A systematic comparison of continuous and discrete mixture models

Stephane Hess[*] Michel Bierlaire[†] John W. Polak[‡]

November, 17, 2006

Report TRANSP-OR 061117
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

[*]Centre for Transport Studies, Imperial College London, stephane.hess@imperial.ac.uk, Tel: +44(0)20 7594 6105, Fax: +44(0)20 7594 6102

[†]Transport and Mobility Laboratory, School of Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne, michel.bierlaire@epfl.ch, Tel: +41(0)21 693 25 37, Fax: +41(0)21 693 55 70

[‡]Centre for Transport Studies, Imperial College London, j.polak@imperial.ac.uk, Tel: +44(0)20 7594 6089, Fax: +44(0)20 7594 6102

Abstract

Modellers are increasingly relying on the use of continuous random coefficients models, such as Mixed Logit, for the representation of variations in tastes across individuals. In this paper, we provide an in-depth comparison of the performance of the Mixed Logit model with that of its far less commonly used discrete mixture counterpart, making use of a combination of real and simulated datasets. The results not only show significant computational advantages for the discrete mixture approach, but also highlight greater flexibility, and show that, across a host of scenarios, the discrete mixture models are able to offer comparable or indeed superior model performance.

1 Introduction and context

Allowing for variations in behaviour across decision makers 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 typical way of allowing for such variation is through a deterministic approach, linking the taste heterogeneity to variations in socio-demographic factors such as income or trip purpose.

While 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 structures, such as the Mixed Multinomial Logit (MMNL) model, which allow 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 (see for example Hess et al., 2006) 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. Here, there is a major risk of producing misleading results when making an inappropriate choice of distribution, as discussed by Hess et al. (2005).

In this paper, 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. Mathematically, the model structure of a DM model is a special case of a latent class model (Kamakura and Russell, 1989; Chintagunta et al., 1991, cf.), 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 (2003) and Lee et al. (2003). 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.

Thus far, there have seemingly been only two applications of this approach in the area of transport research, by Gopinath (1995), in the context of mode choice for freight shippers, and by Dong and Koppelman (2003), 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 DM models have been discussed by several other authors (Wedel et al., 1999, e.g.), 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 DM models, with the hope of encouraging their more widespread use. Additionally, the paper aims to offer a systematic comparison of the performance of discrete and continuous mixture models across a host of situations, making use of simulated data.

The remainder of this paper is organised as follows. The next section sets out the theory behind DM models. Section 3 presents a case study using real data, while Section 4 uses four different simulated datasets in a systematic comparison of discrete and continuous mixture models. Finally, Section 5 presents the conclusions of the paper.

2 Methodology

We begin by introducing some general notation, which is used throughout the remainder of this paper. Specifically, let x_{in} 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_{jn} 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 $P_n(i \mid x_n, C_n, \gamma, \beta)$ 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.

In a discrete mixture context, the number of possible values for the taste coefficients β is finite. Here, we 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[1]

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

and

$$\sum_{j=1}^{m_k} \pi_k^j = 1, \quad k = 1, \dots, K. \tag{2}$$

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

$$P_n \left(i \mid x_n, C_n, \gamma, \beta = \langle \bar{\beta}, \widehat{\beta}_1^{j_1}, \dots, \widehat{\beta}_K^{j_K} \rangle \right), \tag{3}$$

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:

$$\begin{aligned} &P_n \left(i \mid x_n, C_n, \gamma, \bar{\beta}, \widehat{\beta}, \pi \right) \\ &\quad = \sum_{j_1=1}^{m_1} \cdots \sum_{j_K=1}^{m_K} P_n \left(i \mid x_n, C_n, \gamma, \beta = \langle \bar{\beta}, \widehat{\beta}_1^{j_1}, \dots, \widehat{\beta}_K^{j_K} \rangle \right) \pi_1^{j_1} \cdot \ldots \cdot \pi_K^{j_K}, \end{aligned} \tag{4}$$

where $\bar{\beta}$, $\widehat{\beta}$ and π ($\pi = \langle \pi_1^1, \dots, \pi_1^{m_1}, \dots, \pi_K^1, \dots, \pi_K^{m_K} \rangle$) 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 (3) used inside the mixture has a closed form, then so does the DM 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 (4) 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. In this case, the

[1]These constraints can be avoided by setting $\pi_i = \frac{e^{\alpha_i}}{\sum_{j=1}^{J} e^{\alpha_j}}$, where α_j with $j = 1, \dots, J$ are estimated without constraints. While avoiding the need for constraints, this formulation becomes highly non-linear and difficult to handle in estimation.

vector γ would contain parameters that determine the nesting structure of the model. 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[2]. This however leads to a requirement to use simulation, as with all continuous mixture models.

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 (4) by probabilities for sequences of choices, and by using the resulting DM term inside the log-likelihood function.

Several issues arise in the estimation of DM 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 (1) and (2). Thirdly, clustering of mass points (for example around the mode of the true distribution) is a frequent phenomenon with DM models, and the use of additional bounds on the 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 (Bierlaire, 2003), 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, a concept discussed for example by Cirillo and Axhausen (2006).

[2]This approach can then also be used to include error components for correlation or heteroscedasticity.

3 VTTS case study

In this section, we present the findings of an analysis making use of real world data. We first give a brief description of the data in Section 3.1, before looking at model specification in Section 3.2. The estimation results are presented in Section 3.3.

3.1 Data

The study presented here makes use of Stated Preference (SP) data collected as part of a recent value of time study undertaken in Denmark (Burge and Rohr, 2004). Specifically, we make use of data describing a binary choice process for car travellers, with alternatives described only in terms of travel cost and travel time. Each respondent was presented with 9 choice situations, including one with a dominating alternative.

After eliminating the observations with a dominating alternative, as well as additional data cleaning (removing non-traders and respondents who did not choose the dominating alternative), a sample of 13,386 observations from 1,723 respondents was obtained. This equates to 3,037 observations from 392 commuters, 1,081 observations from 142 respondents travelling for education purposes, 1,767 observations from 230 people on shopping trips, 3,155 observations from 404 people travelling to visit friends or relatives, 1,752 observations from 224 general leisure travellers and 2,594 observations from 331 respondents travelling for other purposes.

To allow us to gauge the stability of the results, random subsamples of around 80% of the original sample size were generated for each of the above listed six purpose segments[3], where in each case, 10 such subsamples were created.

3.2 Model specification

The models used in this paper were estimated in log-WTP (willingness to pay) space (Fosgerau, 2004, cf.), avoiding the effect of heterogenous

[3] The selection was performed at the individual-specific level, rather than the observation-specific level.

scale (Fosgerau and Bierlaire, 2006, cf.), while allowing us to represent random variations in the VTTS without the issue of calculating the VTTS on the basis of separate randomly distributed coefficients for travel time and travel cost. Some modifications of the utility functions are required before estimation. Specifically, let T_i and C_i define the time and cost attributes of alternative i, and let us rearrange the data such that $T_1 > T_2$ and $C_1 < C_2$, i.e., the first alternative is slower but cheaper than the second alternative. Then, a very basic specification of utility is given by:

$$U_i = \beta_T T_i + \beta_C C_i, \tag{5}$$

where β_T and β_C represent time and cost coefficients respectively, and where $i = 1, 2$.

The first alternative is then chosen if the respondent is not willing to pay $C_2 - C_1$ to obtain a reduction in travel time by $T_1 - T_2$. This equates to:

$$P_1 = P\left(\beta_T (T_1 - T_2) > \beta_C (C_2 - C_1)\right). \tag{6}$$

With both β_T and β_C forced to take on negative values (Hess et al., 2005, cf.), and with the above detailed relationships between the cost and time attributes for the two alternatives, equation (6) can be rewritten as:

$$P_1 = P\left(-\frac{\Delta_C}{\Delta_T} > \frac{\beta_T}{\beta_C}\right), \tag{7}$$

where $\Delta_C = C_1 - C_2$ and $\Delta_T = T_1 - T_2$. After noting that the VTTS is given by $\frac{\beta_T}{\beta_C}$, and after a further change to equation (7), we obtain the choice probabilities in log-WTP space as follows:

$$P_1 = P\left[\ln\left(-\frac{\Delta_C}{\Delta_T}\right) > \alpha_{LV}\right], \tag{8}$$

where $\alpha_{LV} = \ln(VTTS)$.

By noting that the absence of an estimated coefficient in the utility for alternative 1 leads to a need to explicitly estimate the scale, the utility functions for alternative 1 and 2 are given by:

$$U_1 = \lambda \ln\left(-\frac{\Delta_C}{\Delta_T}\right) + \varepsilon_1 \tag{9}$$

and

$$U_2 = \lambda \, \alpha_{LV} + \varepsilon_2, \tag{10}$$

where λ is estimated in addition to α_{LV}, and where ε_1 and ε_2 give the usual type I *iid* extreme value terms. With travel costs given in Danish Krona (DKK) and travel times given in minutes, the actual VTTS in DKK per hour is obtained by $60 \cdot \exp(\alpha_{LV})$.

The specification set out above can now be used in a standard discrete choice framework, with either a fixed estimate for α_{LV}, or with random variation across respondents. At this point, it should also be noted that attempts to estimate models with an additional constant associated with the first of the two SP alternatives[4], hence accounting for a *left-right* reading effect, did not lead to any significant differences in the VTTS estimates.

3.3 Model results

During the analysis, four different types of model were estimated on the data; a simple MNL model, a MMNL model using a Normal distribution, and two DM specifications, one with two support points, DM(2), and one with three support points, DM(3)[5]. In the MMNL and DM models, the repeated choice nature of the data was taken into account by specifying the likelihood function with the integration (respectively summation in the DM models) outside the product over replications for the same respondent.

Each of these models was estimated across the six population segments and the ten subsets of the data, leading to 240 estimated models. Given this wealth of results, we presented detailed results only for shopping trips (Section 3.3.1), and give summary results for the remaining five population segments (Section 3.3.2).

3.3.1 Detailed results for shopping trips

The results for the various models estimated on the data for shopping trips are summarised in Table 1. Several differences arise across models in the

[4] In preference space.

[5] Models with more than three support points collapsed back to the more basic specifications.

Model	Parameter		1	2	3	4	5	6	7	8	9	10	
		Respondents:	185	186	180	190	182	179	188	179	180	175	
		Observations:	1421	1429	1382	1457	1403	1376	1446	1377	1383	1343	Overall
MNL		Final LL:	-880.96	-877.38	-861.61	-909.25	-869.03	-862.35	-891.57	-863.82	-840.02	-837.69	
		adj. ρ^2:	0.1036	0.1122	0.0985	0.0977	0.1043	0.0938	0.1085	0.0929	0.1216	0.0980	0.1031
		Estimation time (s):	1	1	1	1	1	1	1	1	1	1	1
	α_{LV}	est.	-1.1100	-1.1100	-1.0500	-1.0800	-1.0800	-1.1000	-1.0500	-1.0600	-1.1500	-1.0200	
		asy. t-ratio	-14.20	-14.80	-13.60	-13.80	-14.00	-13.20	-14.60	-13.20	-15.00	-13.30	
	λ	est.	0.8380	0.8800	0.8390	0.8210	0.8490	0.7950	0.8890	0.8070	0.8960	0.8460	
		asy. t-ratio	11.50	12.00	11.40	11.50	11.70	10.90	12.10	11.10	12.00	11.40	
		VTTS (DKK/hour):	19.77	19.77	21.00	20.38	20.38	19.97	21.00	20.79	19.00	21.64	20.37 (0.77)
MMNL		Final LL:	-849.65	-851.11	-828.27	-872.44	-842.02	-831.17	-862.25	-826.92	-818.61	-805.16	
		adj. ρ^2:	0.1343	0.1377	0.1322	0.1332	0.1311	0.1254	0.1367	0.1305	0.1429	0.1319	0.1336
		Estimation time (s):	75	81	74	73	80	68	76	71	80	70	74.8
	$\alpha_{LV,\mu}$	est.	-1.0800	-1.0800	-1.0200	-1.0500	-1.0500	-1.0700	-1.0300	-1.0300	-1.1200	-0.9910	
		asy. t-ratio	-11.30	-12.00	-10.70	-10.90	-11.30	-10.50	-11.60	-10.30	-12.40	-10.40	
	$\alpha_{LV,\sigma}$	est.	0.8950	0.8260	0.9070	0.9440	0.8590	0.9420	0.8360	0.9700	0.7800	0.9000	
		asy. t-ratio	8.75	8.53	8.80	9.04	8.43	8.53	8.72	8.85	8.07	8.75	
	λ	est.	1.0300	1.0500	1.0400	1.0200	1.0200	0.9810	1.0700	1.0100	1.0500	1.0600	
		asy. t-ratio	12.10	12.40	12.00	12.10	12.10	11.50	12.60	11.80	12.30	12.00	
		Mean VTTS (DKK/hour):	30.41	28.66	32.64	32.78	30.36	32.07	30.38	34.29	26.54	33.39	31.15 (2.35)
		VTTS standard deviation	33.70	28.35	36.88	39.31	31.72	38.34	30.55	42.86	24.29	37.30	34.33 (5.65)
Discrete mixture (2 pts.)		Final LL:	-845.40	-847.15	-826.74	-868.64	-838.90	-828.01	-859.40	-820.48	-817.15	-803.60	
		adj. ρ^2:	0.1366	0.1397	0.1317	0.1349	0.1322	0.1266	0.1376	0.1351	0.1424	0.1314	0.1348
		Estimation time (s):	1	1	1	1	1	1	1	1	1	1	1
	$\alpha_{LV,1}$	est.	0.5410	0.4590	0.3780	0.4120	0.5750	0.4240	0.4920	0.7600	0.0685	0.1840	
		asy. t-ratio	1.89	1.60	1.22	1.60	1.72	1.42	1.60	2.79	0.22	0.74	
	π_1	est.	0.2130	0.2020	0.2620	0.2670	0.1940	0.2570	0.2070	0.2110	0.2680	0.3330	
		asy. t-ratio $^{(i)}$	3.73	3.40	3.18	4.03	3.13	3.51	3.24	4.42	2.56	3.56	
	$\alpha_{LV,2}$	est.	-1.4700	-1.4300	-1.4900	-1.5400	-1.4000	-1.5400	-1.3900	-1.4500	-1.5400	-1.5500	
		asy. t-ratio	-13.40	-13.40	-10.70	-12.20	-12.60	-11.30	-12.80	-14.10	-9.82	-9.60	
	π_2	est.	0.7870	0.7980	0.7380	0.7330	0.8060	0.7430	0.7930	0.7890	0.7320	0.6670	
		asy. t-ratio $^{(i)}$	13.80	13.40	8.94	11.10	13.00	10.20	12.40	16.60	6.99	7.11	
	λ	est.	1.0100	1.0400	1.0200	1.0100	1.0100	0.9680	1.0500	1.0100	1.0400	1.0400	
		asy. t-ratio	12.30	12.60	12.10	12.30	12.30	11.60	12.70	12.00	12.40	12.10	
		VTTS (DKK/hour):	32.81	30.64	32.92	33.62	32.61	33.12	32.16	38.18	26.64	32.51	32.52 (2.83)
		VTTS standard deviation	36.55	32.36	32.56	34.39	36.31	34.44	33.71	46.61	22.76	27.99	33.77 (6.13)
Discrete mixture (3 pts.)		Final LL:	-844.60	-846.55	-824.79	-867.64	-837.76	-827.06	-858.06	-819.82	-816.17	-802.06	
		adj. ρ^2:	0.1354	0.1383	0.1317	0.1339	0.1313	0.1255	0.1369	0.1337	0.1413	0.1309	0.1339
		Estimation time (s):	3	3	3	3	3	3	4	4	3	2	3.1
	$\alpha_{LV,1}$	est.	0.7770	0.7060	0.8160	0.7170	0.9070	0.7630	0.8160	0.9170	0.6420	0.7410	
		asy. t-ratio	2.26	1.88	2.31	2.11	2.34	1.98	2.30	2.98	1.26	1.85	
	π_1	est.	0.1560	0.1430	0.1460	0.1750	0.1290	0.1610	0.1340	0.1750	0.1050	0.1470	
		asy. t-ratio $^{(i)}$	2.69	2.25	2.53	2.48	2.56	2.21	2.56	3.51	1.38	1.92	
	$\alpha_{LV,2}$	est.	-1.7800	-0.8940	-0.7890	-0.7920	-0.8690	-0.7890	-0.8380	-1.7800	-1.7800	-0.6650	
		asy. t-ratio	-5.40	-1.81	-2.71	-1.87	-2.50	-1.76	-2.63	-4.87	-6.63	-2.00	
	π_2	est.	0.4550	0.3810	0.4490	0.3460	0.4560	0.3600	0.4500	0.4330	0.4860	0.4110	
		asy. t-ratio $^{(i)}$	1.53	1.09	2.82	1.75	1.76	1.72	2.04	1.30	2.29	2.79	
	$\alpha_{LV,3}$	est.	-0.9100	-1.7000	-1.8800	-1.8100	-1.7900	-1.8300	-1.7700	-0.9670	-0.7450	-1.8300	
		asy. t-ratio	-2.09	-4.92	-6.86	-6.76	-5.17	-6.34	-5.88	-2.24	-1.93	-7.40	
	π_3	est.	0.3890	0.4760	0.4050	0.4800	0.4150	0.4790	0.4160	0.3920	0.4080	0.4420	
		asy. t-ratio $^{(i)}$	1.35	1.31	2.43	2.34	1.55	2.20	1.81	1.20	2.04	2.88	
	λ	est.	1.0300	1.0600	1.0500	1.0300	1.0300	0.9890	1.0800	1.0300	1.0600	1.0700	
		asy. t-ratio	12.10	12.50	12.00	12.20	12.20	11.50	12.60	11.90	12.30	12.00	
		VTTS (DKK/hour):	34.29	31.94	35.78	35.68	34.86	35.14	34.11	39.60	28.41	35.42	34.52 (2.87)
		VTTS standard deviation	41.86	37.14	42.15	40.92	44.34	41.75	40.62	51.22	30.55	38.81	40.93 (5.24)

Table 1: Estimation results on Danish shopping data

presentation of the results. As such, for the MNL model, only α_{LV} and λ are estimated. For the MMNL model, α_{LV} follows a Normal distribution, with mean $\alpha_{LV,\mu}$ and standard deviation $\alpha_{LV,\sigma}$. For the two DM models, the value of α_{LV} is spread across several support points $\alpha_{LV,k}$ with associated probabilities $0 \leq \pi_k \leq 1$, such that $\sum_{k=1}^{K} \pi_k = 1$, with $K = 2$ and $K = 3$ in DM(2) and DM(3) respectively. In addition, the table shows the calculated VTTS. For the MNL model, the mean VTTS is simply obtained through $60 \cdot \exp(\alpha_{LV})$. However, for the three mixture models, the non-linearity in the exponential means that a different approach is required. With $\alpha_{LV} \sim N(\mu_\alpha, \sigma_\alpha)$ in the MMNL model, the actual VTTS follows a log-normal distribution with mean $\mu_{VTTS} = \exp\left(\mu_\alpha + \frac{\sigma_\alpha^2}{2}\right)$ and standard deviation $\sigma_{VTTS} = \mu\sqrt{\exp(\sigma_\alpha^2) - 1}$. Both μ_{VTTS} and σ_{VTTS} can then be multiplied by 60 to obtain hourly values. For the DM models, a slightly different approach was used. As such, with K support points $\alpha_{LV,k}$ and associated probabilities π_k, a sequence of draws was generated that contained $\pi_k \cdot N$ points with a value equal to $\exp(\alpha_{LV,k})$, with $k = 1, \ldots, K$. The sample mean and standard deviation from this sequence were then used as estimates of the mean and standard deviation for the actual VTTS. For the results presented here, the value of N was set to $100,000$, beyond which no visible differences were observed for σ_{VTTS}. Finally, along with the results for individual subsamples, the table also shows some overall measures, namely the average of the adjusted ρ^2 measure, the average estimation time, and the average for μ_{VTTS} and σ_{VTTS} (together with a standard deviation of this mean across subsamples).

The first observation that can be made from Table 1 is that all three mixture models offer significant improvements in model fit over the base MNL model, across all ten subsamples. Given the structural differences between the continuous and discrete mixture models, the comparison between these models is carried out using the adjusted ρ^2 measure rather than the log-likelihood function. Here, we can see that, overall, DM(2) offers the best performance, ahead of DM(3) and the MMNL model. While the model with three support points always obtains slightly better model fit than the model with two support points, the gains are not large enough to be significant when taking into account the additional cost in terms of the number of parameters. In other words, the model with three support points is not able to retrieve significant amounts of additional heterogeneity when compared to the model with two support points. This can partly be seen as a reflection of the success of the model with two support points, but is also an illustration of the difficulties of estimating models with more than two support points, as alluded to in Section 2.

With three exceptions (samples 3, 9 and 10), the DM(2) model obtains the best performance across the three structures. Overall, the differences in performance between the DM(2) model and the MMNL model are very small, such that we now focus on other factors. Here, the first observation relates to the much lower estimation cost for the DM(2) model, with an average estimation time of one second, compared to seventy-five with the MMNL model. This much lower estimation cost would give the DM models a significant advantage in the case of larger datasets, where the absolute estimation times would be more substantial. Furthermore, the estimation time for the MMNL model was in this case kept low through the use of only 250 Halton draws in the estimation.

In terms of substantive results, the mean VTTS measures obtained by the three mixture models are significantly higher than the point estimate obtained with the MNL model. This is at least partly a result of the asymmetrical distribution of the VTTS in the mixture models. While there are also some differences between the three mixture models in the estimates for μ_{VTTS}, these are much smaller than the difference when compared to the MNL estimates. Finally, the estimate for σ_{VTTS} is much higher in the DM(3) model, while the estimate for the DM(2) model and the MMNL model are very similar.

3.3.2 Other results

Table 2 summarises the results for the various models estimated on the remaining five purpose segments. With very little variation across the ten subsamples, only the overall results are shown here. These in turn are very similar to those obtained on the data for shopping trips. As such, all three mixture models outperfom the MNL model, where the best performance is

		Commuters	Education	Leisure	Other	Visit
MNL	adj. ρ^2:	0.1017	0.1282	0.1102	0.0888	0.1007
	estimation time (s):	1	1	1	1	1
	Mean VTTS (DKK/hour):	29.08	29.32	26.40	22.73	23.82
MMNL	adj. ρ^2:	0.1263	0.1599	0.1395	0.1127	0.1294
	estimation time (s):	131	51	74	107	127
	Mean VTTS (DKK/hour):	39.51	37.28	37.62	34.76	35.83
	Std.dev. VTTS	35.90	29.24	38.43	39.85	39.61
DM(2)	adj. ρ^2:	0.1291	0.1609	0.1433	0.1156	0.1337
	estimation time (s):	2	1	1	2	2
	Mean VTTS (DKK/hour):	39.78	36.96	37.03	34.43	37.36
	Std.dev. VTTS	30.50	24.01	28.03	29.17	36.28
DM(3)	adj. ρ^2:	0.1279	0.1576	0.1412	0.1142	0.1326
	estimation time (s):	4	1	2	3	4
	Mean VTTS (DKK/hour):	39.78	37.04	37.03	34.43	37.18
	Std.dev. VTTS	30.50	24.36	28.03	29.17	36.16

Table 2: Summary of results for commuters, education trips, leisure trips, other purposes and visits

consistently obtained by the DM(2) model. Again, the DM(3) model is not able to retrieve significant levels of additional taste heterogeneity to warrant the estimation of two additional parameters. In fact, the estimates for μ_{VTTS} and σ_{VTTS} are almost universally equivalent across the two models[6]. As in the case of shopping trips, the advantages of the DM models in terms of estimation time are again very significant, across all five purpose segments. Finally, while there are almost no differences in the estimates for μ_{VTTS} between the three different mixture models (where the estimates are again significantly higher than those for the MNL models), the estimates for σ_{VTTS} are now lower in the DM models, something that was not the case in the shopping segment.

[6]It is worth noting that, with the exception of the education segment, the adjusted ρ^2 measure is higher for the DM(3) model than for the MMNL model.

4 Simulated data case studies

The application presented in Section 3 has shown the potential advantages of using a discrete mixture approach. However, it is clearly impossible to generalise these results, which could well be specific to the data at hand. For this, a systematic comparison between discrete and continuous mixture models is required; this is the topic of this section, which presents the findings of four case studies making use of simulated data.

In each of the four case studies, the generation of the data is based on the Danish VOT data used in the case study described in Section 3. Specifically, we use $10,776$ observations from $1,347$ respondents, and generate choices based on the attributes used in the original survey data. For each of the four different true models, ten sets of choices are generated for each observation, allowing us to gauge the stability of results across different samples. Unlike in the case study described in Section 3, we now work in preference space, with separate coefficients for travel time and travel cost. In each case, the travel cost coefficient is kept fixed while some random distribution is used for the travel time coefficient. Finally, the data generation was in each case carried out under the assumption of constant tastes across replications for the same individual, and the same approach was later used in model estimation.

In the first two case studies, the true model is a discrete mixture, while in the final two case studies, the true model is a continuous mixture. This allows us to gauge the relative difficulties of the two types of model in dealing with data for which the other model type is more appropriate.

Before proceeding to the discussion of the results, it should be noted that all MMNL models presented here make use of a Normal distribution. Attempts to use alternative continuous distribution functions, such as Johnson's S_B, did not lead to consistent results on the data used here. While the findings from this analysis are thus limited to a comparison between a discrete mixture and a normal mixture, it should be remembered that the vast majority of MMNL studies make use specifically of this Normal distribution, such that the results are still relevant.

4.1 Case study 1: discrete mixture with two support points

The first case study makes use of data generated with the help of a discrete mixture model with two mass points for β_T, at -1 and 0.5, with probabilities of 0.25 and 0.75 respectively. The travel cost coefficient is fixed at a value of -1, such that we obtain a mean VTTS of 37.5 DKK per hour with a standard deviation of 13.33 DKK per hour.

The estimation results obtained on this dataset are presented in two parts. Table 3 presents detailed results for the first of the ten subsamples, while Table 4 summarises the results obtained across all ten subsamples. In addition to a basic MNL model, we estimated a MMNL model using a Normal distribution and a discrete mixture model with two support points on this dataset[7]. In both cases, we allowed for random variations in β_C as well as β_T. Consistent with the true model, no variations were observed for β_C in the discrete mixture model, labelled $DM(2)_A$, such that a second model, $DM(2)_B$, was estimated, in which β_C was kept fixed.

In a comparison between the three remaining models, MNL, MMNL and $DM(2)_B$, we observe that the discrete mixture model outperforms the continuous mixture model, which in turn outperforms the MNL model. In terms of estimation time, $DM(2)_B$ has clear advantages over the MMNL model, and the higher estimation cost when compared to MNL is well justified on the basis of the improvements in model performance. All three models offer very good performance in retrieving the mean VTTS, while the two mixture models additionally offer good performance in the estimation of the standard deviation.

A final point deserves some special attention. As mentioned above, we initially allowed for random variation in β_C as well as β_T. The estimation of the first discrete mixture model, $DM(2)_A$, offered no evidence of such heterogeneity, such that the model was replaced by $DM(2)_B$. However, for the continuous mixture model, MMNL, we retrieved significant heterogeneity for β_C as well as for β_T, despite the fact that β_C was kept fixed in

[7] No further gains in model performance were obtained by allowing for more than two support points.

	MNL		MMNL		$DM(2)_A$		$DM(2)_B$	
Final LL	-4565.42		-4122.22		-4007.05		-4007.05	
par.	2		4		8		5	
adj. ρ^2	0.3885		0.4476		0.4625		0.4629	
est.time (s)	2		234		17		6	
	est.	asy.t-rat.	est.	asy.t-rat.	est.	asy.t-rat.	est.	asy.t-rat.
β_T	-0.4081	-36.16	-	-	-	-	-	-
$\beta_{T,\mu}$	-	-	-0.6409	-36.28	-	-	-	-
$\beta_{T,\sigma}$	-	-	0.1553	10.31	-	-	-	-
$\beta_{T,1}$	-	-	-	-	-0.5050	-40.99	-0.5050	-40.99
$\pi_{T,1}$	-	-	-	-	0.7258	50.22	0.7258	50.22
$\beta_{T,2}$	-	-	-	-	-1.0231	-40.81	-1.0231	-40.81
$\pi_{T,2}$	-	-	-	-	0.2742	18.97	0.2742	18.97
β_C	-0.6424	-34.09	-	-	-	-	-1.0083	-42.20
$\beta_{C,\mu}$	-	-	-1.0613	-36.64	-	-	-	-
$\beta_{C,\sigma}$	-	-	0.2071	9.66	-	-	-	-
$\beta_{C,1}$	-	-	-	-	-1.0083	-12.20	-	-
$\pi_{C,1}$	-	-	-	-	0.3035	0.00	-	-
$\beta_{C,2}$	-	-	-	-	-1.0083	-24.03	-	-
$\pi_{C,2}$	-	-	-	-	0.6965	0.00	-	-
μ_{VTTS}	38.11		37.81		38.50		38.50	
σ_{VTTS}	-		12.75		13.75		13.75	

Table 3: Detailed estimation results for first subsample for first simulated dataset

the generation of the data. This offers clear evidence of confounding; by being unable to retrieve the correct patterns of heterogeneity for β_T, the MMNL model explains part of the remaining error in the model through heterogeneity in β_C. As such, while the model is able to correctly retrieve the mean and standard deviation of the VTTS, it does so by incorrectly indicating a variation across respondents in the sensitivity to changes in travel cost.

The findings from Table 3 are confirmed by a graphical analysis of the shape for the distribution of β_T in Figure 1, where this comparison is made possible by the fact that the mean estimate for β_C is essentially equal to -1 in all models.

Due to space considerations, no detailed results are presented for the remaining nine subsamples. The results are available on request. Never-

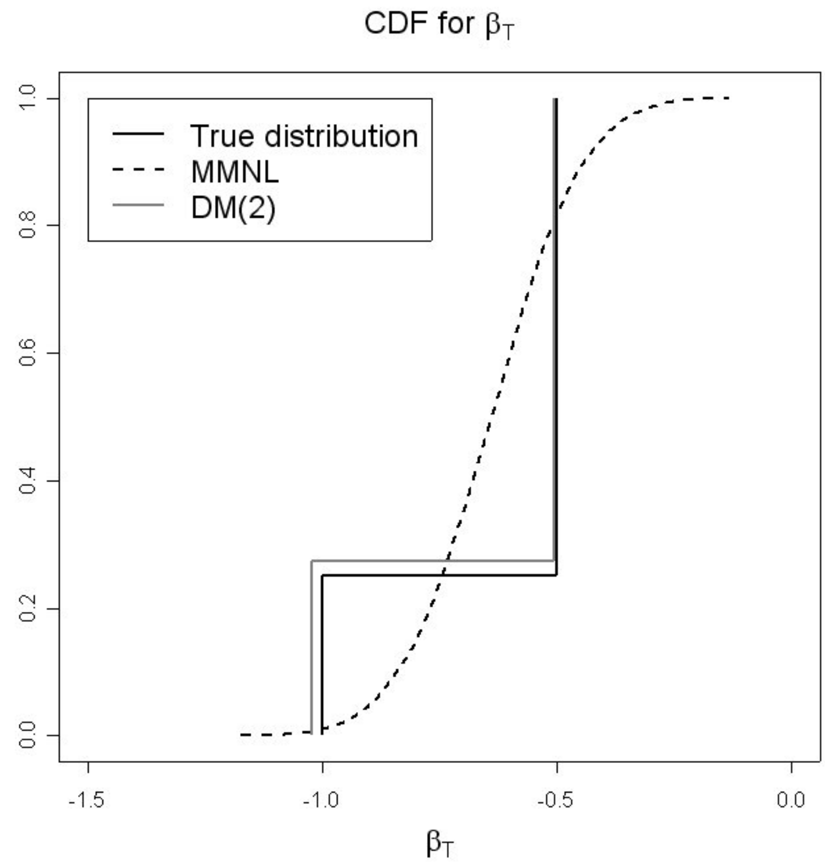

Figure 1: Cumulative distribution function for β_T first subsample for first simulated dataset

theless, the results presented in Table 4 give an indication of the stability of the results across the ten samples. As such, there is very little variation in terms of model performance (adj. ρ^2), where the advantages of the DM model clearly remain, with the same applying in the case of estimation time. Finally, while for the mean VTTS, the results are very stable across datasets and models, the estimation of the MMNL models led to very high standard deviations for the VTTS measures in some of the subsamples, which is reflected in a higher mean value for σ_{VTTS}, along with greater variation across samples. This is a direct result of the incorrect patterns of

		MNL	MMNL	$DM(2)_A$	$DM(2)_B$
adj.ρ^2	mean	0.3919	0.4475	0.4601	0.4605
	std.dev.	0.0053	0.0057	0.0057	0.0057
Est.time (s)	mean	1.8	258.6	19.5	5.4
	std.dev.	0.42	16.96	4.72	0.52
μ_{VTTS}	mean	37.94	37.88	38.22	38.22
	std.dev.	0.25	0.41	0.31	0.31
σ_{VTTS}	mean	-	21.89	13.25	13.25
	std.dev.	-	16.43	0.38	0.38

Table 4: Summary of results across subsamples for first simulated dataset

heterogeneity retrieved for β_C in these models, leading to a wider range for the VTTS.

4.2 Case study 2: discrete mixture with three support points

In the second case study, the true model is again a discrete mixture of a MNL model, where this time, three support points are used for β_T, at -1, -0.7 and -0.4, with probabilities of 0.3, 0.35 and 0.35. This leads to a true mean VTTS of 41.1 DKK per hour, with a standard deviation of 14.48 DKK per hour. Four different models were estimated on these data; along with the usual MNL and MMNL models, we estimated a DM with two support points, and a DM with three support points[8]. Again, the DM models were estimated with two different specifications, using a randomly distributed β_C coefficient in $DM(2)_A$ and $DM(3)_A$, and a fixed β_C coefficient in $DM(2)_B$ and $DM(3)_B$. The detailed results for the first sample are presented in Table 5, while the overall results are summarised in Table 6.

The results show major improvements for the MMNL and various DM models when compared to the MNL model. All six models perform very well in terms of retrieving the mean VTTS, while the five mixture models also obtain a good approximation to the true standard deviation of the

[8]No further gains could be made by using more than three support points.

	MNL		MMNL		$DM(2)_A$		$DM(2)_B$		$DM(3)_A$		$DM(3)_B$	
Final LL	-4721.69		-4155.65		-4126.23		-4227.43		-4120.96		-4120.99	
par.	2		4		8		5		12		7	
adj. ρ^2	0.3676		0.4431		0.4465		0.4334		0.4467		0.4473	
est.time (s)	1		346		16		6		151		13	
	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat
β_T	-0.3925	-33.72	-	-	-	-	-	-	-	-	-	-
$\beta_{T,\mu}$	-	-	-0.6817	-34.78	-	-	-	-	-	-	-	-
$\beta_{T,\sigma}$	-	-	0.2423	25.15	-	-	-	-	-	-	-	-
$\beta_{T,1}$	-	-	-	-	-0.8561	-36.56	-0.4005	-36.62	-0.3930	-29.72	-0.7015	-34.17
$\pi_{T,1}$	-	-	-	-	0.6210	27.38	0.5028	27.72	0.3185	14.88	0.4069	14.02
$\beta_{T,2}$	-	-	-	-	-0.4221	-28.99	-0.8084	-39.33	-0.7031	-32.76	-0.3927	-29.84
$\pi_{T,2}$	-	-	-	-	0.3790	16.71	0.4972	27.41	0.4093	13.50	0.3187	14.89
$\beta_{T,3}$	-	-	-	-	-	-	-	-	-1.0262	-32.13	-1.0234	-34.26
$\pi_{T,3}$	-	-	-	-	-	-	-	-	0.2723	10.94	0.2744	11.64
β_C	-0.5732	-33.39	-	-	-	-	-0.8783	-40.98	-	-	-1.0084	-39.31
$\beta_{C,\mu}$	-	-	-0.9965	-37.48	-	-	-	-	-	-	-	-
$\beta_{C,\sigma}$	-	-	0.0591	4.51	-	-	-	-	-	-	-	-
$\beta_{C,1}$	-	-	-	-	-1.2023	-35.79	-	-	-1.0015	-23.66	-	-
$\pi_{C,1}$	-	-	-	-	0.5357	13.28	-	-	0.8114	0.69	-	-
$\beta_{C,2}$	-	-	-	-	-0.8469	-33.57	-	-	-1.0454	-6.67	-	-
$\pi_{C,2}$	-	-	-	-	0.4643	11.51	-	-	0.1886	0.16	-	-
$\beta_{C,3}$	-	-	-	-	-	-	-	-	-1.2583	0.00	-	-
$\pi_{C,3}$	-	-	-	-	-	-	-	-	0.0000	0.00	-	-
μ_{VTTS}	41.08		41.18		41.22		41.25		41.18		41.09	
σ_{VTTS}	-		14.86		14.68		13.94		14.41		14.44	

Table 5: Detailed estimation results for first subsample for second simulated dataset

		MNL	MMNL	$DM(2)_A$	$DM(2)_B$	$DM(3)_A$	$DM(3)_B$
adj.ρ^2	mean	0.3690	0.4429	0.4457	0.4351	0.4460	0.4467
	std.dev.	0.0051	0.0035	0.0039	0.0038	0.0037	0.0037
Est.time (s)	mean	1.9	338.6	17.5	6.3	133.5	13.6
	std.dev.	1.1	8.4	2.32	1.7	56.4	2.4
μ_{VTTS}	mean	40.93	40.74	40.87	40.75	40.79	40.78
	std.dev.	0.16	0.25	0.24	0.29	0.22	0.23
σ_{VTTS}	mean	-	14.62	14.43	13.77	14.35	14.30
	std.dev.	-	0.25	0.25	0.27	0.21	0.24

Table 6: Summary of results across subsamples for second simulated dataset

VTTS. We now look in more detail at the differences between the various mixture models. As was the case in the case study discussed in Section 4.1, the MMNL model again falsely recovers some random variation for β_C, where the level of variation is however much lower than was the case in the first case study. When only allowing for two support points, the DM models also retrieve significant variation for β_C, as reflected in the drop in model fit observed from $DM(2)_A$ to $DM(2)_B$ when constraining β_C to a fixed value. This is no longer the case when using three support points. Finally, as was the case in Section 4.1, the DM models again have a significant advantage over the MMNL model in terms of estimation cost.

Figure 2 shows the cumulative distribution functions for β_T in the MMNL model, as well as in $DM(2)_A$ and $DM(2)_B$. The advantages of the DM models are again very obvious, especially in the case of the model with three support points.

The results from Table 6 show very stable performance across the ten samples, for all four indicators. The fact that, unlike in the first case study (cf., Table 4), the estimate for σ_{VTTS} in the MMNL model is now very stable can be explained by the lower coefficient of variation for β_C in the MMNL estimates in the second case study.

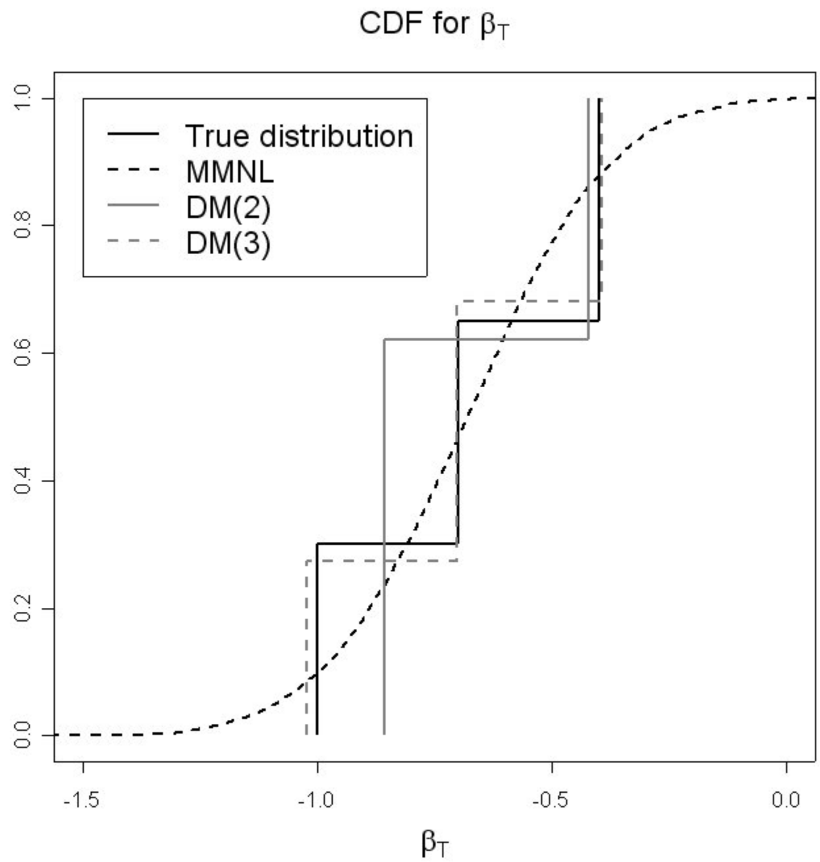

Figure 2: Cumulative distribution function for β_T first subsample for second simulated dataset

4.3 Case study 3: Normal mixture

For the third case study, a MMNL model with a normally distributed travel time coefficient was chosen as the true model. Specifically, β_C is still fixed to a value of -1, while β_T now follows a Normal distribution with mean of -0.8 and a standard deviation of 0.3, leading to a mean VTTS of 48 DKK/hour, with a standard deviation of 18 DKK.

The results for the first subsample of the third simulated dataset are summarised in Table 7. A slightly different strategy was employed in the

	MNL		$MMNL_A$		$MMNL_B$		$DM(5)_A$		$DM(5)_B$		$DM(6)_A$		$DM(6)_B$	
Final LL	-4742.06		-3912.57		-3913.9		-3910.2		-3913.54		-3908.43		-3908.61	
par.	2		4		3		14		11		16		13	
adj. ρ^2	0.3649		0.4756		0.4756		0.4746		0.4746		0.4746		0.4750	
est.time (s)	1		341		233		143		41		141		59	
	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat
β_T	-0.4008	-32.11	-	-	-	-	-	-	-	-	-	-	-	-
$\beta_{T,\mu}$	-	-	-0.8359	-36.67	-0.8329	-36.69	-	-	-	-	-	-	-	-
$\beta_{T,\sigma}$	-	-	0.3134	25.27	0.3113	25.03	-	-	-	-	-	-	-	-
$\beta_{T,1}$	-	-	-	-	-	-	-0.1343	-1.99	-0.1867	-4.82	-0.0859	-1.55	-0.1293	-1.88
$\pi_{T,1}$	-	-	-	-	-	-	0.0372	2.34	0.0566	3.82	0.0245	2.55	0.0362	2.32
$\beta_{T,2}$	-	-	-	-	-	-	-0.4585	-13.87	-0.5071	-17.37	-0.3621	-8.79	-0.4449	-14.26
$\pi_{T,2}$	-	-	-	-	-	-	0.1838	6.55	0.2326	9.14	0.0742	1.90	0.1710	6.66
$\beta_{T,3}$	-	-	-	-	-	-	-0.7021	-22.49	-0.7904	-28.13	-0.7102	-20.65	-0.6800	-24.56
$\pi_{T,3}$	-	-	-	-	-	-	0.2231	3.96	0.3529	9.63	0.2026	3.28	0.2247	4.71
$\beta_{T,4}$	-	-	-	-	-	-	-1.1872	-29.66	-1.1208	-26.75	-0.9006	-19.61	-0.8830	-20.63
$\pi_{T,4}$	-	-	-	-	-	-	0.2905	7.70	0.3296	9.67	0.2653	5.11	0.2597	5.79
$\beta_{T,5}$	-	-	-	-	-	-	-0.8964	-19.79	-1.6253	-20.78	-0.5177	-10.19	-1.1502	-29.93
$\pi_{T,5}$	-	-	-	-	-	-	0.2654	5.25	0.0283	2.85	0.1441	3.01	0.2818	7.47
$\beta_{T,6}$	-	-	-	-	-	-	-	-	-	-	-1.1902	-29.58	-1.6423	-21.24
$\pi_{T,6}$	-	-	-	-	-	-	-	-	-	-	0.2893	7.61	0.0267	3.10
β_C	-0.4999	-30.13	-	-	-1.0267	-38.53	-	-	-1.0135	-37.67	-	-	-1.0213	-37.66
$\beta_{C,\mu}$	-	-	-1.0254	-38.58	-	-	-	-	-	-	-	-	-	-
$\beta_{C,\sigma}$	-	-	0.0080	0.47	-	-	-	-	-	-	-	-	-	-
$\beta_{C,1}$	-	-	-	-	-	-	-0.7467	-18.57	-	-	-0.7485	-18.50	-	-
$\pi_{C,1}$	-	-	-	-	-	-	0.0862	2.67	-	-	0.0861	2.68	-	-
$\beta_{C,2}$	-	-	-	-	-	-	-1.0545	-34.46	-	-	-1.0569	-34.41	-	-
$\pi_{C,2}$	-	-	-	-	-	-	0.9138	28.31	-	-	0.9139	28.42	-	-
μ_{VTTS}	48.10		48.93		48.68		48.96		48.72		48.77		48.81	
σ_{VTTS}	-		18.34		18.20		18.08		18.15		18.15		18.15	

Table 7: Detailed estimation results for first subsample for third simulated dataset

model estimation in this case study. From the experience of the first two case studies, it had to be assumed that some of the distribution of β_T would erroneously be picked up as heterogeneity in β_C. This would apply especially in the discrete mixture models with a low number of support points. As such, alongside the MNL model, two different MMNL models were estimated, one with β_C kept fixed, and one with a randomly distributed β_C. In the discrete mixture models, 2 support points were used for β_C, while the number of support points for β_T was gradually increased up to the point where no heterogeneity was retrieved for β_C, i.e. the random taste heterogeneity in the data is captured correctly by β_T on its own. It was found that this point was reached between five and six support points for β_T. No further gains in model performance could be obtained by increasing the number of support points for β_T any further, independently of the treatment of β_C.

Again, all the different models offer good performance in retrieving the true mean value of the VTTS, while the various mixture models additionally offer a good approximation to the true standard deviation. The six mixture models offer significant improvements in model performance when compared to the MNL model. As in the other examples, the DM models again have computational advantages over the MNL model. Given the results from the other case studies, it is of interest to look at the issue of confounding between the heterogeneity for β_T and β_C. In the MMNL model and the DM model with six support points, the reductions in model fit resulting from using a fixed β_C coefficient are not significant. With only five support points, the drop in model fit is slightly more visible ($\mathrm{DM}(5)_\mathrm{A}$ *vs* $\mathrm{DM}(5)_\mathrm{B}$), yet still not significant when taking into account the cost of estimating three additional parameters. However, in earlier models, using fewer than five support points for β_T, this was not the case, and there was significant confounding[9].

Finally, it is of interest to look at the specific patterns of heterogeneity retrieved by the discrete mixture models, where we focus on MMNL_B, $\mathrm{DM}(5)_\mathrm{A}$ and $\mathrm{DM}(6)_\mathrm{B}$. Here, it can be seen from Figure 3 that the two DM models offer a very good approximation to the Normal distribution.

[9] Detailed results available on request.

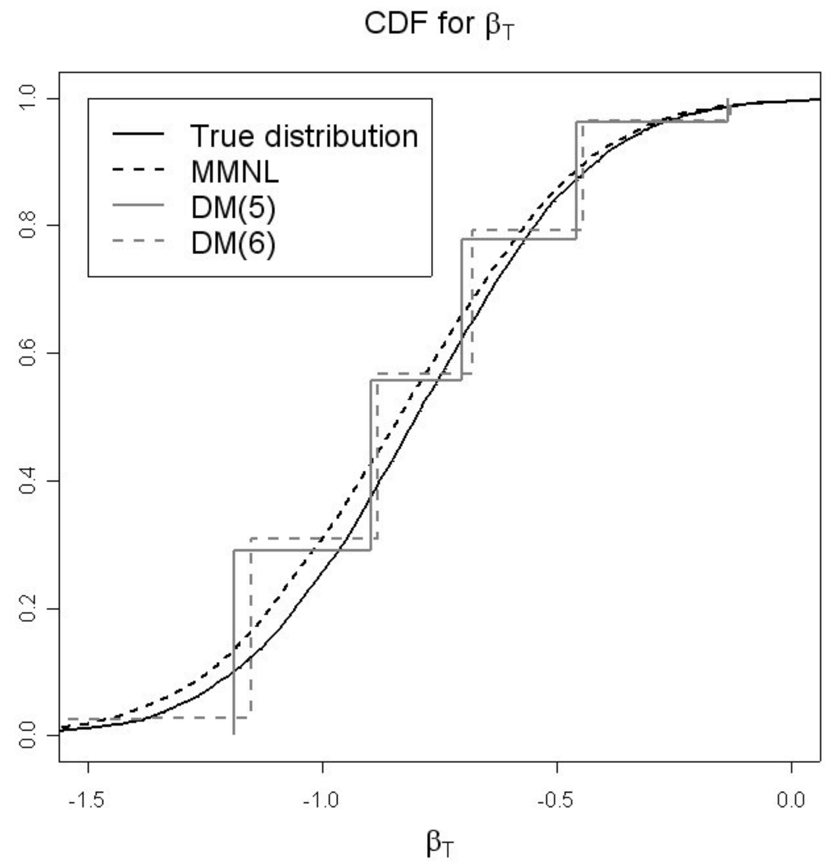

Figure 3: Cumulative distribution function for β_T first subsample for third simulated dataset

The average results across the ten subsamples are summarised in Table 8. The results show that, on average (and unlike in the first subsample), not allowing for heterogeneity in β_C leads to a minor drop in the adjusted ρ^2 measure for the MMNL model and the DM(5) model. This is however again not the case for DM(6), showing that six support points are sufficient to retrieve the true heterogeneity in the data. In terms of estimation time, the DM mixtures retain their advantage, even with a higher number of support points. Finally, the results for the mean and standard deviation of the VTTS are very stable across subsamples.

		MNL	MMNL_A	MMNL_B	DM(5)_A	DM(5)_B	DM(6)_A	DM(6)_B
adj. ρ^2	mean	0.3701	0.4724	0.4722	0.4715	0.4710	0.4714	0.4718
	std.dev.	0.0057	0.0057	0.0057	0.0057	0.0055	0.0057	0.0056
Est.time (s)	mean	1	338.1	242.9	124.8	45.3	172	63.7
	std.dev.	0.0	7.9	25.5	31.27	8.3	30.4	14.4
VTTS μ	mean	47.71	48.71	48.58	48.70	48.61	48.64	48.60
	std.dev.	0.35	0.40	0.41	0.48	0.49	0.52	0.49
VTTS σ	mean	-	17.63	17.63	17.68	17.46	17.60	17.60
	std.dev.	-	0.48	0.48	0.45	0.49	0.43	0.45

Table 8: Summary of results across subsamples for third simulated dataset

4.4 Case study 4: Mixture of two Normals

For the fourth case study, a more complex mixture was used. As such, the true distribution is now a mixture of two Normal distributions, where $\beta_T = \pi_1 \beta_{T_1} + \pi_2 \beta_{T_2}$, with $\pi_1 = \pi_2 = 0.5$, and with $\beta_{T_1} \sim N(-0.8, 0.2)$ and $\beta_{T_2} \sim N(-0.3, 0.1)$. The cost coefficient β_C was again kept fixed at -1. With this, we obtain a true mean VTTS of 33 DKK/hour, with a standard deviation of 17.76 DKK. In model estimation, the strategy from the third case study was again adopted, gradually increasing the number of support points for β_T in the DM models, while maintaining the number of support points for β_C fixed at 2. Again, the issue of confounding largely disappeared when using five or more support points.

The results for the first subsample are presented in Table 9, with Table 10 presenting a summary of the results across all ten subsamples. Along with the MNL model, two MMNL models were estimated, where MMNL_A and MMNL_B again differ by using a randomly distributed and fixed β_C coefficient respectively. Although the standard deviation for β_C is significantly different from zero in model MMNL_A, it is very small compared to the mean value, such that it is no surprise that the effect of using a fixed coefficient is very small, with very similar model performance for MMNL_B. In the DM models, we experience a very small, and insignificant drop in model fit when constraining β_C to a single value. Here, two further observations can be made. In model DM(5)_A, the difference between $\beta_{C,1}$ and $\beta_{C,2}$ is not significant beyond the 48% level of confidence, while, in model

	MNL		$MMNL_A$		$MMNL_B$		$DM(5)_A$		$DM(5)_B$		$DM(6)_A$		$DM(6)_B$	
Final LL	-5296.84		-4405.54		-4406.11		-4359.07		-4363.23		-4359.03		-4363.23	
par.	2		4		3		14		11		16		13	
adj. ρ^2	0.2906		0.4096		0.4097		0.4145		0.4144		0.4143		0.4141	
est.time (s)	7		341		213		197		33		174		75	
	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat
β_T	-0.2153	-27.80	-	-	-	-	-	-	-	-	-	-	-	-
$\beta_{T,\mu}$	-	-	-0.5115	-30.23	-0.5155	-30.62	-	-	-	-	-	-	-	-
$\beta_{T,\sigma}$	-	-	0.2930	29.23	0.2954	29.82	-	-	-	-	-	-	-	-
$\beta_{T,1}$	-	-	-	-	-	-	-0.0844	-1.33	-0.0787	-1.27	-0.0855	-1.36	-0.0787	-1.27
$\pi_{T,1}$	-	-	-	-	-	-	0.0377	1.42	0.0363	1.47	0.0385	1.41	0.0363	1.47
$\beta_{T,2}$	-	-	-	-	-	-	-0.2831	-7.90	-0.2713	-16.19	-0.2835	-7.01	-0.2713	-16.19
$\pi_{T,2}$	-	-	-	-	-	-	0.3461	0.98	0.3761	6.32	0.3395	0.77	0.3761	6.32
$\beta_{T,3}$	-	-	-	-	-	-	-0.9833	-20.76	-0.3788	-11.94	-0.3216	-3.95	-0.3788	-11.94
$\pi_{T,3}$	-	-	-	-	-	-	0.1900	4.71	0.0921	1.46	0.1102	0.25	0.0921	1.46
$\beta_{T,4}$	-	-	-	-	-	-	-0.3253	-4.58	-0.6543	-28.15	-0.7247	-9.82	-0.4676	0.00
$\pi_{T,4}$	-	-	-	-	-	-	0.1047	0.29	0.2848	10.16	0.0749	0.33	0.0000	0.00
$\beta_{T,5}$	-	-	-	-	-	-	-0.6690	-20.61	-0.9546	-30.00	-0.6555	-11.72	-0.6543	-28.15
$\pi_{T,5}$	-	-	-	-	-	-	0.3215	8.36	0.2107	8.94	0.2508	1.13	0.2848	10.16
$\beta_{T,6}$	-	-	-	-	-	-	-	-	-	-	-0.9842	-21.79	-0.9546	-30.00
$\pi_{T,6}$	-	-	-	-	-	-	-	-	-	-	0.1860	4.77	0.2107	8.94
β_C	-0.4194	-31.89	-	-	-0.9721	-37.89	-	-	-0.9737	-37.40	-	-	-0.9737	-37.40
$\beta_{C,\mu}$	-	-	-0.9718	-38.17	-	-	-	-	-	-	-	-	-	-
$\beta_{C,\sigma}$	-	-	0.0352	1.97	-	-	-	-	-	-	-	-	-	-
$\beta_{C,1}$	-	-	-	-	-	-	-1.0781	-22.66	-	-	-0.8725	-15.68	-	-
$\pi_{C,1}$	-	-	-	-	-	-	0.5965	4.03	-	-	0.3572	1.70	-	-
$\beta_{C,2}$	-	-	-	-	-	-	-0.8801	-20.83	-	-	-1.0662	-18.58	-	-
$\pi_{C,2}$	-	-	-	-	-	-	0.4035	2.73	-	-	0.6428	3.06	-	-
μ_{VTTS}	30.80		31.60		31.82		32.60		32.49		32.64		32.49	
σ_{VTTS}	-		18.15		18.23		17.39		17.10		17.38		17.10	

Table 9: Detailed estimation results for first subsample for fourth simulated dataset

$\mathrm{DM}(6)_A$, it is not significant beyond the 50% level of difference. It can also be seen that, on average, when moving from $\mathrm{DM}(5)_A$ to $\mathrm{DM}(5)_B$ and from $\mathrm{DM}(6)_A$ to $\mathrm{DM}(6)_B$, the standard errors associated with the various $\pi_{T,k}$ parameters decrease. Finally, model $\mathrm{DM}(6)_B$ can be seen to reduce to model $\mathrm{DM}(5)_B$; the additional support point, as well as its associated probability, are not significantly different from zero. All seven models again offer good performance in the retrieval of the true mean VTTS, where the six mixture models also perform well for the standard deviation. The DM models maintain their advantages in terms of estimation cost, where these are naturally smaller than before given the higher number of parameters. In terms of model performance, the MMNL models clearly outperform the MNL model, while the various DM models have a small advantage over the MMNL models. The results from Table 10 again show stable performance over the ten subsamples.

When looking at the retrieval of the true shape for the distribution of β_T, it can be seen that the MMNL models using a single Normal distribution produce a mean that is the weighted average of the mean of the two Normal distributions. The DM models on the other hand do recover the multi-modality of the true distribution[10]. These findings are reflected in the shape of the distributions for β_T in Figure 4, where the DM models ($\mathrm{DM}(5)_A$ and $\mathrm{DM}(6)_B$) are better able to account for the multi-modality of the true distribution.

In closing, it should be noted that, in this example, the uni-modal MMNL model still manages to retrieve the true mean and standard deviation of the multi-modal true distribution of the VTTS. This can be explained by the fact that the probabilities for the two Normal distributions were set evenly to 0.5, where the difference in the standard deviation for β_{T_1} and β_{T_2} was also rather small. Different patterns could be expected in a more asymmetrical scenario.

[10] It should be noted that, in the retained DM model, $\mathrm{DM}(5)_B$, two of the probabilities for support points, $\pi_{T,1}$ and $\pi_{T,3}$, are only significant at the 85% level of confidence.

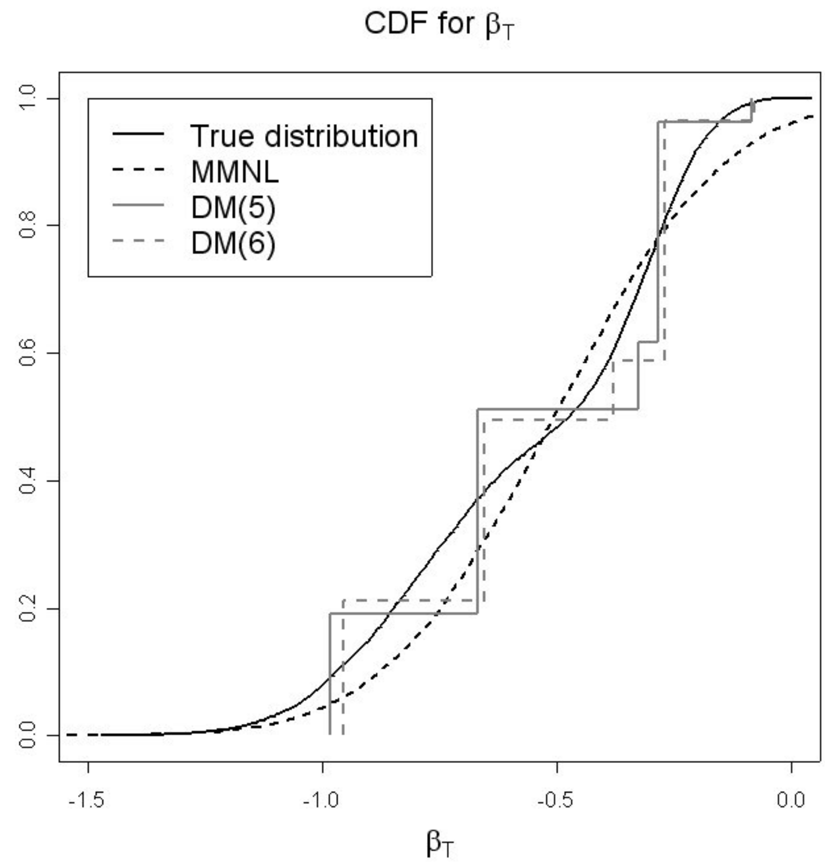

Figure 4: Cumulative distribution function for β_T first subsample for fourth simulated dataset

5 Summary and Conclusions

With the availability of powerful computers and estimation tools, researchers and practitioners are increasingly making use of continuous mixture structures, such as Mixed Logit, in the representation of random taste heterogeneity across respondents. Despite the gains in estimation power, the cost of using such mixture models remains high, especially in large scale studies. Furthermore, several issues arise due to the models' reliance on specific distribution functions, whose shape is not necessarily consistent with that

		MNL	$MMNL_A$	$MMNL_B$	$DM(5)_A$	$DM(5)_B$	$DM(6)_A$	$DM(6)_B$
adj.ρ^2	mean	0.2896	0.4082	0.4082	0.4118	0.4115	0.4116	0.4118
	std.dev.	0.0036	0.0041	0.0041	0.0045	0.0040	0.0046	0.0045
Est.time (s)	mean	8.8	363.7	235.4	142.1	40.8	168	59.7
	std.dev.	2.4	14.8	10.9	37.28	10.0	18.0	11.8
VTTS μ	mean	31.32	31.90	32.27	32.87	32.74	32.83	32.76
	std.dev.	0.38	0.48	0.44	0.51	0.46	0.46	0.47
VTTS σ	mean	-	18.48	18.67	17.71	17.58	17.73	17.66
	std.dev.	-	0.33	0.31	0.31	0.35	0.29	0.34

Table 10: Summary of results across subsamples for fourth simulated dataset

of the *true*, unobserved distribution.

In this paper, we have discussed an alternative approach for the representation of random taste heterogeneity, making use of discrete mixtures instead of continuous mixtures. Although several issues can also arise in the estimation of such models, they have the advantage of a closed form solution, and can hence be estimated and applied without relying on simulation processes. Furthermore, the models are free from a priori assumptions as to the shape of the *true* distribution.

The paper presents several case studies offering an in-depth comparison of the two modelling approaches, making use of real data as well as four separate simulated datasets. The results of these analyses clearly show the major advantage of the discrete mixture approach in terms of estimation cost. They also show that, across scenarios, the discrete mixture models are able to attain similar or indeed better performance than their continuous counterparts. Finally, they are better able to deal with complicated *true* distributions, such as the presence of multiple modes.

Although further comparisons between the two modelling approaches are required, the results from this paper do suggest that discrete mixture models present a viable alternative, partly thanks to their lower cost in estimation and application, but also due to the absence of a priori shape assumptions, which is of great interest in the context of recent discussions of the issue of the specification of continuous heterogeneity by Hess et al. (2005).

Acknowledgements

Part of the work described in this paper was carried out during a guest stay by the first author in the Institute of Transport and Logistics Studies at the University of Sydney.

References

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.

Burge, P. and Rohr, C. (2004). *DATIV: SP Design: Proposed approach for pilot survey*, Tetra-Plan in cooperation with RAND Europe and Gallup A/S.

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.

Cirillo, C. and Axhausen, K. W. (2006). Evidence on the distribution of values of travel time savings from a six-week diary, *Transportation Research Part A: Policy and Practice* **40**(5): 444–457.

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.

Fosgerau, M. (2004). *Nonparametric and semiparametric estimation of the distribution of the value of travel time*, paper presented at the European Transport Conference, Strasbourg.

Fosgerau, M. and Bierlaire, M. (2006). *Discrete choice models with multiplicative error terms*, Technical Report TRANSP-OR 060831, Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne.

Gopinath, D. (1995). *Modeling Heterogeneity in Discrete Choice Processes: Application to Travel Demand*, PhD thesis, MIT, Cambridge, MA.

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

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

Hess, S., Train, K. and 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: Methodological* **40**(2): 147–163.

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

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 10$^{\text{th}}$ International Conference on Travel Behaviour Research, Lucerne.

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.