Contribution to Web-Based Conjoint Analysis for Market Research

by

Julio M. Faura

Ingeniero de Telecomunicación
Escuela Técnica Superior de Ingenieros de Telecomunicación
Universidad Politécnica de Madrid, Spain, 1993

Submitted to the Alfred P. Sloan School of Management in Partial Fulfillment of the Requirements of the Degree of

Master of Science in Management of Technology

at the

Massachusetts Institute of Technology
June 2000

© 2000 Julio M. Faura
All rights reserved
The author hereby grants to MIT permission to reproduce and to distribute publicly copies of this thesis document in whole or in part.

Signature of Author

MIT Alfred P. Sloan School of Management
May 19, 2000

Certified by

John R. Hauser
Thesis Supervisor

Accepted by

David A. Weber
Director, Management of Technology Program
Contribution to Web-based Conjoint Analysis for Market Research

by

Julio M. Faura

Submitted to the Alfred P. Sloan School of Management in Partial Fulfillment of the Requirements of the Degree of Master of Science in Management of Technology at the Massachusetts Institute of Technology

ABSTRACT

The topic of this thesis is focused around the new possibilities in the field of market research opened with the advent of the internet, in particular around the use of web interfaces to perform conjoint analysis, a market research technique based on comparisons between pairs of potential product configurations to compute the perceived relative utility of each of several product design attributes. We first overview conjoint analysis and the online market research industry in general, so see how the use of the internet for this purpose can reduce the cost of these analyses in more than an order of magnitude due to a better accessibility to test customers, a cleaner and faster interface, and the possibility of reducing the number of questions necessary to compute the utility functions using an adaptive technique that generates optimal questions with dynamic web content as the test proceeds. We then study the issues related with the automation of web sites for performing this type of analyses without having to redesign the interface, with the introduction of dynamic content web technology for adaptive conjoint analysis, and with the possibilities offered by this technique towards fast segmentation of incoming customers. Finally, we report the first implementation - to the best of our knowledge - of an actual web architecture that uses a novel adaptive conjoint algorithm and automates the whole analysis setup process.

Thesis Supervisor: John R. Hauser

Title: Kirin Professor of Marketing
Table of contents

1.- Introduction and Statement of the Objectives

1.1 Introduction
1.2 Statement of the objectives
1.3 Organization of this thesis

2.- Conjoint Analysis and Online Market Research: an Overview

2.1 Introduction

2.2 Introduction to conjoint analysis
   2.2.1 Card sorting
   2.2.2 Constant-Sum Pair Comparisons (CSPC)
   2.2.3 Choice-Based Conjoint Analysis
   2.2.4 Self-explicated and hybrid types

2.3 Adaptive conjoint analysis
   2.3.1 Sawtooth's approach
   2.3.2 MIT's approach

2.4 Online market research
   2.4.1 Market research techniques used over the web
   2.4.2 Benefits of online market research
   2.4.3 Challenges of online market research
   2.4.4 Web-based adaptive conjoint analysis

2.5 Conclusions

3.- Web-based Conjoint Analysis: Feasibility Study

3.1 Introduction

3.2 A generalized architecture for web-based conjoint analysis
   3.2.1 Administrator access
   3.2.2 Respondent access
   3.2.3 Dynamic HTML templates
   3.2.4 Computational engine
   3.2.5 Database
   3.2.6 CGI scripts

3.3 Technologies
3.4 Conclusions

4.- Implementation Case Study

4.1 Introduction

4.2 Description
   4.2.1 Internal architecture
   4.2.2 Administrator access for analysis setup
4.2.3 Respondent access
4.2.4 Administrator access for analysis monitoring

4.3 Conclusions

5.- Conclusions and Future Work

5.1 Main conclusions

5.2 Future research lines

References
Acknowledgments

First of all I would like to thank John Hauser, my thesis advisor, for the opportunity of working on this thesis under his guidance. It has been a pleasure and an honor to work with him and the rest of the team at CIPD, with special mention to Olivier Toubia, Duncan Simester and Ely Dehan.

Thanks of course to La Caixa and to Jose Maria Insenser and the rest of the team at SIDSA in Spain, without whose financial support my stay at MIT would have been impossible.

Thanks to my friends Ken and Noor and all the rest of the MOT gang for so much support, advice and fun during the past twelve months. Also kudos to Dave Weber and Jennifer Mapes for making this year at MIT run so smoothly.

And obviously this thesis is dedicated to Natalia, my wife. Without you, all this would just make no sense.

Julio Faura
May 19, 2000
Chapter 1: Introduction and Statement of the Objectives

1.1. Introduction

The advent of the internet age and the electronic commerce is changing the way consumers and businesses interact with each other. Every day more products seem to be sold well over the net, and more parts of the business processes are increasingly migrating to be partly or totally carried out through the web.

We believe that the same can happen to market research because of several reasons: first, the internet is a formidable way to reach customers in a very cheap, efficient and geographically distributed way. Therefore, it can realize significant time and money savings while acquiring respondents, one of the major problems and more important costs for market researchers; second, the web is a very good interface between humans and computers, and therefore market research can be performed over the web in a much more interactive manner assisted by actual computers running calculations in real time. This is a huge advantage compared to paper-based analyses with a single off-line post-analysis data interpretation, and even to computer-based analyses that needed the respondent to physically move to an especial location where the computer was; and third, the web dramatically accelerates traditionally cumbersome tasks such as analysis design and setup, or documentation. Web automation can provide a very easy way to set up an analysis just by outlining what the considered parameters will be. This, linked to the easiness to reach a variety of respondents, entail two main benefits: the overall experiment cycle time can be dramatically reduced, and experimentation for analysis setup refinement is suddenly possible,
compared to the large and cumbersome studies necessary in the past to determine what would be the perfect analysis for each situation.

In this thesis we will overview the online market research industry as we find it today, and we will try to envision further possibilities for the incoming years. We will focus though in a particular technique that has become increasingly popular in the marketing and management literature in the last decades: conjoint analysis. We believe that this discipline is especially suitable for its use over the web since it is purely quantitative, and therefore it does not really require a strong human interaction with the respondent such as in the case of focus groups.

The possibility of automating the whole process of market research, from experiment design and setup to respondent selection to data analysis to documentation, is highly interesting. An integrated approach to the whole process could be provided under the form of an internet ASP (Application Service Provider), which would make serious market research available to a wider range of public, would dramatically decrease its cost, and would make the whole cycle many times faster, alleviating the time-to-market constraints faced by most corporations these days.

One of the objectives of the thesis is to prove this automation concept. We will specifically work on the automation of conjoint analyses, but we believe that the concept itself can be easy generalized to other market research techniques.

1.2. Statement of the objectives

We may state the main objectives of this thesis as follows:

- To study the feasibility of an integrated automated approach to web-based market research, using conjoint analysis as a particular case. We will synthesize a generalized
architecture for a web-based market research server based on conjoint analysis suitable to be used in a variety of problems. The methodology can be generalized in a straightforward manner to support other fundamental market research techniques.

- To prove this concept by physically implementing a particular realization of one of these architectures. The selected case uses a novel variation of adaptive conjoint analysis being developed at MIT for web automation.

- To provide a background introduction to conjoint analysis in general and to online market research techniques currently used in the electronic marketplace. We will try to assess what are the benefits and challenges posed by the advent of the internet when it comes to market research as opposed to the traditional methodologies.

1.3. Organization of this thesis

In this first chapter we have stated what we believe will be the main contributions of this thesis. The second chapter aims to be an introduction to online market research and the techniques available in this industry as of today. We will pay special attention to conjoint analysis, especially in its adaptive variations since as we will argue the technique is especially suited to online use.

The third chapter addresses the feasibility of the general idea of an integrated web-based analysis server. Although focused on conjoint analysis, the methodology outlined there should be easily extendible to other market research techniques. Throughout the chapter we will describe the general architecture of such as server and we will study each of the building blocks. We will also mention some guidelines about the required web technologies to use.
Chapter fourth reports a first particular implementation of the generalized architecture outlined in the previous chapter. Our implementation was preliminary tested with successful results, which can be considered as a proof of the concept developed up to that point. Further tests with actual respondents will promptly follow.

The fifth chapter summarizes the contributions of the thesis and mentions some of the interesting research lines to be addressed in the near future.
Chapter 2: Conjoint Analysis and Online Market Research: an Overview

2.1. Introduction

This chapter aims to serve both as an introduction to conjoint analysis in general and as an overview of the market research industry in the electronic marketplace. As we stated in the introductory chapter, our aim is to explore the feasibility and adequacy of conjoint analysis for automating web-based market research. We will first introduce the main concepts of conjoint analysis, a relatively well known and traditional technique that has been used for decades, along with some of the variations routinely used in the industry these days. We will then outline the principles of adaptive conjoint, a particular variety of conjoint that aims to adapt to the particular respondent as the survey progresses. An overview of the online market research industry will follow, although drawing such a picture is challenging due to the fast pace at which this incipient discipline is moving as the internet technology progresses and becomes more pervasive. Finally, we will find some reasons that explain why conjoint analysis techniques are especially suitable for the online medium. We believe that web-based conjoint analysis can play a decisive role in product and concept development in the internet age.

2.2. Introduction to conjoint analysis

Conjoint analysis (CA) has been an increasingly important and popular technique for assessing consumer preferences in market research (Urban et al 1996, Cattin et al 1982, Wittink
et al 1989, Wittink et al 1993). The technique is especially well suited for product development, since it allows the researcher to find out how much potential customers value each of the possible features to be included in a final product. It also lets them find sensible market segmentations for which particular product configuration tradeoffs can be made.

To understand the basics of conjoint analysis, we may consider a particular product with a number of attributes. For example, if we consider cars, possible attributes would be horsepower, color, number of doors, or price. Each attribute would have a number of possible levels, which are normally considered in discrete steps rather than in a continuous fashion for analysis convenience.

Given a particular structure of the problem, that is, the number of attributes and the numbers of levels per attribute, we may generate a full factorial design, that is, a set of product configurations with all the possible combinations of different levels on each of the attributes. Obviously the complexity of the factorial design grows rapidly as we consider more attributes and levels, so in principle it is important to pick the right attributes and levels in order to keep the experiment design practical.

Once the factorial design is known, the respondent would be asked to rate her preference for each of the product configurations in a given relative scale. Each product configuration in the factorial design would be designated by a set of indices, one per attribute, with each index set to an ordinal number designating the particular level for the corresponding attribute. The basic idea of conjoint analysis consists on running a multivariate regression on these indices using the consumer preferences for each product as the independent variable. The result is a set of estimates of the relative utilities or part worths for each of the attributes at each level, which is a measure of how much do consumer value each of the attributes for a given product. Finding
ratios of each of the attributes to price (which is another attribute), one can find actual dollar utilities for each of the attributes of the product. Therefore tradeoffs in the levels of the attributes can be found for each given price point.

A number of concerns arise from this basic idea. First, if the respondent is required to rank her preference for all the product configurations of the full factorial design, then the experiment is only practical for relatively simple cases with few attributes and levels. For example, if we consider ten attributes with two levels each, the number of possible cases in the full factorial design would be $2^{10}=1024$, which is humanly impossible to handle. A first simplification would use an orthogonal design rather than a full factorial, and would include only linearly independent combinations rather than all of them.

The complexity of orthogonal designs is smaller than full factorials but typically in the same order of magnitude. To further make the experiment manageable for respondents, another typical approach in these cases is to assume that respondents are all alike, and therefore the rating can be done collectively by distributing the products of the factorial design (or just an orthogonal one) among several respondents. The problem of considering homogeneous respondent distributions is that we would lose the possibility of estimating individual partial utility vectors, which would be very useful afterwards for identifying market segments using vector clustering techniques.

A second concern with this approach is the inherent independence of the attributes considered in the regression, while this may not be the case. For example, consumers may prefer cars with two doors if the horsepower is high but four doors if horsepower is medium, reflecting two market segments centered in sporty cars vs familiar ones. Such dependence between
attributes are called *interactions*, and conjoint experiments typically deal with them by using more dummy attributes used in fact as extra variables in the multivariate regression.

At a more practical level, there are several varieties of conjoint analysis typically used in market research experiments. We summarize them in the paragraphs below.

### 2.2.1. Card sorting

The traditional way of conducting conjoint analysis experiments involves the use of cards, each one describing a particular product configuration (that is, a given combination of levels on the considered attributes). The respondent is then asked to sort the cards in order of preference. The preference level is assessed as the sorting number given for each card, and is then used as the independent variable for the multivariate regression.

Card sorting has been used for decades in product development and related activities. As we mentioned before the number of attributes and levels to be considered are limited by what a human respondent can perceive. The fact that the respondents do not see the actual product prototype but a card describing it (either textually or graphically) also limits the number of attributes to be used, although in some cases actual prototypes are built and used for this kind of analysis. In general terms, making tradeoffs with more than six unrelated attributes are challenging for typical respondents.

### 2.2.2. Constant-Sum Pair Comparisons (CSPC)

Partly to overcome the problems posed by traditional card sorting methods, a variety of conjoint analysis called *constant-sum pair comparisons* (CSPC) has been proposed and has been widely used in the last years. The basic idea involves asking the respondent to rate her preference
between pairs of products from the factorial design rather than sorting them all. The preferences are collected in a similar scale to be able to be compared (a constant number of points is allocated among the two products, hence the constant sum of points, that is, relative product preferences). A multivariate regression using these preferences as the independent variable can then be used to estimate the part worths of each attribute at each level.

CSPC makes the task of answering questions more manageable for respondents, since respondent fatigue has been always a concern among market researchers. Even if a great number of attributes and levels were to be considered, each comparison pair can be limited to vary only a number of attributes (four is customary), which is easier to grasp for typical respondents. Aggregate designs can also be used here to split the experiment between a number of respondents in order to reduce the number of questions per respondent. The same concerns about losing the possibility of distinguishing between individual respondents apply if that is considered.

2.2.3. **Choice-Based Conjoint Analysis**

In choice-based conjoint analysis the respondent is asked to choose from a list of products rather than to rate her preference on a given scale (Huber et al 1992, Oliphant et al 1992, Sawtooth 1995). Several of these choice questions are given to each respondent, and each question and the overall design is determined using orthogonality considerations. As the number of respondents grow, A quantitative choice model, such as the logit transform, is used to estimate the part worths.

Choice-based conjoint analysis has advantages and disadvantages compared to traditional conjoint and CSPC. The main strength relies on the fact that the respondent is asked to make a
choice much in the same way that it happens in the actual marketplace, where consumers do not explicitly rank possible choices but end up making a single decision. Another advantage is that it is possible to include product specific information, that is, characteristics that apply to only one or several of the product pairs of the factorial design.

The main disadvantage of choice-based conjoint analysis is that the respondent has to process a lot of information to make the actual decision, which makes the technical impractical for large number of attributes and also makes it inadequate for estimating utilities of the individual respondents. This type of analysis is only valid in the aggregate level and therefore limits the ability of the researcher to identify market segments with cluster analysis, for example.

2.2.4. Self-explicated and hybrid types

In self-explicated analysis, the researcher directly asks the respondent to rate attributes and levels in a direct manner, directly obtaining the part worths rather than estimating them indirectly through regression upon observations of product comparisons. Although it seems counter-intuitive, studies show that these methods are quite robust in practice.

Finally, hybrid analysis types (Green 1984) try to mix self-explicated analysis with more traditional conjoint techniques. Sometimes the partial utilities are estimated weighing the data obtained with each of these two methods. For other variations a self-explicated part of the analysis is initially used to find out what are the most significant attributes in an attempt to reduce complexity towards a more elaborate (traditional) conjoint analysis in cases with a large number of attributes.
2.3. Adaptive conjoint analysis

Traditionally, conjoint analysis, much as the rest of the market research techniques widely used in the industry, has been performed in a paper and pencil basis, where respondents had to sort cards or respond to questionnaires by filling holes in forms. However, as computers have become more pervasive in the industry, a new range of computer-aided market research techniques have become available. One of such techniques is Adaptive Conjoint Analysis (Tumbusch 1999, Green et al 1991).

In adaptive conjoint, each CSPAN question is created on the fly or selected from the full factorial design depending on what the previous questions and answers were. Therefore the actual questions asked to each respondent is not deterministic and greatly varies on her answers as the survey evolves. The beauty of this approach is that each question can be designed to maximize the amount of information to gather. Therefore, the overall benefit hinges upon the fact that the total number of questions can be reduced. Another potential benefit is that each answer may be audited to check its consistency according to the previous questions and answers. Therefore all sorts of "checks" can be done on the fly to evaluate the quality (that is, consistency) of the respondent, being able to assess her general seriousness and her relevancy to the current study.

Due to the reduced number of questions needed to extract the same amount of information than in the "fixed" case, adaptive conjoint is believed to be the analysis of choice for problems with a large number of attributes, where traditional approaches with fixed designs result clearly impractical. We will cite two main initiatives of adaptive conjoint analysis developed or being developed today:
2.3.1. *Sawtooth's approach*

Sawtooth Software Inc. developed a first approach to adaptive conjoint analysis using in fact a hybrid technique especially geared to deal with problems with a relatively large number of attributes (Sawtooth 1996, Johnson and Olberts 1991, McLauchlan 1991). The methodology is divided in four parts: first, the respondent enters her preference order for the levels of each attribute, eliminating those levels considered unacceptable; next, the respondent is directly asked to rate the importance of the attributes themselves. Then, an adaptive conjoint design in generated considering only those attributes ranked as significant by the respondent. Finally, some specially significant product configurations are built and presented one at a time and an actual "likelihood to buy" is asked to the respondent, as a sort of a final check.

The strength of this approach relies on its capability for handling a large number of attributes and levels, which may be cited by most researchers as a major challenge in today's market analyses. The main problems of the technique are that a) it relies on direct assessments from the respondents of what the most significant attributes and levels are (that is, it relies on self-directed analysis), b) it needs a large number of initial questions for setting up the actual final CSPC conjoint design, which can be cumbersome and certainly increases the length of the interview (although the number of questions keeps within practical limits as compared to fixed designs for the same number of attributes), and c) it is not useful to deal with interactions since it assumes that all the attributes have independent effects on the general product preference.

2.3.2. *MIT's approach*

An enhanced approach to conjoint analysis is being leaded by O. Toubia, D. Simester and J. Hauser at MIT (Toubia 2000) and a preliminary version has been implemented for its use in an
automated web-based market research server as described in this thesis. The methodology used in our approach is different, since it relies on multidimensional algebra as opposed to the statistical methodology used by Sawtooth. In our case, the utility gains from level to level for each attribute is represented by a dimension in a multidimensional hyper-space, and an estimate of an utility vector for a particular respondent would be represented as a vector (that is, a point) within that hyper-space. With an appropriate representation, each question represents a hyper-plane direction, and the actual answer would fix the constant term of the hyper-plane (that is, it would fix which of all the infinite parallel hyper-planes is the one that actually contains the utility vector). Therefore, each time a question is asked, one dimension is lost in the cumulative hyper-plane until it eventually becomes a single point or the center of a feasibility region (a multidimensional hyper-polyhedron) is estimated, which then becomes the utilities estimate.

At the moment of writing this thesis, preliminary tests have been conducted that show the superiority of this approach for a smaller number of questions compared to both Sawtooth's methodology and the fixed cases. A full set of characterization results will soon be reported.

2.4. **Online market research**

With the advent of the internet, a full new range of possibilities are unleashed in the market research industry. As the internet itself is still evolving at a very fast pace, online market research can still be considered as an incipient discipline. To this moment most of the market research initiatives carried over the web are mainly adaptations of traditional methodologies, and we are yet to see the birth of novel techniques that fully take advantage of the particular characteristics of the new medium in a fundamental manner.
In this section we will outline what we believe are the pros and cons of online market research, and we will describe what we think may be one of these new techniques that could revolutionize the online market research arena: web-based adaptive conjoint analysis.

2.4.1. Market research techniques used over the web

According to the Council of American Survey Research Organizations (CASRO 2000), there are basically three methods for conducting online surveys: e-mail surveys, HTML forms, and downloadable interactive survey applications. E-mail surveys are conducted remotely by sending a questionnaire (and this can be as easy as a plain text file) to the potential respondent, and then waiting for the answers to the questions to come back. Even fixed conjoint analysis could be performed in this way, but although it is probably the cheapest and fastest way to create a survey, it does not allow much interaction between the respondent and the researcher (or at least with the computational engine used to interact with the respondent).

HTML forms are the usual way to collect information in web sites. They allow respondents to specify choices through pull-down menus, radio buttons and check buttons. They also let the respondent enter text in special fields or text areas. The data collected from the forms are normally submitted to CGIs, which are programs that are capable of processing this input information and generating dynamic HTML content for the next screen, thus providing interaction to some extent.

Finally, respondents may be asked to download a program as a separate application through FTP or as a plug-in to their navigator. In this case the program is fully interactive with the respondent and can perform advanced mathematical calculations as the survey goes on.
Opinion polls, feedback forms and other simple questionnaires are quite pervasive in today's online marketplace. Some specialized firms have developed chats (interactive pages that echo to the screen what all the users connected to them type on a form) to be used for conducting focus groups over the web.

Finally, conjoint analysis is just starting to be used to conduct product development and tradeoff analysis over the web (McArdle 2000, Dahan and Srinivasan 2000). Early experiments in this area have used fixed designs. Our objective is to use adaptive conjoint analysis to leverage the interaction possibilities between the analysis engine and the actual respondent.

2.4.2. Benefits of online market research

Among the benefits of online market research we may count the following:

- It can be significantly cheaper compared to traditional methods. Paper-based techniques incur significant respondent compensation costs, and phone-based surveys require a great deal of respondent filtering (that is, a large number of calls) to get an actual meaningful sample. The cost of an online respondent can be more than an order of magnitude smaller than for the other types. Also, the cost of setting up the analysis is typically smaller in the electronic medium.

- It is fast. As electronic commerce becomes ubiquitous in our lives, large numbers of online consumers can access the same web site in parallel and a whole survey can be filled out in a matter of hours. This is a significant improvement compared to paper and phone based types of analysis. As a consequence, product development cycles can be reduced in orders of magnitude.
It can interact with computers much more easily. Therefore the questionnaire can become *adaptive* in the sense that it can be redesigned on the fly to extract the maximum or more valuable information for each particular respondent. The traditional benefits of computerized versus paper and pencil methods (Finkbeiner and Platz 1986) also apply here.

It could use integrated multimedia applications to render high quality graphics that may very efficiently communicate the different attributes of a particular product. This type of "virtual reality" type of prototyping can dramatically accelerate the market analysis - product design cycle, since the analysis can be done by rendering data being developed in the CAD system before any actual prototype is constructed (Dahan and Srinivasan 2000). Also, multimedia technologies can automate and dramatically simplify the documentation process.

It provides much more possibilities for experimentation. One of the bottlenecks of traditional analyses such as conjoint is designing and setting up the analysis itself. Since respondents are expensive and the overall survey cycle is slow (months), researchers must devote a lot of attention to select what attributes and levels are going to be tested and how will the information be interpreted afterwards. The continuous availability of respondents and the easiness to change the parameters of the analysis itself on a web implementation (such as the attributes or their levels in a conjoint analysis) lets researchers try different designs in a much more dynamic way.

2.4.3. *Challenges of online market research*

We may mention the following drawbacks of online market research:
Obviously the human interaction with the actual respondents is lost. Purely quantitative techniques such as conjoint analysis are not affected by this constraint, but some other techniques may lose part of the very rich information that can be gathered from actual human interaction. As an example, and although there is a number of commercial firms currently offering versions of online focus groups implementations using chats, all the consumers' gestures, voice tones, faces, and general "feelings" that may be perceived in a "real" focus group would be lost over the web.

Online respondents do not have the social pressure exerted through a personal interview or even through the phone. Therefore they are more prone to be unmotivated, dishonest and give dummy answers. Since they are much more anonymous than "real" respondents, it is more difficult to assess their seriousness and their real profile. This problem can be partly overcome by running internal consistency checks on the responses and by measuring the time spent for answering each question to eliminate these respondents that are paying less attention.

Online surveys may be too "virtual" to actually convey the right message about the products to be evaluated. In an in-person experiment, the researcher can show real prototypes or mock-ups or at least can describe the product with words, photographs, hand language and gestures. Not all of these resources are available in the internet, although multimedia technologies can actually be an advantage in some cases to reconstruct virtual prototypes of potential products in a cost and time-efficient manner.
2.4.4. *Web-based adaptive conjoint analysis*

Finally, we mention the special case of web-based adaptive conjoint analysis since we believe that it is illustrative of how the information technology revolution can impact the current market research practices.

Adaptive conjoint was introduced in a previous section as the natural evolution of conjoint analysis as researchers gained access to computers as market research tools. In the internet, all kinds of adaptive analysis are especially interesting since the CGI technology allows to intermingle computing programs with data gathering from forms and page layout dynamic generation, all in an integrated fashion transparent to the respondent. The main motivation is to maximize the information gathered from each respondent, with the aim to minimize the number of questions as much as possible. In the electronic arena, is not only the traditional concern on respondent fatigue what calls for an smaller number of questions; it is also the difficulty to retain a respondent for the necessary time to complete the survey, since bail-out rates from commercial web sites are believed to be rather high.

Adaptive conjoint analysis is also very suitable for analysis setup automation, which is one of the main themes of this thesis. Fixed designs would require human attention to generate the factorial and pick what orthogonal questions to ask and how to shuffle them between groups of respondents; but adaptive conjoint can be designed to be smart enough to generate all the questions from scratch, greatly simplifying the analysis setup phase.

Finally, a potential use of adaptive conjoint is to arrive to a very fast segmentation of an impinging user by finding what cluster does she belong to (or is closer to). The motivation to achieve a reliable segmentation with a very few questions (for example linking collaborative filtering and adaptive conjoint in order to get a meaningful segmentation by asking apparently
unrelated questions whose correlations to the actual segmentation variables is strong) is huge, since it could dramatically boost the hit rates of the banner advertising in commercial portals.

2.5. Conclusions

In this chapter we have introduced the different techniques used in online market research, mentioning some benefits and drawbacks compared to traditional off-line approaches. Special emphasis has been put on conjoint analysis since the possibility to carry it out over the web is especially promising. In particular we have roughly described a novel approach to adaptive conjoint analysis begin developed at MIT which seems promising as the results from the first practical implementation suggest. Finally, we have discussed the interesting possibilities derived from performing adaptive conjoint analysis over the internet, which could dramatically reduce product design cycles compared to traditional approaches while reducing the costs of designing, setting up, executing and documenting this kind of analyses. The next two chapters deal with the feasibility of a system for implementing web-based adaptive conjoint analysis and the details of a first implementation of such a system, respectively.
Chapter 3: Web-based Conjoint Analysis: a Feasibility Study

3.1. Introduction

In this chapter we explore the feasibility of an integrated web server for Conjoint Analysis. This server would allow market researchers to set up customized analyses in an straightforward manner just by specifying all the parameters of the analysis such as what algorithm to use, the number of questions, number and names of all the attributes and levels, even what graphic layout to use and what formatted help texts to display. Respondents would interact with the server through a different access gate that would only allow them to answer the questions of the survey designed by the researcher. The "administrator" access would also let researchers monitor in real time the evolution of the experiment, showing the number of respondents so far, averages of partial utility vectors, clusters of vectors, deviations to the averages, or even displaying the whole database if so required.

In the reminder of this chapter we address the feasibility of such a web server. We outline the architecture of the server at the block level and we study what technologies would be useful in this context. We start with the particular case of conjoint analysis but the ultimate idea would be to create a "one stop shop" for market research set up as an ASP (Application Service Provider) that would take care of all the market research needs of potential customers at any level, including different techniques (different conjoint types, perceptual maps, clustering), survey design service, respondent provision, data analysis, and final documentation generation.
3.2. A generalized architecture for web-based conjoint analysis

Figure 3.1 shows the overall architecture of the web server

![Web server block diagram](image)

Fig. 3.1 - Web server block diagram

The server is composed of four main blocks (database, CGI, computational engine and HTML templates) which interact with different types of users through two types of access (respondent and administrator). At the heart of the web site there is a set of CGI scripts which are responsible for the coordination of all the different actions: they interpret the templates to show questions and help screens to respondents; they interface with the computational engine to process the answers and obtain new questions; and they interact with the database to save the results and to retrieve and show information through the administrator access for the researchers. In the following paragraphs we will separately describe each of those blocks in more detail.

3.2.1. Administrator access

From the administrator access, the researcher would be able to configure the whole server specifying what is the product, what attributes we will be varying, their types and ranges, what algorithm should be used and with which parameters, etc. For example, if a fixed conjoint is to be used, the researcher would need to specify which exact pairs are to be shown, and how many
questions are to be made (if we want to reduce the number of questions at the risk of blurring the
market segments by considering similar utility vectors). Other less fundamental variables could
also be chosen here, such as the layout type for the respondent pages including colors, fonts,
positions of items, type of data entry device (row of switches vs "speed meter" vs scroll bar,
etc.), and so on. Providing several prefixed formats would be practical at this point, although
custom ones should also be easily specified, for example by entering a background GIF or JPEG
file plus coordinates for the active areas of the different elements.

After entering all the desired options, the administration page would run a CGI script that
would initialize the database for the particular analysis to be run. All the entered parameters will
be saved since they will be used later by the respondent access CGIs to customize the templates
and interact with the engine and database. All these parameters, which utterly define the analysis
itself, should be saved in a text file with a known format to enable other ways of automation, and
also to provide the possibility of manually patching it. An XML DTD (Document Type
Definition) seems quite appropriate in this case, as it will be explained in section 3.3.

The other possible use of the administrator access is to dynamically show the results of
the survey by directly probing the database. A comprehensive tool set could be developed and
plugged in to see the current estimation of dollar utilities, look for clusters of utility vectors to
determine market segments, run correlation analyses with demographics vectors, or even run
correlation analyses with previously estimated dollar utility vectors for cross-segmenting (to
segment an individual by asking apparently unrelated questions). It would be useful to provide a
set of fundamental tools (regressions, correlations, clustering analyses) to be able to let the
researcher define the type of analysis to be done on the data. The data analysis definition would
be saved in an XML file similar to the one used for the analysis setup itself, and could be
rendered in a variety of ways depending on the actual situation. In particular it would be useful to render the analysis results online in real time, refreshing the screen as the data is gathered.

3.2.2. **Respondent access**

The respondent would interact with the site from the respondent access page. After directing her to the appropriate analysis (there may be several analyses running at the same time on the server), a set of introductory pages would be shown to explain the product, the attributes, and the mechanics of the analysis (how to enter answers, etc.) . Then a set of questions would be presented and the answers would be collected.

The particular sequence of introductory pages and the layout of all the pages shown to the respondent are defined in an HTML template that would be chosen or even designed by the administrator (the market researcher) in analysis setup time. The next paragraph explains in more details what the possibilities are for these templates.

3.2.3. **Dynamic HTML templates**

In order to speed up the setup process of an ordinary conjoint analysis, it is useful to create generalized templates that would define the style, graphic layout and general flow of operation (introductory pages, questions and farewell page) of the respondent access. The idea is to provide several of these templates for general purposes, although a clear API (Application Programming Interface) should be created and documented in order to design custom ones. Ideally, a proprietary tool could be created to design custom templates through the administrator access.
The templates provide the information to generate the HTML code that will render the product pairs and will obtain the preferences input from the respondents. Figure 3.2 shows a typical architecture of the template: a frame for general navigation, a frame (or two sub-frames) for product pair presentation and a frame for preference data entry.

Fig. 3.2 - HTML template overview

The frame for typical navigation can be used to access help and to render information about the analysis: reminding what product is being considered, what is the current question number, how many questions are left, or even what is the prize to be won after the answers are given.

The pairs of products being considered can be shown in the two middle frames. Each product would be described using words and graphics, if available. Help about each of the attributes that are different in each pair could also be shown here if requested by the respondent. There would be a number of ways to actually show these product pairs, and even simple tables could be used instead of frames, as in the implementation described in the next chapter.

Finally, a frame for data entry would be placed at the bottom. In principle the user would specify a relative preference for one of the products of the pair. Several typical kinds of entry
devices would be provided here, such as sliding rules, lines of radio buttons, pop-up menus, "db" bars as in a boom box, a needle as in a speed meter, and so on.

3.2.4. *Computational engine*

In order to be able to use the adaptive conjoint techniques introduced in the preceding chapters, it is necessary to be able to perform substantial computational work on each of the answers just collected from each respondent. The specificity of these tasks does not allow (in general) the use of a general purpose scripting language such as Perl, and therefore we do not consider the computational engine as part of the CGI itself. Besides, it would be useful to be able to select what particular algorithm or variation should be used for each particular analysis, and separating it from the general CGI becomes then useful.

At every iteration, the CGI sends to the engine all the questions and answers so far and receives in return the next question to ask, which will be interpreted by the CGI to construct the next question page using the template information.

3.2.5. *Database*

The database is where all the information about each particular respondent is stored, including the questions and answers (in the adaptive algorithms the questions are different for each respondent since they depend on the actual answers), the demographic information if asked (age, postal code, income level, ethnicity, etc.), and any other data that may be useful in the future for off-line analysis, such as time taken to answer each question (which may be useful to determine the seriousness with which the respondent was taking the survey), number of times
she used the online help (reconstructing a hit trajectory within the site may help to debug and improve it) and so forth.

The database needs to be initialized from the administrator access through special CGIs that will save the parameters of each analysis in an appropriate manner so that it knows how to interact with the computational engine when the actual analysis is running.

3.2.6. **CGI scripts**

The heart of the web server is a set of CGI scripts which are responsible for all the interaction between the different constituent blocks mentioned above involved with data collection or analysis. Whenever the user submits data, it is the CGI's responsibility to gather it and pipe it to the appropriate data repository or computational engine. In an adaptive conjoint analysis the data is received and sent to the computational engine for processing. The output of the engine (which is the next question to ask) is read by the CGI to customize the dynamic HTML template with the new product attributes.

If a traditional (not adaptive) conjoint analysis is used, the computational engine is simplified to extracting questions from a list either sequentially or randomly.

3.3. **Technologies**

In this section we will explore what could be the most appropriate technologies to use in the implementation of the architecture described above. We assume that all the system, including the computational engine and the database are going to be hosted by the server, and therefore the client (both for administrator and respondent access) is only going to be used as an interface (monitoring, rendering questions and directions, and obtaining input data). We do not intend to
make a recommendation about what underlying operating system or hardware platform to use,
since one of our concerns is the portability of the server across platforms. Therefore we should
restrict ourselves to popular and highly extended technologies (such as most of Perl's libraries)
available for all sorts of hardware/software platforms. Among these we may count Linux/PC,
Solaris/Sparc and NT/PC as the most popular ones, and our goal would be to design applications
that could run interchangeably in all these three platforms. The case of Linux looks especially
appealing since most of the software technologies mentioned in the remainder of the section and
the Linux operating system itself are not only available but also free and well documented and
maintained. Therefore it is not surprising that Linux has turned into the most popular technology
for web applications nowadays.

Perhaps the least standard part of the whole server is the computational engine. An
obvious solution to implement this part is writing a custom C program using an extensive math-
oriented library. Another possibility would be to use an specialized mathematical computation
package such as Mathlab or Mathematica. The advantage of this last approach hinges on a much
easier development process (these packages do provide extended debugging capabilities) and a
quick time-to-release since they usually provide easy mechanisms for direct compilation from
native code into general C code that can be easily interfaced with the rest of the CGIs. A final
possibility would be setting up a "computation server" under the form of a living process in a
multitasking environment (such as Linux, Solaris or even NT) with which the client CGIs would
communicate through some kind of IPC (Inter Process Communication) mechanism, namely
sockets, shared memory, messages, pipes or even plain files. This last solution would probably
be the fastest and in some respects the most efficient although it would also be very operating
system-specific.
There are also several possibilities towards the implementation of the database part. The easiest solution would be to store all the data in a file or set of files in a proprietary format, especially if that format can be read from the tool to be used to render and interpret the data, typically the same one used for the computational engine. A more elaborate solution would entail the utilization of an standardized database server with a known interface protocol such as OBDC, NDBM, SDBM, GDBM or BSD-DB. A typical choice would be OBDC using SQL as a communication interface. This implementation would be very powerful since a great number of software tools these days include the possibility to use SQL syntax to interface databases. All kinds of statistical studies could be run upon the data saved this way in the database, linking demographics with the actual results arising from the conjoint analysis.

A good choice for CGI scripts would probably be Perl, a very extended and popular package routinely used in web development and automation. Perl can interact with the database, has built-in IPC mechanisms, has dozens of modules especially written for web design, and is free and well documented and maintained.

Another useful technology for the general automation of the server is XML (eXtensible Markup Language). XML was originally designed to overcome the inherent limitations of HTML as a markup language by being extensible in the sense that new tags could be added to those available in original HTML syntax. For our purposes, XML is a very elegant way to specify and store analysis definitions, and therefore it allows us to easily separate the definition of the analysis and the way of rendering (executing) it, that is, using the template. We first would define what the syntax of a generalized analysis is through a DTD (Document Type Definition). Then, a particular analysis would be specified and stored as an XML document using the syntax defined in the DTD. For example, the analysis setup pages in the administrator access would
result in writing an XML file according to the DTD and what the researcher specifies through the pages. Finally, what the respondent would see is the output of the execution of a "template object" using both the DTD and the XML file for parsing the particular parameters of the analysis being considered. A template could directly be written to parse these XML documents without knowing what the particular analysis is really about, and the analyses written in XML (either manually or assisted by the administrator access pages) could be interpreted and rendered using whatever template is selected. Tools for parsing and debugging these XML definitions are becoming ubiquitous in the web development community these days.

Finally, a number of technologies are available for enhancing the graphical aspects of the respondent web interface. Between them we may count the use of advanced features of the HTML 4 specification (such as refresh meta-tags, useful to refresh the screen when showing the results online in real time to the market researcher through the administrator access), the use of client server scripts (typically in JavaScript, VBscript or JScript) embedded in the HTML body, or even several browser-specific advanced features (such as dynamic HTML generation). Obviously using browser-specific features is not advisable before they become widely accepted and implemented by the other browsers, since it significantly reduces the possible audience.

3.4. Conclusions

The creation of a web server capable to perform the functions of an ASP (Application Service Provider) as a "one stop shop" for market research is feasible using a range of web technologies available today. An especially interesting example is the case of conjoint analysis for which the ACA (Adaptive Conjoint Analysis) algorithms mentioned in chapter 2 are particularly useful. In this chapter we have outlined what the general architecture of such a web
server would be; we have described each separate component and their interaction mechanisms; and we have mentioned what technologies available in today's marketplace would be useful in its implementation. In the following chapter we report the creation of a first real implementation case of such a system.
Chapter 4: Implementation Case Study

4.1. Introduction

In this chapter we introduce a first implementation of the architecture of a web server for market research as described in the preceding chapter. We will show practical results obtained from our particular implementation which, although preliminary, we nevertheless believe are significant enough to become a proof of concept for the whole methodology. The web server has been installed in an actual Linux machine and has been tested with sample respondents to check its functionality and robustness, with satisfactory results.

4.2. Description

4.2.1. Internal architecture

The heart of the server consists on a set of CGI scripts written in Perl 5.0. Altogether these scripts amount to about a thousand lines of code. The templates have also been created using Perl code with a well defined API (Application Programming Interface) to the rest of the scripts. Each template now accounts for about 700 lines of Perl code.

The engine is written in Matlab native code, using more that 1200 lines of code. Interface scripts to Matlab subroutines have been written and are called from the Perl CGI scripts. Only one adaptive conjoint algorithm (the one developed at MIT as mentioned in chapter 2) is available at the moment of writing this thesis. A comprehensive API has been defined and used
in the coding of the algorithm, and therefore if a new one is written respecting this API its utilization by the rest of the CGI scripts would be straightforward.

The database is the first implementation is a set of Matlab files suitable to be interactively monitored from the Matlab console, although some automatic reporting mechanisms are provided as described in the sections below.

4.2.2. **Administrator access for analysis setup**

Figure 4.1 shows the administrator access screen.

![Administrator Access](image_url)

**Administrator Access**

In this page you initialize and destroy databases for Adaptive Conjoint Analysis. First you have to specify the number of levels per attribute in the "Problem Definition" box, separated with ","s. For example, 2,2,4 would be entered for a problem with three attributes and two, two and four levels per attribute respectively. You also have to enter the number of attributes that will vary (at most) in a given question. Attributes that are at the same level for the two products on a given question will not be displayed.

Then you specify the relative path for the analysis. Several analyses may be running at the same time, each one with different parameters. For example, if you enter "conj2_2_4" then a directory named "conj2_2_4" would be created in the analyses subdirectory and you would have to direct your respondents to http://hoover.mit.edu/conj/analyses/conj2_2_4.

You may remove the parameters of an ongoing analysis by clicking on the "remove" buttons and entering the directory for the analysis (the one you specified when you created the analysis).

You may also monitor the results of an ongoing analysis by clicking on the "monitor" button and entering the directory for the analysis.

<table>
<thead>
<tr>
<th>Problem Definition</th>
<th>Number of questions</th>
<th>Number of varying attributes per question</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,2,4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Create ☑ Remove ☑ Monitor ☑ OK ☑ Reset ☑

This is version alpha--

© Julio Faura 2000 (faura@ieee.org)

Fig 4.1 - Administrator access screen
As it can be seen, a researcher can easily create, destroy or monitor a custom conjoint analysis. A first step is specifying the number of questions to be asked, the number of attributes and levels, and the number of attributes to be allowed to vary per question. A second step allows to specify names and units for the attributes, the levels to be considered, and whether the order of the levels for each particular attribute should be specified by the user or not (for example an attribute such as price would not need to be re-sorted, while color would. An attribute such as weight may need to sort just the two extremes to assess what the direction of the preference is). Figures 4.2 and 4.3 show attribute and level definition screens.

Fig 4.2 - Attributes and levels definition screen
As it can be seen in the figures, even the welcome and help texts can be customized using the tool. Actually everything entered will be interpreted as HTML source so it is possible to use tags, embed scripts (in JavaScript for example) and images, or point to other web pages, both locally and remotely.
4.2.3. **Respondent Access**

Once the analysis has been set, the administrator would tell respondents to point their URLs to the respondent entry page, shown in Figure 4.4.

![Image of respondent entry page](image)

**Fig 4.4 - Respondent entry page**

The respondent is then greeted with the welcome message and help screen specified by the researcher through the administrator access, as shown in figure 4.5. In fact a whole series of HTML pages with embedded images or JavaScript scripts could be placed here since the help text and the welcome note themselves are interpreted as HTML code as we mentioned before.

Then, a preference assessment page would ask the respondent to sort the attributes levels if required (the particular re-sorting requirements were specified in setup time by the researcher), as shown in figure 4.6.
Welcome to the car survey!

Welcome to the car survey. We need your help to design a car that suits you best, and therefore we would like to ask you some questions about your preferences. The following screens will describe two different cars. Each of them may have a different color or feature, and it is important that you mentally picture how does the car look like in your mind and then tell us which one do you prefer and in what degree. Thanks for taking the time to respond to this survey. Your input is greatly appreciated.

Click here to start.

Please click on the level of Air bag you prefer:

- Air Bag included
- No Air Bag included

Speaking of Color you prefer:

- Black
- Navy blue
- Orange
- Cow type black & white

Submit:

Fig 4.5 - Welcome screen

Fig 4.6 - Preference assessment screen
Once the required levels have been re-sorted, the respondent gets the actual questions and answers them, as shown in figure 4.7. The questions are automatically generated (the very first one is randomly generated for the very first respondent) and displayed using the template CGI scripts.

![Question screen](image)

Fig 4.7 - Question screen

The respondent can always click on the corresponding help icon (the question mark on a caption) when she needs help about the meaning of a particular attribute. In that case, a new window pops up with the text specified in the analysis setup phase, which is also interpreted as HTML code.
Finally the respondent would sequentially submit all the answers, which would be saved to the database. Demographic information could also be asked at this (or any) point in time.

4.2.4. **Administrator access for analysis monitoring**

By clicking on the "monitor" radio button in the administrator access, the researcher can browse the results of the survey in real time, as shown in figure 4.8. The amount of information given through this page could be in fact easily customized for the particular needs of each researcher. In our particular implementation, the information show in the screen is refreshed every ten seconds.

---

**Interactive analysis monitor**

Path for this analysis: demo  
Problem Definition: 2;2;2;4  
Number of questions: 4  
Number of varying attributes per question: 4  
Summary of attributes:

- Air Conditioning \( (\text{A/C included}, \text{No A/C included}) \)
- Anti-Blocking System \( (\text{ABS included}, \text{No ABS included}) \)
- Air bag \( (\text{Air Bag included}, \text{No Air Bag included}) \)
- Color \( (\text{Black, Navy blue, Orange, White}) \)

Number of respondents so far: 5  
Aggregate initial estimate: \(-1.7079e-10, 8.034, 10.7401, 41.7093, 39.5166\)  
Mean of individual estimates: \(3.0861, 14.5027, 8.1211, 36.7215, 37.5086\)  
Individual error to aggregate estimate: 10.8339  
First question: 1-1;1;1;1

Fig 4.8 - Analysis monitoring screen
4.3. Conclusions

In this chapter we have shown the practical results of a first real implementation of the general architecture described in the preceding chapter. In this particular implementation we have coded and used the novel conjoint analysis algorithm developed at MIT mentioned in chapter 2, although the general methodology we have used would allow us to extend the capabilities of the server to other conjoint analysis techniques or even different types of market research analysis.
Chapter 5: Conclusions and Future Work

5.1. Main conclusions

In section 1.2 we stated what the objectives of the thesis were. We now explain how we believe we have reached those objectives:

- We have found that an integrated automated approach to web-based market research servers is definitely feasible. In chapter three we have outlined what the block diagram of a generalized server of this kind may be, and we have found solutions to the design and implementation of all of these blocks and their interdependence and interconnection. We have also addressed the use of current internet technologies to determine what role would they play in the implementation of this kind of system. We have focused on the use of adaptive conjoint analysis but since this technique is one of the ones that require most interaction between the computational engine and the respondent, our methodology should be easily generalized to a wide range of additional market research techniques in an straightforward manner.

- We have proven this concept by physically implementing a particular realization of one of these architectures. Our first realization uses a novel variation of adaptive conjoint analysis being developed at MIT for web automation. We strictly follow the guidelines and general structure synthesized in chapter 3. The web server has been tested with sample conjoint analyses using actual sample respondents, with excellent results.
In chapter two we provided a background introduction for conjoint analysis and we reported our conclusions about the emerging field of online market research. We found some benefits and drawbacks compared to the traditional approach, and all in all we can say that online market research enhances and extends in many ways the techniques traditionally available to market researchers. We are still to see what truly novel methodologies will be in place as the internet technology pervades into all the different parts of the business processes of the different industries.

5.2. **Future research lines**

We believe that the actual possibilities for online market research have only begun to unravel. Among the most urgent and promising future work we may mention the following:

- Actually extending the methodology to other market research techniques (such as choice-based conjoint or perceptual maps) by implementing case studies as we have done in the adaptive conjoint analysis case. An integrated web server capable of providing all these techniques for the market researcher to choose would be extremely powerful.

- Further automating data analysis and documentation. As an ultimate objective, the market research would just specify the parameters of the design and would just have to interpret the numerical figures we would directly get from the automatic documentation generator.

- Extend the methodology for clustering analysis to determine and document market segments. Extremely fast segmentation schemes could be obtained in a matter of days with this methodology.
Linking conjoint analysis and clustering analysis to collaborative filtering as a way of fast consumer segmentation. The idea would be to ask apparently unrelated questions that the user would be typically willing to answer with the idea estimating preferences on other attributes by previously running regressions and estimating correlations between attributes. This technique would enable internet marketers to dramatically increase the hit rates in banner advertising. As market researchers, we could also qualify respondents in a much more efficient manner.
References


Toubia, Olivier, MsC Thesis, MIT (to be published).


Urban, Glen L., Bruce D. Weinberg and John R. Hauser, "Premarket Forecasting of Really-New Products", Journal of marketing, 60 (1996), 47-60


Mgt
3207
2000
S.M.M.O.T.
FAURA, Julio M.