Discrete Mixtures Models

Stephane Hess, ETHZ
Michel Bierlaire, EPFL
John W. Polak, Imperial College

Discrete Mixtures Models

Stephane Hess
Institute for Transport Planning and Systems
ETHZ
Zürich

stephane.hess@ivt.baug.ethz.ch
February 2006

Michel Bierlaire
Inst. of Mathematics

EPFL
Lausanne

michel.bierlaire@epfl.ch

John W. Polak
Centre for Transport Studies
Imperial College
London

j.polak@imperial.ac.uk

Abstract

Allowing for variations in behaviour across respondents is one of the most fundamental principles in discrete choice modelling, given that the assumption of a purely homogeneous population cannot in general be seen to be valid. Two approaches have classically been used to address this problem; the use of deterministic segmentations of the population, and the use of a random continuous representation of variations in tastes across respondents. In this paper, a revised version of [10], we discuss an alternative approach, based on the use of discrete mixtures of underlying choice models over a finite set of distinct support points. The applied part of this paper shows how the resulting model structure can be used to test the validity of hypotheses such as the presence of individuals with zero valuations of travel-time changes.

Keywords

Discrete choice - Mixtures – STRC 2006 – Monte Verita

1 Introduction and context

Allowing for variations in behaviour across respondents is one of the most fundamental principles in discrete choice modelling, given that the assumption of a purely homogeneous population cannot in general be seen to be valid. The most basic approach for representing such variations is through a segmentation of the population into mutually exclusive subsets, either in the form of separate models for different population segments, or separate coefficients within the same model for different population segments. These approaches can for example be used to differentiate between different journey purposes, or different income classes. In the case of continuous attributes, such as income, such segmentations can however be seen to be very arbitrary, and it is in this case preferable (though computationally more expensive) to use a continuous variation in tastes as a function of the concerned attribute.

Deterministic variations in tastes, such as those described above, can be accommodated within the standard random utility framework, and are applicable for all known model structures. However, although the use of such deterministic variations is appealing from the point of view of interpretation (and especially for forecasting), it is often not possible to represent all variations in tastes in a deterministic fashion, for reasons of data quality, but also due to inherent randomness in choice behaviour. For this reason, random coefficient models, such as the Mixed Multinomial Logit (MMNL), which allows for random variations in behaviour across respondents, have an important advantage in terms of flexibility. In general, such models have the disadvantage that their choice probabilities take on the form of integrals that do not possess a closed-form solution, such that numerical processes, typically simulation, are required during estimation and application of the models. This greatly limited the use of these structures for many years after their initial developments. Over recent years, gains in computer speed and the efficiency of simulation-based estimation processes [c.f. 12] have however led to increased interest in the MMNL model in particular, by researchers and, to a lesser degree also practitioners.

Despite the improvements in estimation capability, the cost of using the MMNL model remains high. While this might be acceptable in many cases, another important issue remains, namely the choice of distribution to be used for representing the random variations in tastes across respondents. This issue can be divided into several sub-issues.

Firstly, it is important to reconcile the choice of distribution with theoretical or intuitive preconceptions regarding what constitute reasonable or plausible patterns of variation in parameter values across a population. As such, a strictly positive distribution would not be used for a coefficient where positive as well as negative values are expected in the population. On the other hand, in the case of a coefficient with a strong sign assumption (such as a negative cost coefficient), the use of strictly bounded distributions can lead to an inability to uncover problems with the data or utility specification that would manifest themselves as counter-intuitively signed coefficients for part of the population [c.f. 11].

Secondly, even with the use of the most flexible distributions available, it seems almost inevitable that there will be some discrepancy between the true and postulated distribution; cases

will arise in which real-world behaviour cannot be characterised adequately by one of a set of standard statistical distributions. One case in point arises in the modelling of tastes which may theoretically have a significant mass at zero but be exclusively positively or negatively signed elsewhere [c.f. 4]. The situation becomes even more complicated in the case of an attribute which some individuals value positively and some individuals value negatively, with a remaining part of the population being indifferent to the attribute. This applies for example in the case of attributes describing discrete qualitative features of an alternative, such as a distinction between forward and backward facing seats for rail-travel. Representing this situation is not possible with the use of standard continuous distributions, where the notion of a mass at a specific point (especially if not at the extremes of the domain) does not apply, such that the results obtained with such distributions may lead to unwarranted conclusions. Another example of such a parameter that can take on positive, negative and zero valuations is an Arrow-Pratt absolute risk aversion parameter.

Given these problems, it is of interest to explore alternative ways of representing random variations in tastes across respondents, avoiding some of the issues discussed above.

One possible solution is to use Kernel densities of individual-specific coefficient values in the search for an appropriate distribution. The most basic approach consists of estimating individual-specific MNL models, which is only possible in the presence of multiple observations per individual, and to infer information about the true distribution from plotting the Kernel density of the hence obtained coefficient values. This causes significant problems in practice, given the potential lack of information in the resulting small datasets. In this context, Hensher and Greene [9] advocate the use of a jackknife-style procedure that starts with the full sample, and proceeds by eliminating individuals one-by-one, each time estimating a new model. The resulting set of estimates can then be used to produce a Kernel density function. In practice, the applicability of such methods is often limited by high computational cost and data requirements.

A second approach is to use empirical distributions, based on estimating a set of support points with corresponding masses, with linear segments between support points. The success of this approach however not only depends crucially on the number of support-points used, but important issues of implementation need to be faced in the estimation of the support points, where problems arise because of the non-differentiability of the likelihood function.

A final approach comes in the use of non-parametric approaches, which are free of a priori assumptions about the shape of the true distribution. The application of such approaches to the estimation of the value of travel-time savings (VTTS) is described by Fosgerau [6]. The results show that the non-parametric approaches outperform a set of parametric approaches, but the fact that such approaches are very data-hungry leads to problems in recuperating the distribution in the tails of the population, a situation that Fosgerau addresses through the use of a semi-parametric approach, where part of the distribution is accounted for through a set of covariates. While very promising, non-parametric and (to a lesser extent) semi-parametric regression approaches can be difficult to apply in practice, and more work is required to allow widespread application.

The three approaches described above are in principle able to deal with the main issue described

by Hess, Bierlaire and Polak [11], namely the behaviour in the tails of the distribution. Similarly, they do, unlike most standard continuous distributions, have the ability to allow for a multi-modal distribution of a specific taste coefficient. However, it seems that neither of the three approaches can deal adequately with the presence of a heightened mass at a given point, such as a zero VTTS. While the use of Kernel densities can signal the presence of such mass-points, the issue of how to incorporate them in the final model remains.

In this paper, a revised version of [10], we explore an alternative approach, based on the idea of replacing the continuous distribution functions by discrete distributions, spreading the mass among several discrete values. Theoretically, such discrete mixtures allow modellers to deal with each of three issues described above (tail-behaviour, multiple modes, inflated mass), although certain issues, notably in estimation, need to be addressed, as described in Section 2.

Mathematically, the model structure of a discrete mixture model is a special case of a latent-class model [c.f. 13, 3], assigning different coefficient values to different parts of the population of respondents, a concept discussed in the field of transport studies for example by Greene and Hensher [8] and Lee et al. [15]. The work of Gopinath [7] especially is of interest in the context of the case-study described in this paper, as it makes use of a latent-class model in the analysis of variations in the VTTS across respondents, showing the presence of multiple subgroups in the population. Latent-class approaches make use of two sub-models, one for class-allocation, and one for within-class choice. The former models the probability of an individual being assigned to a specific class as a function of attributes of the respondent and possibly of the alternatives in the choice-set. The within-class model is then used to compute the class-specific choice-probabilities for the different alternatives, conditional on the tastes within that class. The actual choice probability for individual n and alternative i is given by a sum of the class-specific choice probabilities, weighted by the class-allocation choice probabilities for that specific individual.

The latent-class approach is appealing from the point of view that it allows for differences in sensitivities across population groups, where the group-allocation can be related to socio-demographic characteristics. However, in practice, it may not always be possible to explain group-allocation with the help of a probabilistic model relating the outcome to observed variables. This situation is similar to the case where taste heterogeneity cannot be explained deterministically, leading to a requirement for using random coefficients models. As such, in this paper, we explore the use of models in which the class-allocation probabilities are independent of explanatory variables, and are simply given by constants that are to be estimated during model calibration. As such, the resulting model exploits the class-membership concept in the context of random coefficients models, with a limited set of possible values for the coefficients. In theory, existing discrete distributions (e.g. Poisson) could be used; however, this comes at the cost of flexibility and again leads to the problem of reconciling the theoretical and empirical/practical characteristics of the mixing distribution. This problem does not exist in the case where a fixed set of coefficient values are used that each have an associated probability, but where the values and associated probabilities are free from any a priori constraints.

Thus far, there have seemingly been only two applications of this approach in the area of transport research, by Gopinath [7], in the context of mode-choice for freight shippers, and by Dong

and Koppelman [5], who made use of discrete mixtures of MNL models in the analysis of mode-choice for work trips in New-York, referring to the resulting model as the "Mass Point Mixed Logit model". Although the properties of discrete mixture models have been discussed by several other authors [e.g. 16], the model structure does not seem to have received widespread exposure or application, despite its many appealing characteristics.

Given the above discussion, part of the aim of this paper is to re-explore the potential advantages of discrete mixture models, with the hope of encouraging their more widespread use. However, the main aim, and contribution of this paper, is to demonstrate how the model structure can be exploited to allow for a part of the population in which people are indifferent to changes in a specific attribute, a treatment that is not generally possible with the use of continuous mixture structures. Although the discussion in this paper looks specifically at the case of zero valuations of changes in travel-time (leading to zero VTTS), the same principle obviously applies in the case of other attributes. Finally, the analysis also aims to investigate the potential bias in coefficient estimates that can result from not allowing for the presence of individuals with such zero valuations.

The remainder of this paper is organised as follows. The next section sets out the theory behind discrete mixture models. Section 3 describes a set of tests of the validity of the model structure conducted with the help of simulated data, while Section 4 presents the main case-study testing for the presence of respondents with zero VTTS. Finally, Section 5 summarises the contents of the paper and presents the conclusions of the study.

2   Methodology

We will begin by introducing some general notation, which is used throughout the remainder of this paper. Specifically, let $x_{i,n}$ be a vector defining the attributes of alternative i as faced by respondent n (potentially including interactions with socio-demographic variables), and let β be a vector defining the tastes of the decision-maker, where, in purely deterministic models, β is constant across respondents. Let x_n be a vector grouping together the individual vectors $x_{j,n}$ across the alternatives contained in the choice-set of respondent n, and let γ represent an additional set of parameters, which can for example contain the structural parameters (and possibly allocation parameters) used to represent inter-alternative correlation in a Generalised Extreme Value (GEV) context. In a very general form, we can then define $P_n \left(i \mid x_n, C_n, \gamma, \beta \right)$ to give the choice probability of alternative i for individual n, with a choice-set C_n, conditional on the observed vector x_n, and for given values for the vectors of parameters β and γ (to be estimated). Due to the potential inclusion of socio-demographic attributes in x_n, this notation allows for deterministic variations in tastes across respondents.

This notation can now be used as the building block for models allowing for a distribution of tastes across respondents. In a continuous mixture model, the choice probabilities are then

given by:

$$P_n\left(i \mid x_n, C_n, \gamma, \Omega\right) = \int_\beta \left[P_n\left(i \mid x_n, C_n, \gamma, \beta\right) f\left(\beta \mid \Omega\right)\right] \mathrm{d}\beta, \tag{1}$$

where the vector β is distributed according to $f\left(\beta \mid \Omega\right)$, with vector of parameters Ω. With $P_n\left(i \mid x_n, C_n, \gamma, \beta\right)$ giving MNL choice probabilities, equation (1) represents the choice probabilities in a MMNL model; however, any other GEV-type choice probability can be used for $P_n\left(i \mid x_n, C_n, \gamma, \beta\right)$, with an explicit role for the vector γ, leading to a more general GEV mixture model.

From a statistical point of view (in the context of mixture densities), the MMNL model is a continuous mixture of MNL models over the distribution of β. In this context, it is clear that discrete mixtures are also possible, a notion that we exploit in this paper by limiting the number of possible values for β. As such, we now divide the set of parameters β into two sets; $\bar{\beta}$ represents a part of β containing deterministic parameters, while $\widehat{\beta}$ is a set of K random parameters that have a discrete distribution. Within this set, the parameter $\widehat{\beta}_k$ has m_k mass points $\widehat{\beta}_k^j$, $j = 1, \ldots, m_k$, each of them associated with a probability π_k^j, where we impose the conditions that

$$0 \leq \pi_k^j \leq 1, \quad k = 1, \ldots, K; \; j = 1, \ldots, m_k, \tag{2}$$

and

$$\sum_{j=1}^{m_k} \pi_k^j = 1, \quad k = 1, \ldots, K. \tag{3}$$

For each realisation $\widehat{\beta}_1^{j_1}, \ldots, \widehat{\beta}_K^{j_K}$ of $\widehat{\beta}$, the choice probability is given by

$$P_n\left(i \mid x_n, C_n, \gamma, \beta = \langle \bar{\beta}, \widehat{\beta}_1^{j_1}, \ldots, \widehat{\beta}_K^{j_K} \rangle\right), \tag{4}$$

where the deterministic part of $\bar{\beta}$ stays constant across realisations of the vector $\widehat{\beta}$.

The unconditional (on a specific realisation of β, not on the distribution of $\widehat{\beta}$) choice probability for alternative i and decision-maker n can now be written straightforwardly as a mixture over the discrete distributions of the various elements contained in $\widehat{\beta}$ as:

$$P_n\left(i \mid x_n, C_n, \gamma, \bar{\beta}, \widehat{\beta}, \pi\right)$$
$$= \sum_{j_1=1}^{m_1} \cdots \sum_{j_K=1}^{m_K} P_n\left(i \mid x_n, C_n, \gamma, \beta = \langle \bar{\beta}, \widehat{\beta}_1^{j_1}, \ldots, \widehat{\beta}_K^{j_K} \rangle\right) \pi_1^{j_1} \cdot \ldots \cdot \pi_K^{j_K}, \tag{5}$$

where $\bar{\beta}$, $\widehat{\beta}$ and π ($\pi = \langle \pi_1^1, \ldots, \pi_1^{m_1}, \ldots, \pi_K^1, \ldots, \pi_K^{m_K} \rangle$) are vectors of parameters to be estimated in a regular maximum likelihood estimation procedure. An obvious advantage of this approach is that, if the model (4) used inside the mixture has a closed form, then so does the discrete mixture itself.

In this paper, we mainly focus on the simple case where the underlying choice model is of MNL form; however, the form given in equation (5) is appropriate for any underlying model, where, with an underlying GEV structure, the resulting model obtains a closed-form expression, avoiding the need for simulation in estimation and application. The approach can easily be extended to the case of combined discrete and continuous random taste variation, by partitioning β into three parts; the above defined parts $\bar{\beta}$ and $\widehat{\beta}$, and an additional part $\widetilde{\beta}$, whose elements follow continuous distributions. This however leads to a requirement to use simulation, as with all continuous mixture models. Allowing for continuous random terms in addition to discrete random terms not only increases flexibility from the point of view of random taste heterogeneity, but also allows for the use of error-components to represent heteroscedasticity and inter-alternative correlation, where the latter is however also possible with the use of an underlying GEV structure.

Finally, independently of the additional treatment of random variations in tastes, a treatment of repeated choice observations analogous to the standard continuous mixture treatment, with tastes varying across individuals, but not across observations for the same individual, is made possible by replacing the conditional choice probabilities for individual observations in equation (5) by probabilities for sequences of choices, and by using the resulting discrete mixture term inside the log-likelihood function.

The approach we use in this paper clearly offers greater modelling flexibility than an approach based on fixed-point estimates, by allowing for random as well as deterministic variations in tastes. It may also seem tempting to see the approach as an alternative to models using continuous distributions. However, this is many cases impractical, notably because of the resulting over-specification in terms of the number of parameters, which can lead to problems in estimation. In the remainder of the paper, we therefore rely mainly on the notion that the approach is an extension of a fixed point model, while a detailed comparison between continuous and discrete mixture models, across a number of different datasets, is an important topic for further research.

Several issues arise in the estimation of discrete mixture models. Firstly, the non-concavity of the log-likelihood function does not allow the identification of a global maximum, even for discrete mixtures of MNL. Given the potential presence of a high number of local maxima, performing several estimations from various starting points is advisable. Also, it is good practice to use starting values other than 0 or 1 for the π_k^j parameters. Secondly, constrained maximum likelihood must be used to account for constraints (2) and (3). Here, it should be noted that eliminating (3) by replacing π_k^1 with

$$\pi_k^1 = 1 - \sum_{j=2}^{m_k} \pi_k^j \tag{6}$$

does not help, as the constraint $0 \leq \pi_k^1 \leq 1$ now leads to the new condition $0 \leq \sum_{j=2}^{m_k} \pi_k^j \leq 1$. Thirdly, clustering of mass points (for example around the mode of the true distribution) is a frequent phenomenon with discrete mixture models, and the use of additional bounds on the

8

mass points can be useful, based on the definition of (potentially mutually exclusive) a priori intervals for the individual mass points. In this context, a heuristic is needed to determine the optimal number of support points in actual applications.

For the purpose of this analysis, the model was coded into BIOGEME [2], where various constraints on the parameters can be imposed to address the issues described above. This also allows modellers to test the validity of specific assumptions, such as a mass at zero for the VTTS.

3   Testing the validity of the discrete mixture structure

Before proceeding to the use of discrete mixture models in practice, it is important to investigate the validity of the approach as well as its implementation in BIOGEME, by testing its performance on synthetic data where the *true* values of the parameters are known. For this, a quasi-simulated dataset was produced on the basis of a sample of $1,242$ observations taken from a binomial mode-choice survey (car *vs* rail) conducted in the context of the analysis of the VTTS in Switzerland [1, 14]. For the present analysis, the sample size was augmented from $1,242$ to $5,000$ through minor random variations on the observed attributes.

The utility specification in this model uses travel-cost, travel-time, frequency, and the number of interchanges as explanatory variables, where linear specifications are used for all attributes, and where the ASC for rail is normalised to zero. In order to generate the synthetic choices, we assume that, except for the travel time coefficient for the car alternative, the true parameters are fixed as shown in Table 1, giving a true VTTS for rail-travel of 14CHF/hour[1].

In the first experiment, we assume that the population is divided into two segments. The VTTS for car-travel in the first segment, composed of 50% of the sample, is assumed to be 16CHF/hour (car travel-time coefficient at -0.08), while it is 6CHF/hour for the second segment (car travel-time coefficient at -0.03).

The resulting dataset was then used in the estimation of a discrete mixture model with an underlying MNL structure and two support points for the car travel-time coefficient, where the results are shown in Table 2. The results show a near-perfect recovery of the $50\% - 50\%$ market share, where the upper VTTS is slightly underestimated, at 14.72CHF/hour, while the lower one is overestimated, at 7.13CHF/hour. The VTTS for rail is also slightly overestimated, at 14.84CHF/hour. These slight biases are however well within acceptable bounds.

In the second experiment, we assume that the segment with the lower VTTS represents only 30% of the population. The estimation results for this dataset are summarised in Table 3, showing that the $30\% - 70\%$ split is reproduced almost perfectly. Both VTTS measures are slightly underestimated, at 4.70CHF/hour and 13.75CHF/hour, instead of 6CHF/hour and 16CHF/hour respectively. The rail VTTS is estimated at 13.73CHF/hour, instead of 14CHF/hour. Again, these biases are acceptable.

---

[1]CHF1 $\approx$ €0.65

Although more testing is required, the two experiments described here have shown that the discrete mixture models are indeed able to recover the values and market-shares of discretely distributed coefficients. The extension to cases with more than two mass-points is possible, although the estimation becomes significantly more complicated, with the presence of several local maxima, and possible degeneracy, that is convergence of two points toward a common value.

4   VTTS case-study

We now turn our attention to the analysis exploiting the discrete mixture structure to allow for the presence of individuals with zero VTTS. For these experiments, SP data from the Swiss VTTS study were used, in the form of a binomial route-choice survey for rail travellers. The sample used in the present analysis includes 315 observations from business travellers, $1,881$ observations from leisure travellers, and 288 observations from travellers on shopping trips. The relatively small sample sizes for the business and shopping groups could decrease reliability of the results in these two groups, although problems with significance were only observed in one case, as detailed later on.

Again, the final utility specification uses travel-cost, travel-time, frequency, and the number of interchanges as explanatory variables, where linear specifications are used for all attributes. No significant ASCs could be identified in the present model. The analysis first looks at a simple MNL model, estimated separately for each of the three subgroups, with results summarised in Table 4. The results show that all estimates are of the correct sign, and significant, with the exception of the travel-time coefficient for respondents on shopping trips, which is significant only at the 74% level. In terms of substantive results, the estimation does, as expected, show higher VTTS for business travellers, with very low VTTS for shopping trips, where the value does however need to be put into context by noting the high standard error for the travel-time coefficient.

We next estimate discrete mixtures of the three MNL models, with results summarised in Table 5. With the aim of investigating the presence of individuals with zero valuations of travel-time changes, the models are specified with two travel-time coefficients, of which one is fixed at zero, while the other is initialised to zero, but estimated freely. Here, it should be noted that the implementation of the models used in the present analysis does not allow for a treatment of the repeated choice nature of the dataset, such that intra-agent and inter-agent variations in tastes are treated in the same way. As in the continuous mixture case, this can be expected to yield consistent estimates, while the use of the *panel* approach produces efficient estimates.

The results show that, at the cost of one additional *estimated* parameter, the discrete mixture models offer improvements in log-likelihood by 1.30, 17.23 and 1.32 units for respondents on business, leisure and shopping trips respectively. As such, in the present case, the discrete mixture approach leads to significant improvements only in the case of leisure travellers. However, important insights are also gained in the remaining two population segments.

The results show significant differences across the three population groups in terms of the presence of respondents with a zero VTTS. Indeed, in the model for business travellers, the share is very low, at 9.63%, while for leisure travellers, and respondents on shopping trips, the shares are a very high 65.63% and 84.59% respectively. In the case of business-travellers, the share is different from 0% at the 75% level, while, for shopping trips, it is different from 100% at the 89% level.

Any *non-traders* (e.g. respondents always choosing the cheapest or fastest alternative) had been removed from the data prior to estimation, such that these results should not be seen as a simple effect of estimation bias due to captivity. The fact that a much lower share of travellers with zero VTTS is observed in the business models is consistent with intuition. Although it is realistic to assume that, in the absence of a binding time constraint, a non-trivial part of respondents travelling for leisure or shopping purposes are indeed indifferent to travel-time changes (either positive or negative), the high shares observed in these two population groups are still striking, and call for a closer investigation, in terms of a comparison with an unconstrained model.

Before proceeding to these additional tests, it is worth looking at the findings in terms of VTTS in the share of the population associated with $\beta_{TT}(A)$. In the model for business travellers, the results are roughly similar to those observed in the model using a fixed travel-time coefficient (increase by 16.09%), which was to be expected, given the low probability associated with $\beta_{TT}(B)$. On the other hand, in the models for leisure and shopping trips, the VTTS in the share of the population associated with $\beta_{TT}(A)$ increases dramatically in comparison with the fixed coefficients model, and in fact yield VTTS higher than those observed in the model for business travellers. This however needs to be put into context by noting that the present model specification in effect groups the population into two very crude groups, one for respondents with a zero VTTS, and one for all remaining respondents. Further insights could be expected with the use of a higher number of support points, but this requires additional work to deal with identification issues.

Two interesting further observations can be made from these models. The first observation relates to the model for respondents on shopping trips. Here, the fixed travel-time coefficient in the simple MNL model was significant only at the 74% level (c.f. Table 4). However, when allowing for the presence of respondents with a zero valuation of travel-time changes, the coefficient in the remainder of the population is significant at the 95% level, although it should be noted that the associated mass is significantly different from zero only at the 89% level. Again, the findings need to be put into context by the small sample size, but the results do suggest that the estimation of a significant common coefficient for the entire population is hampered by the presence of respondents with a zero VTTS. The second observation deals with a related point. In the presence of significant variations in a given coefficient across respondents, the use of a common fixed coefficient can be seen to yield an approximate average value of this coefficient across respondents. In the present case, the simple MNL model is clearly unable to explicitly represent the presence of a part of the population with a zero VTTS, and as such, can be expected to produce a biased fixed-point estimate. This notion is supported by a calculation of the weighted average on the basis of the results from the discrete mixture model. Indeed,

using $\pi_{TT}^{(A)} \cdot \beta_{TT}(A) + \pi_{TT}^{(B)} \cdot \beta_{TT}(B)$, we obtain values of 25.95, 14.27 and 6.73 CHF/hour in the models for business, leisure and shopping trips respectively, where these values are indeed very close to the fixed-point VTTS obtained with the simple MNL model. Here, it is important to note that, because of the non-linearity of the model, this comparison is meaningful at a qualitative level only.

We now turn our attention to the comparison between the constrained and unconstrained model. The aim of this process was to test the hypothesis that there is a significant mass at zero, by comparing the model estimated with $\beta_{TT}(B)$ fixed at zero to its unconstrained counter-part. For this, the three models shown in Table 5 were re-estimated as shown in Table 6, where both $\beta_{TT}(A)$ and $\beta_{TT}(B)$ were estimated freely from the data.

The results are highly interesting. They show that, in the model for business travellers, the unconstrained model leads to a statistically significant improvement in log-likelihood by 3.12 units, at the cost of one additional parameter, hence rejecting the constrained model. Furthermore, both estimated support-points are significantly different from zero, at high levels of confidence. The distribution of the mass between the two support-points is very even, and not significantly different from a $50\% - 50\%$ split. Furthermore, the VTTS in group (A) is higher than that produced by the constrained model (c.f. Table 5), while the weighted average, at 25.68 CHF/hour is almost identical to that from the constrained model, and again close to the MNL value. Overall, these results reject the hypothesis of a significant mass at zero for the travel-time coefficient in this population segment, such that the mass of 9.63% obtained with the constrained model can be explained on the grounds that it captures mass from values close to zero. However, the results also provide proof of heterogeneity, with two different support points for β_{TT}, and better model fit than the MNL model.

While the above process thus rejects the hypothesis of a significant share of travellers with a zero VTTS in the business segment, the situation is very much different in the leisure and shopping segments. Here, the unconstrained model achieves gains in log-likelihood by 0.72 and 0.64 units in log-likelihood respectively, neither of which is significant, coming at the cost of one additional estimated parameter. Additionally, the estimated values for $\beta_{TT}(B)$ are not significantly different from zero, with confidence levels of 75% and 62% respectively. As such, the positive estimate for the two coefficients is of little importance, and should in no case be seen as a proof of the presence of respondents with a negative VTTS (see also Hess, Bierlaire and Polak 11). The VTTS for respondents in group (A) is quite close to that observed in the constrained models. Overall, the results show that, in these two groups, the unconstrained model does not reject the constrained model, such that the test does not offer convincing proof to suggest that the findings with regards to the high shares for a zero VTTS in the constrained models were incorrect.

5   Summary and Conclusions

In this paper, we have discussed an alternative approach for representing inter-agent variations in tastes, and by extension, choice behaviour. The approach is based on the use of discrete mixtures of choice models, replacing the fixed-parameter choice probabilities by a weighted sum of choice-probabilities calculated on the basis of different values for the specific coefficients for which taste heterogeneity is to be introduced. The weights associated with the different *support-points* reflect the market shares of the respective coefficient-values in the sample population. This approach has certain conceptual advantages over continuous mixtures, by being free from any a priori assumption with regards to the shape of the true distribution. Additionally, discrete mixtures can clearly serve as a starting point in the search for an appropriate continuous specification.

The main aim and contribution of this paper is to demonstrate how discrete mixture models can be used to test for the presence of respondents with zero valuations of changes in a specific travel-attribute, where, in the present case, we look specifically at the case of zero VTTS in a route-choice experiment. The results, and subsequent *validation* thereof, show that, while no evidence of a significant share of such individuals exists in the case of business travellers, a share of 66% was found for leisure travellers, with a corresponding share of 85% for respondents on shopping trips.

These results are striking, and are possibly in part specific to the data at hand, such that more testing is required. Additionally, it should be noted that, in the case of SP data, another potential reason for results showing zero valuations for changes in a given attribute for some individuals is the design of the surveys, for example in the case of a lack of variation for the concerned attribute for these individuals (i.e. insufficient stimuli). A similar issue arises in the presence of non-traders. As such, further tests should also be conducted on RP data. However, it should be noted that, while, with SP data, multiple possible explanations for zero valuations arise, discrete mixture models maintain their advantage, in terms of being able to highlight the impact of such problems.

Even though the results of this research cannot be generalised without further investigation, certain observations can be made. Indeed, the comparisons between the MNL and discrete mixture models have shown that a failure to account for the presence of individuals with a zero valuation of changes in a travel-attribute can lead to significant bias in the estimated coefficients, and by extension the willingness-to-pay indicators, possible resulting in misguided policy-measures. This problem has seemingly not been addressed in the existing literature, at least not in the context of discrete mixture models. Clearly, the ramifications of this issue are very serious indeed, and the results presented in this paper call for a thorough investigation into the prevalence of zero valuations, across a host of variables, datasets and data-sources (i.e. RP *vs* SP). Indeed, although the discussion in this paper was limited to the case of changes in travel-time, zero valuations potentially play a role for a whole range of attributes, such as for example frequency, and qualitative attributes. Additionally, problems with survey design can potentially also lead to apparent "zero-valuations" in the case of attributes such as cost, where a consistent negative

effect would be expected.

In closing, it should be noted that the same issues in terms of biased results can be seen to apply in the case of continuous mixture models when relying on the use of distributions that are not able to represent a heightened share at zero. Here, the presence of individuals with zero valuations for changes in a specific attribute can potentially also lead to biased results in terms of the existence of a share of respondents with counter-intuitively signed coefficients, a point that is related to the issue of an asymmetrical *true* distribution with a mean close to zero, as discussed by Hess, Bierlaire and Polak [11]. In this context, important work remains to be done in terms of exploring the use of model structures allowing for a variation in tastes in the non-zero domain, in addition to the presence of a significant mass at zero, in the spirit of the theoretical distribution discussed by Cirillo and Axhausen [4], who propose the use of a Normal distribution with a heightened mass at zero. Such an approach can in fact be used in combination with any type of continuous distribution, where a *discrete mixture* is used across two values, one of them equal to zero, while the second value in addition follows a continuous distribution. While straightforward from a conceptual point of view, the approach causes considerable problems in estimation, such that the search for efficient ways of implementing such combined distributions in estimation packages is an important topic for further research.

References

[1] Axhausen, K. W., Koenig, A., Abay, G., Bates, J. and Bierlaire, M. [2004], *Swiss Value of Travel Time Savings*, paper presented at the European Transport Conference, Strasbourg, France.

[2] Bierlaire, M. [2003], *BIOGEME: a free package for the estimation of discrete choice models*, Proceedings of the 3rd Swiss Transport Research Conference, Monte Verità, Ascona, Switzerland.

[3] Chintagunta, P., Jain, D. and Vilcassim, N. [1991], 'Investigating heterogeneity in brand preference in logit models for panel data', *Journal of Marketing Research* **28**, 417–428.

[4] Cirillo, C. and Axhausen, K. W. [2004], *Evidence on the distribution of values of travel time savings from a six-week diary*, Arbeitsbericht Verkehrs- und Raumplanung 212, IVT, ETH Zurich.

[5] Dong, X. and Koppelman, F. S. [2003], *Mass Point Mixed Logit Model: Development and Application*, paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, Switzerland.

[6] Fosgerau, M. [2005], *Investigating the distribution of the value of travel time savings*, Transportation Research Part B: Methodological, forthcoming.

[7] Gopinath, D. [1995], Modeling Heterogeneity in Discrete Choice Processes: Application to Travel Demand, PhD thesis, MIT, Cambridge, MA.

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

[9] Hensher, D. and Greene, W. [2003], 'The Mixed Logit Model: The State of Practice', *Transportation* **30**(2), 133–176.

[10] Hess, S., Bierlaire, M. and Polak, J. [2005*a*], Discrete mixtures of GEV models, *in* 'Proceedings of the 5th Swiss Transportation Research Conference'. www.strc.ch.

[11] Hess, S., Bierlaire, M. and Polak, J. W. [2005*b*], 'Estimation of value of travel-time savings using mixed logit models', *Transportation Research Part A: Policy and Practice* **39**(2-3), 221–236.

[12] Hess, S., Train, K. and Polak, J. W. [2005], *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: Methodological, forthcoming.

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

[14] Koenig, A., Abay, G. and Axhausen, K. W. [2004], *Zeitkostenansaetze im Personenverkehr*, Schriftenreihe 1065, final report for SVI 2001/534, Bundesamt für Strassen, UVEK, Bern, Switzerland.

[15] Lee, B. J., Fujiwara, A., Zhang, J. and Sugie, Y. [2003], *Analysis of Mode Choice Behaviours based on Latent Class Models*, paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, Switzerland.

[16] Wedel, M., Kamakura, W., Arora, N., Bemmaor, A., Chiang, J., Elrod, T., Johnson, R., Lenk, P., Neslin, S. and Poulsen, C. S. [1999], 'Discrete and continuous representations of unobserved heterogeneity in choice modeling', *Marketing Letters* **10**(3), 219–232.

Parameter	Value
ASC for car	4
Interchanges	-1.15
Travel-cost (CHF)	-0.3
Frequency (per hour)	0.9
Rail travel-time (min.)	-0.07

Table 1: Parameter values used in generation of simulated data

Sample size:	5,000
Final log-likelihood:	-868.45
Adjusted ρ^2:	0.7471

Parameter	est.	t-stat.
ASC for car	4.0265	15.76
Interchanges	-1.2306	-12.70
Travel-cost (CHF)	-0.3138	-19.62
Frequency (per hour)	0.9282	14.15
TT_{rail} (min.)	-0.0776	-13.58
$TT_{car}(A)$ (min.)	-0.0770	-5.73
$TT_{car}(B)$ (min.)	-0.0373	-3.54
Mass at $TT_{car}(A)$	0.5149	2.55
Mass at $TT_{car}(B)$	0.4851	2.40

Table 2: Results for first synthetic data experiment

16

Sample size:	5,000	
Final log-likelihood:	-906.99	
Adjusted ρ^2:	0.7360	

Parameter	est.	t-stat.
ASC for car	4.1307	16.38
Interchanges	-1.2055	-12.90
Travel-cost (CHF)	-0.3203	-20.17
Frequency (per hour)	0.9600	14.91
TT_{rail} (min.)	-0.0733	-14.04
$TT_{car}(A)$ (min.)	-0.0251	-2.35
$TT_{car}(B)$ (min.)	-0.0734	-7.25
Mass at $TT_{car}(A)$	0.2704	2.19
Mass at $TT_{car}(B)$	0.7296	5.91

Table 3: Results for second synthetic data experiment

	Business		Leisure		Shopping	
Sample size	315		1,881		288	
Final log-likelihood	-124.69		-925.36		-139.30	
Adjusted ρ^2	0.4106		0.2872		0.2822	
Parameter	est.	t-stat.	est.	t-stat.	est.	t-stat.
Interchanges	-1.1285	-6.60	-1.1737	-18.95	-0.9394	-6.30
Travel-cost (CHF)	-0.3051	-5.08	-0.1335	-5.36	-0.5658	-3.75
Frequency (per hour)	0.5970	6.11	0.4188	12.19	0.6603	7.22
Travel-time (min.)	-0.1258	-7.71	-0.0300	-7.48	-0.0465	-1.13
VTTS (CHF/hour)	24.73		13.50		4.93	

Table 4: VTTS case-study: MNL estimation results

17

	Business		Leisure		Shopping	
Sample size	315		1,881		288	
Final log-likelihood	-123.39		-908.13		-137.98	
Adjusted ρ^2	0.412		0.2996		0.2838	
Parameter	est.	t-stat.	est.	t-stat.	est.	t-stat.
Interchanges	-1.3298	-5.33	-1.5022	-15.21	-1.1109	-5.58
Travel-cost (CHF)	-0.3416	-4.39	-0.2268	-5.86	-0.7190	-3.32
Frequency (per hour)	0.6862	5.00	0.5237	11.14	0.7603	6.00
$\beta_{TT}(A)$ (min.)	-0.1635	-4.07	-0.1570	-4.52	-0.5236	-1.99
$\beta_{TT}(B)$ (min.)	0	-	0	-	0	-
Mass for $\beta_{TT}(A)$	0.9037	10.81	0.3437	7.70	0.1541	1.62
Mass for $\beta_{TT}(B)$	0.0963	1.15	0.6563	14.71	0.8459	8.87
VTTS (A) (CHF/hour)	28.71		41.54		43.69	
VTTS (B) (CHF/hour)	0		0		0	

Table 5: VTTS case-study: Discrete mixture MNL estimation results, with one support point fixed at zero

	Business		Leisure		Shopping	
Sample size	315		1,881		288	
Final log-likelihood:	-120.27		-907.41		-137.34	
Adjusted ρ^2:	0.4217		0.2994		0.2820	
Parameter	est.	t-stat.	est.	t-stat.	est.	t-stat.
Interchanges	-1.5835	-5.21	-1.5256	-14.66	-1.2352	-4.89
Travel-cost	-0.5208	-3.54	-0.2130	-5.71	-0.7048	-3.00
Frequency	0.9033	4.18	0.5292	11.12	0.8573	4.98
Mass for $\beta_{TT}(A)$	0.4718	3.49	0.4121	5.02	0.2852	1.39
Mass for $\beta_{TT}(B)$	0.5282	3.91	0.5879	7.16	0.7148	3.48
$\beta_{TT}(A)$	-0.3408	-3.38	-0.1377	-3.89	-0.4905	-1.92
$\beta_{TT}(B)$	-0.1176	-4.09	0.0119	1.15	0.1157	0.87
VTTS (A) (CHF/hour)	39.26		38.79		41.76	
VTTS (B) (CHF/hour)	13.55		-		-	

Table 6: VTTS case-study: Discrete mixture MNL estimation results, with both support points estimated from the data