The implications of incorrect utility function specification for welfare measurement in choice experiments

Cati Torres*, Nick Hanley⊥, Antoni Riera*

Abstract

Despite the vital role of utility functional form in welfare measurement, the implications of working with incorrect utility specifications have not been examined in the choice experiments literature. This paper addresses the importance of the specification of both non-monetary attributes and the marginal utility of income. Monte Carlo experiments have been conducted wherein different attribute specifications and assumptions for the *Cost* parameter –that is, different functional forms of utility– have been assumed to generate simulated choices on which Multi-Nomial Logit and Mixed Logit models have been estimated under correct and incorrect assumptions about the *true*, underlying utility function. The inferred values have been compared with the *true* ones directly calculated from the *true* utility specifications. Results show that working with simple experimental designs and continuous-linear specifications makes attribute specification irrelevant for measuring attribute marginal values regardless of the *true* effects the attribute has on utility.

Keywords: utility specification, attributes, welfare measurement, accuracy, efficiency, choice experiments, Monte Carlo analysis.

***JEL classification**:* C51, D69, C99, C15.

*Corresponding author. Centre de Recerca Econòmica (UIB · SA NOSTRA), 07122-Palma de Mallorca, Spain. Email: cati.torres@uib.es. Telephone: +34 971 171381. Fax: +34 971 172389.

⊥ Department of Economics, University of Stirling, Stirling FK9 4LA, United Kingdom. Email: n.d.hanley@stir.ac.uk. Telephone: +44 1786 466410.

*Centre de Recerca Econòmica (UIB · SA NOSTRA), 07122-Palma de Mallorca, Spain. Email: antoni.riera@uib.es. Telephone: +34 971 171381. Fax: +34 971 172389.

I. Introduction

Since the 1990's, choice experiments (CE) have been increasingly used in environmental valuation. One of their most important advantages is their ability to estimate willingness-to-pay (WTP) for changes in an attribute (i.e. implicit prices). In this context, decisions the researcher makes concerning the effects non-monetary attributes have on utility (i.e. linear or non-linear effects) are of interest. These decisions have to do with the specification of attributes, that is, with their nature (i.e. continuous or discrete attributes) and their number of levels. In this sense, if a discrete attribute is assigned three or more levels it is supposed to have non-linear effects on utility, whereas if it has two levels only a linear relationship can be represented (Hensher et al., 2005). Likewise, linear effects on utility can also be depicted with a continuous attribute entering linearly the utility function, and a non-linear relationship can be shown if, for instance, the continuous attribute has a quadratic specification and is assigned at least three levels. Given that attribute specification determines econometric and experimental design issues -and consequently, has an impact on the efficiency of marginal and total WTP estimates- decisions concerning the effects attributes have on utility are non-trivial.

However, despite the role of utility specification in welfare measurement, CE studies concerned with the precision of benefit estimates have been centered on the implications for welfare calculation of different experimental design strategies. In this context, the interest in the functional form of utility has been restricted to the analysis of the impacts of alternative experimental designs under different utility specifications. Despite preference specification issues lying at the core of discrete choice models, little attention has also been paid to utility specification issues in research around other valuation approaches based on random utility models, which has been mainly focused on the implications for benefit estimates of the specification of the recreation demand function, the estimation model and the WTP elicitation approach.

Given that decisions about the nature and the number of levels of attributes must be taken in a context of uncertain knowledge about the *true* preferences of individuals, this raises the question of the implications of working with incorrect utility specifications: that is, it raises the issue of how important the specification of attributes is for welfare measurement. This question appears to be largely unexplored in the CE literature, and can be applied to both monetary and non-monetary attributes. As is well known, the parameter of the *Cost* variable, usually interpreted as the marginal utility of income, plays a key role in welfare measurement. Problems related to the assumption of a random *Cost* parameter (i.e. the inappropriateness of the normal distribution or the probability of working on extreme values for some individuals if a lognormal distribution is assumed) have led many researchers to consider it constant when specifying the utility function. However, this is unlikely to be true. In other words, if it is expected that rich and poor people assign a different value to one monetary unit, it should be expected that the marginal utility of income is different among individuals. Therefore, assuming homogeneous a parameter that is likely to be heterogeneous, as traditionally done, could have important implications for welfare estimates. In this context, the relevance of attribute specification for the estimation of attribute values cannot be examined without simultaneously undertaking an analysis of the effects derived from mistaking assumptions about the *Cost* parameter.

This article addresses the importance of attribute specification for CE valuation estimates with a special emphasis on the effects of mistaking assumptions about the marginal utility of income. To do that, Monte Carlo (MC) experiments have been conducted wherein different attribute specifications and assumptions for the *Cost* parameter –that is, different functional forms of utility– have been assumed to generate simulated choices on which Multi-Nomial Logit (MNL) and Mixed Logit (MXL) models have been estimated under correct and incorrect assumptions about the *true*, underlying utility function. The inferred values have been compared with the *true* ones directly calculated from the *true* utility specifications. This procedure has been repeated 1,000 times to examine the robustness of results.

The structure of the article is as follows. The next section provides a review of the environmental valuation literature focussing on the analysis of factors affecting welfare estimates, in an attempt to show the lack of studies dealing with attribute specification issues. Section III discusses the methodology used, based on MC analysis, and the data employed for the experiments. Results are reported in section IV. Conclusions are drawn in section V.

II. Accuracy and precision of welfare estimates in the literature

Over the recent past, the importance of examining bias and precision of welfare estimates has been stressed in the literature on CEs. In this regard, most studies have been focused on analyzing the effects derived from the use of different experimental design strategies. Thus, Ferrini & Scarpa (2007) use MC analysis to compare simple, shifted (orthogonal) designs with D-efficient designs and cast light on both the use of prior information in undertaking experimental design, and the issue of whether the nature of the actual data generating process is consistent with that assumed by the analyst in choosing their econometric approach. Scarpa & Rose (2008) analyze the performance of different design strategies, undertaken under the assumption that a prior belief on the range of values for the utility parameters can plausibly be defined, with a focus on efficiency of WTP estimates from a MNL model. Although they prioritize the use of some designs over another ones, they finally conclude that the analyst need not worry about the experimental design if the budget for a study allows working on high sample sizes. Carlsson & Martinsson (2003) use MC analysis to compare three kinds of experimental design (orthogonal, cyclical and D-optimal) in terms of bias and mean squared error for three different *true* utility functions. In a similar vein is a paper by Lusk & Norwood (2005) who also use MC experiments to compare the effects of specifying utility as a continuous function of attributes, with a step-wise specification, in terms of the implications of alternative experimental designs. Their main finding is that *true* and estimated WTP are insignificantly different for all experimental designs considered, and that higher sample sizes always improve the fit of actual and estimated WTP.

Indeed, recognition of the need for analyzing the accuracy of welfare estimates (Kling, 1991; Kling & Sexton, 1990) has also led researchers working on valuation methods other than CEs to investigate issues such as the specification of the recreation demand function, and the WTP elicitation approach. Early studies concerned about the factors affecting welfare measurement emerge in the field of revealed preference (RP) methods and deal with the effects of different approaches to travel cost (TC) modelling. Thus, Kling (1987) looks at the impacts on WTP estimates for quality changes in the Chesapeake Bay from the use of four different recreation demand models: single equation, pooled demand, varying parameter and logit. Parameters from recreation surveys are combined with a utility function to simulate a TC data set to which the four alternative approaches are applied. Kling (1987) finds that all approaches underestimate the *true* mean welfare change. A related paper is Kling (1988), who again uses MC analysis to compare three different *true* utility specifications with alternative functional forms for the demand function in TC models. Comparisons of results are done in terms of the errors in estimating *true* welfare changes. Interestingly –a finding that echoes our own- the paper shows that rather simple specifications for TC models can actually yield relatively small errors in welfare estimation. Issues of functional form choice in TC models are also central to papers by Adamowicz et al (1989) and Kling (1989). The former article looks at effects on the variance of welfare estimates, comparing linear, semilog, log-log and restricted Box-Cox forms, and finds that impacts on both variance and mean can be substantial. In the latter paper, Kling (1989) focuses on the magnitude of errors in WTP estimates from incorrect choice of functional form and finds that the choice of functional form is less important for small relative to big price changes, but that goodness-of-fit tests are a relevant aspect of recreational demand modelling. A related area of concern in the RP framework is decisions over appropriate nesting structures in multiple site recreation demand models. In this sense, Kling & Thomson (1996) show that parameter estimates depend on both nesting structure and estimation method (sequential or Full Information Maximum Likelihood), whilst Herriges & Kling (1997) report the sign and size of bias from inappropriate nesting structures and analyze the ability of conventional goodness of fit tests to identify the best model.

Concerns about the accuracy and precision of welfare estimates can also be found in the field of SP choice approaches other than CEs. Thus, Kling (1997) uses MC analysis to investigate the advantages of combining TC and contingent valuation (CV) data in terms of the bias and precision of welfare measures, and Alberini (1995) analyzes, by undertaking MC experiments, the gains from using a double-bounded discrete choice model in the CV context, relative to a bivariate probit model and finds the double-bounded approach to produce gains in terms of lower bias and greater precision. Scarpa & Bateman (2000) also use MC methods to analyse the design of follow-up questions in multiple-bounded question formats, and to investigate the efficiency gains from asking such follow-up questions, whereas Park et al (1991) investigate the effects of functional form on WTP estimates within a discrete choice set-up.

In the light of this background, the question of how important the specification of attributes in the utility function is for welfare measurement has not been fully answered. Given the role of utility specification in welfare calculation, efforts need to be made to fulfil this gap. In recent years, some authors have argued that addressing the effects of misspecifying the underlying utility function – for example, using a linear form when *true* utility is non-linear – is an important area for future research (Lusk & Norwood, 2005). In this context, this paper analyzes the implications of attribute specification for CE welfare estimates with a special emphasis on the assumptions about the parameter of the *Cost* attribute. Although results of this article are restricted to the assumptions made in the

MC experiments, they may provide some insights into the relevance of attribute specification for calculation of attribute values in CEs more widely.

III. Designing MC experiments to examine the importance of attribute specification

To test for the relevance of utility function specification, MC analysis has been applied. For this reason, three different *true* specifications for a non-monetary attribute (X_1) – and, hence, three different *true* functional forms of utility– and an error structure have been considered, under two assumptions about the marginal utility of income, to simulate choices on which MNL and MXL models have been estimated. The estimated marginal value of X_1 has been compared with the *true* marginal value. Therefore, four factors have been taken into account in designing the MC experiments: experimental design, *true* attribute specification, attribute specification in the estimation models, and assumed marginal utility of income. A detailed description of these factors is presented in the next sections.

3.1 The experimental designs

The attribute data employed to create experimental designs come from a CE study on recreational beach use in Santa Ponça Bay, a small Mallorcan tourism area. We consider three non-monetary attributes (X_1, X_2 and X_3) and one monetary attribute (X_4), each at three levels.[1] To examine the importance of the specification of attributes in terms of their number of levels and their continuous or discrete nature, two and five levels have also been assigned to X_1. This has led to create three different types of experimental designs with a universe of $(2x3^3)x(2x3^3)$ possible pairs combinations for the first, 2-level design, 3^4x3^4 for the second, 3-level design and $(5x3^3)x(5x3^3)$ for the third, 5-level design.[2] The designs have been generated under a D-efficiency criterion and allowing for main effects (ME) only. According to Louviere et al. (2000), this kind of

[1] For a detailed description of the attributes and their levels, see Torres et al. (In press).

[2] Because the attribute levels of the business-as-usual (BAU) option are constant across the choice sets, only pair combinations have been optimized when creating the experimental design, the BAU alternative being added to the generated choice sets after the optimization process.

design typically explains about 70-90% of the variance in choice. The final designs have consisted of 72, 36 and 180 pairs of attribute combinations for the 2-level, the 3-level and the 5-level designs, respectively. These have been then blocked into different versions, each of 6 choice sets of 2 alternatives plus the business-as-usual (BAU) option. The main features of the designs are shown in Table I.

Table I. Main features of the ME only designs

Experimental design factors		2-level design	3-level design	5-level design
Attribute levels	**X_1**	2 6*	2 4 6*	2 3 4 5 6*
	X_2	3 6 8*	3 6 8*	3 6 8*
	X_3	0.3 1* 2	0.3 1* 2	0.3 1* 2
	X_4	3 10.5 24 (0*)	3 10.5 24 (0*)	3 10.5 24 (0*)
Alternatives		2+BAU	2+BAU	2+BAU
Choice sets per individual		6	6	6
Blocks		12	6	30
Block replications		20	40	8
Total observations[a]		1,440	1,440	1,440

* Starred numbers correspond to the levels for the BAU option.
[a] Total observations are the number of choice sets x the number of blocks x the number of block replications.

3.2 The *true* attribute specification and the *true* attribute marginal value

At the first stage of the MC analysis, three different generic utility functions with the same explanatory variables (X_1, X_2, X_3 and X_4) and known parameters have been specified to compute the *true* marginal value for X_1. Linear and non-linear effects on utility have been considered for X_1. Thus, for a scenario in which X_1 has *true* linear effects a linear specification has been employed (Equation 1), and to consider non-linear effects two different specifications have been used: a quadratic one (Equation 2) and a stepwise function (Equation 3) where the marginal utility of X_1 takes three constant values between 0 and c_2.[3]

$$U_{ji} = \alpha_{11}X_1 + \beta_2 X_2 + \gamma_3 X_3 + \omega_4 X_4 + \varepsilon_{ji} \tag{1}$$

$$U_{ji} = \alpha_{11}X_1 + \alpha_{12}X_1^2 + \beta_2 X_2 + \gamma_3 X_3 + \omega_4 X_4 + \varepsilon_{ji} \tag{2}$$

[3] Note that all utility specifications are linear-in-parameters.

$$\text{If } X_1 < c_1 \qquad U_{ji} = \alpha_{11} + \beta_2 X_2 + \gamma_3 X_3 + \omega_4 X_4 + \varepsilon_{ji} \tag{3}$$

$$\text{If } c_1 \leq X_1 < c_2 \qquad U_{ji} = \alpha_{12} + \beta_2 X_2 + \gamma_3 X_3 + \omega_4 X_4 + \varepsilon_{ji}$$

$$\text{If } X_1 \geq c_2 \qquad U_{ji} = \alpha_{13} + \beta_2 X_2 + \gamma_3 X_3 + \omega_4 X_4 + \varepsilon_{ji}$$

where U_{ji} is the indirect utility of alternative *j* for individual *i*, α_{11}, α_{12}, α_{13}, β_2, γ_3, ω_4 are the known parameters of the attributes -ω_4 being the marginal utility of income-, c_1 and c_2 are the critical attribute values delimiting the three steps of the stepwise marginal utility of X_1 and ε_{ji} is the error term associated with alternative *j* and individual *i*.[4]

All the parameters have been considered constant for each generic utility specification, although an additional assumption about the value of ω_4 has been made to consider not only a constant marginal utility of income but also a non-constant one. In this sense, when ω_4 has been assumed constant, the marginal value of X_1 has been equal for all the simulated individuals making choices (homogenous preferences), whereas when it has been considered non-constant each individual have assigned a different marginal value to X_1.[5] To represent this heterogeneity of the marginal utility of income, the values of ω_4 have been randomly drawn from a lognormal distribution. Two-hundred and forty simulated individuals have been considered. Following Hanemann (1984), the *true* marginal value of X_1, defined as the WTP for a change in the attribute from the BAU scenario, has been calculated for the linear, quadratic and stepwise utility specification, as shown in Equations (4), (5) and (6), respectively.

[4] For simplicity reasons, subscript *j* for explanatory variables has been omitted.
[5] In this case, the *true* marginal value of X_1 has been obtained by averaging the sum of the *true* marginal values for each individual over all the individuals of the sample.

$$CV = -\frac{1}{\omega_4}\left[\alpha_{11}\left(X_{1_1} - X_{1_0}\right)\right] \quad (4)$$

$$CV = -\frac{1}{\omega_4}\left[\alpha_{11}\left(X_{1_1} - X_{1_0}\right) + \alpha_{12}\left(X_{1_1}^2 - X_{1_0}^2\right)\right] \quad (5)$$

$$CV = -\frac{1}{\omega_4}\left(\alpha_{1y} - \alpha_{1z}\right); \quad y, z=1, 2, 3 \quad (6)$$

where CV is the compensating variation, X_{1_1} and X_{1_0} represent the attribute levels of X_1 for the policy-on and the BAU situation, respectively, and y and z represent one of the three ranges of the three-stepwise function and depend on the values of c_1 and c_2.

Table II contains the *true* utility specification, the known parameters, the critical values c_1 and c_2 for the stepwise function and the *true* marginal value for a hypothetical change in X_1 from the BAU level (6, see Table I) to a situation in which it takes the level 2.

Table II. *True* attribute specifications and *true* marginal values

Parameter values[a]	Constant *Cost* parameter			Random *Cost* parameter[b]		
	***True* utility specification**			***True* utility specification**		
	Linear	**Quadratic**	**3-Stepwise**	**Linear**	**Quadratic**	**3-Stepwise**
α_{11}	-1.8	-2	-3.6	-1.8	-2	-3.6
α_{12}		0.1	-6		0.1	-6
α_{13}			-8			-8
β_2	-0.7	-0.7	-0.7	-0.7	-0.7	-0.7
γ_3	0.4	0.4	0.4	0.4	0.4	0.4
ω_4	-0.8	-0.8	-0.8	LogN(0.8,0.2)	LogN(0.8,0.2)	LogN(0.8,0.2)
c_1			3			3
c_2			5			5
True marginal value of X_1	9	6	5.5	11.8	7.9	7.2

[a] The value of the known parameters of X_1 have been chosen in such a way that the marginal utility of X_1 for level 2 is equal for all the *true* utility specifications.
[b] The parameter of the *Cost* attribute is lognormally-distributed with 0.8 mean and 0.2 (0.8x0.25) standard deviation.

3.3 The MC experiments and the estimated attribute marginal values

At the second stage of the analysis, MC experiments have been undertaken to estimate the marginal value of X_1 and compare it with the *true* marginal value. Therefore, choices have been simulated for each type of *true* utility specification (linear, quadratic and stepwise) and experimental design (2-level design, 3-level design and 5-level design) under two different data generating processes (DGP) derived from the assumptions about the marginal utility of income (a constant value for ω_4 -or MNL-DGP- and a lognormally-distributed value for ω_4 -or MXL-DGP). This has generated 18 different sets of simulated choices (3x3x2). To obtain these sets, the utility of each alternative in each choice occasion has been calculated by combining the known parameters of the utility function with the attribute levels and adding an error term. The error terms have been generated from a type I extreme value distribution and a unique error has been randomly drawn not only for each alternative but also for each observation in the sample. The simulated choice has been assigned to that alternative in the choice set providing the highest utility level. Because 240 individuals have been considered in the simulation and each of them has faced 6 choice sets, 1,440 observations (240x6) have been generated by this process for each of the 18 sets of simulated choices.

Using these simulated samples, MNL and MXL models have been estimated to infer the marginal value of X_1. To examine the importance of the specification of X_1 for measuring its marginal value at this stage, X_1 has been codified as continuous both entering linearly the utility function to match the specification in Equation 1 (i.e. continuous-linear assumed specification) and having a quadratic specification to match the specification in Equation 2 (i.e. continuous-quadratic assumed specification).[6] On the other hand, a dummy-coding structure without interactions between dummy variables has been used to match the attribute specification in Equation 3 (i.e. discrete-linear utility specification).

In this context, to test for potential effects derived from mistaking assumptions about the marginal utility of income, two different possible erroneous assumptions that an analyst might make about the *Cost* parameter (ω_4) have been considered under the hypothesis that the marginal utility of income is different for all individuals: (i) the assumption of homogeneity in ω_4 and (ii) the assumption of a distribution for ω_4 other than the *true* one. Then, the 18 sets of simulated choices have been divided into two groups of 9 sets each according to the type of DGP followed to simulate them (i.e. MNL-DGP and MXL-DGP). To test for the effects from mistaking assumptions about the marginal utility of income, a MNL and a MXL model, both under the three different specifications of X_1, have been applied to each of the 9 sets of choices derived from the MXL-DGP, a triangular distribution being assigned to ω_4 when estimating the MXL model.[7] Put another way, erroneously applying a MNL model when the *Cost* attribute is heterogeneous and estimating a MXL model by incorrectly assigning a triangular distribution to ω_4 when the *true* one is lognormal have served to test for the effects from mistaking assumptions about the marginal utility of income. In this way, 27 different MC experiments have been undertaken for each scenario of mistaken assumptions about ω_4 (9 sets of choices x 3 types of estimation model according to the attribute specification). The results obtained by incorrectly applying the MNL and MXL models have been compared with those derived from a correct application of the models, that is, derived from two scenarios of correct assumptions about ω_4, these latter

[6] The use of ME only designs when the *true* utility has higher order effects (i.e. it is quadratic in the attribute) has not been a problem since sufficient degrees of freedom derived from repeating a given design have been used (Lusk & Norwood, 2005).

[7] Like the lognormal distribution, the triangular distribution can be constrained to have the same sign for the parameter of interest. This is why it can also be assigned to a random *Cost* parameter when the lognormal distribution is not assumed. Given that the Matlab code by Kenneth Train to estimate the MXL model has been used in this paper, the triangular distribution for the *Cost* parameter has been defined as follows: ω_4 ~Triangular (μ+σt) where t is triangular between -1 and 1, and μ and σ are estimated.

consisting of applying a MNL to the 9 sets of choices generated from the MNL-DGP and of estimating a MXL model assuming a lognormally-distributed ω_4 on the 9 sets generated from the MXL-DGP. Again, this has led to 27 different MC experiments for each scenario of correct assumptions about ω_4 (i.e. 9 sets of choices x 3 types of estimation model according to the attribute specification). Table III lists the four assumptions about the marginal utility of income considered to undertake the MC experiments in terms of the DGP followed to simulate choices and the estimation model finally applied on the simulated sample.

Table III. Assumptions about the marginal utility of income used

Description	DGP	Estimation model	"Mistaken assumption"	MC experiments
MNL-MNL	MNL-DGP	MNL	None	27
MXL-MNL	MXL-DGP	MNL	ω_4 is constant	27
MXL-MXL/LogN	MXL-DGP	MXL	None	27
MXL-MXL/TriangN	MXL-DGP	MXL	ω_4 is triangular-distributed	27

Therefore, each type of *true* attribute specification (linear, quadratic, stepwise), experimental design (2-level design, 3-level design and 5-level design), attribute specification in the estimation model (continuous-linear, continuous-quadratic and discrete-linear) and assumption about the marginal utility of income (MNL-MNL, MXL-MNL, MXL-MXL/LogN, MXL-MXL/Triang) has led to undertake 108 different MC experiments (3x3x3x4 or 27x4). The marginal value of X_1 has been estimated for each MC experiment following Equations (4), (5) and (6) according to the attribute specification assumed in the model. This process has been repeated 1,000 times, this leading to a distribution of 1,000 estimated marginal values for X_1 for each MC experiment. From each distribution, the estimated marginal value of X_1 has been calculated as the average of the sum of the values obtained in each MC experiment over 1,000 repetitions.[8]

The importance of attribute specification for measuring attribute marginal values has been examined by quantifying the errors in the estimated marginal values for X_1 through the calculation of two commonly used accuracy measures: bias and mean squared error (MSE) in the estimated marginal WTP. As shown in Equations (7) and (8), bias has been defined as the average over 1,000 repetitions of the difference between the estimated and the *true*, simulated marginal WTP for X_1, whereas MSE represents the average over 1,000 repetitions of the square of the bias.

$$BIAS = \frac{1}{R}\left[\sum_{r=1}^{R}\left(CV_r^e - CV^t\right)\right] \tag{7}$$

$$MSE = \frac{1}{R}\left[\sum_{r=1}^{R}\left(CV_r^e - CV^t\right)^2\right] \tag{8}$$

where R is the number of repetitions of each MC experiment, CV_r^e is the estimated compensating variation in repetition r and CV^t is the *true* compensating variation.

The variance of the estimated compensating variation can be obtained by applying Equation (9), this allowing computing the significance of bias.

[8] Like for the calculation of the simulated marginal value of X_1, the estimated marginal value for each MC experiment when a MXL model is applied has been obtained by averaging the sum of the individual marginal values over all the individuals of the sample.

$$BIAS^2 = MSE - Var\left(CV^e\right) \tag{9}$$

IV. Results

The results of bias in the estimated marginal value of X_1 for each MC experiment are presented in Table IV in terms of the *true* attribute specification (linear, quadratic, stepwise), the experimental design (2-level design, 3-level design and 5-level design), the attribute specification assumed in the estimation model (continuous-linear, continuous-quadratic and discrete-linear) and the assumption about the marginal utility of income (MNL-MNL, MXL-MNL, MXL-MXL/LogN, MXL-MXL/Triang).

Table IV. Bias in the estimated marginal value of X_1 (over 1,000 repetitions)[a]

***True* attribute specification**	**Assumed attribute specification**	**Type of design**	**MNL-MNL**	**MXL-MNL**	**MXL-MXL/LogN**	**MXL-MXL/Triang**
Linear	**Continuous-linear**	**2-level**	-0,0107	0,1027*	-0,1765*	0,3466*
		3-level	-0,0047	-0,6137*	0,0279*	0,6383*
		5-level	-0,0038	-0,5302*	0,0453*	0,5663*
	Continuous-Quadratic	**2-level**	-0,0107	0,1027*	-0,1766*	0,3466*
		3-level	-0,0033	-0,7086*	-0,0452*	0,5881*
		5-level	-0,0024	-0,4997*	0,0725*	0,6337*
	Discrete-linear	**2-level**	-0,0107	0,1027*	-0,1765*	0,3466*
		3-level	-0,0033	-0,7086*	-0,0453*	0,5881*
		5-level	0,0006	-0,6123*	0,0499*	0,6009*
Quadratic	**Continuous-linear**	**2-level**	0,0007	0,1415*	-0,0623*	0,1725*
		3-level	-0,1528*	-0,5205*	-0,2482*	0,0363*
		5-level	-0,3650*	-0,7506*	-0,3695*	-0,1302*
	Continuous-Quadratic	**2-level**	0,0007	0,1414*	-0,0624*	0,1725*
		3-level	-0,0043	-0,4033*	-0,0742*	0,2762*
		5-level	0,0039	-0,2005*	0,0714*	0,3740*
	Discrete-linear	**2-level**	0,0007	0,1414*	-0,0624*	0,1725*
		3-level	-0,0043	-0,4033*	-0,0742*	0,2761*
		5-level	0,0150	-0,3589*	0,0467*	0,3652*
3-Stepwise	**Continuous-linear**	**2-level**	0,0081	0,1175*	-0,0605*	0,1300*
		3-level	-0,0783*	-0,4096*	-0,1429*	0,1182*
		5-level	-1,0851*	-1,648*	-1,2268*	-1,0531*
	Continuous-Quadratic	**2-level**	0,0081	0,1175*	-0,0605*	0,1300*
		3-level	-0,0026	-0,3494*	-0,0735*	0,2428*
		5-level	-0,309*	-0,6230*	-0,4055*	-0,1865*
	Discrete-linear	**2-level**	0,0081	0,1175*	-0,0605*	0,1300*
		3-level	-0,0026	-0,3494*	-0,0735*	0,2427*
		5-level	0,0188*	-0,4439*	0,0202	0,2660*

* The starred values mean that bias is significant at the 95% of the confidence level. The t-statistic has been computed as the ratio of the bias to its standard error, where this latter has been calculated as the standard deviation of the estimates (i.e. squared root of the variance from Equation (9)) divided by the squared root of *R* (1,000).

Although the consideration of all the experimental factors has led to obtain 108 estimates of the marginal value of X_1 and, hence, 108 values of bias, what is most relevant in this analysis is the effects of attribute misspecification on welfare estimates. In this sense, most attention will be spent discussing the results from the MC experiments where the attribute specification assumed in the model does not match the *true* one. In this sense, note that under all the scenarios of *true* attribute specification and assumptions about the marginal utility of income, the discrete-linear specification approaches well the continuous-quadratic one (i.e. both leading to the same value of bias) when X_1 takes 2 and 3 levels, this implying that, if X_1 has *true* non-linear effects, only the continuous-linear specification will be of interest.

As shown in Table IV, when the *true* marginal utility of income is constant among the simulated individuals and a MNL model is estimated on their choices (MNL-MNL), biases are not significantly different from zero at the 95% level of confidence under a *true* linear specification for X_1, which means attribute misspecification in the estimation model leads to no significant errors in the attribute marginal value. In contrast, assuming a continuous-linear specification when X_1 has *true* non-linear effects (i.e. quadratic and stepwise marginal utility) provokes significant biases when 3 or 5 levels are assigned to the attribute causing in both cases an underestimate of the marginal value of X_1. Note that, under a *true* stepwise specification, the use of a 5-level design also leads to a significantly underestimated marginal value of X_1. Only when 2 levels are assigned to the attribute it can be ensured that attribute misspecification is irrelevant for measuring attribute marginal values regardless of the *true* effects X_1 has on utility.

When choices are simulated under the MXL-DGP and where the heterogeneity in the marginal utility of income is correctly captured by estimating a MXL model with a lognormally-distributed *Cost* parameter (MXL-MXL/LogN), biases are significant at 5% level in almost all the cases, even when the attribute specification assumed in the model matches the *true* one (except for a *true* stepwise marginal utility when 5 levels are assigned to X_1 and a discrete-linear specification is assumed for estimation). In this scenario, attribute misspecification always leads to higher magnitudes of bias when X_1 takes 3 and 5 levels, which are especially marked when the attribute has *true* non-linear effects: all the marginal values are underestimated when a 3-level design is used. Interestingly, under each *true* attribute specification, when X_1 only varies across 2 levels the three specifications assumed in the model lead to identical underestimated marginal values (i.e. same negative value of bias), this suggesting that, in spite of bias significance, misspecification of X_1 when it is assigned 2 levels does not have any effect on its estimated marginal value.[9]

Looking at the results from incorrect applications of the MNL and the MXL models, it is easy to see that decisions about the nature and number of levels of X_1 gain importance for welfare measurement when the assumptions about the marginal utility of income are mistaken, as not only all biases are significant at 5% level but also are higher in magnitude for almost all the cases when compared with the results from correct applications of the models (MXL-MNL vs. MNL-MNL and MXL-MXL/Triang vs. MXL-MXL/LogN). In this sense, if the marginal utility of income is erroneously considered constant (MXL-MNL), misspecifying the attribute always leads to biases higher in magnitude when assigning 3 and 5 levels to X_1, except when a continuous-quadratic specification is assumed for estimation and 5 levels are assigned to X_1 in a context of *true* linear effects, showing underestimated marginal WTPs in all these cases (MXL-MNL vs. MNL-MNL). However, under each *true* attribute specification, when a 2-level design is used, all the specifications assumed in the model lead not only to equally overestimated marginal values (i.e. same positive value of bias), indicating again that

[9] Surprisingly, the highest value of bias is obtained when the attribute is codified as continuous entering linearly the utility function in a context of *true* linear effects.

attribute misspecification does not have any effect on welfare measurement, but also to the most precise estimates.

Additionally, if the marginal utility of income is correctly assumed non-constant but a triangular distribution is assigned to the *Cost* parameter (MXL-MXL/Triang), although all biases are higher in magnitude when there are *true* linear effects (MXL-MXL/Triang vs. MXL-MXL/LogN), the values derived from models with a misspecified attribute are lower than that from models in which X_1 has been well specified and a 3-level design has been used causing an overestimation of the marginal WTP for X_1. Likewise, when the attribute has *true* non-linear effects on utility, the biases obtained when the attribute is misspecified and 3 or 5 levels are assigned to X_1 are not only lower than those derived from models with well specified attributes –being the marginal values overestimated and underestimated, respectively- but also lower than those derived from correct assumptions about the distribution of the *Cost* parameter. Surprisingly, there seems to be that mistaking the distribution of ω_4 compensates the effects of attribute misspecification when X_1 has non-linear effects (MXL-MXL/Triang vs. MXL-MXL/LogN). Again, if 2 levels are assigned to X_1 there are no attribute misspecification effects as the marginal value is equally overestimated by all the specifications assumed in the model under each *true* attribute specification.

Results regarding bias are confirmed in Table V where the values of MSE in the estimated marginal WTP for X_1 are reported.

V. MSE in the estimated marginal value of X_1 (over 1,000 repetitions)

True attribute specification	Assumed attribute specification	Type of design	MNL-MNL	MXL-MNL	MXL-MXL/LogN	MXL-MXL/Triang
Linear	Continuous-linear	2-level	0,0366	0,0905	0,1163	0,2608
		3-level	0,0549	0,4605	0,1152	0,6383
		5-level	0,0520	0,3662	0,1239	0,5138
	Continuous-Quadratic	2-level	0,0366	0,0905	0,1163	0,2608
		3-level	0,0594	0,5870	0,1242	0,5458
		5-level	0,0633	0,3561	0,1473	0,6140
	Discrete-linear	2-level	0,0366	0,0905	0,1163	0,2608
		3-level	0,0594	0,5870	0,1242	0,5459
		5-level	0,0661	0,4935	0,1630	0,6111
Quadratic	Continuous-linear	2-level	0,0369	0,0837	0,0656	0,1116
		3-level	0,0583	0,3270	0,1279	0,0929
		5-level	0,1854	0,6417	0,2193	0,1253
	Continuous-Quadratic	2-level	0,0369	0,0837	0,0656	0,1116
		3-level	0,0398	0,2220	0,0803	0,1761
		5-level	0,0726	0,1347	0,1120	0,2674
	Discrete-linear	2-level	0,0369	0,0837	0,0656	0,1116
		3-level	0,0399	0,2220	0,0803	0,1761
		5-level	0,0920	0,2378	0,1332	0,2953
3-Stepwise	Continuous-linear	2-level	0,0363	0,0721	0,0589	0,0875
		3-level	0,0388	0,2188	0,0773	0,0912
		5-level	1,2285	2,7923	1,5767	1,1998
	Continuous-Quadratic	2-level	0,0363	0,0721	0,0589	0,0875
		3-level	0,0372	0,1747	0,0677	0,1422
		5-level	0,1688	0,476	0,2557	0,1414
	Discrete-linear	2-level	0,0363	0,0721	0,0589	0,0875
		3-level	0,0372	0,1747	0,0677	0,1422
		5-level	0,0970	0,2965	0,1174	0,2082

To facilitate interpretation of results and help to see the effects from mistaking assumptions about the marginal utility of income, Figures I, II and III present, for each *true* specification scenario and assumption about ω_4, the values of MSE derived from all the attribute specifications assumed in the models under each type of design. In all of them, *LIN* indicates that the assumed specification for X_1 is continuous-linear, *QUAD* means it is continuous-quadratic and *DIS* represents it is discrete-linear, whereas *2L*, *3L* and *5L* corresponds to the use of a 2-level, 3-level and 5-level design, respectively.

Figure I. MSE in the estimated marginal value of X_1 when the *true* attribute specification is continuous-linear

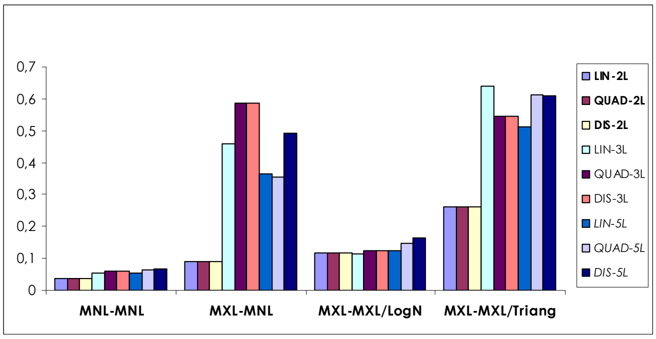


Figure II. MSE in the estimated marginal value of X_1 when the *true* attribute specification is continuous-quadratic

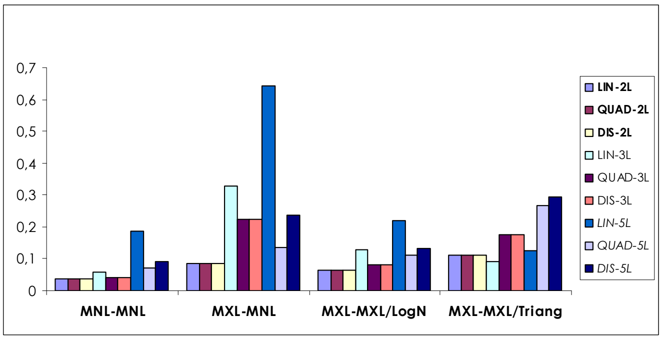
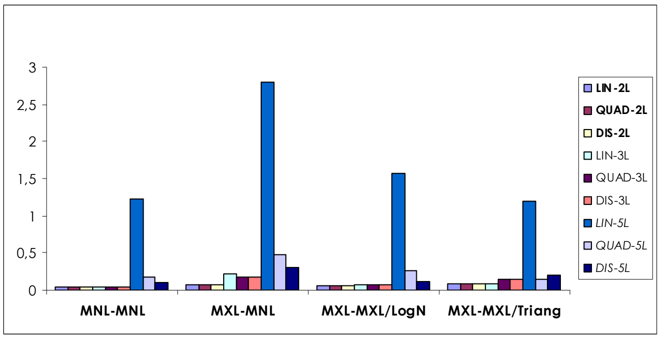


Figure III. MSE in the estimated marginal value of X_1 when the *true* attribute specification is discrete-linear



By looking at the figures, the importance of working under correct assumptions about the marginal utility of income is easily perceived, as indicated by the higher values of MSE obtained when the assumptions about the *Cost* parameter are mistaken. We can see that this is especially so when X_1 has *true* linear effects (Figure I), in which case the increases in MSEs in relation to the values from correct applications of the MNL and MXL models are higher (MXL-MNL vs. MNL-MNL and MXL-MXL/Triang vs. MXL-MXL/LogN). Another straightforward conclusion from Figures II and III is that, when X_1 has *true* non-linear effects on utility, assuming a constant marginal utility of income when it actually varies across individuals provokes higher increases in the values of MSE than mistaking the distribution of the *Cost* parameter (MXL-MNL vs. MNL-MNL and MXL-MXL/Triang vs. MXL-MXL/LogN), specifically when the attribute is misspecified in the models, as already outlined when analyzing results from Table V.

Thus, despite mistaking assumptions about the marginal utility of income worsens precision of estimates in almost all the cases, the key result from the analysis of accuracy measures is that, under each *true* attribute specification, using a 2-level design not only leads to the same value of bias and MSE for all the specifications assumed in the models but also to the most precise estimates of the marginal value of X_1 in three of the four scenarios of assumptions about the marginal utility of income -in all four scenarios when looking at MSE under a *true* stepwise specification. In the light of this, results seem to suggest that working on both simple designs (i.e. 2 attribute levels) and continuous-linear attribute specifications is the best option when designing CEs, because it always avoids attribute misspecification effects and gives the most accurate estimates under almost all the assumptions about the marginal utility of income.[10]

To analyze the sensitivity of results to the magnitude of welfare change –and consequently, examine the robustness of results outlined above– the MC experiments have been repeated considering a hypothetical change in X_1 from the BAU level (6, see Table I) to a level of 4, that is, a smaller attribute change.[11] In general, results show that the effects from mistaking the assumptions about the marginal utility of income are less critical, as the magnitudes of bias are smaller in most cases than those obtained for a higher attribute change. Although the value and sign of some biases have changed, the most important result is that, again, under each *true* attribute specification, the use

[10] Note that when assuming a quadratic specification for an attribute that only takes 2 values, its parameter estimates will be non-significant. Therefore, decisions about attribute specification when using a 2-level design should be restricted to the choice between a continuous-linear and a discrete-linear specification.
[11] Results of bias and MSE for the smaller change in X_1 are available from the authors upon request.

of a 2-level design leads to the same values of bias –if they are significant- and to unbiased estimates –if they are not significantly different from zero- for all the specifications assumed in the models and under the four assumptions about the marginal utility of income, this indicating that attribute misspecification does not have any effect on the value of X_1.[12] As in the case of a high attribute change, the values of MSE for a small change in X_1 again confirm these results.

V. Conclusions

By applying MC analysis, this paper has investigated the importance of the specification of non-monetary attributes in a CE –in terms of their continuous or discrete nature and their number of levels- for estimating their marginal value, under different assumptions about the marginal utility of income. Results show that, although attribute specification generally has effects on the accuracy of estimates that are especially marked when the assumptions about the *Cost* parameter are mistaken, opting for simple specifications can actually yield relatively small errors in welfare estimation. More precisely, when the attribute takes 2 levels, all the attribute specifications assumed in the estimation models lead not only to the same inferred marginal values but also to the most precise estimates in most of cases. In other words, results indicate there is no real justification for an attribute taking more than 2 levels and being specified in a more sophisticated way than a continuous-linear fashion.

These results, however, are subject to the data employed in these MC experiments, that is, to the specific experimental designs, *true* attribute specifications, known parameters and error structures, attribute specifications assumed in the models and assumptions about the marginal utility of income considered. Although this suggests results may not be generalizable to all cases, the experimental designs and methods of analysis used

[12] An interesting result differing from the case of a high attribute change is that when X_1 has *true* non-linear effects the biases obtained when a MXL model is correctly applied by assigning a lognormal distribution to ω_4 are non-significant.

here are hardly un-common. Indeed, continuous and discrete non-monetary attributes with a number of levels lower than 5 and ME only designs are features that can be found in many CE studies reported in the literature. Additionally, the analysis of effects derived from the assumption of an homogeneous *Cost* parameter under the hypothesis that the marginal utility of income actually varies across individuals makes a relevant contribution, as it gives evidence of the magnitude of bias that can be obtained when environmental valuation studies are built, as traditionally done, on the assumption of a constant marginal utility of income.

In a context in which utility specification issues have been largely overlooked in economic valuation studies, this paper is only a first step on the long path to fulfil this gap. Although our results seem promising, it would be interesting to analyze which would happen under different values of known parameters. Likewise, the results under alternative specifications or number of levels of attributes, more than one non-monetary attribute varying across utility specifications and experimental designs constructed on different efficiency criteria or allowing for interactions effects remains to be tested. Further research on these issues could help to examine the robustness of the conclusions drawn here. It would also provide insights into the explanation of unexpected results as obtaining, under a well specified model capturing the heterogeneity of the marginal utility of income, significant biases when the assumed attribute specification matches the *true* one. Therefore, it is time for researchers to take advantage from the increasing power of computers and the development of sophisticated software and to apply MC simulation methods in an attempt to test for these and other related empirical questions that, although being at the core of discrete choice studies, have been largely ignored in the economic valuation literature to date.

VI. References

Adamowicz, W. L., Fletcher, J. J., & Graham-Tomasi, T. (1989). Functional form and the statistical properties of welfare measures. *American Journal of Agricultural Economics, 71*(2), 414-421.

Alberini, A. (1995). Efficiency vs bias of willingness-to-pay estimates: Bivariate and interval-data models. *Journal of Environmental Economics and Management, 29*, 169-180.

Carlsson, F., & Martinsson, P. (2003). Design techniques for stated preference methods in health economics. *Health Economics, 12*(4), 281-294.

Ferrini, S., & Scarpa, R. (2007). Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study. *Journal of Environmental Economics and Management, 53*, 342-363.

Hanemann, W. M. (1984). *Applied welfare analysis with quantitative response models*. Working paper no. 241. University of California, Berkeley.

Hensher, D. A., Rose, J. M., & Greene, W. H. (2005). *Applied choice analysis: a primer*. Cambridge, UK: Cambridge University Press.

Herriges, J. A., & Kling, C. L. (1997). The performance of nested logit models when welfare estimation is the goal. *American Journal of Agricultural Economics, 79*, 792-802.

Kling, C. L. (1987). A simulation approach to comparing multiple site recreation demand models using Chesapeake Bay survey data. *Marine Resource Economics, 4*, 95-109.

Kling, C. L. (1988). The reliability of estimates of environmental benefits from recreation demand models. *American Journal of Agricultural Economics*.

Kling, C. L. (1989). The importance of functional form in the estimation of welfare. *Western Journal of Agricultural Economics, 14*(1), 168-174.

Kling, C. L. (1991). Estimating the precision of welfare measures. *Journal of Environmental Economics and Management, 21*, 244-259.

Kling, C. L. (1997). The gains from combining travel cost and contingent valuation data to value nonmarket goods. *Land Economics, 73*(3), 428-439.

Kling, C. L., & Sexton, R. J. (1990). Bootstrapping in applied welfare analysis. *American Journal of Agricultural Economics, 72*, 406-418.

Kling, C. L., & Thomson, C. J. (1996). The implications of model specification for welfare estimation in nested logit models. *American Journal of Agricultural Economics, 78*, 103-114.

Louviere, J. J., Hernsher, D. A., & Swait, J. D. (2000). *Stated choice methods: analysis and application.* Cambridge, UK: Cambridge University Press.

Lusk, J. L., & Norwood, F. B. (2005). Effect of experimental design on choice-based conjoint valuation estimates. *American Journal of Agricultural Economics, 87*(3), 771-785.

Park, T., Loomis, J. B., & Creel, M. (1991). Confidence intervals for evaluating benefits estimates from dichotomous choice contingent valuation studies. *Land Economics, 67*(1), 64-73.

Scarpa, R., & Bateman, I. (2000). Efficiency gains afforded by improved bid designs versus follow-up valuation questions in discrete-choice CV studies. *Land Economics, 76*(2), 299-311.

Scarpa, R., & Rose, J. M. (2008). Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. *The Australian Journal of Agricultural and Resource Economics, 52*, 253-282.

Torres, C. M., Riera, A., & García, D. (In press). Are preferences for water quality different for second-home residents? *Tourism Economics*.