Accommodating coefficient outliers in discrete choice modelling: a comparison of discrete and continuous mixing approaches

Danny Campbell
Institute for a Sustainable World, Queen's University Belfast
E-mail: d.campbell@qub.ac.uk

Stephane Hess
Institute for Transport Studies, University of Leeds
E-mail: s.hess@its.leeds.ac.uk

Riccardo Scarpa
Economics Department, University of Waikato
E-mail: rscarpa@mngt.waikato.ac.nz

John M. Rose
Institute of Transport and Logistics Studies, University of Sydney
E-mail: j.rose@itls.usyd.edu.au

To appear in *Choice modelling: the state-of-the-art and the state-of-practice*,
edited by Stephane Hess and Andrew Daly

October, 2009

Abstract

The presence of respondents with apparently extreme sensitivities in choice data may have an important influence on model results, yet their role is rarely assessed or even explored. Irrespective of whether such outliers are due to genuine preference expressions, their presence suggests that specifications relying on preference heterogeneity may be more appropriate. In this paper we compare the potential of discrete and continuous mixture distributions in identifying and accommodating extreme coefficient values. To test our methodology we use five stated preference datasets (four simulated and one real). The real data were collected to estimate the existence value of rare and endangered fish species in Ireland.

1 Introduction

The analysis of discrete choice data is nowadays routinely conducted by means of specifications which accommodate random taste variation (e.g., Revelt and Train, 1998; Train, 1998; McFadden and Train, 2000; Hensher and Greene, 2003). This growing interest in understanding and explaining preference heterogeneity has yet to produce a systematic investigation of the role of extreme heterogeneity, where the term 'extreme' refers the values of taste coefficient outliers. In the present paper, we distinguish such type of outliers from choice prediction outliers (e.g., see Ben-Akiva and Lerman, 1985) and work by (Daly and Zachary, 1975) in the context of choice prediction for binary logit and probit models[1].

While the role of extreme heterogeneity is rarely assessed or even explored, it is well-known that observations in the tails of distributions of attribute coefficients may exert undue influence on estimates (Kanninen, 1995; Lewbel, 1997) and ultimately on the estimator's numerical performance. They may also affect the estimated asymptotic standard errors, with consequences for inference, interpretation of statistical significance and hypotheses testing (Louviere *et al.*, 2003). Finally, observations of outlying coefficient values are behaviourally relevant because they may provide evidence that not all individuals conform to established conceptions of economic behaviour. The above concerns motivate the present study on how to appropriately identify and accommodate outlying taste-intensities in discrete choice analysis.

In fact, when dealing with outliers, their identification represents only a first step. After ruling out the possibility of outliers being due to coding errors, the second step relates to how to adequately handle their presence in the sample. Assessing the legitimacy of outliers requires a potentially controversial judgement and needs to be evaluated on a case-by-case basis. The option of removing them from the active sample used in estimation is prone to criticism. Therefore, improvements to the model specification that can accommodate them should first be sought, whenever possible (Ben-Akiva and Lerman, 1985; Sælensminde, 2006).

Irrespective of whether or not extreme values are genuine preference statements, their presence suggests that model specifications relying on preference heterogeneity are more appropriate. Interacting observed respondent-specific variables with the attributes provides a first step to uncovering the 'observed' component of heterogeneity. 'Unobserved' preference heterogeneity, on the other hand,

[1] For an analysis and comparison of choice prediction and coefficient outliers see Campbell and Hess (2009).

can be accommodated by treating preferences as varying according to some distributional assumption. However, the presence of extreme coefficient values raises potential problems when deciding an appropriate mixing distribution (see Hess and Axhausen, 2005).

We illustrate the impacts of a small number of respondents with extreme sensitivities on sample population level estimates of heterogeneity with the following example. Let us consider a population in which 95 percent of respondents have a cost coefficient (β) distributed $\mathcal{N}(-0.5, 0.1)$. Assume further that the remaining 5 percent of respondents form two equally sized outlying groups, one with very low, the other with very high cost sensitivity. The low cost sensitivity group has the cost coefficient distributed $\mathcal{N}(-0.1, 0.05)$ while the high cost sensitivity group are distributed $\mathcal{N}(-1.5, 0.1)$. This leads to a true distribution with three separate modes. However, erroneously fitting a single Normal distribution to this population implies $\mathcal{N}(-0.516, 0.196)$, which exaggerates the density of respondents in the intervals between the highest mode and the two smaller modes of the outlying groups, as shown in Figure 1. These problems are likely to also arise with other typically used continuous distributions, which, with a few exceptions, are unimodal and make strict shape assumptions. It is important to note that the degree of heterogeneity (as defined by standard deviation) in the two distributions in the

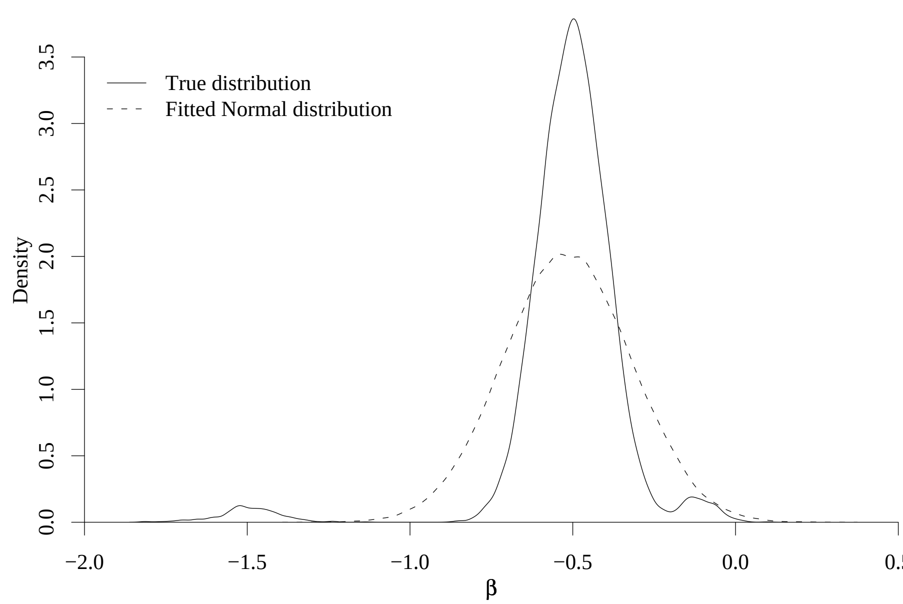

Figure 1: Example distribution affected by extreme sensitivities

example is identical; the difference lies in the shape of the distribution, and in the fact that the *fitted* distribution incorrectly assigns a large amount of probability between the middle mode and the two outlying groups.

For the above reasons, in datasets where coefficient outliers are expected, the use of discrete mixing distributions may provide a more realistic and flexible representation of the underlying form of unobserved preference heterogeneity. Indeed, in a discrete mixture environment, each coefficient takes on a finite number of different values, with no *a priori* assumptions on the shape of the resulting distribution. In this paper, we assess the role of extreme coefficient values in discrete choice analysis and the relative aptness of continuous and discrete mixture distributions to identify and accommodate them. Within our discrete mixture approach, we give particular emphasis to explaining the extreme lower and upper elements of the distribution (i.e., the 'outliers'). In order to do so in estimation we specify three mass points for each coefficient.

We compare this modelling approach with a standard MNL model and a mixed MNL (MMNL) model, where continuous Normal distributions are assumed. In addition, we use the means of the conditional distributions (from the continuous MMNL) of each respondent in the sample to assign them to either a lower, middle or upper class, and estimate models with separate coefficients for the three classes. To test our methodology, we use four simulated datasets and a stated choice dataset designed to determine the existence value of a number of rare and endangered fish species in Ireland.

The remainder of the paper is structured as follows: Section 2 describes our econometric approach; Section 3 outlines the simulated data and its results; our empirical case study is presented in Section 4 along with the results; and Section 5 provides a discussion and conclusion.

2 Methodology

Starting with the conventional specification of utility, where respondents are indexed by n, preferred alternatives by i, choice occasions by t and the vector of attributes are represented by x, we have:

$$U_{nit} = \beta_n' x_{nit} + \epsilon_{nit}, \tag{1}$$

where β are parameters to be estimated, ϵ is an *iid* Gumbel distributed error term, with constant variance $\pi^2/6$, and we assumed a linear-in-parameters specification of the indirect utility. Since we are interested in identifying respondents with outlying values of taste-intensity we treat the vector β_n as random across respondents, denoted by n. In this case the vector β_n is fixed over the sequence of t choices by each respondent. Denoting the respondent's chosen alternative in choice occasion t as y_{nt} and their sequence of choices over the T_n choice occasions as $y_n = \langle y_{n1}, y_{n2}, \ldots, y_{nT_n} \rangle$, then, conditional on β_n, the probability of respondent n's sequence of choices is the product of logit formulas:

$$\text{Prob}\,(y_n|\beta_n, x_n) = \prod_{t=1}^{T_n} \frac{\exp\left(V_{ny_{nt}t}\,(\beta_n)\right)}{\sum_{j=1}^{J} \exp\left(V_{njt}\,(\beta_n)\right)}, \tag{2}$$

where $V_{nit} = \beta_n' x_{nit}$. Further, we denote attributes by k and seek to identify and accommodate the extreme lower and upper (i.e., outlying) values of β_{nk}. To assess the aptness of the mixing distributions used to uncover these outliers, we compare discrete and continuous distributions.

2.1 Discrete mixture approach

In a discrete mixture (henceforth abbreviated as DM) context, the number of possible values for the parameter coefficients is finite (e.g., see Hess *et al.*, 2007; Train, 2008, for a review). To address coefficient outliers, we place particular emphasis on explaining the extreme lower and upper elements of the distribution. In estimation, we impose each of the K random parameters to take three mass points—the first ($\beta_{nk,L}$) and third ($\beta_{nk,H}$) of which are respectively associated with lower and upper outlying parameters, whilst the middle support point ($\beta_{nk,M}$) represents the intermediate and most common preference intensities (i.e., different from the outlying minorities). The mass points for the lower, middle and upper classes are associated with the probabilities $\pi_{k,L}$, $\pi_{k,M}$ and $\pi_{k,H}$ respectively.

The overall choice probability is given by:

$$\text{Prob}\,(y_n|\beta_n, x_n) = \prod_{t=1}^{T_n} \left[\sum_{s=1}^{S} \omega_s \frac{\exp\left(V_{ny_{nt}t}\,(\beta_{n,s})\right)}{\sum_{j=1}^{J} \exp\left(V_{njt}\,(\beta_{n,s})\right)} \right], \tag{3}$$

where $s = 1, ..., S$ is an index over all possible combinations of values for the K taste coefficients given their three values each (i.e., $S = 3^K$). As an example, with two coefficients, say α and γ, we would have that $S = 9$, and with $s = 1$ relating to the case where the lower values are used for both coefficients, we would have that $\omega_1 = \pi_{\alpha_L}\pi_{\gamma_L}$ and $\beta_1 = (\alpha_L, \gamma_L)$.

Unlike the approach used in Train (2008), where the densities are estimated for fixed parameters, we fix the values of $\pi_{k,L}$, $\pi_{k,M}$ and $\pi_{k,H}$ and estimate the coefficients. Imposing an ordering restriction (i.e., $\beta_{nk,L} \leq \beta_{nk,M} \leq \beta_{nk,H}$) to ensure monotonicity, provides a means for identifying coefficients at different points in the distribution. Given our interest in the extreme lower and upper ends of the distribution we set the values of $\pi_{k,L}$ and $\pi_{k,H}$ to be relatively small.

2.2 Continuous mixture approach

When using continuous mixing distributions, we assume that the set of K random parameters β_{nk} are Normally distributed (i.e., $\beta_{nk} \sim \mathcal{N}\left(\mu_k, \sigma_k^2\right)$). Dropping the subscript k to unclutter notation, we note that as β_n is random, the unconditional choice probability is obtained by integrating the logit probability $L(y_n|\beta_n, x_n)$ over the range of values for β_n, using as weights the draws from the assumed distribution $f(\beta|\mu, \Omega)$, where μ and Ω are the vectors of population hyper-parameters:

$$\text{Prob}(y_n|\mu, \Omega, x_n) = \int_{\beta_n} L(y_n|\beta_n, x_n) f(\beta_n|\mu, \Omega) d\beta_n. \tag{4}$$

In this paper we use a two-stage estimation approach. First we estimate the population vector of means $\hat{\mu}$ and the associated variance-covariance matrix $\hat{\Omega}$. Secondly, we derive $\hat{\mu}_n$ conditional on each respondent's pattern of observed choice $y_{1n}, y_{2n}, \cdots, y_{Tn}$ and choice-task attributes $x_{1n}, x_{2n}, \cdots, x_{Tn}$ for each of the K random coefficients. Then we use the means of these respondent-specific distributions to assign each respondent to either an outlier class (lower or upper) or to the main middle class. While disregarding the fact that these conditional parameters themselves follow a random distribution around this mean, this approach nevertheless gives us some information about the likely position of a respondent on the distribution. It further allows a rudimentary assessment of the presence of extreme sensitivities, and the range of these sensitivities. The work in this paper is exploratory and is only a first step in the direction of an appropriate treatment of coefficient outliers. An important area for future work is to use the entire conditional distribution in assigning respondents to outlier groups, i.e., incorporating the uncertainty in the conditional distribution into the allocation process.

On the basis of this class partition we then proceed with a second stage estimation to obtain separate parameters for the three classes (i.e., $\beta_{nk,L}$, $\beta_{nk,M}$ and

$\beta_{nk,H}$). This is achieved by specifying each attribute parameter as a function of respondent-specific dummy variables:

$$U_{nit} = \left(\delta'_{n,L}\beta_{n,L}\right)' x_{nit} + \left(\delta'_{n,M}\beta_{n,M}\right)' x_{nit} + \left(\delta_{n,H}\beta_{n,H}\right)' x_{nit} + \epsilon_{nit}, \tag{5}$$

where $\delta_{n,L}$, $\delta_{n,M}$ and $\delta_{n,H}$ are the dummy variables, which are specified independently for each coefficient as follows:

$$\delta_{n,L} = \begin{cases} 1 & \text{if the respondent's conditional mean is a lower outlier;} \\ 0 & \text{if otherwise.} \end{cases}$$
$$\delta_{n,M} = \begin{cases} 1 & \text{if the respondent's conditional mean is not an outlier;} \\ 0 & \text{if otherwise.} \end{cases}$$
$$\delta_{n,H} = \begin{cases} 1 & \text{if the respondent's conditional mean is an upper outlier;} \\ 0 & \text{if otherwise.} \end{cases}$$

In estimation the choice probabilities for the MMNL models are approximated by simulating the sample log-likelihood with pseudo-random draws.

We recognise that the goal of MMNL estimation is to obtain population estimates of distribution parameters, yet propose a method to assess their sensitivity to a differential treatment of those few sample respondents whose panel of choices reveals conditional distribution of taste which are anomalous. The method is useful for different reasons. Firstly, it should help us identify respondents who are more likely to have extreme sensitivities. Secondly, the second stage estimation allows us to evaluate the model performance in the presence of a specific treatment of these outlying sensitivities. Thirdly, it allows us to judge whether using a special treatment for these respondents leads to a significant change (and, arguably, correction) in the shape of the retrieved distribution, compared to the base specification. Finally, as we will see later in the paper, it helps us assess issues of confounding under the DM approach.

3 Simulated data experiments

This section of the paper discusses the results of four simulated data experiments. Each experiment makes use of the same underlying data, but with different assumptions on the choice data generating processes (DGPs). The underlying data is based on an orthogonal design, with two alternatives, each described by travel time (TT) and travel cost (TC). A sample of 720 respondents was used, with ten choice sets per respondent.

In each case study, six models were estimated. The first four models make use of the data in its original form, with the following specifications being used:

MNL Standard multinomial logit (MNL) model, with two marginal utility parameters, $\beta_{\text{TT,M}}$ and $\beta_{\text{TC,M}}$;

DM$_1$ Discrete mixture (DM) model, with three support points each for the TT and TC coefficients, identified as $\beta_{\text{T}\cdot,\text{L}}$, $\beta_{\text{T}\cdot,\text{M}}$ and $\beta_{\text{T}\cdot,\text{H}}$, with associated probabilities of $\pi_{\text{T}\cdot,\text{L}}$, $\pi_{\text{T}\cdot,\text{M}}$ and $\pi_{\text{T}\cdot,\text{H}}$, fixed to values of 0.025, 0.95 and 0.025 respectively. An order constraint was imposed on $\beta_{\text{T}\cdot,\text{L}}$, $\beta_{\text{T}\cdot,\text{M}}$ and $\beta_{\text{T}\cdot,\text{H}}$;

DM$_2$ Like DM$_1$, but with estimated values for $\pi_{\text{T}\cdot,\text{L}}$, $\pi_{\text{T}\cdot,\text{M}}$ and $\pi_{\text{T}\cdot,\text{H}}$ to show the effect of the restricting the probabilities for the three support points in DM$_1$; and

MMNL Standard mixed multinomial logit (MMNL) model, with univariate Normal distributions used for the two coefficients, leading to the estimation of mean coefficients $\beta_{\text{T}\cdot,\text{M}}$ and standard deviations $\sigma_{\text{T}\cdot,\text{M}}$[2].

The two DM models and the MMNL model were estimated with consideration to the repeated choice nature of the data, with variation in tastes across respondents, but not across choices for the same respondent.

After estimation of the MMNL model, individual-specific parameter distributions were generated for the two marginal utility coefficients, and the means of these distributions were used to assign each individual to one of three classes for each of the two coefficients, namely a *middle* class with a weight of at least 95%, and a *lower* and *upper* class with maximum weights of 2.5%[3]. On the basis of this, two additional models were estimated:

MNL$_C$ MNL model making use of separate coefficients in the *lower* (L), *middle* (M) and *higher* (H) group; and

MMNL$_C$ MMNL model, with Normal distributions used for the separate coefficients in the L, M and H group.

[2]The use of univariate Normal distributions is justified given the assumption of independence between coefficients in the DGPs.

[3]The use of strict $2.5\%, 95\% and 2.5\%$ splits was not always possible due to some respondents having the same mean values for the conditional distributions.

3.1 Simulated data case study I

Table 1 reports the details of the DGP and the estimation results for the first simulated dataset. The DGP for this experiment splits respondents into three groups, with no additional heterogeneity within groups. For each attribute, all respondents are simulated as having one of three distinct values. The three values represent in growing order: (*a*) the lower tail outlier, (*b*) a non-outlier parameter value in the middle class, and (*c*) the upper tail outlier. Looking at the output from DM_1 we find that, with the possible exception of the lower tail outlier for TC, the means are sufficiently close to those used to simulate the data. DM_2 also retrieves the correct means and, importantly, the estimated values of π are not dissimilar to those used to generate the data, and no major change in model fit is observed.

Interestingly, while the means of the random parameters in the MMNL model are close to the support points specified for the middle group, the standard deviations are significantly different from zero. This indicates that the MMNL model retrieves random taste heterogeneity as a result of the presence of extreme sensitivities. With the use of standard continuous distributions, this would lead to incorrect patterns of heterogeneity in between the three distinct groups used in the DGP. Furthermore, in terms of model fit, the MMNL does not perform as well as any of the DM models.

The two models with conditional segmentation provide superior fits. Both of these models retrieve correct mean values, but the problems associated with the lower outlier for the TC attribute remain. We speculate this might be due to the small size of this group and the fact that the value we used in the data generation process was quite extreme. While, for this parameter, preference heterogeneity is retrieved by the $MMNL_C$ specification, it is reassuring that the coefficient of variation is very small. Notwithstanding this idiosyncrasy (and the data construct), the fact that the estimated standard deviations are insignificant when outliers are accommodated in estimation provides further evidence that the presence of just a handful of respondents with lower (N=15 and N=6 for TT and TC respectively) and upper (N=18 for both TT and TC) outliers can result in retrieving significant patterns of preference heterogeneity which may be misinterpreted depending on the distributional assumptions.

3.2 Simulated data case study II

The DGP and results for the second study on simulated data are given in Table 2. This data was simulated based on the same mean values and weights for outliers

Table 1: Results for first study on simulated data

	Simulation	MNL		DM_1		DM_2		MMNL		MNL_C		$MMNL_C$	
	settings	est.	|t-rat.|	est.	|t-rat.|	est.	|t-rat.|	est.	|t-rat.|	est.	|t-rat.|	est.	|t-rat.|
$\pi_{TT,L}$	0.025	-	-	0.025	-	0.038	0.57	-	-	-	-	-	-
$\beta_{TT,L}$	-0.800	-	-	-0.559	2.11	-0.427	1.18	-	-	-0.750	9.54	-0.750	9.53
$\sigma_{TT,L}$	0.000	-	-	-	-	-	-	-	-	-	-	0.001	0.30
$\pi_{TT,M}$ †	0.950	-	-	0.950	-	0.923	1.18	-	-	-	-	-	-
$\beta_{TT,M}$	-0.150	-0.135	34.80	-0.155	32.87	-0.156	18.32	-0.163	28.89	-0.155	36.03	-0.155	35.41
$\sigma_{TT,M}$	0.000	-	-	-	-	-	-	0.078	10.21	-	-	0.004	0.38
$\pi_{TT,H}$	0.025	-	-	0.025	-	0.039	3.75	-	-	-	-	-	-
$\beta_{TT,H}$	0.100	-	-	0.101	5.81	0.087	3.70	-	-	0.114	6.36	0.114	7.87
$\sigma_{TT,H}$	0.000	-	-	-	-	-	-	-	-	-	-	0.000	0.47
$\pi_{TC,L}$	0.025	-	-	0.025	-	0.025	2.09	-	-	-	-	-	-
$\beta_{TC,L}$	-4.000	-	-	-10.600	2.45	-8.960	0.91	-	-	-28.200	113.69	-33.300	50.59
$\sigma_{TC,L}$	0.000	-	-	-	-	-	-	-	-	-	-	0.310	6.79
$\pi_{TC,M}$ †	0.950	-	-	0.950	-	0.942	3.80	-	-	-	-	-	-
$\beta_{TC,M}$	-0.750	-0.668	35.75	-0.749	33.49	-0.759	28.46	-0.796	29.28	-0.766	36.65	-0.766	36.84
$\sigma_{TC,M}$	0.000	-	-	-	-	-	-	0.353	8.68	-	-	0.011	0.20
$\pi_{TC,H}$	0.025	-	-	0.025	-	0.034	2.96	-	-	-	-	-	-
$\beta_{TC,H}$	0.500	-	-	0.536	4.40	0.453	2.93	-	-	0.673	6.48	0.673	8.50
$\sigma_{TC,H}$	0.000	-	-	-	-	-	-	-	-	-	-	0.001	0.31
$N_{TT,L}$ ‡		-		-		-		-		15		15	
$N_{TT,M}$		-		-		-		-		687		687	
$N_{TT,H}$		-		-		-		-		18		18	
$N_{TC,L}$		-		-		-		-		6		6	
$N_{TC,M}$		-		-		-		-		696		696	
$N_{TC,H}$		-		-		-		-		18		18	
Final LL		-3,259.04		-3,158.32		-3,156.68		-3,188.65		-2,912.88		-2,912.87	
par.		2		6		10		4		6		12	
adj. ρ^2		0.347		0.366		0.366		0.36		0.415		0.414	

† t-ratio calculated with respect to a base value of 1

‡ N denotes the number of respondents with the separate coefficient

Table 2: Results for second study on simulated data

	Simulation settings	MNL est.	MNL |t-rat.|	DM$_1$ est.	DM$_1$ |t-rat.|	DM$_2$ est.	DM$_2$ |t-rat.|	MMNL est.	MMNL |t-rat.|	MNL$_C$ est.	MNL$_C$ |t-rat.|	MMNL$_C$ est.	MMNL$_C$ |t-rat.|
$\pi_{TT,L}$	0.025	-	-	0.025	-	0.084	2.30	-	-	-	-	-	-
$\beta_{TT,L}$	-0.800	-	-	-2.560	4.19	-0.519	3.48	-	-	-0.626	10.83	-0.701	11.25
$\sigma_{TT,L}$	0.000	-	-	-	-	-	-	-	-	-	-	0.003	0.48
$\pi_{TT,M}$†	0.950	-	-	0.950	-	0.853	4.12	-	-	-	-	-	-
$\beta_{TT,M}$	-0.150	-0.121	33.72	-0.141	29.61	-0.143	18.62	-0.155	26.34	-0.136	34.78	-0.149	28.49
$\sigma_{TT,M}$	0.050	-	-	-	-	-	-	0.089	12.09	-	-	0.051	6.90
$\pi_{TT,H}$	0.025	-	-	0.025	-	0.063	4.01	-	-	-	-	-	-
$\beta_{TT,H}$	0.100	-	-	0.110	7.94	0.076	3.35	-	-	0.136	7.03	0.142	11.44
$\sigma_{TT,H}$	0.000	-	-	-	-	-	-	-	-	-	-	0.000	0.15
$\pi_{TC,L}$	0.025	-	-	0.025	-	0.154	1.44	-	-	-	-	-	-
$\beta_{TC,L}$	-4.000	-	-	-3.450	4.92	-1.700	3.11	-	-	-28.700	92.01	-32.500	33.50
$\sigma_{TC,L}$	0.000	-	-	-	-	-	-	-	-	-	-	0.390	0.02
$\pi_{TC,M}$†	0.950	-	-	0.950	-	0.779	2.21	-	-	-	-	-	-
$\beta_{TC,M}$	-0.750	-0.592	33.92	-0.694	28.64	-0.674	10.09	-0.768	26.54	-0.689	35.62	-0.767	28.84
$\sigma_{TC,M}$	0.300	-	-	-	-	-	-	0.448	12.51	-	-	0.287	8.64
$\pi_{TC,H}$	0.025	-	-	0.025	-	0.067	3.36	-	-	-	-	-	-
$\beta_{TC,H}$	0.500	-	-	0.492	6.04	0.325	2.78	-	-	0.676	7.15	0.701	8.68
$\sigma_{TC,H}$	0.000	-	-	-	-	-	-	-	-	-	-	0.014	0.41
$N_{TT,L}$‡		-		-		-		-		18		18	
$N_{TT,M}$		-		-		-		-		684		684	
$N_{TT,H}$		-		-		-		-		18		18	
$N_{TC,L}$		-		-		-		-		7		7	
$N_{TC,M}$		-		-		-		-		694		694	
$N_{TC,H}$		-		-		-		-		19		19	
Final LL		-3477.02		-3350.06		-3327.20		-3346.16		-3153.36		-3128.85	
par.		2		6		10		4		6		12	
adj. ρ^2		0.303		0.328		0.331		0.329		0.370		0.371	

† t-ratio calculated with respect to a base value of 1

‡ N denotes the number of respondents with the separate coefficient

as in the first simulated case study, but with heterogeneity around the means in the middle and most numerous (95%) group of respondents. Both DM models have superior fits compared to the standard MNL model and, with the exception of the lower outlier for TT, the retrieved mean values are relatively close to those of the true DGP. However, DM_2 picks up more outliers, especially for the lower TC parameter, which is potential evidence of confounding with the heterogeneity used for the middle group in the DGP. Additionally, DM_2 outperforms DM_1 in terms of model fit. We find that the MMNL model has inferior model fit *vis-à-vis* DM_2. It retrieves a significant amount of random taste heterogeneity, but the confounding between the heterogeneity in the middle group and the two outlying groups is evident in the inflated degree of heterogeneity. Again, the models with conditional segmentation provide superior model fits. While some problems remain with the lower outlier for TC, both models adequately retrieve the mean values. Furthermore, the $MMNL_C$ specification correctly uncovers random variation for the middle group only with the derived standard deviations being very close to those used in the DGP.

3.3 Simulated data case study III

In Table 3, we report the DGP and results for the third case study. This dataset was simulated on a similar basis as the second simulated case study, but with less extreme values for the lower outliers and with heterogeneity in all three groups (i.e., essentially a mixture of Normals). Aside from recurring issues associated with the lower outliers, the DM models recover support points that are not dissimilar to those used in the true DGP. Both DM models have better fits than the standard MNL model, but DM_2 does marginally better. This is a result of DM_2 having a more flexible specification, allowing it to deal with the heterogeneity within groups (again illustrating the confounding of heterogeneity within groups and the very presence of the extreme groups). In terms of model fit, while MMNL is found to outperform DM_1, it is inferior to DM_2. The means retrieved under the MMNL model are of a comparable magnitude to those specified for the middle group. While the MMNL specification correctly retrieves random variation, it is clearly once again affected by confounding between the two *types* of heterogeneity. The performance of the models based on conditional segmentation is again superior. The means recovered from both models are relatively close to those of the true DGP. However, we do note some inconsistencies in the standard deviations for the outlier groups, but the heterogeneity in the middle group is retrieved very accurately.

Table 3: Results for third study on simulated data

	Simulation settings	MNL		DM_1		DM_2		MMNL		MNL_C		$MMNL_C$	
		est.	|t-rat.|	est.	|t-rat.|	est.	|t-rat.|	est.	|t-rat.|	est.	|t-rat.|	est.	|t-rat.|
$\pi_{TT,L}$	0.025	-	-	0.025	-	0.128	0.74	-	-	-	-	-	-
$\beta_{TT,L}$	-0.400	-	-	-2.510	1.39	-0.496	2.72	-	-	-0.439	8.05	-0.579	4.98
$\sigma_{TT,L}$	0.200	-	-	-	-	-	-	-	-	-	-	0.622	2.13
$\pi_{TT,M}$ †	0.950	-	-	0.950	-	0.786	1.31	-	-	-	-	-	-
$\beta_{TT,M}$	-0.150	-0.124	34.09	-0.141	29.50	-0.145	17.01	-0.154	26.99	-0.136	35.07	-0.150	28.79
$\sigma_{TT,M}$	0.050	-	-	-	-	-	-	0.082	11.47	-	-	0.052	6.78
$\pi_{TT,H}$	0.025	-	-	0.025	-	0.086	3.08	-	-	-	-	-	-
$\beta_{TT,H}$	0.100	-	-	0.102	5.57	0.060	2.57	-	-	0.073	4.40	0.074	3.15
$\sigma_{TT,H}$	0.050	-	-	-	-	-	-	-	-	-	-	0.063	1.49
$\pi_{TC,L}$	0.025	-	-	0.025	-	0.070	1.69	-	-	-	-	-	-
$\beta_{TC,L}$	-2.000	-	-	-3.040	3.06	-1.480	2.12	-	-	-2.030	4.31	-2.750	2.83
$\sigma_{TC,L}$	0.500	-	-	-	-	-	-	-	-	-	-	1.140	2.22
$\pi_{TC,M}$ †	0.950	-	-	0.950	-	0.866	3.45	-	-	-	-	-	-
$\beta_{TC,M}$	-0.750	-0.608	34.46	-0.699	29.34	-0.718	8.20	-0.761	27.56	-0.679	35.49	-0.760	29.34
$\sigma_{TC,M}$	0.300	-	-	-	-	-	-	0.399	12.33	-	-	0.278	8.10
$\pi_{TC,H}$	0.025	-	-	0.025	-	0.064	3.45	-	-	-	-	-	-
$\beta_{TC,H}$	0.500	-	-	0.357	6.78	0.151	1.39	-	-	0.347	3.93	0.388	5.95
$\sigma_{TC,H}$	0.200	-	-	-	-	-	-	-	-	-	-	0.015	1.14
$N_{TT,L}$ ‡		-		-		-		-		15		15	
$N_{TT,M}$		-		-		-		-		687		687	
$N_{TT,H}$		-		-		-		-		18		18	
$N_{TC,L}$		-		-		-		-		6		6	
$N_{TC,M}$		-		-		-		-		696		696	
$N_{TC,H}$		-		-		-		-		18		18	
Final LL		-3,429.81		-3,334.14		-3,315.82		-3,330.08		-3,221.64		-3,191.68	
par.		2		6		10		4		6		12	
adj. ρ^2		0.312		0.331		0.334		0.332		0.353		0.358	

† *t*-ratio calculated with respect to a base value of 1

‡ N denotes the number of respondents with the separate coefficient

3.4 Simulated data case study IV

Table 4 reports the DGP and results for the fourth case study. In this case study the same means and spreads are used for the middle group as those used in the second and third case studies, but no explicit outliers are defined. Our motivation here is to examine the risk of confounding between outliers and heterogeneity under the DM approach. Both DM models have superior fits than the MNL model, and retrieve some of the heterogeneity used in data generation, with fits almost as high as the MMNL model. We note that the values found for the outliers in DM_1 are quite extreme, especially the lower outliers, given the values used for the Normal distribution, showing some risk of confounding[4]. Here, the estimation of the DM_2 model is an important check for analysts to carry out; this model produces a visibly less extreme distribution of sensitivities.

The MMNL model recovers means and standard deviations that are very similar to those used in the true DGP. While superior fits are achieved under MNL_C and $MMNL_C$, both models are found to have equivalent fits. These are also found to produce relatively extreme values given the true DGP, along with no random variation in the $MMNL_C$ model. These findings further highlight the potential issues with confounding and the difficulty of disentangling the extreme sensitivities from the more commonly assumed type of heterogeneity. Indeed, we acknowledge the fact that the $MMNL_C$ model in this case insinuates the presence of extreme outliers rather than standard random taste heterogeneity. In this context, using the two stage approach with different sizes for outlying groups may be a helpful way of assessing this *type* of heterogeneity.

4 Empirical case study

To illustrate the proposed methodology on an empirical case study we use stated preference data collected to estimate the existence value of a number of rare and endangered fish species in the Lough Melvin Catchment in Ireland. Lough Melvin is a freshwater lake in the North West of Ireland which straddles the border between the Republic of Ireland and Northern Ireland. With a unique population of native fish species, the Lough Melvin Catchment has an internationally important conservation status. Lough Melvin and its associated river system supports the only remaining population of Arctic char (AC) in Northern Ireland and contains Atlantic salmon (AS) and three genetically distinct populations of brown trout

[4]As an example, the true lower 2.5 percentile point for the TT coefficient is -0.248.

Table 4: Results for fourth study on simulated data

	Simulation settings	MNL		DM_1		DM_2		MMNL		MNL_C		$MMNL_C$	
		est.	|t-rat.|	est.	|t-rat.|	est.	|t-rat.|	est.	|t-rat.|	est.	|t-rat.|	est.	|t-rat.|
$\pi_{TT,L}$ †	-	-	-	0.025	-	0.893	1.57	-	-	-	-	-	-
$\beta_{TT,L}$	-	-	-	-0.714	3.23	-0.163	16.08	-	-	-0.634	11.69	-0.643	11.93
$\sigma_{TT,L}$	-	-	-	-	-	-	-	-	-	-	-	0.002	1.30
$\pi_{TT,M}$	1.000	-	-	0.950	-	0.107	1.57	-	-	-	-	-	-
$\beta_{TT,M}$	-0.150	-0.136	35.69	-0.145	31.62	-0.024	0.73	-0.154	29.12	-0.147	36.01	-0.148	35.74
$\sigma_{TT,M}$	0.050	-	-	-	-	-	-	0.060	8.78	-	-	0.003	0.16
$\pi_{TT,H}$	-	-	-	0.025	-	0.000	<0.01	-	-	-	-	-	-
$\beta_{TT,H}$	-	-	-	0.046	2.25	3.030	<0.01	-	-	0.072	3.86	0.072	6.32
$\sigma_{TT,H}$	-	-	-	-	-	-	-	-	-	-	-	0.000	1.67
$\pi_{TC,L}$ †	-	-	-	0.025	-	0.739	1.94	-	-	-	-	-	-
$\beta_{TC,L}$	-	-	-	-2.440	5.37	-0.888	10.85	-	-	-31.400	126.00	-31.300	12.70
$\sigma_{TC,L}$	-	-	-	-	-	-	-	-	-	-	-	0.151	0.01
$\pi_{TC,M}$	1.000	-	-	0.950	-	0.261	1.94	-	-	-	-	-	-
$\beta_{TC,M}$	-0.750	-0.667	36.05	-0.717	31.63	-0.317	2.68	-0.759	30.01	-0.742	36.49	-0.750	32.99
$\sigma_{TC,M}$	0.300	-	-	-	-	-	-	0.293	8.85	-	-	0.097	1.62
$\pi_{TC,H}$	-	-	-	0.025	-	0.000	<0.01	-	-	-	-	-	-
$\beta_{TC,H}$	-	-	-	0.118	1.91	14.700	<0.01	-	-	0.316	3.96	0.317	8.99
$\sigma_{TC,H}$	-	-	-	-	-	-	-	-	-	-	-	0.000	0.11
$N_{TT,L}$ ‡		-		-		-		-		18		18	
$N_{TT,M}$		-		-		-		-		684		684	
$N_{TT,H}$		-		-		-		-		18		18	
$N_{TC,L}$		-		-		-		-		11		11	
$N_{TC,M}$		-		-		-		-		690		690	
$N_{TC,H}$		-		-		-		-		19		19	
Final LL		-3,254.90		-3,222.44		-3,220.30		-3,219.07		-3,013.12		-3,012.82	
par.		2		6		10		4		6		12	
adj. ρ^2		0.347		0.353		0.353		0.354		0.395		0.394	

† t-ratio calculated with respect to a base value of 1

‡ N denotes the number of respondents with the separate coefficient

known as ferox (F), gillaroo (G) and sonaghan (S). Since the habitat of these fish populations is recognised as being vulnerable, there is a need to assess the extent to which the general public supports the prevention of their extinction.

The discrete choice experiment consisted of a panel of sixteen repeated choice sets. Each choice set outlined three possible outcomes. The first two outcomes—labelled as 'Option A' and 'Option B'—described the conservation status of each of the fish species after the implementation of two experimentally designed conservation schemes. At the end of these schemes, the fish species would either be 'Conserved' or 'Extinct'. While a particular scheme described under either 'Option A' or 'Option B' may have been unable to prevent some of the fish species from becoming extinct, they both ensured against the extinction of all fish species (i.e., at least one species was conserved under each scheme). The final outcome—labelled as 'Do Nothing'—showed the expected outcome if nothing was done to protect the fish species. In this case, the respondents were informed that all five fish species would become extinct. 'Option A' and 'Option B' were both described to respondents as available at a positive cost (CST). The payment vehicle used was the amount that they would personally have to pay per year—through an increase in their Income Tax and/or Value Added Tax contributions—to implement the scheme. The 'Do Nothing' (or status-quo) option had zero cost to the respondent.

The population of interest was the adult population of the Republic of Ireland and Northern Ireland. The study adopted a stratified random sample to reflect the geographic distribution of the adult population; the approximate rural/urban spilt; the approximate socio-economic status of the regional population; and the approximate gender and age profile of the populations within both jurisdictions. A final sample of 624 useable responses was obtained which, with each respondent answering 16 choice tasks, resulted in 9,984 observations for model estimation. The survey was administered via a 'paper and pencil' questionnaire by professional interviewers and the responses were manually entered into a data file[5].

4.1 Discrete mixtures modelling results

Table 5 reports the standard MNL model, with marginal utility parameters for the six attributes and an alternative specific constant for the status-quo option (SQ).

[5]While we acknowledge that this method of data collection can increase the risk of incorrect coding, every attempt was made to ensure that there were no errors in the recording and entry of respondent's answers. We are, therefore, confident that any extreme heterogeneity uncovered in estimation is not due to coding errors.

Table 5: MNL and discrete mixtures models for the empirical data

	MNL		$DM_{1\%}$		$DM_{2.5\%}$		$DM_{5\%}$	
	est.	|t-rat.|	est.	|t-rat.|	est.	|t-rat.|	est.	|t-rat.|
π_{L}	-	-	0.010	-	0.025	-	0.050	-
$\beta_{AC,L}$	-	-	-1.150	6.99	-0.999	7.24	-0.906	-6.72
$\beta_{AS,L}$	-	-	-1.370	3.60	-1.000	4.89	-0.792	6.20
$\beta_{F,L}$	-	-	-1.560	6.53	-1.210	6.25	-0.920	5.79
$\beta_{G,L}$	-	-	-1.020	6.25	-0.767	4.45	-0.553	3.59
$\beta_{S,L}$	-	-	-1.760	7.36	-1.460	5.96	-1.030	5.11
$\beta_{CST,L}$	-	-	-0.205	14.30	-0.205	14.22	-0.209	15.52
$\beta_{SQ,L}$	-	-	-28.400	37.98	-28.300	38.61	-28.200	41.03
π_{M}	-	-	0.980	-	0.950	-	0.900	-
$\beta_{AC,M}$	0.304	13.05	0.332	8.90	0.344	8.49	0.359	8.26
$\beta_{AS,M}$	0.642	24.29	0.681	12.62	0.693	11.77	0.711	11.32
$\beta_{F,M}$	0.297	12.57	0.382	11.31	0.400	10.59	0.428	9.88
$\beta_{G,M}$	0.355	15.32	0.526	14.24	0.555	13.43	0.582	12.64
$\beta_{S,M}$	0.508	20.11	0.484	10.15	0.504	9.49	0.529	9.01
$\beta_{CST,M}$	-0.018	-19.83	-0.013	5.46	-0.014	5.58	-0.015	-5.89
$\beta_{SQ,M}$	-0.974	-20.59	-2.340	14.05	-2.470	9.25	-2.810	-8.15
π_{H}	-	-	0.010	-	0.025	-	0.050	-
$\beta_{AC,H}$	-	-	2.960	12.88	2.760	10.43	2.530	9.22
$\beta_{AS,H}$	-	-	4.370	12.83	4.100	16.06	3.920	19.13
$\beta_{F,H}$	-	-	4.150	5.59	1.960	1.13	1.350	3.48
$\beta_{G,H}$	-	-	1.860	16.24	1.810	14.8	1.720	13.28
$\beta_{S,H}$	-	-	4.620	16.23	4.460	16.19	4.350	16.24
$\beta_{CST,H}$	-	-	0.063	8.03	0.062	8.14	0.059	8.21
$\beta_{SQ,H}$	-	-	1.890	5.36	1.500	3.54	1.030	3.45
Final LL	-8,727.64		- 6,930.94		- 6,658.60		- 6,442.65	
par.	7		21		21		21	
adj. ρ^2	0.204		0.366		0.391		0.411	

This model is compared against three DM models, where the probabilities of the support points for lower and upper coefficients are fixed. As a test of sensitivity, a series of models with different predefined densities are estimated:

$\mathbf{DM}_{1\%}$ Lower and upper outliers fixed to 0.01;

$\mathbf{DM}_{2.5\%}$ Lower and upper outliers fixed to 0.025; and

$\mathbf{DM}_{5\%}$ Lower and upper outliers fixed to 0.05.

The coefficients estimates for the middle group are close to those from the MNL model. We find that the DM models produce extreme values at the lower and upper support points, possibly providing evidence of coefficient outliers. We remark, however, in line with expectations, that as one moves from $\mathrm{DM}_{1\%}$ to $\mathrm{DM}_{5\%}$ the support points of the outliers become less extreme, although for some of the parameters, the support points remain relatively close to the original values. Additionally, as we increase the size of the outlying groups, we observe an increase in model performance. This suggests that the models with larger groups offer a better representation of the heterogeneity, possibly caused by a large degree of heterogeneity in the data. The outlying values found for the lower and upper groups would, however, suggest that some of the retrieved heterogeneity is caused by extreme sensitivities rather than the more standard type of heterogeneity. Indeed, the DM models could be seen to be picking up *both* coefficient outliers *and* more standard type of preference heterogeneity, and as the size of the outlying groups increases, more of the standard variation is explained, resulting in an improvement in model fit.

A further point needs noting. All DM models suggest the presence of marginal estimates that are intuitively inconsistent with rational economic behaviour. Indeed, the marginal utility of increases in species conservation and cost would be expected to be positive and negative respectively, and this is not always the case. We note that these counter intuitive preferences are directly derived from the data, which we argue, especially the *evidence* of positive cost sensitivity, may be due to the effects of 'warm-glow' (i.e., yeah saying).

4.2 MMNL modelling results

In estimation, we replicate the approach used for the simulated data and estimate a panel MMNL model in which all parameters are assumed random. After testing various distributional assumptions we settled on multivariate Normal distributions[6].

The MMNL model (reported in Table 6) is estimated by simulating the log-likelihood with 500 pseudo-random draws. The estimates for the means of the random parameters are statistically significant and of the expected sign. The MMNL model provides evidence of heterogeneity in the taste-intensities across the sample of respondents for the random parameters. We also observe that the MMNL model gives a higher model fit than the DM models, even after taking into account the higher number of parameters.

Individual-specific parameter distributions of the MMNL model were retrieved and the kernel densities for the conditional means were produced, where Figure 2 illustrates this process for the gillaroo (G) attribute[7]. As observed earlier in this section, using a DM approach puts the analyst at risk of confounding between outliers and heterogeneity. For example, if the true distribution has a high variance, the mass points representing the lower and upper limits of the distribution may be more extreme than the true limits. It is hence of interest to compare the support points of the DM distributions against the continuous distribution. So, we include

[6]While we recognise that we did not allow for correlation amongst the coefficients in the DM models, our decision to allow for correlated coefficients in the MMNL specifications was based on the fact that they afforded much superior model fits, from which we could retrieve more reliable conditional distributions. In contrast, allowing for correlation amongst the coefficients in the DM models is not straightforward.

[7] Distributions for the remaining attributes are available from the lead author.

Table 6: Base model for conditional distributions

	μ			σ	
	est.	$\|t$-rat.$\|$		est.	$\|t$-rat.$\|$
AC	0.525	10.68		0.763	12.53
AS	1.029	14.58		1.229	16.96
F	0.513	11.65		0.516	9.17
G	0.641	14.38		0.592	8.18
S	0.935	12.65		1.44	16.88
CST	-0.042	12.41		0.076	17.38
SQ	-3.588	12.68		2.948	14.12
Final LL			-6180.64		
par.			35		
adj. ρ^2			0.436		

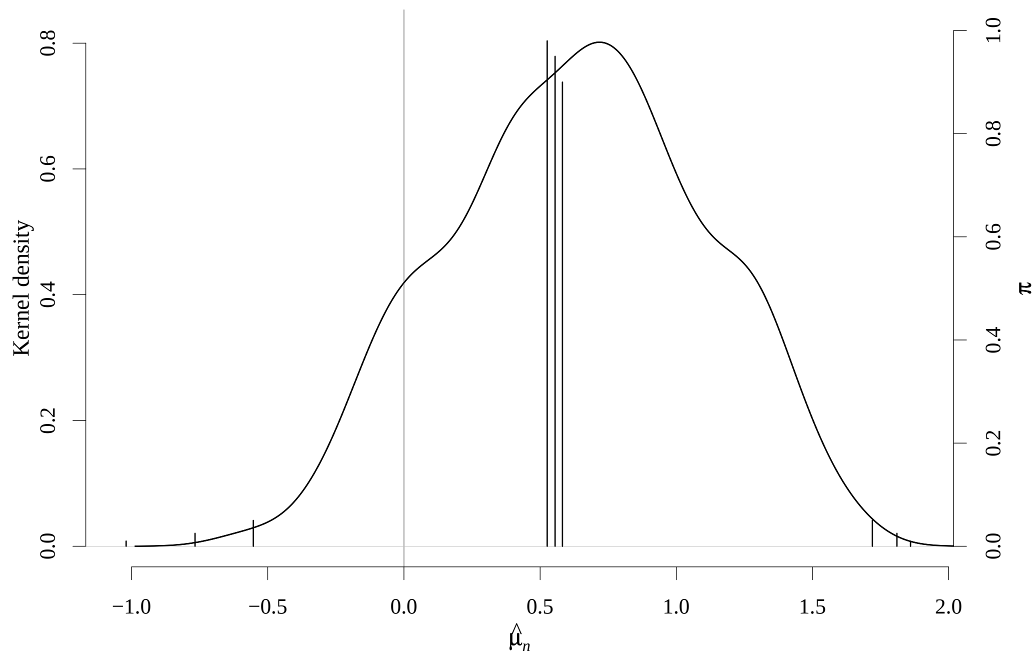

Figure 2: Conditional distributions of $\hat{\mu}_{n\mathrm{G}}$ and the support points retrived under the DM models

a secondary y-axis in Figure 2 showing the probability of the support points of the three DM models. To illustrate, the rightmost vertical mark has a value of -1.02 and probability of 0.01, representing the estimated lower outlier for the G attribute in $\mathrm{DM}_{1\%}$. Examination of the location of the lower and upper support points highlights that they generally lie within the extreme tails of the distribution. This is an important finding as it indicates that our DM modelling approach identifies extreme values that are consistent with those approximated using continuous Normal distributions. We do note, however, that in a few cases the DM support points are more extreme than those produced under the continuous distribution.

The means of the individual-specific parameter distributions obtained from the MMNL model were used to assign respondents to either a lower, middle or upper class. On the basis of this, further MMNL models were estimated, with separate means and spreads for the lower, middle and upper classes. Again, to tease out the influence of coefficient outliers at various levels, we estimated three models, each with a different proportion of outliers. For consistency, the percentiles used for classifying outlying respondents corresponded with those used in the DM models in Table 5.

The conditioned MMNL models specified that all taste-parameters are distributed independent Normal and probabilities of choice were simulated in estimation with 500 pseudo-random draws[8]. Results from these are presented in Table 7. In line with *a priori* expectations, we find that the magnitude of the estimated coefficients in the outlying classes generally becomes less extreme as class sizes increase, where the larger classes increasingly capture heterogeneity from the middle group. Nevertheless, we find further evidence that these are quite extreme, and in some cases theoretically inconsistent. Parameter estimates for the middle class are relatively stable across the six models. There is little evidence of heterogeneity amongst outliers, especially in the lower group. But, as expected, the degree of heterogeneity in taste intensities amongst outliers does slightly increase as the outlying groups increases in size, again given that these larger classes increasingly capture heterogeneity from the middle group. Correspondingly, there is a general decline in the coefficients of variation amongst the middle group as one moves from $\text{MMNL}_{\text{C},1\%}$ to $\text{MMNL}_{\text{C},5\%}$—indicating relatively smaller heterogeneity.

We find that using the conditional means to assign separate parameters for lower, middle and upper classes leads to better model fits, which is to be expected. Further improvements in model fit are achieved as the size of lower and upper groups increases, where this is again a result of capturing more heterogeneity.

An examination of the conditional means obtained from the estimates in Table 7 reveals that as more respondents are assigned to outlier groups, the distributions become increasingly different from those produced by the base MMNL model. As expected, we observe that the density of the outliers increases as more respondents are assigned as having extreme values. Associated with this finding is a general move from unimodal distributions under $\text{MMNL}_{\text{C},1\%}$ to increasingly apparent tri-modal distributions under $\text{MMNL}_{\text{C},5\%}$. Moreover, as the proportion of outliers increases, the position of the outliers become considerably less extreme, as they converge towards the median. While the location of conditional means for the non-outlying group do not vary much, it is clearly evident that the peaks become much more pronounced, given that the two outlying groups now capture more of the heterogeneity. This is consistent with the earlier finding that the extent of heterogeneity declines for this group as the outlying groups get larger. Importantly, the proportion of the distribution in the intuitively inconsistent domain diminishes as one moves from the conditional means obtained from $\text{MMNL}_{\text{C},1\%}$ to

[8]Multivariate distributions were not used in these models since three coefficients were estimated for each attribute and is not possible to allow for correlation between these.

Table 7: Models taking conditional distributions into account

	$MMNL_{C,1\%}$		$MMNL_{C,2.5\%}$		$MMNL_{C,5\%}$	
	est.	|*t*-rat.|	est.	|*t*-rat.|	est.	|*t*-rat.|
$\beta_{AC,L}$	-1.513	3.86	-1.43	2.57	-1.203	7.22
$\sigma_{AC,L}$	0.205	0.38	0.087	0.07	0.034	0.15
$\beta_{AS,L}$	-1.429	3.37	-1.444	3.91	-1.245	7
$\sigma_{AS,L}$	0	<0.01	0.097	0.07	0.076	0.34
$\beta_{F,L}$	-0.598	1.86	-0.252	0.78	-0.794	4.77
$\sigma_{F,L}$	0.202	0.46	0.045	0.04	0.243	0.9
$\beta_{G,L}$	-0.53	1.26	0.018	0.07	-0.64	3.22
$\sigma_{G,L}$	0.467	0.79	0.236	0.22	0.445	1.11
$\beta_{S,L}$	-2.105	4.12	-1.985	4.69	-1.611	8.3
$\sigma_{S,L}$	0.157	0.25	0.012	<0.01	0.012	0.04
$\beta_{CST,L}$	-0.593	5.09	-0.433	6.04	-0.404	11.38
$\sigma_{CST,L}$	0.183	1.7	0.119	0.97	0.067	1.55
$\beta_{SQ,L}$	-56.028	<0.01	-27.059	<0.01	-26.393	<0.01
$\sigma_{SQ,L}$	0	<0.01	0	<0.01	0	<0.01
$\beta_{AC,M}$	0.467	10.33	0.397	8.87	0.486	11.32
$\sigma_{AC,M}$	0.611	10.54	0.581	11.05	0.463	7.69
$\beta_{AS,M}$	1.073	15.9	1.43	19.75	1.032	18.1
$\sigma_{AS,M}$	1.179	16.8	1.363	18.79	0.808	13.22
$\beta_{F,M}$	0.484	11.59	0.425	9.75	0.477	11.75
$\sigma_{F,M}$	0.42	6.32	0.48	7.22	0.328	4.73
$\beta_{G,M}$	0.643	14.85	0.474	7.79	0.641	15.68
$\sigma_{G,M}$	0.45	6.8	0.82	11.09	0.289	3.42
$\beta_{S,M}$	-0.856	13.05	0.718	11.93	0.823	14.24
$\sigma_{S,M}$	1.249	18.15	1.043	18.85	0.925	15.5
$\beta_{CST,M}$	-0.04	11.27	-0.049	15.33	-0.041	12.75
$\sigma_{CST,M}$	0.081	20.17	0.075	23.87	0.062	20.19
$\beta_{SQ,M}$	-3.054	16.52	-3.187	22.44	-3.2	17.66
$\sigma_{SQ,M}$	2.176	16.13	1.533	16.85	1.667	12.67
$\beta_{AC,H}$	2.259	4.78	2.285	6.07	2.213	9.31
$\sigma_{AC,H}$	0.092	0.18	0.06	0.02	0.584	2.05
$\beta_{AS,H}$	8.723	2.54	5.452	9.24	7.935	6.34
$\sigma_{AS,H}$	0.904	0.22	0.001	<0.01	2.828	2.9
$\beta_{F,H}$	1.736	4.18	1.85	4.27	1.763	8.99
$\sigma_{F,H}$	0.056	0.13	0.066	0.03	0.014	0.05
$\beta_{G,H}$	2.357	5.01	2.496	5.27	2.195	10.64
$\sigma_{G,H}$	0.052	0.1	0.033	0.04	0.08	0.29
$\beta_{S,H}$	69.631	<0.01	30.212	<0.01	8.076	5.1
$\sigma_{S,H}$	0	<0.01	0	<0.01	1.536	1.04
$\beta_{CST,H}$	1.799	7.21	1.042	8.53	0.079	7.04
$\sigma_{CST,H}$	0.065	0.31	2.412	34.45	0.008	0.36
$\beta_{SQ,H}$	-3.569	7.32	3.172	6.05	2.217	10.6
$\sigma_{SQ,H}$	0.019	0.03	0.067	0.03	0.348	0.89
Final LL	-6,060.55		-5,926.44		-5,445.38	
par.	42		42		42	
adj. ρ^2	0.446		0.459		0.503	

$\text{MMNL}_{\text{C},5\%}$. These findings further highlight the important role that a few extreme coefficient values can have on the degree of unobserved heterogeneity retrieved using MMNL models.

5 Conclusion

In this paper, we build on a surprisingly sparse literature regarding the role of coefficient outliers in discrete choice data. We explore two separate ways of representing random taste heterogeneity. In our first approach we assume that random taste heterogeneity for each coefficient can be adequately described by three discrete points: one for the majority of agents, while the other two for a minority taking up extreme values. To implement this partitioning we constrain the point estimates to be monotonic, so that we can derive lower, middle and upper estimates on the distribution. We further impose that the densities of the lower and upper classes are relatively small so that we capture the extreme lower and upper elements of the distribution. This approach is then compared with the common continuous mixed logit assumption where random parameters are continuously distributed. From these models we retrieve the individual-specific distributions and use the conditional means to assign the values of three dummy variables denoting either lower, middle or upper class values. By specifying the attribute parameters as a function of these dummy variables we estimate separate coefficients for the three classes. To test our methodology we first use four simulated datasets and then apply it on a stated choice dataset designed to estimate the existence value of a number of rare and endangered fish species in Ireland.

In our simulated datasets we find that the DM models retrieve the correct mean values and densities of the outliers used in the data generating processes for the simulations. We also find that in the presence of extreme coefficient values the MMNL models generally overstate the degree of random taste heterogeneity in the *middle* group. In other words, while the models may retrieve the correct overall degree of heterogeneity, they misrepresent the shape of the distribution and assign too much weight away from a central group. This finding suggests some caution when selecting an appropriate mixing distribution for datasets believed to include extreme coefficient values.

Throughout the paper, we acknowledge the fact that there may be issues of confounding between outliers and more standard heterogeneity, and that our DM approach is potentially more vulnerable to such confounding. Indeed, in our fourth simulated dataset, we demonstrate the difficultly in separating outliers from remaining heterogeneity. Nevertheless, we find that in our empirical application the outlying support points of the DM models are generally consistent with the outlying conditional means derived from the MMNL. This is reassuring, as it provides some evidence that the DM approach is relatively robust to this potential source of confounding in datasets where there are a handful of respondents with extreme sensitivities. In applying the approach to our empirical dataset we find further evidence that extreme coefficient outliers may lead to misleading findings in terms of the patterns of heterogeneity. We also find that accommodating these outliers in our empirical application reduces the proportion of extreme and intuitively inconsistent point estimates.

Whilst deciding on the legitimacy of coefficient outliers is a difficult judgement and is ultimately an empirical issue to be evaluated case-by-case, the fact that some of them are found to be behaviourally inconsistent with *a priori* expectations suggests some caution for analysts engaged in discrete choice modelling. Indeed, the findings indicate the importance of testing for the presence of extreme coefficient values and that their detection and evaluation should become part of the standard course of action in practice.

References

Ben-Akiva M. and Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand*. MIT Press, Massachusetts.

Campbell, D. and Hess, S. (2009). Outlying sensitivities in discrete choice data: causes, consequences and remedies, *European Transport Conference*, Amsterdam.

Daly, A. J. and Zachary, S. (1975). Commuters' value of time, *Local Government Operational Research Unit*, Report T55, Reading.

Hensher, D. A. and Greene, W. H. (2003). The mixed logit model: the state of practice, *Transportation* **30**: 133–176.

Hess, S. and Axhausen, K. W. (2005). Distributional assumptions in the representation of random taste heterogeneity, *Swiss Transport Research Conference*, Zurich.

Hess, S., Bierlaire, M. and Polak, J. W. (2007). A systematic comparison of continuous and discrete mixture models, *European Transport* **37**: 35–61.

Kanninen, B. J. (1995). Bias in discrete response contingent valuation, *Journal of Environmental Economics and Management* **28**: 114–125.

Lewbel, A. (1997). Semiparametric estimation of location and other discrete choice models, *Econometric Theory* **13**: 32–51.

Louviere, J. J., Hensher, D. A. and Swait, J. D. (2003). *Stated choice methods: analysis and application*. Cambridge University Press, Cambridge.

McFadden, D. and Train, K. E. (1998). Mixed MNL Models for discrete response, *Journal of Applied Econometrics* **15**: 447–470.

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

Sælensminde, K. (2006). Causes and consequences of lexicographic choices in stated choice studies, *Ecological Economics* **59**: 331–340.

Train, K. E. (1998). Recreation demand models with taste differences over people, *Land Economics* **74**: 185–194.

Train, K. E. (2008). EM Algorithms for nonparametric estimation of mixing distributions, *Journal of Choice Modelling* **1**: 40–69.