University of Warsaw
Faculty of Economic Sciences

Working Papers

No. 20/2020 (326)

EX-ANTE AND EX-POST MEASURES TO MITIGATE HYPOTHETICAL BIAS. ARE THEY ALTERNATIVE OR COMPLEMENTARY TOOLS TO INCREASE THE RELIABILITY AND VALIDITY OF DCE ESTIMATES?

Sergio Colombo
Wiktor Budziński
Mikołaj Czajkowski
Klaus Glenk

Warsaw 2020

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES
WORKING PAPERS

Ex-ante and ex-post measures to mitigate hypothetical bias. Are they alternative or complementary tools to increase the reliability and validity of DCE estimates?

Sergio Colombo[a] *, Wiktor Budziński[b], Mikołaj Czajkowski[b], Klaus Glenk[c]

[a] *Department of Agricultural Economics, IFAPA*
[b] *Faculty of Economic Sciences, University of Warsaw*
[c] *Scotland's Rural College, SRUC*
* *Corresponding author: sergio.colombo@juntadeandalucia.es*

Abstract: Hypothetical bias remains at the heart of controversy about the reliability and validity of value estimates from discrete choice experiments (DCEs). This especially applies to environmental valuation, where typically no market value exists for the good under study. This has motivated a large body of literature that investigates possible approaches to test for and mitigate this bias. Our study provides further evidence to this debate by testing whether the use of ex-ante or ex-post mitigation strategies is effective in reducing hypothetical bias in DCEs. Specifically, we use individual and multiple ex-ante reminders along with ex-post data analysis to test whether their individual or joint use improves the quality of the willingness to pay (WTP) estimates. The analysis is carried out with the use of the state-of-the-art mixed logit model, as well as using the innovative semi-parametric logit-mixed-logit, which can capture non-standard heterogeneity distributions. The empirical study focuses on preferences for environmental and social impacts of organic olive production. Comparing three experimental treatments to a control treatment, we test whether cheap talk addressing hypothetical bias, a scale reminder or a combination of both affect stated WTP. In addition, we use ex-post data analysis aimed at correcting WTP estimates. Results show that cheap talk scripts and scale reminders, alone or in conjunction, did not significantly influence the results obtained from a sub-sample with standard budget constraint reminders. The ex-post approach outperforms the ex-ante approach and provides a significant reduction on mean WTP estimates.

Keywords: Choice experiment, Hypothetical bias, mitigation strategies

JEL codes: C1, Q1

Acknowledgements: This research was financed by project PP.AVA.AVA2019.031 funded by Andalusian Institute of Agricultural Economics and Training (IFAPA) and co-funded by FEDER fund. MC gratefully acknowledges the support of the National Science Centre of Poland (Opus, 2017/25/B/HS4/01076). WB gratefully acknowledges the support of the Foundation for Polish Science (FNP) and the National Science Centre of Poland (Preludium, 2016/21/N/HS4/02094). KG gratefully acknowledges funding through the Scottish Government Rural Affairs and the Environment Portfolio Strategic Research Programme 2016–2021 (Theme 1: Natural Assets).

1. Introduction

In the past decade, stated preferences have been the main valuation approach employed to elicit the value of environmental goods (Hanley and Czajkowski, 2019). Notwithstanding, stated preference studies have been widely criticized for often failing validity and reliability tests (Zawojska and Czajkowski 2017, Bishop and Boyle 2019). As an increasingly popular stated preference method, the potential of discrete choice experiments (DCEs) to provide credible information for decision making must still be scrutinized and procedures should be explored that improve the validity of welfare estimates derived from DCEs (Johnston et al., 2017).

The academic community has paid much attention to the issue of hypothetical bias (HB), given the inherent hypothetical nature of the valuation task that has been frequently found to impede that respondents behave as they would in real markets (Lusk and Schroeder, 2004; Vossler et al., 2012; Loomis, 2014). In particular, hypothetical stated preference questions may not be understood as payment consequential by respondents. As a result, respondents may overstate their Willingness To Pay (WTP). Therefore, a large body of HB literature focused on mechanisms to reduce HB, or on explaining its underlying behavioral reasons (Kang et al., 2011; Grebitus et al. 2013; Jacquement et al. 2013), investigated for private (Lusk and Schroeder, 2004; Tonsor and Shupp, 2011; Moser et al., 2014; Doyon et al. 2015; Liebe et al. 2019) and for public goods (Carson and Groves, 2007; Czajkowski et al. 2017; Zawojska et al. 2019). Despite these efforts, mitigating HB remains a concern for stated preference valuation studies (Penn and Hu, 2018; Loomis, 2014; Murphy et al., 2005).

Approaches to reduce HB from stated preference-based valuations can be classified into ex-ante and ex-post mitigation strategies1. Ex-ante approaches aim at reducing HB at the survey design stage by emphasizing the consequences of respondents' choices, for example in terms of additional payments, or by reminding them to behave as they would in a real choice or purchasing situation (e.g., to consider budget restrictions, the existence of substitutes, or to avoid socially desirable responses). Amongst the ex-ante tools often employed, cheap talk scripts (CTS) have been widely used that ask respondents consider responding as if payments were real. Despite the simplicity of CTS, the empirical evidence about their effectiveness is mixed. Some studies found CTS to be

[1] For a complete review of this approaches see Loomis (2011) and Loomis (2014).

successful in mitigating HB (Carlsson et al. 2005; List et al. 2006; Chowdhury et al. 2011; Tonsor and Shupp, 2011, Ladernburg and Olsen, 2014); others observed no effects on HB (Bosworth and Taylor, 2012; Moser et al. 2013; Varela et al., 2014). The use of CTS did not reduce HB in Doyon et al. (2015), who nevertheless suggested other positive effects such as the increasing the level of participation in the market. Several studies found CTS to be effective only for specific groups of people (Aadland and Caplan, 2003; Champ, Moore and Bishop, 2009; Barrage and Lee, 2010; Ami et al., 2011). Multiple reasons underpin these results. Through a meta-analysis on CTS performance, Penn and Hu (2019) recently demonstrated that CTS tends to be more effective for public goods (compared to private goods), and in cases where the initial extent of HB is larger, concluding that failure to detect reductions in HB through the use of CTS may be due to HB being small in magnitude in the first place.

Ex-post approaches address HB at the data analysis stage by means of procedures that screen the data for implausible responses, often based on responses to questions asked after the valuation tasks. These may include respondents' stated maximum WTP for the good in question following a choice experiment (Bush et al. 2009; Colombo et al., 2016) or respondents' certainty about their choice (Champ et al., 2009; Ready et al. 2010; Akter and Bennett, 2013). Results of ex-post approaches generally assume that hypothetical bias exists and that follow-up questions can be used to obtain WTP estimates that better reflect true preferences, although an incorrect calibration of the responses may result in further bias (Beck et al., 2016). An alternative to follow-up questions is the combination of data from revealed preferences studies (if possible) with stated preference data (Azevedo et al., 2003; Brooks and Lusk, 2010), or to use revealed preference data to calibrate stated WTP (Fox et al., 1998).

Simultaneous application of more than one HB mitigation technique may enhance HB reduction. Ladenburg and Olsen (2014) and Varela et al. (2014), for example, combined two ex-ante approaches by augmenting CTS with an Opt-Out reminder. This was, however, only found to be effective in the case of Ladenburg and Olsen (2014). As pointed out by Loomis (2014), ex-ante and ex-post approaches may also be combined in a single application. In a contingent valuation (CV) study, Whitehead and Cherry (2007) concluded that WTP estimates are similar when either ex-ante or ex-post approaches are employed, suggesting that both approaches should be considered as complements rather than substitutes. To our knowledge, there are no previous studies that investigated the joint effect of using both ex-ante and ex-post approaches in DCEs.

This study addresses this research gap by drawing on data from a DCE on the environmental and social impacts of organic olive oil production to investigate the joint effect of ex-ante and ex-post approaches to mitigating HB. For the ex-ante approach, we test for sensitivity of WTP values to different CTS formats. Apart from the typical CTS that informs respondents about a common propensity to exaggerate stated WTP, we also consider a CTS which refers to the relative scale of the proposed project, and test whether the HB is affected by either CTS format. Additionally, we test whether the use of the combination of both CTS reduces stated WTP further. Regarding the ex-post approach, we follow the methodology proposed by Colombo et al. (2016), who reduce HB by allowing respondents to revise those choices in the DCE that were found to be inconsistent with responses to a follow-up question. Owing to multiple experimental treatments, our empirical data has four times the number of observations used by Colombo et al. (2016). Finally, to shed light on whether the mixed results observed in the literature on the performance of various HB mitigation strategies are due to the modelling approach, we use both a standard multinomial mixed (random parameters) logit (MXL; Revelt and Train, 1998) model and the more recent semi-parametric Logit Mixed Logit (LML; Train, 2016) model. The LML is arguably a more flexible approach, which allows estimation of the shape of the preference heterogeneity distribution without imposing restrictive assumptions regarding its parametric specification. This is one of the first applications of the LML model, thus providing further insights regarding its potential superiority over the standard MXL model (Franceschinis et al., 2017; Bazzani et al., 2018; Bansal et al., 2018).

The paper is organized as follows. The subsequent section introduces the study design focusing on the ex-ante and ex-post procedures employed to mitigate HB and provides an overview of the case study. In section 3, we outline the methodology used for data analysis. We then present the results and discuss the implications for decision-making in section 4, finishing the article with a set of conclusions.

2. Study design

An online questionnaire was developed to elicit respondents' preferences towards environmental and social impacts of organic olive growing in the sloping areas of the Andalucía region, South of Spain. In the survey, relevant information regarding the studied good was introduced by means of short and clear pieces of information in order to keep respondents' attention. We employed several graphical illustrations to describe the olive growing production systems and their environmental and social impacts, which constitute the attributes of the CE. We clearly explained that the four different olive growing production systems (marginal, traditional, intensive and super-intensive) are associated with specific environmental and social impacts, and that among the four systems this study focused on marginal olive production. The CE attributes were i) reducing climate change impact, ii) biodiversity, iii) risk of pollution of water resources, iv) soil erosion, v) agricultural employment, and vi) an increase in tax. All the attributes except increase in tax were treated as qualitative and coded as dummy variables2. In all cases, detailed information about the levels each attribute could take was provided to respondents.

Table 1 lists the attributes, their levels and the short labels used to identify them in the remainder of this article. An example of choice card is shown in Figure 1. Each choice task had three alternatives. Two alternatives described impacts of potential changes in olive growing in the area and were associated with an increase in tax. A third status quo alternative was available at no extra cost. A fractional factorial experimental design was determined by minimising D-error for a MNL specification using Bayesian techniques with priors determined from an earlier pilot study.

Table 1: Attributes and levels of the choice experiment

Attribute	Levels	Label
Tackling climate change	Low	TCC1
	Medium	TCC2
	High	TCC3
Biodiversity	**Low**	BD1
	Medium	BD2
	High	BD3
Risk of pollution of water resources	**High**	WP1
	Moderate	WP2
	Low	WP3
Soil erosion	**High**	SE1
	Moderate	SE2
	Low	SE3
Agricultural employment	**0 %**,	AE1
	+5 %	AE2
	+10 %	AE3
Tax	**0**, 2, 7, 14, 23, 35, 51 €/year	TAX

Note: Levels of the current situation are shown in bold.

[2] Further information about the choice experiment design in terms of attribute selection and description and the experimental design can be found in Colombo et al. (2016).

Impact		Current situation	Alternative 1 with organic farming	Alternative 2 with organic farming
	*Tackling climate change*	Neither alternative 1, nor alternative 2 compensate the tax increase. I prefer the current situation	low	medium (−25 % Co_2)
	*Biodiversity (density of animals and plants)*		low	medium (+40 %)
	*Risk of pollution of water resources*		elevated	reduced
	*Soil erosion*		moderate (−22 %)	moderate (−22 %)
	*Increase in agricultural employment*		+ 10 %	+ 10 %
	*Increase in your taxes (€/year)*		14 €	23 €
Which option do you prefer?		☐	☐	☐

Figure 1: Example choice card

We aim to test the effect of HB mitigation approaches on the premise that HB is likely to be present, given the public goods context of the DCE (Penn and Hu, 2019). Our hypothesis is that ex-ante and ex-post approaches are effective in achieving a positive reduction in WTP. Adopting the terminology of Penn and Hu (2019), we interpret a reduction in WTP through the use of ex-ante and ex-post approaches as a reduction in the extent of potential HB. For the experimental treatments associated with different (combinations of) ex-ante devices tested, and their combination with an ex-post approach, our null hypothesis is therefore that (mean) WTP based on these treatments is different from (mean) WTP based on a control treatment in which no HB reduction is undertaken.

In order to test the effect of ex-ante approaches to mitigate HB, four different versions (treatments T1 to T4) of the questionnaire were implemented. In all treatments, we carefully described the choice task and included a typical reminder about respondents' budget constraints and the existence of alternative goods they may prefer to consume. Special effort was dedicated to explain the consequentiality of the study in terms of the tax increase associated with the choice of alternatives. In particular, we informed respondents that the survey was commissioned by the Andalusian government, which is well known to be the authority that holds the local competence in the agricultural policy field and can coercively impose taxes for the purposes under study.

We employed available features of online surveys in order to encourage respondents to think carefully about their choices. These included delaying the availability of advancing in the survey when important information was conveyed, introducing access to pop up windows to explain the choice task and a definition of the attribute levels which were available throughout the completion of the choice tasks. Treatment 1 (T1) served as a control treatment where additional ex-ante HB mitigation strategies were absent. In treatment 2 (T2), respondents received a CTS that explicitly reminded them prior to the choice tasks about the consequences that their choices may have. The CTS script read:

"*Previous research shows that, sometimes, respondents selected an alternative which they would not choose if they had to pay for it in reality. That is to say: they chose an alternative ignoring the associated cost to them, because this cost is not incurred immediately. This type of behavior can lead to erroneous conclusions from the study and may result in the application of a higher tax increase than society is actually willing to pay to see the proposed changes in olive orchards implemented. Therefore, we ask you to only choose an alternative if you are willing to pay the associated increase in your taxes, in exchange for the effects described. Otherwise, simply choose the current situation*."

To help respondent to make the choices consequential, we also reminded that results of the study would be used to feed the forthcoming public support to organic olive growing. Additionally, following Ladenburg and Olsen (2014) and Varela et al. (2014), we included an opt-out budget reminder which reminds respondents about the possibility to opt-out when the cost associated to any alternative is greater than their WTP.

"When choosing the preferred alternative, remember that there are no correct or incorrect responses and keep in mind your budget constraint and that you may prefer to spend your money on other things rather than agriculture that you consider more important (education, health etc.). Hence, if you think that the cost associated with the proposed alternatives is too high, you rather choose the current situation, which has not cost for you."

In treatment 3 (T3), we included a CTS aimed at considering the relative scale of the project. This treatment serves to investigate the likely spatial dimension of HB. Given the hypothetical nature of the project described in the CE, respondents may not focus on the dimension of the proposed changes and express a WTP measure having the entire olive grove area in mind instead of only focusing on the area of marginal olive groves. In this treatment, a pie chart accompanied a textual reminder to convey information about the proportion of marginal olive orchard area affected by the project relative to the total olive orchard area (24%) and the total agricultural area (9%). The pie chart is shown in Figure 2, and the accompanying CTS that reminded respondents that the tax payments associated with each choice card alternative would only provide the environmental and social benefits described in the affected area read as follows:

"*All the changes we show on the cards refer only to the mountainous olive grove that cannot be mechanized and that occupies 24% of the area of the entire Andalusian olive grove, and only 9% of the agricultural land in Andalusia. Please keep in mind that the increase in tax associated with each alternative is only intended to finance policies for the promotion of organic olive growing in the mountainous olive grove areas of Andalusia*."

In treatment 4 (T4), we combined both types of CTS (i.e., those shown in T2 and T3) to test whether there is an effect of joint presentation of CTS.

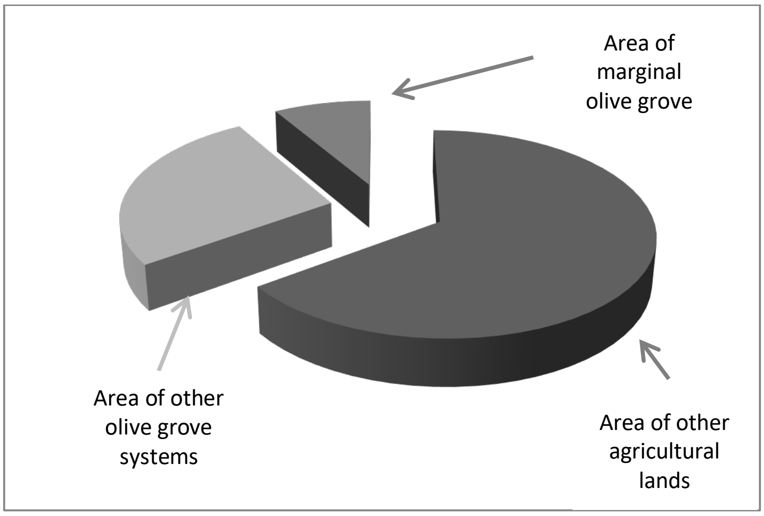


Figure 2: Olive grove area affected by the project in relation to total agricultural area.

Between December 2012 and February 2013, 200 completed surveys were achieved for each of the treatments by a specialised market research company through a random sampling procedure. For each treatment, we recorded the time spent by each respondent for completing the choice tasks and the percentage of inconsistent choices determined using the procedure described below.

After completing the choice tasks, i.e. ex-post, respondents' choices were screened by means of an iterative procedure in line with Colombo et al. (2016). In a follow up question, we asked respondents to state their maximum WTP for what arguably constitutes the best possible outcome according to non-monetary attribute levels. Based on expected monotonic preferences for the non-monetary attributes, this outcome is characterized by the highest level of each non-monetary attribute. The stated maximum WTP was compared to the tax increase of the chosen alternatives to detect whether respondents' choices were inconsistent with maximum WTP stated ex-post. That is, we compared if respondents had chosen an alternative in the choice tasks that represents a worse environmental and social outcome at a higher cost than the stated maximum WTP for the best outcome. In cases where this occurred, we asked respondents to review their decisions, allowing them to revise their choices if they wished doing so. This provided information for an ex-post analysis of choices, in which revised choices replace initially "inconsistent choices". It is important to point out that we did neither prompt nor force respondents to alter their initial choice: respondents could choose to retain their initial choice or to revise it. We stored both initial responses to choice tasks (ex-ante) and responses to choice tasks following potential revision (ex-post), thus enabling an investigation of the effect of ex-post revision on WTP estimates. In the last part of the questionnaire, we gathered respondents' socio-economic data and other information about their current consumption of organic food.

3. Econometric framework

According to random utility theory, the utility respondent i obtains from alternative j at choice occasion t relies on a deterministic term $\boldsymbol{X}_{ijt}\boldsymbol{\beta}_i$, and a random term ε_{ijt} that follows a Gumbel distribution:

$$U_{ijt} = \boldsymbol{X}_{ijt}\boldsymbol{\beta}_i + \varepsilon_{ijt} \tag{1}$$

where $\boldsymbol{X}_{ijt}$ is the vector of k attributes describing alternative j faced by respondent i at time occasion t and $\boldsymbol{\beta}_i$ is the individual-specific vector of k preference parameters. In the mixed logit model (MXL; Revelt and Train, 1998, McFadden and Train, 2000, Train, 2009), elements of $\boldsymbol{\beta}_i$ are modeled as random, following a parametric probability distribution selected *a priori* by the researcher. The MXL appears to be the state-of-practice in the econometric analysis of discrete choice data. In addition, we apply the semi-par ametric logit-mixed logit model (LML; Train, 2016) as an alternative way to model preference heterogeneity.

The MXL model allows accounting for preference or WTP heterogeneity following a particular parametric distribution. The multivariate (parametric) distribution of these parameters in the sample is $\boldsymbol{\beta}_i \sim \boldsymbol{f}(\boldsymbol{b}, \boldsymbol{\Sigma})$, where $\boldsymbol{b}$ is a vector of sample means and $\boldsymbol{\Sigma}$ is a variance-covariance matrix. A convenient way of accounting for preference differences associated with information treatments is $\boldsymbol{\beta}_i \sim \boldsymbol{f}(\boldsymbol{b} + z_i\boldsymbol{\delta}, \boldsymbol{\Sigma})$, where z is a binary indicator for treatment effects and $\boldsymbol{\delta}$ is a vector of its estimated attribute-specific effects.

To facilitate interpretation of the results we specify the model in WTP-space (Train and Weeks, 2005):

$$U_{ij} = \beta_i^m\left(X_{ijt}^m + \boldsymbol{X}_{ijt}^{-m}\boldsymbol{\beta}_i^{-m}\right) + \varepsilon_{ij}, \tag{2}$$

where X_{ij}^m is the monetary attribute with respect to which all marginal rates of substitution (WTP) are expressed, and $\boldsymbol{X}_{ij}^{-m}$ are all other attributes. In this specification, parameter estimates $(\boldsymbol{\beta}_i^{-m})$ can be readily interpreted as marginal WTP for the non-monetary attributes. Here, we can also define $\boldsymbol{\beta}_i^{-m} \sim \boldsymbol{f}(\boldsymbol{b}^{-m} + z_i\boldsymbol{\delta}^{-m}, \boldsymbol{\Sigma}^{-m})$, which conveniently allow us to interpret $\boldsymbol{b}^{-m}$ as a mean WTPs for a base treatment and $\boldsymbol{b}^{-m} + z_i\boldsymbol{\delta}^{-m}$ as a mean WTP for other treatments.

Estimation of the MXL requires calculation of the k -dimensional integral for a likelihood function of individual i:

$$L_i = \int p\left(\boldsymbol{y}_i | \boldsymbol{X}_{ijt}, \beta_i^m, \boldsymbol{\beta}_i^{-m}\right) f(\beta_i^m, \boldsymbol{\beta}_i^{-m} | \Omega) d(\beta_i^m, \boldsymbol{\beta}_i^{-m}) \tag{3}$$

where $f(\beta_i^m, \boldsymbol{\beta}_i^{-m} | \Omega)$ is a density function of random parameters, whose distributions depend on parameters to be estimated, Ω, and $p\left(\boldsymbol{y}_i | \boldsymbol{X}_{ijt}, \beta_i^m, \boldsymbol{\beta}_i^{-m}\right)$ is conditional probability of making choices, $\boldsymbol{y}_i$, given by

$$p\left(\boldsymbol{y}_i | \boldsymbol{X}_{ijt}, \beta_i^m, \boldsymbol{\beta}_i^{-m}\right) = \prod_t \left(\sum_j y_{ijt} \frac{exp\left(\beta_i^m\left(X_{ijt}^m + \boldsymbol{X}_{ijt}^{-m}\boldsymbol{\beta}_i^{-m}\right)\right)}{\sum_l exp\left(\beta_i^m\left(X_{ilt}^m + \boldsymbol{X}_{ilt}^{-m}\boldsymbol{\beta}_i^{-m}\right)\right)}\right). \tag{4}$$

As the analytical formula for integral in (3) is usually not known it has to be approximated. Usually, researchers employ Maximum Simulated Likelihood (MSL) method, in which R random draws from distribution described by $f(\beta_i^m, \boldsymbol{\beta}_i^{-m} | \Omega)$ has to be generated for each individual, and then (3) can be approximated as

$$L_i \approx \frac{1}{R}\sum_r p\left(\boldsymbol{y}_i|\boldsymbol{X}_{ijt}, \beta_{ir}^m, \boldsymbol{\beta}_{ir}^{-m}\right), \tag{5}$$

where additional index r, denotes r -th draw. To make estimation more precise we use 10,000 scrambled Sobol draws (Czajkowski and Budziński, 2015).

The LML model is a novel semi-parametric approach proposed by Train (2016), which allows to estimate the shape of preference heterogeneity distribution without imposing restrictive assumptions regarding its parametric specification. Initial research suggests that it may be a promising new direction in discrete choice modelling, as it allows to recover multimodal and asymmetric distributions (Franceschinis et al. 2017) and recover induced means of respondents' WTP better than the standard MXL model (Bazzani et al., 2018).

The econometric specification of LML is not much different from the one described above for the MXL, although instead of assuming some parametric, continuous distribution for random parameters, such as $f(\beta_i^m, \boldsymbol{\beta}_i^{-m}|\Omega)$, the true (continuous) distribution is approximated using a discrete distribution. Specifically, we assume that k -th random parameter lies within some interval, $[LOW_k, UP_k]$, and we choose N points dividing this interval into $N-1$ smaller intervals of equal lengths. This creates a grid of N^K vectors of parameter values. The likelihood function is then given by:

$$L_i = \sum_{n=1}^{N^K} W(\beta_{in}^m, \boldsymbol{\beta}_{in}^{-m}|\boldsymbol{\alpha}, \boldsymbol{\lambda}) p\left(\boldsymbol{y}_i|\boldsymbol{X}_{ijt}, \beta_{in}^m, \boldsymbol{\beta}_{in}^{-m}\right), \tag{6}$$

where $W(\beta_{in}^m, \boldsymbol{\beta}_{in}^{-m}|\boldsymbol{\alpha}, \boldsymbol{\lambda})$ is the probability of the vector of parameters values,$\beta_{in}^m, \boldsymbol{\beta}_{in}^{-m}$, which depends on parameters to be estimated, $\boldsymbol{\alpha}, \boldsymbol{\lambda}$. The formula for the choice probability is given by a standard multinomial logit

$$W(\beta_{in}^m, \boldsymbol{\beta}_{in}^{-m}|\boldsymbol{\alpha}, \boldsymbol{\lambda}) = \frac{exp(\boldsymbol{Z}(\beta_{in}^m, \boldsymbol{\beta}_{in}^{-m})\boldsymbol{\alpha} + \boldsymbol{V}(\beta_{in}^m, \boldsymbol{\beta}_{in}^{-m}, \boldsymbol{T}_i)\boldsymbol{\lambda})}{\sum_d exp(\boldsymbol{Z}(\beta_{id}^m, \boldsymbol{\beta}_{id}^{-m})\boldsymbol{\alpha} + \boldsymbol{V}(\beta_{id}^m, \boldsymbol{\beta}_{id}^{-m}, \boldsymbol{T}_i)\boldsymbol{\lambda})}. \tag{7}$$

$\boldsymbol{Z}(\beta_{in}^m, \boldsymbol{\beta}_{in}^{-m})$ in (7) denotes some flexible transformation of values of the random parameters vector. The transformations we consider here are Legendre polynomials, step functions and four versions of splines (linear spline, cubic spline, piecewise cubic spline, and piecewise cubic Hermite interpolating spline). To incorporate correlations between random parameters we included first-order interactions between elements of vectors $(\beta_{in}^m, \boldsymbol{\beta}_{in}^{-m})$.

$\boldsymbol{V}(\beta_{in}^m, \boldsymbol{\beta}_{in}^{-m}, \boldsymbol{T}_i)$ in (7) denotes some transformation of values of the random parameters vector and individual-specific treatment, T_i. Incorporation of additional individual-specific explanatory variables into LML framework did not attract much attention yet. Approaches which were considered to date include incorporating the additional interaction between socio-demographic/treatment variable and attribute directly into the utility function (Bansal et al., 2017) or estimating a separate model for each value of socio-demographic/treatment variable (Caputo et al., 2017). We consider both approaches to be suboptimal, as the former requires fixing the interaction parameter in the utility function, which reportedly makes estimation 20-40 times longer, and the latter may be infeasible if socio-demographic/treatment variables take multiple values. Here we incorporate socio-demographic/treatment variables as an interaction with values of random parameters, namely $\boldsymbol{V}(\beta_{in}^{m}, \boldsymbol{\beta}_{in}^{-m}, \boldsymbol{T}_i) = (\boldsymbol{T}_i\beta_{in}^{m}, \boldsymbol{T}_i\boldsymbol{\beta}_{in}^{-m})$. Obviously, more complex functions can be defined, such as interaction with polynomials of $\beta_{in}^{m}, \boldsymbol{\beta}_{in}^{-m}$ of a higher order than one, but this also requires estimation of a greater number of coefficients.

Similarly to the MXL case, estimation of the model can be performed using the Maximum Simulated Likelihood method. We use R random draws from the grid (each point is drawn with the same probability) for each individual and use them to approximate the likelihood function:

$$L_i \approx \sum_{r=1}^{R} W(\beta_{ir}^{m}, \boldsymbol{\beta}_{ir}^{-m}|\boldsymbol{\alpha}, \boldsymbol{\lambda}) p\left(\boldsymbol{y}_i|\boldsymbol{X}_{ijt}, \beta_{ir}^{m}, \boldsymbol{\beta}_{ir}^{-m}\right). \tag{8}$$

As in the case of MXL, we use scrambled Sobol draws to make estimation more efficient. Approximation in (8) can be used to calculate mean WTP, as a sum of R random draws from the grid, weighted by the estimated probability that they will occur in the population, $W(\beta_{ir}^{m}, \boldsymbol{\beta}_{ir}^{-m}|\boldsymbol{\alpha}, \boldsymbol{\lambda})$:

$$MWTP \approx \sum_{r=1}^{R} W(\beta_{ir}^{m}, \boldsymbol{\beta}_{ir}^{-m}|\boldsymbol{\alpha}, \boldsymbol{\lambda})\, \boldsymbol{\beta}_{ir}^{-m}. \tag{9}$$

Working with the LML model requires choosing an appropriate specification – there are multiple options available for the specification of $\boldsymbol{Z}(\beta_{in}^{m}, \boldsymbol{\beta}_{in}^{-m})$ function. Most of existing studies used information criteria to guide the specification choice. We also employ this approach.[3] Finally, the specification of the model requires selecting the values of the random distribution bounds. Most studies use the estimates from the MXL

[3] Similar to Bazzani et al. (2018), we find that the LML does not provide an improvement over the MXL model if the Bayesian Information Criterion (BIC) is used as a basis for comparison. Bansal et al. (forthcoming) note that this is most likely in small samples and recommend considering significant changes in histograms of parameters distribution and choosing minimal number of parameters so that any additional parameters would not change shape of the distribution substantially. We found this approach hard to implement in practice, as it was difficult to assess whether observed changes in shape should be considered significant. Additionally, following this approach would likely exponentially increase the number of model specifications to consider. As a result, in what follows we used the Akaike Information Criterion (AIC) for comparing the fit of different specifications.

model with bounds defined as two standard deviations above and below the mean. This was the approach used first by Train (2016) and, as far as we know, only Caputo et al. (2017) experimented with different settings by taking three standard deviations or extending parameters bounds based on visual inspection. The results they obtained are mixed – for some specifications of $\boldsymbol{Z}(\beta_{in}^{m}, \boldsymbol{\beta}_{in}^{-m})$ extending parameters bounds increases the model fit, but for other specifications it decreases the model fit. We also test for sensitivity of model performance depending on the range of bounds, which are defined with reference to parameters of an MXL model without correlation.

4. Results

The overall sample is representative for Andalusia population with respect to gender (chi squared=0.12; p-value=0.73). However, it differs in terms of age, education and income, being composed by younger (chi squared=2399,0; p-value=0.00) and higher educated citizens (chi squared=1125.5; p-value=0.00). Relative to income, the sample is formed by a representative number of low income people but not representative in terms of medium and high income respondents, who are over and underrepresented respectively (chi squared=52.5; p-value=0.00). These differences are maintained across the four treatments, which are all different relative to the general population with respect to the dimensions described above. That is, there are no significant differences in the above-mentioned socio-economic characteristics across treatments. Treatments do not affect the response time related to choosing the preferred option in choice tasks. In all treatments, respondents became more rapid in completing the choice tasks as they move through the sequence of choices, as observed previously in other studies (Bonsall and Lythgoe, 2009; Carlsson et al., 2012). The decline in completion times is particularly pronounced between the first and the second choice task. Treatment did not affect the percentage of "inconsistent" choices, with values close to 22% for ex-ante choices and 12% for ex-post choices.

Table 2 presents respondents' estimated WTP for the attribute levels elicited for each of the four CTS treatments using ex-ante data (the initial choices, before respondents were given a chance to revise inconsistencies with the follow up maximum WTP question). The models assume that all attributes were dummy-coded and that their parameters are normally distributed, with the exception of the (negative) cost parameter, which was assumed log-normally distributed. The left panel of Table 2 presents the mean WTP estimates based on the MXL model. The right panel of Table 2 presents the equivalent results based on the best fitting LML model. Results presented in Table 2 refer to models that combine observations from all treatments ('All treatments jointly') or allow for treatment specific WTPs. To find the best fitting LML model specification, we first applied a grid-search procedure to examine the sensitivity of the estimated log-likelihood at convergence to the specification of the parameter bounds. We found that using bounds specified as 1.5 to 2.5 MXL-based standard deviations below and above the mean resulted in relatively similar results, with the optimum identified at 1.8. This lends support to using the rule of thumb of approximately 2 standard deviations below and above the mean (Train, 2016). However, it must be noted that this approach did not differentiate the bounds for each parameter, did not implement asymmetric bounds, or generally other reference points for determining bounds (i.e., not based on the results of the MXL model without correlations). The results for the effects of selecting different bound ranges are provided in Appendix A. Next, we compared the performance of various LML model specifications (asymptotic normal, polynomial, step function, and four types of spline function of the degree 2 to 10) with and without correlated parameters. The comparison of fit of various LML model specifications are available in Appendix B. Based on Aikake Information Criterion (AIC), we selected the 8-knot piece-wise cubic spline as the best fitting specification of the LML model.[4] It is worth to say that for the models with correlated parameters, based on Bayesian Information Criterion, which is more restrictive in terms of penalizing models for the number of parameters, none of the LML models with correlated outperformed the MXL specification in terms of model fit.

The first thing to note about the results presented in Table 2 is that there are some differences between mean WTP estimates implied by the MXL and the LML model. These differences are within approximately 10% of the WTP derived from the MXL model and are not statistically different according to z-tests with the exception of the Status quo, for which WTP resulting from the LML model was 40% lower. As an aside, we found that the status quo parameter estimate was relatively the least stable across various LML model specifications. For the rest of parameter estimates, LML standard errors are significantly lower than MXL indicating a higher accuracy of parameter estimates in this model.

[4] The models presented here were estimated using a DCE package developed in Matlab and available at https://github.com/czaj/DCE. The code and data for estimating the specific models presented in this study, as well as full and supplementary results, are available from http://czaj.org/research/supplementary-materials.

Looking at the differences in WTP associated with treatments, we find that in almost all of the cases, estimates of mean WTP resulting from treatments 2, 3 and 4 are not significantly different from WTP estimates inferred from the control treatment (T1). In the case of the MXL model, the only significant difference is observed for a 10% increase in agricultural employment for T2 and T3. Irrespective of statistical significance of differences, there is no consistent trend of lower mean WTP estimates arising from single or joint treatment with CTSs. For LML model results, significant differences are observed in more cases than for MXL model results. This is primarily due to the lower standard errors of parameter estimates of the LML model relative to the MXL model. Significant differences represent either an increase (e.g., Soil erosion – high and T3) or a decrease (e.g., Biodiversity – high: T2 and T3) in the attribute levels' WTP estimates.

Table 2: Mean WTP for policy attributes estimated using joint and treatment-specific data and the MXL and LML models using *ex ante* data [EUR / year]

	MXL					**LML**				
	All treatments jointly	Control: Standard remainders	Treatment 2: HB cheap talk	Treatment 3: Scale of the project reminder	Treatment 4: HB cheap talk and scale of the project reminder	All treatments jointly	Control: Standard remainders	Treatment 2: HB cheap talk	Treatment 3: Scale of the project reminder	Treatment 4: HB cheap talk and scale of the project reminder
Status quo (alternative specific constant)	31.90*** (3.55)	27.60 (4.90)	27.63 (4.96)	37.73 (5.67)	37.12 (6.54)	18.74*** (1.32)	18.30 (2.36)	16.53 (2.35)	23.13* (2.28)	20.72 (2.07)
Tackling climate change – medium (vs. low)	15.22*** (1.60)	16.34 (2.36)	11.35 (2.22)	16.03 (2.88)	17.71 (2.62)	13.64*** (0.12)	13.55 (0.21)	13.78 (0.22)	13.30 (0.20)	14.14** (0.21)
Tackling climate change – high (vs. low)	19.88*** (1.70)	18.82 (2.50)	16.91 (2.26)	21.45 (2.77)	21.71 (2.64)	17.59*** (0.37)	17.87 (0.65)	16.84 (0.56)	17.91 (0.57)	17.41 (0.63)
Biodiversity – medium (vs. low)	15.16*** (1.57)	17.77 (2.56)	14.82 (2.33)	15.46 (2.97)	12.14 (2.78)	13.15*** (0.28)	13.16 (0.47)	14.19 (0.53)	13.64 (0.48)	13.11 (0.62)
Biodiversity – high (vs. low)	21.10*** (1.71)	23.14 (2.57)	19.27 (2.37)	21.10 (2.79)	20.76 (2.69)	18.98*** (0.41)	19.92 (0.64)	18.22* (0.78)	18.03** (0.69)	19.68 (0.80)
Risk of pollution of water resources – moderate (vs. high)	18.67*** (1.62)	17.08 (2.58)	17.40 (2.33)	23.25 (3.01)	18.40 (2.75)	16.04*** (0.08)	16.09 (0.15)	16.05 (0.15)	16.20 (0.17)	15.99 (0.16)
Risk of pollution of water resources – low (vs. high)	28.10*** (2.09)	28.19 (2.94)	24.56 (2.91)	32.51 (3.55)	27.86 (3.26)	26.29*** (0.22)	26.54 (0.39)	26.25 (0.38)	26.32 (0.39)	25.92 (0.40)
Soil erosion – moderate (vs. high)	12.05*** (1.57)	10.32 (2.49)	11.35 (2.44)	13.20 (2.79)	13.50 (2.67)	10.89*** (0.02)	10.85 (0.04)	10.89 (0.03)	10.89 (0.04)	10.93* (0.03)
Soil erosion – low (vs. high)	17.84*** (1.57)	16.79 (2.52)	17.79 (2.21)	19.71 (2.74)	17.53 (2.64)	15.28*** (0.22)	15.12 (0.41)	15.45 (0.47)	16.34** (0.46)	14.61 (0.36)
Agricultural employment – 5% increase (vs. no change)	15.50*** (1.63)	17.27 (2.76)	13.56 (2.29)	14.68 (2.96)	17.61 (2.72)	15.80*** (0.03)	15.83 (0.05)	15.75 (0.05)	15.80 (0.05)	15.81 (0.05)
Agricultural employment – 10% increase (vs. no change)	27.24*** (1.98)	32.56 (3.46)	21.67** (2.84)	24.26* (3.41)	30.81 (3.26)	24.42*** (0.75)	24.82 (1.17)	24.05 (1.24)	22.51 (1.19)	25.64 (1.11)

Notes: *, **, *** indicate statistical significance (Wald test) of the difference of WTP between control treatment (standard remainders) and the other treatments (additional measures aimed at reducing hypothetical bias) at the 0.1, 0.05 and 0.01 level, respectively. For the 'all treatments jointly', the asterisks indicate statistical significance with respect to 0.

The analysis of WTP after respondents could review their “inconsistent” choices (ex-post) provides similar results. The results, presented in Table 3, show that in most cases the additional cheap talk and scale reminders used in T2–T4 did not result in statistically different WTP estimates for the MXL model, relative to T1 that only used standard budget constraint reminders. Again, a greater number of significant differences is found for LML model results, but there is no uniform trend of lower WTP estimates for T2–T4 relative to T1 across all attributes. Furthermore, the share of inconsistent choices is not statistically different between any of the four treatments, revealing that the use of CTS does not affect respondents’ degree of choice inconsistency. Overall, we conclude that the inclusion of cheap talk scripts and scale reminders had a limited effect on WTP estimates, neither for respondents’ initial choices nor for their revised choices.

Table 3: Mean WTP for policy attributes estimated using joint and treatment-specific data and the MXL and LML models using *ex post* data [EUR / year]

	MXL					**LML**				
	All treatments jointly	Treatment 1 Standard remainders – control	Treatment 2 HB cheap talk	Treatment 3 Scale of the project reminder	Treatment 4 HB cheap talk and scale of the project reminder	All treatments jointly	Treatment 1 Standard remainders – control	Treatment 2 HB cheap talk	Treatment 3 Scale of the project reminder	Treatment 4 HB cheap talk and scale of the project reminder
Status quo (alternative specific constant)	15.11*** (2.09)	16.93 (3.40)	11.88 (3.26)	14.11 (3.70)	17.06 (3.30)	18.45*** (1.89)	15.23 (2.44)	10.29* (1.92)	16.07 (1.97)	17.37 (2.27)
Tackling climate change – medium (vs. low)	8.80*** (0.99)	10.88 (1.80)	8.12 (1.69)	6.88 (2.01)	10.75 (1.76)	8.04*** (0.31)	8.74 (0.36)	8.61 (0.41)	7.86* (0.34)	9.16 (0.27)
Tackling climate change – high (vs. low)	11.64*** (1.11)	12.82 (1.89)	11.06 (1.73)	11.96 (1.97)	11.83 (1.73)	10.08*** (0.62)	12.82 (0.79)	10.25*** (0.65)	10.92** (0.68)	10.65** (0.62)
Biodiversity – medium (vs. low)	9.33*** (1.06)	11.44 (1.83)	9.27 (1.66)	8.89 (1.96)	7.72 (1.76)	12.33*** (0.28)	12.26 (0.43)	9.79*** (0.45)	11.89 (0.30)	11.00** (0.37)
Biodiversity – high (vs. low)	11.97*** (1.13)	13.01 (1.82)	11.37 (1.67)	11.01 (2.01)	11.98 (1.78)	13.89*** (0.74)	11.98 (0.71)	11.10 (0.94)	13.88* (0.74)	14.33** (0.74)
Risk of pollution of water resources – moderate (vs. high)	11.78*** (1.24)	13.60 (1.87)	11.22 (1.82)	12.40 (2.10)	10.17 (1.89)	10.56*** (0.06)	10.73 (0.07)	10.63 (0.06)	10.64 (0.08)	10.82 (0.05)
Risk of pollution of water resources – low (vs. high)	17.80*** (1.36)	20.60 (2.15)	17.74 (2.13)	19.51 (2.46)	13.96** (2.09)	18.51*** (0.51)	18.94 (0.70)	19.11 (0.58)	19.16 (0.53)	18.67 (0.51)
Soil erosion – moderate (vs. high)	6.51*** (1.16)	7.46 (1.83)	5.53 (1.79)	4.89 (1.99)	6.76 (1.84)	5.64*** (0.26)	5.37 (0.19)	5.29 (0.20)	6.02*** (0.17)	6.07*** (0.16)
Soil erosion – low (vs. high)	9.81*** (1.10)	10.63 (1.79)	10.39 (1.63)	7.74 (1.98)	9.55 (1.78)	8.88*** (0.47)	7.78 (0.57)	8.84 (0.50)	8.88 (0.43)	10.11*** (0.38)
Agricultural employment – 5% increase (vs. no change)	9.98*** (1.15)	12.29 (1.84)	9.22 (1.74)	6.88** (2.04)	10.41 (1.81)	9.13*** (0.00)	9.13 (0.01)	9.13 (0.00)	9.14 (0.00)	9.12** (0.00)
Agricultural employment – 10% increase (vs. no change)	14.76*** (1.48)	18.14 (2.35)	13.52 (2.29)	9.32*** (2.42)	17.35 (2.39)	11.69*** (1.08)	14.26 (1.27)	10.53** (1.28)	10.71** (1.17)	12.05 (1.17)

Notes: *, **, *** indicate statistical significance (Wald test) of the difference of WTP between control treatment (standard remainders) and the other treatments (additional measures aimed at reducing hypothetical bias) at the 0.1, 0.05 and 0.01 level, respectively. For the 'all treatments jointly', the asterisks indicate statistical significance with respect to 0.

However, we do find that allowing respondents to revise their choices leads to significant reductions of WTP. This result is illustrated in Figure 3, which presents estimates of mean WTP resulting from MXL and LML models for ex-ante (unrevised) and ex-post (revised) data along with the 95% confidence interval. According to z-tests, mean WTP estimates for all attributes are statistically different at the 5 % significance level for MXL and LML models apart from the ASC and the medium level of the biodiversity attribute in the LML model. Allowing respondents to reconsider their choices leads, on average across all attributes, to a WTP decrease of 43% (MXL model) or 33% (LML model). This effect is considerable given that respondents were neither prompted nor forced to change their initial choices in the revision process. The percentage of "inconsistent" choices dropped from 21.8% before the revision to 12.5% after the revision, indicating that a significant part of the sample opted to retain their initial choices. Colombo et al. (2016) proposed different explanations for the confirmation of "inconsistent" choices. In line with the soft cutoffs approach (Swait, 2001; Bush et al. 2009), it is possible that individuals deliberately violate their price cutoffs (maximum WTP), because the disutility of the violation is lower than the disutility of choosing the second-best option. Another reason is that respondents' preferences are vague and affected by a degree of uncertainty or fuzziness. In this case, the maximum WTP should not be considered as a fixed amount, but as a distribution with a variance that is proportional to the degree of fuzziness or uncertainty. Carson et al. (2012) observed that in this case individual choice in a public good context can be expected to diverge significantly from what standard utility theory predicts if preferences are well defined. Finally, the two elicitation methods employed may have disclosed different values. Roe et al. (1996) and Salensminde (2003) observed that DCE, compared to open ended CV, tends to capture respondents' relative valuation rather than their absolute valuation, which is more in line with their budget constraints. As such, it should be expected that the values obtained from open ended CV would be lower than those derived through DCE.

Whatever the reason may be, our results clearly indicate that the ex-post treatment of the choices, which in this case means asking individuals to review their choices if implied WTP exceeds their (stated) WTP for a 'full' improvement program, has a considerable effect on WTP estimates, potentially reducing HB.

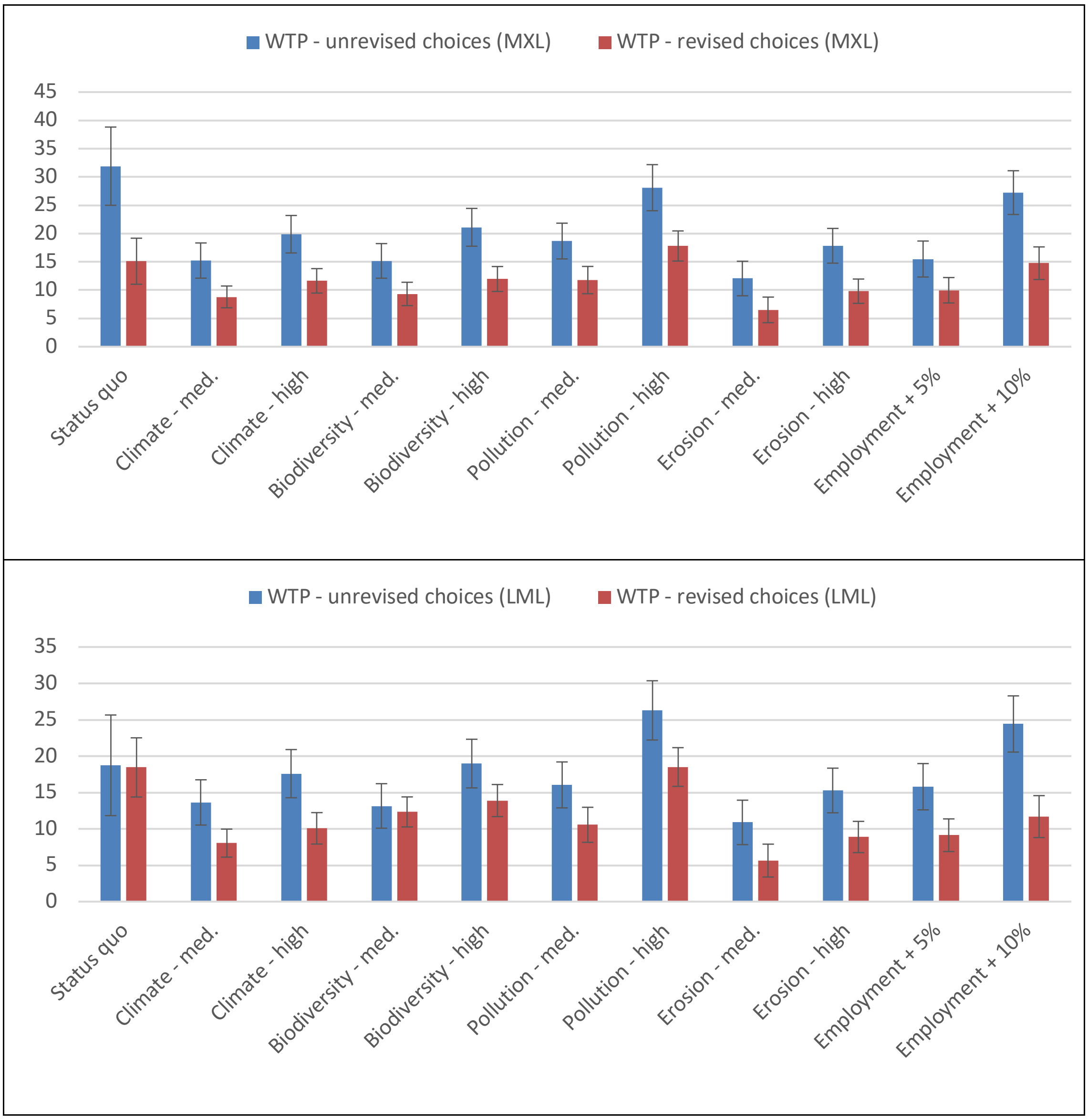


Figure 3. Comparison of WTP means between the estimated models

5. Discussion and conclusions

The hypothetical nature of CE lies at the heart of the controversies about to the reliability and validity of willingness to pay (WTP) estimates for non-marketed goods. Because observing the "true" value of environmental goods and services is impossible, it is important to consider the inclusion of measures aimed at testing and reducing hypothetical bias (HB). Results of this paper confirm this. In our data, HB potentially exists with significant implications for WTP estimates. Therefore, it is important to include devices in the survey to reduce HB.

The use of cheap talk scripts (CTS) as an ex-ante tool to reduce HB has a limited effect on the WTP estimates for the environmental goods valued in this study. This is observed despite the CTS was shown for 30 seconds on the respondents' screen without the option to proceed and despite CTS were augmented with a single opt-out reminder.

There may be several reasons explaining this result. First, there may have been a lack of perceived consequentiality in the online surveys. This observation is in line with Penn and Hu (2019), who find that CTS tends to be relatively ineffective for internet-based surveys compared to mail surveys. Despite efforts to inform respondents that the results of the study will be used for tailoring future agri-environmental policies and that the opinion of the respondent matters for the development of these policies, we cannot be certain that individuals have indeed understood believed this information. Unfortunately, we have not collected ex-post opinions about perceived consequentiality to obtain indication if this was the case or not. We acknowledge that it may not be sufficient to remind respondents about consequentiality, and that it is necessary to expend greater effort to test whether and how consequentiality with respect to payment and policy has been perceived (Barbier et al., 2017, Zawojska et al., 2018). Second, the definition of the cost vector may have included tax amounts that were too low to choke off demand. The values were based on in person qualitative pretests (focus groups), which may differ from on-line settings. Defining the levels of the price vector is still a pending issue in discrete choice experiments (DCE), especially considering the importance it has on WTP estimates (Glenk et al., 2019). Finally, the CTS effect may have been diluted along the survey as suggested by Ladenburg and Olsen (2014).

The CTS related to the scale of the project (T3) also proved to be relatively ineffective. This is remarkable given that respondents evaluated the attribute level intensity as expected, thus passing an internal scope test (i.e., WTP is found to be higher for greater benefits in all attributes). However, as pointed out by Rolfe and Wang (2011), scale and scope are often intertwined issues that are not always appropriately separated by respondents. The ineffectiveness of a scale-related CTS while internal scope is demonstrated may also be attributed to respondents expressing relative values that result in internally consistent choices, but that may appear inconsistent if compared across treatments. This is in line with the idea of coherent arbitrariness (Ariely et al., 2003). Proximity and loyalty effects can be also present for a locally iconic crop such as olives (Granado et. al., 2020). Respondents may have higher values for goods provided at greater proximity (local goods) than good provided further away, especially for environmental goods that may inspire a sense of identification and intimacy to people (Faccioli et al., 2020). Thus, independently of the scale of the good, a good that elicits an emotional response may be highly valued (LaRiviere et al., 2014). Alternatively, respondents may be simply be insensitive to the scale as previously observed by other authors (Rolfe and Windle, 2003).

Finally, the joint presentation of reminders (T4) proved ineffective. Therefore, using multiple CTS may just add complexity to the survey, without providing additional benefits. This outcome is in line with Varela et al. (2014), who find that the use of multiple ex-ante mitigation strategies does not impact WTP estimates. At the same time, it contrasts with results from other studies, who find that the use multiple ex-ante mitigation strategies can reduce HB (Jacquemet et al., 2013, Lademburg and Olsen, 2014). All of these studies test different ex-ante measures. Varela et al. (2014) employed CTS plus a single opt-out reminder; Ladenburg and Olsen (2014) compared the effect of CTS with multiple opt-out reminders; Jacquemet et al. (2013) joined CTS and solemn oath. In our paper, we employ two different CTS and a single opt-out reminder. In summary, the effect of CTS seems to depend on the way it is designed in each specific condition.

The lack of CTS to reduce HB also contrast with previous results of Penn and Hu (2019), who found that CTS work better if actual HB is large. Previous studies may therefore fail to identify a reduction in HB through CTS, because actual HB is small in magnitude rather than because CTS are ineffective. In our case, the ex-post WTP estimates are significantly lower than the initial estimates, suggesting that HB exists and was considerable in magnitude. Therefore, in our case CTS likely failed to reduce HB for reasons other than the actual extent of HB present in the data.

Approaches that revise respondents' choice after the choice task (ex-post) appear better positioned to reduce HB. Estimates of WTP decrease on average by 43% in the case of MXL model and 33% in the LML model. This effect is statistically significant and reveals the quantitative importance of HB in the estimated WTP values, pointing to the potential for substantial error in the design of public policies of non-marketed good if mitigation measures to reduce HB are not adopted. Thus far, ex-post approaches to mitigate HB are far less widespread than ex-ante approaches. Results of this study clearly reveal that research on ex-post mitigation instruments, at least in public good contexts, deserve greater attention.

The results also beg the question why ex-post mitigation measures outperformed ex-ante mitigation measures, a result which has also been observed in previous research

(Penn and Hu, 2018). In our opinion, there are several reasons to consider. First, ex-ante approaches do not reveal any information about respondents' attention to and understanding of the information provided, unless specifically inquired. Thus, the use of CTS should be accompanied by questions that allow the analyst to appreciate respondents' understanding of the CTS. This assessment would, however, increase the length of the questionnaire, and represents a challenge for future research about how to reliably gather this information. Second, answering a choice experiment is an unfamiliar task for most respondents involving the realization of many trade-offs, often regarding an unfamiliar good, which may easily induce errors. Here, it is important to consider that respondents often face a DCE for the first time, and thus it is more difficult for them to fully comprehend information about something they have not experienced yet. Because of that, information provided through ex-ante instruments may be misleading and ineffective. In this context, ex-post mitigation gives respondents a "second chance" to rate or rethink how they performed in the task, thus allowing them to scrutinize their initial choices. This is an advantage given that it places respondents in a better situation to notice possible issues that may have arisen while completing the choice tasks. Also, in the specific ex-post mechanism employed in this article, respondents may be driven by several intrinsic or extrinsic motivations guiding their revisions such as moral commitments, morality or fairness (Hollander-Blumoff, 2011). Finally, ex-post analysis is an instrument for the analyst to determine the quality of respondents' choices. By obtaining information about respondents' "understanding and confidence" regarding their choices, the analyst can apply different analytical and methodological tools to either correct, weight or even exclude unreliable responses.

Our study is one of the first applications of the LML model. We find that our conclusions are robust to allowing for more flexible model specifications than the standard MXL model. Despite differences between WTP estimates derived from the two models are overall not statistically significant, we note that LML-based estimates are associated with lower standard errors than their MXL counterparts. At the same time, our experiences with the new model call for caution. Despite relatively quick estimation, the model had to be estimated multiple times to investigate its stability and assure correct specification and convergence. These limitations and the critical influence of arbitrary decisions in the estimation process deserves future research before the LML model can become state-of-practice for discrete choice models. Furthermore, under the BIC decision criterion the statistical performance of the LML models is inferior to the MXL, revealing that the latter should be preferred when a larger penalization for intricate parametrization is assumed.

We acknowledge several limitations in our study that should be investigated in future research. First, as discussed above in the context of ineffectiveness of CTSs in our study, the absence of follow-up questions on consequentiality and the use of CTS limits our understanding of how these were perceived by respondents. Such questions should be considered in future surveys.

Second, in our experiment we did not include a “real” treatment that may be used as a benchmark for disclosing the respondents’ true preferences. In this sense, we cannot be sure about the value of the true unbiased WTP values. We therefore had to resort to assuming that lower WTP estimates are closer to the unbiased value. Although some authors have used “non-hypothetical” treatments, we argue that it is never possible to mimic the real consumers’ behavior. For instance, Loomis et al. (2009) and Chowdhury et al. (2011) provided respondents with an initial monetary endowment that they had to spend according to the choices made. Alemu and Olsen (2018) gave a lump sum to respondents prior to the CE and informed them that they were welcome to keep the money for their own use later on. In this context, evidence suggest that respondents’ behavior may differ on whether wealth is “windfall” or “earned” (Cherry et al., 2002). To our knowledge, only Moser et al. (2013) and Liebe et al. (2019) carried out a field experiment where respondents used their own money, with Liebe et al. (2019) implementing the experiment in an online format. However, this was possible because both studies analyzed preferences for existing consumer (market) goods of relatively low value. Obviously, it would not be possible in the case of non-market goods. Furthermore, in the case of public goods, where an increase in taxes is used as payment, providing a monetary incentive to respondents would equate to some form of a “tax-rebate”, which is implausible and may create distrust in respondents.

Third, we did not investigate the reasons why respondents decided to retain their initial choices in the revision process. Colombo et al. (2016) showed that by removing these “inconsistent” choices from the sample, WTP estimates were further reduced by 18%. Future research should therefore investigate the reasons underlying this behavior and determine which observations, if any, should be retained in the analysis.

To conclude, we affirm that hypothetical bias may significantly affect WTP estimates, especially if they relate to unfamiliar goods. Ex-ante and ex-post mitigation strategies have been of varying effectiveness in the stated preference literature.

Irrespective of the approach used to reduce hypothetical bias, our findings demonstrate a need to extend the research effort beyond employing ex-ante scripts in experimental tests of their effectiveness and gathering ex-post information to investigate their potential to assist with WTP adjustments. Specifically, there is a need to develop a common understanding of ex-ante and ex-post instruments across respondents, thus allowing for more rigorous tests of their effectiveness based on theoretical expectations.

References

Aadland, D., Caplan, A.J., 2003. Willingness to pay for curbside recycling with detection and mitigation of hypothetical bias. American Journal of Agricultural Economics 85(2):492–502.

Akter, S., and Bennett. J., 2013. Preference uncertainty in stated preference studies: facts and artefacts. Applied Economics 45:2107-2115.

Alemu, M.H., Olsen, S., 2018. Linking Consumers' Food Choice Motives to their Preferences for Insect-based Food Products: An Application of Integrated Choice and Latent Variable Model in an African Context. Journal of Agricultural Economics 70(1):241-258.

Ami, D., Aprahamian, F., Chanel, O. and Luchini, S., 2011. A Test of Cheap Talk in Different Hypothetical Contexts: The Case of Air Pollution. Environmental and Resource Economics 50:111-130.

Ariely, D., Loewenstein, G. and Prelec D., 2003. Coherent arbitrariness: stable demand curves without stable preferences. Quarterly Journal of Economics 118 (1):73-105

Azevedo, C. D., Herriges, J. A. and Kling, C. L., 2003. Combining Revealed and Stated Preferences: Consistency Tests and Their Interpretations. American Journal of Agricultural Economics 85:525-537.

Bansal, P., Daziano, R. A., and Achtnicht, M., 2017. Extending the logit-mixed logit model for a combination of random and fixed parameters. Journal of Choice Modelling 27:88-96.

Bansal, P., Daziano, R. A., and Achtnicht, M., 2018. Comparison of parametric and semiparametric representations of unobserved preference heterogeneity in logit models. Journal of Choice Modelling 27:97-113.

Barbier, E. B., M. Czajkowski, and N. Hanley., 2017. Is the income elasticity of the willingness to pay for pollution control constant? Environmental and Resource Economics 68:663–682.

Barrage, L., and Lee, M. S., 2010. A penny for your thoughts: Inducing truth-telling in stated preference elicitation. Economics Letters 106(2):140-142.

Bazzani, C., Palma, M. A., and Nayga, R. M., 2018. On the use of flexible mixing distributions in WTP space: an induced value choice experiment. Australian Journal of Agricultural and Resource Economics 62(2):185-198.

Bishop, R.C., Boyle, K.J., 2019. Reliability and Validity in Nonmarket Valuation. Environmental and Resource Economics 72:559-582.

Beck, M.J., Fifer, S. and Rose, J.M., 2016. Can you ever be certain? Reducing hypothetical bias in stated choice experiments via respondent reported choice certainty. Transportation Research Part B 89:149-167.

Bonsall, P. and Lythgoe, B., 2009. Factors affecting the amount of effort expended in responding to questions in behavioural choice experiments. Journal of Choice Modelling 2(2):216-236.

Brooks, K. and Lusk. J. L., 2010. Stated and Revealed Preferences for Organic and Cloned Milk: Combining Choice Experiment and Scanner Data. American Journal of Agricultural Economics 92 (4):1229-1241.

Bosworth, R. and Taylor, L. O., 2012. Hypothetical Bias in Choice Experiments: Is Cheap Talk Effective at Eliminating Bias on the Intensive and Extensive Margin of Choice. The B. E. Journal of Economic Analysis and Policy 12:1-26.

Bush, G., Colombo, S. and Hanley, N., 2009. Should all Choices Count? Using the Cut-Offs Approach to Edit Responses in a Choice Experiment. Environmental and Resource Economics 44 (3):397-414.

Caputo, V., Scarpa, R., Nayga, R. M., and Ortega, D. L., 2017. Are Preferences for Food Quality Attributes Really Normally Distributed? An Analysis Using Flexible Mixing Distributions. Journal of Choice Modelling 28:10-27.

Carlsson, F., Frykblom, P. and Lagerkvist, C. J., 2005. Using cheap talk as a test of validity in choice experiments. Economics Letters 89:147-152.

Carlsson, F., Mørkbak, M. and Olsen S., 2012 The first time is the hardest: A test of ordering effects in choice experiments. Journal of Choice Modelling 5(2):19-37.

Carson, R. T., and Groves, T., 2007 Incentive and Informational Properties of Preference Questions. Environmental and Resource Economics 37:181–210.

Champ, P. A., R. Moore, and Bishop R. C., 2009 A Comparison of Approaches to Mitigate Hypothetical Bias. Agricultural and Resource Economics Review 38:166–180.

Cherry, T.L., Frykblom, P., Shogren, J.F., 2002. Hardnose the dictator, American Economic Review 92(4):1218–1221.

Chowdhury, S., Meenakshi, J. V., Tomlins, K. and Owori, C., 2011. Are consumers in developing countries willing-to-pay more for micronutrient-dense biofortified foods? Evidence from a field experiment in Uganda. American Journal of Agricultural Economics 93(1):83–97.

Colombo, S., Glenk, K., Rocamora-Montiel, B., 2016 Analysis of choice inconsistencies in on-line choice experiments: impact on welfare measures. European Review of Agricultural Economics 43(2), 271–302.

Czajkowski, M., and Budziński, W., 2015. An insight into the numerical simulation bias – a comparison of efficiency and performance of different types of quasi Monte Carlo simulation methods under a wide range of experimental conditions. In: Environmental Choice Modelling Conference, Copenhagen.

Czajkowski, M., Vossler, C. A., Budziński, W., Wiśniewska, A., and Zawojska, E., 2017 Addressing empirical challenges related to the incentive compatibility of stated preference methods. Journal of Economic Behavior and Organization 142:47–63.

Doyon, M., Saulais, L., Ruffieux, B. and Bweli, D., 2015. Hypothetical Bias for Private Goods: Does Cheap Talk Make a Difference? Theoretical Economics Letters 5:749-756.

Franceschinis, C., Scarpa, R., and Thiene, M., 2017. A Monte Carlo Evaluation of the Logit-Mixed Logit under Asymmetry and Multimodality. University of Waikato Working Papers in Economics 17/23.

Fox, J. A., J. F. Shogren, D. J. Hayes, and Kliebenstein, J. B, 1998. CVM-X: Calibrating Contingent Values with Experimental Auction Markets. American Journal of Agricultural Economics 80:455–465.

Glenk, K., Meyerhoff, J., Akaichi, F., and Martin-Ortega, J., 2019: Revisiting Cost Vector Effects in Discrete Choice Experiments. Resource and Energy Economics 57, 135–155

Granado-Díaz, R., Gómez-Limón, J.A., Rodríguez-Entrena, M., and Villanueva, A.J., 2020. Spatial analysis of demand for sparsely located ecosystem services using alternative index approaches. European Review of Agricultural Economics 47(2):752-784.

Grebitus, C., Lusk, J. L. and Nayga Jr, R. M., 2013. Explaining differences in real and hypothetical experimental auctions and choice experiments with personality. Journal of Economic Psychology 36:11-26.

Hanley, N., Czajkowski, M., 2019. The Role of Stated Preference Valuation Methods in Understanding Choices and Informing Policy. Review of Environmental Economics and Policy 13:248-266.

Hollander-Blumoff, R., 2011. Intrinsic and Extrinsic Compliance Motivations: Comment on Feldmand. Washington University Journal of Law & Policy 35:53-67.

Faccioli, M., Czajkowski, M., Glenk, K., Martin-Ortega, J., 2020. Environmental attitudes and place identity as determinants of preferences for ecosystem services. Ecological Economics, 174 (Forthcoming, August).

Jacquemet, N., R.-V. Joule, S. Luchini, and Shogren. J.F., 2013. Preference elicitation under oath. Journal of Environmental Economics and Management 65:110-132.

Johnston, R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., Hanemann, W. M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R. and Vossler, C. A., 2017. Contemporary Guidance for Stated Preference Studies. Journal of the Association of Environmental and Resource Economists 4:319-405.

Kang, M.J., A. Rangel, M. Camus, and C.F. Camerer., 2011. Hypothetical and real choice differentially activate common valuation areas. The Journal of Neuroscience 31:461-468.

Ladenburg, J. and Olsen, S., 2014. Augmenting short Cheap Talk scripts with a repeated Opt-Uut Reminder in Choice Experiment surveys. Resource and Energy Economics 37:39-63.

LaRiviere, J., M. Czajkowski, N. Hanley, M. Aanesen, J. Falk-Petersen, and Tinch. D., 2014. The value of familiarity: Effects of knowledge and objective signals on willingness to pay for a public good. Journal of Environmental Economics and Management 68:376–389.

Liebe, Ulf., Glenk, K., Von Meyer-Hofer, M. and Spiller A., 2019. A web survey application of real choice experiment. Journal of Choice Modelling, 33:100150

List, J. A., Sinha, P. and Taylor, M. H., 2006. Using Choice Experiments to Value Non Market Goods and Services: Evidence from Field Experiments. Advances in Economic Analysis and Policy 6:1-37.

Loomis, J., 2011. What's to know about hypothetical bias in stated preference valuation studies? Journal of Economic Surveys 25:363-370.

Loomis, J. B., 2014. Strategies for overcoming hypothetical bias in stated preference surveys. Journal of Agricultural and Resource Economics 39:34-46.

Lusk, J. L. and Schroeder, T. C., 2004. Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks. American Journal of Agricultural Economics 86:467 -482.

McFadden, D., and Train, K., 2000. Mixed MNL Models for Discrete Response. Journal of Applied Econometrics 15(5):447-470.

Moser, R., Raffaelli, R. and Notaro. S., 2013.. European Review of Agricultural Economics 41:25-46.

Murphy, J. J., P. G. Allen, T. H. Stevens, and Weatherhead. D., 2005. A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation. Environmental & Resource Economics 30:313–325.

Penn, J. M., and Hu, W., 2018. Understanding Hypothetical Bias: An Enhanced Meta-Analysis. American Journal of Agricultural Economics 100(4):1186–1206.

Penn, J., Hu, W., 2019. Cheap talk efficacy under potential and actual Hypothetical Bias: A meta-analysis. Journal of Environmental Economics and Management 96, 22-35.

Ready, R.C,, Champ, P. A. and Lawton, J.L., 2010. Using Respondent Uncertainty to Mitigate Hypothetical Bias in a Stated Choice Experiment. Land Economics 86(2):363-381.

Revelt, D., and Train, K., 1998. Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level. Review of Economics and Statistics 80(4):647-657.

Rocamora-Montiel, B., Glenk, K., Colombo, S., 2014. Territorial Management Contracts as a tool to enhace the sustainability of sloping and mountainous olive groves: Evidence from a case study in Southern Spain. Land Use Policy 41:313-324.

Roe, B., Boyle, K.J. and Teisl, M.F., 1996 Using conjoint analysis to derive estimates of compensating variation. Journal of Environmental Economics and Management 31:145–159.

Rolfe, J. and Windle, J., 2003. Valuing the protection of aboriginal cultural heritage sites. The Economic Record 79:(Special issue) s85 – s95.

Rolfe, J. and Wang, X., 2011. Dealing with scale and scope issues in stated preference experiments. International Handbook on Non-Market Environmental Valuation. J. Bennett. Cheltenham, Edward Elgar Publishing.

Sælensminde, K., 2003 Embedding effect in valuation of non-market-goods. Transport Policy, 10:59-72.

Swait J., 2001 A non-compensatory choice model incorporating attribute cut-offs Transportation Research B 35:903-928.

Tonsor, G. T. and Shupp, R. S., 2011. Cheap talk scripts and online choice experiments: looking beyond the mean. American Journal of Agricultural Economics 93(4):1015–1031.

Train, K. E., 2009. Discrete Choice Methods with Simulation. 2 Ed., Cambridge University Press, New York.

Train, K. E., and Weeks, M., 2005. Discrete Choice Models in Preference Space and Willingness-to-pay Space. In: Applications of Simulation Methods in Environmental and Resource Economics, R. Scarpa and A. Alberini, eds., Springer, Dordrecht, 1-16.

Train, K., 2016. Mixed logit with a flexible mixing distribution. Journal of Choice Modelling 19:40-53.

Varela, E., Mahieu, P.-A., Giergiczny, M., Riera, P. and Soliño, M., 2014. Testing the single opt out reminder in choice experiments: An application to fuel break management in Spain. Journal of Forest Economics 20:212-222.

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

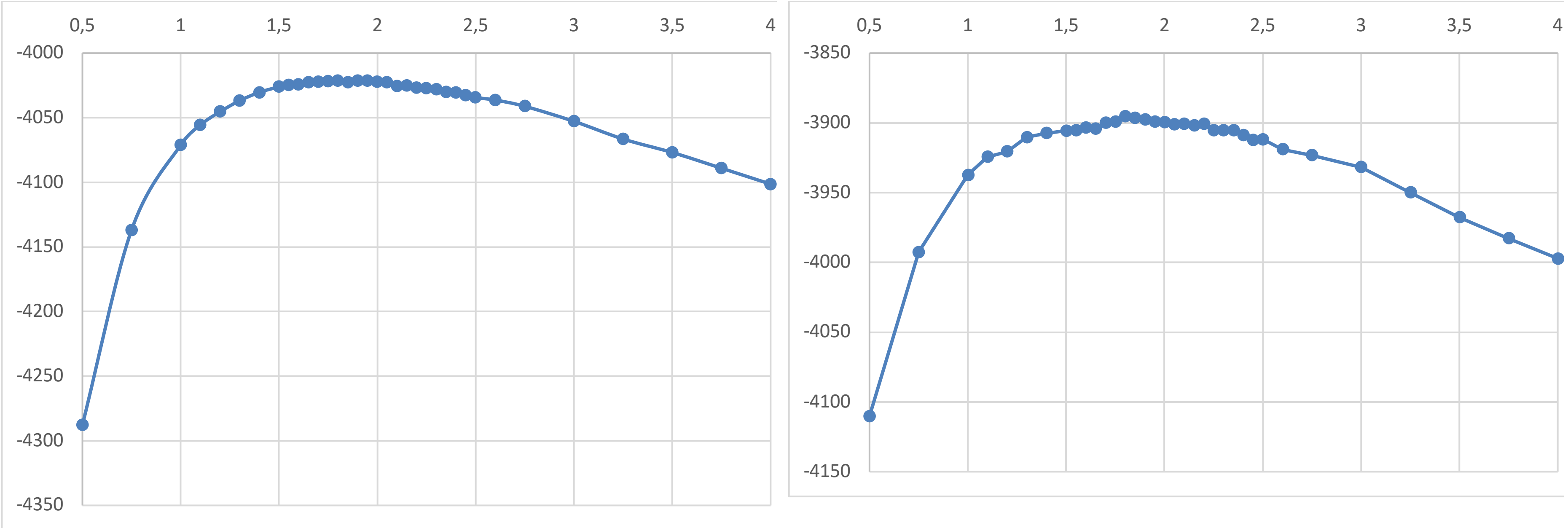
Whitehead, J. C., and Cherry, T. L., 2007. Willingness to Pay for a Green Energy program: A Comparison of Ex-Ante and Ex-Post Hypothetical Bias Mitigation Approaches. Resource and Energy Economics 29:247–261.

Zawojska, E., and Czajkowski, M., 2017. Re-examining empirical evidence on stated preferences: Importance of incentive compatibility. Journal of Environmental Economics and Policy 6:374-403.

Zawojska, E., A. Bartczak, and Czajkowski, M., 2018. Disentangling the effects of policy and payment consequentiality and risk attitudes on stated preferences. Working Papers no. 1 / 2018 (260), Faculty of Economic Sciences, University of Warsaw, Warsaw.

Zawojska, E., A. Bartczak, and Czajkowski, M., 2019. Disentangling the effects of policy and payment consequentiality and risk attitudes on stated preferences. Journal of Environmental Economics and Management 93:63-84.

Appendix A. The effect of bounds range on the average fit of LML models



Notes: Vertical axes show the mean log-likelihood over 56 LML model specifications without (left panel) or with (right panel) correlated parameters. The specifications included used approximate normal distributions, Legendre polynomials, step functions, linear splines, cubic splines, piece-wise cubic splines, and piece-wise cubic Hermite interpolating splines with 2 to 10 levels (orders of polynomials, step function segments or knots of splines used to represented distribution of WTP). Horizontal axes show the range of the bounds expressed as the number of standard deviations above / below the mean of each distributions. Means and standard deviations used to generate bounds were based on a MXL model with normally distributed coefficients without correlations. The lower bound for the monetary attribute was always anchored at 0.

Appendix B. The comparison of fit of various LML model specifications

Distribution	Level	Uncorrelated parameters				Correlated parameters			
		Log-likelihood	Number of parameters	AIC/n	BIC/n	Log-likelihood	Number of parameters	AIC/n	BIC/n
MXL	-	-4104.41	24	1.7180	1.7504	-3939.19	90	1.6767	1.7980
AN	2	-4088.28	24	1.7113	1.7437	-3941.12	90	1.6775	1.7988
Poly	2	-4088.28	24	1.7113	1.7437	-3936.00	90	1.6754	1.7967
Step	2	-4119.87	12	1.7195	1.7356	-3990.34	78	1.6930	1.7982
Spline L	2	-4071.76	36	1.7094	1.7580	-3922.85	102	1.6749	1.8124
Spline C	2	-4074.31	36	1.7105	1.7590	-3933.82	102	1.6795	1.8170
Spline PW-C	2	-4071.51	36	1.7093	1.7578	-3921.79	102	1.6745	1.8120
Spline PW-CHI	2	-4075.08	36	1.7108	1.7593	-3924.51	102	1.6756	1.8131
AN	3	-4079.13	36	1.7125	1.7610	-3927.83	102	1.6770	1.8145
Poly	3	-4079.13	36	1.7125	1.7610	-3929.99	102	1.6779	1.8154
Step	3	-4094.83	24	1.7140	1.7464	-3968.09	90	1.6888	1.8101
Spline L	3	-4046.56	48	1.7039	1.7686	-3907.53	114	1.6735	1.8272
Spline C	3	-4055.14	48	1.7075	1.7722	-3909.11	114	1.6742	1.8279
Spline PW-C	3	-4056.35	48	1.7080	1.7727	-3914.93	114	1.6766	1.8303
Spline PW-CHI	3	-4048.60	48	1.7048	1.7695	-3920.12	114	1.6788	1.8324
AN	4	-4051.42	48	1.7060	1.7707	-3915.10	114	1.6767	1.8303
Poly	4	-4053.09	48	1.7067	1.7713	-3907.25	114	1.6734	1.8271
Step	4	-4082.16	36	1.7138	1.7623	-3946.50	102	1.6848	1.8222
Spline L	4	-4027.03	60	1.7008	1.7817	-3884.30	126	1.6689	1.8387
Spline C	4	-4032.40	60	1.7030	1.7839	-3911.63	126	1.6802	1.8501
Spline PW-C	4	-4030.30	60	1.7022	1.7830	-3890.07	126	1.6713	1.8411
Spline PW-CHI	4	-4024.27	60	1.6997	1.7805	-3906.53	126	1.6781	1.8479
AN	5	-4025.63	60	1.7002	1.7811	-3899.88	126	1.6754	1.8452
Poly	5	-4027.79	60	1.7011	1.7820	-3894.36	126	1.6731	1.8429
Step	5	-4067.20	48	1.7125	1.7772	-3939.02	114	1.6867	1.8403
Spline L	5	-4016.33	72	1.7013	1.7984	-3867.37	138	1.6668	1.8528
Spline C	5	-4017.11	72	1.7017	1.7987	-3872.85	138	1.6691	1.8551
Spline PW-C	5	-4009.68	72	1.6986	1.7956	-3891.88	138	1.6770	1.8630
Spline PW-CHI	5	-4022.04	72	1.7037	1.8008	-3878.75	138	1.6716	1.8576
AN	6	-4006.06	72	1.6971	1.7941	-3891.46	138	1.6768	1.8628
Poly	6	-4017.22	72	1.7017	1.7988	-3882.75	138	1.6732	1.8592
Step	6	-4044.45	60	1.7081	1.7889	-3919.19	126	1.6834	1.8532
Spline L	6	-3991.19	84	1.6959	1.8091	-3850.43	150	1.6648	1.8669
Spline C	6	-3989.13	84	1.6950	1.8082	-3861.35	150	1.6693	1.8715
Spline PW-C	6	-3984.01	84	1.6929	1.8061	-3868.21	150	1.6722	1.8743
Spline PW-CHI	6	-3995.64	84	1.6977	1.8109	-3830.30	150	1.6564	1.8586
AN	7	-4002.15	84	1.7004	1.8137	-3873.35	150	1.6743	1.8765
Poly	7	-4005.57	84	1.7019	1.8151	-3870.76	150	1.6732	1.8754
Step	7	-4034.24	72	1.7088	1.8058	-3878.13	138	1.6713	1.8573
Spline L	7	-3980.06	96	1.6962	1.8256	-3841.10	162	1.6659	1.8842
Spline C	7	-3982.44	96	1.6972	1.8266	-3830.53	162	1.6615	1.8798
Spline PW-C	7	-3972.30	96	1.6930	1.8224	-3842.33	162	1.6664	1.8847

Spline PW-CHI	7	-3978.74	96	1.6957	1.8251	-3826.20	162	1.6597	1.8780
AN	8	-3998.54	96	1.7039	1.8333	-3870.26	162	1.6780	1.8964
Poly	8	-4002.54	96	1.7056	1.8350	-3869.87	162	1.6778	1.8962
Step	8	-4010.27	84	1.7038	1.8170	-3864.46	150	1.6706	1.8728
Spline L	8	-3967.51	108	1.6960	1.8416	-3807.52	174	1.6569	1.8914
Spline C	8	-3933.43	108	1.6818	1.8274	-3815.53	174	1.6602	1.8947
Spline PW-C	8	-3958.03	108	1.6921	1.8376	**-3784.35**	**174**	**1.6473**	**1.8818**
Spline PW-CHI	8	-3957.92	108	1.6920	1.8376	-3813.01	174	1.6592	1.8937
AN	9	-3997.33	108	1.7084	1.8540	-3863.41	174	1.6802	1.9147
Poly	9	-3970.70	108	1.6973	1.8429	-3867.03	174	1.6817	1.9162
Step	9	-3982.44	96	1.6972	1.8266	-3846.80	162	1.6682	1.8866
Spline L	9	-3941.77	120	1.6903	1.8520	-3793.73	186	1.6562	1.9068
Spline C	9	-3930.21	120	1.6855	1.8472	-3796.57	186	1.6573	1.9080
Spline PW-C	9	-3932.64	120	1.6865	1.8482	-3799.17	186	1.6584	1.9091
Spline PW-CHI	9	-3943.40	120	1.6910	1.8527	-3793.00	186	1.6558	1.9065
AN	10	-3994.34	120	1.7122	1.8739	-3856.33	186	1.6822	1.9329
Poly	10	-3961.57	120	1.6985	1.8603	-3828.53	186	1.6706	1.9213
Step	10	-3976.28	108	1.6997	1.8452	-3838.08	174	1.6696	1.9041
Spline L	10	-3913.32	132	1.6834	1.8614	-3782.27	198	1.6564	1.9232
Spline C	10	-3906.73	132	1.6807	1.8586	-3793.36	198	1.6610	1.9279
Spline PW-C	10	-3917.85	132	1.6853	1.8632	-3783.55	198	1.6569	1.9238
Spline PW-CHI	10	-3906.19	132	1.6805	1.8584	-3774.21	198	1.6530	1.9199

Notes: MXL refers to the MXL model (included for comparison only). Distribution types considered were: AN – approximate normal, Poly – Legendre polynomial, Step – step function, Spline L – linear spline, Spline C – cubic spline, Spline PW-C – piece-wise cubic spline, Spline - PW-CHI – piece-wise cubic Hermite interpolating spline. AN and Poly assumed normal distributions for all non-monetary attributes and log-normal distribution for the cost. All non-linear splines were derived using linear extrapolation for bound extension, which results in smoother functions in the boundary segments. Level refers to the order of a polynomial (AN, Poly), the number of step-function segments (excluding the reference segment) or the number of knots of splines (excluding the boundary kno

University of Warsaw
Faculty of Economic Sciences
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl