Individual-specific posterior distributions from Mixed Logit models: Properties, limitations and diagnostic checks☆

Mauricio Sarrias

IDEAR, Department of Economics, Universidad Católica del Norte, Antofagasta, Avenida Angamos 0610, Chile

ARTICLE INFO

ABSTRACT

Individual-specific posterior distributions are an attractive tool for disentangling the tastes for each person in the sample. However, there exists some risks and certain limitations regarding their use. This study reviews and summarizes the theoretical literature about the individual-specific posterior distributions derived from the Mixed Logit model, focusing on their properties, limitations and common pitfalls. It also reviews and analyzes the behavior of some diagnostic checks proposed in the literature for the reliability of such estimates in applied works using Monte Carlo experiments. Finally, this article provides reasonable guidelines for the correct use of individual-specific posterior distributions.

1. Introduction

Since the Revelt and Train (2000)'s work and Train (2009)'s book, the use of conditional estimators derived from conditional posterior distributions has been an attractive tool for disentangling the tastes for each individual in the sample. Conditional estimates, put simply, allow us to know something about respondents' tastes based on their previous choices providing their most likely location on the population distribution.

Due to the attractiveness of eliciting the tastes for each person, individual-specific estimates (or conditional estimates) derived from the Mixed Logit (MIXL) Model have received a lot of attention from the applied literature in different fields. For example, they have been used to compute willingness-to-pay (WTP) measures at the individual level (Sillano and de Dios Ortúzar, 2005; Greene et al., 2005; Hensher et al., 2003, 2006; Hess, 2007; Sandorf et al., 2017; Dumont et al., 2015), to analyze the spatial dependence of tastes (Campbell et al., 2009; Budziński et al., 2018; Abildtrup et al., 2013), to retrieve individual-specific attribute processing strategies (Hess and Hensher, 2010), to create clusters or segments of individuals (Richter and Pollitt, 2018; Huber and Train, 2001) and for predicting the future behavior of the individuals (Train, 2009; Dumont et al., 2015).

Unfortunately, there exists some risks and certain limitations regarding the use of conditional estimates. For example, Revelt and Train (2000) have already shown that individual-specific estimates are consistent if and only if, given a fixed number of individuals, the number of choice situations increases without bound.[1] That is, we need several choice situations per individual to learn something about the preferences of each respondent: if we could observe infinitely how individuals react to changes in attributes and their choices in terms of these changes, in theory, we could elucidate the specific tastes of individuals. However, there exist published articles using individual-specific estimates with a very low number of choice situations without reporting any type of diagnostic check to analyze

---

[1] For a similar study in the Latent Class Multinomial Logit model see Sarrias and Daziano (2018).

whether the estimates are reliable to be used in a second-step procedure.

In addition, it seems that there is still some confusion regarding the relationship between the variance of the individual-specific estimates and the variance of the population distribution of the random parameters. Some researchers use individual-specific estimates arguing that they show more plausible results in terms of their domain. For example, it is common to find in the applied literature claims such that "*conditional estimates show more reasonable estimates* or comments such as "*using conditional means sign violations of the coefficients disappears*", especially when computing individual-specific WTP. But, as briefly mentioned by Daly et al. (2012), such claims ignore the fact that the variance of the conditional mean will be lower than the variance of the population distribution of the random parameters (or unconditional population), specially when the number of choice situations per individual is low. In other words, the apparent better fit of the individual estimates in terms of sign and values might be an artifact of the statistical behavior of the conditional estimates when the number of choice situations is not large enough.

Another problem not yet analyzed in depth, and pointed out by Hess (2010), is the potential impact of the misspecification of the parametric population distribution on the individual estimates. As argued by Hess (2010), researchers should analyze the impact of assumptions made for the unconditional distributions on the shape of the conditional distributions. Failing to choose the most adequate distribution for the random parameters might invalidate the use of conditional estimates and lead to misleading conclusions.

Thus, the first objective of this paper is to review and summarize the properties, limitations and common pitfalls when using individual-specific estimates in the context of the MIXL model and reinforce their understanding. Relying on the previous work of Revelt and Train (2000), Daly et al. (2012) and Hess (2010), this study revisits the theoretical properties of individual-specific estimator regarding to their consistency and the relationship between the conditional and unconditional distribution of tastes. The second objective is to provide reasonable guidelines for the correct use of individual-specific tastes under a well-specified specification and analyze the potential problems under misspecification, extending the previous work done by Hess (2010). To achieve this goal, the study extends the Monte Carlo experiment carried out by Revelt and Train (2000) by analyzing the behavior of the conditional estimates under misspecification and by including new measures as diagnostic tools.

The rest of the paper is organized as follows. Section 2 briefly reviews the MIXL model. Section 3 summarizes the main theoretical properties of the individual-specific estimates. Section 4 explains the Monte Carlo setup, whereas Section 5 presents the results. Finally, Section 6 discusses the results and concludes.

2. The Mixed Logit model

Assume that each individual ($i = 1, ..., N$) faces a choice among J alternatives in each of T choice situations.[2] Then, the utility associated with each alternative $j = 1, ..., J$ for individual i in choice situation t is:

$$U_{ijt} = \mathbf{x}_{ijt}`\boldsymbol{\beta}_i + \varepsilon_{ijt},$$

where $\mathbf{x}_{ijt}$ is a $K \times 1$ vector of attributes of the alternatives. The important characteristic of the MIXL model is the vector of tastes $\boldsymbol{\beta}_i$, which is assumed to vary across individuals according to $g(\boldsymbol{\beta}_i|\boldsymbol{\Omega})$, where $g(\cdot)$ is some well-behaved parametric distribution with parameters given by $\boldsymbol{\Omega}$. In general, $\boldsymbol{\Omega}$ contains the mean and variance of such distribution.

Let $\mathbf{y}_i = \{y_{i1}, y_{i2}, ..., y_{iT}\}$ be the sequence of choices made by individual i, and $\mathbf{X}_i = \{\mathbf{x}_{i1}`, \mathbf{x}_{i2}`, ..., \mathbf{x}_{iT}`\}$ be a matrix collecting of the variables for individual i. If ε_{ijt} is IID Type 1 extreme value, then the conditional probability (conditional on the value of $\boldsymbol{\beta}_i$) of the respondent i's sequence of choices is:

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

where $y_{ijt} = 1$ if individual i chooses alternative j in choice occasion t, and 0 otherwise.

The unconditional probability for each individual is obtained by weighting Equation (1) over all the possible values of $\boldsymbol{\beta}_i$:

$$f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\Omega}) = \int f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}_i)g(\boldsymbol{\beta}_i|\boldsymbol{\Omega})d\boldsymbol{\beta}_i. \tag{2}$$

Equation (2) is basically the expected probability over all possible values of $\boldsymbol{\beta}_i$. Thus, the unconditional probability that individual i will choose alternative j given the specific characteristics of their choice set and the underlying parameters is equal to the expected value of the conditional probability as it ranges over the possible values of $\boldsymbol{\beta}_i$.

Then, the exact log-likelihood function is:

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{N}\log f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\Omega}) = \sum_{i=1}^{N}\log\left[\int f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}_i)g(\boldsymbol{\beta}_i|\boldsymbol{\Omega})d\boldsymbol{\beta}_i\right],$$

where $\boldsymbol{\theta}$ is a vector that collects all the parameters of the model. Since the integrals do not have a close-form solution, they are approximated by simulation using Monte Carlo integration. For a given value of the parameters $\boldsymbol{\theta}$, a value of $\boldsymbol{\beta}_i$ is drawn from $g(\boldsymbol{\beta}_i|\boldsymbol{\Omega})$.

---

[2] For simplicity in the notation, we assume that all individuals face the same number of choice situations.

Using this draw the unconditional probability in Equation (2) is computed. This process is repeated for many draws, and the average over the draws gives the simulated probability (or Monte Carlo approximation of the integrals). In particular, the simulated probability for individual i is:

$$\widehat{f}\left(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\Omega}\right) = \frac{1}{R}\sum_{r=1}^{R} f(\mathbf{y_i}|\mathbf{X}_i, \boldsymbol{\beta}_{ir}),$$

And the Simulated Maximum Likelihood (SML) is (Train, 2009):

$$\log L_s(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log \left[\frac{1}{R}\sum_{r=1}^{R} f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}_{ir})\right],$$

where $\boldsymbol{\beta}_{ir}$ is the r-th draw of $\boldsymbol{\beta}_i$, and R is the total number of draws.

After the parameters are estimated, researchers can carry out a series of analyzes regarding the heterogeneity of the tastes (or preferences) of the individuals for certain attributes. For example, if the normal distribution is used we can obtain information on the share of the population that places a positive value on some attributes using $\Phi(\widehat{\beta}/\widehat{\sigma}_\beta)$, where $\widehat{\beta}$ and $\widehat{\sigma}_\beta$ are the mean and standard deviation of the distribution of β_i, and $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution. We might also plot the estimated unconditional distribution of the parameters by taking (several) draws from the estimated distribution $g(\boldsymbol{\beta}_i|\widehat{\boldsymbol{\Omega}})$.

Although the MIXL model is quite flexible to capture the aggregated heterogeneity of tastes with respect to attributes, it does not tell us where each individual lies in the distribution of tastes in the population. The next section describes how this can be done using the individual-specific estimators.

3. Individual-specific estimates

3.1. Estimator

The individual-specific parameters can be obtained by deriving the individual's conditional distribution based on the choices observed for that individual using Bayes's formula (Revelt and Train, 2000; Train, 2009). Explicitly, the conditional distribution for individual i is given by:

$$h(\boldsymbol{\beta}_i|\mathbf{y}_i, X_i, \boldsymbol{\theta}) = \frac{f(\mathbf{y}_i|X_i, \boldsymbol{\beta}_i)g(\boldsymbol{\beta}_i|\boldsymbol{\Omega})}{f(\mathbf{y}_i|X_i, \boldsymbol{\Omega})} = \frac{f(\mathbf{y}_i|X_i, \boldsymbol{\beta})g(\boldsymbol{\beta}_i|\boldsymbol{\Omega})}{\int f(\mathbf{y}_i|X_i, \boldsymbol{\beta}_i)g(\boldsymbol{\beta}_i|\boldsymbol{\Omega})d\boldsymbol{\beta}_i}. \tag{3}$$

Thus, $h(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta})$ is the conditional (on the observed choices) distribution of the individual i's tastes, whereas $g(\boldsymbol{\beta}_i|\boldsymbol{\Omega})$ is the unconditional distribution in the population. Specifically, $h(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta})$ gives us all potential values of β for individual i, given what we observe about that individual (data) and the underlying parameters. It can also be interpreted as the density of β_i in the subpopulation of individuals who would choose the sequence of choices that individual i made, $\mathbf{y}_i$, when facing $\mathbf{X}_i$ (Huber and Train, 2001; Train, 2009).

The individual-specific coefficients are obtained by computing the average value among all potential values for individual i. Thus, taking the expectation of $h(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta})$ yields:

$$\overline{\boldsymbol{\beta}}_i = \mathbb{E}[\boldsymbol{\beta}_i|\mathbf{y}_i, X_i, \boldsymbol{\theta}] = \int \boldsymbol{\beta}_i h(\boldsymbol{\beta}_i|\mathbf{y}_i, X_i, \boldsymbol{\theta})d\boldsymbol{\beta}_i = \frac{\int \boldsymbol{\beta}_i f(\mathbf{y}_i|X_i, \boldsymbol{\beta}_i)g(\boldsymbol{\beta_i}|\boldsymbol{\Omega})d\boldsymbol{\beta}_i}{\int f(\mathbf{y}_i|X_i, \boldsymbol{\beta}_i)g(\boldsymbol{\beta_i}|\boldsymbol{\Omega})d\boldsymbol{\beta}_i}. \tag{4}$$

It is important to mention that Equation (3) is also known as the posterior distribution, whereas Equation (4) is known as the posterior mean in a Bayesian setting with quadratic loss. Prior knowledge about the individual estimates is given by the prior distribution $g(\boldsymbol{\beta}_i|\boldsymbol{\Omega})$ and the sampling distribution, or likelihood function, is given by $f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}_i)$.

Another statistic used in applied research is the variance of the conditional distribution, which is given by:

$$\overline{\boldsymbol{\sigma}}_i^2 = Var(\boldsymbol{\beta}_i|\mathbf{y}_i, X_i, \boldsymbol{\theta}) = \mathbb{E}[\boldsymbol{\beta}_i^2|\mathbf{y}_i, X_i, \boldsymbol{\theta}] - (\mathbb{E}[\boldsymbol{\beta}_i|\mathbf{y}_i, X_i, \boldsymbol{\theta}])^2 = \left(\frac{\int \boldsymbol{\beta}_i^2 f(\mathbf{y_i}|X_i, \boldsymbol{\beta}_i)g(\boldsymbol{\beta}_i|\boldsymbol{\Omega})d\boldsymbol{\beta}_i}{f(\mathbf{y_i}|X_i, \boldsymbol{\beta}_i)g(\boldsymbol{\beta}_i|\boldsymbol{\Omega})d\boldsymbol{\beta}_i}\right) - \left(\frac{\int \boldsymbol{\beta}_i f(\mathbf{y_i}|X_i, \boldsymbol{\beta}_i)g(\boldsymbol{\beta}_i|\boldsymbol{\Omega})d\boldsymbol{\beta}_i}{\int f(\mathbf{y_i}|X_i, \boldsymbol{\beta}_i)g(\boldsymbol{\beta}_i|\boldsymbol{\Omega})d\boldsymbol{\beta}_i}\right)^2. \tag{5}$$

Thus, the conditional (posterior) distribution of tastes for individual i is given by $h(\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\theta})$, with expected value (or conditional mean) given by Equation (4) and variance given by Equation (5).

One key problem when obtaining estimates for $\overline{\beta}_i$ and $\overline{\sigma}_i^2$ is that they do not have a closed-form solution. Again, we can rely on simulations to approximate the integrals and obtain the Bayes point estimator. For the conditional mean, the procedure consists of taking several draws of $\boldsymbol{\beta}$ from the population density $g(\boldsymbol{\beta}_i|\widehat{\boldsymbol{\Omega}})$, say $r = 1, ..., R$ and compute (Train, 2009):

$$\widehat{\overline{\boldsymbol{\beta}}}_{\mathbf{i}} = \widehat{\mathbb{E}}[\boldsymbol{\beta}_i|\boldsymbol{y}_i, \boldsymbol{X}_i, \widehat{\boldsymbol{\theta}}] = \frac{\sum_{r=1}^{R}\widehat{\boldsymbol{\beta}}_{ir}f(\boldsymbol{y}_i|\boldsymbol{X}_i, \widehat{\boldsymbol{\beta}}_{ir})}{\sum_{r=1}^{R}f(\boldsymbol{y}_i|\boldsymbol{X}_i, \widehat{\boldsymbol{\beta}}_{ir})}, \tag{6}$$

where $\widehat{\boldsymbol{\beta}}_{ir}$ is the rth draw from its distribution, and $f(\boldsymbol{y}_i|\boldsymbol{X}_i, \widehat{\boldsymbol{\beta}}_{ir})$ is the estimated likelihood function for individual i given $\widehat{\boldsymbol{\beta}}_{ir}$.[3]

Likewise, $\overline{\boldsymbol{\sigma}}^2$ can be estimated using (Hess, 2010; Greene, 2012):

$$\widehat{\overline{\boldsymbol{\sigma}}}_i^2 = \widehat{Var}(\boldsymbol{\beta}_i|\boldsymbol{X}_i, \widehat{\boldsymbol{\theta}}) = \frac{\sum_{r=1}^{R}(\widehat{\boldsymbol{\beta}}_{ir} - \widehat{\overline{\boldsymbol{\beta}}}_i)^2 f(\boldsymbol{y}_i|\boldsymbol{X}_i, \widehat{\boldsymbol{\beta}}_{ir})}{\sum_{r=1}^{R}f(\boldsymbol{y}_\mathbf{i}|\boldsymbol{X}_i, \widehat{\boldsymbol{\beta}}_{ir})}. \tag{7}$$

3.2. Consistency of individual estimates and implications

3.2.1. T-asymptotic

Researchers interested in the tastes for each sampled individual usually compute the conditional mean using Equation (6).[4] Since the conditional means are generally used for second-stage procedures such as targeted product design, spatial analysis of tastes, compute WTP measures, etc., a natural question that arises is whether $\widehat{\overline{\beta}}_i$ can be considered an estimate of β_i. In other words, the question is: when is the estimated mean of the conditional distribution a reliable estimate of individual i's coefficient? As Revelt and Train (2000) and Train (2009) have already shown, the only way that we can be confident that $\widehat{\overline{\beta}}_i$ is a consistent estimate is when the number of choice situations that an individual face is large enough. Strictly speaking, as $T \to \infty$ the mean of the conditional distribution converges to the true specific individual coefficient, that is, $\widehat{\overline{\beta}}_i \overset{p}{\to} \beta_i$: intuitively, if we could observe infinitely how individuals react to changes in attributes, and their choices in terms of these changes, in theory we could elucidate the specific tastes of individuals.[5] This result can also be obtained by considering the frequentist view of the asymptotic behavior of the Bayesian estimator, which is summarized by the Bernstein-von Mises Theorem. This theorem states that under mild conditions, the mean of the posterior distribution, $\widehat{\overline{\beta}}_i$, is asymptotically equivalent to the maximum likelihood estimator of β_i as the sample size (T in this case) grows (see Train, 2009, for a formal exposition of this).

To understand the mechanism behind the convergence of the conditional mean, consider Fig. 1. The black distribution represents the unconditional distribution $g(\beta_i|\boldsymbol{\Omega})$— the aggregated population distribution of tastes—, whereas the blue distributions represent the conditional distributions, $h(\beta_i|\boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{\theta})$, for some specific individual and different number of choice situations ($T = 5, 10, 50$). Assume that the red line, $\beta = 1$, represents the true coefficient for this particular individual in $g(\beta_i|\boldsymbol{\Omega})$. We can observe that as T increases (i.e., we observe how this particular individual choose under different changes of the attributes), the conditional mean of $h(\beta_i|\boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{\theta})$, $\widehat{\overline{\beta}}_i$, converges to the true individual's coefficient. In other words, $h(\beta_i|\boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{\theta})$ should collapse onto β_i as $T \to \infty$. Finally, it is important to point out that often the number of choice situations per individual is less than the number of individual-level coefficients for the individual, and so individual-level coefficients are unidentified by the individual's choices.[6]

Another important point that can be observed in Fig. 1 is that the variance of the conditional distribution shrinks as more information about individual i's choices are available. That is, the variance of the conditional distribution $\overline{\sigma}_i^2$ tends to zero as T increases without bound. Thus, more observations per individual reduces the uncertainty in the determination of the conditional mean reducing $\mathrm{Var}(\beta_i|\boldsymbol{y}_i, \boldsymbol{X}, \boldsymbol{\theta})$.

In sum, there are two important conclusions regarding the mean and variance of the conditional distribution. Formally:

$$\widehat{\overline{\boldsymbol{\beta}}}_i = \widehat{\mathbb{E}}\left[\boldsymbol{\beta}_i|\boldsymbol{y}_i, \boldsymbol{X}_i, \widehat{\boldsymbol{\theta}}\right] \overset{p}{\to} \boldsymbol{\beta}_i \quad \text{and} \quad \widehat{\overline{\boldsymbol{\sigma}}}_i^2 = \widehat{\mathrm{Var}}(\boldsymbol{\beta}_i|\boldsymbol{y}_i, \boldsymbol{X}_i, \widehat{\boldsymbol{\theta}}) \overset{p}{\to} \boldsymbol{0} \quad \text{as} \quad T \to \infty. \tag{8}$$

---

[3] As an aside, since the conditional estimates are estimated for each individual based on their choice set, predictions outside of the sample is not possible. The unconditional estimates are more suitable for this purpose (Jones and Hensher, 2004).

[4] Revelt and Train (2000) offer two additional estimator depending on how the simulation is performed. Instead of taking the population parameters θ as given, and use the point estimates $\widehat{\theta}$ as in Equation (6), a different approach is to take into consideration the sampling distribution of θ. Let $N(\theta|\overline{\theta}, \Sigma_\theta)$ be the multivariate normal density (asymptotic distribution) of θ with mean $\overline{\theta}$ and covariance Σ_θ. Then, the expected value of the individual-level preference parameters β_i conditional on the population parameters $\overline{\theta}, \Sigma_\theta$ and on the individual choices $\mathbf{y}_i$, is: $\mathbb{E}[\beta_i|\mathbf{y}_i, \overline{\theta}, \Sigma_\theta] = \int_\theta \mathbb{E}[\beta_i|\mathbf{y}_i,$ $\mathbf{X}_i, \theta] \mathcal{N}(\theta|\overline{\theta}, \Sigma_\theta)d\theta$. A Monte Carlo approximation to this expectation is obtained by calculation of the empirical mean evaluated at draws of θ from the asymptotic distribution of the estimator. Revelt and Train (2000) and Sarrias and Daziano (2018) have shown that estimators based on point and sampling distribution deliver similar results in Monte Carlo experiments. The second estimator, instead of taking draws from the asymptotic distribution, takes draws from the posterior distribution of θ. Since the denominator in Equation (3) is a constant, we can write the following proportionality relation: $h(\beta_i|\mathbf{y}_i, \mathbf{X}_i, \theta) \propto f(\mathbf{y}_i|\mathbf{X}_i, \beta_i)g(\beta_i|\boldsymbol{\Omega})$. Draws from the posterior $h(\beta_i|\mathbf{y}_i, \mathbf{X}_i, \theta)$ can then be obtained using the Metropolis-Hasting algorithm (Sillano and de Dios Ortúzar, 2005), which allows to obtain the sampling distribution of $h(\beta_i|\mathbf{y}_i, \mathbf{X}_i, \theta)$.

[5] The reader familiar with panel data estimation can make an analogy with the consistency of the FE estimator of the individual effects. If T is fixed and $N \to \infty$ as typical in short panels, then only the FE estimator of β is consistent, whereas the estimator of the individual effects are not consistent since the number of these parameters increases as N increases.

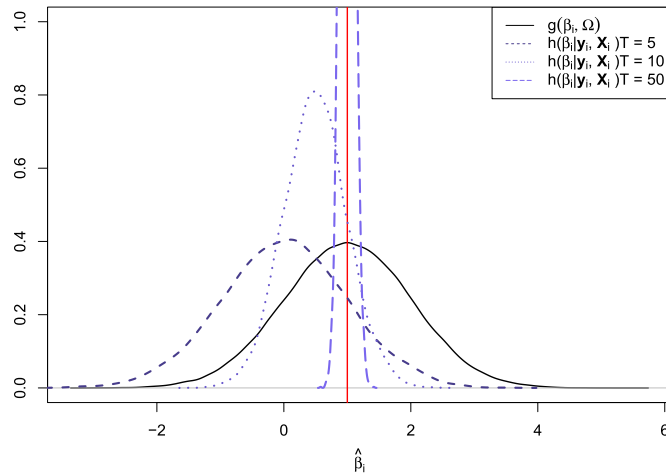[6] The author is grateful to Kenneth Train for pointing this out.

Fig. 1. Conditional and unconditional distributions.

3.2.2. N-asymptotic

A final question remains: What about the asymptotic in terms of the sample size, N? Does increasing the number of individuals help to achieve consistency of individual-level estimators? When the number of choice situations is fixed, but the number of individuals goes to infinity, the population parameters θ are estimated with less bias improving the estimate of $\widehat{\theta}$, and hence, the point estimates of $\widehat{\overline{\beta}}_i$ and $\widehat{\overline{\sigma}}_i^2$. However, increasing the sample size and keeping the number of choice situations fixed prevents the mean of the conditional distribution from converging to the true individual's parameter (Revelt and Train, 2000). In other words, a large sample size with a low number of choice situation per individual does not guarantee consistency of the conditional means.

3.3. Relationship between unconditional and conditional distributions

The result in (8) has several consequences, which can be used as a diagnostic tool in applied research. For a correctly specified model at the true population parameters, the conditional distributions aggregated over all individuals should be equal to the unconditional (population) distribution (Revelt and Train, 2000; Hensher et al., 2005),[7] and consequently the moments of the aggregated conditional means should be equal to the moments of $g(\beta_i|\Omega)$. This implies that: first, the sample average of the conditional means over individuals should converge to the mean of the population distribution. The intuition is that since the conditional distributions collapse in the true individual parameter, so does their expected values. Therefore, for a sufficiently large T, if we add up the conditioned expectations, which converge to the most probable position of each individual in $g(\beta_i|\Omega)$, its distribution must be equal to $g(\beta_i|\Omega)$. Formally, let $\widehat{\mu}_{\overline{\beta}_i} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\overline{\beta}}_i$ be the sample average of the estimated conditional means and μ be the mean of the unconditional distribution. Then $\widehat{\mu}_{\overline{\beta}_i} \xrightarrow{p} \mu$, as T increases without bound.

Second, the variance of the conditional means should converge to the population variance of the unconditional distribution. To see this point, recall that the unconditional variance of β_i can be written as:

$$\mathrm{Var}(\beta_i) = \mathbb{E}\left[\mathrm{Var}(\beta_i|\mathbf{y}_i, \mathbf{X}_i, \theta)\right] + \mathrm{Var}[\mathbb{E}\left[\beta_i|\mathbf{y}_i, \mathbf{X}_i, \theta\right]]\,\sigma^2 = \mathbb{E}\left[\overline{\sigma}_i^2\right] + \mathrm{Var}\left[\overline{\beta}_i\right],$$

where $\mathrm{Var}(\beta_i) = \sigma^2$ is the variance of the unconditional distribution $g(\beta_i|\Omega)$, $\mathrm{Var}(\beta_i|\mathbf{y}_i, \mathbf{X}_i, \theta)$ and $\mathbb{E}\left[\beta_i|\mathbf{y}_i, \mathbf{X}_i, \theta\right]$ are the variance and expectation of the conditional distribution, respectively. Since $\widehat{\overline{\sigma}}_i^2 \xrightarrow{p} 0$ as $T \to \infty$, then $\mathrm{Var}(\widehat{\overline{\beta}}_i) \xrightarrow{p} \sigma^2$ as $T \to \infty$. In words, as the number of choice situations per individual increases without bound, then the variance of the conditional means should converge to the variance of the unconditional distribution.

Moreover, since in general $\widehat{\overline{\sigma}}_i^2 \geq 0$ when T is not sufficiently large, such that $\mathbb{E}\left[\widehat{\overline{\sigma}}_i^2\right] \geq 0$, then it is true that:

$$\mathrm{Var}(\beta_i) \geq \mathrm{Var}[\widehat{\overline{\beta}}_i].$$

That is, the distribution of the conditional means has generally a narrower range than the unconditional distribution. Although this property has been pointed out by several authors (see for example Revelt and Train, 2000; Hensher et al., 2003; Greene et al., 2005; Hess, 2010; Daly et al., 2012; Sarrias and Daziano, 2018), has not been considered in most of the empirical studies that rely on individual-specific estimates. For example, several articles have found that the tastes of individuals show more plausible and less

---

[7] For a proof of this result see Revelt and Train (2000) or Hensher et al. (2005, section 8.1).

counter-intuitive results in terms of signs when conditioned means are used. This might be true, or it can be a result of the above. As Daly et al. (2012) state: "*The procedure mistakenly ignores the fact that conditional distributions are derived from, and must aggregate to, the unconditional distribution. In particular, the unconditional population variance is equal to the variance of the conditional means plus the variance of the unconditional distribution around the conditional means. The variance of conditional means is less than the unconditional variance not because it is more reasonable estimate, but rather because **it incorrectly excludes the variance around the conditional means**.*" Therefore, the conditional estimates are estimated with uncertainty, which is usually ignored in practice, and this uncertainty decreases as more choice per individual are observed.

These results provide three potential tools to diagnose how reliable the estimates at the individual level are (Allenby and Rossi, 1998; Revelt and Train, 2000; Hess, 2010). All of them aim to evaluate how similar is the distribution of the conditional means across respondents to the population distribution. Consider the following two ratios: $\hat{\mu}_{\bar{\beta}_i}/\hat{\mu}$ and $sd(\hat{\bar{\beta}}_i)/\hat{\sigma}$. If the model is correctly specified and accurately estimated, then the moments of the conditional means should be equal to the moments of the population distribution and, consequently, these two ratios should be very close to 1. If they are not, the difference could be due to (1) misspecification of the population distribution, (2) insufficient number of draws, (3) an inadequate sample size, (4) the maximum likelihood converging at a local rather than global maximum, (5) not enough choices per person, and/or (6) bad experimental design (Revelt and Train, 2000). A more formal method to evaluate whether both distribution are equal is to use, for example, the Kolmorogov-Smirnov (KS) test where the null hypothesis is that both distributions are equal.

Finally, it is important to note that all these results apply also to conditional means estimated from models re-parameterized in WTP space, as suggested by Train and Weeks (2005).

4. Monte Carlo setup

To understand how well the diagnostic measures for individual-specific estimates perform under a well-specified model and different number of choice situations, a Monte Carlo experiment with a similar setup as in Revelt and Train (2000) is carried out.[8] Specifically, it is assumed that the true utility for individual i when choosing alternative j in choice occasion t is given by:

$$U_{ijt} = \beta_1 x_{1,ijt} + \beta_{2i} x_{2,ijt} + \beta_{3i} x_{3,ijt} + \beta_{4i} x_{4,ijt} + \beta_{5i} x_{5,ijt} + \varepsilon_{ijt}, \tag{9}$$

where ε_{ijt} is a iid type 1 Extreme Value error term; the five attributes $x_{k,ijt}$, $k = 1, \ldots, 5$ are generated as independent standard normal variables; β_1 is a fixed coefficient with true value equal to 1; and the number of alternatives is equal to 3, $(J = 3)$. The true population densities for the random parameters are set as follows:

- $\beta_{2i} \sim N(1, 1)$, so that about 84% of the individuals have a positive coefficient,
- $\beta_{3i} \sim N(1, 3^2)$, so that about 63% of the individuals have a positive coefficient,
- $\beta_{4i} = \exp(N(0.5, 0.5^2))$, such that $\beta_{4i} \sim \text{logN}(0.5, 0.5^2)$. Thus $\mathbb{E}(\beta_{4i}) = \exp(0.5 + 0.5^2/2) = 1.868$, and $sd(\beta_{4i}) = 1.868 \times \sqrt{(\exp(0.5^2) - 1)} = 0.996$,
- $\beta_{5i} = \exp(N(0.5, 1))$, such that $\beta_{4i} \sim \text{logN}(0.5, 1)$. Thus $\mathbb{E}(\beta_{5i}) = \exp(0.5 + 1/2) = 2.718$, and $sd(\beta_{5i}) = 1.868 \times \sqrt{(\exp(1) - 1)} = 3.563$.

To understand the behavior of the conditional estimates and the diagnostic measures, the experiment is run for different number of choice situations. In particular, similar to Revelt and Train (2000) and Sarrias and Daziano (2018), we choose $T = \{1, 5, 10, 20, 50\}$.

To keep the estimate time manageable, in each experiment $S = 100$ samples of size $N = 300$ are created, and in each sample ε_{ijt} and the five attributes are randomly created. The alternative selected for each of the 300 individuals is the one that delivers the highest utility using Equation (9). The fixed and random parameters are created before running the experiment and held fixed in each trail.

As shown by Revelt and Train (2000), the number of Halton draws used in the simulation does not seem to affect the results if more than 100 Halton draws are used. Thus, 300 Halton draws for each random parameter and individuals are used during the experiment.

For each experiment $s = 1, \ldots, S$, the conditional mean for each random parameter, $\hat{\bar{\beta}}_{is}$, using Equation (6), and the standard deviation of the conditional distribution, $\hat{\sigma}_{is}$, using the square root of Equation (7) are computed. After running the experiments, we compute the following statistics averaged over the S samples:

- Sample mean of the conditional means:

$$\hat{\mu}_{\bar{\beta}_i} = \frac{1}{S} \sum_{s=1}^{S} \left(\frac{1}{N} \sum_{i=1}^{300} \hat{\bar{\beta}}_{is} \right).$$

- Standard deviation of the conditional means:

---

[8] The program and code for the experiment can be downloaded at https://data.mendeley.com/datasets/k8v7ns2xv3/1.

$$sd(\widehat{\overline{\beta}}_i) = \frac{1}{S} \sum_{s=1}^{S} \left(\frac{1}{(N-1)} \sum_{i=1}^{300} [\widehat{\overline{\beta}}_{is} - \widehat{\overline{\overline{\beta}}}_s] \right)^{1/2}.$$

- The sample mean of the standard deviation of the conditional distribution:

$$\overline{\widehat{\sigma}}_i = \frac{1}{S} \sum_{s=1}^{S} \left(\frac{1}{N} \sum_{i=1}^{300} \widehat{\sigma}_{is} \right).$$

- The percentage of positive conditional means:

$$\% > 0 = \frac{1}{S} \sum_{s=1}^{S} \left(\frac{1}{N} \sum_{i=1}^{300} 1 \left[\widehat{\overline{\beta}}_{is} > 0 \right] \right).$$

- As additional diagnostic tool, for each trial we compute two tests that evaluate how similar is the distribution of the conditional means and the true population distribution. The first test is based on the Lepage's D statistic (Lp) which compares the location and scale between two distributions. In particular, let $\delta = \widehat{\mu}_{\overline{\beta}_i} - \mu$ and $\gamma^2 = \frac{\mathrm{Var}(\widehat{\mu}_{\overline{\beta}_i})}{\sigma^2}$. Then the null hypothesis is $H_0 : \delta = 0$ and $\gamma^2 = 1$. The second test is the traditional Kolmogorov-Smirnov (KS) test, which quantifies the distance between both empirical distributions. The null hypothesis is that the samples are drawn from the same distribution. The p-values for both tests are saved for each experiment and the averaged p-values over S are reported.
- The absolute difference bias:

$$Bias = \frac{1}{S} \sum_{s=1}^{S} \left(\frac{1}{N} \sum_{i=1}^{300} |\widehat{\overline{\beta}}_{is} - \beta_i| \right).$$

5. Results

5.1. Revisiting Revelt and Train (2000)

Similar in spirit to Revelt and Train (2000)'s work, this Section analyzes the behavior of the conditional means and the diagnostic measures under a well-specified model.

Table 1 presents the Monte Carlo results averaged over S, where S is the total number of samples for which the SML converged.[9] The estimations were carried out using package gmnl in R (Sarrias and Daziano, 2017).

First, as in Revelt and Train (2000) and Train (2009), it can be observed that as T increases the sample mean of the conditional means converges to the true mean of the unconditional distribution for our 4 parameters: $\widehat{\mu}_{\overline{\beta}_i} \overset{p}{\to} \mu$. For example, for both normal parameters (β_{i2}, β_{i3}), the mean of the conditional means across individuals converges to 1. The same occurs for β_{i4} and β_{i5} which converge to 1.87 and 2.72, respectively. As mentioned before, under a well-specified model the average conditional means should be close to the unconditional mean. Thus, a good diagnostic tool is to compute $\widehat{\mu}_{\overline{\beta}_i}/\mu$ (Allenby and Rossi, 1998; Revelt and Train, 2000; Train, 2009; Hess, 2010). Considering $T = 10$— which is the most typical number of choice situations in real choice experiments—the sample mean of the conditional means represents the 98%, 94%, 97% and 96% of the true population mean, respectively. Therefore, if the model is well-specified and consistently estimated, $\widehat{\mu}_{\overline{\beta}_i}$ should be very close to μ (with low bias) for reasonable number of choice situations.

Second as T increases the mean of the conditional standard deviation, $\overline{\widehat{\sigma}}_i$, decreases, whereas the standard deviation of the sample mean, $sd(\widehat{\overline{\beta}}_i)$, increases and converges to the standard deviation of the population distribution, σ. Computing $sd(\widehat{\overline{\beta}}_i)/\sigma$ for $T = 10$, the standard deviation of the conditional means is about 73%, 90%, 62% and 84% of the true standard deviation of the unconditional distribution, respectively.

The results obtained for the conditional mean and standard deviation emphasize two important points. First, the rate of convergence of the mean and variance of the conditional means is not the same. The mean of the conditional mean is a consistent estimator of the population mean even with a small T, while the variance of the conditional means is biased downward for the population variance for any finite T and becomes consistent if and only if T rises without bound.[10] Second, the moments of the distribution of the conditional means should converge to the moments of the unconditional distribution as the number of choice situations increases. This can also be observed by looking at Fig. 2 which shows the true population distribution alongside the distribution of the conditional means for the 300 individuals (averaged over the S datasets) for each T. For example, for a few number of choice situations the conditional variance for each individual $\widehat{\sigma}_i^2$ is high, resulting in more concentrated conditional means. But, as T increases, $\widehat{\sigma}_i^2$ decreases and $sd(\widehat{\overline{\beta}}_i)$

---

[9] Samples for which the iteration limit (300) was exceeded or the optimization algorithm (BHHH) found a flat region are dropped.
[10] I am grateful to one of the reviewers for pointing this out.

Table 1
Simulation results for MIXL: Misspecification. The results are averaged over the number of samples *S* for which the SML converged.

|	$N(1,1)$								$N(1,3^2)$							
	$\widehat{\mu}_{\bar{\beta}_i}$	$sd(\widehat{\bar{\beta}}_i)$	$\overline{\widehat{\sigma}}_i$	*%>0*	*KS*	*Lp*	*Bias*	*S*	$\widehat{\mu}_{\bar{\beta}_i}$	$sd(\widehat{\bar{\beta}}_i)$	$\overline{\widehat{\sigma}}_i$	*%>0*	*KS*	*Lp*	*Bias*	*S*
T = 1	1.773	0.685	1.821	0.978	0.000	0.000	1.542	81	1.776	3.248	4.335	0.727	0.029	0.016	3.104	81
T = 5	0.986	0.587	0.832	0.941	0.005	0.000	0.680	99	0.964	2.595	1.592	0.643	0.317	0.444	1.239	99
T = 10	0.981	0.727	0.699	0.904	0.069	0.004	0.574	100	0.943	2.724	1.184	0.636	0.677	0.722	0.959	100
T = 20	0.963	0.827	0.533	0.871	0.314	0.081	0.478	99	0.923	2.786	0.845	0.632	0.871	0.843	0.757	99
T = 50	0.917	0.871	0.316	0.845	0.429	0.132	0.384	98	0.886	2.678	0.479	0.631	0.843	0.640	0.594	98
	$LN(0.5,0.5^2)$								$LN(0.5,1)$							
	$\widehat{\mu}_{\bar{\beta}_i}$	$sd(\widehat{\bar{\beta}}_i)$	$\overline{\widehat{\sigma}}_i$	*%>0*	*KS*	*Lp*	*Bias*	*S*	$\widehat{\mu}_{\bar{\beta}_i}$	$sd(\widehat{\bar{\beta}}_i)$	$\overline{\widehat{\sigma}}_i$	*%>0*	*KS*	*Lp*	*Bias*	*S*
T = 1	3.366	0.501	1.797	1.000	0.000	0.000	2.147	81	5.714	3.257	8.779	1.000	0.000	0.000	4.760	81
T = 5	1.837	0.457	0.821	1.000	0.000	0.000	0.685	99	2.700	2.657	2.104	1.000	0.003	0.005	1.480	99
T = 10	1.810	0.615	0.711	1.000	0.019	0.008	0.603	100	2.608	2.999	1.438	1.000	0.038	0.021	1.121	100
T = 20	1.793	0.742	0.560	1.000	0.084	0.016	0.527	99	2.549	3.186	0.977	1.000	0.276	0.138	0.882	99
T = 50	1.722	0.808	0.341	1.000	0.232	0.084	0.449	98	2.418	3.173	0.543	1.000	0.685	0.453	0.692	98
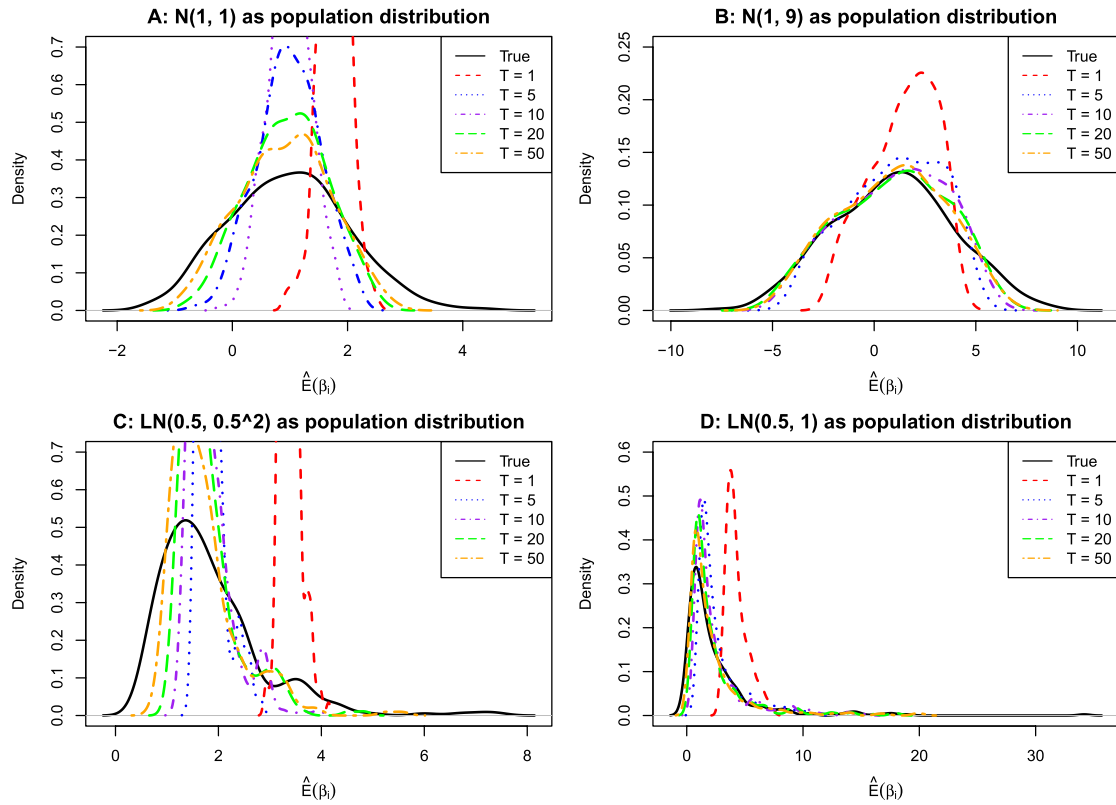
Fig. 2. Kernel densities of conditional means estimates. The true distribution corresponds to the unconditional (population) distribution. The kernel densities for each number of choice situations show the estimated conditional mean for each individual averaged over *S* constructed data sets.

converges to *σ*. It is also important to note that when *T* is low, the percentage of individuals with positive coefficients tends to be overestimated when the true population distribution is normal. Table LABEL:tableind shows that the percentage of positive conditional means converges to the true percentage of positive coefficients in the population distribution—84% and 63%, respectively— when we are able to observe more choice situations per individual. This result is very important for practitioners since the usual claim that *"the conditional means are more reasonable estimates because they do not present sign violation"* is misleading (Daly et al., 2012). This result is due to the fact that when the number choice situations is low the variance of the population distribution will always be greater than the variance of the conditional means: $\widehat{\sigma}^2_{\beta_i} > \mathrm{Var}(\widehat{\overline{\beta}}_i)$.

Although Fig. 2 gives us a graphic idea of how fast the conditional distribution converges to the unconditional distribution, it is necessary to have a statistical measure that allows us to say whether both distributions are the same or not under a well-specified model. Table LABEL:tableind shows the p-value for both *KS* and *Lp* tests averaged over *S*. The results depend on the type of distribution and the degree of individual heterogeneity in the population distribution. For example, if we consider the normally distributed coefficients, the necessary number of choice situations to obtained a p-value sufficiently large to not reject the null is lower when the true population is $N(1, 9)$ instead of $N(1, 1)$: $T = 5$ vs. $T = 50$. If the population distribution is lognormal, both test require a high number of choice situations to not reject the null. Thus, in applied works, including a test to elucidate if both distributions are equal is a conservative diagnostic tool.

Next, we focus our attention in the bias of the conditional means. We can observe that the bias of the individual-specific estimates decreases as the number of choice situation increases. For example, with $T = 10$ the absolute difference of $\widehat{\overline{\beta}}_i$ and β_i is about 0.574, 0.959, 0.603 and 1.121, respectively. Fig. 3 presents the conditional means estimates for each individual (averaged over the *S* samples) at different choice-situation scenarios and the true unconditional distribution for our 4 random parameters. In general, it can be
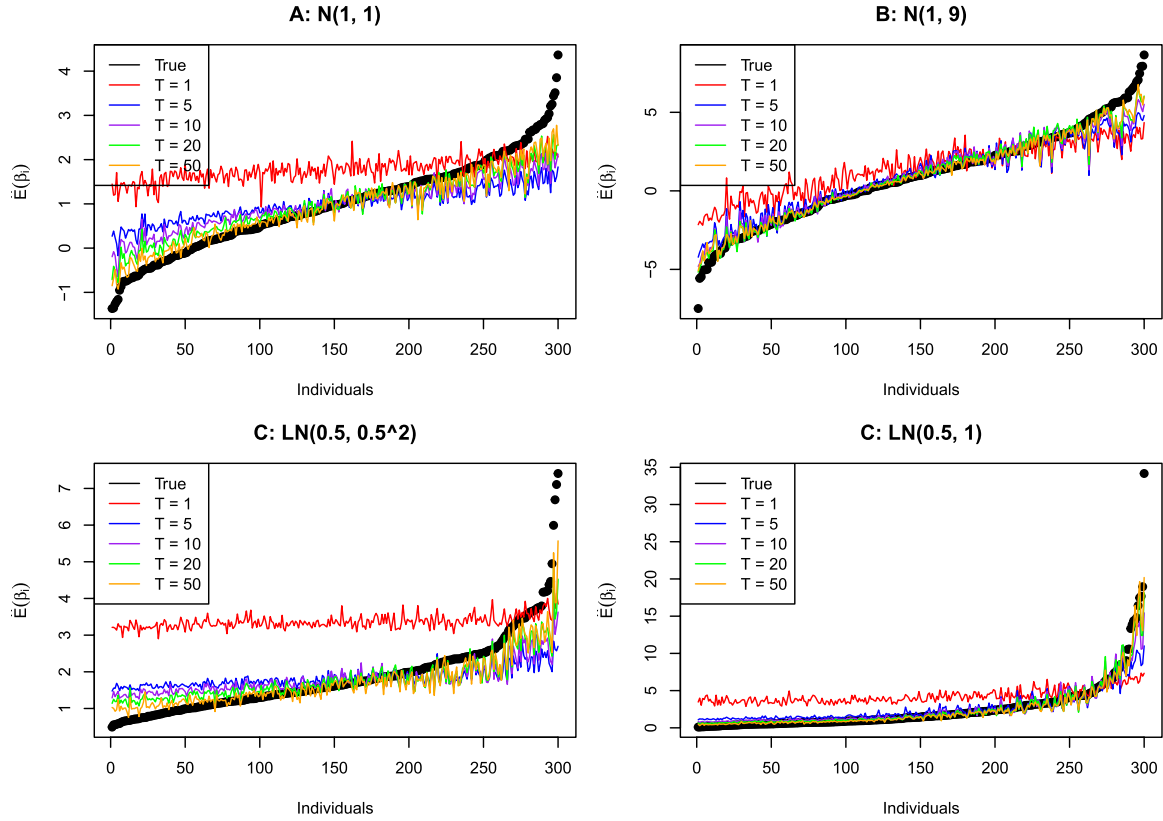
Fig. 3. Conditional means estimates for each individual. The true distribution corresponds to the unconditional (population) distribution. Each point shows the conditional mean for each individual averaged over S constructed data sets.

observed that: 1) with just one choice occasion is almost impossible to disentangle the true individuals' tastes, 2) it is difficult to estimate the true individual-specific estimate when the true coefficient is in the tails of the distribution.[11]

5.2. Misspecification

So far, we have looked at the behavior of the conditional means under a well-specified and consistent model under the true unconditional distribution. However, another issue raised by Hess (2010) is the impact of unconditional distributions on the shape of the conditional distributions. We can consider the impact of misspecification under the Bayesian or frequentist approach. Under a Bayesian perspective, the likelihood function should dominate the population (prior) distribution as the number of choice situations increases. Hence, the relevant question from a Bayesian point of view is not whether the population distribution is misspecified, but rather how large T should be so that the selected prior distribution becomes irrelevant. From a frequentist point of view, if the population distribution is misspecified, then the likelihood function will be biased. Therefore, and considering the Bernstein-von Mises Theorem, the individual coefficients will be inconsistent no matter how much T increases. However, the objective of this section is not to establish which of the two paradigms is the correct one to analyze the impact of population distribution on individual estimates, but rather to understand the reliability of the diagnostic measures to detect a potential problem of poor specification.[12]

---

[11] The monte Carlo experiments in this study were also carried out using $N = 1000$. As explained in Section 3.2, when $N \to \infty$ the consistency of $\widehat{\theta}$ increases, thus the conditional means are estimated with less bias. Yet, if the sample size rises, but the choice situations faced by individuals are fixed, then the conditional distribution and its mean do not change. Our results, nor reported here, confirm this: holding fixed T, but increasing N from 300 to 1000, reduces the bias of the conditional means by about 20% for low levels of T and up to 50% holding fixed T at 50; however, increasing the sample size and holding fixed T prevents the convergence of the conditional means to the true individuals' parameters. These results are in line to those reported by as in Sarrias and Daziano (2018) who conducted a similar study to analyze the behavior of the individual the individual-specific estimates for the LC-MNL model.

[12] To evaluate how large T should be so that the prior is irrelevant would require a Monte Carlo experiment with values for T greater than those used in this study. Therefore, although it may be an interesting theoretical question, it would have little practical relevance, since real choice experiments have a fairly small number of choice situations.

Table 2

Simulation results for MIXL: Misspecification. The results are averaged over the number of samples for which the SML converged.

	$N(1,1)$ as Triangular								$N(1,3^2)$ as Log-normal							
	$\widehat{\mu}_{\beta_i}$	$sd(\widehat{\overline{\beta}}_i)$	$\widehat{\overline{\sigma}}_i^2$	%>0	KS	Lp	Bias	S	$\widehat{\mu}_{\beta_i}$	$sd(\widehat{\overline{\beta}}_i)$	$\widehat{\overline{\sigma}}_i^2$	%>0	KS	Lp	Bias	S
T = 1	0.656	0.207	0.582	0.974	0.000	0.000	0.861	81	4.846	8.515	22.047	1.000	0.000	0.000	5.113	81
T = 5	0.662	0.400	0.559	0.925	0.000	0.000	0.749	99	5.114	15.500	6.532	1.000	0.000	0.000	5.427	99
T = 10	0.773	0.606	0.516	0.865	0.000	0.033	0.635	100	2.856	9.680	2.259	1.000	0.000	0.000	2.844	100
T = 20	0.773	0.654	0.366	0.874	0.000	0.028	0.572	99	1.829	4.407	0.845	1.000	0.000	0.000	1.716	99
T = 50	0.776	0.690	0.202	0.887	0.000	0.040	0.530	98	1.516	2.189	0.345	1.000	0.000	0.000	1.337	98
	$LN(0.5,0.5^2)$ as Normal								$LN(0.5,1)$ as Truncated Normal							
	$\widehat{\mu}_{\beta_i}$	$sd(\widehat{\overline{\beta}}_i)$	$\widehat{\overline{\sigma}}_i^2$	%>0	KS	Lp	Bias	S	$\widehat{\mu}_{\beta_i}$	$sd(\widehat{\overline{\beta}}_i)$	$\widehat{\overline{\sigma}}_i^2$	%>0	KS	Lp	Bias	S
T = 1	1.107	0.132	0.449	0.998	0.000	0.000	0.919	81	1.373	0.385	0.972	1.000	0.000	0.000	1.859	81
T = 5	1.169	0.284	0.526	0.997	0.000	0.000	0.829	99	1.457	0.825	0.885	1.000	0.002	0.025	1.610	99
T = 10	1.378	0.479	0.559	0.994	0.001	0.008	0.710	100	1.671	1.173	0.803	1.000	0.007	0.068	1.396	100
T = 20	1.380	0.561	0.446	0.995	0.002	0.001	0.671	99	1.617	1.172	0.567	1.000	0.008	0.019	1.290	99
T = 50	1.390	0.590	0.278	0.997	0.001	0.000	0.622	98	1.608	1.233	0.346	1.000	0.020	0.011	1.190	98

To analyze the impact of misspecifying the population distribution on the individual-specific estimates we choose incorrect distributions for our 4 parameters during the Monte Carlo experiment. Specifically, we choose the following distributions:

- For $\beta_{2i} \sim N(1, 1)$ we specify a triangular distribution without any type of constraint on the parameters,
- for $\beta_{3i} \sim N(1, 9)$ we specify a log-normal distribution,
- for $\beta_{4i} \sim LN(0.5, 0.5^2)$ we specify a normal distribution,
- for $\beta_{5i} \sim LN(0.5, 1)$ we specify a truncated normal distribution.

Again, the sensitivity of the diagnostic measures are analyzed, but now under a misspecified model. We first look at the ratios $\widehat{\mu}_{\bar{\beta}_i} / \mu$ and $sd(\bar{\widehat{\beta}}_i)/\sigma$. Table 2 shows that for the sample average of the conditional means, and $T = 10$, the ratios are 77%, 285%, 74% and 62%, respectively, whereas that for the standard deviation of the conditional means the ratios are 60%, 323%, 48%, and 33%. As expected, when the unconditional distribution is poorly specified, the sample moments of the conditional means do not match those of the population distribution (given a number of choice situations). Another important and interesting result is that for any T, the average p-values for the *KS* and *Lp* tests are always lower than 0.05 implying that we always reject the null hypothesis that both distributions are equal at 5% significance level.

These results reveal two important conclusions. First, researchers should always analyze the robustness of the results through different choices of population distribution and report diagnostic tools to reveal if there exists a potential misspecification problem of the individual-specific estimates.[13] Second, even if the researcher is not interested in the individual-specific estimates, these tools can be used to detect a misspecification problem in the population distribution. However, it is important to emphasize that failure of diagnostic tests can also indicate another type of problem and not necessarily misspecification.

The results also reveal that a better knowledge of the potential domain of tastes results in a lower bias of the real coefficient for each individual. That is, the population distribution assumed for the parameters (priors) has an impact on the magnitude of the bias. Although these results are not completely generalizable to other specifications, it can be observed that if a triangular unconditional distribution is assumed, when the actual distribution is a normal one, the bias is approximately 1.1 times larger for any number of choice situations. However, if a log-normal distribution is assumed (which restricts the real domain of the coefficients) when the true distribution is normal, the bias is 2.7 times larger. Finally, in all scenarios under misspecification the bias of the conditional means decreases as T increases, which might indicate that the role of the population distribution diminishes in determining the true coefficient as more choices we observe for each individual.

6. Conclusion

The MIXL model has been a revolution in the field of discrete choice models during the last two decades due to its capability to accommodate heterogeneity in tastes by assuming that the marginal utility coefficients are distributed randomly across respondents. However, in some practical cases, just knowing that a coefficient varies across individuals is not enough (Revelt and Train, 2000; Hess, 2010). Given this, individual-specific estimators have been gaining more attention from researchers. Their ability to determine the tastes of each individual in the sample is an attractive property from the applied point of view since it allows richer and more detailed analyzes, which can be used to retrieve individual-specific attribute processing strategies or create segments of customers, among other uses. Yet, there are some issues about their properties that have not generally been considered in the applied literature, which may cast doubt on the potential conclusions derived from their use.

This study reviews and summarizes the literature on the theoretical properties of individual-specific estimators focusing on: 1) their consistency, 2) the relationship between unconditional and conditional distribution, and 3) potential diagnostic measures for their correct use. Then, a Monte Carlo experiment is run to analyze the behavior of the individual-specific estimates focusing on the required conditions for true parameter recovery and the sensitivity of some diagnostic measures under both a well-specified and misspecified model. The focus is to assess whether the diagnostic tools are able to discriminate whether the distribution of the conditional means equals the population (unconditional) distribution under these two scenarios.

The results from the Monte Carlo experiment allow to generate some reasonable guidelines for the correct use of individual specific-estimates. Researchers using either a preference- or WTP-space model should analyze and report the following diagnostic tools to be confident that the conditional estimates are reliable:

1. If the model is correctly specified and accurately estimated, then we should observe that both the distribution of the estimated conditional means and the estimated population distribution are very similar (Revelt and Train, 2000). Thus:

(a) It is always advisable to analyze whether the average of conditional means across all individuals, $\widehat{\mu}_{\bar{\beta}_i}$, is similar to the estimated mean of the population distribution, $\widehat{\mu}$. Considering 10 choice situations, and under a well-specified model, $\widehat{\mu}_{\bar{\beta}_i}$ should represent at least the 90% of $\widehat{\mu}$.

---

[13] See for example Hess (2010) for an application.

(b) In addition to the above, researchers should analyze whether the variance of the conditional means, $\text{Var}(\widehat{\bar{\beta}}_i)$, equals the estimated variance of the population distribution, $\widehat{\sigma}^2$. Again, considering 10 choice situations, and under a well-specified model, $\text{Var}(\widehat{\bar{\beta}}_i)$ should represent at least the 60% of $\widehat{\sigma}^2$.

(c) It is also recommended to test whether the distribution of the conditional means equals the estimated unconditional distribution using, for example, the KS test. Failure to reject the null hypothesis is a conservative indicator that the distribution selected for the parameter is correct.

It is important to carry out these diagnostic checks for different choices of the population distribution of each of the parameters and select that distribution that shows a better fit to the distribution of the conditional means. Another important point to consider is that if the distribution of the individual-specific coefficients is not equal to the estimated population distribution cannot only be due to misspecification, but also to: insufficient number of draws, an inadequate sample size, bad experimental design, and/or the maximum likelihood converging at a local rather than global maximum. Therefore, researchers must analyze whether the estimation procedure may suffer from these problems to ensure a good use of the diagnostic tool.

Once it has been verified that the individual estimates are reliable, care should be taken when using them in the following cases:

1. When using unbounded distributions (e.g, normal) and the number of choice situation is not large enough, it is not advisable to use the estimated conditional means to capture individuals with positive or negative tastes to certain attributes, since the sample variance of the conditional means tend to be underestimated.
2. Since the variance of the conditional distribution tend to decrease as $T \to \infty$, it should not be used to construct confidence intervals for the conditional means.
3. It is much more difficult to estimate (with greater precision) the specific tastes of individuals with extreme tastes (or the tails of the distribution), specially with a few number of choice situations. In other words, it is expected that estimates for individuals with extreme (very high or low) tastes to be downward biased.
4. When computing individual-specific WTP, researchers must check whether the unconditional distribution has well-defined moments using Daly et al. (2012)'s theorems.

This work is not without shortcomings, which can be improved in future work. First, the design of the experiment does not consider more complex MIXL models that include correlation among the parameters and observed heterogeneity, or other models highly used in applied studies such as the Generalized Multinomial Logit model or the Mixed-Mixed Multinomial Logit model (see for example Keane and Wasi, 2013). Since all these models add additional layers of unobserved heterogeneity, the convergence criterion of individual-level estimators is expected to be more demanding in terms of the number of choice situations. Second, it would be interesting in a future work to analyze the behavior of two additional estimators proposed by Revelt and Train (2000). The first takes into consideration the sampling distribution of θ by drawing several draws from its asymptotic distribution, whereas the second estimator takes draws from the posterior distribution of θ using, for example, the Metropolis-Hastings algorithm as in Sillano and de Dios Ortúzar (2005) (see footnote 4). Finally, a less explored issue is to analyze different approaches to generate interval estimators of individual-specific parameters for the MIXL model. For example, Sarrias and Daziano (2018) assessed empirical coverage from two procedures to derive standard errors that had been proposed in the literature in the context of the Latent Class Multinomial Logit model. A similar study can be carried out for the MIXL model.

Declaration of competing interest

I declare that I do not have conflict of interest.

CRediT authorship contribution statement

Mauricio Sarrias: Conceptualization, Methodology, Software, Writing - original draft, Funding acquisition.

Acknowledgements

References

Abildtrup, J., Garcia, S., Olsen, S.B., Stenger, A., 2013. Spatial preference heterogeneity in forest recreation. Ecol. Econ. 92, 67–77.

Allenby, G.M., Rossi, P.E., 1998. Marketing models of consumer heterogeneity. J. Econom. 89 (1–2), 57–78.

Budziński, W., Campbell, D., Czajkowski, M., Demšar, U., Hanley, N., 2018. Using geographically weighted choice models to account for the spatial heterogeneity of preferences. J. Agric. Econ. 69 (3), 606–626.

Campbell, D., Hutchinson, W.G., Scarpa, R., 2009. Using choice experiments to explore the spatial distribution of willingness to pay for rural landscape improvements. Environ. Plann. 41 (1), 97–111.

Daly, A., Hess, S., Train, K., 2012. Assuring finite moments for willingness to pay in random coefficient models. Transportation 39 (1), 19–31.

Dumont, J., Giergiczny, M., Hess, S., 2015. Individual level models vs. Sample level models: contrasts and mutual benefits. Transportmetrica: Transport. Sci. 11 (6), 465–483.

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, seventh ed. Prentice Hall.

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. In: Working Paper, Manuscript. Institute of Transport Studies, University of Sydney.

Hensher, D.A., Greene, W.H., Rose, J.M., 2006. Deriving willingness-to-pay estimates of travel-time savings from individual-based parameters. Environ. Plann. 38 (12), 2365.

Hensher, D.A., Rose, J.M., Greene, W.H., 2005. Applied Choice Analysis: A Primer. Cambridge University Press.

Hess, S., 2007. Posterior analysis of random taste coefficients in air travel behaviour modelling. J. Air Transport. Manag. 13 (4), 203–212.

Hess, S., 2010. Conditional parameter estimates from mixed logit models: distributional assumptions and a free software tool. J. Choice. Model. 3 (2), 134–152.

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.

Huber, J., Train, K., 2001. On the similarity of classical and bayesian estimates of individual mean partworths. Market. Lett. 12 (3), 259–269.

Jones, S., Hensher, D.A., 2004. Predicting firm financial distress: a mixed logit model. Account. Rev. 79 (4), 1011–1038.

Keane, M., Wasi, N., 2013. Comparing alternative models of heterogeneity in consumer choice behavior. J. Appl. Econom. 28 (6), 1018–1045.

Revelt, D., Train, K., 2000. Customer-specific taste parameters and mixed logit: households' choice of electricity supplier. In: Working Paper. Department of Economics, UCB.

Richter, L.-L., Pollitt, M.G., 2018. Which smart electricity service contracts will consumers accept? The demand for compensation in a platform market. Energy Econ. 72, 436–450.

Sandorf, E.D., Campbell, D., Hanley, N., 2017. Disentangling the influence of knowledge on attribute non-attendance. J. Choice. Model. 24, 36–50.

Sarrias, M., Daziano, R., 2017. Multinomial logit models with continuous and discrete individual heterogeneity in r: the gmnl package. J. Stat. Software 79 (2), 1–46.

Sarrias, M., Daziano, R.A., 2018. Individual-specific point and interval conditional estimates of latent Class logit parameters. J. Choice. Model. 27, 50–61.

Sillano, M., de Dios Ortúzar, J., 2005. Willingness-to-pay estimation with mixed logit models: some new evidence. Environ. Plann. 37, 525–550.

Train, K., Weeks, M., 2005. Discrete choice models in preference space and willingness-to-pay space. In: Applications of Simulation Methods in Environmental and Resource Economics, 1–16. Springer.

Train, K.E., 2009. Discrete Choice Methods with Simulation, second ed. Cambridge university press.