Science of the Total Environment 633 (2018) 967–980

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

A multiple indicator solution approach to endogeneity in discrete-choice models for environmental valuation

Petr Mariel [a,b,*], David Hoyos [a,b,c], Alaitz Artabe [b,d], C. Angelo Guevara [e]

[a] Department of Applied Economics III (Econometrics and Statistics), University of the Basque Country (UPV/EHU), Spain
[b] Avda. Lehendakari Aguirre, 83. E48015 Bilbao, Spain
[c] EKOPOL, Research Group on Ecological Economics and Political Ecology, Spain
[d] Department of Applied Economics II (Public Finances and Tax Law), University of the Basque Country (UPV/EHU), Spain
[e] Departamento de Ingeniería Civil, Universidad de Chile, Casilla 228-3, Santiago, Chile

HIGHLIGHTS

- The paper applies the multiple indicator solution method in a discrete choice model for environmental valuation.
- It compares two different approaches to deal with endogeneity arising from omitted explanatory variables.
- The multiple indicator solution method and the hybrid model approach provide similar results in terms of welfare estimates.
- The multiple indicator solution method is more parsimonious and notably easier to implement but less efficient and flexible.

GRAPHICAL ABSTRACT

Environmental Valuation Studies

How shall we deal with endogeneity arising from omitted explanatory variables in discrete choice models for environmental valuation?

Hybrid Choice Model	Multiple Indicator Solution
Similar parameter estimates	
(-) Higher estimation cost	(+) Easier estimation
(-) Specific code or econometric software package needed	(+) Easily applicable in a general econometric software
(+) Higher efficiency of the estimation method	(-) Lower efficiency of the estimation method
(+) More flexible method. Latent variable can be incorporated to any part of the model.	(-) Attitudinal latent variable can be incorporated only at individual level (allocation function)

ARTICLE INFO

Article history:
Received 13 December 2017
Received in revised form 21 March 2018
Accepted 21 March 2018
Available online 28 March 2018

Editor: F Frederic Coulon

Keywords:
Discrete choice
Valuation
Multiple indicator solution
Choice modelling
Endogeneity
Stated preferences

ABSTRACT

Endogeneity is an often neglected issue in empirical applications of discrete choice modelling despite its severe consequences in terms of inconsistent parameter estimation and biased welfare measures. This article analyses the performance of the multiple indicator solution method to deal with endogeneity arising from omitted explanatory variables in discrete choice models for environmental valuation. We also propose and illustrate a factor analysis procedure for the selection of the indicators in practice. Additionally, the performance of this method is compared with the recently proposed hybrid choice modelling framework. In an empirical application we find that the multiple indicator solution method and the hybrid model approach provide similar results in terms of welfare estimates, although the multiple indicator solution method is more parsimonious and notably easier to implement. The empirical results open a path to explore the performance of this method when endogeneity is thought to have a different cause or under a different set of indicators.

1. Introduction

Discrete choice experiments (DCEs) are increasingly being used to elicit preferences for environmental and natural resources. Many

* Corresponding author at: Department of Applied Economics III (Econometrics and Statistics), University of the Basque Country (UPV/EHU), Spain.
E-mail addresses: petr.mariel@ehu.eus (P. Mariel), david.hoyos@ehu.eus (D. Hoyos), alaitz.artabe@ehu.eus (A. Artabe), crguevar@ing.uchile.cl (C.A. Guevara).

https://doi.org/10.1016/j.scitotenv.2018.03.254

methodological issues have been addressed in the development of DCEs in the field of environmental valuation, including optimal experimental design, econometric models, attribute non-attendance, and so forth. However, the issue of endogeneity and how to deal with it has received little attention (Hoyos, 2010). This apparent non-interest is especially surprising when most authors agree that endogeneity has the potential to distort inferences about preferences and the policy recommendations that could be derived from the analysis of the choice data (Louviere et al., 2005).

Endogeneity refers to the existence of correlation between the deterministic part and the error term of a regression or choice model. In the presence of endogeneity, parameter estimates are inconsistent thus invalidating any inference obtained from the model. Although this issue may be unavoidable in many practical cases, the severe consequences that the existence of endogeneity may cause, requires special attention to investigate how to deal with it in the field of DCEs for environmental valuation. The main source of endogeneity in environmental valuation may be found in the omission of contextual conditions in the choice situations, either due to the impossibility of the researcher to measure omitted attributes in the choice tasks, or to extract the exact measure that was inferred by the respondents when making their choices. For instance, we may find endogeneity when modelling environmental decisions due to the omission of latent environmental attitudes. When selecting an alternative, the decision maker may take into consideration his/her pro-environmental beliefs or attitudes. As a result, if endogeneity is not properly addressed, the estimated coefficients will likely be biased.

Dealing with endogeneity in classical regression models is well established in the econometric literature (see e.g. Bun and Harrison, 2014; Wooldridge, 2010). However, many aspects of how to deal with this problem in the framework of non-linear models, like discrete choice models, are still under development and have received scarce attention in areas such as transportation and environmental economics. The control function (CF) method has been found to properly address endogeneity in discrete choice models when this problem arises at the alternative level (Guevara and Ben-Akiva, 2006; Guevara, 2015; Petrin and Train, 2010). As with the case of linear models, one critical step when applying the CF method in discrete choice models is finding a valid instrument for each endogenous variable. A proper instrument needs to be correlated with the endogenous variable but, at the same time, uncorrelated with the error term. In practice, finding a valid instrument is sometimes problematic, thus motivating the search for alternative methods such as the multiple indicator solution (MIS).

The MIS procedure was originally proposed in the late 1960s by Blalock Jr. and Costner (1969), and Costner (1969), for sociological models. More recently, Wooldridge (2010) formalised the method for linear models, and Guevara and Polanco (2016) adapted it for DCEs. The MIS method requires a minimum of two indicators and is applied in two steps. In the first step, one of the indicators is added to the structural equation of the model as an additional explanatory endogenous variable. In the second step, the endogeneity is dealt with using the latter indicator as an instrumental variable for the former one. While Wooldridge (2010) uses the two-stage least squares (2SLS) method to address endogeneity in linear models, the MIS method has been extended to discrete choice models by using the CF method in the second stage (Guevara and Polanco, 2016).

In this paper, we present an exploratory analysis of the performance of the MIS method to deal with endogeneity arising from omitted attributes in a DCE for environmental valuation. To our knowledge, this is the first application of the MIS method to correct for endogeneity in the context of discrete-choice models for environmental valuation. The performance of this method is compared with another treatment of endogeneity recently used in environmental valuation, namely hybrid choice models.

Hybrid choice models allow for the incorporation of latent behavioural constructs within the traditional choice models, and therefore, they can be seen as a solution for endogeneity caused by the omission of a relevant variable. Hybrid choice models were first proposed by McFadden (1986) and Train et al. (1987) and have been increasingly used in the last decade. Despite some criticism (Chorus and Kroesen, 2014), their applications can be found in transportation (Paulssen et al., 2014; Bhat et al., 2015; Thorhauge et al., 2015), environment (Hess and Beharry-Borg, 2012; Hoyos et al., 2015) or health economics (Kløjgaard and Hess, 2014). We find that the MIS method and the hybrid model approach provide similar results in terms of model fit and parameter interpretation, but MIS is more parsimonious and notably easier to implement.

The rest of the paper is structured as follows: Section 2 addresses the methodological issues behind the MIS method to deal with endogeneity in discrete choice models; Section 3 describes the empirical study in which this methodology will be tested along with previous results; Section 4 provides the main results of the paper; and Section 5 finishes by discussing the main findings of this investigation and suggesting future lines of research in the area.

2. Methodology

We depart from a classical structural equation for a choice model given by the random utility theory, which is used to link the deterministic model with a statistical model of human behaviour. Under this framework, the utility of alternative i for respondent n is given by[1]:

$$\begin{aligned} u_{in}^* &= x_{in}^{'}\beta + \beta_q q_{in}^* + e_{in}^* = x_{in}^{'}\beta + \varepsilon_{in}^* = v_{in} + \varepsilon_{in}^*, \\ y_{in} &= 1\left[u_{in}^* \geq u_{jn}^*;\ \forall j \in c_n\right], \end{aligned} \tag{1}$$

where, utility u_{in}^* is a latent variable that cannot be observed by the researcher but, instead, an indicator y_{in} is observed, which takes the value one if the utility of alternative i is the largest among those in the choice set c_n, and takes the value zero otherwise. The latent utility that an individual obtains from an alternative depends linearly, with coefficients β, on a set of explanatory variables or attributes collected in a row vector x_{in}, on a latent (omitted) variable q_{in}^* and on an error term e_{in}^*. In the set of explanatory variables x_{in} we may find a set of attributes of alternative i for respondent n, the first element of this vector being a one (for all but one alternative), accounting for an alternative specific constant. In the case where the omitted variable q_{in}^* is correlated with any variable included in x_{in}, there is an endogeneity problem in the model because ε_{in}^* is correlated with v_{in} in this case.

The CF approach is a common method used in linear regression analysis that can address endogeneity problems at the level of each alternative in discrete choice models. However, as with the case of linear regression, a critical requirement for applying this method is finding valid instruments for the endogenous variables. An instrument for an endogenous variable is valid if it is, at the same time, correlated with the endogenous variable and independent (not only uncorrelated as in the 2SLS method for linear models) of the error term of the model (Guevara and Polanco, 2016). The MIS method provides valid instruments that can then be used in the CF approach.

Let us assume that, instead of the latent variable q_{in}^*, the researcher observes two indicators q_1 and q_2 generated by the following equations:

$$\begin{aligned} q_{1in} &= \alpha_{10} + \alpha_{11} q_{in}^* + e_{q1in}^*, \\ q_{2in} &= \alpha_{20} + \alpha_{21} q_{in}^* + e_{q2in}^*, \end{aligned} \tag{2}$$

[1] For clarification purposes, we denote with an asterisk those variables that are latent in order to distinguish them from those that are observed by the researcher. Model coefficients are denoted with Greek letters.

where,

$$\alpha_{11} \neq 0, \alpha_{21} \neq 0, \tag{3}$$

and the following pairs of variables are independent:

$$\left(q^*_{in}, e^*_{q1in}\right); \left(x_{in}, e^*_{q1in}\right); \left(q^*_{in}, e^*_{q2in}\right); \left(x_{in}, e^*_{q2in}\right); \left(e^*_{q1in}, e^*_{q2in}\right). \tag{4}$$

Then, if we replace $q^*_{in} = 1/\alpha_{11}(q_{1in} - \alpha_{10} - e^*_{q1in})$ from Eq. (2) in Eq. (1), the new error term of the model would be ξ^*_{in}, as shown in the following equation:

$$\begin{aligned} u^*_{in} &= x'_{in}\beta + \frac{\beta_q}{\alpha_{11}}\left(q_{1in} - \alpha_{10} - e^*_{q1in}\right) + e^*_{in} = \\ &= x'_{in}\beta + \gamma_{q1} q_{1in} + \left(-\frac{\beta_q \alpha_{10}}{\alpha_{11}} - \frac{\beta_q e^*_{q1in}}{\alpha_{11}}\right) + e^*_{in} = \\ &= x'_{in}\beta + \gamma_{q1} q_{1in} + \xi^*_{in}. \end{aligned} \tag{5}$$

This shows that including one of the indicators q_{1in} in the utility function does not solve the endogeneity problem, but changes the source of it, as the only endogenous variable in this new model is precisely q_{1in}, which is correlated with e^*_{q1in}, because of Eq. (2). However, as shown also in Eq. (2), q_{2in} is a proper instrument for q_{1in}, allowing the correction of endogeneity. Indeed, firstly, q_{2in} is correlated with q_{1in} as both q_{1in} and q_{2in} depend on q^*_{in}; and secondly, q_{2in} is independent of ξ^*_{in} because e^*_{q1in} and e^*_{q2in} are independent error terms according to assumption (4).

The MIS method for discrete choice models proposed by Guevara and Polanco (2016) provides consistent estimation of the parameters of interest, β in Eq. (1), following a two-stage procedure. In the first stage, the residuals $\hat{\eta}_{1in}$ from the linear regression of the first indicator q_{1in} on the vector of explanatory variables x_{in} and a second indicator q_{2in}, are obtained:

$$q_{1in} = x'_{in}\hat{\theta} + \hat{\theta}_{q2} q_{2in} + \hat{\eta}_{1in}. \tag{6}$$

In the second stage, the CF approach is used in order to estimate the choice model including the set of explanatory variables x_{in}, the first indicator q_{1in} and the residuals $\hat{\eta}_{1in}$ of the linear regression estimated in the first stage (6) in the utility function.

$$u^*_{in} = x'_{in}\beta + \gamma_{q1} q_{1in} + \gamma_{\hat{\eta}}\, \hat{\eta}_{1in} + \zeta^*_{in}. \tag{7}$$

It is important to bear in mind that the standard errors of the two-stage procedure cannot be directly obtained from the information matrix. Instead they can be calculated by bootstrapping or using, for example, the delta method proposed by Karaca-Mandic and Train (2003).

3. Data

The case study refers to a DCE for an environmental valuation conducted in 2008 in the province of Gipuzkoa, Spain (see Fig. 1). The valuation study aimed to analyse the social preferences for different land-use options in a special protection area known as Garate-Santa Barbara (GSB). Detailed information about the survey design can be found in Hoyos et al. (2012).[2]

The survey asked respondents to choose from different land-use options characterised by a set of five environmental attributes (see Table 1): percentage of land area covered by native tree species (NAT), vineyards (VIN) and exotic tree plantations (FOR), biodiversity protection – number of endangered species of flora and fauna (BIO), the level of conservation of recreational and cultural facilities (REC); and a payment attribute, namely the cost of the conservation program (COST). Table 1 presents assumed levels of each attribute.

A main effects fractional factorial design with second order interactions was used to simplify the construction of choice sets (Louviere et al., 2000). The final version of the questionnaire included 120 choice sets (blocked into 20 groups of 6 choice sets); each formed by the status quo option plus two alternative protection programmes for GSB (programme A and programme B). For a better understanding of the trade-offs between the attributes and alternatives, the choice sets included maps and percentage values (see Fig. 2). The proposed payment vehicle was an annual contribution by all Basque citizens to a foundation exclusively dedicated to protecting the site. The complexity of the choice task was satisfactorily pre-tested in focus groups and through pilot surveys.

In addition to the choice data and socioeconomic information, environmental attitudinal information was also collected in the survey. Respondents were asked a series of attitudinal questions following the typical awareness of consequences (AC) psychometric scale. This scale has been used extensively in environmental psychology as a measure for the value-belief-norm (VBN) theory (Stern et al., 1993, 1995). This popular behavioural theory proposes that egoistic (EGO), altruistic (ALT) and biospheric (BIO) value orientations influence the way in which individuals formulate and structure environmental beliefs (Stern, 2000). However, empirical research has uncovered some limitations in measuring AC beliefs, mainly poor dimensionality and theoretically inconsistent subscale correlations (Hansla et al., 2008; Snelgar, 2006). So, following Ryan and Spash (2012), the scale is reinterpreted as a measure of beliefs supporting environmental action and inaction (BSEAI scale). Beliefs supporting environmental action can also be divided into beliefs in the positive consequences of environmental protection and in the seriousness of environmental damages (see Table 2).

Table 2 presents four groups of items classified according to the BSEAI scale. Groups 1A and 1B represent beliefs supporting environmental action and groups 2A and 2B represent environmental inaction. The first column shows the label of the original AC classification. Specifically, items prefixed by EGO, ALT, and BIO aim to capture egoistic, altruistic, and biospheric value orientations, respectively.

Finally, the survey was administered through in-person computer-aided individual home interviews. The population considered relevant was that of the Basque Autonomous Community, comprised of 1.8 million people aged at least 18. A stratified random sample of 400 individuals was selected from this population. The strata used included age, gender and size of the town of residence, following official statistical information by the Basque Statistics Office (EUSTAT). The questionnaire was distributed using random survey routes in each of the locations in the Basque Country. The data analysis involved 221 completed questionnaires, yielding 1326 observations, as each respondent was given six choice sets (an example of a choice card can be found in Fig. 2).

4. Results

4.1. Descriptive statistics

Table 3 provides a complete description of the full set of variables used in the econometric modelling subsection along with their descriptive statistics. The mean age, gender structure and disposable income of respondents are in line with the average age, gender structure and income composition of the population (40.15 years, 45% and 1029 €, respectively). Apart from the six attributes, native forest, vineyards, forest plantations, biodiversity, recreation and cost, other explanatory variables were also considered. These variables were: *recreationalist* (taking the value 1 if the respondent was a recreationalist and 0 otherwise); *gender* (taking the value 1 if respondent was a male and 0 otherwise); *number of adults* (the number of adults in the family); *number of children* (the number of children in the family); *education* (for respondent's level of education with 1 being the lowest and 5 the highest); and *environmental NGO* (taking the value 1 if respondent was a member of an environmentalist organisation and 0 otherwise).

[2] A copy of the survey instrument is available from the authors upon request.

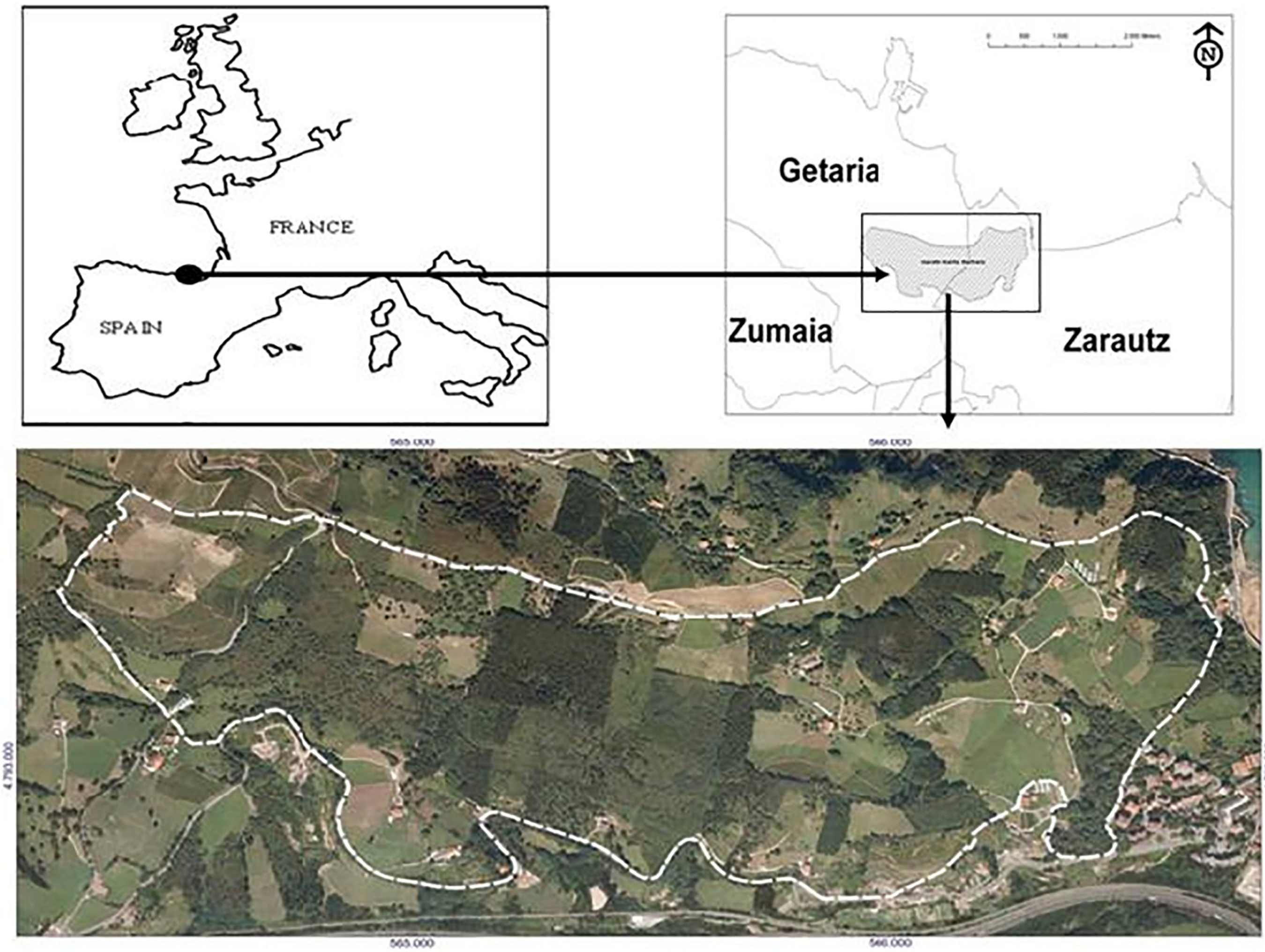

Fig. 1. Location of Garate-Santa Barbara N2000 site (Basque Country, Southern Europe).

Table 4 shows the response distributions in a 5-point Likert scale for the answers to the BSEAI scale items. For each statement, values closer to five would equate to strong agreement while values closer to one would equate to strong disagreement. As shown in this table, respondents are generally aware of the increasing environmental degradation of the Earth and are worried about the environment that future generations will have. For example, 91% of the respondents agreed with item 1 (EGO 1 - environmental protection will provide a better world for me and my children) and 87% disagreed with item 6 (ALT3 - we do not need to worry much about the environment because future generations will be better able to deal with these problems than us).

Table 1
Attributes and levels considered.

Attribute	Level					
Native forest (NAT)	2%*	10%	20%	30%		
Vineyard (VIN)	40%*	30%	20%	10%		
Exotic tree plantations (FOR)	40%*	30%	25%	15%		
Biodiversity (BIO)	25*	15	10	5		
Recreation (REC)	Low*	Medium	High	Very High		
Cost of programme (COST)	0€*	5€	10€	30€	50€	100€

(*) Levels with asterisk represent the status quo scenario.

4.2. Model specification

The structural equation for the choice model has a representative utility, v_{int} from Eq. (1), specified in our case as a function of the attributes:

$$v_{int} = ASC_i + \beta_1 NAT_{int} + \beta_2 VIN_{int} + \beta_3 FOR_{int} + \beta_4 BIO_{int} + \beta_5 REC_{int} + \beta_6 COST_{int} = x'_{int}\beta, \tag{8}$$

where: NAT, VIN, FOR, BIO, REC and $COST$ are the choice attributes native forest, vineyards, forest tree plantations, biodiversity, recreation and cost, respectively. Subscript t is added to denote a sequence of choice task typically included in a DCE with a panel data structure. For example, the variable NAT_{int} represents the value of the native forest attribute corresponding to the level for alternative i in a given choice situation t for respondent n. The remaining attributes are coded according to the levels described in Table 2. Next, we present five different discrete choice models, starting with the simplest multinomial logit (MNL) model specification (model 1) and then moving to more flexible specifications allowing for preference heterogeneity and the incorporation of attitudinal information under a latent class model (LCM) framework (models 2 to 5).

LCMs are very commonly used econometric approaches to model unobserved heterogeneity of preferences. Therefore, the following models (models 2, 3, 4 and 5) are latent class models aimed at capturing

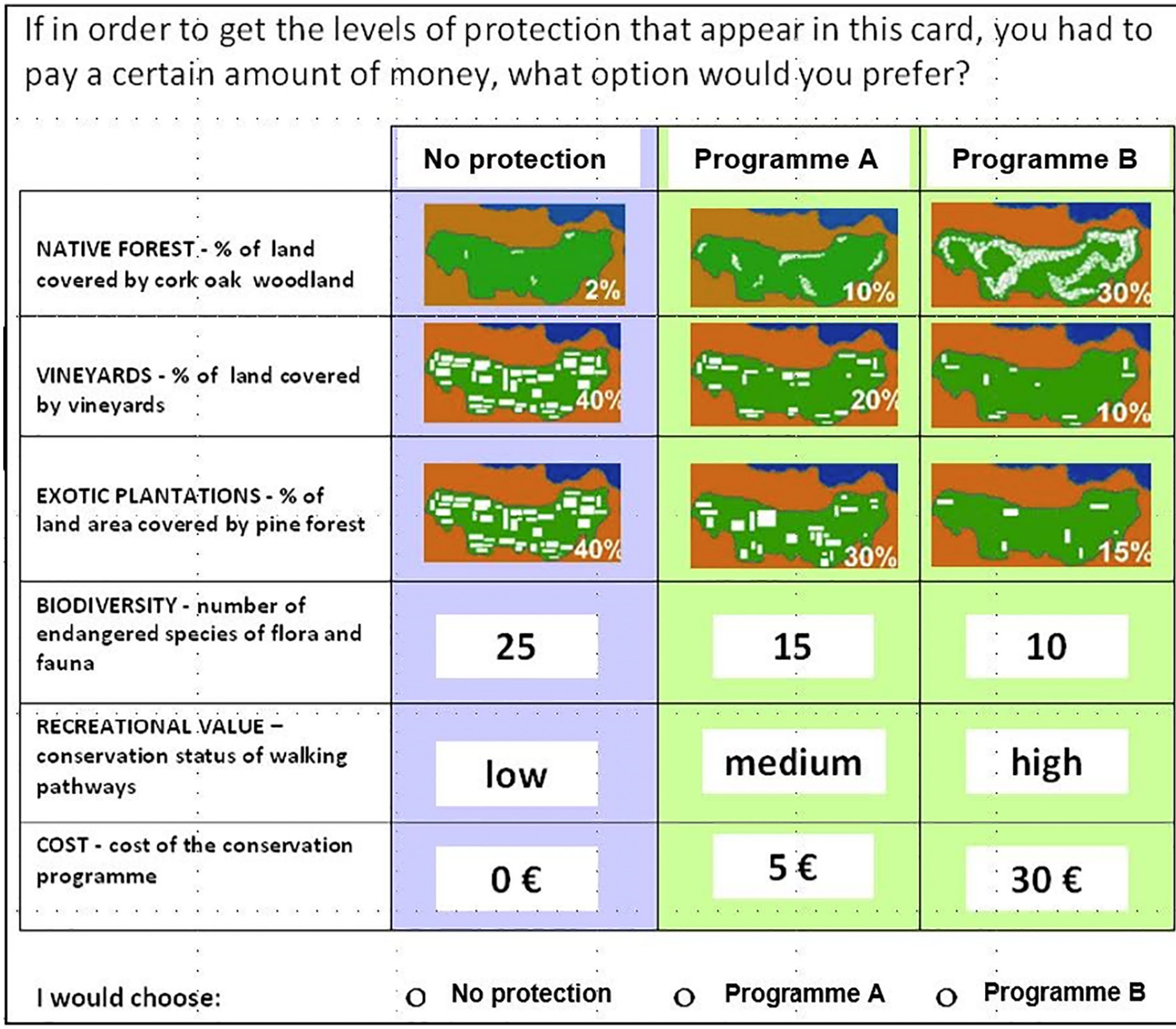

Fig. 2. Example of a choice set with different protection alternatives used in the valuation exercise, translated into English.

respondent heterogeneity in a two-class framework. The determination of the number of classes was achieved using goodness of fit and consistency within behavioural models (Hoyos et al., 2015). As a classical LCM, model 2 attempts to disentangle the preference heterogeneity by the inclusion of socio-demographic characteristics in the class allocation function. In addition to modelling preference heterogeneity, the remaining models (models 3, 4 and 5) incorporate into the allocation function separate socio-demographic variables and also attitudinal information. Model 3 does it under a hybrid latent class modelling framework, while model 4 includes attitudinal indicators directly in the class allocation function without specific treatment which, by definition, creates endogeneity. Finally, Model 5 deals with endogeneity of the indicator with the MIS method.

4.2.1. Model 1: Multinomial logit model

The first model specified is a simple MNL model with generic coefficients for all attributes, with a linear representative utility function as specified in Eq. (8). As is usually the case, this model is used as a benchmark model for the more flexible models that follow.

4.2.2. Model 2: Latent class model (LCM)

Following a latent class model specification (Greene and Hensher, 2003), in model 2 we assume that individuals can be sorted into a set of C classes (2 classes in this case), each of which is characterised by unique class-specific utility parameters β_C. Given membership of class c_s, the probability of respondent n's sequence of choices is given by:

$$P_n = \Pr(y_n^t | c_s, x_n) = \prod_{t=1}^{T_n} \frac{\exp\left(ASC_i + \beta'_{c_s} x_{int}\right)}{\sum_{j=1}^{J} \exp\left(ASC_i + \beta'_{c_s} x_{jnt}\right)}, \tag{9}$$

where y_n^t is the sequence of choices over the T_n choice occasions for respondent n and ASC_i is an alternative specific constant for alternative i normalised to zero for one of the J alternatives. Eq. (9) is a product of MNL probabilities. The LCM framework recognises that actual membership of a class is not observed, it is latent. If the probability of membership of a latent class c_s of respondent n is defined as $\pi_{n,\,c_s}$, the unconditional probability of a sequence of choices can be derived by taking the expectation over all C classes, that is:

$$P_n = \Pr(y_n^t | x_n) = \sum_{s=1}^{C} \pi_{n,c_s} \prod_{t=1}^{T_n} \frac{\exp\left(ASC_i + \beta'_{c_s} x_{int}\right)}{\sum_{j=1}^{J} \exp\left(ASC_i + \beta'_{c_s} x_{jnt}\right)}. \tag{10}$$

The class allocation probabilities $\pi_{n,\,c_s}$ are usually modelled using a logit structure, where the utility of a class is a function of the socio-demographics of the respondent SD_n and parameters λ_s, in addition to a constant, μ_{0s}, for class c_s. Let us consider $\pi_{n,\,c_s}$ the allocation probability for class c_s for individual n. Class probabilities are specified assuming the

Table 2
Beliefs supportive of environmental action and inaction (BSEAI scale).

Item	Description
	GROUP 1: Beliefs supporting environmental action
	Group 1A Beliefs that environmental protection has positive consequences
EGO1	Environmental protection will provide a better world for me and my children
EGO2	Environmental protection is beneficial to my health
EGO5	A clean environment provides me with better opportunities for recreation
ALT1	Environmental protection benefits everyone
ALT2	Environmental protection will help people have a better quality of life
BIO4	Tropical rain forests are essential for maintaining a healthy planet Earth
	Group 1B Beliefs that the environment is being seriously harmed
ALT4	The effects of pollution on public health are worse than we realise
ALT5	Pollution generated here harms people all over the Earth
BIO2	Over the next several decades, thousands of species will become extinct
BIO5	Modern development threatens wildlife
	GROUP 2: Beliefs supporting environmental inaction
	Group 2A Beliefs that environmental protection has negative consequences
EGO3	Protecting the environment will threaten jobs for people like me
EGO4	Laws to protect the environment limit my choice and personal freedom
	Group 2B Beliefs that the environment is not being seriously harmed
ALT3	We do not need to worry much about the environment because future generations will be better able to deal with these problems than us
BIO1	While some local plants and animals may have been harmed by environmental degradation, over the whole Earth there has been little effect
BIO3	Claims that current levels of pollution are changing Earth's climate are exaggerated

MNL form:

$$\pi_{n,c_s} = \frac{\exp(\mu_{0s} + \lambda_s' SD_n)}{\sum_{s=1}^{C} \exp(\mu_{0s} + \lambda_s' SD_n)}, \tag{11}$$

where μ_{0s} and λ_s' are parameters to be estimated. The sign of these parameters determines whether increases in their value lead to an increased or decreased probability of a specific class. In our case, the socio-demographic variables were dummy variables for recreationalists, gender, NGO, number of adults and number of children in the household. See Table 3 for variable definition and summary statistics.

4.2.3. Model 3: Hybrid latent class model (HLCM)

Previous models were compared with a hybrid latent class model specification recently proposed by Hoyos et al. (2015). In this case, attitudes are considered to be latent variables and are used in the class allocation function of a classical LCM. This hybrid modelling framework describes how attitudes affect choices through class allocation probabilities, and at the same time uses observed choices as feedback for the estimation of the latent attitudinal variables. The aim of this approach is to adequately capture individual taste heterogeneity through attitudinal indicators. Some of the heterogeneity may be related to the socio-demographic characteristics of respondents but non-observed attitudes may in fact be the main cause of heterogeneity (Small et al., 2005, 2006). In line with Hess and Beharry-Borg (2012) and Daly et al. (2012), both the repeated choice nature of the data and the ordinal nature of the attitudinal indicators are taken into account.

Table 3
Summary of statistics and socioeconomic variables.

Variable	Description	Mean	Std. Dev.	Min	Max
NAT	Native forest attribute	14.13	10.87	2	30
VIN	Vineyard attribute	26.79	11.46	10	40
FOR	Forest attribute	28.90	9.39	15	40
BIO	Biodiversity attribute	15.02	7.8	5	25
REC	Recreation attribute	−0.34	2.31	−3	3
COST	Cost	26.22	33.91	0	100
MALE	Gender (1 if male)	0.47	0.5	0	1
ADULT	Number of adults	2.56	0.92	1	5
CHILD	Number of children	0.31	0.66	0	4
EDUC	Education	2.73	1.16	1	5
NGO	Environmental NGO	0.03	0.16	0	1
RECR	Recreationalist	0.50	0.5	0	1

Table 4
Responses to the environmental attitudinal questions.

Item	1	2	3	4	5
Group 1A Beliefs that environmental protection has positive consequences					
EGO1	0.45%	0.45%	7.69%	28.96%	62.44%
EGO2	0.90%	0.90%	7.24%	26.24%	64.71%
EGO5	0.45%	2.26%	9.05%	32.13%	56.11%
ALT1	0.00%	1.81%	5.43%	28.51%	64.25%
ALT2	0.45%	2.71%	9.50%	28.05%	59.28%
BIO4	1.36%	2.26%	12.67%	29.86%	53.85%
Group 1B Beliefs that the environment is being seriously harmed					
ALT4	1.36%	4.52%	15.84%	33.94%	44.34%
ALT5	4.52%	7.69%	20.81%	35.75%	31.22%
BIO2	4.98%	3.17%	15.38%	33.94%	42.53%
BIO5	3.17%	1.81%	14.48%	36.20%	44.34%
Group 2A Beliefs that environmental protection has negative consequences					
EGO3	56.56%	14.93%	11.76%	10.86%	5.88%
EGO4	46.61%	28.05%	16.29%	7.69%	1.36%
Group 2B Beliefs that the environment is not being seriously harmed					
ALT3	62.90%	23.98%	7.24%	4.07%	1.81%
BIO1	35.75%	32.13%	18.10%	11.31%	2.71%
BIO3	35.29%	23.08%	22.62%	14.03%	4.98%

Note: Awareness of consequences scale attitudinal indicators (EGO1–EGO5, ALT1–ALT5 and BIO1–BIO5) were framed in a 5 point Likert scale, with 1 indicating total disagreement and 5 total agreement.

So, the structural equation for the choice model was the same one as described in Eq. (8). In addition, the structural equation for the q-th latent variable model is given by the following formula:

$$LV_{qn}^* = \gamma_{q,Eus} EUS_n + \gamma_{q,Recr} RECR_n + \gamma_{q,Male} MALE_n + \gamma_{q,Adult} ADULT_n + \gamma_{q,Child} CHILD_n + \gamma_{q,Educ} EDUC_n + \gamma_{q,NGO} NGO_n + \omega_{qn} \tag{12}$$

where ω_q is a random disturbance, which is assumed to be normally distributed with a zero mean and standard deviation σ_q. Following the psychological framework proposed by Ryan and Spash (2012), two latent variables were defined: the first latent variable, LV_1^*, aimed to capture beliefs supporting environmental action; and the second latent variable, LV_2^*, aimed to capture beliefs supporting environmental inaction.

Both latent variables, LV_1^* and LV_2^*, were linked to the remaining part of the model through class allocation probabilities. They are, therefore, respondent-specific and a function of the latent variable:

$$\pi_{n,c_s} = \frac{\exp(\mu_{0s} + \mu_{1s} LV_{1n}^* + \mu_{2s} LV_{2n}^* + \lambda_s' SD_n)}{\sum_{s=1}^{C} \exp(\mu_{0s} + \mu_{1s} LV_{1n}^* + \mu_{2s} LV_{2n}^* + \lambda_s' SD_n)}, \tag{13}$$

where μ_{0s}, μ_{1s}, μ_{2s}, and λ_s are parameters to be estimated. The sign of μ parameters determines whether increases in the value of the latent variable lead to an increased or decreased probability of a specific class allocation function, while λ parameters determine the influence of certain socio-demographic characteristics of respondents on the class allocation function. In our case, these socio-demographic variables were dummy variables for recreationalists, gender, NGO, education, number of adults and number of children in the household.

Finally, measurement equations use the values of the attitudinal indicators as dependent variables, and explain their values with the help of the latent variables. The ℓ^{th} indicator (of total L_q indicators) for respondent n is therefore defined as

$$I_{q\ell n} = m\left(LV_{qn}^*, \zeta_q\right) + v_{qn}, \tag{14}$$

where the indicator $I_{q\ell}$ is a function of latent variable LV_q^* and a vector of parameter ζ_q. The specification of v_q determines the behaviour of the measurement model and depends on the nature of the indicator. In our case the first latent variable LV^*_{1n}represents groups 1A and 1B from Table 2 and captures therefore, the beliefs supporting environmental action. Its indicators are items from groups 1A and 1B as defined in Table 2. Similarly, the second latent variable LV^*_{2n} represents groups 2A and 2B from Table 2, corresponding to the beliefs supporting environmental inaction and using the indicators corresponding to groups 2A and 2B.

4.2.4. Model 4: Latent class with direct inclusion of indicators in allocation probabilities (LCM IND)

Departing from model 3, model 4 tries to mimic the hybrid model by direct incorporation of two indicators in the allocation function so that it resembles eq. (13) and is defined as:

$$\pi_{n,c_s} = \frac{\exp(\mu_{0s} + \gamma_{1s} IND_{11n} + \gamma_{2s} IND_{21n} + \lambda'_s SD_n)}{\sum_{s=1}^{C} \exp(\exp(\mu_{0s} + \gamma_{1s} IND_{11n} + \gamma_{2s} IND_{21n} + \lambda'_s SD_n))}. \quad (15)$$

It is important to note that responses to attitudinal questions (indicators) cannot be directly included in the class allocation function of a LCM without specific treatment, for the same reason that a new source of endogeneity was shown to emerge in Eq. (5). This is due to the fact that expression (15) is a multinomial logit formula corresponding to a model with alternative specific coefficients (in this case classes are alternatives) that is based on a latent variable similar to Eq. (1). The only difference is that in Eq. (1) we have information of explanatory variables on alternative level and generic coefficients, but in a utility equation underlying the allocation function we have information only on the individual level. That is why the coefficients λ_s, γ_{1s} and γ_{2s} are class specific. If the indicators IND_{11n} and IND_{21n} are generated by a similar process to Eq. (14), they are by definition endogenous and likely to be correlated with the error of the underlying utility equation for the allocation function. The main purpose of this model is to precisely investigate the impact of this theoretically incorrect approach of incorporating attitudinal data in a discrete choice model at the individual level.

As the main goal is to compare different ways of including underlying environmental beliefs into a choice model, we define the allocation function as close as possible to the hybrid approach. That is why we include two indicators into Eq. (15) to mimic the allocation function (13). As in the previous case, the first indicator in Eq. (15) represents beliefs supporting environmental action resembling the first latent variable LV^*_{1n} and similarly the second indicator represents the beliefs of environmental inaction resembling the second latent variable LV^*_{2n}. The difference between the two approaches is that the latent variables in the hybrid framework are dependent variables in another structural Eq. (12) and, at the same time, explanatory variables in measurement Eq. (14), which makes the whole model very complex. The direct inclusion of indicators of the allocation function, as done in Eq. (15), simplifies the model considerably but it suffers from an endogeneity problem as explained above. The question is how big this bias is in estimated coefficients and welfare measures. Thus, as explained above, model 4 aims to analyse the effect of untreated endogeneity both in estimated coefficients and welfare estimates.

4.2.5. Model 5: Latent class with MIS correction in allocation probabilities (LCM MIS)

The aim of model 5 is to include the responses of the attitudinal questions in the class allocation function and treat their possible endogeneity by the MIS approach described in Section 2. As mentioned before in model 4, the first indicators IND_{11n} and IND_{21n} represent the beliefs supporting environmental action and inaction, respectively. According to Section 2, in our case, two latent variables, denoted q^*_{in} in Eq. (1), are included into the underlying utility equation for the allocation function. That is why two indicators for each latent variable are needed in order to apply the MIS correction. Let us assume that there are two pairs of indicators IND_{11n}, IND_{12n} and IND_{21n}, IND_{22n} for the first and second latent variables, respectively. Then, according to Eq. (6), the two auxiliary regressions for the two latent variables would be:

$$\begin{aligned} IND_{11n} &= \hat{\theta}_{10} + \hat{\theta}'_{11} SD_n + \hat{\theta}_{12} IND_{12n} + \hat{\eta}_{1n} \\ IND_{21n} &= \hat{\theta}_{20} + \hat{\theta}'_{22} SD_n + \hat{\theta}_{22} IND_{22n} + \hat{\eta}_{2n}. \end{aligned} \quad (16)$$

Then, according to Eq. (13) the corrected allocation function of model 5 is defined as:

$$\pi_{n,c_s} = \frac{\exp\left(\mu_{0s} + \gamma_{1s} IND_{11n} + \gamma_{\hat{\eta}_1} \hat{\eta}_{1n} + \gamma_{2s} IND_{21n} + \gamma_{\hat{\eta}_2} \hat{\eta}_{2n} + \lambda'_s SD_n\right)}{\sum_{s=1}^{C} \exp\left(\exp\left(\mu_{0s} + \gamma_{1s} IND_{11n} + \gamma_{\hat{\eta}_1} \hat{\eta}_{1n} + \gamma_{2s} IND_{21n} + \gamma_{\hat{\eta}_2} \hat{\eta}_{2n} + \lambda'_s SD_n\right)\right)}. \quad (17)$$

4.3. Selection of indicators

Given the importance that the selection of indicators has in the MIS procedure, in this section we propose a specific procedure to choose these indicators based on the theoretical foundations and results of a multivariate analysis applied on the responses to the attitudinal questions. As a first step, an exploratory factor analysis was conducted on the responses to the attitudinal questions presented in Table 2. The exploratory factor analysis employed principal axis factor analysis. There are two latent variables in model 3 (HLCM), model 4 (LCM IND), and model 5 (LCM MIS). The first latent variable represents groups 1A and 1B and the second latent variable represents groups 2A and 2B from Table 2. It is expected that these latent variables are linked to the two groups of variables representing the beliefs supporting environmental action (1A and 1B) and inaction (2A and 2B) as classified according to the BSEAI scale.

As shown in Table 5, almost 40% of the variance is represented by the first two factors. Fig. 3 presents the projection of all variables in the plane defined by the first two factors. This clearly shows the first big group (orientated to the right hand part of the graph) of variables belonging to groups 1A and 1B conforming the set of beliefs supporting environmental action and, the second, smaller group (orientated to the left hand upper corner) of variables from groups 2A and 2B, representing the set of beliefs supporting environmental inaction.

As explained before, the MIS method requires two indicators to address the endogeneity coming from each omitted factor. Each first indicator is used as an explanatory variable in the allocation function, and each second indicator is used as explanatory variable in the respective auxiliary regression (6).

We therefore propose that the choice of the two indicators for each latent variable to be made both on the theoretical definition of the BSEAI scale and the results of the exploratory factor analysis. The specific choice is, therefore, based on the factor loadings presented on the right hand side part of Table 5. The highest factor loadings of first factor correspond to the variables ALT2 and EGO2 and these at the same time belong to the first group of the BSEAI scale corresponding to the environmental action. Similarly, highest factor loadings of second factor correspond to the variables EGO3 and BIO3 and these at the same time belong to the second group of the BSEAI scale, that of environmental inaction. That is why these two pairs of variables are chosen to be indicators of the first and second latent variable, respectively, corresponding to IND_{11n}, IND_{12n} and IND_{22n}, IND_{22n} in Eqs. (16).

4.4. Estimation results

Results of the MNL (model 1), LCM (model 2) and HLCM (model 3) are replicated from Hoyos et al. (2015), and that is why they are included in the Appendix A (Table A1). Using these results as a starting

Table 5
Results of the principal axis factor analysis.

Eigenvalues and percentages				Factor loadings					
Factor	Eigenvalue	% of variance	Cumulative %	Variable	Factor 1	Factor 2	Factor 3	Factor 4	Group
Factor 1	4.56	30.39	30.39	BIO2	0.31	−0.11	−0.28	0.23	1A
Factor 2	1.37	9.15	39.54	EGO1	**0.64**	−0.08	0.23	0.04	1A
Factor 3	1.16	7.73	47.27	ALT4	**0.64**	0.12	−0.01	0.26	1A
Factor 4	1.08	7.17	54.44	ALT2	**0.78**	0.04	0.05	0.12	1A
Factor 5	1.01	6.73	61.17	EGO2	**0.76**	0.05	−0.02	0.20	1A
Factor 6	0.87	5.79	66.96	ALT3	−0.37	**0.43**	0.33	−0.10	1A
Factor 7	0.75	5.03	71.98	EGO4	−0.18	0.22	**0.72**	−0.01	1B
Factor 8	0.71	4.70	76.68	EGO5	**0.67**	0.22	0.08	0.19	1B
Factor 9	0.66	4.40	81.08	BIO1	**−0.50**	0.34	0.13	0.49	1B
Factor 10	0.59	3.96	85.04	ALT5	**0.52**	0.23	0.16	−0.12	1B
Factor 11	0.59	3.92	88.96	BIO3	−0.36	**0.56**	−0.34	0.40	2A
Factor 12	0.50	3.32	92.29	BIO4	**0.50**	0.29	0.07	**−0.49**	2A
Factor 13	0.46	3.10	95.38	ALT1	**0.75**	0.20	0.08	0.14	2B
Factor 14	0.37	2.48	97.86	EGO3	**−0.42**	**0.55**	−0.17	−0.21	2B
Factor 15	0.32	2.14	100.00	BIO5	**0.44**	0.38	**−0.43**	−0.37	2B

point allows for direct comparison to the results obtained in the new models proposed in this paper (model 4 and model 5). The two classes obtained in the LCM and the HLCM can be characterised as follows. The respondent utility in class 1 increases if the surface area covered by native forest increases and if the number of endangered species decreases. The cost sensitivity is lower (in absolute value) than in class 2 and, moreover, the significant negative coefficient for ASC1 suggests that, all else being equal, respondents in this class tend to move away from the status quo (i.e. they prefer to implement a protection programme). That indicates that class 1 corresponds to individuals with environmental concerns and are willing to pay to protect native forest and endangered species (individuals supporting environmental action). In class 2 the cost sensitivity is much higher than in class 1 and only the coefficient accompanying the forest attribute is significant apart from the cost attribute. Individuals in this class prefer more surfaces covered by forest tree plantations, indicating their business focussed orientation regarding the studied area, together with higher price sensitivity leading to a lower willingness to pay (individuals supporting environmental inaction).

The key elements of the models presented in Hoyos et al. (2015) and in this paper are, however, the allocation functions (11), (13), (15) and (17). That is, the goal is how the allocation functions can incorporate some latent attitudes. The allocation function of the LCM (model 2) in Table A1 does not include any latent behaviour and presents only two significant socio-demographic variables, showing that recreationalists have a lower probability of belonging to class 2. The number of adults in the household, on the other hand, increases the probability of being in this class.

The allocation function of the HLCM (model 3) in Table A1 does not present any significant socio-demographic variables because their influence is represented by the latent variables and their corresponding coefficients. As explained in Hoyos et al. (2015) the first latent variable

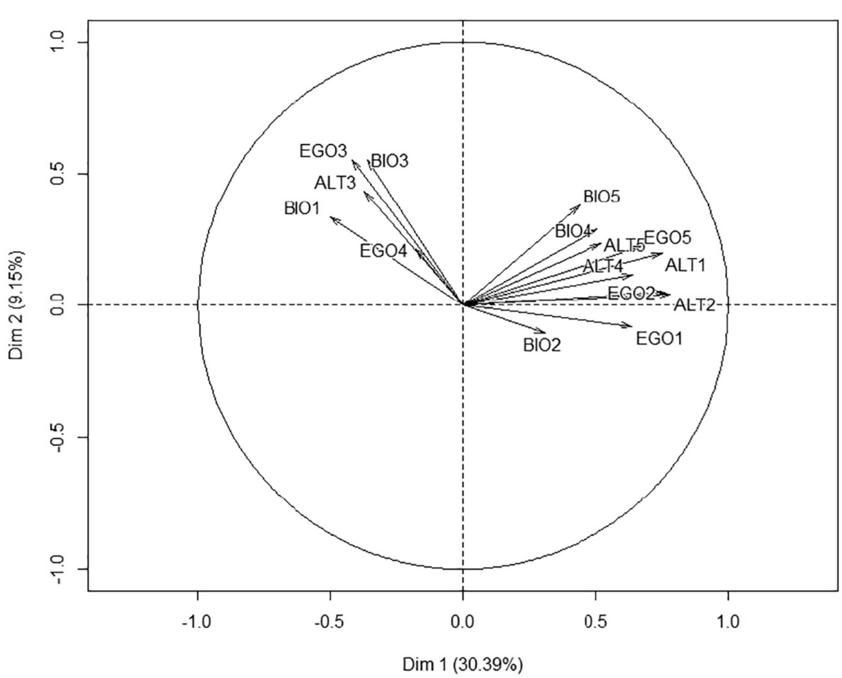

Fig. 3. Variables factor map (principal component analysis).

Table 6
Estimations of LCMs with direct inclusion of indicators in the allocation function without (LCM IND) and with correction of endogeneity (LCM MIS).

	LCM IND (model 4)			LCM MIS (model 5)		
Variable	*Est.*	*p-value*		*Est.*	*p-value*	
Class 1						
ASC_1	−1.585	<0.01	***	−1.590	<0.01	***
ASC_2	0.083	0.27		0.083	0.27	
β_{NAT}	0.052	<0.01	***	0.052	0.00	***
β_{VIN}	0.007	0.19		0.007	0.19	
β_{FOR}	−0.011	0.10		−0.011	0.10	
β_{BIO}	−0.054	<0.01	***	−0.054	<0.01	***
β_{REC}	0.033	0.21		0.033	0.20	
β_{COST}	−0.017	<0.01	***	−0.017	<0.01	***
Class 2						
ASC_1	−1.091	0.32		−1.084	0.32	
ASC_2	0.519	0.11		0.521	0.10	*
β_{NAT}	0.023	0.26		0.024	0.24	
β_{VIN}	0.018	0.36		0.017	0.36	
β_{FOR}	0.068	0.01	**	0.068	0.01	**
β_{BIO}	0.018	0.63		0.019	0.62	
β_{REC}	−0.118	0.24		−0.118	0.24	
β_{COST}	−0.095	0.00	***	−0.094	0.00	***
Class allocation						
μ_{02}	0.656	0.61		−2.824	0.13	
γ_{12}	−0.707	<0.01	***	−0.599	0.05	**
γ_{22}	0.109	0.47		1.340	0.01	***
$\gamma_{\hat{\eta}_1}$				−0.122	0.77	
$\gamma_{\hat{\eta}_2}$				−1.291	0.01	***
$\lambda_{12,\ Recr}$	−0.595	0.10		−0.776	0.04	**
$\lambda_{22,\ Male}$	0.236	0.51		0.089	0.81	
$\lambda_{32,\ Adult}$	0.282	0.14		0.202	0.31	
$\lambda_{42,\ Child}$	0.179	0.53		0.448	0.15	
$\lambda_{52,\ Educ}$	0.077	0.63		0.221	0.18	
$\lambda_{62,\ NGO}$	0.173	0.88		0.572	0.63	
N		1326			1326	
K		25			27	
lnL		−895.86			−893.38	
AIC		1841.72			1840.77	
BIC		2151.22			2175.02	

***, **, * denote significance at the 1%, 5%, and 10% level respectively.

(pro-action) is smaller for males and higher for recreationalists and families with children. The second latent variable (pro-inaction) is higher for males and households with more adults, though smaller for families with children, more educated people and environmentalists (these results correspond to the estimation of Eq. (12) and are presented in Hoyos et al., 2015). The influence of these latent variables is represented by the coefficients in the class allocation model (lower part of Table A1). Respondents with a more negative value for the first latent variable (pro-action) are less likely to belong to class 2, while respondents with a more positive value of the second latent variable (pro-inaction) are more likely to belong to class 2. The probability of belonging to a specific class is thus driven by the latent attitudes represented by the two latent variables, with individuals supporting environmental action more likely to belong to class 1 and individuals supporting environmental inaction more likely to belong to class 2.

Table 6 presents the results of LCMs with direct inclusion of indicators in the allocation function without (model 4) and with correction of endogeneity (model 5). The estimated attribute coefficients in both classes are very similar and very close to the estimation of the plain LCM (model 2) presented in Table A1. That means that the characterisation of the two classes remains the same. The novelty emerges in the allocation function. As explained above and defined in (15), the coefficients γ_{12} and γ_{22} should represent the effect of latent attitudes represented by two indicators on the allocation function.

Similar to the HLCM, no socio-demographic variable is significant in the allocation function in model 4 when the two indicators are included. As expected, the effect of the first indicator (γ_{12}) is negative indicating that the people with beliefs supporting environmental action have a lower probability to be in the second class labelled group as individuals supporting environmental inaction. The second indicator is not significant.

Finally, model 5 includes the MIS correction of possible endogeneity by the means of inclusion of residuals of the auxiliary regressions (presented in Table 7) into the allocation function defined in Eq. (17). The significant coefficient $\gamma_{\hat{\eta}_2}$ indicates that the second indicator suffers endogeneity (Rivers and Voung, 1988), implying that model 4 provides inconsistent estimators. Additionally, the two coefficients γ_{12} and γ_{22} are significant at 5% and their signs are in line with the previous interpretation of classes and indicators. The first indicator represents the set of beliefs supporting environmental action and its negative coefficient γ_{12} indicates that individuals with those beliefs have a lower probability of belonging to Class 2. Similarly, the second indicator represents the set of beliefs supporting environmental inaction and, that is why, its coefficient γ_{22} is positive making the probability of belonging to class 2 bigger. There is only one socio-demographic variable significant in the allocation function of model 5 showing that recreationalists have a lower probability of belonging to class 2 (individuals supporting environmental inaction).

The HLCM allows deeper analysis of the existing preference heterogeneity than a plain choice model through the link of allocation probabilities to socio-demographic variables by the use of underlying attitudes. In the HLCM, the latent variables depend on the socio-demographic variables as defined in Eq. (12) and at the same time the latent variables enter the allocation function (13). This interesting

Table 7
Auxiliary regressions.

Dependent variable:	ALT2 (Indicator 1 of LV_1^*)			Dependent variable:	EGO3 (Indicator 1 of LV_2^*)		
Explanatory variables	*Est.*	*p-value*		*Explanatory variables*	*Est.*	*p-value*	
Constant	1.51	<0.01	***	*Constant*	2.08	<0.01	***
EGO2 (Indicator 2 of LV1)	0.67	<0.01	***	*BIO3 (Indicator 2 of LV2)*	0.24	<0.01	***
Recreationalists	0.19	<0.01	***	*Recreationalists*	0.08	0.20	
Gender (male)	−0.04	0.20		*Gender (male)*	0.05	0.48	
Education	−0.06	<0.01	***	*Education*	−0.13	<0.01	***
Number adults in household	−0.03	0.09	*	*Number adults in household*	0.04	0.25	
Number children in household	0.10	<0.01	***	*Number children in household*	−0.14	<0.01	***
Environmental NGO	−0.03	0.78		*Environmental NGO*	−0.23	0.26	
R^2	0.43			R^2	0.10		
Number of observations	1326			Number of observations	1326		

influence of the socio-demographic variables on the allocation function, described in Hoyos et al. (2015), can be summarised as follows. The expected value of the first latent variable (pro-action) decreases with an increase in number of adults in the household and for males, but it increases for recreationalists and families with children. The expected value of the second latent variable (pro-inaction) increases for males and number of adults in the household and decreases for families with children, more educated people and environmentalists. Very similar effects can be found in the auxiliary regressions presented in Table 7 assuming that variable ALT2 represents the first latent variable and EGO3 the second latent variable.

Moreover, the estimation results of the attribute coefficients of the LCM with MIS correction (model 5) and HLCM (model 3) are very similar: pro-environmental individuals are less likely to be found in class 2, where we are more likely to find individuals showing a higher sensitivity to agricultural development attributes, such as forest tree plantation extensions. In contrary to this, individuals belonging to class 1 show a higher sensitivity to environmental attributes (native forest, biodiversity and recreation). Quite interestingly, the MIS method seems to be more parsimonious, as it is capable of reaching almost identical results with significantly fewer parameters, 29 (or 47 if the coefficients of the auxiliary regressions are included), as compared with the more complicated hybrid model that requires the estimation of 113 parameters to offer an equally rich interpretation.

4.5. Impact on welfare

Next, we compare the implications of the previous results in terms of welfare measures. In the case of welfare measures, it is convenient to conduct this comparison by analysing willingness to pay (WTP) values. Compensating surplus (CS) welfare estimates may be obtained from Hanemann (1984) and Train (1998):

$$CS = -\frac{1}{\alpha}\left(\ln\left(\sum exp\left(\beta X_{ij}^{0}\right)\right) - \ln\left(\sum exp\left(\beta X_{ij}^{1}\right)\right)\right), \tag{18}$$

where α is the marginal utility of income (usually represented by the coefficient of the payment attribute) and X_{ij}^0 and X_{ij}^1 represent the vector of environmental attributes at the initial level (status quo) and after the change levels, respectively. Simplifying the above equation, the WTP for a marginal change in the level of provision of each environmental attribute is obtained by dividing the coefficient of the attribute by the coefficient of the cost attribute.

Marginal WTP values were simulated following the Krinsky and Robb (1986) procedure. Fig. 4 shows the box-plots of the simulated WTP distributions derived from models 2 (LCM), 3 (HLCM), 4 (LCM IND) and 5 (LCM MIS). Each WTP distribution is characterised by its minimum, 25th percentile, median, 75th percentile and maximum. At first glance, the distributions of Fig. 4 show relative robustness regardless of the model used. The noteworthy result is the wide spread of the distributions based on LCM MIS model. In similarity with the instrumental variable approach in a linear regression, the MIS approach is not efficient and can lead to relatively wide distributions (see e.g. Guevara, 2015).

But what really matters is whether there are significant differences between simulated WTP distributions among the four analysed models. We tested the difference between the simulated WTP distributions depicted in Fig. 4 using the complete combinatorial method (Poe et al., 2005). This method is designed to calculate the difference between two independent empirical distributions and conduct a statistical test on that difference. Table 8 presents *p-values* of all pair comparisons between models for all attributes. As can be seen in Table 8, the null hypothesis that the difference is zero cannot be rejected in any comparison.

The results in Fig. 4 and Table 8 are based on marginal changes. An interesting question is whether the result of no differences in the obtained outcomes among the four different approaches remains if the attribute changes are bigger. That is why the compensating variation (CV) measures (Adamowicz et al., 2011) are computed corresponding to four hypothetical scenarios. These scenarios presented in Table A2 in the Appendix A are described in detail in Hoyos et al. (2012). They were developed taking into account ecologically feasible land use changes: (1) enhancement of vineyard activity causing the area of vineyard plantations to increase; (2) moderate enhancement of ecological values; (3) high enhancement of ecological values; (4) maximum enhancement of ecological values.

Fig. 5 shows, in a similar fashion to Fig. 4, the simulated CV distributions derived from models 2 (LCM), 3 (HLCM), 4 (LCM IND) and 5 (LCM MIS). Despite the fact that the assumed changes in the attribute levels are bigger than those assumed in Fig. 4, the general conclusions remain. The distributions of the CV for the four hypothetical scenarios obtained by model 3 (HLCM), 4 (LCM IND) and 5 (LCM MIS) are, firstly, very similar with respect to median values and, secondly, the MIS approach can lead to relatively wide distributions (see e.g. Guevara, 2015).

Similar to Table 8, we also tested the difference between the simulated CV distributions depicted in Fig. 5 using the complete combinatorial method (Poe et al., 2005). A significant difference is found only for models 2 (LCM) and 3 (HLCM) in Scenario 2 (Table 9). Generally, model 2 (LCM) can be seen as a model with missing explanatory variables in the allocation functions. This shortcoming is solved in 3 (HLCM), 4 (LCM IND) and 5 (LCM MIS). This can be an explanation of the similar behaviour of the corresponding three distributions in Fig. 5 and slight departures of the distributions of model 2 (LCM).

This result highlights the conclusion reached in the previous subsection: the MIS correction to endogeneity provides similar welfare estimates as the HLCM though with a significantly smaller set of

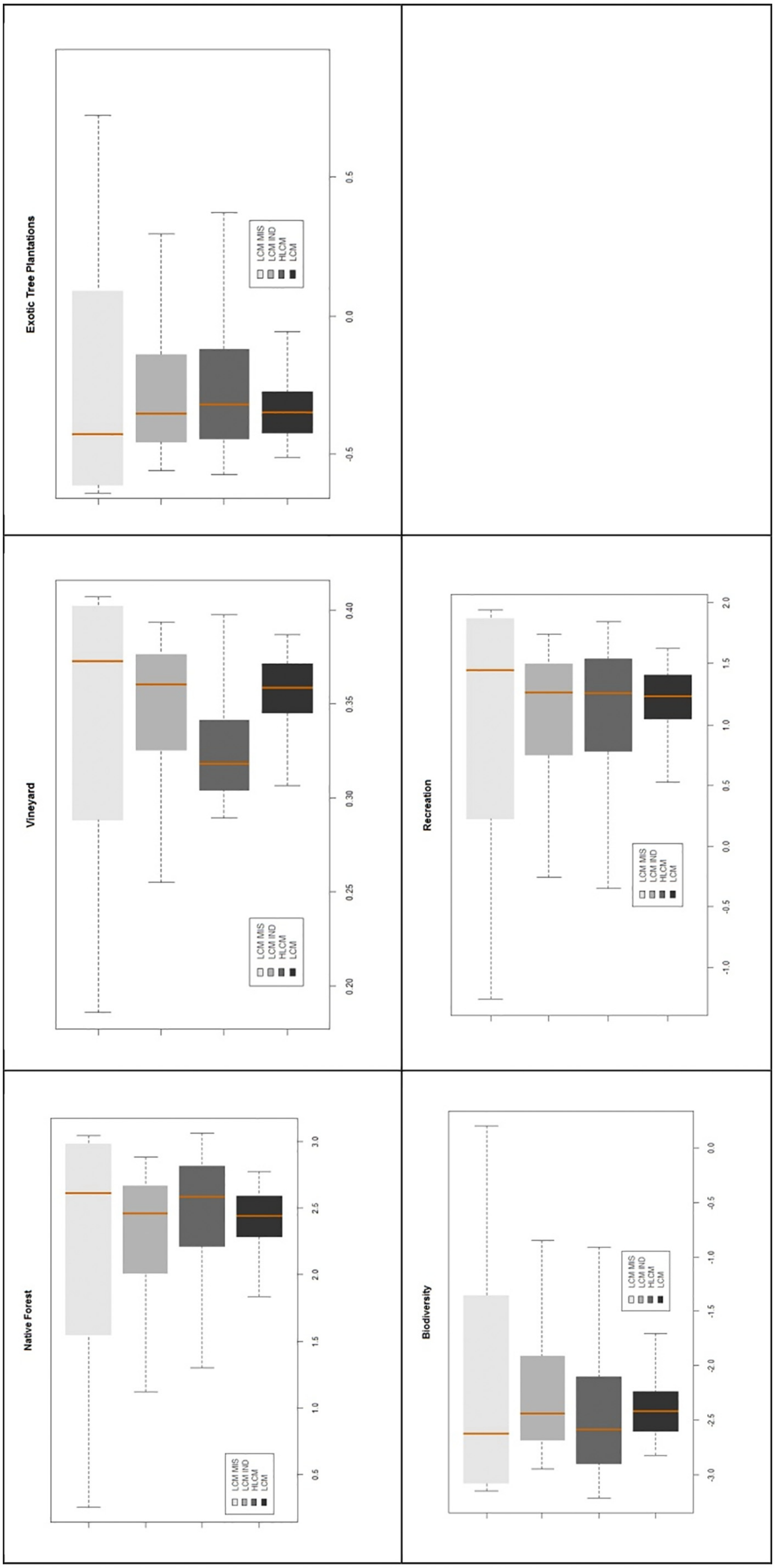

Fig. 4. Comparison of welfare estimates (WTP).

Table 8
Poe test for difference between two independent empirical distributions (WTP).

	Native forest	Vineyard	Exotic tree plantations	Biodiversity	Recreation
LCM-HLCM	0.40	0.18	0.44	0.41	0.50
LCM-LCM IND	0.49	0.49	0.48	0.49	0.49
LCM-LCM MIS	0.45	0.44	0.45	0.45	0.44
HLCM-LCM IND	0.38	0.32	0.48	0.38	0.47
HLCM-LCM MIS	0.49	0.37	0.42	0.49	0.43
LCM IND -LCM MIS	0.41	0.41	0.41	0.41	0.41

Table 9
Poe test for difference between two independent empirical distributions (CV).

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
LCM-HLCM	0.24	0.02	0.21	0.43
LCM-LCM indicators	0.48	0.49	0.48	0.49
LCM-LCM MIS	0.44	0.44	0.45	0.44
HLCM-LCM indicators	0.36	0.19	0.33	0.42
HLCM-LCM MIS	0.39	0.29	0.37	0.45
LCM indicators -LCM MIS	0.41	0.41	0.42	0.41

parameters. In addition, our results show that, although we find evidence of endogeneity, the bias it may produce both on parameter and welfare estimates seems to be almost negligible, mostly noticeable only at the level of the class allocation model.

5. Discussion and conclusions

This paper shows that the application of the MIS method to correct for endogeneity in discrete choice models for environmental valuation may be considered as a valuable tool for practitioners dealing with this issue given its relative simplicity. Another advantage to practitioners is that this method can be easily applied in any econometric software. Although in this particular case, the existence of endogeneity does not seem to considerably bias the estimation results, this method may be considered a useful tool for testing the existence of endogeneity due to omitted attributes in DCEs for environmental valuation. Given that researchers usually collect attitudinal information from respondents, the robustness of the estimation results to the presence of endogeneity can be easily tested following the MIS methodology presented in this paper. In this regard, this research contributes to the development of the MIS method by proposing and illustrating a factor analysis procedure for the selection of indicators.

The paper also allows the comparison of the performance of the MIS method and the recently proposed hybrid choice models in order to address the endogeneity problem that may be found due to the omission of latent environmental attitudes. Firstly, the results suggest that, in this particular application, the MIS method seems to successfully address the omission of latent variables in a more simplistic way than that of the hybrid choice model. This result is in line with the Monte Carlo evidence provided by Guevara (2015). The MIS method basically

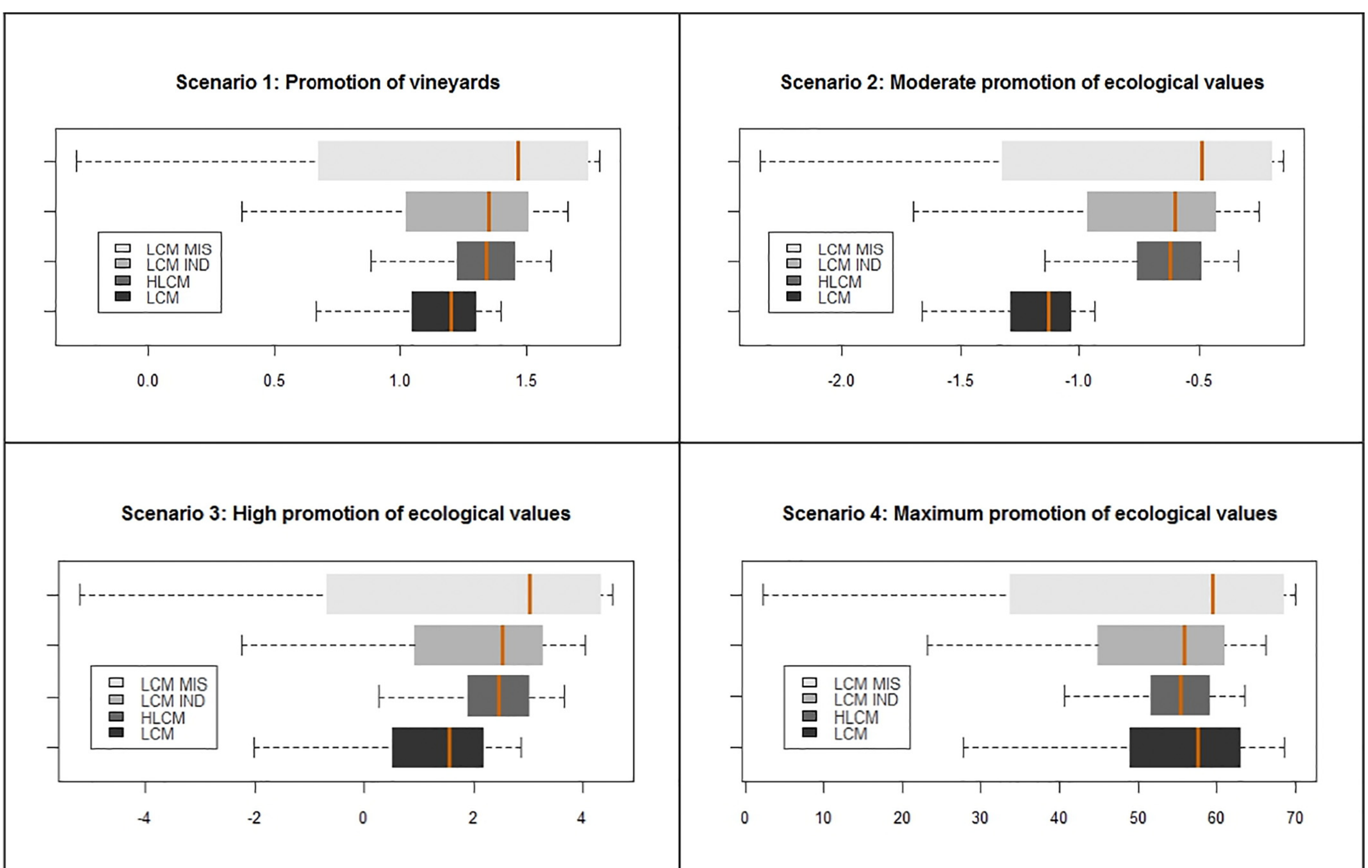

Fig. 5. Comparison of compensating variations for different hypothetical scenarios.

requires an auxiliary linear regression and a standard discrete choice model, whereas the hybrid model involves simultaneous estimation of more complicated structural and measurement equations. Secondly, the simplicity of the MIS method leads to a notably lower computational burden. The estimation of hybrid models involves maximisation of complex likelihood functions and related computational issues. Thirdly, we find that in our case both estimation methods produce similar parameter estimates. However, researchers should bear in mind that these models are not directly comparable since they are based on different assumptions. The main assumption for the MIS approach relates to the independency required in Eq. (4), while hybrid modelling relies on the specification of valid structural and measurement Eqs. (12), (13) and (14). In this sense, despite the fact the MIS method may be less efficient in general than the HLCM, it can be said to be more robust because it requires milder modelling assumptions.

It is important to note that the MIS method applies only to cases where the researcher believes that endogeneity is due to omitted attributes (latent environmental attitudes, in our case) in the allocation function, and that the hybrid model remains a valuable tool to deal with this problem in other possible cases. For example, when the omission of the latent environmental attitude occurs in the utility function. From this perspective, the hybrid choice model framework remains as a more flexible method.

To the best of our knowledge, there are no previous applications of any method aimed to detect and/or correct for endogeneity in DCE for environmental valuation using the MIS approach. However, various DCE studies in other fields have identified endogeneity as a critical problem, including, but not limited to: choice of wine (Palma et al., 2018), airline itinerary choice (Lurkin et al., 2017), mode choice (Fernández-Antolín et al., 2016), passenger booking timing (Wen and Chen, 2017), learning models of route choice (Guevara et al., 2017), mobility data collection (Zegras et al., 2018), demand for electric vehicles (Helveston, 2016), valuation of public transport attributes (Guevara et al., 2018), residential choice (Guevara, 2005, 2010; Guevara and Ben-Akiva, 2006; Guevara and Polanco, 2016) and automobile choice (Petrin and Train, 2010). Among those, Palma et al. (2018), Fernández-Antolín et al. (2016) and Guevara and Polanco (2016) have used the MIS method.

Moreover, it is important to bear in mind that the properties of the MIS method critically depend on the quality of the indicators used. As shown by Guevara (2015), the bias for the MIS method with proper indicators is negligible, but the results can be extremely poor when the two indicators do not fulfil the required conditions (2) and (4), or they are weak indicators of the underlying unobserved construct.

The empirical results reached in this paper open a path to further explore the issue of endogeneity in discrete choice modelling for environmental valuation. More applications are needed in order to establish a body of literature sufficient enough to determine the magnitude of the problem and the best way to deal with it.

Acknowledgements

P. Mariel, D. Hoyos and A. Artabe acknowledge the funding provided by the Department of Education of the Basque Government through grant IT-642-13 (UPV/EHU Econometrics Research Group), IT-783-13 (BIRE-Bilbao Research Economics), by Spanish Ministry of Economy and Competitiveness through grant ECO2017-82111-R and by University of the Basque Country (UPV/EHU) under research grant US15/11. C. Angelo Guevara acknowledges the funding provided by CONICYT, FONDECYT 1150590 and the Complex Engineering Systems Institute, Chile (CONICYT FB0816).

Appendix A

Table A1
Estimations of MNL, LCM and HCLM.

	MNL (model 1)	LCM (model 2)	HLCM (model 3)
Number of individuals:	221	221	221
Number of observations:	1326	1326	1326
Log-likelihood:	−1208.705	−902.60	−4265.168
AIC	2433.41	1851.19	8756.34
BIC	2532.45	2135.93	8869.34
Parameters:	8	23	113

	MNL (model 1)		LCM (model 2) Class 1		LCM (model 2) Class 2		HLCM (model 3) Class 1		HLCM (model 3) Class 2	
	Est.	*p-val.*	Est.	*p-val.*	Est.	*p-val.*	Est.	*p-val.*	Est.	*p-val.*
ASC_1	0.266	0.30	−1.550***	<0.01	−0.896	0.35	−1.960***	<0.01	−0.447	0.64
ASC_2	0.094	0.17	0.085	0.30	0.540 *	0.07	0.086	0.30	0.438*	0.08
β_{NAT}	0.046***	<0.01	0.052***	<0.01	0.025	0.13	0.053***	<0.01	0.021	0.11
β_{VIN}	0.007	0.12	0.007	0.23	0.016	0.39	0.006	0.28	0.019	0.20
β_{FOR}	−0.007	0.27	−0.011	0.13	0.067 ***	<0.01	−0.010	0.14	0.048 **	0.04
β_{BIO}	−0.043 ***	<0.01	−0.053 ***	<0.01	0.015	0.66	−0.056 ***	<0.01	0.016	0.61
β_{REC}	0.015	0.52	0.033	0.20	−0.124	0.12	0.032	0.22	−0.082	0.26
β_{COST}	−0.017 ***	<0.01	−0.017 ***	<0.01	−0.095 ***	<0.01	−0.017 ***	<0.01	−0.055 ***	0.01

	Class allocation (LCM)		*Class allocation* (HLCM)	
	Est.	*p-val.*	Est.	*p-val.*
μ_{02}	−2.280 ***	<0.01	−2.400 ***	<0.01
μ_{12}			0.703 ***	0.01
μ_{22}			0.649 **	0.05
$\lambda_{12,\ Recr}$	−0.702 *	0.06	−0.551	0.21
$\lambda_{22,\ Male}$	0.402	0.27	−0.180	0.70
$\lambda_{32,\ Adult}$	0.325 *	0.08	0.068	0.66
$\lambda_{42,\ Child}$	0.008	0.97	0.436	0.19
$\lambda_{52,\ Educ}$	0.093	0.55	0.214	0.22
$\lambda_{62,\ NGO}$	−0.049	0.97	0.892	0.50

Table A2
Management scenarios based on the share of land use area.

	Status quo	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Cork oak tree	11.59%	11.59%	14.71%	19.81%	36.10%
Heathland and bushes	17.13%	17.13%	17.13%	18.48%	2.09%
Other native tree species	13.09%	13.09%	15.12%	17.03%	29.19%
Tree plantations	15.99%	14.91%	10.83%	2.47%	0.00%
Meadows, gardens and crops	31.00%	29.39%	31.00%	31.00%	23.85%
Vineyard	11.21%	13.90%	11.21%	11.21%	8.78%
Total	100.00%	100.00%	100.00%	100.00%	100.00%

References

Adamowicz, W., Dupont, D., Krupnick, A., Zhang, J., 2011. Valuation of cancer and microbial disease risk reductions in municipal drinking water: an analysis of risk context using multiple valuation methods. J. Environ. Econ. Manag. 61 (2), 213–226.

Bhat, C.R., Dubey, S.K., Nagel, K., 2015. Introducing non-normality of latent psychological constructs in choice modeling with an application to bicyclist route choice. Transp. Res. B 78, 341–363.

Blalock Jr., J.H.M., Costner, H.L., 1969. Multiple indicators and the causal approach to measurement error. Am. J. Sociol. 75, 264–273.

Bun, M.J.G., Harrison, T.D., 2014. OLS and IV Estimation of Regression Models Including Endogenous Interaction Terms. (UvA-Econometrics Discussion Paper; No. 2014/02). University of Amsterdam, Amsterdam.

Chorus, C., Kroesen, M., 2014. On the (im-) possibility of deriving transport policy implications from hybrid choice models. Transp. Policy 36, 217–222.

Costner, H.L., 1969. Theory, deduction and rules of correspondence. Am. J. Sociol. 75, 245–263.

Daly, A., Hess, S., Patruni, B., Potoglou, D., Rohr, C., 2012. Using ordered attitudinal indicators in a latent variable choice model: a study of the impact of security on rail travel behaviour. Transportation 39:267–297. https://doi.org/10.1007/s11116-011-9351-z.

Fernández-Antolín, A., Guevara, C.A., de Lapparent, M., Bierlaire, M., 2016. Correcting for endogeneity due to omitted attitudes: empirical assessment of a modified MIS method using RP mode choice data. J. Choice Model. 20, 1–15.

Greene, W.H., Hensher, D.A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transp. Res. B Methodol. 37:681–698. https://doi.org/10.1016/S0191-2615(02)00046-2.

Guevara, C.A., 2005. Addressing Endogeneity in Residential Location Models Master Thesis. Massachusetts Institute of Technology.

Guevara, C.A., 2010. Endogeneity and Sampling of Alternatives in Spatial Choice Models Doctoral Dissertation. Massachusetts Institute of Technology.

Guevara, C.A., 2015. Critical assessment of five methods to correct for endogeneity in discrete-choice models. Transp. Res. A Policy Pract. 82:240–254. https://doi.org/10.1016/j.tra.2015.10.005.

Guevara, C.A., Ben-Akiva, M., 2006. Endogeneity in residential location choice models. Transp. Res. Rec. 1977, 60–66.

Guevara, C.A., Polanco, D., 2016. Correcting for endogeneity due to omitted attributes in discrete-choice models: the multiple indicator solution. Transportmetrica A 12: 458–478. https://doi.org/10.1080/23249935.2016.1147504.

Guevara, C.A., Tang, Y., Gao, S., 2017. The initial condition problem with complete history dependency in learning models for travel choices. Transp. Res. B Methodol. https://doi.org/10.1016/j.trb.2017.09.006 (Forthcomming).

Guevara, C.A., Tirachini, A., Hurtubia, R., Dekker, T., 2018. Correcting for Endogeneity due to Omitted Crowding in Public Transport Choice Using the Multiple Indicator Solution (MIS) Method. Working Paper. Universidad de Chile.

Hanemann, M.W., August 1984. Welfare evaluations in contingent valuation experiments with discrete responses. Am. J. Agric. Econ. 66, 332–341.

Hansla, A., Gamble, A., Juliusson, A., Gärling, T., 2008. The relationships between awareness of consequences, environmental concern, and value orientations. J. Environ. Psychol. 28:1–9. https://doi.org/10.1016/j.jenvp.2007.08.004.

Helveston, J.P., 2016. Development and Adoption of Plug-in Electric Vehicles in China: Markets, Policy, and Innovation. PhD Thesis. Carnegie Mellon University.

Hess, S., Beharry-Borg, N., 2012. Accounting for latent attitudes in willingness-to-pay studies: the case of coastal water quality improvements in Tobago. Environ. Resour. Econ. 52, 109–131.

Hoyos, D., 2010. The state of the art of environmental valuation with discrete choice experiments. Ecol. Econ. 69:1595–1603. https://doi.org/10.1016/j.ecolecon.2010.04.011.

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. J. For. Econ. 18:329–344. https://doi.org/10.1016/j.jfe.2012.05.002.

Hoyos, D., Mariel, P., Hess, S., 2015. Incorporating environmental attitudes in discrete choice models: an exploration of the utility of the awareness of consequences scale. Sci. Total Environ. 505:1100–1111. https://doi.org/10.1016/j.scitotenv.2014.10.066.

Karaca-Mandic, P., Train, K., 2003. Standard error correction in two-stage estimation with nested samples. Econ. J. 6:401–407. https://doi.org/10.1111/1368-423X.t01-1-00115.

Kløjgaard, M.E., Hess, S., 2014. Understanding the formation and influence of attitudes in patients' treatment choices for lower back pain: testing the benefits of a hybrid choice model approach. Soc. Sci. Med. 114:138–150. https://doi.org/10.1016/j.socscimed.2014.05.058.

Krinsky, I., Robb, A.L., 1986. On approximating the statistical properties of elasticities. Rev. Econ. Stat. 68, 715–719.

Louviere, J., Hensher, D.A., Swait, J., 2000. Stated Choice Methods: Analysis and Applications. Cambridge University Press, Cambridge.

Louviere, J., Train, K., Ben-Akiva, M., Bhat, C., Brownstone, D., Cameron, T.A., Carson, R.T., Deshazo, J.R., Fiebig, D., Greene, W., Hensher, D., Waldman, D., 2005. Recent progress on endogeneity in choice modeling. Mark. Lett. 16:255–265. https://doi.org/10.1007/s11002-005-5890-4.

Lurkin, V., Garrow, L.A., Higgins, M.J., Newman, J.P., Schyns, M., 2017. Accounting for price endogeneity in airline itinerary choice models: an application to continental US markets. Transp. Res. A Policy Pract. 100, 228–246.

McFadden, D.L., 1986. The choice theory approach to marketing research. Mark. Sci. 5, 275–297.

Palma, D., Hess, S., Guevara, C.A., Rizzi, L.I., de Dios Ortúzar, J., Casaubon, G., 2018. When Expensive Is Good: Addressing Price-Quality Association in Choice Modelling. Working Paper. University of Leeds http://www.stephanehess.me.uk/publications.html.

Paulssen, M., Temme, D., Vij, A., Walker, J.L., 2014. Values, attitudes and travel behavior: a hierarchical latent variable mixed logit model of travel mode choice. Transportation 41, 873–888.

Petrin, A., Train, K., 2010. A control function approach to endogeneity in consumer choice models. J. Mark. Res. 47, 3–13.

Poe, G.L., Giraud, K.L., Loomis, J.B., 2005. Computational methods for measuring the difference of empirical distributions. Am. J. Agric. Econ. 87, 353–365.

Rivers, D., Voung, Q.H., 1988. Limited information estimators and exogeneity tests for simultaneous probit models. J. Econ. 39, 347–366.

Ryan, A.M., Spash, C.L., 2012. The awareness of consequences scale: an exploration, empirical analysis, and reinterpretation. J. Appl. Soc. Psychol. 42:2505–2540. https://doi.org/10.1111/j.1559-1816.2012.00951.x.

Small, K., Winston, C., Yan, J., 2005. Uncovering the distribution of motorists' preferences for travel time and reliability. Econometrica 73, 1367–1382.

Small, K., Winston, C., Yan, J., 2006. Differentiated Road Pricing, Express Lanes, and Carpools: Exploiting Heterogeneous Preferences in Policy Design. Brookings–Wharton Papers on Urban Affairs. pp. 53–96.

Snelgar, R.S., 2006. Egoistic, altruistic, and biospheric environmental concerns: measurement and structure. J. Environ. Psychol. 26:87–99. https://doi.org/10.1016/j.jenvp.2006.06.003.

Stern, P.C., 2000. Toward a coherent theory of environmentally significant behavior. J. Soc. Issues 56, 407–424.

Stern, P.C., Dietz, T., Kalof, L., 1993. Value orientations, gender, and environmental concern. Environ. Behav. 25:322–348. https://doi.org/10.1177/0013916593255002.

Stern, P.C., Dietz, T., Guagnano, G.A., 1995. The new ecological paradigm in social-psychological context. J. Environ. Econ. Manag. 26, 271–292.

Thorhauge, M., Rich, J., Cherchi, E., 2015. Departure Time Choice: Modelling Individual Preferences, Intention and Constraints (Doctoral Dissertation). Technical University of Denmark, Department of Transport.

Train, K.E., 1998. Recreation demand models with taste differences over people. Land Econ. 74, 230–239.

Train, K.E., McFadden, D.L., Goett, A.A., 1987. Consumer attitudes and voluntary rate schedules for public utilities. Rev. Econ. Stat. 69, 383–391.

Wen, C.H., Chen, P.H., 2017. Passenger booking timing for low-cost airlines: a continuous logit approach. J. Air Transp. Manag. 64, 91–99.

Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. 2nd Ed. MIT Press, Cambridge, MA.

Zegras, P.C., Li, M., Kilic, T., Lozano-Gracia, N., Ghorpade, A., Tiberti, M., Aguilera, A., Zhao, F., 2018. Assessing the representativeness of a smartphone-based household travel survey in Dar es Salaam, Tanzania. Transportation 1–29.