**Title** A latent class nested logit model for rank-ordered data with application to cork oak reforestation

**Authors** José L. Oviedo[†] and Hong Il Yoo[‡,*]

[†] Institute of Public Goods and Policies (IPP), Consejo Superior de Investigaciones Cientficias (CSIC). Madrid, Spain.
[‡] Durham University Business School, Durham University. Durham, United Kingdom.
[*] Contact author at: Durham University Business School, Mill Hill Lane, Durham DH1 3LB, United Kingdom. Phone: +44 191 33 45544. Email: h.i.yoo@durham.ac.uk.

**Abstract** We analyze stated ranking data collected from recreational visitors to the Alcornocales Natural Park (ANP) in Spain. The ANP is a large protected area which comprises mainly cork oak woodlands. The visitors ranked cork oak reforestation programs delivering different sets of environmental (reforestation technique, biodiversity, forest surface) and social (jobs and recreation sites created) outcomes. We specify a novel latent class nested logit model for rank-ordered data to estimate the distribution of willingness-to-pay for each outcome. Our modeling approach jointly exploits recent advances in discrete choice methods. The results suggest that prioritizing biodiversity would increase certainty over public support for a reforestation program. In addition, a substantial fraction of the visitor population are willing to pay more for the social outcomes than the environmental outcomes, whereas the existing reforestation subsidies are often justified by the environmental outcomes alone.

**JEL classification** C33, C35, C51, Q23, Q51, Q57

**Keywords** Discrete choice, stated preference, willingness-to-pay, forest, land use

**Funding** Oviedo's involvement in this study was funded by the Spanish Ministry of Economy and Competitiveness (VEABA ECO2013-42110-P, I + D National Plan).

**Conflict of interest** We declare that we have no conflict of interest.

**Acknowledgements** We thank Alejandro Caparrós and Pablo Campos for allowing us

to access the data used in this study. We wish to thank Editor Christian Vossler and two anonymous referees for helpful and constructive comments. All views expressed herein are our own.

# 1 Introduction

The valuation of non-market ecosystem services has attracted growing academic interest, derived from the demand for their integration into economic decision-making (Loomis, 2005). Evidence to date suggests that non-market services account for a substantial share of ecosystem values (Caparrós, Campos and Montero, 2003; Wusteman et al., 2014) and often outweigh market services (Bateman et al., 2011).

From the society's perspective, the optimal allocation of land to competing uses is an economic decision task which requires non-market valuation. Forestry, for example, produces multiple ecosystem services, many of which are like public goods and provided without explicit market transactions (e.g. landscape, biodiversity, carbon sequestration). Replacing this land use with, say, croplands or grasslands would result in other non-market services (e.g. pollination, open space) which may be valued differently.

Stated preference techniques facilitate the valuation of non-market services from different land uses, by allowing the researcher to consider a range of prospective ecosystem scenarios for which no suitable revealed preference data exist (Layton and Brown, 2000). The discrete choice experiment (DCE) is probably the most popular stated preference technique for studying willingness-to-pay (WTP) for non-market ecosystem services (Layton and Brown, 2000; Huber, Hunziker and Lehmann, 2011; Johnson et al., 2012; Schulz, Breustedt and Latacz-Lohman, 2013). This technique enables estimating trade-offs across different levels and types of ecosystems services associated with competing land uses.

Mediterranean forestry ecosystems provide an empirical research context which may benefit from the application of the DCE technique. Here, both human interventions and natural processes drive land use competition that has been recognized as a contemporary policy challenge (Doblas-Miranda et al., 2014). For instance, land management promoting hunting uses is expected to result in replacement of the current grasslands by shrublands over time. In addition, the abandonment of traditional agroforestry activities and the natural mortality of trees are expected to lead to gradual replacement of forestlands by shrublands. Since the main effects of these land use changes will not be

observed within a few decades, it is important to employ a technique that can enable the valuation of prospective ecosystem scenarios. In the United States, the DCE technique has already been applied in the analysis of land use competition, very often with a focus on the conversion of land from agricultural to urban uses (e.g. Duke and Ilvento, 2004). It is difficult, however, to transpose the related evidence to Europe where land zoning regulations are much stricter, making this type of conversion less relevant.

A well-known policy affecting Mediterranean forests is the European Common Agricultural Policy (CAP). Under Council Regulation No. 2080/92, the CAP subsidizes reforestation of agricultural lands to increase the provision of forestry-related ecosystem services. In the Mediterranean area, it has preferentially subsidized reforestation using native oaks which are likely to provide higher biodiversity than other species (Santos et al., 2006). In particular, Council Regulation No. 2080/92 sets an additional premium for the renovation and improvement of cork oak stands.

There is, however, little evidence that can facilitate a holistic economic assessment of reforestation in this context. Several features of a reforestation program may affect the future appearance and biodiversity of forests, and induce variations in landscape and biodiversity values across programs. Reforestation programs can also have non-environmental spillover effects on the local economy, for example by providing more outdoor recreation opportunities and boosting employment. There is a need for non-market valuation studies which cover such a broad range of potential outcomes that may influence public preferences for a reforestation program.

We develop an integrative empirical strategy to analyze data from a choice experiment involving cork oak reforestation programs in the Alcornocales Natural Park (ANP). Located in the southern Spanish region of Andalucía, the ANP is a large protected area which mostly comprises cork oak woodlands. The experiment sampled on-site visitors to this natural park. The program attributes include environmental (reforestation technique, biodiversity, forest surface) and social (jobs and recreation sites created) outcomes. Our modeling approach brings together four discrete choice methods which have not been jointly exploited before, though each of them has received a degree of attention in the recent years. We explore their joint usage to conduct an in-depth analysis

of public preferences for reforestation. We anticipate that our approach can be broadly exploited in other stated and revealed preference studies in non-market valuation.

First, instead of eliciting one choice out of several alternatives, the experiment elicits the full ranking of alternatives from most to least preferred. The extra information thus gained can improve statistical precision (Beggs, Cardell and Hausman, 1981) and facilitate empirical identification of flexible choice models (Layton, 2000; Berry, Levinsohn and Pakes, 2004; Train and Winston, 2007). As Scarpa et al. (2011) point out, the former advantage makes rankings especially attractive as an elicitation format for environmental valuation studies that often require sampling from very specific populations (e.g. visitors to a particular location as in our study): a small target population size makes it inherently difficult to obtain an adequate sample size. It is therefore important to consider developing econometric methods for rank-ordered data, to complement the continual effort to understand better and improve ranking survey designs (Caparrós, Oviedo and Campos, 2008; Chang, Lusk and Norwood, 2009; Scarpa et al., 2011; Akaich, Nayga, and Gil, 2013; Louviere, Flynn and Marley, 2015). The existing approach to analyzing rank-ordered data usually exploits extensions and variants of the exploded logit (Chapman and Staelin, 1982), both within (Chang, Lusk and Norwood, 2009; Scarpa et al., 2011; Resano, Sanjuan and Albisu, 2012; Othman and Rahajeng, 2013; Varela et al., 2014) and outside (Fok, Paap and Van Dijk, 2012; Yoo and Doiron, 2013) the environmental valuation literature. Our strategy adds to the empirical practitioner's toolkit an approach building on the nested rank-ordered logit of Dagsvik and Liu (2009) that allows for more plausible substitution patterns. Unlike the exploded logit, the nested rank-ordered logit does not exhibit the independence-of-irrelevant-alternatives property (even without incorporating a mixture specification).

Second, we augment the baseline nested rank-ordered logit model with a finite mixture or latent class specification of person-specific random utility parameters. The resulting model operationalizes Train's (2009, pp.167-168) conceptualization of "mixed nested logit." The finite mixture accommodates interpersonal heterogeneity in systematic tastes for the observed attributes describing alternatives, while the nested logit structure captures correlated residual tastes for a nested subset of alternatives. The

model thus relaxes the common stochastic assumption that once interpersonal heterogeneity in systematic tastes is taken into account, behavioral errors can be treated as independently and identically distributed.

Third, the underlying random utility function is parameterized in the WTP space of Train and Weeks (2005). This allows estimating the population distribution of WTP for each outcome directly, instead of transforming initial preference parameter estimates to derive that distribution. The WTP space makes it convenient to apply and interpret finite mixture models. These models have received growing attention in the non-market valuation literature due to their non-parametric appeal (Claassen, Hellerstein and Kim, 2013; Shulz, Breustedt and Latacz-Lohman, 2013), but proliferate the number of parameters in return for the increased flexibility.

Finally, we compute individual-level statistics to develop further insight into (i) where on the population WTP distribution each person in our sample lies, and (ii) the expected demographic profile of individuals in each preference class. The first issue is addressed by individual-specific coefficients (Train, 2009, Ch 11), which measure what each individual's WTP is expected to be. Examining the distribution of these coefficients over all respondents facilitates the discussion of which specific reforestation outcomes would appeal to more individuals. The second issue is addressed by computing the weighted average of sample characteristics where each individual's posterior probability of membership in a particular class serves as weights, in a similar manner to Hess et al. (2011, p.13). We believe this is a promising way to investigate the association between taste heterogeneity and observed demographic characteristics in the context of heavily parameterized finite mixture models, which are not amenable to repeated estimation of several demographic specifications.

Our preferred model identifies four classes of preferences. The level of status quo aversion and the WTP coefficients on "environmental" and "social" outcomes vary quite distinctively across all classes, with one exception: the WTP for biodiversity is comparable across two largest classes which make up a dominant majority of the ANP visitor population. The latent class nested logit specification allows us to verify that accounting for those four classes almost fully explains why someone who ranks a reforestation

program above the status quo also tends to rank another reforestation program above the status quo, without relying on the within-nest error correlation which cannot be as readily interpreted as unobserved preference heterogeneity. There are apparent variations in the ANP visitors' characteristics that distinguish two minority classes from the dominant majority, as well as from each other.

The remainder of this paper is organized as follows. Section 2 summarizes the relevant policy background and the stated preference data to be analyzed. Section 3 describes our modeling strategy and estimation method. Section 4 presents the empirical results. Section 5 concludes.

## 2 Application and data

We analyze rank-ordered data which originate from the stated preference survey of Caparrós, Oviedo and Campos (2008). The data source study, however, did not exploit the rank-ordered data in full as it focused on the statistical comparison of implied choices from a rank ordering experiment with actual choices from a parallel (pick-one) choice experiment. The authors recoded the data from the former experiment as if only the most preferred alternative in each choice scenario were observed. We exploit the full rank ordering of alternatives in each scenario to estimate a richer behavioral model, which provides a more complete description of public preferences for reforestation. The remainder of this section presents the selected features of the underlying survey that are immediately relevant to our analysis.

### *2.1 Survey background*

The immediate context of the survey is cork oak reforestation in the Alcornocales Natural Park (ANP) in Spain. Located in the southwest of the region of Andalucía, the ANP is a large protected forest (1,677 $km^2$) comprising mainly cork oak woodlands. The ANP is a popular outdoor recreation site, and forestry activities in the ANP (e.g. cork harvesting, grazing and hunting) make a significant contribution to the local economy.

The cork oak is a native species from the Mediterranean basin and has a particularly large presence in the south of Portugal and Spain, though it can be also found in other

countries including Morocco and Tunisia. In the south of Spain, the cork oak faces a regeneration and aging problem due to natural mortality accentuated by a disease known as *La seca*. This problem will eventually lead to the loss of cork oak woodlands, as dead trees will be gradually replaced by shrublands or bare lands. At the time of the survey, it was forecast that the ANP would lose 20% of its cork oak woodlands in 30 years.

Under the Common Agricultural Policy (CAP), the regional government of Andalucía has subsidized landowners who reforest their lands, and provided larger subsidies to promote the use of a native species like the cork oak. Ovando et al. (2007) estimate that between 1993 and 2000, a total of 83,435 hectares of mixed and pure cork oak stands were planted in Spain. The survey we analyze was conducted mainly to study public preferences, quantified as willingness-to-pay (WTP), for conserving and increasing the cork oak forest area in the ANP.

In a recent update to the principles of administering the CAP programs, potential environmental benefits (e.g. landscape, biodiversity, climate change mitigation) are highlighted as major justifications for the reforestation subsidies (European Commission, 2014). Although the survey we analyze was conducted in 2002-2003, our empirical findings are still relevant to the ANP where the regeneration problem continues, and more broadly to contemporary reforestation policies which intend to integrate public preferences into the evaluation of programs with non-market benefits. There are several non-market valuation studies on alternative forestry management practices (e.g. Boyle et al., 2001; Hoyos et al., 2012), but these practices do not affect non-market ecosystem services as much as reforestation does. In particular, given a particular land use (forestry), landscape and biodiversity tend to remain relatively similar across alternative management practices. Moreover, land use changes like reforestation usually cost more and involve bigger implications for the local economy. Studies analyzing non-market benefits of reforestation tend to have a specialist focus, for example on the number of trees planted (Cameron et al., 2002; Kraczwyck, 2012) or on carbon sequestration and soil erosion alongside recreational aspects of forests (Mogas et al. 2006).

We analyze data from an experiment that considers the cork oak. This native

Mediterranean species has been experiencing a regeneration problem, and faces potential land use competition from shrub encroachment and from other species which are easier to establish (e.g. Stone pine). Our valuation study simultaneously addresses a broad array of attributes relevant to assessing reforestation programs. These include environmental outcomes associated with landscape and biodiversity values; the empirical results will thus be highly relevant to policy-making based on the recent update to the CAP. The attributes also include non-environmental social outcomes that may be considered as additional justifications for providing reforestation subsidies. Our study thus expands the available information on non-market values that can be used in land use policy formulations. Our findings can be viewed both as direct evidence from the ANP, and indicative evidence for other Mediterranean forest ecosystems where native tree species are facing similar problems.

### *2.2 Survey design*

The rank ordering experiment consisted of eight different choice scenarios per respondent, and was completed by 450 recreation visitors who were recruited on-site at the ANP. In each scenario, the respondent faced a choice set of three alternatives (two reforestation programs and the status quo) and ranked them from most to least preferred.

The decision to sample on-site visitors was motivated by concerns over the potential relevance of reforestation programs in the ANP to a broader population. Extending the sampling frame, for example to the general population of the Andalucía region or that of Spain, would have weakened the relevance of reforestation programs to many respondents, increasing the non-response rates and, more importantly, the probability that the respondents do not find the choice task consequential. As Carson and Groves (2007) argue, for consequentiality to hold in a stated preference survey, it is important that the respondents care about the potential benefits of the program in question. When the reforestation program focuses on a relatively limited area (a natural park) and on a single tree species (cork oaks) as in our study, active users are likely to make up most of the concerned population; reforestation programs are more likely to be relevant (e.g. in a cost-benefit analysis) to passive users of the general population when the programs

target larger and/or several areas within a region or a country.

Before completing the experiment, each respondent received an information booklet describing the current situation (the status quo) of the cork oak forest area in the ANP and different types of reforestation programs. In addition, the respondents were told that the regional government was planning to carry out a reforestation program to stop the current trend in the loss of the cork oak forest area in the ANP, and that the survey was going to collect information about their preferences to inform the regional government's program design. Such scripts were inserted to increase the respondents' perception that their responses to the hypothetical experiment would influence the actual design of the program, which is another condition for consequentiality (Carson and Groves, 2007).[1]

The booklet described five attributes characterizing a reforestation program and each attribute's possible levels. Then, it announced that the respondents would face a series of choice scenarios where each scenario comprises two alternative reforestation programs and the status quo (which had been already described to them). It also explained that each alternative program included a one-off increase in the regional income tax that would finance implementation of the program; and that the status quo option involved no increase in the regional income tax and no reforestation program. After reading the booklet, the respondents faced eight choice scenarios with different reforestation alternatives resulting from different combinations of attribute levels. In each scenario, they were asked to rank three alternatives (two reforestation programs and the status quo) from most to least preferred. All questionnaires were administered by trained interviewers who clarified any question about the booklet and the survey.

The use of the regional government's income tax as payment-vehicle was to increase the perception that the agency in charge of the reforestation program could enforce payment.[2] In Spain, regional governments are not only responsible for managing natural parks, but also have authorities to impose and collect a share of income taxes. We note that most of the ANP's visitors are from local areas and potentially pay some income taxes to the regional government of Andalucía: indeed, only 2.7% of our sample lived outside Andalucía.

Table 1 summarizes the definition and possible levels of each program attribute.

Three of the five attributes describe environmental outcomes of reforestation: BIO measures the biodiversity of the new forest area, NAT influences what the new forest will look like (a natural forest vs. a homogeneous plantation) and SUR measures the new forest's surface area. The other two attributes describe "social" outcomes, referring to more recreation sites (REC) and new jobs in the local economy (EMP). Recall that at the time of the survey, it was forecast that the ANP would lose 20% of today's surface area in 30 years; the levels of SUR measure the new forest's surface area in terms of percentage points gained relative to this status quo level of loss.

[Insert Table 1 about here]

The status quo is fully defined as: no trees, no technique, no additional recreational area, no new job, 20% reduction of the current forest surface and no tax raise. This definition is operationalized by setting all of the five attributes to zero. Our model specifications will include a reforestation program intercept, allowing "no technique" to be distinguished from "artificial plantation."

The experiment had a main effects design for the five attributes by selecting sixteen treatments from the universe of 1,024 possible combinations ($4^4 \times 2^2$) of the attribute levels in Table 1. A full design of 120 choice sets ($C_2^{16}$) was obtained by forming pairwise combinations of these sixteen treatments. This design permits identification of the main effects of attributes and all effects between treatments.

## 3 Model and estimation

Our latent class nested rank-ordered logit (LC-NROL) model exploits the modeling framework of Dagsvik and Liu (2009) for rank-ordered data, and operationalizes the mixed nested logit approach that Train (2009, pp.167-168) has conceptualized. The resulting model is like a mixed multinomial logit model (McFadden and Train, 2000) in that it incorporates person-specific random parameters to capture panel correlation over 8 choice scenarios and interpersonal heterogeneity in systematic tastes. Once the random parameters are realized, our model is like a nested multinomial logit model (McFadden, 1978) in that it allows correlated unsystematic tastes for the reforestation

alternatives. Our model is parameterized in the willingness-to-pay space of Train and Weeks (2005), to combine the resulting convenience of interpretation with the flexibility of a non-parametric mixing distribution of the random parameters.

### *3.1 Nested rank-ordered logit (NROL)*

In what follows, we initially focus on the nested rank-ordered logit (NROL) of Dagsvik and Liu (2009) that forms the kernel of our mixed model. Let $n = 1, 2, \cdots, N$ index a person; $t = 1, 2, \cdots, T_n$ a choice scenario; and $j \in \mathbf{J} = \{1, 2, 3\}$ an alternative. In our data, $N = 450$, and $j = 1, 2$ are reforestation programs which comprise a nest while $j = 3$ is the status quo. 447 respondents completed all 8 scenarios ($T_n = 8$) while 3 respondents completed only 6 scenarios ($T_n = 6$).

Following the random utility maximization model of McFadden (1974), suppose that in scenario $t$, person $n$ derives utility $U_{nt,j}$ from alternative $j \in \mathbf{J} = \{1, 2, 3\}$

$$U_{nt,j} = V_{nt,j} + \varepsilon_{nt,j} \tag{1}$$

where $V_{nt,j}$ is the systematic utility component which depends on the observed attributes and $\varepsilon_{nt,j}$ is the unsystematic utility component or error term. When $\varepsilon_{nt,j}$ is i.i.d. type 1 extreme value, the resulting probabilistic choice model is the multinomial logit (MNL) model that exhibits the independence of irrelevant alternatives (IIA) property (McFadden, 1974) which, in a nutshell, implies implausibly that a reforestation program and the status quo are equally attractive as substitutes for another reforestation program.

As Herriges and Phaneuf (2002) illustrate extensively in another context, McFadden's (1978) nested multinomial logit (NMNL) model can capture the notion of the reforestation programs being better substitutes for each other than the status quo. The IIA property is relaxed by accommodating positively correlated tastes for the reforestation programs. Specifically, $\varepsilon_{nt,j}$ in (1) are postulated as draws from a generalized extreme value distribution with the joint distribution function

$$\begin{aligned} \Pr(\varepsilon_{nt,1} &\leq \epsilon_{nt,1}, \varepsilon_{nt,2} \leq \epsilon_{nt,2}, \varepsilon_{nt,3} \leq \epsilon_{nt,3}) = e^{w} \\ \text{where } w &\equiv -(e^{-\epsilon_{nt,1}/\tau} + e^{-\epsilon_{nt,2}/\tau})^{\tau} - e^{-\epsilon_{nt,3}}. \end{aligned} \tag{2}$$

The probability that $i \in \mathbf{J}$ is the utility-maximizing alternative, $P_{nt,i}(\mathbf{J}) = \Pr(U_{nt,i} = \max_{j \in \mathbf{J}} U_{nt,j})$, then becomes:

$$P_{nt,i}(\mathbf{J}) = \frac{e^{V_{nt,i}/\tau}(e^{V_{nt,1}/\tau} + e^{V_{nt,2}/\tau})^{\tau-1}}{\left(e^{V_{nt,1}/\tau} + e^{V_{nt,2}/\tau}\right)^{\tau} + e^{V_{nt,3}}} \text{ for } i \in \{1,2\} \tag{3}$$

$$P_{nt,3}(\mathbf{J}) = \frac{e^{V_{nt,3}}}{\left(e^{V_{nt,1}/\tau} + e^{V_{nt,2}/\tau}\right)^{\tau} + e^{V_{nt,3}}} \text{ for } i = 3. \tag{4}$$

Parameter $\tau$ in (3) and (4) is often called the dissimilarity coefficient because $(1 - \tau^2) = corr(\varepsilon_{nt,1}, \varepsilon_{nt,2})$ measures unobservable similarity between the nested alternatives (Ben-Akiva and Lerman, 1985, p.289).[3] Whether $\tau$ lies in the $[0, 1]$ interval can be used to test for whether the modeled behavior is generally consistent with the random preferences that (1) and (2) specify (Train, 2009, p.81).

The random utility maximization model in (1) can also serve as a conceptual basis for the econometric analysis of rank-ordered data, as shown by Beggs, Cardell and Hausman (1981). Suppose that in scenario $t$, person $n$ ranked alternatives $i, k, l$ as the most, second-most, and least preferred respectively, where $i, k, l \in \mathbf{J}$ and $i \neq k \neq l$. Let $r_{nt,ikl}$ denote such a rank ordering. Much as the probability of a choice can be derived as that of a particular realization of possible utility-maximizing choices, $P_{nt,ikl}$ or the probability of $r_{nt,ikl}$ can be derived as that of a particular realization of possible preference orderings: $P_{nt,ikl} = \Pr(U_{nt,i} > U_{nt,k} > U_{nt,l})$. The main challenge has been that apart from the case of the i.i.d. extreme value $\varepsilon_{nt,j}$ implying IIA, the probability of a rank ordering is difficult to derive analytically. Dagsvik and Liu (2009) propose the following way around this challenge to obtain the rank-ordered data counterpart to NMNL.[4]

In obtaining the probability of a rank ordering $r_{nt,ikl}$, $P_{nt,ikl} = \Pr(U_{nt,i} > U_{nt,k} > U_{nt,l})$, NROL exploits logical links between *choice probabilities* that exist regardless of which type of joint distribution is postulated for the error terms. To begin with, consider a pair of alternatives $k$ and $l$. There are three rank orderings in which alternative $k$ is preferred to another alternative $l$ so as to make the statement $U_{nt,k} > U_{nt,l}$ true. First, $k$ is the most preferred and $l$ is the second most preferred: $U_{nt,k} > U_{nt,l} > U_{nt,i}$. Second, $k$ is the most preferred, and $l$ is the least preferred: $U_{nt,k} > U_{nt,i} > U_{nt,l}$. Finally, $k$

is the second most preferred and $l$ is the least preferred: $U_{nt,i} > U_{nt,k} > U_{nt,l}$. Next, note that the first and second of these rank orderings collectively describe the case when $k$ is the most preferred out of 3 available alternatives: i.e. $U_{nt,k} = \max_{j \in \mathbf{J}} U_{nt,j}$. The binary choice probability, $\Pr(U_{nt,k} > U_{nt,l})$, can therefore be decomposed in terms of multinomial choice and rank-ordering probabilities as follows

$$\Pr(U_{nt,k} > U_{nt,l}) = \Pr(U_{nt,k} = \max_{j \in \mathbf{J}} U_{nt,j}) + \Pr(U_{nt,i} > U_{nt,k} > U_{nt,l}) \tag{5}$$

since $\Pr(U_{nt,k} = \max_{j \in \mathbf{J}} U_{nt,j}) = \Pr(U_{nt,k} > U_{nt,l} > U_{nt,i}) + \Pr(U_{nt,k} > U_{nt,i} > U_{nt,l})$. By rearranging (5), one can express the probability of a rank ordering $r_{nt,ikl}$ as a difference between binary and multinomial choice probabilities

$$\Pr(U_{nt,i} > U_{nt,k} > U_{nt,l}) = \Pr(U_{nt,k} > U_{nt,l}) - \Pr(U_{nt,k} = \max_{j \in \mathbf{J}} U_{nt,j}). \tag{6}$$

Equation (6) is a very general result that applies to any internally consistent system of choice probabilities, and does not rely on any particular property of a generalized extreme value distribution. To complete the derivation of NROL, the random utility function in (1) and the associated stochastic assumption in (2) need be used to obtain specific functional forms of $\Pr(U_{nt,k} > U_{nt,l})$ and $\Pr(U_{nt,k} = \max_{j \in \mathbf{J}} U_{nt,j})$. The latter equals the NMNL probability in (3) and (4) above, and can be written as $P_{nt,k}(\mathbf{J})$ using the earlier notation. Next, consider the former or binary probability that alternative $k$ is preferred to another alternative $l$ when the choice set comprises $k$ and $l$ alone: $P_{nt,k}(\{k,l\}) = P_{nt,k}(\{l,k\}) = \Pr(U_{nt,k} > U_{nt,l})$. In case the pair $\{k,l\}$ comprises one reforestation program ($j = 1$ or $j = 2$) and the status quo ($j = 3$), this probability takes the usual binary logit functional form:

$$\begin{aligned} P_{nt,k}(\{1,3\}) = P_{nt,k}(\{3,1\}) &= \frac{e^{V_{nt,k}}}{e^{V_{nt,1}} + e^{V_{nt,3}}} \text{ for } k \in \{1,3\} \text{ and} \\ P_{nt,k}(\{2,3\}) = P_{nt,k}(\{3,2\}) &= \frac{e^{V_{nt,k}}}{e^{V_{nt,2}} + e^{V_{nt,3}}} \text{ for } k \in \{2,3\} \end{aligned} \tag{7}$$

since then two underlying error terms $\varepsilon_{nt,k}$ and $\varepsilon_{nt,l}$ are independent, meaning $\varepsilon_{nt,k} - \varepsilon_{nt,l}$ follows the standard logistic distribution. In case the pair comprises two reforestation

programs, $j = 1$ and $j = 2$, this probability still takes the binary logit form but the index function is now scaled by dissimilarity coefficient $\tau$

$$P_{nt,k}(\{1,2\}) = P_{nt,k}(\{2,1\}) = \frac{e^{V_{nt,k}/\tau}}{e^{V_{nt,1}/\tau} + e^{V_{nt,2}/\tau}} \text{ for } k \in \{1,2\} \tag{8}$$

since a positive correlation (i.e. $1 - \tau^2 > 0$) between $\varepsilon_{nt,1}$ and $\varepsilon_{nt,2}$ makes the variance of their difference smaller than that of a standard logistic random variable. Then, based on (6), one can immediately obtain the functional form of the probability of a rank ordering $r_{nt,ikl}$ as a difference between two well-known quantities, namely binary logit and NMNL probabilities:

$$P_{nt,ikl} = P_{nt,k}(\{k,l\}) - P_{nt,k}(\mathbf{J}). \tag{9}$$

The NROL formula (9) is a closed-form expression since its right-hand side terms can be replaced with closed-form expressions in (3), (4), (7), and (8). The sample log-likelihood function of the NROL model can be constructed in the usual manner by summing the log of the probability of an actually observed rank ordering across the sample, much as the sample log-likelihood function of the NMNL model is constructed by summing the log of the probability of an actually observed choice across the sample. For instance, suppose that person $n$ ranked alternatives 2, 3, and 1 respectively as the most, second-most and least preferred alternatives in scenario $t$. Then, this particular observation $r_{nt,231}$ contributes $\ln(P_{nt,231}) = \ln(P_{nt,3}(\{1,3\}) - P_{nt,3}(\mathbf{J}))$ to the sample log-likelihood where $P_{nt,3}(\{1,3\})$ and $P_{nt,3}(\mathbf{J})$ are as defined in (7) and (4). As the right-hand side of (9) suggests, coding a maximum likelihood estimation routine for this baseline NROL model requires only about as much programming effort as coding a self-written routine for the NMNL model. In our experience, this baseline NROL model can be readily estimated using any of usual gradient-based optimization methods including Newton, BHHH and BFGS.

### *3.2 Latent class NROL in the willingness-to-pay space*

We now turn to a more specific discussion of our LC-NROL model which mixes (9) over a discrete distribution. The systematic utility $V_{nt,j}$ is often parameterized in what Train and Weeks (2005) call the preference space. In this space, our specification is

$$V_{nt,j} = \alpha_n ASC_j^{RF} + \boldsymbol{x}_{nt,j}^{'}\boldsymbol{\beta}_n - \lambda_n TAX_{nt,j} \text{ for } j \in \mathbf{J} \tag{10}$$

where $ASC_j^{RF} = 1[j \neq 3]$ is the nest dummy for the reforestation programs, $TAX_{nt,j}$ is the tax payment for $j$, and $\boldsymbol{x}_{nt,j}$ is the vector of other observed attributes. As described earlier, these variables are set to 0 for the status quo, meaning that $V_{nt,3}$ is normalized to 0. All attributes enter $\boldsymbol{x}_{nt,j}$ linearly, except the forest surface area (SUR) for which we also include its square: since a larger forest area leaves less land for other uses, more of this attribute may not always be preferred. $\alpha_n$, $\boldsymbol{\beta}_n$ and $\lambda_n$ are person-specific random parameters. Following the recent land use studies of Claassen, Hellerstein and Kim (2013) and Schulz, Breustedt and Latacz-Lohman (2013), the joint distribution of these parameters, i.e. mixing distribution, is specified as discrete, to approximate the interpersonal taste distribution without imposing a particular shape on it. This finite mixture or latent class approach has a nonparametric flavor (Train, 2008), much as it does in the context of duration analysis (Heckman and Singer, 1984).[5]

As in other non-market valuation studies, the main parameters of interest include the willingness-to-pay (WTP) for specific attributes, or $\boldsymbol{\omega}_n = \boldsymbol{\beta}_n/\lambda_n$. All results to be reported in the next section are obtained by specifying $V_{nt,j}$ in the WTP space of Train and Weeks (2005) to estimate the distribution of $\boldsymbol{\omega}_n$ directly. Specifically, the utility function (10) is re-parameterized as:

$$\begin{aligned} V_{nt,j} &= \lambda_n\left(\frac{\alpha_n}{\lambda_n} ASC_j^{RF} + \boldsymbol{x}_{nt,j}^{'}(\boldsymbol{\beta}_n/\lambda_n) - \frac{\lambda_n}{\lambda_n} TAX_{nt,j}\right) \\ &= \lambda_n(\kappa_n ASC_j^{RF} + \boldsymbol{x}_{nt,j}^{'}\boldsymbol{\omega}_n - TAX_{nt,j}) \end{aligned} \tag{11}$$

so that $\kappa_n$ and $\boldsymbol{\omega}_n$ become relevant random parameters in lieu of $\alpha_n$ and $\boldsymbol{\beta}_n$. In the case of finite mixture models, moving from the preference space to the WTP space makes

it more convenient to obtain the distribution of $\boldsymbol{\omega}_n$ while preserving the postulated preference structure: the optimal mass points in the parametric space remain invariant to such a move so long as $\lambda_n$ is non-zero at each point.[6] This invariance contrasts with the case of popular continuous mixture models which imply two different preference structures in the two different spaces (Train and Weeks, 2005; Scarpa, Thiene and Train, 2008).

As noted above, our model specifies the error term $\varepsilon_{nt,j}$ to follow a generalized extreme value distribution that allows for a positive correlation between the two reforestation program's error terms, $\varepsilon_{nt,1}$ and $\varepsilon_{nt,2}$. In the standard mixed logit framework, the error term $\varepsilon_{nt,j}$ is i.i.d. type 1 extreme value and it is sometimes the desire to approximate this type of correlation that motivates inclusion of a random alternative-specific constant like $\alpha_n$ in (10) (Herriges and Phaneuf, 2002). Even though $\varepsilon_{nt,j}$ is i.i.d. over alternatives, the resulting composite error terms for the reforestation programs, $(\alpha_n + \varepsilon_{nt,1})$ and $(\alpha_n + \varepsilon_{nt,2})$, would exhibit a positive correlation when the correlation is computed over the population of decision makers; the shared error component $\alpha_n$ ensures that someone with a larger composite error for one reforestation program tends to have a larger one for the other program too. The positive correlation thus induced is a property which pertains to the population and arises from interpersonal taste heterogeneity, and does not pertain to an individual decision maker. In the standard mixed logit framework, an individual decision maker tends to consider reforestation programs and the status quo equally attractive once their observed attributes are taken into account, as one can see by noting that conditional on person-specific utility parameters, the probability of a rank ordering takes the form of the ROL formula that exhibits the IIA property.

Accommodating a positive correlation in $\varepsilon_{nt,j}$ directly, instead of relying on person-specific utility parameter $\alpha_n$ to approximate it, allows our model to provide a more general description of individual choice behavior. Our modeling approach allows for that an individual decision maker tends to find reforestation programs better substitutes for each other than the status quo even after taking into account their observed attributes: conditional on person-specific utility parameters, our approach has the prob-

ability of a rank ordering specified as the NROL formula that does not exhibit the IIA property. As a stochastic assumption, a positive correlation in $\varepsilon_{nt,j}$ over reforestation alternatives which leads to the NROL formula is arguably not only more general but also more natural than independence, since behavioral noises associated with evaluating each reforestation program's attractiveness relative to the status quo are likely to be positively correlated. In addition, a positive correlation in $\varepsilon_{nt,j}$ over reforestation alternatives may reflect inadequate specification of interpersonal taste heterogeneity. Since person-specific random utility parameters multiply alternative-specific attributes $(ASC_j^{RF}, \boldsymbol{x}_{nt,j}, TAX_{nt,j})$, unless taste heterogeneity is adequately modeled, some aspects of taste heterogeneity would enter alternative-specific idiosyncratic errors $\varepsilon_{nt,j}$ like error components (Train, 2009, pp.139-141). Such error components vary with the observed attributes, in terms of which reforestation programs are positively correlated in that they deliver improvement (relative to the status quo) in social and environmental outcomes in return for an increase in taxes.

Of course, the presence and extent of such a positive within-nest correlation are an empirical question, and likely to vary from application to application. We note that much as the NMNL model nests the MNL model, the NROL model nests the ROL model as a special case arising when the dissimilarity coefficient $\tau = 1$ (equivalent to zero within-nest correlation). Our modeling approach allows one to test directly, via $\tau$, for whether there is a within-nest correlation that remains after modeling interpersonal taste heterogeneity, whereas the standard mixed logit approach assumes such a correlation away. In other words, our approach allows the presence of correlated evaluative noises, and also a possible misspecification of taste heterogeneity, to be treated genuinely as an empirical question that can be explored using the usual methods of statistical inference.

For estimation, the sample log-likelihood function is constructed by bringing (9) and (11) together. Let $\boldsymbol{\theta}_n = (\lambda_n, \kappa_n, \boldsymbol{\omega}_n)$, and $\boldsymbol{\theta}_c = (\lambda_c, \kappa_c, \boldsymbol{\omega}_c)$ for $c = 1, 2, \cdots, C$ where $C$ is the number of mass points, or "classes", to be pre-specified. Replacing $\boldsymbol{\theta}_n$ in (11) with $\boldsymbol{\theta}_c$, and then substituting the resulting expression into the NROL formula (9) gives $L_{nt,ikl}(\tau, \boldsymbol{\theta}_c)$, the likelihood of observing person $n$'s rank ordering in scenario $t$ conditional on that she is in class $c$. The corresponding likelihood of observing her

sequence of rank orderings over $T_n$ scenarios is then

$$L_{n|c}(\tau, \boldsymbol{\theta}_c) = \prod_{t=1}^{T_n} \prod_{\{i,k,l\} \in \mathbf{Q}} L_{nt,ikl}(\tau, \boldsymbol{\theta}_c)^{r_{nt,ikl}} \tag{12}$$

where $\mathbf{Q}$ is the set of all possible rank orderings i.e. permutations of $\{1, 2, 3\}$; and $r_{nt,ikl}$ equals 1 if her actual rank ordering in scenario $t$ is as listed in the second set of subscripts and 0 otherwise. For later reference, we define $\mathbf{r}_n$ as a collection of such rank ordering indicators over $T_n$ scenarios.

Since her class is not known, the unconditional likelihood of this sequence is obtained by averaging or mixing $L_{n|c}(\tau, \boldsymbol{\theta}_c)$ over $c$. Let $\pi_c = \Pr(\boldsymbol{\theta}_n = \boldsymbol{\theta}_c)$ denote the population share of class $c$. Person $n$'s contribution to the sample likelihood is then

$$L_n(\tau, \boldsymbol{\theta}, \boldsymbol{\pi}) = \sum_{c=1}^{C} \pi_c L_{n|c}(\tau, \boldsymbol{\theta}_c) \tag{13}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_C)$ and $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_{C-1})$ with $\pi_C \equiv 1 - \sum_{c=1}^{C-1} \pi_c$ by normalization. Formula (13) is the likelihood of person $n$'s rank orderings, $\mathbf{r}_n$, under our LC-NROL model. It is functionally identical to the likelihood of the usual latent class MNL model (see for example, Claassen, Hellerstein and Kim, 2013), apart from that summand $L_{n|c}(\tau, \boldsymbol{\theta}_c)$ is the likelihood of a class-specific NROL model, instead of that of a class-specific MNL model.

All estimates of $(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$ in Section 4 have been computed by applying the method of maximum likelihood. As usual, the maximum likelihood estimates of those parameters can be obtained by maximizing the sample log-likelihood $\ln L = \sum_{n=1}^{N} \ln L_n(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$ with respect to $(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$. The estimates that maximize $\ln L$ can also be obtained by maximizing the expected complete data log-likelihood $Q = \sum_{n=1}^{N} \sum_{c=1}^{C} h_{nc}[\ln \pi_c + \ln L_{n|c}(\tau, \boldsymbol{\theta}_c)]$, which lends itself more easily to numerical optimization by allowing the estimates of $(\tau, \boldsymbol{\theta})$ and $\boldsymbol{\pi}$ to be updated in two separate tasks from iteration to iteration. This latter approach is called the expectation-maximization (EM) algorithm (Bhat, 1997; Train, 2008). We follow the hybrid estimation strategy of Bhat (1997) by using the final EM solution as starting values for maximizing $\ln L$ directly, to double-check convergence and

compute standard errors.[7] Appendix 1 summarizes the EM algorithm for estimating our LC-NROL model and other computational details.

In the EM algorithm's objective function $Q$, $h_{nc}$ refers to person $n$'s *posterior* probability of membership in class $c$. It measures how likely she is in that class given her observed behavior $\mathbf{r}_n$, and is a derived statistic based on model parameters being estimated $(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$ and the data: it is not an extra parameter to be estimated directly. It is defined as

$$h_{nc} = \Pr(\boldsymbol{\theta}_n{=}\boldsymbol{\theta}_c|\mathbf{r}_n) = \frac{\pi_c L_{n|c}(\tau, \boldsymbol{\theta}_c)}{\sum_{s=1}^{C} \pi_s L_{n|s}(\tau, \boldsymbol{\theta}_s)} = \frac{\pi_c L_{n|c}(\tau, \boldsymbol{\theta}_c)}{L_n(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})}. \tag{14}$$

Person $n$'s posterior probability of membership in class $c$ or $h_{nc}$ is larger (smaller) than the population share of that class or $\pi_c$, which may be viewed as her *prior* probability, when her behavior can be fitted better (worse) by assuming that she is in class $c$ than by mixing her likelihood over all classes: that is, by $L_{n|c}(\tau, \boldsymbol{\theta}_c)$ than by $L_n(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$. Note that $\pi_c$ is primarily a quantity that pertains to the population since it is the population share of class $c$, and does not vary across individuals $n = 1, 2, ..., N$. But then there is a $\pi_c$ chance that a randomly sampled individual from the population is a member of class $c$, and it is in this sense that $\pi_c$ can be interpreted as a quantity pertaining to an individual, more specifically as her prior probability of membership in class $c$. By contrast, $h_{nc}$ is inherently a quantity that pertains to a particular individual $n$, since it is this individual's *posterior* probability of membership in class $c$ that is conditioned on the observed sequence of her rank orderings in several choice scenarios.

Since the EM algorithm entails evaluation of each individual's posterior probabilities, $h_{nc}$ for $n = 1, 2, ..., N$ and $c = 1, 2, ...C$, the use of the EM algorithm makes it convenient to compute individual-specific coefficients (Train, 2009, Ch 11) as follows.

$$E(\boldsymbol{\omega}_n|\mathbf{r}_n) = \sum_{c=1}^{C} h_{nc}\boldsymbol{\omega}_c \tag{15}$$

Given our WTP-space parameterization, these coefficients can be interpreted as the expected values of person $n$'s WTP that our model allows us to infer from her behavior. Note that since the posterior probabilities $h_{nc}$ vary across individuals, these coefficients

vary across individuals too. The distribution of these coefficients across 450 people in our sample will be examined to complement the analysis of taste heterogeneity in the population captured by $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$.

# 4 Empirical findings

Following the common practice (Train, 2008; Claassen, Hellerstein and Kim, 2013; Keane and Wasi, 2013; Shulz, Breustedt and Latacz-Lohman, 2013; Yoo and Doiron, 2013), we use the Bayesian Information Criterion (BIC) to choose the optimal number of classes. Our preferred model features four classes and 36 parameters. To facilitate the subsequent interpretation of this model, our discussion begins with the estimation results for more restrictive models involving fewer parameters.

## *4.1 Baseline model estimates*

The first two columns of Table 2 report the nested multinomial logit (NMNL) model and the nested rank-ordered logit (NROL) model, which assume interpersonal taste homogeneity. Each coefficient measures the marginal WTP for a particular attribute in €s, except for the coefficients on SUR and its square $\text{SUR}^2$ which jointly measure the marginal WTP for the forest surface area.[8] To facilitate interpretation, Figure 1 plots the cumulative willingness-to-pay (WTP) for changing the forest surface area from the status quo (SUR = 0) to shown levels. For the NMNL estimation, the data have been recoded as if we only observed the most preferred alternative in each scenario.

[Insert Table 2 about here]

[Insert Figure 1 about here]

In both models, the dissimilarity coefficient $\tau$ lies between 0 and 1, satisfying the sufficient condition for consistency with random utility maximization (Train, 2009, p.81). The implied within-nest correlation, $1 - \tau^2$, is quite large: 0.97 in NMNL and 0.93 in NROL.[9]

The WTP estimates are only slightly smaller in NROL and practically the same in both models. Including an extra native tree species (BIO) in the program is valued at

some €20, almost as much as an extra recreation site (REC) and 30 new jobs (30 × EMP). The use of natural regeneration instead of artificial plantations (NAT) is valued at some €30. The quadratic term of SUR is negative, indicating a decreasing marginal valuation of the forest surface area: Figure 1 shows that the marginal WTP is mostly positive over the proposed range of increases, with the cumulative WTP peaking at some €60 near the highest increase.

While both models arrive at the same substantive results in terms of attribute valuation, NROL delivers more precise estimates than NMNL. The marginal utility of money (TAX × -1), the alternative-specific constant for reforestation programs measuring the aversion to the status quo ($ASC^{RF}$) and the dissimilarity coefficient ($\tau$) are significant at the 1% level in NROL, and by contrast insignificant at any conventional level in NMNL.[10] Both $ASC^{RF}$ and $\tau$ parameters can explain why when someone's utility from one reforestation option is higher than the status quo, her utility from the other reforestation option is also likely to be higher. Berry, Levinsohn and Pakes (2004) and Train and Winston (2007) report similar findings on the substitution parameters of normal error-component logit models for rank-ordered data.

The precise estimate of $\tau$ in NROL (0.26) is also about 60% larger than its imprecise counterpart in NMNL (0.16). But the implied within-nest correlation is still quite large at 0.93. In a revealed preference study, observing all relevant attributes is difficult and it is natural to consider unobserved attributes shared by nested alternatives as the main source of such a correlation (Ben-Akiva and Lerman, 1985, pp.285-294). In a stated preference study, however, it may be more natural to consider omitted interpersonal taste variation as the main source. A careful experimental design would have most of key attributes observed as the descriptors of choice scenarios, leaving less room for omitted attributes that could induce subjective evaluative noises to be correlated over alternatives. As Train (2009, pp.139-141) explains, omitted taste heterogeneity can be observationally equivalent to omitted shared attributes.

In Table 2, the 2-class version of the latent class NROL (LC-NROL) model is reported in the third (Class 1) and fourth (Class 2) columns. The average of class-specific estimates, weighted by their population shares, is reported in the last column (Mean).

This average can be interpreted as the estimated population mean of each random utility parameter in the WTP space. Specifically, let $\theta_{n,k}$ be one of random utility parameters in $\boldsymbol{\theta}_n$, and $\theta_{c,k}$ be the corresponding class-specific parameter in $\boldsymbol{\theta}_c$, where $\boldsymbol{\theta}_n$ and $\boldsymbol{\theta}_c$ are as defined in the previous section. The population mean of $\theta_{n,k}$, then, can be estimated by plugging the estimates of $\pi_c$ and $\theta_{c,k}$ into the following formula

$$E(\theta_{n,k}) = \sum_{c=1}^{C} \Pr(\boldsymbol{\theta}_n = \boldsymbol{\theta}_c)\theta_{c,k} = \sum_{c=1}^{C} \pi_c \theta_{c,k} \tag{16}$$

where $\pi_c$ is the population share of class $c$ and the total number of classes $C = 2$ in the present context.

The results present prima facie evidence that the high within-nest correlation resulted from omitted taste heterogeneity as suspected. Capturing two preference segments leads to an increase in $\tau$ from 0.26 to 0.89, implying a substantial decline in the within-nest correlation from 0.93 to 0.21. Class 2 makes up 95% of the population who value both environmental and "social" outcomes of reforestation. They have a similar preference structure as found in NROL. Class 1 captures the remaining 5% who have evidently distinct preferences. This minority is unlikely to find any reforestation program more attractive than the status quo. They have preferences for the status quo ($\text{ASC}^{\text{RF}} < 0$, c.f. $\text{ASC}^{\text{RF}} > 0$ for the majority), are much more tax-sensitive (TAX $\times -1 = 0.035$, c.f. 0.014 for the majority), and accordingly have smaller WTP, most of which are insignificant both practically and statistically.

### *4.2 Preferred 4-class LC-NROL model estimates*

Table 3 reports our preferred 4-class LC-NROL model, and Figure 2 plots the implied cumulative WTP for changing the forest surface area from the status quo to shown levels. The last column of Table 3, as in Table 2, presents the estimated population mean of each parameter that is derived using equation (16), where $C = 4$ now. This model appears to capture the main aspects of the underlying taste heterogeneity well enough to nearly eliminate the within-nest correlation.

[Insert Table 3 about here]

[Insert Figure 2 about here]

The dissimilarity coefficient $\tau$ is now 0.99, implying a very modest correlation of about 2%. Put another way, consider the question of why someone who prefers a reforestation program to the status quo is also likely to prefer another reforestation program to the status quo. The present model addresses this question with reference to her systematic tastes for the observed attributes, without relying on the residual correlation in $\varepsilon_{njt}$ which cannot be as readily interpreted. The estimated heterogeneity features a "complex" structure, in the sense of Keane and Wasi (2013, p.1032), as each class is quite distinct from another in terms of its valuation of different attributes.

Class 1 makes up 4% of the population and resembles their counterpart in the 2-class LC-NROL model of Table 2. They show preferences for the status quo, and have small and insignificant WTP for most attributes. Their only significant WTP is for the forest surface area. Figure 2 illustrates that for maintaining the current size of the forest surface area (SUR = 20), this class is willing to pay €38, more than others except Class 3. The amount grows to €47 for increasing the size by 20% (SUR = 40) around at which the marginal WTP turns negative, but even then it is not large enough to reverse their preferences for the status quo ($\text{ASC}^{\text{RF}}$ = -€116).

The remaining three classes exhibit preferences for reforestation ($\text{ASC}^{\text{RF}} > 0$), and may be viewed as a breakdown of the dominant majority segment in the 2-class model. The two largest segments are Class 2 and Class 3, making up 52% and 41% of the population respectively. Both classes are willing to pay more than €18 for the use of an additional tree species (BIO), but their preferences for the other attributes are remarkably different.

Class 2 represents those who are primarily concerned with what we have called "social" outcomes. Only this class has statistically significant WTP for an extra recreation site (REC) and an additional job in the local economy (EMP). Moreover, their WTP of €45 for REC and €1.38 for EMP are evidently larger than any other class's, including Class 3 who are willing to pay €5 for REC and €0.29 for EMP. Interestingly, Class 2 also has rather large and negative WTP of -€18 for the use of natural regeneration (NAT), meaning that their preferred implementation technique is artificial plantation.

This estimate agrees with their preferences for social outcomes, though it is statistically insignificant. The use of artificial plantation may be associated with a boost to the local economy over and above the impact of EMP, as it requires more initial investment and subsequent maintenance expenditure than natural regeneration.

In contrast, Class 3 represents those who are primarily concerned with the natural appearance and size of the forest area, or the "environmental" outcomes of reforestation. They value the use of natural regeneration at €84, which is much larger than the second largest amount of €11 that Class 4 is willing to pay. As Figure 2 shows, over the whole range of proposed increases, they are also willing to pay visibly more than Class 2 for any given increment of the forest surface area; and over most of this range, they are also willing to pay more than any other class. Thus, people in this class place greater weight on the potential landscape values of reforestation which are to be realized in the future, and less on the direct social benefits.

Class 4, making up 3% of the population, captures those who have some preferences for reforestation but are also highly cost-sensitive: $\text{ASC}^{\text{RF}}$ is still positive but much smaller, and TAX × -1 is much larger, than in Class 2 and Class 3. In line with their large TAX × -1 or marginal utility of money, they have small WTP for most attributes, though they qualitatively resemble Class 3 in weighting the future environmental outcomes more than the immediate social outcomes. They are thus likely to consider the costs of some reforestation programs unjustified by the proposed outcomes.

For comparison with our preferred model, we have also estimated two standard mixed ROL models in the WTP space: "independent" and "correlated" mixed ROL models. Apart from the WTP space parameterization, these models are like the mixed ROL models of Layton (2000), Calfee, Winston and Stempski (2001), Train and Winston (2007) and Train (2008) which conceptualize an observed rank ordering as a realization of random preference orderings. Following Train and Weeks (2005), our specification assumes a multivariate normal distribution of random utility parameters $\ln(\lambda_n)$, $\kappa_n$ and $\boldsymbol{\omega}_n$, where the notations refer to the random utility function in equation (11). The independent mixed ROL model assumes away all correlations between 8 random utility parameters, and requires estimating 16 coefficients comprising each parameter's

population mean and standard deviation. The correlated mixed ROL model comes much closer to our preferred model in terms of flexibility of the mixing distribution, as it accommodates the full set of such correlations, thereby requiring estimation of 44 coefficients (8 means; 8 standard deviations; 28 covariances).[11] Both models are estimated via the method of maximum simulated likelihood, using 1000 Halton draws for Monte Carlo integration. Table A1 in Appendix 2 summarizes the results.

In the present application, all three models (LC-NROL, independent and correlated mixed ROL) turn out to be similar in terms of the estimated mean WTP for reforestation outcomes. But for an analyst who would like to impose a minimal set of a priori assumptions on the distribution of random utility parameters and that of error terms, our modeling approach provides distinctive advantages in terms of flexibility, and potentially also in terms of computational convenience. Since the ROL formula implicitly imposes $\tau = 1$ a priori, neither of the mixed ROL models allows verifying directly whether that model eliminates the within-nest correlation which is a key feature of the NMNL and NROL results in Table 2. We also note that using our self-written program in TSP International, it took slightly over 26 minutes (including the time spent on the EM algorithm's iterations) to execute the production run for the LC-NROL model; using the default settings of Stata command -mixlogitwtp- (Hole, 2015), it took about 68 minutes to estimate the independent mixed ROL model and almost 31 hours to estimate the correlated mixed ROL model.[12] Making direct comparisons of the three models in terms of run time is difficult partly because we used two different software packages, and inherently because each model involves different sets of parameters. In our view, the results nevertheless provide a convincing indication that the LC-NROL model provides a computationally attractive alternative to the mixed ROL models.

### *4.3 Individual-specific WTP coefficients*

Our preferred LC-NROL model has identified distinct preference segments which are four major mass points in the population taste distribution. This does not mean that one of the four segments will be able to describe every individual's tastes exactly. As Clarke and Muthen (2009) suggest, it may be appropriate to think of a specific individual as

having "fractional membership" in all classes, with her tastes possibly lying somewhere between the four major mass points. We present individual-level statistics which use the posterior probability $h_{nc}$ in equation (14) as a measure of each person $n$'s fractional membership in class $c$.

Table 4 summarizes individual-specific WTP in our sample of 450 respondents, computed using formula (15) and the estimated 4-class model. For the forest surface area (SUR), the individual-specific cumulative WTP for improving this attribute from 0 to a specified level has been computed in an analogous manner to Figure 2. Since each of 450 respondents has her own individual-specific WTP, the summary statistics are computed over 450 data points where each data point pertains to a particular respondent.

[Insert Table 4 about here]

As discussed earlier, each respondent $n$'s individual-specific WTP can be viewed as what that respondent's WTP is expected to be, on the basis of how she actually ranked alternatives in each scenario. Put another way, consider a hypothetical case where one can present each respondent with a very large number of choice scenarios, say 150, without inducing fatigue and heuristics. This would allow one to estimate the NROL model separately for each of 450 respondents using 150 observations on that respondent, thereby obtaining 450 sets of NROL estimates in total. Then, one may proceed to analyze within-sample taste heterogeneity by studying the distribution of these 450 sets of individual-specific estimates. In realistic settings, of course, each respondent can be asked to complete only so many choice scenarios (in our case, 8 scenarios) making it difficult to operationalize and justify such individual-level estimation. By combining the estimated population taste distribution (that is, the results presented in Table 3) with each person's observed sequence of rank orderings in 8 scenarios, nevertheless, one can still obtain individual-specific WTP and analyze its distribution as one would analyze the distribution of individual-specific NROL results in the hypothetical case.

The distribution of individual-specific WTP calls for some caution against interpreting the earlier class-specific results as though every person's tastes could be exactly described by one of the four classes. For most attributes, the interquartile range ($Q_3 - Q_1$) of individual-specific WTP is narrower than what the direct comparisons of Class 2

and Class 3, together making up 93% of the population, would suggest. This means that many respondents cannot be exactly classified into one particular class, though most of respondents have quite a large posterior probability of being in one of the four classes: over the 450 respondents, the first quartile, median and third quartile of the maximum posterior membership probability, $\max(h_{n1}, h_{n2}, h_{n3}, h_{n4})$, are 0.68, 0.86 and 0.96 respectively.

But otherwise, this distribution reinforces the main policy implications of the class-specific results. For instance, our earlier findings showed that both Class 2 and Class 3 valued biodiversity but were very different in their valuation of other outcomes. Here, an extra tree species (BIO), the use of natural regeneration (NAT), and an additional recreation site (REC) are comparable in terms of the median and mean of the individual-specific WTP. But interpersonal variations in the WTP are only minimal for BIO, whereas they are sizable for REC and even larger for NAT. The policymaker could anticipate that a reforestation program prioritizing the improvement of biodiversity is likely to attract wider support than those prioritizing that of the other attributes. Also, in line with the diminishing marginal WTP for the forest surface area (SUR) found in all classes except Class 4, interpersonal variations in the cumulative WTP for a given increase in SUR are smallest at the highest level of increase as the marginal WTP approaches 0. In this case, the policymaker could expect wide agreement when the proposed surface area increase is close to the highest level and less agreement when it is smaller.

### *4.4 Preference segments and individual characteristics*

We now turn to the question of the expected characteristics of individuals whose preferences are best approximated by a particular mass point in the population taste distribution, i.e. class of preferences. Such an analysis provides an opportunity to check the plausibility of preference parameters estimated from stated preference data, by comparing them against characteristics related to the environmental goods being valued (e.g. in our application, the use and knowledge of the natural park in question). In addition, when a large number of demographic characteristics are at disposal, it is rather com-

putationally cumbersome to search for observed heterogeneity in preference parameters by repeatedly estimating random parameter model specifications that incorporate different subsets of interaction terms involving those characteristics and attributes. For a finite mixture model like ours, such demographic specification search is also complicated by that the optimal number of classes as indicated by a model selection criterion may vary depending on which subsets of the interaction terms are included.[13] In such cases, variations in the expected characteristics across classes may provide practically useful insight into potential observed heterogeneity in preferences.

The approach we take is to produce the weighted sample characteristics as in Hess et al. (2011, p.13). Specifically, suppose that $z_n$ is a personal characteristic of interest, say person $n$'s age. Using a sample of $N$ people and a model specifying $C$ classes, the expected age of someone in class $c$ can be computed as

$$\sum_{n=1}^{N} \left( \frac{h_{nc}}{\sum_{n=1}^{N} h_{nc}} \right) z_n. \tag{17}$$

where $h_{nc}$ refers to person $n$'s posterior probability of membership in class $c$ as defined in equation (14). The weighting of $z_n$ in formula (17) is based on the notion of fractional class membership: in total, $\sum_{n=1}^{N} h_{nc}$ individuals in the sample belong to class $c$ since a fraction $h_{nc}$ of person $n$ belongs to class $c$.

Table 5 reports the actual mean characteristics of our sample, followed by the expected characteristics of each class. In Appendix 2, Table A2 provides the full definition of these characteristics. The deviations from the actual mean tend to be much more pronounced for the two smallest segments, Class 1 and Class 4, than Class 2 and Class 3. Many of those deviations provide a further insight into our earlier findings on the former two classes.

[Insert Table 5 about here]

As discussed earlier, both Class 1 and Class 4 have small and insignificant WTP for most outcomes of reforestation. Now, this finding may be associated with their relatively tight budget constraints. Someone in either class is expected to have a smaller per capita monthly family income than the sample mean. For Class 1, the likelihood of being in

employment is much smaller too. While both classes also feature higher likelihoods of having a university degree and visiting the Alcornocales Natural Park (ANP) for active tourism (e.g. hiking, climbing, cycling), it is difficult to speculate on a specific mechanism through which those characteristics may influence the WTP.

Recall that Class 1 stood out from other classes including Class 4 in terms of preferring the status quo to reforestation. Now Class 1 stands out in terms of the use and knowledge of the ANP: they visit the park more at both extensive and intensive margins, and are more likely to be satisfied with their current visit. Moreover, their expected spending on the current visit is close to the sample mean, in contrast to that of Class 4 which is much smaller, despite both classes facing relatively tight budget constraints. The use and knowledge variables can be viewed as measures of preferences that the respondents directly reported, whereas our estimated model yields measures of preferences inferred from their ranking responses. The expected profile of Class 1 in terms of use and knowledge suggests that they like the ANP as it is today and may see a little need for any substantial change. This agrees with their estimated preferences for the status quo and insignificant WTP for all outcomes but the forest surface area (SUR), a 20-point increase in which is required to maintain today's surface area 30 years later. The results support that our model provides a good approximation to the underlying public preferences.

That the expected profiles of Class 2 and Class 3 resemble each other so closely is rather surprising because their preferences are remarkably different: the immediate social outcomes are of primary concern to Class 2 whereas the long-term environmental outcomes are to Class 3. Considering that these two classes make up 93% of the population, it appears to be the case here that as usual, much of interpersonal taste heterogeneity cannot be explained with reference to observed individual characteristics. In a contingent valuation study, Nunes and Schokkaert (2003) find that variables measuring attitudes towards donation and specific environmental issues explain interpersonal variations in the WTP, even when most of usual observed characteristics do not. Such attitudinal information, however, is not available in our data.

# 5 Conclusion

We have specified and estimated a latent class nested rank-ordered logit model to analyze preferences for cork oak reforestation in the Alcornocales Natural Park (ANP), Spain. Our preferred 4-class model shows that a vast majority of the ANP visitors have positive willingness-to-pay (WTP) for some subsets of five reforestation outcomes under consideration. Accounting for those four classes almost fully explains why someone who ranks a reforestation program above the status quo also tends to rank another reforestation program above the status quo, without resorting to the within-nest correlation in the unsystematic component of utility that is more difficult to interpret than heterogeneous WTP. There is much heterogeneity in the valuation of environmental and social outcomes across the two largest preference segments in our study. The two other preference segments have statistically insignificant and also often practically small WTP for most outcomes, but they make up less than 10% of the population. The empirical findings have broad implications for policy making and future research as follows.

Two aspects of the estimated public preferences call for more integrative evaluation of tax-financed reforestation programs, such as those promoted by the European Common Agricultural Policy. First, a substantial fraction of the ANP visitors place greater weight on the social outcomes of a reforestation program (new jobs and recreation areas in our application) than environmental outcomes per se. At least when the area to be reforested makes a significant contribution to the local economy as the ANP does, social benefits would deserve due consideration alongside environmental benefits. Second, biodiversity (measured as the number of native species used in reforestation) emerges as an important attribute for most visitors, with small variations in the WTP distribution at both the population and sample levels. Prioritizing biodiversity would therefore increase certainty over public support for a reforestation program. This finding is also of particular relevance to the growing interest in the use of reforestation to deter climate changes, which takes carbon sequestration as the main objective. Deploying a single fast-growing species would help meeting this goal rapidly and perhaps at lower monetary costs. However, public preferences for biodiversity and landscape may not

be compatible with the plantation of species with high potential for carbon sequestration (e.g. eucalyptus). This implication also applies to bioenergy policies promoting the plantation of exotic species for biomass production (e.g. the species Paulownia, which originates from China, has been recently tested in the southern forestlands of Spain for its potential use in bioenergy production).

In the presence of land use competition, designating any area for reforestation is likely to generate both assenting and dissenting voices. The estimation results suggest that public opinion on a reforestation project is more likely to be homogeneous when the affected area is larger: the dispersion of the individual-specific WTP for the forest surface area becomes smaller relative to its mean when the incremental area becomes larger, due to the diminishing marginal WTP. An immediate practical implication is that a possible controversy surrounding a small scale pilot project would not be necessarily carried over to a full scale project.

Understanding the association between the individual's background and preferences will be a step towards reconciling studies involving different samples and policy contexts. In the present study, we find that the expected profiles of individuals in the two minor segments with small WTP are quite distinct from that of the rest, in terms of experiences with the ANP and socioeconomic characteristics. No evident variation, however, exists across individuals with large WTP for social outcomes and those with large WTP for environmental outcomes. It is to be seen if collecting deeper background information on each person, such as metrics of their general attitudes towards environmental and other public goods, could help developing insights into the sources of heterogeneity in public preferences for reforestation.

We conclude with further remarks on the directions for future research. The respondent population in the present study comprises the recreational users of the ANP. These users represent an important part of the relevant population that benefit from cork oak reforestation in the ANP. Our analysis is partly limited because the sampling scope does not incorporate passive users from a more general population, although it is unclear what the optimal scope of an extended sampling framework should be. In the context of our latent class approach, it is possible that further research using more

broadly sampled data with passive users identifies more nuanced preference segments. Biodiversity in the present study was a count of different native tree species to inhabit the reforested area. Our finding on its suitability as a target attribute warrants the analysis of other aspects of biodiversity too, by incorporating outcome measures in terms of wildlife and flora. Finally, following up on the earlier discussion in relation to climate change policies, the relative valuation of biodiversity and the time required to complete reforestation can be advanced as a research question of contemporary policy relevance.

## Notes

[1]Several empirical studies have recently shown the importance of consequentiality by comparing stated preference surveys (Herriges et al., 2010); financially binding experiments with different provision rules with a stated preference survey (Vossler and Evans, 2009; Vossler et al., 2012); and a real referendum with a stated preference survey (Vossler and Watson, 2013). These studies find that WTP differs depending on whether or not respondents believe that the survey is minimally consequential, and that hypothetical WTP and real WTP converge when respondents believe that their responses will have an impact on policy design. Although these studies focused on discrete choice referendum questions, the implications of consequentiality may apply to the hypothetical scenarios in other survey-based discrete choice methods.

[2]Carson and Groves (2007) recommend the use of coercive payment vehicles, such as a tax, in a preference elicitation survey as they are incentive-compatible.

[3]Note that when $\tau = 1$, formulas (3) and (4) simplify to the MNL model.

[4]The idea of using choice probabilities to derive a rank-ordering probability has been proposed by McFadden (1986), who acknowledges the related contributions of Falmagne (1978) and Barberá and Pattanaik (1986). Layton and Levine (2003) exploit it to derive a probit model for partially ranked (best-worst) data in the presence of a large number of alternatives, and develop an accompanying Bayesian estimation algorithm. Layton and Lee (2006a; 2006b) exploit it to develop likelihood ratio tests for the poolability of different formats of stated preference responses, including rankings and ratings. To our best knowledge, however, the study of Dagsvik and Liu (2009) is the first one to derive and estimate the nested rank-ordered logit model which relaxes IIA of the exploded logit while maintaining a closed-form likelihood.

[5]The latent class approach can be exploited for a number of distinctive modeling uses. Building on the idea of Heckman and Singer (1984), our usage of the latent class approach is to specify a type of mixed logit specification which is called, inter alia, finite mixture logit, discrete mixture logit, or non-parametric mixed logit (Train, 2008; Claassen, Hellerstein and Kim, 2013; Keane and Wasi, 2013; Yoo and Doiron, 2013). As

the last name, due to Train (2008), makes explicit, this specification uses a discrete distribution with $C$ support points as a tool to obtain a non-parametric approximation to the unknown population distribution of random utility parameters, without committing to a particular parametric form (e.g. multivariate normal) of the population distribution a priori. Given this objective, it is desirable to specify as many support points as compatible with a model selection criterion (e.g. Bayesian Information Criterion) to accomplish a better approximation. For ease of exposition, it is common practice to adopt the typical latent class parlance and interpret each support point as preference "class" or "segment", and the probability mass at each point as "class share" or "prior probability of class membership".

An alternative use of the latent class approach is to specify an endogenous segmentation model, a la Bhat (1997). Instead of viewing a discrete distribution as a tool to obtain a non-parametric approximation, this model makes a stronger structural assumption that there are indeed $C$ different preference classes and decision makers are probabilistically assigned to different classes. The assignment probabilities are sometimes called class shares and often modeled as a function of demographic characteristics.

The use of the latent class approach to obtain a finite mixture model or an endogenous segmentation model thus has distinct conceptual foundations. Moreover, while the finite mixture model has class shares (i.e. probability mass points) invariant with respect to demographic characteristics, in practice it is not necessarily a restricted functional form of the endogenous segmentation model. As Bhat (1997) points out, the number of classes, $C$, that one can empirically identify from a data set tends to be smaller when class shares are allowed to vary with demographic characteristics: estimating more class share parameters entails estimating fewer utility parameters. It is therefore difficult to recast a preferred finite mixture model as an endogenous segmentation model: such recasting often requires reducing the number of classes, which is at odds with the initial non-parametric approximation motive of specifying the finite mixture model.

[6]We have, however, found it useful to follow the advice of Shachar and Nalebuff (2004, p.388) on numerical optimization and estimate our model in both spaces to double-check convergence. Over many sets of starting values, the WTP space model often resulted in

a worse log-likelihood than the equivalent preference space model.

[7]The estimation routine has been written in TSP International 5.1 and is available from the authors upon request.

[8]This latter marginal WTP equals $\omega_{SUR} + 2\omega_{SUR^2}\text{SUR}$.

[9]Following Layton (2000), Calfee, Winston and Stempski (2001), Berry, Levinsohn and Pakes (2004), Train and Winston (2007), and Dagsvik and Liu (2009) among others, we focus on modeling an observed rank ordering as a realized preference ordering, instead of modeling it as a sequence of choices. When there are three alternatives per choice set and the ROL model is the true model, a single observation on a rank-ordering can be exploded into two pseudo-observations on the best and second-best choices made in sequence (Beggs, Cardell and Hausman, 1981; Train, 2009, pp.156-158). Based on this type of result, some studies contend that one should test for the poolability of pseudo-observations on the best and second-best choices a la Chapman and Staelin (1982) and Ben-Akiva, Morikawa and Shiroish (1992), before making use of rank-ordered data. But except when one adopts a sequential choice model as the underlying behavioral framework so that a rank ordering is viewed as a sequence of repeated choices instead of a realized preference ordering (Giergiczny et al., 2013), such testing procedures entail a stringent maintained assumption that the ROL model is the true model so that the probability of a rank ordering equals a product of marginal choice probabilities (Hausman and Ruud, 1987). The Monte Carlo study of Yan and Yoo (2014) shows that such testing procedures are highly sensitive to the maintained assumption: even when the ROL model is a slightly misspecified model (e.g. because the true error distribution is i.i.d. normal instead of i.i.d. type 1 extreme value), popular tests falsely reject poolability almost always in plausible sample size configurations. Both our NMNL and NROL estimation results reject $H_0 : \tau = 1$ at the 1% level, soundly suggesting that the error terms are not i.i.d. type 1 extreme value.

[10]Strictly speaking, $\text{ASC}^{\text{RF}}$ measures the constant utility change from choosing a reforestation program using artificial plantation. For simplicity, our discussion treats it as that from choosing a reforestation program; this is adequate for our purpose because in all estimation results, the sum of $\text{ASC}^{\text{RF}}$ and the WTP for natural regeneration

(NAT) has the same sign as $ASC^{RF}$. While TAX × -1 is a composite parameter capturing both the marginal utility of money and the overall scale of utility, we abstract from the conceptual distinction between the two components as they are observationally indistinguishable.

[11] A discrete mixing distribution implicitly allows for an unrestricted pattern of correlations between random utility parameters, as no restriction is placed on how different one support point (i.e. class-specific preference vector) should be from another.

[12] All our estimation results were obtained using a Windows 7 PC running on Intel i7-4790 CPU and 32 GB of RAM.

[13] One may also incorporate demographic characteristics by specifying the population class shares, $\pi_c$ for $c = 1, 2, ..., C$, directly as a function of demographic characteristics. From a statistical perspective, this may be the most appealing way to explore the association between demographic characteristics and preference classes. But, as we discuss in footnote 5, such an "endogenous segmentation model" specification is conceptually distinct from our use of the latent class approach. Besides, we note that our preferred 4-class model cannot be directly "generalized" by specifying the population shares to vary with the characteristics listed in Table 5, as the maximum likelihood estimation routine then fails to achieve convergence, suggesting that the resulting model is not empirically identified. For empirical identification, the dimension of the model may need to be curtailed by reducing the number of classes, thereby compromising the quality of the non-parametric approximation and comparability with our preferred model; and also by reducing the number of demographic characteristics, which would require cumbersome specification search especially because the optimal number of classes is likely to vary with which particular subset of characteristics are allowed to affect class shares.

# Appendix 1: Summary of EM algorithm

This appendix summarizes the expectation-maximization (EM) algorithm for estimating our latent class nested rank-ordered logit (LC-NROL) model. The use of the EM algorithm has been motivated by generic numerical difficulties associated with estimating finite mixture or latent class models via direct maximization of the sample log-likelihood function, not by any peculiar issue arising from estimating the LC-NROL model. While our own program was written in TSP International 5.1, it does not rely on any of TSP International's specialized features. Programming the EM algorithm entails setting up a rudimentary loop over updating tasks (20) and (21) to be explained below, and can be implemented in any software package that allows the user to supply a self-written log-likelihood function e.g. Stata (Pacifico and Yoo, 2013).

Unless specified otherwise, we follow the notations introduced in Section 3, and let $n = 1, 2, ..., N$ index individuals and $c = 1, 2, ..., C$ index classes. Our objective is to estimate three types of parameters: dissimilarity coefficient $\tau$, class-specific preference parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_C)$ in the willingness-to-pay space, and the population share of each class $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_{C-1})$. The population share of class $C$, $\pi_C$, is not a free parameter to be estimated since the class shares must add up to 1 and hence $\pi_C \equiv 1 - \sum_{c=1}^{C-1} \pi_c$. As discussed at the beginning of Section 4, the total number of classes $C$ needs be set by the researcher: the common empirical practice is to estimate several specifications that vary in $C$, and choose one that results in the best Bayesian Information Criterion (BIC).

The sample log-likelihood function, $\ln L(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$, can be constructed in the usual manner by summing the log of each person's likelihood, $L_n(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$ in equation (13):

$$\ln L(\tau, \boldsymbol{\theta}, \boldsymbol{\pi}) = \sum_{n=1}^{N} \ln L_n(\tau, \boldsymbol{\theta}, \boldsymbol{\pi}) = \sum_{n=1}^{N} \ln \left( \sum_{c=1}^{C} \pi_c L_{n|c}(\tau, \boldsymbol{\theta}_c) \right). \tag{18}$$

As explained in detail around equation (12), the kernel function $L_{n|c}(\tau, \boldsymbol{\theta}_c)$ uses class $c$'s preference parameters to evaluate the baseline NROL model's likelihood of observing person $n$'s actual sequence of rank orderings over choice scenarios. Since $L_{n|c}(\tau, \boldsymbol{\theta}_c)$ has a closed-form expression, the sample log-likehood $\ln L(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$ also has a closed-form expression. It is therefore possible to proceed in the usual manner to compute the maximum likelihood estimates, say $\{\tau^{ML}, \boldsymbol{\theta}^{ML}, \boldsymbol{\pi}^{ML}\}$, by applying gradient-based optimization methods (e.g. Newton method or quasi-Newton methods like BFGS) to maximize $\ln L(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$ with respect to $\{\tau, \boldsymbol{\theta}, \boldsymbol{\pi}\}$. But, as Bhat (1997) and Train (2008) note in the context of latent class multinomial logit models, maximizing the sample log-likelihood of a finite mixture model tends to be susceptible to convergence

failures which often arise as a numerical optimizer gets trapped in flat regions of the sample log-likelihood function.

It turns out that parametric values $\{\tau^{ML}, \boldsymbol{\theta}^{ML}, \boldsymbol{\pi}^{ML}\}$ that maximize $\ln L(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$ should, in theory, also maximize another objective function $Q(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$ and vice versa. The EM algorithm is an estimation strategy that aims at obtaining $\{\tau^{ML}, \boldsymbol{\theta}^{ML}, \boldsymbol{\pi}^{ML}\}$ by maximizing this alternative objective function which is specified as

$$\begin{aligned} Q(\tau, \boldsymbol{\theta}, \boldsymbol{\pi}) &= \sum_{n=1}^{N}\sum_{c=1}^{C} h_{nc}(\tau, \boldsymbol{\theta}, \boldsymbol{\pi}) \times \big(\ln \pi_c + \ln L_{n|c}(\tau, \boldsymbol{\theta}_c)\big) \\ &= \sum_{n=1}^{N}\sum_{c=1}^{C} h_{nc}(\tau, \boldsymbol{\theta}, \boldsymbol{\pi}) \ln \pi_c + \sum_{n=1}^{N}\sum_{c=1}^{C} h_{nc}(\tau, \boldsymbol{\theta}, \boldsymbol{\pi}) \ln L_{n|c}(\tau, \boldsymbol{\theta}_c). \end{aligned} \tag{19}$$

$h_{nc}(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$ refers to person $n$'s posterior probability of membership in class $c$ or $h_{nc}$ defined in equation (14), but we change the notation slightly here to emphasize that it is derived from the set of parameters being estimated. $Q(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$ may be interpreted as an expected complete data log-likelihood that views a set of indicators, $1[\boldsymbol{\theta}_n = \boldsymbol{\theta}_c]$ for $n = 1, 2, ..., N$ and $c = 1, 2, ..., C$, as missing data. Maximizing $Q(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$ is computationally easier and more numerically stable than maximizing $\ln L(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$ directly because, as we shall summarize shortly, $Q(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$ can be maximized with respect to $\{\tau, \boldsymbol{\theta}\}$ and $\boldsymbol{\pi}$ in two separate tasks. While the use of the EM algorithm to estimate a finite mixture model is common outside discrete choice modeling too, it is Bhat (1997) who introduced this estimation strategy into the discrete choice modeling literature. Train (2008; 2009) masterfully summarizes the conceptual foundations and operational aspects of the EM algorithm for discrete choice models.

Our implementation of the EM algorithm builds on Bhat (1997) and Train (2008; 2009). Specifically, let superscript $s$ denote candidate estimates obtained at the $s^{th}$ iteration of this algorithm. Then, at iteration $s+1$, the estimates are updated as follows.

$$\{\tau^{s+1}, \boldsymbol{\theta}^{s+1}\} = \arg\max_{\{\tau, \boldsymbol{\theta}\}} \sum_{n=1}^{N}\sum_{c=1}^{C} h_{nc}(\tau^s, \boldsymbol{\theta}^s, \boldsymbol{\pi}^s) \ln L_{n|c}(\tau, \boldsymbol{\theta}_c) \tag{20}$$

$$\boldsymbol{\pi}^{s+1} = \arg\max_{\boldsymbol{\pi}} \sum_{n=1}^{N}\sum_{c=1}^{C} h_{nc}(\tau^s, \boldsymbol{\theta}^s, \boldsymbol{\pi}^s) \ln \pi_c \tag{21}$$

Each person's posterior class membership probabilities are evaluated at the $s^{th}$ estimates, thereby influencing computation of the $(s+1)^{th}$ estimates only as any other type of known sampling weight would be. Both (20) and (21) therefore represent relatively simple maximization tasks. The algebraic structure of task (20) is just like that of estimating the baseline NROL model (not the LC-NROL model) using $C$ years of data, allowing for year-specific coefficients and accounting

for sampling weights: it can be readily solved by a maximum likelihood estimation program coded for the baseline NROL model. Our own program for estimating the baseline NROL model initially uses the BHHH method, a quasi-Newton method which is the default optimizer for maximum likelihood estimation in TSP International, and double-checks convergence by supplying the BHHH solution as starting values for executing the Newton method. The second updating task (21) is even easier to solve since it does not require any numerical optimization. An analytic solution to (21) can be derived and coded as

$$\pi_c^{s+1} = \frac{\sum_{n=1}^{N} h_{nc}(\tau^s, \boldsymbol{\theta}^s, \boldsymbol{\pi}^s)}{\sum_{n=1}^{N}\sum_{l=1}^{C} h_{nl}(\tau^s, \boldsymbol{\theta}^s, \boldsymbol{\pi}^s)} \text{ for } c = 1, 2, ..., C$$

using that $\pi_C \equiv 1 - \sum_{c=1}^{C-1} \pi_c$.

Once starting values are provided for initial estimates at $s = 0$, the EM algorithm proceeds by repeatedly updating the candidate estimates as above until $\Delta \ln L^{s+1} = \ln L(\tau^{s+1}, \boldsymbol{\theta}^{s+1}, \boldsymbol{\pi}^{s+1}) - \ln L(\tau^s, \boldsymbol{\theta}^s, \boldsymbol{\pi}^s)$ is small enough. Our own program uses the following set of starting values. To initialize the dissimilarity coefficient and class shares, we assume no within-nest correlation and equal class sizes i.e. $\tau^0 = 1$ and $\pi_c^0 = 1/C$ for all $c = 1, 2, ..., C$. To initialize class-specific preference parameters, $\boldsymbol{\theta}^0 = (\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0, \cdots, \boldsymbol{\theta}_C^0)$ , we randomly partition the sample into equally sized $C$ subsamples and estimate the ROL model on each subsample: then, the ROL estimates from the $c^{th}$ subsample are used as starting values for class $c$'s parameters, $\boldsymbol{\theta}_c^0$. Our program takes the $(s+1)^{th}$ estimates as the final estimates if $\Delta \ln L^{s+1}$ is smaller than 0.0001% of $\ln L(\tau^s, \boldsymbol{\theta}^s, \boldsymbol{\pi}^s)$: call these final estimates $\{\tau^{EM}, \boldsymbol{\theta}^{EM}, \boldsymbol{\pi}^{EM}\}$.

There are two drawbacks inherent in the EM algorithm as implemented here. First, as one may infer from the updating tasks (20) and (21), it does not produce valid standard errors of the final estimates, $\{\tau^{EM}, \boldsymbol{\theta}^{EM}, \boldsymbol{\pi}^{EM}\}$. Second, while $\{\tau^{EM}, \boldsymbol{\theta}^{EM}, \boldsymbol{\pi}^{EM}\}$ are equivalent to $\{\tau^{ML}, \boldsymbol{\theta}^{ML}, \boldsymbol{\pi}^{ML}\}$ in theory, they may diverge in practice since the EM algorithm may declare convergence prematurely, or even in case the model is empirically unidentified: its stopping criterion is based on $\Delta \ln L^{s+1}$ and does not execute checks on the gradient and Hessian of $\ln L(\tau^{s+1}, \boldsymbol{\theta}^{s+1}, \boldsymbol{\pi}^{s+1})$. To obtain standard errors and double-check convergence, therefore, we have followed the hybrid estimation strategy of Bhat (1997) that uses $\{\tau^{EM}, \boldsymbol{\theta}^{EM}, \boldsymbol{\pi}^{EM}\}$ as starting values for direct maximization of $\ln L(\tau, \boldsymbol{\theta}, \boldsymbol{\pi})$. The main manuscript reports the resulting estimates $\{\tau^{ML}, \boldsymbol{\theta}^{ML}, \boldsymbol{\pi}^{ML}\}$ and associated standard errors. In our experience, this direct maximization step always achieves convergence within a very small number of iterations, since the use of a stringent stopping criterion (0.0001% change in the sample log-likelihood) ensures

that $\{\tau^{EM}, \boldsymbol{\theta}^{EM}, \boldsymbol{\pi}^{EM}\}$ are almost identical to $\{\tau^{ML}, \boldsymbol{\theta}^{ML}, \boldsymbol{\pi}^{ML}\}$ in practice as well as in theory. Since starting values $\{\tau^{EM}, \boldsymbol{\theta}^{EM}, \boldsymbol{\pi}^{EM}\}$ tend to be close to the final solution $\{\tau^{ML}, \boldsymbol{\theta}^{ML}, \boldsymbol{\pi}^{ML}\}$, the use of a quasi-Newton method does not bring in practical benefits and our own program immediately uses the Newton method for this direct maximization step.

## Appendix 2: Supplementary tables

[Insert Table A1 about here]

[Insert Table A2 about here]

## References


Akaich F, Nayga RM, Gil JM (2013) Are results from non-hypothetical choice-based conjoint analyses and non-hypothetical recoded-ranking conjoint analyses similar? American Journal of Agricultural Economics 95: 946-963.

Barberá S, Pattanaik PK (1986) Falmagne and the rationalizability of stochastic choices in terms of random orderings. Econometrica 54: 707715

Bateman IJ, Mace GM, Fezzi C, Atkinson G, Turner K (2011) Economic analysis for ecosystem service assessments. Environmental and Resource Economics 48: 177-218

Beggs S, Cardell S, Hausman J (1981) Assessing the potential demand for electric cars. Journal of Econometrics 17: 1-19

Ben-Akiva M, Morikawa T, Shiroish F (1992) Analysis of the reliability of preference ranking data. Journal of Business Research 24: 149-164.

Ben-Akiva M, Lerman SR (1985) Discrete choice analysis: theory and application to travel demand. MIT Press, Cambridge

Berry S, Levinsohn J, Pakes A. (2004) Differentiated products demand system from a combination of micro and macro data. Journal of Political Economy 112: 68-105

Bhat C (1997) An endogenous segmentation mode choice model with an application to intercity travel. Transportation Science 31: 34-48

Boyle KJ, Holmes TP, Teisl MF, Roe B (2001) A comparison of conjoint analysis response formats. American Journal of Agricultural Economics 83: 441-454

Calfee J, Winston C, Stempski R (2001) Econometric issues in estimating consumer preferences from stated preference data: a case study of the value of automobile travel time. Review of Economics and Statistics 83: 699-707.

Cameron TA, Poe GL, Ethier RG, Schulze WD (2002) Alternative non-market value-elicitation methods: are the underlying preferences the same? Journal of Environmental Economics and Management 34: 391-425

Caparrós A, Campos P, Montero G (2003) An operative framework for total Hicksian income measurement: application to a multiple use forest. Environmental and Resource Economics 26: 173-198

Caparrós A, Oviedo JL, Campos P (2008) Would you choose your preferred option? Comparing choice and recoded ranking experiments. American Journal of Agricultural Economics 90: 843-855

Carson RT, Groves T (2007) Incentive and Information Properties of Preference Questions. Environmental and Resource Economics 37: 181-210.

Chang JB, Luck JL, Norwood (2009) How closely do hypothetical surveys and laboratory experiments predict field behavior? American Journal of Agricultural Economics 91: 518534

Chapman RG, Staelin R (1982) Exploiting rank ordered choice set data within the stochastic utility model. Journal of Marketing Research 19: 288-301

Claassen R, Hellerstein D, Kim SG (2013) Using mixed logit in land use models: can expectation-maximization (EM) algorithms facilitate estimation? American Journal of Agricultural Economics 95: 419-425

Clark SL, Muthén B (2009) Relating latent class analysis results to variables not included in the analysis. *mimeo*.
https://www.statmodel.com/download/relatinglca.pdf. Cited 24 Mar 2015

Dagsvik JK, Liu G (2009) A framework for analyzing rank-ordered data with application to automobile demand. Transportation Research Part A 43: 1-12

Doblas-Miranda E, Martínez-Vilalta J, Lloret F, Álvarez A, Ávila A, Bonet FJ, Brotons L, Castro J, Curiel Yuste J, Díaz M, Ferrandis P, Garca-Hurtado E, Iriondo JM, Keenan TF, Latron J, J. Llusiá, Loepfe L, Mayol M, Moré G, Moya D, Peñuelas J, Pons X, Poyatos R, Sardans J, Sus O, Vallejo VR, Vayreda J, Retana J (2014) Reassessing global change research priorities in Mediterranean terrestrial ecosystems: how far have we come and where do we go from here? Global Ecology and Biogeography 24: 25-43

Duke JM, Ilvento TW (2004) A conjoint analysis of public preferences for agricultural land preservation. Agricultural and Resource Economics Review 33: 209-219

European Commission (2014) Comission Regulation (EU) N0 702/2014 of 25 June 2014. Official Journal of The European Union Vol. 57 L193

Falmagne JC (1978) A representation theorem for finite scale systems. Journal of Mathematical Psychology 18: 5272

Fok D, Paap R, Van Dijk B (2012) A rank-ordered logit model with unobserved heterogeneity in ranking capabilities. Journal of Applied Econometrics 27: 831-846

Giergiczny M, Hess S, Dekker T, Chintakayala PK (2013) Testing the consistency (or lack thereof) between choices in best-worst surveys. Paper presented at the 3rd International Choice Modelling Conference, Sydney, 3-5 July 2013

Hausman J, Ruud P (1987) Specifying and testing econometric models for rank-ordered data. Journal of Econometrics 34: 83-104.

Heckman J, Singer B (1984) A method for minimizing the impact of distributional assumptions in econometric models for duration data. Econometrica 52: 271-320

Herriges JA, Phaneuf DJ (2002) Inducing patterns of correlation and substitution in repeated logit models of recreation demand. American Journal of Agricultural Economics 84: 1076-1090

Herriges J, Kling C, Liu C-C, Tobias J (2010) What are the consequences of consequentiality? Journal of Environmental Economics and Management 59: 6781

Hess S, Ben-Akiva M, Gopinath D, Walker J (2011) Advantages of latent class over continuous mixture of logit models. *mimeo*. http://www.stephanehess.me.uk/papers/Hess_Ben-Akiva_Gopinath_Walker_May_2011.pdf. Cited 3 Mar 2016

Hole AR (2015) MIXLOGITWTP: Stata module to estimate mixed logit models in WTP space. Statistical Software Components S458037, Boston College Department of Economics.

Hoyos D, Mariel P, Pascual U, Etxano I (2012) Valuing a Natura 2000 Network site

to inform land use options using a discrete choice experiment: an Illustration from the Basque Country. Journal of Forest Economics 18: 329-344

Huber R, Hunziker M, Lehmann B (2011) Valuation of agricultural land-use scenarios with choice experiments: a political market share approach. Journal of Environmental Planning and Management 54: 93-113

Johnson KA, Polasky S, Nelson E, Pennington D (2012) Uncertainty in ecosystem services valuation and implications for assessing land use tradeoffs: an agricultural case study in the Minnesota River Basin. Ecological Economics 79: 71-79

Keane MP, Wasi N (2013) Comparing alternative models of heterogeneity in consumer choice behavior. Journal of Applied Econometrics 28: 1018-1045

Krawczyk M (2012) Testing for hypothetical bias in willingness to support a reforestation program. Journal of Forest Economics 18: 282-289

Layton DF (2000) Random coefficient models for stated preference surveys. Journal of Environmental Economics and Management 40: 21-36

Layton DF, Brown G (2000) Heterogeneous preferences regarding global climate change. Review of Economics and Statistics 82: 616-624

Layton DF, Lee ST (2006a) From ratings to rankings: the econometric analysis of stated preference ratings data. In: Halvorsen R, Layton DF (eds) Explorations in environmental and natural resource economics: essays in honor of Gardner M. Brown, Jr. Edward Elgar Publishing, MA

Layton DF, Lee ST (2006b) Embracing model uncertainty: strategies for response pooling and model averaging. Environmental and Resource Economics 34: 51-85

Layton DF, Levine RA (2003) How much does the far future matter? A Hierarchical Bayesian analysis of the public's willingness to mitigate ecological impacts of climate change. Journal of the American Statistical Association 98: 533-544

Loomis J (2005) Economic values without prices: the importance of nonmarket values and valuation for informing public policy debates. Choices 20: 179-182

Lubowski RN, Plantinga AJ, Stavins RN (2008) What drives land-use change in the

United States? A national analysis of landowner decisions. Land Economics 84: 529-550

Louviere JJ, Flynn TN, Marley AAJ (2015) Best-Worst Scaling: Theory, Methods and Applications. Cambridge University Press.

McFadden D (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) Frontiers in econometrics. Academic Press, New York

McFadden D (1978) Modeling the choice of residential location. In: Karlqvist A, Lundqvist L, Snickars F, Weibull J (eds) Spatial interaction theory and planning models. North Holland, Amsterdam

McFadden D (1986) The choice theory approach to market research. Marketing Science 5: 275297

McFadden D, Train KE (2000) Mixed MNL models of discrete response. Journal of Applied Econometrics 15: 447-470

Mogas J, Riera P, Bennett J (2006) A comparison of contingent valuation and choice modeling with second-order interactions. Journal of Forest Economics 12:5-30

Nunes PALD, Schokkaert E (2003) Identifying the warm glow effect in contingent valuation. Journal of Environmental Economics and Management 45: 231-245

Othman J, Rahajeng A (2013) Economic valuation of Jogjakarta's tourism attributes: a contingent ranking analysis. Tourism Economics 19: 187-201

Ovando P, Campos P, Montero G (2007) Forestaciones con Encinas y Alcornoques en el Área de la Dehesa en el Marco del Reglamento (CEE) 2080/92 (1993-2000). Revista Española de Estudio Agrosociales y Pesqueros 214: 173-186

Pacifico D, Yoo HI (2013) lclogit: A Stata command for fitting latent-class conditional logit models via the expectation-maximization algorithm. Stata Journal 13: 625-639

Rashford BS, Walker JA, Bastian CT (2011) Economics of grassland conversion to cropland in the Prairie Pothole region. Conservation Biology 25: 276-284

Resano H, Sanjuan AI, Albisu LM (2012) Consumers response to the EU Quality policy allowing for heterogeneous preferences. Food Policy 37: 355365

Santos T, Tellería JL, Díaz M, Carbonell R (2006) Evaluating the benefits of CAP reforms: can afforestations restore bird diversity in Mediterranean Spain? Basic and Applied Ecology 7: 483-495

Scachar R, Nalebuff B (2004) Verifying the solution from a nonlinear solver: a case study: a comment. American Economic Review 94: 382-390

Scarpa R, Notaro S, Louviere J, Raffaelli R (2011) Exploring scale effects of best/worst rank ordered choice data to estimate benefits of tourism in Alpine Grazing Commons. American Journal of Agricultural Economics 93: 813-828

Scarpa R, Thiene M, Train KE (2008) Utility in willingness-to-pay space: a tool to address confounding random scale effects in destination choice to the Alps. American Journal of Agricultural Economics 90: 994-1010

Schulz N, Breustedt G, Latacz-Lohman U (2013) Assessing farmers' willingness to accept "greening": insights from a discrete choice experiment in Germany. Journal of Agricultural Economics 65: 26-48

Train KE (2008) EM algorithms for nonparametric estimation of mixing distributions. Journal of Choice Modelling 1: 40-69

Train KE (2009) Discrete choice methods with simulation, 2nd edition. Cambridge University Press, New York

Train KE, Weeks M (2005) Discrete choice models in preference space and willingness-to-pay space. In: Alberini A, Scarpa R (eds) Applications of simulation methods in environmental resource economics. Springer, Dordrecht.

Train KE, Winston C (2007) Vehicle choice behavior and the declining market share of U.S. automakers. International Economic Review 48: 1469-1496

Varela E, Giergiczny M, Riera P, Mahieu P-A, Soliño M (2014) Social preferences for fuel break management programs in Spain: a choice modelling application to prevention of forest fires. International Journal of Wildland Fire 23 (2): 281-289

Vossler CA, Evans MF (2009) Bridging the Gap between the Field and the Lab: Environmental Goods, Policy Maker Input, and Consequentiality. Journal of Environmental Economics and Management 58: 338-345.

Vossler CA, Doyon M, Rondeau D (2012) Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments. American Economic Journal: Microeconomics 4: 145-171.

Vossler CA, Watson SB (2013) Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field. Journal of Economic Behavior & Organization 86: 137-147

Wustemann H, Meyerhoff J, Ruhs M, Schafer A, Hartje V (2014) Financial costs and benefits of a program of measures to implement a national strategy on biological diversity in Germany. Land Use Policy 36: 307-318

Yan J, Yoo HI (2014) The seeming unreliability of rank-ordered data as a consequence of model misspecification. MPRA Paper No. 56285.
https://mpra.ub.uni-muenchen.de/56285/. Cited 3 March 2016.

Yoo HI, Doiron D (2013) The use of alternative preference elicitation methods in complex discrete choice experiments. Journal of Health Economics 32: 1166-1179

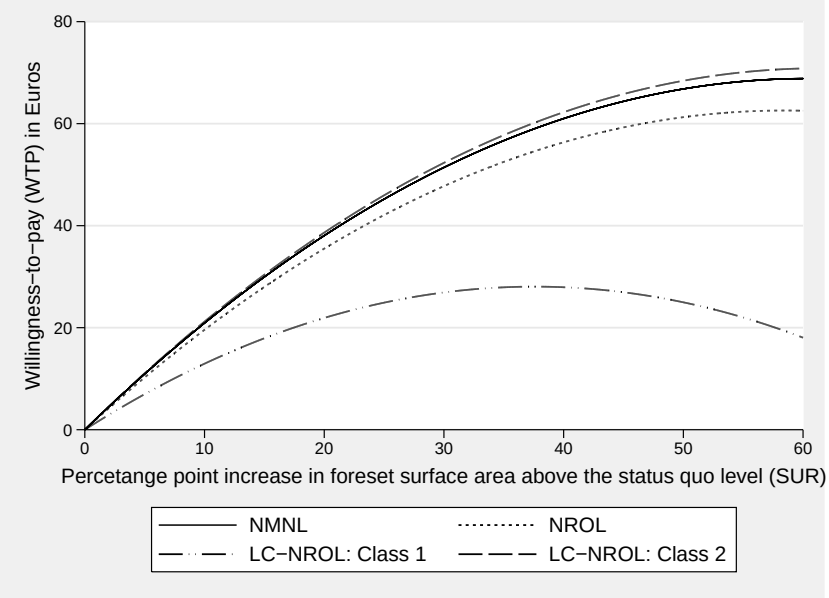


**Figure 1. Baseline results on WTP for increasing cork oak forest area**

Note: This figure is based on the results in Table 2. The vertical axis measures the willingness-to-pay (WTP) for changing SUR from 0 (the status quo level or 20% decline from today's forest area) to shown levels. The levels of 20 and 40 in SUR, for example, indicate 0% change and 20% increase from today's forest area, respectively.

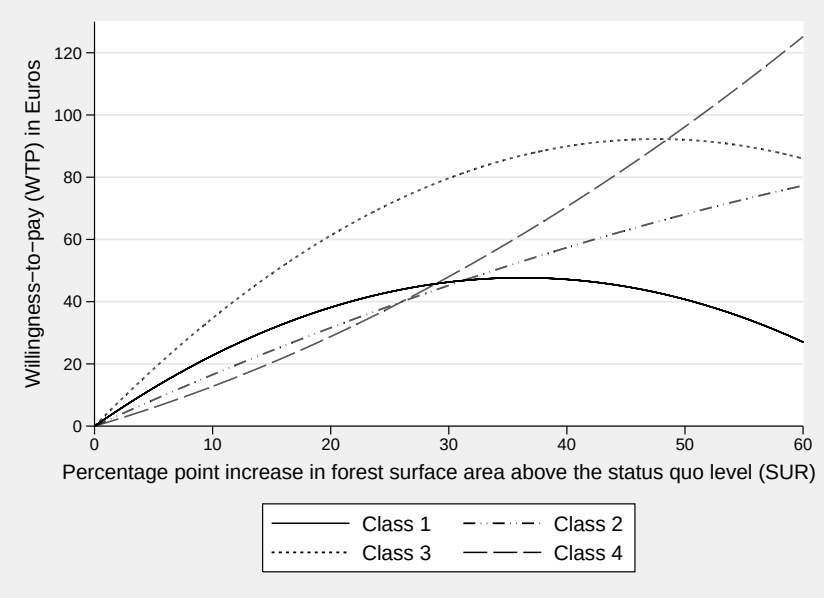


**Figure 2. Preferred results on WTP for increasing cork oak forest area**

Note: This figure is based on the results in Table 3. All other information is the same as provided in the note to Figure 1.

**Table 1. Reforestation program attributes and associated levels**

| Acronym | Definition of attribute | Levels |
|---|---|---|
| BIO | Number of native tree species used in reforestation, always including the cork oak. | =1,2,3,4. |
| NAT | Reforestation technique to be used. | =0 for artificial plantation, =1 for natural regeneration. |
| SUR | Resulting forest surface area. | =10,20,40 and 60 for 10% less, the same, 20% more, 40% more than what is available today. |
| REC | Number of new recreational areas to be created. | =0 for no new area, =2 for two new areas. |
| EMP | Number of permanent jobs created. | =20,40,60,80. |
| TAX | One-off increase in the regional government's income tax for this year. | =6,12,24,48 in €s. |

**Table 2. Baseline estimation results**

| | | | LC-NROL | | |
|---|---|---|---|---|---|
| | NMNL | NROL | Class 1 | Class 2 | Mean |
| BIO | 20.55*** | 18.06*** | 1.20 | 21.10*** | 20.13*** |
| | (2.99) | (2.55) | (2.30) | (2.63) | (2.51) |
| NAT | 32.47*** | 29.50*** | 2.71 | 33.61*** | 32.10*** |
| | (5.81) | (5.10) | (5.42) | (4.68) | (4.47) |
| SUR | 2.28*** | 2.14*** | 1.49** | 2.31*** | 2.27*** |
| | (0.51) | (0.47) | (0.76) | (0.51) | (0.49) |
| $SUR^2$ | -0.019*** | -0.018*** | -0.020* | -0.019*** | -0.019*** |
| | (0.006) | (0.006) | (0.010) | (0.007) | (0.006) |
| REC | 20.14*** | 17.60*** | 0.44 | 20.41*** | 19.43*** |
| | (4.34) | (3.89) | (5.24) | (3.84) | (3.66) |
| EMP | 0.75*** | 0.66*** | -0.12 | 0.77*** | 0.72*** |
| | (0.13) | (0.11) | (0.12) | (0.11) | (0.10) |
| TAX × -1 | 0.003 | 0.004*** | 0.035*** | 0.014*** | 0.015*** |
| | (0.003) | (0.001) | (0.007) | (0.002) | (0.002) |
| $ASC^{RF}$ | 1080.65 | 572.19*** | -45.30** | 247.66*** | 233.32*** |
| | (1682.11) | (203.36) | (18.22) | (44.50) | (42.32) |
| $\tau$ | 0.16 | 0.26*** | 0.89*** | | |
| | (0.22) | (0.08) | (0.09) | | |
| share | | | 0.05*** | 0.95*** | |
| | | | (0.01) | (0.01) | |
| logL | -2632.16 | -3034.04 | -2532.84 | | |
| BIC | 5319.30 | 6123.07 | 5175.64 | | |
| # param | 9 | 9 | 18 | | |

Note: Standard errors are in parentheses. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels respectively. The estimates BIO through EMP measure willingness-to-pay (WTP) for relevant attributes in €s. TAX × -1 is the scale parameter ($\lambda_n$), $ASC^{RF}$ is the alternative-specific constant ($\kappa_n$) for reforestation programs, and $\tau$ is the dissimilarity coefficient: see equation (11). Share is the population share of each class ($\pi_c$). For all parameters except class shares, the null hypothesis is zero. For class shares, the null hypothesis is 1/2. In LC-NROL, column Mean is derived as the weighted average of class-specific estimates, wherein class shares are used as weights.

**Table 3. Preferred estimation results (LC-NROL)**

| | Class 1 | Class 2 | Class 3 | Class 4 | Mean |
|---|---|---|---|---|---|
| BIO | 3.03 | 26.27** | 19.20** | 0.90 | 21.78*** |
| | (3.11) | (11.91) | (9.67) | (2.27) | (3.60) |
| NAT | 0.57 | -17.55 | 83.77** | 11.24* | 25.81 |
| | (7.08) | (11.86) | (35.91) | (6.30) | (18.07) |
| SUR | 2.64** | 1.73 | 3.88** | -1.11 | 2.56*** |
| | (1.09) | (1.18) | (1.67) | (0.77) | (0.57) |
| $SUR^2$ | -0.036** | -0.007 | -0.041** | 0.016 | -0.021*** |
| | (0.015) | (0.013) | (0.018) | (0.011) | (0.008) |
| REC | 1.89 | 45.45** | 4.89 | -3.73 | 25.69*** |
| | (6.90) | (20.94) | (7.82) | (5.96) | (9.93 ) |
| EMP | -0.10 | 1.38** | 0.29 | 0.001 | 0.84*** |
| | (0.16) | (0.63) | (0.24) | (0.126) | (0.27) |
| TAX × -1 | 0.038*** | 0.014** | 0.019** | 0.047*** | 0.018*** |
| | (0.010) | (0.005 ) | (0.008) | (0.009) | (0.002) |
| $ASC^{RF}$ | -116.94*** | 322.39** | 187.29** | 48.25*** | 242.88*** |
| | (39.67) | (144.55) | (83.90) | (15.30) | (55.99) |
| $\tau$ | 0.99*** | | | | |
| | (0.06) | | | | |
| share | 0.04*** | 0.52*** | 0.41*** | 0.03*** | |
| | (0.01) | (0.07) | (0.07) | (0.01) | |
| logL | -2413.81 | | | | |
| BIC | 5047.54 | | | | |
| # param | 36 | | | | |

Note: For class shares, the null hypothesis is 1/4. All other information is the same as provided in the note to Table 2.

**Table 4. Individual-level WTP coefficients: descriptive statistics**

| | $Q_1$ | $median$ | $Q_3$ | $mean$ | $SD$ | $Q_3 - Q_1$ | $\frac{SD}{mean}$ | $\frac{Q_3-Q_1}{mean}$ |
|---|---|---|---|---|---|---|---|---|
| BIO | 20.32 | 23.11 | 25.38 | 21.78 | 5.62 | 5.06 | 0.26 | 0.23 |
| NAT | -5.03 | 19.18 | 54.02 | 25.81 | 33.81 | 59.05 | 1.31 | 2.29 |
| REC | 11.69 | 27.30 | 40.33 | 25.69 | 15.03 | 28.64 | 0.59 | 1.11 |
| EMP | 0.47 | 0.89 | 1.25 | 0.84 | 0.43 | 0.78 | 0.51 | 0.93 |
| SUR = 10 | 18.45 | 22.72 | 29.37 | 23.49 | 8.14 | 10.93 | 0.35 | 0.47 |
| SUR = 20 | 34.73 | 41.17 | 52.56 | 42.69 | 13.85 | 17.83 | 0.32 | 0.42 |
| SUR = 40 | 59.76 | 67.86 | 80.36 | 68.21 | 18.60 | 20.61 | 0.27 | 0.30 |
| SUR = 60 | 77.94 | 80.01 | 83.32 | 76.55 | 17.64 | 5.38 | 0.23 | 0.07 |

Note: This table summarizes the distribution of 450 sets of individual-specific willingness-to-pay (WTP) coefficients. The individual-specific WTP coefficients have been computed for each of 450 respondents in our sample, by using the estimates in Table 3 to evaluate equation (15). $Q_1$ and $Q_3$ refer to the first and third quartiles respectively. $SD$ is the standard deviation. For SUR, the reported figures refer to the individual-specific WTP for changing that attribute from 0 to shown levels.

**Table 5. Actual and weighted sample mean characteristics**

| | Obs | Actual | Class 1 | Class 2 | Class 3 | Class 4 |
|---|---|---|---|---|---|---|
| *A. Use and knowledge of ANP* | | | | | | |
| times visited in a year[a] | 431 | 1.60 | 1.88 | 1.65 | 1.54 | 1.60 |
| hours spent visiting[a] | 450 | 27.82 | 33.51 | 28.47 | 26.92 | 28.59 |
| money spent visiting[a] | 450 | 19.73 | 17.48 | 20.33 | 19.93 | 11.63 |
| don't find it congested | 450 | 0.81 | 0.67 | 0.82 | 0.80 | 0.82 |
| satisfied with it | 444 | 0.46 | 0.60 | 0.47 | 0.44 | 0.44 |
| active tourist | 450 | 0.31 | 0.49 | 0.35 | 0.26 | 0.44 |
| *B. Socioeconomic characteristics* | | | | | | |
| age[a] | 446 | 34.44 | 35.39 | 33.43 | 34.75 | 41.39 |
| family income[a] | 427 | 1615.22 | 1677.70 | 1642.07 | 1601.42 | 1460.11 |
| family size[a] | 448 | 3.61 | 4.17 | 3.49 | 3.68 | 3.51 |
| per capita family income[a] | 425 | 525.24 | 422.93 | 555.41 | 511.92 | 466.52 |
| employed | 450 | 0.77 | 0.60 | 0.77 | 0.78 | 0.75 |
| postgraduate | 450 | 0.39 | 0.62 | 0.42 | 0.35 | 0.56 |

Note: ANP stands for Alcornocales Natural Park. The weighted mean for Class $c$ has been computed by using each person's posterior probability of membership in class $c$ to weight that person's characteristics: see equation (17). Each person's posterior probabilities have been computed by using the estimates in Table 3 to evaluate equation (14). Superscript a identifies non-dichotomous variables; all others are zero-one variables.

**Table A1. Comparison of LC-NROL with mixed ROL estimation results**

| | LC-NROL | Independent Mixed ROL | | Correlated Mixed ROL | |
|---|---|---|---|---|---|
| | Mean | Mean | Std Dev | Mean | Std Dev |
| BIO | 21.78*** | 19.83*** | 13.91*** | 19.25*** | 11.30*** |
| | (3.60) | (2.31) | (2.46) | (1.89) | (1.44) |
| NAT | 25.81 | 32.83*** | 59.48*** | 38.10*** | 53.75*** |
| | (18.07) | (5.07) | (7.10) | (4.98) | (5.20) |
| SUR | 2.56*** | 2.19*** | 0.91*** | 2.33*** | 3.33*** |
| | (0.57) | (0.44) | (0.20) | (0.47) | (0.38) |
| $SUR^2$ | -0.021*** | -0.018*** | 0.004 | -0.019*** | 0.044*** |
| | (0.008) | (0.006) | (0.004) | (0.006) | (0.005) |
| REC | 25.69*** | 21.57*** | 44.36*** | 22.32*** | 36.11*** |
| | (9.93) | (4.01) | (5.95) | (3.83) | (3.65) |
| EMP | 0.84*** | 0.70*** | 0.96*** | 0.78*** | 0.78*** |
| | (0.27) | (0.10) | (0.14) | (0.09) | (0.08) |
| TAX × -1 | 0.018*** | -3.78*** | 0.01 | -3.58*** | 0.89*** |
| | (0.002) | (0.09) | (0.20) | (0.09) | (0.09) |
| $ASC^{RF}$ | 242.88*** | 607.03*** | 373.52*** | 432.29*** | 318.17*** |
| | (55.99) | (105.65) | (59.61) | (41.10) | (32.63) |
| logL | -2413.8 | -2397.1 | | -2358.2 | |
| BIC | 5047.54 | 4891.88 | | 5062.94 | |
| # param | 36 | 16 | | 44 | |

Note: In LC-NROL, column Mean is taken from our preferred 4-class model in Table 3. In Independent/Correlated Mixed ROL, column Mean (Std Dev) reports the population mean (standard deviation) of a normally distributed random coefficient; in case of the log-normal coefficient $\lambda_n$ on TAX × -1, Mean (Std Dev) reports the mean (standard deviation) of $\ln \lambda_n$. Correlated Mixed ROL also includes 28 unreported parameters; the results are available upon request. All other information is the same as provided in the note to Table 2.

**Table A2. Individual characteristics and definitions**

| Characteristic | Definition |
|---|---|
| *A. Use and knowledge of ANP* | |
| times visited in a year | number of times the respondent has visited the Alcornocales Natural Park (ANP) in the past 12 months. |
| hours spent visiting | hours that the respondent has spent in total in his/her visits to the ANP. |
| money spent visiting | money spent by the respondent during his/her current visit to the ANP (e.g. gas, food and parking). |
| don't find it congested | =1 if thinks that there was a small or appropriate number of visitors in the ANP area he/she visited on the interview date. |
| satisfied with it | =1 if satisfaction from his/her current visit to ANP has surpassed his/her initial expectations. |
| active tourist | =1 if the main reason for the current visit to the ANP is any kind of active tourism (e.g. biking, hiking, climbing). |
| *B. Socioeconomic characteristics* | |
| age | age in years |
| family income | monthly family income in €s. |
| family size | number of members of the respondent's family. |
| per capita family income | monthly family income in €s divided by family size. |
| employed | =1 if has a permanent job or a permanent source of income. |
| postgraduate | =1 if has a university/college degree. |