Journal of Choice Modelling 27 (2018) 50–61

Contents lists available at ScienceDirect

Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

Individual-specific point and interval conditional estimates of latent class logit parameters

Mauricio Sarrias[a,*], Ricardo A. Daziano[b]

[a] Department of Economics, IDEAR, Universidad Catolica del Norte, Antofagasta, Chile
[b] School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14853, USA

ARTICLE INFO

JEL classification:
C25
C53

Keywords:
Discrete choice
Discrete heterogeneity
Random parameters
Standard errors

ABSTRACT

Within the realm of logit-type random parameter models to address unobserved heterogeneity in preferences there are two dominant approaches: the mixed logit model, which assumes parametric and continuous heterogeneity distributions, and the latent class logit model, which is a discrete and semiparametric counterpart of mixed logit. In addition to offer flexibility benefits, random parameter models allow researchers to make conditional (posterior) inference on preference parameters at the individual-specific level. In this paper we extend the individual-specific experimental approach, that was conducted by Revelt and Train (2000) for the continuous heterogeneity distributions of a mixed logit, to the discrete case of the latent class logit model. Our Monte Carlo study results confirm the expectation that for a given number of individuals, the density of the conditional means converges to the conditional population as the number of choice situations increases. We also add to the analysis the behavior of interval estimates using two methods for the derivation of standard errors of the individual-specific estimates. In general, as we have more information of the choices made by the individuals, we are in better shape to identify individual-specific preferences. Our main conclusion is that accurate individual-specific estimation is possible – including correct assignment to classes, but a large number of choice situations is needed to correctly approximate the true underlying distribution.

1. Introduction

Markets of highly differentiated goods are usually characterized by high levels of heterogeneity of preferences among consumers. The use of logit-type random parameter models is now standard practice in choice modeling research to address unobserved heterogeneity in preferences (Train, 2009). These random parameter choice models are usually characterized by a conditional logit kernel (MNL) (McFadden, 1974) and a mixing distribution that represents how preferences vary in the population. The mixing distribution is most commonly assumed to be parametric and continuous, leading to mixed (multinomial) logit models (MMNL) (McFadden and Train, 2000; Hensher and Greene, 2003). If the mixing distribution is semiparametric and discrete, then the resulting model is a latent class logit (LC-MNL) (Kamakura and Russell, 1989; DeSarbo et al., 1995; Greene and Hensher, 2003; Shen, 2009). Finally, the benefits of both MMNL and LC-MNL can be combined producing a mixture of both models. This double-mixture is known as the ‘mixed-mixed’ logit model (Keane and Wasi, 2013).

Even though the analyst may be interested in using disaggregate choice models to forecast aggregate market shares, random

* Corresponding author.
E-mail addresses: mauricio.sarrias@ucn.cl (M. Sarrias), daziano@cornell.edu (R.A. Daziano).

https://doi.org/10.1016/j.jocm.2017.10.004
Received 26 May 2017; Received in revised form 24 October 2017; Accepted 24 October 2017
Available online 4 November 2017
1755-5345/

parameter models offer –at least in theory– the flexibility of recovering parameters for each individual, conditional on past or observed choices. Individual-specific estimates can be exploited for customized marketing efforts and targeted product design (Allenby and Rossi, 1998).

In particular, individual-specific inference is performed by moving from the unconditional –sample– distribution to a conditional/posterior distribution of preferences. The derived point individual-specific estimates can be used to develop segments, identify outliers, and simulate market choices (Huber and Train, 2001). In the same vein, Hess and Rose (2007) pointed out the advantages –flexibility gains while reducing the impact of the unconditional parametric assumptions– and disadvantages –out sample forecasting– of using the conditional approach as opposed to the unconditional approach. In the context of LC-MNL, individual estimates have been mainly used to compute willingness-to-pay measures (Scarpa and Thiene, 2005; Hensher and Greene, 2010; Beharry-Borg and Scarpa, 2010) and/or other welfare measures at the individual level (Shen, 2009; Hynes et al., 2008).[1]

In addition to point estimates at the individual level, it is equally important to make inference on the precision of the individual-specific estimates. Consider an individual's parameter of 0.001 for a given attribute. From a frequentist standpoint, a question that arises is: is this coefficient for that particular individual truly positive or is it statistically not different from zero? These simple questions reveal the importance of interval inference (cf. Craig et al., 2005; Daziano and Achtnicht, 2014; Greene et al., 2014). Even though the few studies that have looked into individual-specific estimates mostly focused on point estimation, two approaches to derive precision metrics of the conditional means have been actually proposed. The first approach uses the Bayes precision, i.e. an estimator of the conditional variance of the conditional distribution based on point estimates (Hensher et al., 2003; Greene, 2012), whereas the second approach takes into consideration the sampling distribution of the parameters using a re-sampling procedure (Revelt and Train, 2000; Greene et al., 2014) for deriving standard errors.

Little is known about the statistical and asymptotic behavior of these interval estimators. For example, Revelt and Train (2000) derive and analyze the asymptotic behavior of the individual-specific MMNL estimates as the number of choice situations increases without bound, but the authors did not look into standard errors. This paper attempts to revisit the problem of point and interval estimates or random parameter logit models at the individual level by performing a full Monte Carlo study for discrete heterogeneity distributions (LC-MNL). The idea is to understand the asymptotic behavior of both interval estimators and also investigate their properties under differing scenarios, such as correct specification and misspecification. Misspecification of preferences is not just a technical problem: biased confidence intervals of the individual-specific estimates lead to poor policy decisions. In this respect, it is expected that the results of this paper will help to improve and understand methods for statistical inference on choices at the individual level, and thus contribute to better decisions by diverse stakeholders, including policymakers, firms, and researchers and analysts.

The following sections are organized as follows. In Section 2 we briefly review the LC-MNL model. In Section 3 we explain the main issues regarding individual-specific estimates and both methods used for deriving precision/interval estimators. Section 4 explains the Monte Carlo setup and Section 5 presents the main results. Finally, Section 6 discusses the results and concludes.

2. Latent class logit models

Consider the following random parameter logit specification:

$$U_{ijt}^{*} = \mathbf{x}_{ijt}^{\top}\boldsymbol{\beta}_i + \varepsilon_{ijt},\quad i = 1,\dots,N; j = 1,\dots,J, t = 1,\dots,T_i,$$

where U_{ijt}^{*} is the latent indirect utility for individual i when choosing alternative j in choice situation t; $\mathbf{x}_{ijt}$ is a $K \times 1$ vector of observed alternative attributes; ε_{ijt} is the idiosyncratic taste shock, and is i.i.d. Type 1 Extreme Value; the parameter vector $\boldsymbol{\beta}_i$ is unobserved for each i and is assumed to vary in the population following some distribution $g(\cdot)$. Different assumptions about $g(\cdot)$ gives rise to different logit models.

Let $y_{ijt} = 1$ if individual i chooses j on occasion t, and 0 otherwise. Given a specific value of the preference parameters $\boldsymbol{\beta}_i = \boldsymbol{\beta}_q$, the conditional joint density of choices made by consumer i is:

$$f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}_q) = \prod_{t=1}^{T_i}\prod_{j=1}^{J}\left[\frac{\exp\left(\mathbf{x}_{ijt}^{\top}\boldsymbol{\beta}_q\right)}{\sum_{j=1}^{J}\exp\left(\mathbf{x}_{ijt}^{\top}\boldsymbol{\beta}_q\right)}\right]^{y_{ijt}},$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iT_i})$. The unconditional joint density of individual choices depends on the assumptions of the mixing distribution.

Unlike MMNL, where $g(\cdot)$ is parametric and continuous, LC-MNL (Kamakura and Russell, 1989) assumes that preferences are distributed following a discrete distribution.[2] Unobserved preference heterogeneity is then accommodated in LC-MNL by a discrete number Q of separate (and unobserved) classes or segments of individuals with different values for the preference parameters within each class.[3] Note that individuals in each segment share homogeneous preferences (parameters are fixed within a class), but heterogeneity in

[1] For some applications of the individual-specific estimates in the context of MMNL, see Greene et al. (2005); Sillano and de Dios Ortúzar (2005); Hensher et al. (2006); Hess and Hensher (2010).

[2] For a deeper review of the LC-MNL model, see Hess (2014). For some comparison between the LC-MNL and MMNL see Shen (2009); Greene and Hensher (2003).

[3] Subsets of parameters can be constrained to be the same across classes.

preferences exists across classes. Formally, the population distribution of the parameters is specified as:

$$g(\boldsymbol{\beta}_i|\boldsymbol{\gamma}) = \begin{cases} \boldsymbol{\beta}_1 & \text{with probability } w_{i1}(\boldsymbol{\gamma}) \\ \boldsymbol{\beta}_2 & \text{with probability } w_{i2}(\boldsymbol{\gamma}) \\ \vdots & \vdots \\ \boldsymbol{\beta}_Q & \text{with probability } w_{iQ}(\boldsymbol{\gamma}) \end{cases}, \tag{1}$$

where individual i belongs to class q with probability w_{iq} $(q = 1, \dots, Q)$, such that $\sum_q w_{iq} = 1$ and $w_{iq} > 0$; $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_Q)$ is the set of parameters that describe the stochastic assignment to classes. The discrete mixing distribution in (1) (or class assignment/membership probability) is unknown to the analyst (as is the number of classes). Given Q, the most widely used formulation for w_{iq} is the semi-parametric multinomial logit (Shen, 2009; Greene and Hensher, 2003):

$$w_{iq}(\boldsymbol{\gamma}) = \frac{\exp\left(\mathbf{h}_i^\top \boldsymbol{\gamma}_q\right)}{\sum_{q=1}^{Q} \exp\left(\mathbf{h}_i^\top \boldsymbol{\gamma}_q\right)};\; q = 1, \dots, Q,\; \boldsymbol{\gamma}_1 = 0,$$

where $\mathbf{h}_i$ denotes a vector of socio-economic characteristics that determine assignment to classes. The parameters of the first class –in this case[4]– are normalized to zero for identification of the model. Note that one could omit any socio-economic covariate as a determinant of the class assignment probability. Under this scenario of constant assignment, the class probabilities simply become constants of the form:

$$w_{iq}(\gamma) = \frac{\exp\left(\gamma_q\right)}{\sum_{q=1}^{Q} \exp\left(\gamma_q\right)};\; q = 1, \dots, Q,\; \gamma_1 = 0, \tag{2}$$

where γ_q $(q = 1, \dots, Q)$ is a set of constants used to compute class probabilities (Scarpa and Thiene, 2005).

The unconditional probability of the sequence of choices made by individual i is given by:

$$f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\theta}) = \sum_{q=1}^{Q} w_{iq}\left(\gamma_q\right) \left\{ \prod_{t=1}^{T_i} \prod_{j=1}^{J} \left[\frac{\exp\left(\mathbf{x}_{ijt}^\top \boldsymbol{\beta}_q\right)}{\sum_{j=1}^{J} \exp\left(\mathbf{x}_{ijt}^\top \boldsymbol{\beta}_q\right)} \right]^{y_{ijt}} \right\}.$$

where $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\beta})$ is the vector of parameters of interest at the population level $(\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_Q))$.

Since this probability does not require integration, the estimation for the sample of consumers can be undertaken using the standard maximum likelihood estimator (MLE). However, for a large number of classes quasi-Newton methods may exhibit convergence problems; for these situations, the iterative Expectation-Maximization (EM) algorithm (Bhat, 1997; Train, 2008) can be implemented to retrieve maximum likelihood estimates.

3. Individual-specific estimates

3.1. Conditional parameters at the individual level

For discrete mixing distributions in general, the unconditional[5] probability of the sequence of choices of consumer i is:

$$f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\theta}) = \sum_{q=1}^{Q} f\left(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}_q\right) g(\boldsymbol{\beta}_i|\boldsymbol{\gamma}),$$

One would like to know where each individual parameter $\boldsymbol{\beta}_i$ lies within the population heterogeneity distribution $g(\boldsymbol{\beta}_i|\boldsymbol{\gamma})$. Exploiting Bayes' theorem it is possible to obtain (Revelt and Train, 2000):

$$f(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta}) = \frac{f\left(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}_q\right) g(\boldsymbol{\beta}_i|\boldsymbol{\gamma})}{f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\theta})} = \frac{f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}) g(\boldsymbol{\beta}_i|\boldsymbol{\gamma})}{\sum_{q=1}^{Q} f\left(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}_q\right) g(\boldsymbol{\beta}_i|\boldsymbol{\gamma})},$$

which is the posterior distribution of the individual part-worths (given $\boldsymbol{\theta}$). Note that whereas $g(\boldsymbol{\beta}_i|\boldsymbol{\gamma})$ is the unconditional distribution of preferences in the population, the posterior $f(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta})$ is the conditional distribution of the individual parameter $\boldsymbol{\beta}_i$ –conditional on the sequence of choices $\mathbf{y}_i$ when facing a design matrix of attributes $\mathbf{X}_i$ (i.e. conditional on the observed data) and on the parameters of the distribution of preferences in the population $\boldsymbol{\theta}$.[6]

[4] Any single class can be set as baseline.

[5] Unconditional on the individual parameters.

[6] Train (2009, chap. 11) interprets this result as the density of $\boldsymbol{\beta}$ in the subpopulation of people who would choose sequence $\mathbf{y}_i$ when facing $\mathbf{X}_i$.

The **population conditional expectation** of $\boldsymbol{\beta}_i$ is the **posterior mean**:

$$\overline{\boldsymbol{\beta}}_i = \mathbb{E}[\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta}] = \frac{\sum_{q=1}^{Q} \boldsymbol{\beta}_q f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}_q) g(\boldsymbol{\beta}_i|\boldsymbol{\gamma})}{\sum_{q=1}^{Q} f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}_q) g(\boldsymbol{\beta}_i|\boldsymbol{\gamma})}. \tag{3}$$

The conditional expectation of the preference parameters $\boldsymbol{\beta}_i$ specific to consumer i in (3) are generally different from the mean $\boldsymbol{\beta}$ of the unconditional distribution $g(\boldsymbol{\beta}_i|\boldsymbol{\theta})$.[7]

In a Bayesian setting with a quadratic loss, the Bayes decision for a point estimate of the posterior distribution is the posterior mean. Considering the LC-MNL assumptions, a **Bayes point estimator** of the posterior (3) is then the conditional expectation:

$$\widehat{\overline{\boldsymbol{\beta}}}_i = \widehat{\mathbb{E}}\left[\boldsymbol{\beta}_i \middle| \mathbf{y}_i, \mathbf{X}_i, \widehat{\boldsymbol{\theta}}\right] = \frac{\sum_{q=1}^{Q} \widehat{\boldsymbol{\beta}}_q \widehat{w}_{iq}(\boldsymbol{\gamma}) \prod_{t=1}^{T_i} \prod_{j=1}^{J} \left[\frac{\exp\left(\mathbf{x}_{ijt}^{\top}\widehat{\boldsymbol{\beta}}_q\right)}{\sum_{j=1}^{J}\exp\left(\mathbf{x}_{ijt}^{\top}\widehat{\boldsymbol{\beta}}_q\right)}\right]^{y_{ijt}}}{\sum_{q=1}^{Q} \widehat{w}_{iq}(\boldsymbol{\gamma}) \prod_{t=1}^{T_i} \prod_{j=1}^{J} \left[\frac{\exp\left(\mathbf{x}_{ijt}^{\top}\widehat{\boldsymbol{\beta}}_q\right)}{\sum_{j=1}^{J}\exp\left(\mathbf{x}_{ijt}^{\top}\widehat{\boldsymbol{\beta}}_q\right)}\right]^{y_{ijt}}}. \tag{4}$$

Similarly, an estimator of the **posterior membership probability** is (Kamakura and Russell, 1989; Greene and Hensher, 2003):

$$\widehat{\pi}_{iq}(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta}) = \frac{\widehat{w}_{iq}(\boldsymbol{\gamma}) \prod_{t=1}^{T_i} \prod_{j=1}^{J} \left[\frac{\exp\left(\mathbf{x}_{ijt}^{\top}\widehat{\boldsymbol{\beta}}_q\right)}{\sum_{j=1}^{J}\exp\left(\mathbf{x}_{ijt}^{\top}\widehat{\boldsymbol{\beta}}_q\right)}\right]^{y_{ijt}}}{\sum_{q=1}^{Q} \widehat{w}_{iq}(\boldsymbol{\gamma}) \prod_{t=1}^{T_i} \prod_{j=1}^{J} \left[\frac{\exp\left(\mathbf{x}_{ijt}^{\top}\widehat{\boldsymbol{\beta}}_q\right)}{\sum_{j=1}^{J}\exp\left(\mathbf{x}_{ijt}^{\top}\widehat{\boldsymbol{\beta}}_q\right)}\right]^{y_{ijt}}},$$

which gives the probability for individual i belonging to class q given observed choices. An empirical strategy for assigning individuals to specific segments is to use the class with the highest posterior $\widehat{\pi}_{iq}(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta})$ (DeSarbo et al., 1995).

In sum, individual parameters in general $\boldsymbol{\beta}_i$ are realizations of a random process with conditional distribution $f(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta})$ –which is specific to the consumer– and unconditional distribution $g(\boldsymbol{\beta}_i|\boldsymbol{\gamma})$. The conditional point estimator $\widehat{\overline{\boldsymbol{\beta}}}_i$ (posterior mean) is just an estimator of the expected value of the conditional distribution $f(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta})$ (posterior distribution of the individual part-worths). Note that the LC-MML conditional point estimates can be written as an explicit function of the posterior membership probabilities:

$$\widehat{\overline{\boldsymbol{\beta}}}_i = \sum_{q=1}^{Q} \widehat{\boldsymbol{\beta}}_q \widehat{\pi}_{iq}(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta}), \tag{5}$$

which can be easily implemented to derive estimates of the conditional expectation of individual parameters. Revelt and Train (2000) and Train (2009, chap. 11), basically recurring to the Bernstein-von Mises theorem, discuss the conditions under which the posterior mean $\widehat{\overline{\boldsymbol{\beta}}}_i$ can be considered a valid estimate of $\boldsymbol{\beta}_i$. Traditional asymptotics ($N\to\infty$) but keeping the number of choice situations T fixed is not sufficient for statistical consistency of the conditional expectation: without new information about the choices made by consumer i one cannot retrieve the true parameter $\boldsymbol{\beta}_i$.[8] However, if T rises without bound, then $\widehat{\overline{\boldsymbol{\beta}}}_i$ is a consistent estimator of the true $\boldsymbol{\beta}_i$. The intuition behind this convergence process is that $f(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta})$ will tend to move toward $\boldsymbol{\beta}_i$ as T rises, becoming more concentrated. In fact, as $T\to\infty$, the conditional distribution, and hence its expected value, converges to the true value of $\boldsymbol{\beta}_i$. Thus:

$$\widehat{\overline{\boldsymbol{\beta}}}_i \xrightarrow{p} \boldsymbol{\beta}_i \quad \text{as} \quad T\to\infty.$$

3.2. Standard errors of the conditional expectation of individual parameters

In addition to deriving point estimates, one may also desire to make interval inference for the conditional expectation $\widehat{\overline{\boldsymbol{\beta}}}_i$. For example, an analyst may need confidence intervals at the individual level to tell whether elements in $\widehat{\overline{\boldsymbol{\beta}}}_i$ are positive (desired feature), negative (undesired feature), or zero (irrelevant feature). In the MMNL setting, two approaches to interval inference have been proposed: 1) exploiting the conditional variance of the posterior $f(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta})$, and 2) accounting for the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$. In fact, the use of the asymptotic distribution was proposed by Revelt and Train (2000) as an alternative method to derive posterior means. We

[7] It is worth mentioning that the conditional expectation can be computed for any statistic of $\boldsymbol{\beta}_i$, such as willingness-to-pay and marginal effects (see for example Scarpa and Thiene, 2005). Formally, if $k(\boldsymbol{\beta}_i)$ is a function of $\boldsymbol{\beta}_i$, then

$$\mathbb{E}[k(\boldsymbol{\beta}_i)|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta}] = \frac{\sum_{q=1}^{Q} k(\boldsymbol{\beta}_q) f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}_q) g(\boldsymbol{\beta}_i|\boldsymbol{\gamma})}{\sum_{q=1}^{Q} f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}_q) g(\boldsymbol{\beta}_i|\boldsymbol{\gamma})}.$$

[8] For a larger N but fixed T, a better estimated $\boldsymbol{\theta}$ yields a better estimated mean of the posterior, but that posterior mean is not equal to the parameter of the individual.

note that in a full Bayesian setting, the posterior variance of the individual estimates can be used for (credible) interval estimation (see Daziano and Achtnicht, 2014).

3.2.1. Precision of the posterior mean (conditional variance)

From the general definition of variance, given the point estimate $\widehat{\boldsymbol{\theta}}$ it is possible to derive the **posterior variance** of the conditional distribution of $\boldsymbol{\beta}_i$ from: (cf. Greene, 2012, chap. 15 where this expression is used for MMNL)

$$\widehat{V}_i = \widehat{\mathrm{Var}}\left(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \widehat{\boldsymbol{\theta}}\right) = \widehat{\mathbb{E}}\left[\boldsymbol{\beta}_i^2|\mathbf{y}_i, \mathbf{X}_i, \widehat{\boldsymbol{\theta}}\right] - \left(\widehat{\mathbb{E}}\left[\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \widehat{\boldsymbol{\theta}}\right]\right)^2. \tag{6}$$

This conditional variance is related to the Bayes precision with a quadratic loss, which is the posterior variance that can be used to construct credible intervals. For example, a highest density posterior (HDP) 95% credible interval is generally found by identifying the narrowest possible interval that contains 95% of the posterior probability mass. In a frequentist setting, Craig et al. (2005) constructed approximate 95% confidence intervals from the posterior distribution $f(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta})$ by taking the conditional expectation of a specific parameter plus and minus 2.5 conditional standard deviations of the same attribute.

We recall that for a low number of choice situations T the conditional expectation is not a good representation of the individual parameter of interest $\boldsymbol{\beta}_i$. Thus, $\widehat{\overline{\boldsymbol{\beta}}}_i$ may be far from $\boldsymbol{\beta}_i$ for a low T and the interval constructed around the posterior mean using the conditional variance may not contain the true $\boldsymbol{\beta}_i$. Even though the conditional expectation does approach the true parameter of the individual as T increases, more observations per individual also reduce uncertainty in the determination of the posterior (which is measured by the conditional variance). At the limit, and as a result of $\widehat{\overline{\boldsymbol{\beta}}}_i \xrightarrow{p} \boldsymbol{\beta}_i$ as $T\to\infty$, the conditional variance converges to $\mathbf{0}$:

$$\widehat{\mathrm{Var}}\left(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \widehat{\boldsymbol{\theta}}\right)\to\mathbf{0} \quad \text{as} \quad T\to\infty, \tag{7}$$

leading to conditional standard deviations that become too low for a larger number of choice situations.

In addition to the issues discussed in the previous paragraph, there is another problem in the use of the conditional approach alone to the derivation of standard errors and actual confidence intervals. Both Equations (3) and (6) condition on $\widehat{\boldsymbol{\theta}}$. From a frequentist point of view, using a different sample will lead to a different point estimate of the population-level parameters of preference heterogeneity. The conditional variance, calculated as a function of a single $\widehat{\boldsymbol{\theta}}$, thus neglects sampling variability that is the base for the derivation of standard errors and the construction of confidence intervals. This fact is acknowledged in Greene et al. (2014).

Note that the unconditional variance of $\boldsymbol{\beta}_i$ can be written as the sum of the expected conditional variance of $\boldsymbol{\beta}_i$ and the variance of the expectation:

$$\mathrm{Var}(\boldsymbol{\beta}_i) = \mathbb{E}\left[\widehat{\mathrm{Var}}\left(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \widehat{\boldsymbol{\theta}}\right)\right] + \mathrm{Var}\left[\widehat{\mathbb{E}}\left[\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \widehat{\boldsymbol{\theta}}\right]\right]. \tag{8}$$

At the limit, when $T\to\infty$, $\widehat{\mathrm{Var}}(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \widehat{\boldsymbol{\theta}})\to 0$ and $\mathrm{Var}[\widehat{\mathbb{E}}[\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \widehat{\boldsymbol{\theta}}]]\to\mathrm{Var}(\boldsymbol{\beta}_i)$. That is, the sample variance of the estimated conditional expectation converges to the unconditional variance in the population. Furthermore, because variances are by definition non-negative, the variance of the conditional expectation is smaller than the variance of $\boldsymbol{\beta}_i$ calculated from the unconditional distribution of the parameter.[9]

3.2.2. Standard errors using the sampling distribution of the population parameter θ

Because of the neglect of the sampling variability in the conditional approach, Revelt and Train (2000) proposed, for MMNL, another estimator of the conditional expectation that takes into consideration the sampling distribution of $\boldsymbol{\theta}$. If $\mathcal{N}(\boldsymbol{\theta}|\overline{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ denotes the multivariate normal density of $\boldsymbol{\theta}$ with mean $\overline{\boldsymbol{\theta}}$ and covariance $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, then for any random parameter model the expected value of the individual-level preference parameters $\boldsymbol{\beta}_i$ conditional on the population parameters $\overline{\boldsymbol{\theta}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ (but not on a specific value of $\boldsymbol{\theta}$)[10], and on the individual choices $\mathbf{y}_i$, is:

$$\mathbb{E}\left[\boldsymbol{\beta}_i|\mathbf{y}_i, \overline{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right] = \int_{\boldsymbol{\theta}} \mathbb{E}[\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta}]\mathcal{N}\left(\boldsymbol{\theta}|\overline{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)d\boldsymbol{\theta}. \tag{9}$$

A Monte Carlo approximation to this expectation (that integrates out $\boldsymbol{\theta}$) is obtained by calculation of the empirical mean evaluated at pseudo or quasirandom draws of $\boldsymbol{\theta}$ from the asymptotic distribution of the estimator. Furthermore, the authors propose to generate the draws from the asymptotic distribution evaluated at the maximum (simulated) likelihood estimates $\widehat{\overline{\boldsymbol{\theta}}}$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}$, which is basically exploiting the method of Krinsky and Robb (Krinsky and Robb, 1986, 1990). The steps for this Krinsky-Robb (KR) procedure –which were recently also formalized by Greene et al. (2014) for MMNL– are the following for a general logit-type random parameter model:

- Estimate the model using the maximum (simulated) likelihood estimator. Consider $r \in \{1, \ldots, R\}$, and start with $r = 1$.

[9] For some implications of using the conditional variance for computing willingness-to-pay measures see Daly et al. (2012).

[10] We recall that $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\beta})$ in LC-MNL, where $\boldsymbol{\beta}$ contains all Q preference vectors (i.e. each $\boldsymbol{\beta}_q$), and $\boldsymbol{\gamma}$ contains all $Q-1$ vectors of the parameters of the unconditional stochastic assignment to classes (cf. $\boldsymbol{\beta}$ in MMNL that represents the vector of random preferences in the population).

1. Take a random draw of $\boldsymbol{\theta}^r$ from $\mathcal{N}(\widehat{\overline{\boldsymbol{\theta}}}, \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}})$, which is the estimated asymptotic distribution of $\widehat{\boldsymbol{\theta}}$.
2. Use Equation (5), but substituting $\boldsymbol{\theta}^r$ for $\widehat{\boldsymbol{\theta}}$, to calculate $\widehat{\overline{\boldsymbol{\beta}}}_i^r = \widehat{\mathbb{E}}[\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta}^r]$
3. Update $r = r + 1$, and go back to step 1

- Repeat for a large number of repetitions R (.e.g., $R = 1000$)
- Calculate the **empirical mean** of the individual-level parameters:

$$\widehat{\mathbb{E}}\left[\boldsymbol{\beta}_i|\mathbf{y}, \overline{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right] = \frac{1}{R}\sum_{r=1}^{R}\widehat{\mathbb{E}}[\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta}^r]. \tag{10}$$

Given the R KR-estimates and the empirical mean, confidence intervals (and standard errors) can be computed in two ways. The first procedure is to directly derive the variance over the R KR-estimates:

$$\widehat{V}_{KR,i} = \frac{1}{R-1}\sum_{r=1}^{R}\left(\widehat{\overline{\boldsymbol{\beta}}}_i^r - \widehat{\mathbb{E}}\left[\boldsymbol{\beta}_i|\mathbf{y}, \overline{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right]\right)^2. \tag{11}$$

A valid estimate of the standard error of the individual-level parameter of the kth attribute is then the following empirical standard deviation (cf. Train, 2009; Greene et al., 2014, for MMNL):

$$se_{KR}\left(\widehat{\overline{\beta}_{ki}}\right) = \left[\widehat{V}_{KR,k}\right]^{1/2},$$

which unlike the conditional variance, does take into consideration frequentist sampling variability. This standard error can be used as usual for constructing confidence intervals at the individual level.

The second procedure consists in sorting all $\widehat{\overline{\boldsymbol{\beta}}}_i^r$ values in ascending order, and then dropping the top and bottom 2.5% of the sorted realizations of the individual-level parameters. This procedure is known as the percentile (parametric) bootstrap in the statistics literature.[11]

4. Monte Carlo study

To assess and compare the performance of the two interval estimation approaches to conditional individual-specific preference parameters, we ran a full Monte Carlo study. We assumed that hypothetical individuals are faced to different sets of scenarios in which they must choose among 3 alternatives ($J = 3$). Each alternative is characterized by two attributes. In the simulated data generating process, the true latent utility for individual i, alternative j, and choice situation t is:

$$U_{ijt}^* = \beta_{1i}x_{1ijt} + \beta_{2i}x_{2ijt} + \varepsilon_{ijt},$$

where ε_{ijt} is distributed Type 1 Extreme Value; x_1 is assumed to be independent and standard normally distributed, whereas x_2 is dummy variable created from a uniform distribution: $1(u < 0.5)$, where $u \sim U[0, 1]$.

To assess the impact of the number of classes in the estimation of the confidence intervals we generated two experiments. In the first one, the random parameters are distributed following a discrete distribution with 2 classes, namely:

$$\beta_{1i} = \begin{cases} -2 & \text{with probability} \quad 0.6 \\ 2 & \text{with probability} \quad 0.4 \end{cases}, \tag{12}$$

and

$$\beta_{2i} = \begin{cases} -0.5 & \text{with probability} \quad 0.6 \\ 0.5 & \text{with probability} \quad 0.4 \end{cases}. \tag{13}$$

The second experiment assumes 3 classes according the following discrete distribution:

$$\beta_{1i} = \begin{cases} -2 & \text{with probability} \quad 0.25 \\ 0 & \text{with probability} \quad 0.5 \\ 2 & \text{with probability} \quad 0.25 \end{cases}, \tag{14}$$

and

$$\beta_{2i} = \begin{cases} -0.5 & \text{with probability} \quad 0.25 \\ 1 & \text{with probability} \quad 0.5 \\ 1.5 & \text{with probability} \quad 0.25 \end{cases}. \tag{15}$$

[11] It is should be noted that the percentile approach is likely to perform better if the sampling distribution is non-symmetrical. See for example Hole (2007) for a deeper discussion.

Table 1
Simulation results for LC-MNL model with 2 Classes. The results are averaged over S.

	Point Estimate					Sampling Distribution					
	Mean	*Bias*	*SD*	*Cov*	*S*	*Mean*	*Bias*	*SD*	*Cov A*	*S*	*Cov B*
	Panel A: N = 300										
T = 1	−0.389	0.862	1.779	0.807	277	−0.386	0.866	1.726	0.717	276	0.711
T = 5	−0.403	0.117	1.955	0.193	300	−0.403	0.118	1.954	0.934	300	0.932
T = 10	−0.399	0.064	1.965	0.010	300	−0.399	0.064	1.965	0.950	300	0.947
T = 20	−0.401	0.047	1.965	0.000	300	−0.401	0.047	1.965	0.949	300	0.948
T = 50	−0.400	0.029	1.966	0.000	300	−0.400	0.029	1.966	0.938	300	0.938
	Panel B: N = 1000										
T = 1	−0.384	0.728	1.668	0.894	300	−0.383	0.733	1.653	0.629	300	0.623
T = 5	−0.403	0.071	1.950	0.257	300	−0.403	0.071	1.950	0.929	300	0.927
T = 10	−0.398	0.036	1.963	0.015	300	−0.398	0.036	1.963	0.950	300	0.947
T = 20	−0.400	0.025	1.960	0.000	300	−0.400	0.025	1.960	0.949	300	0.943
T = 50	−0.400	0.017	1.961	0.000	300	−0.400	0.017	1.961	0.939	300	0.940

Databases were constructed of size $N = 300$, so that we can compare our results with those in Revelt and Train (2000), as well as a larger sample size of $N = 1000$. The number of choice situations per consumer is such that $t \in \{1, 5, 10, 20, 50\}$. In each simulation scenario, $N \times J \times T$ values were randomly generated for x_{kijt} and ε_{ijt}. We created $S = 300$ databases (trials) by repeating the process and taking a new set of random draws for each iteration, whereas the true individual parameters β_{i1} and β_{i2} were held fixed.

The analysis is focused on individual-level point estimates and intervals. For a given simulated sample $s \in \{1, \ldots, S\}$, individual-level parameters $\widehat{\overline{\boldsymbol{\beta}}}_{is}$ are derived either using the conditional expectation (posterior mean) of Equation (5), or Equation (10) with $R = 1000$ when accounting for the sampling variance.

For each case in the simulation plan, we compute the following statistics for each parameter:

$$\text{(Sample) Mean} = \overline{\widehat{\overline{\beta}}}_s = \frac{1}{S}\sum_{s=1}^{S}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{\overline{\beta}}_{is}\right), \tag{16}$$

$$\text{Bias} = \frac{1}{S}\sum_{s=1}^{S}\left(\frac{1}{N}\sum_{i=1}^{N}\left|\widehat{\overline{\beta}}_{is} - \beta_i\right|\right), \tag{17}$$

$$\text{SD} = \frac{1}{S}\sum_{s=1}^{S}\left[\frac{1}{N-1}\sum_{i=1}^{N}\left(\widehat{\overline{\beta}}_{is} - \overline{\widehat{\overline{\beta}}}_s\right)^2\right]^{1/2}. \tag{18}$$

In addition to the statistics above, empirical coverage (*Cov*) of the interval estimates was calculated. An empirical coverage probability is the proportion of simulated samples for which the estimated 95% interval includes the true individual-level parameter. If interval inference is correct, then one should observe a 95% coverage. If coverage is lower than 95%, then the estimated intervals are too narrow, on average. If coverage is larger than 95%, then the estimated intervals are too wide, on average.

For the results using the KR procedure, we compute the confidence interval in the two described ways: 1) we compute the standard deviation of the resampling procedure and then construct the CI for each individual (*Cov A*), 2) we drop the top and bottom 2.5% of the observations from the resampling procedure (*Cov B*).

5. Results

5.1. Conditional individual-level preference parameters

In this section we revisit the Monte Carlo results of Revelt and Train (2000) to assess the ability of the latent class logit model for retrieving the true individual parameters. Table 1 displays the Monte Carlo results for the sample means of the conditional expectations (Equation (16)) using point and sampling distribution estimates for the experiments with $Q = 2$: panel A shows the results for $N = 300$ and panel B shows the results for $N = 1000$. We only report the results for β_1, averaged over N and S, where S is the total number of trials that either converged or did not have atypical high values of the posterior means $\widehat{\overline{\beta}}_{1i}$.[12] Estimation was carried out using the 'gmnl' package in R (Sarrias and Daziano, 2017).

Recall that we should observe $\widehat{\overline{\beta}}_i \rightarrow \beta_i$ as $T \rightarrow \infty$. This convergence pattern implies that the conditional expectation for each individual should collapse to its true coefficient and the distribution of the conditional expectations should converge to the true unconditional distribution. Given the true parameters in the simulation study with two classes, the population mean and standard deviation of β_{1i} are

[12] Trials with individual estimates over $|-4|$ were considered unusual.

Table 2
Simulation results for LC-MNL model with 3 Classes. The results are averaged over S.

	Point Estimate					Sampling Distribution					
	Mean	*Bias*	*SD*	*Cov*	*S*	*Mean*	*Bias*	*SD*	*Cov A*	*S*	*Cov B*
	N = 1000										
T = 1	−0.236	1.954	2.286	0.900	300	−0.146	2.267	2.603	0.601	292	0.566
T = 5	0.001	0.412	1.279	0.946	300	0.001	0.414	1.276	0.455	300	0.450
T = 10	0.001	0.159	1.371	0.900	300	0.001	0.159	1.370	0.720	300	0.718
T = 20	−0.017	0.086	1.384	0.414	300	−0.007	0.060	1.402	0.877	286	0.876
T = 50	−0.003	0.021	1.413	0.010	300	−0.001	0.015	1.417	0.952	297	0.950

−0.4 and 1.9595, respectively.[13]

The results of Table 1 are consistent with those found by Revelt and Train (2000) for the mixed logit data generating process. Coverage is discussed in subsection 5.3. As the number of choice situations increases, the average and standard deviation of the individual-specific estimates converges to both the true mean and standard deviation of the population discrete distribution. Moreover, both estimators in terms of the sample means (and conditional expectations at the individual level) are very similar to each other. It is also important to notice that the bias of the individual parameters decrease with the number of choice situations. For example, with 50 choice situations, the difference between the true value and point-estimate of the conditional mean for each individual is only 0.029 ($N = 300$) and 0.017 ($N = 1000$). Increasing the number of individuals improves the results. When comparing the absolute difference of $\mathbb{E}(\beta)$ from β for $N = 300$ with those for 1000 individuals, we can observe that bias is reduced by nearly 55%.

Table 2 shows the results for the experiment which the number of classes is increased to 3 ($Q = 3$) and the number of individuals per sample is $N = 1000$. For this setting, the true population mean and standard deviation are 0 and 1.4142, respectively. Similarly to the case with $Q = 2$, the results from Table 2 show that the means and standard deviations are very close to those of the population distribution when the number of choice situations is 50. Furthermore, the magnitude of the bias in the case with 3 classes is similar in magnitude to the bias with 2 classes. We note, however, that the number of samples with atypical values for the sampling distribution estimates increases with a higher number of classes. This reveals that the KR procedure is more stringent to detect atypical values for each simulation and draw r.

To get some visual insights, Fig. 1 displays the (continuous) kernel density of the true unconditional distribution $g(\beta_i)$ along with the distribution of the conditional expectation of the individual part-worths for β_{1i}, for different settings of T. Panels A and B show the results for $Q = 2$, whereas panels C and D show the results for $Q = 3$ given $N = 1000$. By looking at the figures one can confirm that the density of the conditional expectations converges to the conditional population as the number of choice situations increases, as expected. In other words, as we have more information of the choices made by the individuals, we can better identify each individual-specific estimate.

5.2. Class assignment

Another interesting question is whether the posterior membership probability can be used effectively to assign individuals into classes. Since the estimated probabilities give us the most likely allocation of individuals to the classes, as discussed when introducing the model individuals could be allocated into classes by assigning them into the class with the highest posterior probability.

Panel A of Table 3 reports the class-assignment using the posterior probabilities. Column 1 reports the percentage of individual that was correctly allocated into the true class using the experiment with 3 classes and 1000 individuals. Again, the results are the average over S. Interestingly, with only 10 choice situations per individual we obtain, in average, a 95% of correct allocation of individuals and with 50 choice situations the correct prediction is almost 100%. Similarly, when the number of choice situation increases the shares of the individuals in each class converges to the true share (25%, 50% and 25%).

Another approach for computing the shares for each class consists on using the structural parameters (population) instead of the posterior probabilities. Panel B shows the results using Equation (2). Note that these shares are based on the regression parameters instead of individual estimates. The results from Panel B are very close to those based on posterior probabilities. However, the main drawback is that these are aggregate class sizes: we are not able to identify to which class each individual belongs.

5.3. Standard errors

In this section we evaluate the performance of both methods for interval inference of the individuals' conditional means. In particular, we use empirical coverage probabilities to assess standard error performance, using a confidence level of 95%. The results for the coverage probabilities are shown in Tables 1 and 2.

The first result is that the coverage probability for the point-estimate method is too low and it gets lower as the number of choice situations increases. This is in line with our theoretical discussion in Section 3.2.1. As stated, since the conditional expectation will collapse to the true individual parameter, the sample variance of the estimated individual posterior mean computed using the point

[13] This values are calculated as: $\mathbb{E}(\beta_{1i}) = -2{\cdot}0.6 + 2{\cdot}0.4$ and $\mathrm{Var}(\beta_{1i}) = (-2+0.4)^2{\cdot}0.6 + (2+0.4)^2{\cdot}0.4 = 3.84$.

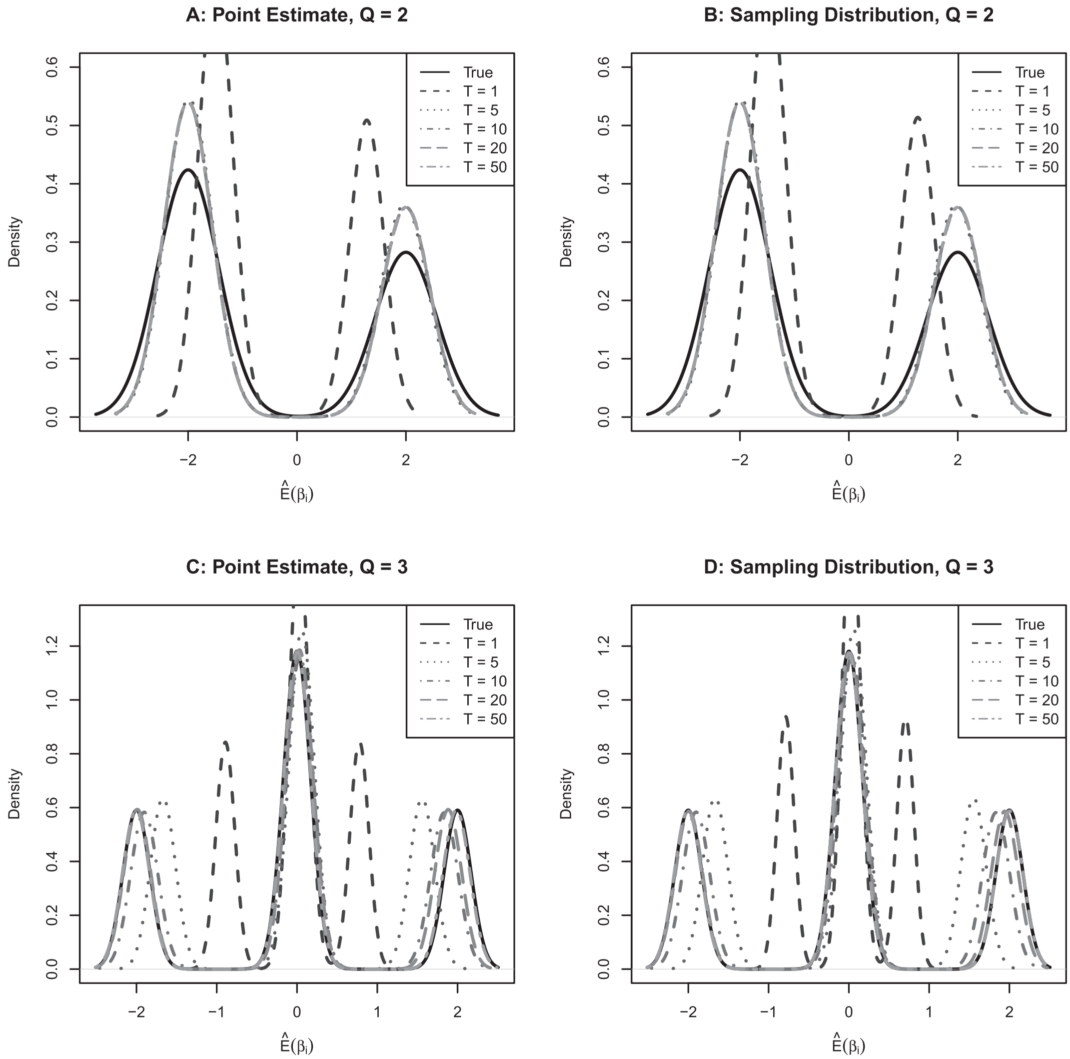

Figure 1. Kernel densities of Individual-specific estimates. The true distribution corresponds to the unconditional (population) distribution. The unconditional distributions corresponds to the estimated distribution of the conditional means. The densities were computed using the average over S constructed data sets.

estimate of $\widehat{\boldsymbol{\theta}}$ will approach to zero as the number of choice situations increases without bound. The standard errors using the sampling distribution of $\widehat{\boldsymbol{\theta}}$, either using *Cov A* or *Cov B*, do not present this problem as corroborated by the Monte Carlo results. With 2 classes the 95 percent coverage is achieved with around 20 choice situations per individuals. However, with 3 classes the number of choice situations required to achieve the same nominal coverage probability is higher. This may be due to the increase in the number of estimated parameters and consequent increase in the draws from the joint distribution of the parameters. Note that using either the standard deviation or the quantiles of the draws gives approximately the same results.

5.4. *Misspecification*

Using the experiment with 3 classes and $N = 1000$, we analyze the effects of misspecification on the conditional estimates when we force the estimation of a LC-MNL model with just 2 classes (instead of the underlying 3). The results are presented in Table 4. As expected, the bias is higher when the model is misspecified and does not substantially decrease when the number of choice situations increases. This result can be explained by looking at Fig. 2. Individuals for class 3 are basically forced to belong to class 2 as T increases. The interval estimates show the same pattern as in the previous tables, and the Krinsky and Robb confidence intervals are less accurate in this case, which is to be expected since results are based on biased estimates.

6. Discussion and conclusions

In this paper we have applied point and interval estimators of individual-specific parameters of logit models with discrete

Table 3
Predicted shares for each class. Panel A based on posterior probabilities. Panel B based on structural parameters. Simulation results for LC-MNL Model with 3 classes. Results averaged over S.

	% Correct prediction	% Class 1	% Class 2	% Class 3
	Panel A: Based on posterior probability			
T = 1	0.512	0.278	0.419	0.302
T = 5	0.860	0.254	0.490	0.256
T = 10	0.953	0.252	0.495	0.253
T = 20	0.921	0.250	0.493	0.257
T = 50	0.996	0.250	0.502	0.248
	Panel B: Based on Logit formula			
T = 1		0.262	0.443	0.295
T = 5		0.249	0.501	0.250
T = 10		0.250	0.500	0.250
T = 20		0.250	0.499	0.250
T = 50		0.250	0.502	0.248

Table 4
Simulation results for LC model

	Point Estimate					Sampling Distribution					
	Mean	*Bias*	*SD*	*Cov*	*S*	*Mean*	*Bias*	*SD*	*Cov*	*S*	*Cov*
	N = 1000										
T = 1	−0.038	1.001	0.718	0.700	295	−0.032	0.998	0.672	0.261	288	0.272
T = 5	−0.106	0.819	0.848	0.321	300	−0.106	0.817	0.845	0.044	300	0.046
T = 10	−0.183	0.692	0.951	0.150	300	−0.183	0.692	0.950	0.045	300	0.045
T = 20	−0.210	0.615	1.017	0.030	300	−0.210	0.615	1.017	0.203	286	0.202
T = 50	−0.218	0.600	1.032	0.000	300	−0.218	0.600	1.032	0.238	297	0.237

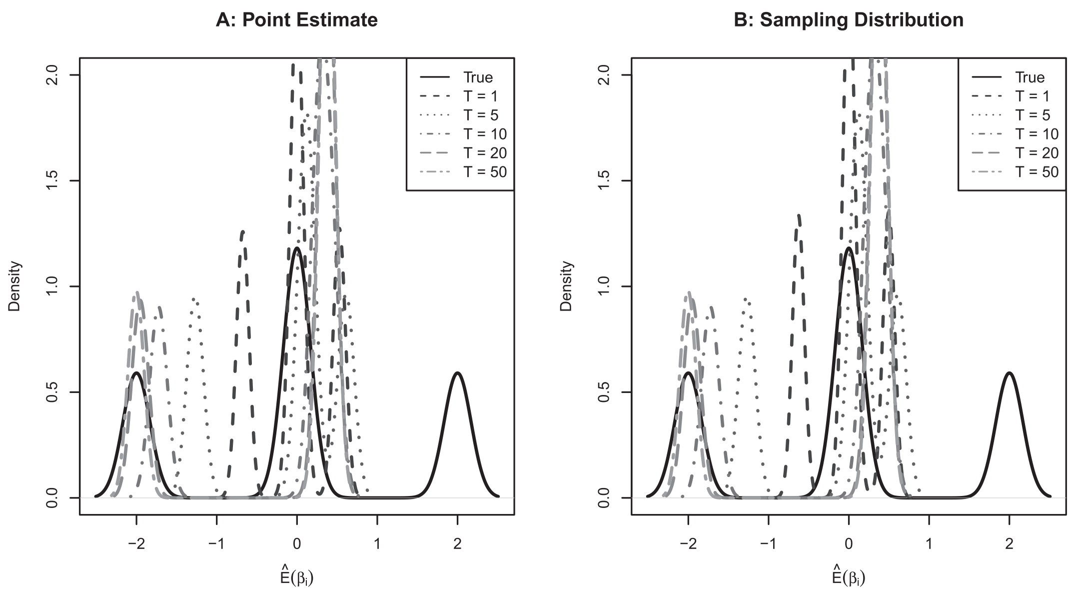

Figure 2. Kernel densities for misspecified individual-specific estimates. The true distribution corresponds to the unconditional (population) distribution. The unconditional distributions corresponds to the estimated distribution of the conditional means. The densities were computed using the average over *S* constructed data sets.

heterogeneity distributions. Using a full Monte Carlo study, we have analyzed the behavior of the conditional estimates focusing on the assessment of the required conditions for true parameter recovery, including actual class membership. We have also assessed empirical coverage from two procedures to derive standard errors that had been proposed in the literature for mixed logit.

Overall, our results are similar to those obtained by Revelt and Train (2000) in the context of MMNL: we show that conditional estimates for LC-MNL are consistent as the number of choice situations increases for any of the two estimators. Yet, we note that the level of bias is lower for LC-MNL. For example, Revelt and Train (2000) show that the absolute difference of $\overline{\overline{\beta}}_i$ from β_i in the context of

MMNL, and using $N = 300$, $T = 50$ and 10,000 Halton draws is about 0.243, whereas our results under a similar setting indicate a bias of about 0.029. However, it should be considered that the MMNL model is much more complex since it requires simulation-aided inference to recover the parameters. Therefore, the difference in bias, among other things, may be due to simulation noise and/or the amount of variability of the individual parameters in the experiment setup.

The results from our experiment has several implications for practitioners using individual-specific estimates from an LC-MNL model. First, it is not reliable to use individual-specific estimates when the number of choice situations is lower than 10, and weak identification is likely for a very low number of choice situations—in particular if a cross-sectional dataset is used ($T = 1$). Below $T = 10$, the amount of bias of the individual-specific estimates is considerable and the variance of the conditional mean is less than the unconditional variance. This is also critical when computing welfare measures from the individual-specific estimates as argued by Daly et al. (2012). Usually, researchers use the conditional estimates instead of the estimated population density for computing willingness-to-pay measures (see for example Scarpa and Thiene, 2005; Shen, 2009; Bujosa et al., 2010; Liao et al., 2015) arguing that this procedure does not show counterintuitive signs for some attributes. However, if T is low the conditional individual-specific estimates, and hence the willingness-to-pay measures computed as the ratio of two parameters, is biased. In fact, in such cases it is not clear whether the individual-specific estimates of WTP actually reflect real sensitivities or reflect simply the bias due to a low number of choice situations. In Daly et al. (2012) words: *"The variance of the conditional means is less than the unconditional variance not because it is a more reasonable estimate, but rather because it incorrectly excludes the variances around the conditional means"*

However there exists a diagnostic tool that can give one some insights about the degree of bias of the conditional expectations (Allenby and Rossi, 1998; Revelt and Train, 2000). As we stated previously and confirmed by our results, the variance of the conditional means should converge to the population variance of the unconditional distribution. Thus, if the average standard deviation of the individuals' conditional preference distribution is similar to the estimated population distribution, the model is correctly specified, accurately estimated and the amount of bias is expected to be low. For example, the results with $N = 1000$ and $T = 10$ show that the average standard deviation of the conditional expectations is 1.370, whereas the true population standard deviation is 1.4142. Thus, variation in the expected coefficients capture more than 96% of the total estimated variation in this coefficient.

Second, even though the individual-specific estimates show some level of bias for a low number of choice situations, the individuals can be allocated to the right class even in these cases. For example, with 10 choice situations, we are able to correctly allocate an average of 95% of the individuals into the right class, and the aggregate shares are precisely estimated.

Third, our results confirm that, at least for the LC-MNL model, the KR procedure gives accurate estimates of the standard errors for the individual conditional parameters. The main problem with the standard errors using point estimates is that they converge to 0 as the number of choice situations increases, therefore it tends to produce standard errors that are too small as T increases without bound. In applied research this implies that using point-estimate standard errors might give the false impression that almost all individuals' parameters are significant, whereas in fact it is a misspecification problem of the standard errors. We also note that the number of choice situations that are required to obtain reliable and accurate standard errors using the KR procedure is large (much larger) compared to those used in empirical work. This in our opinion is a critical point for practitioners: *using the standard errors of the KR procedure requires many choice situations per individual that perhaps in practice is difficult to obtain.*

Another important issue is the misspecification of the number of classes. It is well documented that the most difficult part from the point of view of the analyst when estimating LC-MNL is to choose the number of classes (Greene and Hensher, 2003; Shen, 2009). Since Q is not a free parameter, hypothesis testing cannot be performed and the selection of the optimal Q depends on information criterion such as AIC, BIC or AIC. In terms of the individual parameters, our results confirm that the choice of the number of classes is a process that must be analyzed in more detail in the empirical works. Using an incorrect number of classes leads to bias in the individual coefficients and therefore in the confidence intervals.

Finally, it is important to stress that the simulation results were all based on the LC-MNL model which does not require any simulation technique. Therefore, extrapolating the results regarding the standard errors of the individual-specific estimates to the MMNL context should be done with caution due to simulation error in the latter.

Acknowledgments

This research is based upon work supported by the National Science Foundation CAREER Award CBET-1253475 and FONDECYT Grant 11160104.

References

Allenby, G.M., Rossi, P.E., 1998. Marketing models of consumer heterogeneity. J. Econ. 89 (1), 57–78.
Beharry-Borg, N., Scarpa, R., 2010. Valuing quality changes in caribbean coastal waters for heterogeneous beach visitors. Ecol. Econ. 69 (5), 1124–1139.
Bhat, C., 1997. An endogenous segmentation mode choice model with an application to intercity travel. Transp. Sci. 3, 34–48.
Bujosa, A., Riera, A., Hicks, R.L., 2010. Combining discrete and continuous representations of preference heterogeneity: a latent class approach. Environ. Resour. Econ. 47 (4), 477–493.
Craig, C.S., Greene, W.H., Douglas, S.P., 2005. Culture matters: consumer acceptance of US films in foreign markets. J. Int. Mark. 13 (4), 80–103.
Daly, A., Hess, S., Train, K., 2012. Assuring finite moments for willingness to pay in random coefficient models. Transportation 39 (1), 19–31.
Daziano, R.A., Achtnicht, M., 2014. Accounting for uncertainty in willingness to pay for environmental benefits. Energy Econ. 44, 166–177.
DeSarbo, W., Ramaswamy, V., Cohen, S., 1995. Market segmentation with choice-based conjoint analysis. Mark. Lett. 6, 137–147.
Greene, W., Harris, M.N., Spencer, C., 2014. Estimating the Standard Errors of Individual-specific Parameters in Random Parameters Models. Bankwest Curtin Economics Centre. Working paper.
Greene, W., Hensher, D., Rose, J., 2005. Using classical simulation-based estimators to estimate individual WTP values. In: Scarpa, R., Alberini, A. (Eds.), Applications of Simulation Methods in Environmental and Resource Economics, Volume 6 of the Economics of Non-market Goods and Resources. Springer Netherlands, pp. 17–33.

Greene, W.H., 2012. Econometric Analysis, 7th edition. Prentice Hall.
Greene, W.H., Hensher, D.A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transp. Res. Part B: Methodol. 37 (8), 681–698.
Hensher, D., Greene, W., Rose, J., 2003. Taking Advantage of Priors in Estimation and Posteriors in Application to Reveal Individual-specific Parameter Estimates and Avoid the Potential Complexities of WTP Derived from Population Moments. Institute of Transport Studies, University of Sydney. Working paper, Manuscript.
Hensher, D.A., Greene, W.H., 2003. The mixed logit model: the state of practice. Transportation 30 (2), 133–176.
Hensher, D.A., Greene, W.H., 2010. Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification. Empir. Econ. 39 (2), 413–426.
Hensher, D.A., Greene, W.H., Rose, J.M., 2006. Deriving willingness-to-pay estimates of travel-time savings from individual-based parameters. Environ. Plan. A 38 (12), 2365.
Hess, S., 2014. Latent class structures: taste heterogeneity and beyond. In: Handbook of Choice Modelling, p. 311.
Hess, S., Hensher, D.A., 2010. Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. Transp. Res. Part B: Methodol. 44 (6), 781–790.
Hess, S., Rose, J.M., 2007. Effects of Distributional Assumptions on Conditional Estimates from Mixed Logit Models. ETH, Eidgenössische Technische Hochschule Zürich, Institut für Verkehrsplanung und Transportsysteme. Working paper.
Hole, A.R., 2007. A comparison of approaches to estimating confidence intervals for willingness to pay measures. Health Econ. 16 (8), 827–840.
Huber, J., Train, K., 2001. On the similarity of classical and bayesian estimates of individual mean partworths. Mark. Lett. 12 (3), 259–269.
Hynes, S., Hanley, N., Scarpa, R., 2008. Effects on welfare measures of alternative means of accounting for preference heterogeneity in recreational demand models. Am. J. Agric. Econ. 90 (4), 1011–1027.
Kamakura, W., Russell, G., 1989. A probabilistic choice model for market segmentation and elasticity structure. J. Mark. Res. 26, 379–390.
Keane, M., Wasi, N., 2013. Comparing alternative models of heterogeneity in consumer choice behavior. J. Appl. Econ. 28 (6), 1018–1045.
Krinsky, I., Robb, A.L., 1986. On approximating the statistical properties of elasticities. Rev. Econ. Stat. 68 (4), 715–719.
Krinsky, I., Robb, A.L., 1990. On approximating the statistical properties of elasticities: a correction. Rev. Econ. Stat. 72 (1), 189–190.
Liao, F.H., Farber, S., Ewing, R., 2015. Compact development and preference heterogeneity in residential location choice behaviour: a latent class analysis. Urban Stud. 52 (2), 314–337.
McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–142.
McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. J. Appl. Econ. 15 (5), 447–470.
Revelt, D., Train, K., 2000. Customer-specific Taste Parameters and Mixed Logit: Households' Choice of Electricity Supplier. Department of Economics, UCB. Working paper.
Sarrias, M., Daziano, R., 2017. Multinomial logit models with continuous and discrete individual heterogeneity in r: the gmnl package. J. Stat. Softw. Articles 79 (2), 1–46.
Scarpa, R., Thiene, M., 2005. Destination choice models for rock climbing in the northeastern Alps: a latent-class approach based on intensity of preferences. Land Econ. 81 (3), 426–444.
Shen, J., 2009. Latent class model or mixed logit model? a comparison by transport mode choice data. Appl. Econ. 41 (22), 2915–2924.
Sillano, M., de Dios Ortúzar, J., 2005. Willingness-to-pay estimation with mixed logit models: some new evidence. Environ. Plan. A 37, 525–550.
Train, K., 2008. EM algorithms for nonparametric estimation of mixing distributions. J. Choice Model. 1 (1), 40–69.
Train, K.E., 2009. Discrete Choice Methods with Simulation, 2nd edition. Cambridge University Press.