# Latent class structures: taste heterogeneity and beyond

Stephane Hess*

Institute for Transport Studies, University of Leeds, s.hess@its.leeds.ac.uk

## 1 Introduction

The treatment of heterogeneity across individual decision makers is one of the key topics of research in choice modelling, as evidenced by many of the chapters in this book. While part of this heterogeneity can in many cases be linked to differences in key socio-demographic characteristics across agents, there has long been a recognition that often a non-trivial share of it cannot be explained in this manner. A number of reasons exist, on the one hand an inability to capture all possible socio-demographic characteristics that may be relevant, and on the other hand the existence of idiosyncratic differences in preferences across decision makers.

Limiting ourselves to a purely deterministic treatment of taste heterogeneity can result in a loss of explanatory power, a lack of insights into the true extent of preference heterogeneity, and, depending on the shape and extent of the omitted heterogeneity, potential bias in key model outputs. With the significant increase in performance of personal computers and the availability of easy to use software, a majority of academic studies as well as a large share of applied work now allow for some degree of random preference heterogeneity in their models.

The key principle in any model aiming to capture random heterogeneity is to allow for a distribution in sensitivities across decision makers. Two main approaches exist, making use of either a discrete or a continuous distribution. The former generally relies on the notion of individual latent classes of decision makers, although this chapter also briefly looks at discrete mixtures at the level of individual coefficients. The latter relies on the specification of a multivariate continuous distribution for the coefficients in a choice model. In recent years

*The author wishes to acknowledge the role of Moshe Ben-Akiva, Joan Walker and Dinesh Gopinath in earlier work (Hess et al., 2009) which led to the developments in Sections 2.2.3 and 2.2.4.

especially, the continuous specification, often simply referred to as mixed logit (despite latent class and discrete mixtures also being mixtures of logit models), has come to dominate in many fields, notably in transport. It should be acknowledged that this chapter focusses solely on continuous mixed logit rather than probit (cf. Daganzo, 1979), which is seeing a resurgence thanks to the work by Bhat (2011); Bhat and Sidharthan (2011) - many of the same distinctions discussed here apply to probit, albeit that the distributional assumptions are stricter than in mixed logit. Similarly, the latent class discussions focus on a logit kernel, i.e. not discussing for example latent class probit structures.

The theoretical differences between continuous mixed logit and latent class logit were set out in detail by Greene and Hensher (2003), with empirical comparisons for example in Andrews et al. (2002); Hanley et al. (2002); Scarpa et al. (2005); Provencher and Bishop (2004); Shen (2009). Aside from providing further detail relating to the general structure, notably in terms of the correlation structure findings of Hess et al. (2009), a key focus of the present chapter is to look at important developments in latent class models since the work by Greene and Hensher (2003). First, in the ten years that have passed, a number of analysts have sought to combine the relative advantages of the two structures in hybrid models. Second, a larger body of (mainly empirical) research has made use of latent class structures with a view to capturing patterns of heterogeneity going beyond taste coefficients, looking at information processing, heuristics and heterogeneity in decision rules. Finally, there have also been further advances in terms of estimation performance for continuous mixture models in the last ten years, developments relating to the flexibility of mixing distributions, and growing use of continuous mixtures for capturing phenomena going beyond simple taste heterogeneity. Throughout the chapter, we do not seek to come to clear conclusions as to one model being superior to others, in fact, we rather highlight that the choice of an appropriate approach may be situation specific, in line with a number of past empirical comparisons.

# 2 Contrasts between model structures

## 2.1 Background methodology

Let $P_{nit}(\beta)$ give the probability of individual $n$ choosing alternative $i$ in choice situation $t$, conditional on a vector of taste coefficients $\beta$. In a multinomial logit (MNL) model (cf. McFadden, 1974), we have:

$$P_{nit}(\beta) = \frac{e^{V_{nit}}}{\sum_{j=1}^{J} e^{V_{njt}}}, \tag{1}$$

where $J$ is the total number of alternatives, and where the deterministic utility $V_{nit}$ is given by $f(x_{nit}, \beta, z_n)$, which is a function of the attributes of alternative $i$ as faced by individual $n$ in task $t$, $x_{nit}$, the vector of taste coefficients $\beta$[1], and the vector of socio-demographic characteristics $z_n$. With $ni^*t$ referring to the alternative chosen by individual $n$ in choice task $t$, the contribution by this individual to the likelihood function (across his/her $T_n$ choices) is simply given by $L_n(\beta) = \prod_{t=1}^{T_n} P_{ni^*t}$, where the aim is to find values of $\beta$ that maximise this function at the sample level, where simple maximum likelihood (ML) is the most commonly used approach. In this specification, deterministic heterogeneity is accommodated through the interaction between the vectors $\beta$ and $z_n$, allowing potentially for a mixture between continuous interactions and segmentations. We now look at the treatment of random heterogeneity in three different approaches.

#### 2.1.1 Continuous mixed logit

The first applications mixing logit probabilities across an assumed continuous distribution of elements in $\beta$ are generally credited to Boyd and Mellman (1980) and Cardell and Dunbar (1980), though widespread use of the model was to take almost two more decades, largely owing to computational complexity. In-depth discussions of the resulting model structure are given for example in McFadden and Train (2000), Hensher and Greene (2003) and Train (2009).

We now allow the vector $\beta$ to follow a random distribution with parameters $\Omega$, and the choice probabilities are given by:

$$P_{nit}(\Omega) = \int_{\beta} P_{nit}(\beta) f(\beta \mid \Omega) \, \mathrm{d}\beta, \tag{2}$$

where $P_{nit}$ is the MNL choice probability from Equation 1 and where $f(\beta \mid \Omega)$ gives the density function for the vector of taste coefficients $\beta$, which could allow for some fixed elements as well as correlation between individual random elements. Clearly, there is also scope for still incorporating deterministic heterogeneity through interaction between $\beta$ and $z_n$, whether at the level of the means or the dispersion parameters (cf. Greene et al., 2006).

Equation 2 would mean that the taste heterogeneity applies at the level of individual tasks. In the case of multiple observations per individual, we instead generally work with the assumption that sensitivities vary across individual decision makers, but stay constant across choices for the same individual, notwithstanding an interest in additional within-individual heterogeneity in some work (e.g Hess and Rose, 2009). Following the work of Revelt and Train (1998), we

[1]The inclusion of any alternative specific constants is not made explicit here.

then write the likelihood of the observed sequence of choices for decision maker $n$ as:

$$L_n(\Omega) = \int_\beta \left[\prod_{t=1}^{T_n} P_{ni^*t}(\beta)\right] f(\beta \mid \Omega)\,\mathrm{d}\beta. \tag{3}$$

The integral in Equation 3 (and Equation 2) does not have a closed form solution and the model is typically estimated using maximum simulated log-likelihood (MSL), i.e. the simulated analog of the ML typically used for MNL, averaging $\prod_{t=1}^{T_n} P_{ni^*t}(\beta)$ across a sufficiently large number of draws from $f(\beta \mid \Omega)$. Improvements in computer performance as well as the way in which draws from $f(\beta \mid \Omega)$ can be generated to better represent the distribution (see e.g. Bhat, 2001, 2003; Hess et al., 2006) have led to widespread use of the model in many fields. A growing number of studies also rely on Bayesian techniques, which are especially useful when the dimensionality of $\beta$ is large (see Train 2009, chapter 12 for an overview), though much of the work to date has been on datasets with limited sample size and limited numbers of alternatives.

Before proceeding, it should be noted that this discussion has centred on using mixed logit to accommodate heterogeneity in sensitivities across respondents, often referred to as random parameters logit. A mathematically equivalent specification, referred to as error components logit (cf. Walker et al., 2007), uses the random terms to capture phenomenae such as correlation between alternatives or choices, as well as heteroscedasticity. Capturing these effects in a latent class approach is less straightforward (or even possible), and this is a motivation for combining the approaches, as discussed later in the chapter.

### 2.1.2 Simple discrete mixtures

An alternative to the use of continuous distributions for individual elements in $\beta$ is to allow for a finite number of possible values for each element in $\beta$, with an associated probability. This gives rise to what is variably called a discrete mixture model or a mass point logit model, with discussions in Gopinath (1995); Dong and Koppelman (2003); Wedel et al. (1999); Hess et al. (2007); Train (2008).

Let us assume that $\beta$ has $K$ different elements, where we allow for $S_k$ different values for $\beta_k$, where $S_k$ needs to be specified by the analyst. With different weights for the different possible values for $\beta_k$ given by $\pi_{sk}$, we would then have that:

$$L_n(\beta, \pi) = \sum_{s_1}^{S_1} \sum_{s_2}^{S_2} \dots \sum_{s_K}^{S_K} \pi_{s_1} \cdot \pi_{s_2} \cdot \dots \cdot \pi_{s_K} \prod_{t=1}^{T_n} P_{ni^*t}(\beta_{s_1}, \beta_{s_2}, \dots, \beta_{s_K}) \tag{4}$$

i.e. a weighted average across all the possible combinations of values in $\beta$, with the weight for each combination being given by a product of the respective weights for the individual elements in $\beta$, with $\pi$ grouping together all individual weights, where $0 \leq \pi_{s_k} \leq 1$, $\forall s, k$ and $\sum_{s_k}^{S_k} \pi_{s_k} = 1$, $\forall k$. The likelihood for this model has a closed form solution and no simulation is thus required in estimation. However, it can be seen straightaway that even with a low number of elements ($K$) in $\beta$ and modest values for the number of possible values ($S_k$) for each $\beta_k$, the number of combinations rapidly becomes very large and leads to computational complexity not dissimilar from the estimation of a continuous mixed logit model. As an example, many applications using mixed logit rely on fewer than say 250 draws in simulation based estimation even with as many as 5 random coefficients. This would mean that $P_{ni^*t}(\beta)$ in Equation 3 would need to be evaluated 250 times. If we estimated a discrete mixture analog with $S_k = 3, \forall k$, we would need to evaluate 243 terms in the weighted sum in Equation 4.

Choosing an appropriate value of $S_k \, \forall k$ is down to the analyst, and is a non-trivial task. A key component of this is that in the estimation of discrete mixture models, in common with latent class structures, we see a rapid explosion in the number of parameters and the often observed phenomenon of multiple elements for $\beta_k$ collapsing to the same value, which is especially likely in the case of strongly peaked distributions. The latter issue can be addressed to some extent by moving away from simple maximum likelihood estimation and making use of EM algorithms, with in-depth discussions in Train (2008). In terms of the explosion in the number of parameters and the question of improvements in fit justifying such increases, it is wise to move to model fit criteria which penalise the inclusion of additional parameters more strongly, with typical approaches being the the Akaike information criterion (AIC) or the Bayesian information criterion (BIC); see for example Mittelhammer et al. (2000, section 18.5).

### 2.1.3 Latent class structures

Latent class models have a long tradition in choice modelling. Their development is often traced back to work by Kamakura and Russell (1989) and Gupta and Chintagunta (1994), with important work also in Swait (1994), Gopinath (1995) and Bhat (1997). The heterogeneity in sensitivities across individuals is now accommodated by making use of separate classes with different values for the vector of taste coefficients $\beta$ in each class. The distinction from a simple discrete mixture as discussed above is that the classes capture joint distribution of the individual elements in $\beta$. Specifically, in a model with $S$ classes, we would have $S$ instances of the vector $\beta$, say $\beta_1$ to $\beta_S$, with a possibility of some of the elements in $\beta$ staying constant across some of the classes. As with discrete mixture models, the number of classes $S$ needs to be specified by the analyst.

A Latent Class model uses a probabilistic class allocation model, where individual $n$ belongs to class $s$ with probability $\pi_{ns}$, and where $0 \leq \pi_{ns} \leq 1 \; \forall n, s$ and $\sum_{s=1}^{S} \pi_{ns} = 1, \; \forall n$. Latent class models are generally specified with an underlying MNL model, but can easily be adapted for more general underlying structures such as nested or cross-nested logit - the same clearly also applies to continuous mixtures (cf. Garrow, 2004; Hess et al., 2005a) or discrete mixtures.

Let $P_{nit}(\beta_s)$ give the probability of individual $n$ choosing alternative $i$ in choice task $t$, conditional on $n$ falling into class $s$. The likelihood of the observed set of choices for $n$, working on the assumption of intra-individual homogeneity in sensitivities, is then given by:

$$L_n(\beta, \pi) = \sum_{s=1}^{S} \pi_{ns} \left( \prod_{t=1}^{T_n} P_{ni^*t}(\beta_s) \right) \tag{5}$$

with $P_{ni^*t}(\beta_s)$ again being given by Equation 1.

In common with the discrete mixture model, no simulation is required in the estimation of latent class models of the form above, so that for example simple ML estimation can be used. However, in contrast with the discrete mixture model, the number of combinations of values is a function only of $S$ and not of the number of elements $(K)$ in $\beta$. The issue of choosing an appropriate value for $S$ remains.

In the most basic version of a latent class logit model (Kamakura and Russell, 1989), the class allocation probabilities are constant across individuals such that $\pi_{ns} = \pi_s, \; \forall n$. The real flexibility however arises when the class allocation probabilities are not constant across individuals but when a class allocation model is used to link these probabilities to characteristics of the individuals (Gupta and Chintagunta, 1994). Typically, these characteristics would take the form of socio-demographic variables, such as income, age and employment status. With $z_n$ giving the concerned vector of characteristics for individual $n$, and with the class allocation model taking on a logit form (this is a common specification rather than an absolute requirement), the probability of individual $n$ falling into class $s$ would be given by:

$$\pi_{ns} = \frac{e^{\delta_s + g(\gamma_s, z_n)}}{\sum_{l=1}^{S}, e^{\delta_l + g(\gamma_l, z_n)}}, \tag{6}$$

where $\delta_s$ is a class-specific constant[2], $\gamma_s$ is a vector of parameters to be estimated and $g(\cdot)$ gives the functional form of the *utility* function for the class allocation

[2]In a model with generic class allocation probabilities, such as in Kamakura and Russell (1989), only these constants would be estimated.

model - appropriate normalisation is to be used for both $\delta$ and $\gamma$. The class allocation model allows us to probabilistically allocate individuals to different classes depending on their socio-demographic characteristics.

We have already discussed the issue of the proliferation of parameters above in the context of discrete mixtures, and the same issues apply in latent class models. Similarly, estimation with larger numbers of classes can be problematic with parameters collapsing to the same values across classes or some classes obtaining very small probabilities, and here, the EM algorithm can once again be one possible solution, discussed in Train (2008) but also earlier on by Bhat (1997). Nevertheless, it remains almost unavoidable that with a large number of classes, some of the coefficient values may not be significant across classes, or lend themselves to easy interpretation.

## 2.2 Contrasts

This section provides some theoretical contrasts between model structures, focussing on continuous mixed logit models and latent class structures. This extends on work by Bhat (1997) who derived elasticity expressions as well as on the discussions in Greene and Hensher (2003), and complements a substantial body of empirical comparisons between the structures, for example in Andrews et al. (2002); Hanley et al. (2002); Greene and Hensher (2003); Scarpa et al. (2005); Provencher and Bishop (2004); Shen (2009). The evidence in these empirical comparisons is mixed, highlighting that both models have their advantages and that the choice of an appropriate structure will depend on the data at hand.

### 2.2.1 Taste heterogeneity

The main emphasis in discussing mixed logit and latent class logit is on their ability to capture random heterogeneity across individuals in addition to deterministic heterogeneity such as also allowed for in simple MNL models. The two structures do this in very different ways, as already outlined in Section 2.1. In the basic specification of the continuous mixed logit model, the random heterogeneity is entirely random, and while such a specification is common in most of the empirical work, it is clearly possible (and indeed desirable) to link the random heterogeneity to observable individual characteristics, typically through making the parameters of the random distribution a function of such characteristics (cf. Greene et al., 2006). A specification not linking the random heterogeneity to individual characteristics is similarly possible in a latent class framework (Kamakura and Russell, 1989), though here, the typical specification does rely on a parameterisation of the class allocation probabilities on socio-demographics such as in Equation 6, meaning that the class allocation probabilities (and hence the implied sensitivities) vary also as a function of these individual characteristics.

In both models, the assumptions made at the specification stage can have important influences on parameter estimates and substantive model results such as willingness-to-pay measures. It is well documented that the need to determine which coefficients should be allowed to vary across individuals, and what distributions are to be used is a key issue facing analysts using continuous mixed logit models. There is a strong influence of these assumptions on model results (see e.g. Hess et al., 2005b), and while much progress has been made since the discussions by Greene and Hensher (2003) with flexible and non-parametric (Fosgerau, 2006, 2007; Fosgerau and Bierlaire, 2007) distributions, numerous applications continue to rely on misguided specifications, also in relation to ensuring the existence of moments for ratios of coefficients (Daly et al., 2012), notwithstanding the possible solution of working in willingness-to-pay space (Train and Weeks, 2005).

A key limitation of most parametric distributions is a strong shape assumption and general uni-modality. In theory, the same does not apply with latent class structures as no assumptions are made on the relationship between the values for a given coefficient across classes, thus allowing for flexible shapes and multi-modality. This is often touted as an advantage of latent class models. In practice however, the decision by the analyst on the number of classes to use has major implications for the shape of the distribution, and the shape of the *true* underlying distribution, for example in terms of the relative importance of different modes, will have impacts on the ability to retrieve sensitivities in less well represented parts of the distribution. With both models, the ability to retrieve the *true* patterns of heterogeneity in the data thus depends both on the shape of that heterogeneity and the specification used by the analyst.

#### 2.2.2 Posterior analysis

The estimation of either type of models provides information relating to the sample level patterns of heterogeneity. By making the parameters of the continuous distribution in mixed logit models a function of socio-demographics or by incorporating socio-demographics in the class allocation model in a latent class structure, we can obtain further insights into the likely location of a given type of individual on that sample level distribution. This however treats two individuals who are identical on those socio-demographics as also having identical sensitivities, contrary to the notion of random heterogeneity. Further insights can be obtained post estimation in a Bayesian manner, by calculating information relating to a given individual's sensitivities on the basis of the sample level model estimates and that individual's observed choices.

In a continuous mixed logit context, these calculations are straightforward, as discussed for example by Train (2009, chapter 12). Specifically, we have from Equation 3 that the likelihood of the observed sequence of choices for person $n$ is given by:

$$L_n(\Omega) = \int_\beta L_n(\beta) f(\beta \mid \Omega) \, \mathrm{d}\beta. \tag{7}$$

where $L_n(\beta) = \prod_{t=1}^{T_n} P_{ni^*t}(\beta)$.

Using Bayes' rule, we can then rewrite this as:

$$L(\beta_n \mid C_n) = \frac{L_n(\beta) f(\beta \mid \Omega)}{L_n(\Omega)} \tag{8}$$

This gives us the probability of given values for $\beta_n$, conditional on the observed choices ($C_n$) for individual $n$, where it is important to remember that $\beta_n$ is not observed but is distributed. It is then straightforward to for example calculate a conditional mean for $\beta_n$ as:

$$\bar{\beta_n} = \int_{\beta_n} \beta_n L(\beta_n \mid C_n) \, \mathrm{d}\beta_n, \tag{9}$$

with similar calculations to obtain the corresponding variance or other measures.

It is similarly possible to calculate a number of posterior measures from latent class models. A key example comes in the form of posterior class allocation probabilities, where the posterior probability of individual $n$ for class $s$ is given by:

$$\widehat{\pi_{ns}} = \frac{\pi_{ns} L_n(\beta_s)}{L_n(\beta, \pi_n)}, \tag{10}$$

where $L_n(\beta_s)$ gives the likelihood of the observed choices for individual $n$, conditional on class $s$.

To explain the benefit of these posterior class allocation probabilities, let us assume that we have calculated for each class in the model a given measure $w_s = \frac{\beta_{s1}}{\beta_{s2}}$, i.e. the ratio between the first two coefficients. Using $\overline{w_n} = \sum_{s=1}^{S} \pi_{ns} w_s$ simply gives us a sample level mean for the measure $w$ for an individual with the specific observed characteristics of person $n$. These characteristics (in terms of socio-demographics used in the class allocation probabilities) will however be common to a number of individuals who still make different choices, and the most likely value for $w$ for individual $n$, conditional on his/her observed choices, can now be calculated as $\widehat{w_n} = \sum_{s=1}^{S} \widehat{\pi_{ns}} w_s$.

Finally, it might also be useful to produce a profile of the membership in each class. From the parameters in the class allocation probabilities, we know which class is more or less likely to capture individuals who posses a specific characteristic, but this is not taking into account the multivariate nature of these characteristics. Let us for example assume that a given socio-demographic characteristic $z_c$ is used in the class allocation probabilities, with associated parameter $\gamma_c$, and using a linear parameterisation in Equation 6. We can then calculate the likely value for $z_c$ for an individual in class $s$ as:

$$\widehat{z_{cs}} = \frac{\sum_{n=1}^{N} \widehat{\pi_{ns}} z_{cn}}{\sum_{n=1}^{N} \widehat{\pi_{ns}}}, \tag{11}$$

where we again use the posterior probabilities to take into account the observed choices. Alternatively, we can also calculate the probability of an individual in class $s$ having a given value $\kappa$ for $z_c$ by using:

$$P\left(\widehat{z_{cs} = \kappa}\right) = \frac{\sum_{n=1}^{N} \widehat{\pi_{ns}} \left(z_{cn} = \kappa\right)}{\sum_{n=1}^{N} \widehat{\pi_{ns}}}. \tag{12}$$

### 2.2.3 Correlation between coefficients

In models without random taste heterogeneity, any correlation in the distribution of individual coefficients can solely arise as a result of interactions with socio-demographic attributes and specifically where multiple coefficients interact with the same socio-demographic characteristics. As an example, one could imagine a situation where cost sensitivity decreases with income while time sensitivity increases with income, resulting in negative correlation between the time and cost coefficients across the sample.

In a continuous mixture model, additional correlation can be accommodated by specifying a joint distribution for the random taste coefficients. While most estimation packages allow users to specify multivariate Normal distributions, the vast majority of continuous mixture applications continue to make use of independently distributed taste coefficients, despite the obvious simplification and likely lack in performance this engenders. Correlation is rarely introduced in models not based on the Normal distribution, one exception being given in Walker (2001), while flexible correlation structures in continuous mixtures are also a benefit of the GMNL specification of mixed logit (Fiebig et al., 2010).

In a latent class model, correlation between coefficients is an inherent characteristic of the model structure as long as the two coefficients in question take on more than one value across the $S$ classes. As highlighted repeatedly earlier in the chapter, the nature of the distribution of sensitivities in a latent class model is a function of both the estimates of the class specific $\beta$ vectors as well as the individual specific class allocation probabilities. A characterisation of these distributions at the level of individuals should thus use the posterior probabilities to encompass the information gained from observed choices. Drawing on Hess et al. (2009), we can then easily see that:

$$
\begin{aligned}
cov\left(\beta_{n1},\beta_{n2}\right) &= E\left[\left(\beta_{n1}-E\left(\beta_{n1}\right)\right)\left(\beta_{n2}-E\left(\beta_{n2}\right)\right)\right] \\
&= E\left(\beta_{n1}\,\beta_{n2}\right)-E\left(\beta_{n1}\right)E\left(\beta_{n2}\right) \\
&= \sum_{s=1}^{S}\widehat{\pi_{ns}}\beta_{1,s}\beta_{2,s}-\left(\sum_{s=1}^{S}\widehat{\pi_{ns}}\beta_{1,s}\right)\left(\sum_{s=1}^{S}\widehat{\pi_{ns}}\beta_{2,s}\right)
\end{aligned} \tag{13}
$$

For ease of notation, let $\alpha=\beta_1$ and $\gamma=\beta_2$ in which case Equation 13 can be written as:

$$
cov\left(\alpha_n,\gamma_n\right)=\sum_{s=1}^{S}\widehat{\pi_{ns}}\alpha_s\gamma_s-\left(\sum_{s=1}^{S}\widehat{\pi_{ns}}\alpha_s\right)\left(\sum_{s=1}^{S}\widehat{\pi_{ns}}\gamma_s\right) \tag{14}
$$

A special situation arises when $S=2$, in which case the class allocation probabilities have no effect on the sign of the correlation. Indeed, with the notation from Equation 14, we then have:

$$
\begin{aligned}
cov\left(\alpha_n,\gamma_n\right) &= \widehat{\pi_{n1}\pi_{n2}}\left[\alpha_1\left(\gamma_1-\gamma_2\right)+\alpha_2\left(\gamma_2-\gamma_1\right)\right] \\
&= \widehat{\pi_{n1}\pi_{n2}}\left[\left(\alpha_1-\alpha_2\right)\left(\gamma_1-\gamma_2\right)\right],
\end{aligned} \tag{15}
$$

where the sign of $cov\left(\alpha_n,\gamma_n\right)$ only depends on the changes in the two elements in $\alpha$ and $\gamma$ across the two classes.

It should be noted that, using Equation 8, we also obtain individual specific distributions for the coefficients in a continuous mixed logit model, where any correlation between these will be a function of the choices (leading to the posterior distributions), the assumptions in relation to the sample level covariance structure, and any incorporation of socio-demographic characteristics in the specification of the distributions. Unlike with a latent class structure, a simple analytic solution such as shown here is not straightforward.

#### 2.2.4 Disaggregate elasticities

As a final step, we briefly contrast disaggregate point elasticities in the different models (see also Bhat, 1997). With both types of mixtures having a logit kernel, it is worth restating the well known MNL elasticities (see e.g. Ben-Akiva and Lerman 1985), with the direct elasticity in relation to attribute $x$ given by:

$$
E_{i,x_{ni}}=\frac{\partial V_{ni}}{\partial x_{ni}}x_{ni}\left(1-P_{ni}\left(\beta\right)\right), \tag{16}
$$

where, with a linear in attributes specification, $\frac{\partial V_{ni}}{\partial x_{ni}} = \beta_x$. The corresponding cross-elasticity is given by:

$$E_{i,x_{nj}} = -\frac{\partial V_{nj}}{\partial x_{nj}} x_{nj} P_{nj}\left(\beta\right), \tag{17}$$

exhibiting the *IIA* characteristic at the disaggregate level - note this does not imply *IIA* in the aggregate elasticities (Louviere et al., 2000).

In a continuous mixed logit model, the direct elasticity (see e.g. Train 2009) is given by:

$$E_{i,x_{ni}} = \frac{\int_{\beta} \frac{\partial V_{ni}}{\partial x_{ni}} x_{ni} \left(1 - P_{ni}\left(\beta\right)\right) P_{ni}\left(\beta\right) f\left(\beta \mid \Omega\right) \mathrm{d}\beta}{\int_{\beta} P_{ni}\left(\beta\right) f\left(\beta \mid \Omega\right) \mathrm{d}\beta}, \tag{18}$$

with the cross-elasticity being:

$$E_{i,x_{nj}} = -\frac{\int_{\beta} \frac{\partial V_{nj}}{\partial x_{nj}} x_{nj} P_{nj}\left(\beta\right) P_{ni}\left(\beta\right) f\left(\beta \mid \Omega\right) \mathrm{d}\beta}{\int_{\beta} P_{ni}\left(\beta\right) f\left(\beta \mid \Omega\right) \mathrm{d}\beta}, \tag{19}$$

where this varies across alternatives, such that it does not exhibit the *IIA* property. Here, it can be seen that the elasticities are given by an integration of logit elasticities.

In a latent class logit model, the direct elasticity is given by:

$$\begin{aligned} E_{i,x_{ni}} &= \frac{\partial P_{ni}\left(\beta\right)}{\partial x_{ni}} \frac{x_{ni}}{P_{ni}\left(\beta\right)} \\ &= \left(\sum_{s=1}^{S} \pi_{ns} \frac{\partial P_{ni}\left(\beta_s\right)}{\partial x_{ni}}\right) \frac{x_{ni}}{P_{ni}\left(\beta\right)} \\ &= \left(\sum_{s=1}^{S} \pi_{ns} \frac{\partial V_{nis}}{\partial x_{ni}} P_{ni}\left(\beta_s\right)\left(1 - P_{ni}\left(\beta_s\right)\right)\right) \frac{x_{ni}}{P_{ni}\left(\beta\right)} \\ &= \sum_{s=1}^{S} \frac{\pi_{ns} P_{ni}\left(\beta_s\right)}{P_{ni}\left(\beta\right)} \left[\frac{\partial V_{nis}}{\partial x_{ni}} x_{ni}\left(1 - P_{ni}\left(\beta_s\right)\right)\right]. \end{aligned} \tag{20}$$

It can be seen that the term in square brackets corresponds to a MNL direct elasticity for a specific class in the latent class model. This means that the direct elasticities are a weighted sum of MNL elasticities, with the weights being given by multiplying the class membership probability with the class specific conditional probability and by dividing this product by the marginal probability.

It can similarly be seen that the cross-elasticities are given by a weighted sum of MNL cross-elasticities, with:

$$\begin{aligned} E_{i,x_{nj}} &= \frac{\partial P_{ni}(\beta)}{\partial x_{nj}} \frac{x_{nj}}{P_{ni}(\beta)} \\ &= \sum_{s=1}^{S} \pi_{ns} \left( -\frac{\partial V_{njs}}{\partial x_{nj}} P_{ni}(\beta_s) P_n(j \mid \beta_s) \right) \frac{x_{nj}}{P_{ni}(\beta)} \\ &= \sum_{s=1}^{S} \frac{\pi_{ns} P_{ni}(\beta_s)}{P_{ni}(\beta)} \left[ -\frac{\partial V_{njs}}{\partial x_{nj}} x_{nj} P_{nj}(\beta_s) \right]. \end{aligned} \tag{21}$$

The contrasts and similarities between the continuous mixed logit and latent class logit elasticities are clear. Both are a function of MNL elasticities, and both avoid the *IIA* assumption. The mixture in the continuous model means a reliance on integration/simulation, while the latent class model uses weighted summation. In all of the models, the elasticities vary as a function of the attribute levels of the alternatives and hence the probabilities, but also as a function of any socio-demographic interactions with $\beta$. In the latent class model, we have the additional influence of socio-demographics through the class allocation probabilities, where, in the continuous mixed logit model, the same is the case if the parameters of the distribution are a function of decision maker characteristics.

## 3 Combining continuous mixed logit and latent class

The discussion in the previous section has highlighted the contrasts between continuous mixed logit and latent class logit models. Both structures have strengths and weaknesses and it should thus come as no surprise that a number of researchers have put forward structures that combine the two approaches.

The first published such application seems to be the work of Walker and Li (2006), who add additional continuous variation into a latent class structure in the form of error component terms aimed at capturing correlation across alternatives and across choices for the same decision maker. Specifically, their model takes the general form of:

$$L_n(\beta, \pi, \sigma) = \sum_{s=1}^{S} \pi_{ns} \int_{\eta} \prod_{t=1}^{T_n} P_{ni^*t}(\beta_s, \eta) f(\eta \mid \sigma) \, \mathrm{d}\eta \tag{22}$$

In this specification, the continuous random components $\eta$ follow Normal distributions with a mean of zero and with standard deviations captured in the vector $\sigma$. With a view to capturing correlation across alternatives as well as across choices for the same decision maker, these error components are generic across classes within the overarching latent class structure.

A different direction in combining the two structures uses the continuous component to allow for additional heterogeneity in sensitivities within given classes, where this heterogeneity varies across classes. In effect, this can be described most straightforwardly as a latent class mixed logit, using a continuous mixed logit model inside each class to capture heterogeneity. In particular, we would write:

$$L_n\left(\Omega,\pi\right)=\sum_{s=1}^{S}\pi_{ns}\int_{\beta_s}\prod_{t=1}^{T_n}P_{ni^*t}\left(\beta_s\right)f\left(\beta_s\mid\Omega_s\right)\mathrm{d}\beta_s \tag{23}$$

In this model, we have that the vector of coefficients $\beta_s$ is specific to class $s$ and contains at least some components that are distributed randomly across decision makers within that class, according to $f\left(\beta_s \mid \Omega_s\right)$, where $\Omega=\langle\Omega_1,\ldots,\Omega_S\rangle$. Such a specification has been used by Bujosa et al. (2010) on revealed preference data (with $T_n=1$, $\forall n$) and Greene and Hensher (2013) on stated preference data.

In a different direction, there has in recent years been growing interest in allowing for intra-agent heterogeneity in addition to inter-agent heterogeneity (Bhat and Sardesai, 2006; Hess and Rose, 2009) making use of a specification such as:

$$L_n\left(\Omega_\gamma,\Omega_\alpha\right)=\int_{\alpha}\prod_{t=1}^{T_n}\left[\int_{\gamma}P_{ni^*t}\left(\beta=\alpha+\gamma\right)f\left(\gamma\mid\Omega_\gamma\right)\mathrm{d}\gamma\right]h\left(\alpha\mid\Omega_\alpha\right)\mathrm{d}\alpha, \tag{24}$$

where $\beta=\alpha+\gamma$ with $\alpha$ distributed across decision makers and $\gamma$ distributed across individual choices for the same decision maker. Models of this type have proven to be very difficult to estimate due to the double layer of integration, and this raises the question whether replacing one layer with weighted summation through a latent class structure would be beneficial, in essence adapting Equation 23 by moving the position of the integral to the level of an individual choice:

$$L_n\left(\Omega,\pi\right)=\sum_{s=1}^{S}\pi_{ns}\prod_{t=1}^{T_n}\int_{\beta_s}P_{ni^*t}\left(\beta_s\right)f\left(\beta_s\mid\Omega_s\right)\mathrm{d}\beta_s. \tag{25}$$

This specification would now mean that the latent class structure captures the variation in sensitivities across individual decision makers through the class structure, while the integration over class specific random coefficients captures additional heterogeneity across choices for individual decision makers.

Finally, the focus above has solely been on allowing for additional continuous random heterogeneity for the choice model parameters within individual latent classes. However, the drivers of the class allocation model could similarly include other latent factors (such as attitudes) that should be explicitly captured in the model specification. Such a specification, as discussed by Walker and Ben-Akiva (2002) and Hess et al. (2013a), relies on specifying a set of latent variables $\alpha_n = h(\theta, z_n) + \eta_n$ where $\eta_n$ is a vector of standard normal random variables. These $\alpha_n$ terms, which can for example represent underlying attitudes and perceptions, are then used in parameterising the class allocation probabilities, rewriting Equation 26 to:

$$\pi_{ns} = \frac{e^{\delta_s + g(\gamma_s, z_n) + \tau_s \alpha_n}}{\sum_{l=1,}^{S} e^{\delta_l + g(\gamma_l, z_n) + \tau_l \alpha_n}}. \tag{26}$$

At the same time, $\alpha_n$ is used to explain answers by decision maker $n$ to a set of attitudinal questions, grouped together in $I_n$, with e.g.: $I_n = \zeta \alpha_n + \nu$ where $\nu$ is a vector of random disturbances. The estimation then jointly maximises the likelihood of the observed choices and answers to the attitudinal questions, through having:

$$L_n(\beta, \gamma, \theta, \delta, \tau) = \int_{\eta_n} \sum_{s=1}^{S} \pi_{ns} \left( \prod_{t=1}^{T_n} P_{ni^*t}(\beta_s) \right) P(I_n \mid \alpha_n) \, \phi(\eta_n) \, \mathrm{d}\eta_n \tag{27}$$

where $\pi_{ns}$ is now also a function of $\alpha_n$.

## 4 Confirmatory latent class structures: recent developments and future research needs

The discussion of latent class models thus far has centred on a form of the model which is particularly accessible as there are well-established estimation software programs to estimate such models. This model can be referred to as an *exploratory* latent class model - the analyst merely specifies the number of classes and selects the attributes which are to be used in the class allocation model, and the rest is left to model estimation. This will, with a suitably robust estimation approach, lead to a well fitting structure for a model of the specified size, but there is no guarantee that it will lead to reasonable results or meaningful insights into behaviour, much the same way as when just estimating a continuous mixed logit model with standard distributions.

An alternative approach is to use what can be termed a *confirmatory* approach, imposing different a-priori restrictions on the specifications of the class membership models and on the class specific choice probabilities, and estimating parameters subject to these constraints. This applies for example when the latent classes are based on a priori behavioural hypotheses. An example of such a confirmatory approach is given in Gopinath (1995), while the work by Train (2008) in the context of estimating weights for fixed points in a distribution is also an example of a confirmatory approach.

An added reason for discussing confirmatory approaches in the present chapter is a strong stream of research activity making use of such models in two related but distinct contexts in recent years, namely the domains of information processing and decision rule heterogeneity.

### 4.1 Attribute processing strategies

The field of information processing strategies (IPS) or attribute processing strategies (APS) is a burgeoning area of work, especially in the context of stated choice surveys. The main emphasis has been on the question whether some decision makers may actually make their choices based on only a subset of the attributes that describe the alternatives at hand. This phenomenon is typically referred to as attribute non-attendance or attribute ignoring, and an in-depth review of work in this area is given in Hensher (2010), and also the Hensher contribution in the present volume. The interest in this topic in this chapter comes in the context of ways to accommodate attribute non-attendance in models.

A key role in this area was played by the early discussions in Hess and Rose (2007), who proposed the use of a latent class approach to accommodate attribute non-attendance, a method since adopted by numerous other studies (e.g. Hensher et al., 2012; Scarpa et al., 2009; Hensher and Greene, 2010; Hole, 2011; Campbell et al., 2010). With this approach, different latent classes relate to different combinations of attendance and non-attendance across attributes. For each attribute treated in this manner, there exists a non-zero coefficient (to be estimated), which is used in the *attendance classes*, while the attribute is not employed in the *non-attendance* classes, i.e. the coefficient is set to zero. In a complete specification, covering all possible combinations, this would thus lead to $2^K$ classes, with $K$ being the number of attributes, where a given coefficient will take the same value in all classes where that attribute is included. A simplification so as to avoid estimating $2^K$ separate class allocation probabilities is to use a multiplicative approach, i.e. treating non-attendance independent across attributes, much as in the discrete mixture discussions in Section 2.1.2, and as discussed in Hole (2011).

In addition to the vector $\beta$, we now have a $S$x$K$ matrix $\Lambda$, in which each row contains a different combination of 0 and 1 elements, where $S = 2^K$. Next, let

$A \circ B$ be the element-by-element product of two equally sized vectors $A$ and $B$, yielding a vector $C$ of the same size, where the $k^{th}$ element of $C$ is obtained by multiplying the $k^{th}$ element of $A$ with the $k^{th}$ element of $B$. Using this notation, the specific values used for the taste coefficients in class $s$ are then given by the vector $\beta_s = \beta \circ \Lambda_s$. The likelihood for decision maker $n$ is then given by:

$$L_n\left(\beta, \pi\right) = \sum_{s=1}^{S} \pi_s \prod_{t=1}^{T} P_{ni^*t}\left(\beta_s = \beta \circ \Lambda_s\right). \tag{28}$$

The overall findings of the growing body of work using the latent class specification point towards a significant portion of people ignoring attributes, including cost variables. In later work, Hess et al. (2013b) argue that an important shortcoming of this simple latent class approach is the reliance on only two possible values for each coefficient, one of which is fixed to zero, where the latter might capture sensitivities close to (rather than equal to) zero, while the two class structure might simply be a proxy for more general taste heterogeneity. Hess et al. (2013b) put forward a model which combines the confirmatory latent class structure with additional continuous heterogeneity in the non-zero coefficient values, aiming to reduce the risk of the class at zero capturing low sensitivities. The likelihood function for decision maker $n$ is simply rewritten as:

$$L_n\left(\Omega, \pi\right) = \sum_{s=1}^{S} \pi_s \int_{\beta} \prod_{t=1}^{T} P_{ni^*t}\left(\beta_s = \beta \circ \Lambda_s\right) f\left(\beta \mid \Omega\right) \mathrm{d}\beta. \tag{29}$$

Empirical evidence by Hess et al. (2013b) on multiple datasets reveals major improvements in fit by the specification in Equation 29 over the model in Equation 28, along with a reduction in the implied rates of non-attendance, which crucially however remains above zero for many attributes. Further work on this structure was subsequently conducted by Collins et al. (2013).

### 4.2 Decision rule heterogeneity and other mixtures of models

Although structures belonging to the family of random utility models have come to dominate, it is important to recognise that alternative paradigms for decision making have been proposed, for example the elimination by aspects model of Tversky (1972), but also more recent work based on the concepts of happiness (Abou-Zeid and Ben-Akiva, 2010) and regret (Chorus et al., 2008). The evidence in the literature is that which paradigm works best is very much dataset specific. Hess et al. (2012) put forward the hypothesis that variations in decision rules may be across decision makers with a single dataset, not just across datasets, and propose the use of a confirmatory latent class approach in this context.

Specifically, let $L_n(\beta_m, m)$ give the probability of the observed sequence of choices for decision maker $n$, conditional on using a choice model identified as $m$, where this uses a vector of parameters $\beta_m$. The Hess et al. (2012) framework is based on the idea that $M$ different behavioural processes are used in the data. The probability for the sequence of choices observed for decision maker $n$ is now given by:

$$L_n(\beta, \pi) = \sum_{m=1}^{M} \pi_{nm} L_n(\beta_m, m), \tag{30}$$

where we use different behavioural processes in different classes, with the probability of decision rule class $m$ for decision maker $n$ given by $\pi_{nm}$. Hess et al. (2012) additionally allow for random heterogeneity in parameters within individual decision rule classes, such that:

$$L_n(\Omega, \pi) = \sum_{m=1}^{M} \pi_{nm} \int_{\beta_m} L_n(\beta_m, m) f(\beta_m, \Omega_m) \, \mathrm{d}\beta_m, \tag{31}$$

where $\beta_m \sim f(\beta_m, \Omega_m)$ and $\Omega_m = \langle \Omega_1, \ldots, \Omega_M \rangle$.

Hess et al. (2012) use the model to allow for mixtures between random utility maximisation, random regret minimisation and elimination by aspects. In later work, Hess and Stathopoulos (2012) use an approach as in Walker and Ben-Akiva (2002) and Hess et al. (2013a), making the class allocation a function of a latent factor, which in this case also explains decision makers' real world choices.

At this stage, it should be noted that a latent class model mixing various decision rules is just one example of a wider set of structures that combine different models. A further possibility for example would be a model using different GEV nesting structures in different latent classes, somewhat similar in aims to the work of Ishaq et al. (2013). Finally, a separate body of work looks at using different choice sets in different classes, in the context of choice set generation work (see e.g. Swait and Ben-Akiva 1985; Ben-Akiva and Boccara 1995 and Gopinath 1995, section 2.7).

# 5 Summary and conclusions

This chapter has revisited the topic of contrasting continuous mixed logit models and latent class structures, ten years on from the work by Greene and Hensher (2003). The key distinction between the models clearly remains that the former uses continuous distributions of sensitivities while the latter uses a finite number of classes of sets of coefficient values. Both models allow for deterministic heterogeneity, along with an influence of observed components such as socio-demographics on the nature of the random heterogeneity, albeit that this is arguably done less frequently with continuous mixtures. While latent class models lead to reduced computational costs compared to continuous mixtures, they are characterised by a rapid increase in the number of parameters. Post analysis calculations of measures of heterogeneity, correlation and elasticities are relatively straightforward in both models, again with the distinction between simulation and averaging across classes, where this chapter provides some additional insights for correlation in latent class models. A further point not touched on thus far is that of using the models in application/forecasting, where the computational cost of latent class models is lower, which is important especially in the case of micro-simulation uses.

The key motivation for extending on the discussions in Greene and Hensher (2003) can be found in the many methodological developments that have taken place in the last ten years. On the continuous mixed logit side, progress has been made in estimation capabilities, the flexibility of parametric and non-parametric distributions, and the treatment of phenomena such as inter-alternative correlation and heteroscedasticity. Especially the latter two are not as straightforward to capture in a latent class framework, and this, along with a desire for more flexible specifications of heterogeneity, has motivated work on combining the two approaches, for example in Walker and Li (2006); Bujosa et al. (2010); Greene and Hensher (2013); Hess et al. (2013b). Similarly, the major interest in modelling attitudes and perceptions (cf. Ben-Akiva et al., 2002) has led to hybrid models in which the class allocation is in part driven by these latent psychological constructs (see e.g. Walker and Ben-Akiva, 2002; Hess et al., 2013a).

The other key focus of the chapter has been the added interest in latent class structures in recent years in the context of attribute processing strategies (see the summary in Hensher, 2010) and decision rule heterogeneity (cf. Hess et al., 2012). A substantial number of studies now make use of confirmatory latent class approaches which estimate allocation probabilities for classes characterised by specific behavioural assumptions. With growing interest in ever richer specifications of heterogeneity, the uptake of latent class structures in this context is bound to increase further, likely in conjunction with continuous layers of heterogeneity, especially given the hype of activity on treatments of latent psychological factors such as attitudes and perceptions, as evidence for example in Hess and Stathopoulos (2012).

There remains substantial scope for future work in this area, both theoretical and empirical. A key avenue for work especially with some of the most complex structures is that of estimation. Notwithstanding the work on EM algorithms by

Bhat (1997) and Train (2008), issues with dominant peaks in distributions persists, and the importance of starting values is not to be underestimated. Finally, on the empirical side, substantially more effort needs to go into the specification of the class allocation models and the search for appropriate observable and latent drivers of heterogeneity, be it in sensitivities, processing rules or decision rules. It remains up to the analyst to make an informed choice between the two structures, where hybrid approaches combining the benefits of both add an important further level of flexibility.

## Acknowledgements

This chapter is partly based on work conducted during a stay as a visiting research scholar in the Department of Civil & Environmental Engineering at the Massachusetts Institute of Technology. The author is grateful to Moshe Ben-Akiva, Joan Walker and Dinesh Gopinath for inputs into the work on correlation and elasticities, and also wishes to thank John Rose and Charisma Choudhury for helpful comments on an earlier draft.

## References

Abou-Zeid, M., Ben-Akiva, M., 2010. A model of travel happiness and mode switching. In: Hess, S., Daly, A. (Eds.), Choice Modelling: The State-of-the-Art and the State-of-Practice. Emerald Publishing, UK, pp. 289–305.

Andrews, R. L., Ainslie, A., Currim, I. S., 2002. An empirical comparison of logit choice models with discrete versus continuous representations of heterogeneity. Journal of Marketing Research, 479–487.

Ben-Akiva, M., Boccara, B., 1995. Discrete choice models with latent choice sets. International Journal of Research in Marketing 12 (1), 9–24.

Ben-Akiva, M., Lerman, S. R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge, MA.

Ben-Akiva, M., Walker, J., Bernardino, A., Gopinath, D., Morikawa, T., Polydoropoulou, A., 2002. Integration of choice and latent variable models. In: Mahmassani, H. (Ed.), In Perpetual motion: Travel behaviour research opportunities and application challenges. Pergamon, Ch. 13, pp. 431–470.

Bhat, C., 1997. An endogenous segmentation mode choice model with an application to intercity travel. Transportation Science 31 (1), 34–48.

Bhat, C. R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial Logit model. Transportation Research Part B 35 (7), 677–693.

Bhat, C. R., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transportation Research Part B 37 (9), 837–855.

Bhat, C. R., 2011. The maximum approximate composite marginal likelihood (macml) estimation of multinomial probit-based unordered response choice models. Transportation Research Part B: Methodological 45 (7), 923–939.

Bhat, C. R., Sardesai, R., 2006. The impact of stop-making and travel time reliability on commute mode choice. Transportation Research Part B 40 (9), 709–730.

Bhat, C. R., Sidharthan, R., 2011. A simulation evaluation of the maximum approximate composite marginal likelihood (macml) estimator for mixed multinomial probit models. Transportation Research Part B: Methodological 45 (7), 940–953.

Boyd, J., Mellman, J., 1980. The effect of fuel economy standards on the US automotive market: a hedonic demand analysis. Transportation Research Part A 14 (5-6), 367–378.

Bujosa, A., Riera, A., Hicks, R., 2010. Combining discrete and continuous representation of preference heterogeneity: a latent class approach. Environmental & Resource Economics 47 (4), 477–493.

Campbell, D., Lorimer, V., Aravena, C., Hutchinson, W. G., 2010. Attribute processing in environmental choice analysis: implications for willingness to pay. 84th Annual Conference, March 29-31, 2010, Edinburgh, Scotland 91718, Agricultural Economics Society.

Cardell, S., Dunbar, F., 1980. Measuring the societal impacts of automobile downsizing. Transportation Research Part A 14 (5-6), 423–434.

Chorus, C., Arentze, T., Timmermans, H., 2008. A random regret minimization model of travel choice. Transportation Research Part B 42 (1), 1–18.

Collins, A. T., Rose, J. M., Hensher, D. A., 2013. Specification issues in a generalised random parameters attribute nonattendance model. Transportation Research Part B: Methodological 56 (0), 234 – 253.

Daganzo, C. F., 1979. Multinomial Probit: The Theory and its Applications to Demand Forecasting. Academic Press, New York.

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

Dong, X., Koppelman, F. S., 2003. Mass Point Mixed Logit Model: Development and Application. paper presented at the $10^{th}$ International Conference on Travel Behaviour Research, Lucerne.

Fiebig, D. G., Keane, M., Louviere, J. J., Wasi, N., 2010. The generalized multinomial logit: accounting for scale and coefficient heterogeneity. Marketing Science 29 (3), 393–421.

Fosgerau, M., 2006. Investigating the distribution of the value of travel time savings. Transportation Research Part B 40 (8), 688–707.

Fosgerau, M., 2007. Using nonparametrics to specify a model to measure the value of travel time. Transportation Research Part A 41 (9), 842–856.

Fosgerau, M., Bierlaire, M., 2007. A practical test for the choice of mixing distribution in discrete choice models. Transportation Research Part B 41 (7), 784–794.

Garrow, L., 2004. Comparison of models representing correlation and random taste variation: An application to airline passengers' rescheduling behavior. Published doctoral dissertation. Northwestern University, Evanston, Illinois.

Gopinath, D., 1995. Modeling heterogeneity in discrete choice processes: Application to travel demand. Ph.D. thesis, MIT, Cambridge, MA.

Greene, W. H., Hensher, D. A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B 37 (8), 681–698.

Greene, W. H., Hensher, D. A., 2013. Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. Applie 45 (14), 1897–1902.

Greene, W. H., Hensher, D. A., Rose, J. M., 2006. Accounting for heterogeneity in the variance of unobserved effects in mixed logit models. Transportation Research Part B 40 (1), 75–92.

Gupta, S., Chintagunta, P., 1994. On using demographic variables to determine segment membership in logit mixture models. Journal of Marketing Research 31 (1), 128–136.

Hanley, N., Wright, R., Koop, G., 2002. Modelling recreation demand using choice experiments: climbing in scotland. Environmental and Resource Economics 22, 449–466.

Hensher, D. A., 2010. Attribute processing, heuristics and preference construction in choice analysis. In: Hess, S., Daly, A. J. (Eds.), State-of Art and State-of Practice in Choice Modelling: Proceedings from the Inaugural International Choice Modelling Conference. Emerald, Bingley, UK, Ch. 3, pp. 35–70.

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. Empirical Economics 39 (4), 413–426.

Hensher, D. A., Rose, J. M., Greene, W. H., 2012. Inferring attribute non-attendance from stated choice data: implications for willingness to pay estimates and a warning for stated choice experiment design. Transportation 39 (2), 235–245.

Hess, S., Ben-Akiva, M., Gopinath, D., Walker, J., 2009. Advantages of latent class models over continuous mixed logit. paper presented at the $12^{th}$ International Conference on Travel Behaviour Research, Jaipur, India.

Hess, S., Bierlaire, M., Polak, J. W., 2005a. Capturing taste heterogeneity and correlation structure with Mixed GEV models. In: Scarpa, R., Alberini, A. (Eds.), Applications of Simulation Methods in Environmental and Resource Economics. Springer Publisher, Dordrecht, The Netherlands, Ch. 4, pp. 55–76.

Hess, S., Bierlaire, M., Polak, J. W., 2005b. Estimation of value of travel-time savings using mixed logit models. Transportation Research Part A 39 (2-3), 221–236.

Hess, S., Bierlaire, M., Polak, J. W., 2007. A systematic comparison of continuous and discrete mixture models. European Transport 36, 35–61.

Hess, S., Rose, J. M., 2007. A latent class approach to recognising respondents' information processing strategies in SP studies. paper presented at the Oslo Workshop on Valuation Methods in Transport Planning, Oslo.

Hess, S., Rose, J. M., 2009. Allowing for intra-respondent variations in coefficients estimated on repeated choice data. Transportation Research Part B 43 (6), 708–719.

Hess, S., Shires, J., Jopson, A., 2013a. Accommodating underlying pro-environmental attitudes in a rail travel context: Application of a latent variable latent class specification. Transportation Research Part D 25, 42–48.

Hess, S., Stathopoulos, A., 2012. Linking the decision process to underlying attitudes and perceptions: a latent variable latent class construct. paper presented at the $13^{th}$ International Conference on Travel Behaviour Research, Toronto.

Hess, S., Stathopoulos, A., Campbell, D., O'Neill, V., Caussade, S., May 2013b. It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity. Transportation 40 (3), 583–607.

Hess, S., Stathopoulos, A., Daly, A. J., 2012. Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. Transportation 39 (3), 565–591.

Hess, S., Train, K., Polak, J. W., 2006. On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit model for vehicle choice. Transportation Research Part B 40 (2), 147–163.

Hole, A. R., 2011. A discrete choice model with endogenous attribute attendance. Economics Letters 110 (3), 203–205.

Ishaq, R., Bekhor, S., Shiftan, Y., 2013. A flexible model structure approach for discrete choice models. Transportation 40 (3), 60–624.
URL http://link.springer.com/article/10.1007/s11116-012-9431-8

Kamakura, W. A., Russell, G., 1989. A probabilistic choice model for market segmentation and elasticity structure. Journal of Marketing Research 26, 379–390.

Louviere, J. J., Hensher, D. A., Swait, J., 2000. Stated Choice Models: Analysis and Application. Cambridge University Press, Cambridge.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behaviour. 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. Journal of Applied Econometrics 15 (5), 447–470.

Mittelhammer, R., Judge, R., Miller, D., 2000. EconometricFoundations. Cambridge University Press, New York.

Provencher, B., Bishop, R. C., 2004. Does accounting for preference heterogeneity improve the forecasting of a random utility model? a case study. Journal of Environmental Economics and Management 48 (1), 793–810.

Revelt, D., Train, K., 1998. Mixed Logit with repeated choices: households' choices of appliance efficiency level. Review of Economics and Statistics 80 (4), 647–657.

Scarpa, R., Gilbride, T., Campbell, D., Hensher, D. A., 2009. Modelling attribute non-attendance in choice experiments for rural landscape valuation. European Review of Agricultural Economics 36 (2), 151–174.

Scarpa, R., Willis, K., Acutt, M., 2005. Individual-specific welfare measures for public goods: a latent class approach to residential customers of yorkshire water. In: Koundouri, P. (Ed.), Econometrics Informing Natural Resource Management. Edward Elgar Publisher, Cheltenham, pp. 316–337.

Shen, J., 2009. Latent class model or mixed logit model? a comparison by transport mode choice data. Applied Economics 41 (22), 2915–2924.

Swait, J., 1994. A structural equation model of latent segmentation and product choice for cross-sectional revealed preference choice data. Journal of retailing and consumer services 1 (2), 77–89.

Swait, J., Ben-Akiva, M., 1985. Incorporating random constraints in discrete models of choice set generation. Transportation Research Part B 21 (2), 91–102.

Train, K., 2008. Em algorithms for nonparametric estimation of mixing distributions. Journal of Choice Modelling 1 (1), 40–69.

Train, K., 2009. Discrete Choice Methods with Simulation, second edition Edition. Cambridge University Press, Cambridge, MA.

Train, K., Weeks, M., 2005. Discrete choice models in preference space and willingness-to-pay space. In: Scarpa, R., Alberini, A. (Eds.), Application of simulation methods in environmental and resource economics. Springer, Dordrecht, Ch. 1, pp. 1–16.

Tversky, A., 1972. Elimination by aspects: a theory of choice. Psychological Review 79, 281–299.

Walker, J., 2001. Extended discrete choice models: Integrated framework, flexible error structures, and latent variables. Ph.D. thesis, MIT, Cambridge, MA.

Walker, J., Ben-Akiva, M., 2002. Generalized random utility model. Mathematical Social Sciences 43 (3), 303–343.

Walker, J., Li, J., 2006. Latent lifestyle preferences and household location decisions. Journal of Geographical Systems 9 (1), 77–101.

Walker, J. L., Ben-Akiva, M., Bolduc, D., 2007. Identification of parameters in normal error component logit-mixture (neclm) models. Journal of Applied Econometrics 22 (6), 1095–1125.

Wedel, M., Kamakura, W. A., Arora, N., Bemmaor, A., Chiang, J., Elrod, T., Johnson, R., Lenk, P., Neslin, S., Poulsen, C. S., 1999. Discrete and continuous representations of unobserved heterogeneity in choice modeling. Marketing Letters 10 (3), 219–232.