	+	0.70	-2287.231	3011.340	0.396	-1676.401
Attribute P	+	0.00	+	0.00	-3881.827	221.401	**0.027**	−	0.96	−	0.43	-3895.521	194.014	0.024	-2294.072
Global	+	0.12	+	0.11	-18,418.623	19,058.142	0.341	−	0.00	−	0.00	-17,683.355	20,528.679	**0.367**	-17645.384

Note that each model for each attribute has different number of observations since they are available depending of the waves. Thus they all have different initial log-likelihood and we omit to write it for clarity.

Table 5.2: Comparison of *lead* and $|lead|$ for Cortina data using MNL Scaled by Category.

	Using *lead*							Using $\lvert lead \rvert$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
Attribute 1	+	0.57	+	0.77	-1209.310	2771.666	0.533	−	0.15	+	0.00	-1202.266	2785.755	**0.536**
Attribute 2	+	0.10	+	0.48	-814.903	2362.721	0.590	+	0.59	−	0.01	-812.916	2366.695	**0.591**
Attribute 3	−	0.37	+	0.03	-1391.447	4004.403	0.589	−	0.06	+	0.29	-1392.543	4002.212	0.589
Attribute 4	−	0.03	−	0.17	-1396.558	1997.917	0.416	−	0.04	+	0.00	-1389.177	2012.678	**0.419**
Attribute 5	−	0.01	+	0.62	-660.998	1073.521	0.446	−	0.01	+	0.23	-660.696	1074.124	0.446
(Attribute 6)	−	0.02	+	0.00	-847.765	1498.493	**0.467**	−	0.03	+	0.77	-851.977	1490.068	0.465
Attribute 7	−	0.60	+	0.33	-1742.888	2902.268	0.453	−	0.22	−	0.00	-1719.153	2949.738	**0.461**
Attribute 8	+	0.00	+	0.00	-2781.754	2022.294	0.266	+	0.79	−	0.00	-2685.580	2214.643	**0.291**
(Attribute 9)	−	0.00	+	0.10	-2287.704	3010.394	0.396	−	0.00	+	0.10	-2287.736	3010.330	0.396
Attribute P	+	0.00	+	0.00	-3883.659	217.737	**0.027**	−	0.40	−	0.34	-3894.463	196.130	0.024
Global	+	0.11	+	0.24	-18,418.992	19,057.405	0.341	−	0.00	−	0.00	-17,625.123	20,645.144	**0.369**

Table 5.3: Comparison of *lead* and $|lead|$ for Cortina data using MNL Scaled by Category and Wave.

	Using *lead*							Using $\|lead\|$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
Attribute 1	$-$	0.70	$+$	0.40	-1209.041	2772.204	0.533	$-$	0.32	$+$	0.00	-1204.709	2780.868	**0.535**
Attribute 2	$+$	0.25	$-$	0.77	-814.187	2364.154	0.591	$+$	0.78	$-$	0.00	-811.976	2368.575	**0.592**
Attribute 3	$-$	0.16	$+$	0.07	-1391.479	4004.339	0.589	$-$	0.03	$+$	0.98	-1392.483	4002.332	0.589
Attribute 4	$-$	0.02	$-$	0.25	-1396.770	1997.494	0.416	$-$	0.03	$+$	0.02	-1394.940	2001.154	0.416
Attribute 5	$-$	0.06	$-$	0.59	-662.725	1070.066	0.444	$-$	0.03	$+$	0.28	-661.483	1072.550	**0.445**
Attribute 6	$-$	0.12	$-$	0.56	-853.140	1487.742	0.464	$-$	0.07	$-$	0.65	-852.529	1488.964	0.464
Attribute 7	$+$	0.73	$+$	0.84	-1743.862	2900.320	0.453	$-$	0.12	$-$	0.00	-1708.634	2970.776	**0.464**
Attribute 8	$+$	0.00	$+$	0.00	-2782.523	2020.758	0.266	$+$	0.74	$-$	0.00	-2681.980	2221.843	**0.292**
(Attribute 9)	$-$	0.00	$+$	0.69	-2289.341	3007.121	0.396	$-$	0.00	$-$	0.34	-2288.800	3008.202	0.396
Attribute P	$+$	0.00	$+$	0.00	-3869.078	246.901	**0.030**	$-$	0.22	$-$	0.03	-3891.572	201.912	0.025
Global	$+$	0.00	$+$	0.00	-18,408.504	19,078.381	0.341	$-$	0.00	$-$	0.00	-17,713.444	20,468.500	**0.366**

Table 5.4: Comparison of *lead* and $|lead|$ for Cortina data using mixed logit estimates.

	Using *lead*							Using $\lvert lead \rvert$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
Attribute 1	$-$	0.38	$+$	0.00	-1198.187	2793.911	0.537	$-$	0.03	$+$	0.00	-1197.576	2795.133	0.537
Attribute 2	$+$	0.02	$+$	0.14	-813.492	2365.544	**0.591**	$+$	0.86	$-$	0.55	-816.263	2360.003	0.590
Attribute 3	$+$	0.49	$+$	0.00	-1387.876	4011.546	0.590	$-$	0.52	$+$	0.00	-1381.109	4025.080	**0.592**
Attribute 4	$-$	0.07	$+$	0.00	-1387.804	2015.425	0.419	$-$	0.02	$+$	0.00	-1382.396	2026.241	**0.422**
Attribute 5	$-$	0.59	$+$	0.00	-657.057	1081.403	0.449	$-$	0.07	$+$	0.00	-657.514	1080.489	0.449
Attribute 6	$-$	0.81	$+$	0.01	-851.045	1491.932	0.465	$-$	0.20	$+$	0.00	-846.745	1500.533	**0.470**
Attribute 7	$-$	0.89	$+$	0.00	-1733.594	2920.856	0.456	$-$	0.01	$-$	0.00	-1721.260	2945.1525	**0.460**
Attribute 8	$+$	0.00	$+$	0.00	-2749.349	2087.104	0.274	$-$	0.00	$-$	0.00	-2727.931	2129.941	**0.280**
Attribute 9	$-$	0.91	$+$	0.00	-2294.929	2995.945	**0.394**	$-$	0.45	$+$	0.10	-2297.834	2990.134	0.393
Attribute P	$+$	0.50	$+$	0.00	-3795.847	393.362	**0.049**	$-$	0.00	$+$	0.09	-3805.257	374.542	0.046
Global	$+$	0.50	$+$	0.00	-18,109.689	19,676.010	0.352	$-$	0.00	$-$	0.00	-17,954.487	19,986.414	**0.357**

Table 5.5: Comparison of *lead* and $|lead|$ for Wind Power data without Panel Specification.

	Using *lead*							Using $\|lead\|$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
(Size of farms)	−	0.37	−	0.42	-1205.712	2487.740	**0.507**	−	0.57	+	0.95	-1206.046	2487.073	0.506
(Maximum height)	−	0.09	+	0.68	-1432.023	2035.118	0.414	−	0.10	−	0.36	-1431.684	2035.796	0.414
Red kite	−	0.49	−	0.70	-1368.248	2162.668	0.440	−	0.63	−	0.54	-1368.105	2162.955	0.440
Minimum distance	−	0.04	−	0.04	-1206.131	2486.903	0.506	−	0.05	+	0.01	-1203.404	2492.357	**0.508**
Price	−	0.72	−	0.23	-1221.994	2455.176	0.500	+	0.16	−	0.06	-1220.474	2458.216	**0.501**
				Init. log-likelihood:	-2449.582					Init. log-likelihood:	-2449.582			
Global	−	0.01	−	0.01	-6451.054	11,593.713	0.473	−	0.21	+	0.08	-6452.652	11,590.517	0.473
				Init. log-likelihood:	-12,247.911					Init. log-likelihood:	-12,247.911			

Table 5.6: Comparison of *lead* and |*lead*| for Wind Power data with Panel Specification.

	Using *lead*							Using $\lvert lead \rvert$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
(Size of farms)	−	0.41	+	0.48	-1205.485	2488.193	0.507	−	0.32	+	0.66	-1205.687	2487.790	0.507
(Maximum height)	−	0.17	+	0.04	-1430.186	2038.791	**0.415**	−	0.17	−	0.50	-1432.237	2034.691	0.414
Red kite	−	0.68	−	0.15	-1367.045	2165.074	0.441	+	0.28	−	0.26	-1367.424	2164.317	0.441
Minimum distance	−	0.34	−	0.78	-1207.488	2484.189	0.506	−	0.55	+	0.03	-1205.896	2487.372	0.506
Price	+	0.84	−	0.86	-1222.755	2453.654	0.500	+	0.25	−	0.00	-1220.474	2458.216	**0.501**
				Init. log-likelihood:	-2449.582						Init. log-likelihood:	-2449.582		
Global	−	0.08	−	0.13	-6453.311	11,589.199	0.473	−	0.85	+	0.31	-6454.279	11,587.264	0.473
				Init. log-likelihood:	-12,247.911						Init. log-likelihood:	-12,247.911		

Table 5.7: Comparison of *lead* and $|lead|$ for Wind Power data with individual betas.

	Using *lead*							Using $\lvert lead \rvert$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
(Size of farms)	−	0.10	+	0.01	-1196.799	2500.021	0.510	−	0.01	+	0.00	-1197.072	2499.475	0.510
(Maximum height)	−	0.01	+	0.81	-1421.116	2051.386	0.418	−	0.01	+	0.39	-1421.763	2050.092	0.418
Red kite	−	0.12	+	0.16	-1362.509	2168.601	0.442	−	0.02	+	0.00	-1361.725	2170.168	0.442
Minimum distance	−	0.77	+	0.06	-1203.888	2485.842	0.507	−	0.23	+	0.00	-1200.982	2491.654	**0.508**
Price	−	0.55	+	0.02	-1214.146	2465.328	0.503	−	0.00	−	0.05	-1210.588	2472.443	**0.504**
				Init. log-likelihood:	-2446.810					Init. log-likelihood:	-2446.810			
Global	−	0.00	+	0.00	-6412.965	11,642.165	0.476	−	0.00	+	0.00	-6413.512	11,641.071	0.476
				Init. log-likelihood:	-12,234.048					Init. log-likelihood:	-12,234.048			

The initial log-likelihood is slightly different in this model compared to the two previous ones (without and with panel specification) because we removed the observation of one respondents who answered only four questions instead of five for technical reasons.

Table 5.8: Comparison of *lead* and $|lead|$ for Potatoes data using MNL.

	Using *lead*							Using $\|lead\|$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
Texture	–	0.17	–	0.00	-444.108	317.861	**0.259**	–	0.32	+	0.54	-450.605	304.866	0.248
Price	–	0.35	–	0.85	-254.888	696.300	0.572	–	0.05	–	0.39	-252.604	700.868	**0.576**
(Colour of flesh)	–	0.10	–	0.85	-480.607	244.862	0.198	–	0.12	–	0.67	-480.491	245.094	0.198
Production	+	0.46	+	0.92	-280.174	645.728	0.530	–	0.77	–	0.56	-280.330	645.416	0.530
Nutrition	–	0.04	–	0.17	-255.885	694.305	0.571	–	0.07	–	0.30	-255.802	694.473	0.571
Country	–	0.14	–	0.94	-395.262	415.552	0.340	–	0.03	–	0.77	-394.515	417.046	**0.341**
				Init. log-likelihood:	-603.038						Init. log-likelihood:	-603.038		
Global	–	0.00	–	0.44	-2197.836	2840.785	0.392	–	0.00	+	0.00	-2193.651	2849.155	**0.393**
				Init. log-likelihood:	-3618.228						Init. log-likelihood:	-3618.228		

Table 5.9: Comparison of *lead* and $|lead|$ for Potatoes data using Mixed Logit.

	Using *lead*							Using $\|lead\|$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
Texture	−	0.22	−	0.84	-450.729	304.618	0.248	−	0.32	+	0.12	-449.782	306.512	**0.249**
Price	−	0.11	+	0.05	-252.368	701.340	**0.577**	−	0.29	−	0.91	-254.891	696.294	0.572
(Color of flesh)	−	0.02	+	0.02	-475.679	254.719	0.206	−	0.01	+	0.01	-475.827	254.423	0.206
Production	+	0.49	+	0.00	-272.968	660.141	0.542	−	0.15	+	0.00	-273.211	659.653	0.542
Nutrition	−	0.03	+	0.97	-254.469	697.137	**0.573**	−	0.06	+	0.16	-255.195	695.687	0.572
Country	−	0.08	+	0.30	-392.717	420.642	0.344	−	0.03	−	0.04	-391.502	423.072	**0.346**
				Init. log-likelihood:	-603.038						Init. log-likelihood:	-603.038		
Global	−	0.00	+	0.00	-2176.584	2883.289	**0.398**	−	0.00	+	0.00	-2183.769	2868.918	0.396
				Init. log-likelihood:	-3618.228						Init. log-likelihood:	-3618.228		

In the attribute attendance probability model using estimates from the choice probability model scaled by category and wave (Table 5.3) *lead* is significant, but with opposite signs to those for $|lead|$. Moreover, fewer attributes are significant in this attribute attendance probability models. This could be explained by the weight of the price attribute in the global evaluation. Indeed the price attribute is always available to respondents so it is has much more observations than other attributes.

The panel mixed model (Table 5.4) gives different values and level of significance for the attributes compared to previous attribute attendance probability models. However, even if the results are different, $|lead|$ is still more significant globally. Thus the results for Cortina data suggests the use of $|lead|$ rather than *lead*.

For the wind power datas the two first attribute attendance probability models, without and with panel specification, respectively Tables 5.5 and 5.6, provide disappointing results. For most of the attributes, *lead* and $|lead|$ are not significantly different from zero. It is thus difficult to derive an indication as to one is to be prefered to the other. When using the utilities computed at the individual specific means for the distributions of β_n's, the results are clearly better with more significant factors. In this case, we have a very slight preference for the attribute attendance probability model with $|lead|$ since $\overline{\rho}^2$ for two attributes are a bit higher in this case. Note that in this model, we also have the predicted signs by Cameron's theory.

With the potatoes data, using a classic MNL choice probability model, we have poor results at the attribute level. So it is difficult to draw a conclusion about the use of *lead* or its absolute value. In a global analysis we note the expected signs and again a slightly better fit for $|lead|$. The regressions on the utilities computed on individual specific estimates from the panel mixed logit give expected signs for $|lead|$ and also for *lead*. At the attribute level, two attributes have better fit in each different specification of *lead*. Globally, *lead* fits the data slightly better.

Why is this comparison between *lead* and $|lead|$ a preliminary condition to their use of a more general model of attendance? First, because we obviously cannot use both of them at the same time. They are highly correlated and it has no sense to use both of them. But mostly because using *lead* or $|lead|$ has different interpretations about the behaviour of choice makers if we follow Cameron's theory. Note that *lead* is supposed to have a positive β, i.e. to be an important determinant in attendance to attribute. In the own-attribute case, considering a bounded utility function made of one attribute, if the difference in utility between the two leading alternatives is small, there is little reason for attendance to this attribute. If the value of lead is close to zero attendance is in theory going to be less likely since this attribute does not help the choice maker to identify the best alternatives. If this difference, *lead*, is positive, i.e., if the attribute confirm the *a priori* knowledge of the choice maker about his choice, going in the same direction, using $|lead|$ or *lead* does not change anything. But, and this is the interesting case, if this difference, *lead*, is negative, it means that the choice maker is facing an attribute that supports the second leading alternative, an attribute that is perhaps going to create a doubt in the mind of the choice maker. In this particular case, using the absolute value $|lead|$ to explain attendance means that the choice maker is going to attend all attributes that could give him more information for his choice assuming his *a priori* knowledge represented by the full-information choice probability model. He is going to attend an attribute even if this attribute is giving him a preference for the second leading alternative and not the first one. In the contrary, using *lead* without absolute value means in the case of a negative difference that the larger this difference is the less likely it is to attend attribute. The interpretation would be that choice makers do not attend attributes going to make them change their decision or going in the opposite direction than their decision before considering the attribute.

On the other hand, in the case of other-attribute dissimilarities, using a utility function with all attributes except the one we consider, the theory suggests that if this difference is null or small, the attribute not considered is deserving attention because considering it may help decide between the two alternatives. And if this difference is – instead – large and positive, the attribute is unlikely to be attended because the choice maker knows that it is unlikely that attending to the missing attribute can change the preferred alternative. In the particular case of a negative difference, it means that the missing attribute is pivotal and hence we expect the attention to this attribute to be more likely when this attribute is pivotal.

In our different tests, we observe that the absolute value $|lead|$ is fitting the observed attendance slightly better. Thus we are going to use this specification with absolute value in our next attribute attendance probability models instead of *lead* without absolute value.

5.2 Bounded leading alternatives

Now we know that $|lead|$ is a better explanatory variable than *lead* for attention to attributes.

We still need to know if it is better to use a local specification of *lead*, $lead_{local}$, based on the bounded utility, as described in Section 4.2 about explanatory variables, or a specification using the complete utility function in the choice probability models. Remember here that in the local specification, $lead_{local}$ is always positive, so we don't need to make the previous analysis of the better specification, like in Section 5.1.

To test these two different specifications, we estimate again two different attribute attendance probability models using only $lead(x_k\beta_k)$ and $lead(x'_{-k}\beta_{-k})$ as explanatory variables, once using bounded utility function and once with the complete utility function. The first attribute attendance probability model is:

$$y = ASC + \beta_{other} * lead_{local}(x'_{-k}\beta_{-k}) + \beta_{own} * lead_{local}(x_k\beta_k)$$

And the second attribute attendance probability model, with absolute value, is:

$$y = ASC + \beta_{other} * |lead(x'_{-k}\beta_{-k})| + \beta_{own} * |lead(x_k\beta_k)|$$

for each attribute k and where y is the dependent variable with value 1 if attribute is attended and 0 otherwise and ASC is an constant.

Attribute attendance probability model results can be seen in tables 5.10, 5.11, 5.12, 5.13, 5.14, 5.15, 5.16, 5.17 and 5.18 using estimates of each different choice probability models for different data sets. Results for the first attribute attendance probability model, with local specification, are in the left half of the tables and results for the second attribute attendance probability model, using a complete utility function to define leading alternatives, are in the right half of the tables.

For the Cortina data, all attribute attendance probability models fit the data better when using $lead_{local}$.

In the wind power case all attribute attendance probability models fit the data similarly well. In the attribute level, there is a slight preference for $|lead|$ but we can observe that neither $lead_{local}$ and $|lead|$ are significant in most cases. In the last model, with individual specific means for the distributions of β_n there is a clear improvement but both specifications are still globally equivalent.

With the potatoes data, as for the Cortina data, the specification with $lead_{local}$ is better for both attribute attendance probability models globally and for the attribute level we also have better $\overline{\rho}^2$.

Remember that we use the choice probability models, about potatoes or wind power policy in our examples, to have a linear utility function giving the importance weights of each attribute on the probability of choice. We use this weights as an *a priori* knowledge by the choice maker about the importance of each attribute in his choice. In particular, *lead* is the utility difference between leading alternatives and is assumed to motivate, at least in part, the attention to attributes. Comparing these two different specifications of lead, $lead_{local}$ and $|lead|$ casts some light on the use by choice makers of this *a priori* knowledge. In the local specification, the choice maker is considering the utility difference between the two locally leading alternatives. It suggests that he has an *a priori* knowledge of the effect of this attribute in the utility function on its own but maybe not much knowledge of the overall utility function and consequently of his general choice. Using the general specification $|lead|$ we assume that the choice maker has some knowledge not only of the effect of the attribute on its own but also of the complete utility function *before* attending to all attributes.

The results of Tables 5.10 to 5.18 seem to give a preference for the local specification $lead_{local}$. Thus we are going to use it in place of the previously

Table 5.10: Comparison of $lead_{local}$ and $|lead|$ for Cortina data using basic MNL.

	Using $lead_{local}$							Using $\|lead\|$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
Attribute 1	–	0.12	+	0.21	-1207.692	2774.901	0.533	–	0.14	+	0.01	-1204.862	2780.562	**0.535**
Attribute 2	+	0.90	–	0.59	-816.272	2359.983	0.590	+	0.41	–	0.02	-813.302	2365.924	**0.591**
Attribute 3	–	0.08	+	0.00	-1387.896	4011.505	**0.591**	–	0.07	+	0.64	-1393.056	4001.186	0.589
Attribute 4	–	0.03	+	0.91	-1396.926	1997.180	0.416	–	0.02	+	0.00	-1390.632	2009.769	**0.418**
Attribute 5	–	0.02	+	0.14	-660.971	1073.575	**0.446**	–	0.04	+	0.38	-662.138	1071.240	0.445
(Attribute 6)	–	0.03	+	0.47	-851.934	1490.155	**0.465**	–	0.04	–	1.00	-852.210	1489.602	0.464
Attribute 7	–	0.09	–	0.00	-1717.303	2953.438	0.461	–	0.16	–	0.00	-1715.249	2957.547	**0.462**
Attribute 8	–	0.16	–	0.00	-2582.106	2421.590	**0.318**	+	0.30	–	0.00	-2705.536	2174.730	**0.286**
(Attribute 9)	–	0.00	–	0.05	-2285.259	3015.285	**0.397**	–	0.00	+	0.70	-2287.231	3011.340	0.396
Attribute P	+	0.00	–	0.01	-3865.432	254.192	**0.031**	–	0.96	–	0.43	-3895.521	194.014	0.024
Global	–	0.08	+	0.00	-17471.433	20952.522	**0.375**	–	0.00	–	0.00	-17,683.355	20,528.679	0.367

Table 5.11: Comparison of $lead_{local}$ and $|lead|$ for Cortina data using MNL Scaled by Category.

	Using $lead_{local}$							Using $\lvert lead \rvert$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
Attribute 1	−	0.08	+	0.34	-1207.719	2774.848	0.533	−	0.15	+	0.00	-1202.266	2785.755	**0.536**
Attribute 2	+	0.84	−	0.52	-816.180	2360.167	0.590	+	0.59	−	0.01	-812.916	2366.695	**0.591**
Attribute 3	−	0.11	+	0.00	-1387.206	4012.886	**0.590**	−	0.06	+	0.29	-1392.543	4002.212	0.589
Attribute 4	−	0.05	+	0.57	-1397.317	1996.399	0.415	−	0.04	+	0.00	-1389.177	2012.678	**0.419**
Attribute 5	−	0.01	+	0.15	-660.310	1074.896	0.446	−	0.01	+	0.23	-660.696	1074.124	0.446
(Attribute 6)	−	0.03	−	0.17	-851.070	1491.882	0.465	−	0.03	+	0.77	-851.977	1490.068	0.465
Attribute 7	−	0.13	−	0.00	-1722.074	2943.897	0.460	−	0.22	−	0.00	-1719.153	2949.738	**0.461**
Attribute 8	−	0.08	−	0.00	-2580.790	2424.223	**0.319**	+	0.79	−	0.00	-2685.580	2214.643	0.291
(Attribute 9)	−	0.00	−	0.11	-2287.774	3010.255	0.396	−	0.00	+	0.10	-2287.736	3010.330	0.396
Attribute P	−	0.14	−	0.00	-3868.684	247.687	**0.030**	−	0.40	−	0.34	-3894.463	196.130	0.024
Global	+	0.11	+	0.24	-17443.279	21008.831	**0.376**	−	0.00	−	0.00	-17,625.123	20,645.144	0.369

Table 5.12: Comparison of $lead_{local}$ and $|lead|$ for Cortina data using MNL Scaled by Category and Wave.

	Using $lead_{local}$							Using $\|lead\|$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
Attribute 1	−	0.33	+	0.33	-1208.645	2772.996	0.533	−	0.32	+	0.00	-1204.709	2780.868	**0.535**
Attribute 2	+	0.83	−	0.16	-815.349	2361.829	0.590	+	0.78	−	0.00	-811.976	2368.575	**0.592**
Attribute 3	−	0.06	+	0.00	-1382.500	4022.297	**0.592**	−	0.03	+	0.98	-1392.483	4002.332	0.589
Attribute 4	−	0.07	−	0.15	-1396.774	1997.486	0.416	−	0.03	+	0.02	-1394.940	2001.154	0.416
Attribute 5	−	0.02	+	0.22	-661.109	1073.298	**0.446**	−	0.03	+	0.28	-661.483	1072.550	0.445
Attribute 6	−	0.11	+	0.67	-853.032	1487.959	0.464	−	0.07	−	0.65	-852.529	1488.964	0.464
Attribute 7	−	0.02	−	0.00	-1716.230	2955.585	0.462	−	0.12	−	0.00	-1708.634	2970.776	**0.464**
Attribute 8	−	0.17	−	0.00	-2583.600	2418.602	**0.318**	+	0.74	−	0.00	-2681.980	2221.843	0.292
(Attribute 9)	−	0.00	−	0.00	-2283.227	3019.348	**0.397**	−	0.00	−	0.34	-2288.800	3008.202	0.396
Attribute P	+	0.00	−	0.00	-3875.762	233.532	**0.028**	−	0.22	−	0.03	-3891.572	201.912	0.025
Global	−	0.00	−	0.00	-17572.097	20751.194	**0.371**	−	0.00	−	0.00	-17,713.444	20,468.500	0.366

Table 5.13: Comparison of $lead_{local}$ and $|lead|$ for Cortina data using mixed logit estimates.

	Using $lead_{local}$							Using $\|lead\|$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
Attribute 1	−	0.03	+	0.00	-1197.948	2794.390	0.537	−	0.03	+	0.00	-1197.576	2795.133	0.537
Attribute 2	−	0.88	+	0.90	-816.439	2359.649	0.590	+	0.86	−	0.55	-816.263	2360.003	0.590
Attribute 3	−	0.91	+	0.00	-1364.161	4058.975	**0.597**	−	0.52	+	0.00	-1381.109	4025.080	0.592
Attribute 4	−	0.02	+	0.00	-1391.652	2007.729	0.418	−	0.02	+	0.00	-1382.396	2026.241	**0.422**
Attribute 5	−	0.09	+	0.00	-655.915	1083.686	**0.450**	−	0.07	+	0.00	-657.514	1080.489	0.449
Attribute 6	−	0.28	+	0.00	-844.559	1504.905	0.469	−	0.20	+	0.00	-846.745	1500.533	**0.470**
Attribute 7	−	0.01	−	0.00	-1731.419	2925.206	0.457	−	0.01	−	0.00	-1721.260	2945.1525	**0.460**
Attribute 8	−	0.00	−	0.00	-2658.086	2269.632	**0.298**	−	0.00	−	0.00	-2727.931	2129.941	0.280
Attribute 9	−	0.50	−	0.07	-2297.615	2990.574	0.393	−	0.45	+	0.10	-2297.834	2990.134	0.393
Attribute P	−	0.00	−	0.00	-3833.919	317.217	0.039	−	0.00	+	0.09	-3805.257	374.542	**0.046**
Global	−	0.00	−	0.00	-17927.998	20039.392	**0.358**	−	0.00	−	0.00	-17,954.487	19,986.414	0.357

Table 5.14: Comparison of $lead_{local}$ and $|lead|$ for Wind Power data without Panel Specification.

	Using $lead_{local}$							Using $\lvert lead \rvert$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
(Size of farms)	−	0.59	+	0.52	-1205.829	2487.506	**0.507**	−	0.57	+	0.95	-1206.046	2487.073	0.506
(Maximum height)	−	0.09	+	0.59	-1431.957	2035.251	0.414	−	0.10	−	0.36	-1431.684	2035.796	0.414
Red kite	−	0.46	−	0.45	-1367.919	2163.326	0.440	−	0.63	−	0.54	-1368.105	2162.955	0.440
Minimum distance	−	0.15	−	0.70	-1207.319	2484.526	0.506	−	0.05	+	0.01	-1203.404	2492.357	**0.508**
Price	+	0.65	−	0.04	-1220.977	2457.211	0.500	+	0.16	−	0.06	-1220.474	2458.216	**0.501**
				Init. log-likelihood:	-2449.582						Init. log-likelihood:			
											-2449.582			
Global	−	0.15	+	0.78	-6453.758	11588.306	0.473	−	0.21	+	0.08	-6452.652	11,590.517	0.473
				Init. log-likelihood:	-12,247.911						Init. log-likelihood:			
											-12,247.911			

Table 5.15: Comparison of $lead_{local}$ and $|lead|$ for Wind Power data with Panel Specification.

	Using $lead_{local}$							Using $\|lead\|$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
(Size of farms)	−	0.28	+	0.36	-1205.416	2488.333	0.507	−	0.32	+	0.66	-1205.687	2487.790	0.507
(Maximum height)	−	0.16	−	0.23	-1431.745	2035.674	0.414	−	0.17	−	0.50	-1432.237	2034.691	0.414
Red kite	+	0.38	−	0.34	-1367.787	2163.591	0.440	+	0.28	−	0.26	-1367.424	2164.317	0.441
Minimum distance	−	0.77	−	0.74	-1208.225	2482.715	0.506	−	0.55	+	0.03	-1205.896	2487.372	0.506
Price	+	0.45	−	0.05	-1220.788	2457.589	0.500	+	0.25	−	0.00	-1220.474	2458.216	**0.501**
				Init. log-likelihood:	-2449.582						Init. log-likelihood:	-2449.582		
Global	−	1.00	+	0.77	-6454.774	11,586.273	0.473	−	0.85	+	0.31	-6454.279	11,587.264	0.473
				Init. log-likelihood:	-12,247.911						Init. log-likelihood:	-12,247.911		

Table 5.16: Comparison of $lead_{local}$ and $|lead|$ for Wind Power data with individual betas.

	Using $lead_{local}$							Using $\|lead\|$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
(Size of farms)	−	0.01	+	0.06	-1198.903	2495.813	0.509	−	0.01	+	0.00	-1197.072	2499.475	**0.510**
(Maximum height)	−	0.01	+	0.25	-1421.459	2050.701	0.418	−	0.01	+	0.39	-1421.763	2050.092	0.418
Red kite	−	0.01	+	0.03	-1362.098	2169.424	0.442	−	0.02	+	0.00	-1361.725	2170.168	0.442
Minimum distance	−	0.25	+	0.01	-1202.620	2488.379	0.507	−	0.23	+	0.00	-1200.982	2491.654	**0.508**
Price	−	0.00	−	0.03	-1208.183	2477.253	**0.505**	−	0.00	−	0.05	-1210.588	2472.443	0.504
				Init. log-likelihood:	-2446.810						Init. log-likelihood:	-2446.810		
Global	−	0.00	+	0.00	-6412.857	11,642.382	0.476	−	0.00	+	0.00	-6413.512	11,641.071	0.476
				Init. log-likelihood:	-12,234.048						Init. log-likelihood:	-12,234.048		

Table 5.17: Comparison of $lead_{local}$ and $|lead|$ for Potatoes data using MNL.

	Using $lead_{local}$							Using $\lvert lead \rvert$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
Texture	−	0.15	−	0.10	-449.523	307.030	**0.250**	−	0.32	+	0.54	-450.605	304.866	0.248
Price	−	0.01	−	0.87	-251.218	703.640	**0.578**	−	0.05	−	0.39	-252.604	700.868	0.576
(Colour of flesh)	−	0.12	−	0.14	-479.570	246.937	**0.200**	−	0.12	−	0.67	-480.491	245.094	0.198
Production	+	0.89	−	0.71	-280.549	644.978	0.530	−	0.77	−	0.56	-280.330	645.416	0.530
Nutrition	−	0.04	−	0.53	-255.728	694.620	0.571	−	0.07	−	0.30	-255.802	694.473	0.571
Country	−	0.02	−	0.42	-394.461	417.155	0.341	−	0.03	−	0.77	-394.515	417.046	0.341
				Init. log-likelihood:	-603.038						Init. log-likelihood:	-603.038		
Global	−	0.00	+	0.00	-2189.997	2856.462	**0.394**	−	0.00	+	0.00	-2193.651	2849.155	0.393
				Init. log-likelihood:	-3618.228						Init. log-likelihood:	-3618.228		

Table 5.18: Comparison of $lead_{local}$ and $|lead|$ for Potatoes data using Mixed Logit.

	Using $lead_{local}$							Using $\|lead\|$						
	β_{other}		β_{own}					β_{other}		β_{own}				
	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$	Value	p-value	Value	p-value	Final logL	L ratio test	$\overline{\rho}^2$
Texture	–	0.24	+	0.01	-448.144	309.789	**0.252**	–	0.32	+	0.12	-449.782	306.512	0.249
Price	–	0.01	–	0.84	-251.252	703.572	**0.578**	–	0.29	–	0.91	-254.891	696.294	0.572
(Colour of flesh)	–	0.01	+	0.02	-475.127	255.821	**0.207**	–	0.01	+	0.01	-475.827	254.423	0.206
Production	–	0.04	+	0.00	-273.171	659.734	0.542	–	0.15	+	0.00	-273.211	659.653	0.542
Nutrition	–	0.02	+	0.15	-254.396	697.285	**0.573**	–	0.06	+	0.16	-255.195	695.687	0.572
Country	–	0.03	+	0.96	-394.125	417.826	0.341	–	0.03	–	0.04	-391.502	423.072	**0.346**
				Init. log-likelihood:	-603.038						Init. log-likelihood:	-603.038		
Global	–	0.00	+	0.00	-2161.079	2914.298	**0.402**	–	0.00	+	0.00	-2183.769	2868.918	0.396
				Init. log-likelihood:	-3618.228						Init. log-likelihood:	-3618.228		

defined *lead* and $|lead|$. It means that the assumption of an *a priori* knowledge using the results of the complete information models is maybe valid but only, in the case of *lead*, from a "local" point of view.

5.3 Complete utility function for attendance

Similarly to the two previous sections, we estimate attribute attendance probability models. We now use all the available factors described in Section 4.2: (i) the own- and other-attribute utility dissimilarities for *lead* (and we use, as explained in Section 5.2, $lead_{local}$ specification), (ii) entropy, (iii) standard deviation, (iv) skewness, (vi) the effect of being pivotal, (vii) the effect of the number of attributes and – when possible – ($viii$) that of their order.

We estimate an attribute attendance probability model with all of these factors and an intercept:

$$\begin{aligned} y \ = ASC \quad & +\beta_{own}^{lead} \cdot lead_{local}(x_k\beta_k) + \beta_{other}^{lead} \cdot lead_{local}(x'_{-k}\beta_{-k}) \\ & +\beta_{own}^{entropy} \cdot H_k + \beta_{other}^{entropy} \cdot H_{-k} \\ & +\beta^{pivotal} \cdot Pivotal + \beta^{\#} \cdot \# \\ & +\beta_{own}^{sd} \cdot SD(x_k\beta_k) + \beta_{other}^{sd} \cdot SD(x'_{-k}\beta_{-k}) \\ & +\beta_{own}^{skew} \cdot Skew(x_k\beta_k) + \beta_{other}^{skew} \cdot Skew(x'_{-k}\beta_{-k}) \end{aligned}$$

for each attribute k, with # the number of attributes. An example based on Cortina data using the basic MNL model estimates for β's can be seen in Table 5.19 with for each factors the sign of the estimates and their p-values.

Most of the factors are not significant and we have significant correlations between some of them[1]. To avoid as much as possible side effects of these correlations on our attribute attendance probability models, we systematically remove from each model the factor with highest $\overline{\rho}^2$ value until having only significant factors. The final results for Cortina data using basic MNL estimates for β's can be seen in Table 5.20 for comparison with Table 5.19. For all other different attribute attendance probability models and data we present only this final version, in Tables 5.21, 5.22, 5.23, 5.24, 5.25, 5.26, 5.27 and 5.28.

We first describe the expectation we have about the different factors, and then will describe the results.

We expect a positive sign for own-attribute $lead_{local}$ since if the own-attribute utility difference between the two leading alternatives is large it means that this attribute has a large potential to change the choice of the decision maker. On the contrary, little attention stems from a small difference across alternatives.

For other-attribute $lead_{local}$ a negative sign is expected. Indeed, when it is small it means that without the attribute the utility functions of the different

[1]As an example, in this particular case, 47 of the 55 possible couples of parameters have correlation significantly different from 0.

Table 5.19: Full model of attention to attributes for Cortina data using basic MNL

			$lead_{local}$				Entropy								Standard deviation				Skewness						
	ASC		own		other		own		other		Pivotal		# attributes		own		other		own		other				
	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Final logL	L ratio test	$\overline{\rho}^2$
Attribute 1	−	0.07	−	0.39	+	0.10	+	0.08	+	0.10	−	0.11	−	0.68	+	0.03	+	0.51	+	0.00	−	0.04	-1195.830	2798.626	0.535
Attribute 2	+	0.39	−	0.72	−	0.44	+	0.49	−	0.25	−	0.00	−	0.27	+	0.22	−	0.18	+	0.95	+	0.36	-807.842	2376.843	0.590
Attribute 3	−	0.65	+	0.00	−	0.01	+	0.62	−	0.07	−	0.14	+	0.00	−	0.03	−	0.43	−	0.00	+	0.07	-1358.194	4070.909	0.597
Attribute 4	+	0.14	−	0.00	+	0.57	−	0.12	+	0.01	+	0.09	+	0.54	+	0.25	+	0.01	+	0.00	+	0.52	-1381.599	2027.835	0.419
Attribute 5	−	0.46	+	0.30	−	0.74	+	0.42	−	0.83	−	1.00	−	0.00	−	0.36	−	0.56	−	0.31	−	0.86	-652.794	1089.928	0.446
Attribute 6	+	1.00	−	0.29	−	0.25	+	1.00	−	0.66	+	0.11	−	0.11	+	0.93	+	0.97	+	0.10	+	0.40	-847.209	1499.604	0.463
Attribute 7	−	0.32	−	0.00	−	0.20	+	0.32	+	0.52	+	0.29	+	0.00	+	0.47	+	0.12	+	0.02	+	0.03	-1688.401	3011.243	0.468
Attribute 8	−	0.00	+	0.16	−	0.36	+	0.00	+	0.20	+	0.76	+	0.00	+	0.03	+	0.08	−	0.00	+	0.10	-2469.264	2647.274	0.346
Attribute 9	−	1.00	−	0.36	−	0.14	+	1.00	−	0.31	−	0.92	+	0.01	+	0.90	−	0.09	+	0.48	+	0.31	-2275.850	3034.103	0.397
Attribute P	+	0.05	−	0.00	−	0.32	−	0.00	+	0.27	+	0.94	+	0.00	−	0.00	+	0.02	−	0.76	+	0.04	-3542.889	899.278	0.110
Global	+	0.00	−	0.00	+	0.41	−	0.00	+	0.00	+	0.72	+	0.00	−	0.00	+	0.23	+	0.01	+	0.01	-16,936.906	22,021.576	0.394

Table 5.20: Model of attention to attributes without non-significant factors for Cortina data using basic MNL

	ASC		$lead_{local}$ own		$lead_{local}$ other		Entropy own		Entropy other		Pivotal		# attributes		Standard deviation own		Standard deviation other		Skewness own		Skewness other				
	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Final logL	L ratio test	$\overline{\rho}^2$
Attribute 1	−	0.01					+	0.01							+	0.01			+	0.00			-1202.124	2786.038	0.535
Attribute 2	+	0.00									−	0.01					−	0.05					-811.912	2368.704	0.592
Attribute 3			+	0.00			+	0.00					+	0.00	−	0.00			−	0.00			-1362.653	4061.991	0.597
Attribute 4	+	0.00	−	0.00			−	0.00	+	0.01									+	0.00			-1388.257	2014.520	0.418
Attribute 5					−	0.03	+	0.00					−	0.00									-654.810	1085.896	0.451
Attribute 6					−	0.04	+	0.00															-852.201	1489.621	0.465
Attribute 7			−	0.00	−	0.00	+	0.00					+	0.00					+	0.00	+	0.01	-1691.162	3005.720	0.469
Attribute 8	−	0.00					+	0.00	+	0.00			+	0.00	+	0.01			−	0.00			-2472.380	2641.043	0.347
Attribute 9	+	0.00	−	0.01									+	0.00			−	0.00					-2278.401	3029.002	0.398
Attribute P			−	0.00			−	0.00	+	0.00			+	0.00	−	0.00	+	0.00			+	0.00	-3544.965	895.125	0.110
Global	+	0.00	−	0.00			−	0.00	+	0.00			+	0.00	−	0.00			+	0.01	+	0.00	-16,937.682	22,020.024	0.394

Table 5.21: Model of attention to attributes without non-significant factors for Cortina data using MNL Scaled by Category

	ASC		$lead_{local}$ own		$lead_{local}$ other		Entropy own		Entropy other		Pivotal		# attributes		Standard deviation own		Standard deviation other		Skewness own		Skewness other				
	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Final logL	L ratio test	$\overline{\rho}^2$
Attribute 1	−	0.01					+	0.01							+	0.01			+	0.00			-1201.900	2786.485	0.535
Attribute 2							+	0.00			−	0.00			+	0.1	−	0.03					-809.574	2373.379	0.592
Attribute 3			+	0.00	−	0.0	+	0.00			−	0.03	+	0.00	−	0.00			−	0.00			-1339.135	4109.027	0.603
Attribute 4			−	0.00					+	0.00					+	0.00			+	0.00			-1386.484	2018.065	0.420
Attribute 5							+	0.00					−	0.00									-657.074	1081.368	0.450
Attribute 6	+	0.00			−	0.03																	-852.019	1489.984	0.465
Attribute 7			−	0.00	−	0.01			+	0.00			+	0.00			+	0.00	+	0.00	+	0.01	-1693.138	3001.768	0.468
Attribute 8	−	0.00					+	0.00	+	0.00			+	0.00	+	0.02			−	0.00			-2472.281	2641.240	0.347
Attribute 9	+	0.00			−	0.00							+	0.01							+	0.00	-2281.479	3022.845	0.397
Attribute P			−	0.00			−	0.00	+	0.00			+	0.00	−	0.00	+	0.00			+	0.00	-3544.086	896.883	0.111
Global	+	0.00	−	0.00			−	0.00	+	0.00			+	0.00	−	0.00					+	0.00	-16,907.111	22,081.167	0.395

Table 5.22: Model of attention to attributes without non-significant factors for Cortina data using MNL Scaled by Category and Wave

			$lead_{local}$				Entropy								Standard deviation				Skewness						
	ASC		own		other		own		other		Pivotal		# attributes		own		other		own		other				
	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Final logL	L ratio test	$\overline{\rho}^2$
Attribute 1									+	0.00	−	0.01							+	0.00			-1202.812	2784.663	0.535
Attribute 2					+	0.00			+	0.00													-815.882	2360.763	0.590
Attribute 3			+	0.00	−	0.00	+	0.00					+	0.00	−	0.00			−	0.00			-1343.428	4100.442	0.602
Attribute 4	−	0.00			+	0.02	+	0.00	+	0.00					+	0.00	+	0.00					-1380.893	2029.248	0.421
Attribute 5					−	0.04	+	0.00					−	0.00									-654.961	1085.594	0.451
Attribute 6	+	0.00																					-854.359	1485.304	0.464
Attribute 7			−	0.00	−	0.02	+	0.00					+	0.00					+	0.00			-1693.202	3001.641	0.468
Attribute 8	−	0.00			−	0.01	+	0.00					+	0.00	+	0.00			−	0.00			-2472.496	2640.812	0.347
Attribute 9					−	0.02			+	0.00			+	0.00					−	0.00			-2278.043	3029.717	0.398
Attribute P	+	0.02	−	0.00			−	0.00	+	0.00			+	0.00	−	0.00	+	0.00					-3546.271	892.513	0.110
Global	+	0.00	−	0.00	−	0.00	−	0.00			−	0.01	+	0.00	−	0.00			+	0.00	+	0.00	-17,074.522	21,746.346	0.389

Table 5.23: Model of attention to attributes without non-significant factors for Cortina data using mixed logit estimates

			$lead_{local}$				Entropy								Standard deviation				Skewness						
	ASC		own		other		own		other		Pivotal		# attributes		own		other		own		other				
	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Final logL	L ratio test	$\overline{\rho}^2$
Attribute 1	+	0.00					−	0.00	+	0.00							+	0.02					-1193.639	2803.009	0.539
Attribute 2					+	0.00	−	0.00	+	0.00	−	0.01									−	0.00	-805.990	2380.547	0.594
Attribute 3	+	0.00					−	0.00					+	0.00					+	0.00			-1335.428	4116.150	0.605
Attribute 4	+	0.00					−	0.00	+	0.01			+	0.00	−	0.04							-1382.318	2026.397	0.421
Attribute 5	+	0.00					−	0.00					−	0.00									-651.838	1091.841	0.453
Attribute 6	+	0.00					−	0.00							−	0.00							-834.900	1524.221	0.475
Attribute 7	+	0.00	−	0.00	+	0.00	−	0.00	+	0.00			+	0.00	−	0.00					−	0.00	-1697.003	2994.038	0.466
Attribute 8	+	0.00	−	0.00			−	0.00	+	0.00			+	0.00	−	0.00	+	0.00	−	0.00			-2527.748	2530.306	0.331
Attribute 9					+	0.00			+	0.00			+	0.02							−	0.00	-2295.783	2994.237	0.394
Attribute P	+	0.01	−	0.00	+	0.00	−	0.00	+	0.00			+	0.00	−	0.00					−	0.00	-3303.338	1378.380	0.171
Global			−	0.00	+	0.00	−	0.00	+	0.00	+	0.00	+	0.00	−	0.00	+	0.00			−	0.00	-17,259.746	21,375.896	0.382

Table 5.24: Model of attention to attributes without non-significant factors for Wind Power data without Panel Specification

			$lead_{local}$				Entropy						Standard deviation				Skewness						
	ASC		own		other		own		other		Pivotal		own		other		own		other				
	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Final logL	L ratio test	$\overline{\rho}^2$
(Size of farms)	+	0.00			−	0.02			−	0.02					−	0.04					-1203.428	2492.308	0.507
(Maximum height)	+	0.00			−	0.00			−	0.00					−	0.00					-1426.223	2046.719	0.416
Red kite	+	0.00			−	0.01			−	0.01					−	0.01					-1364.740	2169.685	0.441
Minimum distance	+	0.00																			-1208.316	2482.533	0.506
Price					−	0.02	+	0.00							+	0.00					-1213.982	2471.200	0.503
Global					−	0.00	+	0.00	−	0.00			+	0.00							-6445.981	11,603.860	0.473

Table 5.25: Model of attention to attributes without non-significant factors for Wind Power data with Panel Specification

			$lead_{local}$				Entropy						Standard deviation				Skewness						
	ASC		own		other		own		other		Pivotal		own		other		own		other				
	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Final logL	L ratio test	$\overline{\rho}^2$
(Size of farms)									+	0.00					+	0.00					-1205.834	2487.496	0.507
(Maximum height)									+	0.00					+	0.00			−	0.03	-1429.973	2039.219	0.415
Red kite							+	0.00													-1368.998	2161.168	0.441
Minimum distance									+	0.00					+	0.00					-1207.284	2484.595	0.506
Price	+	0.00	−	0.04																	-1221.082	2457.001	0.501
Global									+	0.00					+	0.00					-6451.132	11,593.558	0.473

Table 5.26: Model of attention to attributes without non-significant factors for Wind Power data with individual betas

			$lead_{local}$				Entropy						Standard deviation				Skewness						
	ASC		own		other		own		other		Pivotal		own		other		own		other				
	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Final logL	L ratio test	$\overline{\rho}^2$
(Size of farms)									+	0.00			+	0.00	+	0.00					-1188.251	2517.117	0.513
(Maximum height)			+	0.00					+	0.00			−	0.01	+	0.00	−	0.00			-1412.748	2068.122	0.421
Red kite	+	0.00	−	0.02	−	0.01	−	0.00													-1352.582	2188.455	0.446
Minimum distance							+	0.00					+	0.00							-1201.115	2491.389	0.508
Price			−	0.04	+	0.00			+	0.00									−	0.03	-1203.816	2485.987	0.506
Global	+	0.00							+	0.00			+	0.00			−	0.02			-6394.000	11,680.095	0.477

Table 5.27: Model of attention to attributes without non-significant factors for Potatoes data using MNL

			$lead_{local}$				Entropy								Standard deviation				Skewness						
	ASC		own		other		own		other		Order		Pivotal		own		other		own		other				
	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Final logL	L ratio test	$\overline{\rho}^2$
Texture	+	0.00			−	0.00															+	0.00	-445.646	314.784	0.256
Price	+	0.00									+	0.00											-243.465	719.145	0.293
(Color of flesh)									+	0.00													-479.441	247.194	0.203
Production							+	0.00															-281.785	642.505	0.531
Nutrition									+	0.00	+	0.00											-251.231	703.614	0.280
Country									+	0.00													-395.227	415.622	0.343
Global					−	0.00	+	0.00			+	0.02			+	0.00					+	0.00	-2169.894	2896.669	0.399

Table 5.28: Model of attention to attributes without non-significant factors for Potatoes data using Mixed Logit

			$lead_{local}$				Entropy								Standard deviation				Skewness						
	ASC		own		other		own		other		Order		Pivotal		own		other		own		other				
	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Val	p-val	Final logL	L ratio test	$\overline{\rho}^2$
Texture	+	0.00	−	0.01			−	0.00			+	0.05							+	0.02			-425.498	355.081	0.286
Price			+	0.00					+	0.00	+	0.00											-239.480	727.116	0.598
(Color of flesh)	+	0.00			−	0.04									+	0.00			+	0.04			-465.090	275.896	0.222
Production					+	0.01			+	0.00					+	0.00					−	0.00	-260.393	685.291	0.562
Nutrition	+	0.00			−	0.01	−	0.01			+	0.01									+	0.02	-243.565	718.946	0.588
Country	+	0.00	−	0.00	−	0.00	−	0.00	−	0.01					−	0.02			+	0.03			-369.626	466.825	0.375
Global	+	0.00	−	0.03	−	0.00					+	0.00			+	0.00			+	0.00			-2099.218	3038.021	0.418

alternatives are close and the attribute could make a difference. When this distance is large, it is less likely that attendance to the attribute would change the decision. The alternative specific constant should be positive since by default when the choice maker as no information about the potential effect of an attribute there is a greater likelihood that he is going to attend it. For standard deviation and skewness the sign should be positive for the own-attribute specification and negative for the other-attribute specification since they measure the closeness of the utility function (using a similar reasoning than with $lead_{local}$ about the sign). The sign of entropy for own- and other-attribute utility functions should be conversed compared to other factors since entropy is large when utilities are similar, i.e., close from each other. So the sign of entropy should be negative for the own-attribute specification and positive for the other-attribute specification. Being pivotal should have a positive sign on the probability of attendance, since considering the attribute would change the decision compared to the suboptimal choice made when it is ignored.

Finally, we use two factors that are not available for all choice sets. The order of attributes in the list presented to choice maker in the survey is partially available for Potatoes data. The data of the first day of surveying are not available so we introduced in our attribute attendance probability model only the data collected in the following days. We assume that this factor should have a negative sign since when the order is small, the attribute is at the beginning of the sequence of choices and the choice maker could pay more attention on these attributes. Conversely, when the order is high, the value of the factor is high and the attention is low. About the number of attributes ("# attributes" in the tables), we expect a negative sign since the more attributes the choice maker has to consider, the less he is going to attend each one.

We know describe the results of the attribute attendance probability models and compare them with our expectation.

It is difficult to provide an overall interpretation of all these results since they vary substantially. We can first observe that using individual specific means for the distribution of β_n's from panel mixed choice probability models produces more significant factors in the attribute attendance probability models. In particular with the Cortina data, the use of individual means improves even the goodness of fit of the global attribute attendance probability model, as measured by $\overline{\rho}^2$. However, this is not the case for the other data sets, Wind and Potatoes.

About the different signs, we observe that for the alternative specific constant, it is clearly correct. It is always positive or non significant considering all attributes together ("global" row), with just a few negative signs considering each attribute individually. The entropy of the own-attribute utility function is mostly correct in the global analysis but the results are not so clear and in particular at the attribute level. The results for other-attribute entropy are clearer. The sign is globally positive (or non-significant) in all cases but one. At the attribute level the results are good.

The own-attribute $lead_{local}$ is slightly more often significant than non significant in the global results. However, it has the wrong sign. It suggests that choice makers are less likely to attend an attribute when this one is a clear support for the leading alternative. This analysis is not intuitive and not coherent with the theoretic model.

Another unexpected sign is found on the number of attributes, which is found to be positive implying that choice makers are more likely to attend each attribute when there are many than when there are few. Since these results come from the Cortina data, we have to think about how attendance was reported by choice makers. They had to write the attributes they did not attend. So when the list was long we could imagine they put just a few names and were lazy to declare more attributes. From another viewpoint, when the list was small, the choice of declaring which attributes were ignored was easier. Finally, the last unexpected sign is found for the order of attributes, which is found mostly positive. It means that choice makers would pay more attention to the last attributes in the list than to those at the beginning, in contrast with the left-to-right and the top-to-bottom expected effect.

All other attributes expected to explain attendance are either mostly non significant (own- and other-attribute standard deviation and skewness) or rarely significant, with both negative and positive signs, without any clear tendency.

When observing the predicted results of our different attribute attendance probability models and comparing them with observed attendance to attributes, we see that these factors predict attendance quite poorly. With all different data sets and all different estimates of β's, our attribute attendance probability models always predict attendance. The only exceptions are found in Cortina data attribute 8 (i.e., congestion) and price attribute. For attribute 8 the attribute attendance probability model correctly predicts 81% of observations [2]. But an attribute attendance probability model predicting always attendance would be correct for 79% of observations since 79% of choice makers have attended this attribute.

Concerning the price attribute, it was attended in 59% of the choice processes. The attribute attendance probability model correctly predicts between 65% to 70% of observations[3].

For all other attributes with Cortina data and for all attributes of the other two data sets, the all different attribute attendance probability models always predict attendance. The observed attendance to attributes varies between 76% and 92%, so the model doesn't predict attendance at all.

With the results based on our data sets and different choice probability mod-

---

[2] The basic MNL model and the two scaled ones predict 81% of observations correctly while the mixed logit with individual specific β's predict only 80% of observations correctly.

[3] 65% of observations are correctly predicted using β's from the basic MNL model, 66% for the scaled by category model, 65% for the scaled by category and wave model and 70% for the mixed model with individual β's.

els used to derive the $\hat{\beta}$ necessary to compute the different factors expected to be determining attention, little evidence was found of any general effect on attendance when compared with stated and observed attention to attributes. It seems thus difficult to use them to improve choice models to that correct for degrees of attention to attributes, as suggested by Cameron [6]. Our results tend to confirm the strategy of using observed data about attention to correct the conventional assumption of generalized attention. It also cast some doubts on the possibility of predicting attention to attributes using the estimates of a complete information model and the factors used here.

Conclusion

This Master Thesis was mainly about testing an assumption of Cameron's article [6]: attention to attributes stems from the expected value of the avoided lost utility of a suboptimal choice resulting of ignoring attributes. This assumption brings us to use other- and own-dissimilarities to predict attention to attributes. In general, when estimating a full-information model, we get a parameter for each attribute. Using a mixed logit model, we can get a parameter for each attribute and each respondent. Whatever, then we multiplied this parameter with the value of the attribute to obtain the part of utility due to this attribute. We can use it directly or use complete utility without it to define factors of attendance.

Using different data sets and different models to estimate parameters, no evidence is found of any prediction power for the different factors we used. It suggests that using these dissimilarities to correct for the conventional assumption of complete attention is not very efficient.

Before using any factor to correct for complete attribute attendance, it seems necessary to test it empirically. We know that attention is not likely to be complete, we know that choice-task attendance is probably better than only considering serial attendance. We still need to test if "box attendance" is significantly better than choice-task attendance. It could be done using an information display board and defining two groups, opening in one group all boxes for an attribute and in the other group each single box independently. It would be also interesting to test if – ceteris paribus – information display board and statements of attendance by respondents are equivalent, i.e., if respondents are truly able to report the way they attended to attributes.

These reflections lead me to a broader thinking about interfaces. I discovered that a stated preference survey with a sheet of paper and a pen is sometimes not enough, in particular when you study attention to attributes. If you think about it, choosing between two apples as described by written attributes on a sheet of paper is definitively not a similar experience compared to apple choice in a grocery shop, with all the different aromas of fruits and with different ways of labeling price on the shelves. And we know the potential effects of these attributes and environment characteristics on choice from market research studies. Different authors also emphasize about the role of emotions in behavior, and emotions you feel in front of your sheet of paper is probably close to zero.

As soon as we try to go beyond the outcome of models and to describe cognition processes, we need more extensive data. Information display board is a promising evolution, but we can even imagine to use eye-tracking or even virtual reality, like in Bateman [2]. Serious gaming is probably a very interesting opportunity. It uses the concepts of video games and allows to control the environment and test only what the analyst needs to test, with different scenarios. At the same time, it gives to the agent a broad experience and makes it easier for him to imagine the different alternatives.

I began this thesis with Simon and Popper and their tragic vision of humankind. I will finish it with a remark about these ideas. In my opinion, Simon's bounded rationality is not only a toolbox to sub-optimize decisions but it was also a criticism of neoclassical economic theory and a call to empirically test our knowledge, to challenge it with observed and real world, to "falsify" it as Popper said.

Bibliography

[1] J. Antonakis and O. Dalgas. Predicting Elections: Child's Play! *Science*, 323(5918):1183, February 2009.

[2] I. J. Bateman, B. H. Day, A. P. Jones, and S. Jude. Reducing gain-loss asymmetry: A virtual reality choice experiment valuing land use change. *Journal of Environmental Economics and Management*, 58(1):106–118, 2009.

[3] M. Bierlaire. An introduction to BIOGEME Version 1.6.

[4] M. Bierlaire. BIOGEME: A free package for the estimation of discrete choice models. 2003.

[5] M. Bierlaire. Modélisation mathématique du comportement. Unpublished course, EPFL, 2007.

[6] T.A. Cameron and J. R. DeShazo. Differential Attention to Attributes in Utility-theoretic Choice Models. *Under Review*, 2008.

[7] D. Campbell, W. G. Hutchinson, and R. Scarpa. Incorporating discontinuous preferences into the analysis of discrete choice experiments. Technical report, University of Waikato, Hamilton, New Zealand, Department of Economics, September 2007.

[8] F. Carlsson, M. Pataria, and E. Lampi. Ignoring Attributes in Choice Experiments. Technical report, Göteborg University, 2008.

[9] J. Conslisk. Why Bounded Rationality? *Journal of Economic Literature*, XXXIV:669–700, June 1996.

[10] X. Gabaix and D. Laibson. A Boundedly Rational Decision Algorithm. *American Economic Review*, 90(2):433–438, 2000.

[11] X. Gabaix, D. Laibson, G. Moloche, and S. Weinberg. Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model. *American Economic Review*, 96(4):1043–1068, September 2006.

[12] T. J. Gilbride. A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules. *Marketing Science*, 23(3):391–406, Summer 2004.

[13] D. A. Hensher. How do respondents process stated choice experiments? attribute consideration under varying information load. *Journal of Applied Econometrics*, 21(6):861–878, 2006.

[14] D. A. Hensher, J. Rose, and W. Greene. The implications on willingness to pay of respondents ignoring specific attributes. *Transportation*, 32(3):203–222, May 2005.

[15] D. Kahneman. Maps of bounded rationality: psychology for behavioral economics. *The American Economic Review*, 93(5):1449–1475, 2003.

[16] W. Kaye-Blake, W. Abell, and E. Zellman. Respondents' ignoring of attribute information in a choice modelling survey. *Australian Journal of Agricultural and Resource Economics*, forthcoming.

[17] M. Klaes and E.-M. Sent. A Conceptual History of the Emergence of Bounded Rationality. *History of Political Economy*, 37:27–59, 2005.

[18] C. Lawrence, J. Zhou, and A. Tits. User's Guide for CFSQP Version 2.5: A C Code for Solving (large scale) Constrained Nonlinear (minimax) Optimization Problems, Generating Iterates Satisfying all Inequality Constraints. Technical Report TR-94-16r1, Institute for Systems Research, University of Maryland, College Park, 1997.

[19] J. Meyerhoff and U. Liebe. Discontinuous Preferences in Choice Experiments: Evidence at the Choice Task Level. *To be published*, 2009.

[20] S. M. Puckett and D. A. Hensher. The role of attribute processing strategies in estimating the preferences of road freight stakeholders. *Transportation Research Part E: Logistics and Transportation Review*, 44(3):379 – 395, 2008.

[21] S. M. Puckett and D. A. Hensher. Revealing the extent of process heterogeneity in choice analysis: An empirical assessment. *Transportation Research Part A: Policy and Practice*, 43(2):117 – 126, 2009.

[22] D. Rigby and M. Burton. Modeling Disinterest and Dislike: A Bounded Bayesian Mixed Logit Model of the UK Market for GM Food. *Environmental & Resource Economics*, 33:485–509, 2006.

[23] A. Rubinstein. *Modeling Bounded Rationality*. MIT Press, 1998.

[24] R. Scarpa, M. Thiene, and D.A. Henscher. Monitoring Choice Task Attribute Attendance in Non-Market Valuation of Multiple Park Management Services : Does it matter ? 2009.

[25] H.A. Simon. A behavioural model of Rational Choice. *The Quarterly Journal of Economics*, 69(1):99–118, February 1955.

[26] H.A. Simon. *Models of a Man.* J. Wiley, New York, 1957.

[27] H.A. Simon. *Models of bounded rationality : Empricially Grounded Economic Reason*, volume 3. Cambridge, Mass. : MIT Press, 1997.

[28] J. Swait and W. Adamowicz. The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching. *Journal of Consumer Research*, 28:135–148, June 2001.

[29] N.N. Taleb. *Fooled by Randomness.* Penguin Books, 2e edition, 2007.

Appendix

Example of computation of measures

Let's consider a choice-task in the context of Cortina data. The respondent had to choose between three alternatives and we observed he choosed the first one. We also have the information about the different levels of each attribute. Considering only the first attribute in the data set about Cortina, we know that for this first attribute, the level was 2 for alternative A, 1 for alternative B and 0 for alternative C, i.e., statu quo.

We estimated a model and so we have the corresponding estimates. With the levels and the estimates, we can compute $x'_k\beta_k$ for each attribute k. With all these $x'_k\beta_k$, we can sum them to have the complete utility function and we can remove the first one, $x'_1\beta_1$, in order to have the "other-attribute utility", $x'_{-1}\beta_{-1} = \sum_{k=2}^{K} x'_k\beta_k$, with K the number of attributes. We can do this development for each alternative. In our particular example, showing the information only for the first attribute without lost of generality, we have:

A	$(x_1\beta_1)^A$	$(x_1\beta_1)^B$	$(x_1\beta_1)^C$	$(x'_{-1}\beta_{-1})^A$	$(x'_{-1}\beta_{-1})^B$	$(x'_{-1}\beta_{-1})^C$
0	0.04	0.38	0	-3.37	-4.52	-3.11

where A represent attendance to the first attribute, and 1 meaning the respondent said he attended this attribute.

Leadership

In order to compute own- and other-attribute utility *lead* we first need to have the complete utility of each alternative to know what are the two globally leading alternatives. In our example, with Cortina data, we have 9 attributes. Thus we need to sum for alternative A all $(x\beta_i)^A$ for $i = 1, ...9$ and doing similarly for alternative B and C we get the three complete utilities:

U^A	U^B	U^C
-3.33	-4.15	-3.11

We see that alternative C is the leading one and alternative A is the second leading one. Thus $lead(x_1\beta_1) = (x_1\beta_1)^C - (x_1\beta_1)^A = 0 - 0.04 = -0.04$ and $lead(x'_{-1}\beta_{-1}) = (x'_{-1}\beta_{-1})^C - (x'_{-1}\beta_{-1})^A = -3.11 - (-3.37) = 0.26$.

$|lead|$ is of course the absolute values of these results.

The local version consider only the own- and other-attribute utility functions to define which alternatives are the two leading ones. For the own-attribute utility function, considering only attribute 1, we have utility functions $(x_1\beta_1)^A$, $(x_1\beta_1)^B$ and $(x_1\beta_1)^C$. The two leading alternatives are clearly B and A with 0.38 and 0.04 respectively. Thus $lead_{local}(x_1\beta_1) = (x_1\beta_1)^B - (x_1\beta_1)^A = 0.38 - 0.04 = 0.33$. In the case of other-attribute utility, the utility functions without the first attribute but with all other ones are $(x'_{-1}\beta_{-1})^A$, $(x'_{-1}\beta_{-1})^B$ and $(x'_{-1}\beta_{-1})^C$ and the two leading alternatives are C and then A. Thus $lead_{local}(x'_{-1}\beta_{-1}) = (x'_{-1}\beta_{-1})^C - (x'_{-1}\beta_{-1})^A = -3.11 - (-3.37) = 0.26$. Note that in this case $lead_{local}(x'_{-1}\beta_{-1})$ and $lead(x'_{-1}\beta_{-1})$ are similar.

Standard deviation

Standard deviations used in the models in this report is defined with the own- and other-attribute utility functions and we have:

$$\begin{aligned} Sd(x_1\beta_1) &= \sigma\big((x_1\beta_1)^A, (x_1\beta_1)^B, (x_1\beta_1)^C\big) \\ &= \sigma(0.04, 0.38, 0) = 0.21 \end{aligned}$$

$$\begin{aligned} Sd(x'_{-1}\beta_{-1}) &= \sigma\big((x'_{-1}\beta_{-1})^A, (x'_{-1}\beta_{-1})^B, (x'_{-1}\beta_{-1})^C\big) \\ &= \sigma(-3.37, -4.52, -3.11) = 0.75 \end{aligned}$$

Skewness

The skewness variables are simply the skewnesses of in one hand the skewness of $x'_k\beta_k$ and in the other hand of $x'_{-k}\beta_{-k}$. In our example, considering attribute 1, the skewness variables of attribute 1 are:

$$\begin{aligned} Skew_1 &= skewness\big((x'_1\beta_1)^A, (x'_1\beta_1)^B, (x'_1\beta_1)^C\big) \\ &= skewness(0.04, 0.38, 0) = 1.65 \end{aligned}$$

$$\begin{aligned} Skew_{-1} &= skewness\big((x'_{-1}\beta_{-1})^A, (x'_{-1}\beta_{-1})^B, (x'_{-1}\beta_{-1})^C\big) \\ &= skewness(-3.37, -4.52, -3.11) = -1.5 \end{aligned}$$

Entropy measure

Before computing any entropy measure we need the probability estimated by the model for a choice maker to choose the considered alternative. As explained in section 4.2 we get for own-attribute utility the following probabilities using attribute 1 as an example:

$$\pi_1(A) = \frac{e^{(x_1\beta_1)^A)}}{\sum_{j' \in \{A,B,C\}} e^{(x_1\beta_1)^{j'}}} = 0.3$$

And $\pi_1(B) = 0.42$ and $\pi_1(C) = 0.29$ (whose sum should be 1 but it is not the case because of rounding). And identically for other-attribute utility:

$$\pi_{-1}(A) = \frac{e^{(x'_{-1}\beta_{-1})^A)}}{\sum_{j' \in \{A,B,C\}} e^{(x'_{-1}\beta_{-1})^{j'}}} = 0.38$$

And $\pi_{-1}(B) = 0.12$ and $\pi_{-1}(C) = 0.5$.

And finally we get the own- and other-attribute entropy:

$$\begin{aligned} H_1 &= -\sum_{j \in \{A,B,C\}} \pi_1(j) \log \pi_1(j) = 1.08 \\ H_{-1} &= -\sum_{j \in \{A,B,C\}} \pi_{-1}(j) \log \pi_{-1}(j) = 0.97 \end{aligned}$$

Pivotality

The pivotal factor is 1 if the attribute is pivotal and 0 otherwise. By definition the attribute is pivotal if when we consider it, it changes the choice predicted by the model. Thus we need to consider the utility of the complete model and the other-attribute utility, ignoring one particular attribute. We have seen just before that using other-attribute utility, the leading alternative is C ($\max\left((x'_{-1}\beta_{-1})^A, (x'_{-1}\beta_{-1})^B, (x'_{-1}\beta_{-1})^C\right) = (x'_{-1}\beta_{-1})^C$) and using the complete utility we have the same result ($\max(U^A, U^B, U^C) = U^C$). So in this case the first attribute doesn't change the choice in the decision process as we model it. The pivotal factor is 0.

Number of attributes

In this case there are 10 attributes available to the choice maker. It is here possible, with Cortina data, to estimate this factor since some other choice tasks had 8, 6 or 4 attributes.

Order of attributes

In the data about Cortina the order of the attributes is always the same. Thus it is impossible to estimate its effect. In data about Potatoes we just add the number representing the order of the attribute.