College des Economistes de la Sante- Health Economics Study Group Workshop

Paris 14-16 January 2004

Work in progress. Please do not quote or cite without permission

Complexity effects in Stated Preferences Choice Experiments: a simulation exercise.

Mabelle Amaya-Amaya[*], Momodou Kalleh, Mandy Ryan and Maria Odejar

Health Economics Research Unit Institute of Applied Health Sciences

Polwarth Building, Foresterhill, Aberdeen, Scotland

Tel 01224 553733, FAX 01224 550926

[*]Corresponding author: m.amaya@abdn.ac.uk

Abstract

Stated Preferences Choice Experiments have increasingly gained popularity as a benefits valuation tool in Health Economics. They are based on the economic notion of value (utility) maximization whereby individuals are assumed to be able to assign values to all alternatives and choose the highest utility alternative, independent of the choice context and complexity. More behaviorally oriented research on consumer decision-making on the other hand acknowledges that consumers do not always behave in such a perfectly rational manner. In particular, consumers have been found to change their decision-making "heuristics" as choice complexity changes. This paper presents an analysis of complexity effects on stated preference choice tasks using a number of simulated datasets. It is hypothesised that complexity impacts the variance of the utility distributions. As individuals face increasing complexity they will respond with increasing information about their tradeoffs (decreasing variance) but, beyond some point of complexity, greater inconsistency across individuals will be found, and so error variance increases. The two complexity dimensions varied across datasets include the number of alternatives the individual is choosing from and the number of attributes under consideration. A heteroskedastic multinomial logit is used where scale parameters are allowed to vary with choice task complexity, as represented by approximate entropy. Our results provide suggestive evidence of the existence of an "information overload" effect and argue for the development of design principles seeking to maximize the information content of the data to be collected, subject to constraints related to respondents' cognitive abilities and "cognitive budgets".

Acknowledgements.

The first author would like to thank the Chief Scientific Office of the Scottish Executive Health Department for her PhD studentship and the Health Economics Study Group for funding her assistance to this conference. Financial support from the University of Aberdeen and Health Foundation is also acknowledged. HERU is supported by the Chief Scientific Office of the Scottish Executive Health Department. The usual disclaimer applies.

1.Introduction

Recent years have seen an increased interest in the use of Stated Preference Choice Experiments (SPCE) in many contexts in health economics, ranging from eliciting preferences for health care, modelling priority setting choices in clinical services to willingness to pay measurement. During this time there has also been an increase in the number of empirical studies addressing methodological issues in conducting SPCE. (see Ryan and Gerard 2003 for a review of studies).
The most common format of SPCE present respondents with the task of choosing one option from among two or more alternatives, each described in terms of a set of attributes, which are varied according to an experimental design. The modelling of SPCE responses rests upon the traditional assumption of value (utility) maximization used in Economics, in which individuals are assumed to adopt fully rational, fully compensatory optimising behaviour. In every choice situation, individuals are able to compare and evaluate all available alternatives and information in a perfectly costless information processing sense, and then choose the alternative with the highest utility, independent of context, learning, fatigue, etc.

However, the fact that respondents' decision-making and choice behaviour is affected by the complexity of choice task (e.g. number of alternatives and/or number of attributes, time pressure, alternative similarity) as well as the decision environment and person characteristics has been well documented in the literature on Behavioral Decision Theory. Here a large body of experimental and simulation research using process tracing methods (e.g. verbal protocols and/or information boards) has demonstrated that individuals behave as "cognitive misers" and adapt the amount of effort they invest in making a *good* (or accurate) decision to their context and resources (Shugan, 1980; Payne et al, 1993), often adopting simplifying, non-compensatory strategies to make their decision task more manageable (see Ford et al. (1989) and Payne et al (1993) for comprehensive literature reviews of this stream of research). Further, it has been suggested that higher choice complexity may lead to higher propensity to avoid conflict (Hogarth, 1987), or the choice itself by deferring choice, choosing the status quo (Tversky and Shafir, 1992) or even choosing not to choose at all (Dhar, 1999). The implication may be that whilst preferences and utility maximisation may play a role in how decisions are made, they seem to compete with other (possibly non-compensatory) decision protocols for defining and solving the choice task at hand (Loewenstein 2001).

Economists have also considered the potential limitations of individuals' ability to process information and its implications for choice behavior. Simon (1955) first argued that that, due to the bounds of rationality, consumers develop partial information, "satisficing" decision rules to avoid the full cognitive cost of choosing. (see also March, 1978). Heiner (1983) also argued the complexity and uncertainty surrounding a choice situation often leads consumers to adopt simplified strategies. In a more formal examination, De Palma et al. (1994) proposed that individuals with lower "ability to choose" will make more errors in comparisons of marginal utilities.

These findings above may have implications for the development of SPCE. Given that these experiments often involve repeated choices among multiple alternatives described by multiple attributes, they may result in a very complex task for some or all respondents. Furthermore, this cognitive challenge might be increased for SPCE designed so as to maximise *statistical efficiency* (Zwerina et al 1997; Louviere et al, 2000), which may well lower *respondent efficiency* (Severin, 2001). Indeed, research interest on the impact of task complexity (e.g the length of a survey, the levels and ranges of values, the number of choice sets as well as the information available about them and the way in which they may be correlated) on SPCE responses is growing in different applied fields, including marketing (Elrod et al, 1992; Dellaert et al, 1999), transportation (Wang et al, 2001; Saelensminde (2001), Hensher et al 2002, Arentze et al 2003), environmental valuation (Mazzotta and Opaluch 1995, Swait and Adamowicz, 2001a, Hanley et al. 2002, Deshazo and Fermo, 2002; Foster and Mourato (2002)) and health economics (Cairns et al. 2002, Maddala et al 2003, Porter et al (2003); Amaya-Amaya et al. 2003). Empirical evidence from this research suggests that increased choice complexity can affect choice consistency increasing the "noise" (error) of data, hence distorting parameters estimates and welfare measures. Nevertheless this evidence remains limited to the extent that it comes from either a single dataset or a variety of datasets where other confounding factors apart from complexity may be affecting choice consistency.

The purpose of this paper is to investigate this issue of complexity effects on stated preference choice tasks in more detail. To do so, we make use of the model developed by Swait and Adamowicz (2001a) which extends the traditional consumer optimisation problem to recognize the role and impact of decision context as represented by task complexity. This model is applied to a number of simulated datasets in which two dimensions of complexity are purposely varied: the number of alternatives the individual is choosing from and the number of attributes under consideration. The remainder of the paper is organised as follows. Section 2 outlines the theoretical framework and the empirical model for testing the impact of complexity, as developed by Adamowicz and Swait. Section 3 describes the data generation process. In Section 4 the estimation results from both the traditional multinomial logit (MNL) and the model incorporating complexity effects are compared for the different datasets. Finally, section 5 concludes with a discussion of the findings and their potential implications of the results and lines for future research.

2. Incorporating complexity effects in choice models

2.1. Traditional framework

The traditional discrete choice problem formulation is shown in (1) (Hanemann, 1984).

$$\begin{aligned} & MaxU(y_1,...y_J,x_1,...,x_J,z) \\ & s.t.\sum_{i=1}^{J} p_i y_i + z \le M \\ & \quad y_i y_j = 0 \ \forall i \ne j \\ & \quad y_i = y_i^* \forall i \end{aligned} \tag{1}$$

Here it is assumed that the consumer maximises utility *U* by choosing among is a discrete set $y_1,...,y_J$ of "goods", characterised by a vector of attributes $x_1,...,x_J$ and corresponding prices $p_1,...,p_J$, and a numeraire good *z,* subject to a budget constraint $\sum_{i=1}^{J} p_i y_i + z \le M$ where *M* is income. Further constraints ensure that one alternative is chosen ($y_i y_j = 0 \ \forall i \ne j$) and mutual exclusivity of the alternatives ($y_i = y_i^* \forall i$ with y_i^* the optimal quantity of *i)*

Within the traditional framework, the standard tool for analysing responses to SPCE is McFadden's (1974) discrete choice model based on the random utility hypothesis. The idea behind Random Utility Models (RUM) is that researchers cannot observe all factors affecting preferences (represented by utilities). Therefore, the latent utility U_{in} that an individual *n* associates with alternative *i* is considered decomposable in two additively separable parts, namely a systematic (or explainable) component, V_{in}, and a random (or unexplainable) component ε_{in} representing unmeasured variation in preferences

$$U_{in} = V_{in} + \varepsilon_{in} \tag{2}$$

Given the random component is inherently stochastic, the probability $P_n(i \setminus k)$ that an individual *n* will choose an alternative *i* from a choice set *k* of J total alternatives available is estimated. Assuming utility maximizing behaviour, this can be expressed as follows:

$$P_n(i \setminus k) = \Pr[(V_{in} + \varepsilon_{in}) > (V_{2n} + \varepsilon_{2n}) > ... > (V_{Jn} + \varepsilon_{Jn})] \tag{3}$$

Solutions to equation 3 depend on assumptions about the form of systematic utilities V_{in} and distributional and statistical properties of ε_{in}. For a traditional linear-index model (i.e. $V_{in} = \beta' X_{ink}$) and assuming independent and identically distributed Extreme Value Type I disturbances[1], the familiar Multinomial Logit (MNL) model is derived (Ben-Akiva et al, 1985). That is

$$P_n(i \setminus k) = \exp(\mu\beta' X_{ink}) / \sum_{j=1}^{J} \exp(\mu\beta' X_{jnk}) \tag{4}$$

where X_{jnk} is a vector of attribute levels corresponding to an alternative j ($j = 1,2,...,i,J$) in the choice set k, β' is the vector of indirect marginal utilities, and μ is a scale factor inversely related to the variance σ^2 of the disturbances ε_{in} ($\mu = \pi / \sqrt{6}\sigma$). In any one data-set, the scale factor μ cannot be uniquely identified; hence, means and variances are perfectly confounded and only the combined effect of μ and β can be estimated (Swait and Louviere, 1993). However, the assumption of identically distributed disturbances allows the utility function to be scaled by an arbitrary factor without affecting the choice probabilities. Thus the scale parameter, μ, typically is set by the analyst to a value that provides a convenient estimate of the standard deviation of disturbances σ, usually one (Ben-Akiva and Lerman, 1985).

2.2 Extending the traditional framework.

This paper adopts an extended formulation of the traditional discrete choice utility maximisation problem proposed by Swait and Adamowicz (2001a). Given individuals' scarce information-processing resources, an individual's ability to choose the utility maximizing alternative depends on the level of effort applied and the complexity of the choice task at hand. Equation 1 is extended to become

$$\begin{aligned}
& Max U(y_{11},...y_{1J}, x_{11},...,x_{J1}, E_1C_1;...;y_{1K}...y_{1J}, x_{1K},...,x_{JK}, E_KC_K; z, E_{K+1}C_{K+1}) \\
& s.t. \sum_{k=1}^{K}\sum_{i=1}^{J} p_{ik}y_{ik} + z \le M \\
& \sum_{k+1}^{K+1} E_k \le B \\
& y_{ik}y_{jk} = 0 \ \forall i \ne j \\
& y_{ik} = y_{ik}^{*} \forall i
\end{aligned} \tag{5}$$

---

[1] That is, ε_{in} for all i,n is distributed as

$F(\varepsilon) = \exp[-e^{-\mu(\varepsilon-\eta)}], \ \ \mu > 0$

$f(\varepsilon) = \mu e^{-\mu(\varepsilon-\eta)} \exp[-e^{-\mu(\varepsilon-\eta)}]$

where η is a location parameter and μ is a strictly positive scale parameter

Where E_k represents the unobservable effort needed to differentiate between alternatives in choice set k (k = 1,…,K) and choose the utility-maximizing one amongst them, C_k is the eventually measurable complexity of the choice task in choice set k, and B is the unobservable effort budget allocated over all K choice tasks. The total effort will be distributed amongst choice tasks depending on the utility associated with them: higher expected utility choices will attract more effort. Effort will also be influenced by the complexity associated with the choice task. For any given utility level, a complex choice will result in lower expected returns, hence attract a lower effort. Simply put, a complex choice task will be worth the necessary high effort if it concerns an important good (yielding high utility), but not if it concerns a less important good (yielding low utility).

Whilst this theoretical framework suggests that both effort and complexity should be incorporated within the consumer choice problem, the empirical model presented below suppresses the unobservable effort element for simplicity reasons. As with Swait and Adamowicz (2001a), this paper focuses on the response of consumers to different levels of complexity, in terms of their apparent ability to choose, which is endogenously determined and influenced by the complexity of the choice task.

Swait and Adamowicz (2001a) demonstrated that by making different assumptions about the distribution of the error term within a random utility framework, namely still independent but <u>not identically</u> distributed Extreme Value I errors, the augmented direct utility function in their extended formulation will translate into MNL-like probabilities. However, the scale parameter is no longer constant, but is a function of choice complexity C_n

$$P(i \setminus K) = \exp(\mu_n(C_n)(\beta' X_{iK})) / \sum_{i=1}^{J} \exp(\mu_n(C_n)(\beta' X_{iK})) \qquad (6)$$

It is hypothesised that differences in complexity generate differential consistency levels in preferences across individuals. This will be reflected in equation (6) only by influencing the variances of the assumed distribution for the disturbances. As in de Palma et al. (1994), this scale factor is interpreted as a representation of a respondent's ability to choose. As the scale will vary amongst individuals, the model is heteroskedastic in nature.

To estimate this model we need a measure of complexity. Following Swait and Adamowicz (2001a), we use an overall summary measure, known as "entropy", as a formal parsimonious representation of choice complexity C_n. In communication theory, entropy has been used to quantify the amount of information in an experiment (Soofi, 1994; Shannon, 1948)[2]. Given a set of outcomes (or alternatives in this case), $\left\{x_j, j = 1,...,J\right\}$, that are described by a probability distribution $\pi(x)$, the entropy (or uncertainty) of the choice situation is defined as:

$$H(X) = H(\pi_x) = -\sum_{j=1}^{J} \pi(x_j) \log \pi(x_j) \geq 0 \tag{7}$$

To operationalize entropy as a complexity measure, the probability of selection of the alternatives $\pi(x)$ is required. Swait and Adamowicz (2001a) constructed an a priori estimate of this from a MNL with the following form:

$$\tilde{\pi}(X) = \frac{\exp(\varpi X_i)}{\sum_{j \in k} \exp(\varpi X_j)} \tag{8}$$

where ϖ is a vector of unknown attribute weights. In the absence of knowledge about the true parameters, a signed "flat prior" over attribute levels may be used, in which weights are given the expected sign by theory or the analyst experience (e.g. high price is worse that low or high quality care is more attractive than low quality) and all main attribute effects are given equal weights. This assumption is equivalent to the Majority of Confirming Dimensions (MCD) heuristic (Russo and Dosher, 1983). The resulting approximate level of uncertainty about a choice set (entropy) will therefore be given by:

$$\tilde{H} = -\sum_{j=1}^{K} \tilde{\pi}(x_j) \log \tilde{\pi}(x_j) \tag{9}$$

This empirical measure of entropy increases with the complexity of a SPCE, as measured by factors such as number of attributes, number of replications, similarity between alternatives, familiarity with the choice situation. This is shown in Figures 1-2.[3] For example, in a SPCE with J alternatives in a choice set, entropy reaches its maximum if each of the alternatives are equally likely (indicating that alternatives are very similar hence more difficult choices). This is clearly seen in Figure 1 for the case of 2 alternatives (J=2). Entropy reaches its maximum when all the alternatives in the choice set with are equally likely (indicating that alternatives are very similar hence more difficult choices) and it is minimised at zero when one

[2] It should be noted that the concept of entropy has been used in many contexts in Economics, for example in information and entropy econometrics (Golan, 2002) or as a descriptor of biodiversity (Simpson, 2002).

alternative has a probability of one (so is dominant, and therefore seen to be a relatively easy choice) and the others have probabilities of zero. If the number of equally likely alternatives (J) increases, entropy also increases.

Further, as the number of attributes included in the systematic part of the utility function (V) increases, the variance of disturbances (ε) would be expected to decrease. Given the inverse relationship between variance and scale, this will result in an increase in the scale μ and parameter estimates ($\mu\beta$). Thus as the number of attributes increases, the estimated attribute weights ϖ will increase, affecting the estimated selection probabilities $\tilde{\pi}(x)$ and hence $\tilde{H}$. As seen in Figure 2, an increase in the number of attributes (i.e. the scale factor) will cause the symmetric shape of $\tilde{H}$ to get tighter and tighter around the indifference point (V1-V2=0) of maximum complexity (entropy). A similar behaviour would be expected for other factors potentially affecting error variances such the number of replications in a SPCE, similarity between alternatives or familiarity with the choice.

Finally the scale factor is defined as an exponential function, which ensures non-negativity and results in a highly non-linear-in-parameters model, yet one with excellent convergence properties (Swait and Adamowicz, 2001a):

$$\mu_n(C_n) = \exp(\theta_1 \tilde{H}_n + \theta_2 \tilde{H}_n^2) \tag{10}$$

The quadratic form in (10) allows testing for a hypothesised U-shape relationship between complexity and decision effectiveness due to an information overload effect (Keller and Staelin, 1987). That is, with increasing complexity, consumers apply more effort to making decisions, leading to a greater degree of preference consistency, (lower variance) up to a certain point of complexity. After this they resort to a plethora of simplifying decision heuristics that generate greater unobserved variability in responses (see also Jacoby et al, 1974). If the kind of information overload effect suggested by Keller and Staelin (1987) is found empirically, we would expect that $\theta_1 \leq 0$ and $\theta_2 \geq 0$. A reasonable alternative hypothesis would be that $\theta_1 \geq 0$ and $\theta_2 \geq 0$, which implies that error variances are constantly decreasing in complexity. This would lend empirical support to Bettman et al's (1998) suggestion that for more difficult choices respondents' processing effort may be greater, leading to greater preference consistency hence to richer information on preferences

[3] More information on the plotting of these graphs is available from the authors upon request.

Figure 1:Behaviour of the entropy measure as a function of choice probability and number of alternatives

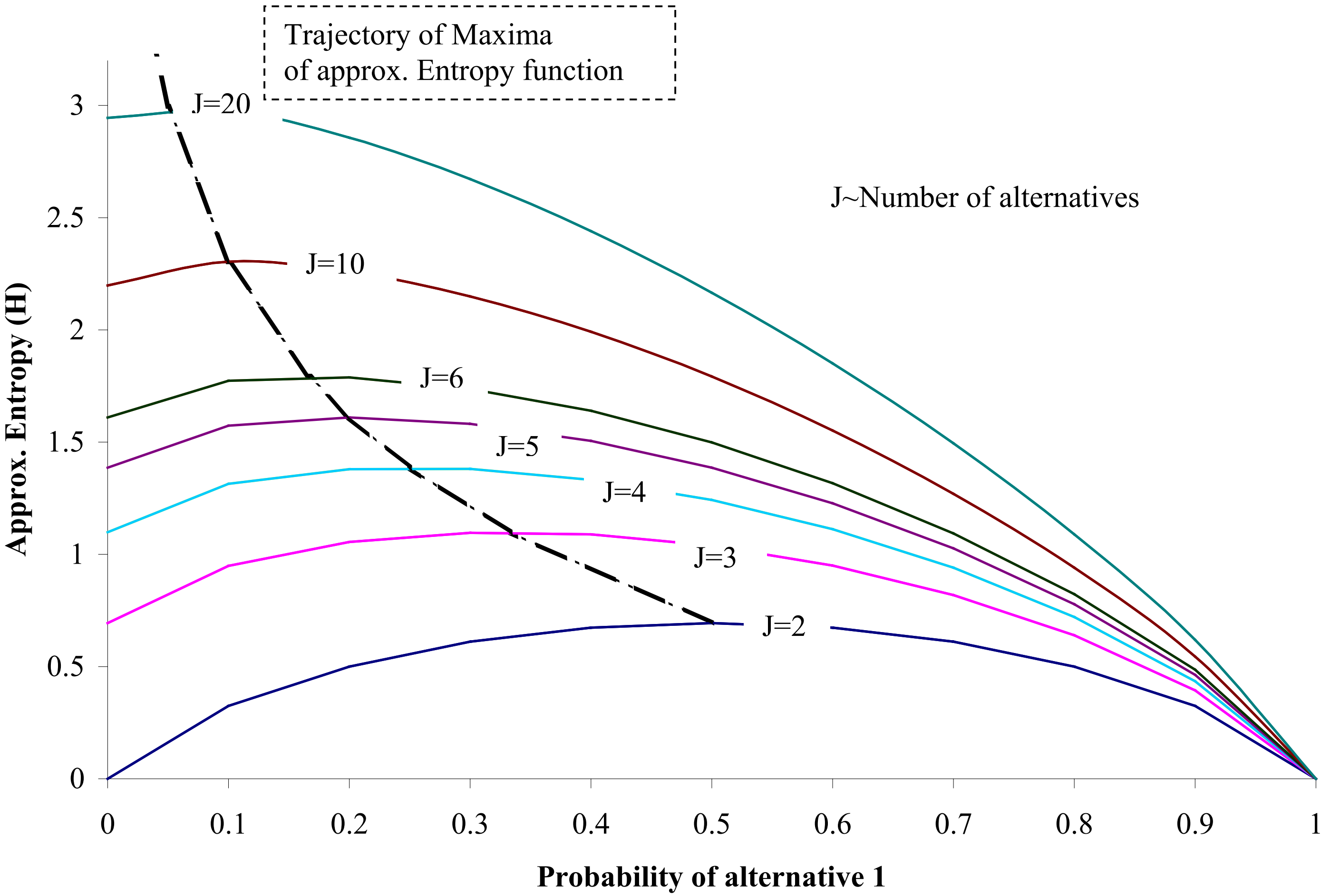
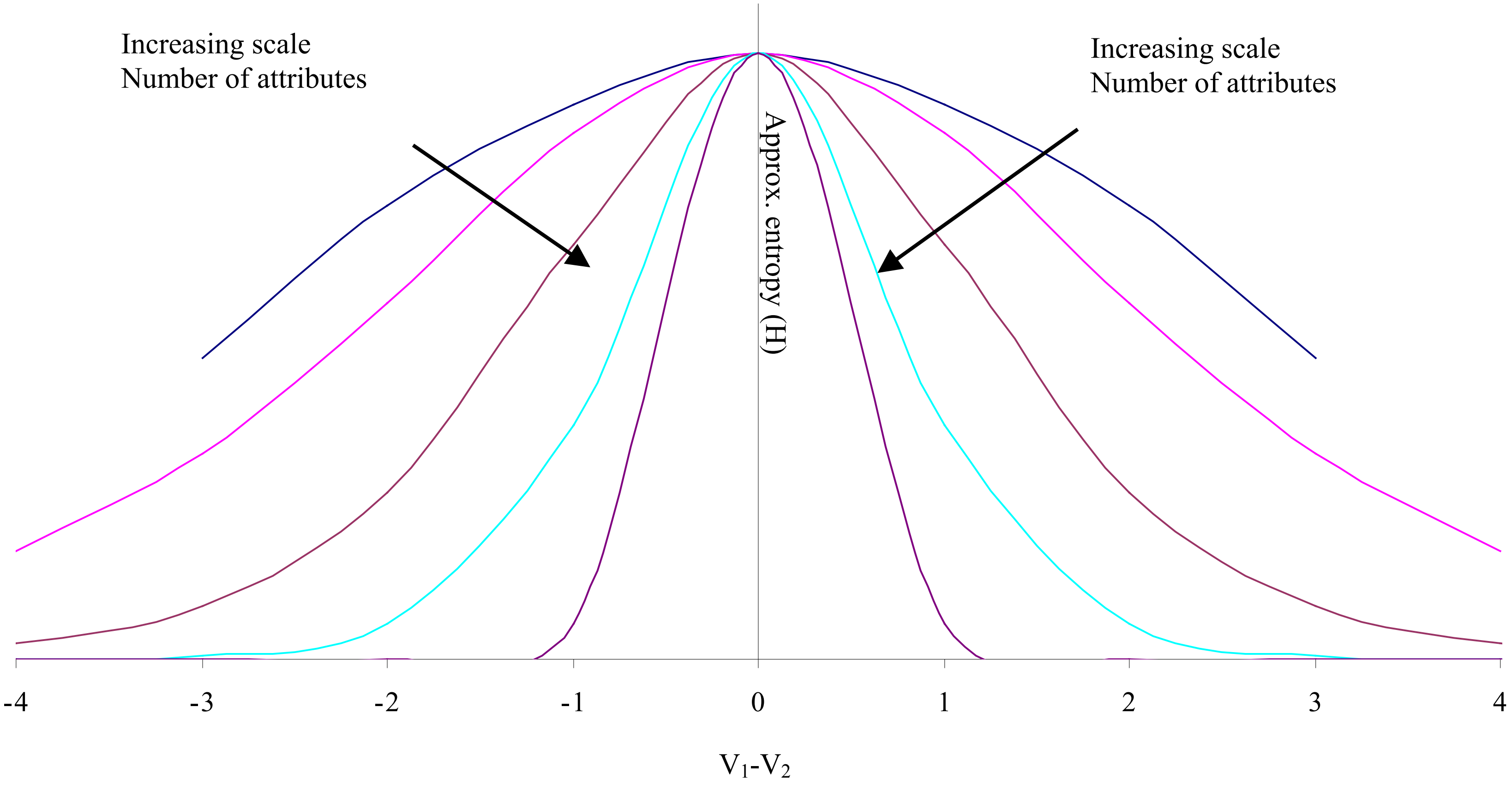


Figure 2: The impact of increasing scale (e.g. number of attributes) in the measure of entropy



3. Data Generation Process

To empirically investigate the relationship between choice complexity and variance (scale factor), thus the impact of complexity on utility estimates, we generated several datasets from simulated responses of 50 hypothetical individuals to a total of 10 experimental designs. In these experiments two dimensions of complexity were varied: the number of alternatives the individual is choosing from and the number of attributes under consideration (see Table 1). In experiment 1 to 4, each choice set had 2 alternatives and the number of attributes defining each alternative varies from 2 to 5. In experiment 5 to 7, the number of alternatives was increased to 3 and attributes vary from 2 to 4. Finally experiments 8 to 10 include 4 alternatives described by 2 to 4 attributes respectively. Generally as these factors increase so to would task complexity. For simplicity reasons only main effects were considered in all experiments and the number of levels is kept constant and equal to 2 across datasets. All attributes were considered "discrete" or qualitative.

The experimental designs were generated as follows. Using the software Stated Preference Experiment Editor Designer (SPEED Version 2.1) and based on catalogues (see e.g. Kocur et al (1982)) we obtained as many different fractions of the full factorial design (profiles) as alternatives compounded the choice set in each experiment (2, 3 or 4). The profiles in these fractions were then randomly grouped together. It should be noted that given the number of alternatives and attributes included in the different experimental designs, these result in a different number of questions each of the 50 hypothetical individuals would have to answer if faced with the questionnaire. For example, while a participant in experiment 1 would have to answer 6 questions, a person involved in experiment 4 would have to face up to 64 questions. Given we selected a fixed number of 50 hypothetical respondents to take part in each of the experiments, this made it impossible to create uniform sized datasets for each experimental design scenarios (see column 4 in Table 1). However, arguably, the greater the number of choices an individual is asked to answer, the more complex the task and the higher the likelihood of individuals becoming fatigue or bored increasing the number of choice inconsistencies. Thus a priori a significant impact of complexity on error variances would be expected more likely to occur the greater the number of total observations in the sample.

Columns 2 to 4 in Table 1 summarise the characteristics of each experimental setting. Column 5 indicates the maximum level entropy that can be achieved in each dataset; based on the number of alternatives J in the choice problem (lnJ). We would expect that the higher the maximum level of entropy, the more complex (uncertain) the dataset. The range of maximum entropy across datasets is [0.69-1.386].

Given the alternatives and attributes in the choice set, 50 hypothetical respondents must determine which of the options they deemed most desirable. To determine the desiderability of each alternative, the hypothetical respondents' true utility function was generated. For each attribute, effect codes rather than dummy variables were used. Thus in each case a dichotomous variable was created which took the value of 1 if the attribute took the second level and –1 otherwise. The utility function of each individual is given by equation 11

$$U_{ij} = \sum_{j=1}^{J} \alpha_{ij} + \beta_{1a} x_{1a} + \beta_{2a} x_{2a} + \ldots + \beta_{Ma} x_{Ma} + \varepsilon_{ij} \tag{11}$$

Where α_{ij} are "true" alternative specific constants, $x_{1a} \ldots x_{Ma}$ are the dummy variables corresponding to the levels in attributes 1 to M respectively and β_k are "true" utility parameters, which were assumed arbitrarily equal to 1 or –1 for all attributes.

Extreme Value Type I (EV-I) random errors were obtained by inverting the cumulative density function of the EV-I distribution (see footnote 1) at different draws of a standard uniform distribution. Thus, for each observation, an independent EV-I random error ε_{ij} was drawn by applying the following transformation, *-ln(-ln(u))* where *u* is a random number generated from a uniform distribution between 0 and 1 (Train, 2002).

The deterministic and error term are added to get the generated "true" utility derived by any hypothetical respondent from each alternative. Finally the dependent variable was created by assigning a value of 1 to the alternative that had the highest utility and a value of zero to the remaining less desirable alternatives in the choice set.

4. Estimation

4.1. Methods

For each dataset, the classical (homocedastic) MNL model (expression 4) and the alternative heterocedastic MNL (HMNL) (expression 6), allowing the scale factor μ to vary according to the complexity of the decision context, were estimated by full information maximum likelihood (FIML), using specialized code in Gauss 4 (2002). For the MNL we repeated the estimation 200 times and then averaged the result. This mean vector was used as starting values for estimation the HMNL. In this case, the log-likelihood of the model is simultaneously maximised with respect to the utility parameters, β and the θ parameters in the scale factor. The resulting log-likelihood of the model is:

$$LogLikelihood = \sum_{K}\sum_{i} d_{iK} \ln(P_{iK}) \qquad (11)$$

where $P_{iK} = \exp^{\exp(\theta_1^{'}\tilde{H}_n + \theta_2^{'}\tilde{H}_n^2)V_{iK}} / \sum_{j=1}^{J} \exp^{\exp(\theta_1^{'}\tilde{H}_n + \theta_2^{'}\tilde{H}_n^2)V_{jK}}$ is the probability of an individual n choosing alternative i in choice set K, and $V_{iK} = \sum_{m} \beta_m x_{mjK}$ is the systematic portion of the indirect utility function corresponding to alternative i in choice set K described by m attributes.

The hypothesis of interest is that both the linear and quadratic terms in the scale function in the HMNL are simultaneously equal to zero (i.e H_0: $\theta_1 = \theta_2 = 0$), meaning insignificant complexity effects. In this case, the scale parameter could be considered constant, hence the HMNL collapses into the (homocedastic) MNL model. A likelihood ratio test with a test statistic –2[LL(r) – LL(ur)], where LL(r) and LL(ur) are the log-likelihood for the MNL (restricted) and the HMNL (unrestricted) models respectively, was used to compare these two nested models. This statistic is Chi-squared distributed with 2 degrees of freedom and the critical values of $\chi^2_{(d.f=2,\alpha\%)}$ at the 95% and 99% level of significant are 5.99 and 9.21 and respectively.

4.2. Results

Columns 6-10 in Table 1 summarise the estimation results[4] - columns 6-7 show the log likelihood values for the MNL and the HMNL respectively, column 8 the chi-squared statistic for the null hypothesis that scale parameter is constant so the MNL applies, and columns 9-10 the estimated value of the parameters for the scale as a function of complexity and corresponding p-values. At the 95% confidence level the null hypothesis of homocedastic preferences is highly rejected for all experiments, exception made number 7, where complexity appears not to affect choice consistency (variance)[5]. Also, as expected the greater the number of attributes and alternatives included in the experimental design and the greater the number of questions individuals have to answer, the more significant the impact of complexity in error variances.

Consider first the experiments with 2 alternatives (1 to 4), with a theoretical entropy range of [0, 0.693]. For these experiments the value of the χ^2 statistics are greater than the critical value at the 5% level (5.991) for experiments 1 and 2 and even greater than that at 1% (9.21) for experiments 3 and 4. Thus the heterocedastic model is strongly favoured statistically in all four cases. However, taken individually the estimated parameters for the scale factors in such model of HMNL are not significant in all cases at the traditional levels. Moreover, they do

[4] Additional parameter estimates are available form the authors upon request.

not show the expected sign for an U-shaped relationship complexity-variance ($\theta_1 \leq 0$ and $\theta_2 \geq 0$), except for experiment 4. Interestingly, whilst this result could be justified, as this latter experiment is the more complex, in fact in this case the coefficients are highly insignificant. Thus overall the idea of an information overload effect is not supported in this group of experiments.

Moving now to the results for experiments with 3 alternatives (5 to 7), we find again highly significant complexity effects for experiment 5 and 6 (χ^2 statistics significant at 1% level). The result in experiment 7 runs counter to our initial idea that increases in the number of attributes will result in an increase in the "noise" in the data. However in this case the null hypothesis cannot be rejected, thus the MNL is favoured. Similar to the results in experiment 4, the estimated scale parameters in experiments 5 and 6 have signs supporting the idea of an information overload effect, i.e. an inverted U-shaped relationship complexity variance. Nonetheless, these are significant only for experiment 5 but not for experiment 6. It is interesting to observe that in experiment 7, despite the hypothesis of homocedastic preferences cannot be rejected; the estimated individual coefficients are statistically significant at 90% confidence level.

The final set of experiments (8, 9 and 10) has the highest number of alternatives and as indicated by theoretical entropy measure, these experiments are considered to be more complex. The estimation results for all these three experiments strongly reject the MNL specification indicating heteroscedastic preferences. It is worth noting that both for experiments 9 and 10 the coefficients on the scale function are statistically significant at 5% significant level and have the expected sign supporting an inverse U-shaped relationship between complexity (entropy) and the estimated error variances.

It is worth noting experiment 8 displays significant constantly decreasing (increasing) patterns of variance (scale) with increasing complexity. This may indicate support for an interpretation in harmony with Bettman et al's (1998) suggestion that more difficult choices may lead to richer information on preferences as respondent processing effort increases with complexity.

[5] We use White's method to robustly estimate our covariance matrix.

Table 1 Characteristics of Experimental Data set and the Estimated Result of the MNL and the HMNL

Experiment	No of Alternatives	No of attributes	No of Observations[a]	Maximum entropy	LogL-MNL	LogL -HMNL	χ^2 for Ho	θ_1 (p-value)	θ_2 (p-value)
1	2	2	300	0.693147	-110.700	-107.176	7.049*	10.8309 (0.2434)	-10.7946 (0.2285)
2	2	3	800	0.693147	-309.308	-304.72	9.176**	10.0900 (0.1156)	-9.7990 (0.1115)
3	2	4	1600	0.693147	-562.122	-542.488	39.268**	3.7282 (0.1044)	3.39671 (0.136)
4	2	5	3200	0.693147	-1030.265	-1017.408	25.719**	-5.1623 (0.4906)	4.9342 (0.4860)
5	3	2	600	1.098612	-351.173	-323.651	55.044**	-14.7717 (0.006)**	8.7704 (0.0041)**
6	3	3	1200	1.098612	-786.108	-728.3436	115.528**	-2.3153 (0.5270)	1.0891 (0.6033)
7	3	4	2400	1.098612	-1423.466	-1421.532	3.868	4.2213 (0.0730)	-2.2575 (0.0940)
8	4	2	800	1.386294	-812.055	-782.612	58.886**	5.1047 (0.0141)	2.6791 (0.0072)**
9	4	3	1600	1.386294	-1417.135	-1387.8768	58.516**	-9.7372 (0.0023)**	4.1946 (0.0025)**
10	4	4	3200	1.386294	2328.606	2266.906	123.401**	-1.3726 (0.0265)*	5.499 (-0.0681)

H_0: $\theta_1 = \theta_2 = 0$ Critical Value chi-square at 95% and 99% confidence level are 5.991 and 9.21 respectively
* significant at 5% ** highly significant at 1%
a. This results from multiplying the number of hypothetical respondents (50) by the number of questions obtained for each experimental design

5. Tentative conclusions and further research.

Despite the growing popularity of Stated Preference Choice Experiments (SPCE) for valuing non-market goods and services such as publicly provided health care interventions, a number of issues remain unresolved. A crucial one relates to the extent to which consumer choice behaviour is affected by context and the complexity of choice environment, as advanced by behavioral researchers. Empirical evidence is growing in different fields including environmental valuation and health economics. This paper attempted to explore this issue in more detail using a total of 10 simulated datasets where two dimensions of complexity were purposely varied: the number of alternatives the individual is choosing from and the number of attributes under consideration. Swait and Adamowicz's (2001a) model incorporating complexity effects was used to test the hypothesis that complexity impacts the variance of the utility distributions. Scale parameters are allowed to vary with choice task complexity, as represented by approximate entropy. In line with previous research (e.g. Swait and Adamowicz (2001a); Amaya-Amaya et a (2003)), in half the experiments we found an inverted U-shaped relationship complexity –variance, supporting the idea of an information overload effect. We may then be tempted to conclude that in SPCE as individuals face increasing complexity they respond with increasing information about their tradeoffs (decreasing variance) but, beyond some point of complexity, greater inconsistency across individuals is found, and so error variance increases. Further, we found for one of the experiments a significant inverse relationship complexity-variance, which may suggest that respondents' effort may be greater for more complex decisions, leading to less choice inconsistencies (i.e. less variance). Furthermore, we found strong empirical support for the idea that heterokedasticity is an important factor to consider when modelling choice behaviour. In all but one experiment the null hypothesis of homocedastic preferences is strongly rejected. A well-known result is that ignorance of the problem of heterokedasticity in probit/logit models will result not just in a loss of efficiency, as is the case in the general linear model, but also in bias (Yatchew and Griliches, 1984). Hence, the implications of our findings for choice modelling practice are potentially serious (see Munizaga et al. (2000) for representation of heteroskedasticity in discrete choice models)

We recognise the limitations bearing on the interpretation the results (Swait and Adamowicz, 2001a). First, the sensitivity of results to the empirical calculation of entropy should be investigated. Swait and Adamowicz (2001b) argue that the "flat prior" approximates the true level of information uncertainty, providing an index that characterises task demands on respondents. However, they recognise that this "may be not always be effective". Future work should explore this issue in more depth by testing more informative priors, for example, "borrowing" attribute weights from another studies or using estimates from pilot data.

Second, the fact that task demands are partly defined by the current choice set and partly by the prior effort expended was ignored. The possibility of including "cumulative entropy" in the scale function as an approximation of the "cumulative cognitive burden" experienced by an individual up to a determined choice set in the experiment should be explored.

Third, in this paper we have considered entropy as an overall measure of complexity. A very interesting possibility in the future would be to explore defining the scale as a function of the basic elements of complexity, e.g. number of attributes and/or number of alternatives, rather than using an overall measure (Deshazo and Fermo (2001))

Fourth, we have not investigated the effect of potential indicators of different processing capabilities across individuals (e.g. level of expertise or socio-demographics). Nor have we explored the possibility that some (or all) the effect being attributed to scale might actually be due to taste heterogeneity. These issues are also left for future research.

Fifth, we continue to assume that respondents always adopt a compensatory decision process (as embodied in the linear specification of the utility of choosing each alternative). Consideration is starting to be given to choice models recognizing the possibility that individuals may adopt a number of different decision strategies as a function (Horowitz and Louviere, 1995; Swait and Adamowicz, 2001b, Swait, 2001, Amaya-Amaya and Ryan, 2003).

Finally, the focus on this paper was on the relative performance of the alternative specification of the choice model hence parameter estimates were not reported. Nevertheless it would be very appealing to perform a montecarlo simulation to investigate how parameter estimates deviate from the true values in the different experiments. Given the estimated parameters are actually estimates of the scaled true parameter values, we would expect the average absolute deviation of estimates from true parameter values to increase as complexity, hence variance, increases.

The implications of complexity effects for design strategies also deserve further research efforts. To date there is little empirical evidence available to guide researchers wishing to design not "overly complex" SPCE. Swait and Adamowicz (2001a) advocated the need to develop design principles that seeking to maximize the signal-to-noise ratio (i.e. information content) of the data to be collected, subject to constraints related to respondents' cognitive abilities and "cognitive budgets". Bayesian experimental design may prove fruitful for such endeavours (Sebastiani and Wynn 2001).

To sum up, complexity issues, and context effects more generally (Swait et al, 2002) open a Pandora's Box of challenging topics for future research to improve our understanding of individuals' preferences for health care for better resource allocation decisions.

References

Amaya-Amaya M, Ryan M, An exploratory analysis of the impact of complexity in stated preference choice experiments in health care, Paper presented at Scottish Economic Society Annual Conference, University of Glasgow, March 2003.

Amaya-amaya, M and Ryan, M, Decision Strategy Switching in choice experiments: preliminary evidence from health care. Paper presented at the 2nd international workshop in discrete choice experiments, Oxford, September 2003.

Arentze T, Borgers A, Timmermans H, and DelMistro R. (2003) Transport stated choice responses: effect of task complexity, presentation format and literacy, *Transport research part E* 39: 229-2244

Ben-Akiva, M. and Lerman S. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press: Cambridge, MA

Bradley, M. (1991) *Users manual for SPEED version 2.1*. Hague Consulting Group, Hague.

Cairns, J. M. van der Pol and Lloyd A. J. (2002) Decision making heuristics and the elicitation of preferences: being fast and frugal about the future, *Health Economics,* 11: 655-658.

De Palma, A. G.M. Myers and Papageorgiou Y.Y. (1994) Rational Choice Under an Imperfect Ability to Choose. *American Economic Review*, 84:419-440.

Dellaert, B., J.D. Brazell and J. J. Louviere (1999), "The Effect ofAttribute Variation on Consumer Choice Consistency," Marketing Letters, 10, 2,139-147.

Deshazo, J.R. and Fermo, G. (2002) Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency, *Journal of Environmental Economics and Management,* 44(1): 123-143.

Dhar, R "Choice Deferral," *In The Elgar Companion to Consumer Research and Economic Psychology* (ed. P. Earl and S. Kemp), 1999.

Elrod, T, J.J. Louviere and K.S. Davey, An empirical comparison of rating-based and choice-based conjoint models, *Journal of Marketing Research*, 1992, 29, 368-377.

Ford, J.K; N. Schmitt; S.L. Schechtman, Hults, B.M. and Doherty M.L (1989) Process tracing methods: contributions, problems and neglected research questions. Organizational behaviour and human decision process 43, 75-117

Foster, V. and Mourato, S, 2002, Testing for consistency in contingent raking experiements. *Journal of Environmental Economics and Management,* 44: 302-328.

Gauss V.4.0, Aptech Systems (2002).

Golan, A (ed.), 2002, Information and Entropy Econometrics, *Journal of Econometrics,* 107

Hanemann, W. (1984) Welfare evaluations in contingent valuation experiments with discrete responses: reply, *American Journal of Agricultural Economics*, 69, 332-41.

Hanley, N., G. Koop and Wright R.E (2002) Modelling Recreational Demand Using Choice Experiments: Rock Climbing in Scotland. *Environmental and Resource Economics*, vol. 21, no. 3, pp. 449-466

Heiner, R.A. (1983). The Origin of Predictable Behavior. *American Economic Review*. 73:560-595.

Hensher, D.A., P.S. Stopher and J.J. Louviere, 2001, An exploratory analysis of the number of choice sets in designed choice experiments: an airline choice application. Journal of *Air Transport Management*, 7: 373-379.

Hogarth, Robin M. *Judgement and Choice: The Psychology of Decision* (2nd edition), 1987, New York: Wiley

Jacoby, J., D.E. Speller and C.K. Berning, , 1974, "Brand choice behavior as a function of information load," *Journal of Marketing Research*, 11, 63-69.

Johnson, E., Payne, J. (1985) Effort and Accuracy in Choice, *Management Science*, 31:395-414.

Keller, K.L., and R. Staelin. (1987) Effects of Quality and Quantity of Information on Decision and Effectiveness. *Journal of Consumer Research*, 14:200-213.

Kocur, G., Adler, T., Hyman,W. and Aunet B(1982) Guide to forecasting travel demand with direct utility assessment. United Stated Department of Transportation, Urban Mass Transportation Administration, Report UMTA-NH-11-0001-82-1, Washington DC.

Loewenstein, G. 2001, The creative deconstruction of decision research, *Journal of consumer research*, 28, 499-505.

Louviere, J.J. (2001) What if consumer experiments impact variances as well as means? Response variability as a behavioral phenomenon, *Journal of Consumer Research*, 28, 506-511.

Louviere, J.J., D. Hensher and J. Swait (2000) *Stated choice methods. Analysis and Applications*, Cambridge University Press.

Maddala, T., K.A. Phillips and F. R. Johnson, An experiment on simplifying conjoint analysis designs for measuring preferences, *Health Economics, forthcoming.*

March, J.G. (1978) "Bounded rationality, ambiguity and the engineering of choice", *Bell journal of Economics*, 9:587-609

Mazzotta, M. and J. Opaluch Decision Making When Choices Are Complex: A Test of Heiner's Hypothesis. *Land Economics*. 1995; 71(4):500-515.
McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behaviour. In: P. Zarembka (ed.), *Frontiers in Econometric*s. New York: Academic Press
Munizaga, M.A.; Heydecker, B.G; and Ortuzar, J.D. (2000) Representation of heteroskedasticity in discrete choice models. *Transportation Research Part B* 34: 219-240
Payne, J.W., Bettman, J.R., and Johnson, E.J. *The Adaptive Decision Maker*, New York:Cambridge, 1993.
Porter K, Ryan M, and Amaya-Amaya M. Is one reason enough? Comparing compensatory and non-compensatory decision-making strategies in discrete choice experiments. Paper presented to *Health Economic Study Group meeting*, University of Leeds, January 2003.
Russo, J.E. and Dosher, B. Strategies for multiattribute binary choice, *Journal of experimental psychology,* 1983
Ryan M. and Gerard, K., 2003, Using discrete choice experiments to value health care: current practice and future prospects. *Applied Health Economics and Policy Analysis*
Saelensminde, K (2001). Inconsistent choices in stated choice data, *Transportation*, 2001 28:269-296.
Sebastiani, P. and Wynn, H.P. Experimental design to maximize information. Paper presented at MaxEnt 2000Twentieth International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering CNRS, Gif sur Yvette, France, July 8-13, 2000
Severin, V. Comparing statistical and respondent efficiency in choice experiments, 2001, Unpublished PhD dissertation, Department of marketing, University of Sydney.
Shannon, C.E. A Mathematical Theory of Communication. *Bell System Technical Journal*. 1948; 27:379-423.
Shugan, S.M, 1980, The cost of thinking, *Journal of consumer research,* 7(2): 99-111.
Simon, H.A. 1955, Behavioral Model of Rational Choice. *Quarterly Journal of Economics*; 69:99-118.
Simpson, R.D. 2002, Definitions of Biodiversity and Measures of Its Value, Resources for the future, discussion paper 02-62.
Download: www.rff.org/disc_papers/PDF_files/0262.pdf
Soofi, E.S. (1994). Capturing the Intangible Concept of Information. *Journal of the American Statistical Association*, 89:1243-1254.
Swait, J. and J. Louviere (1993). The Role of the Scale Parameter in The Estimation and Use of Multinomial Logit Models, *Journal of Marketing Research*, 30:305-314
Swait J.(2001) A non-compensatory choice model incorporating attribute cut-offs. *Transportation Research* 2001; 35:903-928.
Swait, J. and Adamowicz, W. 2001a, Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice, *Organizational behaviour and human decision processes*, 86(2):141–167.
Swait, J. and Adamowicz, W. 2001b, The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching. *Journal of Consumer Research* 28: 135-148
Swait, J., Adamovicz, W., Hanemann, M. et al. Context Dependence and aggregation in disaggregate choice analysis, *Marketing Letters*, 2002; 13(3): 195–205
Tversky, A., and E. Shafir 1992. The disjunction effect in choice under uncertainty. Psychological Science, 3, 5, 305-309.
Wang, D, J. Li and H. Timmermans, (2001) "Reducing Respondent Burden, Information Processing Effort and Incomprehensibility in Stated Preference Surveys: Principles and Properties of the Pairwise Design Strategy", Transportation Research Record, 1768, pp. 71-78.
Yatchew, A. and Z. Griliches. 1984. Specification Error in Probit Models. Review of Economics and Statistics. 66:134-139.
Zwerina, K., J. Huber, and W. Kuhfeld, A General Method for Constructing Efficient Choice Designs, 1996, SAS Working Paper. http://citeseer.nj.nec.com/rd/89088587,376089,1,0.25. Download/http:qSqqSqftp.sas.comqSqtechsupqSqdownloadqSqtechnoteqSqts629.pdf