Choice set formation for outdoor destinations: the role of motivations and preference discrimination in site selection for the management of public expenditures on protected areas

July 2016

Forthcoming in *Journal of Environmental Economics and Management*

Mara Thiene

University of Padua

Joffre Swait

University of South Australia

Riccardo Scarpa

Durham University Business School, University of Waikato, and University of Verona

Abstract

Effective public expenditure currently dominates the management focus of many protected areas. This calls for explicit modelling of constraints and motivations that, respectively, obstruct and stimulate visits to selected outdoor destinations. Choice set formation is the result of screening and/or inclusion of specific sites (alternatives) to form the set of sites considered in real choices. Evidence shows that the omission of a structural representation of choice set formation is harmful to econometric inference. Yet, the literature has largely ignored the underlying behavioural phenomenon. We show, using a discrete choice experiment involving selection among seven recreational sites in an Italian national park, that choice set formation is behaviourally relevant, even after controlling for preference discrimination. Motivations (why visit?) are important determinants of preliminary site screening for choice set inclusion, as well as site selection, justifying the additional value of such modelling extension.

Keywords: discrete choice modelling, demand for outdoor recreation, site selection, travel choice, nonmarket valuation, choice set formation, efficient public expenditure, local finance.

Acknowledgements: The research was funded by the Dolomiti Bellunesi National Park.

1. Introduction

This paper focuses on the additional insights that a multi-layered destination choice model can convey in driving effective public expenditure in the management of protected areas. While access fees are one means to raise funds for the conservation of protected areas, implementing such fees is often too costly, either administratively or politically. Inevitably, in these cases, the bulk of the management funds still comes from general taxation. With the on-going squeeze in public finances ensuing from the 2008 financial crisis, the management of conservation areas has increased its focus on making expenditures more effective. We show how destination choice models can be extended to address a variety of features that can inform public expenditures for the conservation of protected areas in two important aspects. The first is the selective spatial allocation of specific services, which is a form of site-specialization. The second is the increase in monitoring efforts on selected site attributes to trace out the effectiveness of expenditure. To adequately measure effectiveness we extend the conventional destination choice model with heterogeneous preferences to account for choice set formation and preference discrimination.

Since the early work by Bockstael, Hanemann and Strand (1987) and Bockstael, Hanemann and Kling (1987), random utility models (henceforth RUMs) have been employed to study demand for outdoor recreation (amongst others, Morey, Rowe, and Watson 1993; Herriges and Kling 1997; Provencher and Bishop 1997, 2004) and the associated demand for environmental quality. These models explain observed choices over a finite set of mutually exclusive outdoor destinations, but typical applications tend to ignore certain behavioral processes that may act as substantive determinants of choice. We focus on two such aspects, the first of which is choice set formation and its determinants; the second is the ability of the data to discriminate between preference signals over random noise from the idiosyncratic error component. This latter phenomenon is sometimes referred to as ‘preference discrimination’ (Swait and Erdem 2007), ‘choice uncertainty’ and ‘choice consistency’. In choice models it takes the form of heteroscedasticity in stochastic utility, a topic which has been explored before in an environmental or resource economics setting (e.g. De Shazo and Fermo 2002), albeit not in conjunction with choice set formation. The omission of relevant variables leads to misspecification and biased welfare estimates and so does the omission of relevant behavioral processes. Hence, the exploration of substantive behavioral issues is of interest on its own account in terms of adding insight and realism to conventional choice models.

The theoretical importance of “choice set generating processes” was emphasized as early as 1977 by Manski, who also alerted economists to the consequences of the curse of dimensionality: as the number of alternatives increases, latent choice set generation models become quickly intractable, posing an obstacle to their application in contexts with many alternatives. In practice, the problem of defining choice sets, or the subset actually considered (the so-called “consideration” set), has often been solved by appealing to assumptions (a process termed ‘choice set imputation’)[1], which have been supported by arguments with varying degrees of plausibility. This commonly held assumption of “exogeneity” of choice sets from survey data is in stark contrast with the behavioral framework of random utility maximization. Endogenizing this process, in the sense of “making it dependent on data”, poses several challenges. Despite the paucity of formal econometric models for this important component of choice analysis, the random utility paradigm and its significant extensions to discrete-continuous demand analysis (e.g. Phaneuf, Kling, and Herriges 2000) has been very

[1] We distinguish between “choice set imputation” and “choice set formation”. The former is used to describe the exercise of assigning a specific set of alternatives to a decision maker (e.g., sites visited in the past year), whereas the latter is reserved for the modelling of a probability distribution reflecting the likelihood that members of a collection of choice sets is the true choice set.

effective in the profession, with literally hundreds of applications to date.

A review of the existing literature in environmental economics reveals that only a few attempts have been made to explore the policy implications of endogenous choice sets in recreation demand models. In particular, these have focused on the importance of alternative assumptions on choice sets for the estimates of interest and their consequent role in policy and management decision for outdoor activities. To date, substantially less emphasis has been placed on the determinants of inclusion of individual sites in choice sets; this is therefore the first topic to which we wish to contribute with this paper.

The dependence of welfare estimates and visitation share forecasts on the assumptions concerning the size and composition of choice sets has been well-documented in nonmarket valuation for some time (Peters et al., 1995; Haab and Hicks 1997; Parsons and Hauber 1998; Parsons et al. 2000a, 2000b; Hicks and Strand 2000). Very early applications, such as Caulkins, Bishop, and Bouwes (1986), made some efforts to individualize choice sets by including for each respondent only the sites actually visited. However, Peters, Adamowicz, and Boxall (1995) were the first to truly "endogenize" the choice set using data collected in the Southern Alberta Sportfishing survey in 1991. They compared MNL models and their welfare estimates from three separate choice set imputations: (1) the set of all sites known to the researcher, (2) the answer to the survey question "which of these sites they had visited in the past or would consider when choosing a site to go fishing", and (3) randomly generated choice sets. The last set was determined on the basis of the results from McFadden (1978)[2] and repeated in recreation demand by Parsons and Kealy (1992). Welfare change estimates for site closures, tree planting and trout stocking all showed substantial sensitivity to the definition of choice sets.

Haab and Hicks (1997) would seem to be the only paper published in environmental economics that actually makes an attempt at modeling the determinants of the probability of inclusion of a candidate site into a visitor's choice set. This probability is integrated in the computation of the site selection probability by using a variant of the Manski's model (1977). This method relies on the sequential decomposition of the choice probability into the probability of including the site in the choice set and the probability of the same site providing maximum utility. As recognized by the authors, this is a rather restrictive assumption that might not be generally applicable, but it is nevertheless similar to assumptions made in other fields (e.g., Swait and Ben-Akiva 1987; Horowitz and Louviere, 1995). The curse of dimensionality forced Haab and Hicks (1997) to implement the model in choice studies with a small number of destination sites (5 beaches in New Bedford and 12 beaches in Chesapeake Bay). Their results show a substantial impact of accounting for choice set formation on estimates of both selection probability and welfare due to water quality improvement.

Hicks and Strand (2000) also made an attempt at endogenizing choice sets based on available data by conditioning on respondent's self-reported statements of knowledge of destination sites. Ignorance of the existence of a site by a given respondent would imply deletion of this site from the respondent's choice set. Hicks and Strand (2000) also assessed the effect of making the choice set a function of distance from the residential location of the respondent, using different travel-time cut-offs (a spatial criterion) as well as the inclusion of only "familiar" sites to the visitor. By focusing on comparisons of estimates for mean

[2] McFadden (1978) shows that the Independence of Irrelevant Alternatives (IIA) Property of MNL models allows for consistent (though not efficient) estimation of utility function parameters using random subsets of alternatives (plus the chosen one) from the full set. This result is often misunderstood since it does not in any way address the topics of choice set imputation or formation. In fact, the whole point of that result is that choice is assumed to be made from among all alternatives, but the parameters of the utility function can be consistently estimated using a subsample of all alternatives; the result in no way implies that the choice set can be imputed to be a random subset, nor can one estimate a choice set formation model using multiple random samples of alternatives.

compensating variation associated with three separate policy actions (40 percent decrease in Fecal Coli and closure of two sites: Bay Ridge and Sandy Point), Hicks and Strand find that more restrictive criteria (e.g. cut-off at 1 hour travel time and inclusion of only familiar sites) induce the largest differences in welfare change estimates. These differences range from several orders of magnitude (1 hour cut-off, closure of Sandy Point) to 40 percent (familiar set, closure of Sandy Point). This form of sensitivity is obviously due to the lower availability of substitute sites in smaller choice sets, and despite the somewhat arbitrary nature of the cutoffs, their results are indicative of the potential dimension of the bias arising from choice set misspecification. The notion of familiarity was used also by Parson, Massey and Tomasi (2000), who also made choice sets individual-specific.

Parson, Plantinga and Boyle (2000) take a different approach to this issue. Their criteria for choice set composition were based on exogenous spatial aggregation (four choice sets) and one endogenous criterion: the degree of popularity of the site (one set). Welfare estimates (mean per trip compensating variation for loss of 5 sites) were obtained for all five choice sets under analysis. They were expressed as percent change from the welfare estimates of the standard choice model including all sites that is almost universally used. Alternative choice sets caused welfare estimates to vary from 43 to 60 percent. A similar range of welfare estimate bias (30-50%) emerged in a recent Monte Carlo study (Li, Adamowicz and Swait, 2015) that used a two stage decision (choice set formation first, alternative selection second) data generating process, in the absence of taste heterogeneity, and compared the bias across an array of commonly employed specifications. This study did not seek to address the bias versus efficiency tradeoff previously raised by von Haefen (2008).

Bias due to inadequate assumptions on choice sets has been a concern for even a longer time in other disciplines. Interest in this issue was started by the pioneering work done in transport by Swait and Ben Akiva in the mid to late eighties (see 1985, 1987a,b). In this strand of the literature the study by Swait and Ben-Akiva (1985) is particularly noteworthy because it contains a theoretical analysis of the bias introduced on utility function parameters if choice set formation is assumed away when it is in fact present. This study is among the first to jointly estimate endogenously individualized choice sets and choice selection probabilities. Somewhat later similar investigations were started in consumer research (Fotherimgam 1988) and progressed all the way through the proposal of modeling the inclusion of alternatives into consideration sets on the basis of marginal cost and benefit of consideration by Roberts and Lattin (1991).

Several literature reviews on the subject were produced in various fields: e.g., by Thill (1992) in destination choice modeling, Roberts and Nedungadi (1995) in consumer research, Haab and Hicks (2000) in recreation demand, and more recently, Hauser (2014) in consumer research. Haabs and Hicks (2000) concluded by launching the following challenge to the profession:

"Without careful attention to issues such as the horizontal and geographic extent of the market, perceptions versus measurable behavior, and familiarity with sites versus consideration of sites, econometric models only serve to allow the researcher more modeling flexibility. Future efforts into the understanding of choice set issues in recreation demand modeling should take the empirical results described in this special issue and apply those to new survey design and data collection efforts" (Haab and Hicks 2000, 279-80)

Yet, to our knowledge, with the exception of von Haefen (2008), in which goals and motivations do not play an explicit role in the choice formation stage of the model, no further attempts have since been made by environmental economists to address this cogent issue in empirical data. More specifically, the challenge posed by Haab and Hicks nearly fifteen years ago remains unheeded.

Other disciplines in which discrete choice models are in common use have behaved differently. In transport and consumer research, for example, throughout the 00's there was a flurry of contributions revolving around the issue of consideration sets. To review, note the choice set Generation logit (or GenL) model proposed by Swait (2001a), the cutoff approximation method using a Lagrangian relaxation of the direct utility function suggested by Swait (2001b), the "inclusion function" approach proposed by Cascetta and Papola (2001), to the model retrieving unobservable consideration sets from panel data proposed by Van Nierop et al. (2010).

Compelling theoretical reasons were put forward a long time ago for both extending and restricting individual choice sets. Uncertainty over future preferences was used to rationalize flexibility, and expansion in the size of consideration sets (Kreps 1979) to produce benefits to choice agents. On the other hand, Richardson (1982) argued that the search cost associated with inclusion of additional alternatives into an agent's consideration sets would be the prime motivators for size reduction, an argument similar to the conclusion drawn from the bounded rationality argument by Simon (1991). Yet, the theory of "choice overload" due to the presence of too many options is still causing controversy (Scheibehenne et al. 2010, Chernev 2010) in consumer research, as a growing body of empirical studies yields mixed results.

Our paper examines a classic problem in natural resource management for recreation: the impact on welfare of a population of recreationists from different management policies at outdoor destinations. Results from studies outside environmental economics have persuasively shown that goals and individual constraints act not so much at the level of attribute preference, but at the earlier stage of choice set formation. We propose an independent availability logit (IAL) model to assess the importance of these drivers on choice set formation, to revive interest in an important topic that we feel has lain dormant for too long. Specifically, we report on a study conducted to address and embrace the challenge originally posed by Haab and Hicks (2000). We use an organizational principle for choice set formation based on motivations and apply it to purpose-collected data from visitors to alpine locations in an environmentally protected area managed for outdoor recreation. In modeling choice set formation we hope to derive insight that can increase efficiency in public expenditure via spatial specialization of the supply of amenities across destinations within the park. This should respond to the pressing demand by management authorities of new tools to rationally prioritize expenditures.

2. Motivations and barriers underlying choice set formation

The behavioral rationale for the existence of choice set formation rests on the idea that decision makers are subject to constraints (e.g., financial, time, social, risk – see, e.g., Swait and Ben-Akiva 1987, Swait 2001b) and limitations (e.g., cognitive, decision time, knowledge and/or awareness – see Hauser 2014) which lead them to use heuristics in decision making. Among such simplifying heuristics used to reduce decision effort are those that lead to the elimination or "screening out" of alternatives from further deliberation. That is, "choice sets" are rational constructs that a decision maker adopts to account for constraints (Hauser 2014). Our earlier citations from the environmental and resource economics literature have implicitly or explicitly taken this constraint-driven perspective.

A second perspective is, however, entirely possible: decision-makers' motivations (i.e., decision objectives) lead them to create choice sets as a deliberate means of leading to tradeoffs among alternatives known to satisfy one or more important objectives. For example, an outdoor enthusiast may initially desire to develop specific climbing skills (the main objective to be pursued) by training at certain sites (the choice set) where he will be "pushed" as much as possible without endangering himself overly much (risk mitigation being a second objective, which acts as a constraint to, rather than a driver of, his behavior). While some sites may be removed from the choice set because they are too easy, unsuitable for climbs or too risky, tradeoffs between the remaining sites (the choice set) allow further influences to come to the fore in his decision making, while insuring that the main motivation of skills improvement is well-served throughout the decision process. This approach can be more informative to the management of protected areas as it identifies scope for expenditure specialization.

The decision making literature in psychology has come to interpret behavior in terms of goals and plans (Weber and Johnson 2009). Motivations are high level goals that antecede, initiate and direct decision making by serving as the basis for selectivity, a central characteristic of goal-directed behavior. "Selectivity" is a broad term intended to encompass mindsets, attitudes and intermediate actions that serve to implement resource allocations and priority setting arising from the motivations guiding behavior, and eventually lead to the choice of an alternative. Selectivity applies to activation of objectives, attention to information, input for decision making (time and effort), evaluation processes and decision rule selection, and hence underpins choice itself. Swait and Argo (2011) use survey data to show that in multiple decision contexts (job interview preparation, restaurant menu item selection, candy bar choice) respondents self-report pursuing multiple goals simultaneously; Krantz and Kunreuther (2007) have demonstrated that which goals are activated strongly influence what is chosen, but also help determine *how* a decision is made. Thus, motivations establish antecedent volitions (Li, 2013) that determine the strategy of decision making in a given context. These antecedent volitions encompass a portfolio of actions that a decision maker can take to establish the parameters of the decision making process: whether or not to screen alternatives, what information is relevant to discriminate between alternatives, what preferences to employ, what decision rule to employ, among others.

As we will explain in greater detail below, we have taken a specific approach to incorporate both self-reported motivations and constraints into the choice set formation and choice processes. The former are intended to capture the positive motivations, as it were, and the latter the negative barriers that might underpin destination choice set formation in the empirical context we examine. Details of our approach follow in coming sections.

3. The data and the survey

3.1 Description of the Dolomiti Bellunesi National Park (DBNP)

The Dolomiti Bellunesi National Park (henceforth DBNP) is located in the northeastern Italian Alps, covers 32,000 hectares and is the only nationally protected area of the region. Since 2009 it has been a UNESCO World Heritage site due to its biodiversity and to the remains of ancient human activities, which include pre-historical remains, a mining centre of over five hundred years of age, a Middle Ages monastery, the Christian chapels of the piedmont belt, a medieval hospice and the more recent "military roads" built to connect the Serenissima Republic of Venice to the rest of Europe.

Already in the 18th century, its peaks Vette di Feltre and Mt. Serva were renowned amongst botanists for the flora biodiversity. The vascular flora (plants with flowers and others, such as ferns, having roots, stems, and leaves) consists of about 1,400 species (1/4 of those inside Italy), among which are many species deserving of mention, either because they are endemic, rare, or have great phytogeographical value. The southern part of the Park has highest biodiversity because least impacted by glaciations, and hence hosts the highest rate of survival of ancient species.

The Dolomiti Bellunesi are the southeastern district of the Dolomitic Alps. It is a complex mountain range overlooking one of the largest alpine valleys (media valle del Piave). The structural complexity and relative variety of rocks give rise to an impressive orographic fragmentation and a great variety of landscapes. The park's watercourses flow in a dense network of valleys and dells, often through narrow ravines. There are many artesian springs in woodlands, accompanied by showy cushions of musk. Foamy waterfalls and spectacular potholes are common. In fact, the karstic nature of the rocks has allowed a subterranean landscape to develop: potholes, cracks, halls, tunnels, and abysses penetrate into the bowels of the earth. The karstic complex, which generated over 30 Km of tunnels, is the largest in the Dolomites, and one of the most extensive cave systems in the Veneto Region and in Italy, frequently visited by amateur and expert speleologists.

3.2 The Data

Data were collected during the autumn of 2013 (November-December) by means of a web-based survey fielded by a specialised market research firm. Respondents were randomly sampled from a representative panel of the population of the Veneto region (Italy), and were segmented to match the socio-demographic characteristics of the 2011 Italian census. The Veneto is a populous (5 million) region located in the northeast part of Italy, with seven cities with population over fifty thousand and a variety of terrains (mountains, hills, alluvial planes and coastlines). Focus groups and a pilot study were conducted to test and calibrate the survey instrument. The survey was broadly aimed at the whole potential population of visitors DBNP, as the sample was extended to those who had not visited the Park.

The questionnaire had three sections: i) the first explored the outdoor recreational profile of respondents, by asking, for example, the number of years of engagement in hiking, climbing and mountain-biking activities; ii) the second was a discrete choice experiment, iii) the third addressed motivations and constraints (i.e., antecedent volitions) affecting behaviour by eliciting the motivations that would drive visitation decisions, personal constraints (e.g., mobility restrictions) and their perceived association to each destination site at the park, scale items to characterize maximizer vs. satisficer tendencies (see Schwartz et al., 2002), graded according to a 5-point Likert scale, plus conventional socio-demographics of respondents.

The data consists of 1,452 completed interviews. Summary statistics show that the average age for women is 38 (s.d. 11.8) and 40 (s.d. 11.2) for men. Fifty-four percent of women are high school graduates (32% for men) and one third are university graduates (55% for men). Thirty-five percent of women have annual household after-tax incomes considered to be low (less or equal to €20,000), whereas only 26% of men are in this stratum. Only 13% of women and 17% of men declared a yearly income in excess of €35,000.

Seventy percent of respondents visited the DBNP at least once within the last five years. Among those, 27% visited 2-4 times, thereby suggesting that most respondents are familiar with the area, and that at least one fourth appreciate the area enough to make repeat visits. Forty-one percent of respondents defined themselves as hikers, 20% had been hiking for at least ten years. Only 10.5% started hiking in the last three years. Almost 20% of respondents described themselves as Mountain Bikers (MTBs). Interest in MTB has recently increased; in fact, 38% of MTBs report taking it up recently. Participation in alpine mountaineering, which mainly focuses on climbing, has also recently increased. Less than 10% of the sample engage in these risky and challenging activities, but almost 60% took up this activity within the last three years.

3.3 Sites, Attributes & Levels

The park management authority has had an active role in this research since its inception, informing the selection of both sites and attributes used to characterize destinations. Their goal was to obtain useful information to guide the effective implementation of sustainable management policies. The seven selected sites are at the boundaries of the DBNP. Each can easily be accessed by means of private vehicles. Destinations include four valleys (Val di Lamen, Val Canzoi, Val del Mis and Val dell'Ardo), one mountain pass (Passo Croce d'Aune) and two sites located along the main road crossing the park along a north-south axis (refer to

Figure A1, on-line appendix).

The aim of the empirical study was to address the largest variety of outdoor activities of interest to different categories of visitors. So, both general and activity-specific attributes were selected. We used ten attributes, with varying number of levels (see Table 1). Some attribute levels were not available at all sites, i.e., there are site-specific levels for several attributes.

--- Table 1 about here ---

Park authorities currently do not levy any *entrance fee* but seek to explore the visitors' willingness to pay for site access as well as identify sites where expenditures for specific activities should be targeted. Given the lack of public funding, the introduction of an entrance fee might be a realistic and efficient option to finance park infrastructure, park upkeep, and new facilities. In our discrete choice experiment (DCE) the levels for the fee attribute were set to €0, €2, €6, and €10 per person per visit.

The second attribute deals with *bivouacs*. These are quite important facilities located at high altitude; they provide shelter to hikers, climbers and MTBs in case of bad weather conditions. They are located throughout the park, but their real availability is influenced by some key local factors, which were taken into account when defining the four levels used in the DCE (see table 1).

The third attribute is *site access*. At two of the seven sites (Val Canzoi and Val del Mis) the park authority is specifically interested in exploring the option of denying private car access during the week-end. These two sites are amongst the most visited and tend to suffer from road traffic and trail congestion (see table 1 for the levels).

Attribute four refers to *crowding*, which affects selected sites. Four levels (Table 1) of congestion are described in terms of numbers of visitors encountered during the visit.

The fifth attribute concerns the availability of *picnic sites* ranging from none to seven. These are much appreciated facilities, particularly (although not exclusively) by visitors looking for relaxation and who wish to avoid strenuous activities.

The sixth attribute is *wildlife sites*, a new type of facility for which the park has no information in terms of visitors' appreciation or willingness to pay. Wildlife sites are large areas delimited by fences in which wild animals, e.g., wolves (*Canis lupus*) and rock goats (chamois or *Rupicapra rupicapra*) can be observed within a fairly natural habitat. These enclosures provide visitors the opportunity to enjoy a direct wildlife sighting experience without necessarily engaging in long and challenging hikes. In this case, the levels simply describe the availability (or not) of accessible wildlife areas.

The seventh attribute describes improvements of safety features of *'via ferratas'*, which are equipped trails along exposed areas. By allowing hikers to fasten onto an iron cable along the most challenging tracts of the trails even visitors with minimal skills are allowed to reach mountain peaks or other locations from which they can enjoy spectacular viewscapes. Levels for this attribute refer to structural and technical aspects of this feature, since the length of tracts equipped with iron cable can be varied. Another climber-specific attribute (the eighth) involves the provision of additional climbing itineraries in crags and cliffs.

Despite the recent increase in interest, there is currently no specific itinerary devoted to mountain bikers. The park authority is interested in understanding the impact on visitors of developing up to three *MTB trails* at specific sites and this was the ninth attribute.

The tenth and last site attribute of the DCE is the implementation of additional *thematic itineraries* specifically focused on cultural and historical aspects, wild flora and fauna. The levels for this attribute range between one and three itineraries.

To ensure realism, the experimental design process must take a significant number of restrictions into account because of site-specific limitations. For example, via ferratas are feasible at only two of the seven destinations; similarly, MTB trails are only possible at four. In fact, except for fee and crowding levels, all attributes described above have at least one exclusion constraint. These constraints naturally eliminate the use of orthogonal designs, and also bring into question the applicability of efficient designs. With respect to the latter, it is important to note that using an efficient design assumes that the analyst has some probabilistic knowledge of the data generation process (dgp). In our approach we are explicitly calling into doubt that the dgp describes a single-stage utility maximizing agent (such as would be assumed in a simple MNL model), we believe a more flexible design strategy is called for. Thus, the design method we employ is a hybrid that recognizes that optimizing a design for the MNL or RPL model is unsuitable for the choice set formation that may be present; rather, we allow for the existence of choice sets by adopting an availability design to overlay the basic identification of the utility function parameters. The availability design is particularly important in this situation because it explores choice set formation in a controlled fashion by removing sites systematically. The basic design strategy was as follows:

1) we generated 192 candidate choice set profiles (runs) assuming the MNL model holds for all seven sites, with constraints imposed at a site level as required by the context. The design criterion we employed was average maximum entropy over the runs, using parameter priors defined by the researcher;

2) we then used an availability design of eight runs, the first seven being limited to three sites and the last inclusive of all seven sites, each run establishing which of the seven sites would be shown;

3) we selected three or seven sites from among eight randomly selected runs from the 192 runs to combine into choice sets according to the availability design (examples of choice-tasks are reported in Figure 1).

In general, the procedure for generating a design optimized for a choice set formation model is quite a complex problem, which to our best knowledge has not been studied. Future research on this matter may be very valuable. At this point in time we feel that our hybrid approach is a reasonable attempt at a design method that combines utility function parameter identification with a challenge to choice set formation via the availability design overlay.

--- Figure 1 about here ---

3.4 Motivations, constraints and maximization tendency scale

As described in Section 2, we allowed for a variety of unusual variables to be accounted for in our model of choice behavior. The third section of the survey asked respondents their reasons for visiting the DBNP, by means of questions investigating their motivations and goals and the associated types of activities they would practice during the visit in order to accomplish their goals. These were specifically explored in the preceding pilot study, which collected a total of 238 questionnaires by interviewing visitors intercepted at the DBNP. Based on these results, a list of goals and related activities were included in our main study, identifying motivations for visiting the park listed in Table 2.

--- Table 2 about here ---

In the main survey, respondents were asked to match these motivations with outdoor activities and with the seven focal sites of the study. To explore the importance of these motivations, respondents were also asked to indicate the importance of each on a scale from 1 (=most important) to 5 (=least important) when visiting DBNP. The purpose of this question was to establish the ideal goal mix for each respondent, to be used to (partially) explain choice set formation. The survey also elicited personal constraints (Swait 2001b, Morey and Thiene

2012), as it was assumed that health problems, lack of time or similar issues could potentially limit the visitation of some park sites. For this reason, respondents were asked to indicate their perceptions of being impeded by different constraints at each site (see Table 2).

In order to place respondents into a maximizer tendency scale, we used a series of scale items often used in the psychology literature. It has been shown that some individuals consistently seek to choose the "best" option in most (if not all) contexts, whereas others tend to "satisfice" and settle for options that they consider good enough (Simon, 1955; Schwartz et al., 2002). It is important to note that this scale only measures a behavioral tendency towards a type of outcome and it does not detect some inviolable rule of behavior. Recently Nenkov et al. (2008) demonstrated that a shorter, 6-item maximization scale (see Table A1 in the on-line appendix at the link ****) performs as well at classifying as the original 13-items scale by Schwartz et al. (2002). These six items were used in our survey; respondents were asked to indicate (on a scale from 1 to 5, where 1=does not describe me at all, 5=describes me very well) the degree to which each statement describes him/her.

4. The model

Our model specification is an extension of the independent availability logit (IAL) model first presented by Swait and Erdem (2007), which in turn builds on Swait and Ben-Akiva (1987) (for applications of the IAL see also Andrews and Srinivasan 1995, Ben-Akiva and Boccara 1995). It is a two-stage decision model (simultaneously estimated) with a latent choice set formation in the first stage, followed by a probability of site selection in the second stage conditional on the selected choice set. The novelty of our extension is that we address a series of behavioral phenomena that have recently interested choice modelers. These include (*i*) the impact of antecedent volitions (Li 2013; Swait 2013; Swait and Feinberg 2013), which accounts for heterogeneity in motivations for the visit; (*ii*) personal constraints and individual characteristics that might impact decision making behaviors (Morey and Thiene 2012); (*iii*) heteroscedasticity of random utility to represent preference discrimination (Swait and Erdem 2007); (*iv*) impact of choice complexity on the site availability function through the inclusion of a systematically presented set size variation; (*v*) preference heterogeneity of random utility parameters (e.g., Train 2003) in the site selection equation are also addressed, maintaining the panel structure of the choice data throughout.

We assume respondents screen alpine sites based on the motivations or goals they want to pursue, which in turn determine the activities they wish to practice at a destination. Such activities are the means to achieve their goals. For example, if a respondent wishes to spend time with the family because of small children, he/she may only include in the choice set for the visit those sites suitable for spending time with family, such as sites where picnics can be held and/or short hikes are available, thereby excluding other sites from the choice set. On the other hand, a climber who desires to improve her skills would exclude from her choice set those sites not offering itineraries with some sort of climbing routes. One of the aims of our study is to test if and to what extent the choice set formation process is affected by these antecedent volitions of potential visitors. Based on the six basic motivations identified for visiting the park, we make the availability (or the inclusion propensity in the respondent's choice set) equation for each of the seven alpine sites a direct function of motivation self-reports and personal constraints. If one suffers from some forms of health problems or mobility restrictions, or has money or time restrictions, this will certainly affect his/her choice set, independently from his/her structure of preferences. Further to motivational drivers and barrier to participation, site availability was parameterized to be a function of whether or not the individual was a "day tripper", i.e. someone who tends to visit the park for a day as opposed to making multi-day trips. Because the park is reasonably accessible from several major population centers in northeastern Italy, this is an important discriminator among park visitors.

Besides capturing site screening, or choice set formation, the model includes a conditional site selection probability model. The utility function for site selection is based on the ten site attributes described above, which are known to be site-specific characteristics relevant to potential visitors. To characterize it as a whole, the model specification is a heteroscedastic conditional mixed logit model, applied to a choice set determined by a latent independent availability choice set formation model dependent on contextual complexity (presented set size), and personal characteristics, constraints and motivations. Among the particular models we test are included variants of this basic framework with random tastes and random availability effects, assumed to be independent random normal across respondents, using a panel specification. To our knowledge this is the first study to address all these issues in the environmental economics literature, and perhaps in others.

Heteroscedasticity in the stochastic component of utility across alternatives has been termed "preference discrimination" by Swait and Erdem (2007). The rationale for this terms is that when the scale parameter (which is inversely related to the variance of the stochastic utility) is large, this will translate into a more discriminating choice behavior across alternatives (i.e., more extreme conditional choice probabilities). This occurs because the stochastic component of utility becomes relatively less important vis-à-vis the systematic indirect utility component, thereby implying that respondents rely more on the latter.

The probability that visitor n chooses site i in choice scenario t, conditional on random taste β and random availability coefficient δ, is given by the expression below:

$$P_{\text{int}|\beta,\delta} = \sum_{C_t \in \Delta_{M(t)}} P_{\text{int}|C_t}(X_{nt}, Z_n \mid \beta, \theta)\, Q_{nt}(C_t, W_{nC} \mid \delta), \forall i \in M(t), t = 1,...,T, \tag{1}$$

where C_t is a choice set in $\Delta_{M(t)}$, which is the set of all possible choice sets in the available set $M(t)$ of sites eligible for choice; $P_{int|C_t}$ is the conditional probability of choosing i from set C at time t of the sequence of T choices, and $Q_{nt}(C_t, W_{nC}|\delta)$ is the likelihood of C_t being the true choice set. In the first term X_n and Z_n refer to utility determinants and scale function variables, respectively. These may be made conditional on a vector of individual-specific taste parameter β_n conformable with the vector of X_n for site attributes, and on θ, a vector of scale parameters conformable with Z_n. The second (multiplicative) term, Q, defines the probability that the sites in set C_t are collectively the choice set, relative to other sets in $\Delta_{M(t)}$. It depends on the specific choice set C_t and on W_{nC}, which is a vector of site- and person-specific characteristics, and δ_n is an individual-specific conformable parameter vector to be estimated.

Assuming that the stochastic utility terms are independent across alternatives – but not identically distributed due to scale differences – we obtain that the choice probability at point t of the panel is a multinomial logit (MNL) with individual-specific scale functions λ_n, thus:

$$P_{\text{int}|C_t}(X_{nt}, Z_n \mid \beta, \theta) = \begin{cases} \dfrac{\exp[\lambda(Z_n \mid \theta) \cdot V_{in}(X_{\text{int}} \mid \beta)]}{\sum_{j \in C_t} \exp[\lambda(Z_n \mid \theta) \cdot V_{jn}(X_{jnt} \mid \beta)]} & i \in C_t \\ 0 & i \notin C_t \end{cases} \tag{2}$$

where V_{in} is the deterministic utility, defined below as a function of site attributes:[3]

$$V_{in} = \sum_{i-1} \alpha_{in} + \beta'_{2n} SiteAttr_{mi} + \beta_{3n} TCDT_{ni} + \beta_{4n} TCnDT_{ni}, i \epsilon M \tag{3}$$

where:

[3] In previous versions of this study we also estimated models on the entire sample in which goals and motivations were interacted with attributes in the indirect utility function. Such models are reported in the Appendix available on-line to the interested reader.

α_{in} = random alternative-specific constants for site *i*
β_{rn} = random parameters or parameter vectors to be estimated conformable to their respective variable vectors, *r*=1, …, 4
TCDT = travel cost, round trip, for day trippers
TCnDT = travel cost, round trip, for non-day trippers
m = denotes the number of site attribute parameters (26 in total)
$SiteAttr_m$ = list of m = 1, ..., 26 site attribute variables (see Table 1).

The travel cost for each visitor encompasses the roundtrip vehicular costs from home to the access town nearest the park, as well as entrance fees.

To ensure non-negativity the scale functions λ_n are exponentiated latent linear-in-parameter factors defined on the vector Z_n of person-specific characteristics, the activity level index and the maximizer score index. In our case, the details of the specification are as follow:

$$\lambda_n(Z_n|\theta) = \exp(\theta_1 Male + \theta_2 Age + \theta_3 Age\_sq + \theta_4 ActLevel + \theta_5 MaxScore + \theta_6 MaxScore\_sq + \theta_7 Income + \theta_8 IncMiss) \tag{4}$$

where *Male*, *Age*, *Age_sq*, *Income* are socio-economic variables; *ActLevel* is an indicator of engagement in mountain activities; *MaxScore*, and *MaxScore_sq* are the linear and quadratic maximization tendency scale described above; and *IncMiss* is a dummy variable denoting missing income. Note that we do not associate stochastic heterogeneity with θ.

Hence, preference discrimination is parameterized by making the scale parameter a function of *i*) an index describing the level of engagement in mountain activities, *ii*) a maximizer tendency score index describing the respondent's general propensity to seek the best option versus being satisfied with one that's good enough, and *iii*) certain demographic characteristics that we consider *a priori* to be related to consistency in behavior; in particular gender, age and income. The activity level index is obtained by summing the number of years each respondent was engaged in hiking, mountaineering and MTB and dividing the total by sixty. High values of this index would indicate well-experienced and well-trained visitors. The maximizer index is the average score across answers to the six questions in Table A1. After mean-scaling, a high value would indicate a "maximizer" rather than a "satisficer".

The probability of the set *C* being the true choice set is based on the assumption that the availability of any individual alternative is probabilistically independent of the (un)availability of any other alternative. Swait and Ben-Akiva (1987) dubbed this the Independent Availability choice set formation model:

$$Q_{nt}(C_t, W_{nC} \mid \delta) = \frac{\left(\prod_{j \in C_t} A_{jn}(W_{nC} \mid \delta)\right)\left(\prod_{k \in M(t)-C_t} [1 - A_{kn}(W_{nC} \mid \delta)]\right)}{1 - \prod_{k \in M(t)} [1 - A_{kn}(W_{nC} \mid \delta)]} \tag{5}$$

where $(A_{kn}(W_{kCn}|\delta))$ is a binary logistic availability (or inclusion) probability parameterized as follows:

$$A_{in}(W_{nC}|\delta) = [1 + \exp(-(\delta_{i1} + \delta_2{}' Constraints_n + \delta_{3i} Relax_n + \delta_{4i} Family_n + \delta_{5i} Skills_n + \delta_{6i} KnowTerr_n + \delta_{7i} NaturCtc_n + \delta_{8i} OtherG_n))]^{-1}, \ \ \forall i \in M \tag{6}$$

where A_{in} is the probability that alternative *i* is included in the choice set *C* by respondent *n*, W_{iCn} is a vector of choice set-, site- and person-specific characteristics, and δ is an individual-specific parameter vector. The normalization constant in the denominator of (5) excludes the possibility of a null choice set. $Constraints_n$ is a vector of six individual constraints (see also Table 2; coding is specified subsequently), and δ_2 is the vector of associated parameters. The remaining coefficients are associated with the motivations of respondents (see Table 2), interacted with site indicators.

In several of the models estimated we have assumed one or both of the β (utility) and δ (availability) parameters to be randomly distributed. We have specified the joint density distribution across the components of these vectors as multivariate normal with a diagonal covariance matrix. Thus, over the panel of T scenarios given to each respondent we specify the probability of observing the collection of responses as

$$P_n(\beta,\theta,\delta) = \int_{\beta,\delta} \left(\prod_{t=1}^{T} P_{i_{nt}^*|\beta,\delta} \right) f(\beta,\delta)\, d\beta d\delta\,, \qquad (7)$$

where i_{nt}^* is the chosen alternative for person n, scenario t, $f(\beta,\delta)$ is the joint density function described above, other quantities as previously defined. (Consonant with the literature, we assume in (7) that choice responses are independent across scenarios.) The log likelihood for the sample is then simply

$$LL(\beta,\theta,\delta) = \sum_n \ln P_n(\beta,\theta,\delta)\,. \qquad (8)$$

Parameter estimates are obtained by maximization of (8) using simulation methods.

5. Results and discussion

Of the many model specifications we estimated from the data, we present seven in Table 3: models 1 and 2 are MNL, while 3 and 4 are panel mixed logit models. Models 2 and 4 are the heteroskedastic versions of models 1 and 3. (Note the model mnemonics mentioned in the header of Table 3; these will be used interchangeably with the model numbers during discussion.) Models 5, 6 and 7 are the core models of this paper, as they introduce the choice-set formation stage via the independent availability equations (IAv) and maintain the parameterization of the scale function. Model 5 has no random coefficients, while models 6 and 7 do so, hence use panel estimators. Model 7 differs from model 6 because random coefficients are also included in the site availability function, which instead are fixed parameters in model 6. Note that both models 6 and 7 have been "tested down" by fixing to zero all variance coefficients that were individually insignificant, in order to avoid parameter proliferation.

--- Table 3 about here ---

Estimation for all model specifications was carried out based on a random subsample of the entire dataset (~90% of respondents, or 1,304 out of 1,452), with the aim of using the estimated parameters to evaluate forecasts on the observed choices of the subsample held out (~10% of original data or 148 randomly selected respondents).

From Table 3 we observe that, with the exclusion of model 5, there is a gradual improvement in terms of log likelihood values, with model 7 being the best performing in terms of fit to the data. When the number of parameters is accounted for, the AIC information criterion supports the selection of Model 7 (Stochastic IAv & Het-MXL), whereas the BIC information criteria supports the selection of Model 4 (Het-MXL) – though note that Model 7 is ranked second according to the BIC.[4]

Allowing for scale heterogeneity (preference discrimination) always improves fit, and so does allowing for coefficient variation across respondents for both site selection and site availability equations. Importantly for the main theme of this paper, allowing for choice set formation always improves on their counter-parts, regardless of whether coefficients are random or fixed (note, model 5 should be compared with models 1 and 2).

[4] Using a criterion that penalizes more for parameter proliferation such as the corrected AIC (Hurvich and Tsai, 1989) the selection of model 3 (MXL) is supported.

The signs and the significance of the indirect utility parameter estimates is stable across all models. In general, the addition of choice set formation equation decreases the precision of the beta estimates. Of course, this is partially offset by the benefits of recognizing the role of constraints, motivations and site-specific effects in the probabilities of site inclusion.

5.1 Indirect utility coefficients

We start by focusing on the results for the utility function parameters. Across destination sites, Val del Mis and Val dell'Ardo seem to be consistently the most appreciated sites, followed at some distance by Passo Croce. The coefficients for travel cost effects are negative and significant. The site attributes coefficients are identified at the single attribute level and tend to have the correct order of magnitude. Visitors show a clear preference towards bivouacs that are always open as well as equipped with food and wood; in contrast, destinations with bivouacs with "access upon request" (which implies having to ask ahead for the key) as well as those with "no access" are unattractive options. Respondents are definitely against restrictions of vehicular access to Val del Mis and Val Canzoi during the entire weekend, even if a shuttle service is offered. But they appreciate the opportunity to observe wild animals (like wolves and chamois) from close up, as shown by the positive and significant coefficient related to the wildlife sites variable. Generalist visitors tend to have little appreciation for technical features, such as cable extensions and climbing routes, so it is unsurprising that their coefficients are insignificant. The same applies for the constraints, which are likely not binding for generalists' activities, although the large standard deviation estimates suggest a strong heterogeneity of preference for this attribute. A similar comment applies to the estimates for additional MTB routes.

5.2 Preference discrimination (heteroscedasticity) coefficients

Being male decreases preference discrimination (equivalently, increases variance) significantly, as does having a high activity level and not reporting household income. *Age* has no effect in model 4 (Het-MXL) but it does in all the heteroscedastic models that include choice set formation. *Household income* only has a statistically significant effect in model 4, for those who reported income; not reporting income, however, is consistently associated with a decrease in scale in all heteroscedastic models. Somewhat unexpectedly, we find that the *maximizer score* is insignificant for preference discrimination (though we direct the reader to our online appendix, which includes some model specifications in which this individual characteristic is a significant predictor of preference discrimination).

5.3 Availability coefficients

Models 5, 6 and 7 add choice set formation (via availability functions) into the choice framework. As described earlier, the availability model takes as arguments the individual constraints plus the motivations respondents associate with each site, including the site-goal interactions. Model 6 adds preference heterogeneity in the utilities explaining site selection to Model 5. Model 7, with 82 estimated parameters, further adds heterogeneity to the availability coefficients explaining site inclusion in the individual choice sets, which are fixed in Models 5 and 6.

There are two facts worth noting when moving from the fixed coefficient Model 5 to its random coefficient counterparts (Models 6 and 7): *i*) the signs for all the site intercepts in the availability function change; *ii*) most of the significant coefficients for the motivations-site interactions lose significance once availability heterogeneity is addressed and little explanatory power is added even when heterogeneity is accounted for in the site availability function. It would seem that adding stochastic heterogeneity to the utility function, while substantially improving overall fit, markedly detracts from the insights provided by the site availability estimates, though not from the qualitative results of the site selection propensity. This effect is further exacerbated in Model 7, which further adds stochastic heterogeneity to the site availability functions. In this data, it seems that accounting for unobserved heterogeneity masks away much of the information related to choice set formation. This result would seem to place the analyst on the horns of a dilemma: should one sacrifice better fit for more detailed information on choice set formation in the availability equation? This is analogous to the dilemma often encountered in latent class analysis in which the addition of a class (i.e., adding taste heterogeneity), while improving model fit as measured by standard information criteria, often detracts from the significance and interpretability of utility coefficients reported in models with fewer classes.

In model 5 all ASCs have significant and negative coefficients in the availability function, indicating that the average propensity to include or exclude any site is, ceteris paribus, well below 50%. All else equal, the seven sites in this park are not very likely to be in individual choice sets until we start factoring in socio-demographic effects, individual attitudes and motivations; these factors then begin to contribute or detract to the propensity to include a given site. *Health problems* and *small kids* have negative and significant effect on the probability of inclusion, while *lack of money* has a positive and significant effect. Turning our attention to models 6 and 7 now, we note that we should draw the opposite conclusion, since all site intercepts are positive, even when random (model 7) and only few coefficients on motivations are significant. What this tells us is that when accounting for heterogeneity, the effect of motivations can be masked by the specification of stochastic taste variation.

In model 5, motivations (goals) are found to have a significant impact in choice set formation, especially in our extended availability function specification that separates effects by destination. The motivations *Contact with nature* and *Relax* display the largest single positive effects, whereas the largest single negative one is related to *acquire and/or improve skills*. It is apparent that the same goal can show quite different size effects depending on the specific destination. Overall, *Other goals, Contact with nature* and *Relax* are the ones that most commonly affect site selection significantly. Note that despite having carried out a specific pilot survey to identify the list of most relevant goals as perceived by visitors, the relative high incidence of *other goals* indicates that goals not included in our list are still only partially accounted for by this variable. This suggests that a more extensive set of goals should be considered in future applications for generalist visitors.

To aid readers in the interpretation of the availability functions, we note that these are simply logistic regressions. This means that we can interpret *exp*(coefficient) as the rate of change of the odds of inclusion, to wit,

$$\frac{\partial}{\partial X_{ik}}\left(\frac{A_i}{1-A_i}\right)=\exp(\beta_k). \tag{9}$$

So, if the coefficient for “Daytripper” in an availability function (e.g. model 5) is -0.12, that means that the same site is only [*exp*(-0.12)≈] 88.7% as likely (equivalently, 11.3% less likely) to be included in a day tripper choice set than in the choice set of a non-day tripper, ceteris paribus. In model 7 this effect is estimated to be -0.45 or [*exp*(-0.45)≈] 63.8%.

5.4 Elasticity and welfare estimates

Elasticity estimates, in the form of average percent change in visitation with reference to the base case, are obtained for models 4 to 7 and reported in Figure 2. The site attributes with strongest influence on visitation probabilities are *bivouacs* and *wildlife sites*, with *climbing routes* in third position. All other variables show little impact. The IAv & Het-MNL (Model 5) gives lower elasticity estimates for high impact attributes than equivalent estimates from mixed logit models. The likely reason for this was given by Swait and Ben-Akiva (1985), who pointed out that the erroneous inclusion of alternatives that were screened from the choice set leads to the inference of a *weakened* impact of attributes on utility. The intuition behind this observation is that changes in attributes of omitted alternatives are irrelevant to parameter inferences in the true dgp; if these alternatives are nonetheless included, as they would be in the mixed logit models, it would imply that attribute sensitivity (say, of price) is weaker than if the alternatives were properly omitted.

Sample average values (over people and sites) for compensating variation (CV in Euros per person-visit) of each attribute level are computed for models 5-7. For a given choice set of sites, CV is the difference in the log-sums converted to a monetary unit using the estimated marginal utility of income (i.e., the coefficient of the travel cost variable). In the case of the choice set formation models, the reported values are actually weighted averages of CVs for individual choice sets, with choice set probabilities used as weights. These values are reported in Figure 3. Differences across models are not particularly noteworthy, and the most valuable attributes obviously mirror those with highest elasticity. For example, mixed logit models predict for bivouacs with food and wood availability that are always open have a value of €7-9 (per person-trip) more than the current state of being open only upon demand. However, fixed coefficient IAL only estimates this at €4. Similarly, creating fenced wildlife sites, currently unavailable, would increase welfare by approximately €4-5 (per person-trip) according to mixed logit models, but around €3 according to fixed coefficient IAL. Finally, a negative impact of allowing access on weekends only by shuttle services would decrease welfare by approximately €1 (per person-trip).

--- Figures 2 & 3 about here ---

5.5 Change in visitation shares from policy scenarios

We explored changes in visitation probabilities of three policies for all four models. The first policy (Figure 4) would produce a scenario in which *congestion* would reach maximum levels in Val Del Mis and Candaten, which are sites with high level of visitors with different goals. Model 7 (Stochastic IAv & Het-MXL) predicts the largest shifts across sites. Unexpectedly, model 4 (Het-MXL) predicts a small increase in visitation at Val del Mis and decreases in Val di Lamen and Val Canzoi. Model 5 (IAv & Het-MNL) predicts the mildest changes, but along with Model 7 it coherently predicts a decrease in trip share in the two sites with additional congestion, albeit of much lower dimensions.

The second scenario (Figure 5) concerns the construction of three new MTB trails at Passo Croce, Val di Lamen and Val Canzoi. New trails should attract more visitors to these sites, which are chosen because of their relative diversity and the current lack of trails. In this case only model 5 (IAv & Het-MNL) predicts changes in visitation shares consistent with expectation, while all other models predict changes of much larger magnitude and in the unexpected direction.

--- Figures 4, 5 & 6 about here ---

The final scenario (Figure 6) involves making access to Val Canzoi and Val del Mis ten percent more costly, on the basis of the total travel cost. These are chosen because they are two main gateways to the park and both are car-accessible and positioned in valleys so that access fees are easy to administer. All models provide predictions of changes consistent with expectation, but the size of the change is two to three times larger in the three panel models, with the largest change predicted by the best fitting model 7 (Stochastic IAv & Het-MXL) of nearly 25 percent decrease in shares for each site subject to the change.

5.6 Forecasting out-of-sample

Model validation via out-of-sample forecasting is often a persuasive argument in specification selection. To start with, some insight can be derived by looking at the sorted sample distributions of the contributions to the sample likelihood in the holdout sample, as estimated by each of the 7 models. These are reported in Figure 7 and show that mixed logit models accounting for choice set formation and scale variation display a more extensive range of sample likelihood values. The improvement in the range is stark for all random coefficient models compared to those with fixed coefficients, but it is further increased, especially in the tails, by the IAv models with random coefficients. We take this as good evidence of improvement of out-of-sample forecasting when accounting explicitly for choice set formation.

Additional evidence can be derived by looking at the percentages of correctly forecasted choices by the subsample held out from estimation at each site by each model. These are reported in Table 4, which also reports the log-likelihoods and pseudo R-square (rho-sq). In terms of likelihood the two mixed logit models do best, but we note that adding choice set formation via the IAv equation improves the holdout likelihood, whereas this is not so when moving from the het-MNL (model 2) to the IAv & Het-MNL (model 5). This suggests that the recognition of the panel structure may itself improve inference about choice set formation, leading to better out-of-sample forecast. This is an issue worth exploring in the future.

All models with availability functions tend to forecast better at those sites (Candaten, Val Cordevole, Valle dell'Ardo) that are geographically clustered. Candaten, in particular, is the site favored for picnic areas and family activities. However, apart from the obvious forecast improvement when moving from fixed coefficient models to those with random coefficients, there does not appear to be a clear winner in out of sample forecast.

6. Conclusions

Increasing the effectiveness of public expenditures for the management of conservation areas with recreational use is arguably amongst the most challenging tasks currently faced by area managers. This is particularly true in locations in which large scale implementation of access fees is still a politically unviable proposition due to a sense of entitlement broadly held across the population of visitors. Accurate prediction of social effects of park management changes requires sophisticated tools. We expanded the degree of realism of simple destination choice models to provide a composite picture of a multi-layered preference structure. From such a picture, important insights are derived that can drive better targeted public expenditures in protected areas with recreational use. Determinants of choice set composition can explain the probability of inclusion and exclusion of sites from choice sets actually used in site selection decisions, above and beyond the mere preference intensity for site attributes. Motivations and personal constraints can be used as levers to attract a wider number of visitors of a certain type (e.g., families, or activity-specialized visitors) or even to calibrate crowding. For example, managers can use this information to plan the provision of a bundle of specialized sites catering to different categories of visitors and market the differential supply using location-brand strategies. This can be accompanied by a plan that builds on local vocations or/and complementarities of sites to cater for certain segments driven by similar motivations (such as relaxation and skill improvements, for example).

A further contribution of the increased realism of the proposed site availability models can be seen in the role the additional insights can have to improve the cost-effectiveness of monitoring systems for site access. Such systems are costly to plan, develop and maintain. They are nevertheless necessary to evaluate the management activity of protected areas with respect to their various institutional goals, one of which is obviously as an attraction for visitors. Our model can be used to provide a clear indication of what is worth monitoring at each of the sites while also providing guidelines for visitor evaluation surveys. These can be matched to visitor types as determined by visit motivations and personal constraints, something that can also guide promotions across the population of visitors and inform territorial marketing.

Should endogenous choice set formation be an integral part of destination choice models? We believe that it is necessary to also undertake a case-specific analysis of the trade-off between credibility of assumptions and effects of their use in the derivation of estimates (see Manski 2013). When moving from the fixed coefficient heteroskedastic IAv model to the standard mixed logit, the issue is one of (subjective) degrees of credibility between distributional assumptions of taste and imposing more structure to the choice process by assigning a role to choice set formation (more generally, screening of sites). If either a priori reasoning or qualitative work indicate a potential role for choice set formation in site selection, and its identification would be helpful to policy makers, we believe it warranted and recommended to explore the kinds of models estimated in this paper despite the recognized practical difficulties associated with choice set formation modeling. While random coefficient models (e.g., mixed logit) are known to be quite flexible in adapting themselves to manifold underlying data generation processes, they are not to be recommended if screening processes are present in the data. We believe that analysts should be open to sacrificing a small improvements in fit to gain credibility in model structure.

Our results demonstrate the significant role of choice set formation in site selection by using simultaneously estimated two-stage models with and without panel structure. Results show it to be a phenomenon not to be ignored in stated preference elicitation methods, just as it was shown to be the case in revealed preference data. We believe that future research needs to be conducted about this first stage of choice. While the added complexity of allowing for choice set formation is undeniable, *per se* this is no justification for not seeking econometric specifications that enable us to better approximate the phenomenon. Swait (2001a), for example, indicates that the GenL model can use a limited number of choice set candidates, motivating interest in the possibility that we can seek interesting ways to justify that not all choice sets are possible; e.g., mental maps of an area may be a way of capturing consumer heterogeneity in (limiting) choice sets. Swait and Feinberg (2013) suggest that approximations and dimensional reduction may be two strategies to model specification of choice set formation that may address the complexity of the screening stage.

Finally, we have sought to show that understanding people's motivations towards consumption of environmental "goods" helps analysts predict their choice set formation. Our modeling approach to motivations has been relatively unsophisticated, and calls for research to help us understand how these motivations are activated and what gives them importance. This kind of knowledge will aid the profession in predicting how policy impacts may, in turn, affect goal activation and importance. Future research should explore further the important role of motivations and compare endogenous choice set formation models with other exogenous choice set formation rules.

7. References

Adamowicz, W. & Boxall P.T., (1995), The influence of choice set considerations in modelling the benefits from improved water quality. *Water Resources Research* 31(7):1781-1787.

Andrews, R. L. & Srinivasan, T. (1995), 'Studying consideration effects in empirical choice models using scanner panel data', Journal of Marketing Research, 30-41.

Ben-Akiva, M. & Boccara, B. (1995), 'Discrete choice models with latent choice sets', *International Journal of Research in Marketing* **12**, 9-24.

Bockstael, N. E.; Haneman, W. M. & Kling, C. L. (1987), 'Estimating the value of water quality improvements in a recreational demand framework', *Water Resources Research,* **23**, 951-960.

Bockstael, N. E.; Haneman, W. M. & Strand, I. (1987), 'Measuring the benefit of water quality improvements using recreation demand models', *Draft Report presented to the U.S. Environmental Protection Agency* under cooperative agreement CR-811043-01-0. Washington D.C.

Cascetta, E. & Papola, A. (2001), 'Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand', *Transportation Research Part C: Emerging Technologies* **9**(4), 249--263.

Chernev, A.; Bückenholt, U. & Goodman, J. (2010), 'Commentary on Scheibehenne, Greifeneder, and ToddChoice Overload: Is There Anything to It?', *Journal of Consumer Research* **37**(3), 426--428.

Caulkins, P., R.C. Bishop, and N.W. Bouwes. (1986). The Travel Cost Model for Lake Recreation: A Comparison of Two Methods for Incorporating Site Quality and Substitution Effects. *American Journal of Agricultural Economics* 68(2):291–97.

DeShazo, J. R. & Fermo, G. (2002), 'Designing choice sets for stated preference Methods: the effects of complexity on choice consistency', *Journal of Environmental Economics and Management* **44**, 123-143.

Fotheringham, S. A. (1988), 'Consumer Store Choice and Choice Set Definition', *Marketing Science* **7**, 299-310.

Haab, T. C. & Hicks, R. L. (1997), Accounting for choice set endogeneity in random utility models of recreation demand, *Journal of Environmental Economics and Management*, 34, 127-14

Haab, T. C. & Hicks, R. L. (2000), Choice Set Considerations in Models of Recreation Demand: History and Current State of the Art, *Marine Resource Economics*, 14, 271-28.

Hauser, J. (2014), 'Consideration-Set Heuristics', *Journal of Business Research* 67, 1688-1699.

Herriges, J. A. & Kling, C. L. (1997), 'The Performance of Nested Logit Models when Welfare Estimation is the Goal', *American Journal of Agricultural Economics* **79**(3), 792--802.

Hicks, R. L. & Strand, I. E. (2000), 'The extent of information: its relevance for random utility models', *Land Economics*, 374--385.

Horowitz, J. & Louviere, J. (1995) What is the role of consideration sets in choice modelling *International Journal of Research in Marketing*, 12, 39-54

Hurvich, M., and C. Tsai. 1989. 'Regression and Time Series Model Selection in Small Samples', *Biometrika* 76:297–307.

Krantz D. & Kunreuther, H. (2007), 'Goals and plans in decision making,' *Judgment & Decision Making* **2**, 137--168.

Kreps, D. M. (1979), 'A representation theorem for “preference for flexibility"', *Econometrica*, 47(3):565--577.

Li, L. (2013), 'The Role of Antecedent Volition on Consumer Evaluative Processes and Choice Behavior,' Doctoral Dissertation, School of Business, University of Alberta, Fall 2013.

Li, L., Adamowicz, W. & Swait J. (2015), The effect of choice set misspecification on welfare measures in random utility models, *Resource and Energy Economics*, 42:71-92.

Nenkov G. Y., Morrin M., Ward A. Schwartz. B. Hulland. J. (2008), 'A short form of the Maximization Scale: Factor structure, reliability and validity studies', *Judgment and Decision Making* **3**(5), 371-388.

Manski, C. F. (1977), 'The structure of random utility models', *Theory and Decision* **8**, 229-254.

Manski, C. F. (2013), Public Policy in an Uncertain World: Analysis and Decisions, Harvard University Press, Boston, Massachusetts, USA.

McFadden, D. (1978), Modeling the Choice of Residential Location, *in* A. Karlqvist; L. Lundqvist; F. Snickars & J. Weibull, ed., 'Spatial Interaction Theory and Planning Models', North-Holland, Amsterdam, pp. 75-96.

Morey, E.; Rowe, R. & Watson, M. (1993), 'A Repeated Nested-Logit Model of Atlantic Salmon Fishing', *American Journal of Agricultural Economics* **75**, 578-592.

Morey, E. & Thiene, M. (2012), 'A parsimonious, stacked latent-class methodology for predicting behavioral heterogeneity in terms of life-constraint heterogeneity', *Ecological Economics* **74**, 130–144.

Parsons, G. R. & Hauber, A. B. (1998), 'Spatial boundaries and choice set definition in a random utility model of recreation demand', *Land Economics* **74**(1), 32--48.

Parsons, G. R. & Kealy, M. J. (1992), 'Randomly drawn opportunity sets in a random utility model of lake recreation', *Land Economics* **68**(1), 93--106.

Parsons, G. R.; Massey, D. M. & Tomasi, T. (2000a), 'Familiar and favorite sites in a random utility model of beach recreation', *Marine Resource Economics* **14**, 299--315.

Parsons, G. R.; Plantinga, A. J. & Boyle, K. J. (2000b), Narrow choice sets in a random utility model of recreation demand, *Land Economics*, 86-99

Peters, T., Adamowicz, W. L. & Boxall, P. C. (1995), 'Influence of choice set considerations in modeling the benefits from improved water quality', *Water Resources Research* **31**(7), 1781--1787.

Phaneuf, D. J.; Kling, C. L. & Herriges, J. (2000), 'Estimation and Welfare Calculations in a Generalized Corner Solution Model with an Application to Recreation Demand', *Review of Economics and Statistics* **82**, 83-92.

Provencher, B. & Bishop, R. (2004), 'Does accounting for preference heterogeneity improve the forecasting of a random utility model? A case study', *Journal of Environmental Economics and Management* **48**(1), 793-810.

Provencher, B. & Bishop, R. C. (1997), 'An Estimable Dynamic Model of Recreation Behavior with an Application to Great Lakes Angling', *Journal of Environmental Economics and Management* **33**, 107-127.

Richardson, A. (1982), 'Search models and choice set generation', *Transportation Research Part A: General* **16**(5), 403-419.

Roberts, J. & Nedungadi, P. (1995), 'Studying consideration in the consumer decision process: Progress and challenges', *International Journal of Research in Marketing* **12**(1), 3--7.

Roberts, J. H. & Lattin, J. M. (1991), 'Development and Testing of a Model of Consideration Set Composition.', *Journal of Marketing Research (JMR)* **28**(4).

Scheibehenne, B.; Greifeneder, R. & Todd, P. M. (2010), Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload Journal of Consumer Research, 37, 409-42

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: Happiness is a matter of choice. Personality and Social Psychology, 83, 1178–1197.

Simon, Herbert (1991). 'Bounded Rationality and Organizational Learning'. *Organization Science* **2** (1): 125–134. doi:10.1287/orsc.2.1.125.

Swait, J. (2001a), 'Choice set generation within the Generalized Extreme Value family of discrete choice models', *Transportation Research* **35**(7), 643--666.

Swait, J. (2001b), 'A Non-Compensatory Choice Model Incorporating Attribute Cutoffs', *Transportation Research Part B*, **35**(10):903-928.

Swait, J. & Argo, J. (2012), 'The Pervasive Role of Goals in Choice Behavior', Working Paper, School of Business, University of Alberta, January 2012, 59pp.

Swait, J. & Ben-Akiva, M. (1985), 'An Analysis of the Effects of Captivity on Travel Time and Cost Elasticities', Annals of the 1985 International Conference on Travel Behavior, April 16-19, Noordwijk, Holland, 113-128.

Swait, J. & Ben-Akiva, M. (1987), 'Incorporating random constraints in discrete models of choice set generation', *Transportation Research Part B: Methodological* **21**(2), 91--102.

Swait, J. & Erdem, T. (2007), ' Brand Effects on Choice and Choice Set Formation Under Uncertainty', *Marketing Science*, **26**(5), 679–697.

Swait, J. & Feinberg, F. (2013), 'Deciding How to Decide: An Agenda for Multi-Stage Choice Modeling Research in Marketing,' *Handbook of Choice Modelling*, Stephane Hess and Andrew Daly, Editors, Edward Elgar:London, UK.

Thill, J.-C. (1992), 'Choice set formation for destination choice modelling', *Progress in Human Geography* **16**(3), 361--382.

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, New York.

Van Nierop, E.; Bronnenberg, B.; Paap, R.; Wedel, M. & Franses, P. H. (2010), 'Retrieving unobserved consideration sets from household panel data', *Journal of Marketing Research* **47**(1), 63--74.

von Haefen, R. H. (2008), 'Latent Consideration Sets and Continuous Demand System Models', *Environmental and Resource Economics* **41**, 363-379.

Weber, E. & Johnson, E. (2009), 'Mindful Judgment and Decision Making,' *Annual Review of Psychology*, **60**, 53-85.

Table 1 - Attributes and levels of DCE

N.	Attribute	Levels
1	Entrance fee	No fee / €2 / €6 / €10
2	Bivouacs	Unavailable / open upon request / always open / always open with facilities (food, wood)
3	Vehicular Access	Always open / open Mon-Sat with shuttle / open Mon-Fri with shuttle
4	Crowding	Less than 10 visitors / 10-20 / 21-40 / >40
5	Picnic sites	None available / 1 / 2 / 3 / 4 / 5 / 6 / 7
6	Wildlife sites	Not available, available
7	Via Ferratas	None available Iron cable along part of the path (baseline) Iron cable along the whole path Iron cable along the whole path plus artificial holds
8	Climbing routes along cliffs and crags	No routes / 10 / 20 / 30
9	Trails for MTBike	None available / 1 / 2 / 3
10	Thematic itineraries	None available / 1 / 2 / 3

Table 2 - List of goals, activities and constraints

Goals	Activities	Constraints
Relax	Hiking	Walking disability
Spend time with the family	Via-ferrata	Health problems
Acquire and/or improve skills	Climbing	Small kids
Knowledge of the territory	Mountain-biking	Lack of training
Contact with nature	Picnic	Lack of technical skills
	Photography	Constraints due to other people
	History interest	Lack of free time
	Religious interest	Lack of money
	Geology interest	
	Research & study	
	Wildlife observation	
	Flora & vegetation observ.	
	Landscape observation	

Table 3 – Parameter Estimates (cases=15,648; respondents=1,304)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	MNL	Heteroscedastic MNL (Het-MNL)	Mixed Logit (MXL)	Heteroscedastic Mixed Logit (Het-MXL)	Independent Availability & Heteroscedastic MNL (IAv & Het-MNL)	Independent Availability & Het-MXL (IAv & Het-MXL)	Stochastic Independent Availability & Het-MXL (Stochastic IAv & Het-MXL)
Log likelihood	-17746.3	-17718.7	-16891.1	-16840.1	-17567.4	-16837.5	-16798.7
Rho-Squared	0.030	0.032	0.077	0.080	0.040	0.080	0.082
Rho-Squared (Akaike)	0.028	0.029	0.074	0.076	0.036	0.076	0.077
Number of Parameters	34	42	53	61	73	75	82
Number of Respondents	1304	1304	1304	1304	1304	1304	1304
Deviance	35492.6	35437.3	33782.2	33680.3	35134.8	33675.1	33597.5
AIC	35560.6	35521.3	33888.2	33802.3	35280.8	33825.1	33761.5
BIC	35736.4	35738.6	34162.3	34117.9	35658.4	34213.1	34185.7
cAIC	35628.2	35647.5	34140.3	34186.2	35940.1	34540.4	34698.6
Utility Function			*Independent Normals [mean , std]*		*p-values*: *(0.05-0.10), **(0.01-0.05), ***(≤0.01)		
Site Constants							
Passo Croce d'Aune	-0.2781***	-0.2739**	[-0.440*** , 0.7492***]	[-0.462***, 0.734***]	-1.0764**	[-0.545***,0.589***]	[-1.121** ,0.673**]
Val di Lamen	-0.7415***	-0.7483***	[-1.0502***, 0]	[-1.0769***, 0]	-1.3589**	[-1.0961***, 0]	[-1.6747** ,0.310*]
Val Canzoi	-0.4515***	-0.4612**	[-0.6566***, 0.7487***]	[-0.681***, 0.735***]	-0.9402**	[-0.791***,0.565***]	[-1.1728** ,0.695**]
Val del Mis	0.0235	0.0644	[-0.1206 , 1.1099***]	[-0.1584, 1.073***]	-0.347*	[-0.288***,0.888***]	[-0.6543**,1.0858**]
Candaten	-0.5838***	-0.6074**	[-0.924*** , 0.8162***]	[-0.932***, 0.778***]	-0.7792**	[-0.915***,0.619***]	[-1.2124**,0.7487**]
Val Cordevole (Partenza Bianchet)	-0.3511***	-0.4007**	[-0.4898** , 0.400***]	[-0.492***, -0.370***]	-0.5102**	[-0.51***,-0.302***]	[-0.8249**, 0]
Valle dell'Ardo	-0-	-0-	-0-	-0-	-0-	-0-	-0-

Bivouacs (base: Not available)							
Open upon request	-0.1408***	-0.1529***	[-0.1925***, 0]	[-0.1852***, 0]	-0.1054**	[-0.1551***, 0]	[-0.1943**, 0]
Always open	0.1244***	0.1471**	[0.1858*** , 0.1454**]	[0.182***, 0.150***]	0.0875*	[0.1472***, 0]	[0.1952**, 0]
Always open & facilities available (food, wood)	0.2974***	0.3275***	[0.3907*** , 0.5142***]	[0.370*** , 0.5183***]	0.2399**	[0.309***,0.418***]	[0.4104**,0.5466**]
Vehicular Access (base: always open)							
Closed Sunday (shuttle service)	-0.0735***	-0.089**	[-0.1022** , 0.2371**]	[-0.095**, 0.201***]	-0.0518	[-0.08** , 0.1919***]	[-0.1074* , 0]
Closed Saturday-Sunday (shuttle service)	-0.0323	-0.0403	[-0.0904* , 0.3639***]	[-0.078*, 0.3596***]	-0.0653*	[-0.068**, 0.270***]	[-0.1331* , 0.375**]
Crowding (base: no visitors)							
10-20 visitors	0.0329**	0.032	[0.0632**, 0]	[0.0584***, 0]	0.0202	[0.052***, 0]	[0.0646* , 0]
21-40 visitors	0.0124	0.0139	[0.0163 , 0.1933**]	[0.019 , 0.1886***]	-0.0081	[0.0118, 0.1594***]	[0.0217, 0]
More than 40 visitors	0.0143	0.0248	[-0.007 , 0.4359***]	[-0.0044, 0.438***]	-0.0102	[-0.0076, 0.347***]	[-0.0218 , 0.4565**]
Picnic sites (base: 0 sites)							
1 site	-0.0597	-0.0794	[-0.1059, 0]	[-0.0961, 0]	-0.0766	[-0.095* , 0]	[-0.1432, 0]
2 sites	-0.0177	-0.0341	[-0.0023, 0]	[-0.0168, 0]	-0.0529	[-0.0179, 0]	[0.0023, 0]
3 sites	-0.0335	-0.0312	[-0.0156, 0]	[-0.0231, 0]	-0.0118	[-0.0161, 0]	[-0.0067, 0]
4 sites	0.0131	0.013	[0.048, 0]	[0.0507, 0]	0.037	[0.0447, 0]	[0.0831, 0]
5 sites	0.0594**	0.068*	[0.0848**, 0]	[0.0851**, 0]	0.0509	[0.073**, 0]	[0.0886* , 0.3525**]
6 sites	0.0259	0.0336	[0.0234, 0]	[0.0294, 0]	0.0656	[0.0313, 0]	[0.0187, 0]
7 sites	0.1847***	0.2105**	[0.2067** , 0.3151***]	[0.205***, 0.326***]	0.1667*	[0.172***, 0.262***]	[0.2363**, 0]
Wildlife sites (base: not available)							
Available	0.2386***	0.2669***	[0.2987***, 0.3911***]	[0.284***, 0.394***]	0.2069**	[0.241*** , 0.317***]	[0.3373** , 0.411**]
Via ferrata (base: iron cable along part of the path -baseline)							
Iron cable along the whole path	-0.0628**	-0.0607	[-0.0305, 0]	[-0.0326, 0]	-0.022	[-0.021, 0]	[-0.0195, 0]

Iron cable along the whole path plus artificial holds	0.0208	0.0128	[0.0014 , 0.4966***]	[-0.003 , 0.4779***]	0.0256	[0.0005 , 0.394***]	[0.0111 , 0.4817**]
Number of climbing routes							
# Climbing Routes (L)	-0.0762***	-0.0875**	[-0.1134***, 0.1436***]	[-0.1118***, 0.155***]	-0.0044	[-0.086***,0.113***]	[-0.096**, 0]
# Climbing Routes (Q)	0.0404	0.0419	[0.0708 , 0.3777***]	[0.0757 , 0.3867***]	0.0215	[0.068*, 0.3046***]	[0.0905, 0]
Number of mountain bike trails (base: no routes)							
10 routes	0.0342	0.0377	[0.0567, 0]	[0.0483, 0]	0.0202	[0.0482* , 0]	[0.0409, 0]
20 routes	-0.0099	-0.0208	[-0.0217 , 0.2671***]	[-0.017 , 0.253***]	0.0128	[-0.0103, 0.204***]	[0.0042, 0]
30 routes	-0.0116	-0.0311	[-0.023 , 0.3922***]	[-0.017, 0.397***]	0.0437*	[-0.0124, 0.328***]	[-0.0069 , 0.4161**]
Number of thematic itineraries (base: none)							
1 itinerary	-0.0121	-0.0094	[-0.0037, 0]	[-0.0031, 0]	0.0016	[-0.0006, 0]	[-0.0032, 0]
2 itineraries	0.0155	0.0129	[0.0166, 0]	[0.0175, 0]	0.0051	[0.0097, 0]	[0.0205, 0]
3 itineraries	0.0565***	0.0502**	[0.0855**,0.2456***]	[0.082***,0.244***]	0.0632*	[0.072***, 0.190***]	[0.1085**,0.2646**]
Travel cost							
Travel Cost (€/trip), Daytripper	-5.0399***	-5.2009***	[-7.4982***, 0]	[-7.405***, 0]	-8.9099**	[-6.4823***, 0]	[-10.4508**, 0]
Travel Cost (€/trip), non-Daytripper	-5.7978***	-5.9661***	[-8.2421***, 0]	[-8.1996***, 0]	-8.8166**	[-7.1505***, 0]	[-10.8776**, 0]
ln(Scale Function)							
Male	-0-	-0.2405***	-0-	-0.0011	-0.2875***	-0-	-0.1178**
Age (L)	-0-	1.1253**	-0-	0.6719	4.3567***	0.3437***	1.2001**
Age (Q)	-0-	-0.5264*	-0-	-0.152	-2.2979***	-0-	-0.427*
Activity Level	-0-	-0.8841***	-0-	-0.1011	-1.3795***	-0-	-0-
Maximization Propensity (L)	-0-	-1.0892	-0-	-1.702	-0-	-0-	-1.2732
Maximization Propensity (Q)	-0-	-0.9833	-0-	-1.1255	-0-	-0-	-2.8132
Household Income (€/year)	-0-	-0.2461	-0-	0.4136**	-0-	-0-	0.2818
Household Income (Missing)	-0-	-0.2529***	-0-	-0.1371*	-0.3625***	-0.2739***	-0.2079**

Site Availability Functions			
Intercepts			
Passo Croce d'Aune	-0.5595***	2.1963***	[3.861*** , 1.5803***]
Val di Lamen	-0.6115***	2.7476***	[2.504*** , 1.637***]
Val Canzoi	-0.7289***	3.7462***	[2.8657***, 0]
Val del Mis	-1.0553***	3.0223***	[3.9217***, 0]
Candaten	-0.9398***	3.517***	[1.622***, 0]
Val Cordevole (Partenza Bianchet)	-0.9509***	2.2991***	[2.4944***,1.0112**]
Valle dell'Ardo	-1.0734***	1.2691***	[1.1147***, 0]
Daytripper?	-0.1182***	-0-	[-0.4543***, 0]
Personal Constraints			
Health problems	-0.1636**	-0.9929***	-0-
Small kids	-0.171***	-0-	[0 , 1.9049***]
Lack of technical skills	-0-	-0-	[0 , 1.1733***]
Constraints due to other people	-0-	-0-	[0 , 1.5936***]
Lack of free time	-0-	-0-	[0 , 2.2286***]
Lack of money	0.2926***	0.765*	[2.0855**,3.224***]
Motivations: Passo Croce d'Aune			
Acquire and/or improve skills	-0.5439***	-1.0089*	-0-
Establish contact with nature	0.7717***	1.2768**	-0-
Motivations: Val di Lamen			
Spend time with the family	-0-	-0-	[0.7362**, 0]
Acquire and/or improve skills	-0.4975***	-1.3204*	-0-
Establish contact with nature	0.6028***	1.7056*	-0-
Other	-0.1565**	-1.0286	-0-
Motivations: Val Canzoi			
Acquire and/or improve skills	-0.3609***	-0-	-0-
Acquire knowledge of the	-0.1986***	-0-	[0 , 1.3772***]

territory			
Establish contact with nature	0.5211^{***}	-0-	-0-
Other	-0.0996^{*}	-0-	-0-
Motivations: Val del Mis			
Relax	0.1536^{***}	-0-	-0-
Acquire and/or improve skills	-0.3579^{***}	-0.8842	-0-
Acquire knowledge of the territory	-0.1553^{**}	-0-	-0-
Establish contact with nature	0.6496^{***}	-0-	$[0 , 1.8193^{***}]$
Other	-0.1159^{**}	-0.653	-0-
Motivations: Candaten			
Spend time with the family	0.2196^{***}	0.4475	$[0.805^{***} , 0]$
Acquire and/or improve skills	-0.7337^{***}	-2.3229^{*}	$[-0.994^{***} , 1.483^{***}]$
Establish contact with nature	0.4939^{***}	-0-	-0-
Other	-0.0782^{*}	-0-	-0-
Motivations: Val Cordevole (Partenza Bianchet)			
Spend time with the family	-0-	-0-	$[0 , 1.166^{***}]$
Acquire and/or improve skills	-0.2479^{***}	-0-	-0-
Establish contact with nature	0.4053^{***}	-0-	-0-
Other	-0-	-0-	$[0 , 1.3961^{**}]$
Motivations: Valle dell'Ardo			
Relax	-0-	-0-	$[0 , 1.3155^{***}]$
Acquire and/or improve skills	-0.2297^{***}	-0-	-0-
Acquire knowledge of the territory	-0.0802	-0-	-0-
Establish contact with nature	0.6375^{***}	0.7959^{***}	$[0.7286^{***} , 0]$
Other	-0-	-0-	$[0 , 1.2675^{***}]$

Table 4 – Percentages of correct site selections in the holdout sample (N=148) by the estimated models.

		% Correct Predictions						
	Observed choices	MNL	Het-MNL	MXL	Het-MXL	IAv & Het-MNL	IAv & Het-MXL	Stochastic IAv & Het-MXL
Passo Croce d'Aune	251	34.1	33.9	34.4	34.5	29.8	34.2	31.7
Val di Lamen	270	35.6	35.8	35.9	36.0	26.5	34.6	35.7
Val Canzoi	235	32.4	32.2	32.3	32.2	29.2	30.3	30.4
Val del Mis	204	29.0	28.6	28.5	28.4	31.4	27.8	25.6
Candaten	252	37.1	38.0	37.0	36.9	36.8	37.8	41.0
Val Cordevole (Bianchet)	269	32.5	32.4	32.6	32.6	38.8	34.0	34.1
Valle dell'Ardo	295	34.8	35.2	35.6	35.6	35.2	39.3	42.1
Log-likelihood		-2023.77	-2025.46	-1899.48	-1904.91	-2082.94	-1900.24	-1929.06
Rho-sq		0.025	0.025	0.085	0.083	-0.003	0.085	0.071
Pearson χ^2		0.438	0.439	0.441	0.443	0.450	0.451	0.481

Figure 1 – Examples of three- and seven-site choice tasks

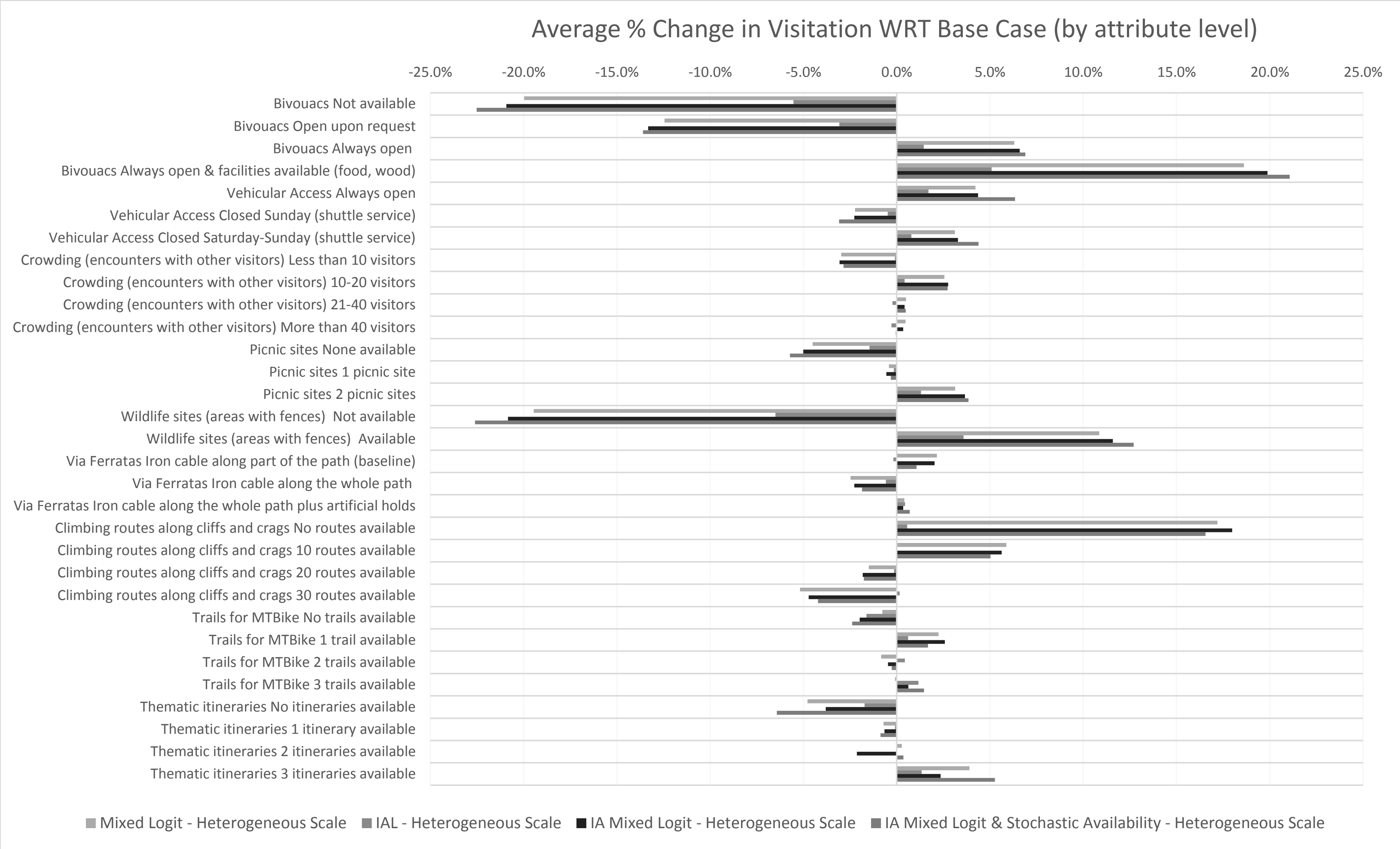
Choose one of the three sites. Assume those are the only available sites. (help)

Look at the map of the sites	**Val di Lamen**	**Partenza Bianchet**	**Val del Mis**
Bivouacs availability (?)	always open + food and fuel	Open upon request (key)	no
Access to Val Canzoi and Val del Mis (?)	Always open	Always open	Closed on Sunday (shuttle service)
Picnic areas (n) (?)	3	no	4
Via Ferratas features (?)	no	no	no
Number of visitors encountered (?)	Between 21 and 40 visitors	Between 10 and 20 visitors	Less than 10 visitors
Climbing itineraries (n) (?)	no	30	20
Mountain biking trails (?)	3	no	2
Thematic itineraries (n) (?)	no	2	no
Wildlife sites (?)	no	Available	no
Entrance fee (€) (?)	€2	€0	€2

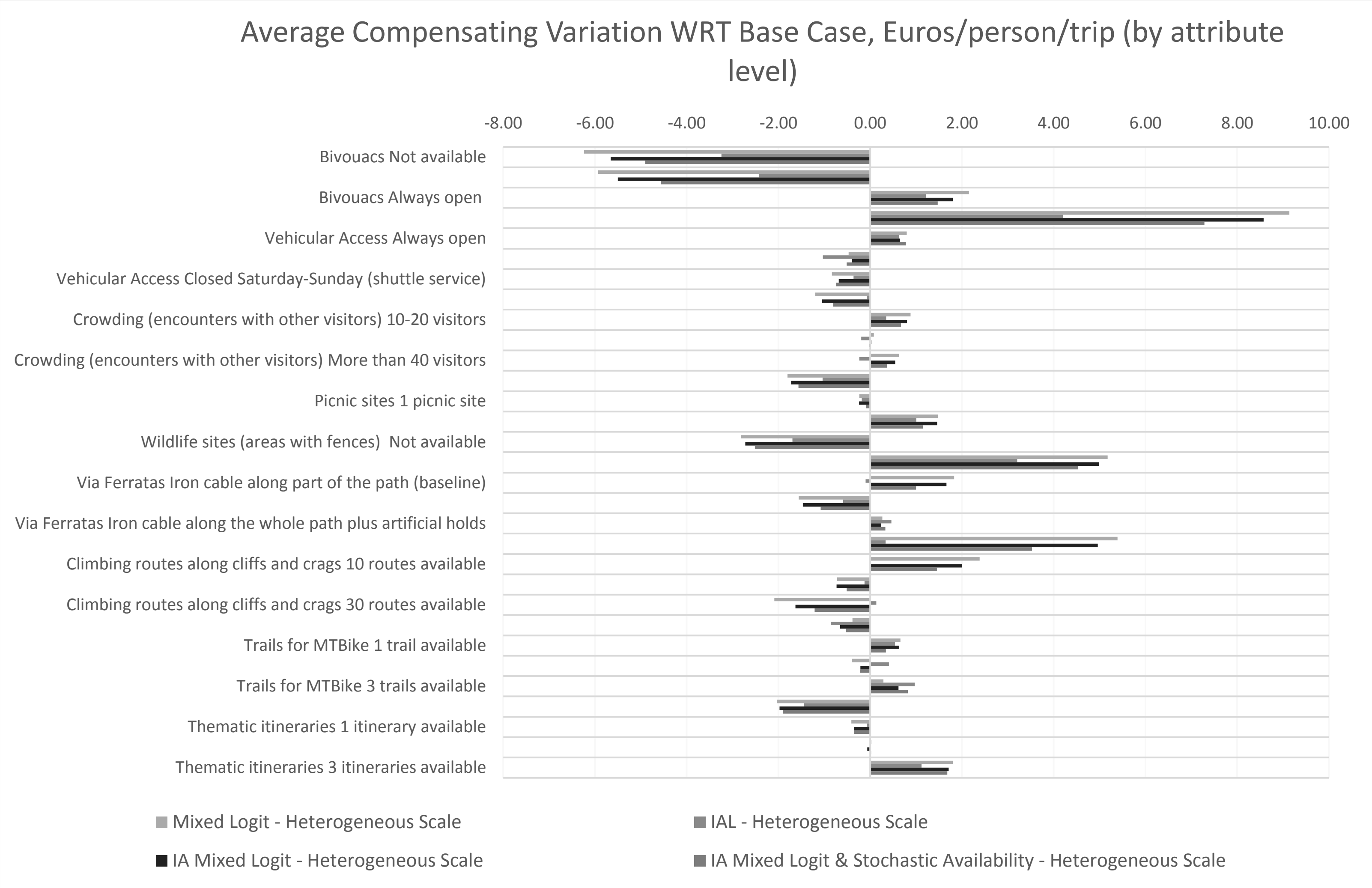
Choose one of the seven sites. Assume those are the only available sites. (help)

Look at the map	**Val Canzoi**	**Passo Aune**	**Partenza Bianchet**	**Valle dell'Ardo**	**Candaten**	**Val del Mis**	**Val di Lamen**
Bivouacs availability (?)	no	Always open	Open upon request (key)	Always open	no	no	no
Access to Val Canzoi and Val del Mis (?)	Closed on Sunday (shuttle service)	Always open	Always open	Always open	Always open	Always open	Always open
Picnic areas (n) (?)	6	2	no	3	7	3	2
Via Ferratas features (?)	no	no	Iron cable only where strictly necessary	Iron cable only where strictly necessary	no	no	no
Number of visitors encountered (?)	Less than 10 visitors	Less than 10 visitors	Between 21 and 40 visitors	Between 10 and 20 visitors	Between 10 and 20 visitors	Between 10 and 20 visitors	Between 21 and 40 visitors
Climbing itineraries (n) (?)	30	30	20	30	no	30	no
Mountain biking trails (?)	3	2	no	no	no	no	2
Thematic itineraries (n)(?)	no	2	no	2	no	1	no
Wildlife sites (?)	no	no	Available	no	Available	no	no
Entrance fee (€) (?)	€0	€2	€0	€10	€0	€2	€0

Figure 2 – Average % Change in Visitation with reference to Base Case (by attribute level)



995 **Figure 3 – Average Compensating Variation (€/site visit/person) with reference to Base Case (by attribute level)**



996

997

Figure 4 – Policy Simulation 1: Maximum Congestion at Val Del Mis and Candaten

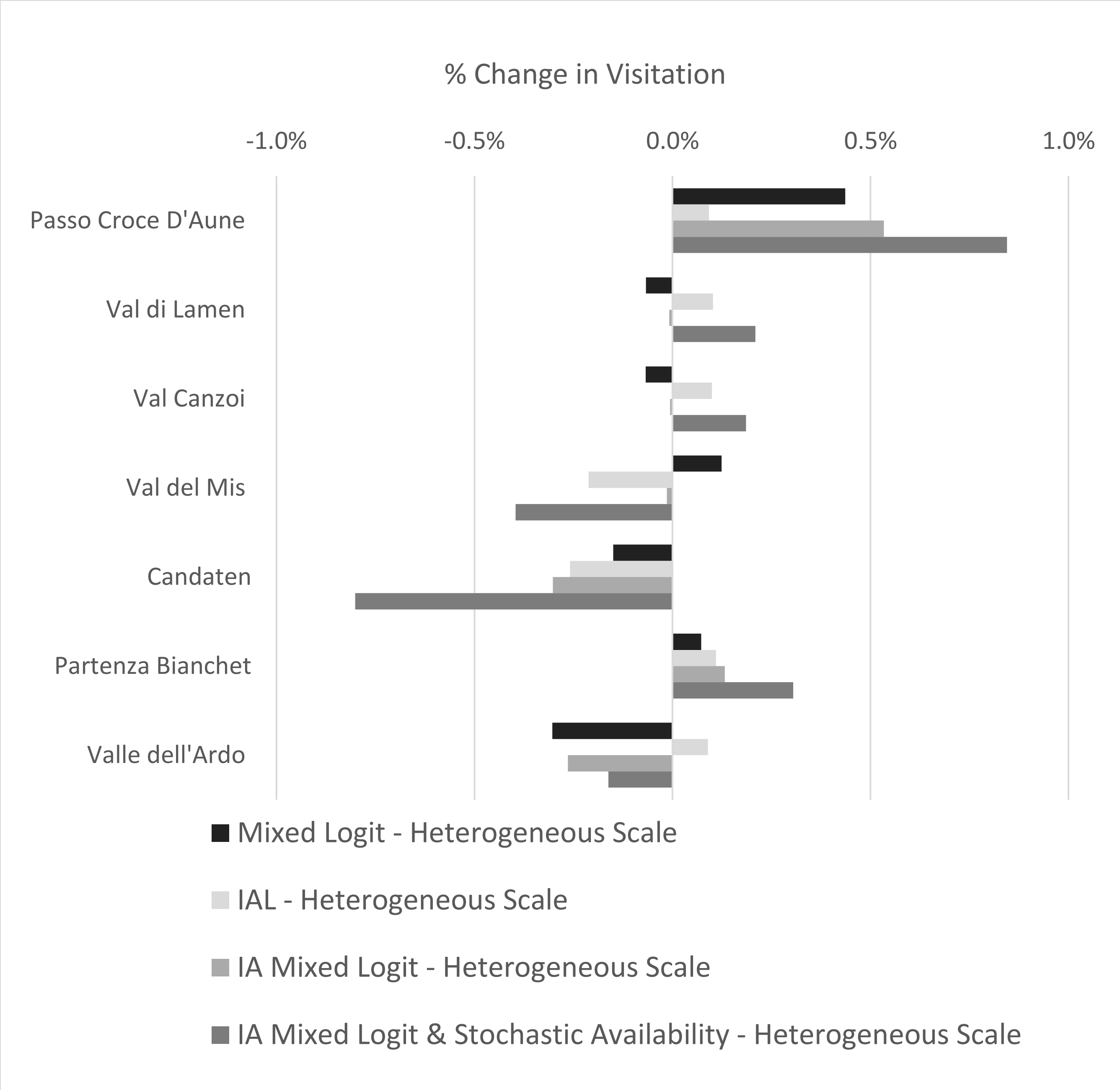


Figure 5 – Policy Simulation 2: Three MTB trails added at each of the three sites: Passo Croce d'Aune, Val di Lamen and Val Canzoi

Figure 6 – Policy Simulation 3: Increase travel cost by 10% to access Val Canzoi and Val del Mis

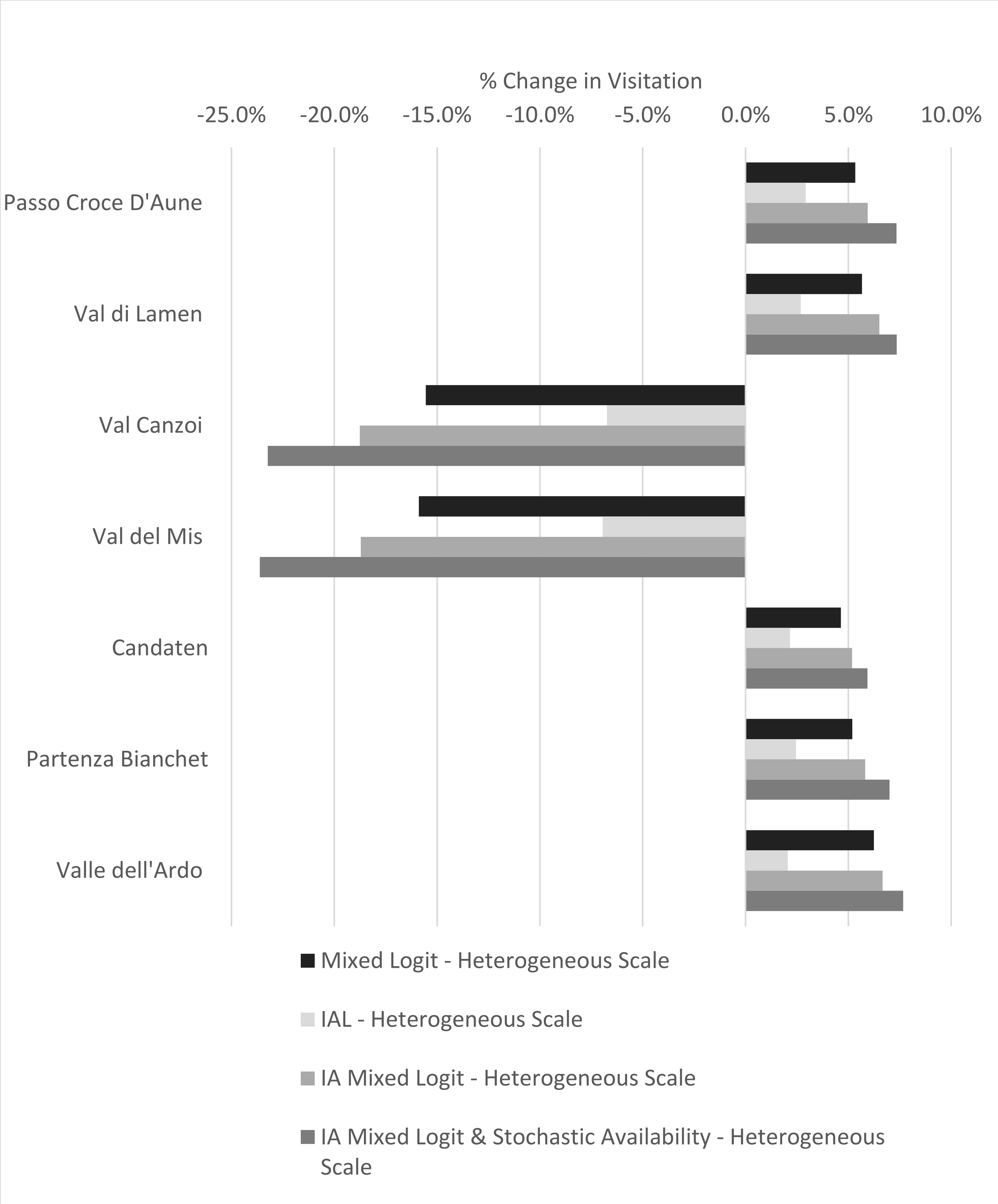
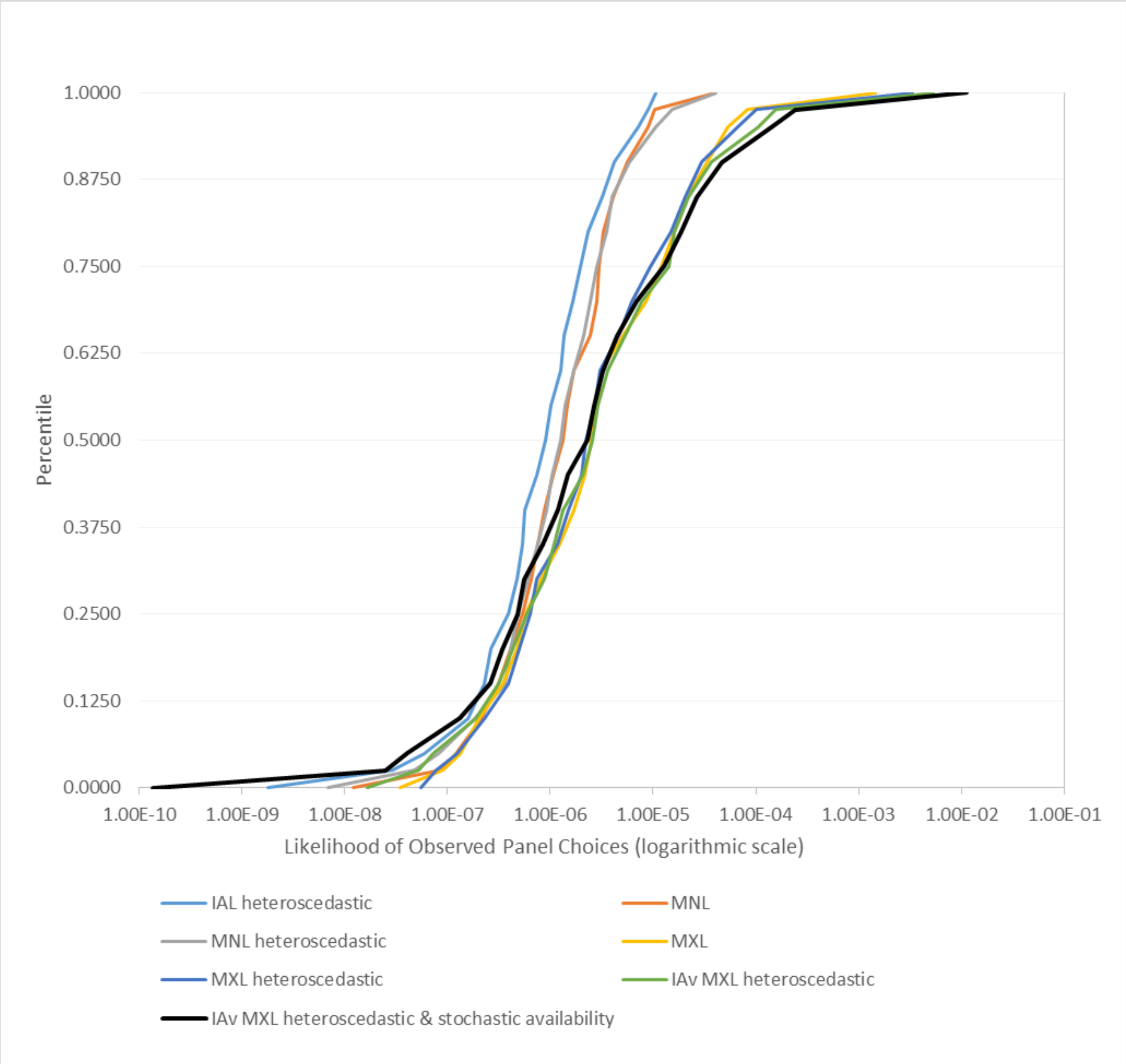


Figure 7 - Sample distribution of the contributions to the sample likelihood in the holdout sample for all estimated models



Online Appendix

Table A1 – Scale Items for the Maximizer Tendency behavioral profile.

Questions
1. I often read information on tourism destinations just out of curiosity.
2. I get bored with visiting the same locations even if they are good places to visit.
3. I shop around a lot for places to spend my vacations and outings just to find out more about my country.
4. I like introducing new places to visit to my family and friends.
5. I enjoy going to new places just to get some variety in my outings.
6. When choosing my destination I never consider a second option.

Table A2 – Models with goals and motivation interactions: cases=17,424, respondents=1,452.

(*** p<=0.01,** p<=0.05,* p<=0.1, blank p>0.1)

	Standard MNL Model	Heteroscedastic MNL Model	Independent Availability Logit Model (Homoscedastic)	Independent Availability Logit Model (Heteroscedastic)
Goodness-of-fit				
LL(Convergence)	-19467.40	-19435.70	-19359.10	-19323.50
Rho-Squared	0.0444	0.046	0.0497	0.0515
Akaide Rho-Squared	0.0406	0.0418	0.0431	0.0444
Number of Parameters	78	86	136	144
ChiSq (DF) wrt Model 1	-0-	63.4 (8)	216.6 (58)	287.8 (66)
Number of Respondents	1452	1452	1452	1452
Number of Choices	17424	17424	17424	17424
Utility Functions				
Passo Croce	0.0701	0.0214	1.0642 ***	0.2111 *
Val di Lamen	-0.1024	-0.0233	0.008	-0.0261
Val Canzoi	-0.0359	-0.0048	0.2368	0.0858
Val del Mis	0.1131	0.0474	1.124 ***	0.2467 *
Candaten	-0.1959 *	-0.0548	-0.3775 **	-0.1174
Partenza Bianchet	0.0401	0.0148	0.6693 ***	0.1216
Valle dell'Ardo	-0-	-0-	-0-	-0-
Fee L	-0.2299 ***	-0.0784 **	-0.2792 ***	-0.0771 **
Fee Q	0.1773 ***	0.0567 **	0.2157 ***	0.0563 *
Open on request	-0.1505 ***	-0.0498 **	-0.1854 ***	-0.0497 *
Always Open	0.1205 ***	0.0424 **	0.1492 ***	0.0421 *
Always Open & Fac.	0.3467 ***	0.1184 **	0.4202 ***	0.1173 **
Closed Sunday	-0.0499 *	-0.0204	-0.0636 *	-0.0209
Closed Sat & Sun	-0.122 ***	-0.0449 **	-0.1341 ***	-0.0395 *
10-20 Visitors	0.0513 ***	0.0161 *	0.0588 ***	0.015 *
21-40 Visitors	0.0105	0.0024	0.0112	0.0033
40 Plus Visitors	-0.0653 ***	-0.0208 *	-0.0701 ***	-0.0197 *
1 Picnic site	-0.1667 ***	-0.0584 *	-0.2156 ***	-0.0608 *
2 Picnic site	-0.0638	-0.0208	-0.1004 **	-0.0262
3 Picnic site	-0.0367	-0.0102	-0.0433	-0.0096
4 Picnic site	0.0427	0.0199	0.0759	0.0249
5 Picnic site	0.1008 ***	0.0329 *	0.1367 ***	0.0356 *
6 Picnic site	0.1321 **	0.0447	0.1713 ***	0.0448
7 Picnic site	0.2271 ***	0.0759 **	0.2786 ***	0.0742 *
Wildlife site	0.2487 ***	0.0866 **	0.2995 ***	0.0854 **
Cable along all	-0.0305	-0.011	-0.0428	-0.0115
Cable along all+Hold	0.0425	0.0143	0.052	0.0139
Climbing Routes L	0.0171	0.0071	0.0106	0.0049
Climbing Routes Q	0.011	-0.0016	0.034	0.0027

MBTrails=1	-0.0005	0.0005	0.0041	0.0019
MBTrails=2	0.0384 *	0.011	0.0442 *	0.01
MBTrails=3	0.0856 ***	0.027 *	0.0976 ***	0.0266 *
ThemItine=1	-0.0225	-0.0066	-0.0315	-0.0073
ThemItine=2	0.0394 **	0.0144 *	0.0482 **	0.0132
ThemItine=3	0.105 ***	0.0345 **	0.132 ***	0.036 *
Mobility Restric	-0.0449	-0.0177	-0.0941	-0.0471
Health Problems	-0.1482 **	-0.0515 *	-0.17 **	-0.0477
Small Children	-0.0758	-0.0244	-0.0747	-0.0185
Out of shape?	-0.0113	-0.0122	-0.0167	-0.0081
Lack training	-0.0406	-0.0112	-0.0556	-0.014
Other people	-0.0775	-0.0308	-0.071	-0.0263
Lack time	-0.0565	-0.0273	-0.0285	-0.0162
Lack money	0.0968	0.0261	0.1224	0.0316
PC:Goal:Relax	0.0112	0.0122	-0.6476 ***	-0.0549
PC:Goal:Stare fam	-0.0145	-0.0029	-0.0208	-0.008
PC:Goal:Acq/Mig abil	0.02	-0.0029	0.1296	0.0337
PC:Goal:Territorio	-0.0136	-0.0059	0.1688	0.0618
PC:Goal:ContattoNatu	-0.0957	-0.0335	-0.4631 ***	-0.1776 *
PC:Goal:AltriObietti	-0.0068	0.0006	-0.1265	0.0056
VL:Goal:Relax	0.1625 **	0.0548	0.1899 **	0.0555
VL:Goal:Stare fam	0.0035	0.0084	-0.0871	-0.0078
VL:Goal:Acq/Mig abil	-0.0564	-0.0193	0.0821	0.0254
VL:Goal:Territorio	-0.0059	-0.0024	0.1281	0.0516
VL:Goal:ContattoNatu	-0.1524 *	-0.0561	-0.4423 ***	-0.116 *
VL:Goal:AltriObietti	-0.0504	-0.017	0.0287	0.0103
VC:Goal:Relax	0.1216	0.0418	0.1538	0.039
VC:Goal:Stare fam	-0.0849	-0.0186	-0.1067	-0.015
VC:Goal:Acq/Mig abil	0.0225	-0.0074	0.0521	-0.0042
VC:Goal:Territorio	-0.1204	-0.038	-0.0921	-0.0207
VC:Goal:ContattoNatu	-0.2123 ***	-0.0743 *	-0.7173 ***	-0.2286 *
VC:Goal:AltriObietti	-0.0271	-0.0103	-0.0295	-0.0086
VM:Goal:Relax	0.2531 ***	0.0846 *	0.8195 ***	0.2165 *
VM:Goal:Stare fam	-0.1276 *	-0.0409	-0.5303 ***	-0.1055
VM:Goal:Acq/Mig abil	0.0092	-0.001	0.234	0.0485
VM:Goal:Territorio	-0.0604	-0.0265	-0.0601	-0.0251
VM:Goal:ContattoNatu	-0.0934	-0.0257	-0.4403 *	-0.1286
VM:Goal:AltriObietti	-0.0999 *	-0.026	0.2328	0.0621
Ca:Goal:Relax	0.0442	0.0213	-0.159	-0.0289
Ca:Goal:Stare fam	0.1835 ***	0.0627 *	0.3093 ***	0.1047 *
Ca:Goal:Acq/Mig abil	-0.4087 ***	-0.138 **	-0.4075 ***	-0.0811 *
Ca:Goal:Territorio	0.064	0.017	0.0486	0.0245
Ca:Goal:ContattoNatu	-0.0954	-0.0355	-0.0551	-0.0232
Ca:Goal:AltriObietti	-0.0636	-0.0166	-0.1153	-0.0149
PB:Goal:Relax	0.0012	0.0044	-0.1717	-0.0284
PB:Goal:Stare fam	-0.0411	-0.0114	-0.108	-0.0083

PB:Goal:Acq/Mig abil	0.0254	0.0025	0.4084 ***	0.1054 *
PB:Goal:Territorio	0.0459	0.0144	0.0404	0.0305
PB:Goal:ContattoNatu	-0.1924 **	-0.0648 *	-0.5957 ***	-0.1621 *
PB:Goal:AltriObietti	-0.0144	-0.0031	-0.1523	-0.0438
Scale Functions (natural logarithm)				
Male	-0-	-0.1456 ***	-0-	-0.195 ***
Age L	-0-	0.822	-0-	0.5111
Age Q	-0-	-0.5023 *	-0-	-0.3819
ActivityLevel	-0-	-0.493 **	-0-	-0.5075 **
Maximizer L	-0-	8.1923 ***	-0-	11.8384 ***
Maximizer Q	-0-	-16.8182 ***	-0-	-23.4211 ***
HH Income Missing	-0-	-0.2001 **	-0-	-0.3689 ***
HH Income '000s Eu	-0-	0.1491	-0-	0.1121
Availability Functions				
Passo Croce			1.5978 **	4.0375 ***
Val di Lamen			3.11 ***	5.611 ***
Val Canzoi			2.7006 ***	4.6305 ***
Val del Mis			1.4592 **	3.8834 ***
Candaten			6.2772 ***	7.4875 ***
Partenza Bianchet			2.104 ***	4.6067 ***
Valle dell'Ardo			3.6117 ***	5.3904 ***
Scenario Size			-0.1407 *	-0.1451 *
Male			0.2492 ***	0.4964 ***
Age L			-1.3146	-1.1432
Age Q			0.8291 *	1.0201
ActivityLevel			1.092 ***	1.6538 ***
Maximizer L			-2.5137	-22.7045 ***
Maximizer Q			3.5518	42.0609 ***
HH Income Missing			-0.2015	0.2136
HH Income '000s Eu			-0.5846 *	-0.85 *
PC:Goal:Relax			1.0174 ***	0.3981 *
PC:Goal:Stare fam			0.0177	-0.0062
PC:Goal:Acq/Mig abil			-0.2339	-0.3478
PC:Goal:Territorio			-0.5598 **	-0.72 **
PC:Goal:ContattoNatu			0.7125 ***	1.1148 ***
PC:Goal:AltriObietti			0.1925	-0.1002
VL:Goal:Relax			-0-	-0-
VL:Goal:Stare fam			0.5475	0.3132
VL:Goal:Acq/Mig abil			-1.0633 **	-1.0834 ***
VL:Goal:Territorio			-1.4611 ***	-1.6881 ***
VL:Goal:ContattoNatu			2.0183 ***	1.7656 ***
VL:Goal:AltriObietti			-0.7212 *	-0.6663 **
VC:Goal:Relax			-0-	-0-
VC:Goal:Stare fam			-0.028	-0.0998

VC:Goal:Acq/Mig abil	-0.4849	-0.2915
VC:Goal:Territorio	-0.5773	-0.5328
VC:Goal:ContattoNatu	2.7295 ***	2.4423 ***
VC:Goal:AltriObietti	0.05	-0.018
VM:Goal:Relax	-0.5537 **	-0.6383 **
VM:Goal:Stare fam	0.382 **	0.2713
VM:Goal:Acq/Mig abil	-0.2031	-0.2246
VM:Goal:Territorio	-0.1261	-0.0828
VM:Goal:ContattoNatu	0.3005	0.4085 *
VM:Goal:AltriObietti	-0.4403 ***	-0.4753 ***
Ca:Goal:Relax	3.3722 **	1.5142 ***
Ca:Goal:Stare fam	-2.9118	-2.3549 *
Ca:Goal:Acq/Mig abil	-1.3105 **	-1.464 ***
Ca:Goal:Territorio	-0.0471	-0.4428
Ca:Goal:ContattoNatu	-0.187	0.1957
Ca:Goal:AltriObietti	0.7831	-0.0841
PB:Goal:Relax	0.2744	0.2153
PB:Goal:Stare fam	0.1435	-0.0831
PB:Goal:Acq/Mig abil	-1.0418 ***	-1.0838 ***
PB:Goal:Territorio	-0.0811	-0.3248
PB:Goal:ContattoNatu	0.9065 ***	1.0227 ***
PB:Goal:AltriObietti	0.2954	0.3464 *
VA:Goal:Relax	0.2044	-0.319
VA:Goal:Stare fam	0.0157	0.5782 **
VA:Goal:Acq/Mig abil	0.0588	0.0951
VA:Goal:Territorio	0.4346	0.1771
VA:Goal:ContattoNatu	-0.3102	0.1697
VA:Goal:AltriObietti	0.4033 *	0.4202 *
Distance/100	-0.665	-1.112 **
(Distance/100)^2	0.066	0.2426

Figure A1 - Map of Dolomiti Bellunesi National Park