Are ranking preferences information methods comparable with the choice experiment information in predicting actual behavior?

Yangui, A[1]., Akaichi, F[2]., Costa-Font, M[1]., and Gil, J.M[1].

[1] CREDA-UPC-IRTA
Parc Mediterrani de la Tecnologia Edifici ESAB C/Esteve Terrades, 8 08860 Castelldefels, Barcelona, Spain.

[2] Scotland's Rural College (SRUC),
Land Economy, Environment & Society, King's Buildings, West Mains Road, Edinburgh EH9 3JG, UK

Paper prepared for presentation at the EAAE 2014 Congress
'Agri-Food and Rural Innovations for Healthier Societies'

August 26 to 29, 2014
Ljubljana, Slovenia

Are ranking preferences information methods comparable with the choice experiment information in predicting actual behavior?

1. Introduction

Since its introduction conjoint analysis (CA) has become one of the most popular marketing research tools (Lusk et al., 2008). The most widely used CA format to elicit consumers' preferences for market and non-market goods is choice experiment (CE). CE gained popularity thanks to its ability to mimic the real market setting where consumers who are faced with competing products purchase the product that fit most their preferences. However, it is informationally inefficient, since it only allows the observation of the most preferred option (Lusk et al., 2008).

Accordingly to Lancsar et al. (2013), there have been three ways to get more insights to explore individual preferences: 1) to increase the sample size; and/or 2) to ask the respondents to evaluate more choice sets; or 3) instead of more choice sets, to ask more questions per choice set. In this context, contrary to CE, participants in a ranking conjoint analysis (RCA) are also provided with a set of product concepts but they are asked to rank them from the most to the least preferred. Therefore, it might be more accurate since it provides information about consumers' preferences for all the product concepts included in a choice set. In same line, recently, Louviere et al. (2004) introduced another CA format named best worst scaling (BWS). The BWS approach consists in asking respondents to firstly choose the best option in each choice set, then 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 options is obtained. BWS tasks seem to be easy for people due to the human skills at identifying extremes (Flynn and Marley, 2012).

Despite the widely application of the different CA formats (i.e. CE, RCA and BWS) over the last two decades, few researchers, however, have compared their performance in terms of estimated marginal partworths, the predictive power of the derived models, and the reliability of the Willingness to Pay (WTP) values deduced from the estimated partworths. Most of the past literature provide good examples about assessing the incentive compatibility of CA and propose modified CA formats to incentivize subjects to truthfully reveal their preferences (Ding et al., 2005; Lusk et al., 2008; among others) or about comparing the predictive power of real purchase behavior of different CA formats (Caparros et al., 2008; Chang et al., 2009; Akaichi et al., 2013). Caparros et al. (2008) found that participants behave similarly when compared their responses in CE and RCA using just the most preferred option. Akaichi et al. (2013) confirmed this result and pointed out that differences in participants' responses increased when the number of alternatives in each choice set also increased. Taking into account all preferences ranking information of the participants, Chang et al. (2009) commented that new non-hypothetical RCA introduced by Lusk et al. (2008) significantly outperformed the hypothetical and non-hypothetical CE methods in predicting retail sales. However, none of the published studies on CA have compared, at the same time, the performance of the three CA formats (CE, RCA and BWS) which is one of the main novelties of this paper. The comparison is done in terms of the estimated partworths, predictive power and estimated WTP values taking into consideration not only the most preferred option but also the additional information obtained generated in RCA and BWS.

Despite the empirical findings that back up the criticism associated of the hypothetical bias, it is worth noting that almost of the studies that assessed the comparability of CA formats reported results obtained from economic experiments conducted in hypothetical settings (with the exception of Chang et al., 2009; and Akaichi et al., 2013). Additionally, the experiments involving BWS as the preference elicitation method were implemented in hypothetical conditions (Louviere et al. 2008; Scarpa et al. 2011; Lancsar et al. 2013). Therefore, due to the skepticism surrounding the validity of values obtained from hypothetical CA experiments, we have carried out the CE, the RCA and the BWS in a non-hypothetical setting. In any case, the hypothetical CE will be used as the benchmark.

Furthermore, to compare the external validity of the three CA formats, we have included a non-hypothetical holdout choice task in the experimental design in the three cases. Finally, one of the main assumption underlying stated preference methods is that respondents know their preferences and that these preference are stable and coherent (Brown et al., 2008). In this study, we compare the decision consistency of the respondents and the internal validity across the three response formats used. To tackle with this issue, respondents were presented with the same choice task at the beginning and at the end of the choice experiment.

To sum up, our study stands out by comparing the ability of three non-hypothetical CA response formats (CE, RCA, and BWS) in terms of estimated partworths, internal and external predictive power, estimated WTP, and participants' response consistency, in two contexts: 1) the additional information obtained in RCA and BWS is not taken into account and, hence, only the most preferred option is considered (RRCA and RBWS); and 2) the additional information is included and the corresponding econometric models are estimated (NHRCA and NHBWS). This will allow us to assess the comparability of the three CA formats before and after considering the additional information.

Our paper is structured into five sections. In the next section, the experiment design and the experimental procedures are described. The econometric model used to estimating the partworths is outlines in section 3. The results are discussed in the fourth section and we finish by drawing some concluding remarks.

2. Experiment design

In this study, four treatments were carried out, hypothetical CE (HHCE) (as the benchmark), non-hypothetical CE (NHCE), non-hypothetical RCA (NHRCA) and non-hypothetical BWS (NHBWS). To assess the differences between these treatments a sample of 220 Barcelona' citizens was recruited to evaluate their preferences towards the purchase of olive oil. The participants were randomly and equally distributed over the different treatments[1]. The main attributes and attribute levels were first identified from literature review and two focus groups carried out among highly experienced and low experienced olive oil consumers. Four attributes were selected, three with three levels: type of olive oil (virgin extra, virgin, and refined olive oil)[2], origin (Andalucía, Catalonia, and rest of Spain) and price (2.20 €/liter, 3.50 €/liter, and 4.80 €/liter, which account for 85% of the price distribution in retail outlets), and one with two levels: brand (Manufacturer brand and private brand).

The combination of attribute levels generates a total of 54 (3^3*2) one liter bottles of olive oil. Following the Street and Burgess (2007), an orthogonal fractional factorial design, taking into account only main effects, was generated to reduce the number of combinations, resulting in 9 choice products, which will be considered as the first option in each choice set. Four additional options were offered to respondents in each choice set (plus the no choice option) applying the following generators (1000), (1111), (2121), and (2122) on the orthogonal design obtained. This resulted in a 100% efficient design.

During each treatment, the participants did two main tasks. The first task consists of the main of either treatments (HHCE, NHCE, NHRCA or NHBWS). Respondents were offered 10 choice sets one by one (9 choice sets from the experimental design plus the fifth card which was presented at the end to assess the consistency and the internal validity). In each

[1] The participants were randomly recruited. Across the treatments, no significant differences at the 5% level were found in relation to gender and age. However, significant differences were found in terms of education level and self-reported income levels.

[2] The three types of olive oil were defined according to the International Olive Council (IOC). In this context, the refined olive oil is defined as the olive oil obtained from virgin olive oils by refining methods which do not lead to alterations. This are marketed as "olive oil". Respondents were aware about differences across levels.

choice set, the participant was asked to choice their preferred option or to rank the options based on their preferences, that is, taking into account their real purchase habits. The second task, named the holdout task, was carried out to determine the external validity of the estimated parameters obtained from the different elicitation methods. The holdout task is a choice exercise in nature in which each participant had to choose just one product from a choice set including 10 alternatives generated from the full factorial design and different to those used in the first task. Each treatment of the experiment was conducted over 5 sessions throughout both different days of the week and different hours of day. Each session includes a maximum of 10-15 persons. After the two tasks, the participants fulfilled a short questionnaire aimed at collecting socio-demographic and lexicographic characteristics of respondents as well as on attitudes and olive purchasing and consumption habits.

At the beginning of the experiment, participants were informed that they would receive 15 Euros in cash at the end of the experiment. Additionally, we explained them the functional mechanism of the assigned treatment. In the next section details about the experimental procedure of each treatment are presented.

Hypothetical (HHCE) and non-hypothetical choice experiment (NHCE)

In hypothetical CE, we ask the participant to assume that each choice set as a real shopping situation while (s)he had not to pay for any of the chosen products. In each choice set, the participants were asked to choose their most preferred option taking into account their real purchase habits. If participant did not like any product they can choose the "none of them" option. The non-hypothetical CE experiment was similar, but in this case we informed the participants that each choice set was a real shopping scenario. In fact, the participants could receive any of options they had selected across all choice sets and they should pay for it the posted price. In both cases, before to start the second task, participants were informed that this was a real shopping scenario.

After completing the two tasks and the survey, we asked for a volunteer to draw randomly a number between 1 and 2 to selecting the binding task. If the binding task was the main task, in the hypothetical setting, all the participants receive their money and the experiment finished. In the non-hypothetical CE, another volunteer was selected to randomly draw one of the 9 choice set[3] to determine which of the choice set will be the binding one. Hence, each participant receives his money and will buy the chosen option paying the corresponding price. In case, the participant chose the "none of them" option, (s)he received the money and did not buy any product. If the binding task is the holdout task, regardless the type of treatments (HHCE or NHCE), each participant had to buy the chosen option, paying the corresponding price. If the chosen option is the "none of them", the experiment finished for him(her).

Non-hypothetical rank conjoint analysis (NHRCA)

The same 10 choice sets were presented to each participant, who was asked to rank the options in each choice set from most to the least preferred option. In case the participant did not like any of presented alternatives, (s)he could choose the "none of them" option. The non hypothetical nature of the experiment was also revealed to participants since the beginning. After completing the main and the holdout task, a draw was made from a volunteer to select the binding task. If the main task was chosen as the binding task a volunteer draws the binding choice set. Following Lusk et al. (2008), to ensure us that the ranking treatment will be incentive compatible, the participant had to purchase the binding product with a probability proportional to the rank (s)he assigned. Then, each participant who did not choose the "none of them" option draws a number from 1 to 50 to select the biding product. If the number drawn was between 1 and 17, the participant should purchase the most preferred option and

[3] The last choice set (the number 10) was the same fifth choice set of the experimental design and was repeated at the end for assess the consistency and the internal validity. Therefore, for the equal probability to draw any choice set we will remove the tenth choice set.

pay for it the posted price, if the number drawn was between 18 and 30 the second most preferred option will be the biding product; if between 31 and 40 the participant should be purchase the third option in (her)his preference ranking; if number drawn was between 41 and 47, the participant bought (her)his fourth preferred option; and between 40 and 50, the participant should buy the least preferred option. If the binding task was the holdout task, the procedure was similar than in the HHCE and HHCE treatments.

Non-hypothetical best worst scaling (NHBWS)

Consistently with the previous treatments, the same 10 choice sets were presented to each participant in the main task. In this case, the participant was asked to choose firstly the most preferred option within the choice set, followed by the worst option of the four remaining options, followed by the second best option of the three remaining options, followed by the second worst option of the two remaining options. At the end of the day we obtains the preference ranking of each participant from the BWS treatment classifying the best option as the first option of the ranking, the second best option as the second option in the ranking, the third option of the ranking will be the remained option, the fourth option will be the second worst option, finally the last option of the ranking will be the first worst option. Once, the holdout task finished, the same procedure than in the NHRCA was followed to get the binding product.

3. Methodological approach

In our empirical specification, the deterministic component of the utility function is given by:

(1)

In (1) attributes levels (extra virgin olive oil (EVOO), olive oil (OO), Manufacturer Brand (BrManf), Catalonia (CAT) and rest of Spain (RSp)) were effect coded (-1, 0, 1)[4], except fir the price that was coded as a linear variable. The constant ASC represents the “none of them” option and has been coded as a dummy variable that takes the value 1 when the option “none of them” was chosen by the participant; and 0, otherwise. To estimate the utility function (1), the random parameter logit model (RPL) has been used. According to Lusk et al. (2008) to estimate the partworth estimates of the RPL on the ranking data (i.e. NHRCA and NHBWS), the exploded ranking data were converted into choices obtaining the so called rank order random parameter logit (RO-RPL) model.

This model assumes that the probability of a particular ranking of alternatives from an individual is the product of the multinomial choice probability for choosing always the best of the remaining options. That is, the probability (that an individual *i* rank the five options as A> B> C> D> E from a choice set of five options (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 involved (A, B, C, D, E), the probability of choosing B as the best from the remaining options (B, C, D, E), the probability of choosing C as the best from (C, D, E) and the probability of choosing D as the best option from (D, E). In other words:

(4)

It is worth noting that the estimation of (4) takes into account the unobserved effect of the correlation that could be exist between the attributes due the number of choices faced by each participant (Hensher et al., 2005). In fact, we assume that all the taste partworths of our empirical model are random and follow a normal distribution with mean and variance-covariance matrix , as they are not independently distributed.

[4] The attribute levels virgin olive oil (VOO), private brand (BrPRV), and Andalucía (AND) were considered as the baseline for the attributes type of olive oil, brand, and origin, respectively.

Taking into account that our study has considered alternative treatments that have been applied using different samples, it was imperative to test for preference regularity across treatments. As in Lusk and Schroeder (2004), the null hypothesis of the test is preference equality across treatments. The test statistic is of a likelihood ratio type (, and it is distributed as a with K(M-1) degrees of freedom, where is the log likelihood values of pooled data (e.g. HHCE plus NHCE data), is the log likelihood values of the estimated model for each treatment, K is the number of restrictions, and M is the number of treatments (Louviere et al., 2000).

Willingness to pay

WTP estimates are simply calculated as the negative of the ratio of the marginal utilities of specified non-monetary attribute to the price coefficient. To test for differences between estimated marginal average WTP values for different attributes across treatments, the combinatorial non parametric test of Poe et al. (2005) was used.

Consistency, internal and external validity

To assess the participants' responses consistency and their stability across the different treatments, as mentioned above, the fifth choice set was repeated at the end of the main task[5]. We have calculated the proportion of the participants who repeated the same choice/rank. To check for the existence of statistically significant differences among hit rates[6] the Z-test is used. To assess the internal validity of the estimated parameters we have used the estimated partworths to predict the respondent's choice in the main task. Then, a hit rate[7] is calculated by comparing the predicted participants' decisions using the maximum utility, to their real decision done in the fifth choice set for each treatment. Finally, in respect to the external validity, the estimated partworths from the main task are used to predict participants' decisions in the holdout task. The predicted and the actual decision in the holdout task are compared to determine the hit rate. A Z-test will be used to assess the difference between hit rates of both internal and external validity across treatments.

4. Results

Table 1 reports the means and standards deviations of random parameter estimates and show that all random parameters reveals the expected sign and were highly statistical significant as well as their standards deviations. Results reveal higher statistical significance of the parameter No-option and his negative effect in every treatment. That is, most of the respondents tried to participate in experiments by choosing or ranking their preferred alternatives instead to choose the no-option alternative. Furthermore, the results reveal that the most preferred type of olive oil is the virgin extra one, and that the olive oil refined is the least preferred from the consumers in every treatment. Catalan consumers' perceive highly the local origin (Catalan Origin) followed by Andalusia. Also, we found that the price is the main obstacle for buying olive oil and that Catalan consumers perceive more manufacturer brand than private label. Moreover, Table 2 provides the posterior consumers' willingness to pay (WTP). The results show that consumers' willingness to pay for extra virgin olive oil in respect to virgin one vary between 0.76 euro and 1.48 euro across the treatments. Also, Table 2 show that the second most important prime that Catalan consumers willing to pay is associated at the local origin of the product which vary between 0.61 euro and 0.86 euro across the treatments.

---

[5] The fifth choice set was repeated at the end instead the first one for the reliability of the test. This approach mitigates the associated bias related to participants repeating their responses not for consistency but due to the memory effect for repeating the first choice.

[6] Hit rate is ratio of the number of the participants who have repeated the same choice or rank about the total number of participants in each treatment.

[7] The hit rate in this case corresponds to the ratio of the total number of hits about the sample size in each treatment. The hit is defined as the success when the model correctly predicts the respondent's actual choice in the main task as well as in the holdout task.

Table 1. RPL and RO-RPL models estimates for elicitation methods.

Traitment	HHCE	NHCE	RRCA	RBWS	NHRCA	NHBWS
Models	**RPL**	**RPL**	**RPL**	**RPL**	**RO-RPL**	**RO-RPL**
			Random parameters estimators			
No-option	-4.822***	-3.340***	-5.050***	-5.015***	-4.641***	-4.437***
(SE)	(0.472)	(0.353)	(0.465)	(0.464)	(0.321)	(0.288)
EVOO	1.555***	0.925***	0.794***	0.988***	0.665***	0.687***
	(0.347)	(0.211)	(0.172)	(0.200)	(0.072)	(0.068)
VOO[1]	-0.202	0.096	0.088	-0.112	0.063	3.750
	(----)	(----)	(----)	(----)	(----)	(----)
OO	-1.352***	-1.021***	-0.883***	-0.876***	-0.728***	-0.656***
	(0.276)	(0.243)	(0.177)	(0.237)	(0.068)	(0.075)
BrManf	0.347***	0.213**	0.158**	0.187**	0.131**	0.145***
	(0.091)	(0.088)	(0.074)	(0.085)	(0.043)	(0.037)
BrPRV[1]	-0.347	-0.213	-0.158	-0.187	-0.131	-0.145
	(----)	(----)	(----)	(----)	(----)	(----)
AND[1]	-0.657	-0.334	0.113	-0.394	0.150	0.027
	(----)	(----)	(----)	(----)	(----)	(----)
CAT	1.308***	0.911***	0.553***	0.940***	0.274***	0.612***
	(0.192)	(0.193)	(0.135)	(0.164)	(0.065)	(0.073)
Rsp	-0.650***	-0.577***	-0.666***	-0.545***	-0.424***	-0.639***
	(0.178)	(0.173)	(0.139)	(0.132)	(0.063)	(0.075)
Price	-1.602***	-1.213***	-0.697***	-1.089***	-0.447***	-0.834***
	(0.165)	(0.112)	(0.114)	(0.113)	(0.051)	(0.058)
			Standards deviations of random parameters			
EVOO	2.495***	1.772***	1.460***	2.195***	1.048***	1.129***
	(0.312)	(0.220)	(0.179)	(0.252)	(0.082)	(0.084)
OO	2.946***	1.715***	1.365***	2.638***	0.930***	1.324***
	(0.436)	(0.212)	(0.193)	(0.320)	(0.079)	(0.097)
BrManf	0.141	0.283**	0.224**	0.410**	0.350***	0.156***
	(0.131)	(0.114)	(0.086)	(0.138)	(0.050)	(0.048)
CAT	2.140***	1.329***	0.689***	1.143***	0.630***	0.597***
	(0.283)	(0.183)	(0.138)	(0.189)	(0.080)	(0.061)
Rsp	0.779***	0.992***	0.502**	0.639***	0.596***	0.489***
	(0.192)	(0.181)	(0.168)	(0.165)	(0.097)	(0.067)
Price	1.620***	0.497***	1.153***	1.128***	0.945***	0.894***
	(0.201)	(0.077)	(0.115)	(0.133)	(0.052)	(0.055)
Number of observations	2970	2970	2970	2970	7155	7110
Log-likelihood	-523.2198	-591.4546	-619.1057	-575.7293	-1745.575	-1070.695

[1] base line; (***) (**) (*) Statistically significant at 1%, 5%, and 10% level.

Table 2. Estimated Willingness to pay for each level of attribute.

Traitment	HHCE	NHCE	RRCA	RBWS	NHRCA	NHBWS
Models	**RPL**	**RPL**	**RPL**	**RPL**	**RO-RPL**	**RO-RPL**
EVOO	0.970	0.762	1.139	0.907	1.488	0.823
[CI]	[0.55; 1.38]	[0.38; 1.14]	[0.51; 1.76]	[0.52; 1.28]	[1.01; 1.96]	[0.63; 1.01]
OO	-0.8439	-0.842	-1.266	-0.804	-1.630	-0.786
	[-1.15; -0.53]	[-1.24; -0.43]	[-1.86; -0.66]	[-1.23; -0.37]	[-2.10; -1.15]	[-0.98; -0.58]
BrManf	0.216	0.176	0.227[a]	0.171	0.294	0.174
	[0.10; 0.33]	[0.02; 0.32]	[-0.001; 0.45]	[0.01; 0.33]	[0.08; 0.50]	[0.08; 0.26]
CAT	0.816	0.751	0.793	0.863	0.613	0.734
	[0.54; 1.08]	[0.43; 1.07]	[0.36; 1.21]	[0.56; 1.16]	[0.3; 0.92]	[0.54; 0.92]
Rsp	-0.406	-0.475	-0.956	-0.501	-0.950	-0.766
	[-0.63; -0.18]	[-0.75; -0.19]	[-1.41; -0.49]	[-0.74; -0.26]	[-1.28; -0.61]	[-0.95; -0.57]

[CI]: Confidence Interval; [a] this value is not statistically significant at 5% level.

Before moving fellow, we tested the preference regularity across the treatments. Table 3 reports the hypothesis tested and the results of the LR test. The results pointed out the significant difference existing between the estimates across all treatments. Therefore, it is appropriate to examine whether this difference between parameters estimates across the samples due the preference heterogeneity or due the error variance derived from the difference in cognitive process of the treatments. Consistent with Caparrós et al. (2008), the error variance varies across the treatment but it is constant within each sample. When calculating the WTPs as the negative ratio of non-price and price attributes estimated parameters automatically the error variance will be simplified. Hence, whether there is significant difference between the estimated mean WTPs across the treatments, therefore we can compare the parameters estimated based on the preference heterogeneity in the sample. In the contrary case, the difference between the parameters estimated due the error variance derived from the cognitive process of the different elicitation methods.

Table 3. Hypothesis of preference regularity test across the treatments.

Hypothesis test of preference regularity	Number of observations	LL	LR	df	p-value
All treatments	5940	-1139.1138			
HHCE	2970	-523.2198			
NHCE	2970	-591.4546			
H_0:test of equality between hypothetical and non-hypothetical CE			49.05	12	P<0.005
All treatments	11880	-2413.8433			
HHCE	2970	-523.2198			
NHCE	2970	-591.4546			
RRCA	2970	- 619.1057			
RBWS	2970	- 575.7293			
H_0:test of equality between hypothetical and non-hypothetical first choice option			208.66	72	P<0.005
All treatments	8910	-1868.23			
NHCE	2970	-591.45			
RRCA	2970	- 619.10			
RBWS	2970	- 575.72			
H_0:test of equality between non-hypothetical first choice option			163.89	36	P<0.005
All treatments	5940	-1228.15			
RRCA	2970	- 619.10			
RBWS	2970	- 575.72			
H_0:test of equality between non-hypothetical RRCA and RBWS			66.64	12	P<0.005
All treatments	14265	-3457.67			
RCA	7155	-1745.57			
BWSDCE	7110	-1678.66			
H_0:test of equality between non-hypothetical RCA and BWSDCE			66.87	12	P<0.005

Table 4 reports results of the combinatorial test suggested by Poe, Giraud, and Loomis (2005) used to compare the mean WTPs. The results reveal that WTPs were not statistically different across treatments. Therefore, the statistical difference between the parameters estimates is indeed underlying of difference in error variance and that the preferences revealed from the samples across the treatments is statistically similar[8]. In line with Caparrós et al. (2008), that first ranking (i.e RRCA and RBWS) implies more difficult cognitive process than the choice task, which could be traduced in higher error variance than the choice.

Table 4. Hypothesis test of equality WTPs across the treatments.

Hypothesis	**Ext. Virgin**	**Olive Oil**	**Br. Manuf**	**Cataluña**	**R. España**
	p-value of Complete combinatorial test				
HHCE vs NHCE	0.471	0.439	0.488	0.486	0.458
HHCE vs RRCA	0.486	0.479	0.425	0.483	0.499
HHCE vs RBWS	0.439	0.438	0.398	0.483	0.467
HHCE vs NHRCA	0.478	0.499	0.349	0.387	0.394
HHCE vs NHBWS	0.462	0.471	0.413	0.466	0.436
NHCE vs RRCA	0.449	0.454	0.420	0.453	0.442
NHCE vs RBWS	0.404	0.372	0.382	0.457	0.419
NHCE vs NHRCA	0.442	0.435	0.326	0.357	0.363
NHCE vs NHBWS	0.425	0.404	0.405	0.437	0.497
RRCA vs RBWS	0.456	0.416	0.448	0.498	0.463
RRCA vs NHRCA	0.456	0.448	0.380	0.388	0.395
RRCA vs NHBWS	0.487	0.456	0.487	0.480	0.446
RBWS vs NHRCA	0.487	0.454	0.424	0.411	0.428
RBWS vs NHBWS	0.479	0.463	0.472	0.474	0.438
NHRCA vs HBWS	0.482	0.495	0.385	0.414	0.368

The notion that accompanied the explosion of preference elicitation methods is the hypothetical bias. That is, the WTP elicited from hypothetical decision tasks almost exceeds WTP elicited from non-hypothetical ones (Chang et al. 2009; Ding et al. 2005; Hensher 2010). Nevertheless, consistent with Carlsson and Martinsson (2001), Lusk and Schroeder (2004), and among others, our findings did not reveal evidence of difference in the marginal WTP between hypothetical and non-hypothetical treatments. There are three explanations: 1) the inclusion of the null option in our experimental design for not forced the participants to make a choice for which could be included some level of payment (Hensher 2010); 2) the effect of a well-scripts presentation explaining the objectives and the conducting of the each experiment, at the beginning of every session. In line with Hensher (2010) and Murphy et al. (2005) this mechanism may be source to eliminate hypothetical bias; 3) the fact to conducting the different sessions throughout the week and at the different hours of day could be raison to enhance the consistency of the subjects' response due the fact to be closer the most possible at their real purchase habits.

Moreover, Table 4 show that there is not statistical difference en marginal WTPs through the non-hypothetical treatments. Consistent with Capparós et al. (2008) and Akaichi et al. (2013), our findings show there is similarity in participants' preferences when they are asked to choose the most preferred option or to take the most preferred into the participants' alternatives ranking. Additionally, the estimated marginal WTPs derived from RCA and BWS elicitation methods are statistical similar when only the ranking first option taking into account (RRCA and RBWS) that follow pertinent also when all information available was considered (NHRCA and NHBWS). Therefore, our results pointed out the importance to use elicitation methods such as RCA and BWS. That is, this methods not only provides similar

[8] Additionally, a combinatorial test of Poe, Giraud, and Loomis (2005) carried out to test the equality distribution of the preferences across treatments. The results confirm the similarity between the consumers' preferences across the treatments and that statistical differences detected between the parameter estimated by LR test is related to error variance derived from the difference in cognitive process of the treatments.

results in respect the CE in terms of consumers' preferences but also make available to the researcher important additional information related to consumers' preferences apart from the ranking first. However, remains to compare whether there is statistical difference in terms of response consistency of the participants, internal and external validity of estimates parameters across treatments.

The results of response consistency, internal and external validity analysis are displayed in Table 5. The results reveal that the hit rate of participants repeat the same choice exceeds the 76% and can be take 87% across first choice treatments (HHCE, NHCE, RRCA, and RBWS), and at 5% significance level there are not statistical differences between these treatments. Consistent with Brouwer et al. (2010), the high participants' choices certainty or stability could be due to learning effect which could be taking place during the course of the experiments. However, the participants' response consistency decreases when moving at NHRCA and NHBWS, supporting the hypothesis that the stability of ranking information decreases with decreasing rank due to the difficult cognitive process of the ranking mechanism (Ben-Akiva et al. 1992). Additionally, The results show that all treatments have a statistically similar internal predictive power and that the highest hit rate were associated to NHCE and RRCA treatments.

Regarding the external validity, the estimates parameters in the hypothetical treatment lowly correctly predicted participants' responses in the holdout task respect the non-hypothetical treatments. The external validity hit rate of the HHCE is near 24%, however for NHCE, RRCA, RBWS, NHRCA, and NHBWS this hit rate are 44%, 38%, 53%, 40%, and 62%, respectively. The p-value shows that there is statistically difference between them. It illustrates that incentivizing the participants to behave truthfully enhance the predictive power of the treatments used (Chang et al. 2009; Lusk et al. 2008). On the other hand, the p-value for the external validity shows that there is statistical similarity between NHCE, RRCA and RBWS, although the estimates parameters in RBWS better correctly predicted participants' responses (53%) in the holdout task than NHCE (44%) and RRCA (38%). That is, participants behave similarly whether there are asked to choose or to state their most preferred through two ranking elicitation mechanism. Additionally, when we take into account all the preferences ranking information the difference in predictive power of the estimates parameters of NHRCA and NHBWS became statistically significant. This is the most striking finding in our study. It is important to note, the advantage of BWS in respect to RCA by providing natural process to identify extremes which seem to be easy to understand by the people, leads to enhancement in predictive power in respect to RCA. The fact to take into account all ranking preferences information, the external validity hit rate pass from 38% to 40% for RRCA and NHRCA. However, this rise has been accentuated in the case of BWS (i.e. the external validity hit rate pass from 53% to 62% for RBWS and NHBWS).

Table 5. Consistency, internal and external validity tests across treatments.

	Consistency				Internal validity			External validity		
Treatments	**T.Nº of choice**	**Nº correct of prediction**	**Hit rate (%)**	***p-value***	**Nº correct of prediction**	**Hit rate (%)**	***p-value***	**Nº correct of prediction**	**Hit rate (%)**	***p-value***
HHCE vs	55	48	87.27	0.103	35	63.63	0.153	13	23.63	0.013
NHCE	55	43	78.18		40	72.72		24	43.63	
HHCE vs	55	48	87.27	0.069	35	63.63	0.153	13	23.63	0.049
RRCA	55	42	76.36		40	72.72		21	38.18	
HHCE vs	55	48	87.27	0.069	35	63.63	0.421	13	23.63	0.000
RBWS	55	42	76.36		34	61.81		29	52.72	
HHCE vs	55	48	87.27	0.000	35	63.63	0.272	13	23.63	0.032
NHRCA	55	27	49.09		38	69.09		22	40	
HHCE vs	55	48	87.27	0.000	35	63.63	0.344	13	23.63	0.000
NHBWS	55	25	45.45		37	67.27		34	61.81	
NHCE vs	55	43	78.18	0.410	40	72.72	0.5	24	43.63	0.280
RRCA	55	42	76.36		40	72.72		21	38.18	
NHCE vs	55	43	78.18	0.410	40	72.72	0.111	24	43.63	0.170
RBWS	55	42	76.36		34	61.81		29	52.72	
NHCE vs	55	43	78.18	0.001	40	72.72	0.337	24	43.63	0.349
NHRCA	55	27	49.09		38	69.09		22	40	
NHCE vs	55	43	78.18	0.000	40	72.72	0.266	24	43.63	0.028
NHBWS	55	25	45.45		37	67.27		34	61.81	
RRCA vs	55	42	76.36	0.5	40	72.72	0.111	21	38.18	0.062
RBWS	55	42	76.36		34	61.81		29	52.72	
RRCA vs	55	42	76.36	0.001	40	72.72	0.337	21	38.18	0.422
NHRCA	55	27	49.09		38	69.09		22	40	
RRCA vs	55	42	76.36	0.000	40	72.72	0.266	21	38.18	0.006
NHBWS	55	25	45.45		37	67.27		34	61.81	
RBWS vs	55	42	76.36	0.001	34	61.81	0.211	29	52.72	0.090
NHRCA	55	27	49.09		38	69.09		22	40	
RBWS vs	55	42	76.36	0.000	34	61.81	0.274	29	52.72	0.167
NHBWS	55	25	45.45		37	67.27		34	61.81	
NHRCA vs	55	27	49.09	0.351	38	69.09	0.418	22	40	0.011
NHBWS	55	25	45.45		37	67.27		34	61.81	

5. Conclusions

Parting from an important issue in experimental economics what stated preference method can be predicted best the real consumer behavior, using the same experimental design, we compared the consistency, the internal and external predictive power of individuals' responses derived from four elicitation methods (hypothetical CE, non-hypothetical CE, non-hypothetical RCA, and non-hypothetical BWS). This paper contributes the literature existed, firstly, by making BWS incentive compatible following the Lusk et al. (2008)'s conjoint ranking incentive compatible process. Secondly, recoded the sequential BWS treatment to provide a preference ranking and to empirically compare it with the Lusk et al. (2008)'s conjoint ranking incentive compatible. Thirdly, parting from the hypothesis that ranking exercise recoded and analyzed as a choice using only the ranking first option is similar than a choice task, we tested whether the difference in cognitive process between the RCA and BWS lead us towards a different results.

Overall, our results suggest that applied a ranking experiments (RCA and BWS) may be a safe practice, independently whether the researcher focus only on the first rank and analyze it as CE or whether all preference ranking information took into account. In incentive compatible context, and compared with choice experiment, these methods not only provide similar results regarding to predictive power in-sample and out-of-sample, and individual response consistency when just only the first rank data is analyzed, but also make available to the researcher important additional information on consumer's preferences for the no chosen profiles. Additionally, it is worth noting the importance to recode the sequential BWS data to get a preferences ranking information. We found that the easily understanding cognitive process from the participants by detecting the extreme values significantly increase the predictive power of the NHBWS in respect to NHRCA, therefore open more future insights for the researcher.

Our findings should come as welcome relief to agribusiness researchers, but remains important to generalize and to confirm our results through to predicting actual grocery store sales of different categories of the products. Furthermore, our experimental design include nine choice each one include five alternatives plus the no-option choice. Therefore, the results holds consistent if increasing the number of the alternatives in choice sets and even the number of the repeated choice sets?

References

Akaichi, F., R. M. Nayga, Jr. and J. Gil. (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.

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

Brouwer, R., Dekker, T., Rolfe, J., and Windle, J. (2010). Choice certainty and consistency in repeated 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.

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.

Carlsson, F., and Martisson, P. (2001). Do hypothetical and actual marginal willingness to pay differ in choice experiments? *Journal of Environment Economics and Management* 41: 179-192.

Chang, J.B., Lusk, J.L., and Norwood, B. (2009). How closely do hypothetical surveys and laboratory experiments predict field behavior? *American Journal of Agriculture Economics* 96: 518-534.

Ding, M., Grewal, R., and Liechty, J. (2005). Incentive-Aligned Conjoint Analysis. *Journal of Marketing Research* 42: 67–83.

Flynn, T., and Marley, A.J. (2012). Best worst scaling: Theory and methods. Sydney, CenSoC Working Paper Nº 12-002.

Hensher, D.A. (2010). Hypothetical bias, choice experiment and willingness to pay. *Transport Reseach Part B* 44: 735-752.

Hensher, D.A., Rose, J.F., Greene, W.H. (2005). *Applied Choice Analysis: A Primer*. Cambridge University Press, Cambridge.

Lancsar, E., Louviere, 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.

Louviere, J., Street, A., Burgess, L., Wasi, N., Islam, T., and Marley, A. (2004). Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. Sydney, CenSoC Working Paper Nº 04-005.

Louviere, J.J. and Street, D. (2000). Stated-preference Methods. In Hensher, D.A., and Kenneth, J.B. (eds), *Handbook of Transport Modelling*. Pergamon Press, Amsterdam, Netherlands, 131-143.

Louviere, J.J., Street, D., Burgess, L., Wasi, N., Towhidul, I., Anthony, A.J.M. (2008). Modelling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. *Journal of Choice Modelling* 1: 128-163.

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: 467–82.

Lusk, J.L., Fields, D., and Prevatt. W. (2008). An Incentive Compatible Conjoint Ranking Mechanism. *American Journal of Agricultural Economics* 90: 487–98.

Murphy, J.J., Geoffrey, J.A., 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.

Poe, G.L., Giraud, K.L., and Loomis, J.B. (2005). Computational Methods for Measuring the Difference of Empirical Distributions. *American Journal of Agricultural Economics* 87: 353-365.

Scarpa, R., Nataro, S., Louviere., 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.

Street, D., and Burgess, L.B. (ed.) (2007). *The Construction of Optimal Stated Choice Experiments: Theory and Methods*. John Wiley and Sons, Hoboken: New Jersey.