

ABSTRACT

Title of dissertation: IMPLICATIONS OF HETEROGENEITY
 IN DISCRETE CHOICE ANALYSIS

Adán L. Martínez-Cruz,
Doctor of Philosophy, 2013

Dissertation directed by: Professor Kenneth E. McConnell
 Department of Agricultural
 and Resource Economics

This dissertation carries out a series of Monte Carlo simulations seeking the implications for welfare estimates from three research practices commonly implemented in empirical applications of mixed logit and latent class logit.

Chapter 3 compares welfare measures across conditional logit, mixed logit, and latent class logit. The practice of comparing welfare estimates is widely used in the field. However, this chapter shows comparisons of welfare estimates seem unable to provide reliable information about the differences in welfare estimates that result from controlling for unobserved heterogeneity. The reason is that estimates from mixed logit and latent class logit are inherently inefficient and inaccurate.

Researchers tend to use their own judgement to select the number of classes of a latent class logit. Chapter 4 studies the reliability of welfare estimates obtained under two scenarios for which an empirical researcher using his/her judgement would arguably choose less classes than the true number of classes. Results show that models with a number of classes smaller than the true number tend to yield down-

ward biased and inaccurate estimates. The latent class logit with the true number of classes always yield unbiased estimates but their accuracy may be worse than models with the smaller number of classes.

Studies implementing discrete choice experiments commonly obtain estimates of preference parameters from latent class logit models. This practice, however, implies a mismatch: discrete choice experiments are designed under the assumption of homogeneity in preferences, and latent class logit search for heterogeneous preferences. Chapter 5 studies whether welfare estimates are robust to this mismatch. This chapter checks whether the number of choice tasks impact the reliability of welfare estimates. The findings show welfare estimates are unbiased regardless the number of choice tasks, and their accuracy increases with the number of choice tasks. However, some of the welfare estimates are inefficient to the point that cannot be statistically distinguished from zero, regardless the number of choice tasks.

Implications from these findings for the empirical literature are discussed.

IMPLICATIONS OF HETEROGENEITY
IN DISCRETE CHOICE ANALYSIS

by

Adán L. Martínez-Cruz

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:

Professor Kenneth E. McConnell, Chair/Advisor

Professor Anna Alberini

Professor Erik Lichtenberg

Professor Charles Towe

Professor Maureen L. Cropper, Dean's Representative

© Copyright by
Adán L. Martínez-Cruz
2013

Dedication

*To Liz,
my perfect plan.*

Acknowledgments

Ted, thank you. For reading through so many bad written drafts. For your patience, your kindness, your support.

Anna, thank you for your interest in this dissertation. For your comments, and your support.

Erik, thank you for your help in my attempts to study deforestation, investment and land rights.

I thank Charles Towe and Maureen Cropper for being part of my advisory committee, and for their comments to this dissertation.

I thank Katherine, Jeff and Pin Wan Lin. You make life at our Department much more easier and enjoyable.

I thank my friends. I am so lucky that I can count you guys on the dozens (the list does not reflect any personal preference, ok?): Gabo and Maga, Kabir and Pini, Will and Rebecca, Jaime and Jazmin, Denny and Kote, Tim and Liz, Vicente, Upu, Benjamin, Geret and Jill, John, Jeff, Seth, Lucija and Stephan, Magda and George, Amparo and Michael, Fabian and Adriana, Kohei, Paulina and John, Fanqing, Carlos and Carolina, Taka and Clara, Aoife, Beat, Ariel, Anand, Joaquin, Cristobal, Rosita and Pedro, Simone and Priscila, Pancho and Kiyomi, Yoanna, Rubaba, Danielle, Beatriz, Ana Laura, Sara, Tania and Ruti, Andrea and Corita, Luisa and Ivonn. A full list of reasons to thank you is a lengthy list.

I thank the people who adopted me and Liz as part of their family: Denny and Kote, Fabian and Adriana, and Gabo.

I thank my adopted sisters: Amparo, Magda, Pini.

I really have to thank Denny for disc-golf.

Lucija, thank you for English 101.

I thank all the support and help from my in-laws: Deb, Corky, Linda, Gordon, Talia, Katherine, and Ne.

Gracias a Chela, mi madre. Tú plantaste la semilla, jefa. Ahora sólo resta que coseches. Gracias a Lalo, mi padre. Por obligarme a recoger los documentos de la preparatoria.

Thank you Liz. For the songs, the road trips, the IPA, the Jazz. Thank you so much for the Gifford scholarship.

CONTENTS

1. <i>Introduction</i>	1
1.1 What can we learn from comparing welfare estimates across econometric specifications?	6
1.2 Implications from using researchers' own judgement in selecting the number of classes in latent class logit models	7
1.3 Welfare implications from a mismatch: inference of heterogeneous preferences from experiments designed under the assumption of ho- mogeneity	7
2. <i>Welfare measures from discrete choice models</i>	9
2.1 Welfare measures in the RUM	10
2.2 Welfare measures in presence of unobserved preference heterogeneity .	14
2.3 True versus estimated welfare measures	17
3. <i>Welfare estimates from conditional logit, mixed logit, and latent class logit</i>	19
3.1 Introduction	19
3.2 Relative magnitude of welfare estimates in empirical applications . . .	22
3.2.1 Studies under review	23
3.2.2 Relative magnitude of welfare estimates	26
3.2.3 Meta-analysis on relative magnitude of welfare estimates . . .	30
3.3 Simulation strategy	34
3.3.1 Utility-generating processes	41
3.4 Results	46
3.5 Conclusions and discussion	52
4. <i>Welfare implications from misspecification of latent class logit models</i> . . .	57
4.1 Introduction	57
4.2 Strategies to select latent classes	59
4.3 Simulation strategy	63
4.4 Results	70
4.5 Conclusions and discussion	79
5. <i>Homogeneous discrete choice experiments and heterogeneous logit models: Implications for welfare estimates</i>	81
5.1 Introduction	81
5.2 Current practices in discrete choice experiments	84
5.2.1 Background	84

5.2.2	Current practices	88
5.2.3	The issue	92
5.3	Simulation strategy	93
5.3.1	Discrete choice experiment	96
5.3.2	True and estimated WTP measures	97
5.4	Results	99
5.5	Conclusions and discussion	105
6.	<i>General conclusions</i>	109
	 <i>Appendix</i>	 113
A.	<i>Description of empirical applications reporting welfare estimates from conditional logit, mixed logit and latent class logit</i>	114
B.	<i>Densities of estimated WTP measures by preference scenario</i>	119
C.	<i>Description of studies using researcher’s own judgement in selecting number of classes</i>	127
D.	<i>95% confidence intervals of estimated WTP measures for the one small class scenario</i>	132

LIST OF FIGURES

3.1	Steps of Monte Carlo simulation studying reliability of welfare estimates from conditional logit, mixed logit, and latent class logit	37
3.2	Snapshot on WTP to avoid the loss of alternative 2 by econometric method (discrete, correlated preferences scenario)	47
4.1	Steps of Monte Carlo simulation studying reliability of estimates welfare measures to number of classes	64
4.2	Snapshot of WTP for 25% improvement in Q of alternative 1 by econometric method (close-to-zero price parameter scenario)	71
4.3	Snapshot of WTP for marginal improvement in Q by econometric method (close-to-zero price parameter scenario)	73
4.4	Snapshot of WTP to avoid loss of alternative 2 by econometric method (close-to-zero price parameter scenario)	75
5.1	Steps of Monte Carlo simulation studying robustness of welfare estimates to number of choice tasks	95
5.2	95% confidence interval of WTP for 25% improvement in T of alternative 1 by number of choice tasks)	100
5.3	95% confidence interval of WTP for marginal change in T by number of choice tasks)	101
5.4	95% confidence interval of WTP to avoid the loss of alternative 2 by number of choice tasks)	102
B.1	Snapshot on WTP for 25% increase in Q of alternative 1 by econometric method (discrete, correlated preferences scenario)	119
B.2	Snapshot on WTP for a marginal change in Q by econometric method (discrete, correlated preferences scenario)	120
B.3	Snapshot on WTP for 25% increase in Q of alternative 1 by econometric method (Normal-normal, correlated preferences scenario) . . .	121
B.4	Snapshot on WTP to avoid the loss of alternative 2 by econometric method (Normal-normal, correlated preferences scenario)	122
B.5	Snapshot on WTP for a marginal change in Q by econometric method (Normal-normal, correlated preferences scenario)	123
B.6	Snapshot on WTP for 25% increase in Q of alternative 1 by econometric method (Normal-normal, uncorrelated preferences scenario) . .	124
B.7	Snapshot on WTP to avoid the loss of alternative 2 by econometric method (Normal-normal, uncorrelated preferences scenario)	125

B.8	Snapshot on WTP for a marginal change in Q by econometric method (Normal-normal, uncorrelated preferences scenario)	126
D.1	Snapshot of WTP for 25% improvement in Q of alternative 1 by econometric method (one small class scenario)	132
D.2	Snapshot of WTP for marginal improvement in Q by econometric method (one small class scenario)	133
D.3	Snapshot of WTP to avoid loss of alternative 2 by econometric method (one small class scenario)	134

1. INTRODUCTION

This dissertation carries out a series of Monte Carlo simulations designed to learn the implications for welfare estimates from research practices implemented in studies that incorporate unobserved preference heterogeneity in discrete choice models.

Discrete choice analysis refers to the study of how economic agents choose among a set of exclusive alternatives. Discrete choice models are widely used among economists, sociologists, psychologists, political scientists, and policy analysts (Heinrich and Wenger, 2002). This popularity is due to the contributions made by Daniel McFadden during the 1960's and 1970's. McFadden formulated a statistical model for discrete choice analysis, called conditional logit (McFadden, 1975, 1976), and provided a direct connection to consumer theory, linking unobserved preference heterogeneity to a description of the distribution of demands (McFadden, 1974).

Unobserved preference heterogeneity refers to the component of an individual's utility that is unobservable from the researcher's perspective. This unobservable component may or may not be correlated with the observable portion of an individual's utility. Uncorrelated unobserved heterogeneity is not of major concern for researchers because the conditional logit incorporates this heterogeneity in the form of a generalized extreme value error term. Unobserved correlated heterogeneity in

preferences, however, is a major concern to researchers. Unobserved preference heterogeneity that is correlated with the observed utility arises, for instance, when the unobserved utility includes an individual-specific term that interacts with variables captured in the observed utility. For example, a common concern in non-market valuation studies is that attributes enjoyed by individuals visiting parks systematically vary depending on whether individuals are scenic-lovers or not. This case illustrates the presence of unobserved correlated heterogeneity because researchers can not observe whether an individual is a scenic-lover or not, and suspect the preference parameters systematically vary depending on whether the individual is a scenic-lover or not.

The rest of this dissertation refers to unobserved correlated heterogeneity in preferences simply as unobserved preference heterogeneity or unobserved heterogeneity, unless otherwise is stated. This convention obviates the reference to the correlation between unobserved and observed utilities because the research practices under study here are concerned with correlated unobserved heterogeneity, and not with uncorrelated unobserved heterogeneity.

The term mixed logit encompasses a wide variety of statistical models incorporating unobserved preference heterogeneity in discrete choice analysis. The distinguishing feature of a mixed logit is that preference parameters are modeled as random variables, i.e. described by a statistical distribution. Preference parameters may or may not be assumed jointly distributed, with the corresponding implications for correlation patterns. Preference parameters may be modeled as continuously distributed or as discretely distributed. By convention, the term mixed logit is reserved

for models describing continuously distributed parameters, and the term latent class logit is used for models describing discretely distributed parameters. This convention is followed hereafter.

The incorporation of both mixed logit and latent class logit in the toolbox of the economic research is direct consequence of the advances in computer speed and simulation methods. Currently, the most popular and promising strategies to incorporate unobserved preference heterogeneity are both the mixed logit and the latent class logit.

Theoretically, any random utility function can be approximated to any degree of accuracy by a mixed logit specification, including the possibility that the best approximation is a latent class logit specification (McFadden and Train, 2000). This result holds when the transformations of observed variables and the random distributions are sufficiently flexible (McFadden, 2001). This result is, however, a blessing and a curse. A blessing because, once simulation methods have been incorporated in the maximization of the likelihood functions, researchers have optimistically engaged in the search of mixed logit specifications that flexibly characterize unobserved heterogeneity. A curse because of the lack of theoretical guidance. McFadden and Train's theorem is an existence proof, and does not provide guidance for finding the distributions attaining an arbitrarily close approximation to the true utility-generating process (Train, 2008).

Researchers have developed a series of recommended practices for practitioners engaged in the estimation of mixed logits. These practices have been under permanent revision, as illustrated by the periodicity of the papers describing the state of

the art in discrete choice analysis: Nevo (2000), Hensher and Greene (2003), Ortúzar (2006), and Cherchi (2009). However, despite major advances in model specification made during the last decade, there is currently no guarantee a specific mixed logit or latent class logit can effectively be estimated. Researchers do not know whether the estimated structure is correctly reproducing the true utility-generating process (Cherchi, 2009). Thus the reliability of both mixed and latent class logits is an issue under study.

In the spirit to contribute to a better understanding of the strengths and limitations of the mixed logit and latent class logit, this dissertation studies the implications from research practices that are common in empirical applications of mixed logit and latent class logit.

The research strategy in this dissertation relies on Monte Carlo simulations. These simulations use pseudo-datasets. A pseudo-dataset is created according to a utility-generating process completely known by the analyst. The use of simulated data allows the analyst to isolate possible confounding effects. Confounding effects are present when an econometric specification cannot unequivocally identify a true data-generating process (Cherchi, 2009). Alternatively, the presence of confounding effects implies that, given the available data, no econometric specification can control for all relevant factors explaining the true data-generating process. Controlling for possible confounding effects is pertinent in this dissertation because previous research has shown presence of unobserved heterogeneity may be confused with heteroskedasticity (see Cherchi, 2009). By carrying out Monte Carlo simulations on pseudo-datasets for which homoskedasticity is imposed, the experiments in this

dissertation solely study the implications from varying structures of unobserved heterogeneity.

Ultimately, an incorrect incorporation of unobserved heterogeneity impacts the reliability of welfare estimates, and therefore the reliability of public policy recommendations. Thus this dissertation seeks the implications from research practices on welfare estimates. Three welfare measures are of interest: (i) willingness to pay for a marginal change in an attribute; (ii) willingness to pay for a non-marginal change in an alternative's attribute; and (iii) willingness to pay to avoid the loss of an alternative. These measures are at the core of cost-benefit analysis exercises that inform public policy recommendations.

This dissertation seeks the implications for welfare estimates from three research practices: (i) the comparison of welfare estimates from a conditional logit versus welfare estimates from a mixed logit or a latent class logit; (ii) the use of researcher's own judgement when selecting the number of classes of a latent class logit specification; and (iii) the estimation of latent class logit specifications on data gathered through discrete choice experiments that rely on the assumption of homogeneity in preferences. These research practices are studied in chapters 3, 4, and 5 respectively. Because the estimation of welfare measures is a common feature across chapters, chapter 2 describes the strategy to estimate welfare measures when estimates of preference parameters are obtained through a mixed logit or a latent class logit. The rest of this chapter presents a summary of the empirical chapters in this dissertation.

1.1 What can we learn from comparing welfare estimates across econometric specifications?

Comparison of welfare estimates between conditional logit and mixed or latent class logit is a common practice in applied research. Researchers compare welfare estimates under the assumption that the better statistical fit provided by mixed logit and latent class logit generates more accurate welfare estimates. Thus significant differences in the welfare estimates are expected.

Chapter 3 provides a literature review showing an unexpected empirical regularity: estimates from conditional logit tend to be statistically indistinguishable from estimates obtained through mixed logit and latent class logit. Then a meta-analysis strongly suggests features of the econometric specifications do not explain variations in the relative magnitudes of welfare estimates. Confidence intervals of estimates from mixed logit and latent class logit are usually large, and may be the main reason behind the similarity in welfare estimates. Actually, the Monte Carlo simulations developed in chapter 3 strongly suggest this is the case. The results show that conditional logit yields biased welfare estimates with relatively small confidence intervals, and mixed logit and latent class logit yield unbiased welfare estimates with relatively large confidence intervals, specially mixed logit. As in the empirical literature, the point estimates from conditional logit, mixed logit and latent class logit specifications tend to be close in absolute value. Thus these findings support the notion that large confidence intervals are an inherent feature of the welfare estimates obtained from mixed logit and latent class logit.

1.2 Implications from using researchers' own judgement in selecting the number of classes in latent class logit models

There are no standard strategies to select the number of classes in empirical applications of latent class logit models. Current practices in applied research include the use of the researchers' own judgement when likelihood-based criteria provide conflicting evidence about the number of classes. The prominence of this practice is illustrated by the 40% of applications that rely only in the researcher's own judgement to select the number of classes (see section 4.2).

Chapter 4 raises the question of whether the strategies to implement a researcher's own judgement ultimately impact the reliability of welfare estimates. Accordingly, the Monte Carlo simulations in chapter 4 are designed to learn whether welfare estimates from latent class logit models are robust to the number of classes in the estimated model. Results show that the reliability of welfare estimates crucially depends on the estimated number of classes: latent class logit specifications yield biased welfare estimates when estimated with a number of classes different from the true number of classes.

1.3 Welfare implications from a mismatch: inference of heterogeneous preferences from experiments designed under the assumption of homogeneity

Current practices applied research include the estimation of latent class logit models on data gathered through a discrete choice experiment. This practice, however, relies on a mismatch of assumptions about preferences: discrete choice experiments are designed under the assumption of homogeneity in preferences, and latent

class logit is carried out to infer heterogeneous preferences.

Chapter 5 looks at the reliability of welfare estimates when homogeneous discrete choice experiments and latent class logit are combined in a study. Chapter 5 first identifies the most common discrete choice experiment implemented in empirical applications of latent class logit — an orthogonal fractional-factorial design that identifies only main effects. Accordingly, the Monte Carlo simulations carried out in chapter 5 study the reliability of the welfares estimates obtained from using in the same study a latent class logit and an orthogonal fractional-factorial design that identifies only main effects. The findings are straightforward: welfare estimates are unbiased regardless the number of choice tasks, and their accuracy increases with the number of classes.

2. WELFARE MEASURES FROM DISCRETE CHOICE MODELS

This chapter describes how welfare measures are estimated when unobserved preference heterogeneity is incorporated in a random utility maximization model (RUM). The description in this chapter heavily borrows from McFadden (1995).¹ The description starts with welfare expressions in the simplest random utility model. Then the corresponding additions are made to incorporate unobserved heterogeneity. Both continuous and discrete unobserved heterogeneities are motivated by an error components approach. Discrete unobserved heterogeneity is usually motivated by a random parameters interpretation. However, the motivation from an error components approach is useful for the comparisons carried out in chapter 3 of this dissertation.

The expressions to calculate welfare measures presented in this chapter are useful in chapters 3, 4, and 5 because these chapters carry out comparisons of welfare estimates against true welfare measures. Thus this chapter ends with an explanation of how true and estimated welfare measures are calculated through this dissertation.

¹ Expressions to estimate welfare measures from discrete choice models were first provided by Hanemann (1982), and Small and Rosen (1982). Then McFadden (1995) provided a generalization to the case in which the error term is distributed according to a generalized extreme value distribution, and McConnell (1995) showed that the same expressions for welfare measures can also be derived from a discrete choice model that does not rely on the utility theory.

2.1 Welfare measures in the RUM

The random utility maximization model (RUM) assumes individual i chooses among J mutually exclusive alternatives. An individual's indirect utility from each alternative is denoted as U_{ij} for $i = 1, 2, \dots, I$ and $j = 1, 2, \dots, J$. The individual is assumed to know his own utilities with certainty. The researcher, however, cannot fully observe each U_{ij} . Assuming a linear indirect utility function, U_{ij} can be expressed as

$$\begin{aligned} U_{ij} &= V_{ij} + \epsilon_{ij} \\ V_{ij} &= \mathbf{x}_{ij}^\top \beta \end{aligned} \tag{2.1}$$

where V_{ij} is the component of utility observed by the researcher; \mathbf{x}_{ij} is a $(M + 1) \times 1$ column vector denoting M alternative-specific attributes² and one alternative-specific dichotomous variable; β is a $(M + 1) \times 1$ column vector denoting one alternative-specific intercept and marginal utilities from the M attributes; and ϵ_{ij} captures the purely random heterogeneity, unobserved by the researcher.

Individuals are assumed to choose the alternative associated with the highest utility. That is, individual i chooses U_i^{max} , where

$$U_i^{max} \equiv \max\{U_{i1}, U_{i2}, \dots, U_{iJ}\} \tag{2.2}$$

However, due to the presence of ϵ_{ij} , the researcher does not observe U_i^{max} and

² Alternative-specific attributes may include measures of alternative attributes, and individual-specific characteristics interacted with alternative-specific attributes.

can make statements only in terms of expected maximum utilities. Expectations are calculated over the error term ϵ_{ij} , i.e.

$$E_{\epsilon}(U_i^{max}) = E_{\epsilon}[\max\{U_{i1}, U_{i2}, \dots, U_{iJ}\}] \quad (2.3)$$

Under the assumption that ϵ_{ij} is distributed according to a type I extreme value distribution, the expected maximum utility is calculated with the logsum formula:³

$$E_{\epsilon}(U_i^{max}) = \ln \sum_{j=1}^J \exp(V_{ij}) \quad (2.4)$$

Accordingly, when the researcher is interested in estimating welfare measures, he/she can only make statements in terms of expected welfare measures. A welfare measure provides the amount a person is willing to pay to avoid a change in an alternative's attribute. Alternatively, a welfare measure is the compensation that a person needs to receive when a change in an alternative occurs so that this person's utility does not change when an alternative is exogenously modified.

To derive an expression to calculate expected welfare measures, assume individual i chooses his/her maximum utility under two scenarios. These scenarios are labeled before (b) and after (a), meaning that individual i maximizes his/her utility before and after an alternative has exogenously been modified. Under the assumptions that ϵ_{ij} is distributed according to a type I extreme value distribution

³ Pioneer derivations of the logsum formula were independently developed by Ben-Akiva (1972), McFadden (1974), and Domencich and McFadden (1975).

and utility is linear in income, the expected value of the compensating variation (CV) from the change in individual i 's utility is expressed as

$$\begin{aligned}
E_{\epsilon}(CV_i) &= \frac{1}{-\beta_c} \left(E_{\epsilon}(U_i^{max,a}) - E_{\epsilon}(U_i^{max,b}) \right) \\
&= \frac{1}{-\beta_c} \left(\ln \sum_{j=1}^J \exp(V_{ij}^a) - \ln \sum_{j=1}^J \exp(V_{ij}^b) \right) \quad (2.5)
\end{aligned}$$

where β_c is the price preference parameter (c stands for travel cost). The term inside parentheses represents the change in expected maximum utility once an alternative has been modified. The division by the negative of the price preference parameter monetizes the change in expected maximum utility because, under the assumption that the utility is linear in income, the marginal utility from price is identical to the negative of the marginal utility from income.

This dissertation is concerned with estimating the expected compensating variation from three events: (i) a marginal change in an attribute; (ii) a non-marginal change in an alternative's attribute; and (iii) the loss of an alternative.

The compensating variation for the loss of an alternative measures the willingness to pay to avoid the loss of an alternative (WTPL). Assume alternative 1 is lost. According to expression (2.5), the expected value of WTPL is calculated as

$$\begin{aligned}
E_{\epsilon}(WTPL_i) &= \frac{1}{-\beta_c} \left(E_{\epsilon}(U_i^{max,a}) - E_{\epsilon}(U_i^{max,b}) \right) \\
&= \frac{1}{-\beta_c} \left(\ln \sum_{j=2}^J \exp(V_{ij}) - \ln \sum_{j=1}^J \exp(V_{ij}) \right) \\
&= \frac{1}{-\beta_c} \ln \left(\frac{\sum_{j=2}^J \exp(V_{ij})}{\sum_{j=1}^J \exp(V_{ij})} + \frac{\exp(V_{i1})}{\sum_{j=1}^J \exp(V_{ij})} - \frac{\exp(V_{i1})}{\sum_{j=1}^J \exp(V_{ij})} \right)
\end{aligned}$$

$$= \frac{1}{-\beta_c} \ln(1 - Pr[U_i^{max} = U_{i1}]) \quad (2.6)$$

where $Pr[U_i^{max} = U_{i1}] = \exp(V_{i1} / \sum_j \exp(V_{ij}))$ is the probability that individual i chooses the alternative 1.

The marginal willingness to pay for a marginal change in an attribute (MWTP) is also derived from equation (2.5). Assume an attribute changes in a non-marginal fashion across all alternatives. Denote this attribute by q , and $q^a = q^b + \Delta q$ is the level of q after Δq has been added to q^b , where b stands for before. To calculate the corresponding compensating variation, q^a is introduced in equation (2.5). The change in q can be factored because it occurs across all alternatives.⁴ Thus the expected compensating variation takes the following form:

$$E_\epsilon(CV[\Delta q]_i) = -\Delta q \frac{\beta_q}{\beta_c} \quad (2.7)$$

where β_q is the marginal utility from q . Expression (2.7) reduces to the willingness to pay for a marginal change across alternatives when $\Delta q = 1$, i.e. when the change in q is marginal,

$$E_\epsilon(MWTP_i) = -\frac{\beta_q}{\beta_c} \quad (2.8)$$

Expression 2.8 can be interpreted as the ratio of the marginal utility from the attribute that changes and the negative of the marginal utility from income.

Equation (2.5) does not reduce to an easy-to-calculate expression for the case

⁴ Further details can be found in Haab and McConnell (2002).

of the willingness to pay for a non-marginal change in an alternative's attribute (WTPA). In contrast to the derivation of expression (2.8), the change in the attribute cannot be factored because the attribute changes only for one alternative. Then for the case of WTPA, equation (2.5) is simply re-expressed as

$$E_{\epsilon}(WTPA_i) = \frac{1}{-\beta_c} \left(\ln \sum_{j=1}^J \exp(V_{ij}^a) - \ln \sum_{j=1}^J \exp(V_{ij}^b) \right) \quad (2.9)$$

So far, the RUM assumes the unobserved component of the indirect utility, ϵ_{ij} , captures purely random behavior. Arguably, the unobserved component of utility may include unobserved preference heterogeneity that induces correlation between observed and unobserved components of the utility function. The simplest RUM requires modifications to account for this type of unobserved preference heterogeneity.

2.2 Welfare measures in presence of unobserved preference heterogeneity

Changes in the assumptions about the unobserved component of the utility function are introduced when the researcher suspects unobserved preference heterogeneity induces correlation between observed and unobserved components of the utility function. A RUM incorporating unobserved preference heterogeneity may equivalently be motivated by either a random parameters representation or an error components representation. This section motivates the incorporation of unobserved preference heterogeneity in a RUM by means of the error components representation

because this approach place the emphasis on the correlation between observed and unobserved components of the utility function (Train, 2003). By doing so, the error components representation provides insights that will prove useful when designing Monte Carlo experiments in chapter 3.

The error components representation of a RUM assumes the presence of omitted attributes systematically impacting the utility function. That is,

$$\begin{aligned}
 U_{ij} &= V_{ij} + \eta_{ij} \\
 V_{ij} &= \mathbf{x}_{ij}^\top \beta \\
 \eta_{ij} &= \mathbf{o}_{ij}^\top \zeta_i + \epsilon_{ij}
 \end{aligned} \tag{2.10}$$

where \mathbf{o}_{ij} is a $L \times 1$ column vector denoting L omitted attributes, and ζ_i is a $L \times 1$ column vector representing individual i 's deviations from the average preference parameters, β ; η_{ij} is the portion of the utility that is unobserved from the researcher's perspective.

Thus the unobserved utility in equation (2.10) comprises two components: effects from omitted attributes, $\mathbf{o}_{ij}^\top \zeta_i$, and the purely random term, ϵ_{ij} . Different correlation patterns between unobserved and observed utilities can be induced through different assumptions about the nature of the omitted attributes, \mathbf{o}_{ij} , and their statistical association with the observed attributes, \mathbf{o}_{ij} . In this study, correlation between the observed utility, V_{ij} , and the unobserved utility, η_{ij} , is induced by assuming the set of omitted attributes is identical to the set of observed attributes, i.e. $\mathbf{o}_{ij} \equiv \mathbf{x}_{ij}$. Under this assumption, η_{ij} remains unobservable because the re-

searcher does not observe the individual i 's deviation from the average preferences, ζ_i . The presence of \mathbf{x}_{ij} in the observed and unobserved utilities implies systematic association between the observed and unobserved utilities.

When calculating welfare measures from the RUM in equation (2.10), the researcher must add one layer of randomness to the calculation of welfare measures. This additional layer is consequence of ζ_i being random. Thus expected compensating variation in equation (2.5) is re-expressed as

$$E_{\zeta}E_{\epsilon}(CV_i) = \int \frac{1}{-\beta_c} \left(\ln \sum_{j=1}^J \exp(V_{ij}^a) - \ln \sum_{j=1}^J \exp(V_{ij}^b) \right) f(\zeta) d\zeta \quad (2.11)$$

where $f(\zeta)$ is the distribution assumed for the deviations from the average preferences. These deviations may be continuously distributed, producing continuous unobserved heterogeneity, or may also be distributed in a discrete fashion, generating discrete unobserved heterogeneity. Equation (2.11) presents the most general formula to calculate welfare measures when unobserved heterogeneity is continuous. Equation (2.11) does not reduce to a closed solution. Therefore, computation of equation (2.11) is carried out by simulation. This simulation involves two steps: (i) taking a draw from $f(\zeta)$, and (ii) evaluating equation (2.11) at the values drawn in step (i). Steps (i) and (ii) are sequentially repeated S times. The simulated expected compensating variation equals the average value of the S computed values (see Train, 2003, for details). Derivation of expressions for WTPL, MWTP, and WTPA from equation 2.11 results in the integration of expressions (2.6), (2.8), and (2.9), respectively.

For the case in which unobserved heterogeneity is discrete, equation (2.11) can be re-expressed as

$$E_{\zeta}E_{\epsilon}(CV_i) = \sum_{g=1}^G \pi_i^g \frac{1}{-\beta_c^g} \left(\ln \sum_{j=1}^J \exp(V_{ij}^{a,g}) - \ln \sum_{j=1}^J \exp(V_{ij}^{b,g}) \right) \quad (2.12)$$

where the price parameter take value β_c^g with probability π_i^g . In this case, the deviations from the average preference parameters vary according to a finite set of values. Derivation of expressions for WTPL, MWTP, and WTPA from equation 2.11 results in finite mixture versions of equations (2.6), (2.8), and (2.9), respectively. The mixture in these equations is done according to probabilities defined by π_i^g and their corresponding preference parameters.

2.3 True versus estimated welfare measures

Chapters 3, 4, and 5 carry out comparisons of true welfare measures against welfare estimates. This section explains how these quantities are calculated.

Welfare calculations vary depending on from whom's perspective we are approaching the estimation. From the perspective of an empirical researcher, U_{ij} is not fully known. Thus an empirical researcher calculates expected welfare measures, carrying out expectations over the unobserved component of the utilities. Expression (2.5) is used to calculate expected welfare measures when the researcher assumes the unobserved utility only captures pure randomness. Expression (2.11) is used to calculate expected welfare measures when the researcher assumes the unobserved utility captures both pure randomness and unobserved heterogeneity that produces

correlation between observed and unobserved utilities.

In this dissertation, true and estimated welfare measures are calculated from the empirical researcher's perspective. True welfare measures are derived from equation (2.5), i.e. true welfare measures from an empirical researcher's perspective is an expectation over pure randomness. In this calculation, true preference parameters are used. This way to estimate true welfare measures assume the empirical researcher estimates preference parameters identical to true preference parameters.

Calculation of estimated welfare measures varies depending on the empirical discrete choice specification. When a conditional logit is estimated, expression (2.5) is used to estimate welfare measures. In contrast to the calculation of true welfare measures, for which expression (2.5) is used as well, estimated preference parameters are used when calculating welfare measures from a conditional logit. Thus true and estimated welfare measures are identical when preference parameters estimated through a conditional logit are identical to true preference parameters.

When a mixed logit or a latent class logit are estimated, expression (2.11) is used to estimate welfare measures. For this calculation, preference parameters estimated through mixed logit or latent class logit are used to estimate welfare measures from a mixed logit or a latent class logit. Expression (2.11) requires the empirical researcher assumes a joint distribution for the individual deviations from the average preferences. Different distributions, both continuous and discrete, are assumed and justified in chapter 3. Chapters 4 and 5 estimate welfare measures assuming discrete distributions because these chapters carry out only latent class specifications.

3. WELFARE ESTIMATES FROM CONDITIONAL LOGIT, MIXED LOGIT, AND LATENT CLASS LOGIT

3.1 Introduction

Comparison of welfare estimates between conditional logit (CL), and mixed logit (ML) or latent class logit (LCL) is a common practice in economics (e.g. Beharry-Borg and Scarpa, 2010; Birol et al., 2006; Kosenius, 2010; Westerberg et al., 2010). Researchers compare welfare estimates under the assumption that the better statistical fit provided by ML and LCL generates more accurate welfare estimates. Thus significant differences in the welfare estimates are expected. However, the literature review presented in section 3.2 strongly suggests the estimates from CL tend to be statistically indistinguishable from estimates obtained through ML and LCL. In addition, a meta-analysis presented in section 3.2 suggests that the relative magnitude of the welfare estimates — measured as a ratio of estimates— is statistically identical to one, regardless of the features of the econometric specification. Confidence intervals of estimates from ML and LCL are usually large, and may be the main reason behind the statistical similarity of welfare estimates.

Consequently, section 3.3 tests whether the confidence intervals of estimates from ML and LCL reported in empirical applications are actually an inherent feature of these estimates. A series of Monte Carlo simulations are designed to compare

willingness to pay (WTP) estimates from CL, ML, and LCL. The WTP measures under study are (i) WTP for a marginal change in an attribute; (ii) WTP for a non-marginal change in an alternative's attribute; and (iii) willingness to pay to avoid the loss of an alternative. Average WTP estimates over the Monte Carlo replications are compared against true WTP. True WTP is calculated with the logsum equation — expression (2.5) in chapter 2. That is, true WTP is calculated from an empirical researcher's perspective, carrying out an expectation over pure randomness. True preference parameters are used in the calculation of true WTP. True preference parameters are available because they are imposed by the analyst.

True indirect utility is assumed linear in two attributes and, implicitly, in income. True utility is simulated under three unobserved heterogeneity scenarios. These scenarios vary the distribution and correlation of the preference parameters: (i) independently normally distributed; (ii) jointly normally distributed; and (iii) jointly discretely distributed.

Preference parameters and welfare measures for the three pseudo-datasets are estimated through CL, ML and LCL specifications. Thus the experimental set up in this study allows for conclusions with respect to performance of (i) the CL in presence of continuous unobserved preference heterogeneity; (ii) the CL in presence of discrete unobserved preference heterogeneity; (iii) the ML in presence of discrete unobserved preference heterogeneity; and (iv) the LCL in presence of continuous unobserved preference heterogeneity.

Performance is evaluated in terms of (i) unbiasedness, i.e. whether the true average value falls within the 95% confidence interval of the average estimates; (ii)

efficiency, i.e. which specification yields the smallest 95% confidence interval; and (iii) accuracy, i.e. how large is the average relative difference between the estimates and the true values according to the absolute value of the mean relative error.

The comparisons in this paper reveal that CL yields biased WTP estimates with relatively small confidence intervals, and ML and LCL yield unbiased WTP estimates with relatively large confidence intervals, specially ML. As in the empirical literature, the point estimates from CL, and ML and LCL specifications tend to be close in absolute value.

These findings support the notion that large confidence intervals are an inherent feature of the welfare estimates obtained from ML and LCL. Implications from these findings for the empirical literature include (i) the comparison of welfare estimates across econometric specifications seems unable to provide reliable information about the differences in welfare estimates resulting from controlling for unobserved heterogeneity; and (ii) the use of ML and LCL represents an overlooked trade-off between the gains in statistical fit and the inefficiency in welfare estimates.

This study contributes to the growing literature seeking a better understanding of the strengths and limitations of the ML and LCL. The literature specifically studying the relative magnitude of welfare estimates has offered mostly results from case studies (see Greene and Hensher, 2003; Hynes et al., 2008; Provencher and Bishop, 2004; Shen, 2009). However, comparisons in case studies may be contaminated by confounding effects such as the documented differences in scale parameter (Cherchi, 2009). Monte Carlo simulations are better equipped to make comparisons that experimentally vary one factor at a time. This study has designed Monte Carlo

simulations varying the structure of the unobserved preference heterogeneity.

A few papers have used Monte Carlo simulations to study the relative magnitude of welfare estimates across ML and LCL, with no consideration of the CL (see Cherchi et al., 2009; Hess et al., 2007; Torres et al., 2011a,b).¹ This study takes a step forward in regards to the experimental design: to the best of my knowledge, this is the first Monte Carlo study comparing CL, LCL and ML that designs an experimental set up in which correlation among preference parameters is identical across discrete and continuous unobserved heterogeneity scenarios. This experimental feature eliminates a possible confounding factor when comparing the LCL with ML because (i) correlation among parameters is an inherent characteristic in the estimation of a LCL (Hess et al., 2011), and (ii) correlation among parameters determines the correlation of unobserved utilities across alternatives (Train, 2003).

3.2 Relative magnitude of welfare estimates in empirical applications

This section first reviews the relative magnitude of welfare estimates reported in the literature. Then a meta-analysis is carried out. This meta-analysis seeks for factors explaining the variation in the relative magnitude of welfare measures.

The applications reviewed in this section refer to discrete choice applications reporting point WTP estimates from (i) CL specifications, and (ii) at least one specification incorporating unobserved preference heterogeneity. A list of closely related applications not covered in this review include (i) applications reporting point esti-

¹ A related literature has used Monte Carlo simulations to study the impact of design of discrete choice experiments on welfare measurement. See Torres et al. (2011b) for references on this literature.

mates from either CL specifications or ML and LCL but not from both (e.g. Breffle et al., 2011; Brouwer et al., 2010; Colombo et al., 2009; Garrod et al., 2012; Ouma et al., 2007; Scarpa and Thiene, 2005; Train and Weeks, 2005); (ii) LCL applications not reporting enough information to infer average WTP measures (e.g. Scarpa et al., 2003); (iii) applications reporting other measures of economic behavior such as elasticities (e.g. Richards, 2000); and (iv) applications graphically comparing WTP estimates across econometric specifications or groups of respondents (e.g. Beharry-Borg and Scarpa, 2010; Domanski and von Haefen, 2012; Hoyos et al., 2009). The last set of applications reports no point estimates that may allow for comparisons of average WTP measures.

3.2.1 *Studies under review*

Twenty studies are reviewed in this section. Table A.1 describes these applications in terms of type of application, objective, and type of elicited preferences. With respect to the type of application, 42% estimate recreational demands; 16% focus on non-market valuation of water quality or atmospheric-nuisance reductions; 21% carry out non-market valuations of wetland ecosystems; and 21% study mode transportation choices. Fourteen studies (70%) compare welfare estimates as a by-product of the main objective; and 30% of the studies are designed to exclusively carry out welfare comparisons across econometric specifications. With respect to the type of elicited preferences, 37% analyze revealed preferences, and 63% focus on stated preferences.²

² Table A.1 also provides information about the surveying method, population under study, and features of the dataset such as number of alternatives, sampled individuals, and number of choice tasks.

Forty econometric specifications incorporating unobserved preference heterogeneity are reported in the 20 studies under review. Table 3.1 describes these econometric specifications in terms of price parameter distribution, whether correlation among parameters is assumed, and the reported welfare measure. Eleven (28%) are LCL specifications, and 29 (72%) are ML specifications. From the ML specifications, 23 assume preference parameters are uncorrelated (MLU), and 6 assume preference parameters are correlated (MLC). From the MLC subset, 2 specifications assume correlation among the full set of parameters. With respect to the price parameter distribution, 9 (23%) assume a finite mixture distribution; 8 (20%), a lognormal distribution; 2 (5%), a normal distribution; and 21 (52%) assume price parameter is fixed. Twenty-two studies (55%) report the WTP for a non-marginal change in an attribute (WTPA); 6 studies (17%) report the WTP to avoid the loss of an alternative (WTPL); and 21 studies (53%) report WTP for a marginal change in an attribute (MWTP). By making inter-column inferences, from columns 1 and 2 table 3.1 we notice that 17 studies (43%) estimate ML specifications that assume both a fixed price parameter and uncorrelated parameters.

Thus a general profile of the specifications under study can be depicted as follows: ML specifications account for around three quarters of the 40 applications, and LCL account for a quarter. Around half of the specifications assume either a fixed price parameter or uncorrelated parameters. Around 40% assume both a fixed price parameter and uncorrelated parameters. Twenty percent of the specifications assume a lognormally distributed price parameter. Normally distributed price parameters are uncommon. Half of the applications report WTP for a marginal change

in an attribute; half report WTP for non-marginal changes in an attribute; and 17% studies report WTP to avoid the loss of an alternative. This last set of percentages do not add to 100 because some studies report two welfare measures.

Tab. 3.1: Description of mixed logit and latent class logit specifications in studies under review

Econometric specification ^{a,b}	Price parameter distribution	Attributes ^c / random parameters ^d	Correlation among parameters	Interactions among attributes	Reported welfare measure ^e
Studies reporting welfare estimates as by-product					
Train (1998)					
MLU	Log-normal	7/6	No	No	WTPA, WTPL
MLC	Log-normal	7/6	Subset	No	WTPA, WTPL
McConnell and Tseng (1999)					
MLU	Fixed	3/1	No	No	WTPA, WTPL
MLU	Fixed	3/3	No	No	WTPA, WTPL
Brefle and Morey (2000)					
MLU	Fixed	10/2	No	Yes	WTPA, WTPL
Boxall and Adamowicz (2002)					
MLU	Fixed	17/17	No	No	WTPA, WTPL
LCL	Finite mixture	17/17	Full set	No	WTPA, WTPL
Provencher et al. (2002)					
MLU	Fixed	11/11	No	Yes	WTPA
LCL	Finite mixture	11/11	Full set	Yes	WTPA
Carlsson et al. (2003)					
MLC	Fixed	11/6	Subset	No	MWTP
Nahuelhual et al. (2004)					
MLU	Fixed	2/2	No	No	MWTP
MLU	Fixed	5/2	No	Yes	MWTP
Sillano and Ortúzar (2005)					
MLU	Fixed	4/3	No	No	MWTP
MLU	Normal	4/3	No	No	MWTP
MLU	Log-normal	4/3	No	No	MWTP
Birol et al. (2006)					
MLU	Fixed	5/5	No	No	WTPA, MWTP
MLU	Fixed	29/5	No	No	WTPA, MWTP
LCL	Finite mixture	5/5	Full set	No	WTPA, MWTP
Hanley et al. (2006)					
MLU	Fixed	5/3	No	No	MWTP
Milon and Scrogin (2006)					
LCL	Finite mixture	7/7	Full set	No	WTPA
Scarpa et al. (2008)					
MLU	Log-normal	6/6	No	No	MWTP
MLC	Log-normal	6/6	Full set	No	MWTP
MLU ^f	Log-normal	6/6	No	No	MWTP
MLC ^f	Log-normal	6/6	Full set	No	MWTP
Kosenius (2010)					
MLU	Fixed	10/5	No	Yes	WTPA

Continued on next page

Table 3.1 – *Continued from previous page*

Econometric specification ^{a,b}	Price parameter distribution	Attributes ^c / random parameters ^d	Correlation among parameters	Interactions among attributes	Reported welfare measure ^e
LCL	Finite mixture	5/5	Full set	No	WTPA
Westerberg et al. (2010)					
MLU	Fixed	19/5	No	Yes	MWTP
Studies primarily comparing welfare estimates					
Greene and Hensher (2003)					
MLU	Fixed	7/3	No	No	MWTP
LCL	Finite mixture	7/7	Full set	No	MWTP
Provencher and Bishop (2004)					
MLU	Fixed	10/10	No	Yes	WTPA
MLC	Fixed	7/7	Subset	Yes	WTPA
LCL	Fixed	10/10	Subset	Yes	WTPA
Hess et al. (2007)					
MLU ^f	Normal	1/1	No	No	MWTP
LCL ^f	Finite mixture	1/1	Full set	No	MWTP
Hynes et al. (2008)					
MLC	Log-normal	19/6	Subset	Yes	WTPA
LCL	Finite mixture	11/11	Full set	No	WTPA
Cherchi et al. (2009)					
MLU	Fixed	6/1	No	No	WTPA
LCL	Fixed	6/1	No	No	WTPA
MLU	Fixed	6/3	No	No	MWTP
LCL	Finite mixture	6/6	Full set	No	MWTP

^a MLC: mixed logit with correlated parameters; MLU: mixed logit with uncorrelated parameters; LCL: latent class logit.
^b All econometric specifications are estimated in preference space with the exception of those reported in bold. ^cExcluding price. ^dExcluding price parameter.
^eWTPA: WTP for a non-marginal change in an attribute; WTPL: WTP to avoid loss of an alternative; and MWTP: willingness to pay for a marginal change in an attribute.
^f Specification estimated in the willingness to pay space.

3.2.2 Relative magnitude of welfare estimates

The 20 studies under review report 204 pairs of average welfare estimates. A pair of average welfare estimates includes (i) a welfare estimate obtained through a CL, and (ii) a welfare estimate obtained through either a ML or a LCL. The welfare estimates of each pair are calculated for the same sample.

A measure of relative magnitude is the ratio of welfare estimates. That is

$$R_t = W_{cl}/W_t \tag{3.1}$$

where W refers to a welfare estimate; the subscript cl stands for conditional logit; R refers to the ratio of welfare estimates; and the subscript $t = mlu, mlc, lcl$, where mlu , mlc , and lcl refer to mixed logit with uncorrelated parameters (MLU), mixed logit with correlated parameters (MLC), and latent class logit (LCL), respectively. Thus R_{mlu} is the relative magnitude of a welfare estimate from a CL with respect to the paired welfare estimate from a MLU; R_{mlc} is the relative magnitude of a welfare estimate from a CL with respect to the paired welfare estimate from a MLC; and R_{lcl} is the relative magnitude of a welfare estimate from a CL with respect to the paired welfare estimate from a LCL.

Table 3.2 presents descriptive statistics of R_{mlc} , R_{mlu} , and R_{lcl} . The top panel of table 3.2 summarizes all three ratios together. From a total of 204 ratios, 79 (39%) are ratios with respect to WTP for a non-marginal change in an attribute (WTPA); 11 (5%) are with respect to WTP to avoid the loss of an alternative (WTPL); and 114 (56%) are with respect to WTP for a marginal change in an attribute (MWTP). The average ratios are 0.95, 0.96 and 0.90 for WTPA, WTPL and MWTP, respectively. The medians are 0.87, 1.04, and 0.83 for WTPA, WTPL and MWTP, respectively. The standard deviations are large, particularly for the case of WTPA (1.38). The means and medians are consistently around one, and the standard deviations are large as well when splitting the ratios by econometric

specification (second to fourth panels in table 3.2).

Tab. 3.2: Descriptive statistics of welfare ratios reported in studies under review (n=204)

WTP						
measure ^a	n	mean	median	std dev	min	max
All ratios (R_{mlc} , R_{mlu} , R_{lcl}) ^b						
WTPA	79	0.95	0.87	1.38	-7.96	5.25
WTPL	11	0.96	1.04	0.13	0.72	1.09
MWTP	114	0.91	0.83	0.68	-0.47	6.77
Ratio with respect to mixed logit with correlated parameters (R_{mlc}) ^b						
WTPA	13	1.18	0.58	1.35	0.18	4.70
WTPL	2	1.06	1.06	0.01	1.05	1.07
MWTP	31	0.94	0.69	1.14	0.31	6.77
Ratio with respect to mixed logit with uncorrelated parameters (R_{mlu}) ^b						
WTPA	33	1.00	0.87	0.45	0.37	2.00
WTPL	8	0.92	0.98	0.14	0.72	1.08
MWTP	68	0.88	0.89	0.38	-0.47	2.69
Ratio with respect to latent class logit (R_{lcl}) ^b						
WTPA	33	0.82	0.89	1.93	-7.96	5.25
WTPL	1	1.09	1.09	–	1.09	1.09
MWTP	15	0.95	0.79	0.5	0.58	2.21

^a WTPA: WTP for a change in an alternative's attribute;
WTPL: WTP to avoid the loss of an alternative; MWTP:
WTP for an attribute marginal change across alternatives.
^b $R_{mlu} = W_{cl}/W_{mlu}$, $R_{mlc} = W_{cl}/W_{mlc}$, and
 $R_{lcl} = W_{cl}/W_{lcl}$, where W stands for welfare estimate,
and cl, mlc, mlu and lcl stand, respectively for conditional
logit, mixed logit with correlated parameters, mixed logit
with uncorrelated parameters and latent class logit.

Basic descriptive statistics suggest that, regardless whether ML allows for correlation across parameters or not, WTP estimates from CL are similar in magnitude to estimates from both LCL and ML.

Summaries in table 3.2 refers to welfare measures obtained by two types of studies: (i) studies reporting welfare measures as a by-product of their declared main objectives, and (ii) studies declaring their main goal is comparing welfare estimates across econometric specifications. Arguably, the relative magnitude of

welfare measures may differ depending on whether a paper’s declared main goal is the comparison of welfare measures or not. A possible reason for this difference is that studies primarily focused on comparing welfare measures may pay more attention to methodological aspects. For instance, assuring that mixed logit and latent class logit specifications are comparable by including interactions in the mixed logit specification (e.g. Hynes et al., 2008). To explore the possibility of a systematic difference in the relative magnitude of welfare measures, table 3.3 reports the relative magnitude of welfare measures from studies primarily comparing welfare measures across econometric specifications.

The five studies included in 3.3 declare their main goal is the comparison of welfare measures across econometric specification. These studies report 51 pairs of welfare estimates. Table 3.3 presents the point welfare estimate from CL, and the point welfare estimate from the specifications incorporating unobserved heterogeneity and its standard errors or 95% confidence interval. The point estimates from CL fall in the 95% confidence interval of the heterogeneous specification in all pair comparisons but two reported by Hynes et al. (2008).³ Thus empirical studies seeking for differences in welfare estimates have mostly found no differences. Despite this similarity, a tendency can be observed in table 3.3: the point estimates from CL are smaller than estimates from heterogeneous specifications in all but two pairs of welfare measures. The large standard errors of estimates from heterogeneous specifications explain the systematic no rejection of the null hypothesis of equality in

³ Welfare estimates are correlated because they are extracted from the same dataset. Thus a more accurate comparison strategy should take correlation into account. However, methods such as the convolution approach suggested by Poe et al. (2005) require access to the original dataset in order to implement re-sampling techniques.

welfare measures.

In summary, basic descriptive statistics on the relative magnitude of welfare estimates suggest an empirical regularity: while point welfare estimates from CL tend to be smaller than point welfare estimates from ML and LCL, statistical differences are seldom found. Estimates obtained through ML and LCL usually have large confidence intervals. These confidence intervals seem to be responsible of the statistical similarity of welfare estimates across econometric specifications.

3.2.3 *Meta-analysis on relative magnitude of welfare estimates*

Basic descriptive statistics may hide the possibility that, after controlling for features explaining the variation in the relative magnitude of welfare measures, the relative magnitude is actually different from one. This section tests for this possibility. That is, this section reports the results of a meta-analysis seeking to explain the variation of the relative magnitude of welfare measures in terms of features of the empirical applications under review.

This section uses ordinal least squares (OLS) regressions to model the relative magnitude of welfare measures, R_{st} , in terms of the features of the empirical applications, z_{st} , i.e.

$$R_{st} = \gamma z_{st} + \nu_{st} \tag{3.2}$$

where R is defined by expression (3.1); subscripts s and t refer, respectively, to study and econometric specification used to estimate the welfare measure; z_{st} are the features of the study s and the econometric specification t ; and ν_{st} stands for a

normally distributed error term with mean zero and standard deviation σ .

The features in z_{st} include to the type of elicited preferences, the reported welfare measure, the distribution assumed for the price parameter, whether the parameters are correlated in the econometric specification, whether interaction of alternative- and individual specific attributes are included in the econometric specifications, the number of individuals analyzed in each application, the number of alternatives faced by respondents in each application, and fixed effects by study.

Elicited preferences may be either stated or revealed. Stated preferences are inferred from choices made by respondents when faced to hypothetical scenarios. Revealed preferences are inferred from actual choices made by respondents. The reported welfare measures may be the WTP for a marginal change in an attribute (MWTP), the WTP for a non-marginal change in an attribute (WTPA), and the WTP to avoid the loss of an alternative (WTPL). The price parameter may be fixed or distributed as a lognormal, a normal or a finite mixture distribution. The finite mixture distribution is the distribution assumed when a latent class logit model is estimated.

Tab. 3.3: Welfare ratios in papers whose declared main goal is the comparison of welfare measures across econometric specifications

WTP for a marginal change in an attribute (MWTP)				WTP for a non-marginal change in an attribute (WTPA)			
Heterogeneous specification (I) ^a	Estimate from (I)	Std errors or 95% CI ^b	Estimate from CL ^c	Heterogeneous specification (I) ^a	Estimate from (I)	Std errors or 95% CI ^b	Estimate from CL ^c
Greene and Hensher (2003)				Provencher and Bishop (2004)			
MLU	7.36	3.01	2.52	MLU	1037	379	578
MLU	6.06	2.41	2.20	MLU	1467	546	1030
MLU	6.11	2.48	1.74	MLU	1233	447	711
LCL	3.54	2.45	2.52	MLU	1735	686	1221
LCL	3.46	1.69	2.20	MLC	998	316	578
LCL	2.19	1.71	1.74	MLC	2122	728	1030
Hess et al. (2007)				MLC			
MLU	30.41	33.70	19.77	MLC	1332	456	711
LCL	32.81	36.55	19.77	LCL	2821	1103	1221
LCL	34.29	41.86	19.77	LCL	945	285	578
Shen (2009)				LCL			
MLU	2042	–	1684	LCL	1569	472	1030
MLU	704	–	589	LCL	1173	355	711
MLU	302	–	344	LCL	2057	685	1221
LCL	2039	1057	1684	MLU	1564	781	642
LCL	624	163	589	MLU	1980	984	1062
LCL	362	161	344	MLU	1397	705	522
MLU	1929	–	1501	MLU	1861	913	864
MLU	967	–	727	MLC	1180	375	642
MLU	795	–	394	MLC	1576	526	1062
LCL	2211	1473	1501	MLC	1123	368	522
LCL	1023	575	727	MLC	1432	486	864
LCL	679	309	394	LCL	791	280	642
				LCL	1172	414	1062
				LCL	712	272	522
				LCL	981	362	864
				Hynes et al. (2008)			
				MLC	1.89	-2.22 to 9.72	0.34
				LCL	1.19	0.42	0.34
				MLC	0.67	-0.29 to 3.40	3.15
				LCL	0.61	0.22	3.15
				MLC	-0.39	-0.87 to -0.09	-1.36
				LCL	-0.33	0.12	-1.36

^a CL: conditional logit; MLC: mixed logit with correlated parameters; MLU: mixed logit with uncorrelated parameters; LCL: latent class logit. ^b A few studies report empirical confidence intervals instead of standard errors. ^c Estimate from CL falls in the 95% confidence interval of heterogeneous estimate for all cases but those reported in bold font

Table 3.4 presents the results from linear regressions seeking whether variation in welfare ratios depends on the features of the strategy used in empirical applications. Specification I explores whether welfare ratios depend on the type of elicited preferences. Revealed preferences are used as reference category. Specification II explores whether welfare ratios depend on the welfare measure that is reported. WTP for a marginal change is used as reference category. Specification III looks for differences explained by features of the econometric specification. Dichotomous variables are defined to consider three features: (i) price parameter distribution, (ii) correlation among parameters, and (iii) interaction between alternative- and individual-specific factors. Specification IV controls for number of individuals and alternatives. Specification V adds fixed effects by study and excludes number of alternatives. Ideally, specification V should control for both number of alternatives as well. However, evidence of strong collinearity between number of alternatives and type of elicited preferences is observed when adding fixed effects. Collinearity between elicited preferences and number of alternatives become apparent when adding fixed effects because of the lack of variation in the number of alternatives across elicited preferences. For instance, studies analyzing revealed preferences tend to use either only two alternatives or a relatively large number of alternatives (e.g. 59 alternatives in Train, 1998). In contrast, studies analyzing stated preferences use mostly 3 or 4 alternatives. To check whether exclusion of number of alternatives impact the regression results, specification VI includes number of alternatives and excludes elicited preferences while controlling for fixed effects.

Results from linear regressions (I) to (VI) in table 3.4 coincide in the absence

of statistical significance from study characteristics on the relative magnitude of the welfare estimates. The only statistically significant parameter is the intercept, with a value that is not statistically different from one.

Consistent with the absence of significance from study characteristics, the r-squared is smaller than 0.01 for all specifications that do not include fixed effects. When fixed effects are included the r-squared is 0.30. By controlling only for fixed effects, specification VII formally checks the explanatory power of the fixed effects. That is, no features of the study matters in the relative magnitude of the welfare measures. The only variables that affect relative magnitude of welfare measures are study-specific variables that are not associated with the features of the empirical strategy, i.e. fixed effects.

In summary, both the basic descriptive statistics and the meta-analysis presented in this section suggest that (i) welfare estimates from CL are statistically indistinguishable from the welfare estimates from ML and LCL; (ii) the relative magnitude of welfare estimates is not impacted by features of the econometric specifications used on their estimation; and (iii) the average relative magnitude of the welfare estimates is not statistically different from one.

3.3 Simulation strategy

The evidence from section 3.2 suggests that welfare estimates from CL are most frequently statistically indistinguishable from estimates obtained through ML and LCL. The features of the econometric specifications used in the estimation of the

Tab. 3.4: Ordinal least square regressions of welfare ratio on study characteristics (21 studies, 204 pairs of welfare measures)

	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
Intercept	0.89 ^a (0.10)	0.86 ^a (0.13)	0.98 ^a (0.28)	1.01 ^b (0.31)	1.23 (1.26)	1.15 ^a (0.31)	1.14 ^a (0.19)
Stated preferences ^{c,d}	0.07 (0.14)	0.08 (0.14)	0.05 (0.24)	0.05 (0.24)	0.05 (0.74)	— —	— —
Welfare measure ^e ...							
WTPA ^{c,f}	—	0.06 (0.15)	0.10 (0.20)	0.04 (0.21)	-0.01 (0.39)	-0.01 (0.39)	—
WTPL ^{c,f}	—	0.09 (0.32)	0.06 (0.34)	-0.11 (0.40)	-0.33 (0.53)	-0.33 (0.53)	—
Price parameter distribution ^g ...							
lognormal ^c	—	—	-0.20 (0.27)	-0.33 (0.36)	0.09 (0.54)	0.09 (0.54)	—
normal ^c	—	—	-0.09 (0.53)	-0.09 (0.53)	0.14 (0.59)	0.14 (0.59)	—
finite mixture ^c	—	—	-0.32 (0.27)	-0.29 (0.27)	0.14 (0.31)	0.14 (0.31)	—
Correlated parameters ^c	—	—	0.14 (0.19)	0.14 (0.19)	-0.01 (0.20)	-0.01 (0.20)	—
Interaction of alternative- and individual-specific factors ^c	—	—	-0.22 (0.22)	-0.19 (0.24)	-0.15 (0.36)	-0.15 (0.36)	—
Individuals/1000	—	—	—	-0.09 (0.30)	-0.20 (2.45)	-0.20 (2.45)	—
Alternatives/1000	—	—	—	7.98 (10.94)	—	7.91 (10.53)	—
Fixed effects by study	No	No	No	No	Yes	Yes	Yes
N	204	204	204	204	204	204	204
R ²	0.00	0.00	0.01	0.01	0.30	0.30	0.29
Log-likelihood	-288	-288	-287	-287	-252	-252	-253

Standard errors in parentheses. ^a Significant at 99% of confidence.

^b Significant at 95% of confidence.

^c Dichotomous variable: 1 if characteristic is observed.

^d Reference category: revealed preferences.

^e Reference category: WTP for marginal change in an attribute (MWTP).

^f WTPA: WTP for a non-marginal change in an attribute.

WTPL: WTP to avoid the loss of an alternative.

^g Reference category: fixed price parameter.

welfare measures seemingly do not impact the variation of the relative magnitude of welfare measures. Confidence intervals of welfare estimates are usually very large, and may seem the main reason explaining why empirical applications can not reject the null hypothesis that welfare measures across econometric specifications are equal. This possibility is tested in section 3.4 through a series of Monte Carlo simulations. This section describes the design of the Monte Carlo simulation.

A flow chart describing the simulation strategy is presented in figure 3.1. The Monte Carlo simulation has been designed under the following reasoning: (i) an empirical researcher has access to a dataset including both individual- and alternative-specific attributes describing the alternatives from which an individual chooses; (ii) an empirical researcher has means to estimate the preferences an individual has over attributes describing the alternatives. Then the empirical researcher can estimate the observed component of the individual's utility function. A portion of the individual's utility, however, always remain unobserved to the empirical researcher. The empirical researcher can at best assume a probabilistic distribution for the unobserved component, and then estimate the preference parameters. In the context of this Monte Carlo simulation, a dataset is not available to the empirical researcher.

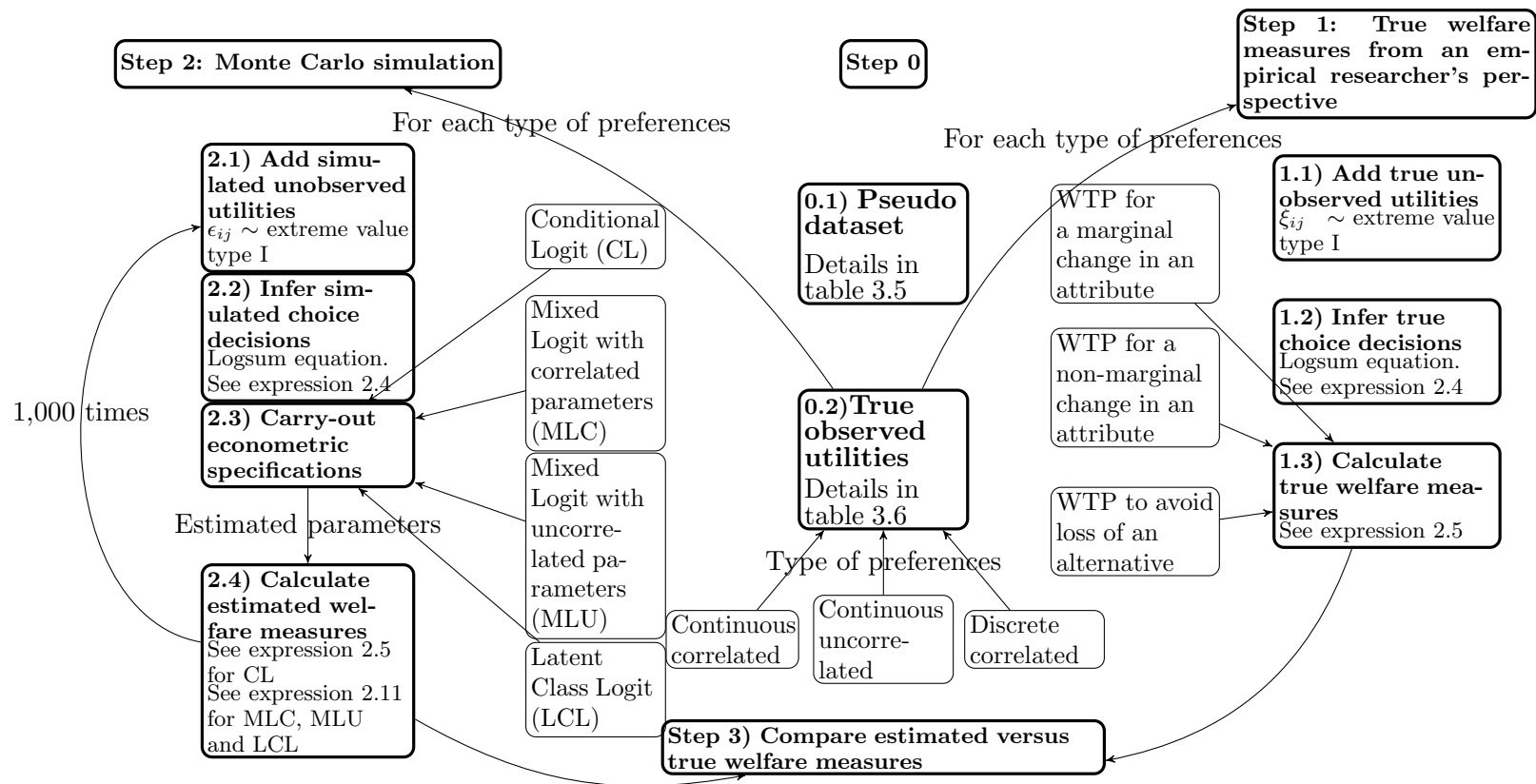


Fig. 3.1: Steps of Monte Carlo simulation studying reliability of welfare estimates from conditional logit, mixed logit, and latent class logit

Thus the first step consists in making available a pseudo-dataset to our imaginary empirical researcher. This step is labeled *step zero* to highlight that this step is carried out as a pre-requisite to carry out the Monte Carlo simulation. The pseudo-dataset is comprised of 2,000 pseudo-individuals who engage in two choice tasks. Within each choice task, the individual chooses among 3 alternatives. These alternatives are described by two attributes, C and Q. C is intended to resemble a travel cost variable. Thus C is log-normally distributed. Q is intended to resemble a quality index. Thus Q is normally distributed. Both C and Q vary across alternatives and individuals. The parameters of the respective distributions are presented in table 3.5.

Tab. 3.5: Pseudo-dataset (2,000 individuals, 3 alternatives, 2 choice tasks)

Variable	Distribution / true value	Description
C	$\mathcal{LN}(2, 1)$	Varies across alternatives and individuals
Q	$\mathcal{N}(2, 2.25)$	Varies across alternatives and individuals
ϵ	Type I extreme value $(1, \pi^2/6)$	Varies across alternatives and individuals

As described in figure 3.1, the generation of true observed utilities is also included in step zero. In order to calculate an individual's true observed utility, a set of true preference parameters is required. Three sets of true preference parameters are designed, each representing a type of preferences — discrete correlated, continuous correlated, and continuous uncorrelated. More details about the true preferences parameters are provided in section 3.3.1.

True observed utilities, i.e. observed utilities calculated using true preference parameters, are used in step one (see figure 3.1). This step consists in calculating

the true welfare measures to be used as reference to evaluate the performance of four econometric methods. True welfare measures are calculated from an empirical researcher's perspective: (i) true unobserved utilities are added to the true observed utilities, assuming unobserved utilities are distributed according to an extreme value type I distribution; (ii) true choice decisions are inferred; and (iii) true welfare measures are calculated. Because the empirical researcher does not know the true unobserved utilities, welfare measures are, strictly speaking, expected true welfare measures. Expectations are taken over the extreme value distributed term, using expression (2.5) which calculates the difference in expected maximum utilities.

Notice the expectations to calculate expected true welfare measures are calculated assuming the only source of randomness arises from the extreme value term. This is consistent with the first step of step one: it takes the three types of true observed utilities — discrete, continuous correlated, continuous uncorrelated—, and adds an extreme value term. This extreme value term is labeled *true unobserved utility*. In the second step of the simulation strategy, an extreme value term is also added to true observed utilities but this term is labeled *simulated unobserved utility*. This difference in labels is a convention. Both true and simulated unobserved utilities are distributed according to an extreme value distribution. The true unobserved utilities result from the first draw of extreme value distributed variables.

As described in figure 3.1, the second step of the simulation strategy consists in carrying out the Monte Carlo simulation. The goal of each of the 1,000 replications is the estimation of welfare measures that will be compared against true welfare measures. Within each replication, (i) a simulated unobserved utility is added to the

true observed utility; (ii) choice decisions are inferred; (iii) preference parameters are estimated through four econometric models; and (iv) welfare measures are estimated, using welfare expression consistent with the corresponding econometric model used to estimate the preference parameters. For each of the three types of observed utilities, four econometric models are estimated: (i) a conditional logit (CL); (ii) a latent class logit (LCL); (iii) a mixed logit with correlated parameters where both parameters are jointly normally distributed (MLC); and a mixed logit with uncorrelated parameters where both parameters are normally distributed (MLC).

Step three compares estimated welfare measures against true welfare measures, and evaluates the performance of each econometric model in retrieving the true welfare measure. Performance is evaluated in terms of (i) unbiasedness, (ii) efficiency and (iii) accuracy. An estimate categorized as unbiased if its 95% confidence interval includes the true value. The most efficient estimate is the one with the smallest 95% confidence interval. Efficiency comparison is restricted to unbiased estimates. Accuracy refers to the magnitude of the relative difference between the estimates and the true values, measured as the absolute value of the mean relative error.

Three WTP measures are compared in step three: (i) WTP for a marginal change in Q; (ii) WTP for a 25% improvement in Q of alternative 1; and (iii) willingness to pay to avoid the loss of alternative 2.

Average WTP estimates over the Monte Carlo replications are compared against average true WTP. Average true WTP results from averaging true WTP over the 2,000 pseudo-individuals. True WTP is calculated in step one, as summarized in figure 3.1. Average estimated WTP results from averaging the mean

estimated WTP over 1,000 Monte Carlo replications. That is, in each replication, the average estimated WTP is calculated over the 2,000 pseudo-individuals, and stored. Thus the average of the 1,000 average estimated WTP is compared against the average true WTP.

3.3.1 *Utility-generating processes*

For illustration purposes, pseudo-data are assumed describing the choices of individuals deciding among 3 alternatives: staying at home, visiting natural park A, and visiting natural park B. Individuals make this decision twice per period. Individuals have preferences over two attributes: travel costs to a natural park (C) and quality (Q). To fix ideas, Q can be thought as an index of natural scenery or wildlife abundance. Individuals receive utility from Q and disutility from C. Marginal utility from Q is represented by β_q . Marginal disutility from C is represented by β_c . Utilities are assumed linear on attributes. Implicitly, utilities are assumed linear in income, with $-\beta_c$ representing the marginal utility from income.

Preferences, i.e. marginal utilities β_q and β_c , are assumed heterogeneous across individuals. Simulated heterogeneity structures are designed to resemble realistic heterogeneity patterns. Arguably, a realistic pattern must account for the possibility that the unobserved utility is correlated with the observed utility. The incorporation of unobserved heterogeneity through an error components representation, as in section 2.2, facilitates the conceptualization of the correlation between unobserved and observed utilities. Correlation is incorporated by assuming the attributes describing the alternatives determine both the observed and the unobserved utilities

(see equation (2.10 and the corresponding explanation). In an errors components representation of the random utility model (RUM), the preference parameters in the observed utility reflect the average preferences in the population. The preference parameters in the unobserved utility reflect the deviation of each individual from the average preferences.

Preference parameters may or may not be correlated. Correlation among parameters is not necessary to induce correlation between observed and unobserved utilities. That is, observed and unobserved utilities may be correlated even when preferences are not correlated. However, in an errors component representation of the RUM, correlated preferences imply that observed and unobserved utilities are correlated.

This study analyzes heterogeneity scenarios for which correlated and uncorrelated preferences are assumed. Controlling for correlated preferences has been a main motivation to use mixed logit models since pioneer applications of these models (see Train, 1998). There are two options to generate pseudo-individuals for which preferences are correlated. One possibility is assuming individuals can be grouped into a finite number of classes. These classes are defined by the preferences of their corresponding members. Thus members of the same class have identical preferences but preferences differ across classes. This strategy, for instance, allows for the identification of two stylized individuals: a scenery lover, price indifferent individual and a scenery indifferent, price focused individual. The scenery lover, price indifferent individual obtains a relatively large marginal utility from Q and a relatively small disutility from C. This individual enjoys the natural scenery in

natural parks and pays less attention to the travel cost. The scenery indifferent, price-focused individual obtains a relatively small utility from Q and a relatively large disutility from C. This individual pays attention to travel cost and less attention to the natural scenery. Preferences captured by a dataset generated according to a grouping strategy are labeled *discrete, correlated* preferences.

The second possibility to generate pseudo-individuals with correlated preferences is assuming preferences vary in a continuous fashion. In this case, individuals can be described as being anywhere in a continuous spectrum that goes from scenery lover, price indifferent to scenery indifferent, price focused. Preferences captured by a dataset generated according to continuous variation in preferences are labeled *continuous, correlated* preferences.

Pseudo-individuals may have uncorrelated preferences. Uncorrelated preferences are simulated only under the continuous scenario, and are labeled *continuous, uncorrelated* preferences.

Table 3.6 describes the functional form of the indirect utility under the three utility-generating process simulated in this study: (i) independently normally distributed; (ii) jointly normally distributed; and (iii) jointly discretely distributed. The first two utility-generating processes assume continuously distributed preferences, and the third utility-generating process assumes discretely distributed preferences.

Indirect utilities in table 3.6 are expressed according to the error components interpretation explained in section 2.2. The observed utilities, $V_{ij} = \beta^j + \beta_c C_{ij} + \beta_q Q_{ij}$, are interpreted as reflecting the average preferences, β_c and β_q . The unob-

Tab. 3.6: Functional form of true indirect utility under the three utility-generating processes simulated in this study (see section 2.3 for details).

True utilities: $U_{ij} = V_{ij} + \eta_{ij}$, where $V_{ij} = \beta^j + \beta_c C_{ij} + \beta_q Q_{ij}$				
$\eta_{ij} = S_{ij} + \epsilon_{ij}$, $\epsilon_{ij} \stackrel{iid}{\sim}$ Type I extreme value $(1, \pi^2/6)$				
Preferences ^a	S_{ij}	Variables	Distributions	Covariance ^b
Normal-normal, uncorrelated	$\theta_i^c C_{ij} + \theta_i^q Q_{ij}$	θ^c, θ^q	$\mathcal{N}(0, 2.31)$, $\mathcal{N}(0, 1.44)$	$cov(\beta_c, \beta_q) =$ $cov(\theta^c, \theta^q) = 0.00$
Normal-normal, correlated	$\lambda_i^c C_{ij} + \lambda_i^q Q_{ij}$	λ^c, λ^q	$\mathcal{N}(0, 2.31)$, $\mathcal{N}(0, 1.44)$	$cov(\beta_q, \beta_c) =$ $cov(\lambda^c, \lambda^q) = 1.82$
Discrete, correlated	$\gamma_c^1 d_i^1 C_{ij} + \gamma_q^1 d_i^1 Q_{ij} +$ $\gamma_c^3 d_i^3 C_{ij} + \gamma_q^3 d_i^3 Q_{ij}$	d^1, d^2, d^3	Multinomial $(1, \pi)$, $\pi =$ $(0.32, 0.36, 0.32)$	$cov(\beta_c, \beta_q) =$ $\pi^1 \gamma_c^1 \gamma_q^1 + \pi^3 \gamma_c^3 \gamma_q^3 =$ 1.82

^a Normal-normal and discrete refers to the distribution of the individual deviations from average preferences. Average preferences are β_c and β_q . Individual deviations are θ_i^c and θ_i^q for the case of normal-normal uncorrelated preferences; λ_i^c and λ_i^q for the case of normal-normal correlated preferences; and $\gamma_c^1, \gamma_c^3, \gamma_q^1$, and γ_q^3 for the case of discrete correlated preferences. For the cases where preferences are correlated, correlation between individual deviations is imposed. Correlation between individual deviations, e.g., $cov(\lambda^c, \lambda^q) = 1.82$ translates into correlation in preferences, i.e. $cov(\beta_c, \beta_q) = 1.82$.

^b For the case of discrete, correlated preferences, true values for γ_c , and γ_q are in table 3.7.

served utilities, η_{ij} , are assumed as composed by two components: $\eta_{ij} = S_{ij} + \epsilon_{ij}$, where ϵ_{ij} reflects the purely random component, and S_{ij} reflects the part of the unobserved utility that generates correlation between observed and unobserved utilities. S_{ij} is interpreted as reflecting individual deviations from the average preferences. Inclusion of individual deviations can be done in a number of ways. This study generates three scenarios by assuming three different structures for S_{ij} .

As summarized in table 3.6, the three heterogeneity scenarios differ in the assumptions made about the nature of the distribution of the individual deviations and whether they are correlated or not. The normal-normal, uncorrelated preferences result from assuming individual deviations, θ_i^c and θ_i^q , are normally distributed with zero means, variances 2.31 and 1.44 respectively, and zero covariance, i.e. $cov(\theta^c, \theta^q) = 0.00$. The normal-normal, correlated preferences result from as-

suming individual deviations, λ_i^c and λ_i^q , are normally distributed with zero means, variances 2.31 and 1.44 respectively, and covariance 1.82, i.e. $cov(\lambda^c, \lambda^q) = 1.82$. The discrete, correlated preferences result from assuming individual deviations, γ_c^1 , γ_c^3 , γ_q^1 and γ_q^3 , are distributed according to a multinomial distribution. The parameters of this multinomial distribution reflect the probabilities that three events occurs in one trial, $\pi = (0.32, 0.36, 0.32)$. This vector of probabilities implicitly determines the correlation in preferences, i.e. $cov(\beta_c, \beta_q) = \pi^1 \gamma_c^1 \gamma_q^1 + \pi^3 \gamma_c^3 \gamma_q^3 = 1.82$ (see Hess et al., 2011, for details about this expression). Values of γ deviations are listed in table 3.7. True average preferences, β_c and β_q , are also listed in table 3.7.

Tab. 3.7: True preference parameters

Variable	Distribution / true value	Description
Continuous preferences		
β_c	-6.00	Marginal utility from C
β_q	4.00	Marginal utility from Q
β^1	2.00	Intercept for alternative 1
β^2	-2.00	Intercept for alternative 2
β^3	0.00	Intercept for alternative 3
Discrete preferences		
γ_c^1	-1.90	Additional marginal utility from C in class 1 with respect to class 2
γ_q^1	-1.50	Additional marginal utility from Q in class 1 with respect to class 2
β_c	-6.00	Marginal utility from C in class 2
β_q	4.00	marginal utility from Q in class 2
γ_c^3	1.90	Additional marginal utility from C in class 3 with respect to class 2
γ_q^3	1.50	Additional marginal utility from Q in class 3 with respect to class 2

In designing the heterogeneity scenarios, particular attention has been paid to three features: (i) assuring identical true average preference parameters across the three heterogeneity scenarios, i.e. $\beta_c = -6.00$ and $\beta_q = 4.00$ (see table 3.7); (ii) assuring identical dispersion of the deviations from the average preferences, i.e. 2.31 and 1.44 for deviations with respect to β_c and β_q respectively⁴ (see table 3.6);

⁴ The variance of the individual deviations in the discrete, correlated scenario may not be obvi-

and (iii) assuring identical covariance between preferences in the two scenarios with correlated preferences, i.e. $cov(\beta_c, \beta_q) = 1.82$ (see table 3.6).

Thus the heterogeneity scenarios have been designed so that average preferences, variance of deviations, and covariance in preferences are identical across preference scenarios. By making sure these features are identical across scenarios, we are able to carry out cleaner experiments. That is, we isolate possible confounding factors. In the case that preference scenarios differ in average preferences, or variance of deviations, or covariance in preferences, then differences in the performance of econometric models may be due to the differences in either average preferences, or variance of deviations, or covariance in preferences. Previous studies using simulated datasets have overlooked this designing feature (e.g. Torres et al., 2011a).

3.4 Results

Average true WTP measures are compared against average estimated WTP measures. Estimated WTP measures are calculated with preference parameters obtained through conditional logit (CL), latent class logit with three classes (LCL), mixed logit with jointly normally distributed parameters (MLC), and mixed logit with uncorrelated normally distributed parameters (MLU). The four econometric specifications approximate simulated choices generated under three preference scenarios: (i) discrete correlated preferences; (ii) jointly normally distributed parameters; and (iii) uncorrelated normally distributed preferences.

For instance, $var(\beta_c) = \pi^1(\gamma_c^1)^2 + \pi^2(0)^2 + \pi^3(\gamma_c^3)^2$. That is, $var(\beta_c) = (2)(0.32)(1.9)^2 = 2.31$.

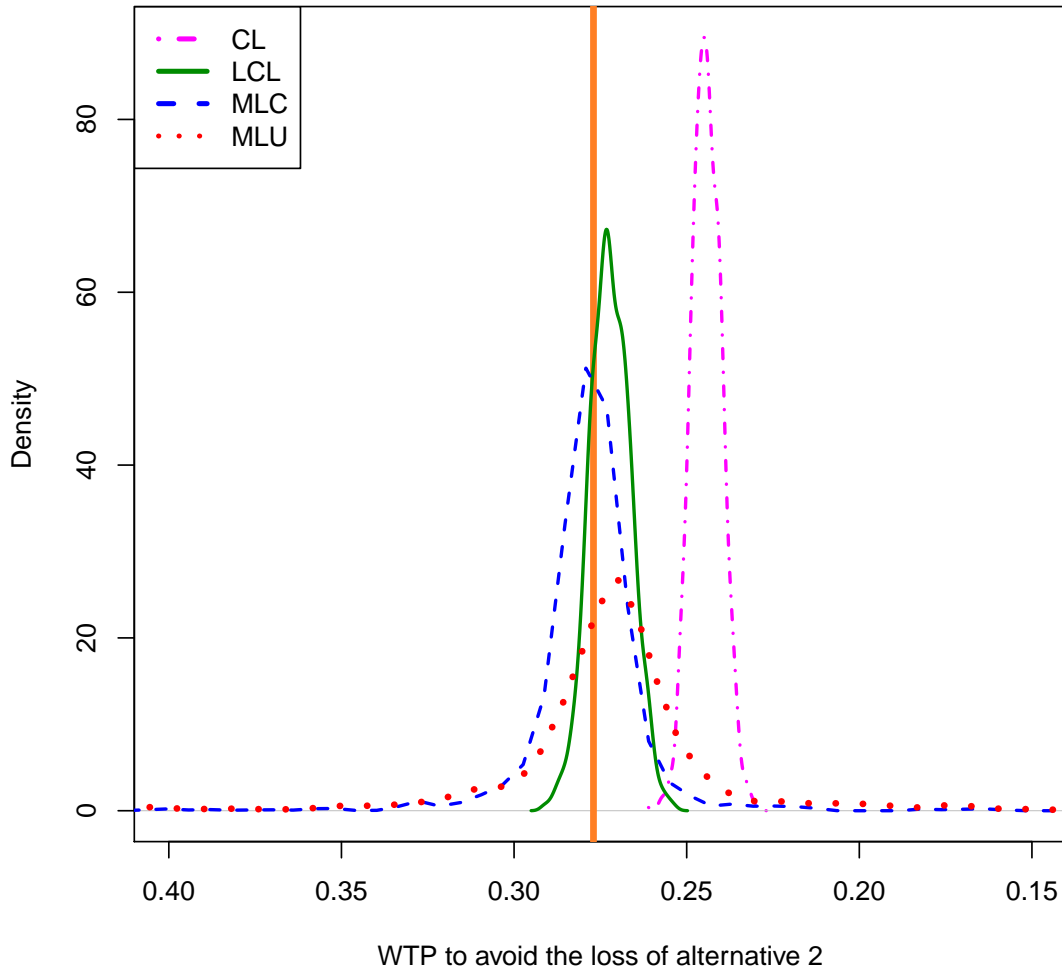


Fig. 3.2: Snapshot on WTP to avoid the loss of alternative 2 by econometric method (discrete, correlated preferences scenario)

Performance is evaluated in terms of (i) unbiasedness, (ii) relative efficiency and (iii) accuracy. Before presenting the specific measures of performance for each econometric methodology, discussion of figure 3.2 will prove useful to understand the main results from the Monte Carlo simulation.

Figure 3.2 presents a snapshot on the empirical distributions of the WTP to avoid the loss of alternative 2 (WTPL) by econometric specification for the case in which discrete, correlated preferences are analyzed. The densities of WTPL obtained from the two mixed logit specifications are not completely presented in figure 3.2 due to their long tails. The vertical straight line is the true value of WTPL. Four features in figure 3.2 are highlighted: (i) CL yields biased estimates of WTPL because the empirical distribution of estimated WTPL yielded by CL does not include the true WTPL; (ii) LCL, MLC, and MLU yield unbiased estimates because their empirical distributions include the true WTPL; (iii) LCL, MLC, and MLU yield distributions with long tails which is particularly true for the case of MLC and MLU; and (iv) the empirical distribution yielded by CL is completely included in the distributions obtained through MLC and MLU, and intersects the distribution obtained through LCL. These four features are also present in figures presenting empirical distributions by econometric specification, regardless of the WTP measure and the preference scenario. Additional figures are not discussed and can be found in the appendix B.

The four features highlighted in figure 3.2 are interpreted as evidence suggesting that although CL yields biased WTP estimates, the large confidence intervals of estimates from MLC, MLU and LCL are the main reason to fail to reject the null

hypothesis that WTP estimates are identical.

Discussion of figure 3.2 is intended as illustration of what is observed across econometric specifications, WTP measures, and preference scenarios. In what follows, performance is summarized in two tables.

Table 3.8 summarizes performance in terms of unbiasedness and relative efficiency of the welfare estimates by econometric specification for each preference scenario. A check mark symbol (\checkmark) indicates the 95% confidence interval of the welfare estimate includes the true value. If this is the case, the estimate is considered unbiased. A plus symbol (+) is reported if the true value is smaller than the lower bound of the 95% confidence interval. A minus symbol ($-$) is reported if the true value is larger than the upper bound of the 95% confidence interval. Thus, according to table 3.8, estimates from CL do not include the true WTP value in its 95% confidence interval, regardless of the WTP measure and the preference scenario. Also, MLC, MLU and LCL always include the true value in their 95% confidence interval. This evidence suggest MLC, MLU and LCL yield unbiased estimates and CL yields biased estimates. These results hold for all WTP measures and preference scenarios.

Table 3.8 identifies with a \star the estimates that, among the unbiased estimates, have the smallest 95% confidence interval for each WTP measure and preference scenario. The WTP estimate with the smallest confidence interval is considered the most efficient estimate. LCL yields the most efficient estimate of the three WTP measures for each utility-generating process. This result strongly suggests that, among the specifications yielding unbiased estimates, LCL yields the most efficient

Tab. 3.8: 95% confidence interval of the welfare estimate includes true value^a, and smallest 95% confidence interval among unbiased estimates^b

	CL ^c	MLC ^c	MLU ^c	LCL ^c
Discrete, correlated preferences				
WTP from 25% increase in Q of alternative 1 (WTPA)	+	✓	✓	✓*
WTP to avoid loss of alternative 2 (WTPL)	-	✓	✓	✓*
Marginal willingness to pay for Q (MWTP)	-	✓	✓	✓*
Normal-normal, correlated preferences				
WTP from 25% increase in Q of alternative 1 (WTPA)	+	✓	✓	✓*
WTP to avoid loss of alternative 2 (WTPL)	-	✓	✓	✓*
Marginal willingness to pay for Q (MWTP)	-	✓	✓	✓*
Normal-normal, uncorrelated preferences				
WTP from 25% increase in Q of alternative 1 (WTPA)	+	✓	✓	✓*
WTP to avoid loss of alternative 2 (WTPL)	-	✓	✓	✓*
Marginal willingness to pay for Q (MWTP)	-	✓	✓	✓*

^a ✓: true value is included; +: true value is smaller than lower bound; -: true value is larger than upper bound.

^b *: Smallest 95% confidence interval among the unbiased estimates.

^c CL: conditinal logit; MLC: mixed logit with two jointly normally distributed parameters; MLU: mixed logit with two uncorrelated normally distributed parameters; LCL: latent class logit.

estimates. The relative efficiency of LCL can be thought as a direct consequence of the larger number of parameters estimated in a LCL specification in comparison to the number of parameters estimated in a ML specification.

Accuracy is measured as the average of the absolute value of the relative errors (AARE), i.e.

$$AARE = M^{-1} \sum_{m=1}^M \left| \frac{(WTP - \hat{WTP})}{WTP} \right| \quad (3.3)$$

The AARE expresses the difference between estimated and true WTP mea-

sures relative to the magnitude of the true WTP. A small AARE reflects accuracy in the estimates. Table 3.9 reports the AARE of welfare estimates by econometric specification for each preference scenario. Three findings are highlighted: (i) LCL yields the smallest AARE, i.e. LCL yields the most accurate estimates regardless of the WTP measure and the heterogeneity scenario; (ii) CL yields the most inaccurate estimates regardless of the WTP measure and the heterogeneity scenario; and (iii) CL yields AARE values similar to those yielded by MLU, specially for the normal-normal correlated scenario. The fact that CL is as inaccurate as MLU is a revealing finding when we consider that, according to table 3.8, MLU yields unbiased estimates and CL yield biased estimates. That is, despite CL yields biased WTP estimates, CL is not much more inaccurate than MLU.

Together, the findings in terms of unbiasedness and accuracy support the notion that mixed logit regularly yields welfare estimates with relatively large confidence intervals. This conclusion applies to LCL despite its relative efficiency. This conclusion holds by preference scenario and WTP under study. This means that ML and LCL regularly yield confidence intervals large enough to include the biased welfare estimates from CL, regardless of the nature of the true utility-generating process and the econometric approach used to approximated the unobserved preference heterogeneity.

Tab. 3.9: Average of absolute value of relative errors (AARE) with respect to true WTP measures.^a

	CL ^b	MLC ^b	MLU ^b	LCL ^b
Discrete, correlated preferences				
WTP from 25% increase in Q of alternative 1 (WTPA)	0.11157	0.03258	0.03119	0.02443
WTP to avoid loss of alternative 2 (WTPL)	0.07227	0.02226	0.03248	0.01947
Marginal willingness to pay for Q (MWTP)	0.28479	0.08383	0.07747	0.06185
Normal-normal, correlated preferences				
WTP from 25% increase in Q of alternative 1 (WTPA)	0.19260	0.06229	0.18084	0.03069
WTP to avoid loss of alternative 2 (WTPL)	0.12966	0.06192	0.12579	0.02522
Marginal willingness to pay for Q (MWTP)	0.50117	0.14550	0.44775	0.08371
Normal-normal, uncorrelated preferences				
WTP from 25% increase in Q of alternative 1 (WTPA)	0.18293	0.06509	0.10870	0.04418
WTP to avoid loss of alternative 2 (WTPL)	0.15016	0.05016	0.11905	0.04233
Marginal willingness to pay for Q (MWTP)	0.49299	0.21328	0.38943	0.12985

^a Measured as $M^{-1} \sum |(WTP - \hat{WTP})/WTP|$, where M is the number of Monte Carlo observations, i.e. 1,000; WTP is the true WTP; and \hat{WTP} is the estimated WTP. ^b CL: conditional logit; MLC: mixed logit with two jointly normally distributed parameters; MLU: mixed logit with two uncorrelated normally distributed parameters; LCL: latent class logit

3.5 Conclusions and discussion

The series of Monte Carlo simulations carried out in this study have sought for evidence supporting the notion that welfare estimates from conditional logit are indistinguishable from estimates obtained through mixed logit and latent class logit simply because methodologies incorporating unobserved heterogeneity yield large confidence intervals. The evidence in terms of unbiasedness and accuracy support this notion: despite the biasedness of the estimates from conditional logit, the accuracy of conditional logit is under some scenarios as good as mixed logit specifications and not much more inaccurate than latent class logit specifications.

The results from this study have two implications for the empirical literature that carries out welfare comparisons across econometric specifications: (i) these comparisons are seemingly unable to provide reliable information about the differences in welfare estimates resulting from controlling for unobserved heterogeneity; and (ii) the use of mixed logit and latent class logit presents a trade-off between gains in statistical fit and efficiency in welfare estimates.

The trade-off between statistical fit and efficiency in welfare estimates has been pointed out previously (e.g. Meijer and Rouwendal, 2006). Actually, this trade-off is at the core of the justification for the use of estimation of discrete choice models in the willingness to pay space (e.g. Scarpa et al., 2008; Train and Weeks, 2005). This literature has focused on the willingness to pay for a marginal change in an attribute, and has overlooked the large standard errors in welfare measures for non-marginal changes, including the loss of an alternative. Arguably, inefficiency in marginal willingness to pay extends to non-marginal changes because marginal willingness to pay is calculated as a simplified version of the expression of the non-marginal changes. Consequently, both marginal and non-marginal changes include the price parameter in the denominator and therefore the distribution of this parameter impacts the standard errors of both types of welfare measures. This argument is behind the justification to keep the price parameter fixed when estimating mixed logits. However, the evidence presented in both the literature review and meta-analysis strongly suggests large standard errors are present even when researchers keep the price parameter fixed. That is, researchers seemingly have not been able to impact the efficiency of welfare estimates.

The trade-off between statistical fit and efficiency in welfare measures opens a question for the literature that combines mixed logit and latent class logit in a single specification (e.g. Greene and Hensher, 2013). This literature allows for an additional layer of continuous preference heterogeneity within each class of a latent class model. The combination of mixed and latent class logits has been developed aiming for an increase in statistical fit. However, this increase has proven poor in several applications (e.g. Bujosa-Bestard et al., 2010; Burton and Rigby, 2009; von Haefen et al., 2005). Considering this poor increase in statistical fit, and the trade-off present in the use of mixed logit and latent class logit, it seems reasonable to wonder whether the researcher is inadvertently giving up efficiency in welfare estimation for relatively poor increase in statistical fit.

Monte Carlo simulations in this study show that welfare estimates from latent class logit are the most efficient among the unbiased estimates. This result holds for both marginal and non-marginal welfare measures (with one exception). This result also holds regardless unobserved heterogeneity is discrete or continuous, correlated or uncorrelated. This finding has implications for the empirical literature: even if researchers strongly suspect continuous heterogeneity, the estimation of a latent class logit may provide more efficient welfare estimates.

The relative performance of latent class logit in this study contrasts with the findings from Torres et al. (2011a), who suggest a mixed logit with lognormally distributed parameters performs better than a latent class specification. They carry out comparisons of non-marginal welfare estimates across latent class logit, mixed logit with two uncorrelated lognormally distributed parameters, and mixed logit

with triangularly distributed parameters. They simulate two utility-generating processes, one with discrete unobserved heterogeneity, and the other with continuous unobserved heterogeneity. It is not clear the distribution of preference parameters in the continuous heterogeneity scenario. The comparisons carried out by Torres et al. (2011a) are essentially different from those carried out here. This study has not carried out mixed logit specifications with lognormally distributed or triangularly distributed parameters. These distributions assure the price parameter never takes a positive value. In contrast, under a normal distribution, a price parameter may take positive values and increase the range of the confidence intervals of the resulting welfare measures. This condition may be driving the poorer efficiency of mixed logit estimates in comparison to latent class logit in this study. Considering the results from the meta-analysis, it seems reasonable to conclude that both mixed logit and latent class logit yield welfare estimates with large confidence intervals, regardless of the relative performance in specific circumstances.

This study has paid attention to keeping correlation between preference parameters fixed across simulated unobserved heterogeneity scenarios. Arguably, this practice increases the reliability of the experimental set up because eliminates a possible confounding factor when comparing latent class logit with mixed logit. The confounding effect arises from the fact that correlation among preference parameters is an inherent feature of the latent class logit but is not an inherent feature of the mixed logit (Hess et al., 2011). The correlation among preference parameters in a latent class logit results from the fact that the preference parameters share the probabilities of occurring. For instance, assume a case in which two classes of visitors

are present. Assume the visitors have preferences over two attributes. Then there are two sets of two parameters to be estimated. The correlation between the two preference parameters arises from the fact that, given the relative size of the classes, they can only be observed by pairs. The correlation among parameters determines the correlation of unobserved utilities across alternatives (Train, 2003) which is a feature a researcher may want to control when designing experiments that compare mixed logit and latent class logit. This paper does not show the impacts from not controlling for correlation among parameters. This issue can be considered a topic for further research.

4. WELFARE IMPLICATIONS FROM MISSPECIFICATION OF LATENT CLASS LOGIT MODELS

4.1 Introduction

Current practices in estimating latent class models include the use of researchers' own judgement when likelihood-based criteria provide conflicting evidence about the number of classes. The prominence of this practice is illustrated by the 40% of applications that rely only in the researcher's own judgement to select the number of classes (see section 4.2). The strategy of using a researcher's own judgement, however, faces a risk: the researcher may not guess the true number of classes. Estimation of a latent class model with an incorrect number of classes may be of concern to economists if the number of classes matters for welfare estimates.

This chapter designs a series of Monte Carlo simulations to learn whether welfare estimates from latent class logit specifications are robust to the number of classes. Simulated choices are generated from utility-generating processes for which individuals belong to one of six different classes. Six latent class logit specifications are estimated on the simulated choices. These specifications differ in the number of classes: from one (conditional logit) to six (six-class latent class logit). Three willingness to pay (WTP) measures are under study: (i) WTP for a marginal change in an attribute; (ii) WTP for a non-marginal change in an alternative's attribute; and

(iii) willingness to pay to avoid the loss of an alternative. Average WTP estimates over the Monte Carlo replications are compared against true WTP.

Monte Carlo simulations are carried out for two utility-generating processes. Each process resembles circumstances under which a researcher would likely use his/her own judgement. The first utility-generating process assumes one of the six classes contains individuals with a close-to-zero price parameter. A model with six classes approximating this simulated data yields, with some probability, price parameter estimates statistically undistinguishable from zero. A common practice in empirical research is choosing models yielding a positive price coefficient. Thus an empirical researcher analyzing choices derived from this utility-generating process would likely choose a latent class logit with five classes if he/she follows the practice of selecting specifications for which the price parameter is positive in all classes. The second utility-generating process assumes the percentage of individuals that belong to one of the six classes is relatively small. Another common empirical practice consists in choosing models with classes that exceed a minimum size. Thus an empirical researcher analyzing choices derived from the second utility-generating process would likely dismiss a latent class logit with six classes if he/she follows the practice of dismissing specifications with small classes.

Reliability of the welfare estimates is evaluated in terms of (i) unbiasedness, i.e. whether the true value falls within the 95% confidence interval of the estimates; (ii) efficiency, i.e. which specification yields the smallest 95% confidence interval; and (iii) accuracy, i.e. how large is the average absolute value of the relative errors between estimated and true WTP values.

Results from the Monte Carlo comparisons show that the reliability of welfare estimates crucially depends on the estimated number of classes: (i) latent class logit specifications yield biased welfare estimates when estimated with a number of classes different from the true number; and (ii) in terms of accuracy, the most inaccurate estimates are yielded by the latent class logit with five classes. Both findings hold for both utility-generating processes simulated in this study. The inaccuracy of the model with five classes is an important finding because empirical researchers analyzing the choices simulated in this study arguably would have preferred models with five classes.

To the best of my knowledge, no previous paper has focussed on whether welfare measures are robust to estimated number of classes. So far, Monte Carlo experiments have focused on welfare measures in the context of latent class conditional logit models have studied the impact from misspecification of the utility function (Torres et al., 2011a), the effect of implementing sampling strategies on large choice sets (Domanski and von Haefen, 2012), and the effect from the design of discrete choice experiments (e.g. Ferrini and Scarpa, 2007).

4.2 *Strategies to select latent classes*

There are no standard strategies to select the number of classes in applications of latent class techniques. Table 4.1 reviews the strategies in 24 empirical applications.¹ Two studies fail to report the criteria used in selecting optimal number of classes. The rest report the use of at least one likelihood-based criterion.

¹ Table C.1 in appendix C describes the goals and methodological strategies in the applications reviewed in this section.

The second column in table 4.1 reports the likelihood criteria favoring the selected number of classes. This criteria may be used together with the researcher's own judgement or not. For instance, six studies (25%) chose a number of classes that is favored by no likelihood-based criterion, and five studies (21%) do not specify whether their selection is supported by a likelihood criterion. In addition, eleven studies (46%) chose the number of classes favored by the Bayesian information criterion (BIC), and two studies (8%) chose the number of classes favored by the Akaike information criterion (AIC).

According to the third column in table 4.1, 13 studies (54%) explicitly report the use of the researcher's own judgement when selecting number of classes. A researcher's own judgement is subjective in nature, and may take the following forms: (i) a priori beliefs about the number of classes (e.g. Beharry-Borg and Scarpa, 2010; Scarpa and Thiene, 2005); (ii) preference for parsimonious specifications (e.g. Boxall and Adamowicz, 2002; Provencher and Bishop, 2004); (iii) preference for specifications with statistically significant variables in most classes, paying particular attention to obtaining positive price parameters (e.g. Boxall and Adamowicz, 2002; Brouwer et al., 2010; Garrod et al., 2012; Hynes et al., 2008; Ruto et al., 2008); (iv) rejecting a specification for which the relative change in information criteria is relatively small when adding classes (e.g. Birol et al., 2006, 2009; Broch and Vedel, 2012; Kosenius, 2010); (v) rejecting specifications with relatively small classes (e.g. Broch and Vedel, 2012); and (vi) any combination of these criteria.

Tab. 4.1: Criteria used to select number of classes in environmental and resource economics studies

Paper	Criteria favoring selected number of classes	Researcher's own judgement is explicitly used to select classes	Additional criteria used in selecting number of classes	Maximum number of classes attempted	Selected number of classes
Richards (2000)	*	No	—	*	2
Boxall and Adamowicz (2002)	BIC	Yes	AIC	6	4
Provencher et al. (2002)	BIC, AIC	No	—	4	3
Greene and Hensher (2003)	*	No	—	5	3
Scarpa et al. (2003)	BIC	No	—	4	2
Provencher and Bishop (2004)	None	Yes	AIC, BIC	5	3
Scarpa and Thiene (2005)	None	Yes	AIC, BIC, crAIC	5	4
Birol et al. (2006)	None	Yes	AIC, BIC	4	2
Milon and Scrogin (2006)	BIC	No	—	4	3
Ouma et al. (2007)	BIC	No	—	4	3
Hynes et al. (2008)	BIC	Yes	AIC, CrAIC, AIC3	11	6
Ruto et al. (2008)	None	Yes	AIC, BIC, AIC3	12	3
Birol et al. (2009)	None	Yes	AIC, BIC	5	3
Colombo et al. (2009)	AIC, CAIC	No	—	*	3
Shen (2009)	AIC, CAIC	No	—	*	3
Beharry-Borg and Scarpa (2010)	*	Yes	BIC, AIC, AIC3	4	2
Brouwer et al. (2010)	None	Yes	AIC, BIC	5	4
Kosenius (2010)	BIC	Yes	AIC	10	5
Breffle et al. (2011)	*	Yes	AIC, AIC3, BIC	5	4
Kikulwe et al. (2011)	BIC, AIC3	No	—	5	2
van Putten et al. (2011)	BIC	No	—	*	3
Broch and Vedel (2012)	BIC	Yes	AIC	5	4
Chung et al. (2012)	BIC, CAIC, Entropy	No	AIC	6	3
Garrod et al. (2012)	*	Yes	AIC	*	4

* means that the feature is not specified in the document.

— means that no additional likelihood criterion is used in choosing the number of classes.

The fourth column in table 4.1 reports whether researchers have used other criteria in addition to the one favoring the selected number of classes and their own judgement. Around 58% of the studies have used more than one likelihood-based criteria.

The fifth column in table 4.1 reports the maximum number of classes attempted in each study. Five studies (21%) have attempted specifications that include 6 or more classes, with a maximum of 12; eight studies (33%) have attempted specifications with 5 classes or less; and five studies (21%) do not report the maximum number of classes attempted.

The sixth column in table 4.1 reports the selected number of classes. The most frequent number of classes is three, selected in 11 studies (46%). Four classes are selected in 6 studies (25%); two classes are selected in five studies (21%); five and six classes are selected in one study each.

Notice that the 13 studies using the researcher's own judgement include the six studies following no likelihood-based criterion and three studies that do not specify whether a likelihood-based criterion favors the selected number of classes (see second column). That is, in 38% of the applications the researcher's own judgement plays the most important role in selecting the number of classes.

Also, applications for which the researcher's own judgement plays the most important role tend to estimate specifications with four classes. Four out of these nine applications, i.e. 44%, have chosen specifications with four classes. Without carrying out a formal statistical comparison, this percentage seems larger than the corresponding percentage observed for the 24 applications (25%).

In summary, around half of the applications have chosen the number of classes favored by the BIC; a quarter of the applications have chosen a number of classes with no support from a likelihood-based criterion; and around 20% have not reported whether a likelihood-based criterion favors the chosen number of classes. Around half of the applications explicitly report the use of the researcher's own judgement. Most of the applications incorporating the researcher's own judgement have relied only on this judgement to decide the number of classes, and have a tendency to choose applications with four classes. These applications represent around 40% of the total number of reviewed applications.

4.3 *Simulation strategy*

The interest in studying whether welfare estimates from latent class logit models are robust to the number of classes originates in a finding of the literature review: around 40% of the reviewed applications rely only on the researcher's own judgement. Following their judgement, researchers tend to disregard either (i) specifications with positive price parameters in one class, or (ii) specifications with a small class, or (iii) both. The practice of using a researcher's own judgement faces the risk of selecting a number of classes different than the true one. This situation becomes an issue for economists if the number of classes matters in terms of welfare estimates. This section describes the Monte Carlo simulations designed to evaluate the reliability of the estimated welfare measures to number of classes.

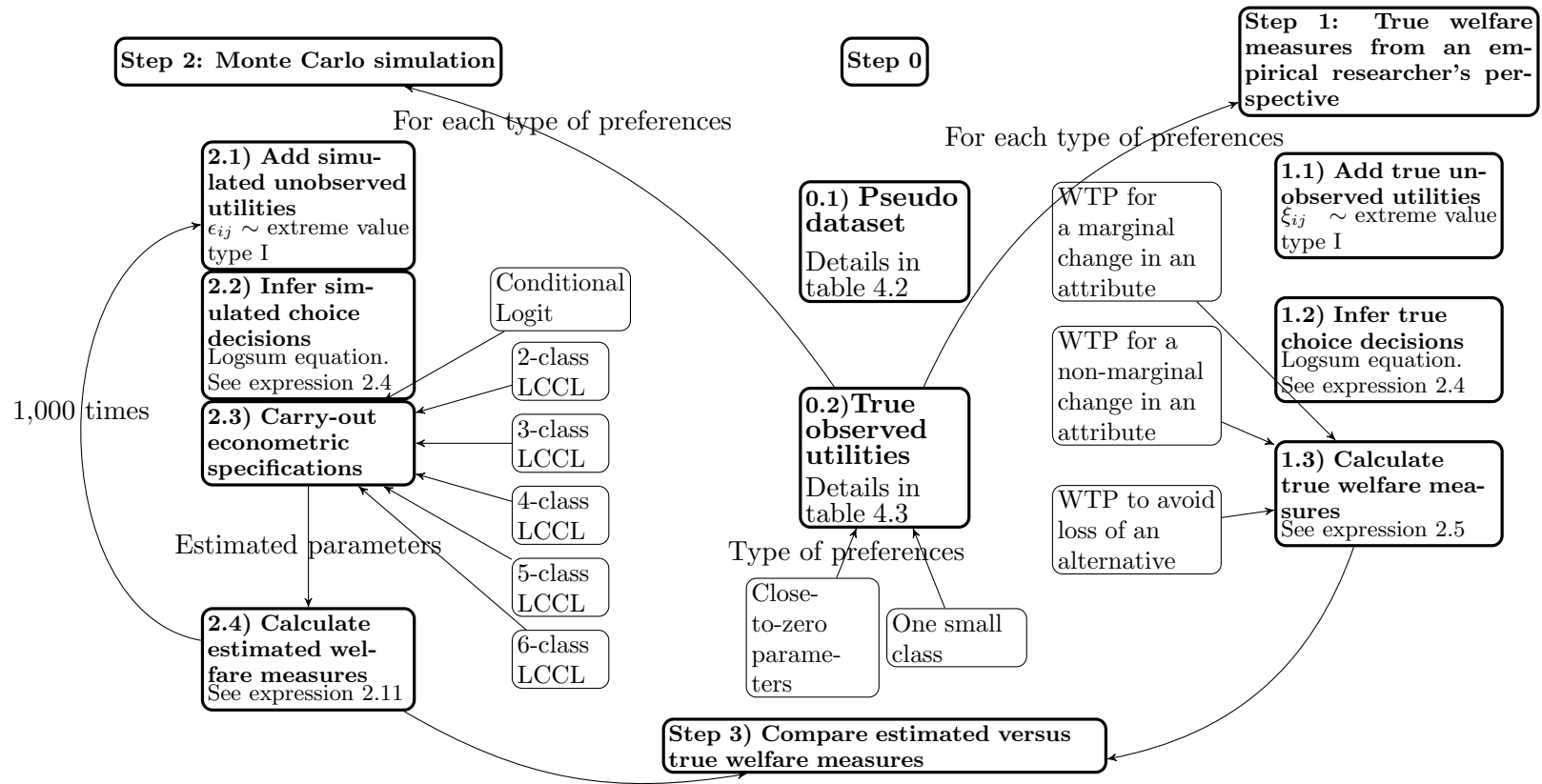


Fig. 4.1: Steps of Monte Carlo simulation studying reliability of estimates welfare measures to number of classes

A flow chart describing the simulation strategy is presented in figure 4.1. The Monte Carlo simulation has been designed under the following reasoning: (i) an empirical researcher has access to a dataset including both individual- and alternative-specific attributes describing the alternatives from which an individual chooses; (ii) an empirical researcher has means to estimate the preferences an individual has over attributes describing the alternatives. Then the empirical researcher can estimate the observed component of the individual's utility function. A portion of the individual's utility, however, always remain unobserved to the empirical researcher. The empirical researcher can at best assume a probabilistic distribution for the unobserved component, and then estimate the preference parameters.

In the context of this Monte Carlo simulation, a dataset is not available to the empirical researcher. Thus the first step consists in making available a pseudo-dataset to our imaginary empirical researcher. This step is labeled *step zero* in figure 4.1 to highlight that this step is carried out as a pre-requisite to carry out the Monte Carlo simulations.

For illustration purposes, the pseudo-dataset is assumed to describe the choices of individuals deciding among 3 alternatives: staying at home, visiting natural park A, and visiting natural park B. Individuals make this decision twice per period. Individuals have preferences over two attributes: travel cost (C) and quality (Q). To fix ideas, Q can be thought as an index of natural scenery or wildlife abundance. Individuals receive utility from Q and disutility from C. Marginal utility from Q is represented by β_q . Marginal disutility from C is represented by β_c . Utilities are assumed linear on attributes. Implicitly, utilities are assumed linear in income, with

$-\beta_c$ representing the marginal utility from income.

The pseudo-dataset is comprised of 2,000 pseudo-individuals who engage in two choice tasks. Within each choice task, the individual chooses among 3 alternatives. These alternatives are described by two attributes, C and Q. C is intended to resemble a travel cost variable. Thus C is log-normally distributed. Q is intended to resemble a quality index. Thus Q is normally distributed. Both C and Q vary across alternatives and individuals. The parameters of the respective distributions are presented in table 4.2.

Tab. 4.2: Pseudo-dataset (2,000 individuals, 3 alternatives, 2 choice sets)

Variable	Distribution / true value	Description
C	$\ln\mathcal{N}(2,1)$	Varies across alternatives and individuals
Q	$\mathcal{N}(2, 2.25)$	Varies across alternatives and individuals
ϵ	Type I extreme value $(1, \pi^2/6)$	Varies across alternatives and individuals

As described in figure 3.1, the generation of true observed utilities is also included in step zero. In order to calculate an individual's true observed utility, a set of true preference parameters is required. Two sets of true preference parameters are simulated. Both sets of parameters resemble situations under which an empirical researcher would likely choose a smaller number classes than the true number of classes. One scenario is labeled *close-to-zero price parameter*, to emphasize that the distinguishing feature of this scenario is the presence of a price parameter close to zero. The second scenario is labeled *one small class*, to emphasize that the distinguishing feature of this scenario is the presence of a relatively small class.

In the close-to-zero price parameter scenario, individuals belong to one of six

different classes. The marginal utility from Q (β_q) is set to zero for class 1, and the marginal (dis)utility from C (β_c) is set to -0.10 — a close-to-zero value— for class 6. All true preference parameters by class, and relative size of each class are listed in table 4.3. A model with six classes approximating the choices simulated according to the close-to-zero price parameter scenario will yield, with some probability, price parameter estimates statistically undistinguishable from zero. Thus an empirical researcher analyzing choices derived from this utility-generating process would likely choose a latent class logit with five classes if he/she follows the practice of selecting specifications for which the price parameter is positive and significant in all classes.

Tab. 4.3: True preference parameters and true relative size of classes

	Scenarios			
	Close-to-zero price parameter ^a		Small class ^a	
	True preference parameters			
	β_c	β_q	β_c	β_q
Class 1	-8.00	0.00	-8.00	0.00
Class 2	-6.50	1.50	-6.50	1.50
Class 3	-5.00	3.00	-5.00	3.00
Class 4	-3.00	5.00	-3.00	5.00
Class 5	-1.50	6.50	-1.50	6.50
Class 6	-0.10	8.00	-0.10	8.00
	True relative size of classes			
Class 1	0.10		0.25	
Class 2	0.15		0.25	
Class 3	0.25		0.20	
Class 4	0.25		0.15	
Class 5	0.15		0.10	
Class 6	0.10		0.05	

^a Alternative-specific parameters are fixed across classes: $\beta^1 = 1.00$, $\beta^2 = -1.00$, and $\beta^3 = 0.00$.

In the one small class scenario, individuals also belong to one of six classes. As shown in table 4.3, the marginal utilities from Q and C are identical under both scenarios. The difference between scenarios consists in the relative size of the six

classes, with particular emphasis on having a relatively smaller class in the one small class scenario. As shown in table 4.3, classes 1 and 6 are the smallest under the close-to-zero price parameter scenario, with a relative size of 0.10. In contrast, the smallest class under the one small class scenario is class 6, with a relative size of 0.05. A common empirical practice consists in choosing models with classes that exceed a minimum size. Thus an empirical researcher following this empirical practice would likely dismiss a latent class logit with six classes when analyzing choices derived from the one small class scenario.

According to the flow chart in figure 4.1, step one of the simulation strategy consists in calculating true welfare measures. True welfare measures are used as reference to evaluate the performance of four econometric methods. True observed utilities are used when calculating true welfare measures. True observed utilities are calculated using true preference parameters as listed in table 4.3.

True welfare measures are calculated from an empirical researcher's perspective: (i) true unobserved utilities are added to the true observed utilities, assuming unobserved utilities are distributed according to an extreme value type I distribution; (ii) true choice decisions are inferred; and (iii) true welfare measures are calculated. Because the empirical researcher does not know the true unobserved utilities, welfare measures are, strictly speaking, expected true welfare measures. Expectations are taken over the extreme value distributed term, using expression (2.5) which calculates the difference in expected maximum utilities. The expectations to calculate expected true welfare measures are calculated assuming the only source of randomness arises from the extreme value term. This is consistent with

the first step of step one: it takes the true observed utilities under both scenarios — close-to-zero price parameter and one small class—, and adds an extreme value term. This extreme value term is labeled *true unobserved utility*. In the second step of the simulation strategy, an extreme value term is also added to true observed utilities but this term is labeled *simulated unobserved utility*. This difference in labels is a convention. Both true and simulated unobserved utilities are distributed according to an extreme value distribution. The true unobserved utilities result from the first draw of extreme value distributed variables.

As described in figure 4.1, the second step of the simulation strategy consists in carrying out the Monte Carlo simulation. The goal of each of the 1,000 replications is the estimation of welfare measures that will be compared against true welfare measures. Within each replication, (i) a simulated unobserved utility is added to the true observed utility; (ii) choice decisions are inferred;(iii) preference parameters are estimated through six econometric models; and (iv) welfare measures are estimated. For each of the two utility-generating scenarios, six econometric models are estimated: a conditional logit, and five latent class specifications. The latent class specifications include five specifications with incorrect number of classes (1 to 5) and the specification with the correct number of classes, i.e. six. Estimates of both preference parameters and the relative size of each class are used in the estimation of WTP measures.

Step three compares estimated welfare measures against true welfare measures, and evaluates the performance of each econometric model in retrieving the true welfare measure. Performance is evaluated in terms of (i) unbiasedness, (iii) efficiency

and (iii) accuracy. An estimate categorized as unbiased if its 95% confidence interval includes the true value. The most efficient estimate is the one with the smallest 95% confidence interval. Efficiency comparison is restricted to unbiased estimates. Accuracy refers to the magnitude of the relative difference between the estimates and the true values, measured as the absolute value of the mean relative error.

Three WTP measures are compared in step three: (i) WTP for a marginal change in Q; (ii) WTP for a 25% improvement in Q of alternative 1; and (iii) willingness to pay to avoid the loss of alternative 2.

Average WTP estimates over the Monte Carlo replications are compared against average true WTP. Average true WTP results from averaging true WTP over the 2,000 pseudo-individuals. True WTP is calculated in step one, as summarized in figure 4.1. Average estimated WTP results from averaging the mean estimated WTP over 1,000 Monte Carlo replications. That is, in each replication, the average estimated WTP is calculated over the 2,000 pseudo-individuals, and stored. Thus the average of the 1,000 average estimated WTP is compared against the average true WTP.

4.4 *Results*

For both utility-generating processes, true WTP measures are compared against welfare estimates obtained through six latent class specifications with, respectively, one (conditional logit), two, three, four, five and six classes. These model are labeled CL, LCL2, LCL3, LCL4, LCL5, and LCL6, respectively. Only LCL6 incorporates

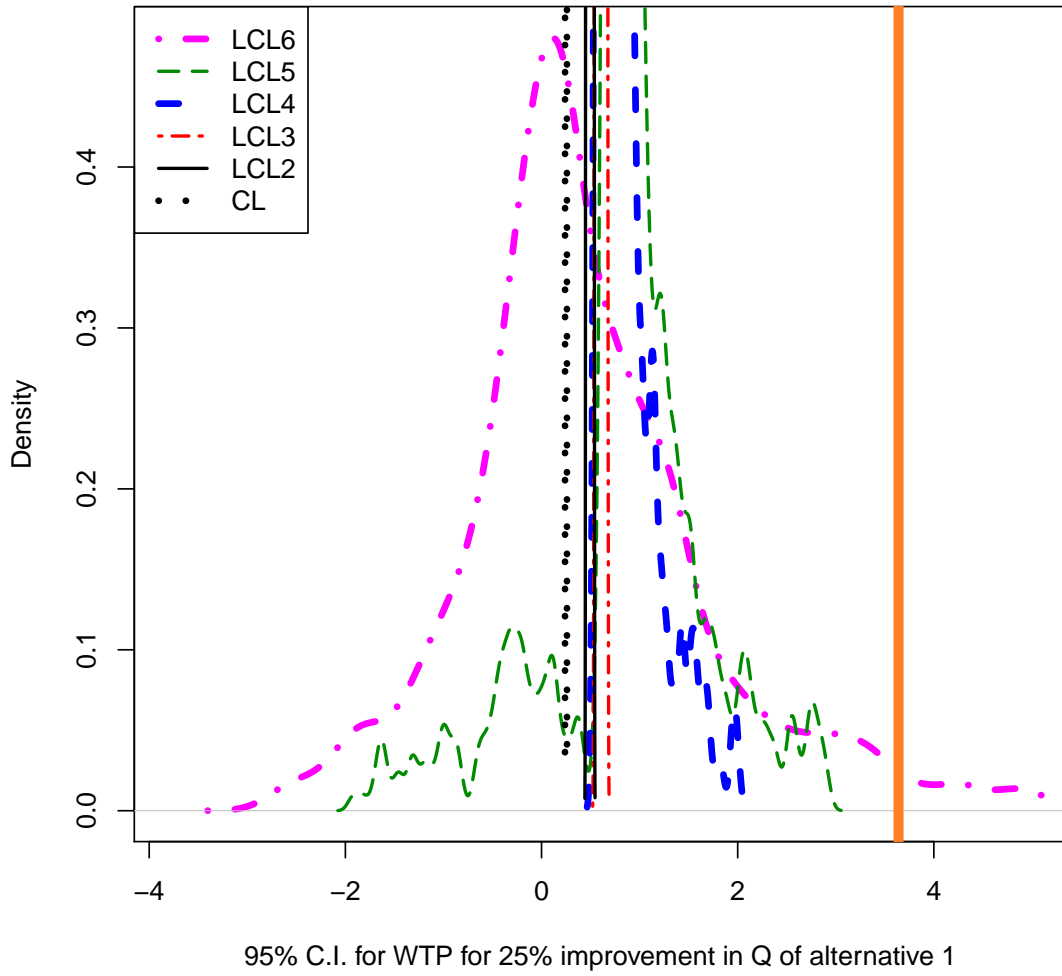


Fig. 4.2: Snapshot of WTP for 25% improvement in Q of alternative 1 by econometric method (close-to-zero price parameter scenario)

the true number of classes.

Performance is evaluated in terms of (i) unbiasedness, (iii) relative efficiency and (iii) accuracy. Before presenting the specific measures of performance for each econometric methodology, discussion of figures 4.2, 4.3, and 4.4 will prove useful to understand the main results from the Monte Carlo simulation.

Figure 4.2 presents a snapshot of the 95% confidence intervals of the WTP for 25% improvement in Q (WTPA) of alternative 1 by econometric specification for the close-to-zero price parameter scenario. The vertical straight line is the true WTPA value. Three features in this figure are highlighted: (i) the true WTPA is included only in one 95% confidence interval — the confidence interval corresponding to LCL6; (ii) the zero is included in two confidence intervals — the corresponding to LCL5 and LCL6; and (iii) the confidence intervals become larger the more classes are estimated. With respect to feature (iii), notice the small confidence interval of the WTPA estimated through CL, LCL2 and LCL3.

The evidence in figure 4.2 suggests that (i) only the LCL6 yields unbiased WTPA estimates; (ii) however, the null hypothesis that WTPA estimates from LCL6 are equal to zero can not be rejected at 95% confidence; (iii) similarly, the null hypothesis that WTPA estimates from LCL5 are equal to zero can not be rejected at 95% confidence; and (iv) WTPA estimates from CL, LCL1, LCL2, LCL3, LC4, although different from zero and with relatively small confidence intervals, are biased.

Figure 4.3 tells an almost identical story than figure 4.3 but for the case of the WTP for a marginal improvement in Q. As shown in figure 4.4, the story is a little

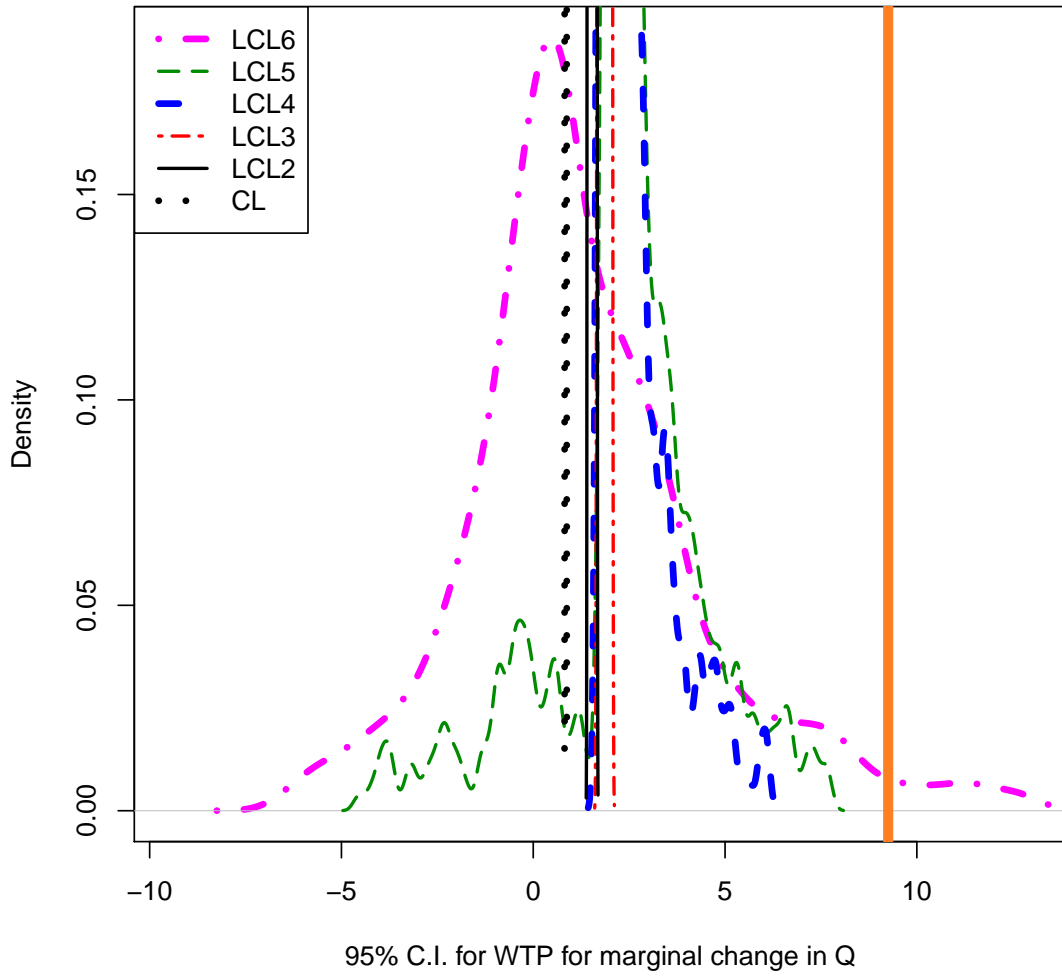


Fig. 4.3: Snapshot of WTP for marginal improvement in Q by econometric method (close-to-zero price parameter scenario)

different for the case of the WTP to avoid the loss of alternative 2 (WTPL). The vertical straight line represents the true WTPL. Three features are highlighted: (i) all but CL and LCL2 econometric specifications yield 95% confidence intervals that include the true WTPL; (ii) LCL6 yield a 95% confidence interval that includes the zero; and (iii) CL, LCL2, and LCL3 yield relatively small confidence intervals.

The evidence in figure 4.4 suggest that (i) WTPL is relatively robust to the number of classes in terms of unbiasedness; (ii) the null hypothesis that WTPL estimates from LCL6 are equal to zero can not be rejected at 95% confidence; (iii) WTPL estimates from CL, LCL1, and LCL2 although different from zero and with relatively small confidence intervals, are biased; and (iv) the best WTPL estimates, in terms of unbiasedness and relative efficiency, seem to be yielded by LCL3.

Relatively similar stories can be told from the corresponding figures presenting the confidence intervals for the case of the one small class scenario. These figures are not discussed and can be found in the appendix D. In what follows, performance is summarized in two tables.

The top panel of table in table 4.4 evaluates unbiasedness and relative efficiency of the welfare estimates by econometric specification for each preference scenario. A check mark symbol (\checkmark) indicates the 95% confidence interval of the welfare estimate includes the true value. If this is the case, the estimate is considered unbiased. A plus symbol (+) is reported if the true value is smaller than the lower bound of the 95% confidence interval. A minus symbol ($-$) is reported if the true value is larger than the upper bound of the 95% confidence interval.

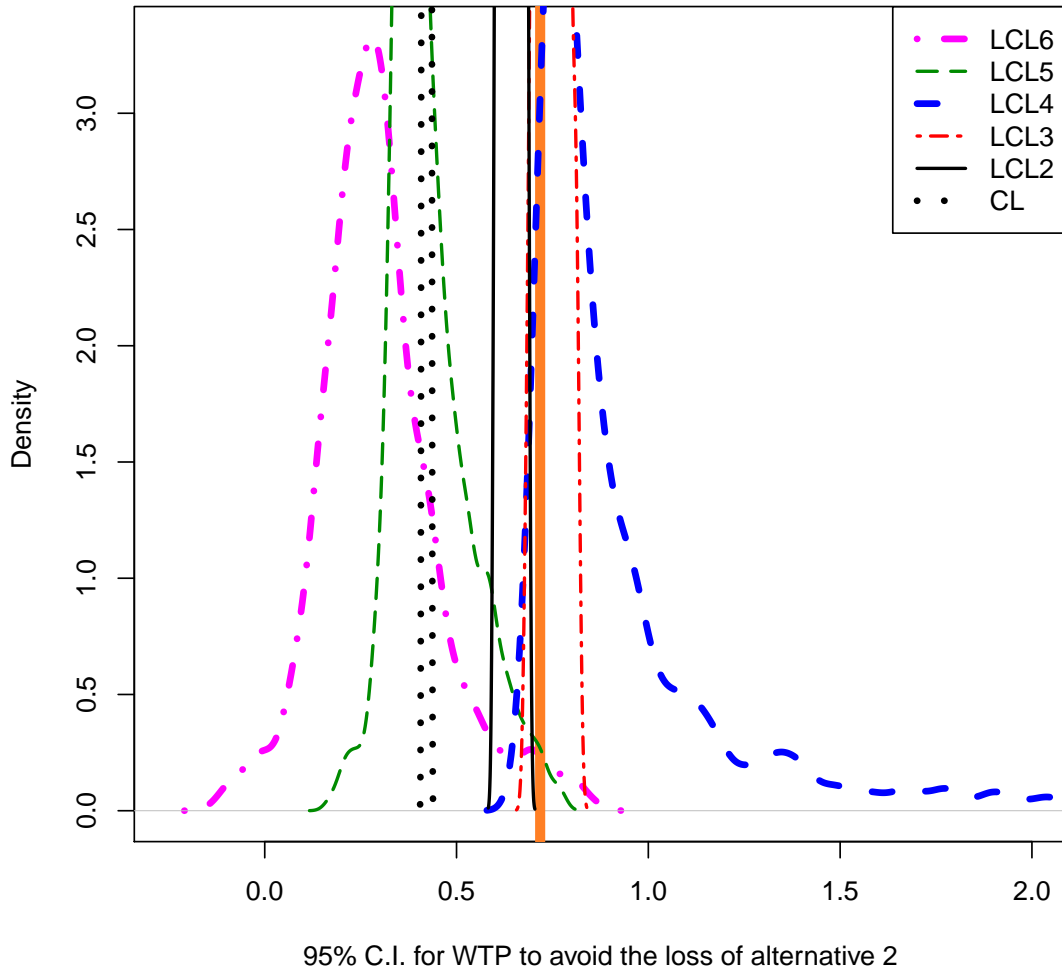


Fig. 4.4: Snapshot of WTP to avoid loss of alternative 2 by econometric method (close-to-zero price parameter scenario)

Tab. 4.4: 95% confidence interval includes true value, and average of absolute value of relative errors (AARE).

	Number of classes in latent class logit					
	One	Two	Three	Four	Five	Six
95% confidence interval includes true value ^a and smallest 95% confidence interval among unbiased estimates ^b						
<i>Close-to-zero parameters scenario</i>						
WTP from 25% increase in Q of alternative 1 (WTPA)	-	-	-	-	-	✓
WTP to avoid loss of alternative 2 (WTPL)	+	+	✓ *	✓	✓	✓
Marginal willingness to pay for Q (MWTP)	-	-	-	-	-	✓
<i>One small class scenario</i>						
WTP from 25% increase in Q of alternative 1 (WTPA)	-	-	-	-	-	✓
WTP to avoid loss of alternative 2 (WTPL)	+	✓ *	-	-	✓	✓
Marginal willingness to pay for Q (MWTP)	-	-	-	-	-	✓
Average of absolute value of relative errors (AARE)						
<i>Close-to-zero parameters scenario</i>						
WTP from 25% increase in Q of alternative 1 (WTPA)	0.931	0.864	0.835	0.924	2.485	1.23
WTP to avoid loss of alternative 2 (WTPL)	0.081	0.021	0.012	0.274	0.320	0.161
Marginal willingness to pay for Q (MWTP)	2.300	2.120	2.035	2.329	6.258	3.075
<i>One small class scenario</i>						
WTP from 25% increase in Q of alternative 1 (WTPA)	0.934	0.840	0.787	0.776	1.028	0.947
WTP to avoid loss of alternative 2 (WTPL)	0.088	0.013	0.038	0.084	0.119	0.134
Marginal willingness to pay for Q (MWTP)	2.321	2.061	1.909	1.883	2.546	2.351

^a ✓: true value is included; +: true value is smaller than lower bound; -: true value is larger than upper bound.

^b ★: Smallest 95% confidence interval among the unbiased estimates.

^c Measured as $M^{-1} \sum |(WTP - \hat{WTP})/WTP|$, where M is the number of Monte Carlo observations, i.e. 1,000.

Thus, according to the top panel of table 4.4, the model with six classes is the only one consistently yielding unbiased welfare estimates. The specifications with a number of classes smaller than the true number yield estimates of WTPA and MWTP with 95% confidence intervals that do not include the true values. These results hold for both the *close-to-zero* scenario and the *one small class* scenario.

For the case of WTPL, some of the misspecified models yield 95% confidence intervals that include the true values under both scenarios. For the *close-to-zero parameters* scenario, specifications with 3, 4, and 5 classes yield confidence intervals including true values of WTPL. The specification with 3 classes yields the most efficient confidence interval. For the *one small class* scenario, specifications with 2 and 5 classes yield confidence intervals including true values of WTPL. The specification with 2 classes yields the most efficient confidence interval, among the unbiased confidence intervals.

Accuracy is measured as the average of the absolute value of the relative errors (AARE), i.e.

$$AARE = M^{-1} \sum_{m=1}^M \left| \frac{(WTP - W\hat{T}P)}{WTP} \right| \quad (4.1)$$

The AARE expresses the difference between estimated and true WTP measures relative to the magnitude of the true WTP. A small AARE reflects accuracy in the estimates. The bottom panel of table 4.4 reports the AARE of welfare estimates by econometric specification for each preference scenario. In general, the six models yield very inaccurate estimates. The specification with five classes yield the

most inaccurate estimates, with AARE reaching 6.258. This value means that the difference between estimated and true WTP is 6 times larger than the magnitude of the true WTP. The model with six classes is either second or third in terms of inaccuracy. These results hold for both scenarios. Consistently with the information in figure 4.4, LCL2, and LCL3 yield relatively accurate estimates of WTPL in both preference scenarios, with AARE reaching values such as 0.013, and 0.012.

The inaccuracy from most of the econometric specifications is a direct consequence of the close-to-zero marginal utility from C assumed for class 6, i.e. $\beta_c^6 = -0.10$. Because β_c^6 is close to zero, a six-class specification may not be able to statistically distinguish it from zero. Thus positive estimates are possible. A positive estimate affects both the numerator and denominator of equation 2.11, used to calculate WTP measures. In the numerator, a positive $\hat{\beta}_c^6$ changes the relative ranking of the alternatives under consideration. In the denominator, a positive and close-to-zero $\hat{\beta}_c^6$ both flips the expected sign and increases the magnitude of the welfare measure. Although positive estimates of β_c^6 may occur infrequently, these outliers impact average welfare measures.

The inaccuracy of the model with five classes is an important finding because empirical researchers analyzing the choices simulated in this study arguably would select models with five classes. This result is explained as follows: a latent class specification with five classes combines the behavior embedded in class 6 with the behavior embedded in a different class. The relative size of the class that results from mixing class 6 with another class is relatively larger than the relative size of class 6 itself. Thus the relative importance of class 6 is over-emphasized in the

five-class specification. In contrast, the relative importance of class 6 is diluted in specifications with less than five classes.

4.5 *Conclusions and discussion*

This chapter has raised the question of whether the strategies to implement a researcher's own judgement ultimately impact the reliability of welfare estimates. The interest in this issue originates in a finding of the literature review presented in this chapter: around 40% of the reviewed applications rely only on the researcher's own judgement.

Reliability of welfare estimates have been studied under two strategies used to incorporate the researcher's own judgement in the selection of number of classes: (i) the practice of selecting specifications for which the price parameter is significant in all classes; and (ii) the practice of dismissing specifications with relatively small classes.

Results from the Monte Carlo comparisons show that the reliability of welfare estimates crucially depends on the estimated number of classes: (i) latent class logit specifications yield biased welfare estimates when estimated with a number of classes different from the true number; and (ii) in terms of accuracy, the most inaccurate estimates are yielded by the latent class logit with five classes. Both findings hold for both practices under study. The inaccuracy of the model with five classes is an important finding because empirical researchers analyzing the choices simulated in this study arguably would have preferred models with five classes.

Only the specification with the true number of classes consistently yields welfare estimates for which the 95% confidence interval includes the true value. These estimates, however, are very inaccurate. This inaccuracy is consequence of the values assumed for the cost parameter in one of the classes. The true value was imposed to be close to zero. In dealing with a close-to-zero negative coefficient, a latent class specification may yield positive estimates. Consequently, the welfare estimates have the opposite sign to what is expected. While positive estimates of cost parameters are outliers, the relative importance of these estimates is large.

These findings have a direct implication for empirical applications: researchers' own judgement in selecting number of classes very likely impacts the reliability of the welfare estimates. This implication does not necessarily translate to the recommendation of avoiding the inclusion of the researcher's own judgement. Particularly for the case of the practice of dismissing classes with zero price parameters. On one hand, a positive or zero price parameter presents both theoretical and empirical challenges to economists. On the other hand, this study shows that the exclusion of the class with an undistinguishable-from-zero price parameter results in biased welfare measures. The practice of eliminating small classes is, however, less defensible. Selection of number of classes remains a difficult issue because there is not a unambiguous likelihood-based criterion helping in the selection.

5. HOMOGENEOUS DISCRETE CHOICE EXPERIMENTS AND HETEROGENEOUS LOGIT MODELS: IMPLICATIONS FOR WELFARE ESTIMATES

5.1 Introduction

Applied research is increasingly relying on discrete choice experiments (DCE) to elicit stated preferences. Two components are essential in a DCE (Carson and Louviere, 2011): (i) a respondent is asked to make a discrete choice among hypothetical alternatives; and (ii) the alternatives are described in terms of strategically-manipulated attributes. Attributes are varied within- and/or between-respondents to avoid collinearity, and therefore, to obtain efficient estimates of preference parameters. Usually, respondents are asked to repeat the choice exercise, selecting from different choice sets each repetition.

Concurrently, applied research is increasingly relying on latent class logit (LCL) specifications to estimate preference parameters when unobserved heterogeneity in preferences is suspected. A LCL model incorporates unobserved heterogeneity by assuming preferences can be categorized into a finite number of classes. The task of a LCL is the identification of who belongs to what class.

Current practices include the estimation of a LCL on data obtained with a DCE (e.g. Broch and Vedel, 2012; Garrod et al., 2012; Kikulwe et al., 2011). This

practice, however, relies on a mismatch of assumptions about preferences: DCE are designed under the assumption of homogeneity in preferences, and latent class logit is carried out to infer heterogeneous preferences. Design strategies incorporating the possibility of continuous unobserved heterogeneity have recently been proposed (see Bliemer and Rose, 2010; Yu et al., 2009, 2011). These innovations have not reached the case in which unobserved heterogeneity is conceptualized as classes of preferences.

This chapter carries out Monte Carlo simulations to study the reliability of the welfare estimates obtained under the described mismatch. Specifically, welfare estimates are obtained from a LCL estimated on data gathered through an orthogonal fractional-factorial design that identifies only main effects. This design is the most commonly in empirical applications of LCL (see section 5.2.2). Choices among two generic alternatives and a status quo alternative are simulated according to a two-class utility-generating process. True utilities are assumed linear in attributes and income. Closely following the design implemented in a published application, alternatives are experimentally generated by manipulating 5 attributes. Three attributes have 2 levels. One attribute has 3 levels, and the price attribute has 4 levels. The attributes are combined according to a main effects orthogonal fractional-factorial design. Choice tasks are created through shifted pairing, and are orthogonally blocked. Choices are simulated for 300 pseudo-respondents. The number of pseudo-respondents is close to the median number of respondents used in published applications (see section 5.2.2). A latent class logit with two classes is carried out on the simulated choices, and welfare estimates are calculated. This

simulation exercise is repeated three times, varying the number of discrete choice tasks (3, 6 and 12).

Three willingness to pay (WTP) measures are under study: (i) WTP for a marginal change in an attribute; (ii) WTP for a non-marginal change in an alternative's attribute; and (iii) WTP to avoid the loss of an alternative. Reliability of the welfare estimates is evaluated in terms of (i) unbiasedness, i.e. whether the true value falls within the 95% confidence interval of the estimates; (ii) efficiency, i.e. which specification yields the smallest 95% confidence interval; and (iii) accuracy, i.e. how large is the average relative difference between the estimates and the true values according to the absolute value of the mean relative error.

The findings from the welfare comparisons are straightforward: welfare estimates are unbiased regardless the number of choice tasks, and their accuracy increases with the number of classes. However, for the case of estimates of WTP for marginal and non-marginal changes in an attribute, the improvement in accuracy does not prevent the presence of zero values in the 95% confidence intervals. The implications of these findings are discussed in section 5.5.

Few studies have researched the consequences on welfare estimates from violations to assumptions underlying the design of DCE (Carson and Louviere, 2011). The exceptions are Carlsson et al. (2003), Lusk and Norwood (2005) and Ferrini and Scarpa (2007). From these studies, only Ferrini and Scarpa (2007) have considered the case of heterogeneity in preferences, specializing in the case of continuous unobserved heterogeneity. They compare a variety of design strategies, from the most rudimentary fractional factorial strategy to the state-of-the-art Bayesian strategy

that incorporates a researcher's a priori beliefs. They find that strategies using poor a priori information perform poorly in comparison to the fractional factorial strategy. In contrast to Ferrini and Scarpa (2007), this paper studies discrete unobserved heterogeneity, controlling for the possibility that number of choice sets impacts welfare estimates.

5.2 *Current practices in discrete choice experiments*

This section describes current practices in applications that estimate a latent class logit (LCL) on data collected through discrete choice experiments (DCE). The lack of a common nomenclature complicates the description of DCE (Carson and Louviere, 2011). Thus the first task is to define the terms used in the description of the experimental designs. These definitions borrow heavily from Kuhfeld (2006, 2010), and Carson and Louviere (2011).

5.2.1 *Background*

Assume a researcher is interested in designing a DCE with two attributes that take three values each. The values an attribute can take are called levels. Designing a DCE consists of two sequential steps. In the first step, the researcher strategically combines the levels of the attributes. Each resulting combination is thought as an alternative that is described in terms of its attributes. In the second step, respondents are asked to choose among the designed alternatives. This decision is called discrete choice.¹ Researchers may ask respondents to make several discrete

¹ The adjective *discrete* is meant to (i) distinguish this type of experiments from the choice experiments used in some natural sciences; and (ii) emphasize the non-continuous nature of the dependent variable in the statistical analysis (Carson and Louviere, 2011).

choices by presenting them to different sets of alternatives. Each set of alternatives is called choice task.

A researcher designs a DCE with several choice tasks when the alternatives are too many to expect respondents can make a credible choice if faced to all the alternatives at once. For instance, two attributes with three levels produce a maximum of nine alternatives. If the researcher considers nine alternatives are too many, he/she may decide to present respondents with three choice tasks, each set composed by three alternatives. Sometimes, choice tasks are too many as well, and the researcher may decide to create sets of choice tasks. Each set of choice tasks is called block, and the action of creating blocks is called blocking.

Once the choice tasks have been defined, a researcher may add a status quo alternative to each choice task. A status quo alternative allows for the possibility that a respondent will not choose among the designed alternatives. The addition of this alternative is not innocuous. Recent evidence suggest that, depending on the number of alternatives offered in addition to the status quo, respondents may choose the status quo option regardless of whether the status quo is the utility maximizing alternative (see Zhang and Adamowicz, 2011, for details on this point). Also, a researcher may decide to label the alternatives presented to the respondents. Unlabeled alternatives are called generic alternatives. The strategies to generate alternatives differ depending on whether alternatives are labeled or generic (see Kuhfeld, 2006, 2010, for details on this point).

Therefore, a description of a DCE lists the strategies used to (i) combine the levels of the attributes, (ii) generate the choice tasks, and (iii) block the choice tasks.

Also, a description of a choice task clarifies whether alternatives have been labeled and whether a status quo alternative has been added.

The strategies to combine the levels of attributes are called factorial designs. The term *factorial* derives from the practice in experimental design theory of calling factors to the attributes (Kuhfeld, 2006). Factorial designs are either full-factorial designs or fractional-factorial designs. A full-factorial design carries out all possible combinations of the levels of the attributes. Full-factorial designs allow for estimation of main effects and all possible interaction effects. A main effect is the impact from one attribute on the discrete choice. An interaction effect is the impact from the interaction of attributes on the discrete choice. If the researcher is manipulating two attributes, the maximum interaction effects are called two-way interaction effects because they imply the interaction of two attributes. Three-way interaction effects are possible when three attributes are being manipulated, and so on.

Fractional-factorial designs select a subset of combinations from the full-factorial design. The advantage from these designs is that a smaller number of alternatives are presented to the respondent. The disadvantage is that some effects become confounded. This disadvantage is not necessarily an issue because a researcher (i) may not be interested in estimating all possible effects, and (ii) has access to an array of strategies that allow for the estimation of the effects he/she may be interested on (e.g. main effects only, main effects and a subset of interaction effects, etc.).

Fractional-factorial designs are either orthogonal or non-orthogonal. An orthogonal fractional-factorial design (OFFD) is both balanced and orthogonal (Kuh-

feld, 2006).² A design is balanced when each level occurs equally often within each attribute. A design is orthogonal when every pair of levels occurs equally often across all pairs of attributes. Orthogonality captures the relationship among attributes, and balance captures the relationship between attributes and the intercept. The intercept is orthogonal to each attribute when a design is balanced.

With respect to strategies to create choice tasks, the simplest strategy is randomly pairing alternatives yielded by a factorial design. This strategy is called random pairing. Shifted (or cyclical) pairing is a more sophisticated strategy. This strategy uses the alternatives from an OFFD as seed alternatives. As first step, each alternative is allocated to different choice tasks. The second step consists in adding a second alternative in each choice set by cyclically adding alternatives based on the attribute levels. The attribute level in the new alternative is the next higher level to the one already included in the choice task. When the highest level is attained, the attribute level is set to its lowest level. The cycling is repeated as many times as number of alternatives in the choice task.³

Blocking of choice tasks can be reached by randomly allocating choice tasks in the corresponding blocks. This strategy is called random blocking. A second strategy, called orthogonal blocking, generates a blocking factor that is orthogonal

² A source of ambiguity that may generate confusion is implicit in this terminology. This ambiguity stems from the convention of naming orthogonal fractional-factorial design to a fractional-factorial design that is both balanced and orthogonal. Strictly speaking a design that is both balanced and orthogonal is called orthogonal array. Implementation of unbalanced, orthogonal fractional-factorial designs is discouraged (Kuhfeld, 2006), and therefore seldom used. Thus in practice the distinction between unbalanced, orthogonal fractional-factorial designs and balance, orthogonal fractional-factorial designs (or orthogonal arrays) is unnecessary. See Kuhfeld (2006, 2010) for details about these terms.

³ Carlsson et al. (2003), Ferrini and Scarpa (2007), and Kuhfeld (2006, 2010) describe additional strategies to generate choice tasks.

to all of the attributes of all of the alternatives.

5.2.2 *Current practices*

Table 5.1 describes how DCE have been designed in applications that estimate a LCL on data collected through a DCE. The second column shows how many alternatives have been included in the choice sets. Fourteen out of 19 applications (74%) have included three alternatives, one of which is a status quo alternative. All applications but Boxall and Adamowicz (2002) have used unlabeled alternatives. The third column shows the number of choice tasks that the respondent is presented to. With a minimum of 4 and a maximum of 16, the most common numbers of choice tasks are 6 and 8. Around 53% of the applications have used either 6 or 8 choice sets.

The fourth column in table 5.1 shows the strategy used to combine the attributes. Twelve applications (63%) have used an orthogonal fractional-factorial design (OFFD). Four applications (21%) have used non-orthogonal fractional-factorial design (NFFD). One application has used a full-factorial design (FFD), and two applications have not specified what type of fractional-factorial design has been used (u-FFD). Fourteen applications (74%) have used a design that identifies only main effects. Three applications have used a design that identifies main and two-way effects, and two applications have not specified the effects.

The fifth column in table 5.1 presents the strategy to generate choice tasks. Ten applications (53%) do not report how the choice tasks are created. Five (26%) applications randomly pair alternatives, and four applications (21%) use shifted

pairing.

The sixth column in table 5.1 presents the strategy to block choice tasks. Eight applications (42%) do not report the strategy to block choice tasks. Three applications (16%) do not use blocking. Six applications (32%) use random blocking. Two applications (10%) use orthogonal blocking.

Tab. 5.1: Current practices in applications that estimate a latent class logit on data collected through discrete choice experiments

Paper	Alternatives ^a	Choice tasks	Strategy to combine attributes ^b	Strategy to generate choice tasks	Strategy to block choice tasks	Attributes / levels	Levels of price attribute	Sample size
Boxall and Adamowicz (2002)	5 + sq	8	OFFD	unspecified	unspecified	5 / $4^4 \times 2$	4	620
Scarpa et al. (2003)	2 + sq	6	OFFD	random	random	5 / $2^3 \times 3 \times 4$	4	300
Birol et al. (2006)	2 + sq	8	OFFD	unspecified	random	5 / $2^3 \times 4^2$	4	407
Milon and Scrogin (2006)	2	7	OFFD	unspecified	unspecified	6 / 3^6	3	240
Milon and Scrogin (2006)	2	7	OFFD	unspecified	unspecified	6 / 3^6	3	240
Ouma et al. (2007)	2 + sq	12	NFFD	unspecified	no blocking	8 / $2^5 \times 3^3$	3	253
Ouma et al. (2007)	2 + sq	11	NFFD	unspecified	no blocking	7 / $2^4 \times 3^3$	3	253
Ruto et al. (2008)	2 + sq	8	OFFD	random	random	5 / $2^3 \times 3^2$	3	311
Birol et al. (2009)	2 + sq	6	OFFD	random	random	5 / $3^2 \times 2^2 \times 5$	5	420
Colombo et al. (2009)	2 + sq	6	OFFD	shifted	unspecified	6 / $3^5 \times 6$	6	300
Beharry-Borg and Scarpa (2010)	2 + sq	8	OFFD	shifted	orthogonal	6 / 3^6	3	86
Beharry-Borg and Scarpa (2010)	2 + sq	9	OFFD	shifted	orthogonal	9 / 3^9	3	198
Brouwer et al. (2010)	2 + sq	4	OFFD	unspecified	unspecified	5 / $2^2 \times 3^2 \times 6$	6	619
Kosenius (2010)	2 + sq	6	NFFD ^c	unspecified	unspecified	5 / $3^4 \times 7$	7	726
Kikulwe et al. (2011)	2 + sq	16	u-FFD	unspecified	no blocking	4 / $3^2 \times 2 \times 6$	6	421
van Putten et al. (2011)	2 + sq	8	u-FFD ^d	shifted	unspecified	5 / $2^4 \times 4$	4	132
Broch and Vedel (2012)	2 + sq	6	NFFD ^d	unspecified	unspecified	4 / $3^3 \times 6$	6	853
Chung et al. (2012)	3 + sq	10	FF ^d	random	random	7 / $2^3 \times 3^2 \times 5 \times 11$	11	873
Garrod et al. (2012)	2	4	OFFD ^c	random	random	5 / 2^5	NA	1,273

^a sq: status quo alternative.

^b OFFD: orthogonal fractional-factorial design; NFFD: non-orthogonal fractional-factorial design; FFD: full-factorial design; u-FFD: unspecified fractional-factorial design.

Applications identify only main effects, with the exception of ^c do not specify the identified effects, and

^d identify main and two-way effects.

The seventh column in table 5.1 presents the number of attributes and levels for each attribute in the experiment. Nine applications (47%) have manipulated 5 attributes. Four applications (21%) have used 6 attributes. Four and seven attributes have been manipulated in two applications each. Eight and nine attributes have been manipulated in one application each. The most recurrent numbers of levels are 2 and 3. Manipulation of the price attribute is of special interest for economists. The eighth column presents the number of levels of the price attribute. Seven applications (37%) manipulate 3 levels. Four applications (21%) manipulate 4 levels. Four applications (21%) manipulate 6 levels. Five, seven and eleven levels are manipulated in one application each. Garrod et al. (2012) do not include a price attribute in their DCE.

The last column in table 5.1 presents the sample size. With a minimum of 86 and a maximum of 1,273, the average sample size is 449. The median is 311. There are three modes, repeated twice each: 240, 253, and 300.

Thus most studies use orthogonal fractional-factorial designs that identify main effects only. From the applications that specify the method used to generate choice tasks, random pairing and shifted pairing account for around 50% each. From the applications that specify the method used to block choice tasks, random blocking is the most common strategy. Most applications manipulate either 5 or 6 attributes, varying the price attribute across 3, 4 or 6 levels. A large majority of studies have used unlabeled alternatives and have added a status quo alternative to the designed strategies. The average number of respondents is 449, with half of the applications interviewing 311 or less respondents. Around half of the applications

have presented respondents to either 6 or 8 choice tasks.

This description of current practices closely resembles the one presented by Ferrini and Scarpa (2007). Focusing on a set of applications published in 2005 or before, Ferrini and Scarpa (2007) find the majority of applications use orthogonal main effects fractional-factorial designs, add an status quo alternative to unlabeled alternatives, manipulate 5 or 6 attributes, present respondents to 4, 6 or 8 choice tasks, and half of the applications interview 350 or less respondents.⁴

5.2.3 *The issue*

The most common design strategies used in the field, as described in the previous section, rely on the assumption that the true model (i) deals with a continuous dependent variable, (ii) is linear in preference parameters, and (iii) captures homogeneous preferences (see Carlsson et al., 2003, for details.). The applications reviewed in this paper obtain parameter estimates through latent class logit models that (i) deal with a discrete dependent variable, (ii) are non-linear in preference parameters, and (iii) search for heterogeneity in preferences.

Thus three mismatches are implicit in the current practices in environmental and resource economics: (i) models that deal with discrete dependent variables are used on data generated under the assumption that the true model deals with a continuous dependent variable; (ii) non-linear models are used to estimate preference parameters from data generated under the assumption that the true model is linear

⁴ Ferrini and Scarpa (2007) also notice that many applications of DCE in environmental and resource economics fail in providing a complete description of the strategy to generate the DCE. The review in this paper suggests this practice remains common in the field.

in parameters (see Carlsson et al., 2003, for reasons of why these mismatches are of potential concern.); and (iii) data is gathered assuming that respondents have homogeneous preferences but the econometric specification inferring these preferences searches for heterogeneity.

This paper seeks for implications from the third mismatch in terms of welfare estimation. Previous studies have focused their attention on the first two mismatches, with findings suggesting that optimal designs for linear models work fine when used to estimate conditional logit models (see Carlsson et al., 2003; Kuhfeld et al., 1994; Lazari and Anderson, 1994; Lusk and Norwood, 2005).

Ferrini and Scarpa (2007) have studied implications for welfare estimation from estimating a mixed logit on a DCE derived under homogeneity assumption. They compare a variety of design strategies, from the most rudimentary fractional factorial strategy to the state-of-the-art Bayesian strategy that incorporates a researcher's a priori beliefs. Their findings suggests strategies using poor a priori information perform poorly in comparison to the fractional factorial strategy.

While Ferrini and Scarpa (2007) specialize in the case of continuous unobserved heterogeneity, this paper focuses on discrete unobserved heterogeneity, checking for the possibility that number of choice tasks impacts welfare estimates.

5.3 *Simulation strategy*

This Monte Carlo simulation is designed to evaluate whether the mismatch of assumptions about heterogeneity of preferences underlying DCE and LCL models

matters for welfare estimation. Choices among two generic alternatives and a status quo alternative are simulated according to a two-class utility-generating process. Following current practices in the field, alternatives are experimentally generated by manipulating 5 attributes. Three attributes have 2 levels. One attribute has 3 levels, and the price attribute has 4 levels. The attributes are combined according to a main effects orthogonal fractional-factorial design (OFFD). Choice tasks are created through shifted pairing, and are orthogonally blocked. Choices are simulated for 300 pseudo-respondents. This number of pseudo-respondents is close to the median number observed in empirical applications. A latent class logit with two classes is carried out on the simulated choices, and welfare estimates are calculated. Average WTP estimates over the Monte Carlo replications are compared against average true values. This simulation exercise is repeated three times, varying the number of discrete choice tasks (3, 6 and 12).

A general description of the simulation strategy is presented in figure 5.1. In step zero, alternatives are experimentally generated, and true indirect utilities are calculated. The label *step zero* is intended to highlight the immutability of both the experimentally generated alternatives, and the true utilities. That is, alternatives designed with a DCE and true utilities are kept fixed through the Monte Carlo simulations.

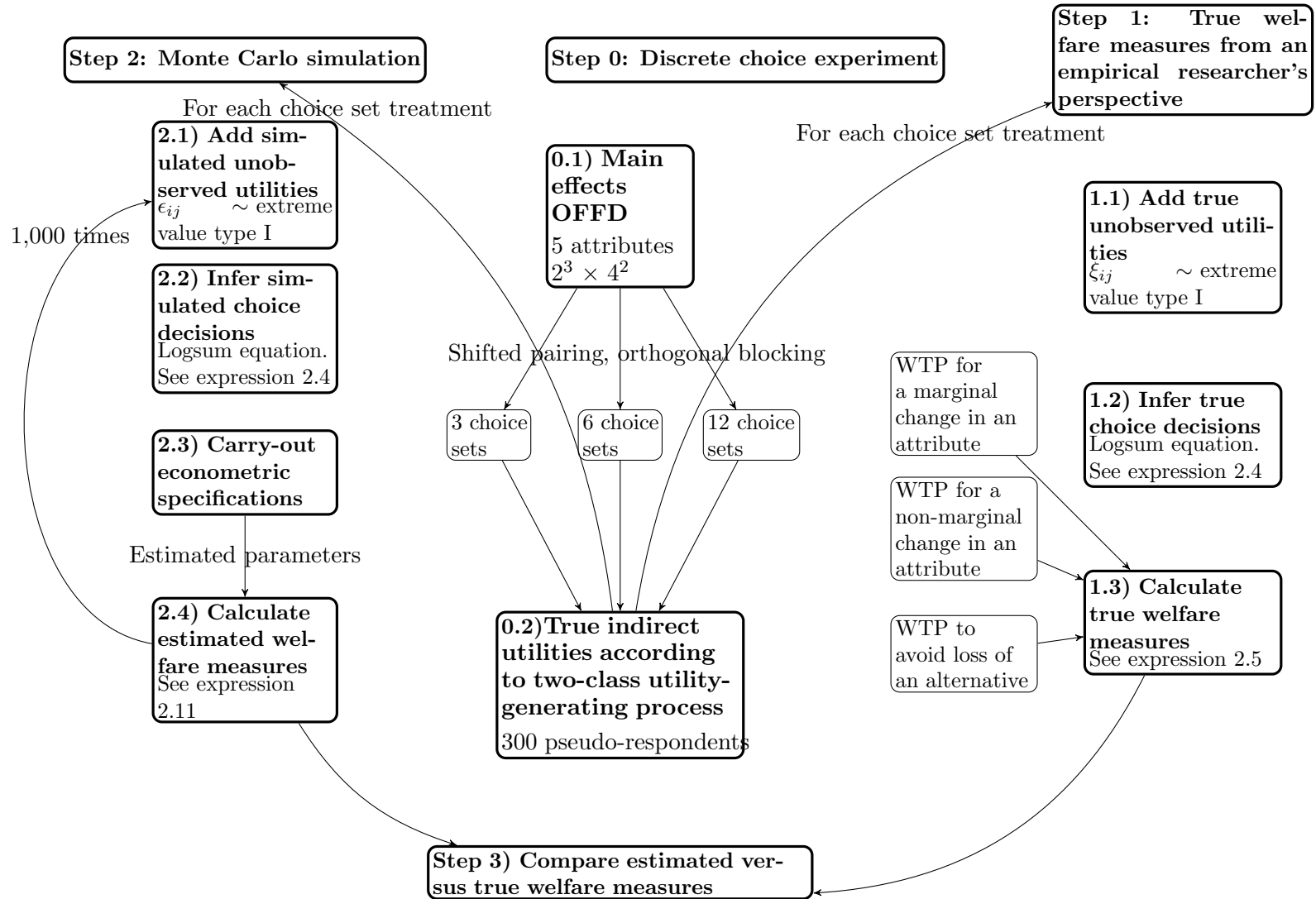


Fig. 5.1: Steps of Monte Carlo simulation studying robustness of welfare estimates to number of choice tasks

Step one calculates the true WTP measures from the simulated observed utilities. Observed utilities result from adding a type I extreme value error term to the true utilities. This error term is labeled *true error term* because it is used in generating the utilities defining the true WTP measures.

Step two consists in carrying out the Monte Carlo simulation. The goal of each of the 1,000 replications is the estimation of WTP measures. Within each replication, a Type I error term is added to the true utilities, choice decisions are simulated and used to inform a latent class logit with two classes. Estimates of preference parameters obtained through the latent class logit are used in the estimation of WTP measures.

Step three compares estimated WTP measures against true WTP, and evaluates the performance of each econometric model in retrieving the true WTP. Performance is evaluated in terms of (i) unbiasedness, (ii) efficiency and (iii) accuracy. An estimate categorized as unbiased if its 95% confidence interval includes the true value. The most efficient estimate is the one with the smallest 95% confidence interval. Efficiency comparison is restricted to unbiased estimates. Accuracy refers to the magnitude of the relative difference between the estimates and the true values, measured as the absolute value of the mean relative error.

5.3.1 *Discrete choice experiment*

In order to carry out a policy-relevant simulation exercise, this study implements a DCE closely resembling the DCE implemented by Birol et al. (2006) who carry out a non-market valuation exercise of a wetland's attributes, and subsequently

a cost-benefit analysis exercise of different management scenarios.

The five manipulated attributes are expressed in terms of variations in the conditions of a wetland with respect to current levels. These attributes and their levels are described in table 5.2. Biodiversity, open water surface area, and research and education are characterized in terms of two levels — high and low. Number of farmers re-trained in environmentally-friendly activities vary according to three levels: 30, 75, and 150. The one-time payment to fund wetland restoration takes 4 values: \$3, \$10, \$40, and \$80.

The five attributes are combined according to a main effects orthogonal fractional-factorial design. Choice tasks are created through shifted pairing, orthogonally blocked, and include two generic alternatives and a status quo alternative. Three versions of the DCE are simulated. The first version presents 3 choice tasks. The second version presents 6 choice tasks. The third version presents 12 choice tasks.

True values of the preferences for the manipulated attributes, by class, are listed in table 5.3. These values closely resemble the estimates obtained from two-class latent class logit specification by Birol et al. (2006). Respondents in class 1 are assumed to represent 70% of the sample size, and 30% are assumed belonging to class 2.

5.3.2 True and estimated WTP measures

Three WTP measures are of interest (i) WTP for a marginal change in Q (MWTP); (ii) WTP for a 25% improvement in Q of alternative 1 (WTPA); and (iii)

Tab. 5.2: Attributes and levels manipulated in the discrete choice experiment.

Attribute	Description	Levels
Biodiversity (B)	Population levels of of different species of plants and animals, the number of different habitats and their size.	Low: deterioration from current level High: a 10% increase in population and size of habitats
Open water surface area (O)	Surface area of the lake that remains uncovered by reed beds.	Low: a 20% decrease from current level High: Increase from current level to 60%
Research and education (R)	Educational, research, and cultural information that may be derived from the existence of the wetland.	Low: deterioration from current level High: improvement from current level by providing better facilities
Re-training (T)	Number of local farmers re-trained in environmentally-friendly activities.	30, 75, 150
Payment (P)	A one-time payment labeled to fund wetland restoration.	\$3, \$10, \$40, and \$80

Tab. 5.3: True preferences for manipulated attributes, by class.

Attribute	Parameter	True values	
		Class 1	Class 2
Status quo	α_{sq}	2.400	-1.200
Biodiversity (B)	α_b	0.270	0.000
Open water surface area (O)	α_o	0.160	0.300
Research and education (R)	α_r	0.140	0.000
Re-training (T)	α_t	0.003	0.003
Payment (P)	α_p	-0.015	-0.045

willingness to pay to avoid the loss of alternative 2 (WTPL).

Average WTP estimates over the Monte Carlo replications are compared against true WTP. The estimated WTP result from averaging WTP estimates over the 1,000 Monte Carlo replications. In each Monte Carlo replication (i) WTP estimates are calculated as explained in section 2.3 for each individual in the pseudo-dataset; and (ii) average WTP over the individuals are obtained, and stored. The true WTP measures result from averaging the individual welfare measures over the individuals in the pseudo-dataset.

5.4 Results

True WTP measures are compared against welfare estimates obtained through a latent class specifications with two classes. Welfare comparisons are carried out three times, varying the number of choice tasks presented to the pseudo-individuals (3, 6, 12). Performance is evaluated in terms of (i) unbiasedness, (ii) efficiency and (iii) accuracy.

Before discussing the measures of performance, figures 5.2, 5.3, and 5.4 are discussed to highlight the main findings of this simulation study. Figure 5.2 presents the 95% confidence interval of the WTP for a 25% improvement in T of alternative 1(WTPA) by choice tasks. The vertical straight line presents the true WTPA. Three features are highlighted: (i) the three confidence intervals include the true WTPA; (ii) the three distributions include the value zero; and (iii) the confidence interval becomes smaller the larger the number of choice tasks. These features are interpreted

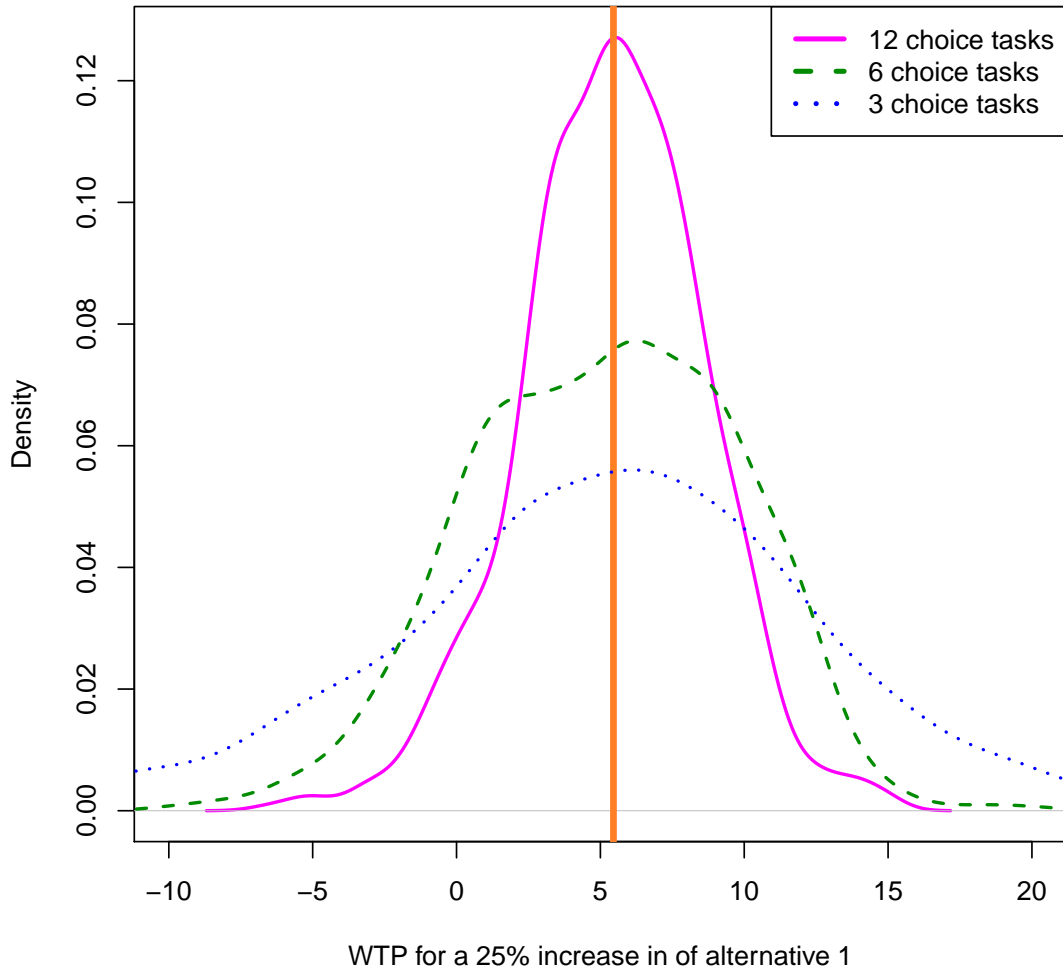


Fig. 5.2: 95% confidence interval of WTP for 25% improvement in T of alternative 1 by number of choice tasks)

as evidence that WTPA estimates are unbiased regardless of the number of choice tasks. However, the large confidence intervals provoke that the null hypothesis that the WTPA estimate is equal to zero can not be rejected. Clearly, the larger the number of choice tasks, the smaller the confidence intervals. A very similar story can be told for the case of the WTP for a marginal change in T, as shown in figure 5.3.

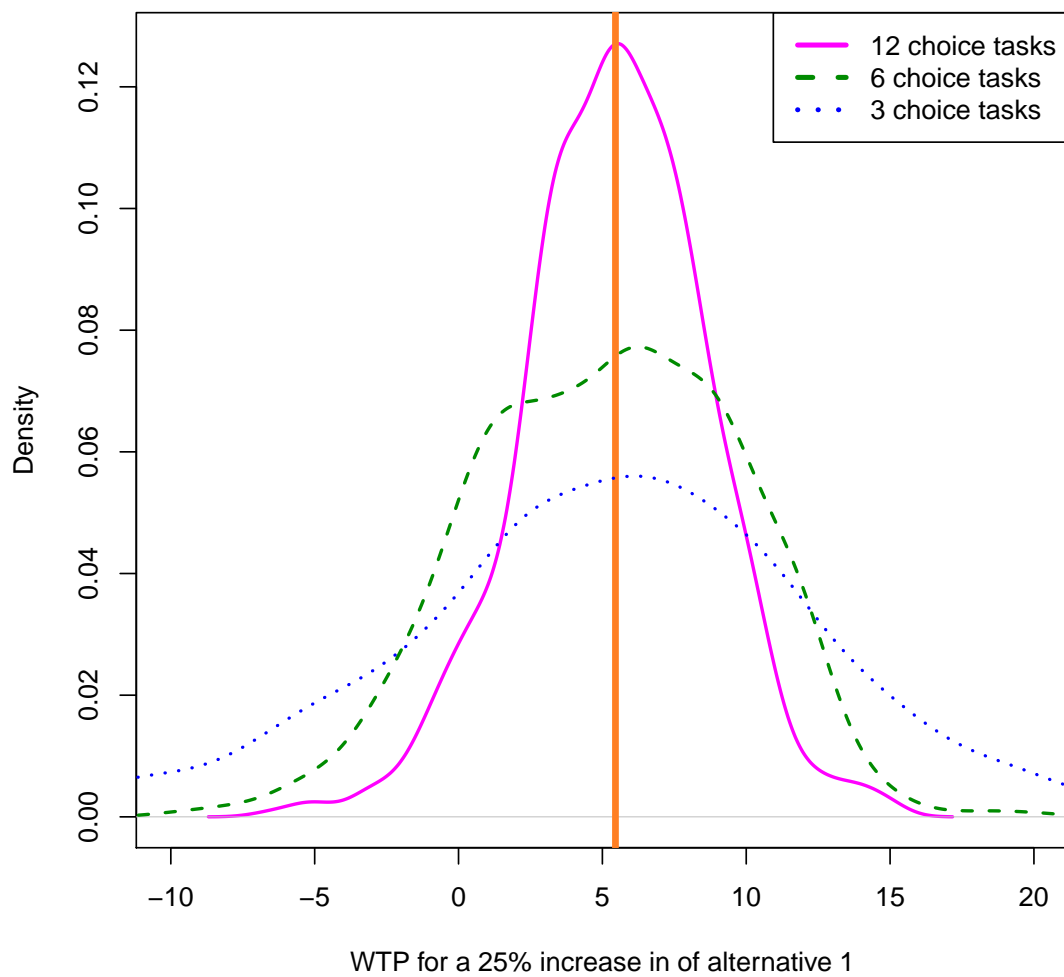


Fig. 5.3: 95% confidence interval of WTP for marginal change in T by number of choice tasks)

An slightly different story can be told from figure 5.4 which the 95% confidence intervals for the WTP to avoid the loss of alternative 2. In this case, the unbiasedness regardless of the number of choice tasks can also be observed. In contrast to figures 5.2 and 5.3, the three confidence intervals do not include the value zero. That is, estimates of WTP to avoid the loss of an alternative are more accurate than estimates for marginal and non-marginal changes in an attribute.

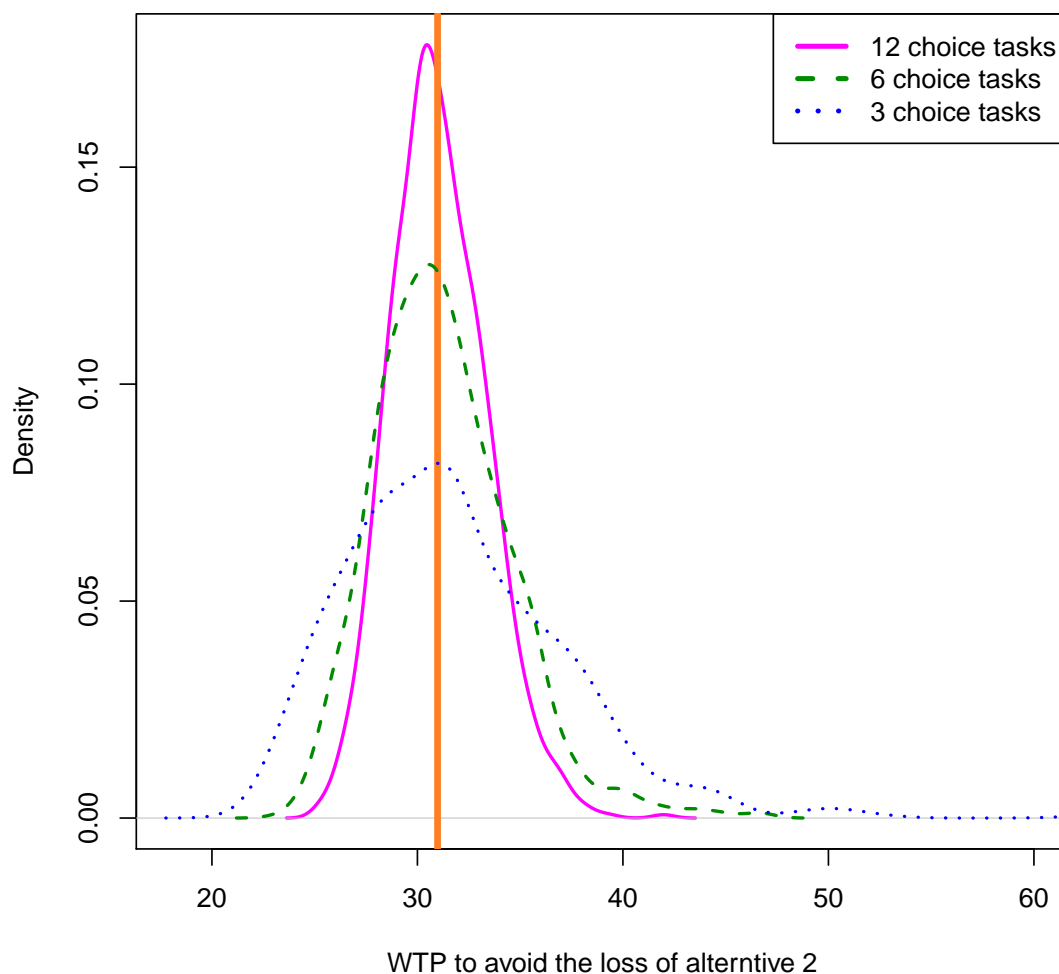


Fig. 5.4: 95% confidence interval of WTP to avoid the loss of alternative 2 by number of choice tasks)

Tab. 5.4: 95% confidence interval includes true value, and average of absolute value of relative errors (AARE).

	Number of choice sets		
	Three	Six	Twelve
	95% confidence interval includes true value ^a		
WTP from 25% increase in T of alternative 1 (WTPA)	✓	✓	✓ *
WTP to avoid loss of alternative 2 (WTPL)	✓	✓	✓ *
Marginal willingness to pay for T (MWTP)	✓	✓	✓ *
	Average of absolute value of relative errors (AARE)		
WTP from 25% increase in T of alternative 1 (WTPA)	1.163	0.691	0.458
WTP to avoid loss of alternative 2 (WTPL)	0.760	0.468	0.333
Marginal willingness to pay for T (MWTP)	0.032	0.019	0.013

^a ✓: true value is included; +: true value is smaller than lower bound; -: true value is larger than upper bound. ^b *: Smallest 95% confidence interval among unbiased estimates. ^c Measured as $M^{-1} \sum |(WTP - \hat{WTP})/WTP|$, where M is the number of Monte Carlo observations, i.e. 1,000.

The top panel of table 5.4 summarizes results in terms of unbiasedness. A check mark symbol (✓) indicates the 95% confidence interval of the welfare estimate includes the true value. If this is the case, the estimate is considered unbiased. A plus symbol (+) is reported if the true value is smaller than the lower bound of the 95% confidence interval. A minus symbol (−) is reported if the true value is larger than the upper bound of the 95% confidence interval.

According to the top panel of table 5.4, welfare estimates are unbiased under the three choice tasks scenarios. This finding holds for the three WTP measures under study. The welfare estimates with the smallest confidence interval are consistently yielded by the case in which respondents are presented to 12 choice tasks.

The bottom panel of table 5.4 summarizes results in terms of accuracy, measured as the average of the absolute value of the relative errors (AARE), i.e.

$$AARE = M^{-1} \sum_{m=1}^M \left| \frac{(WTP - \hat{WTP})}{WTP} \right| \quad (5.1)$$

Three general patterns are highlighted from the bottom panel of table 5.4:

(i) under the three choice tasks scenarios, estimates of marginal WTP are more accurate than estimates of non-marginal measures; (ii) under the three choice tasks scenarios, estimates of WTP for a non-marginal increase in T are the least accurate; and (iii) accuracy of the three welfare estimates increases with the number of choice tasks.

Tab. 5.5: Willingness to pay measures, calculated with parameters estimated through a latent class logit with two classes

	True value	Number of choice tasks		
		Three	Six	Twelve
WTP from 25% increase in T for alternative 1 (WTPA)				
Mean	-5.46	-4.49	-5.13	-5.42
5%	-	-16.89	-12.03	-10.33
95%	-	10.11	2.41	0.09
WTP to avoid loss of alternative 2 (WTPL)				
Mean	-30.98	-31.85	-31.23	-31.06
5%	-	-40.80	-36.85	-34.92
95%	-	-24.53	-26.37	-27.61
Marginal willingness to pay for T (MWTP)				
Mean	0.16	0.17	0.16	0.16
5%	-	-0.19	-0.04	0.02
95%	-	0.56	0.38	0.31

So far, results would suggest good news. That is, welfare estimates are unbiased regardless the number of choice tasks presented to the pseudo-respondents. In

addition, accuracy increases with the number of choice tasks.

However, table 5.5 presents evidence suggesting a drawback. Table 5.5 shows true and estimated welfare measures and their corresponding 95% confidence intervals, by choice task scenario. While the true value is included in all 95% confidence intervals, the value zero is also included in these intervals. This is the case for estimates of WTP for a non-marginal increase in T (WTPA), regardless the number of choice tasks. This is also the case for the marginal WTP (MWTP), under the three choice tasks and six choice tasks scenarios.

Thus a more complete picture of welfare estimates can be depicted as follows: welfare estimates are unbiased regardless the number of choice tasks, and their accuracy increases with the number of classes. However, for the case of WTPA, 12 choice tasks are not enough to obtain estimates with 95% confidence intervals that do not include the zero value. In similar vain, for the case of MWTP, 12 choice tasks are just not enough to obtain estimates with 95% confidence intervals that do not include the zero value.

5.5 *Conclusions and discussion*

Current practices in applied research include the estimation of latent class logit specifications on data gathered through with discrete choice experiments. The literature review presented in this chapter corroborates that this practice implies a mismatch: most empirical applications design discrete choice experiments that assume homogeneity in preferences, and infer heterogeneous preference with the use

of latent class logit models.

This study designs a Monte Carlo simulation to evaluate whether the mismatch of assumptions about heterogeneity of preferences matters for welfare estimation. This simulation exercise tests for the possibility that the number of choice tasks impact the ability of latent class logit to retrieve welfare measures. Three willingness to pay (WTP) measures are under study: (i) WTP for a marginal change in an attribute; (ii) WTP for a non-marginal change in an alternative's attribute; and (iii) willingness to pay to avoid the loss of an alternative.

The findings from the welfare comparisons are straightforward: welfare estimates are unbiased regardless the number of choice tasks, and their accuracy increases with the number of classes. These findings are not surprising because increasing the number of choice tasks is simply increasing the information available to the statistical model. Given that estimates are unbiased, the increase in information increases their efficiency. The increase in efficiency translates to improvements in accuracy.

Despite the obvious nature of these findings, there is a nuance that turns out to have implications for the empirical literature. This nuance refers to the finding that estimates of WTP for non-marginal changes include the value zero in their 95% confidence intervals under the three choice tasks scenarios. Similarly, estimates of WTP for non-marginal changes just barely exclude the value zero for the scenario with 12 choice tasks. These findings imply that, for the specific simulated scenario studied here, more than 12 choice tasks need to be used to obtain estimates statistically different from zero.

These findings suggest that empirical applications estimating WTP for marginal and non-marginal changes may want to consider gathering more information than they usually do. According to the literature review in this chapter, which finds evidence closely resembling a literature review focused on a different subset of studies (see Ferrini and Scarpa, 2007), around half of the studies analyze answers from 300 respondents, and design experiments with 6 or 8 choice tasks. The Monte Carlo simulations in this study assume 300 pseudo-individuals facing 3, 6, and 12 choice tasks. Thus the results from this study arguably suggest that welfare estimates in half of the reviewed applications may incorrectly be statistically undistinguishable from zero.

Empirical researchers can decide between increasing the number of respondents or increasing the number of choice tasks. The selection is not easy. On one hand, financial justifications are usually behind the decision of using several choice tasks for each individual. On the other hand, respondents may not be willing to answer many choice tasks. In taking this decision, empirical researchers may want to consider the recent evidence suggesting that respondents can answer up to 16 or 17 choice tasks without showing symptoms of tiredness (see Bech et al., 2011; Hess et al., 2012). These recent results may potentially be context-dependent. Thus researchers still need to pay attention to designing discrete choice experiments that minimize the mental burden to the respondent.

This study assumes the true underlying heterogeneity in preferences is represented by only two latent classes. Arguably, the increase in unobserved heterogeneity, instrumentalized as a larger number of classes, increases the number of choice

tasks needed to obtain efficient estimates. This possibility stresses the relevance that empirical researchers have in mind at the moment of designing the discrete choice experiment how many classes they expect to estimate.

Researchers may also consider the design strategies that incorporate a priori information about the preferences of respondents in the design of the discrete choice experiment. Some of these strategies do not rely on the assumption of homogeneity in preferences (e.g. Bliemer and Rose, 2010; Ferrini and Scarpa, 2007; Yu et al., 2009, 2011). However, some of these strategies have been proven to be less robust to incorrect a priori information than the simple fractional-factorial designs (e.g. Ferrini and Scarpa, 2007).

6. GENERAL CONCLUSIONS

This dissertation has carried out a series of Monte Carlo simulations seeking the implications for welfare estimates from three research practices commonly implemented in studies that incorporate unobserved preference heterogeneity in discrete choice models. The most popular strategies to incorporate unobserved heterogeneity are the mixed logit and the latent class logit. Thus the focus had been on learning the reliability of welfare measures obtained through mixed logit and latent class logit.

Implications for welfare have been studied under three research practices: (i) the comparison of welfare estimates from a conditional logit versus welfare estimates from a mixed logit or a latent class logit; (ii) the use of researcher's own judgement when selecting the number of classes of a latent class logit specification; and (iii) the estimation of latent class logit specifications on data gathered through discrete choice experiments that rely on the assumption of homogeneity in preferences.

Through the three empirical chapters of this dissertation, three welfare measures have been studied: (i) willingness to pay for a marginal change in an attribute; (ii) willingness to pay for a non-marginal change in an alternative's attribute; and (iii) willingness to pay to avoid the loss of an alternative. Reliability of welfare measures have been measured in terms of biasedness, efficiency and accuracy.

Chapter 3 compares welfare measures across conditional logit, mixed logit, and latent class logit. The practice of comparing welfare estimates is widely accepted in the field. However, this chapter shows that the comparison of welfare estimates across econometric specifications seems unable to provide reliable information about the differences in welfare estimates resulting from controlling for unobserved heterogeneity. The reason behind this finding is the large standard errors from estimates obtained through mixed logit and latent class logit. This result leaves us with more questions than answers: how should an empirical researcher judge the relative magnitude of welfare estimates from conditional logit with respect to estimates from mixed logit or latent class logit? Based on the results of this chapter, the researcher cannot infer whether incorporation of unobserved heterogeneity is actually producing an improvement in welfare estimation. However, this improvement in welfare estimation is somehow an implicit assumption/justification to estimate mixed logit and/or latent class logit instead of a conditional logit.

Chapter 4 studies the reliability of welfare estimates obtained under scenarios for which the empirical researcher would arguably choose the number of classes based on his/her own judgement. Robustness of welfare estimates is studied under two scenarios: (i) a class contains a close-to-zero price parameter, and (ii) a class is relatively small. This chapter shows that welfare estimates are sensitive to the number of classes in the latent class logit. Models with a number of classes smaller than the true number tend to yield biased and inaccurate estimates. The estimates from the latent class with the true number of classes always yield unbiased estimates but their accuracy may be worse than models with an smaller number of classes.

These findings have a straightforward implication: researchers need to be careful in including their own judgement when selecting the number of classes in a latent class model. These findings, however, do not translate directly to the recommendation of avoiding the inclusion of the researcher's own judgement. Specially for the case of the practice of dismissing classes with zero price parameters. On one hand, a positive or zero price parameter presents both theoretical and empirical challenges to economists. On the other hand, this study shows that the exclusion of the class with an undistinguishable-from-zero price parameter results in biased welfare measures. The practice of eliminating small classes is, however, less defensible. Selection of number of classes remains a difficult issue because there is not an unambiguous likelihood-based criterion helping in the selection.

Chapter 5 studies the reliability of welfare estimates under a common mismatch in the literature implementing discrete choice experiments: discrete choice experiments are designed under the assumption of homogeneity in preferences, and latent class logit is carried out to infer heterogeneous preferences. Specifically, the Monte Carlo simulations carried out in chapter 5 study the reliability of the welfare estimates obtained from using in the same study a latent class logit and an orthogonal fractional-factorial design that identifies only main effects. This simulation tests whether number of choice tasks impact the reliability of welfare estimates. The findings from the welfare comparisons are straightforward: welfare estimates are unbiased regardless the number of choice tasks, and their accuracy increases with the number of classes. These results are not surprising. Despite their obvious nature, there is a nuance that turns out to have implications for the empirical lit-

erature: willingness to pay for marginal and non-marginal changes are not different from zero under the three choice tasks scenarios. This finding suggests that empirical applications estimating WTP for marginal and non-marginal changes may want to consider gathering more information than they usually do.

A conclusion can be drawn from the three chapters: mixed logit and latent class yield inefficient and inaccurate welfare estimates under both revealed and stated preferences. This conclusion holds for welfare measures of both marginal and non-marginal changes. Evidence to conclude that inaccuracy and inefficiency hold across type of preferences is provided by the meta-analysis carried out in chapter 3, and is also suggested by the large inaccuracy and inefficiency of estimates across simulated pseudo-datasets. Chapters 3 and 4 simulate data resembling revealed preferences, and chapter 5 simulates data resembling stated preferences. Relative efficiency and accuracy remain poor across both types of simulated preferences.

APPENDIX

A. DESCRIPTION OF EMPIRICAL APPLICATIONS
REPORTING WELFARE ESTIMATES FROM CONDITIONAL
LOGIT, MIXED LOGIT AND LATENT CLASS LOGIT

Tab. A.1: Description of empirical applications reporting welfare estimates from conditional logit, mixed logit and latent class logit

Paper	Type of application		Population	Country	Surveying method	Year data was collected	Elicited preferences	Alternatives/ individuals/ choice sets ^a / choices ^b
Studies reporting welfare estimates as by-product								
Train (1998)	Recreational mand	de-	Montana anglers	USA	Telephone	1993	Revealed	59/258/?/?
McConnell and Tseng (1999)	Recreational mand	de-	Visitors of 10 public beaches on western shore of Maryland	USA	Face-to-face on-site	1984	Revealed	11/388/?/?
Breffle and Morey (2000)	Recreational mand	de-	Maine anglers with license	USA	Mail	1988	Revealed	9/145/100/14,500
Boxall and Adamowicz (2002)	Recreational mand	de-	Visitors of 5 parks in eastern Manitoba, western Ontario and northern Minnesota	Canada and USA	Mail	1995	Stated	6/620/8/4,892
Provencher et al. (2002)	Recreational demand	de-	Wisconsin Lake Michigan anglers	USA	Face-to-face on-site	1996	Revealed	2/192/182/34,944
Carlsson et al. (2003)	Non-market valuation of wetland ecosystem	valuation	Residents of Staffanstorp, age 18 to 78 (13, 000 people)	Sweden	Mail	2001	Stated	3/468/4/1,717
Nahuelhual et al. (2004)	Non-market valuation of public open space	valuation	Residents of Jackson Hole, Wyoming	USA	Mail	?	Stated	2/308/3/924
Sillano and Ortúzar (2005)	Non-market valuation of atmospheric nuisance reductions	valuation	Residents of Santiago	Chile	Face-to-face in-house	2001	Stated	2/75/9/648
Birol et al. (2006)	Non-market valuation of wetland ecosystem	valuation	Residents living around the Cheimaditida wetland	Greece	Face-to-face in-house	2005	Stated	3/407/8/3,256

Continued on next page page

Table A.1 – *Continued from previous page*

Paper	Type of application	Population	Country	Surveying method	Year data was collected	Elicited preferences	Alternatives/ individuals/ choice sets ^a / choices ^b
Hanley et al. (2006)	Non-market valuation river ecosystem	Residents living around river Wear, County Durham, England; and river Clyde, in Central Scotland.	UK	Face-to-face in-house	2001	Stated	3/210/8/1,680
Milon and Scrogin (2006)	Non-market valuation of wetland ecosystem using a functional characterization approach	Residents of five Florida cities (Miami, Fort Myers, Orlando, Tampa, and West Palm Beach)	USA	Face-to-face in-house	1999	Stated	2/240/7/1,680
Milon and Scrogin (2006)	Non-market valuation of wetland ecosystem using a structural characterization approach	Residents of five Florida cities (Miami, Fort Myers, Orlando, Tampa, and West Palm Beach)	USA	Face-to-face in-house	1999	Stated	2/240/7/1,680
Scarpa et al. (2008)	Recreational demand	Rock climbers in the North-eastern Alps (Venetto region)	Italy	Face-to-face on-site	2000	Revealed	19/858/40/?
Kosenius (2010)	No-market valuation of water quality in Gulf of Finland	Residents of Finland, age 18 to 80	Finland	Mail	2006	Stated	3/653/6/3,708
Westerberg et al. (2010)	Non-market valuation of wetland ecosystem	Residents living within a 10 km radius from the Marais des Baux wetland	France	Face-to-face in-house and streets	2008	Stated	3/90/9/810

Studies primarily comparing welfare estimates

Continued on next page page

Table A.1 – *Continued from previous page*

Paper	Type of application	Population	Country	Surveying method	Year data was collected	Elicited preferences	Alternatives/ individuals/ choice sets ^a / choices ^b
Greene and Hensher (2003)	Analysis of choice of long distance travel by car	Car drivers undertaking long-distance trips by car in six cities in New Zealand (Auckland, Hamilton, Palmerston North, Wellington, and both Dunedins)	New Zealand	Face-to-face in streets	2000	Stated	3 or 4/274/16/4,384
Provencher and Bishop (2004)	Recreational demand	Anglers fishing in Milwaukee-Racine waters of Lake Michigan	USA	Face-to-face on-site, telephone, and mail	1996, 1997	Revealed	2/97/270/26,190
Hess et al. (2007)	Non-market valuation of car travelling time	Residents of Sweden	Sweden	Internet and face-to-face in-house	2004	Stated	2/1,723/9/13,386
Hynes et al. (2008)	Recreational demand	Kayakers of 11 whitewater sites in Ireland	Ireland	e-mail, internet, and face-to-face on-site	2003	Revealed	12/279/?/?
Cherchi et al. (2009)	Analysis of mode choice among car, bus and train	Residents of Cagliari	Italy	Telephone, and face-to-face in-house	1998	Stated	3/6,000/8 or 9/?
Shen (2009)	Analysis of mode choice among car, bus and monorail	Residents of the Saito and Onohara Area of northern Osaka	Japan	?	2005	Stated	3/467/8/3,736
Shen (2009)	Analysis of mode choice among car, bus and monorail	Eastern Osaka	Japan	?	2005	Stated	3/453/8/3,624

^a In a stated preference study, the number of choice sets are equivalent to the number of times a respondent states a choice.

In a revealed preference study, the number of choice sets are assumed by the researcher (sometimes, not explicitly reported).

^b Ideally, number of choices results from multiplying individuals times choice sets. However, if no all individuals provided

Continued on next page page

Table A.1 – *Continued from previous page*

Paper	Type of application	Population	Country	Surveying method	Year data was collected	Elicited preferences	Alternatives/ individuals/ choice sets ^a / choices ^b
an answer to all choice sets, the number of choices will be smaller.							

B. DENSITIES OF ESTIMATED WTP MEASURES BY PREFERENCE SCENARIO

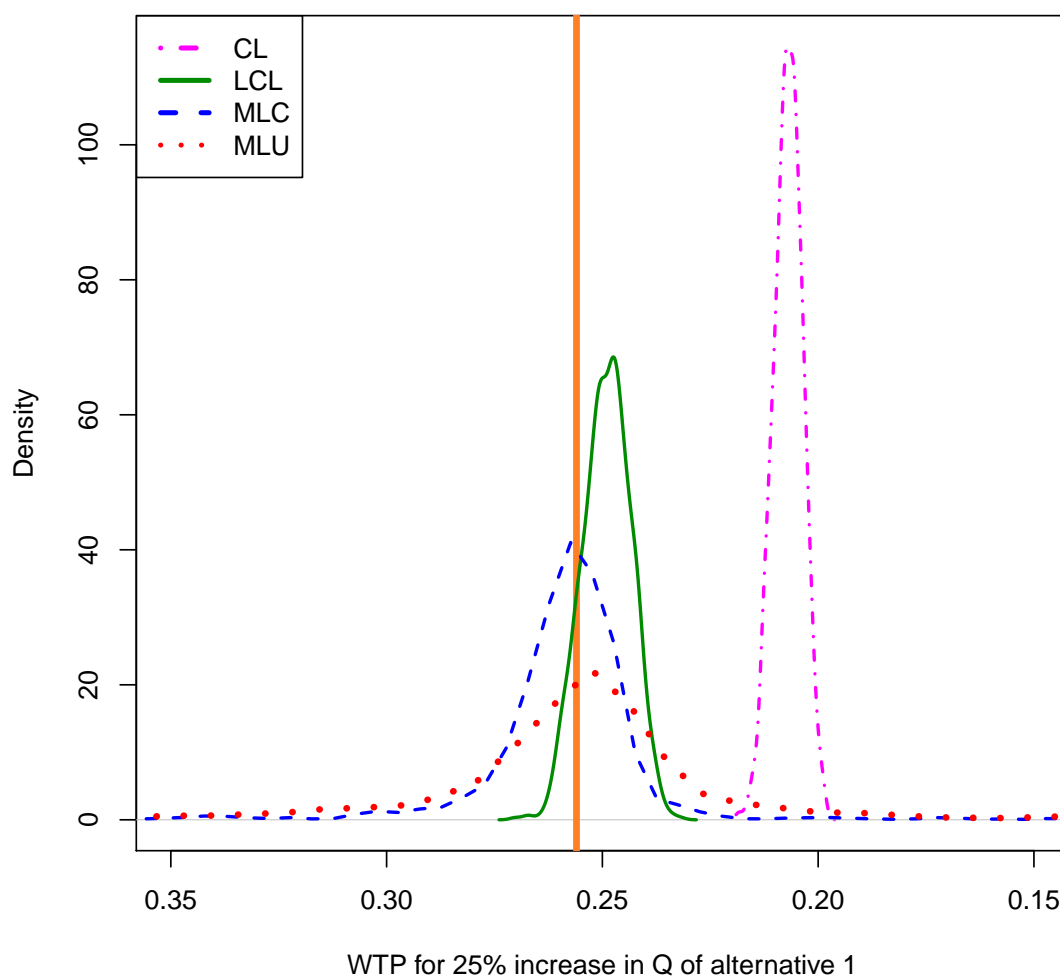


Fig. B.1: Snapshot on WTP for 25% increase in Q of alternative 1 by econometric method (discrete, correlated preferences scenario)

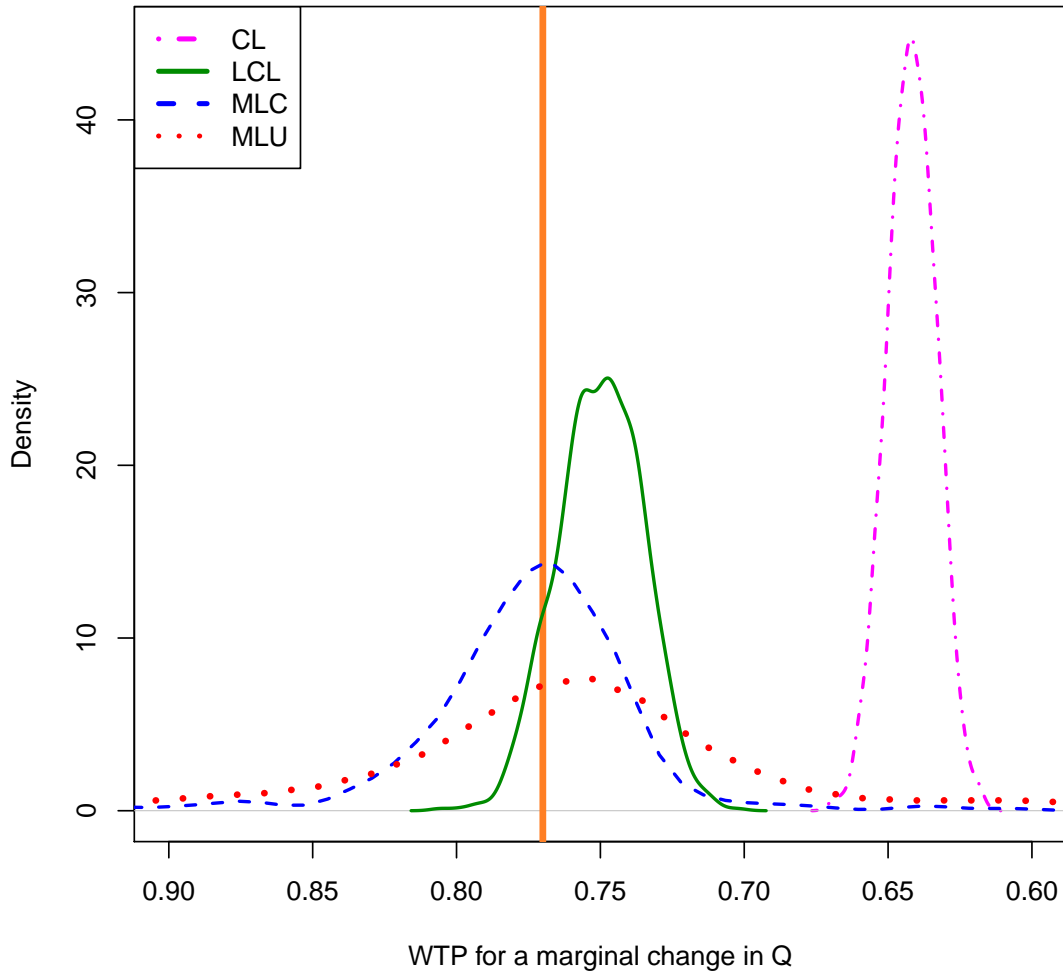


Fig. B.2: Snapshot on WTP for a marginal change in Q by econometric method (discrete, correlated preferences scenario)

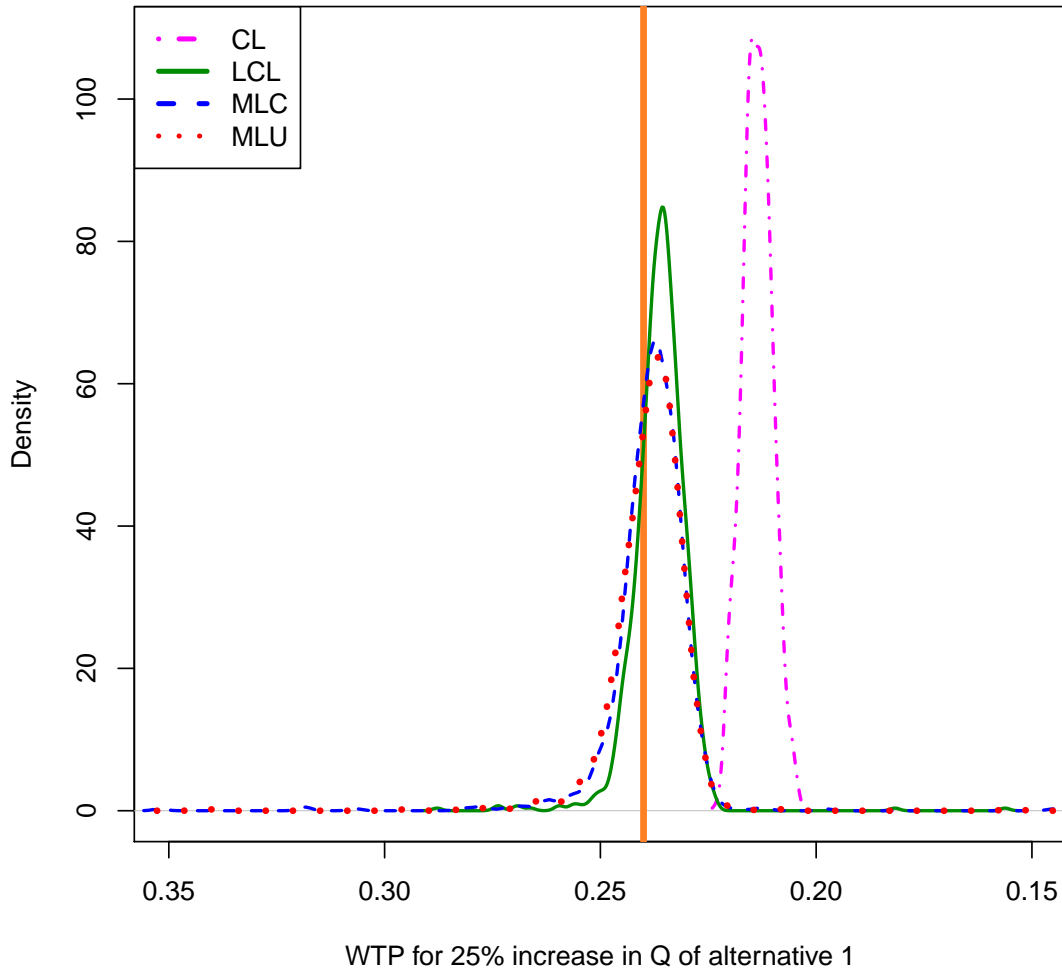


Fig. B.3: Snapshot on WTP for 25% increase in Q of alternative 1 by econometric method (Normal-normal, correlated preferences scenario)

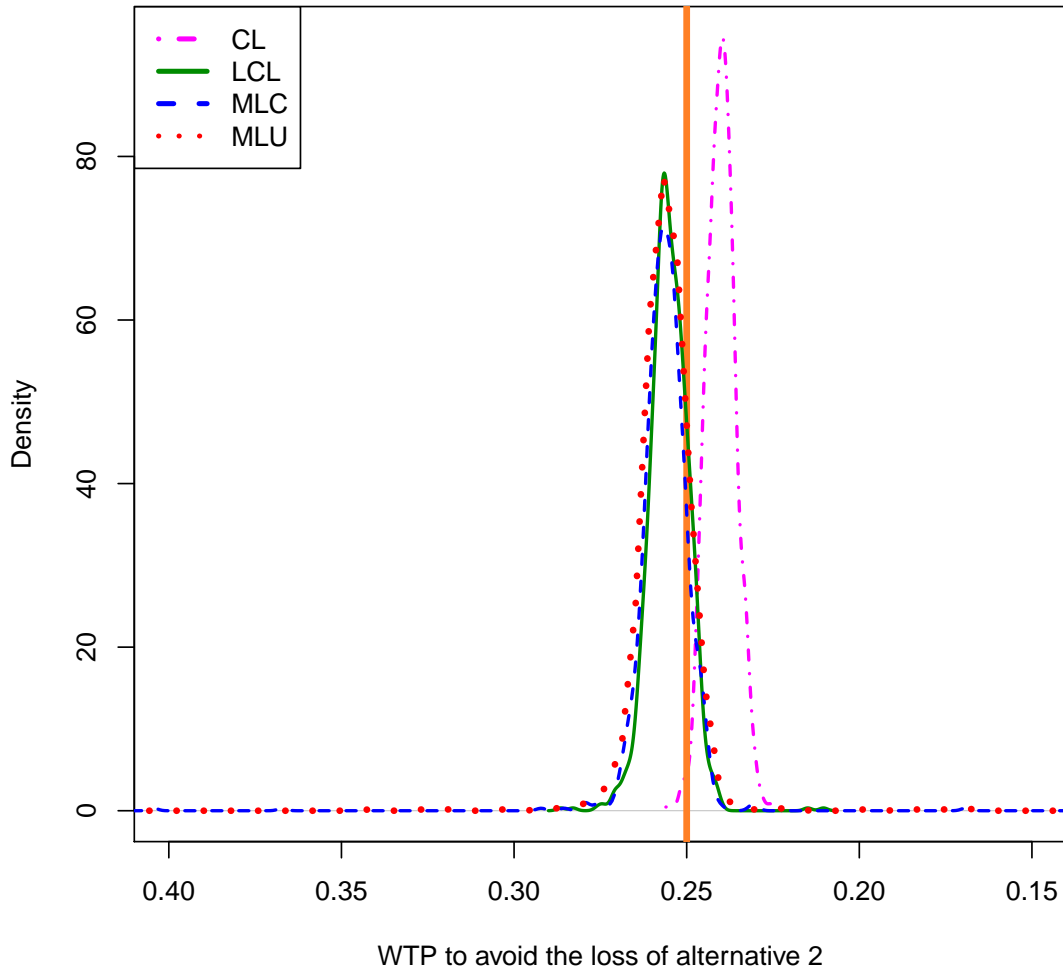


Fig. B.4: Snapshot on WTP to avoid the loss of alternative 2 by econometric method (Normal-normal, correlated preferences scenario)

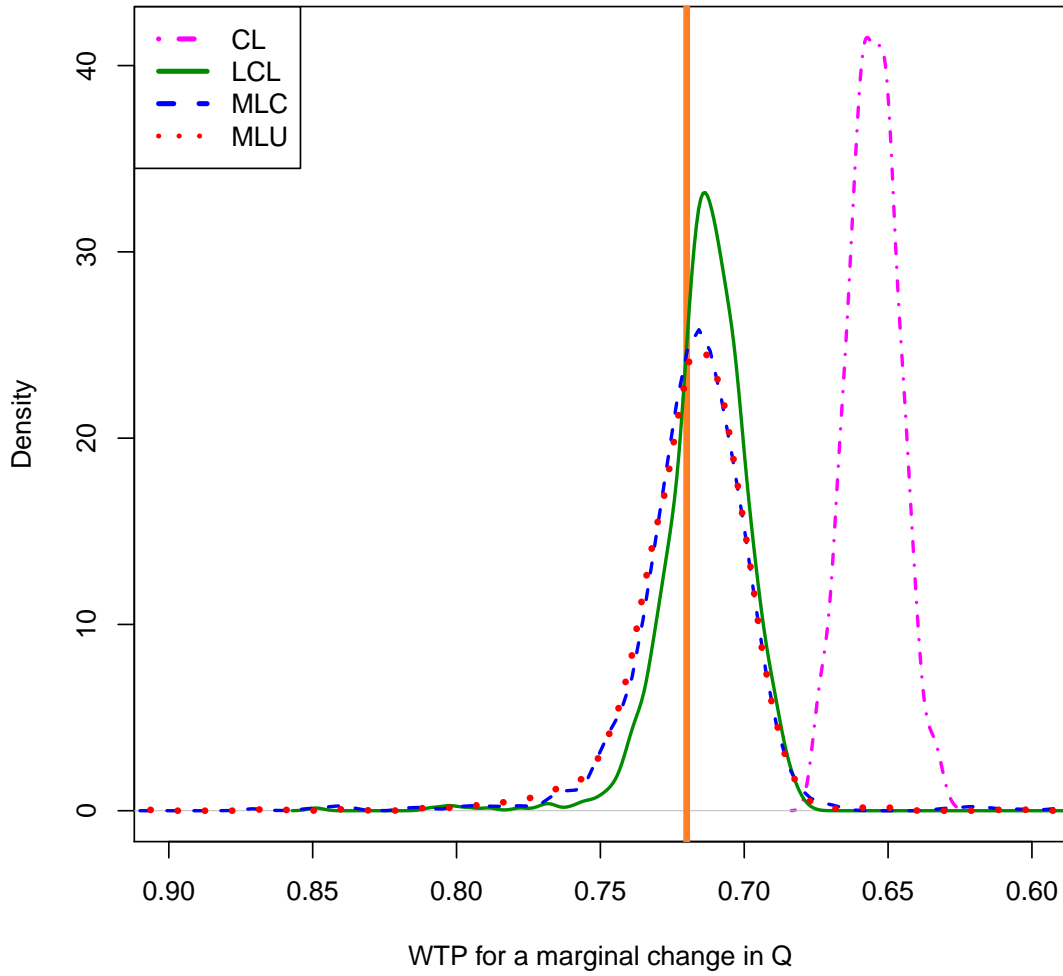


Fig. B.5: Snapshot on WTP for a marginal change in Q by econometric method (Normal-normal, correlated preferences scenario)

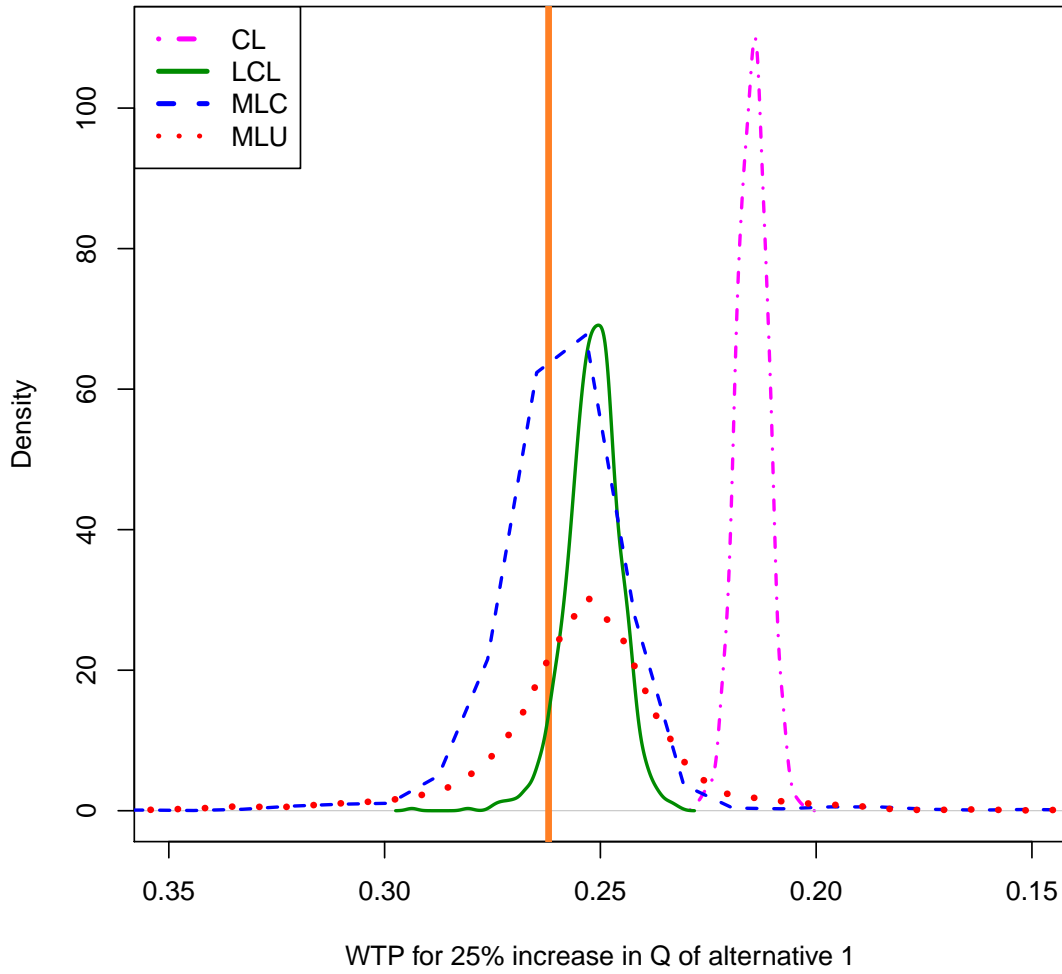


Fig. B.6: Snapshot on WTP for 25% increase in Q of alternative 1 by econometric method (Normal-normal, uncorrelated preferences scenario)

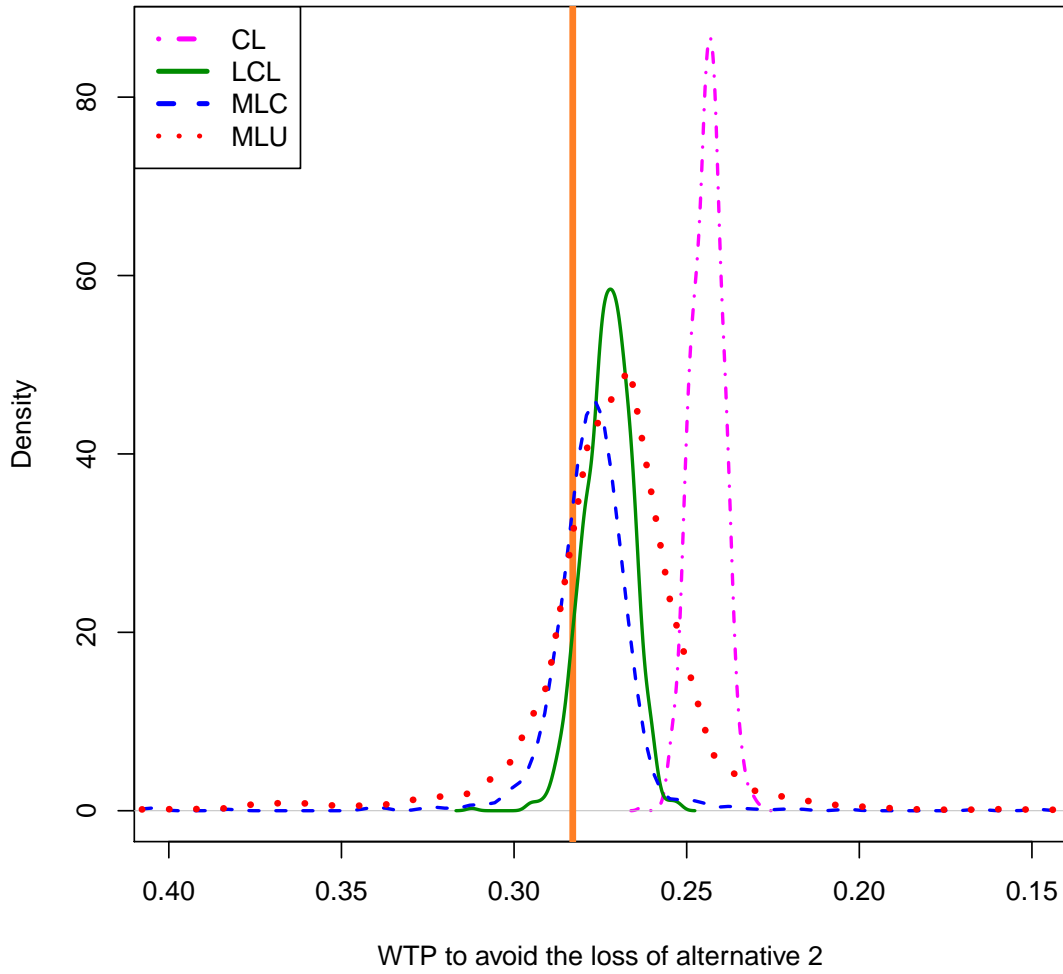


Fig. B.7: Snapshot on WTP to avoid the loss of alternative 2 by econometric method (Normal-normal, uncorrelated preferences scenario)

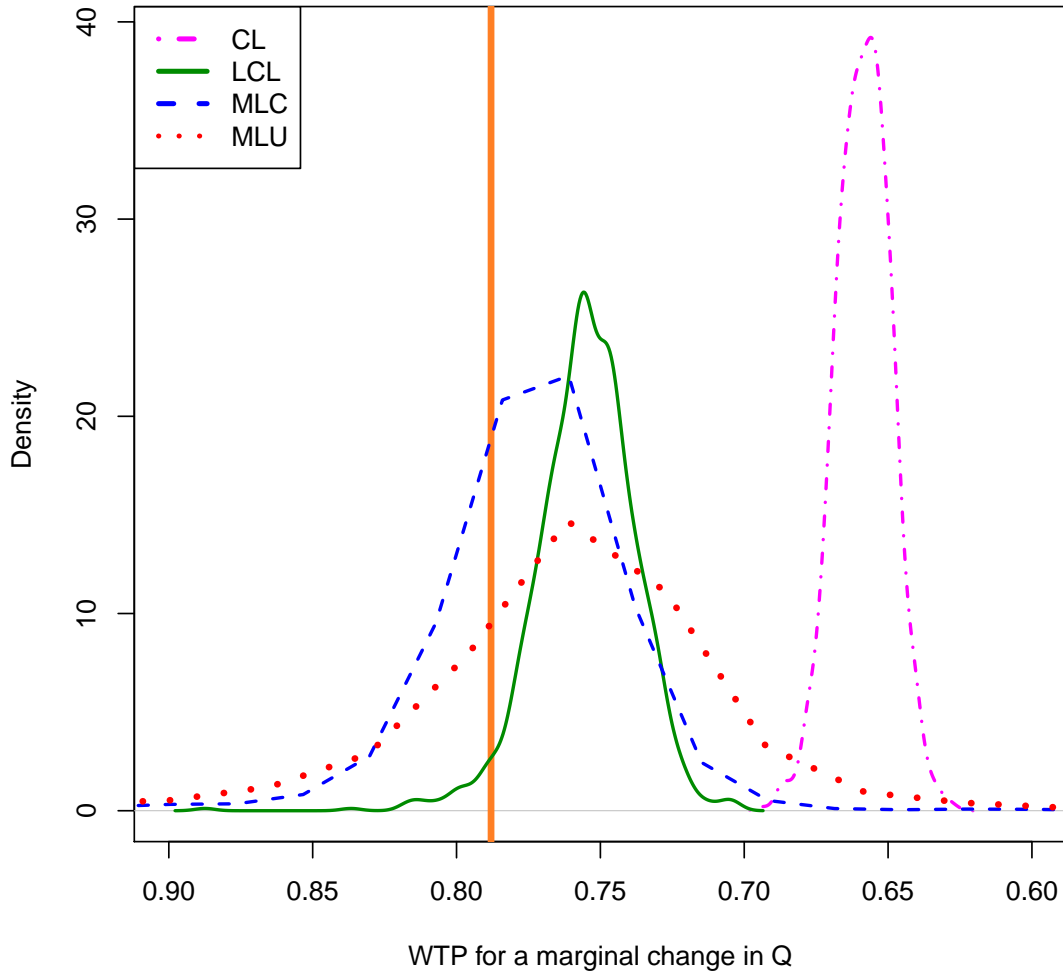


Fig. B.8: Snapshot on WTP for a marginal change in Q by econometric method (Normal-normal, uncorrelated preferences scenario)

C. DESCRIPTION OF STUDIES USING RESEARCHER'S OWN
JUDGEMENT IN SELECTING NUMBER OF CLASSES

Tab. C.1: Description of studies estimating latent class conditional logit specifications in environmental and resource economics

Paper	Type of application	Population	Country	Surveying method	Year data was collected	Elicited preferences	Alternatives/ individuals/ choice sets ^a / choices ^b
Richards (2000)	Segmentation of fruit consumers	Grocery shoppers	USA	Scanning	1997	Revealed	4/9510/1/9510
Boxall and Adamowicz (2002)	Recreational demand	Visitors of 5 parks in eastern Manitoba, western Ontario and northern Minnesota	Canada and USA	Mail	1995	Stated	6/620/8/4,892
Provencher et al. (2002)	Recreational demand	Wisconsin Lake Michigan anglers	USA	Face-to-face on-site	1996	Revealed	2/192/182/34,944
Greene and Hensher (2003)	Analysis of choice of long distance travel by car	Car drivers undertaking long-distance trips by car in six cities in New Zealand (Auckland, Hamilton, Palmerston North, Wellington, and both Dunedins)	New Zealand	Face-to-face in streets	2000	Stated	3 or 4/274/16/4,384
Scarpa et al. (2003)	Non-market valuation of genetically determined pig attributes	Households or backyard producers and small farmers rearing creole pigs in Yucatan	Mexico	Face-to-face in-house	2000	Stated	3/300/6/1800
Provencher and Bishop (2004)	Recreational demand	Anglers fishing in Milwaukee-Racine waters of Lake Michigan	USA	Face-to-face on-site, telephone, and mail	1996, 1997	Revealed	2/97/270/26,190
Scarpa and Thiene (2005)	Recreational demand	Rock climbers members of the Italian Alpine Club	Italy	Face-to-face on-site	1999	Revealed	18/528/?/8,787
Birol et al. (2006)	Non-market valuation of wetland ecosystem	Residents living around the Cheimaditida wetland	Greece	Face-to-face in-house	2005	Stated	3/407/8/3,256

Continued on next page page

Table C.1 – *Continued from previous page*

Paper	Type of application	Population	Country	Surveying method	Year data was collected	Elicited preferences	Alternatives/ individuals/ choice sets ^a / choices ^b
Milon and Scrogin (2006)	Non-market valuation of wetland ecosystem using a functional characterization approach	Residents of five Florida cities (Miami, Fort Myers, Orlando, Tampa, and West Palm Beach)	USA	Face-to-face in-house	1999	Stated	2/240/7/1,680
Milon and Scrogin (2006)	Non-market valuation of wetland ecosystem using a structural characterization approach	Residents of five Florida cities (Miami, Fort Myers, Orlando, Tampa, and West Palm Beach)	USA	Face-to-face in-house	1999	Stated	2/240/7/1,680
Ouma et al. (2007)	Non-market valuation of bull and cows attributes	Cattle-keeping households	Kenya and Ethiopia	Face-to-face in-house	2004, 2005	Stated	3/253/12 or 11/ 2,783 or 3,036
Hynes et al. (2008)	Recreational demand	Kayakers of 11 whitewater sites	Ireland	e-mail, internet, and face-to-face on-site	2003	Revealed	12/279/?/?
Ruto et al. (2008)	Non-market valuation livestock attributes	Livestock markets (cattle producers and traders) in the district of Kajiado	Kenya	Face-to-face on-site	2000	Stated	3/311/8/2488
Birol et al. (2009)	Non-market valuation of maize attributes	Maize-producing farmers	Mexico	Face-to-face in-house	2004	Stated	3/420/6/2,520
Colombo et al. (2009)	Non-market valuation of landscape quality	Residents in the North West region of England	England	Face-to-face in-house	2005	Stated	3/300/6/1,800
Shen (2009)	Analysis of mode choice among car, bus and monorail	Residents of the Saito and Onohara Area of northern Osaka	Japan	?	2005	Stated	3/467/8/3,736

Continued on next page page

Table C.1 – Continued from previous page

Paper	Type of application	Population	Country	Surveying method	Year data was collected	Elicited preferences	Alternatives/ individuals/ choice sets ^a / choices ^b
Beharry-Borg and Scarpa (2010)	Non-market valuation of coastal waters	Snorkellers at Tobago Island (foreign visitors, domestic visitors, and locals)	Trinidad and Tobago	Face-to-face on-site at International Airport	2007	Stated	3/198/9/1,782
Beharry-Borg and Scarpa (2010)	Non-market valuation of coastal waters	Non-snorkellers at Tobago Island (foreign visitors, domestic visitors, and locals)	Trinidad and Tobago	Face-to-face on-site at International Airport	2007	Stated	3/86/ 8 or 9/?
Brouwer et al. (2010)	Non-market valuation of water improvements	Urban and rural residents around Guadalquivir River basin	Spain	Face-to-face in-house	2006	Stated	3/619/4/?
Kosenius (2010)	Non-market valuation of water quality improvements	Residents of Finland, age 18 to 80]	Finland	Mail	2006	Stated	3/726/6/?
Breffle et al. (2011)	Recreational demand	Active Green Bay anglers who purchase fishing licenses	USA	Telephone	1998, 1999	Stated	2/640/8/?
Kikulwe et al. (2011)	Non-market valuation of a disease-resistant genetically modified banana variety	Consumers of banana	Uganda	Face-to-face in-house	2007	Stated	3/421/16/6,736
van Putten et al. (2011)	Preferences for conservation programs	Tasmanian landowners	Australia	?	2004	Stated	3/132/8/?
Broch and Vedel (2012)	Non-market valuation of agri-environmental contracts	Danish farmers	Denmark	e-mail	2009	Stated	3/853/6/5,053
Chung et al. (2012)	Non-market valuation of beef attributes	Grocery shoppers	South Korea	Face-to-face on-site	2007	Stated	4/873/10/?

Continued on next page page

Table C.1 – *Continued from previous page*

Paper	Type of application	Population	Country	Surveying method	Year data was collected	Elicited preferences	Alternatives/ individuals/ choice sets ^a / choices ^b
Garrod et al. (2012)	Non-market valuation of agri-environmental schemes	Residents across England	England	Face-to-face	2010	Stated	2/1,273/4/?

^a In a stated preference study, the number of choice sets are equivalent to the number of times a respondent states a choice. In a revealed preference study, the number of choice sets are assumed by the researcher (sometimes, not explicitly reported).

^b Ideally, number of choices results from multiplying individuals times choice sets. However, if not all individuals provided an answer to all choice sets, the number of choices will be smaller.

D. 95% CONFIDENCE INTERVALS OF ESTIMATED WTP MEASURES FOR THE ONE SMALL CLASS SCENARIO

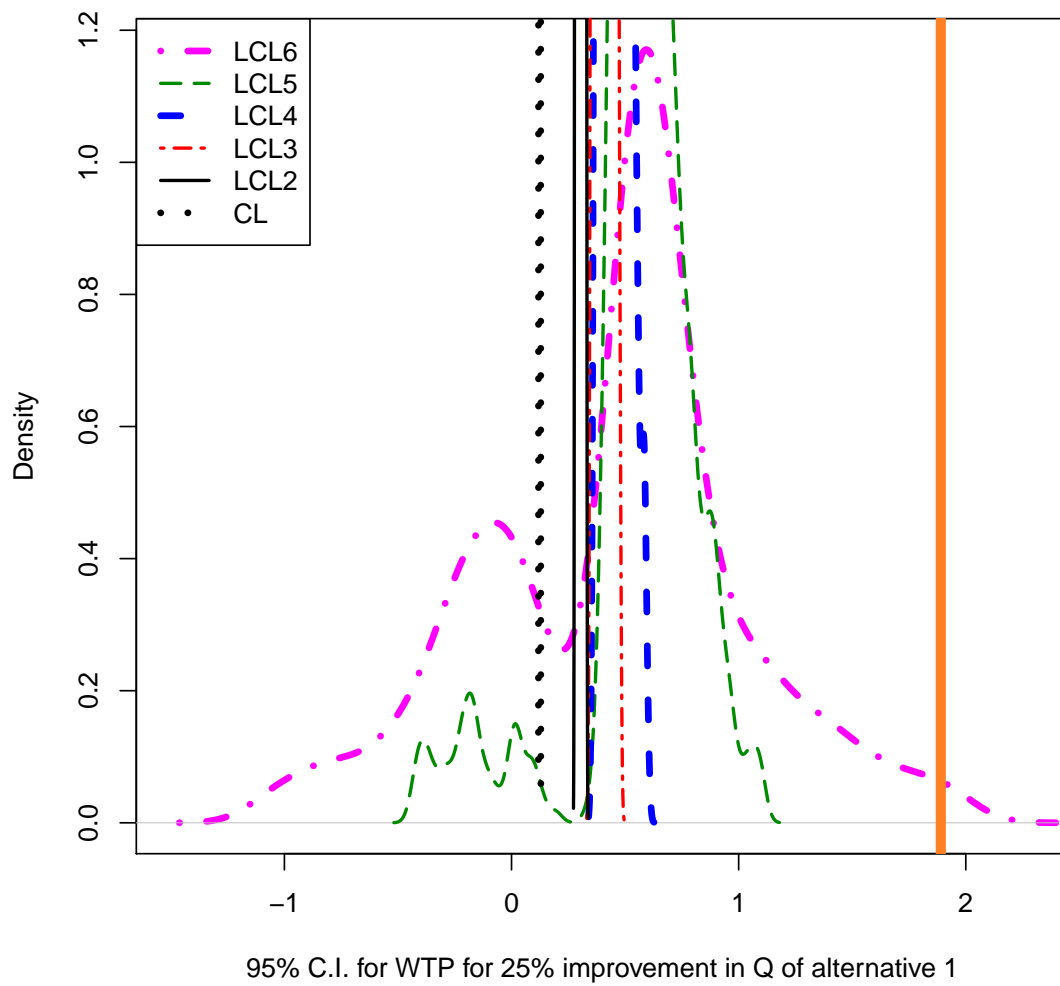


Fig. D.1: Snapshot of WTP for 25% improvement in Q of alternative 1 by econometric method (one small class scenario)

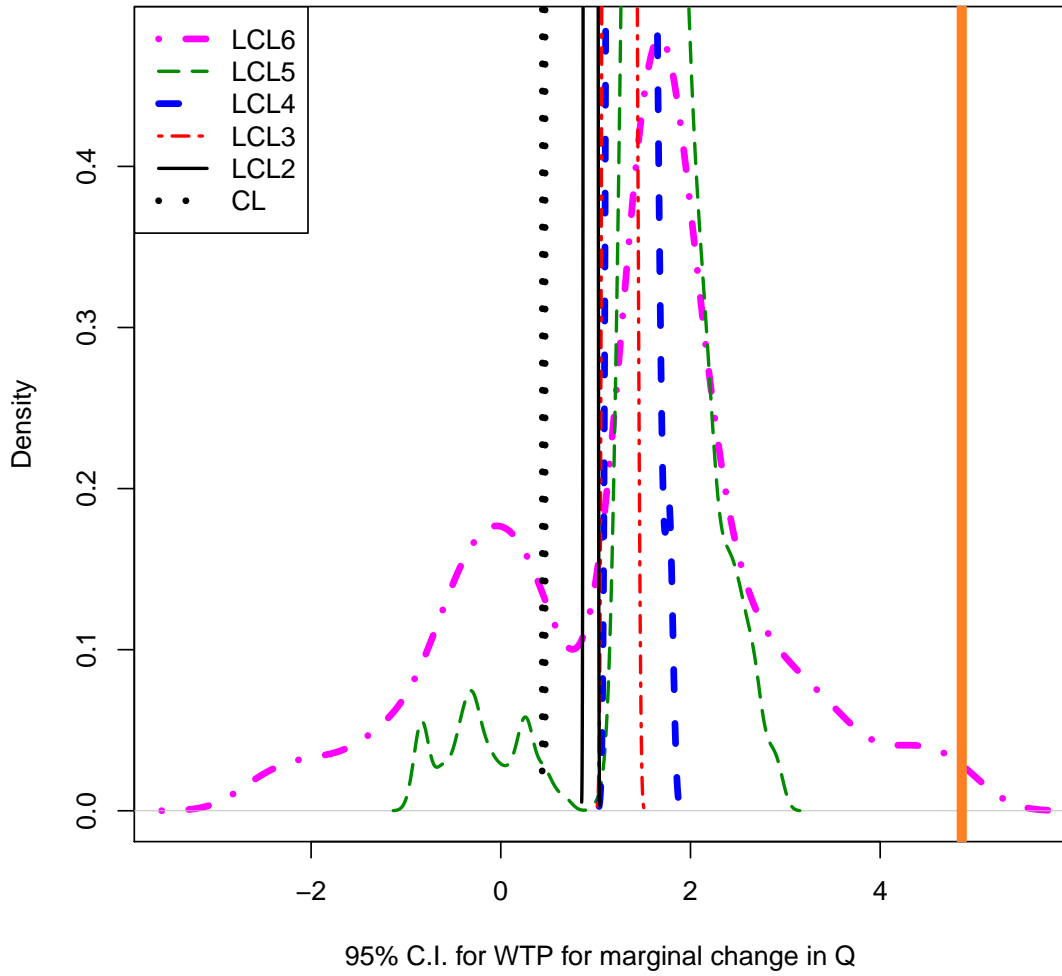


Fig. D.2: Snapshot of WTP for marginal improvement in Q by econometric method (one small class scenario)

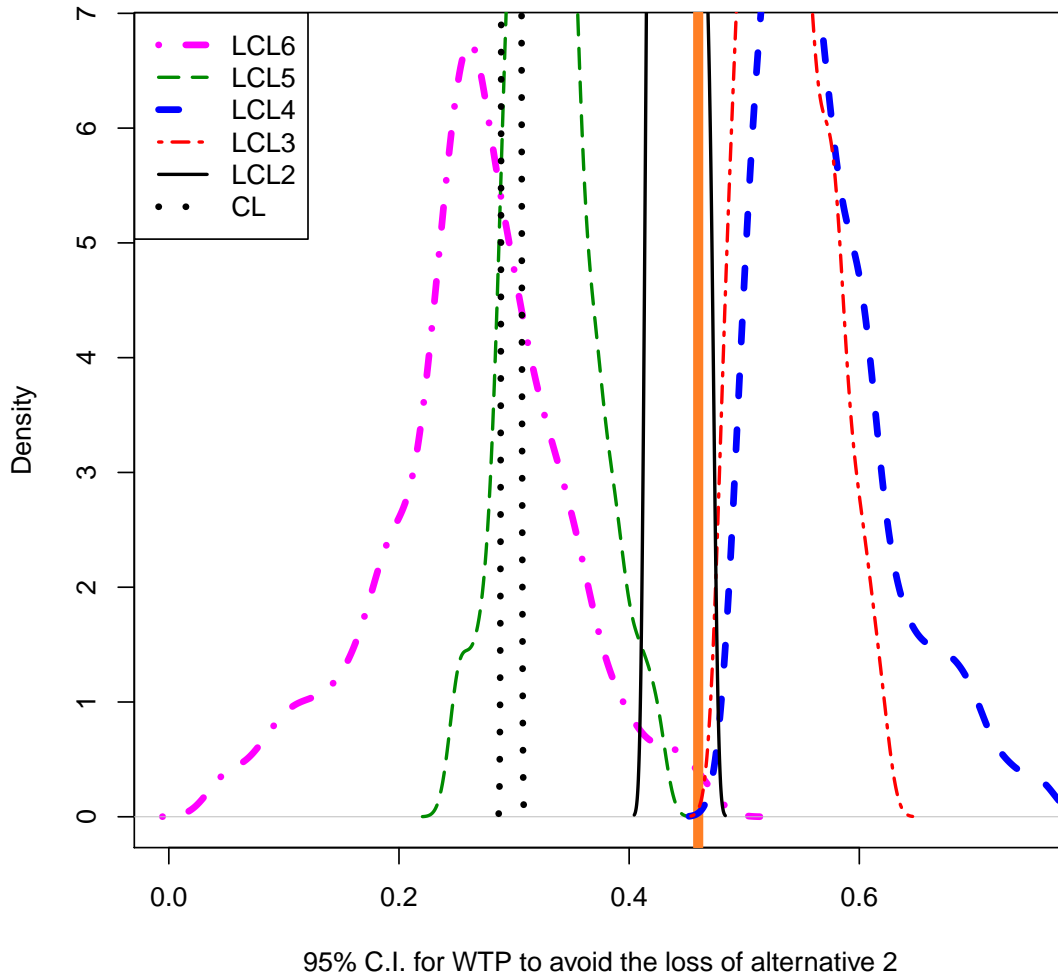


Fig. D.3: Snapshot of WTP to avoid loss of alternative 2 by econometric method (one small class scenario)

BIBLIOGRAPHY

- M. Bech, T. Kjaer, and J. Lauridsen. Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. *Health Economics*, 20: 273–286, 2011.
- N. Beharry-Borg and R. Scarpa. Valuing quality changes in Caribbean coastal waters for heterogeneous beach visitors. *Ecological Economics*, 69:1124–1139, 2010.
- M. E. Ben-Akiva. The structure of travel demand models. Ph. D. dissertation, Sytem Division, Department of Civil Engineering, M.I.T., 1972.
- E. Birol, K. Karousakis, and P. Koundouri. Using a choice experiment to account for preference heterogeneity in wetland attributes: The case of Cheimaditida wetland in Greece. *Ecological Economics*, 60:145–156, 2006.
- E. Birol, E. R. Villalba, and M. Smale. Farmer preferences for milpa diversity and genetically modified maize in Mexico. *Environment and Development Economics*, 14(4):521–540, 2009.
- M. C. Bliemer and J. M. Rose. Construction of experimental design for mixed logit models allowing for correlation across choice observations. *Transportation Research Part B*, 44:720–734, 2010.
- P. C. Boxall and W. L. Adamowicz. Understanding heterogeneous preferences in random utility models: A latent class approach. *Environment and Resource Economics*, 23:421–446, 2002.
- W. S. Breffle and E. R. Morey. Investigating preference heterogeneity in a repeated discrete-choice recreation demand model of Atlantic salmon fishing. *Marine Resource Economics*, 15:1–20, 2000.
- W. S. Breffle, E. R. Morey, and J. A. Thacher. A joint latent-class model: combining Likert-scale preference statements with choice data to harvest preference heterogeneity. *Environment and Resource Economics*, 50(1):83–110, 2011.
- S. W. Broch and S. E. Vedel. Using choice experiments to investigate the policy relevance of heterogeneity in farmer agri-environmental contract preferences. *Environment and Resource Economics*, 51:561–581, 2012.
- R. Brouwer, J. Martin-Ortega, and J. Berbel. Spatial preference heterogeneity: A choice experiment. *Land Economics*, 86(3):552 – 568, 2010.

- A. Bujosa-Bestard, A. Riera-Font, and R. L. Hicks. Combining discrete and continuous representations of preference heterogeneity: A latent class approach. *Environment and Resource Economics*, 47(4):477–493, 2010.
- M. Burton and D. Rigby. Hurdle and latent class approaches to serial non-participation in choice models. *Environment and Resource Economics*, 42(2): 211–226, 2009.
- F. Carlsson, P. Frykblom, and C. Liljenstolpe. Valuing wetland attributes: An application of choice experiments. *Ecological Economics*, 47:95–103, 2003.
- R. T. Carson and J. J. Louviere. A common nomenclature for stated preference elicitation approaches. *Environment and Resource Economics*, 49:539–559, 2011.
- E. Cherchi. Modelling individual preferences, state of the art, recent advances and future directions. paper presented in the 12th International Conference on Travel Behaviour Research, Jaipur, India, 13-18 December, 2009. URL http://iatbr2009.asu.edu/ocs/custom/resource/W5_R1_Modelling%20individual%20preferences,%20State%20of%20the%20art.pdf.
- E. Cherchi, C. Cirilo, and J. W. Polak. Assesment of user benefits in presence of random taste heterogeneity. *Transportation Research Record: Journal of the Transportation Research Board*, 2123:78–86, 2009.
- C. Chung, B. C. Briggeman, and S. Han. Willingness-to-pay for beef quality attributes: A latent segmentation analysis of Korean grocery shoppers. *Journal of Agricultural and Applied Economics*, 44(4):447–459, 2012.
- S. Colombo, N. Hanley, and J. Louviere. Modeling preference heterogeneity in stated choice data: An analysis for public goods generated by agriculture. *Agricultural Economics*, 40:307–322, 2009.
- A. Domanski and R. H. von Haefen. Estimation and welfare analysis from mixed logit models with large choice sets. working paper, 2012. URL <http://aede.osu.edu/sites/drupal-aede.web/files/von%20Haefen%20SU%20presentation%20May%2031%202012.pdf>.
- T. Domencich and D. McFadden. *Urban travel demand: A behavioral analysis*. North-Holland, Amsterdam, 1975.
- S. Ferrini and R. Scarpa. Designs with *a priori* information for nonmarket valuation with choice experiments: A Monte Carlo study. *Journal of Environmental Economics and Management*, 53:342–363, 2007.
- G. Garrod, E. Ruto, K. Willis, and N. Powe. Heterogeneity of preferences for the benefits of Environmental Stewardship: A latent-class approach. *Ecological Economics*, 76:104–111, 2012.

- W. H. Greene and D. A. Hensher. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B*, 37:681–698, 2003.
- W. H. Greene and D. A. Hensher. Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. *Applied Economics*, 45, 2013. URL <http://dx.doi.org/10.1018/00036846.2011.650325>.
- T. Haab and K. McConnell. *Valuing environmental and natural resources. The econometrics of non-market valuation*. Edward Elgar, 2002.
- W. M. Hanemann. Welfare analysis with qualitative response models. Working paper No. 241, California Agricultural Experiment Station, Giannini Foundation of Agricultural Economics, 1982.
- N. Hanley, R. E. Wright, and B. Alvarez-Farizo. 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, 2006.
- C. J. Heinrich and J. B. Wenger. The economic contributions of James J. Heckman and Daniel L. McFadden. *Review of Political Economy*, 14(1):69–89, 2002.
- D. A. Hensher and W. H. Greene. The mixed logit model: The state of the practice. *Transportation*, 30(2):133–176, 2003.
- S. Hess, M. Bierlaire, and J. W. Polak. A systematic comparison of continuous and discrete mixture models. *European Transport*, 37:35–61, 2007.
- S. Hess, M. Ben-Akiva, D. Gopinath, and J. Walker. Advantages of latent class over continuous mixture of logit models. working paper, 2011. URL http://www.stephanehess.me.uk/papers/Hess_Ben-Akiva_Gopinath_Walker_May_2011.pdf.
- S. Hess, D. A. Hensher, and A. Daly. Not bored yet - revisiting respondent fatigue in stated choice experiments. *Transportation Research Part A*, 46:626–644, 2012.
- D. Hoyos, P. Mariel, and J. Fernández-Macho. The influence of cultural identity on the wtp to protect natural resources: Some empirical evidence. *Ecological Economics*, 68:2372–2381, 2009.
- S. Hynes, N. Hanley, and R. Scarpa. Effects on welfare measures of alternative means of accounting for preference heterogeneity in recreational demand models. *American Journal of Agricultural Economics*, 90(4):1011–1027, 2008.
- E. M. Kikulwe, E. Birol, J. Wesseler, and J. Falck-Zepeda. A latent approach to investigating demand for genetically modified banana in Uganda. *Agricultural Economics*, 42:547–560, 2011.

- A.-K. Kosenius. Heterogeneous preferences for water quality attributes: The case of eutrophication in the Gulf of Finland, the Baltic Sea. *Ecological Economics*, 69:528–538, 2010.
- W. F. Kuhfeld. Construction of efficient designs for discrete choice experiments. In R. Grover and M. Vriens, editors, *The handbook of marketing research*, pages 312–329. Sage Publications, 2006.
- W. F. Kuhfeld. Marketing research methods in sas: Experimental design. SAS Institute Inc., Cary, NC, USA, 2010. URL <http://support.sas.com/techsup/technote/ts722.pdf>.
- W. F. Kuhfeld, R. D. Tobias, and M. Garratt. Efficient experimental design with marketing research implications. *Journal of Marketing Research*, 31:132–141, 1994.
- A. G. Lazari and D. A. Anderson. Designs of discrete choice set experiments for estimating both attribute and availability cross effects. *Journal of Marketing Research*, 31:375–383, 1994.
- J. L. Lusk and F. B. Norwood. Effect of experimental design on choice-based conjoint valuation estimates. *American Journal of Agricultural Economics*, 87:771–785, 2005.
- K. E. McConnell. Consumer surplus from discrete choice models. *Journal of Environmental Economics and Management*, 29:263–270, 1995.
- K. E. McConnell and W.-C. Tseng. Some preliminary evidence on sampling of alternatives with the random parameter logit. *Marine Resource Economics*, 14: 317–332, 1999.
- D. McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York, 1974.
- D. McFadden. The revealed preferences of a government bureaucracy: Theory. *The Bell Journal of Economics*, 6(2):401–416, 1975.
- D. McFadden. The revealed preferences of a government bureaucracy: Empirical evidence. *The Bell Journal of Economics*, 7(1):55–72, 1976.
- D. McFadden. Computing willingness-to-pay in random utility models. mimeograph, Department of Economics, University of California, 1995.
- D. McFadden. Economic choices. *American Economic Review*, 91(3):351–378, 2001.
- D. McFadden and K. Train. Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15:447–470, 2000.

- E. Meijer and J. Rouwendal. Measuring welfare effects in models with random coefficients. *Journal of Applied Econometrics*, 21:227–244, 2006.
- J. W. Milon and D. Scrogin. Latent preferences and valuation of wetland ecosystems. *Ecological Economics*, 56:162–175, 2006.
- L. Nahuelhual, M. L. Loureiro, and J. Loomis. Using random parameters to account for heterogeneous preferences in contingent valuation of public open space. *Journal of Agricultural and Resource Economics*, 29(3):537–552, 2004.
- A. Nevo. A practitioner’s guide to estimation of random-coefficients logit model of demand. *Journal of Economics and Management Strategy*, 9(4):513–548, 2000.
- J. Ortúzar. Defining the state of the art in discrete choice modelling. Workshop on transportation and sustainable cities. Universidad de Chile, Santiago, 2006.
- E. Ouma, A. Abdulai, and A. Drucker. Measuring heterogeneous preferences for cattle traits among cattle-keeping households in East Africa. *American Journal of Agricultural Economics*, 89(4):1005–1019, 2007.
- G. L. Poe, K. L. Giraud, and J. B. Loomis. Computational methods for measuring the difference of empirical distributions. *American Journal of Agricultural Economics*, 87(2):353–365, 2005.
- B. Provencher and R. C. Bishop. Does accounting for preferences heterogeneity improve the forecasting of a random utility model? A case of study. *Journal of Environmental Economics and Management*, 48:793–810, 2004.
- B. Provencher, K. A. Baerenklau, and R. C. Bishop. A finite mixture logit model of recreational angling with serially correlated random utility. *American Journal of Agricultural Economics*, 84(4):1066–1075, 2002.
- T. J. Richards. The impact of promotion and advertising: A latent class approach. *Journal of Agricultural and Applied Economics*, 32(3):441–457, 2000.
- E. Ruto, G. Garrod, and R. Scarpa. Valuing animal genetic resources: A choice modeling application to indigenous cattle in Kenya. *Agricultural Economics*, 38: 89–98, 2008.
- R. Scarpa and M. Thiene. Destination choice models for rock climbing in the North-eastern Alps: A latent-class approach based on intensity of preferences. *Land Economics*, 81(3):426–444, 2005.
- R. Scarpa, A. G. Drucker, S. Anderson, N. Ferraes-Ehuan, V. Gómez, C. R. Risopatrón, and O. Rubio-Leonel. Valuing genetic resources in peasant economies: The case of hairless creole pigs in yucatan. *Ecological Economics*, 45:427–443, 2003.

- R. Scarpa, M. Thiene, and K. Train. Utility in willingness to pay space: A tool to address confounding random scale effects in destination choice to the Alps. *American Journal of Agricultural Economics*, 90(4):994–1010, 2008.
- J. Shen. Latent class model or mixed logit mode? A comparison by transport mode choice data. *Applied Economics*, 41:2915–2924, 2009.
- M. Sillano and J. Ortúzar. Willingness-to-pay estimation with mixed logit models: Some new evidence. *Environment and Planning A*, 37:525–550, 2005.
- K. A. Small and H. S. Rosen. Applied welfare economics with discrete choice models. *Econometrica*, 49:105–130, 1982.
- C. Torres, N. Hanley, and S. Colombo. Incorrectly accounting for taste heterogeneity in choice experiments: Does it really matter for welfare measurement? CRE working paper 2011/1, 2011a. URL <https://dspace.stir.ac.uk/handle/1893/2720>.
- C. Torres, N. Hanley, and A. Riera. How wrong can you be? implications of incorrect utility function specification for welfare measurement in choice experiments. *Journal of Environmental Economics and Management*, 62:111–121, 2011b.
- K. Train. Recreation demand models with taste differences over people. *Land Economics*, 74(2):230–239, 1998.
- K. Train. *Discrete choice methods with simulation*. Cambridge University Press, 2003.
- K. Train. EM Algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling*, 1(1):40–69, 2008.
- K. Train and M. Weeks. Discrete choice models in preference space and willingness-to-pay space. In R. Scarpa and A. Alberini, editors, *Applications of simulation methods in environmental and resource economics*, pages 1–16. Springer, 2005.
- I. E. van Putten, S. M. Jennings, J. J. Louviere, and L. B. Burgess. Tasmanian landowner preferences for conservation incentive programs: A latent class approach. *Journal of Environmental Management*, 92:2647–2656, 2011.
- R. H. von Haefen, D. M. Massey, and W. L. Adamowicz. Serial nonparticipation in repeated discrete choice models. *American Journal of Agricultural Economics*, 87(4):1061–1076, 2005.
- V. H. Westerberg, R. Lifran, and S. B. Olsen. To restore or not? A valuation of social and ecological functions of the Marais des Baux wetland in Southern France. *Ecological Economics*, 69:2383–2393, 2010.
- J. Yu, P. Goos, and M. Vandebroek. Efficient conjoint choice designs in the presence of respondent heterogeneity. *Marketing Science*, 28(1):122–135, 2009.

- J. Yu, P. Goos, and M. Vandebroek. Individually adapted sequential Bayesian conjoint-choice designs in the presence of consumer heterogeneity. *International Journal of Research in Marketing*, 28:378–388, 2011.
- J. Zhang and W. L. Adamowicz. Unraveling the choice format effect: A context-dependent random utility model. *Land Economics*, 87(4):730–743, 2011.