# DOCUMENTOS DE TRABAJO

# BILTOKI

D.T. 2010.09

Selecting random parameters in discrete choice experiment for environmental valuation: A simulation experiment.

Petr Mariel, Amaya De Ayala, David Hoyos y Sabah Abdullah

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Facultad de Ciencias Económicas.
Avda. Lehendakari Aguirre, 83
48015 BILBAO.

**Documento de Trabajo BILTOKI DT2010.09**
Editado por el Departamento de Economía Aplicada III (Econometría y Estadística) de la Universidad del País Vasco.

Depósito Legal No.: BI-3161/2010
ISSN: 1134-8984
URN: RePEc:ehu:biltok:201009

# Selecting random parameters in discrete choice experiment for environmental valuation: A simulation experiment

**PETR MARIEL[1], AMAYA DE AYALA[2], DAVID HOYOS[3] and SABAH ABDULLAH[4]**

*[1]Department of Applied Economics III (Econometrics and Statistics), University of the Basque Country, E48015 Bilbao, Spain (email: petr.mariel@ehu.es); [2] Department of Applied Economics III (Econometrics and Statistics), University of the Basque Country, E48015 Bilbao, Spain; [3] Department of Applied Economics III (Econometrics and Statistics), University of the Basque Country, E48015 Bilbao, Spain; [4] Environmental Economics Unit, Institute for Public Economics, University of the Basque Country, E48015 Bilbao, Spain*

**Abstract.**

This paper examines the various tests commonly used to select random parameters in choice modelling. The most common procedures for selecting random parameters are: the Lagrange Multiplier test as proposed by McFadden and Train (2000), the *t*-statistic of the deviation of the random parameter and the log-likelihood ratio test. The identification of random parameters in other words the recognition of preference heterogeneity among population is based on the fact that an individual makes a choice depending on her/his: tastes, perceptions and experiences. A simulation experiment was carried out based on a real choice experiment and the results indicated that the power of these three tests depends importantly on the spread and type of the tested parameter distribution.

**Key words:** choice experiment, preference heterogeneity, random parameter logit, simulation, tests for selecting random parameters.

**JEL classification:** Q51

## 1. Introduction

In the last decade, the use of the discrete choice experiments (DCEs) for the purpose of nonmarket valuation of environmental goods has increased in popularity particularly amongst environmental economists. In essence, this valuation method involves respondents being presented with a series of alternatives characterized by attributes and they are asked to indicate their preferred options from this set. Using the set of observed discrete choices, researchers can estimate separately marginal values for each attribute or attribute level used in describing the project or good alternatives.

Typically, a DCE is characterised by a number of key stages: (1) definition of attributes and their levels of provision, (2) experimental design, (3) questionnaire development, and (4) the estimation procedure. The last stage requires decisions ranging from what type of variables to include in the specification of the models, to the econometric model to be assumed. For instance, in econometric approaches based on a multinomial logit model (MNL), respondents' tastes are assumed to be homogenous throughout the population. But one criticism of the MNL is that everyone is assumed to have similar preferences (Colombo et al., 2005). In real life situations, however, individuals' tastes vary. This variation is based on the fact that an individual makes a choice depending on his/her: tastes, perceptions, attitudes and experiences, which in turn are influenced by socio-economic and demographic factors, such as: income, age and education. In other words, because these factors differ from one person to another, heterogeneity exists in terms of their tastes and preferences. This variation is important to take into account, particularly so as to understand the motivation behind the behaviour and decision making process that affects choice selection. If such variations are ignored when carrying out welfare and preference estimations, then this leads to biased results.

One econometric model that allows for the aforementioned parameters to vary across individuals, thereby accommodating heterogeneity, is the mixed logit model (MXL). The inclusion of heterogeneity provides more information, regarding the influence of socio-economic and demographic factors in respondents' decision making, during the experimental design. The main task when applying this model is to find variables and a mixing distribution that takes into account the other components of utility, which correlate over alternatives or are heteroskedastic (Train, 2003). In uncovering patterns of tastes and respondents' preference heterogeneity among MXL models, it is pertinent to permit wider variation to uncover more influences which affect utility decision making. Moreover, MXL models are known to relax the independence of irrelevant alternatives (IIA) property and can either be random parameter logit (RPL) or the error component logit (ECL) model. The difference between the two specifications is interpretational (Brownstone, 2001) and according to Train (2003), the RPL and ECL models are equivalent. However, the choice regarding which of these models to apply depends on the researcher's interest. In this regard, the RPL, after Revelt and Train (2000), allows for every variable coefficient to vary and be correlated, whereas the ECL model only allows for the errors to be correlated and vary (Brownstone and Train, 1999).

Arising from uncovering heterogeneity among population is relevant for policy implications where estimates are essential in targeting various preferences for groups of interest. In this paper we analyze the empirical size and statistical power of three tests for selecting random parameters for an RPL model under different type and spread of the tested parameter distributions, by using a simulation exercise. This exercise is based on using an actual dataset to obtain conclusions which could be directly applied to DCEs on environmental valuation.

This paper is divided into five sections: This introduction (section 1) is followed by the methodology framework and explanations of the different randomness tests (section 2). Next, in section 3 the simulation exercise is described, which is put forward by presentation of the results and discussion (section 4). Finally, section 5 contains the conclusion.

## 2. Methodology for selecting random parameters

The analysis of DCE choices is based on random utility theory as developed by McFadden (1974), which links the deterministic model with a statistical model of human behaviour. In this regard, the randomness of the utility function suggests that only analysis of the probability of choosing one alternative over another is possible. However, estimable choice models require a distributional assumption for the random component, which has resulted in different random utility models (RUM) being developed. One of the econometric models based on RUM is the MNL form, which specification assumes that the error terms of the utility function are independently and identically distributed (IID) following a type I extreme value (Gumbel) distribution (McFadden, 1974 and Louviere et al., 2000). The power and limitations of the MNL model are as follows (Train, 2003): (1) it can represent systematic taste variations (i.e.

those related to the observed characteristics of the respondents), but not random taste variations (i.e. those that cannot be linked to the observed characteristics of the respondents), (2) it exhibits restrictive substitution patterns, because it implies proportional substitution across alternatives given the specification of the utility function (satisfies IIA property) and (3) it can handle situations where unobserved factors are independent over time, but it cannot be used with panel data when unobserved factors are correlated over time for each respondent.

One possible way to overcome the above limitations is the use of the RPL model. Increasingly, researchers and practitioners are devising sophisticated model taking into account mixtures of revealed preference and stated choice data (Hensher et al., 2003). One reason of the increased growth in the use of the RPL model in recent years can be partly be explained by its inclusion in standard econometric software and partly by their flexible assumptions. Brownstone (2001) states that the RPL popularity has kept growing in spite of some problems related to inference and model selection. Some of the applications can be found across a number of areas, including: transport (Amador et al., 2005, Brownstone et al., 2000, and Bhat, 1997 and Bhat 2000), recreation (Hanley at al., 2002 and Train, 1998) and health (Personn, 2002) among many others. Moreover, other examples covering the environmental areas include studies done by: Colombo et al. (2008), Hanley et al., (2006) and Revelt and Train (1998).

The three main advantages of using an RPL specification namely: allows preference heterogeneity; avoids reliance on the IIA property and incorporates correlation across the responses of the individual who face different choice sets. Hensher et al. (2003) notes that although the theory is relatively clear, estimation and data issues are far from resolved. For instance, the RPL model is where an individual's utility from any alternative in the choice set includes a stochastic part that may be correlated over alternatives and that may be heteroskedastic (Hensher et al., 2005). In this model, preference heterogeneity is directly incorporated into the vector of parameters, so the vector of coefficients of attributes is different for each individual ( $\beta_i$ ) and it is allowed to deviate from the population mean coefficient $\beta$ by the vector of deviation parameters $\eta_i$. Thus, its utility function of an individual *i* and alternative *j* is the following:

$$U_{ij} = \beta x_{ij} + \eta_i x_{ij} + \varepsilon_{ij}, \tag{1}$$

where $x_{ij}$ is an alternative value and $\varepsilon_{ij}$ is the error component.

Importantly, the RPL model handles the case of unobserved heterogeneity by assuming that (some of) the weighting coefficients vary in the population according to some distribution and estimating the parameters of those distributions. Thus, to estimate an RPL model it is necessary to make a few decisions: (1) which coefficients are assumed to be random, (2) the type of distributions to use and (3) the economic interpretation for those coefficients.

In the DCE literature researchers usually determine the random coefficients in an RPL using three procedures: the Lagrange multiplier (LM) test proposed by McFadden and Train (2000), the *t*-statistic of the deviation of the random parameter and the log-likelihood ratio (LR). However, there is limited information for practitioners about the performance of these tests which are used to determine which parameters should be random in the RPL specification.

### *2.1. Lagrange Multiplier (LM) test proposed by McFadden and Train (2000)*

Prior to estimating the RPL model, it is useful to carry out a specification test to determine whether this treatment is appropriate. As proposed by McFadden and Train (2000), a choice from a set *C* = {1,...,J} is considered. Let $x_j$ be a 1×K vector of attributes of alternative *j*. From a random sample *n = 1,...,N*, the test estimates the parameter β in the MNL model:

$$P_j = \frac{e^{\beta' x_j}}{\sum_{h \in c} e^{\beta' x_j}}, \tag{2}$$

where $P_j$ is the choice probability for alternative *j*, using maximum likelihood and constructing artificial variables:

$$z_{tj} = \frac{1}{2}(x_{tj} - x_{tC})^2 \text{ with } x_{tC} = \sum_{j \in C} x_{tj} \cdot P_j, \tag{3}$$

where *t* denotes the parameters that are suspected to be random, *C* is the set of alternatives being offered and $P_j$ is defined in (2). A Wald or Likelihood Ratio Test can be used to test the null hypothesis that the artificial variables $z_{tj}$ should be omitted from the MNL model. The null hypothesis of no random coefficients is therefore rejected if the parameters of the artificial variables are significantly different from zero.

There is dearth of literature that applies McFadden and Train test to specify the model to use. One study that has applied this test is Hoyos et al. (2009), where a choice experiment was conducted in the Basque Country to determine the non-market values of the environmental attributes of the Jaizkibel natural area. In this study 2448 observations in relation to five attributes were considered, with the randomness of all the coefficients being tested using the LM test as proposed by McFadden and Train (2000) and it was found that only two attribute parameters were random. Similarly, Liljenstolpe (2008) evaluated animal welfare in relation to Swedish pig production, with seven attributes and using 1250 observations, concluded that six of the tested parameters were random.

Another study carried out by Brey et al. (2007) estimated economic welfare for an afforestation programme in the Northeast of Spain using six attributes with 730 observations.  Random coefficients of the RPL model were determined combining the LR test and the Lagrange Multiplier test after McFadden and Train (2000). The results under both approaches determined that two coefficients were random.

### *2.2. t-statistic for the deviation*

The *t*-statistic for standard deviation is commonly used to determine the random parameters for its straightforward and simple application. This is a common procedure for most applications of DCE in the literature. For instance, in the area of housing supply, Mohammadian et al. (2008) carried out a DCE using 4 attributes with 1384 observations to analyze what influences the location and housing choices of Canadian homebuilders. Their results showed that the estimated standard deviations of the random parameters of two attributes had significant *t*-statistics, so they were considered random.

With regards to environmental research, various valuation studies have used the *t*-statistic to select random parameters. For example, in Wang et al. (2007) a DCE comprised six attributes with 2730, 2985 and 3045 observations in three different chinese communities to analyse the impact of an environmental improvement. In the three different communities one, two and two parameters respectively were found to be random from ten coefficients of the model. Birol et al. (2006) carried out a DCE on the Cheimaditida wetland in Greece, which contained six attributes and 700 observations. In their RPL model, the estimates revealed significant and large derived standard deviations for three attributes, thus supporting unconditional unobserved heterogeneity for these attributes. Hanley et al. (2006) estimated the value of the improvements of two rivers in the UK, considering six attributes and 420 observations. An RPL was estimated and they noted that only the standard deviation of one attribute was statistically significant, which implied that the major component of preference heterogeneity was only on one attribute.

Similarly, Sillano and Otúzar (2005) derived willingness to pay for reduction in atmospheric pollution in Santiago (Spain) applying a DCE with four attributes and 648 observations. In their study they found that the estimated deviations of all parameters were statistically negligible, hence in the final estimation there were considered as being fixed. In another study, Colombo and Hanley (2008) carried out a DCE for the conservation of mountain agriculture in the Northeast of England involving six attributes and 1275 observations. The significant standard deviations they encountered indicated that two attributes were substantially heterogeneous. Finally, Train (1998) estimated a model of anglers' choice among river fishing sites in Montana taking into account 8 attributes and obtaining 962 observations. The estimated standard deviations of all RPL coefficients were highly significant, signifying that there was variation among the population.

### *2.3. Log-likelihood ratio (LR)*

The log-likelihood ratio test (LR test) is used to compare the values of likelihood functions corresponding to two estimated models where one is nested within the other, i.e. the LR test compares the log likelihood function between the MNL and RPL models.

Some empirical applications in the DCE field have used LR test to evaluate the randomness of the proposed model. Apart from environmental research carried out by Wang et al. (2007) that combined the *t*-statistic and an LR test, as mentioned above, another study by Campbell et al. (2009) estimated the economic benefits of policy measures to improve the rural landscape in Ireland by using 3 attributes. The log-likelihood function was found to be statistically higher under a RPL model specification, where all the attributes are taken as being random, as compared to an MNL specification.

In Revelt and Train (2000), a DCE with four alternative households' electricity suppliers each one with three attributes analyses 4308 observations. The highest log-likelihood value was obtained under a model where three non-price coefficients were normally distributed and two were log-normal. Finally, Hall et al. (2007) apply a DCE with twelve attributes and 3360 observations from the Jewish community and 4176 from the general community, who were participating in genetic testing programmes. All the coefficients of the attributes were considered random and the conventional LR test was applied for testing that.

## 3. Simulation experiment

The simulation experiment approach is applied in other studies such as Bhatta and Larsen (2010) and Fiebig et al. (2009). Bhatta and Larsen (2010) analyzed possible structure and magnitude of biases introduced to the coefficients of a MNL of travel choice due to random measurement errors in two variables using a simulation exercise. They set up a model in which the "true" parameters are based on those estimated in previous studies. Then, they computed the deterministic part of the utilities using the original variables (*X*s, levels of attributes) and "true" parameters. Adding a random Gumbel (0.1) error they obtained hypothetical levels of utilities. Thereafter, the choices are determined by the highest utility. This way, their hypothetical data set is based on real observations and fulfils the assumption of MNL model. In our simulation experiment the same approach was applied.

In another study (Fiebig et al., 2009) a similar approach is also used where two simulation experiments are carried out in order to evaluate the properties of Generalized Multinomial Logit Model (G-MNL). To make the experiments realistic the authors constructed their simulated data sets based on two real data sets. That is, the actual *X*s from the empirical data set are used together with the "true" parameters obtained by estimating G-MNL model to generate hypothetical utilities.

Typically, a choice experiment in environmental valuation involves: three alternatives containing between three and six attributes (including the cost attribute), with at least two possible levels and there being between 1200 and 2500 observations. That is why the reference article for the simulation exercise is Hoyos et al. (2009) which evaluates the Jaizkibel natural area in Guipúzcoa (Spain). In fact, this valuation study is similar to others reported in the DCE literature (e.g.; Campbell et al., 2009, Colombo and Hanley (2008), Brey et al., 2007, Birol et al., 2006, Hanley et al., 2006 or Colombo et al., 2005). The key attributes and levels considered in the study chosen include: (1) Landscape, the percentage of protected surface area in the future; (2) Flora, the future level of protection of today's population of the A. euskadiensis endemism; (3) Avifauna, the future level of protection of today's population of lesser and peregrine falcons; (4) Seabed, the future level of protection of today's extension of red algae; and (5) Annual contribution in Euros, varying from 5€ to 100€. Detailed information about the environmental features of Jaizkibel and the survey design can be found in Hoyos et al. (2009). The considered levels of these attributes are depicted in Table 1.

***Table 1.*** Attributes and levels considered for the DCE of Jaizkibel natural area

| Attribute | Level | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Landscape** | 40%* | 60% | 80% | 100% | | | | |
| **Flora** | 50%* | 70% | 85% | 100% | | | | |
| **AviFauna** | 25%* | 50% | 75% | 100% | | | | |
| **Seabed** | 50%* | 70% | 85% | 100% | | | | |
| **Annual payment** | 0 €* | 5 € | 10 € | 15 € | 20 € | 30 € | 50€ | 100€ |

* Levels with asterisks represent the status quo scenario.

Source: Hoyos et al. (2009)

For the simulation exercises the deterministic part in the utility function ($V_{ij}$) for *i* individual and *j* alternative was defined in equation (4)[1]:

$$V_{ij} = \beta_1 + \beta_2 Payment_{ij} + \beta_3 Landscape_{ij} + \beta_4 Flora_{ij} + \beta_5 Avifauna_{ij} + \beta_6 Seabed_{ij}. \qquad (4)$$

To examine the operating characteristics under the different frameworks of the: McFadden and Train (2000) test, *t*-statistic test and LR test we run simulation exercises by taking into account the context of the economic valuation of the Mount Jaizkibel as in Hoyos et al. (2009). In the study, three alternatives were offered with a random utility function as shown in equation (5):

[1] In this specification, we use an alternative-specific constant β1 only in the equation for the status quo option. The equations for Options A and B do not include any alternative-specific constants because they are both generated from the same experimental design.

$$U_{ij} = V_{ij} + \varepsilon_{ij}, \tag{5}$$

where $V_{ij}$ is defined as in (4) and the disturbances $\varepsilon_{ij}$ were i.i.d. extreme value type I.

We carried out 500 repetitions of the test procedures for a sample of 2448 observations (sample fixed by the actual survey). The three tests procedures were carried out under the null and various alternative hypothesis for the parameters: $\beta_2$, $\beta_3$,...,$\beta_6$ of (3). Under the null hypothesis the values of the "true" parameters were fixed according to Table 2. These are almost equal to the MNL estimated values of the actual study.

***Table 2***. Fixed parameters in $H_0$:

| | |
|---|---|
| [Constant ($\beta_1$)] | -0.700 |
| β_Payment ($\beta_2$) | -0.150 |
| β_Landscape ($\beta_3$) | 0.008 |
| β_Flora ($\beta_4$) | 0.008 |
| β_Avifauna ($\beta_5$) | 0.008 |
| β_Seabed ($\beta_6$) | 0.008 |

During the simulation exercises under the null hypothesis, the McFadden and Train (2000) test was applied in a previously defined MNL model including artificial variables and for the other two tests (*t*-statistic test and LR test) an RPL model was estimated. For the latter tests, the normal distribution was assumed for the tested parameters.

Under the alternative hypothesis for the three tests, distinct types of the distributions were considered: normal, uniform, triangular and lognormal. Furthermore, each type of distribution with three different spreads (from wide to narrow) was assumed.

For the normal distribution, the value of the mean was set to the value presented in Table 2 and the value of the deviation was allowed to change to: half of the mean (σ=μ/2), a quarter of the mean (σ=μ/4) and an eighth of the mean (σ=μ/8). The specification for the uniform distribution was similar to the normal distribution, so its limits *a* and *b* were set to the 2.5% and 97.5% percentile of the normal distribution and the limits *a* and *b* of the triangular distribution were defined in the same manner. Finally, the lognormal distribution was specified as equalling the 5% percentile to the 2.5% percentile of the normal distribution so as to obtain higher resemblance in the shape of the two distributions, as one of the tails does not exist in lognormal distribution.

The following Tables 3 and 4 provide information on the assumed specifications of the different kind of distributions for the considered coefficients β_Payment and β_Avifauna respectively. The specification for the remaining parameters β_Landscape, β_Flora and β_Seabed are the same as for the β_Avifauna because the values of the parameters (see Table 2) are the same.

***Table 3***. Specifications of the distributions for β_Payment

| | | **β _Payment** | | |
|---|---|---|---|---|
| | | Wide spread | Medium spread | Narrow spread |
| **Distribution types** | **Normal (-0.15, σ)** | σ =0.075 | σ =0.0375 | σ =0.01875 |
| | **Uniform (a, b)** | (-0.297, -0.003) | (-0.2235, -0.0765) | (0.1875, -0.11325) |
| | **Triangular (a, b)** | (-0.297, -0.003) | (-0.2235, -0.0765) | (-0.1875, -0.11325) |
| | **Lognormal (ϴ, λ)** | (-2.1, 0.5) | (-1.91, 0.25) | (-1.8832, 0.125) |

*Table 4*. Specifications of the distributions for β _Avifauna

| | | β _Avifauna | | |
|---|---|---|---|---|
| | | Wide spread | Medium spread | Narrow spread |
| Distribution types | **Normal (0.008, σ)** | σ =0.004 | σ =0.002 | σ =0.001 |
| | **Uniform (a, b)** | (0.00016, 0.01584) | (0.00408, 0.0119) | (0.00302, 0.0049) |
| | **Triangular(a, b)** | (0.00016, 0.01584) | (0.00408, 0.0119) | (0.00302, 0.0049) |
| | **Lognormal (ϴ, λ)** | (-5.132, 0.6) | (-4.922, 0.3) | (-4.855, 0.15) |

# 4. Results and discussion

We applied the three tests described above for alternately generated choices under the defined null and alternative hypotheses. The corresponding RPL models were estimated using 200 Halton draws. The empirical size and power of the: McFadden and Train (2000), *t*-statistic and LR tests were obtained for all the attributes' coefficients after applying actual data found in Hoyos et al (2009). Nevertheless, the results presented here only include those for two attributes' coefficients as the results with regards to the other attributes were similar to these two.

Firstly, the results are presented for the coefficient of Payment (β_Payment), an attribute with 8 levels ranging from 0€ to 100€ and corresponding hypothetical parameter with negative sign. Secondly, the results for the coefficient of Avifauna (β_Avifauna) are almost the same as the rest of the attributes, as: all of them have four similar levels (between 25% and 100%), their corresponding parameters under the null hypothesis have the same values and that is why they present similar empirical size and power in the simulations.

Tables 5 and 6 show the empirical size (under $H_0$) and statistical power (under different $H_a$) of the analyzed tests. The first two rows provide information on the empirical size under the null hypothesis for 5% and 10% significance levels. The following rows illustrate histograms of the assumed random parameters under the aforementioned different types of distribution (normal, uniform, triangular and lognormal) and for different distribution spreads (from wide to narrow). Below histograms, the rows contain the power percentages for the 5% and 10% significance levels for the different randomness tests and varied distribution type.

***Table 5***. McFadden and Train (McFT) test, *t* test and LR test - **β_Payment**

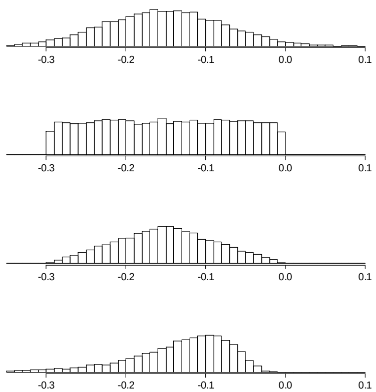
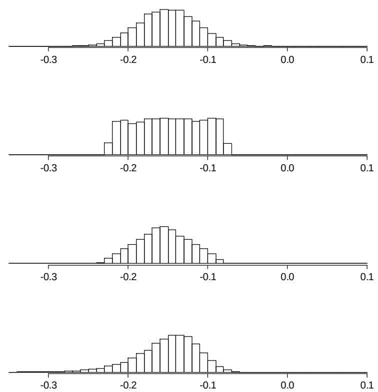
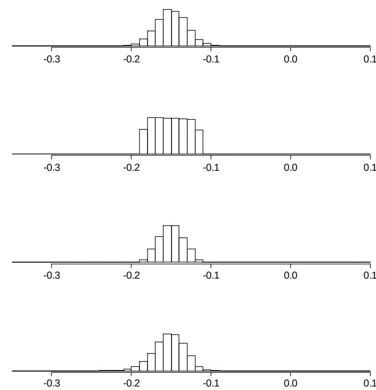
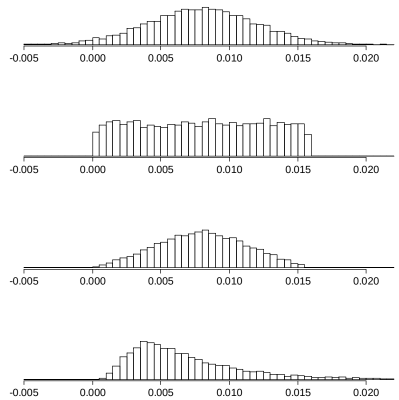
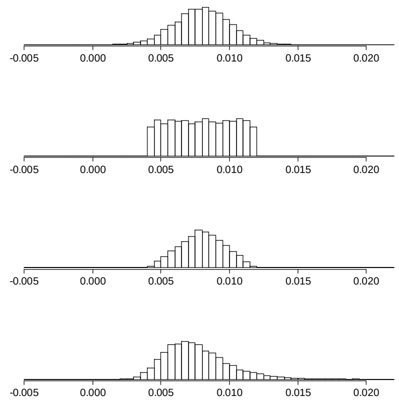
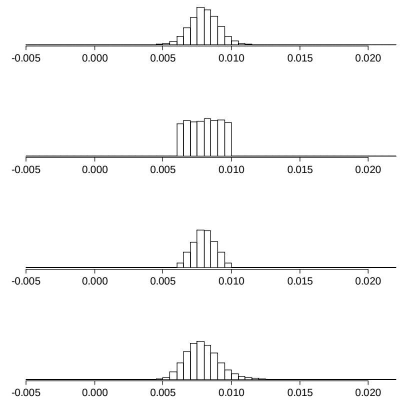
| **Empirical size (under $H_0$)** | | **McFT** | | | **t** | | | **LR** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **5%** | 4.6% | | | 15.4% | | | 1.2% | | |
| | **10%** | 10% | | | 20.8% | | | 3.4% | | |
| | | Wide spread | | | Medium spread | | | Narrow spread | | |
| **Tested distributions** | |  | | |  | | |  | | |
| **Power (under $H_a$)** | | **McFT** | ***t*** | **LR** | **McFT** | ***t*** | **LR** | **McFT** | ***t*** | **LR** |
| ***Normal*** | 5% | 100% | 100% | 100% | 27.6% | 61.6% | 28% | 5.2% | 19.6% | 3.6% |
| | 10% | 100% | 100% | 100% | 37.0% | 67.8% | 38.6% | 12.2% | 27.8% | 8.6% |
| ***Uniform*** | 5% | 100% | 100% | 100% | 20.6% | 56.2% | 28.8% | 6.2% | 22.2% | 6.2% |
| | 10% | 100% | 100% | 100% | 32.6% | 65.0% | 38.8% | 11.6% | 29.2% | 12% |
| ***Triangular*** | 5% | 93.2% | 98.8% | 95.6% | 10% | 36.4% | 12.8% | 4.8% | 17.8% | 2.6% |
| | 10% | 96.2% | 99.4% | 97% | 17.4% | 46.6% | 18.6% | 8.6% | 26% | 6.4% |
| ***Lognormal*** | 5% | 61% | 85.6% | 69.0% | 12.2% | 36.6% | 17.0% | 4.8% | 15.0% | 4.8% |
| | 10% | 70.6% | 90.2% | 78.0% | 23.2% | 45.2% | 24.4% | 9.4% | 21.20% | 9.6% |

***Table 6***. McFadden and Train (McFT) test, *t* test and LR test - **β_Avifauna**

| **Empirical size (under $H_0$)** | | **McFT** | | | **t** | | | **LR** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **5%** | 6.2% | | | 14% | | | 3.4% | | |
| | **10%** | 12% | | | 19.4% | | | 7% | | |
| | | Wide spread | | | Medium spread | | | Narrow spread | | |
| **Tested distributions** | |  | | |  | | |  | | |
| **Power (under $H_a$)** | | **McFT** | ***t*** | **LR** | **McFT** | ***t*** | **LR** | **McFT** | ***t*** | **LR** |
| ***Normal*** | 5% | 3.6% | 13.8% | 3.8% | 3.6% | 12.4% | 3.8% | 4.8% | 11.6% | 3.4% |
| | 10% | 8.6% | 19.4% | 6.4% | 9.6% | 17.6% | 6.2% | 9.2% | 16.8% | 5.2% |
| ***Uniform*** | 5% | 4.4% | 16.8% | 3% | 6.6% | 15.4% | 3.6% | 6.8% | 16% | 3% |
| | 10% | 10.8% | 22.8% | 7.2% | 12.8% | 20.8% | 7% | 14.4% | 19.2% | 7.4% |
| ***Triangular*** | 5% | 4.4% | 12.6% | 2.8% | 4.2% | 12.2% | 2.8% | 5.4% | 12.2% | 3% |
| | 10% | 9% | 20% | 6% | 9.8% | 18.2% | 5.8% | 10.4% | 19% | 6% |
| ***Lognormal*** | 5% | 4% | 5.75% | 1.5% | 4.4% | 5.8% | 1.8% | 4.6% | 6.4% | 2.0% |
| | 10% | 10% | 8.7% | 3.5% | 9.8% | 8.8% | 3.6% | 8.0% | 9.0% | 3.2% |

Concerning the empirical size of the different tests, we can see that whilst the McFadden and Train (2000) test presents an empirical level near to the theoretical levels (around 5% and 10%), the *t*-statistic has much higher empirical size and LR test has lower than the expected theoretical values.

Regarding the statistical power of the tests, the general conclusion is that the *t*-statistic provides us the highest power under all the alternative hypotheses. Nevertheless, as we have seen before, it has a higher power at the expense of misstated empirical size. Similarly, the LR test presents overall the lowest power owing to its low empirical size. The only test with a suitable empirical size is the McFadden and Train (2000) test, which shows high power under all the considered distributions with high deviations, i.e. wide distributions. As long as the spread of the distribution gets narrower the power decreases. We can conclude that this effect applies for the other two tests too. In sum, a wider spread of the assumed distribution implies a better power of the tests.

Turning to the influence of the type of distribution, higher power percentages are found under the normal and uniform distributions whereas lower powers appear under the lognormal distribution. Importantly, it emerges that the distribution type has an effect on the power of the tests and more reliable results are achieved if the underlying distribution of the random parameter is normal or uniform rather than triangular or lognormal.

Prior to selecting random parameters for the RPL specification, it is important to know the following shortfalls of each test. In this regard, the simulation exercises have demonstrated that the power of the McFadden and Train (2000) test, which is the only one with expected empirical size, is low when the distribution of the tested parameter is narrow. Moreover, the results of this exercise shows that given the high empirical size under the null, lower significance level of the testing of random parameters should be used (e.g. 1%) using *t*-statistic test. Conversely, the LR test presents an empirical level below the theoretical one, hence it is recommended to work with a higher significance level when testing for random parameters (e.g. 10%). The recommended step of adjusting the varied significance levels when applying one of the three tests is clear and concise "rule of thumb" for the econometric practitioners as evident by the above simulation experiment.

## 5. Conclusions

One criticism of the standard MNL model is that it assumes homogenous preferences across the population: in other words, everyone is assumed to have "average and identical" preferences (Colombo et al., 2005). By contrast, the RPL formulation of MXL model allows the parameters to vary across individuals, to accommodate this heterogeneity by assuming that (some of) the weighting coefficients vary in the population according to some distribution, and estimating the parameters of those distributions. However, there is a lack of formal procedures for practitioners to select random parameters in RPL model. Three different tests are usually applied in DCE: the *t*-statistic test, the Lagrange Multiplier test as proposed by McFadden and Train (2000) and the LR test.

This paper's main focus is on simulation exercises based on an actual choice experiment found in Hoyos et al. (2009). We argue that the results are relevant and applicable to other valuation studies using choice experiment for environmental valuation.

We have concluded that the power of these tests depends on the spread and types of the distribution assumed. For instance, under a wider spread and a normal or uniform distribution the power of such tests is higher. However, due to the LR test's low empirical size and *t*-statistic test's misstated empirical size, it is suggested that when testing for random parameters for the former a higher significance level should be used and the converse for the latter.

The policy implications emerging from this study are that researchers should acknowledge the tests limitations when selecting and specifying one econometric model over the other. Indeed, selection of appropriate policy programmes depends on right WTP values computed using the estimation of the model chosen. WTP values generated from models that do not consider heterogeneity or erroneous type of heterogeneity as a consequence of inappropriate selection of random parameters affect the evaluation of costs and benefits for specific projects. Consequently, such projects or policy programmes can be erroneously established. Moreover, acknowledging this heterogeneity among the population results to efficient WTP values that can assist policy makers to target effectively appropriate programmes to specific groups as demanded. In sum, this paper offers simple rules for applying random parameter tests in a typical DCE for environmental valuation and helps to fill the gap between theory and practice.

## Acknowledgements

The authors are grateful to the Department of Education of the Basque Government through grant IT-334-07 (UPV/EHU Econometrics Research Group) and from Spanish Ministerio de Educación and FEDER (SEJ2007-61362).

# References


Amador, F.J., Gonzales, R. and Ortuzar, J. (2005), “Preference Heterogeneity and Willingness to Pay for Travel Time Savings”, *Transportation* **32**, 627-647.

Bhat, C.R. (2000), "Incorporating Observed and Unobserved Heterogeneity in Urban Work Travel Mode Choice Modeling”, *Transportation Science* **34**, 228-238.

Bhat, C.R. (1997), “An Endogenous Segmentation Mode Choice Model with An Application to Intercity Travel”, *Transportation Science* **31**, 34-48.

Bhatta B.P., and Larsen, O.I. (2010), “Errors in variables in multinomial choice modeling: A simulation study applied to a multinomial logit model of travel mode choice”, Transport Policy, forthcoming.

Birol, E., Karousakis, K. and Koundouri, P. (2006), “Using A Choice Experiment to Account for Preference Heterogeneity in Wetland Attributes: The Case of Cheimaditida Wetland in Greece”, *Ecological Economics* **60**, 145-156.

Brey, R., Riera, P. and Mogas, J. (2007), “Estimation of Forest Values Using Choice Modelling: An Application to Spanish Forests”, *Ecological Economics* **64**, 305-312.

Brownstone, D. (2001), “Discrete Choice Modelling for Transportation”, in Hensher, D. (ed.), *Travel Behaviour Research: The Leading Edge*, Pergamon Press, Amsterdam, pp. 97-124.

Brownstone, D., Bunch, D. and Train, K. (2000), “Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-Fuelled Vehicles”, *Transportation Research B* **34**, 315-338.

Brownstone, D. and Train, K. (1999), “Forecasting New Product Penetration with Flexible Substitution Patterns,” *Journal of Econometrics* **89**, 109-129.

Campbell, D., Hutchinson, W.G. and Scarpa, R. (2009), “Using Choice Experiments to Explore the Spatial Distribution of Willigness to Pay for Rural Landscape Improvements”, Enviroment and Planning A **41**, 97-111.

Colombo, S. and Hanley, N. (2008), “Econometric Analysis of Individuals Preference Heterogeneity: An Application to the Economic Valuation of Uphill Landscape”, *Economía Agraria y Recursos Naturales* **8**, 103-24.

Colombo, S., Hanley, N. and Calatrava-Requena (2005), “Designing Policy for Reducing the Off-Farm Effects of Soil Erosion Using Choice Experiments”, Journal of Agricultural Economics **56**, 81-95.

Ferrini, S. and 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.

Fiebig, D.G., Keane, M.P., Louviere, J.J. and Wasi N. (2009), "The Generalized Multinomial Model: Accounting for Scale and Coefficient Heterogeneity", *Marketing Science* **29**, 393-421.

Hall, J., Fiebig, D.G., King, M.T., Hossain, I. and Louviere, J.J. (2007), "What Influences Participation in Genetic Carrier Testing? Results from A Discrete Choice Experiment", *Journal of Health Economics* **25**, 520-537.

Hanley, N., Wright R.E. and Alvarez-Farizo, B. (2006), "Estimating the Economic Value of Improvements in River Ecology Using Choice Experiments: An Application to the Water Framework Directive", *Journal of Environmental Management* **78**, 183-193.

Hanley, N., Wright, R. and Koop, G. (2002), "Modelling Recreational Demand Using Choice Experiment: Climbing in Scotland", *Environmental and Resource Economics* **22**, 449-466.

Hensher D. A., Rose , J.M. and Greene, W.H. (2005), *Applied Choice Analysis. A Primer*, Cambridge University Press, New York.

Hensher, D. A. y Greene, W.H. (2003), “The Mixed Logit Model: The State of Practice”, *Transportation* **30**, 133-176.

Hoyos, D., Mariel, P. and Fernández-macho, J. (2009), “The Influence of Cultural Identity on the WTP to Protect Natural Resources: Some Empirical Evidence”, *Ecological Economics* **68**, 2372-2381.

Liljenstolpe, C. (2008), “Evaluating Animal Welfare with Choice Experiments: An Application to Swedish Pig Production”, *Agribusiness* **24**, 67-84.

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

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

McFadden, D. (1974), “Conditional Logit Analysis of Qualitative Choice Behaviour”, in Zarembka, P. (ed.), *Frontiers in Econometrics*, Academic Press, New York, pp. 105-142.

Mohammadian, A., Haider, M. and Kanaroglou, P.S. (2008), “Homebuiders Choice Behaviour Analysis”, *Canadian Journal of Regional Science* **31**, 225-39.

Personn, T.H. (2002), “Welfare Calculations in Models of the Demand for Sanitation”, *Applied Economics* **34**(12), 1509-1518.

Revelt, D. and Train, K. (2000), ``Customer-Specific Taste Parameters and Mixed Logit'', *Department of Economics, University of California at Berkeley*, CA, WP E00-274.

Revelt, D. and Train, K. (1998), "Incentives for Appliance Efficiency in A Competitive Energy Environment: Random Parameter Logit Models of Household's Choices", *Review of Economics and Statistics* **80**, 647-57.

Sillano, M., Otúzar and J.D. (2005), "Willigness to Pay Estimation with Mixed Logit Models: Some New Evidence", Enviroment and Planning A **37**, 525-550.

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, New York.

Train, K. (1998), "Recreation Demand Models with Taste Differences over People", *Land Economics* **74**, 32-48.

Wang, X., Bennett, J., Xie, C., Zhang, Z. and Liang, D. (2007), "Estimating Non-Market Environmental Benefits of the Conversion of Cropland to Forest and Grassland Program: A Choice Modelling Approach", *Ecological Economics* **63**, 114-125.