Latent class structures: taste heterogeneity and beyond

Stephane Hess*

Choice Modelling Centre & Institute for Transport Studies, University of Leeds, s.hess@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, in almost all cases, part of this heterogeneity can (and should) 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 collect data on all socio-demographic characteristics that may possibly 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 computers and the availability of easy to use and powerful 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 discrete or continuous distributions.

Discrete mixtures generally rely 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. Continuous mixtures on the other hand rely on the specification of a (multivariate) continuous distribution for the coefficients in a choice model. Both types of mixtures rely on the use of a kernel, which is the model that gives the choice probabilities conditional on knowing the values of the random parameters. In many cases, this kernel will be a logit model, reflected in the widespread use of the term mixed logit for continuous mixtures. However,

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

this is not a theoretical requirement, and applications can similarly make use of mixtures of other structures, such as nested logit, for example, or indeed models not adhering to the principle of random utility maximisation (RUM).

In the last two decades, the continuous specification has come to dominate in many fields. 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); Provencher and Bishop (2004); Scarpa et al. (2005); Shen (2009). Aside from providing further detail relating to the general structure, notably in terms of correlation structure, 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, a number of analysts have sought to combine the advantages of the two structures in models using both discrete and continuous random heterogeneity (Bujosa et al., 2010; Greene and Hensher, 2013). Second, a larger body of research has made use of latent class structures with a view to capturing patterns of heterogeneity going beyond taste coefficients, looking at information processing and heuristics (see references in Gonçalves et al., 2022), heuristics and heterogeneity in decision rules (building on Hess et al., 2012). Latent class models have also been used inside hybrid choice model structures (see e.g. Motoaki and Daziano, 2015). Finally, there have also been further advances in terms of estimation techniques for both types of mixtures and developments relating to the flexibility of mixing distributions. 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 $P_{int}(\beta)$ give the probability of individual n choosing alternative i in choice situation t, conditional on a vector of parameters coefficients β. In a multinomial logit (MNL) model (cf. McFadden, 1974), we have:

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

where J is the total number of alternatives, and where the deterministic utility for alternative i is given by

$$V_{int} = f(\beta, x_{int}, z_n), \tag{2}$$

which is a function of the vector of parameters β[1], the attributes of alternative i as faced by individual n in choice situation t, x_{int}, and the vector of sociodemographic characteristics z_n. With i^*nt referring to the alternative chosen by

[1]This may include alternative specific constants that multiply 0-1 elements in the vector x_{int}.

individual n in choice situation t, the contribution by this individual to the likelihood function (across his/her T_n choices) is simply given by $L_n(\beta) = \prod_{t=1}^{T_n} P_{i^*nt}$, where the aim is to find values of β that maximise this function at the sample level, where maximum likelihood estimation (MLE) is the most commonly used approach.

In the above specification, deterministic heterogeneity can already be accommodated through interaction between the vectors β and z_n. We now look at the treatment of random heterogeneity in three different approaches.

2.1.1 Continuous mixtures

The first applications mixing logit probabilities across an assumed continuous distribution of elements in β 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).

With continuous mixtures, we allow the vector β to follow a random distribution with parameters Ω, such that $\Omega \sim f(\beta \mid \Omega)$. The choice probabilities are then given by:

$$P_{int}(\Omega) = \int_\beta P_{int}(\beta) f(\beta \mid \Omega) \, \mathrm{d}\beta, \tag{3}$$

i.e. an integral of probabilities over the distribution $f(\beta \mid \Omega)$. The probability is now a function of the vector of parameters Ω that govern the continuous distribution of β.

A number of points can be made at this point. First, while P_{int} is often an MNL choice probability, as in Equation 1, leading to a Mixed Logit model, the same overall approach applies also with other kernels. Second, the vector β could include some fixed elements, while the distributional assumptions can also allow for correlation between individual random elements. Finally, there is also scope for still incorporating deterministic heterogeneity through interaction between β and z_n, whether at the level of the means or the dispersion parameters (cf. Greene et al., 2006).

Equation 3 would mean that the taste heterogeneity applies at the level of individual choice situations. 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 studies (e.g Hess and Rose, 2009). Following the work of Revelt and Train (1998), we then write the likelihood of the observed sequence of choices for decision maker n as:

$$L_n(\Omega) = \int_\beta \left[\prod_{t=1}^{T_n} P_{i^*nt}(\beta) \right] f(\beta \mid \Omega) \, \mathrm{d}\beta. \tag{4}$$

The integral in Equation 4 (and Equation 3) does not have a closed form solution and is typically approximated using numerical integration. This requires averaging $\prod_{t=1}^{T_n} P_{i^*nt}(\beta)$ 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 β is large (see Train 2009, chapter 12 for an overview).

Before proceeding, it should be noted that this discussion has centred on using continuous mixtures to accommodate heterogeneity in sensitivities across respondents, often referred to as random parameters logit (if using a logit kernel). 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. Many of these effects can also be captured by choosing a non-logit kernel (cf. Hess et al., 2005a), and it should also be pointed out that random coefficients and error components are not mutually exclusive.

2.1.2 Discrete mixtures

An alternative to the use of continuous distributions for individual elements in β is to allow for a finite number of possible values for each element in β, with an associated probability for each such value. This gives rise to what is variably called a discrete mixture model or a mass point logit model, with discussions in Dong and Koppelman (2003); Gopinath (1995); Hess et al. (2007); Train (2008); Wedel et al. (1999), with more recent developments in Vij and Krueger (2017).

Let us assume that β has K different elements, where we allow for S_k different values for β_k. The value for S_k needs to be specified by the analyst, can vary across different elements of β, and can also be set to 1 for some elements (meaning a non-random parameter). The different values for β_k have different weights, with $\pi_{k,1}$ for example giving the weight for the first value of β_k, i.e. $\beta_{k,1}$. Again working with heterogeneity at the level of an individual (as opposed to choice situation), we would then have that:

$$L_n(\beta, \pi) = \sum_{s_1=1}^{S_1} \sum_{s_2=1}^{S_2} \ldots \sum_{s_K=1}^{S_K} \pi_{1,s_1} \cdot \pi_{2,s_2} \cdot \ldots \cdot \pi_{K,s_K} \prod_{t=1}^{T_n} P_{i^*nt}\left(\beta_{1,s_1}, \beta_{2,s_2}, \ldots, \beta_{K,s_K}\right) \tag{5}$$

i.e. a weighted average across all the possible combinations of values in β, with the weight for each combination being given by a product of the respective weights for the individual elements in β, with π grouping together all individual weights, where $0 \leq \pi_{k,s} \leq 1$, $\forall k, s$ and $\sum_{s_k=1}^{S_k} \pi_{k,s_k} = 1$, $\forall k$.

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 β and modest settings for the number of possible values (S_k) for each β_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 $P_{i^*nt}(\beta)$ in Equation 4 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 $3^5 = 243$ terms in the weighted sum in Equation 5.

Choosing an appropriate value of $S_k \, \forall k$ is down to the analyst, and is a non-trivial task. A key component in this decision 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 often observe multiple elements for β_k collapsing to the same estimate, which is especially likely in the case of strongly peaked population distributions. The latter issue can sometimes be a sign of convergence to a poor local optimum and 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); Vij and Krueger (2017). In terms of the explosion in the number of parameters and the question of improvements in fit justifying such increases, it is common practice to move to model fit criteria which penalise the inclusion of additional parameters more strongly, with typical approaches being 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 developments 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 β in each class. The distinction from a simple discrete mixture as discussed above is that the classes capture the joint distribution of the individual elements in β. Specifically, in a model with S classes, we would have S instances of the vector β, say β_1 to β_S, with a possibility of some of the elements in β staying constant across some (or all) 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 π_{ns}, and where $0 \leq \pi_{ns} \leq 1\ \forall n, s$ and $\sum_{s=1}^{S} \pi_{ns} = 1,\ \forall n$.

Let $P_{int}(\beta_s)$ give the probability of individual n choosing alternative i in choice situation t, conditional on n falling into class s. A latent class analogue of Equation 3 would be given by:

$$P_{int}(\beta, \pi_n) = \sum_{s=1}^{S} \pi_{ns} P_{int}(\beta_s) \tag{6}$$

The likelihood of the observed set of choices for n, working on the assumption of intra-individual homogeneity in sensitivities, is then given by:

$$L_n(\beta, \pi_n) = \sum_{s=1}^{S} \pi_{ns} \left(\prod_{t=1}^{T_n} P_{i^*nt}(\beta_s) \right) \tag{7}$$

with $P_{i^*nt}(\beta_s)$ again often (but not necessarily) given by Equation 1 - this leads to a latent class logit model, but the model can easily be adapted for more general underlying structures such as nested or cross-nested logit. It should be noted that, just as with continuous mixture models, there is the possibility to accommodate heterogeneity across individual choice situations instead of or in addition to heterogeneity across individual decision makers. An example of such a specification is given in Song et al. (2022).

In common with the discrete mixture model, no simulation is required in the estimation of latent class models of the form above. 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 β. 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 a class allocation model is used to link these probabilities to characteristics of the individuals (Gupta and Chintagunta, 1994). This then allows us to probabilistically allocate individuals to different classes depending on their socio-demographic characteristics. Typically, these characteristics would be socio-demographic variables, such as income, gender or age, to name but a few. 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 a requirement), the probability of individual n falling into class s would be given by:

$$\pi_{ns} = \frac{e^{\delta_s + g(\gamma_s, z_n)}}{\sum_{l=1}^{S} e^{\delta_l + g(\gamma_l, z_n)}}, \tag{8}$$

where δ_s is a class-specific constant[2], γ_s is a vector of parameters to be estimated,

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

and $g(\cdot)$ gives the functional form of the *utility* function for the class allocation model. Appropriate normalisation is needed for both δ and γ, typically setting these parameters to zero for one class. While socio-demographic characteristics are the most common type of variables used in class allocation models, other possibilities arise. Most notably, there has been growing interest in linking class allocation to underlying latent psychometric constructs such as attitudes and perception, leading to a hybrid latent variable - latent class model (cf. Motoaki and Daziano, 2015). Just as in the standard hybrid choice model , the model makes use of additional measurement model components, but the latent variables are now no longer used to capture heterogeneity in the sensitivities to individual attributes, but heterogeneity in class allocation.

[CROSS-REFERENCE TO ABOU-ZEID & BEN-AKIVA CHAPTER]

We earlier discussed the issue of the proliferation of parameters in the context of discrete mixtures, and the same applies in latent class models. Similarly, estimation with larger numbers of classes can be problematic with parameters collapsing to the same values across classes, having the wrong sign, or some classes obtaining very small probabilities. Again, if this is a result of convergence to a poor local optimum, the EM algorithm can be one possible solution, as discussed in Train (2008) and earlier on by Bhat (1997).

2.2 Contrasts

This section provides some theoretical contrasts between continuous and discrete mixtures. 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); Greene and Hensher (2003); Hanley et al. (2002); Provencher and Bishop (2004); Scarpa et al. (2005); 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 Distributional assumptions

The main emphasis in discussing continuous and discrete mixtures is on their ability to capture random heterogeneity across individuals (and/or across choice situations). The two structures do this in very different ways, as already outlined in Section 2.1.

Deterministic and random heterogeneity

In the basic specification of continuous mixture models, the heterogeneity is entirely random. While such a specification is unfortunately all too common in empirical work, it is clearly possible (and indeed desirable) to explain at least part of the heterogeneity by linking it to observed characteristics of the decision maker and choice setting. This can be done by incorporating interactions with these characteristics in the specification of the utility function (Equation 2) and/or making the parameters of the random distributions a function of such characteristics (cf. Greene et al., 2006). The former approach is more widely used.

In latent class models, analysts can similarly incorporate interactions with observed characteristics directly in the specification of the utility functions (Equation 2), with their impact either being generic or varying across classes. More commonly, deterministic heterogeneity is incorporated through the parameterisation of the class allocation model, i.e. Equation 8, meaning that the class allocation probabilities (and hence the implied sensitivities) also vary as a function of these individual characteristics. This relates to the idea of the model allowing for classes of decision makers with specific sets of preferences. It should be noted that the two approaches, i.e. including covariates directly in the utility functions or in the class allocation model, are not mutually exclusive (subject to identification) - this opens up the possibility for an analyst to test the impact of some covariates, such as income, on the sensitivity to individual attributes, such as cost, while using other covariates to explain class allocation.

Shape of the distribution

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, are key issues facing analysts using continuous mixture 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; Fosgerau and Mabit, 2013) 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 applications using continuous mixtures is a strong shape assumption and general uni-modality. In theory, the same does not apply with latent class structures as (typically) 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.

With both models, the ability to retrieve the *true* patterns of heterogeneity in the data thus depends both on the unobserved shape of that heterogeneity and the specification used by the analyst.

Multivariate distributions

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 (see e.g. Train, 2009, chapter 9). While most estimation packages allow users to specify multivariate distributions, the vast majority of continuous mixture applications make use of independently distributed taste coefficients, despite the obvious simplification and likely lack in performance this engenders (cf. Hess and Train, 2017). More importantly, by using univariate distributions, an analyst is at risk of biased estimates of the amount of heterogeneity in individual coefficients, if in reality the coefficients are correlated.

While, with continuous mixture models, analysts need to specify the correlation explicitly, latent class models are a natural way of dealing with multivariate heterogeneity. Indeed, 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. Let us imagine a model estimated on a simple time-money trade-off, and using two classes. If the time sensitivity (say β_T) is higher in class 1 than class 2, but the reverse happens for the cost sensitivity (say β_C) , then the model captures negative correlation between the time and cost sensitivities. This can be easily formalised by noting that:

$$\begin{aligned} cov\left(\beta_{nT}, \beta_{nC}\right) &= E\left[\left(\beta_{nT} - E\left(\beta_{nT}\right)\right)\left(\beta_{nC} - E\left(\beta_{nC}\right)\right)\right] \\ &= E\left(\beta_{nT}\,\beta_{nC}\right) - E\left(\beta_{nT}\right)E\left(\beta_{nC}\right) \\ &= \sum_{s=1}^{S} \pi_{ns}\beta_{T,s}\beta_{C,s} - \left(\sum_{s=1}^{S} \pi_{ns}\beta_{T,s}\right)\left(\sum_{s=1}^{S} \pi_{ns}\beta_{C,s}\right) \end{aligned} \tag{9}$$

This discussion has shown that while the use of correlated distributions is possible and even advisable with continuous mixture models, latent class structures do so without any additional analyst input. The discussion around correlation also relates to the ongoing confusion in the literature about scale heterogeneity. Much effort has gone into attempts to disentangle scale heterogeneity from other heterogeneity, leading to the development of continuous specifications such as GMNL (Fiebig et al., 2010) or discrete ones such as the scale adjusted latent class model Magidson and Vermunt (2005). As shown by Hess and Train (2017), these specifications are in fact restricted versions of mixed logit and latent class models, and analysts should simply allow for correlation without making futile attempts to separate scale heterogeneity from other heterogeneity.

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 making use of the sample level model estimates and an individual's observed choices to infer their most likely position on the population distribution.

In a continuous mixed logit context, these calculations are straightforward, as discussed for example by Train (2009, chapter 12). Specifically, we have from Equation 4 that the likelihood of the observed sequence of choices for person n is given by:

$$L_n(\Omega) = \int_{\beta} L_n(\beta) f(\beta \mid \Omega) \, \mathrm{d}\beta. \tag{10}$$

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

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

$$L(\beta_n \mid C_n) = \frac{L_n(\beta) f(\beta \mid \Omega)}{L_n(\Omega)} \tag{11}$$

This gives us the probability of given values for β_n, conditional on the observed choices (C_n) for individual n. It is then straightforward to for example calculate a conditional mean for β_n as:

$$\bar{\beta}_n = \int_{\beta_n} \beta_n L(\beta_n \mid C_n) \, \mathrm{d}\beta_n, \tag{12}$$

with similar calculations to obtain the corresponding variance or other measures. This highlights an important confusion in the literature. Posterior analysis does not produce estimates at the person level, but allows us to produce moments for the posterior distribution, clearly highlighting the uncertainty in these values, while the impact of the sample level distribution assumptions is clear from Equation 12.

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:

$$\widehat{\pi}_{ns} = \frac{\pi_{ns} L_n(\beta_s)}{L_n(\beta, \pi_n)}, \tag{13}$$

where $L_n(\beta_s)$ gives the likelihood of the observed choices for individual n, conditional on class s, i.e. $L_n(\beta_s) = \prod_{t=1}^{T_n} P_{i^*nt}(\beta_s)$.

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, such as the value of travel time (VTT), say $VTT_s = \frac{\beta_{T,s}}{\beta_{C,s}}$, i.e. the ratio between the time and cost coefficients. Using $\overline{VTT}_n = \sum_{s=1}^{S} \pi_{ns} VTT_s$ simply gives us a sample level mean for VTT 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 VTT for individual n, conditional on his/her observed choices, can now be calculated as $\widehat{VTT}_n = \sum_{s=1}^{S} \widehat{\pi}_{ns} VTT_s$.

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 γ_c, and using a linear parameterisation in Equation 8. We can then calculate the likely value for z_c for an individual in class s as:

$$\widehat{z}_{c,s} = \frac{\sum_{n=1}^{N} \widehat{\pi}_{ns} z_{nc}}{\sum_{n=1}^{N} \widehat{\pi}_{ns}}, \tag{14}$$

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 κ for z_c by using:

$$\widehat{P}\left(z_{c,s} = \kappa\right) = \frac{\sum_{n=1}^{N} \widehat{\pi}_{ns}\left(z_{nc} = \kappa\right)}{\sum_{n=1}^{N} \widehat{\pi}_{ns}}. \tag{15}$$

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 β 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 Equation 16 , we can then easily see that:

$$\widehat{cov}\left(\beta_{nT}, \beta_{nC}\right) = \sum_{s=1}^{S} \widehat{\pi}_{ns} \beta_{T,s} \beta_{C,s} - \left(\sum_{s=1}^{S} \widehat{\pi}_{ns} \beta_{T,s}\right)\left(\sum_{s=1}^{S} \widehat{\pi}_{ns} \beta_{C,s}\right) \tag{16}$$

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

$$\begin{aligned}
\widehat{cov}\left(\beta_{nT}, \beta_{nC}\right) &= \widehat{\pi}_{n,1}\widehat{\pi}_{n,2}\left[\beta_{T,1}\left(\beta_{C,1} - \beta_{C,2}\right) + \beta_{T,2}\left(\beta_{C,2} - \beta_{C,1}\right)\right] \\
&= \widehat{\pi}_{n,1}\widehat{\pi}_{n,2}\left[\left(\beta_{T,1} - \beta_{T,2}\right)\left(\beta_{C,1} - \beta_{C,2}\right)\right],
\end{aligned} \tag{17}$$

where the sign of $\widehat{cov}\left(\beta_{nT},\beta_{nC}\right)$ only depends on the changes in the two elements in β_T and β_C across the two classes.

It should be noted that, using Equation 11, 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 observed choices, 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.

Two final points need to be made in relation to posteriors. First, it should be noted again that posteriors are not point estimates, but distributions with uncertainty, and it is thus incorrect to refer to these as individual-level parameters, or to think that they can be used to deterministically cluster individuals. These types of uses would ignore the uncertainty in pinpointing the location of individuals on the sample level distributions. Second, the degree of uncertainty, and hence the usefulness of posteriors, depends on the number of observations that are available for each individual - with more choices, there is a greater ability to pinpoint the location of individual people. With limited numbers of choices however, or indeed with cross-sectional data, the usefulness of posteriors is substantially lower.

2.2.3 Substitution patterns

It is well known that the MNL model exhibits the independence from irrelevant alternatives (IIA) assumption. This arises as the denominator in Equation 1 is the same for all alternatives, meaning that the ratio of any two probabilities depends only on the utilities (and hence the attributes) of those two alternatives, i.e. $\frac{P_{int}}{P_{jnt}}=\frac{e^{V_{int}}}{e^{V_{jnt}}}$. The impact of the IIA assumption is that the disaggregate cross-elasticities are equal across alternatives, implying proportional substitution effects. Note that this does not imply IIA in the aggregate elasticities (Louviere et al., 2000).

There has been extensive focus in the choice modelling literature on breaking free from the IIA assumption by allowing for greater substitution between some alternatives. This is achieved by placing a structure on the distribution of error terms, in the form of nested (Daly and Zachary, 1978; McFadden, 1978; Williams, 1977) and cross-nested (e.g. Vovsha, 1997) logit models, or through the use of error components in a mixed logit structure (Walker, 2001).

Analysts often make the statement that latent class structures and mixed logit models are not affected by the IIA assumption. This was illustrated for example in the computation of disaggregate elasticities for latent class by Bhat (1997). It can be easily seen from Equations 3 and 6 that the presence of the integral, respectively weighted average, means that the ratio $\frac{P_{int}}{P_{jnt}}$ is no longer a function of only alternatives i and j. This thus directly means that the IIA assumption no longer applies. However, it should be noted that in most applications of mixed logit or latent class, this is not a result of a structural approach to capturing correlation between alternatives, as the kernel of the model remains MNL, and thus itself exhibits the IIA assumption. Of course, incorporating random heterogeneity in a model can introduce correlation in utility across alternatives, thinking for example of the situation where the presence of a random cost coefficient introduces higher correlation between high-cost alternatives. But if an analyst wishes to retain control to prespecify a given correlation structure, then this should be by using a non-logit kernel (such as in a mixed nested logit model, or a latent class nested logit model), or including error components in a mixed logit model.

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:

$$L_n\left(\beta,\pi,\sigma\right)=\sum_{s=1}^{S}\pi_{ns}\int_{\eta}\prod_{t=1}^{T_n}P_{i^*nt}\left(\beta_s,\eta\right)f\left(\eta\mid\sigma\right)\mathrm{d}\eta \tag{18}$$

In this specification, the continuous random components η follow Normal distributions with a mean of zero and with standard deviations given by the vector σ. 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:

$$L_n\left(\Omega,\pi\right)=\sum_{s=1}^{S}\pi_{ns}\int_{\beta_s}\prod_{t=1}^{T_n}P_{i^*nt}\left(\beta_s\right)f\left(\beta_s\mid\Omega_s\right)\mathrm{d}\beta_s \tag{19}$$

In this model, we have that the vector of coefficients β_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 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:

$$L_n\left(\Omega_\gamma,\Omega_\alpha\right)=\int_\alpha \prod_{t=1}^{T_n}\left[\int_\gamma P_{i^*nt}\left(\beta=\alpha+\gamma\right) f\left(\gamma \mid \Omega_\gamma\right) \mathrm{d}\gamma\right] h\left(\alpha \mid \Omega_\alpha\right) \mathrm{d}\alpha, \quad (20)$$

where $\beta=\alpha+\gamma$ with α distributed across decision makers and γ 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 (cf. Hess and Train, 2011), and this raises the question whether replacing one layer with weighted summation through a latent class structure would be beneficial, in essence adapting Equation 19 by moving the position of the integral to the level of an individual choice:

$$L_n\left(\Omega,\pi\right)=\sum_{s=1}^{S}\pi_{ns}\prod_{t=1}^{T_n}\int_{\beta_s} P_{i^*nt}\left(\beta_s\right) f\left(\beta_s \mid \Omega_s\right) \mathrm{d}\beta_s. \quad (21)$$

This specification would now mean that the latent class structure captures the variation in sensitivities across individual decision makers, 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 $\alpha_n = h\left(\theta, z_n\right)+\eta_n$ where η_n is a vector of standard normal random variables. These α_n terms, which can for example represent underlying attitudes and perceptions, are then used in parameterising the class allocation probabilities, rewriting Equation 22 to:

$$\pi_{ns}=\frac{e^{\delta_s+g(\gamma_s,z_n)+\tau_s\alpha_n}}{\sum_{l=1}^{S}, e^{\delta_l+g(\gamma_l,z_n)+\tau_l\alpha_n}}. \quad (22)$$

At the same time, α_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 ν 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:

$$L_n\left(\beta,\gamma,\theta,\delta,\tau\right)=\int_{\eta_n}\sum_{s=1}^{S}\pi_{ns}\left(\prod_{t=1}^{T_n}P_{i^*nt}\left(\beta_s\right)\right) P\left(I_n \mid \alpha_n\right)\phi\left(\eta_n\right)\mathrm{d}\eta_n \quad (23)$$

where π_{ns} is now also a function of α_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. We finally look at model averaging.

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). 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. Campbell et al., 2010; Hensher and Greene, 2010; Hensher et al., 2012; Hole, 2011; Scarpa et al., 2009). 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 β, we now have a SxK matrix Λ, 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:

$$L_n(\beta, \pi) = \sum_{s=1}^{S} \pi_s \prod_{t=1}^{T_n} P_{i^*nt}(\beta_s = \beta \circ \Lambda_s). \tag{24}$$

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:

$$L_n(\Omega, \pi) = \sum_{s=1}^{S} \pi_s \int_{\beta} \prod_{t=1}^{T_n} P_{i^*nt}(\beta_s = \beta \circ \Lambda_s) f(\beta \mid \Omega)\, \mathrm{d}\beta. \tag{25}$$

Empirical evidence by Hess et al. (2013b) on multiple datasets reveals major improvements in fit by the specification in Equation 25 over the model in Equation 24, 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 β_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:

$$L_n(\beta, \pi) = \sum_{m=1}^{M} \pi_{nm} L_n(\beta_m, m), \tag{26}$$

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

$$L_n(\Omega, \pi) = \sum_{m=1}^{M} \pi_{nm} \int_{\beta_m} L_n(\beta_m, m) f(\beta_m, \Omega_m) \, \mathrm{d}\beta_m, \tag{27}$$

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. Ben-Akiva and Boccara 1995; Swait and Ben-Akiva 1985 and Gopinath 1995, section 2.7).

4.3 Model averaging

The discussion thus far on confirmatory models has focussed on the case where an analyst imposes specific behavioural assumptions in different classes and then simultaneously estimates the parameters in those classes along with the class allocation probabilities.

Model averaging on the other hand uses a more sequential approach. A number of different models are estimated separately, with say $L_{n,m}(\beta_m)$ giving the likelihood of the choices for person n, conditional on using model m, with a set of parameters β_m, for which the estimates are given by $\hat{\beta}_m$. In model averaging, we then take these individual models as inputs into a latent class structure where only the class allocation probabilities are estimated, i.e.

$$L_n\left(\hat{\beta}, \pi_n\right) = \sum_{m=1}^{M} \pi_{ns}\, L_{n,m}\left(\hat{\beta}_m\right), \tag{28}$$

where the use of $\hat{\beta}_m$ instead of β_m reflects the fact that the parameters for individual models are not re-estimated. A recent applications of model averaging of this type is given in Hancock et al. (2020).

5 Summary and conclusions

This chapter has revisited the topic of contrasting continuous mixed logit models and latent class structures. 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 and correlationare 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 two decades. On the continuous mixed logit side, progress has been made in estimation capabilities, 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 Bujosa et al. (2010); Greene and Hensher (2013); Hess et al. (2013b); Walker and Li (2006). 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. Hess et al., 2013a; Walker and Ben-Akiva, 2002).

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), or the innovative work of Vij and Krueger (2017) 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 Charisma Choudhury, Andrew Daly and John Rose 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, 9–24.

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. chapter 13, pp. 431–470.

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

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

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

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

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, 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, 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, 423–434.

Chorus, C., Arentze, T., Timmermans, H., 2008. A random regret minimization model of travel choice. Transportation Research Part B 42, 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, 234 – 253.

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

Daly, A., Zachary, S., 1978. Improved multiple choice models, in: Hensher, D.A., Dalvi, Q. (Eds.), Identifying and Measuring the Determinants of Mode Choice. Teakfields, London, pp. 335–357.

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, 393–421.

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

Fosgerau, M., 2007. Using nonparametrics to specify a model to measure the value of travel time. Transportation Research Part A 41, 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, 784–794.

Fosgerau, M., Mabit, S.L., 2013. Easy and flexible mixture distributions. Economics Letters 120, 206 – 210.

Gonçalves, T., Lourenço-Gomes, L., Pinto, L.M.C., 2022. The role of attribute non-attendance on consumer decision-making: Theoretical insights and empirical evidence. Economic Analysis and Policy 76, 788–805.

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, 681–698.

Greene, W.H., Hensher, D.A., 2013. Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. Applied Economics 45, 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, 75–92.

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

Hancock, T.O., Hess, S., Daly, A., Fox, J., 2020. Using a sequential latent class approach for model averaging: Benefits in forecasting and behavioural insights. Transportation Research Part A 139, 429–454. doi:10.1016/j.tra.2020.07.005.

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. chapter 3, pp. 35–70.

Hensher, D.A., Greene, W.H., 2003. The Mixed Logit Model: The State of Practice. Transportation 30, 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, 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, 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. chapter 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, 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, 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., 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, 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, 565–591.

Hess, S., Train, K., 2011. Recovery of inter- and intra-personal heterogeneity using mixed logit models. Transportation Research Part B 45, 973–990.

Hess, S., Train, K., 2017. Correlation and scale in mixed logit models. Journal of Choice Modelling 23, 1–8.

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, 147–163.

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

Ishaq, R., Bekhor, S., Shiftan, Y., 2013. A flexible model structure approach for discrete choice models. Transportation 40, 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.

Magidson, J., Vermunt, J., 2005. "Technical Guide to Latent Gold Sofrware 4.5". Statistical Innovation.

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., 1978. Modelling the choice of residential location, in: Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J.W. (Eds.), Spatial Interaction Theory and Planning Models. North Holland, Amsterdam. chapter 25, pp. 75–96.

McFadden, D., Train, K., 2000. Mixed MNL Models for discrete response. Journal of Applied Econometrics 15, 447–470.

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

Motoaki, Y., Daziano, R.A., 2015. A hybrid-choice latent-class model for the analysis of the effects of weather on cycling demand. Transportation Research Part A: Policy and Practice 75, 217–230. URL: https://www.sciencedirect.com/science/article/pii/S0965856415000592, doi:https://doi.org/10.1016/j.tra.2015.03.017.

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, 793–810.

Revelt, D., Train, K., 1998. Mixed Logit with repeated choices: households' choices of appliance efficiency level. Review of Economics and Statistics 80, 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, 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, 2915–2924.

Song, F., Hess, S., Dekker, T., 2022. Uncovering the link between intra-individual heterogeneity and variety seeking: the case of new shared mobility. Transportation doi:10.1007/s11116-022-10334-4.

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, 77–89.

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

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

Train, K., 2009. Discrete Choice Methods with Simulation. second edition ed., 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. chapter 1, pp. 1–16.

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

Vij, A., Krueger, R., 2017. Random taste heterogeneity in discrete choice models: Flexible nonparametric finite mixture distributions. Transportation Research Part B: Methodological 106, 76–101. URL: https://ui.adsabs.harvard.edu/abs/2018arXiv180202299V, arXiv:1802.02299.

Vovsha, P., 1997. Application of a Cross-Nested Logit model to mode choice in Tel Aviv, Israel, Metropolitan Area. Transportation Research Record 1607, 6–15.

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, 303–343.

Walker, J., Li, J., 2006. Latent lifestyle preferences and household location decisions. Journal of Geographical Systems 9, 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, 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, 219–232.

Williams, H.C.W.L., 1977. On the Formulation of Travel Demand Models and Economic Evaluation Measures of User Benefit. Environment & Planning A 9, 285–344.