DEVELOPMENT AND ESTIMATION OF A SEMI-COMPENSATORY MODEL INCORPORATING MULTINOMIAL AND ORDERED-RESPONSE THRESHOLDS

KAPLAN Sigal, Faculty of Civil and Environmental Engineering, Technion
SHIFTAN Yoram, Faculty of Civil and Environmental Engineering, Technion
BEKHOR Shlomo, Faculty of Civil and Environmental Engineering, Technion

ABSTRACT

Semi-compensatory models represent a two-stage choice process consisting of an elimination-based choice set formation upon satisfying criteria thresholds and a utility-based choice. Current semi-compensatory models assume a purely non-compensatory choice set formation and hence do not support multinomial criteria that involve trade-offs among attributes at the choice set formation stage. This study proposes the development and estimation of a semi-compensatory model that incorporates both multinomial-response and ordered-response criteria. The model development includes the proposition of a novel behavioural paradigm which involves a hybrid compensatory non-compensatory choice set formation process, followed by compensatory choice. The behavioural paradigm is represented by a mathematical model that accommodates a combination of multinomial-response and ordered-response thresholds and a utility-based choice. The proposed semi-compensatory model is applied to a stated preference experiment of off-campus rental apartment choices by students. Results demonstrate the applicability and feasibility of incorporating multinomial-response thresholds into the semi-compensatory framework.

Keywords: simplification by aggregation, multinomial cut-offs, two-stage model

INTRODUCTION

In complex choice situations, individuals engage in a two-stage decision process consisting of a sequence of choice set formation, in which alternatives are acceptable for consideration upon satisfaction of criteria thresholds, and utility-based choice among the considered alternatives (e.g., Ben Akiva and Boccara, 1995). Traditionally, criteria are thought of as one-dimensional continuous or naturally ordered entities reflecting environmental and individual constraints that govern rational choice (Simon, 1955). Classical examples are physical and economic constraints (e.g., speed, size, time and distance, ability to pay). However, some criteria are multidimensional entities that reflect both constraints and individual preferences, and are multinomial since their selection involves trade-offs among different attributes. For example, elimination of dwelling units on the basis of their location may be subject to evaluating the trade-off between accessibility and location amenities of different zones, elimination of air travel options on the basis of airline companies may involve tradeoffs between convenience and brand loyalty, and elimination of car rental agencies on the basis of their gasoline return policy may entail trade-offs among fuel costs, taxes and convenience.

The aforementioned decisions are typically represented by compensatory models with nested structures. Residential choice is a classical example of a decision that is treated as a nested compensatory choice. For example, Yates and Mackay (2006) describe a selection of tenure type, zone and dwelling type as a three-tier nested logit, and Chattopadhyay (2000) represents the selection of city, neighbourhood and dwelling unit as a three-tier nested logit.
Nevertheless, two main limitations of this compensatory approach are noteworthy. The first is the assumption of utility maximization as sole cognitive mechanism, which according to behavioural studies is justified only in choice situations involving up to five alternatives as a result of limitations to human cognitive capabilities (Payne, 1976). The second is the assumption of individuals as being fully informed regarding the universal realm of alternatives, which is unrealistic in choice situations characterized by a large number of alternatives (Shocker et al., 1991).

Semi-compensatory models better comply with human decision making in choice situations entailing many alternatives, since they assume multiple cognitive mechanisms and partial information regarding the universal realm of alternatives (Kaplan et al., 2009). These models include a probabilistic two-stage choice process, consisting of an elimination-based choice set formation upon satisfying criteria thresholds, followed by a utility maximization based choice. A disadvantage of current semi-compensatory models is that they assume a purely non-compensatory choice set formation followed by purely compensatory choice. Specifically, the choice set formation is assumed to be governed by attribute-based elimination heuristics. This behavioural assumption does not support multinomial criteria, since these heuristics do not support trade-offs within the choice set formation process.

Most existing semi-compensatory models are estimated while considering only naturally ordered criteria. Borgers et al. (1986) and Gensch (1987) reduce the universal realm to the viable choice set by using performance criteria evaluated on an ordinal scale. Gensch (1987) uses price, availability constraints and performance criteria. Ben Akiva and Boccara (1995) and Bierlaire et al. (2010) base the choice set formation on car and transit availability constraints. Morikawa (1995) composes the choice set on the basis of information availability and minimum attraction constraints. Başar and Bhat (2004) delimit the universal realm of alternatives by an overall utility threshold. Swait (2001a), Cantillo and Ortúzar (2005) and Cantillo et al. (2006) consider transport mode level-of-service criteria, including in-vehicle travel time, access time, in-terminal time, travel cost and variability of travel time and cost. Zheng and Guo (2008) implement distance-based criteria for choice set formation. Castro et al. (2009) narrow the realm of alternatives by considering walking time and travel time.

Attempts to represent multinomial criteria in addition to naturally ordered ones are conducted only by Swait (2001b) and Kaplan et al. (2009). Swait (2001b) considers car-rental agencies and gasoline return policy as criteria in addition to price and size, and Kaplan et al. (2009) specify apartment sharing and neighbourhood criteria in addition to price. However, both studies transform the multinomial criteria to naturally ordered ones prior to model estimation. Swait (2001b) transforms the multinomial criteria into several binary criteria. For example, the gasoline return policy criterion, which consists of the three alternatives “prepay full tank”, “return level” and “fill premium”, is transformed into three binary criteria, namely “not prepay gas”, “not return gas level rented”, and “not pay gas premium”. Kaplan et al. (2009) transform the multinomial criteria into ordered criteria by considering the performance of the alternatives with respect to a single aspect. For example, in the apartment sharing criterion vacant apartments are assumed to be better than shared apartments of the same features, which is reasonable when considering the aspect of area per person. Both approaches lead to information loss regarding choice set formation, since they do not consider possible trade-offs among different aspects of the same criterion. The present study addresses this limitation by developing a more general model that can accommodate both compensatory and non-compensatory rules in the choice formation process.

This study proposes the development and estimation of a semi-compensatory model that incorporates both multinomial-response and ordered-response criteria. The model development includes the proposition of a novel behavioural paradigm that involves a hybrid compensatory non-compensatory (HCNC) choice set formation process, followed by compensatory choice. The behavioural paradigm is represented by a mathematical model that accommodates a combination of multinomial-response and ordered-response thresholds, and a utility-based choice. The strength of the model lies in allowing trade-offs among attributes of multinomial thresholds at the choice set formation stage in addition to trade-offs among attributes at the choice stage. Hence, the model is capable of representing without any loss of information a choice set formation process that derives from both intrinsic individual constraints and preferences.

The proposed semi-compensatory model is applied to a stated preference experiment of off-campus rental apartment choices by students, as an example of a choice situation involving both multinomial-response and ordered-response criteria. The remainder of the paper is organized as follows. The next section presents the proposed behavioural paradigm and the corresponding mathematical model. The third section describes the data sample for model estimation and the fourth section presents the estimation results. Last, the fifth section draws conclusions and recommends further research.

MODEL FRAMEWORK

Behavioural paradigm

This study proposes a two-stage behavioural paradigm consisting of a hybrid compensatory-non-compensatory (HCNC) choice set formation followed by compensatory choice. The proposed paradigm is an extension of the model suggested by Payne (1976), which assumes a purely non-compensatory choice set formation followed by compensatory choice. The theoretical foundation of the proposed behavioural paradigm lies in the concept of "ecological rationality" suggested by Todd and Gigerenzer (2003). According to this concept, the human mind stores a collection of human cognitive decision-making mechanisms, called "adaptive toolbox", from which decision strategies are retrieved in order to cope with the complexity of the choice environment while balancing decision accuracy and speed. Specifically, under the theorem of "ecological rationality" decision heuristics can be thought of as building blocks that are retrieved from the "adaptive toolbox" and are mixed and matched to form a variety of decision strategies. Although the proposed behavioural paradigm is yet to be proven by formal descriptive decision making experiments, the structure of commercial on-line shopping websites supports its existence. For example, on-line real-estate portals offer the possibility to delimit the universal realm of apartments to a viable choice set by selecting both naturally ordered criteria thresholds, such as price and number of rooms, and multinomial criteria thresholds, such as city and apartment type. The provision of multinomial criteria in addition to naturally ordered criteria implies the co-existence of compensatory and non-compensatory evaluations at the choice set formation stage.

The proposed two-stage behavioural paradigm is illustrated in Figure 1. At the first stage, individuals delimit the universal realm of alternatives to a viable choice set by applying a combination of compensatory and non-compensatory decision rules. At the second stage, individuals choose their preferred alternative from their retained choice set.

The HCNC choice set formation includes three consecutive steps. The first step is the application of the simplification by aggregation heuristic (Simon, 1999) in order to characterize the universal realm of alternatives while balancing accuracy and cognitive effort by reducing the perceived complexity of the universal realm to a manageable level. The simplification by aggregation heuristic involves utilizing the property of near-decomposability in order to reduce the dimensionality of complex systems. Specifically, when individuals identify a hierarchical structure or a similarity pattern across basic units of a complex system with respect to a certain dimension, they can replace the details of the basic units with a small number of aggregate units. This procedure, which is analogous to the statistical procedure of factor analysis, allows individuals to reduce the system complexity to a manageable level, since often individuals are unable to perform the computations that would represent a system in terms of its basic components. The simplification by aggregation heuristic is useful when gathering full information about the universal realm with a high degree of accuracy is costly, time-consuming and cognitively effortful, since it allows the characterization of the complex system at a simplified aggregate level on the basis of partial information. When the complex system is a large universal realm of alternatives, individuals simplify the choice situation by aggregating alternatives into groups that share similar traits. At each aggregation level, the alternatives are discrete and are characterised by a vector of attributes, although the number of attributes decreases as the level of aggregation increases. For example, residential choice is defined by the level of spatial aggregation employed by the individual. Hence, individuals may simplify the universal realm of dwellings by referring to the spatial dimension and aggregating the alternatives by neighbourhoods. While dwelling units are characterized by their structural features and location amenities, only location amenities are retained at the neighbourhood level. Hence, only information regarding location amenities is gathered by the individual in case that the spatial dimension serves to simplify the universal realm of alternatives.

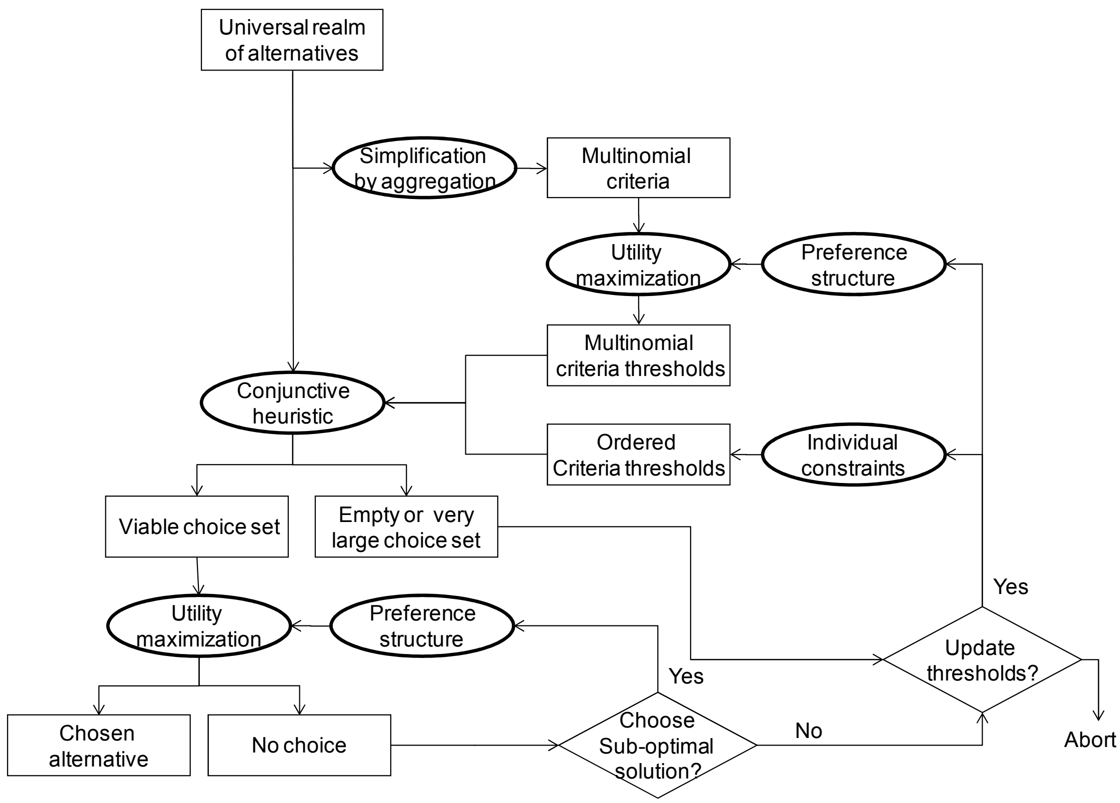

Figure 1- A two-stage behavioural paradigm involving HCNC choice set formation and compensatory choice

The second step is the evaluation of the aggregate groups in a compensatory process in order to account for trade-offs among attributes of aggregate units. The cognitive effort involved in the evaluation of aggregate units is lower than the cognitive effort involved in evaluation the basic units, since both the number of aggregate units and the number of considered attributes decrease as the aggregation level increases. The evaluation of aggregate units reflects the preference structure of individuals based on their knowledge and perceptions regarding the simplified universal realm of alternatives.

The third step is the application of the conjunctive heuristic (Einhorn, 1970) while combining two criteria type. The first type comprises naturally ordered criteria that derive from intrinsic individual constraints (either self-inflicted or externally-imposed). The second type consists of multinomial criteria that derive from individual preferences regarding the simplified universal realm of alternatives on the basis of available information at the choice set formation stage. While criteria that derive from intrinsic constraints are related solely to individual characteristics, criteria that derive from individual preferences are linked to the perception of the simplified universal realm formed in the mind of the each individual at the first and second steps. Specifically, the aggregation dimensions are treated by the individual as multinomial elimination criteria and the aggregate units along each dimension are treated as thresholds. Hence, upon the compensatory evaluation of aggregate units at the second step, the aggregate units that do not meet a tolerable utility threshold are discarded from further consideration.

Mathematical model

The probability $P_q(i|G)$ of individual q $(q=1, 2,\ldots,Q)$ to choose alternative i, as Q individuals face the same universal realm G is given by the following expression (Kaplan et al., 2009):

$$P_q(i \mid G) = P_q(i \mid S) P_q(S \mid G)$$

where $P_q(S|G)$ is the probability that individual q delimits the universal realm G to a viable choice set S, and $P_q(i|S)$ is the probability that individual q chooses alternative i from S. The equation is derived from Manski's formulation (1977). However, the model differs from that expression by relying on observed choice sets in addition to choice outcomes. This allows to reduce the number of possible choice sets to those actually chosen, hence avoiding the summation over all the theoretically possible choice sets in Manski's (1977) original formula.

The proposed model assumes a two-stage process consisting of an HCNC choice set formation followed by utility maximization. The HCNC choice set formation is based on the assumption that individuals apply the simplification by aggregation heuristic in order to represent the universal realm of alternatives in a crude though efficient manner. For parsimony reasons, the current mathematical formulation is based on two assumptions. The first assumption is that the population is homogeneous with respect to the simplified structure of the universal realm, following the application of the simplification by aggregation heuristic. Namely, individuals group the alternatives in the universal realm in the same manner. The second assumption is that, while the simplification by aggregation heuristic is latent, its outcomes are observable. Hence, researchers possess full information regarding the simplified structure of the universal realm as perceived by the population.

The outcomes of the simplification by aggregation heuristic are the aggregate groups of alternatives. Upon conceptualizing a simplified universal realm comprised of aggregate groups of alternatives, individuals can utilize their preference structure regarding the simplified structure of the universal realm in order to delimit it to a viable choice set. Specifically, individuals retain in their viable choice sets only alternatives included within their most preferred aggregate group of alternatives. The utility-based selection of the aggregate group g of alternatives from N aggregate groups depends on attributes of the aggregate groups, individual characteristics, and interactions between attributes of the aggregate groups and individual characteristics. The utility function is assumed to be linear in parameters. The probability P_{qg} to choose group g from N groups of alternatives within the universal realm G is represented by a multinomial logit model:

$$P_{qg} = F\left(\delta' Y_g + \lambda' Z_{qg} + \gamma' Y_g Z_{qg}\right) = \frac{\exp\left(\delta' Y_g + \lambda' Z_{qg} + \gamma' Y_g Z_{qg}\right)}{\sum_{n=1}^{N} \exp\left(\delta' Y_n + \lambda' Z_{qn} + \gamma' Y_n Z_{qn}\right)}$$

where Y_g is a vector of attributes values for group g, Z_{qg} is a vector of characteristics of individual q associated with group g, and $Y_g Z_{qg}$ are interaction terms between the attributes of group g and the vector of individual characteristics. The vectors δ, λ and γ are vectors of parameters to be estimated, and η_{qg} is the vector of error terms.

In addition to narrowing down the universal realm of alternatives according to individual preferences regarding the perceived simplified structure of the universal realm, individuals can further delimit the considered set of alternatives by applying elimination criteria that represent their intrinsic constraints. Individual constraints are naturally ordered variables that depend solely on individual characteristics. For these criteria, the threshold t^*_{kq} is expressed as a function of a vector of individual characteristics Z_{kq}, a vector of coefficients to be estimated α_k, and an error term ε_{kq}:

$$t^*_{kq} = \alpha'_k Z_{kq} + \varepsilon_{kq}$$

Assuming that the error term ε_{kq} for criterion k is identically and independently distributed normal across individuals, the ordered probit model represents the probability that individual q selects the threshold t^*_{kq}:

$$P_q\left(\theta_{(m-1)_k} < t^*_{kq} \leq \theta_{m_k}\right) = \Phi\left(\theta_{(m-1)_k} - \alpha'_k Z_{kq}\right) - \Phi\left(\theta_{m_k} - \alpha'_k Z_{kq}\right)$$

The probability to select a choice set S derives from the selection probability of a combination of the criteria corresponding to individual constraints and aggregate groups of alternatives that represent the preference of the individual regarding the perceived simplified structure of the universal realm. In case that the error terms η_{gq} and ε_{kq} are uncorrelated, the probability to select the choice set S is simply the multiplication of the probabilities to select an aggregate group of alternatives and a threshold of an individual constraint. However, in case that the error terms η_{gq} and ε_{kq} are correlated, the multiplication of the separate single-dimensional models is inappropriate. Considering that the aggregate groups of alternatives follow a multinomial choice structure, and considering a single individual constraint, the joint selection of an aggregate group of alternatives and a naturally ordered threshold to delimit the universal realm to a viable choice set can be represented by the joint multinomial ordered response probit model developed by Bhat (1997).

Following Bhat (1997), the joint selection of an aggregate group and an ordered-response threshold follow a bivariate normal distribution. In order to construct the bivariate normal distribution, the multinomial choice among N aggregate groups of alternatives is transformed into N binary choices. Let R_{qg} be a dummy variable that equals one if the aggregate group of alternatives g is chosen by individual q and zero otherwise. Define:

$$\nu_{qg} = \left\{ \max_{n=1,\dots,N; n\neq g} u_{qn} \right\} - \eta_{qg}$$

Denote U^*_{qg} as the utility of choosing the aggregate group of alternative *g* by individual *q.*

$$U^*_{qg} = \delta' Y_g + \lambda' Z_{qg} + \gamma' Y_g Z_{qg} - \nu_{qg}$$

R_{qg} equals one if U^*_{qg} has a positive value and zero otherwise. The error term v_{gq} is transformed into a standard normal random variable v^*_{gq} as follows:

$$\nu^*_{qg} = \Phi^{-1}\left[F_g\left(\nu_{qg}\right)\right]$$

where Φ is the standard normal distribution function and *F* is the multinomial logit distribution function of v_{gq}. Assuming that both v_{gq} and ε_{kq} are independently distributed across individuals, and considering that both v^*_{gq} and ε_{kq} follow a normal distribution, a bivariate distribution can be specified as follows:

$$\Phi_2\left[\nu^*_{qg}, \varepsilon_{kq}, \rho_q\right] = \Phi_2\left[\Phi^{-1}\left[F\left(\nu_{qg}\right)\right], \varepsilon_{kq}, \rho_q\right]$$

where Φ_2 represents the bivariate normal distribution and ρ_g is the correlation between the aggregate group *g* and the individual constraint. The joint probability of choosing group *g* and threshold t^*_{kq} by individual *q* is expressed as follows:

$$P_q\left(g, \theta_{(m-1)_k} < t^*_{kq} \le \theta_{m_k}\right) = \Phi_2\left(\Phi^{-1}\left[F\left(\delta' Y_g + \lambda' Z_{qg} + \gamma' Y_g Z_{qg}\right)\right], \theta_{(m-1)_k} - \alpha'_k Z_{kq}, \rho_g\right)$$
$$-\Phi_2\left(\Phi^{-1}\left[F\left(\delta' Y_g + \lambda' Z_{qg} + \gamma' Y_g Z_{qg}\right)\right], \theta_{m_k} - \alpha'_k Z_{kq}, \rho_g\right)$$

The corresponding likelihood is:

$$L_q\left(S \mid G\right) = \prod_{n=1}^{N} \left\{ \prod_{m_k=1}^{M_k} \left[\Phi_2\left(\Phi^{-1}\left[F\left(\delta' Y_g + \lambda' Z_{qg} + \gamma' Y_g Z_{qg}\right)\right], \theta_{(m-1)_k} - \alpha'_k Z_{kq}, \rho_g\right) \right.\right.$$
$$\left.\left. -\Phi_2\left(\Phi^{-1}\left[F\left(\delta' Y_g + \lambda' Z_{qg} + \gamma' Y_g Z_{qg}\right)\right], \theta_{m_k} - \alpha'_k Z_{kq}, \rho_g\right) \right]^{d_{qm_k}} \right\}^{d_{qg}}$$

where d_{qmk} is an indicator function that is equal to one if individual *q* selects the threshold category *m* of criterion *k* and zero otherwise, and d_{gq} is an indicator function that is equal to one if individual *q* selects the aggregate group *g* and zero otherwise.

The utility-based choice stage is represented by considering a linear utility function with an identically and independently Gumbel distributed error term. The utility-based selection of alternative *i* from the viable choice set *S* depends on the attribute values of alternative *i*, individual characteristics, and interactions between attributes of alternative *i* and individual characteristics:

$$U_{qg} = \beta' X_i + \omega' Z_{qi} + \mu' X_I Z_{qi} + \zeta_{qi}$$

where X_i is a vector of attribute values for alternative *i*, Z_{qi} is a vector of individual characteristics associated with alternative *i*, and $X_g Z_{qi}$ are interaction terms between the attributes of alternative *i* and a vector of individual characteristics. The vectors *β, ω* and *μ* are vectors of parameters to be estimated, and ζ_{qi} is the vector of error terms. The probability for individual *q* to choose alternative *i* from *J* alternatives within the viable choice set *S* is represented by a multinomial logit model:

$$P_q(i \mid S) = \frac{\exp\left(\beta' X_i + \omega' Z_{qi} + \mu' X_i Z_{qi}\right)}{\sum_{j \in S}^{J} \exp\left(\beta' X_j + \omega' Z_{qj} + \mu' X_j Z_{qj}\right)}$$

The corresponding likelihood is:

$$L_q(i \mid S) = \prod_{j \in S} \left[\frac{\exp\left(\beta' X_i + \omega' Z_{qi} + \mu' X_i Z_{qi}\right)}{\sum_{j \in S}^{J} \exp\left(\beta' X_i + \omega' Z_{qi} + \mu' X_i Z_{qi}\right)} \right]^{d_{qi}}$$

where d_{qi} equals unity if threshold alternative *i* is chosen by individual *q* and zero otherwise.

The error terms of the HCNC choice set formation and the compensatory choice are assumed to be two independent processes. Accordingly, the combined unconditional log-likelihood for a population of *Q* individuals who choose their most preferred alternative *i* from their viable choice set S_q can be written as:

$$LL = \sum_{q=1}^{Q} \ln\left[L_q\left(i \mid S_q\right) L_q\left(S_q \mid G\right) \right]$$

The coefficients of both the HCNC choice set formation and the utility maximization are estimated simultaneously by using maximum likelihood estimation. The maximization of the log-likelihood function is conducted by using a stepwise procedure. In the first step, log-likelihood is maximized with respect to the threshold selection parameters assuming an independent error structure. In the second step, the model parameters are held fixed and the log-likelihood function is maximized with respect to the correlation parameters. In the third step, the parameters of the threshold selection and the correlation parameters are held fixed and the log-likelihood function is maximized with respect to the utility parameters. Finally, the parameters from the third step are used as starting values for the full-information maximum likelihood estimation of the log-likelihood function.

DATA

The data sample for model estimation consists of a population of 858 students studying in the city of Haifa, in the north of Israel, who participated in a web-based experiment of rental apartment choice. Kaplan et al. (2010) thoroughly discuss the survey design, detailing its theoretical foundation, the construction of the synthetic dataset and the design of the questionnaire. This paper briefly presents the main elements of the survey.

Participants delimited the universal realm of alternatives to a viable choice set from which they chose their preferred alternative. Specifically, participants searched a synthetically generated apartment dataset by a list of pre-defined criteria threshold values and ranked their three most preferred apartments from the resulting choice set. The synthetically generated apartment dataset, which was constructed on the basis of a statistical analysis of local real-estate databases, consisted of rental apartments characterized by 18 attributes including location, monthly rent price, structural features, neighbourhood amenities, roommate policy and electrical appliances. The criteria for searching the dataset were apartment sharing, neighbourhood and monthly rent price.

A questionnaire supplemented the experiment by collecting participants' socio-economic characteristics, attitudes and perceptions about issues relevant to rental apartment choice. Participants were also asked to evaluate the perceived location amenities of neighbourhoods that are relevant to students' rental apartment choice, since public records of quantitative information about neighbourhood amenities in Haifa are unavailable. Specifically, questions investigated job accessibility for students, leisure accessibility, public open space availability, campus accessibility by public transport, and perceived car travel time to campus. All the items were expressed on a seven-point Likert scale ranging from 1 (very poor) to 7 (excellent), except the perceived car travel time that was expressed in minutes.

The web-server automatically recorded the participants' answers to the questionnaire and typing actions during the two-stage choice experiment. The three most preferred apartments for each of the 858 students served for model estimation. Hence, the data sample consists of a pool of 2,574 observations of choice outcomes and their corresponding thresholds.

MODEL ESTIMATION RESULTS

Table 1 presents the estimation results for the model combining an HCNC choice set formation and utility maximization with an independent error structure at the utility maximization stage (HCNC-MNL). The model is presented while considering alternatively independent and correlated thresholds.

Three criteria are represented in the estimated model: apartment sharing, neighbourhood and monthly rent price. These criteria were ranked as the most important rental apartment attributes in a preliminary survey and were utilized for searching the database by the entire population sample.

The model structure is based on the assumption that individuals simplified the universal realm by aggregating the alternatives into eight mutually exclusive groups on the basis of apartment sharing and neighbourhood. Apartment sharing differentiates between vacant and shared apartments and four neighbourhoods (i.e., Carmel, Neveshanan, Remez, Nesher) are considered. Figure 2 illustrates the considered aggregate groups of alternatives. The selection of the most preferred aggregate group of alternatives is represented by a multinomial logit model.

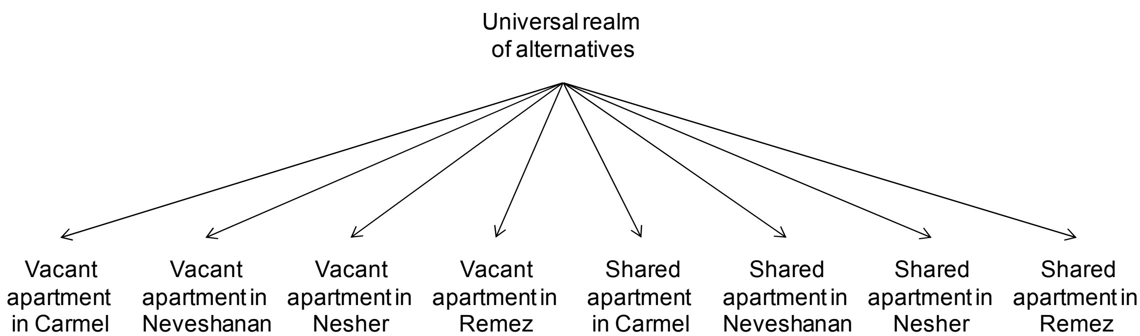

Figure 2 – Aggregate groups simplifying the universal realm of alternatives

The price threshold is best described by an ordered probit model with 11 categories (200, 250,…,700). The choice of an aggregate group of alternatives and prices leads to the formation of 83 non-empty choice sets.

The relevant universal realm for the population sample contains 400 apartments, which are all the available apartments in the generated synthetic database for the four aforementioned neighbourhoods. A priori availability constraints are not imposed on the alternatives. However, for each ranked choice, the alternatives with higher priority ranking are excluded from the choice set prior to model estimation, since respondents could not rank the same alternative twice. It should be noted that repeated choices of the same individual are treated as independent and data are treated as pooled cross-sectional data.

The variable specification for the choice set formation stage is based on the assumption that individual constraints are related to individual characteristics and perceptions, and that the choice among aggregate groups of alternatives are related to their perceived utility. Hence, the perceived utility is related to both the attributes of the aggregate groups of alternatives and individual characteristics.

The selection among the aggregate groups of alternatives is related to students' job accessibility, accessibility to leisure activities, public open space availability, campus accessibility by public transport, and perceived car travel time to campus. These attributes are entered as generic variables in the model (i.e., part-worth utilities are constant across all alternatives). In addition, alternative specific constants are associated with Carmel, Nesher and Remez with respect to Neveshanan. An interaction term between studying in the Faculty of Medicine and residing in Carmel is estimated, since the Carmel neighbourhood provides higher accessibility to the medical campus. Interaction terms between car availability and residence in Carmel, Nesher or Remez are estimated, since Carmel is located far away from the main campus and Nesher and Remez are located on a hilly terrain, while Neveshanan is adjacent to campus and on a level terrain. Finally, individual characteristics are associated with the four aggregate groups entailing vacant apartments. The individual characteristics that explain the choice among the aggregate groups include socio-economic variables (i.e., marital status, gender, age, monthly expenses), transport related variables (i.e., car availability), study-place preferences, and current residential location and arrangements.

The selection of the price threshold is dependent on socio-economic characteristics (i.e., age, marital status, monthly expenses, income source), transport related variables (i.e., car availability, trip frequency to the campus), apartment search experience, current residential location and living arrangement. Latent constructs designed to measure price knowledge and travel preferences also influence the selection of the price threshold. The observed personal characteristics are directly used as explanatory variables, whereas perceptions and preferences are incorporated in the model after performing factor analysis.

The model specification assumes that the utility-based choice depends on the attributes of the apartments included within the choice set. According to Kaplan et al. (2009), the inclusion of interaction terms between individual characteristics and apartment attributes for this particular case study results in only a minimal increase of the model goodness-of-fit. Hence, on the basis of parsimony considerations the current model specification does not include interaction terms between individual characteristics and apartment attributes. Apartment attributes specified in the model include price, structural features (i.e., number of rooms, renovation status, floor, security bars), location amenities (i.e., view, parking, noise level, proximity to campus), electrical appliances (i.e., air conditioning system and solar water heater), number of roommates and smoking policy.

The first three parts of table 1 describe the determinants of the HCNC choice set formation. The fourth part presents the relative importance of apartment attributes at the utility-based choice stage, given the viable choice set. The interpretation of the model results is provided in the next sub-sections.

Table I – Estimation results for the semi-compensatory model with HCNC choice set formation

Variable	Description	HCNC-MNL Independent thresholds est.	HCNC-MNL Independent thresholds t-stat.	HCNC-MNL Correlated thresholds est.	HCNC-MNL Correlated thresholds t-stat.
Apartment type threshold coefficients					
Marital status	Single [a]	-	-	-	-
	Married	2.058	9.39	2.049	9.86
Gender	Female [a]	-	-	-	-
	Male	-0.847	-6.02	-0.839	-6.39
Age	Years	0.024	2.45	0.024	2.54
Car availability	Monthly/weekly [a]	-	-	-	-
	Daily	0.692	4.63	0.691	4.94
Trip frequency to campus	Monthly/weekly [a]	-	-	-	-
	Daily	-0.601	-4.37	-0.606	-4.53
On-campus studying to benefit from teacher-student communication factor		-0.142	-3.61	-0.142	-3.57
Monthly expenses	< $750 [a]	-	-	-	-
	$ 750 - 1000	0.504	2.36	0.511	2.41
	$1000 - 1750	0.777	3.74	0.776	3.87
Current residential arrangement	Dormitories [a]	-	-	-	-
	Parents	-	-	-	-
	Roommates	-1.189	-5.91	-1.156	-6.08
	Alone	1.352	5.14	1.351	5.09
	Spouse	1.730	9.63	1.729	10.47
Current residential location	Haifa city [a]	-	-	-	-
	Haifa suburbs	-0.504	-1.87	-0.504	-1.87
	Haifa outskirts	-0.639	-5.01	-0.631	-2.85
	Center of Israel	-	-	-	-
Neighbourhood specific constant	Neve Shaanan [a]	-	-	-	-
	Carmel	-2.239	-13.78	-2.243	-12.25
	Nesher	-1.585	-13.31	-1.577	-13.20
	Remez	-1.401	-14.03	-1.407	14.82
Accessibility to job opportunities		0.144	5.92	0.141	5.57
Availability of leisure opportunities		0.106	4.04	0.107	4.02
Availability of gardens and open spaces		0.324	14.35	0.325	13.62
Arrival ease to campus by public transport		0.008	0.40	0.007	0.37
Travel time to campus by car		-0.035	-4.58	-0.036	-4.32
Interaction between daily car availability and residence in Carmel		1.378	9.81	1.389	8.89
Interaction between daily car availability and residence Nesher		0.718	5.44	0.725	5.33
Interaction between daily car availability and residence Remez		0.296	2.25	0.294	2.13
Interaction between residence in Carmel and studying in the medical campus		1.348	6.01	1.334	5.81
Interaction between residence Nesher and studying in faculties adjacent to the campus's eastern gate		0.516	4.08	0.521	3.93

Price threshold coefficients					
Marital status	Single [a]	-	-	-	-
	Married	0.588	7.41	0.571	5.83
Gender	Female [a]	-	-	-	-
	Male	-0.201	-3.71	-0.196	-3.48
Age	Years	0.035	4.76	0.034	3.55
Monthly expenses	< $ 500 [a]	-	-	-	-
	$ 500-750	0.366	6.18	0.369	5.07
	$ 750-1500	0.444	5.81	0.457	5.21
	> $1500	0.585	5.74	0.593	4.78
Income source	None/scholarship [a]	-	-	-	-
	Part-time job	0.091	1.82	0.091	1.50
	Full-time job	-	-	-	-
Car availability	Monthly/weekly [a]	-	-	-	-
	Daily	0.243	4.06	0.236	3.66
Price-knowledge factor		0.083	5.64	0.083	4.80
Apartment search experience	≤3 apartment changes [a]	-	-	-	-
	> 4 apartment changes	-0.251	-2.34	-0.252	-2.36
Trip frequency to campus	Monthly/weekly [a]	-	-	-	-
	Daily	-0.347	-5.79	-0.352	-5.27
Current residential arrangement	Dormitories [a]	-	-	-	-
	Roommates	-0.399	-5.11	-0.406	-4.96
	Alone/parents	0.244	3.51	0.244	3.04
	Spouse	0.567	7.30	0.558	6.72
Current residential location	Haifa –low/medium class neighbourhoods [a]	-	-	-	-
	Haifa – upper class neighbourhoods	0.251	3.62	0.249	2.77
	Haifa Suburbs	-	-	-	-
	Haifa outskirts	-	-	-	-
	Center of Israel	0.217	3.05	0.216	2.30
Non-motorized modes preference factor		-0.033	-2.34	-0.033	-1.82
Travel minimization preference factor		-0.041	-2.69	-0.041	-2.13
Cut-off points	200 [a]	-	-	-	-
	250	-0.185	-0.80	-0.184	-0.62
	350	0.277	1.20	0.282	0.95
	350	0.549	2.37	0.548	1.84
	400	0.754	3.26	0.758	2.53
	450	1.209	5.20	1.237	4.10
	500	1.639	7.04	1.677	5.55
	550	2.221	9.56	2.280	7.52
	600	2.448	10.53	2.541	8.33
	650	2.753	11.84	2.993	9.64
	700	2.883	12.39	3.304	10.26
Correlations across thresholds					

Vacant apartment in Carmel and price	-	-	-0.207	-2.39
Vacant apartment in Neveshanan and price	-	-	-0.322	-3.71
Vacant apartment in Nesher and price	-	-	-0.060	-0.71
Vacant apartment in Remez and price	-	-	-0.088	-1.16
Shared apartment in Carmel and price	-	-	0.166	1.07
Shared apartment in Neveshanan and price	-	-	0.485	9.97
Shared apartment in Nesher and price	-	-	0.326	5.39
Shared apartment in Remez and price	-	-	0.384	6.68
Choice stage coefficients given the selected choice set				
Rent price (monthly)	-0.004	-8.46	-0.004	-8.58
Number of rooms	0.686	16.30	0.683	16.21
Number of roommates	-0.570	-8.10	-0.567	-7.96
Walking time to campus	-0.083	-21.19	-0.084	-21.32
Quiet apartment	1.358	29.28	1.341	28.97
Parking	0.408	7.05	0.414	7.13
Floor	-0.116	-5.46	-0.114	-5.34
Smoking allowed	-0.258	-3.87	-0.255	-3.77
Security bars	0.187	4.12	0.184	4.02
Stunning view	0.243	4.88	0.242	4.83
Renovated apartment	0.467	9.59	0.470	9.60
Air conditioner	0.673	14.03	0.666	13.70
Solar water heater	0.573	9.51	0.573	9.33
Number of observations		2574		2574
Number of parameters		67		75
Log-likelihood at zero		-29268.825		-29268.825
Log-likelihood at estimates		-16578.954		-15465.313
McFadden's adjusted R^2		0.431		0.469

a – base category

HCNC choice set formation

Determinants of the choice among aggregate groups of alternatives related to apartment sharing and neighbourhood

The propensity of selecting an aggregate group of alternatives on the basis of apartment sharing and neighbourhood increases according to several neighbourhood characteristics: (i) greater accessibility to job opportunities; (ii) greater availability of leisure activities; (iii) greater availability of gardens and open spaces; (iv) greater ease of arrival to campus by car; (v) lower car travel time from the neighbourhood to campus.

According to the model results, students have an inherent preference for the Neveshanan neighbourhood. The students have also a higher inherent preference for Remez, Nesher and Neveshanan relatively to the Carmel neighbourhood. Possible reasons may be the existence of higher student concentrations in these neighbourhoods, the ease of arrival to campus by non-motorized modes and the lower municipal tax rates.

The interaction terms between car availability and the propensity to delimit the universal realm to apartments in Carmel, Nesher and Remez have positive parameter estimates. The parameter values increase with the distance from campus and the change in terrain characteristics from level to hilly terrain. Specifically, Neveshanan is adjacent to campus and is characterized by a level terrain, Remez and Nesher are adjacent to campus but are characterized by a hilly terrain, and Carmel is both located far away from campus and characterized by a hilly terrain. Hence, students with daily car availability have higher propensity to select less accessible neighbourhoods than other students, and this tendency strengthens with the increase of the difficulty to arrive by non-motorized modes to campus.

A positive parameter estimate is obtained for the interaction term between studying in faculties adjacent to the campus eastern gate and residing in Nesher, which is the only neighbourhood located near this gate. Possibly, due to the hilly terrain near the campus, students who study in faculties located in the eastern side of the campus prefer to reside in proximity of the eastern gate.

Individual characteristics are associated exclusively with the four groups of vacant apartments. The propensity of selecting an aggregate group of alternatives that includes only vacant apartments increases according to the progression of the respondents' lifecycle in terms of age, marital status and monthly expenses, and to travel independence related to daily car availability.

The propensity to delimit the universal realm of alternatives to an aggregate group, which includes only shared apartments, increases with daily trips to campus and the preference to study there in order to benefit from teacher-student interaction. Possibly, respondents who spend more time on campus than in their apartment prefer to ease apartment chores by sharing them.

The propensity to retain a certain apartment type increases with the reference to the status-quo, as students tend to retain apartments that match their current residential arrangements. Namely, respondents who currently reside with roommates have a greater propensity to retain an aggregate group of alternatives including only shared apartments, while respondents who reside alone or with their spouse tend to retain an aggregate group of alternatives consisting only of vacant apartments.

Threshold selection determinants for monthly rent price

The propensity to select higher price thresholds increases according to: (i) progression of the students' lifecycle and socio-economic status, as respondents who are married or have higher monthly expenses tend to select higher price thresholds; (ii) self-reported price knowledge; (iii) reference to the status quo, as respondents who currently rent an apartment alone or with their spouse have a greater propensity to select higher price thresholds than respondents who reside either in the dormitories or with roommates.

The propensity to select higher price thresholds decreases according to: (i) habit to travel daily to campus, likely related to shorter time spent in the apartment with respect to the time spent on campus; (ii) greater apartment search experience, likely reflecting a greater propensity to undergo the burden of replacing the status quo apartment with a more cost-efficient one.

Effect of incorporating correlation between the price threshold and the choice of aggregate alternatives of apartment sharing and neighbourhood

The likelihood ratio test value, which compares the model with an independent error structure at the choice set formation stage versus the model with correlations across the choice of an aggregate group of alternatives and the price threshold, equals 2227.3 and hence exceeds the critical chi-square value for eight correlation parameters at the 0.001 significance level. Thus, the null assumption of an independent error structure at the choice set formation stage can be rejected. As expected, the correlation parameters between the four aggregate groups of vacant apartments and the price threshold are negative, indicating a positive correlation between delimiting the universal realm to vacant apartments and selecting a higher price threshold. Similarly, the correlation parameters between the four aggregate groups of shared apartments and the price threshold are positive, indicating a negative correlation between delimiting the universal realm to shared apartments and selecting a higher price threshold. Five out of eight correlation parameters are significant at the 0.05 significance level. Correlations between the price threshold and delimiting the universal realm to vacant apartments in the neighbourhoods of Nesher and Remez are not statistically significant. The correlation between the price threshold and selecting a shared apartment in Carmel are also not statistically significant, although the reason may be the relatively low frequency in which this aggregate group of alternatives was chosen (only 36 respondents chose this group).

Utility based choice

For each respondent, the choice of an aggregate group of alternatives according to apartment sharing and neighbourhood determines whether the chosen apartment is vacant or shared and the neighbourhood in which it is located. Hence, apartment sharing and neighbourhood do not vary within the viable choice set. Monthly rent price and number of roommates vary within the viable choice set and hence serve as explanatory variables at the utility maximization stage.

The propensity of renting an apartment increases according to an increase in terms of: (i) quality of structural features, such as greater number of rooms and renovation; (ii) availability of security bars; (iii) availability of cost-efficient electrical appliances such as solar water heater; (iv) availability of an air conditioning system, a necessity in the hot Mediterranean climate; (v) parking availability; (vi) environmental location amenities, such as quietness and a nice view.

The propensity of renting an apartment decreases according to an increase in terms of: (i) apartment monthly rent; (ii) floor number, as apartments located on lower floors are preferred, possibly due to the scarcity of elevators in the two neighbourhoods; (iii) number of roommates; (iv) distance from campus, as an increase in walking time to the campus decreases the attractiveness of the apartment; (v) roommates' pro-smoking policy, possibly since 87.0% of the respondents in the sample are non-smokers.

CONCLUSIONS AND FURTHER RESEARCH

This study proposes the development and estimation of a semi-compensatory model that incorporates both multinomial-response and ordered-response criteria. To the authors' knowledge, this model is the first semi-compensatory model that relaxes the assumption of a purely non-compensatory choice set formation and allows trade-offs to occur at both the choice set formation stage and the choice stage.

The model development includes the proposition of a novel behavioural paradigm that involves a HCNC choice set formation process followed by compensatory choice. The HCNC choice set formation includes three consecutive steps. First, individuals apply the simplification by aggregation heuristic in order to reduce the complexity of the universal realm of alternatives and hence the cognitive effort associated with its management. Namely, individuals group the alternatives in the universal realm into a few aggregate groups, thus creating a simplified universal realm with a manageable structure. Second, individuals evaluate aggregate groups with a compensatory process. The cognitive effort involved in the evaluation of aggregate groups is lower than the cognitive effort involved in the evaluation of the basic alternatives, due to the lower number of alternatives and attributes in the simplified universal realm. Third, the conjunctive heuristic is applied. Specifically, upon the compensatory evaluation of the aggregate groups in the second step, aggregate groups that do not meet a tolerable utility threshold are discarded from further consideration. The chosen aggregate group is further delimited by eliminating alternatives that do not meet criteria thresholds that represent individual constraints. Following the HCNC choice set formation, the retained alternatives form the viable choice set from which the most preferred alternative is chosen. The behavioural paradigm is represented by a mathematical model that represents the sequence of HCNC choice set formation and compensatory choice. The mathematical formulation represents the choice of the most preferred alternative from a viable choice set, derived from a joint choice of a viable aggregate group of alternatives and of criteria thresholds that represent individual constraints.

The empirical analysis uses a data set from a stated preference experiment of off-campus rental apartment choices by students. The model is estimated while assuming the following behavioural process: (i) the individual simplifies the universal realm by aggregating the alternatives according to apartment sharing and neighbourhood; (ii) a compensatory evaluation process yields the most preferred aggregate group of alternatives for each individual; (iii) the viable choice set for each individual is formed by eliminating the alternatives that do not belong to the most preferred aggregate group and do not meet the price criteria threshold. Results indicate that: (i) the hypothesized model has an excellent goodness of fit (McFadden's R^2=0.469), thus supporting the assumptions of the behavioural paradigm; (ii) the compensatory evaluation of the aggregate groups of alternatives, according to apartment sharing and neighbourhood, is related to neighbourhood characteristics and to inherent preferences of individuals on the basis of their socio-economic characteristics; (ii) the price threshold selection is explained by individual socio-economic characteristics and price knowledge; (iii) significant correlations exist between the multinomial and the ordered-thresholds. Accounting for these correlations greatly improves the goodness of fit of the model in comparison with the assumption of an independent error structure (McFadden's R^2 equals 0.469 when correlations are accounted for versus 0.431 when an independent error structure is assumed); (iv) the utility-based choice from the viable choice set is related to apartment structural features, number of roommates and their smoking policy, presence of electrical appliances, and location amenities that were not accounted for at the choice set formation stage.

Several possible directions exist for further development of the proposed model. First, estimation results indicate that the aggregate alternatives that represent the perceived simplified universal realm at the HCNC choice set formation stage follow a nested structure instead of an independent structure. Hence, a potential research direction is the use of the nested logit model instead of the multinomial logit model for representing the choice among aggregate groups of alternatives at the choice set formation stage. Second, for parsimony reasons the current model formulation assumes that the population is homogeneous with respect to the simplification by aggregation heuristic and hence to the simplified structure of the universal realm. Furthermore, the model is based on the assumption that researchers possess full information regarding the simplified structure of the universal realm. A possible research direction is to relax these two assumptions by allowing the simplified structure of the universal realm to vary with a probability across the population. Last, the current model formulation adopts the model developed by Bhat (1997) for the joint estimation of multinomial-response and ordered-response models. An alternative approach for representing the HCNC choice set formation could be to estimate copula-based models, due to their ability to represent a wide range of multidimensional distributions involving discrete and continuous variables.

REFERENCES

Başar, G. and C. R. Bhat (2004). A parameterized consideration set model for airport choice: an application to the San Francisco Bay Area. Trans Res Part B, 38 (10), 889-904.

Bhat, C. R. (1997). Work travel mode choice and number of Non-work commute stops. Trans Res Part B, 31 (1), 41-54.

Ben-Akiva, M. E. and B. Boccara (1995). Discrete choice models with latent choice sets. Int J Res Market, 12 (1), 9-24.

Bierlaire, M., R. Hurtubia and G. Flötteröd (2010). An analysis of the implicit choice set generation using the Constrained Multinomial Logit model. Proceedings of the TRB 2010 Annual Meeting, January 10-14, Washington D.C.

Borgers, A., H.Timmermans and J. Veldhuisen (1986). A hybrid compensatory-non-compensatory model for residential preference structures. J Hous Built Environ,1 (3), 227-234.

Cantillo, V. and J.d.D. Ortúzar (2005). A semi-compensatory choice model with explicit attribute thresholds of perception. Trans Res Part B 39 (7), 641-657.

Cantillo, V., B. Heydecker and J. d. D.Ortúzar (2006). A discrete choice model incorporating thresholds for perception in attribute values. Trans Res Part B 40 (9), 807-825.

Castro, M., F. Martinez and M. Munizaga (2009). Calibration of a logit discrete choice model with elimination of alternatives. Proceedings of the International Choice Modelling Conference, March 30-April 1, Harrogate, U.K.

Chattopadhyay, S., (2000). The effectiveness of McFadden's nested logit model in valuing amenity improvement. Reg Sci Urban Econ, 30 (1), 23-43

Einhorn, H. J. (1970). The use of nonlinear, non-compensatory models in decision making. Psychol Bull, 73 (3), 221-230.

Gensch, D. H. (1987). A two-stage disaggregate attribute choice model. Market Sci, 6 (3), 223-239.

Kaplan, S., Y. Shiftan and S. Bekhor (2009). Development and estimation of a semi-compensatory model with flexible error structure. Proceedings of the 12th IATBR conference, December 13-18, Jaipur, India.

Kaplan, S., S. Bekhor and Y. Shiftan (2010). Unraveling semi-compensatory choice: data collection based on two-stage decision protocols. Accepted for presentation at the12th WCTR, July 11-15, Lisbon, Portugal.

Manski, C. (1977). The structure of random utility models. Theor Decis, 8 (3), 229-254.

Morikawa, T. (1995). A hybrid probabilistic choice set model with compensatory and non-compensatory choice rules. Proceedings of the 7th WCTR, Sydney, Australia.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: an information search and protocol analysis. Organ Behav Hum Perform, 16 (2), 366-387.

Simon, H. A. (1955). A behavioral model of rational choice. The Q J Econ, 69 (1), 99-118.

Simon, H. A. (1999). Can there be a science of complex systems? In Y. Bar-Yam (Ed.), Unifying themes in complex systems: Proceedings from the International Conference on Complex Systems 1997 (pp. 4-14). Cambridge, MA: Perseus Press.

Shocker, A. D., M. E. Ben-Akiva, B. Boccara and P. Nedungadi (1991). Consideration set influences on consumer decision-making and choice: issues, models and suggestions. Mark Lett, 2 (3), 187-197.

Swait, J. (2001a). Choice set generation within the generalized extreme value of discrete choice models. Trans Res Part B 35 (7), 643-666.

Swait, J. (2001b). A non-compensatory choice model incorporating attribute cutoffs. Trans Res Part B 35 (10), 903-928.

Todd, P. M. and G. Gigerenzer (2003). Bounding rationality to the world. J Econ Psychol 24, 143–165.

Yates, J. and D. F. Mackay (2006). Discrete choice modeling of urban housing markets: a critical review and an application. Urban Stud, 43 (3), 559-581.

Zheng, J. and J. Y. Guo (2008). A destination choice model incorporating choice set formation. Proceedings of the 87th TRB Annual Meeting, Washington, D.C., USA.