White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/91386/

Version: Accepted Version

Article:
Hess, S, Stathopoulos, A, Campbell, D et al. (2013) It's not that I don't care, I just don't care very much: Confounding between attribute non-attendance and taste heterogeneity. Transportation, 40 (3). pp. 583-607. ISSN: 0049-4488

https://doi.org/10.1007/s11116-012-9438-1

(c) 2012, Springer Science+Business Media, LLC. This is an author produced version of a paper published in Transportation. Uploaded in accordance with the publisher's self-archiving policy. The final publication is available at Springer via http://doi.org/10.1007/s11116-012-9438-1

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity

Stephane Hess* Amanda Stathopoulos† Danny Campbell‡ Vikki O'Neill§
Sebastian Caussade¶

August 30, 2012

Abstract

With the growing interest in the topic of attribute non-attendance, there is now widespread use of latent class (LC) structures aimed at capturing such behaviour, across a number of different fields. Specifically, these studies rely on a confirmatory LC model, using two separate values for each coefficient, one of which is fixed to zero while the other is estimated, and then use the obtained class probabilities as an indication of the degree of attribute non-attendance. In the present paper, we argue that this approach is in fact misguided, and that the results are likely to be affected by confounding with *regular* taste heterogeneity. We contrast the confirmatory model with an exploratory LC structure in which the values in both classes are estimated. We also put forward a combined latent class mixed logit model (LC-MMNL) which allows jointly for attribute non-attendance and for continuous taste heterogeneity. Across three separate case studies, the exploratory LC model clearly rejects the confirmatory LC approach and suggests that rates of non-attendance may be much lower than what is suggested by the standard model, or even zero. The combined LC-MMNL model similarly produces significant improvements in model fit, along with substantial reductions in the implied rate of attribute non-attendance, in some cases even eliminating the phenomena across the sample population. Our results thus call for a reappraisal of the large body of recent work that has implied high rates of attribute non-attendance for some attributes. Finally, we also highlight a number of general issues with attribute non-attendance, in particular relating to the computation of willingness to pay measures.

Keywords: choice modelling; stated choice; attribute non-attendance; attribute ignoring; taste heterogeneity

1 Introduction

Standard practice in choice modelling assumes that all information related to the choice task and available to the respondent is used in evaluating the alternatives. However, a growing body of work across a

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

†Transport and Mobility Laboratory (TRANSP-OR), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédèrale de Lausanne (EPFL), Switzerland, amanda.stathopoulos@epfl.ch

‡Gibson Institute for Land, Food and Environment, Queens University Belfast, d.campbell@qub.ac.uk

§Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge, United Kingdom, vikki.oneill@mrc-bsu.cam.ac.uk

¶LAN Airlines, Caussade Coudeu, Sebastian.Caussade@lan.com

number of different fields has posed the question as to whether some respondents may actually make their choices based on only a subset of the attributes that describe the alternatives at hand. This phenomenon is typically referred to as attribute non-attendance or attribute ignoring, and an in-depth review of work in this area is given in Hensher (2010).

Three quite separate strands of research can be identified in this context. Analysts have focussed on why non-attendance occurs (e.g. Alemu et al., 2011; Cameron and DeShazo, 2011), how it can be accommodated in our models (e.g. Hess and Rose, 2007) and what its impacts are on model results (e.g. Hensher et al., 2005). The latter two points are strongly inter-related, with the impact on results being to a possibly large extent a function of the specific care used by the analyst in model specification and estimation.

Work has gradually shifted from relying on respondent reported information relating to non-attendance (e.g. Puckett and Hensher, 2008; Rose et al., 2011; Scarpa et al., 2010; Carlsson et al., 2010) towards a focus on inferring the actual processing strategies from data on observed or experimental behaviour (e.g Scarpa et al., 2009; Hess and Hensher, 2010). A key role in this context was played by the early discussions in Hess and Rose (2007), who proposed the use of a latent class approach to accommodate attribute non-attendance, a method since adopted by numerous other studies (e.g. Hensher et al., 2012; Scarpa et al., 2009; Hensher and Greene, 2010; Hole, 2011a; Campbell et al., 2010). With this approach, different latent classes relate to different combinations of attendance and non-attendance across attributes. For each attribute treated in this manner, there exists a non-zero coefficient (to be estimated), which is used in the *attendance classes*, while the attribute is not employed in the *non-attendance* classes, i.e. the coefficient is set to zero. In a complete specification, covering all possible combinations, this would thus lead to 2^K classes, with K being the number of attributes.

The overall findings of the growing body of work using the latent class specification point towards a significant portion of people ignoring attributes, including cost variables. As an example, Scarpa et al. (2009) produce results that imply a staggering rate of 90% of respondents not attending to cost when using the standard latent class specification. Significant rates are also retrieved in Campbell et al. (2011) where 61% of respondents are revealed as not attending to cost. In the context of the present paper, albeit using a different method, it is noteworthy to point out that using a more flexible specification, Campbell et al. (2012) later observe a very significant drop in that share.

In the present paper, we argue that an important shortcoming of this simple latent class approach is the reliance on only two possible values for each coefficient, one of which is fixed to zero. The key point is that this specification is a *confirmatory* latent class model rather than an *exploratory* latent class model - the structure is imposed by the analyst, and class allocation probabilities are estimated. This equates to a reduction in model flexibility, and a more flexible specification would estimate the parameter values in each class. This more general specification would allow the model to still yield one class capturing zero sensitivity (i.e. non-attendance) for a given attribute if this has sufficient weight in the data, but could similarly produce two distinct non-zero values. This is not the case for a model which constrains the coefficients in one class to be equal to zero. Our key argument is that if substantial heterogeneity not related to attribute non-attendance exists in the data, then the simple confirmatory model will at best produce inferior model results, by not being able to capture this heterogeneity, but more worryingly, may produce results affected by serious confounding between non-attendance and taste heterogeneity.

The above arguments point to an issue that has not been duly recognised in existing work in this area, which has often taken the estimated probabilities for the class with coefficients fixed to zero as an *exact* measure of the rate of attribute non-attendance. However, as we will discuss in this paper, it is also possible that the *share* of respondents allocated to the *non-attendance* class need not necessarily

be respondents with a zero sensitivity, but simply respondents with a relatively low sensitivity. This is reinforced by recent discussions in Hess and Hensher (2012), Campbell et al. (2012) and Collins et al. (2012) suggesting that real non-attendance may be rarer than commonly assumed and that low attribute importance may be a more plausible explanation of behaviour. In the most extreme scenario, a possibility exists that no respondents actually ignored a given attribute and that the latent class construct simply captures *regular* heterogeneity. Clearly, if this is the case, findings overstating the true degree of non-attendance are thereby misleading.

On the other hand, it is clear that confounding can also arise in the opposite direction, i.e. where respondents with zero sensitivity are not explicitly accommodated in the model and where their behaviour is captured through reducing the overall sensitivity levels. This is likely to arise especially in cases where just a single point value is estimated, but could also be the case in models with more than one value per coefficient, especially when the rate of non-attendance is too low to pull the estimate in one class close enough to zero to become insignificant. Even in models allowing for continuous random taste heterogeneity, respondents with *real* attribute non-attendance have the potential to affect the overall results.

In this paper, we first illustrate the shortcomings of the confirmatory latent class approach by contrasting its performance with an exploratory model, i.e. one where both values are estimated to allow for both high and low sensitivity for each attribute. Each time, we note improvements in model performance, and, for the majority of attributes, we now observe non-zero sensitivity in both classes. This thus highlights confounding between non-attendance and *standard* heterogeneity in the confirmatory model. However, it should be clear that the exploratory model arguably has the reverse weakness, with some of the share of respondents captured in the lower sensitivity class potentially being respondents with a true zero sensitivity. On the other hand, a simple continuous mixture model is not likely to be able to accommodate non-attendance with standard distributional assumptions. To deal with this limitation, we put forward a framework that combines a continuous mixed logit model with a latent class model (see Greene and Hensher 2013 or Bujosa et al. 2010 for recent discussions of such a structure in a more general context). The model still makes use of two values for each coefficient, where one value is still constrained to zero, but where the value in the second class is allowed to vary across respondents using a continuous distribution. The aim of the model is for this second class to capture *regular* heterogeneity, increasing the probability that any behaviour captured by the class with the zero coefficient does indeed relate to attribute non-attendance. We highlight the applicability of the method using three separate datasets. Each time, we observe significant gains in model fit over the simple latent class specification and show a sizeable reduction in the share of respondents assigned to the non-attendance classes. It is worth noting that an alternative approach would be a model in which the non-zero heterogeneity would be captured by discrete rather than continuous mixtures within the non-zero class.

Before proceeding with the remainder of the paper, it should also be noted that any reductions in wrongly inferred non-attendance would serve to ease issues in the computation of willingness-to-pay (WTP) measures, especially when the cost attribute is concerned, albeit that this does not necessarily avoid the extreme WTP estimations. In our view, WTP implications of non-attendance have thus far not received sufficient discussion in the literature. Non-attendance of attributes other than cost simply leads to a reduction in the mean WTP measure for that attribute across the sample population (if the affected respondents are included in sample level calculations). However, what is not always acknowledged is that implied non-attendance to the cost attribute for a part of the sample population leads to major issues in the computation of WTP measures, as the implied WTP for these respondents would be infinite for the remaining attributes.

The remainder of this paper is organised as follows. Section 2 contrasts the standard latent class approach to non-attendance with our proposed method. This is followed in Section 3 by the results from three case studies. Finally, Section 4 presents the conclusions from the research.

2 Methodology

The widely used confirmatory latent class approach to non-attendance relies, in its complete form, on 2^K different classes, where K is the number of attributes in the model. Each of the 2^K different classes makes use of a different combination of estimated coefficients and coefficients fixed to zero. Crucially, a given coefficient will take the same value in all classes where that attribute is included, thus not allowing for additional random heterogeneity, a point we will return to below. While the majority of classes make use of combinations of zero and non-zero coefficient values, the two extreme classes make use of a full set of estimated parameters and a set of all zero coefficients respectively (i.e. a purely random choice process).

In mathematical terms, we make use of a vector β containing a separate element for each of the K attributes. In addition, we have a SxK matrix Λ, in which each row contains a different combination of 0 and 1 elements, where $S = 2^K$. As an example, if the first class equates to full attendance, then the first row of the matrix, say Λ_1, is a vector with each element being equal to 1. Next, let $A \circ B$ be the element-by-element product of two equally sized vectors A and B, yielding a vector C of the same size, where the k^{th} element of C is obtained by multiplying the k^{th} element of A with the k^{th} element of B. Using this notation, the specific values used for the taste coefficients in class s are then given by the vector $\beta_s = \beta \circ \Lambda_s$.

The likelihood of the observed sequence of T choices for respondent n, say y_n, is then given by:

$$L\left(y_n \mid \beta, \pi\right) = \sum_{s=1}^{S} \pi_s \prod_{t=1}^{T} P\left(i_{nt}^{*} \mid \beta_s = \beta \circ \Lambda_s\right), \tag{1}$$

where i_{nt}^{*} is the alternative chosen by respondent n in choice task t. With $\beta_s = \beta \circ \Lambda_s$ where Λ is fixed a priori, we need to estimate the vector β as well as $\pi = \langle \pi_1, \ldots, \pi_S \rangle$, the vector of probabilities for the S different classes.

While the above exposition uses a full set of 2^K classes, it should be acknowledged that, with a few notable exceptions (e.g. Campbell et al., 2010; Hensher et al., 2012), existing studies typically make use of a simplification, generally focussing on just a subset of attributes that are given a latent class treatment. Although this potentially gives rise to further confounding issues, our focus in this paper is on a more general issue with the use of the simple latent class framework for the purpose of accommodating attribute non-attendance.

A final point of importance relates to the estimation of class probabilities. The probability for class s is given by π_s, with $0 \leq \pi_s \leq 1$ and $\sum_{s=1}^{S} \pi_s = 1$. Rather than imposing constraints in estimation, an easier approach is to use $\pi_s = \frac{e^{\delta_s}}{\sum_{m=1}^{S} e^{\delta_m}}$, with one δ_m, i.e. the parameter used in the class allocation probabilities, being fixed to zero. Nevertheless, this specification still involves estimating $2^K - 1$ separate δ terms, of which many will be very negative, equating to very small class probabilities. In the context of the applications presented in this paper, we make use of a simplified approach, by instead setting

$$\pi_s = \prod_{k=1}^{K} \left(\Lambda_{s,k}\left(1 - P_{\text{N-A},k}\right) + \left(1 - \Lambda_{s,k}\right) P_{\text{N-A},k}\right), \tag{2}$$

where $\Lambda_{s,k}$ gives the entry in Λ relating to attribute k in class s. With this specification, previously used by Hole (2011b), we only need to estimated K separate δ elements (with $P_{\text{N-A},k} = \frac{e^{\delta_k}}{e^{\delta_k}+1}$), as opposed to $2^K - 1$, leading to significant reductions in the number of parameters. While this simplification equates to an assumption that the non-attendance is independent across attributes, we observed only small reductions in model fit with this approach, along with increases in model stability. Indeed, the resulting model seemed to be less susceptible to issues with local maxima than its counterpart with separately estimated class allocation probabilities.

As already alluded to in the introduction, the above specification is an example of a *confirmatory* latent class model, with restrictions on parameters instead of freely estimated cofficients. Our key argument is that a class with a given coefficient fixed to zero is likely to capture not just respondents with a zero sensitivity for that attribute but also respondents with a low sensitivity. To illustrate this, we first note that within a given dataset, the values for a given coefficient are likely to vary continuously across respondents, with both high and low sensitivities arising, as well as a possibility of zero sensitivity for some respondents. Let us assume that the true distribution of non-zero sensitivities covers a wide interval, as is commonly observed in applications. A number of possible scenarios can now be imagined.

Firstly, if there are is a sizable share of respondents with zero sensitivities and we do not account for their presence in the model, then they are likely to have a downwards (in absolute terms) effect on the estimated coefficients. In a fixed coefficients model, they will simply reduce the absolute value of the point estimate. In a continuous random coefficients specification, they will reduce the estimated mean and potentially increase the estimated variance. Finally, in a model estimating a finite number of different values for each coefficient, the presence of these respondents is likely to pull at least one of these values towards zero, and, if the share of non-attending respondents is large enough, the coefficient in one class should collapse to zero. This reasoning is the key argument behind using the confirmatory latent class approach discussed above, i.e. it should capture non-attendance even if the share of such respondents in the sample is small, and potentially too small to be retrieved in an exploratory specification.

Conversely, let us imagine a situation where there is no non-attendance for a given attribute, but where we maintain strong residual heterogeneity (i.e. both low and high sensitivities). Such heterogeneity can be captured either in a continuous random coefficient specification or a latent class model with a number of distinct values for each coefficient. It is this latter case that our key concern applies to. If an analyst estimates a model with only two values for a given coefficient, with one of them being fixed to zero, then the presence of continuous heterogeneity in the data is likely to impact on both the estimate in the non-zero class, and the probability of the zero class. In the context of a model with a fixed coefficient for each attribute, the presence of respondents with very low sensitivities will simply be captured in a downwards pull of that point estimate. If we however use two classes for each coefficient, with one value being constrained to zero, then their behaviour could also be captured by this latter class. The relative impact of such respondents on the probability of the class at zero and the estimate in the non-zero class will depend on the shape of the true distribution. The key point is that respondents captured in the non-attendance class may not necessarily have zero sensitivities.

The model with parameters estimated for both classes to reflect higher and lower sensitivity is given by:

$$L\left(y_n \mid \beta, \pi\right) = \sum_{s=1}^{S} \pi_s \prod_{t=1}^{T} P\left(i_{nt}^{*} \mid \beta_s = \left(\beta_1 \circ \Lambda_s + \beta_2 \circ \left(1 - \Lambda_s\right)\right)\right), \tag{3}$$

with:

$$\pi_s = \prod_{k=1}^{K} \left(\Lambda_{s,k} P_{\mathrm{k,1}} + (1 - \Lambda_{s,k})\, P_{\mathrm{k,2}}\right), \tag{4}$$

where $P_{k,1} = \frac{1}{e^{\delta_k}+1}$ and $P_{k,2} = \frac{e^{\delta_k}}{e^{\delta_k}+1}$. This exploratory LC model is applied for each case study and contrasted with the remaining specifications.

This model will be less susceptible to capture low sensitivity as non-attendance, but on the other hand may not be able to reveal true non-attendance if the rate thereof is low compared to remaining heterogeneity.

The solution put forward in the present paper is to make use of a combined latent class - mixed multinomial logit model, hereafter referred to as LC-MMNL, with LC used for the latent class model. Our proposed model still makes use of two values for each coefficient within a latent class framework, much as in Equation 1. However, β is no longer a vector of point values for the different coefficients, but each element in β now follows a distribution in the sample population. This ensures that, for any given class in our LC-MMNL model, those coefficients not fixed to zero allow for additional random heterogeneity. The likelihood function for respondent n is simply rewritten as:

$$L\left(y_n \mid \Omega, \pi\right) = \sum_{s=1}^{S} \pi_s \int_{\beta} \prod_{t=1}^{T} P\left(i_{nt}^{*} \mid \beta_s = \beta \circ \Lambda_s\right) f\left(\beta \mid \Omega\right) \mathrm{d}\beta, \tag{5}$$

where $f\left(\beta \mid \Omega\right)$ gives the multivariate distribution for all elements in β, with a vector of parameters Ω.

The advantage of this approach is that it will allow some (but not necessarily all) of the random heterogeneity to be captured in the randomly distributed non-zero elements within β_s. This added flexibility should reduce the risk of a class where attribute k is not used (i.e. coefficient fixed to zero) simply capturing heterogeneity, including low sensitivity, rather than genuine attribute non-attendance. We specifically use the word *reduce* as some risk of confounding clearly still remains.

3 Empirical work

In this section, we present results from three separate case studies using stated choice data. Our motivation for using three separate datasets is to produce reliable conclusions that are not based on just a single data source as is all too often the case. Each time, we conduct an in-depth comparison between model structures, looking at model fit, patterns of heterogeneity and implied willingness-to-pay measures.

For each study, we first estimate a simple multinomial logit (MNL) model. This is followed by a confirmatory LC specification as given in Equation 1, with class allocation probabilities defined as in Equation 2, and an exploratory LC specification as given in Equation 3, with class allocation probabilities defined as in Equation 4. Finally, we also estimate two random coefficients structures, namely a simple MMNL model, and a LC-MMNL model as given in Equation 5. All models were coded in Ox 6.2 (Doornik, 2001). In the latent class structures, multiple runs with different starting values were used to attempt to avoid issues with inferior local maxima.

For the two mixture models, the estimation made use of 250 MLHS draws (Hess et al., 2006) per respondent and per random parameter, where Lognormal distributions were used. This specific choice of distribution is motivated by the fact that the Lognormal is particularly well suited to capture both low

and high attribute sensitivity, given that it has a low median with a long tail and high mean. This should permit the distribution to capture such low attribute sensitivity rather than it being falsely captured by the non-attendance class. In addition, for the datasets at hand, previous experience showed good performance with the Lognormal distribution.

The results focus on a number of key indicators, in particular the differences in fit between the various models, the implied rates of non-attendance in the LC and LC-MMNL models, and the heterogenity patterns for individual coefficients. As alluded to in the introduction, the computation of WTP indicators is fraught with complications in the presence of a non-zero share for non-attendance, especially when this is the case for the cost attribute. In the present paper, the focus on WTP is reduced, and we give more weight to the above key indicators. Nevertheless, we present summary WTP statistics too, where, in the confirmatory LC model and the LC-MMNL model, we exclude the *non-attendance* class for fare.

3.1 First case study

3.1.1 Data and model specification

Our first case study makes use of data collected in a stated choice (SC) survey for rail and bus commuters, conducted online in the UK, early 2010. A sample of 368 respondents were each faced with ten different choice tasks, each time involving the choice between three unlabelled alternatives, of which the first is a reference trip (invariant across the ten tasks). The three alternatives were described by six attributes, namely travel time (in minutes), fare (in £), the level of crowding (trips out of ten for which you have to stand), the rate of delays (trips out of ten affected by delays), the average length of delays (across delayed trips), and the provision of a free information delay service (via sms text message). Linear effects were specified for the three continuous attributes, while the coefficients for crowding and the rate of delays were multiplied by the rate of affected trips (a division of these two attributes by a factor of ten led to them ranging from 0 to 1). A simple dummy term was included for the delay information service, and alternative specific constants (ASC) were estimated for the first two alternatives.

3.1.2 Results

The main estimation results for the first case study are summarised in Table 1. This shows the point estimates from the MNL and LC models, while, for the MMNL and LC-MMNL models, the use of the Lognormal distribution implies that we are actually estimating the parameters of the normally distributed logarithms of the individual marginal utility coefficients, where a sign change applies to all coefficients except the information service coefficient. For example, $\mu_{\ln(-\beta_{\mathbf{tt}})}$ gives the mean for the normal distribution for the logarithm of the negative of the travel time coefficient, with $\sigma_{\ln(-\beta_{\mathbf{tt}})}$ giving the associated standard deviation. The sign of the mean estimates relates to the absolute size, rather than sign, of the resulting Lognormal coefficient. A negative estimate for $\mu_{\ln(-\beta_{\mathbf{tt}})}$ simply means that the median of the travel time coefficient is greater than -1 but below zero. A positive estimate would imply a median that is smaller than -1. In all models, we also estimate two additional alternative specific constants (ASC_1 and ASC_2).

In the latent class models, we have two sets of values for coefficients, i.e. β_1 and β_2, where, in the confirmatory LC model and in the LC-MMNL model, the elements in β_2 are constrained to zero. For the sake of presentation, the elements for each coefficient were ordered in such a way that the larger value (in absolute terms) is in class 1. In all three latent class structures, we additionally estimate six δ parameters which are used to compute the class allocation probabilities, as shown in Equation 4. As

an example, a negative value for $\delta_{\text{tt},2}$ would mean that the share for the $\beta_{\text{tt},2}$ is smaller than 50%. The results from Table 1 are then used to compute probabilities for the two values for each coefficient, as reported in Table 2, where, in the confirmatory LC model and the LC-MMNL model, the second class corresponds to the *non-attendance* class, while, in the exploratory LC model, it corresponds to the *low sensitivity* class.

We first note that all four *advanced* models comprehensively outperform the MNL model, with gains in log-likelihood by between 538.9 units and 714.6 units, for at worst an increase by 12 in the number of parameters. The exploratory latent class model outperforms the confirmatory latent class model, with an improvement in log-likelihood by 106.46 units, which is highly significant at the cost of six additional parameters. Using the adjusted ρ^2 measure, we note that the MMNL model outperforms both LC models, suggesting that it is able to give a more complete representation of the random heterogeneity. In the combined LC-MMNL, the probabilities for non-attendance collapse to zero (i.e. very negative δ) for all but the final two coefficients (average delay and delay sms). Given the resulting difference in specification, the LC model is no longer nested inside the LC-MMNL model, however the differences in fit are apparent when looking at the adjusted ρ^2 measure. Finally, the LC-MMNL model obtains an improvement in fit over the MMNL model by 10.23 units, which, although modest in terms of absolute improvements, is still statistically significant at the cost of two additional parameters. More sizable differences arise when looking at the actual estimation results, and these justify our recommendation for using the LC-MMNL model.

We observe substantial increases in coefficient magnitude (i.e. scale) when moving from the MNL to the confirmatory LC model, which is consistent with the marginal utility coefficients now relating to the non-ignoring classes. The results from Table 2 show very large shares for the second value for each coefficient in the confirmatory LC model, and thus high rates of implied attribute non-attendance. These are arguably so high as to be considered unrealistic, with e.g. 44% of commuters apparently ignoring fare, where the implied rate of attribute-ignoring is even higher for the quality-of-service attributes. As already mentioned above, the exploratory LC model obtains statistically significant improvements in log-likelihood over the confirmatory LC model. With the values for β_2 now being freely estimated for each coefficient, the class allocation probabilities cannot be compared to the confirmatory model as they relate to different values. Crucially, the estimates in the low importance classes are significantly different from zero with the exception of the delay information attribute. If the high rates of non-attendance in the confirmatory model had been in line with real behaviour in the data, we would have expected the estimates in the second class in the exploratory model not to be significantly different from zero. These observations should serve to confirm our assertions that a two-class LC model will capture the heterogeneity by using classes for both high and low sensitivity, and that classes with coefficients constrained to zero are likely to at least in part capture low attribute sensitivity rather than attribute non-attendance.

Turning our attention to the continuous mixture models, i.e. MMNL and LC-MMNL, we note that both models retrieve significant random heterogeneity for all six coefficients, as highlighted by the estimates for the σ parameters. In the LC-MMNL models, the shares for the second class (i.e. $\beta = 0$) drop to zero for travel time, fare, crowding, and the rate of delays. For the remaining two attributes, namely the average delay and the provision of a free delay sms service, high implied rates of ignoring remain, but they are still lower than in the confirmatory LC model. These findings give further support to our assertion that the confirmatory LC model is likely to produce misleading results for attribute non-attendance, with the latent class construct, specifically the class at zero, serving as a proxy for capturing variations in sensitivity across respondents not related to attribute non-attendance.

We next proceed with a comparison between the estimates for the MMNL and LC-MMNL models.

Table 1: Main estimation results for first case study

	MNL		LC (confirmatory)		LC (exploratory)		MMNL		LC-MMNL	
LL(0)	-4,042.89		-4,042.89		-4,042.89		-4,042.89		-4,042.89	
LL	-3,700.43		-3,161.53		-3,055.07		-2,996.05		-2,985.82	
par	8		14		20		14		16	
adj. ρ^2	0.0827		0.2145		0.2443		0.2555		0.2575	
	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.
ASC_1	0.2026	3.33	0.6697	8.24	0.6761	9.19	0.7900	9.82	0.8257	9.87
ASC_2	0.1314	3.01	0.3084	4.34	0.2917	4.29	0.3281	4.42	0.3361	4.34
$\beta_{\text{tt},1}$	-0.0365	-7.72	-0.1419	-2.39	-0.3615	-3.79	-	-	-	-
$\beta_{\text{fare},1}$	-1.0126	-4.37	-8.6215	-7.51	-11.6830	-13.68	-	-	-	-
$\beta_{\text{crowding},1}$	-1.6805	-7.25	-15.3160	-2.55	-24.5050	-8.81	-	-	-	-
$\beta_{\text{rate of delays},1}$	-1.7777	-7.71	-13.0860	-4.57	-18.5430	-9.47	-	-	-	-
$\beta_{\text{average delay},1}$	-0.0344	-4.87	-0.4712	-4.44	-0.7149	-7.58	-	-	-	-
$\beta_{\text{delay sms},1}$	0.2705	4.48	2.0167	2.52	2.4121	1.22	-	-	-	-
$\beta_{\text{tt},2}$	-	-	0	-	-0.0501	-6.35	-	-	0	-
$\beta_{\text{fare},2}$	-	-	0	-	-0.8452	-5.29	-	-	0	-
$\beta_{\text{crowding},2}$	-	-	0	-	-1.7986	-5.33	-	-	0	-
$\beta_{\text{rate of delays},2}$	-	-	0	-	-2.2140	-6.58	-	-	0	-
$\beta_{\text{average delay},2}$	-	-	0	-	-0.0234	-2.20	-	-	0	-
$\beta_{\text{delay sms},2}$	-	-	0	-	0.2844	1.02	-	-	0	-
$\mu_{\ln(-\beta_{\text{tt}})}$	-	-	-	-	-	-	-2.6615	-22.94	-2.7623	-18.92
$\mu_{\ln(-\beta_{\text{fare}})}$	-	-	-	-	-	-	1.2644	8.89	1.2397	11.19
$\mu_{\ln(-\beta_{\text{crowding}})}$	-	-	-	-	-	-	0.8004	4.01	0.7158	3.64
$\mu_{\ln(-\beta_{\text{rate of delays}})}$	-	-	-	-	-	-	1.0206	6.88	1.0020	5.99
$\mu_{\ln(-\beta_{\text{average delay}})}$	-	-	-	-	-	-	-3.2968	-9.92	-2.2778	-3.88
$\mu_{\ln(-\beta_{\text{delay sms}})}$	-	-	-	-	-	-	-1.0842	-5.60	0.3820	1.81
$\sigma_{\ln(-\beta_{\text{tt}})}$	-	-	-	-	-	-	0.8286	8.78	1.0593	8.98
$\sigma_{\ln(-\beta_{\text{fare}})}$	-	-	-	-	-	-	1.6004	8.69	1.5191	20.03
$\sigma_{\ln(-\beta_{\text{crowding}})}$	-	-	-	-	-	-	1.5719	11.56	1.5241	12.37
$\sigma_{\ln(-\beta_{\text{rate of delays}})}$	-	-	-	-	-	-	1.1127	12.27	1.3873	12.33
$\sigma_{\ln(-\beta_{\text{average delay}})}$	-	-	-	-	-	-	1.3461	10.06	1.2455	3.97
$\sigma_{\ln(-\beta_{\text{delay sms}})}$	-	-	-	-	-	-	1.2028	11.44	0.2962	2.05
$\delta_{\text{tt},2}$	-	-	0.1398	0.22	1.7764	4.37	-	-	-[†]	
$\delta_{\text{fare},2}$	-	-	-0.2563	-1.31	0.0709	0.54	-	-	-[†]	
$\delta_{\text{crowding},2}$	-	-	1.3250	2.46	1.9049	8.83	-	-	-[†]	
$\delta_{\text{rate of delays},2}$	-	-	1.0608	2.69	1.7457	7.06	-	-	-[†]	
$\delta_{\text{average delay},2}$	-	-	1.4462	4.30	2.0609	6.51	-	-	-0.0639	-0.10
$\delta_{\text{delay sms},2}$	-	-	1.1356	1.91	1.7197	1.01	-	-	0.4128	1.13

[†] classes with coefficient at zero (i.e. β_2) eliminated from final model after initial estimates for δ became very negative, implying a zero probability of non-attendance

We first note drops in significance for the parameters relating to average delay and the delay information service, which is expected given that they now only relate to a smaller share of respondents. While

Table 2: Implied probabilities for second class in LC models in first case study

	LC (confirmatory)			LC (exploratory)			LC-MMNL		
	est.	t-rat. (0)	t-rat. (1)	est.	t-rat. (0)	t-rat. (1)	est.	t-rat. (0)	t-rat. (1)
$P_{\text{tt},2}$	0.53	3.45	-3.00	0.86	16.99	-2.88	0	-	-
$P_{\text{fare},2}$	0.44	9.09	-11.75	0.52	15.76	-14.68	0	-	-
$P_{\text{crowding},2}$	0.79	8.84	-2.35	0.87	35.79	-5.33	0	-	-
$P_{\text{rate of delays},2}$	0.74	9.88	-3.42	0.85	27.22	-4.75	0	-	-
$P_{\text{average delay},2}$	0.81	15.61	-3.68	0.89	27.95	-3.56	0.48	2.98	-3.18
$P_{\text{delay sms},2}$	0.76	6.92	-2.22	0.85	3.85	-0.69	0.60	6.84	-4.53

Table 3: Heterogeneity patterns in non-ignoring classes for continuous mixture models in first case study

	MMNL			LC-MMNL		
	μ	σ	cv	μ	σ	cv
β_{tt}	-0.10	0.10	-0.99	-0.11	0.16	-1.44
β_{fare}	-12.74	44.06	-3.46	-10.95	32.95	-3.01
β_{crowding}	-7.66	25.21	-3.29	-6.54	19.83	-3.03
$\beta_{\text{rate of delays}}$	-5.15	8.06	-1.56	-7.13	17.25	-2.42
$\beta_{\text{average delay}}$	-0.09	0.21	-2.26	-0.22	0.43	-1.93
$\beta_{\text{delay sms}}$	0.70	1.26	1.80	1.53	0.46	0.30

the mean estimates for other coefficients remain quite stable, indicating no major scale differences, the estimates for $\mu_{\ln(-\beta_{\text{average delay}})}$ and $\mu_{\ln(-\beta_{\text{delay sms}})}$ are more positive in the LC-MMNL model, indicating a stronger absolute sensitivity in class 1, as would be the case when class 2 captures low or zero sensitivity. It is also of interest to look at the degree of continuous heterogeneity for parameters in the MMNL and LC-MMNL model after transformation to the Lognormal scale. The results are reported in Table 3, where, with the specification in Equation 5, these calculations clearly only relate to the non-ignoring classes for each given coefficient in the LC-MMNL model. By inspecting the coefficients of variation (cv) for the two attributes where a non-zero probability for non-attendance is obtained in the LC-MMNL model (i.e. average delays and delay sms service), we note a drop in heterogeneity when moving from the MMNL model to the LC-MMNL model, where some of what was previously captured as continuous heterogeneity is now accommodated in the non-attendance class. For the remaining four coefficients, we see two increases and two decreases in heterogeneity. This occurs despite the fact that the treatment of these four coefficients is identical in the two models. A possible interpretation is that the specification used for one coefficient can have an impact on the estimated heterogeneity for a different coefficient.

Finally, Table 4 summarises the WTP findings for the different models. In the MNL, exploratory LC, and MMNL models, these relate to sample level estimates[1], while, in the confirmatory LC models, they exclude the non-attendance class for fare. For the three models with sample level estimates (i.e. MNL, exploratory LC, and MMNL), the mean WTP measures are higher in the two mixture models, where the levels of heterogeneity are higher in the MMNL model when compared with the exploratory LC model. When comparing the MNL and confirmatory LC models, the latter clearly obtains lower WTP measures given the inclusion of the zero class for all coefficients other than fare in these calculations. This

[1]For MNL, point estimates are used, for exploratory LC, all four combinations of two coefficients are used in a weighted average of the resulting four ratios, while, in the MMNL models, we obtain a ratio of two distributions.

Table 4: Implied WTP measures for first case study

		LC (conf.)†		LC (expl.)		MMNL		LC-MMNL	
	MNL	μ	σ	μ	σ	μ	σ	μ	σ
Travel time reductions ($£/hr$)	2.16	0.46	0.49	3.73	6.42	6.00	29.86	6.09	33.30
Reductions in crowding ($£$ per $\frac{1}{10}$ reduction)	0.17	0.04	0.07	0.31	0.70	0.78	9.60	0.60	6.04
Reductions in rate of delays ($£$ per $\frac{1}{10}$ reduction)	0.18	0.04	0.07	0.30	0.56	0.52	3.46	0.65	5.39
Average delay reductions ($£/hr$)	2.04	0.63	1.29	3.98	11.70	5.58	49.41	6.33	60.31
Delay sms service ($£$)	0.27	0.06	0.10	0.40	0.73	0.71	5.21	0.56	2.88

†: respondents in *non-attendance* class for fare excluded from calculations

highlights the problematic issue of calculating WTP measures from such models; we obtain a downwards bias in the WTP measures. Indeed, the issue of whether what is picked up is true non-attendance aside, the calculation accounts for the zero WTP for respondents captured in the non-attendance classes for non-fare attributes, but not the *infinite* WTP for respondents captured in the non-attendance class for fare.

Turning to the MMNL and LC-MMNL models, the differences are relatively small. This is in line with having retrieved a positive share for the non-attendance class for only two of the coefficients, and the absence of non-attendance for fare. Additionally, the results would suggest that similar WTP patterns can be obtained by the MMNL model on its own, confirming our hypothesis that the Lognormal distribution is well suited to dealing with low or zero attribute sensitivity.

3.2 Second case study

3.2.1 Data and model specification

Our second case study once again makes use of data collected through an online SC survey conducted in the UK in early 2010, using a sample of 996 rail commuters, each facing 8 choice tasks. In each choice task, a respondent is given a choice between three alternatives, described by attributes with levels pivoted around reported reference values, where the reference alternative was however not included as one of the options in the survey. The five attributes used to describe the alternatives were travel time (minutes), fare (£), the guarantee of a reserved seat (yes/no dummy), the provision of free wifi (yes/no dummy), and whether the ticket had rebooking flexibility (yes/no dummy). The specification is similar to that used in the first case study, with the exception that a log-transform was used for the fare attribute, given strong evidence of decreasing marginal sensitivity to fare.

3.2.2 Results

The results are once again split across four tables, using the same approach to notation as in the first case study. Specifically, the main estimation results are presented in Table 5, with implied probabilities of non-attendance in Table 6, estimated levels of heterogeneity in Table 7, and implied WTP patterns in the non-ignoring classes in Table 8. We use the same five models as in the first case study.

We note that all four *advanced* models once again offer hugely significant improvements in log-likelihood over the simple MNL model, where slightly better fit is again obtained for the models allowing for continuous random heterogeneity. The exploratory LC model comprehensively outperforms the confirmatory LC model, with a gain in log-likelihood by 132.6 units at the cost of five additional parameters.

Table 5: Main estimation results for second case study

	MNL		LC (confirmatory)		LC (exploratory)		MMNL		LC-MMNL	
LL(0)	-8,753.74		-8,753.74		-8,753.74		-8,753.74		-8,753.74	
LL	-6,288.87		-5,461.53		-5,328.97		-5,296.37		-5,250.43	
par	7		12		17		12		17	
adj. ρ^2	0.2808		0.3747		0.3893		0.3936		0.3983	
	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.
ASC_1	0.1075	2.80	0.1432	2.81	0.1697	3.29	0.1800	3.32	0.1828	3.29
ASC_2	0.0017	0.05	-0.0844	-1.68	-0.0746	-1.47	-0.0620	-1.21	-0.0784	-1.45
$\beta_{\text{tt},1}$	-0.0207	-18.13	-0.0571	-17.38	-0.0784	-12.04	-	-	-	-
$\beta_{\text{log-fare},1}$	-5.8199	-24.30	-13.4230	-23.28	-16.0810	-13.95	-	-	-	-
$\beta_{\text{seat},1}$	0.9864	19.95	2.7677	16.87	4.2345	8.05	-	-	-	-
$\beta_{\text{wi-fi},1}$	0.2679	5.65	3.9880	8.89	4.2327	4.33	-	-	-	-
$\beta_{\text{flex},1}$	0.2452	3.26	-2.5579	-5.79	3.2320	9.12	-	-	-	-
$\beta_{\text{tt},2}$	-	-	0	-	-0.0114	-4.32	-	-	0	-
$\beta_{\text{log-fare},2}$	-	-	0	-	-2.7535	-3.77	-	-	0	-
$\beta_{\text{seat},2}$	-	-	0	-	0.7540	5.83	-	-	0	-
$\beta_{\text{wi-fi},2}$	-	-	0	-	0.0184	0.23	-	-	0	-
$\beta_{\text{flex},2}$	-	-	0	-	0.1623	1.42	-	-	0	-
$\mu_{\ln(-\beta_{\text{tt}})}$	-	-	-	-	-	-	-3.3900	-49.54	-3.1611	-34.21
$\mu_{\ln(-\beta_{\text{log-fare}})}$	-	-	-	-	-	-	2.1835	41.55	2.4826	29.08
$\mu_{\ln(\beta_{\text{seat}})}$	-	-	-	-	-	-	0.1494	1.91	0.6925	3.20
$\mu_{\ln(\beta_{\text{wi-fi}})}$	-	-	-	-	-	-	-1.7386	-7.07	0.4899	1.39
$\mu_{\ln(\beta_{\text{flex}})}$	-	-	-	-	-	-	-1.1588	-4.40	0.6724	0.78
$\sigma_{\ln(-\beta_{\text{tt}})}$	-	-	-	-	-	-	1.0617	15.60	0.8921	10.17
$\sigma_{\ln(-\beta_{\text{log-fare}})}$	-	-	-	-	-	-	0.8482	15.90	0.5703	5.28
$\sigma_{\ln(\beta_{\text{seat}})}$	-	-	-	-	-	-	1.2483	14.38	0.7969	2.81
$\sigma_{\ln(\beta_{\text{wi-fi}})}$	-	-	-	-	-	-	1.6180	14.96	0.8283	4.03
$\sigma_{\ln(\beta_{\text{flex}})}$	-	-	-	-	-	-	1.1771	6.98	0.7813	1.63
$\delta_{\text{tt},2}$	-	-	-0.3387	-2.87	0.1711	0.86	-	-	-2.0815	-5.65
$\delta_{\text{log-fare},2}$	-	-	-0.4433	-4.07	-0.1696	-0.77	-	-	-1.6233	-5.55
$\delta_{\text{seat},2}$	-	-	-0.5343	-3.55	0.6278	2.51	-	-	-1.4722	-2.25
$\delta_{\text{wi-fi},2}$	-	-	2.6737	12.53	2.5017	6.98	-	-	1.3724	3.62
$\delta_{\text{flex},2}$	-	-	2.3372	5.81	1.9203	8.30	-	-	1.5140	1.85

Table 6: Implied probabilities for second class in LC models in second case study

	LC (confirmatory)			LC (exploratory)			LC-MMNL		
	est.	t-rat. (0)	t-rat. (1)	est.	t-rat. (0)	t-rat. (1)	est.	t-rat. (0)	t-rat. (1)
$P_{\text{tt},2}$	0.42	14.49	-20.33	0.54	10.99	-9.26	0.11	3.06	-24.49
$P_{\text{log-fare},2}$	0.39	15.08	-23.48	0.46	8.33	-9.87	0.16	4.09	-20.76
$P_{\text{seat},2}$	0.37	10.52	-17.96	0.65	11.49	-6.13	0.19	1.88	-8.19
$P_{\text{wifi},2}$	0.94	72.59	-5.01	0.92	36.83	-3.02	0.80	13.04	-3.30
$P_{\text{flex},2}$	0.91	28.23	-2.73	0.87	33.80	-4.95	0.82	6.77	-1.49

Table 7: Heterogeneity patterns in non-ignoring classes for continuous mixture models in second case study

	MMNL			LC-MMNL		
	μ	σ	cv	μ	σ	cv
β_{tt}	-0.06	0.09	-1.44	-0.06	0.07	-1.10
$\beta_{\text{log-fare}}$	-12.72	13.05	-1.03	-14.09	8.73	-0.62
β_{seat}	2.53	4.90	1.94	2.75	2.59	0.94
$\beta_{\text{wi-fi}}$	0.65	2.32	3.56	2.30	2.28	0.99
β_{flex}	0.63	1.09	1.73	2.66	2.44	0.92

Table 8: Implied WTP measures for second case study, calculated at a fare of $£40$

		LC (conf.)†		LC (expl.)		MMNL		LC-MMNL†	
	MNL	μ	σ	μ	σ	μ	σ	μ	σ
Travel time reductions ($£/hr$)	8.55	5.96	5.03	20.17	25.10	22.94	53.01	13.23	20.71
Reserved seat ($£$)	6.78	5.20	3.98	15.72	20.34	16.34	48.35	8.78	13.06
Free wifi ($£$)	1.84	0.77	2.92	2.70	11.33	4.20	21.89	1.83	6.49
Ticket flexibility ($£$)	1.69	-0.69	2.18	4.44	10.77	4.05	10.88	1.88	6.83

†: respondents in *non-attendance* class for fare excluded from calculations

For the continuous mixture models, the differences are larger than in the first case study, with an improvement in log-likelihood by 45.94 units for the LC-MMNL model over the MMNL model. This is a direct result of the fact that we are able to estimate non-zero probabilities for non-attendance for all five marginal utility coefficients, in contrast with the first case study. This new model also obtains an improvement in log-likelihood by 211.10 units over the LC model. Both improvements are significant at the highest levels of confidence, coming at the cost of 5 additional parameters.

In terms of effects on parameter estimates, we observe increases in scale in the marginal utility coefficients when moving from MNL to the confirmatory LC model, where, rather counter-intuitively, we also obtain a negative estimate for β_{flex}, where it needs to be noted that this value only applies to a small share, with the class probability of just nine percent. In the exploratory LC model, we again retrieve small and large values for all coefficients, where all estimates are of the expected sign in both classes. We note that the estimates for $\beta_{\text{wi-fi,2}}$ and $\beta_{\text{flex,2}}$ are not significantly different from zero. These were also the two coefficients with the highest rates of implied non-attendance in the confirmatory LC model. The share for this class drops only slightly in the exploratory LC model. For those three coefficients where the estimates in the second class are significantly different from zero, we also see some increases in class allocation probability compared to the confirmatory LC model.

When moving from MMNL to LC-MMNL, we observe that the coefficients increase in magnitude (after transforming from the underlying Normal distribution). Also, while there is a shift towards statistical significance for $\mu_{\ln(\beta_{\text{seat}})}$, the opposite applies for $\mu_{\ln(\beta_{\text{wi-fi}})}$ and $\mu_{\ln(\beta_{\text{flex}})}$, where major increases in standard errors are also observed for the standard deviation parameters, especially $\sigma_{\ln(\beta_{\text{flex}})}$. The changes, including increases in magnitude, are most substantial for the final two coefficients which are the ones with the highest implied probabilities for non-attendance. In comparison with the confirmatory LC model, the LC-MMNL results show visibly lower rates of implied non-attendance, which is reassuring in the case of travel time and fare. However, high rates of implied non-attendance are still retrieved for the wifi and

ticket flexibility attributes, where this finding is more realistic, given the likelihood of such quality of service attributes only being relevant to a part of the travelling population. This is also consistent with the results for the exploratory LC model. However, the remaining 16% rate of implied non-attendance for the fare attribute remains problematic. In line with the fact that for each coefficient, some of the heterogeneity in the LC-MMNL model is captured by the non-attendance classes, we observe a reduction in the remaining continuous random heterogeneity for each coefficient in comparison with the MMNL model (cf. Table 7).

Turning our attention finally to the implied WTP patterns (cf. Table 8), we see drops when moving from MNL to confirmatory LC, in line with the implied rates of non-attendance and the fact that the non-attendance group for fare is excluded from these calculations. The negative WTP for flexibility is the result of the negative estimate for β_{flex} in the confirmatory LC model. In the exploratory LC model, we note substantially higher WTP measures, which are a result of the much lower value for $\beta_{\text{log-fare,2}}$ - this model produces evidence of very low as well as very high fare sensitivity. The differences between MMNL and LC-MMNL are more pronounced than was the case in the first study, and this is in line with the higher rates of remaining implied non-attendance. We see decreases for all four measures, as a result of accounting for the presence of the non-attendance class for these attributes but disregarding the 16% probability for $\beta_{\text{log-fare,2}}$ in these WTP calculations.

3.3 Third case study

3.3.1 Data and model specification

Our final case study makes use of data collected in an online SC survey by the Latin American airline LAN in 2009, using a sample of 915 respondents, with ten observations per respondent. In each scenario, a respondent was faced with a choice between three flight options described by six attributes, where there was no reference alternative. The attributes used to describe the alternatives were fare (in thousands of Chilean pesos), the accrual of frequent flyer benefits (LANPASS kms), the ability to reserve a specific seat (no seat reservation/reservation of standard seat/reservation of preferential seat), the ability to change reservation before the departure date, the ability to change reservations after the outbound departure date (yes/no dummies), and the ability to obtain a refund for unused tickets (yes/no dummies). A linear in attributes specification of the utility function was used, with constants for the first two alternatives.

3.3.2 Results

The results are once again split across four tables, using the same approach to notation as in the first two case studies. The main estimation results are presented in Table 9, with implied probabilities of non-attendance in Table 10, estimated levels of heterogeneity in Table 11, and implied WTP patterns in the non-ignoring classes in Table 12. The same five models as in the first two case studies are estimated.

As shown in Table 9, we again have the familiar pattern of differences in model fit. All four *advanced* models comprehensively outperform the MNL model. While the differences between the remaining four models are once again smaller, we see that better performance is obtained by the continuous mixture models, while the exploratory LC model again easily rejects the confirmatory LC model (137.08 units, at the cost of 7 parameters), and the LC-MMNL model also obtains a significant improvement over the MMNL model (51.19 units in log-likelihood, at the cost of six additional parameters). In the combined LC-MMNL, the probability for non-attendance collapsed to zero for the regular seat attribute, meaning that the LC model is no longer nested inside the LC-MMNL model. However, the difference in fit is clearly

Table 9: Main estimation results for third case study

	MNL		LC (confirmatory)		LC (exploratory)		MMNL		LC-MMNL	
LL(0)	-10,711.47		-10,711.47		-10,711.47		-10,711.47		-10,711.47	
LL	-7,483.52		-6,875.05		-6,737.97		-6,733.15		-6,681.96	
par	9		16		23		16		22	
adj. ρ^2	0.3005		0.3567		0.3688		0.3699		0.3741	
	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.	est.	t-rat.
ASC_1	0.3425	9.65	0.4321	9.19	0.38585	7.95	0.4017	8.63	0.4252	8.51
ASC_2	0.2683	7.81	0.3495	7.87	0.32873	7.09	0.3273	7.38	0.3559	7.50
$\beta_{\text{fare},1}$	-0.0944	-33.01	-0.1837	-25.01	-0.25233	-20.00	-	-	-	-
$\beta_{\text{accrkm},1}$	0.2499	13.29	1.0639	9.36	1.7763	6.03	-	-	-	-
$\beta_{\text{regseat},1}$	0.0739	7.64	0.1102	2.58	0.64495	3.34	-	-	-	-
$\beta_{\text{prefseat},1}$	0.1251	10.82	0.2929	7.63	0.60326	3.30	-	-	-	-
$\beta_{\text{chgbefor},1}$	0.8042	20.78	1.9325	14.34	2.4362	8.66	-	-	-	-
$\beta_{\text{chgafter},1}$	0.4895	13.43	1.2550	8.34	0.98547	5.77	-	-	-	-
$\beta_{\text{refund},1}$	0.2513	7.32	1.8707	6.59	0.84912	1.82	-	-	-	-
$\beta_{\text{fare},2}$	-	-	0	-	-0.04764	-10.86	-	-	0	-
$\beta_{\text{accrkm},2}$	-	-	0	-	0.17627	5.13	-	-	0	-
$\beta_{\text{regseat},2}$	-	-	0	-	0.07808	5.29	-	-	0	-
$\beta_{\text{prefseat},2}$	-	-	0	-	0.09528	3.65	-	-	0	-
$\beta_{\text{chgbefor},2}$	-	-	0	-	0.45564	3.66	-	-	0	-
$\beta_{\text{chgafter},2}$	-	-	0	-	0.45332	0.98	-	-	0	-
$\beta_{\text{refund},2}$	-	-	0	-	0.66268	0.86	-	-	0	-
$\mu_{\ln(-\beta_{\text{fare}})}$	-	-	-	-	-	-	-2.0352	-44.41	-1.7816	-32.73
$\mu_{\ln(\beta_{\text{accrkm}})}$	-	-	-	-	-	-	-1.4047	-13.18	-0.2719	-1.79
$\mu_{\ln(\beta_{\text{regseat}})}$	-	-	-	-	-	-	-2.4564	-16.84	-2.5909	-12.57
$\mu_{\ln(\beta_{\text{prefseat}})}$	-	-	-	-	-	-	-2.0939	-17.02	-1.3485	-8.65
$\mu_{\ln(\beta_{\text{chgbefor}})}$	-	-	-	-	-	-	-0.1158	-1.82	0.3821	2.66
$\mu_{\ln(\beta_{\text{chgafter}})}$	-	-	-	-	-	-	-0.5182	-5.59	-0.0326	-0.15
$\mu_{\ln(\beta_{\text{refund}})}$	-	-	-	-	-	-	-1.9476	-6.85	0.2735	0.89
$\sigma_{\ln(-\beta_{\text{fare}})}$	-	-	-	-	-	-	1.0815	23.44	0.8553	14.94
$\sigma_{\ln(\beta_{\text{accrkm}})}$	-	-	-	-	-	-	1.1502	12.72	0.5296	5.33
$\sigma_{\ln(\beta_{\text{regseta}})}$	-	-	-	-	-	-	0.0450	0.18	0.6859	4.60
$\sigma_{\ln(\beta_{\text{prefseta}})}$	-	-	-	-	-	-	0.7430	14.99	0.3813	2.23
$\sigma_{\ln(\beta_{\text{chgbefor}})}$	-	-	-	-	-	-	0.7592	13.42	0.5043	3.99
$\sigma_{\ln(\beta_{\text{chgafter}})}$	-	-	-	-	-	-	0.5629	4.37	0.4614	2.56
$\sigma_{\ln(\beta_{\text{refund}})}$	-	-	-	-	-	-	1.5812	8.12	0.5524	2.29
$\delta_{\text{fare},2}$	-	-	-1.1221	-11.28	-0.34712	-3.08	-	-	-2.3344	-9.47
$\delta_{\text{accrkm},2}$	-	-	0.5688	3.15	1.6081	5.68	-	-	-0.1028	-0.43
$\delta_{\text{regseat},2}$	-	-	-1.8885	-0.71	3.6118	3.69	-	-	-[†]	
$\delta_{\text{prefseat},2}$	-	-	-0.2396	-0.71	1.7913	2.72	-	-	-0.5000	-1.43
$\delta_{\text{chgbefor},2}$	-	-	-0.2893	-1.73	0.54812	1.56	-	-	-1.0805	-2.92
$\delta_{\text{chgafter},2}$	-	-	-0.1252	-0.48	-1.3233	-1.51	-	-	-0.8432	-1.74
$\delta_{\text{refund},2}$	-	-	1.4755	5.47	-0.68782	-0.43	-	-	0.9483	2.89

[†] class with coefficient at zero eliminated from final model after initial estimate for δ became very negative, implying a zero probability of non-attendance

Table 10: Implied probabilities second class in third case study

	LC (confirmatory)			LC (exploratory)			LC-MMNL		
	est.	t-rat. (0)	t-rat. (1)	est.	t-rat. (0)	t-rat. (1)	est.	t-rat. (0)	t-rat. (1)
$P_{\text{fare},2}$	0.25	13.33	-40.92	0.41	15.16	-3.15	0.09	4.45	-45.96
$P_{\text{accrkm},2}$	0.64	15.32	-8.67	0.83	21.18	8.47	0.47	7.91	-8.77
$P_{\text{regseat},2}$	0.13	0.43	-2.87	0.97	0.93	-0.03	0	-	-
$P_{\text{prefseat},2}$	0.44	5.28	-6.71	0.86	10.62	4.43	0.38	4.59	-7.57
$P_{\text{chgbefor},2}$	0.43	10.44	-13.94	0.63	7.77	1.64	0.25	3.62	-10.67
$P_{\text{chgafter},2}$	0.47	7.22	-8.18	0.21	1.44	-1.99	0.30	2.95	-6.86
$P_{\text{refund},2}$	0.81	19.94	-4.56	0.33	0.93	-0.46	0.72	10.91	-4.23

Table 11: Heterogeneity patterns in non-ignoring classes for continuous mixture models in third case study

	MMNL			LC-MMNL		
	μ	σ	cv	μ	σ	cv
β_{fare}	-0.23	0.35	-1.49	-0.24	0.25	-1.04
β_{accrkm}	0.48	0.79	1.66	0.88	0.50	0.57
β_{regseat}	0.09	0.00	0.05	0.09	0.07	0.78
β_{prefseat}	0.16	0.14	0.86	0.28	0.11	0.40
β_{chgbefor}	1.19	1.05	0.88	1.66	0.90	0.54
β_{chgafter}	0.70	0.43	0.61	1.08	0.52	0.49
β_{refund}	0.50	1.66	3.34	1.53	0.91	0.60

Table 12: Implied WTP measures for third case study

		LC (conf.)†		LC (expl.)		MMNL		LC-MMNL†	
	MNL	μ	σ	μ	σ	μ	σ	μ	σ
Accrual of LANPASS kms (1,000s pesos)	2.65	2.09	2.78	4.88	9.06	6.53	21.76	3.95	8.12
Regular seat reservation (1,000s pesos)	0.78	0.52	0.20	1.02	1.48	1.18	1.76	0.81	1.24
Preferential seat reservation (1,000s pesos)	1.34	0.89	0.79	1.85	2.84	2.23	4.78	1.49	2.52
Change before departure (1,000s pesos)	8.52	6.02	5.21	13.01	16.51	16.32	35.50	10.64	17.12
Change after outbound departure (1,000s pesos)	5.19	3.63	3.41	9.62	7.92	9.59	17.73	6.45	10.55
Refund for unused tickets (1,000s pesos)	2.66	1.90	3.96	8.66	6.71	6.84	42.29	3.66	11.04

†: respondents in *non-attendance* class for fare excluded from calculations

significant when looking at the adjusted ρ^2 measure. As in the previous two case studies, we observe some changes in significance for the individual parameters when moving from MMNL to LC-MMNL.

By allowing for a class with zero sensitivities, we observe the usual increases in marginal sensitivities when moving from MNL to LC, and also in the mean sensitivities when moving from MMNL to LC-MMNL. In the latter, the means of the underlying normal distributions are no longer significant for $\mu_{\ln(\beta_{\text{accr. km}})}$ and $\mu_{\ln(\beta_{\text{chg. after}})}$, noting that this does not imply the same to be true for the marginal utilities after transformation.

The results from Table 10 show quite a varied picture in terms of implied attribute non-attendance in the confirmatory LC model, ranging from 13% for the regular seat attribute to 81% for the refund

attribute. The propensity of non-attendance for ticket price is in line with findings from a similar choice setting in Rose et al. (2011). As mentioned above, in the LC-MMNL models, the implied rates of non-attendance drops to zero for the regular seat attribute which, notably, was the attribute with the lowest retrieved rate of non-attendance in the confirmatory LC model. It is also worth noting the high share for β_2 for this coefficient in the exploratory model - this is a reflection of a large share of respondents with low but not zero sensitivity, and in line with the estimate for $\beta_{\text{regseat},1}$ in the confirmatory model. Moreover, we also observe reductions in the rates for all other attributes. With this specific dataset, a non-trivial rate of attribute non-attendance is realistic, given the use a set of qualitative attributes that may not be relevant to all respondents. The 9% rate of implied non-attendance for the fare attribute however still remains an issue. In the exploratory LC model, the estimates in both classes are statistically significant for all marginal utility coefficients except $\beta_{\text{chgafter},2}$ and $\beta_{\text{refund},2}$. This possibly suggests a share of respondents with real non-attendance for these attributes, where this share is however much lower than in the confirmatory LC model, or indeed the LC-MMNL model.

As expected, the results in Table 11 show reduced continuous heterogeneity in the LC-MMNL model compared to the MMNL model, with one exception, namely the regular seat attribute. This is the only attribute for which the implied rate for non-attendance in the LC-MMNL model is zero. For the remaining attributes, part of what was previously captured as heterogeneity in the MMNL model is now captured through the non-attendance class. For β_{regseat}, the only explanation is that the heterogeneity for this attribute was captured by other random coefficients, i.e. the same argument of confounding that we used in the first case study for the three attributes where increases in heterogeneity were observed.

Turning finally to the implied WTP patterns in the non-ignoring classes, the changes between MNL and confirmatory LC are this time slightly less pronounced, except for the accrual of LANPASS kms and the ability to obtain a refund for unused tickets. The most pronounced increase in WTP, is found, similarly to the second case study, for attributes with the highest non-attendance rates. When moving to the exploratory LC model, we observe increases for all WTP measures, including β_{chgafter} and β_{refund} for which a class with zero sensitivity was retrieved. These heightened WTP measures are the result of the low fare sensitivity identified in the second class, with a share of 41%. The WTP measures for the MMNL model are higher, while we see a drop in all mean WTP measures when moving to the LC-MMNL model, as a result of these calculations not taking into account the non-attendance class for fare.

4 Conclusions

In the face of ever more widespread use of confirmatory latent class structures for capturing attribute non-attendance, this paper has asked the question whether results showing high rates of implied non-attendance in such models are possibly misleading. We argue that a confirmatory LC structure where one support point is fixed to zero for each coefficient is vulnerable to confounding, where the phenomena captured by the model may relate more to taste heterogeneity, including low attribute importance, than to actual attribute non-attendance.

We investigate the issue using two different approaches. Firstly, we compare the confirmatory LC approach to an exploratory LC model in which both values are estimated for each coefficient, rather than one being fixed to zero. Secondly, we offer an alternative approach by putting forward the use of a model that additionally allows for continuous random heterogeneity, in the form of a combined LC-MMNL structure, where the coefficients in one class are constrained to zero while the second class allows for continuously distributed coefficients rather than just a point value. The hope here is that the continuous representation of heterogeneity in one class will allow the model to capture both high and

low attribute sensitivity, increasing the chance that what is captured in the *non-attendance* class really is attribute non-attendance, rather than just low sensitivity.

Our findings from three separate empirical case studies seem to give definite support to our assertion that the confirmatory LC model is likely to produce misleading results for attribute non-attendance, with the class at zero serving as a proxy for capturing continuous variations in sensitivity across respondents. Across the three studies, the confirmatory LC model is outperformed by the exploratory LC model, which shows just a single coefficient out of six with value that is not significantly different from zero in one class in the first case study, two out of five in the second case study, and two out of seven in the third case study. This suggests that some of the apparent patterns of non-attendance retrieved in the confirmatory LC model were a result of confounding. Clearly, with such an exploratory LC model, there is also risk of confounding in the opposite direction (cf. Hole, 2011b). Indeed, if the true share of non-attendance is small, and there is substantial remaining (i.e. non-zero) heterogeneity in the data, then it is likely that both classes will retrieve significant sensitivities. To some extent, this could be addressed by gradually increasing the number of classes, so that eventually the weight required for a single class would be small enough to capture the non-attendance. An example of working with up to three attendance and one non-attendance class is given by Campbell et al. (2012). However, increases in the number of classes will often lead to problems with inferior local optima and classes collapsing to the same value.

The experience with the LC-MMNL model in the present paper is that, across all three case studies, and across most coefficients, we see significant reductions in the shares for the non-attendance classes when additionally allowing for random heterogeneity within non-zero classes. In several cases the apparent non-attendance even drops to zero. Across the three LC models (confirmatory LC, exploratory LC, LC-MMNL), the results are consistent across models for some of the coefficients (e.g. wifi and flexibility in the second case study), while in others, the confirmatory LC model is affected by substantial confounding. This is arguably a strong indication that at least a part of what is typically captured by such classes in a confirmatory model is in fact not attribute non-attendance, i.e. zero sensitivity, but simply taste heterogeneity, and reduced attribute importance. This finding also calls for the reappraisal of a large body of empirical work making use of the simple confirmatory LC structure and showing significant shares of attribute non-attendance. As an aside, it should be noted that in our three case studies, the simple MMNL model always gives better fit than the confirmatory LC model (and indeed the exploratory LC model), suggesting that it is indeed able to give a more complete representation of the random heterogeneity.

As always, it is important to acknowledge the limitations of our study. Firstly, the choice of continuous distribution for β is sure to have an impact on the weight of the non-attendance classes. In our case studies, the Lognormal distribution had been found to work well with the data at hand in previous work. Additionally, it is arguably a distribution that is especially well suited for dealing with a mix of low and high attribute importance. Conversely, analysts should also use care as the distribution for the cost coefficient may be pulled towards zero in datasets where non-attendance really arises for this attribute, and this would lead to very long tails for the distribution of WTP measures. Secondly, an argument could be made to compare the LC-MMNL model used here to one with freely estimated β_k in both classes (potentially with random heterogeneity in both), much in the same way as we estimated both values in the exploratory LC model. Our results may thus still be affected by confounding, but hopefully to a lesser extent than applications using the standard confirmatory LC approach. Thirdly, and despite the above comment regarding identification, future work should also look into using more than two classes. Finally, it would also be of interest to revisit the specification of the class probabilities, and in particular, whether an independence assumption such as used in the present paper is justified more broadly.

In closing, it should also be argued that while attribute non-attendance may be an interesting concept from a behavioural perspective, it poses significant complications in the interpretation of results, and especially the computation of WTP indicators when there is a certain share of respondents with implied non-attendance for the cost attribute. Indeed, it is for this reason that we have paid reduced attention to computing WTP measures in the present analysis - the presence of a non-attendance class for cost rules out WTP computation for a share of the sample population. If cost is fully attended, the problems are less severe, and we simply obtain a zero WTP for a given share of respondents for given attributes; in our case, this situation would only occur in the first case study where only non-cost attributes have a risk of non-attendance in the LC-MMNL model. If there is a non-zero probability for a non-attendance class for cost, then the WTP measures produced by an analyst are likely to be underestimated as this class cannot be taken into account in WTP computation; clearly, if this behaviour is simply captured through low sensitivity in a MMNL structure, no such problems arise.

More crucially, the question can be asked whether non-attendance for cost actually makes sense. It is likely to be an SP artifact rather than real life behaviour, so these respondents should not be used in WTP computation anyway. To some extent, the same argument can be made for other continuous attributes. Any attribute non-attendance retrieved in a stated preference context clearly only relates to the ranges of the attributes used in the survey. Finally, we should also note that model estimates may not be caused by actual non-attendance, but by the fact that our designs do not include sufficient scenarios in which the given attribute can influence the choice, a point that applies especially for *less important* attributes.

As usual, there are thus plenty of unanswered questions, but reducing non-attendance classes by capturing heterogeneity in a more adequate manner is certainly a first step to avoiding some of the problem. The evidence from both the exploratory LC model and the LC-MMNL model have shown conclusively that the rates of non-attendance retrieved in the widely used confirmatory LC model are likely to be severely overstated. For a number of reasons, including WTP computation, it could be preferable to simply capture such behaviour within a random heterogeneity specification, losing some insights into behaviour but avoiding a number of other issues.

Acknowledgements

A much earlier version of this paper was presented at the European Transport Conference in Glasgow in October 2011.

References

Alemu, M. H., Mørbak, M. R., Olsen, S. B., Jensen, C. L., 2011. Attending to the reasons for attribute non-attendance in choice experiments. paper presented at the Second International Choice Modelling Conference, Oulton Hall, Leeds.

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

Cameron, T., DeShazo, J., 2011. Differential attention to attributes in utility-theoretic choice models. Journal of Choice Modelling 3 (3), 73–115.

Campbell, D., Hensher, D., Scarpa, R., 2012. Cost thresholds, cut-offs and sensitivities in stated choice analysis: identification and implications. Resource and Energy Economics 34 (3), 396–411.

Campbell, D., Hensher, D. A., Scarpa, R., 2011. Non-attendance to attributes in environmental choice analysis: a latent class specification. Journal of Environmental Planning and Management 54 (8), 1061–1076.

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

Carlsson, F., Kataria, M., Lampi, E., 2010. Dealing with Ignored Attributes in Choice Experiments on Valuation of Sweden's Environmental Quality Objectives. Environmental & Resource Economics 47 (1), 65–89.

Collins, A., Rose, J., Hensher, D., 2012. The random parameters attribute non-attendance model. Paper presented at the 13^{th} International Conference on Travel Behaviour Research, Toronto, Canada.

Doornik, J. A., 2001. Ox: An Object-Oriented Matrix Language. Timberlake Consultants Press, London.

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

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

Hensher, D. A., Greene, W. H., 2010. Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification. Empirical Economics 39 (4), 413–426.

Hensher, D. A., Rose, J. M., Greene, W. H., 2005. The implications on willingness to pay of respondents ignoring specific attributes. Transportation 32 (3), 203–222.

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

Hess, S., Hensher, D., 2012. Making use of respondent reported processing information to understand attribute importance: a latent variable scaling approach. Transportation, in press.

Hess, S., Hensher, D. A., 2010. Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. Transportation Research Part B 44 (6), 781–790.

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

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

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

Hole, A. R., 2011b. Attribute non-attendance in patients' choice of general practitioner appointment. paper presented at the Second International Choice Modelling Conference, Oulton Hall, Leeds.

Puckett, S. M., Hensher, D. A., 2008. The role of attribute processing strategies in estimating the preferences of road freight stakeholders. Transportation Research Part E: Logistics and Transportation Review 44 (3), 379–395.

Rose, J. M., Hensher, D. A., Greene, W. H., Washington, S., 2011. Attribute Exclusion Strategies in Airline Choice: Accounting for Exogenous Information on Decision Maker Processing Strategies in Models of Discrete Choice. Transportmetrica, forthcoming.

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

Scarpa, R., Thiene, M., Hensher, D. A., 2010. Monitoring choice task attribute attendance in nonmarket valuation of multiple park management services: Does it matter? Land Economics 86 (4), 817–839.