[under revision – Please do not cite or quote without permission from the authors]

Demand for forest recreation in Lorraine: Revealed and stated preferences

Jens Abildtrup, INRA, UMR356 LEF, 54000 Nancy, France

Serge Garcia INRA, UMR356 LEF, 54000 Nancy, France

Søren Bøye Olsen, University of Copenhagen, 1958 Frederiksberg C, Denmark

Anne Stenger INRA, UMR356 LEF, 54000 Nancy, France

Abstract

This paper analyses the use of forests for recreational purposes in Lorraine, France. This is a region with many forests and easy access for recreational users. This implies that residents in Lorraine can choose between a large set of forests if they decide to go for a forest visit. The abundance of forests in Lorraine makes identification of the visited forests difficult. To facilitate identification of forests actually visited we incorporate an interactive map in a web-based survey intended to gather both revealed and stated preference data. We compare different sampling schemes to define the choice set used for site selection modeling when the actual choice set considered is unknown and potentially large. The easy access to forest implies also that around half of the visitors walk or bike to the forest. We apply an error component mixed logit model to model the travel mode decision and the site selection decision simultaneously and to combine revealed and stated preference data. Finally, we compare the welfare effects of changes in quality and access to forests based on alternative choice set specifications, model specifications and data sources (revealed and stated preference data).

Keywords: Travel cost method; Choice experiment; Choice set; error component, mixed logit; travel mode choice; recreation; forest;

Introduction

The recreational use of forests is an important non-marketed service provided by forest ecosystems (Hanley et al. 2003). There are many factors influencing the recreational value. They include characteristics of the forests, the access to the forests, and substitute and complementary recreation sites. Forest characteristics having an impact on the users' utility include forest structure attributes , e.g. dominant tree species, and the presence of recreational facilities, e.g. parking places and trekking paths (Scarpa et al. 2000; Termansen et al. 2004; Termansen et al. 2008; Bestard and Font 2009; Abildtrup et al. 2013). An assessment of non-market value and their determinants are important for public forest managers when they have to allocate resources to forest management or to improvement of the recreational quality of forests.

In this study we analyse the determinants of the economic value of forests in a region with relative easy access to forests, applying site selection models (Bockstael et al. 1987) combining revealed and stated preference data. In Lorraine which is our study region about 33 % of the land use is forest and about half of the forest is public-owned with open access. In addition the large majority of private forest is not closed for public use. The public has access if the forest is not fenced or does not have a sign which marks that entering the forest is not allowed.

The easy access to forest raises three challenges for the economic valuation using, site selection models. First of all, there are many alternative forests to choose between for a potential visitor. The random utility model (RUM) is typically considered as appropriate when the access to alternative sites should be considered in the valuation of policy scenarios. Even though the RUM model is capable of coping with a large choice set of several hundred alternatives (Termansen et al. 2004), including all forest considered in Lorraine (more than 5000 forest recreation units) is computationally impractical. Several studies have considered different approaches to reduce the size of the choice set (Haab and Hicks 2000) using, for example, random sampling and grouping of alternative sites (e.g. Parsons and Kealy 1992; Parsons et al 2000). In the present study we apply a strategic sampling scheme suggested by Lemp & Kockelman (2012). They show that this sampling approach, which draws alternatives in proportion to updated choice-probability estimates, provides substantial efficiency benefits and reduces the bias associated with simple random sampling of alternatives when applying mixed logit (MXL) models.

Secondly, applying the RUM we have to identify the specific forest chosen by an individual respondent. The identification of the forest visited is difficult as respondents will not necessarily know the name of the forest visited, and forests in Lorraine do not have unique names. We have, therefore, introduced the use of online maps in our web-based survey. Web-based surveys are an alternative to traditional mail-based surveys (Fleming and Bowden, 2009; Marta-Pedroso et al. 2007; Olsen 2009) and have advantages in terms of implementation costs. However, the use of interactive maps in this survey is an innovative way of giving and collecting more precise information to and from the respondents. Since Doyle et al. (1998) explained the benefits of interactive mapping in a large range of contexts like planning and design issues in space, it seems that our study is the first one giving the respondent the possibility to identify the forest visited using an interactive map.

Thirdly, we were facing the choice of transport mode. The easy access to forests in our case study region implies that a large share of the visitors are either walking or bicycling to a forest. Only a few studies have considered other travel modes than car transport in estimation of the travel model, e.g. Bell and Strand (2003) using a nested logit model. We show how one can include the travel mode choice using a mixed logit error component model (Brownstone et al. 2000).

To better identify the determinants of recreational value of forests we combine revealed and stated preference data. The advantages of combining the revealed preference (RP) data with stated preference (SP) data include data enrichment, reduction of the problem of multicollinearity among characteristics in RP data, and the problem of endogeneity of attributes (Huang et al. 1997; Whitehead et al. 2008).

Our analysis contributes to previous French studies on the recreational use of forests by providing information on the impact of forest attributes on recreation site selection[1]. Examples of forest recreation valuation studies in France using the travel cost methods include Normandin (1998) who analysed the value of recreation in forests in Lorraine in 1997 based on a mail survey, and Peyron et al. (2001) estimated the loss of recreational value of forests in Lorraine due to the 1999 windstorm, and Peyron et al. (2002) and Garcia and Jacob (2009) analysed, on the national level, the demand for forest recreation based on a telephone survey carried out in 2002. However, none of these studies did model the site selection, considering the forest characteristics.

To sum up, the paper contributes to the recreation valuation literature by 1) a new innovative approach to identification of visited recreational sites, i.e. integrating an interactive map in a web-based survey, 2) applying the strategic sampling scheme suggested by Lemp and Kockelman (2012), 3) the modeling of the choice of transport mode explicitly. Finally, 4) this study is the first to combine RP and SP data on forest recreation, applying an error component mixed logit model.

[1] . An overview of French studies on the recreational value of forests can be found in Montagne et al. (2008)

The analysis is based on a survey carried out in 2010 on a random sample of residents in Lorraine. In the following we first describe our methodology. Then we describe the data and the data collection process before we present the results. Finally, we conclude the paper with a discussion of our results.

Methodology

The relative abundance of forest in Lorraine implies that the number of forests which a visitor can choose between, in their selection of a site to visit, is high. The site selection model based on the RUM (McFadden 1974) is therefore considered as an appropriate approach in the present case. This model facilitates the site selection choice taking into account the quality attributes of the forests.

The Kuhn-Tucker demand models applied, e.g., by Phaneuf et al. (2000), are also applicable to the analysis of demand for more than one recreational site, but they are more convenient when the number of sites considered is relative small[2]. An alternative approach based on an incomplete demand system specification suggested by Shonkwiler and Englin (2005) would allow the estimation of the unconditional demand. However, a main objective of the present study is to compare the RP data with data from a choice experiment (CE) focusing on the conditional choice, i.e. individuals who have already visited a forest.

Choice set and econometric models

Even in a RUM framework, estimation may be impractical with very large choice set. There are different approaches to cope with large choice sets, including aggregation of sites (Lupi and Feather 1998; Parsons et al. 2000), defining choice set on subjective measures (Hicks and Strand, 2000), and random sampling in the choice set (Parsons and Kealy, 1992). In our case, aggregation would not make sense due to the relatively even distribution of recreational facilities. Aggregation of sites would reduce the variation of the facility attributes. Furthermore, we consider relative local use of forest, implying that spatial high-resolution data are important. McFadden (1977) showed that random sampling in the choice set provides consistent parameter estimates as long as the independence of irrelevant alternatives property is not violated. However, the computational gains by reducing the sample size should be compared to the loss in estimator efficiency (Nerella and Bhat, 2004). Unfortunately, there is no proof that the sampling of alternatives in the choice set will generate consistent parameter estimates within the MXL framework. Note, that one of the advantages of the MXL model is that it does not require the irrelevant alternatives properties to be met. Empirical analyses indicate that potential bias may be of limited importance when samples are

[2] (Kuriyama et al.2010) considered 52 beaches in a study with trip information on 4367 trips taken by 617 respondents but most often only a few alternative sites have been considered.

more than 25 % of the considered choice set (Nerella and Bhat 2004; Domanski and Haefen 2010). Furthermore, empirical results provided by Lemp and Kockelman (2012) suggest substantial efficiency benefits, applying a strategic sampling strategy. This sampling scheme draws alternatives in proportion to updated choice-probabilities. The initial choice-probabilities are obtained by simple random sampling. To our knowledge, our study is the first where this approach has been implemented in the modeling of recreation. Other studies have applied different importance sampling strategies applying standard conditional logit models (Parsons and Kealy 1992) or applied *ad hoc* sampling strategies (Brownstone et al. 2000) .

The basic idea in the RUM is that the individual chooses from a number of alternatives and selects the one that yield the highest utility level on any given choice occasion. Assume that a forest visitor, *n*, has, *CS* possible multiattribute forest sites from which to choose. The utility perceived by the visitor from visiting forest *i* is assumed to be given by:

$$(1) \quad U_{ni} = \boldsymbol{\beta}'\mathbf{x}_i + \gamma(y_n - p_{ni}) + \varepsilon_{ni}$$

assuming a linear indirect utility function of visiting forest *i.* $\boldsymbol{\beta}$ is a parameter vector, $\mathbf{X}_i$ is a vector of variables describing the forest *i*, y_n is the income of visitor, *n*, and p_{ni} is *n*'s cost of visiting site *i*, and ε_{ni} is the stochastic element of utility. If the error terms are independently and identically drawn from an extreme value distribution, the RUM model is specified as a conditional logit (CL). This implies that the probability of choosing site *i* is the logit

$$(2) \quad P_{ni} = \frac{e^{(\boldsymbol{\beta}'\mathbf{x}_i + \gamma(y_n - p_{ni}))\lambda}}{\sum_{j \in CS} e^{(\boldsymbol{\beta}'\mathbf{x}_j + \gamma(y_n - p_{nj}))\lambda}}$$

where λ is the scale parameter. In the case of a large choice set *CS,* McFadden (1977) have shown that one can sample alternatives from that choice set and obtain consistent parameter estimates. This requires the sampling scheme meets the positive conditioning property, i.e. the sampling scheme must be such that the probability of drawing the subset of alternatives is positive regardless of the actual chosen alternative within that subset (see also Lemp and Kockelman 2012). Define $\pi(D_n \mid i)$ as the probability of generating alternative set D_n given actual choice *i* under the sampling scheme, then it can be shown that the choice probabilities in model estimations (2) must be adjusted as follows:

(3) $$P_{ni} = \frac{e^{\left(\boldsymbol{\beta}'\mathbf{x}_i + \gamma(y_n - p_{ni}) + \ln(\pi(D_n|i))\right)\lambda}}{\sum_{j \in D_n} e^{\left(\boldsymbol{\beta}'\mathbf{x}_j + \gamma(y_n - p_{nj}) + \ln(\pi(D_n|j))\right)\lambda}}$$

In the case where the probability of generating any choice set of non-chosen alternatives is equal the adjustment of the choice probabilities (*P_{ni}*) is not necessary. This would be the case with simple random sampling where all alternatives have the same probability of being included in the choice set.

In the present study we will take into account preference heterogeneity applying a MXL model where we allow the $\boldsymbol{\beta}$ to vary over individuals defined by the distribution $f(\boldsymbol{\beta} | \theta)$.

(4) $$P_{ni} = \int_{-\infty}^{\infty} \frac{e^{\left(\boldsymbol{\beta}'\mathbf{x}_i + \gamma(y_n - p_{ni})\right)\lambda}}{\sum_{j \in CS} e^{\left(\boldsymbol{\beta}'\mathbf{x}_j + \gamma(y_n - p_{nj})\right)\lambda}} f(\boldsymbol{\beta} | \boldsymbol{\theta}) d\boldsymbol{\beta}$$

However, equation (4) cannot be rewritten using a subset of all choice alternatives in the same way as equation (2) can be rewritten due to the introduction of the integral in (4). In the following we will use the approximation suggested by (Lemp and Kockelman 2012):

(5) $$P_{ni} \approx \int_{-\infty}^{\infty} \frac{e^{\left(\boldsymbol{\beta}' b_i + \gamma(y_n - p_{ni}) + \ln(\pi(D_n|i))\right)\lambda}}{\sum_{j \in D_n} e^{\left(\boldsymbol{\beta}'\mathbf{x}_j + \gamma(y_n - p_{nj}) + \ln(\pi(D_n|j))\right)\lambda}} f(\boldsymbol{\beta} | \boldsymbol{\theta}) d\boldsymbol{\beta}$$

which they apply in an iterative strategic sampling scheme. In the first step the choice set is obtained by simple random sampling where each alternative has equal probability of being sampled. Then choice probabilities are calculated based on model (5). Note that with simple random sampling we can ignore the adjustment terms $\ln(\pi(D_n \mid i)$ as the alternatives have equal probability of being selected for the choice set. The estimated choice probabilities are then used in the following first and second iterations of the strategic sampling scheme for selecting individual-specific choice sets. We use the sampling protocol suggested by Frejinger et al. (2009). They use sampling with replacement to make the calculation of the adjustment terms computationally less heavy. For comparison we estimate both the CL and MXL models based on simple random sampling and the strategic sampling scheme.

Integrating site and travel mode choice

To investigate the importance of the transport mode choice, we specify a model which explicitly accounts for transport mode choice based on the access to forests[3]. The intuition behind our model is that individuals living in neighborhoods with attractive forests will choose to walk or bike to the forest while respondents with no close or attractive forests in the neighborhood will go by car. Integrating the choice of transport mode in the choice of recreation site has previously been modeled by Bell and Strand (2003), using a nested model framework. However, we will use the mixed logit error component model which is more flexible (Bhat and Castelar 2002), i.e. we allow for heterogenous preferences.

Let individual *n*'s utility of visiting forest *j* be determined by the conditional indirect utility function[4]:

$$U_{nmj} = \begin{cases} \beta_n{}' x_j - \gamma_n{}' p_{nj}^{car} + \varepsilon_{nj}^{car}, & j = 1,...D_n; \\ \beta_n{}' x_{jt} - \gamma_n{}' p_{nj}^{bike} + \alpha_{bike} + \mu_n^{bike} + \varepsilon_{nj}^{bike}, & j = 1,...D_n; \\ \beta_n{}' x_j - \gamma_n{}' p_{nj}^{walk} + \alpha_{walk} + \mu_n^{walk} + \varepsilon_{nj}^{walk}, & j = 1,...D_n; \end{cases} \quad (6)$$

where β_n is a vector of parameters of person *n*, representing the marginal utility of forest attributes, and x_j is a vector of observed forest attribute variables that relate to alternative *j*. μ_n^m is a normally distributed error component with zero mean and variance σ_μ^m. This allows for travel mode dependent scale factors in the choice of forest. α_{bike} and α_{walk} are alternative specific constants for *bike* an *walk* transport modes, respectively. ε_{nj}^m is identically, independently distributed over alternatives, travel modes *m*={*car*, *bike*, *walk*}, and individuals and is unobserved by the researcher.

Comparing RP and SP data

Finally, we combine the RP data with the SP data from a choice experiment (CE) where respondents choose between two hypothetical forests (h_1^t and h_2^t) and the forest visited most often during the last 12 months (*sq*) (see the following section for the design of the CE). We apply a unified mixed-logit framework for joint analysis for revealed (RP) and stated choice (SP) data suggested by Bhat and

[3] Note, in the analysis of travel mode only the RP data are relevant as we did not consider travel mode in the choice experiment.
[4] Here we have excluded income as it enters with the same value in all alternatives and therefore cancels out.

Castelar (2002) and Hensher et al. (2008) that, among others, accommodates a flexible competition pattern across alternatives, scale differences in the revealed and stated choice context, and heterogeneity across individuals in the intrinsic preferences for alternatives. We formulate below the utility functions for the RP and SP choice alternatives.

$$(7)\quad U_{ntj} = \begin{cases} \beta_n^{sp}{}' x_j^{sp} - \gamma_n{}' p_{ntj}^{sp} + \mu_{nh}^{sp} + \varepsilon_{ntj}, & j = h_1^t, h_2^t,\ t = 1,\dots,T^{sp}; \\ \beta_n^{sp}{}' x_{jt}^{sp} - \gamma_n{}' p_{ntj}^{sp} + \alpha_{sq} + \mu_{nsq}^{sp} + \varepsilon_{ntj}, & j = SQ;\ t = 1,\dots,T^{sp}; \\ \beta_n^{rp}{}' x_j^{rp} - \gamma_n{}' p_{ntj}^{rp} + \varepsilon_{ntj}, & j = 1,\dots D_n,\ t = T^{sp} + 1 \end{cases}$$

where β_n^{sp} and β_n^{rp} are vectors of parameters of individual *n*, representing the marginal utility of forest attributes for the SP and RP alternatives, respectively, and x_j^{sp} and x_j^{rp} are vectors of forest attribute variables of alternative *j*. Note that the vector SP and the RP data may include common as well as dataset specific variables. γ_n is the parameter on the travel cost variable which depends on individual *n*. μ_h^{sp} and μ_{sq}^{sp} are error component terms which are included in utility function for the SP choices. They are constant over SP choice situations but vary over individuals. Define $\mu_{n\kappa}^{sp} = \phi_\kappa E_{n\kappa}, \kappa = h, sq \text{ and} E_\kappa \sim N(0,1)$, where h is representing the hypothetical forests in the choice experiment and *sq* the status quo forest, i.e. the forest actually visited. This implies that the mean of μ_κ^{sp} is restricted to zero. If ϕ_κ is significant different from zero it indicates that the variance of the unobserved utility in the choice experiment (CE) is different from the revealed preference data, i.e. different scale parameter (Brownstone et al. 2000). α_{sq} is a non-random alternative specific constant for the status quo forest, i.e. the forest visited most often during the last 12 months. This parameter captures that the utility of visiting the forest visited in the past may differ systematically from the utility for visiting one of the hypothetical forests and this difference cannot be explained by differences in attribute values. ε_{ntj} is identically, independently distributed over alternatives and individuals and is unobserved by the researcher. *CS* represents the number of choice alternative in the RP choice set. T^{sp} is the number of CE choice situations.

Welfare Analysis

Estimating the WTP of changes in the quality and the access to recreation sites is the main objective of most recreational valuation studies. In this paper, we will consider the implications for welfare measures of the different modelling approaches, including the chosen sampling strategy, travel mode, and combined stated and revealed preference data sources. The welfare changes as measured by the expected WTP from a move from site attribute level x^0 to x^1 and conditional on individual taste β_n take the following familiar form (Bockstael and McConnell 2007):

$$E[WTP_n] = \gamma_n^{-1}\left[\ln\left[\sum_{j\in CS} e^{\left(\boldsymbol{\beta}_n'\mathbf{x}_j^1 - \gamma_n p_{nj}^1\right)}\right] - \ln\left[\sum_{j\in CS} e^{\left(\boldsymbol{\beta}_n'\mathbf{x}_j^0 - \gamma_n p_{nj}^0\right)}\right]\right] \quad (8)$$

Where CS is the total set of forests in Lorraine. The expected measure needs integration over the taste distribution($\varphi(\bar{\beta},\bar{\gamma},\bar{\Omega})$) in the population:

$$\begin{aligned} E[\overline{WTP}] &= \int WTP_n \varphi(\bar{\beta},\bar{\gamma},\bar{\Omega}) d\beta \\ &= \int \gamma_n^{-1}\left[\ln\left[\sum_{j\in CS} e^{\left(\boldsymbol{\beta}_n'\mathbf{x}_j^1 - \gamma_n p_{nj}^1\right)}\right] - \ln\left[\sum_{j\in CS} e^{\left(\boldsymbol{\beta}_n'\mathbf{x}_j^0 - \gamma_n p_{nj}^0\right)}\right]\right]\varphi(\bar{\beta},\bar{\gamma},\bar{\Omega}) d\beta \end{aligned} \quad (9)$$

where $\bar{\beta}, \bar{\gamma},$ and $\bar{\Omega}$ are estimated attribute and cost parameters and the estimated variance of these parameters, respectively. The equation above (9) integrating the travel mode choice is defined:

$$\begin{aligned} E[\overline{WTP}] &= \int WTP_n \varphi(\bar{\beta},\bar{\gamma},\bar{\Omega}) d\beta \\ &= \int \gamma_n^{-1}\left[\begin{aligned} &\ln\left[\sum_{j\in CS} e^{\left(\boldsymbol{\beta}_n'\mathbf{x}_j^1 - \gamma_n p_{nj}^{car}\right)} + \sum_{j\in CS} e^{\left(\boldsymbol{\beta}_n'\mathbf{x}_j^1 - \gamma_n p_{nj}^{bike} + \alpha_{bike} + \mu_n^{bike}\right)} + \sum_{j\in CS} e^{\left(\boldsymbol{\beta}_n'\mathbf{x}_j^1 - \gamma_n - p_{nj}^{walk} + \alpha_{walk} + \mu_n^{walk}\right)}\right] \\ &-\ln\left[\sum_{j\in CS} e^{\left(\boldsymbol{\beta}_n'\mathbf{x}_j^0 - \gamma_n p_{nj}^{car}\right)} + \sum_{j\in CS} e^{\left(\boldsymbol{\beta}_n'\mathbf{x}_j^0 - \gamma_n p_{nj}^{bike} + \alpha_{bike} + \mu_n^{bike}\right)} + \sum_{j\in CS} e^{\left(\boldsymbol{\beta}_n'\mathbf{x}_j^0 - \gamma_n p_{nj}^{walk} + \alpha_{walk} + \mu_n^{walk}\right)}\right] \end{aligned}\right] \\ & \varphi(\bar{\beta},\bar{\gamma},\bar{\Omega}) d\beta \end{aligned} \quad (10)$$

Data and survey implementation

The administration of our questionnaire was Web-based, a survey mode that has gained popularity in CE surveys (Olsen 2009). An email was sent to an email list of inhabitants in Lorraine. A response rate of two percent was projected by the company (*EmailingFrance*) maintaining the applied list. This low response rate raises of course a serious issue of representativeness of the sample if the objective is to generalize results to the general population. However, in the present paper the main objective is to compare different methodological approach and not give a total value of the recreative service of the forests. In the main survey, 53,000 people were sent an e-mail that briefly described the survey and provided a link to the questionnaire on the Web. If the respondents completed the questionnaire, they would be able to participate in a lottery with the chance to win one of 50 USB memory keys. E-mail reminders were sent after two and four weeks. In total, 1837 respondents began to answer the online questionnaire (3.5%), and out of these, 1144 actually completed the questionnaire (2.2%). Compared to other surveys using the same panel, the response rate was relatively high (Bougherara et al. 2011), although compared to most other preference eliciting surveys, in general, the response rate is considered very low. Furthermore, only 816 respondents of the 1144 respondents had in fact residence in Lorraine and of these 526 had visited a forest and given information on which specific forest they had visited. Hence, our final sample used in the subsequent analysis consists of 526 respondents.

The questionnaire had four main sections. The first section concerned basic socio-demographic variables such as age, gender, and the municipality of their residence and how many times they had visited a forest during the last 12 months. The municipality (*"commune")* is the most detailed information on home address we could obtain from the respondents. In a pilot survey the respondents were asked to provide their postal address but the majority refused to do this. Fortunately, the French municipalities are relative small and give a rather precise spatial location. On average, a municipality covers an area of about 10 km^2 and with very few exceptions a municipality consists of one town or one village with its surrounding open space.

In the second section, forest visitors were asked about their visits in the forest (motives, length of visit, mode of transport, etc.) and they were asked to identify the forest they had visited most often during the last 12 months by clicking it on an integrated and interactive version of *Google Maps* showing a satellite image of the Lorraine area. As starting point the map was centered on the commune of their residence or the commune which served as starting point for visiting the forest (from secondary house, hotel, camping or family). To our knowledge, this is the first application of interactive maps in identifying visited forests in a web survey[5]. Based on a focus group test of the questionnaire in a computer lab, it was decided to use the *Google Map.* This map is familiar to most internet users and it allowed also choosing between maps and aerial views. Map indication of forest is appropriate in our case because not all forests have commonly known and unique names. After the identification of the visited forest the respondents were asked to indicate if they believed that they had correctly found the forest on the map that they had actually visited. 89% of the respondents stated that they had found the right forest. Among the 11% who said that they did not find the forest, five percent said that the map was not appropriate for finding the forest visited. We cannot know when a respondent has reported to have found the forest if it was the right forest they indicated on the map. However, only 9 respondents did click on other land uses than forests. Due to uncertainty in clicking with the mouse in relation to the physical resolution of the map, we

[5] The *Google Map* has previously been used in online questionnaires, e.g. by (Horni et al. 2010) in empirical formation of choice sets in analysis of shopping behaviour.

considered clicks which were within a distance of two kilometers from a forest as indicating this forest. We did ask the respondents to give the distance between their starting point and the forest visited. Then, we compared this given distance with the GIS-calculated distance between the centre of the municipality (administrative center) and the nearest entry point to the forest visited. Observations where the difference between these two distances could not be justified by the uncertainty about which entry point used or uncertainty about the exact location of the respondents were excluded. Based on this analysis we excluded 70 respondents.

The third section of the questionnaire included a CE with six choice tasks. Each choice task consisted of a status quo alternative defined as the forest the respondent had visited the most often over the past 12 months and two experimentally designed alternatives. Before they were given the choice tasks, respondents were asked to characterize the forest they had visited most often over the past 12 months according to the same attributes and levels used in the experimental design. Focus group interviews suggested that this way of asking respondents to describe the forest visited, in line with the pre-defined list of attributes and levels, was an effective way of informing them about the attributes and preparing them for the subsequent choice tasks. The choice of the five attributes (dominating tree species, presence of trekking paths, presence of parking and picnic places, presence of lakes and rivers, and distance) was guided by the focus group interviews (Table 1). A pilot test was carried out based on 79 respondents. On the basis of results from this pilot test, an experimental design with an informative Bayesian update to improve design efficiency was constructed using NGENE software (Scarpa et al. 2007). The design and analysis of the stated choices are reported in more detail in Abildtrup et al. (2013). The final section of the questionnaire had attitudinal questions and some additional questions on the socio-economic characteristics of the respondents and their households.

Table 1

To characterize the forests in the choice set of each individual we combined different GIS maps to establish a spatial database of forests (Thirion 2010). Variables describing tree species composition of the forest were obtained from the French National Forest Inventory (IFN). Data describing the presence of hiking paths were obtained from the French Hiking Association (Fédération Française de Randonnée Pédestre), while data concerning the presence of recreational facilities, lakes and rivers in forests were obtained from the French National Geographic Institute (IGN). Basically, forests are defined as continuous land with forest cover of more than 5 hectare. If a forest is very large (typically, greater than 1,000 hectares), it is divided into two forest units that are considered to be a unity in our analysis. The division of forests into units was, among other things, determined by existing structures in the forest, e.g., roads or rivers.

The distance between a respondent and a given forest is the road distance between the town hall of the municipality (*commune*) where the respondent had his/her residence (or the municipality where the respondent was temporarily residing when going to the most visited forest during the past 12 months) and the closest entry point to the forests. The road chosen when transport is by car is based on the road with the shortest distance in time. For people walking, the road chosen is based on shortest distance in kilometers. There is some uncertainty in the calculation of the distance since we do not know the exact place of residence of the respondent in the municipality and we do not know which entry point to the forest has been used by the respondent. The travel costs consist of variable driving costs using a car (fuel and service costs) and alternative costs of time. The driving cost used information about the car type in the questionnaire and car type dependent driving costs from the French automobile club (www.automobile-club.org). When walking or biking to the forest we use only alternative costs of time. We follow the standard approach in the literature where one third of the wage rate is used as alternative cost of time (Cesario 1976)[6]. As a proxy for the hourly wages rate is used the household revenue divided by number of adults in the household and the average number hours working per year.

The sample and descriptive statistics

In Table 2, the main demographic and socioeconomic characteristics of the effective sample used to estimate the site selection model are presented and compared with the total population in Lorraine. The share of female respondents is lower in the sample than in the population and the 40-60-year-old respondents are overrepresented in the sample. The sample exhibits an overrepresentation of people in high-income classes. The relatively high rates of middle-aged people and high-income groups in the sample are not unusual for Internet and mail surveys (Olsen 2009). Thus, even though the response rate might raise some concerns regarding the representativeness of the sample, the skewness of the sample for central socio-demographic characteristics does not seem to be much worse than similar surveys with much higher response rates.

[6] The alternative cost of time is generally a source of discussion in the literature. Whatever the assumption, it will be true that individuals will consider a limited time budget and that the transport time will influence the choice of forest to visit. Even though we find this discussion relevant it is beyond the scope of this paper

Table 2

The majority of the respondents (93%) has visited a forest at least one time during the last 12 month and 90% has visited a forest more than once during the past 12 months, whereas 77% have visited different forests during the period. Forest visitors have visited a forest 27 times during the past year on average. A study carried out at the national level in France in the year 2000 (Peyron et al. 2002) estimated the average forest visits per household in France to be only nine times per year, though this only included car-borne visits. This study also found the percentage of respondents that went to the forest to be 44%. This relatively low percentage at the national level may be due to less accessibility to forests in some other regions in France and to the presence of other non-forest substitute sites.

Table 3 defines the forest attributes. Besides the five attributes used in the CE, the RP data set includes more forest characteristics. Table 3 includes only variables that are kept in the final model presented in the next section.

Table 3

Model estimation and empirical results

In this section we first describe the results based on RP data and conditions on the travel mode choice. That is, we calculate the individual travel costs conditioned on the travel mode used by the respondents. For comparisons we show the results of the simple random sampling of the choice set as well as the first and second iterations of the strategic sampling scheme, using CL and MXL models. Then, we estimate a model based only on RP data where the travel mode and the site selection choices are integrated (equation 6). Next, we compare the RP data and SP data. We estimate a model based only on SP data and a model where we use both datasets and assume equal marginal utilities on common forest attributes variables in the two datasets. Finally, we apply the estimated models to estimate the WTP of two scenarios where the recreational quality of the forest are changed and we compare estimates based on the different approaches for choice set definition, estimation methods, and datasets.

Choice set and estimation

The mixed logit model is estimated assuming that the parameters associated with all forest attributes, except the cost attribute, are normally distributed random parameters[7]. This allows for both negative and positive preferences – something that focus group interviews indicated as being relevant. Quite often in imperical studies, the variable representing the marginal utility of income (travel cost) is kept fixed in order to avoid a number of severe problems associated with specifying a random price parameter (e.g. Train 1998). However, we believe that it may be important in the current case to let the distance be specified as a random variable because the travel costs including alternative cost of time may be associated with significant unobserved elements. Therefore, we have estimated the model, assuming a bounded distribution of the distance parameter, i.e. triangle distribution where the travel cost parameter ($-\gamma$) is restricted to be negative. In this way, we allow for differences in marginal disutility of travel costs but avoid positive preferences for costs.

In the first iteration of the sampling scheme, using simple random sampling, we use a choice set of 30 forests. In the following iteration where forests are sampled using the estimated probability of visiting a forest based on the estimates from the previous iteration we sample 20 forests[8]. Since we use sampling with replacement, a forest may be selected for inclusion in a choice set more than once. Thus, the final choice set may have less than 20 forests. Also, it may have 21 forests if the visited forest is not selected by the sampling procedure and no forests are selected more than once. The sample correction term $\ln(\pi(D_n \mid i)$ (from equation (3) and (5)) is included as a variable in the utility function where the coefficient is restricted to one.

In the analysis of the revealed preference data the selection of the variables to include in the utility function represents an important modeling choice. First of all, for reasons of comparability and consistency across the RP and SP parts of the survey we wanted to include variables describing the attributes which were also used in the CE. Recall that these variables were selected based on *a priori* expectations based on a literature review and experiences gained in focus group interviews. However, to reduce the size of the experimental design we were restricted to relatively few

[7] The log-likelihood function is estimated using Halton draws (150 draws). Estimation of the model was carried out using Nlogit 4.0 software while the strategic sampling scheme was carried out in Matlab.

[8] We use only 20 forests in the strategic sampling scheme because in the modeling of the travel mode choice (equation 6) we have 3*D_n alternatives in the choice set. In this model we only use the choice set generated in the second iteration of strategic sampling scheme. With D_n larger than 20 would make the estimation relatively time consuming in this model. The 30 forest, selected with simple random sampling, was sampled between the forests which could be reached in 30 minutes by car.

attributes in the CE. Other variables, available from the forest attribute data base were included based on our initial beliefs and statistical evidence. Unfortunately, it is not possible to include the dominant species attribute in the RP analysis due to lack of variation within in the choice sets. The coniferous forests are mainly found in the southern part of the region while broadleaved species are found in the center and the north of Lorraine.

First, we investigate the impact of strategic sampling strategy on results given the RP data. We estimate the utility function based on observed choice of forests visited given different assumptions about the sampling strategy and specification of the utility function (conditional and mixed logit). In this analysis the travel costs are conditioned on the choice of travel mode, i.e. for a visitor choosing the car the cost of visiting a forest is based on the estimated travel costs associated with going by car while the alternative cost of time is used for people walking or biking. We assume a biking speed of 10 km/hour and walking speed of 6 km/hour. We use the individual calculated driving cost and alternative costs as described in the previous section.

Table 4 shows the estimates based on a CL model and compares simple random sampling in the choice set with strategic sampling[9]. An overall comparison of the two sampling schemes shows that the parameters have the same sign but the size of the parameters may differ. The respondents have positive preferences for recreational facilities in forests, i.e. trekking paths, parking places, and picnic places. They also prefer a forest with either a lake or a river, however, only significant on a 10 % level and only with simple random sampling. Large forests are preferred to small forests. We use the logarithm of the forest size as this transformation resulted in better model fits than when using the size directly or using other transformations. Forests with a high share of old high-forest is preferred to young forests or forests with coppice management. Forests with a high share of public ownership are preferred to forests with a high share of privately owned land. In the present sample of forests we do not know if the private forest owner has closed the forest for the public which may partly explain this revealed preference for public owned forests. However, the majority of the private forests in Lorraine are open for public access. Forests where it is likely to find blueberries are preferred. However, this variable is not significant when using the strategic sampling scheme. Forest with zones designated as biological reserves have a negative impact on utility when using simple random sampling, but not with strategic sample scheme. We find that the number of forest roads has a negative impact on the choice of a forest. This may indicate that visitors prefer non-managed forests. With both sampling schemes we find that the travel cost variable is negative and highly significant. Considering the preference heterogeneity as revealed by table 5, we only find that the

[9] As in Lemp & Kockelman (2012) the results are an average of 10 replications of the simple and of strategic sampling strategy.

distribution of the distance variable has a significant standard deviation. However, based on the simple random sampling results there seems also to be heterogeneity in the preferences for forest ownership.

For comparison of the coefficients of the utility functions we have normalized them with the cost coefficient. The normalized coefficients represent marginal WTP for the considered attributes. The estimation of marginal WTP (MWTP) in the case of a MXL is based on random simulation where parameters are drawn randomly from the estimated distribution moments (Sillano and Ortúzar 2005). In comparison of the marginal WTPs based on the simple random sampling and based on the strategic sampling scheme we find no systematic differences in the level of WTP. For example, the MWTP for visiting a forest with one trekking path (PATHONE=1) is €0.443 with simple random sampling and €0.394 with strategic sampling while for more than one trekking path (PATHMORE) the MWTP is €0.357 with simple random sampling and €0.406 with strategic sampling (Table 5). On the other hand, comparing the MWTP for the attributes, applying conditional logit (Table 4) and mixed logit (Table5) it seems that the MWTP, in absolute values, is higher for the results using MXL than using CL. This is however only the case when we consider strategic sampling while there is no systematic tendency when using simple random sampling. We only carried out two iterations with the strategic scheme as the results in the first and second iteration was found to be similar (Comparing column five and eight in Table 4 and 5). This corresponds to the results in Lemp and Kockelman (2012).

Table 4 and Table 5

Travel mode modeling

In the model with explicit accounting for the travel mode choice we use average estimates of the costs per km for direct use of cars but use the individual estimates of alternative cost of time. For the visitors who do not go by car we have no information on their direct driving cost if they would decide going by car as we do not know if they have a car, and if they have one what type of car it is. Therefore, we use an average estimate of the direct costs. We estimate the model (6) using the choice set based on the second iteration of the strategic sample scheme. Note that the cost variable in the case of walking or biking to the forest is based on the alternative cost of time only. Table 6 gives the results. The coefficients resemble the estimates conditioned on the travel choice (Table 5) with respect to sign and significance. The estimated MWTPs are also at the same level without systematic differences. The included alternative specific constants (ASCB and ASCW) were only significant for walking mode. This indicates that people walking to the forests have a positive utility of walking compared to driving by car which is not explained by the forest attributes. Note that the present result depends on the definition of the alternative cost of time and the assumed velocity of people walking or biking (here assumed 6 km/h for visitors walking and 10 km/h for bikers).

Table 6

Comparison of RP and SP datasets

Finally we want to compare our RP data with the SP data from the CE. Table 7 presents the utility coefficients for the choice model using only the data from the CE part of the survey (the SP data). Here we use average cost per km as cost variable since we do not have information on how they will travel to a forest if they choose a hypothetical forest alternative in the CE. Applying an average measure of cost corresponds basically to the use of distance as a cost variable[10].

All the five attributes considered in the choice experiment are important for the choice of forests to visit. We find that the visitors prefer broadleaved forest or mixed species forests to forest dominated by coniferous species. All the other attributes are significant and have the expected sign and consistent with the revealed preference results. However, we see that all the parameters are generally more significant in the choice experiment. An exception is the facility attribute where the parameter values are less significant. With the CE we also find that for all parameters there is significant preference heterogeneity between individuals. This was not the case for the RP data. Comparing the MWTP estimates based on the CE with the MWTP in Table 5 we find higher MWTP for the CE data. This may be due to a hypothetical bias as is often found when comparing SP and RP results (Murphy et al. 2005). Even though we included a budget reminder before the choice sets the respondents may have overestimated their willingness to travel to visit attractive forests.

We have combined the RP and SP data applying the econometric model in equation (7). We have included dataset specific error components, implemented as suggested in (Hensher 2008) by including alternative specific constants and restricting the mean to zero. However, in the case of the status quo alternative we allowed the alternative specific constant to differ from zero to account for potential status quo effects. Table 8 presents the results where the coefficients are restricted to be

[10] Bartczak et al. (2012) suggest using the distance as cost variable due to the difficulties associated by estimating the costs.

equal for common variables[11]. In both cases we find that the error components in the choice experiments are significant and indicating that the variance (scale factor) differ between the two datasets. However, a likelihood ratio test (χ^2=367 with DF=10) of the restriction of equal coefficients on common variables indicated that we cannot pool the two dataset even though we allow for different scale factors (Swait and Louviere 1993). Therefore, the common parameters are relative close to the parameters based on CE alone.

Table 7 and 8

Welfare analysis of scenarios

For comparison of the different definitions of the choice sets and modeling strategies we have estimated the per visit WTP for individuals living in Sarrebourg, a town with about 13,000 inhabitants in the county (*département*) Moselle, of changes in the quality of local forests. In the first scenario analyzed, we introduce one trekking path in each of the six forests located closest to Sarrebourg. Presently, there are no trekking paths in these forests. The second scenario analyzed, was defined as the conversion from public to private ownership of the same six forests. Table 9 gives the results conditioned on the travel mode, using equation (8) and (9) for the conditional logit and the mixed logit model, respectively. For the latter we use 700 draws from the parameter distributions to calculate the average impact. Secondly, the WTP for an improvement of the forests (making one trekking path in each of the six forests) is higher for people walking (€0.28) than for people going by car (€0.11)[12]. This is because the substitute forests with a trekking path which are further away, are more expensive to visit for people walking than people who have decided to go by car. We see the same pattern for the scenario where the quality of the forests is reduced by selling them to private owners. The loss is highest for the people who have decided to walk to the forests.

However, we would expect some people to change travel mode as a result of changes in the quality of the forests. In Table 10, we have estimated the impact on welfare measures using equation (10) where the travel modes are endogenously selected. Generally, the WTP for the evaluated changes are smallest in absolute values when the results are based on the model integrating the travel cost mode choice. The average WTP based on travel mode specific estimates is calculated as a weighted average of the travel mode specific WTP estimated in Table 9 using the share of respondents using

[11] We use here the choice set for the RP data based on second iteration of the strategic sampling strategy.

[12] Estimates in parentheses are based on strategic sampling and MXL estimation

the three travel modes as weights (last line in Table 9). We also find that the welfare impact based on the SP data alone iss at the same level as with the RP data for the scenario with establishment of trekking roads in local forests. As ownership was not included as an attribute in the CE we cannot use the SP data to evaluate the change of ownership scenario.

Table 9 and Table 10

Concluding remarks

In this paper we have analyzed the determinants of the recreational value of forests in Lorraine. In particular, we have considered the issue of identifying the visited forest in a questionnaire, handling the issue of large choice sets, integrated the choice of travel mode, and combined RP and SP data.

Lorraine is a relatively densely forested region. Therefore, the residents are always relatively close to a forest and use relatively frequently forests for recreational purposes. For some of the respondents, the forest is where they go daily, for example, to walk the dog. This abundance of forests in Lorraine raises some important issues for recreational modeling. First we show how interactive maps can serve as an appropriate approach in web-based surveys to identify non-iconic forests or recreation sites. The potentially large choice set facing forest visitors in Lorraine challenges also the site selection modeling. We introduce the strategic sampling scheme suggested by Lemp and Kockelman (2012) to the modeling of recreational site selection with large choice set. Their results based on a Monte Carlo simulation showed that this sample scheme makes mixed logit estimation less sensitive to random sampling in the choice set. We compared the results based on simple random sampling and strategic sampling, applying conditional logit and mixed logit estimation. By and large, the results are not sensitive to sampling strategy and the estimation method. This applies both to the estimated marginal WTP and the welfare economic impact of the analyzed scenarios. However, we are not able to compare the estimated values with the “true” values as our data are not generated by simulation as in Lemp and Kockelman (2012) and we are not able to estimate the model with all choice alternatives included for comparison as it was done in for example Nerella and Bhat (2004). The total choice set is simply computationally too large for estimation. One advantage of the strategic sampling scheme was that the likelihood function did converge faster to a maximum when forests included in the choice set were based on the strategic sample scheme[13].

We have shown that the error-component mixed model can be used to integrate the travel mode choice. This is an alternative to the traditional nested model and allows for unobserved individual

[13] With simple random sample the log likelihood function is often flat at optimum.

preference heterogeneity in the population. While we did not find systematical differences in the marginal WTP when taking travel choice into account we find that the welfare impacts of establishing trekking paths or changing ownership was only one half in absolute values when integrating the travel mode choice in the site selection model. The analysis reveals also that the welfare impact depends on the travel mode. Visitors walking to a forest will benefit relative more on local improvement of the recreational quality of the forest compared to visitors going by car. This is because it is less costly for visitors going by car to visit alternative high-quality forest further away. This raises some distributional issues in recreational management of forests. Furthermore, the easy access to attractive forests in the proximity of visitors' residence will increase the likelihood of not using the car to go to the forest. This may be associated with positive externalities like less pollution and higher level of physical activities and consequently improved health of the population. It is important to note that, in the present study, the alternative cost of time was based on household income and are therefore a coarse measure. Future analyses, considering non-car-borne transport modes, should address the visitors alternative cost of time (e.g. Englin and Shonkwiler 1995; Phaneuf 2011).

We find that the results based on stated and revealed preferences are relative similar with respect to sign and significance. However, we reject that the data can be pooled[14]. This may be due to differences in the objective measurement of forest attributes (RP data) and the subjective perception of attributes in the choice experiment. It may also be due to endogeneity of the recreational facility attributes. These facilities may be placed in forest where people go for some other, unobserved, reasons. This may explain why parking and picnic places are highly significant in the RP data but only weakly significant in the SP data. We suggest that the results based on the choice experiment are more reliable when endogeneity of the attribute variable may be a problem. Furthermore, we were not able to account for dominant tree species in the revealed preference dataset due to multicollinarity in the RP dataset. In the choice experiment this is not a problem as it is based on an optimal statistical design. However, we also find that the marginal WTP of the forest attributes where significant higher based on SP data. This may be due to a hypothetical bias in the CE. The respondents did not consider distance in the CE in the same way as when making the real choice. In contrast, the welfare impact of trekking path scenario based on SP data is relative similar to the impact based on the RP data.

The results show rather unambiguously that the average resident in Lorraine prefers visiting forests with trekking paths, with the presence of picnic and parking places, public-owned to private-owned, large forest to small forests and forests with few forest roads. The results also show, but less

[14] Pooling of revealed and stated preference data is also rejected in Huang et al. (1997).

robustly, that respondents have positive preferences for forests with presence of rivers or lakes, blueberries, and forest which are not appointed as biological reserves. It may be surprising that biological reserves have a negative impact on the choice of a visit as such a designation does not imply restrictions on recreational use. However, the results may be due to endogeneity of this variable. Biological reserves are commonly designated where the conservation authorities know that the forests are not often used for recreational purposes. In all models, the share of forests with high trees where always positive but not significant using conventional levels. In the RP based models we only find significant preference heterogeneity with respect to cost, forest area and the number of forest roads. This result is in contrast to the results from the SP data where we find that all parameters have significant variation over individuals. We believe that this is due to the higher amount of information from each individual, i.e. six choices in the SP data set versus one choice in the RP data per individual. This increases the efficiency of the tests based on the SP data.

In the present study we did focus on different methods for site selection modeling and not on the demand for visits. Therefore, the estimated welfare impacts of the considered changes in the forest quality are for one visit. If the total welfare effects of changing the quality of the forest were to be evaluated, one would have to consider the number of visits and the impact on the number of visits of changing the quality of the forests.

Most studies estimating the economic value of forest recreation only consider car-borne visits and are not considering the daily use of forest. The present study has highlighted the importance of considering the travel mode choice when forests are used frequently for recreational purposes and it is shown how this choice can be included in the site selection model. With the increased focus on importance of physical activities for human health, the determinants of such activities will definitely be an important issue in future research. This should also include the determinants of the travel mode choice for recreational trips. We have also shown that combining RP data with SP data provides some advantages. The present study confirms that RP based analysis may be sensitive to endogenous site attributes and that multicollinearity of attributes in RP studies may imply econometric problems. Therefore, we believe that including stated preference questions in surveys on recreational use of forest should be recommended for the evaluation of public policies influencing the recreational service of forests.

References

Abildtrup, J., Garcia, S., Olsen, S. B., Stenger, A. (2013). Spatial preference heterogeneity in forest recreation. *Ecological Economics* (*Forthcoming*).

Bartczak, A., Englin, J. E., Pang, A. (2012). When are Forest Visits Valued the Most? An Analysis of the Seasonal Demand for Forest Recreation in Poland. *Environmental and Resource Economics*, 52(2), 249–264.

Bell, K. P., Strand, I. E. (2003). Reconciling Models of Recreational Route and Site Choices. *Land Economics*, 79(3), 440–454.

Bhat, C. R., Castelar, S. (2002). A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area. *Transportation Research Part B: Methodological*, 36(7), 593–616

Bockstael, N. E., Hanemann, W. M., Kling, C. L. (1987). Estimating the value of water quality improvements in a recreational demand framework. *Water Resources Research*, *23*(5): 951-960

Bockstael, N. E., McConnell, K. E. (2006). Environmental and resource valuation with revealed preferences. A theoretical guide to empirical models. The Economics of non-market good and services Vol 7. Springer, Dordrecht

Bougherara, D., Costa, S., Teisl, M. (2011). Making or Buying Environmental Public Goods: Do Consumers Care? 5emes Journees de Recherches en Sciences Sociales, INRA SFER CIRAD, 8, 9 and 10 Decembre 2011, Dijon.

Brownstone, D., Bunch, D. S., Train, K. E. (2000). Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B: Methodological*, 34(5), 315–338.

Bujosa Bestard, A., Font, A. R. (2009). Environmental diversity in recreational choice modelling. *Ecological Economics*, 68(11), 2743–2750.

Cesario, F. J. (1976). Value of Time in Recreation Benefit Studies. *Land Economics*, 52(1), 32–41.

Domanski, A., Haefen, R. H. Von. (2010). Recreation Demand Models with Large Choice Sets. North Carolina State University, working paper.

Doyle, S., Dodge, M., Smith, A. (1998). The potential of Web-based mapping and virtual reality technologies for modelling urban environments. *Computers, Environment and Urban Systems*, 22(2), 137–155.

Englin, J. E., & Shonkwiler, J. S. (1995). Modeling recreation demand in the presence of unobservable travel costs: toward a travel price model. *Journal of Environmental Economics and Management*, *29*, 368–377.

Fleming, C. M., Bowden, M. (2009). Web-based surveys as an alternative to traditional mail methods. *Journal of environmental management*, 90(1), 284–92.

Frejinger, E., Bierlaire, M., Ben-akiva, M. (2009). Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological*, 43(10), 984–994.

Garcia, S., Jacob, J. (2009). La valeur récréative de la forêt en France: Une approche par les coûts de déplacement. *Revue d’Etudes en Agriculture et Environnement*, 91(1), 43–71.

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*(August), 271–281.

Hanley, N. D., Shaw, W. D., Wright, R. E. (2003). The new economics of outdoor recreation. Cheltenham, Edward Elgar.

Hensher, D. A. (2008). Empirical approaches to combining revealed and stated preference data: Some recent developments with reference to urban mode choice. *Research in Transportation Economics*, 23(1), 23–29.

Hensher, D. A., Rose, J. M., Greene, W. H. (2008). Combining RP and SP data: biases in using the nested logit "trick" – contrasts with flexible mixed logit incorporating panel and scale effects. *Journal of Transport Geography*, 16(2), 126–133.

Hicks, R. L., Strand, I. E. (2000). The extent of information: its relevance for random utility models. *Land Economics*, *76*(3), 374–385.

Horni, A., Charypar, D., Axhausen, K. W. (2010). Empirically Approaching Destination Choice Set Formation. Working paper Transport and Spatial Planning, Institute for Transport Planning and Systems, Swiss Federal Institute of Technology, Zurich.

Huang, J., Haab, T. C., Whitehead, J. C. (1997). Willingness to pay for quality improvements: Should revealed and stated preference data be combined? *Journal of Environmental Economics and Management*, *34*(3), 240–255.

Kuriyama, K., Michael Hanemann, W., Hilger, J. R. (2010). A latent segmentation approach to a Kuhn–Tucker model: An application to recreation demand. *Journal of Environmental Economics and Management*, 60(3), 209–220.

Lemp, J. D., Kockelman, K. M. (2012). Strategic sampling for large choice sets in estimation and application. *Transportation Research Part A: Policy and Practice*, 46(3), 602–613.

Lupi, F., Feather, P. M. (1998). Using partial site aggregation to reduce bias in random utility travel cost models. *Water Resources Research*, 34(12), 3595–3603.

Marta-Pedroso, C., Freitas, H., Domingos, T. (2007). Testing for the survey mode effect on contingent valuation data quality: A case study of web based versus in-person interviews. *Ecological Economics*, 62(3-4), 388–398.

McFadden, D.L. (1974). Conditional logit analysis of qualitative choice behavior. In: Zarembka, P (Ed.). Frontiers in Econometrics, New York, Academic Press: 105-142.

McFadden, D. L. (1977). Modeling the choice of residential location. Cowles Foundation Discussion Paper 477, Yale University, New Haven.

Montagne, C., Harou, P., Garcia, S., Jacob, J., Stenger, A., Navrud, S., Bartczak, A. (2008). Référentiel pour l'Evaluation Economique des Biens, Services et Dommages Environnementaux Liés à la Forêt Française. Rapport, Laboratoire d'Economie Forestière.

Murphy, J. J., Allen, P. G., Stevens, T. H., Weatherhead, D. (2005). A Meta-analysis of Hypothetical Bias in Stated Preference Valuation. *Environmental & Resource Economics*, *30*(3), 313–325.

Nerella, S., Bhat, C. R. (2004). A Numerical Analysis of the Effect of Sampling of Alternatives in Discrete Choice Models. *Transportation Research B*, 1894, 11–19.

Normandin, D. (1998). Une évaluation de la demande sociale de services environnementaux de la forêt. INRA Sciences Sociales, Recherches en Économie et Sociologies Rurales N° 2 -juin 1998, 1–4.

Olsen, S. B. (2009). Choosing Between Internet and Mail Survey Modes for Choice Experiment Surveys Considering Non-Market Goods. *Environmental and Resource Economics*, 44(4), 591–610.

Parsons, George R, & Kealy, M. J. (1992). Randomly Drawn Opportunity Sets Model in a Random Utility of Lake Recreation. Land Economics, 68(1), 93–106.

Parsons, George R, Plantinga, A. J., Boyle, K. J. (2000). Narrow choice sets in a random utility model of recreation demand. *Land Economics*, 76(1), 86–99.

Peyron, J., Bakouma, J., Berthier, A., Colnard, O., Normandin, D., Stenger, A., Tessier, A. (2001). Première évaluation économique globale des dégâts forestiers dus aux tempêtes de décembre 1999. Rapport, Laboratoire d'Economie Forestière, Nancy.

Peyron, J., Harou, P., Niedzwiedz, A., Stenger, A. (2002). National survey on demand for recreation in French forests. Rapport, Laboratoire d'Economie Forestière Nancy.

Phaneuf, D. J. (2011). Can consumption of convenience products reveal the opportunity cost of time? *Economics Letters*, *113*(1), 92–95.

Phaneuf, D. J., Kling, C. L., Herriges, J. A. (2000). Estimation and Welfare Calculations in a Generalized Corner Solution Model with an Application to Recreation Demand. *The Review of Economics and Statistics*, 82(1), 83 – 92.

Scarpa, R., Campbell, D., Hutchinson, W. G. (2007). Benefit Estimates for Landscape Improvements : Sequential Bayesian Design and Respondents ' Rationality in a Choice Experiment. *Land Economics*, 83(4), 619–634.

Scarpa, R., Hutchinson, W. G., Chilton, S. M., Buongiorno, J. (2000). Importance of forest attributes in the willingness to pay for recreation: a contingent valuation study of Irish forests. *Forest Policy and Economics*, 1, 315–329.

Shonkwiler, J., Englin, J. E. (2005). Welfare losses due to livestock grazing on public lands: A count data systemwide treatment. *American Journal of Agricultural Economics*, 87(May), 302–313.

Sillano, M., Ortúzar, J. D. D. (2005). Willingness-to-pay estimation with mixed logit models: some new evidence. *Environment and Planning A*, 37(3), 525–550.

Swait, J., Louviere, J. (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research*, 30(3), 305–314.

Termansen, M., Skov-Petersen, H., McClean, C. J. (2004). Recreational site choice modelling using high-resolution spatial data. *Environment and Planning A*, 36, 1085–1099.

Termansen, M., Zandersen, M., & Mcclean, C. (2008). Spatial substitution patterns in forest recreation. *Regional Science and Urban Economics*, 38(1), 81–97.

Thirion, E. (2010).Utilisation d'un système d'informations géographiques pour caractériser l'usage récréatif des forêts de Lorraine. Rapport LERFoB, AgroParisTech, Nancy.

Train, K. E. (1998). Recreation Demand Models with Taste Differences over People. *Land Economics*, 74(2), 230–239.

Whitehead, J. C., Pattanayak, S. K., Van Houtven, G. L., Gelso, B. R. (2008). Combining Revealed and Stated Preference Data To Estimate the Nonmarket Value of Ecological Services: an Assessment of the State of the Science. *Journal of Economic Surveys*, 22(5), 872–908.

Table 1. Attributes and attribute levels

Attributes	Levels
Dominant tree species	Conifers
	Broadleaves
	Mixed tree species
Hiking paths	No marked hiking paths
	One marked hiking path
	More than one hiking path
Facilities	No facilities
	Parking or picnic places
	Parking and picnic places
Access to water	No water body
	River or lake in the forest
Distance from your home	0.5, 2, 5, 10, 20, 50 km

Table 2. Sample (completed questionnaires) and population characteristics

	Sample	Lorraine
Gender distribution (% women)	39	52
Age distribution (%)		
20 - 39 years	24	34
40 - 59 years	53	37
60 - 74 years	21	18
75- years	1	11
Household income		
€0 – 9,400	5	25
€9,401 – 13,150	6	14
€13,151 – 15,000	5	8
€15,001 – 18,750	4	13
€18,751 – 23,750	10	11
€23,751 – 28,750	13	8
€28,751 - 38,750	24	10
€38,751 – 48,750	15	5
> €48,750	19	6

Source: Age and gender: INSEE – *Population estimations;* Income: Taxable income 2008. www2.impots.gouv.fr/documentation/statistiques/ircom2007/region/region.htm

Table 3 Descriptive statistics and variable definition

Variable	Variable definition	Mean	Std. Dev.	Min	Max
	Forest attributes also in CE				
BROADLEAV	Is 1 if Broadleaves (>70%); otherwise 0	0.707	0.455	0	1
MIXEDSPEC	Is 1 if Mixed tree species (coniferous<70% and broadleaves<70%) ; otherwise 0	0.248	0.432	0	1
PATHONE	Is 1 if one marked hiking path; otherwise 0	0.084	0.277	0	1
PATHMORE	Is 1 if more than one marked hiking path; otherwise 0	0.027	0.162	0	1
FACIL_P	Is 1 if presence of parking or picnic places; otherwise 0	0.040	0.196	0	1
FACIL_PP	Is 1 if presence of parking and picnic places otherwise 0	0.011	0.105	0	1
WATER	Is 1 if presence of lake or river	0.528	0.499	0	1
DIST	Distance to forest (km)	109	56	0	310
	Forest attributes only in RP data				
AREA	Log(Forest recreation unit (m^2))	13.0	1.6	10.4	17.8
PUBLIC	Percentage of forest public owned*0.001	0.040	0.040	0	0.1
HIGHFOR	Percentage of forest with high forest*0.001	0.056	0.042	0	0.1
VAMY	Probability of finding blueberries	0.075	0.129	0.0	0.62
NATURRES	Is 1 if presence of a biological reserve, otherwise 0	0.010	0.102	0	1
FORROADS	Number of forest roads*0.001	0.014	0.033	0	0.488
TC	Individual travel costs	61.09	80.42	0.01	1038.64

The number of observations is *CS***N* (5268 forests *526 respondents= 2770968).

Table 4 RP results with simple random sampling and strategic sample sampling scheme: Conditional logit results.

Variable	Simple random sampling			Strategic sampling scheme					
				First iteration			Second iteration		
	Coefficient	P[|Z|>z]	WTP (€)	Coefficient	P[|Z|>z]	WTP (€)	Coefficient	P[|Z|>z]	WTP (€)
PATHONE	0.42	0.005	0.674	0.33	0.013	0.325	0.351	0.007	0.348
PATHMORE	0.42	0.035	0.669	0.40	0.021	0.400	0.379	0.029	0.375
FACIL_P	0.43	0.016	0.694	0.39	0.010	0.391	0.386	0.011	0.382
FACIL_PP	1.03	<0.001	1.659	0.93	<0.001	0.928	0.985	<0.001	0.976
WATER	0.15	0.314	0.234	0.16	0.195	0.161	0.179	0.151	0.177
AREA	0.85	<0.001	1.366	0.88	<0.001	0.873	0.883	<0.001	0.875
PUBLIC	6.15	0.015	9.865	7.80	0.001	7.754	7.862	0.001	7.793
HIGHFOR	2.37	0.312	3.798	3.09	0.148	3.067	3.095	0.141	3.067
VAMY	1.54	0.110	2.474	1.12	0.219	1.117	1.041	0.260	1.031
NATURRES	-1.21	0.014	-1.939	-0.57	0.181	-0.572	-0.654	0.123	-0.649
FORROADS	-3.11	0.031	-4.999	-4.78	<0.001	-4.759	-4.578	<0.001	-4.538
TC $(-\gamma)$	-0.62	<0.001		-1.01	<0.001	-	-1.009	<0.001	-
CORR				1.00	-	-	1.00	-	-
McFadden Pseudo R-squared		0.60			0.40			0.39	
N=526	Number of choices=		526						

Table 5 RP results with simple random sampling and strategic sample sampling scheme: Mixed logit results.

Variable	Simple random sampling			Strategic sampling scheme					
				First iteration			Second iteration		
	Coefficient	P[|Z|>z]	WTP (€)	Coefficient	P[|Z|>z]	WTP (€)	Coefficient	P[|Z|>z]	WTP (€)
PATHONE	0.427	0.043	0.443	0.345	0.021	0.360	0.372	0.010	0.394
PATHMORE	0.344	0.221	0.357	0.405	0.039	0.423	0.384	0.045	0.406
FACIL_P	0.477	0.051	0.495	0.470	0.007	0.491	0.462	0.005	0.489
FACIL_PP	1.158	0.001	1.201	0.897	0.001	0.937	0.958	<0.001	1.014
WATER	0.253	0.197	0.262	0.213	0.144	0.222	0.224	0.101	0.237
AREA	1.082	<0.001	1.114	0.949	<0.001	1.007	0.950	0.001	1.018
PUBLIC	10.857	0.008	11.173	9.655	<0.001	10.022	9.562	<0.001	10.158
HIGHFOR	3.727	0.236	3.865	3.016	0.216	3.150	3.180	0.174	3.366
VAMY	2.607	0.050	2.704	1.074	0.312	1.122	1.065	0.338	1.127
NATURRES	-1.598	0.063	-1.657	-0.626	0.238	-0.654	-0.737	0.176	-0.780
FORROADS	-5.411	0.051	-5.699	-6.852	0.001	-6.800	-6.250	<0.001	-6.484
TC ($-\gamma$)	-1.302	<0.001	-	-1.301	<0.001	-	-1.294	<0.001	-
Derived standard deviation of parameter distribution							-6.250		
AREA	16.4	0.018		6.335	0.121		6.099	0.144	
PUBLIC	7.19	0.129		0.250	0.451		0.209	0.455	
FORROADS	0.328	0.094		1.566	0.908		0.622	0.958	
TC	1.31	<0.001		1.301	<0.001		1.294	<0.001	
McFadden Pseudo R-squared			0.65		0.41			0.40	
N=526	Number of choices=		526						

Table 6 RP results modelling explicitly the travel mode choice

Variable	Parameter			Derived standard deviation of parameter distribution	
	Coefficient	P[|Z|>z]	MWTP	Coefficient	P[|Z|>z]
PATHONE	0.344	0.019	0.388	-	-
PATHMORE	0.261	0.187	0.294	-	-
FACIL_P	0.512	0.002	0.576	-	-
FACIL_PP	1.000	<0.001	1.126	-	-
WATER	0.200	0.146	0.226	-	
AREA	0.878	<0.001	0.989	0.040	0.889
PUBLIC	9.601	<0.001	10.816	-	-
HIGHFOR	3.206	0.169	3.611	-	-
VAMY	2.091	0.080	2.355	-	-
NATURRES	-1.385	0.028	-1.560	-	-
FORROADS	-6.568	0.001	-7.456	10.812	0.001
TC ($-\gamma$)	-1.168	<0.001		1.168	<0.001
ASCW	1.505	<0.001		2.939	0.001
ASCB	-35.494	0.457		27.856	0.444
McFadden Pseudo R-squared		0.36			
N=526	Number of choices=		526		

Using the choice set based on the second iteration of the strategic sampling scheme

Table 7 Results based on stated preference data alone (choice experiment)

Variable	Parameter			Derived standard deviation of parameter distribution	
	Coefficient	P[|Z|>z]	WTP (€)*)	Coefficient	P[|Z|>z]
BROADLEAV	0.870	<0.001	10.55	0.339	0.165
MIXEDSPEC	1.031	<0.001	12.87	0.723	<0.001
PATHONE	0.365	<0.001	4.03	0.684	<0.001
PATHMORE	0.713	<0.001	9.37	0.549	0.001
FACIL_P	0.182	0.048	2.03	0.453	0.038
FACIL_PP	0.131	0.150	1.58	0.334	0.328
WATER	0.635	<0.001	7.85	0.799	<0.001
TC*) $(-\gamma)$	-0.107	<0.001		0.107	<0.001
ASC	0.393	0.001	4.74	2.167	<0.001
McFadden Pseudo R-squared		0.25			
N=526	Number of choices=	6*526=3156			

*) In the CE the travel costs are based on average travel costs of the individuals in the sample (€0.2434/km)

Table 8 Combining SP (based on CE) and RP data sets (Data enrichments).

Random parameters	SP data		RP data		SP and RP data	
	Coefficient	P[|Z|>z]	Coefficient	P[|Z|>z]	Coefficient	P[|Z|>z]
BROADLEAV	0.788	<0.001				
MIXEDSPEC	0.901	<0.001				
PATHONE					0.558	<0.001
PATHMORE					0.687	<0.001
FACIL_P					0.203	0.071
FACIL_PP					-0.010	0.504
WATER					0.759	<0.001
TC ($-\gamma$)					-0.286	<0.001
AREA			-0.919	<0.001		
PUBLIC			7.481	0.014		
HIGHFOR			0.001	0.558		
VAMY			0.173	0.716		
NATURRES			-0.999	0.148		
FORROADS			-6.097	0.019		
ASC1	0.630	<0.001				
ASC2	0.00	-				
Derived standard deviations of parameter distributions						
BROADLEAV	0.645	0.063				
MIXEDSPEC	1.002	<0.001				
PATHONE					0.600	0.073
PATHMORE					0.621	0.092
FACIL_P					0.820	0.012
FACIL_PP					0.860	0.001
WATER					0.820	<0.001
TC					0.247	<0.001
AREA			0.733	0.004		
FORROADS			6.611	0.288		
ASC1			1.247	0.116		
ASC2			1.515	0.018		
McFadden Pseudo R-squared		0.68				
N 526	Number of choices SP+RP			526*6+1*526=3682		

Using the choice set for RP data based on the second iteration of the strategic sampling scheme

Table 9. Welfare effects of changes of forest in proximity of Sarrebourg conditioned on the actual travel mode: comparison of utility specification and sampling strategy (in € per person per visit)

		WTP car	**WTP bike**	**WTP walk**
			One new trekking path	
Simple random sampling	CL	0.102	0.242	0.441
	MXL	0.114	0.218	0.304
Strategic sampling	CL	0.114	0.225	0.308
	MXL	0.115	0.205	0.278
			State to private forests	
Simple random sampling	CL	-0.073	-0.199	-0.450
	MXL	-0.125	-0.345	-0.610
Strategic sampling	CL	-0.142	-0.346	-0.589
	MXL	-0.172	-0.378	-0.595
Share of respondents (%) with travel mode		51	10	39

Table 10. Welfare effects of changes of forests in proximity of Sarrebourg: comparison of travel mode choice approaches and RP and SP datasets.

	WTP New trekking paths	WTP state to private forests
Weighted average of transport modes	0.187	-0.358
Travel mode choice integrated in site selection model	0.093	-0.136
Stated preference data (Choice experiment)	0.120	-

Based on MXL and second iteration of strategic sampling