

**COMPARING RESULTS OF RANKING CONJOINT ANALYSES, BEST
WORST SCALING, AND DISCRETE CHOICE EXPERIMENTS IN A NON-
HYPOTHETICAL CONTEXT**

Abstract

This study assesses the comparability of discrete choice experiment, ranking conjoint analysis, and multi-profile best worst scaling in a non-hypothetical context in terms of estimated partworths, willingness to pay, response consistency, and external validity. Overall, the results suggest that (1) the conjoint analysis formats that were used in this study provide similar estimated WTP, but different estimated partworths and computed external validity, (2) the inclusion of the full ranking information in the estimation of the parameters of interest affects the estimated partworths, but not the estimated willingness to pay, and (3) it is more appropriate to use multi-profile best worst scaling over discrete choice experiment and ranking conjoint analysis because it has better predictive power of consumers' preferences and provides estimated willingness to pay comparable to those obtained in the others conjoint analysis formats. The best worst scaling' cognitive process could be considered clearness for participants implying significant increment of it predictive power.

Keyword: Discrete choice experiment, Ranking Conjoint Analysis, Best Worst Scaling, External Validity, Willingness to Pay

Since its introduction, conjoint analysis (CA) has become one of the most popular research tools to elicit consumer preference and willingness to pay (WTP). CA is a stated-preference method that requires human participants to rate, rank, or choose between competing products or alternatives (Louviere and Street 2000). CA has been frequently used in a variety of disciplines such as marketing (Ding *et al.* 2005; Alfnes *et al.* 2006; Ding 2007; Chang *et al.* 2009; Dong *et al.* 2010), agriculture economics (Scarpa and Del Giudice 2004; Lusk *et al.* 2008; Corrigan *et al.* 2009; Menapace *et al.* 2011), environmental economics (Hanley and Ruffell 1993; Boyle *et al.* 2001; Caparrós *et al.* 2008; Campbell and Lorimer 2009; Scarpa *et al.* 2011), transport economics (Hensher and Greene 2003; Greene *et al.* 2006; Louviere *et al.* 2008; Greene and Hensher 2013), and health economics (Louviere *et al.* 2010; Flynn *et al.* 2010; De Bekker *et al.* 2012; Lancsar *et al.* 2013).

Currently, discrete choice experiment (DCE) is the most widely used CA format. In CE, respondents are shown a set of combinations of attributes, i.e., profiles, and are asked to indicate which of the profiles they would purchase. DCE gained popularity because it can mimic real-market settings where consumers are offered competing products and asked to purchase the product that aligns with their preferences. However, DCE only allows for the observation of the most-preferred option but does not provide information on consumer' preferences towards the remaining products concepts in the same choice set (Lusk *et al.* 2008; Louviere *et al.* 2008; Lancsar *et al.* 2013). In contrast to DCE, participants in a ranking conjoint analysis (RCA) are provided with a set of product concepts and are asked to rank them from the most to the least-preferred product concept. The use of RCA as an alternative to DCE is gaining popularity since it provides information not only on the most-preferred

product concept, but also on consumer preferences for all the product concepts included in a choice set, which in turn leads to more efficient preference estimates (Chang *et al.* 2009; Louviere *et al.* 2008; Lusk *et al.* 2008).

Despite the extensive use of DCE and RCA over the last two decades, few studies have been published on their performance in terms of estimated partworths, the predictive power of estimated partworths, and the reliability of the WTP values deduced from the estimated partworths (Boyle *et al.* 2001; Holmes and Boyle, 2001; Morrison and Boyle 2001; Siikamäki and Layton 2007; Caparrós *et al.* 2008; Pignone *et al.* 2011; Akaichi *et al.* 2013). To illustrate, Boyle *et al.* (2001) compared DCE, RCA, and recoded RCA (RRCA), which means the participant's response is coded as "1" when the product concept presented in the choice set is ranked first and "0" otherwise. Boyle *et al.* (2001) found differences between the results obtained in DCE and RCA. They argued that these differences could be explained by the various cognitive processes used by subjects in each CA format. Similarly, Caparrós *et al.* (2008) pointed out that DCE and RRCA provide comparable results when a similar experimental design is used for both DCE and RCA. Notably, Akaichi *et al.* (2013) confirmed the results found by Caparrós *et al.* (2008) for small choice sets, i.e., four alternatives; however, they found discrepancies between respondents' preferences in DCE and RRCA when large choice sets, i.e., eight alternatives, were used. Finally, Chang *et al.* (2009) showed that the non-hypothetical RCA outperforms both the hypothetical and non-hypothetical DCE when the full ranking information is considered in the estimation.

Finn and Louviere (1992) introduced another CA format named, Best Worst Scaling (BWS). There are three cases of BWS, which differ in terms of the

complexity of the items or options assessing: BWS object case, BWS profile case and BWS multi-profile case (Flynn and Marley, 2014). The last case of BWS approach (MBWS), which considered in this study, consists of asking respondents to first choose the best and the worst option, then the second best and the second-worst options from the remaining options and so on until a complete preference ordering of all the product concepts included in a choice set is obtained (Scarpa *et al.* 2011). Compared with RCA, MBWS has the advantage of being aligned with the random utility theory (Louviere *et al.* 2010). Furthermore, the choice task in MBWS is easier for respondents to understand thanks to their human skills at identifying extremes (Marley and Louviere, 2005; Vermuelen *et al.* 2010; Potoglou *et al.* 2011; Flynn and Marley, 2014). Similar to RCA, the additional choice information obtained from MBWS has been shown to improve the statistical efficiency of choice models (Lancsar *et al.* 2013). Nevertheless, none of the aforementioned studies that compared different CA formats assessed the comparability of MBWS to both DCE and RCA.

To the best of our knowledge, only two published studies have compared BWS to DCE or to RCA. Potoglou *et al.* (2011) compared welfare estimates obtained in DCE and BWS experiments. They found that the differences in the estimated preference weights between the two CA formats were not statistically significant. Lagerkvist (2013) compared the results obtained in RCA and BWS and found that the rank, choice probability, and attribute dominance obtained in RCA and BWS were significantly different. While their findings informed our research, our study is different from those of Potoglou *et al.* (2011) and Lagerkvist (2013) in five key ways. First, we assessed the comparability of results obtained in the three CA formats, i.e., DCE, RCA, and MBWS. Particularly, we assessed the comparability of the three CA

formats in terms of estimated partworths, predictive power, response consistency, and WTP.

Second, Potoglou *et al.* (2011) and Lagerkvist (2013) conducted the different CA formats in hypothetical settings. Nonetheless, the divergence in valuations between hypothetical and non-hypothetical preference and value-elicitation methods, e.g., discrete choice experiment, contingent valuation, and experimental auction, is well documented in the literature (List and Gallet 2001; Little and Berrens 2004; Murphy *et al.* 2005; Chang *et al.* 2009; Moser *et al.* 2014). Notably, these studies found that participants in hypothetical elicitation methods overestimated their preferences and WTP. This behavior is commonly explained by the fact that in the absence of any moral or monetary cost that prevents subjects from deviating from their actual behavior, participants in hypothetical elicitation methods will not put enough cognitive effort into the elicitation tasks and may not reveal their true preferences and values (Lusk and Shogren 2007). Despite the well-documented negative effect of hypothetical bias, the majority of the studies that assessed the comparability of CA formats reported results from economic experiments conducted in hypothetical settings, with the exception of Chang *et al.* (2009) and Akaichi *et al.* (2013). Accordingly, due to the skepticism surrounding the validity of values obtained from hypothetical CA formats, we conducted the DCE, the RCA, and the MBWS in non-hypothetical settings.

Third, Potoglou *et al.* (2011) and Lagerkvist (2013) used different experimental designs for the DCE, the RCA, and the BWS. Therefore, the divergence of the estimates obtained by Lagerkvist (2013) in RCA and BWS could be a result of the variation in the experimental design used. To rule out the effect of this bias, we used

an identical experimental design for DCE, RCA, and MBWS. As a result, participants in the DCE, the RCA, and the MBWS treatments were provided with identical choice sets, but were asked to reveal their preferences in different ways. In the DCE, respondents were asked to indicate the product they prefer most. Comparatively, in RCA, respondents were asked to rank the product concepts from the most preferred to the least preferred, while in MBWS they were asked to indicate the best and the worst product concepts.

Fourth, with the exception of Akaichi *et al.* (2013), the previous studies that compared different CA formats did not test the external validity of their estimates. Testing the external validity of the estimates in CA is important since it allows for the assessment of the predictive power of the estimated choice models, which in turn constitutes one of the main reasons for using CA formats. To compare the external validity of the three CA formats, we used a non-hypothetical holdout choice task. Following Ding (2007) and Akaichi *et al.* (2013), the holdout task, that is a regular choice task, was held out of the partworths' estimation process since it is only used to assess the out-of-sample predictive power of the estimated partworths in the three CA formats.

Fifth, one of the main assumptions underlying the theory of stated preferences is that respondent preferences are stable and coherent (Brown *et al.* 2008). Nonetheless, according to Hoeffler and Ariely (1999), preference consistency or stability is positively correlated with choice experience and cognitive choice effort. For instance, in repeated choices, respondents are expected to be more precise and hence more consistent in their decisions due to the learning effect (Brouwer *et al.* 2010). On the other hand, when respondents face complex choice tasks, e.g., too many choice sets or

too many product concepts per choice set, they are expected to be less precise and hence less consistent in their choices, when compared with respondents who face an easy choice scenario (Brouwer *et al.* 2010). In this study, we also compared the consistency of respondents' answers in DCE, RCA, and MBWS to find out which of the three CA formats provide the highest level of response consistency. To test the consistency of respondents' answers, one of the choice sets given to respondents was repeated at the end of the main choice task.

Methodology

Design and the implementation of the experiment

In this study, three treatments were conducted in Barcelona, Spain: a non-hypothetical DCE (NHCE), a non-hypothetical RCA (NHRCA), and a non-hypothetical MBWS (NHMBWS). To assess the comparability of the three CA formats, 165 real consumers were recruited. The sample of participants was randomly selected from different locations across the city of Barcelona and its metropolitan area, using a stratified sampling procedure by age and gender. Participants were randomly and equally assigned to the three treatments¹.

¹To calculate the minimum sample size, the software Ngene was used to determine the S-error value of the design obtained from the approach of Street and Burgess (2007). S-error is a measure that has been introduced by Bliemer and Rose (2005a) to determine the necessary sample size for the purpose of obtaining significant estimates of the parameters (see Bliemer and Rose 2009 for more details). In our case, the calculated S-error indicated that the minimum sample size is 26

Olive oil was the food product used in the experiments, and the attributes and attribute levels were identified based on the literature review and the information collected from two focus groups of high to low-experienced consumers of olive oil. Accordingly, four attributes were considered; three of which have three levels including: (1) type of olive oil, i.e., extra virgin, virgin, and olive oil, (2) origin of olive oil, i.e., Andalucía, Catalonia (local olive oil), and the rest of Spain, and (3) price i.e., 2.20 €/liter, 3.50 €/liter, and 4.80 €/liter, which accounts for 85% of the price distribution of olive oil in the retail outlets operating in the city of Barcelona. Brand is the fourth attribute and has two levels: manufacturer label and private label.

Given these attribute levels, a full factorial design of 54 ($3^3 \times 2$) product concepts, i.e., one-liter bottles of olive oil, was generated. Notably, presenting respondents with 54 product concepts could, however, place a high level of cognitive burden on respondents. Therefore, Street and Burgess (2007) approach was followed to reduce the number of product concepts that participants have to evaluate. To explain, first, an orthogonal fractional factorial design of 9 product concepts was generated. These 9 product concepts were considered as the first option in each choice set. Since participants were provided with choice sets of five product concepts each plus a no-choice option, we need to add sets of four generators to the initial profiles using a modular arithmetic such as (1000), (1111), (2121), and (2122) imposing a

participants. A sample size of 55 participants was employed per treatment. Previous studies such as Pignone *et al.* (2011), Magistris *et al.* (2013) and Akaichi *et al.* (2013) used similar sample sizes, ranging between 43 and 55 participants per treatment.

systematic attributes levels change to obtain the other four product concepts (Street and Burgess 2007). Overall, the experiment design consist in nine choice sets of five product concepts each plus the no-choice option which as a 100% efficient main-effects design. Examples of the choice sets used in each treatment are displayed in Figures 1, 2 and 3.

At the beginning of the experiment, participants were informed that they would receive a participation fee of 15 Euros in cash at the end of their participation. Additionally, all participants were informed that they would be participating in non-hypothetical tasks which incorporate a real product and the respondents have to pay for their chosen product concepts at the end of the experiment. Therefore, it is in their best interest to reveal their actual preferences. Particularly, participants were shown how they can lose money if they deviate from their true valuations. Participants were then informed on how each CA format works. Participants in each treatment performed two choice tasks—a main task and a holdout task.

In the main task, and depending on the treatment (NHCE, NHRCA, or NHMBWS), participants were successively provided with 10 choice sets—first, they received the nine choice sets obtained in the efficient design; then they received the fifth choice set again to help assess the consistency of their responses. In each choice set, participants were asked to mark their most-preferred product concept, or to rank all the product concepts, or to choose the best and the worst product concepts, depending on the treatment (NHCE, NHRCA, or NHMBWS).

After collecting each completed choice cards and before distributing the next choice card, participants were asked to answer two questions to collect information on

attribute non-attendance (see Figure 1)². In the first question, each participants was asked to mark the attribute(s) of olive oil that s/he did not consider when evaluating the five alternatives of olive oil. In the second question, each participant was asked to rank the attributes that were not marked in the first question terms of importance (from the most important to the least important) during her/his choice decision.

In the holdout task, participants were given a single choice card of 10 product concepts, which are different from the product concepts provided to participants in the main task, plus a no-choice option (figure 4). Then, they were asked to choose the product concept they prefer most among the 10 product concepts included in the choice set.

Each of the three treatments (NHCE, NHRCA, and NHMBWS) was conducted in five sessions throughout different days of the week and different hours of the day. Ten to fifteen subjects participated in each session. After finishing the two tasks, participants were asked to complete a short questionnaire about their socio-demographic and lexicographic characteristics, as well as their attitudes towards olive oil.

Non-hypothetical discrete choice experiment (NHCE)

In the NHCE treatment, participants were informed that each choice set was a real shopping scenario. Therefore, at the end of the experiment each participant was given

²The information collected from the answers to these two questions was used to assess the comparability of the effect of attribute non-attendance in different conjoint analysis formats. The results on the effect of attribute non-attendance are not reported in this paper.

the product concept they had selected (if different from the no-choice option) and then paid its price. In each choice set, participants were asked to indicate the product concept they preferred most, bearing in mind their real purchase habits. They were also informed that if they did not like any one of the provided olive oil concepts, they could simply choose the no-choice option. After finishing the main task, participants were given a choice set of 11 options, i.e., 10 product concepts and the no-choice option, and they were asked to choose the product concept they preferred most.

After completing the two tasks and the questionnaire, a volunteer among the participants was asked to randomly draw one of the two tasks to be the binding task. If the binding task was the main task, another volunteer was selected to randomly draw one of the nine choice sets³ to determine which of the choice sets was the binding one. Then, each participant was provided with the product concept they had chosen in the binding choice set and received 15 Euros minus the price of the chosen product concept. If the participant chose the no-choice option, they received the 15 Euros and did not buy any product. If the binding task was the holdout task, each participant was given the chosen product concept and was asked to pay its price. Furthermore, if the chosen option was the no-choice option, the participant received 15 Euros and did not buy any product.

Non-hypothetical rank conjoint analysis (NHRCA)

³ The last choice set (the number 10) was the same as the fifth choice set. Therefore, to allow that all the choice sets have the same probability to be drawn we removed the tenth choice set.

The same 10 choice sets that were provided in the NHCE treatment were successively presented to participants in the NHRCA treatment. However, in this treatment they were asked to rank the five product concepts included in each choice set from the most to the least- preferred product concept. In case a participant does not like any one of the five product concepts, they have to mark the no-choice option. The non-hypothetical nature of the experiment was also revealed to participants before performing the ranking task. After completing all of the choice sets in the main task, participants were given a choice set of 11 options, i.e., a holdout task, and were then asked to choose the product concept they preferred most or mark the no-choice option.

After completing the main and holdout tasks, a volunteer among the participants was asked to randomly draw one of the two tasks to be the binding task. If the main task was chosen as the binding task, another volunteer would be approached to draw the binding choice set. As in Lusk, Fields, and Prevatt (2008), the non-hypothetical nature of the RCA requires that each participant purchases the binding product with a probability proportional to the rank they assign to each one of the five product concepts. Particularly, each participant who did not choose the no-choice option was asked to draw a number from 1 to 50 to select the binding product. If the drawn number was between 1 and 17, the participant would purchase the most-preferred product concept and would pay its price. If the drawn number was between 18 and 30, the second most preferred option would be the binding product. If the drawn number was between 31 and 40, the participant would purchase the product concept they ranked third and would pay its corresponding price. If the number drawn was between 41 and 47, the participant would buy their fourth most-preferred product concept. If

the drawn number was between 48 and 50, the participant would have to buy the product concept they ranked fifth and would pay its corresponding price. Finally, if the binding task was the holdout task, the procedure implemented was similar to the one used in NHCE treatment.

Non-hypothetical multi-profile best worst scaling (NHMBWS)

Similar to the previous treatments, the same 10 choice sets were presented to each participant in the main task. In NHMBWS however, each participant was asked to mark the best product concept, i.e., the most-preferred product concept, followed by the worst product concept, i.e., the least-preferred product concept of the four remaining options, followed by the second-best product concept of the three remaining options, and finally the second-worst product concept of the two remaining options. As a result, a complete ranking of the five product concepts can be deduced, i.e., the product concept ranked first, second, third, fourth, and fifth are the best, the second best, the remaining product concept, the second worst, and the worst product concept, respectively.

After finishing the main task and similar to the other two treatments (NHCE and NHRCA), participants in NHMBWS were given a choice set of 11 product concepts, i.e., a holdout task, and were asked to choose the product they preferred most. Once participants finished the main and holdout tasks, a similar procedure to the one applied in the NHRCA was used to determine the binding task, the binding choice set, and the product that the participant must purchase and the price they have to pay.

Data analysis

Estimation of partworths

Participants' choice decisions made in any of the three CA formats considered in this paper are analyzed using random utility theory (McFadden, 1973). The random utility theory postulates that the i^{th} individual's utility function, U_{ijs} , towards an option j from a choice set s can be decomposed into a systematic (observable) component, V_{ijs} , and a stochastic (non-observable) component, ε_{ijs} . The utility of the i^{th} individual is given by:

$$U_{ijs} = V_{ijs} + \varepsilon_{ijs} \quad (1)$$

The systematic component is typically assumed to be a linear relationship of observed attribute levels and respondents' characteristics. In the case of this study, the systematic component is specified as follows:

$$V_{ijs} = \beta_0 NoBuy + \beta_{EVOO} EVOO_{ijs} + \beta_{OO} OO_{ijs} + \beta_{Manf} Manf_{ijs} + \beta_{CAT} CAT_{ijs} + \beta_{RSp} RSp_{ijs} + \beta_{price} price_{ijs} \quad (2)$$

The attribute levels, i.e., extra virgin olive oil (EVOO), olive oil (OO), Manufacturer label (Manf), Catalonian origin (CAT), and the "Rest of Spain" origin (RSp)) were effect coded (-1, 0, 1)⁴, except for the price that was coded as a linear variable. The parameter "NoBuy" represents the no-choice option and has been coded as a dummy variable that takes the value "1" when the no-choice option is chosen by participant, and "0", otherwise.

⁴ The attribute levels virgin olive oil (VOO), private label (PRV), and Andalusia (AND) were set as the baseline levels for the attributes: type of olive oil, brand, and origin, respectively.

In the case of the discrete choice data obtained in the NHCE treatment, the random parameter logit model (RPL) was used to estimate the partworths ($\beta_0, \beta_{EVOO}, \beta_{OO}, \beta_{Manf}, \beta_{CAT}, \beta_{RSp}$ and β_{price}). The conditional logit model (McFadden 1974) has been used as the main choice model for analyzing discrete choice data. Although its popularity, this model has several well-known limitations: (1) it does not control for preference heterogeneity among respondents, and (2) it assumes that the alternatives included in any choice set are independent, which can lead to unrealistic estimates. The RPL model relaxes the assumptions of the conditional logit model. RPL controls for preference heterogeneity among respondents by allowing one or more of the parameters in the model to be randomly distributed. It also relaxes the assumption of independence of the alternatives considered in a choice set by allowing the unobserved factors to be correlated over time (McFadden and Train 2000).

In the RPL, the unconditional probability that consumer i chooses the option j in the choice set s is given by (Train, 2003):

$$Prob_i(j \text{ is chosen}) = \int L_{ij}(\beta_{ij})f(\beta_i/\theta)d\beta_i \quad (3)$$

where $f(\beta_i/\theta)$ is the density function of the coefficients β_i . θ represents the moments (the mean and standard deviation) of the parameters' distributions, and $L_{ij}(\beta_{ij})$ is the conditional probability that individual i chooses the option j . $L_{ij}(\beta_{ij})$ is given by:

$$L_{ij}(\beta_{ij}) = \frac{e^{\beta_{ij}X_{ij}}}{\sum e^{\beta_{ik}X_{ik}}} = \frac{e^{\beta_{ij}X_{ij}}}{\sum e^{\beta_{ik}X_{ik}}}, \text{ with } k \in C_s \quad (4)$$

In the case of the 'choice' data obtained in the NHRCA and NHMBWS treatments, the partworths were estimated using the rank-order random parameter

logit (RO-RPL) model (Lusk *et al.* 2008). The RO-RPL assumes that the probability of a particular ranking of the product concepts presented in a choice set is the product of the multinomial choice probability for always choosing the best of the remaining options. That is, the probability (L_{ij}) that an individual i ranks the five product concepts, A, B, C, D, and E as follows $A > B > C > D > E$, will be modeled as the product of the probability of choosing A as the best option from the choice set (A, B, C, D, E), the probability of choosing B as the best option among the remaining options (B, C, D, E), the probability of choosing C as the best option among the remaining options (C, D, E), and the probability of choosing D as the best option among the remaining options (D, E). Therefore, L_{ij} is given by:

$$L_{ij}(\text{ranking } A, B, C, D, E) = \frac{e^{V_{iA}}}{\sum_{j=A,B,C,D,E} e^{V_{ij}}} * \frac{e^{V_{iB}}}{\sum_{j=B,C,D,E} e^{V_{ij}}} * \frac{e^{V_{iC}}}{\sum_{j=C,D,E} e^{V_{ij}}} * \frac{e^{V_{iD}}}{\sum_{j=D,E} e^{V_{ij}}} \quad (5)$$

In the estimation of RPL and RO-RPL, it was assumed that all the partworths, β_{ij} , of our empirical model were random and followed a normal distribution with mean β and variance-covariance matrix Ω .

Since this study aims to compare participants' preferences and WTP across treatments (between-subjects analysis), it is important to investigate the preference regularity across treatments. Following Lusk and Schroder (2004), the likelihood-ratio test was used to test the null hypothesis of the equality of preferences across treatments ($\mu_{NHCE}\beta_{NHCE} = \mu_{NHRCA}\beta_{NHRCA}$, μ is the scale parameter). The likelihood-ratio test is calculated using the following expression: $-2(LL_j - \sum LL_i)$, which is distributed as a Chi square with $K(M-1)$ degrees of freedom. LL_j is the log likelihood values of the pooled data, e.g., pooling NHCE and NHRCA data, and LL_i is the log likelihood values of the estimated model for each treatment. Furthermore, K is the

number of restrictions and M is the number of treatments (Swait and Louviere 1993; Louviere and Street 2000). If the hypothesis of partworths equality is rejected, then comparing the estimated partworths across treatments is appropriate. By contrast, if the hypothesis of partworths equality is accepted, then comparing the estimated partworths across treatments is not appropriate because the differences in the estimated partworths across treatments could be attributed to the difference in participants' preferences, or to the difference in the error variance, i.e., inverse of the scale parameter, across treatments or a combination of both.

The results of the likelihood ratio test are displayed in Table 1. The results show that the null hypothesis of preferences regularity is rejected even when the full ranking information is considered in the estimation of the partworths. Therefore, comparing the CA formats in terms of partworths is appropriate. To test the comparability of the estimated partworths across the treatments, the estimation procedure used by De-Magistris *et al.* (2013) was followed. First, the data were pooled that correspond to the treatments to be compared. Then, the extended utility function (see Equation 6) was estimated as:

$$\begin{aligned}
 V = & \beta_0 NoBuy + \beta_{price} Price + \beta_{EVOO} EVOO + \beta_{OO} OO + \beta_{CAT} CAT + \beta_{Manf} Manf + \\
 & \beta_{RSp} RSp + \gamma_{EVOO} (EVOO * dtreat) + \gamma_{oo} (OO * dtreat) + \gamma_{cat} (CAT * dtreat) + \\
 & \gamma_{Manf} (Manf * dtreat) + \gamma_{RSp} (RSp * dtreat) + \varepsilon
 \end{aligned} \tag{6}$$

where *dtreat* is coded as “1” for the first treatment and “0” for the second treatment. For example, if NHCE and NHRCA are the treatments to be compared, *dtreat* is coded as “1” if the treatment is NHCE, and “0” if the treatment is NHRCA. In total, 10 extended utility functions were estimated. The significance and the signs of the

estimated γ parameters were then used to test the comparability of the three CA formats in terms of the estimated partworths.

Willingness to pay

In addition to the estimation of partworths, choice data is often used to calculate WTP. WTP is commonly expressed as the negative ratio of the non-price attribute coefficient to the price coefficient:

$$WTP_{non-price\ attribute} = -\frac{\beta_{non\ price\ attribute}}{\beta_{price}} \quad (7)$$

Nonetheless, depending on the distributions of the coefficients, this standard approach of computing WTP can result in heavily-skewed WTP distributions (Train and Weeks 2004). A common approach used to solve this problem is to specify the parameter price as fixed. However, it is often unreasonable to assume that all individuals have the same preferences for price (Meijer and Rouwendal 2006).

To get around this problem Train and Weeks (2004) suggested to estimate the RPL in WTP space rather than in preference space. This is done by re-formulating the model in such a way that the coefficients to be estimated represent the WTP measures and not the partworths. As a results, when the RPL is estimated in WTP space, the a priori assumptions about the distributions of the parameters are made on the WTP rather than on the parameters representing the partworths. The model in WTP space is given in Equation 8:

$$V = \beta_{price} \left[\frac{\beta_0}{\beta_{price}} NoBuy + Price + \frac{\beta_{EVOO}}{\beta_{price}} EVOO + \frac{\beta_{OO}}{\beta_{price}} OO + \frac{\beta_{CAT}}{\beta_{price}} CAT + \frac{\beta_{Mant}}{\beta_{price}} Mant + \frac{\beta_{RSp}}{\beta_{price}} RSp \right] + \varepsilon \quad (8)$$

Equation (8) can be rewritten as:

$V =$

$$\beta_{price}[\theta_0 NoBuy + Price + \theta_{EVOO} EVOO + \theta_{OO} OO + \theta_{CAT} CAT + \theta_{Manf} Manf + \theta_{RSp} RSp] + \varepsilon \quad (9)$$

where $\theta_i = \beta_i / \beta_{price}$ represents the estimated individuals' WTP.

To assess the comparability of the different CA formats in terms of participants' WTP, the choice models were estimated, specifying the utility function as follows:

$V =$

$$\beta_{price}[\theta_0 NoBuy + Price + \theta_{EVOO} EVOO + \theta_{OO} OO + \theta_{CAT} CAT + \theta_{Manf} Manf + \theta_{RSp} RSp] + \gamma_{EVOO}(EVOO * dtreat) + \gamma_{oo}(OO * dtreat) + \gamma_{cat}(CAT * dtreat) + \gamma_{Manf}(Manf * dtreat) + \gamma_{RSp}(RSp * dtreat) + \varepsilon \quad (10)$$

where $dtreat$ is coded as "1" for the first treatment and "0" for the second treatment. For example, if NHCE and NHRCA are the treatments to be compared, $dtreat$ is coded as "1" if the treatment is NHCE, and "0" if the treatment is NHRCA. The significance and the signs of the estimated γ parameters are then used to test the comparability in the different CA formats in terms of the estimated WTP space.

Consistency and external validity

As mentioned previously, to assess participants' responses consistency across treatments, the fifth choice set was repeated at the end of the main task⁵. To measure the consistency of participants' responses in each treatment, the proportion of participants who gave the same response in the fifth and the 10th choice sets, i.e., the repeated choice sets, was calculated. The response was counted as a hit if it was found to be the same in the fifth and the 10th choice sets. Then, the hit rate was calculated by dividing the total number of hits by the total number of participants in each treatment. To compare the hit rates across treatments, the Z-test for independent samples was used. Finally, to assess the external validity of the estimates, the estimated partworths from the main task were used to predict participant responses in the holdout task. The predicted and the actual choice in the holdout task were compared to determine the hit rate. The Z-test for independent samples was used to compare the hit rates across treatments.

Results and discussion

Partworths

The means and the standard deviations of the estimated partworths in the different treatments are displayed in Table 2. In the first three columns of Table 2, we report the estimated partworths in NHCE, NHRRCAs, and NHRMBWS, where only the

⁵ Brower *et al.* (2010) found that respondents were significantly more confident and certain about their choice at the end of the choice task than at the beginning. Therefore, we think that repeating the fifth choice set at the end of the choice task rather at the beginning could increase the reliability of the test.

information on the most-preferred option is considered in the estimation of the partworths. The results displayed in the last two columns of Table 2, are from the estimation of RO-RPL, i.e., in NHRCA and NHMBWS, taking into account the full ranking information.

The results show that participants' preferences across CA formats were comparable in terms of the sign and significance of the estimated partworths. For instance, the results show that participants in all the treatments were more (less) likely to choose extra virgin (virgin) olive oil than non-virgin olive oil. Participants were also found to prefer manufacturer label over private label and local (Catalonia) olive oil over non-local olive oil (olive oil from Andalusia or the rest of Spain). Furthermore, the results show that participants in the different treatments preferred the attribute price to take lower levels (cheaper olive oils are generally preferred). Finally, the negative and significant sign of the "No Buy" coefficient shows that respondents preferred to buy olive oil rather than opting out and choosing the no-choice option. All the estimated standard deviations were statistically significant, showing that participants' preferences in all the treatments were heterogeneous. For example, based on the underlying normal distribution specification (Train 2003), it can be calculated that only about 30%, 29%, 33%, 27%, and 27% of participants in NHCE, NHRCA, NHRMBWS, NHRCA, and NHMBWS, respectively, did not value extra virgin oil more than virgin oil.

Although the estimates from the five models have the same signs, some differences in the means and standard deviations of the partworths can be noticed. As previously mentioned, the procedure used by De-Magistris *et al.* (2013) was followed to test whether the differences in the partworths' values are statistically significant.

The results of the estimated coefficients (γ) that correspond to the variables “dtreat_x” (e.g. dtreat_{NHCE}) are displayed in Tables 3 and 4. The results show that when only the information on the most-preferred option is considered in the estimation of the partworts (see Table 3), few statistically significant differences are detected. Particularly, six out of 18 estimated partworts were statistically different across the treatments—NHCE, NHRRCa and NHRMBWS. Nonetheless, when the full ranking information is considered in the estimation of the partworts, we found that 23 out of 42 estimated partworts were statistically different across treatments.

These aforementioned findings are in line with the findings of Louviere *et al.* (2008); Lusk *et al.* (2008); Chang *et al.* (2009) and Lancsar *et al.* (2013), who showed that considering the additional information collected in RCA and BWS in the estimation of the partworts is likely to lead to different estimates than those obtained in CE. Furthermore, Akaichi *et al.* (2013) did not find differences between NHCE and NHRRCa in terms of the estimated partworts when the choice sets provided to respondents were small (four alternatives); however, they found discrepancies between respondents’ partworts in NHCE and NHRRCa when large choice sets (eight alternatives) were used. In our study, choice sets of five alternatives were used, which may explain the few significant differences found between NHCE, NHRRCa, and NHRMBWS in terms of estimated partworts.

Response consistency and external validity

Results from response consistency and external validity analyses are displayed in Table 5. In general terms, the consistency of participants’ responses is relatively high in the three CA formats. The hit rate ranges from 76.36% to 78.18% when only the most preferred option is considered in the estimation. Results from the one-tailed Z-

test show that the differences between NHCE, NHRCA, and NHRMBWS in terms of response consistency are not statistically significant. Nonetheless, we found that when the full ranking information was considered in the estimation of the partworths, the consistency's hit rate decreased significantly to 49.09% and 45.45%, in NHRCA and NHMBWS, respectively. This phenomenon could be explained by the fact that the stability of the ranking information decreased when the number of options to be ranked increased, due to the higher cognitive effort spent in RCA and BWS in ranking all the different product concepts (Ben-Akiva *et al.* 1992). The results in Table 5 show that response consistency is significantly lower in NHRCA and NHMBWS, than in NHCE. This result implies that in non-hypothetical settings, CE performs better in terms of response consistency, than RCA and BWS. Furthermore, the consistency of participant responses was found to be similar between NHRCA and NHMBWS. Therefore, the way respondents are asked to rank the different options in a choice set does not appear to alter consistency of their responses in NHRCA and NHMBWS.

Regarding the external validity, the results show that the estimated partworths accurately predicted between 38.18% and 61.81% of participant responses in the holdout task across the treatments. Consistent with the findings of Akaichi *et al.* (2013), we found that when only information on the most-preferred/ranked first product concept was used in the estimation of partworths, the external validity in the non-hypothetical choice and ranking CA formats (NHCE, NHRCA and NHRMBWS) is similar. Notably, the most striking result in our study is that when the full ranking information is considered in the estimation of partworths, the external validity is significantly higher in NHMBWS than in NHRCA and NHCE. The additional information gained from ranking of all the options in each choice set as

well as the easier handling of NHMBWS by respondents in comparison with NHRCA (Marley and Louviere 2005; Vermuelen *et al.* 2010; Potoglou *et al.* 2011; Flynn and Marley 2012) are probably the reasons behind the superiority of NHMBWS in terms of external validity. Certainly, more research work is needed to confirm our finding and explain the sources of the differences and the similarities between the three CA formats in terms estimated partworths and external validity

Willingness to pay

One of the main reasons for using discrete choice experiments is to estimate consumers' WTP for specific food attributes. Therefore, we also assessed the comparability of the three CA formats in terms of respondents' WTP. The results that correspond to the estimated WTP space are displayed in Table 6. These results show that the sign and the significance of the estimated WTP are similar across the three CA formats. Particularly, we found that participants in the three treatments were willing to pay a price premium for extra virgin olive oil. This result confirms the results of Bernabeu *et al.* (2009) who found that Spanish consumers are willing to pay, on average, 13% for the extra virgin olive oil over the conventional olive oil. Furthermore, in line with Yangui *et al.* (2014), the results also show that respondents were willing to pay more for olive oil with manufacturer label and local olive oil, than olive oil with private label and non-local olive oil. In all the treatments, participants were found to value the type of olive oil more than its origin, which in turn was more valued than the brand of olive oil (manufacturer vs. private). Importantly, all the standard deviations were statistically significant, implying that participants' WTP were heterogeneous. For example, Based on the underlying normal distribution

specification (Train 2003), it was calculated that the estimated WTP of 28%, 31%, and 34% of participants in NHCE, NHRCA, and NHMBWS, respectively, is lower for extra virgin than for virgin olive oil.

The results displayed in Table 6 show some differences across treatments in terms of estimated WTP, e.g., the price premium for the extra virgin olive oil is 0.98€ in NHCE and 1.95€ in NHRCA, despite the similarity of the estimates' sign and statistical significance. Furthermore, we estimated all the interactions between each one of the non-price attributes and a dummy variable that controls for treatment effect to test whether these apparent differences across the various CA formats were statistically significant. The results are displayed in Tables 7 and 8, showing that when only the information on the most-preferred option is considered in the estimation of participants' WTP (see Table 7), then only one out of 15 estimated WTP was statistically different across the treatments (NHCE, NHRCA, and NHMBWS). These results corroborate those of Akaichi *et al.* (2013) who found that participants in NHCE and NHRCA had statistically similar WTP, even when large choice sets were used; likewise, Caparros *et al.* (2008) found similar results.

Interestingly, the results displayed in Table 8 show that when the full ranking information is considered in the estimation of participants' WTP, then only five out of 35 estimated WTP values were statistically different across treatments. None of the previous studies, that compared different CA formats and considered the full ranking information in the estimation of the parameters, assessed the comparability of the CA formats in terms of participants' WTP. Future research could potentially answer the question of why NHCE, NHRCA, and NHMBWS produce different partworths and have different external validities, but are comparable in terms of estimated WTP.

Conclusions

Consumers' preferences and WTP for private and public goods are increasingly quantified using different conjoint analysis formats such as DCE, RCA, and BWS. The empirical comparison of these preference and value-elicitation methods is important because the three different approaches have strengths and weaknesses—neither can be ruled out a priori as inferior. This study expands the work that has been conducted in previous studies by assessing the comparability of DCE, RCA, and MBWS when used in non-hypothetical settings.

Our results show that the three CA formats provide similar results in terms of the sign and significance of estimated partworths, as well as the estimated WTP values, in almost all cases independently of whether the partworths are estimated by considering only partial or full ranking information. Accordingly, our results suggest that if the estimation of consumers' WTP is the study's main objective, then using any one of the three CA formats would be appropriate. Particularly, the use of NHCE might be preferred because it is simpler to implement and it involves less cognitive burden for respondents, as compared to NHRCA and NHMBWS. Nonetheless, in light of the fact that estimating both consumers' preferences and WTP is of great interest in the majority of studies on consumer choice, our results suggest the use of NHMBWS over NHCE and NHRCA because it has a better predictive power of consumer preference and provides estimated WTP values that are similar to those obtained in the other CA formats.

As in any other empirical study, our work has some gaps that could be filled by future studies. For instance, more research is needed on the extent to which the comparability of the output obtained in NHCE, NHRCA, and NHMBWS is affected

by experimental design parameters such as the size and the number of choice sets, the sample size, and the types of respondents and products. Specifically, further investigation is necessary into the sensitivity of our results to the effects of other factors such as the non-attendance of attributes (Chalak *et al.* 2016), participants' certainty about their choices (Rose *et al.* 2015), and the effect altruism and "free riding" (Lusk and Shogren 2007).

References

- Ajzen, I. and Madden, T.J. (1986). Prediction of goal directed behavior: attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology* 22, 453-474.
- Akaichi, F., Nayga, R.M.Jr. and Gil. J.M. (2013). Are Results from Non-Hypothetical Choice-Based Conjoint Analysis and Non-Hypothetical Recoded-Ranking Conjoint Analysis Similar? *American Journal of Agricultural Economics* 95, 949-963.
- Alfnes, F., Guttormsen, A., Steine, G. and Kolstad, K. (2006). Consumers' willingness to Pay for the Color of Salmon: A Discrete choice experiment with Real Economic Incentives. *American Journal of Agricultural Economics* 88, 1050-1061.
- Ben-Akiva, M., Morikawa, T. and Shiroishi, F. (1992). Analysis of the reliability of preferences ranking data. *Journal of Business Research* 24, 149-164.
- Bernabéu, R., Olmeda, M., Díaz, M. and Olivas, R. (2009). Oportunidades comerciales para el aceite de olive de Castilla-La Mancha. *Grasas y Aceites* 60, 525-533.

- Bliemer, M.C.J. and Rose. J.M. (2009). Efficiency and sample size requirement for stated discrete choice experiments” Paper presented in Transportation Research Board 88th Annual Meeting, January 11-15, Washington, USA.
- Boyle, K., Thomas, J., Holmes, P. Teisl, M.F. and Roe. B. (2001). A Comparison of Conjoint Analysis Response Formats. *American Journal of Agricultural Economics* 83, 441-454.
- Brouwer, R. Dekker, T., Rolfe, J. and Windle, J. (2010). Choice certainty and consistency in repeated discrete choice experiments. *Environmental Resource Economics* 46, 93-109.
- Brown, T.C., Kingsley, D. Peterson, G.L., Flores, N.E., Clarke, A. and Birjulin, A. (2008). Reliability of individual valuations of public and private goods: choice consistency, response time, and preference refinement. *Journal of Public Economic* 92, 1595-1606.
- Campbell, D. and Lorimer. V.S. (2009). Accommodating attribute processing strategies in stated choice analysis: do respondents do what they say they do?” Paper presented at European Association of Environmental and Resources Economist Annual Conference, June 24-27, Amsterdam, Netherland.
- Caparrós, A., Oviedo, J.L. and Campos, P. (2008). Would you Choose your Preferred option? Comparing Choice and Recoded Ranking Experiments. *American Journal of Agricultural Economics* 90, 843-855.
- Chalak, A., Abiad, M. and Balcombe, K. (2016). Joint use of attribute importance rankings and non-attendance data in discrete choice experiments. *European Review of Agricultural Economics* 43, 1-24.

- Chang, J.B., Lusk, J.L. and Norwood, B.F. (2009). How closely do hypothetical surveys and laboratory experiments predict field behavior? *American Journal of Agriculture Economics* 96, 518-534.
- Corrigan, J.R., Depositario, D.P.T., Nayga, R.M.Jr., Wu, X. and Laude, T.P. (2009). Comparing open-ended discrete choice experiments and experimental auctions: An application to golden rice. *American Journal of Agriculture Economic* 91, 837-853.
- De Bekker-Grob, E.W., Ryan, M. and Gerard, K. (2012). Discrete discrete choice experiment in health economics: a review of the literature. *Health Economics* 21, 145-172.
- Dekhili, S. and d'Hauteville, F. (2009). Effect of the region of origin on the perceived quality of olive oil: an experimental approach using a control group. *Food Quality and Preference*. 20, 525-532.
- Dekhili, S., Sirieix, L. and Cohen, E. (2011). How consumers choose olive oil: the importance of origin cues. *Food Quality and Preference*. 22, 757-762.
- De-Magistris, T., Gracia, A. and Nayga, R.M.Jr. (2013). On the use of honesty priming tasks to mitigate hypothetical bias in discrete choice experiments. *American Journal of Agricultural Economics* 95, 1136-1154.
- Ding, M., Grewal R. and Liechty, J. (2005). Incentive-Aligned Conjoint Analysis. *Journal of Marketing Research* 42, 67-83.
- Ding, M. (2007). An incentive-aligned mechanism for conjoint analysis. *Journal of Marketing Research* 44, 214-223.
- Dong, S., Ding, M. and Huber, J. (2010). A simple mechanism to incentive-align conjoint experiment. *International Journal of Research in Marketing* 27, 25-32.

- Finn, A. and Louviere. J.J. (1992). Determining the appropriate response to evidence of public concern: The case of food safety. *Journal of Public Policy & Marketing* 11, 12-25.
- Flynn, T.N. and Marley, A.J. (2014). Best worst scaling: Theory and methods. In: Hess S, and Daly A (eds) *Handbook of choice modeling*, pp. 178-201, Edward Elgar Publishing, Cheltenham.
- Flynn, T.N., Louviere, J.J., Peters, T.J. and Coast, J. (2010). Using discrete discrete choice experiments to understand preferences for quality of life. Variance-scale heterogeneity matters. *Social Science and Medicine* 70, 1957-1965.
- Greene W.H., Hensher, D.A. and Rose. J. (2006). Accounting for heterogeneity in the variance of unobserved effects in mixed logit models. *Transportation Research B* 40, 75-92.
- Greene, W.H., and Hensher, D.A. (2013). Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. *Applied Economics* 45, 1897-1902.
- Gruber, V., and Schlegelmilch. B.B. (2014). How techniques of neutralization legitimize norm-and attitude inconsistent consumer behavior. *Journal of Business Ethics* 121, 29-45.
- Hanley, N.D. and Ruffell, R.J. (1993). The contingent valuation of forest characteristics: two experiments. *Journal of Agricultural Economics* 44, 218-229.
- Hensher, D.A. and Greene, W.H. (2003). Mixed logit models: state of practice. *Transportation* 30, 33-176.

- Hoeffler, S. and Ariely, D. (1999). Constructing stable preferences: a look into dimensions of experience and their impact on preferences stability. *Journal of Consumer Psychology* 8, 113-126.
- Holmes, T.P. and Boyle, K.J. (2001). Cross validation of conjoint ranking and choice data: an application to timber harvesting preferences. Paper presented at EAERE 11th annual conference, June 28-30, Southampton, United Kingdom.
- Lagerkvist, C.J. (2013). Consumer preferences for food labeling attributes: Comparing direct ranking and best-worst scaling for measurement of attribute importance, preference intensity and attribute dominance. *Food Quality and Preference* 29, 77-88.
- Lancaster, K.J. (1966). A new approach to consumer theory. *The journal of political economy* 74, 132-157.
- Lancsar, E., Louviere, J.J., Donaldson, C., Currie, G. and Burgess, L. (2013). Best worst discrete choice experiment in health: Methods and an application. *Social Science and Medicine* 76, 74-82.
- List, J.A. and Gallet, C.A. (2001). What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values? *Environmental and Resources Economics* 20, 241-254.
- Little, J. and Berrens, R. (2004). Explaining Disparities Between Actual and Hypothetical Stated Values: Further Investigation Using Meta-analysis. *Economics Bulletin* 3, 1-13.
- Louviere, J.J. and Street, D. (2000). Stated-preference Methods. In: Hensher, D.A. and Button, K.J. (eds) *Handbook of Transport Modelling*, pp. 131-143, Pergamon Press, Amsterdam.

- Louviere, J.J., Street, D., Burgess, L., Wasi, N., Towdihul I. and Marley, A.J. (2008). Modeling the choices of individual decision-makers by combining efficient discrete choice experiment designs with extra preference information. *Journal of Choice Modelling* 1, 128-163.
- Louviere, J.J., Flynn, T.N. and Carson, R.T. (2010). Discrete discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling* 3, 57-72.
- Lusk, J.L. and Shogren, J.F. (2007). *Experimental auction: Methods and applications in economic and marketing research.* Cambridge University Press. Cambridge, UK.
- Lusk, J.L. and Schroeder, T.C. (2004). Are Discrete choice experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks. *American Journal of Agricultural Economics* 86, 467–82.
- Lusk, J.L., Fields, D. and Prevatt, W. (2008). An Incentive Compatible Conjoint Ranking Mechanism. *American Journal of Agricultural Economics* 90, 487–498.
- Marley, A.J. and Louviere, J.J. (2005). Some probabilistic models of best, worst and best_worstchoices. *Journal of Mathematical Psychology* 49, 464-480.
- McFadden, D. and Train, K. (2000). Mixed MNL models for discrete reponse. *Journal of Applied Econometrics* 15, 447-470.
- McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In Zambreka P. *Frontiers in Econometrics*, pp. 105-142, Academic press, Nueva York.
- Meijer, E. and Rouwendal, J. (2006). Measuring welfare effects in models with random coefficients. *Journal of Applied Econometrics* 21, 227-244.

- Menapace, L., Colson, G., Grebitus, C. and Facendola, M. (2011). Consumers' preferences for geographical origin labels: evidence from the Canadian olive oil market. *European Review of Agricultural Economics* 38, 193–212.
- Morrison, M.D. and Boyle, K.J. (2001). Comparative reliability of rank and choice data in stated preferences models.” Paper presented at EAERE 11th annual conference, June 28-30, Southampton, UK.
- Moser, R., Raffaelli, R. and Nataro, S. (2014). Testing hypothetical bias with a real discrete choice experiment using respondents' own money. *European Review of Agriculture Economics* 41, 25-46.
- Murphy, J.J., Geoffrey, A.P., Stevens, T.H. and Weatherhead, D. (2005). Meta-Analysis of Hypothetical Bias in Stated Preference Valuation. *Journal of Environmental and Resource Economics* 30, 313–325.
- Gollwitzer, P.M. (1999). Implementation intentions. *American Psychologist* 54, 493-503.
- Pignone, M.P., Brenner, A.T., Hawley, S. Sheridan, S.L., Lewis, C.L., Jonas, D.E. and Howard, K. (2011). Conjoint analysis versus rating and ranking for values elicitation and clarification in colorectal cancer screening. *Journal of general Internal Medicine*. 27, 45-50.
- Potoglou, D., Burge, P., Flynn, T. Netten, A., Malley, J., Forder, J. and Brazier, J.E. (2011). Best-worst scaling vs. discrete discrete choice experiments: An empirical comparison using social care data. *Social Science & Medicine* 72, 1717-1727.

- Rose, J.M., Beck, M.J. and Hensher, D.A. (2015). The joint estimation of respondent-reported certainty and acceptability with choice. *Transportation Research Part A: Policy and Practice* 71, 141-152.
- Rose, J.M. and Bliemer, M.C.J. (2005a). Sample Optimality in the Design of Stated Discrete choice experiments. Working paper ITLS-WP-05-13. Institute of Transport and Logistics Studies, University of Sydney.
- Russel, B.W. (1975). The objective situation as a determinant of consumer behavior. *Advances in Consumer research* 2, 427-438.
- Scarpa, R. Notaro, S., Louviere, J.J. and Raffaelli, R. (2011). Exploring scale effects of best/worst Rank ordered choice data to estimate benefits of tourism in alpine grazing commons. *American Journal of Agricultural Economics* 93, 813-828.
- Scarpa, R. and Del Giudice, T. (2004). Market segmentation via mixed Logit: extra-virgin olive oil in urban Italy. *Journal of Agricultural and Food Industrial Organization*. 2, 1-18.
- Sheeran, P., Trafimow, D. and Armitage, C.J. (2003). Predicting behavior from perceived behavioural control: tests of the Accuracy assumption of the theory of planned behaviour. *British Journal of Social Psychology* 42, 393-410.
- Siikamäki, J. and Layton, D.F. (2007). Discrete choice survey experiments: a comparison using flexible methods. *Journal of Environmental Economics and Management* 53, 122-139.
- Street, D. and Burgess, L. (2007). *The Construction of Optimal Stated Discrete choice experiments: Theory and Methods*. A John Wiley & sons, Publication, New Jersey.

- Swait, J. and Louviere, J.J. (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research* 30, 305-14.
- Train, K. and Weeks, M. (2004). Discrete choice models in preference space and willingness-to-pay space. Working paper, Faculty of Economics, University of Cambridge.
- Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, UK.
- Vermeulen, B., Goos, P. and Vandebroek, M. (2010). Obtaining more information from conjoint experiments by best–worst choices. *Computational Statistics & Data Analysis* 54, 1426-1433.
- Yangui, A., Costa Font, M. and Gil, J.M. (2014). Revealing additional preference heterogeneity with an extended random parameter logit model: the case of extra virgin olive oil. *Spanish journal of agricultural research* 12, 553-567.

Table 1. Results from Preference Regularity’s Tests

Test for preference regularity	Number of observations	Log Likelihood	Likelihood Ratio (LR)	Degrees of freedom	p-value
All treatments	8910	-1868.23			
NHCE [□]	2970	-591.45			
NHRRCA [□]	2970	- 619.10			
NHRMBWS [§]	2970	- 575.72			
<i>H₀:test of equality between non-hypothetical first choice option</i>			163.89	36	P<0.005
All treatments	5940	-1228.15			
NHRRCA	2970	- 619.10			
NHRMBWS	2970	- 575.72			
<i>H₀:test of equality</i>			66.64	12	P<0.005

between non-hypothetical NHRCA and NHRMBWS

All treatments	14265	-3457.67
NHRCA [¶]	7155	-1745.57
NHMBWS [□]	7110	-1678.66

H₀:test of equality

<i>between non-hypothetical NHRCA and NHMBWS</i>	66.87	12	P<0.005
---	-------	----	---------

[□] non-hypothetical discrete choice experiment treatment; [□] non-hypothetical ranking conjoint analysis treatment recoded as a traditional discrete choice experiment; [§] non-hypothetical best worst scaling treatment recoded as a traditional discrete choice experiment; [¶] non-hypothetical ranking conjoint analysis treatment; [□] non-hypothetical best worst scaling treatment.

Table 2. Estimated Partworths

Estimated parameters and standard deviations	Partial ranking information			Full ranking information	
	NHCE [□]	NHRRCA [□]	NHRMBWS [§]	NHRCA [¶]	NHMBWS [□]
Random parameters' estimates					
Extra virgin olive oil (EVOO)	0.925 ***	0.794 ***	0.988 ***	0.665 ***	0.687 ***
Olive oil (OO)	-1.021 ***	-0.883 ***	-0.876 ***	-0.728 ***	-0.656 ***
Manufacturer label (Manf)	0.213 **	0.158 **	0.187 **	0.131 **	0.145 ***
Catalonia (CAT)	0.911 ***	0.553 ***	0.940 ***	0.274 ***	0.612 ***
Rest of Spain (RSp)	-0.577 ***	-0.666 ***	-0.545 ***	-0.424 ***	-0.639 ***
Price	-1.213 ***	-0.697 ***	-1.089 ***	-0.447 ***	-0.834 ***
NoBuy	-3.340 ***	-5.050 ***	-5.015 ***	-4.641 ***	-4.437 ***
Standards deviations of the random parameters					
EVOO	1.772 ***	1.460 ***	2.195 ***	1.048 ***	1.129 ***
OO	1.715 ***	1.365 ***	2.638 ***	0.930 ***	1.324 ***
Manf	0.283 **	0.224 **	0.410 **	0.350 ***	0.156 ***
CAT	1.329 ***	0.689 ***	1.143 ***	0.630 ***	0.597 ***
RSp	0.992 ***	0.502 **	0.639 ***	0.596 ***	0.489 ***
Price	0.497 ***	1.153 ***	1.128 ***	0.945 ***	0.894 ***
Number of observations	2970	2970	2970	7155	7110
Log-likelihood	-591.4546	-619.1057	-575.7293	-1745.575	-1678.663
Log-likelihood ratio	283.08	287.53	308.24	899.49	896.86

(***) and (**) denote statistical significance at 1% and (5%) level, respectively; [□] non-hypothetical discrete choice experiment treatment; [□] non-hypothetical ranking conjoint analysis treatment recoded as a discrete choice experiment; [§] non-hypothetical best worst scaling treatment recoded as a discrete choice experiment; [¶] non-hypothetical ranking conjoint analysis treatment; [□] non-hypothetical best worst scaling treatment.

Table 3. Hypothesis Test of Equality of Preferences Parameters across the Treatments Taking into Account only Information on the Most Preferred Option

Treatments	Variables	Estimated γ
NHCE vs. NHRRCA	EVOO x dtreat _{NHCE}	-0.645 *
	OO x dtreat _{NHCE}	0.530
	Manf x dtreat _{NHCE}	0.192 ***
	CAT x dtreat _{NHCE}	0.423
	RSp x dtreat _{NHCE}	0.033
	Price x dtreat _{NHCE}	-0.453 **
NHCE vs. NHRMBWS	EVOO x dtreat _{NHCE}	0.053
	OO x dtreat _{NHCE}	-0.337
	Manf x dtreat _{NHCE}	0.185 ***
	CAT x dtreat _{NHCE}	0.156
	RSp x dtreat _{NHCE}	-0.134
	Price x dtreat _{NHCE}	-0.327 ***
NHRRCA vs. NHRMBWS	EVOO x dtreat _{NHRRCA}	-0.230
	OO x dtreat _{NHRRCA}	-0.004
	Manf x dtreat _{NHRRCA}	0.185 ***
	CAT x dtreat _{NHRRCA}	-0.235
	RSp x dtreat _{NHRRCA}	0.043
	Price x dtreat _{NHRRCA}	0.039

(***), (**) and (*) denote statistical significance at (1%), (5%) and (10%) level, respectively.

Table 4. Hypothesis Test of Equality of Preferences Parameters across the Treatments Taking into Account the Full Ranking Information

Treatments	Variables	Estimated γ
NHRCA vs NHCE	<i>EVOO X</i>	0.257
	<i>dtreat_{NHRCA}</i>	
	<i>OO X dtreat_{NHRCA}</i>	-0.451 ***
	<i>Manf X</i>	0.127 ***
	<i>dtreat_{NHRCA}</i>	
	<i>CAT X dtreat_{NHRCA}</i>	-0.417 ***
	<i>RSp X dtreat_{NHRCA}</i>	0.064
	<i>Price X</i>	0.784 ***
NHMBWS vs NHCE	<i>EVOO X</i>	0.153
	<i>dtreat_{NHMBWS}</i>	
	<i>OO X</i>	-0.036
	<i>dtreat_{NHMBWS}</i>	
	<i>Manf X</i>	0.131 ***
	<i>dtreat_{NHMBWS}</i>	
	<i>CAT X</i>	-0.252 *
	<i>dtreat_{NHMBWS}</i>	
	<i>RSp X</i>	0.030
	<i>dtreat_{NHMBWS}</i>	
<i>Price X</i>	0.673 ***	
NHRCA vs NHRRCA	<i>EVOO X</i>	-0.116
	<i>dtreat_{NHRCA}</i>	
	<i>OO X dtreat_{NHRCA}</i>	-0.083
	<i>Manf X</i>	0.131 ***
	<i>dtreat_{NHRCA}</i>	
	<i>CAT X dtreat_{NHRCA}</i>	-0.197
	<i>RSp X dtreat_{NHRCA}</i>	0.363 ***
	<i>Price X</i>	-0.045
NHMBWS vs NHRRCA	<i>EVOO X</i>	-0.596 ***
	<i>dtreat_{NHMBWS}</i>	
	<i>OO X</i>	0.793 ***
	<i>dtreat_{NHMBWS}</i>	
	<i>Manf X</i>	0.136 ***
	<i>dtreat_{NHMBWS}</i>	
	<i>CAT X</i>	0.347 ***
	<i>dtreat_{NHMBWS}</i>	
	<i>RSp X</i>	-0.020
	<i>dtreat_{NHMBWS}</i>	
<i>Price X</i>	-0.069	
NHRCA vs NHRMBWS	<i>EVOO X</i>	0.165
	<i>dtreat_{NHRCA}</i>	
	<i>OO X dtreat_{NHRCA}</i>	-0.569 ***

	<i>Manf X</i>	0.131	***
	<i>dtreat_{NHRCA}</i>		
	<i>CAT X dtreat_{NHRCA}</i>	-0.513	***
	<i>RSp X dtreat_{NHRCA}</i>	0.255	*
	<i>Price X</i>		
	<i>dtreat_{NHRCA}</i>	0.085	
	<i>EVOO X</i>		
	<i>dtreat_{NHMBWS}</i>	0.054	
	<i>OO X</i>		
	<i>dtreat_{NHMBWS}</i>	-0.245	
	<i>Manf X</i>	0.125	***
	<i>dtreat_{NHMBWS}</i>		
	<i>CAT X</i>	0.077	
	<i>dtreat_{NHMBWS}</i>		
	<i>RSp X</i>		
	<i>dtreat_{NHMBWS}</i>	-0.057	
	<i>Price X</i>		
	<i>dtreat_{NHMBWS}</i>	-0.156	
	<i>EVOO X</i>	0.167	*
	<i>dtreat_{NHRCA}</i>		
	<i>OO X dtreat_{NHRCA}</i>	-0.299	***
	<i>Manf X</i>	0.128	***
	<i>dtreat_{NHRCA}</i>		
	<i>CAT X dtreat_{NHRCA}</i>	-0.236	**
	<i>RSp X dtreat_{NHRCA}</i>	0.157	*
	<i>Price X</i>		
	<i>dtreat_{NHRCA}</i>	-0.097	
NHMBWS vs NHRMBWS			
NHRCA vs NHMBWS			

(***), (**) and (*) denote statistical significance at (1%), (5%) and (10%) level, respectively.

Table 5. Consistency and External Validity Tests

Treatments (Treatment1 vs. Treatment2)	Consistency (hit rate (%))			External validity (hit rate (%))		
	Treatment1	Treatment2	<i>P-value</i>	Treatment1	Treatment2	<i>P-value</i>
NHCE vs. NHRRCA	78.18	76.36	0.41	43.63	38.18	0.28
NHCE vs. NHRMBWS	78.18	76.36	0.41	43.63	52.72	0.17
NHCE vs. NHRCA	78.18	49.09	0.00	43.63	40.00	0.35
NHCE vs. NHMBWS	78.18	45.45	0.00	43.63	61.81	0.03
NHRRCA vs. NHRMBWS	76.36	76.36	0.50	38.18	52.72	0.06
NHRRCA vs. NHRCA	76.36	49.09	0.00	38.18	40.00	0.42
NHRRCA vs. NHMBWS	76.36	45.45	0.00	38.18	61.81	0.00
NHRMBWS vs. NHRCA	76.36	49.09	0.00	52.72	40.00	0.90
NHRMBWS vs. NHMBWS	76.36	45.45	0.00	52.72	61.81	0.17
NHRCA vs. NHMBWS	49.09	45.45	0.35	40.00	61.81	0.01

Table 6. Estimated WTP space in the Different Treatments

Estimated parameters and standard deviations	Partial ranking information			Full ranking information	
	NHCE [□]	NHRRCA [□]	NHRMBWS [§]	NHRCA [¶]	NHMBWS [□]
Random parameters' estimates					
Extra virgin olive oil (EVOO)	0.989 ***	1.959 ***	1.495 ***	1.351 ***	1.052 ***
Olive oil (OO)	-0.854 ***	-2.125 ***	-1.444 ***	-1.579 ***	-0.949 ***
Manufacturer label (Manf)	0.200 **	0.372 **	0.211 *	0.394 **	0.236 ***
Catalonia (CAT)	0.760 ***	0.678 ***	0.945 ***	1.032 ***	0.880 ***
Rest of Spain (RSp)	-0.563 ***	-1.150 ***	-0.807 ***	-1.290 ***	-0.978 ***
Standards deviations of the random parameters					
EVOO	1.684 ***	2.560 ***	2.065 ***	3.278 ***	2.133 ***
OO	1.650 ***	2.531 ***	2.404 ***	2.728 ***	2.592 ***
Manf	0.316 **	0.625 **	0.436 **	1.168 ***	0.471 ***
CAT	1.257 ***	1.406 ***	1.076 ***	1.555 ***	1.127 ***
RSp	0.816 ***	1.132 ***	0.570 ***	1.391 ***	1.033 ***
Number of observations	2970	2970	2970	7155	7110
Log-likelihood	-627.7162	-702.4804	-642.3329	-1983.443	-1921.41

(***), (**), and (*) denote statistical significance at 1%, (5%) and (10%) level, respectively; [□] non-hypothetical discrete choice experiment treatment; [□] non-hypothetical ranking conjoint analysis treatment recoded as a discrete choice experiment; [§] non-hypothetical best worst scaling treatment recoded as a discrete choice experiment; [¶] non-hypothetical ranking conjoint analysis treatment; [□] non-hypothetical best worst scaling treatment.

Table 7. Hypothesis Test of Equality WTP values across the Treatments Taking into Account only Information on the Most Preferred Option

Treatments	Variables	Estimated γ
NHCE vs. NHRRCA	EVOO x dtreat _{NHCE}	-0.624
	OO x dtreat _{NHCE}	1.093 **
	Manf x dtreat _{NHCE}	-0.039
	CAT x dtreat _{NHCE}	0.481
	RSp x dtreat _{NHCE}	0.228
NHCE vs. NHRMBWS	EVOO x dtreat _{NHCE}	-0.467
	OO x dtreat _{NHCE}	0.326
	Manf x dtreat _{NHCE}	-0.008
	CAT x dtreat _{NHCE}	-0.183
	RSp x dtreat _{NHCE}	0.234
NHRRCA vs. NHRMBWS	EVOO x dtreat _{NHRRCA}	-0.243
	OO x dtreat _{NHRRCA}	-0.339
	Manf x dtreat _{NHRRCA}	0.060
	CAT x dtreat _{NHRRCA}	-0.481
	RSp x dtreat _{NHRRCA}	-0.070

(***), (**) and (*) denote statistical significance at (1%), (5%) and (10%) level, respectively.

Table 8. Hypothesis Test of Equality WTP Values across the Treatments Taking into Account the Full Ranking Information

Treatments	Variables	Estimated γ
NHRCA vs NHCE	<i>EVOO X dtreat_{NHRCA}</i>	0.319
	<i>OO X dtreat_{NHRCA}</i>	-0.826 **
	<i>Manf X dtreat_{NHRCA}</i>	0.072
	<i>CAT X dtreat_{NHRCA}</i>	-0.399
	<i>RSp X dtreat_{NHRCA}</i>	-0.345
NHMBWS vs NHCE	<i>EVOO X dtreat_{NHMBWS}</i>	-0.335
	<i>OO X dtreat_{NHMBWS}</i>	0.308
	<i>Manf X dtreat_{NHMBWS}</i>	-0.049
	<i>CAT X dtreat_{NHMBWS}</i>	-0.312
	<i>RSp X dtreat_{NHMBWS}</i>	-0.034
NHRCA vs NHRCA	<i>EVOO X dtreat_{NHRCA}</i>	-0.258
	<i>OO X dtreat_{NHRCA}</i>	0.065
	<i>Manf X dtreat_{NHRCA}</i>	-0.221
	<i>CAT X dtreat_{NHRCA}</i>	0.105
	<i>RSp X dtreat_{NHRCA}</i>	0.380
NHMBWS vs NHRCA	<i>EVOO X dtreat_{NHMBWS}</i>	-0.856
	<i>OO X dtreat_{NHMBWS}</i>	1.383 ***
	<i>Manf X dtreat_{NHMBWS}</i>	-0.207
	<i>CAT X dtreat_{NHMBWS}</i>	0.324
	<i>RSp X dtreat_{NHMBWS}</i>	0.183
NHRCA vs NHRMBWS	<i>EVOO X dtreat_{NHRCA}</i>	-1.045 **
	<i>OO X dtreat_{NHRCA}</i>	0.259
	<i>Manf X dtreat_{NHRCA}</i>	-0.168
	<i>CAT X dtreat_{NHRCA}</i>	-0.765 **
	<i>RSp X dtreat_{NHRCA}</i>	0.234
NHMBWS vs NHRMBWS	<i>EVOO X dtreat_{NHMBWS}</i>	-0.143
	<i>OO X dtreat_{NHMBWS}</i>	-0.202
	<i>Manf X dtreat_{NHMBWS}</i>	-0.106
	<i>CAT X dtreat_{NHMBWS}</i>	-0.110
	<i>RSp X dtreat_{NHMBWS}</i>	0.034
NHRCA vs NHMBWS	<i>EVOO X dtreat_{NHRCA}</i>	-1.146 ***
	<i>OO X dtreat_{NHRCA}</i>	0.613
	<i>Manf X dtreat_{NHRCA}</i>	0.223
	<i>CAT X dtreat_{NHRCA}</i>	-0.128
	<i>RSp X dtreat_{NHRCA}</i>	-0.070

(***), (**) and (*) denote statistical significance at (1%), (5%) and (10%) level, respectively.

Choice card ...			Identification Number:			
	Olive oil 1	Olive oil 2	Olive oil 3	Olive oil 4	Olive oil 5	Option "No Buy"
Type of olive oil	Virgin	Virgin	Olive oil	Extra virgin	Virgin	None of the five olive oils
Brand	Manufacturer label	Manufacturer label	Private label	Manufacturer label	Private label	
Origin	Catalonia	Andalusia	Catalonia	Catalonia	Rest of Spain	
Price (€)	2.20	4.80	2.20	3.50	2.20	
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>Please mark the olive oil you would purchase or mark the option "No buy" if you are not willing to purchase any of the five olive oils.</p>						

Figure 1. An example of a choice set provided to participants in NHCE

Choice card ...			Identification Number:			
	Olive oil 1	Olive oil 2	Olive oil 3	Olive oil 4	Olive oil 5	Option "No Buy"
Type of olive oil	Virgin	Virgin	Olive oil	Extra virgin	Virgin	None of the five olive oils
Brand	Manufacturer label	Manufacturer label	Private label	Manufacturer label	Private label	
Origin	Catalonia	Andalusia	Catalonia	Catalonia	Rest of Spain	
Price (€)	2.20	4.80	2.20	3.50	2.20	
	<input type="text" value="1"/> <input type="text" value="2"/> <input type="text" value="3"/> <input type="text" value="4"/> <input type="text" value="5"/>	<input type="text" value="1"/> <input type="text" value="2"/> <input type="text" value="3"/> <input type="text" value="4"/> <input type="text" value="5"/>	<input type="text" value="1"/> <input type="text" value="2"/> <input type="text" value="3"/> <input type="text" value="4"/> <input type="text" value="5"/>	<input type="text" value="1"/> <input type="text" value="2"/> <input type="text" value="3"/> <input type="text" value="4"/> <input type="text" value="5"/>	<input type="text" value="1"/> <input type="text" value="2"/> <input type="text" value="3"/> <input type="text" value="4"/> <input type="text" value="5"/>	<input type="text"/>
Please rank the five olive oils from the most to the least preferred olive oil or mark the option "No buy" if you don't like any of the five olive oils.						

Figure 2. An example of a choice set presented in NHRCA

Choice card ...			Identification Number:			
	Olive oil 1	Olive oil 2	Olive oil 3	Olive oil 4	Olive oil 5	Option "No Buy"
Type of olive oil	Virgin	Virgin	Olive oil	Extra virgin	Virgin	None of the five olive oils
Brand	Manufacturer label	Manufacturer label	Private label	Manufacturer label	Private label	
Origin	Catalonia	Andalusia	Catalonia	Catalonia	Rest of Spain	
Price (€)	2.20	4.80	2.20	3.50	2.20	
	<input type="checkbox"/> B <input type="checkbox"/> 2B <input type="checkbox"/> W <input type="checkbox"/> 2W	<input type="checkbox"/> B <input type="checkbox"/> 2B <input type="checkbox"/> W <input type="checkbox"/> 2W	<input type="checkbox"/> B <input type="checkbox"/> 2B <input type="checkbox"/> W <input type="checkbox"/> 2W	<input type="checkbox"/> B <input type="checkbox"/> 2B <input type="checkbox"/> W <input type="checkbox"/> 2W	<input type="checkbox"/> B <input type="checkbox"/> 2B <input type="checkbox"/> W <input type="checkbox"/> 2W	<input type="checkbox"/>
<p>From the five options, please indicate your best option marking the box "B". Then, of the remaining four options, please indicates your worst option marking the box "W". Then, of the remaining three options, please indicates your second best option marking the box "2B". Finally, from the remaining two options, please indicate your second worst option marking the box "2W". In case you don't like any of the five options, please mark the option "None".</p>						

Figure 3. An example of a choice set presented in NHBWS

Identification number:						
	Olive oil 1	Olive oil 2	Olive oil 3	Olive oil 4	Olive oil 5	Option "No buy"
Type of olive oil	Extra virgin	Virgin	Olive oil	Extra virgin	Virgin	None of the ten olive oils
Brand	Private	Private	Private	Manufacturer	Private	
Origin	Andalucia	Andalucia	Catalonia	Catalonia	Rest of Spain	
Price (€)	3.50	2.20	4.80	3.50	4.80	
	<input type="checkbox"/>					
	Olive oil 6	Olive oil 7	Olive oil 8	Olive oil 9	Olive oil 10	
Type of olive oil	Olive oil	Virgin	Extra virgin	Virgin	Olive oil	
Brand	Private	Manufacturer	Manufacturer	Manufacturer	Manufacturer	
Origin	Rest of Spain	Catalonia	Rest of Spain	Andalucia	Andalucia	
Price (€)	3.50	2.20	2.20	4.80	3.50	
	<input type="checkbox"/>					
Please mark the olive oil you would purchase or mark the option "No buy" if you are not willing to purchase any the ten olive oils.						

Figure 4. Example of a choice set presented in the Holdout task (training session)

