

**FUZZY ASSOCIATION RULES FOR ESTIMATING CONSUMER  
BEHAVIOUR MODELS AND THEIR APPLICATION TO EXPLAINING  
TRUST IN INTERNET SHOPPING**

J. Casillas

Dept. Computer Science and Artificial Intelligence, Computer Engineering  
School, University of Granada, E-18071, Granada, Spain  
Phone: (+34) 958 240804, Fax: (+34) 958 243317, e-mail:  
casillas@decsai.ugr.es

F.J. Martínez-López

Dept. Marketing, Business Faculty, University of Granada,  
E-18071, Granada, Spain  
Phone: (+34) 958 242350, Fax: (+34) 958 240695, e-mail: fjmlopez@ugr.es

F.J. Martínez

Dept. Financial Economy and Accounting, Business Faculty, University of  
Huelva,  
E-21071, Huelva, Spain  
Phone: (+34) 959 017891, e-mail: francis@uhu.es

**Summary**

Complex models have been traditionally and increasingly used by both marketing academics and practitioners to represent and understand consumer behaviour. Thus, we firstly pose that models of consumer behaviour firms use to make their decisions must be close to what a real Marketing Management Support System should offer, in order to be of use.

In this sense, rather than focusing on analysing the quality of the consumer-related marketing phenomena represented by such models – i.e.: those theoretical issues which support their validity –, we take an in-depth look at the utility of the statistical techniques used to estimate those models theoretically proposed.

Thus, we pose that the use of fuzzy systems as a knowledge discovery tool is of great interest for improving the interpretation and understanding of such consumer models. Moreover, we put forward a new application for consumer behaviour modelling, based on fuzzy association rules (FAS) for adjusting the data, as a complementary alternative to the results obtained by using the classic technique of model estimation based on Structural Equation Modelling (SEM).

With this aim, a behavioural model centred on explaining consumer attitude towards Internet and trust in Internet shopping is presented, being later tested by making use of both FAS and SEM. Finally, a comparative analysis of the results is done, focusing particularly on our proposal of methodological application.

**Keywords:** consumer behaviour, marketing modelling, model estimation, structural equation modelling, fuzzy association rules, knowledge discovery

## 1. INTRODUCTION

The understanding and explanation of consumers' behaviour patterns in a manner increasingly efficiently have been considered, by both marketing academics and practitioners, as an essential question for those firms concerned about being more competitive by performing better market-related decision processes.

In this respect, consumer behaviour models, inasmuch as they are marketing models, are considered as a specific case of Marketing Management Support System (MkMSS), and throughout the time have demonstrated to be a source of transcendental relevance for the development of marketing science [see van Bruggen and Wierenga (2000)]. Notwithstanding, actual models of consumer behaviour do not seem to cover all the necessities that a model which aims to aid marketing decision making should supposedly satisfy. With regard to this, based on Gatignon (2000), future models, considering both their theoretical and technical aspects, which try to explain consumers' decision making will have to be clearly focused on users' (demand side) requirements of such models. That is to say, models must be more complete, flexible, and customized to the strategic singularities of the competitive environment which their users operate in. Thus, as the main problem that actually face firms oriented to consumer markets is not the availability of information (data), but the possession of the necessary level of knowledge to take the right decisions, the use of avant-garde behavioural models able to exploit it may represent an essential source of competitive advantage.

On the other hand, it is expected that MkMSS will tend to improve their performance taking advantage of synergies caused by the integration of modelling estimation techniques based on classic econometrics with expert systems based on artificial intelligence.

In this sense, considering the three pillars on which marketing modelling is based [see Steenkamp (2000); Roberts (2000)], and, more specifically, consumer behaviour modelling, we focus our work on one of them – i.e.: the modelling estimation techniques and their improvement – presenting theoretically and empirically the potentials that methods of estimation based on *fuzzy association rules* (FAS) have for proposing and obtaining estimated models of behaviour that are more exhaustive, complex, flexible and interactive, and which offer much more quantity of qualitative information than preceding estimation techniques used in this field [Gatignon (2000); van Bruggen and Wierenga (2000)]. Thus, FAS can be a plausible alternative to complement, in principle, the results obtained by using Structural Equation Modelling (SEM) techniques which have been the ones usually used in the last decades to estimate complex models of consumer behaviour based on stated responses. Furthermore, even considering the clear utility offered by SEM to test theoretical models proposed, it shows a series of limitations that could be solved by FAS estimation techniques, inasmuch as: (1) it assumes a linear relation among variables of the model, so it does not allow us to analyze nor, therefore, interpret relations among several variables when

such relations are contemplated with different degrees of intensity; (2) it does not provide qualitative information regarding the relations among variables in the model; (3) it usually works with “simple” or recursive causal models, without considering possible reverse relations among variables, so parameters obtained using SEM specific statistical software (e.g.: LISREL, AMOS, etc.) will tend to be biased; and last but not least (4) estimation techniques based on SEM are useful to test theoretical proposals for a consumer behaviour model, though its utility to support marketing decisions (MkMSS) seems to be restricted by its own results derived from its process of estimation [see Laurent (2000); Steenkamp and Baumgartner (2000)]. Therefore, this paper proposes the use of fuzzy systems as a knowledge discovery tool to allow marketing academics and practitioners to improve the understanding of consumer behaviour. With this aim, the paper is organized as follows. Section 2 highlights the relevance that methods based on FAS can have to help satisfy the previous idea when applied to model consumer behaviour. Thus, a methodological application is proposed based on the Cooperative Rules methodology. Section 3 describes a theoretical model of online consumer behaviour that will be used as a benchmark to show how our methodology works. Section 4 shows the results obtained by using both our application and SEM, and it treats the main implications that results arising from our application have for understanding the marketing problem posed. Finally, Section 5 highlights, from a methodological point of view, several interesting findings resulting from our application.

## **2. CONSUMER BEHAVIOR MODELING BY FUZZY ASSOCIATION RULES**

### **2.1. INTRODUCTION TO KNOWLEDGE DISCOVERY BY FUZZY ASSOCIATION RULES**

Fuzzy rule-based systems currently constitute one of the most important areas for the application of fuzzy set theory. These systems are an extension of classical rule-based systems, because they deal with fuzzy rules instead of classical logic rules. They can be considered as a knowledge extraction tool to discover intrinsic relationships contained in a data base [Freitas (2002)]. Thus, by means of association rules (fuzzy rules in our case), we can represent the relation existing among different variables, thus deducing the patterns contained in the examined data. In knowledge discovery, the process to obtain these patterns must be automatic, or semi-automatic, discovered patterns must be comprehensible and they must provide useful information, and data must be invariably presented in substantial quantities [Witten and Frank (2000)].

Useful patterns allow us to make non trivial predictions about new data. There are two extremes to express a pattern: as black boxes, whose internal behaviour is incomprehensible; and as white boxes, whose construction reveals the pattern structure. The difference lies in whether the generated patterns are represented with an easily examined structure, which can be used to reason and to inform of further decisions. In other words, when the patterns are structured in a comprehensible way, they will be able to help in

explaining something about the data. This problem with knowledge discovery, the interpretability-accuracy trade-off, is also being currently tackled in fuzzy modelling [Casillas *et al.* (2003a) (2003b)] and will be considered by our proposal.

The use of fuzzy systems when developing the knowledge discovery process has certain advantages, as follows: they allow us to use uncertainty data; they adequately consider multi-variable relationships; results are easily understandable by a human being; additional information can be easily added by an expert; the accuracy degrees can be easily adapted to the problem necessity; and the process can be highly automatic with low human intervention.

Therefore, we will use fuzzy logic as a tool to structure the information of a consumer behaviour model in a clear legible way that is close to the human being reasoning. The fuzzy system will allow us to properly represent the interdependence of variables and the non-linear relationships that could exist among them. Finally, optimization algorithms will design the fuzzy system to meet the interpretability and accuracy criteria imposed by the expert.

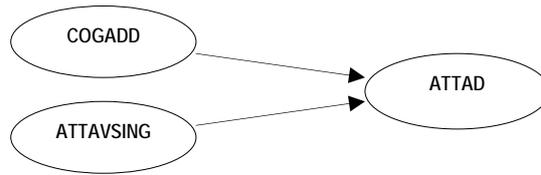
The section below introduces the methodology followed for applying fuzzy systems to consumer behaviour modelling and a brief description of the specific fuzzy system learning method used in this paper.

## **2.2. METHODOLOGY FOR CONSUMER BEHAVIOUR MODELLING WITH FUZZY ASSOCIATION RULES**

This section introduces the process we propose to perform knowledge discovery by fuzzy association rules. Basically, it consists of preparing the data and establishing the scheme we follow to represent the knowledge existing in the data. Once these aspects are defined, a specific learning method is used to automatically design the fuzzy association rule system.

The process is the following:

- **Data collection:** as traditionally done in marketing, it is extracted by means of a questionnaire in a similar way to the models estimated by structural equation modelling.
- **Data processing:** it is necessary to adapt the collected data to a scheme that is easily tractable by fuzzy system learning methods. Thus, when more than one item is used to assess the same concept, the arithmetic mean value is used.
- **Representation:** fuzzy association rules are used to represent the relationships between the variables. Once the structural model is fixed by the marketing expert, a fuzzy rule base is used to relate input with output variables. For example, given the following structural model defined by three latent variables; i.e.: (1) COGAD: cognitive valuations of a certain ad, (2) ATTAVSING: overall opinion about advertising, (3) ATTAD: overall opinion about a certain ad (see Figure 1).

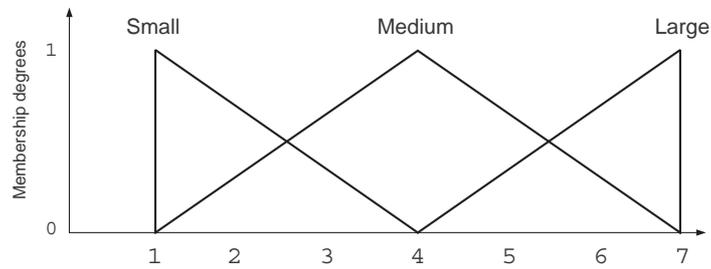


**Figure 1: Example of a structural model**

We use the variables COGADD and ATTA VSING as inputs (antecedent) and ATTAD as output (consequent), producing rules such as the following:

IF COGADD is **high** and ATTA VSING is **low** THEN ATTAD is **medium**

Regarding the membership functions, the numerical scale used in the questionnaire has been translated to a fuzzy semantic as the following example shows (Figure 2) for 3 linguistic terms and a 1 to 7 rating scale:



**Figure 2: Transformation of a rating scale to linguistic terms**

Although some tuning methods have been proposed in the literature to automatically design the membership functions [Karr (1991)], we think that it should be maintained invariable in order to get the maximum interpretability, a very important issue in knowledge discovery.

- **Learning method:** One of the most interesting characteristics of our methodology is that any fuzzy rule learning method for fuzzy rule-based systems can be used as an optimization/design process. In any case, since we force the obtained fuzzy system to be very rigid due to the fact that the membership functions are not optimized, a sophisticated fuzzy rule learning method should be used.

In this paper we propose to use a simple and quick method that has proved to obtain good results in fuzzy modelling, maintaining a high interpretability of the derived fuzzy models: the method COR (cooperative rules) [Casillas *et al.* (2002)]. It arises as an effort to exploit the accuracy ability of linguistic fuzzy rule-based systems by exclusively

focusing on the fuzzy rule base design. In this case, the fuzzy membership functions and the fuzzy model structure remain invariable, thus resulting in the highest interpretability. The method has the final goal of enhancing the *interpolative reasoning* the fuzzy rule-based system develops. This is one of the most interesting features of fuzzy rule-based systems and plays a key role in their high performance, being a consequence of the cooperative action of the linguistic fuzzy rules. To do this, the fuzzy rule base design is made using global criteria that consider the action of the different rules jointly. It is attained by means of a strong smart reduction of the search space. The main advantages of the COR methodology are its capability to include heuristic information, its flexibility to be used with different optimization or search techniques, and its easy integration into other derivation processes.

COR consists of two stages:

1. *Search space construction*, where a set of candidate rules is obtained for each fuzzy input subspace.
2. *Selection of the most cooperative fuzzy rule set*, where a combinatorial search is performed among these sets looking for the combination of rules with the best global accuracy.

In this paper, we also include an enhancement to the original proposal to allow it to eliminate badly defined and conflicting rules with the aim of improving the interpretability (less number of rules) and the accuracy (better cooperation among rules). A wider description of the COR-based rule generation process including the fuzzy rule set reduction mechanism is shown in figure 3. Moreover, a graphical representation of its behaviour is depicted in figure 4.

We use a specific ant colony optimization algorithm as the optimization technique. For a deeper description of the COR methodology and the use of this optimization technique, refer to Casillas *et al.* (2003c).

### **3. AN ONLINE CONSUMER BEHAVIOUR MODEL FOR EXPLAINING ATTITUDE TOWARDS INTERNET AND TRUST IN INTERNET SHOPPING**

First of all, it should be noted that our main aim in this paper will not be focused so much on clarifying problems related with the online consumer behaviour model presented here, as on treating in depth those questions associated with the results provided by our application in comparison with the others usually used, which are based on SEM. Therefore, theoretical issues of this model, as well as those methodological aspects related to the survey applied to obtain the data, are synthetically presented.

The specific literature review in this section concludes the need to develop a specific theoretical framework to better understand consumer behaviour in virtual surroundings. Thus, our model hopes to contribute towards the previous literature by concentrating on the study of certain internal determinants of consumer (real or potential) behaviour on the Internet and on the relationship of consumers with the phenomenon of Internet-based e-shopping.

**Inputs:**

- An input-output data set —  $E = \{e_1, \dots, e_l, \dots, e_N\}$ , with  $e_l = (x_1^l, \dots, x_n^l, y_1^l, \dots, y_m^l)$ ,  $l \in \{1, \dots, N\}$ ,  $N$  being the data set size, and  $n$  ( $m$ ) being the number of input (output) variables — representing the behavior of the problem being solved.
- A fuzzy partition of the variable spaces. Uniformly distributed fuzzy sets are regarded. Let  $\mathcal{A}_i$  be the set of linguistic terms of the  $i$ -th input variable, with  $i \in \{1, \dots, n\}$ , and  $\mathcal{B}_j$  be the set of linguistic terms of the  $j$ -th output variable, with  $j \in \{1, \dots, m\}$ , with  $|\mathcal{A}_i|$  ( $|\mathcal{B}_j|$ ) being the number of labels of the  $i$ -th ( $j$ -th) input (output) variable.

**Algorithm:**

## 1. Search space construction:

- 1.1. Define the fuzzy input subspaces containing positive examples: To do so, we should define the positive example set ( $E^+(S_s)$ ) for each fuzzy input subspace  $S_s = (A_1^s, \dots, A_i^s, \dots, A_n^s)$ , with  $A_i^s \in \mathcal{A}_i$  being a label,  $s \in \{1, \dots, N_S\}$ , and  $N_S = \prod_{i=1}^n |\mathcal{A}_i|$  being the number of fuzzy input subspaces. In this paper, we use the following:

$$E^+(S_s) = \{ e_l \in E \text{ such that} \\ \forall i \in \{1, \dots, n\}, \forall A_i^s \in \mathcal{A}_i, \mu_{A_i^s}(x_i^l) \geq \mu_{A_i^s}(x_i^l) \},$$

with  $\mu_{A_i^s}(\cdot)$  being the membership function associated with the label  $A_i^s$ .

Among all the  $N_S$  possible fuzzy input subspaces, consider only those containing at least one positive example. To do so, the set of subspaces with positive examples is defined as  $S^+ = \{S_h \mid E^+(S_h) \neq \emptyset\}$ .

- 1.2. Generate the set of candidate rules in each subspace with positive examples: Firstly, the candidate consequent set associated with each subspace containing at least an example,  $S_h \in S^+$ , is defined. In this paper, we use the following:

$$C(S_h) = \{ (B_1^{k_h}, \dots, B_m^{k_h}) \in \mathcal{B}_1 \times \dots \times \mathcal{B}_m \text{ such that } \exists e_l \in E^+(S_h) \\ \text{where } \forall j \in \{1, \dots, m\}, \forall B_j^{k_h} \in \mathcal{B}_j, \mu_{B_j^{k_h}}(y_j^l) \geq \mu_{B_j^{k_h}}(y_j^l) \}.$$

Then, the candidate rule set for each subspace is defined as  $CR(S_h) = \{R_{k_h} = [\text{IF } X_1 \text{ is } A_1^h \text{ and } \dots \text{ and } X_n \text{ is } A_n^h \text{ THEN } Y_1 \text{ is } B_1^{k_h} \text{ and } \dots \text{ and } Y_m \text{ is } B_m^{k_h}] \text{ such that } (B_1^{k_h}, \dots, B_m^{k_h}) \in C(S_h)\}$ .

In order to allow COR methodology to reduce the initial number for fuzzy rules, the special element  $R_\emptyset$  (which means "do not care") is added to each candidate rule set, i.e.,  $CR(S_h) = CR(S_h) \cup R_\emptyset$ . If this element is selected, no rules are used in the corresponding fuzzy input subspace.

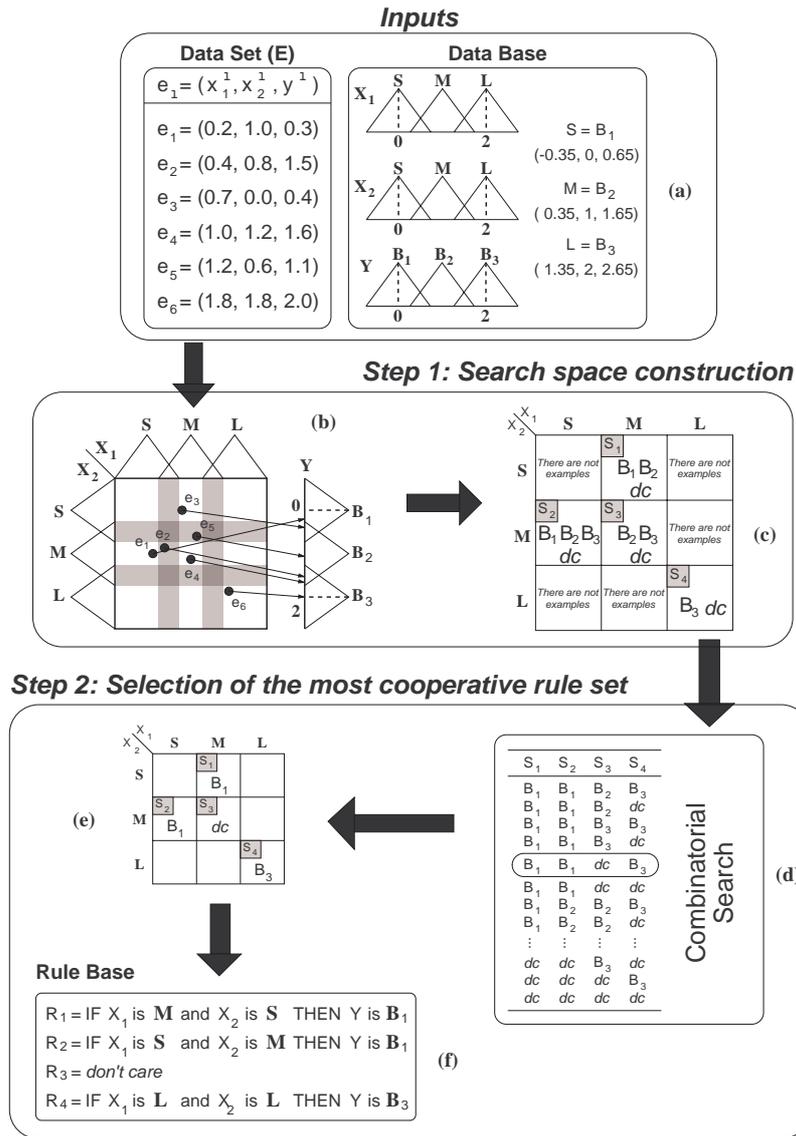
2. Selection of the most cooperative fuzzy rule set — This stage is performed by running a combinatorial search algorithm to look for the combination  $RB = \{R_1 \in CR(S_1), \dots, R_h \in CR(S_h), \dots, R_{|S^+|} \in CR(S_{|S^+|})\}$  with the best accuracy. Since the tackled search space is usually large, approximate search techniques are used.

An index measuring the cooperation degree of the encoded rule set is considered to evaluate the quality of each solution. In our case, the algorithm uses a global error function called *mean square error* (MSE), which is defined as

$$\text{MSE} = \frac{1}{2 \cdot N} \sum_{l=1}^N (F(x_1^l, \dots, x_n^l) - y^l)^2,$$

with  $F(x_1^l, \dots, x_n^l)$  being the output obtained from the fuzzy system when the example  $e_l$  is used, and  $y^l$  being the known desired output. The closer to zero the measure, the higher the global performance and, thus, the better the rule cooperation.

**Figure 3: COR algorithm**



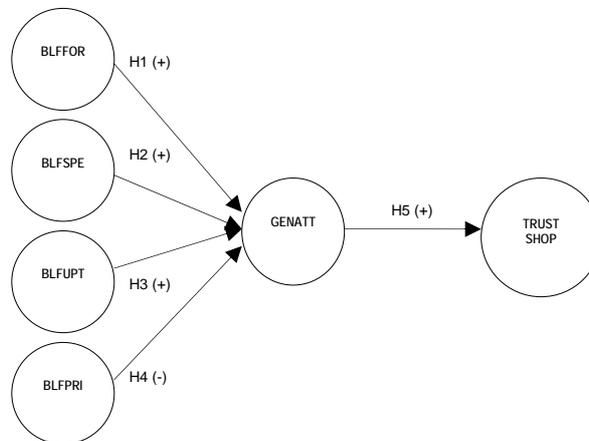
**Figure 4: COR-based learning process for a simple problem with two input variables, one output variable, and three labels in the output fuzzy partition: (a) data set and data base previously defined; (b) the six examples are located in four different subspaces that determine the antecedent combinations and candidate consequents of the rules; (c) set of possible consequents for each subspace, including the special element “don't care”; (d) combinatorial search accomplished within a space composed of 72 different combinations of consequents; (e) rule decision table for the fifth combination; (f) rule base generated from this combination**

In this sense, it should be noted, on the one hand, that knowledge regarding consumer opinions, beliefs and attitudes with respect to a specific object is essential for understanding consumer behaviour towards said object [see, as e.g., Schiffman & Kanuk (1997)]. On the other hand, and with reference to the phenomenon we are dealing with here, users' attitudes as regards certain aspects of the Internet greatly determine their cyber-behaviour [Hoffman, Novak & Peralta (1998)].

Within this context, our model concentrates on the users' general attitude towards Internet and on analysing its influence on the trust consumers have in using it as a way to shop, contributing towards amplifying on the numerous previous studies that look at analysing user attitudes towards a firm's web page and its effects on the trust and purchase intentions consumers might have in such a website [Chen & Wells (1999); Jarvenpaa & Tractinsky (1999); Cheskin Research and Studio Archetype/Sapient (1999); Lin & Lu (2001)].

### 3.1. THEORETICAL SPECIFICATIONS OF THE MODEL AND RESEARCH HYPOTHESES

The model we have put forward for the process of formation of consumer attitudes towards Internet is based on the ABC (Affect, Behaviour and Cognition) attitude model and on the cognitive-affective process of the hierarchy-of-effect model (see Figure 5). Thus, we maintain that a user will develop a series of opinions or beliefs (cognition) regarding the various attributes or characteristics of Internet that will determine, at least in part, his overall attitude towards the medium (affect). Furthermore, this general attitude will determine his attitude in terms of certain behavioural aspects related to its object. More specifically, it is to be expected that this general feeling influences consumer trust in Internet shopping. In consequence, our proposal is based on the following constructs:



**Figure 5: Structural model**

### 3.1.1. Beliefs towards the Internet.

This issue covers the cognitive element of the model, and we define it as the cognitive opinions or responses of the user as regards certain attributes pertaining to Internet and the Web. In this sense, the fact that there is very little literature on this topic made the determination of the basic attributes rather difficult, so we decided to use other prior studies on the factors that determine a consumer's evaluation of a web page and on how the image of a virtual establishment is formed [Crawford and Shern (1998); Farquhar *et al.* (1998); Lohse & Spiller (1999); Helander *et al.* (2000)], on the understanding that, in our case, the object of the attitude was not that of a specific web page but, rather, that of Internet in general.

Specifically, the model we have put forward considers the following beliefs, briefly presented:

*Formal aspects* are related to the consumer's conceptions regarding the design and structure of the Internet information and, more precisely, of the Web. Thus, based on and expanding on Montoya-Weiss, Voss & Grewal's (2003) reflection, which focuses on analyzing the influence of the consumer's design assessment towards a certain web site in subsequent evaluations related to a particular online channel, it is expected that consumers will tend to have a higher overall evaluation of the Internet inasmuch as they better perceive those issues concerning design on the Web.

*Hypothesis 1:* There is a positive relationship between the consumer's perceptions of formal aspects and his overall attitude towards Internet

*Interaction Speed/Time of Response* is defined as the Internet's capacity in general, and of different web-sites in particular, to give a response when required, in a similar way to the response that an individual may receive when holding a conversation with another (Shih 1998). In this sense, Lin and Lu (2000) note that response time is the most important factor in the development of the user's beliefs towards a specific web-site. Thus, if we broaden this relation to the Internet in general (the Web), we can assume that the consumer's perception with regard to the medium's time of response will have a certain influence on the rest of the valuations and opinions about other aspects of the network (the Net), especially on the overall evaluation toward the Internet as a means of communication.

*Hypothesis 2:* There is a positive relation between the consumer's perceptions about interaction speed/time of response and his overall attitude towards Internet.

*Information updates* is related to the consumer's perceptions of content updating of web pages on Internet. In this sense, it is plausible to think that consumers will tend to show better opinions with regard to the Internet inasmuch as they perceive that, in general, information contained on web pages is up-to-date.

*Hypothesis 3:* There is a positive relation between the consumer's beliefs towards updating of the contents on the Internet and his overall attitude towards this medium.

*Privacy* is related to the user's perceptions as regards respect for intimacy when surfing. In this sense, considering the growing relevance that all matters related with consumer privacy have taken on since the nineties, the Internet has been the most criticized medium with regard to its potential to invade a consumer's intimacy (see, e.g., Bartel & Grubbs 2000; Richards 1997; Rust, Kannan & Peng 2002)

*Hypothesis 4:* There is a negative relationship between the consumer's perceptions of invasion of his privacy when navigating and his overall attitude towards Internet.

### **3.1.2. General attitude towards Internet.**

This concept forms the model's affective element in our study and is defined as, adapting the ideas from previous studies (e.g., Ajzen & Fishbein 1980; Oskamp 1991; Schiffman & Kanuk 1997; Solomon 1997), the general or global evaluation that the consumer has of this communication medium. It should be highlighted that consideration of this variable is necessary in any study which endeavours to understand consumer behaviour in the marketplace because, as stated by Hoffman, Novak & Peralta (1999), said attitudes will be determinants of the user's cyber-behaviour.

### **3.1.3. Trust in Internet shopping**

This concept covers the consumer's general perception regarding the credibility of the messages firms send over the Internet and their will to respect commitments reached with the users. Numerous studies have highlighted the particular influence the trust in virtual establishments has on consumers' purchase decisions [Hoffman, Novak & Peralta (1998); Cheskin Research and Studio Archetype/Sapient (1999); Jevons & Gabbott (2000); Lee & Turban (2001); Urban, Sultan & Qualls (2000);]. Furthermore, the possibility has been mentioned that consumer confidence in Internet as a way to shop is a direct antecedent of his trust in shopping on a particular website [McKnight & Chervany (2002); Urban *et al.* (2000)]. Thus, based on the above, we think that the consumer's overall evaluation toward the Internet communication medium, in that it gathers the opinion towards this public electronic infrastructure which supports his online commercial exchanges, should have a direct influence on trust in Internet shopping.

*Hypothesis 5:* There is a positive relationship between the consumer's overall attitude towards Internet and his trust in Internet shopping.

### 3.2. SURVEY RESEARCH METHODOLOGY

With regard to the methodological aspects related with this study:

- **Sample and sampling procedure:** the population targeted by our study was defined as being made up of university students, habitual Internet users, from Spain. The study sample was obtained using a non-probability sampling procedure (convenience sample). After purifying the completed questionnaires, the final sample consisted of 529 individuals.
- **Survey measures:** The measuring instruments used in the questionnaire to assess the study variables on the sample of Internet users were consumer rating scales and differential 7-point semantic scales (from 1 to 7).
- **Construct measurements:** Table 1 shows the measurements taken for each of the previously described constructs.

Table 1: Measures used in establishing the structural model

LABEL	CONSTRUCT	MEASURE
BLFFOR	Beliefs (formal aspects)	The information is well structured on the web pages
		Design of the web pages makes it easier to look for information
		Web pages are attractive
BLFSPE	Beliefs (speed)	Interaction with web pages is slow and boring (R)
		The Internet is quick
		Web pages that I usually visit don't download quick enough (R)
BLFPRI	Beliefs (privacy)	When I surf the Internet, I feel my privacy has been invaded
BLFUPT	Beliefs (update)	Information that I find through the different web sites on the Internet is up-dated
GENATT	General Attitude	Negative-Positive
		I don't like it- I like it
		Unfavourable-Favourable
TRUSHOP	Trust in Internet shopping	The Internet is a reliable source of in-home shopping
		On-line companies fulfil their obligations to the customer properly
		Purchase conditions on the Internet are clear

### 4. APPLICATION OF OUR METHODOLOGICAL APPLICATION TO THE PROPOSED MODEL

As we noted previously, SEM has been the habitual statistical method used to estimate complex models of consumer behaviour like ours. Thus, before applying our methodological approach to modelling estimation based on FAS, we consider it rather convenient to briefly present the main results that would have been obtained, in this case using SEM. Following this, the results obtained by our methodological approach and their analysis are introduced.

#### 4.1. RESULTS WITH STRUCTURAL EQUATION MODELLING

Estimating the proposed model by SEM was tackled using the Weighted Least Squares (WLS) method (LISREL software), which allows us to obtain asymptotically distribution-free estimates and, therefore, does not require the fulfilling of the multi-variate normality assumption. The goodness-of-fit indices resulting from the model's estimation suggest that it represents quite well the real relationships existing between its variables. All these indices obtain values that fall within the generally-accepted limits (see Table 2)

Table 2: Goodness of Fit Statistics

INDEX	VALUE
Root Mean Square Error of Approximation (RMSEA)	0.050
90 Percent Confidence Interval for RMSEA	(0.039;0.061)
Normed Fit Index (NFI)	0.95
Non-Normed Fit Index (NNFI)	0.97
Comparative Fit Index (CFI)	0.97
Incremental Fit Index (IFI)	0.97
Relative Fit Index (RFI)	0.94
Critical N (CN)	307.12
Root Mean Square Residual (RMR)	0.077
Standardized RMR	0.077
Goodness of Fit Index (GFI)	0.98
Adjusted Goodness of Fit Index (AGFI)	0.98

On the other hand, the results presented in Table 3 allow us to individually test the statistical significance of the different relations established previously in the hypotheses formulation section, as well as the degree of variance of the model's endogenous latent variables (GENATT and TRUSHOP) explained by their antecedents ( $R^2$ ). Thus, the whole part of the structural parameters in the model is significantly different to zero, and so the hypotheses have not been rejected, with the exception of the non significant relation BLFUPT – GENATT (hypothesis 3). Furthermore, we verified that the exogenous latent variables accounted for 40% of the variance in general attitudes towards Internet (GENATT), which, in turn, were responsible for 20% of the variance referring to trust in Internet as a way to shop (TRUSHOP). We should also note that two reverse items belonging to BLFSPE were transformed so that all three have the same sense.

Table 3. Structural parameter estimates, (standard error) and *t-value*

DEPENDENT VARIABLE	BLFFOR	BLFSPE	BLFUPT	BLFPRI	GENATT	Error Var.	R <sup>2</sup>
GENATT	<b>0.61</b>	<b>0.16</b>	<b>-0.0038</b>	<b>-0.19</b>	---	<b>0.54</b>	<b>0.40</b>
	(0.075)	(0.077)	(0.056)	(0.046)	---	(0.044)	
	8.14	2.07	-0.069	-4.13	---	12.32	
TRUSHOP	---	---	---	---	<b>0.36</b>	<b>0.45</b>	<b>0.20</b>
	---	---	---	---	(0.042)	(0.053)	
	---	---	---	---	8.46	8.58	

Therefore, this estimation technique has given us information about both the significant relations and the degree of intensity and sense of the relation. For example, as we hypothesized, BLFFOR and BLFSPE have a positive linear relation with GENATT, BLFFOR also being the most influential factor in GENATT (parameter estimate of 0.61), while BLFPRI presents a weaker negative relation with regard to that endogenous variable.

#### 4.2. RESULTS WITH OUR METHODOLOGICAL APPLICATION

However, the application we propose in this paper is able to enrich the previous results obtained by SEM techniques of estimation, allowing us to gain better knowledge about the effects of latent variables jointly considered, as well as a better notion of the behaviour of the relations among latent variables, simplifying and considering them in a isolated (bivariable relation) manner.

In this sense, after applying our methodological approach based on FAS to adjust the data, and previously transforming the marketing scales initially considered to a logical language (linguistic terms), we have obtained the following fuzzy rule set presented in three tables.

Firstly, we focus on GENATT as the consequent and on its behaviour in terms of its four antecedents. So, in Table 4, where a total of 26 rules are shown (this reduced number of rules was determined by the COR method after applying search space and fuzzy rule reduction processes), it can be seen the joint effects caused by the four antecedents – each defined by three linguistic labels, i.e.: small (S), medium (M), large (L)– of GENATT – defined by five linguistic labels, i.e.: very small (VS) , small (S) , medium (M), large (L), and very large (VL)– considered in our model, while in Table 5 we can see the system of rules achieved for the bivariable relations between GENATT and its four antecedents (the linguistic terms used in this case have been defined by five for both antecedents and consequent). Lastly, in Table 6, the system of rules obtained for the bivariable relation GENATT-TRUSTSHOP is shown (both defined by five linguistic labels).

Each table also shows the mean square error (MSE) obtained by each fuzzy association rule set over the treated data. MSE is defined as:

$$MSE = \frac{1}{2N} \sum_{i=1}^N (F(x_1^i, \dots, x_n^i) - y^i)^2 ,$$

with  $F(x_1^i, \dots, x_n^i)$  being the output obtained from the fuzzy system when the example data  $e^i = (x_1^i, \dots, x_n^i, y^i)$  is used, and  $y^i$  being the known desired output. The closer to zero the measure is, the higher the global performance and, thus, the better the rule cooperation.

Table 4: System of rules for GENATT jointly considering the four antecedents

MSE = 0.416517					
Rule	BLFFOR	BLFSPE	BLFUPD	BLFPRI	GENATT
1	L	L	L	S	VL
2	L	M	L	L	VL
3	L	M	L	M	VL
4	L	M	L	S	VL
5	L	M	M	M	VL
6	L	S	L	S	VL
7	L	S	M	L	VL
8	L	S	M	M	L
9	L	S	M	S	VL
10	L	S	S	S	VL
11	M	M	L	L	L
12	M	M	L	M	L
13	M	M	L	S	VL
14	M	M	M	M	L
15	M	M	M	S	L
16	M	M	S	S	VL
17	M	S	L	S	M
18	M	S	M	L	M
19	M	S	M	M	L
20	M	S	M	S	VL
21	M	S	S	M	L
22	M	S	S	S	M
23	S	M	L	L	L
24	S	S	L	M	L
25	S	S	M	M	L
26	S	S	S	L	S

Table 5: System of rules for GENATT considering the four antecedents separately

MSE	0.556464	0.597823	0.588862	0.560186
Antecedent label	BLFFOR-GENATT	BLFSPE-GENATT	BLFUPT-GENATT	BLFPRI-GENATT
VL	L	L	VL	L
L	VL	VL	L	—
M	L	L	L	L
S	L	L	L	L
VS	VS	L	L	VL

Table 6: System of rules for TRUTSHOP

MSE = 0.498801		
Rule	GENATT	TRUTSHOP
1	VL	M
2	L	M
3	M	M
4	S	S
5	VS	VS

### **4.3. ANALYSIS OF THE RESULTS OBTAINED WITH OUR METHODOLOGICAL PROPOSAL**

With the aim of analyzing and interpreting these results, if we follow common practice in marketing modelling testing, we should take a look at the bivariable relations integrating the model (see table 5). Thus, even when we are conscious of both, this approach, on the one hand, underuses the potentials of our application and, on the other, it is not definitely correct due to the fact that, in an isolated relation, variations of the consequent are not solely explained by alterations in certain antecedents, since there are other influencing factors, we consider it can be useful to obtain an initial idea about variable relationships which can aid to clarify the overall results obtained for variables (antecedents and consequent) taken together.

Thus, BLFFOR presents the highest variety of results for the consequent since GENATT shows very low values when beliefs towards formal aspects of the Internet are the lowest, and improves incredibly in that BLFFOR takes values of small or higher. In this sense, it is interesting to note that, as this does not occur with the rest of the antecedents, the consequent GENATT takes the lowest value only when BLFFOR is very small. Furthermore, it should be noted that, as can be seen by observing the graphical representation obtained by combining the system of rules for both variables, it does not look as if a linear relation between such variables fits with real data distribution.

The remaining antecedents do not appear to have such a strong influence as previously mentioned, since GENATT maintains the same good valuation (large), regardless of any modification in their values. Furthermore, their influence on the overall opinion as regards Internet is almost non significant, since variations in them do not cause modifications in the consequent. However, we have perceived that a reaction by GENATT exists when these three factors take certain specific values. In this sense, GENATT takes the highest value (very large) when BLFUPT is at its highest level or when BLFPRI is at its lowest. Furthermore, BLFUPT and BLFPRI have a very similar influence on GENATT if we bear in mind that the former is formulated positively and the latter negatively (invasion of privacy), with the exception, and this is interesting to note, that our application has not considered that there is an L (large) label for BLFPRI (when considering five linguistic labels). Finally, considering the antecedent BLFSPE, GENATT only changes its constant acceptable position of the good value L (large) to the excellent value of VL (very large) when BLFSPE takes the value of L (large).

On the other hand, taking into account the partial model of GENATT and its antecedents as a whole (Table 4), along with the previous results obtained from Table 5, the following main findings should be highlighted:

- Factors BLFFOR and BLFPRI, particularly the former, are the two that really present the potential to considerably discriminate or influence the consequent GENATT. In this sense, it can be observed how high beliefs (Large label) regarding formal aspects are determinant for overall opinion towards the Internet to take the highest value (Very Large label)

regardless of the values shown by the rest of the factors. Thus, examining the system of rules obtained when BLFOR is the highest, it can be seen that, from a total of ten, nine confirm this idea. However, the influence of BLFOR is not so clear when taking intermediate (Medium label) or low (Small label) values. In these cases, the other factors play a more important role as moderators of GENATT.

On the other hand, the factor BLFPRI seems to be more decisive in determining a higher range of values of the consequent GENATT to the extent that, when it is at its lowest (Small label), GENATT usually shows the highest (Very Large label), and when it is in an intermediate position, GENATT is good (Large label) for most of the rules obtained (seven out of nine). In turn, when the user's perception regarding privacy invasion is the highest (Large label), the determining power of BLFPRI over GENATT is not so strong, the consequent being mainly moderated by the values taken by the other three factors.

- Unlike the two previous factors, BLFSPE and BLFUPT do not appear to have such a strong effect on determining GENATT by themselves. In this respect, their influence is weaker than the others and differs when considering their range of values, their role being limited to moderate GENATT in certain combination of values of BLFFOR and BLFPRI (e.g.: rule 17 and 26).
- The great majority of rules seem to confirm the relations hypothesised between the antecedents and consequent GENATT. In this sense, for example, if we take rules 1 and 26, which coincide with the two possible extreme cases for determining GENATT, the former being defined by the best beliefs that the four factors considered can present, and vice versa for the latter, such an endogenous dependent variable takes the best and the worst value. Notwithstanding, we have found certain rules where the effects produced by combinations of values of the four factors do not seem to have a correspondence with the value that, in theory, GENATT should have taken, considering the significant sense of the relationships tested statistically using, firstly, LISREL and then our application. Could it be, therefore, that the results of our FAS-based application are inconsistent? The relations shown by several rules certainly seem to be apparently contradictory, and so a deeper analysis of this is necessary in order to clarify the matter.

In Table 7, several pairs of contradictory rules extracted from table 4 are shown. Furthermore, we reflect on both the expected value that the consequent GENATT should have taken according to previous results and the consistency of the rules system obtained by our application, analysing the behaviour of MSE associated with the model when modifying the consequent label in terms of such an expectation.

Table 7: Contradictory rules found

SETTING	RULE	BLFFOR	BLFSPE	BLFUPD	BLFPRI	GENATT	Model MSE	MSE Variation
Observed	7	L	S	M	L	VL	0.416517	
	8	L	S	M	M	L		
Expected	8	L	S	M	M	VL	0.422708	+0,006191
<hr/>								
Observed	15	M	M	M	S	L	0.416517	
	16	M	M	S	S	VL		
Expected	15	M	M	M	S	VL	0.453628	+0,037111
	16	M	M	S	S	L	0.422980	+0,006463
<hr/>								
Observed	21	M	S	S	M	L	0.416517	
	22	M	S	S	S	M		
Expected	22	M	S	S	S	L	0.419026	+0,002509
	22	M	S	S	S	VL	0.427267	+0,01075

The first contradiction comes from matching rules 7 and 8. To the extent that a negative relation has been detected between BLFPRI and GENATT, it is expected that, *caeteris paribus*, a reduction in BLFPRI will cause an improvement in GENATT or, at least, invariability in such a consequent if the degree of variation of BLFPRI is not enough to cause changes. However, our application has observed that paradoxical situation in rule 8. Under these circumstances, we consider that if a modification in the consequent (expected rule), adjusting it to its expected value, is able to reduce the MSE of the model, then it would be a sign that a modification of the rules system done according to what was expected fits in better with the observed patterns of data. Otherwise, the initial solution would be more appropriate. In this sense, it can be seen that the MSE of the partial model increases with this modification. The same occurs for the other contradictory pairs of rules found, where the MSE variation has depended on the degree of modification considered for the consequent.

Furthermore, it can be concluded, in this respect, that the contradictory effects shown by those pairs of rules in GENATT were just apparent, and that the results are not due to a localised optimum fall of the search algorithm, since better accuracy degrees are obtained with the apparent contradictions. A plausible argument to explain this is that of freeing this application presented here from responsibilities and making a deficient specification of the theoretical marketing model responsible for this. Due to this fact, other relevant factors not considered explicitly in our conceptual model, which can also influence GENATT, would be the answer to contradictions found in our system of rules.

Finally, with respect to TRUSTSHOP –the other endogenous latent variable of the proposed behavioural model– table 6 shows the results after applying our algorithm of approximation (Five linguistic terms have been used for both variables). In this sense, initially considering that users of our sample do not really show a considerable trust in Internet shopping, the pattern of the relation between both variables does not present a linear behaviour. Thus, GENATT seems to be an influential factor in TRUSTSHOP, as the former causes a gradual improvement in its consequent when overall opinion towards Internet increases from low to intermediate positions. However,

GENATT is not influential when taking intermediate or higher values to the extent that TRUSTSHOP maintains a moderate position.

#### **4.4. IMPLICATIONS OF THE PREVIOUS RESULTS BASED ON THE PROPOSED APPLICATION FOR EXPLAINING THE ONLINE CONSUMER BEHAVIOUR MODEL**

Results from using our application allow us to confirm that, on the one hand, the different beliefs towards the Internet contemplated here have a significant influence on the user's general opinion towards the Internet, though their influential power and intensity differ depending on both the factor and the value taken by each of them. On the other hand, attitude towards the Internet has also been a significant factor to determine the user's trust in Internet shopping, though a similar reflection could be made on its degree of influence according to the value it takes.

Thus, opinion about formal aspects of the Internet and web sites has been the most relevant belief to discriminate values taken by the overall attitude towards the Internet. This can be seen more clearly when considering the relation between both in an isolated way. Furthermore, high opinions about formal aspects usually ensure very good attitudes towards the Internet. In other words, when users present the highest opinion about this factor, the remaining beliefs do not seem to be particularly influential. Online businesses should, therefore, take this aspect well into account, since, using a parallel thought, it will have a considerable effect on the user's affective response toward their website. Nonetheless, this factor's degree of influence is not so clear when it has an intermediate or low presence.

With regard to the belief of an invasion of privacy when surfing the web, this does not show a constant degree of influence on general opinion towards the Internet either. It seems to be more influential when it takes low or intermediate values. Thus, the creation of an online environment able to highly respect the user's privacy will also generate good affective responses to a certain website, in that this issue has proved to be more influential when well valued.

Beliefs towards interaction speed, bearing in mind that our sample of users did not really show positive perceptions of this, since the majority of individuals value it on an intermediate or bad level, have a residual influence on attitude towards the Internet. This fact is contradictory with ideas defended by Lin and Lu (2000) when highlighting that this factor is the most influential for the development of the user's beliefs towards a certain web site. On the other hand, a similar influence has been found for the case of beliefs towards content updating of websites. Thus, even when both of these aforementioned factors do not transcendently determine the formation of the overall opinion towards the Internet, we have found certain cases where, when the influence of the formal aspects and invasion of privacy factors is not so clear, one or other of said factors can moderate the value of the dependent variable.

Finally, primarily considering that the user population still shows certain reserves toward becoming involved in Internet shopping, apparent when

observing the overall level of trust, overall attitude towards the Internet seems to be an important variable to generate trust up to a certain point, but beyond that point, this variable does not discriminate such a trust. In this respect, attitude towards the Internet does not improve trust in Internet shopping when taking the former values from the intermediate to the highest level. Therefore, the main problem should be focused on thinking about why the current situation occurs and how to generate higher levels of consumer trust.

## **5. CONCLUDING REMARKS**

As pointed out in the initial section concerning the motivation of this work, a tool for estimating a proposed consumer behaviour model should not only be useful for testing a set of theoretical relationships which form such a model, but it must also be able to help the marketing management function obtain a good perspective of certain marketing problems and take the right decisions. With that aim, we have empirically tested the theoretical relations of our model using both SEM and FAS, and so several conclusions can be drawn when comparing them.

On the one hand, when using SEM we obtain results that allow us to know how well the model fits the data as well as revealing what the significant relations are. Furthermore, we achieve different coefficients (parameters) that give us information about the sense and intensity of the relation.

However, after applying our methodological proposal we have confirmed some previous ideas. In this sense, when real relations between variables are not linear, parameters resulting from SEM are deficient in terms of really explaining and giving information about how a certain variable (antecedent) influences another (consequent). It is not correct to consider that a linear coefficient of the relation with the dependent variable remains the same for all the values that the antecedent may take.

Therefore, our application allows the user's model (e.g.: a marketing manager) to know what the exact behaviour is of the relations among variables for a certain market situation. Results from this application are rather more useful than SEM results for helping MkMSS, to the extent that, not only does it allow the model proposed to be tested, but it is also able to give in-depth qualitative information about the relations according to the degree of intensity of the antecedents.

On the other hand, SEM results can show, as is the case for BLFUPT-GENATT in our example model, that a relation is not statistically significant, while results from our application indicate that such a factor can be decisive in moderating the consequent in certain cases. Thus, a linear approximation to estimate a complex model of consumer behaviour, though helpful, could be too simplistic to give a true perspective of the relations.

Furthermore, a case for future treatment is that of why some factors are statistically significant antecedents and not others, based on SEM results, if when we look at the results from our application it can be seen that they present a similar relational pattern with the consequent. This could be due to the inherent philosophy of adjusting to the data which SEM is based on.

In conclusion, though more exhaustive work is needed to purify the potentials of our application for the consumer marketing modelling issue, based on the results presented here it can be said that our application is able to give higher qualitative information about how the relations among variables behave. Thus, in principle, it is quite convenient to make use of this kind of approximation to explain relations in a complex model of consumer behaviour as a complement to the classic SEM-based techniques used up until now.

## References

- Ajzen, I., Fishbein, M. (1980). *Understanding attitudes and predicting social behaviour*. Prentice Hall, Englewood Cliffs, New Jersey.
- Bartel, K., Grubbs, M. (2000). Dimensions of privacy concern among online consumers. *Journal of Public Policy & Marketing*, 19 (1), 62-73.
- Casillas, J., Cordón, O., Herrera, F. (2002), "COR: A methodology to improve ad hoc data-driven linguistic rule learning methods by inducing cooperation among rules", *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, vol. 32, pp. 526-537.
- Casillas, J., Cordón, O., Herrera, F., Magdalena, L. (Eds) (2003a), *Interpretability issues in fuzzy modelling*, Springer, Heidelberg, Germany.
- Casillas, J., Cordón, O., Herrera, F., Magdalena, L. (Eds) (2003b), *Accuracy improvements in linguistic fuzzy modelling*, Springer, Heidelberg, Germany.
- Casillas, J., Cordón, O., Herrera, F. (2003c), "COR methodology: A simple way to obtain linguistic fuzzy models with good interpretability and accuracy", in Casillas et al. (2003b).
- Chen, Q., Wells, W., 1999. Attitude toward the site. *Journal of Advertising Research* (September/October), 27-37.
- Cheskin Research and Studio Archetype/Sapient, 1999. Ecommerce trust study.
- Crawford, F., Shern, S., 1998. Six Winning Strategies for Web Selling. *Retail news*, Ernst & Young (Spring), 1-3 and 8.
- Farquhar, B., Gordon, L., Balfour, A., 1998. Consumer needs in Global Electronic Commerce. *International Journal of Electronic Markets* 8 (2), 9-12.
- Freitas, A.A. (2002), *Data mining and knowledge discovery with evolutionary algorithms*, Springer, Heidelberg, Germany.
- Gatignon, H. (2000), "Commentary on Peter Leeflang and Dick Wittink's "Building models form marketing decisions: past, present and future"", *International Journal of Research in Marketing*, vol. 17, pp. 209-214.
- Helander, M., Khalid, H., 2000. Modelling the customer in electronic commerce. *Applied Ergonomics* 31, 609-619.
- Hoffman, D., Novak, T., Peralta, M., 1998. Building consumer trust in online environments: the case for information privacy. *Communications of the ACM* 42 (4), 80-85.

- Hoffman, D., Novak, T., & Peralta, M. (1999). Information privacy in the marketplace: implications for the commercial uses of anonymity on the Web. *The Information Society* 15, 129-139.
- Jarvenpaa, S.L., Tractinsky, N., 1999. Consumer trust in an Internet store: a cross-cultural validation. *Journal of Computer-Mediated Communications* 5 (2), <http://www.ascusc.org/jcmc/vol5/issue2/jarvenpaa.html>
- Jevons, C., Gabbott, M., 2000. Trust, Brand equity and brand reality in Internet business relationships: An interdisciplinary approach. *Journal of Marketing Management* 16, 619-634.
- Karr, C.L. (1991), "Genetic algorithms for fuzzy controllers", *AI Expert*, vol. 6, pp. 26-33.
- Laurent, G. (2000), "Improving the external validity of marketing models: a plea for more qualitative input", *International Journal of Research in Marketing*, vol. 17, pp. 177-182.
- Lee, M. & Turban, E. (2001), "A trust model for consumer Internet shopping", *International Journal of Electronic Commerce*, vol 6 (1), pp. 75-91.
- Lin, J., Lu, H. (2000), "Towards an understanding of the behavioural intention to use a website", *International Journal of Information Management* 20, pp. 197-208.
- Lohse, G., Spiller, P., 1999. Internet retail store design: How the user interface influences traffic and sales. *Journal of Computer-Mediated Communications* 5 (2), <http://www.ascusc.org/jcmc/vol5/issue2/lohse.htm>
- McKnight, H., & Chervany, N. (2002), "What trust means in e-commerce customer relationships: an interdisciplinary conceptual typology", *International Journal of Electronic Commerce*, vol. 6 (2), pp. 35-59.
- Montoya-Weis, M., Voss, G. & Grewall, D. (2003). Determinants of online channel use and overall satisfaction with a relational, multichannel service provider. *Journal of the Academy of Marketing Science* 31 (4), 448-458.
- Oskamp, S. (1991). *Attitudes and opinions. Second Edition*. Prentice Hall, Englewood Cliffs, New Jersey.
- Richards, J.I. (1997). Legal potholes on the information superhighway. *Journal of Public Policy & Marketing* 16 (2), 319-326.
- Roberts, J.H. (2000), "The intersection modelling potential and practice", *International Journal of Research in Marketing*, vol. 17, pp. 127-134.
- Rust, R.T., Kannan, P.K., & Peng, N. (2002). The customer economics of the Internet. *Journal of the Academy of Marketing Science* 30 (4), 455-464.
- Schiffman, L., Kanuk, L., 1997. *Consumer Behavior (Fifth Edition)*. Prentice Hall.
- Shih, C. (1998). Conceptualizing consumer experiences in cyberspace. *European Journal of Marketing* 32 (7/8), 655-663.
- Solomon, M., 1997. *Comportamiento del Consumidor (3ª Edición) (Original version: Consumer Behavior Buying, Having and Being)*. Prentice Hall.

- Steenkamp, J.-B.E.M. (2000) "Introduction to the special issue on marketing modelling on the threshold of the 21<sup>st</sup> century", *International Journal of Research in Marketing* , vol. 17, pp. 99-104.
- Steenkamp, J.-B.E.M., Baumgartner, H. (2000) "On the use of structural equation models for marketing modeling", *International Journal of Research in Marketing* , vol. 17, pp. 195-202.
- Urban, G., Sultan, F., Qualls, W., 2000. Placing trust at the center of your Internet strategy. *Sloan Management Review* (Fall), 39-48.
- Van Bruggen, G.H., Wierenga, B. (2000), "Broadening the perspective on marketing decision models", *International Journal of Research in Marketing*, vol. 17, pp. 159-168.
- Witten, I.H., Frank, E. (2000), *Data mining: practical machine learning tools and techniques with Java implementations*, Morgan Kaufmann Publishers, San Francisco, CA, USA.