Accounting for Attribute Non-Attendance and Common-Metric Aggregation in a Probabilistic Decision Process Mixed Multinomial Logit Model: A Warning on Potential Confounding

David A. Hensher
Andrew T. Collins
Institute of Transport and Logistics Studies
The Business School
The University of Sydney, NSW 2006 Australia
David.Hensher@sydney.edu.au
Andrew.Collins@sydney.edu.au

William H. Greene
Department of Economics
Stern School of Business, New York University, New York 10012
wgreene@stern.nyu.edu

27April 2012 (revised August 24 2012, November 3 2012)
To appear in Transportation

Abstract

Latent class models offer an alternative perspective to the popular mixed logit form, replacing the continuous distribution with a discrete distribution in which preference heterogeneity is captured by membership of distinct classes of utility description. Within each class, preference homogeneity is usually assumed, although interactions with observed contextual effects are permissible. A natural extension of the fixed parameter latent class model is a random parameter latent class model which allows for another layer of preference heterogeneity within each class. A further extension is to overlay attribute processing rules such as attribute non-attendance (ANA) and aggregation of common-metric attributes (ACMA). This paper sets out the random parameter latent class model with ANA and ACMA, and illustrates its application using a stated choice data set in the context of car commuters and non-commuters choosing amongst alternative packages of travel times and costs pivoted around a recent trip in Australia. What we find is that for the particular data set analysed, in the presence of attribute processing together with the discrete distributions defined by latent classes, that adding an additional layer of heterogeneity through random parameters within a latent class only very marginally improves on the statistical contribution of the model. Nearly all of the additional fit over the fixed parameter latent class model is added by the account for attribute processing. This is an important finding that might suggest the role that attribute processing rules play in accommodating attribute heterogeneity, and that random parameters within class are essentially a potentially confounding effect. An interesting finding, however, is that the introduction of random parameters increases the probability of membership to full attribute attendance classes, which may suggest that some individuals assign a very low marginal disutility (but not zero) to specific attributes or that there are very small differences in the marginal disutility of common-metric attributes, and this is being accommodated by random parameters, but not observed under fixed parameter latent class model.

1. Introduction

The focus of the majority of discrete choice modelling over the last ten years has been centred on the random parameter (or mixed logit) model, and more recently on extensions of mixed logit to incorporate scale heterogeneity and estimation in willingness to pay space (see for example Train and Weeks 2005, Fiebig *et al.* 2010, Greene and Hensher 2010). The family of mixed logit models assume a continuous distribution for the random parameters across the sampled population.

Latent class models offer an alternative perspective, replacing the continuous distribution with a discrete distribution in which preference heterogeneity is captured by membership in distinct classes of utility description (see Greene and Hensher 2003 and Shen 2009). Within each class, preference homogeneity is usually assumed, although interactions with observed contextual effects are permissible. A natural extension of the fixed parameter latent class model is a random parameter latent class model which allows for another layer of preference heterogeneity within a class (e.g., Greene and Hensher 2012, Bujosi *et al.* 2010, Vij *et al.* 2012, Wasi and Carson 2011). Vij *et al.* (2012) also introduce error components. Wasi and Carson (2011) use the mixture of normals approach from Train (2008).

A further extension is to overlay attribute processing rules in a random parameter latent class model that allows for attribute non-attendance (ANA) and aggregation of common-metric attributes (ACMA). Collins *et al.* (2012) and Hess *et al.* (2011) have recently implemented this model for ANA. Given that each latent class can be defined by a set of parameter restrictions that represent a particular ANA or ACMA processing rule, we refer to the model as a probabilistic decision process model instead of a latent class model, even though the use of a latent class function (with restrictions) provides the econometric method to identify the role, up to a probability, of each assessed attribute processing rule, including full attribute attendance (FAA).

The focus in the current paper is on a very specific situation according to which respondents to a stated choice experiment adopt an attribute processing rule under which specific attributes are either fully attended to, ignored (or "non-attended to" to use the terminology of some environmental economists), and/or are added up where they exhibit a common metric, for all manner of reasons. Earlier efforts of Hensher *et al.* (2005), Hensher (2006), and Layton and Hensher (2010) highlighted the real possibility that such attribute processing strategies do make a difference in estimates of willingness of pay for specific attributes (e.g., the value of travel time savings). Subsequent research by Hensher and Rose (2009), Hess and Hensher (2010), Scarpa *et al.* (2009, 2010), Campbell *et al.* (2008, 2011), Carlsson et al (2010), and Puckett and Hensher (2008) amongst others, has reinforced the view that accounting for "attribute non-attendance" does impact significantly on key behavioural outputs.

A number of the stated choice studies cited above based their identification of ANA or ACMA on supplementary questions designed to establish whether a respondent had ignored an attribute or not, or added up attributes: they could be asked either after each choice set as in Puckett and Hensher (2008) or after completing all choice scenario assessments, as is more common. However, numerous papers have provided evidence that responses to such supplementary questions may not be reliable (Hess and Rose 2007, Hess and Hensher 2010, Carlsson *et al.* 2010). Although the jury is

still out on this issue, there is growing interest in the approach taken in this paper, i.e., identifying the role of ANA or ACMA through model inference, rather than directly asking each respondent. Some recent examples are Scarpa *et al.* (2009), Hensher and Greene (2010), Hole (2011), Campbell *et al.* (2011) and Hensher *et al.* (2012b); all utilising a latent class structure with fixed parameters.

This paper uses ideas from the studies cited above to investigate the potential value adding contribution of integrating latent class models with random parameter overlays within each class under the assumption of full attribute attendance (FAA) in contrast to mixtures of FAA, ANA and ACMA. Although Hess *et al.* (2011) comes closest to investigating similar issues (although focussing only on ANA), Hess *et al.* impose strong restrictions on the parameters identified, namely equality constraints across classes. In the current paper we impose no such constraints and seek out the role that random preference heterogeneity plays in the presence and absence of accounting for specific attribute processing rules. The findings provide evidence on whether additional random parameter layering within a latent class framework that accounts for ANA and ACMA is significant, or whether it is a potentially confounding source. This paper also adds to the growing evidence that the handling of ANA with latent classes and fixed parameters is susceptible to exaggerating the ANA rate by capturing low taste sensitivities as ANA (Campbell *et al.* 2011, Hess *et al.* 2011, Collins *et al.* 2012).

We illustrate its application using a stated choice data set in the context of car commuters and non-commuters choosing amongst alternative packages of travel times and costs (including tolls) pivoted around a recent trip in Australia. Although we investigated more than one data set, we focus herein on one very substantive data set which represents the evidence obtained using other data sets.

2. A Random Parameter Probabilistic Decision Process Model with FAA, ANA and ACMA

Latent class modelling provides an alternative approach to accommodating heterogeneity in contrast to models such as multinomial logit and mixed logit (see Everitt 1988). The natural approach assumes that parameter vectors, β_i are distributed among individuals with a discrete distribution, rather than the continuous distribution that lies behind the mixed (or random parameter) logit model. In our adaptation of the latent class model, it is assumed that the population consists of a finite number, Q, of groups of individuals, where each class is pre-defined by an attribute processing rule (APR). The groups are heterogeneous, with common parameters, β_q, for the members of the group, but the groups themselves are different from one another in line with the APR. We assume that the classes are distinguished by the different parameter vectors (which can, but do not necessarily have to be, equality constrained)[1], though the fundamental data generating process, the probability density for the variable under study, is the same.

---

[1] The set of references have papers that impose equality constraints and also papers which relax this condition. There is no de-facto standard practice; for example Hensher and Greene (2010, Table 2) have different parameters across the classes, for each attribute, as do Campbell et al. (2011, Table 1).

3

The analyst does not know from the data which observation is in which APR class, even if the specific attribute processing rule is imposed on a class, hence the term *latent* classes. The model assumes that individuals are distributed heterogeneously with a discrete distribution in a population. The full specification of the latent class structure for a generic data generating process is (Greene and Hensher 2003):

$$f(y_i|\mathbf{x}_i, \text{class} = q) \; = \; g(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_q) \tag{1}$$

$$\text{Prob}(\text{class} = q) = \; \pi_q(\boldsymbol{\theta}), \, q = 1,...,Q. \tag{2}$$

$\boldsymbol{\theta}$ is the vector of latent class parameters attached to candidate sources of systematic influence on class membership. The unconditional probability attached to an observation is obtained by integrating out the heterogeneity due to the distribution across classes, given in (3).

$$f(y_i|\mathbf{x}_i) \; = \; \Sigma_q \, \pi_q(\boldsymbol{\theta}) g(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_q). \tag{3}$$

To extend the latent class model to allow for two layers of heterogeneity (and the imposition of an APR within a class) both within and across groups, we allow for continuous variation of the parameters within classes (Greene and Hensher 2012). The latent class feature of the model is given as (4) and (5).

$$f(y_i|\mathbf{x}_i, \text{class} = q_{APR}) \; = \; g(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_{i|q_{APR}}) \tag{4}$$

$$\text{Prob}(\text{class} = q_{APR}) = \; \pi_q(\boldsymbol{\theta}), \, q_{APR} = 1,...,Q_{APR}. \tag{5}$$

The within-class heterogeneity, $\mathbf{w}_i$, is structured as

$$\boldsymbol{\beta}_{i|q_{APR}} \; = \; \boldsymbol{\beta}_{q_{APR}} \; + \; \mathbf{w}_{i/q_{APR}} \tag{6}$$

$$\mathbf{w}_{i/q_{APR}} \; \sim \; E[\mathbf{w}_{i/q_{APR}}|\mathbf{X}] \; = \; \mathbf{0}, \; \text{Var}[\mathbf{w}_{i/q_{APR}} \mid \mathbf{X}] \; = \; \boldsymbol{\Sigma}_{q_{APR}} \tag{7}$$

where the $\mathbf{X}$ indicates that $w_{i/q_{APR}}$ is uncorrelated with all exogenous data in the sample. We will assume below that the underlying distribution for the within-class heterogeneity is normal, with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. In a given application, it may be appropriate to further assume that certain rows and corresponding columns of $\boldsymbol{\Sigma}_{q_{APR}}$ equal zero, indicating that the variation of the corresponding parameter is entirely across classes. With random parameters the estimator does a lot of simulation. There is one random parameter model in each class, so, for example, if the model is specified with three random parameters in the model and three latent classes, there are nine random parameters; consequently computation time increases significantly. Overlaying this specification is the APR for each class.

The extension to incorporate ANA and ACMA is relatively simple, in a conceptual sense. To account for possible heuristics defined in the domain of ANA and ACMA, we impose restrictions on the mean and standard deviation parameters defining each random parameter within each latent class, each class representing a particular process heuristic. For example, to impose the condition of non-attendance of a specific attribute, we set its parameter to zero; to impose ACMA we constrain the parameters of two or more attributes to be equal.

4

The contribution of individual i to the log likelihood for the model is obtained for each individual in the sample by integrating out the within-class heterogeneity associated with continuous random parameters and ANA and ACMA restrictions on the mean and standard deviation of the relevant random parameters, and then the class heterogeneity. We allow for a stated choice data setting, hence the observed vector of outcomes is denoted $\mathbf{y}_i$ and the observed data on exogenous variables are collected in $\mathbf{X}_i = [\mathbf{X}_{i1},..,\mathbf{X}_{iT_i}]$. An individual is assumed to engage in T_i choice situations, where $T_i \geq 1$. The generic model is given in (8).

$$f(\mathbf{y}_i|\mathbf{X}_i,\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_{Q_{APR}},\boldsymbol{\theta},\boldsymbol{\Sigma}_1,...,\boldsymbol{\Sigma}_{Q_{APR}})$$
$$= \sum_{q_{APR}=1}^{Q_{APR}} \pi_{q_{APR}}(\boldsymbol{\theta}) \int_{w_i} \prod_{t=1}^{T_i} f[\mathbf{y}_{it}|(\boldsymbol{\beta}_{q_{APR}}+\mathbf{w}_i),\mathbf{X}_{it}]h(\mathbf{w}_i|\boldsymbol{\Sigma}_{q_{APR}})d\mathbf{w}_i. \qquad (8)$$

We parameterise the class probabilities using a multinomial logit formulation to impose the adding up and positivity restrictions on $\pi_{q_{ANA}}(\boldsymbol{\theta})$. Thus, the class probability model becomes:

$$\pi_{q_{APR}}(\boldsymbol{\theta}) = \frac{\exp(\theta_{q_{APR}})}{\Sigma_{q_{APR}=1}^{Q_{APR}} \exp(\theta_{q_{APR}})}, \; q_{APR} = 1,...,Q_{APR}; \; \theta_{Q_{APR}} = 0. \qquad (9)$$

The model employed in this application is a probabilistic decision process latent class, mixed multinomial logit model with FAA, ANA and ACMA. Individual i chooses among J alternatives with conditional probabilities given as (10).

$$f[\mathbf{y}_{it}|(\boldsymbol{\beta}_{q_{APR}}+\mathbf{w}_i),\mathbf{X}_{it}] = \frac{\exp[\Sigma_{j=1}^{J} y_{it,j}(\boldsymbol{\beta}_{q_{APR}}+\mathbf{w}_i)'\mathbf{x}_{it,j}]}{\sum_{j=1}^{J} \exp[(\boldsymbol{\beta}_{q_{APR}}+\mathbf{w}_i)'\mathbf{x}_{it,j}]}, \; j = 1,...,J, \qquad (10)$$

$y_{it,j} = 1$ for the j corresponding to the alternative chosen and 0 for all others, and $\mathbf{x}_{it,j}$ is the vector of attributes of alternative j for individual i in choice situation t.

We use maximum simulated likelihood to evaluate the terms in the log likelihood expression. The contribution of individual i to the simulated log likelihood is the log of equation (11).

$$f^S(\mathbf{y}_i|\mathbf{X}_i,\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_{Q_{APR}},\boldsymbol{\theta},\boldsymbol{\Sigma}_1,...,\boldsymbol{\Sigma}_{Q_{APR}})$$
$$= \sum_{q_{APR}=1}^{Q_{APR}} \pi_{q_{APR}}(\boldsymbol{\theta})\frac{1}{R}\sum_{r=1}^{R} \prod_{t=1}^{T_i} f[\mathbf{y}_{it}|(\boldsymbol{\beta}_{q_{APR}}+\mathbf{w}_{i,r}),\mathbf{X}_{it}] \qquad (11)$$

$\mathbf{w}_{i,r}$ is the rth of R random draws on the random vector $\mathbf{w}_i$. Collecting all terms, the simulated log likelihood is given as (12).

$$\log L^S = \sum_{i=1}^{N} \log\left[\sum_{q_{APR}=1}^{Q_{APR}} \pi_q(\boldsymbol{\theta})\frac{1}{R}\sum_{r=1}^{R} \prod_{t=1}^{T_i} f[\mathbf{y}_{it}|(\boldsymbol{\beta}_{q_{APR}}+\mathbf{w}_{i,r}),\mathbf{X}_{it}]\right] \qquad (12)$$

In the presence of ANA and ACMA, it differs from more familiar formulations of latent class models in that the nonzero elements in $\boldsymbol{\beta}_q$ can be allowed to be the same or free across the classes, and the classes have specific behavioural meaning, as opposed

5

to merely being groupings defined on the basis of responses, as in the strict latent class formulation, hence the reference to a probabilistic decision process model. Estimation of the probabilistic decision process model is straightforward as a latent class MNL model with linear constraints on the coefficients, as suggested above.

Willingness to pay (WTP) estimates for the value of total travel time savings ($/hour) are computed using the familiar result, WTP = -$\beta_{\text{time}}/\beta_{\text{cost}}$. Since there is heterogeneity of the parameters within the classes as well as across classes, the result is averaged to produce an overall estimate. The averaging is undertaken for the random parameters within each class then again across classes using the posterior probabilities as weights. Collecting the results, the procedure is

$$\widehat{WTP} = \frac{1}{N}\sum_{i=1}^{N} \ \sum_{qAPR=1}^{QAPR} \ \left\{\pi_{qAPR}\left(\hat{\theta}\right)\big|i\right\} \left[\frac{\frac{1}{R}\sum_{r=1}^{R} L_{ir|qAPR} \dfrac{-\hat{\beta}_{time,ir|qAPR}}{\hat{\beta}_{cost,ir|qAPR}}}{\frac{1}{R}\sum_{r=1}^{R} L_{ir|q}}\right] \qquad (13)$$

$$= \frac{1}{N}\sum_{i=1}^{N} \ \sum_{qAPR=1}^{QAPR} \ \left\{\pi_{qAPR}\left(\hat{\theta}\right)\big|i\right\}\frac{1}{R}\sum_{r=1}^{R} W_{ir|qAPR}\widehat{WTP_{time,ir|qAPR}},$$

where R is the number of draws in the simulation and r indexes the draws,

$$\hat{\beta}_{time,ir|qAPR} = \hat{\beta}_{time|qAPR} + \hat{\sigma}_{time|qAPR}w_{time,ir|qAPR}$$

and likewise for $\hat{\beta}_{cost,ir|qAPR}$, $L_{ir|qAPR}$ is the contribution of individual i to the class specific likelihood - this is the product term that appears in (11) and (12), and $\pi_{qAPR}\left(\hat{\theta}\right)\big|i$ is the estimated posterior class probability for individual i;

$$\pi_{qAPR}\left(\hat{\theta}\right)\big|i = \frac{\pi_{qAPR}(\hat{\theta})\frac{1}{R}\sum_{r=1}^{R} \ \prod_{t=1}^{T_i} \ f[\mathbf{y}_{it}\,|\,(\hat{\beta}_{q_{APR}}+\mathbf{w}_{i,r}),\mathbf{X}_{it}]}{\sum_{q_{APR}=1}^{Q_{APR}} \ \pi_{qAPR}(\hat{\theta})\frac{1}{R}\sum_{r=1}^{R} \ \prod_{t=1}^{T_i} \ f[\mathbf{y}_{it}\,|\,(\hat{\beta}_{q_{APR}}+\mathbf{w}_{i,r}),\mathbf{X}_{it}]}. \qquad (14)$$

Our model bears some resemblance to Hess *et al.* (2011), and Collins *et al.* (2012), but there are some key differences. These papers impose equality constraints on the fixed and random parameters across the classes, and allow for between class variation solely through constraining these parameters to zero, to represent ANA. This goes some way to reducing the proliferation of parameters as the number of classes increases[2]. The number of class allocation parameters will increase exponentially as the number of classes increases. Hess *et al.* (2011) ameliorates this problem through an alternative, more parsimonious class allocation model, first employed by Hole

---

[2] Hess et al. (2011) use uncorrelated lognormally distributed random parameters. Collins et al. (2012) use lognormal, constrained triangular and Rayleigh distributions, all uncorrelated. The appeal of these two models is that the imposed restrictions reduce the number of free parameters through constraining the random parameter associated with a specific attribute to be generic across classes. If, however the parameters associated with specific attributes under a specific APR are not independent of the role of the other attributes under consideration within a specific APR, then this may be a limiting condition.

(2011) in the context of fixed parameters. Collins *et al.* (2012) present a more flexible class allocation model, with the conventional latent class allocation model and the Hole (2011) approach serving as extremes. The intermediate specifications allow the independence of ANA assumption on which the Hole (2011) approach relies to be selectively relaxed, albeit at a cost in parameters.

3. Application

The data are drawn from a study undertaken in Australia in the context of toll vs. free roads, which utilised a stated choice (SC) experiment involving two SC alternatives (i.e., route A and route B), which are pivoted around the knowledge base of travellers (i.e., the current trip). The trip attributes associated with each route are summarised in Table 1. The data have been used in a number of other applications (e.g., Li *et al.* 2010, Hensher *et al.* 2012a), but not in the methodological context of the current study

Table 1: Trip Attributes in Stated Choice Design

Routes A and B
Free flow travel time
Slowed down travel time
Stop/start/crawling travel time
Minutes arriving earlier than expected
Minutes arriving later than expected
Probability of arriving earlier than expected
Probability of arriving at the time expected
Probability of arriving later than expected
Running cost
Toll Cost

Each alternative has three travel scenarios - 'arriving *x* minutes earlier than expected', 'arriving *y* minutes later than expected', and 'arriving at the time expected'. Each is associated with a corresponding probability[3] of occurrence to indicate that travel time is not fixed but varies from trip to trip. For all attributes except the toll cost, minutes arriving early and late, and the probabilities of arriving on-time, early or late, the values for the SC alternatives are variations around the values for the current trip. Given the lack of exposure to tolls for many travellers in the study catchment area, the toll levels are fixed over a range, varying from no toll to $4.20, with the upper limit determined by the trip length of the sampled trip.

In the choice experiment, the first alternative is described by attribute levels associated with a recent trip; with the levels of each attribute for Routes A and B pivoted around the corresponding level of actual trip alternative with the probabilities of arriving early, on time and late provided. Commuters and non-commuters in a Metropolitan area in Australia were sampled (with employer-related business trips also included in the full data set but excluded from the analysis in this paper[4]). A

---

[3] The probabilities are designed and hence exogenously induced to respondents, similar to other travel time variability studies. We acknowledge the contribution of John Rose in the design of the choice experiment.
[4] We removed the employer business sample because we found that it is so different in the way attributes matter, especially costs that are covered by an employer, and this was causing problems in model estimation compared to the remaining data in which the costs are paid for by the respondent.

telephone call was used to establish eligible participants from households. During the telephone call, a time and location were agreed for a face-to-face Computer Aided Personal Interview (CAPI). In total, 588 commuters and non-commuters (with less than 120 minutes' trip length) were sampled for this study, each responding to 16 choice sets (games), resulting in 9,408 observations for model estimation. The experimental design method of D-efficiency used herein is specifically structured to increase the statistical performance of the models with smaller samples than are required for other less-efficient (statistically) designs such as orthogonal designs (see Rose *et al.* 2008;). An illustrative choice scenario is given in Figure 1.
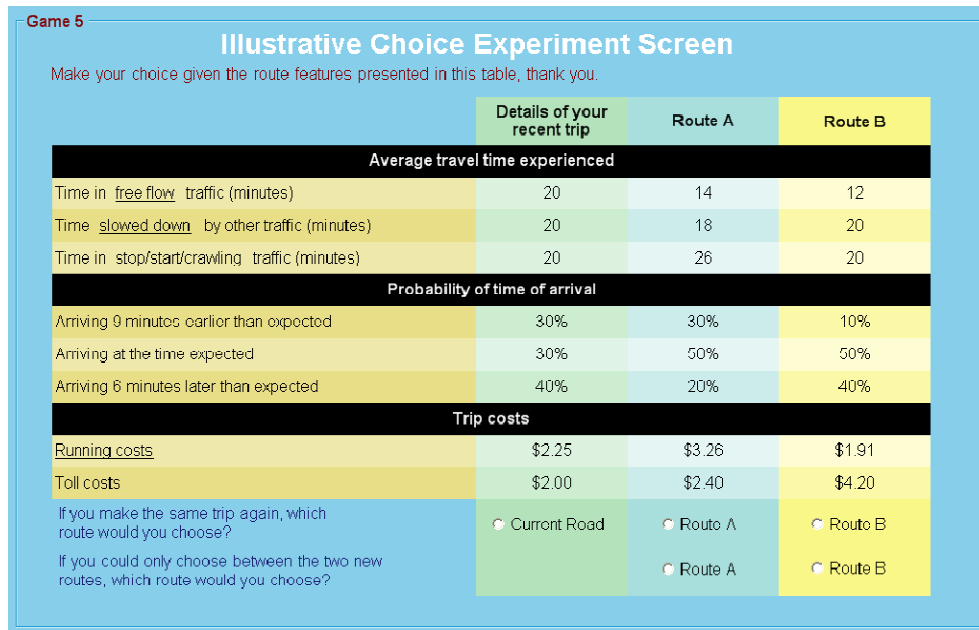

Figure 1: An Illustrative Choice Scenario

4. Results

The findings are presented in Table 2 for four models. The first and third models assume full attribute attendance, while the second and fourth models allow for mixtures of FAA, ANA and ACMA. The first two models assume fixed parameters for all attributes, while the third and fourth models include random parameters for the travel time and cost attributes. The random parameters are defined by a constrained triangular distribution with scale parameter equal to the mean estimate[5].

The development of the models follows a natural sequence from Model 1 through to Model 4. Under the assumption of FAA, selecting the number of classes is explained below, along lines well recognised in the latent class literature. We estimated model 1

---

[5] All models are estimated using Nlogit 5. We investigated unconstrained distributions including lognormal, but models either failed to converge or produced imprecise parameter estimates, most notably on the standard deviations of the random parameters. This is consistent with Collins et al. (2012), who found that constraining the sign of the random parameter distribution is necessary when ANA is handled through latent classes. Literally over 100 hours of model estimation time was undertaken in the estimation of the random parameter versions of the models.

8

under alternative numbers of classes (ranging from two through to seven classes), with five classes having the best overall goodness-of-fit (including AIC). When we move to Model 2, which assumes fixed parameters, but introduces ANA and ACMA, we have to define the ANA and ACMA classes and investigate how many classes should remain as FAA. In earlier studies (e.g., Scarpa *et al.* 2009, Campbell *et al.* 2011, McNair *et al.* 2012), users of latent class models (including Hensher and Greene 2010) imposed only one FAA class when investigating attribute processing rules; however there are good reasons why a number of classes might be considered (just as in Model 1) given that taste heterogeneity can continue to exist between FAA classes in the presence of elements of attribute processing. The introduction of multiple FAA classes may also go some way to reducing the chance that the attribute processing classes end up capturing taste heterogeneity as well as attribute processing, which is a risk when the taste coefficients are not constrained across classes. Model 2 is the final model under fixed parameters, varying the number of FAA classes and specific ANA and ACMA classes. Model 3 overlays Model 1 with random parameters on the four time and cost attributes of the choice experiment; however the number of classes is reduced to four on overall goodness-of-fit. Estimating this model takes many hours. Finally we introduce Model 4, which builds on all previous models, and is also freely defined in terms of the number of FAA classes, the ANA and ACMA classes and the distributional assumptions imposed on each random parameter[6]. Estimation of Model 4 also takes many hours with many models failing to converge (see footnote 4). The number of FAA classes under random parameters is reduced to one compared to three under fixed parameters (Model 2), possibly suggesting that some amount of preference heterogeneity that was accommodated through three classes under fixed parameters in model 2 has been captured in a single class in model 4 through within class preference heterogeneity. The final four models reported below are the outputs of this process.

A question naturally arises, how can the analyst determine the number of classes, *Q*? Since *Q* is not a free parameter, a likelihood ratio test is not appropriate, though, in fact, log *L* will increase when Q increases. Researchers typically use an information criterion, such as AIC, to guide them toward the appropriate value. For Model 1, the AIC was the lowest for five classes, at 1.035 (log-likelihood of -4817.72); whereas for Model 3 with random parameters, we found four classes had the lowest AIC (1.033 and a log-likelihood of -4803.2, slightly better than Model 1). Heckman and Singer (1984) also suggest a practical guidepost in selecting the number of classes; namely that if the model is fit with too many classes, estimates will become imprecise, even varying wildly. Signature features of a model that has been overfit include exceedingly small estimates of the class probabilities, wild values of the structural parameters, and huge estimated standard errors. For the models that account for ANA and ACMA (Models 2 and 4), the number of classes is pre-defined by the number of restrictions on parameters that are imposed to distinguish the attribute processing strategies of interest; however the number of classes with full attribute attendance is free and can be determined along the same lines as Models 1 and 3.

With respect to the ANA and ACMA conditions that might be imposed, authors have suggested that responses to supplementary questions on whether a respondent claims

---

[6] We investigated correlations amongst the random parameters in the unconstrained distributions; however such models were an inferior statistical fit over model 4 report in Table 2.

9

they ignore specific attributes and/or added them up may be useful to signal the possibility of specific attribute processing strategies. For the sample of 588 observations, the following incidence of reported ANA is obtained: free flow time (28%), slowed down and stop start time (27%), running cost (17%), and toll cost (11%). The incidence of ACMA is as follows: total time (60.5%) and total cost (80%). The reliability of such data has been questioned in many papers (see e.g. Hess and Hensher 2010); however, despite the concerns about the believability of such evidence, there is a case to support the presence of heterogeneity in attribute processing. The above incidence rates of ANA motivate the selection of the restrictions imposed on the models that account for ANA and ACMA; although the final classes were based on extensive investigation of alternative restrictions and show that the link with the stated supplementary question responses is tenuous (in line with the general trend in the literature). The number of classes in Models 2 and 4 are determined by the number of attribute processing rules of interest plus an assessment of the number of FAA classes using AIC as a statistical guide, as well as the suggestions of Heckman and Singer.

Table 2 Summary of Models

Sample size N=588 observations. For random parameter models, we used constrained t-distributions and 500 Halton draws. T-ratios are in parentheses. Times in minutes and costs in dollars (AU$2008)
MNL model log-likelihood = -6729.90

Latent Class, Fixed Parameters, No ANA, No ACMA (Model 1)

Attributes	Class 1	Class 2	Class 3	Class 4	Class 5
Free flow time	-0.1945 (-7.81)	-0.0743 (-3.58)	-0.0398 (-4.41)	-0.0312 (-1.45)	-0.0033 (-0.26)
Slowed down and Stop/start/crawling time	-0.2360 (-10.5)	-0.1728 (-7.53)	-0.0782 (-6.46)	-0.1521 (-6.81)	-0.0559 (-4.96)
Running cost	-0.2723 (-3.89)	-2.2544 (-7.89)	-0.4155 (-8.86)	-1.5577 (-7.66)	-0.3854 (-5.53)
Toll cost	-0.2836 (-4.81)	-2.6709 (-8.11)	-0.3309 (-8.61)	-1.2353 (-8.33)	-0.1112 (-2.49)
Reference alt (1,0)	-0.1727 (-0.76)	-0.0823 (-0.38)	0.4211 (2.75)	3.9570 (9.26)	-2.1696 (-6.92)
Commuter trip purpose (1,0)	0.2368 (0.89)	3.2134 (8.72)	-2.8170 (-12.8)	-3.9950 (-7.92)	3.5705 (9.52)
Probability of early arrival	-0.0088 (-1.24)	-0.0303 (-2.88)	-0.0105 (-2.40)	-0.0011 (-0.13)	-0.0190 (-3.17)
Probability of late arrival	-0.0222 (-2.98)	-0.0398 (-3.43)	-0.0198 (-4.17)	-0.0349 (-4.22)	-0.0250 (-3.78)
Stated choice alt 1 (1,0)	-0.1022 (-0.77)	0.2673 (1.34)	-0.0568 (-0.74)	-0.2041 (-1.11)	0.0205 (0.20)
Class membership probability	0.124 (6.39)	0.309 (12.7)	0.184 (9.27)	0.277 (11.3)	0.107 (6.41)
Log-likelihood			-4817.72		
AIC/N			1.035		
McFadden pseudo-R^2			0.5339		

Latent Class, Fixed Parameters, FAA, ANA, ACMA (Model 2)

Attributes	FAA 1	FAA 2	FAA 3	ANA 1	ANA 2	ACMA
Free flow time	-0.0671 (-3.19)	-0.1749 (-6.92)	-0.0378 (-3.25)	fixed	-0.0572 (-3.67)	-0.0419 (-2.22)
Slowed down and Stop/start/crawling time	-0.1657 (-8.46)	-0.2175 (-13.5)	-0.0417 (-3.69)	-0.1590 (-5.10)	fixed	-0.0419 (-2.22)
Running cost	-2.7308 (-10.3)	-0.2396 (-3.61)	-0.7456 (-10.7)	-0.1628 (-2.20)	-0.3965 (-4.39)	-0.3603 (-4.14)
Toll cost	-2.6509 (-11.2)	-0.2831 (-5.75)	-0.5218 (-11.4)	0.0483 (-0.62)	-0.2912 (-4.07)	-0.3603 (-4.14)
Reference alt (1,0)	-0.2306 (-1.35)	-0.0025 (-0.01)	-0.0280 (-0.22)	-1.4529 (-3.50)	-4.1894 (-6.82)	3.8812 (12.0)
Commuter trip purpose (1,0)	0.3193 (1.45)	0.6366 (2.84)	-0.1712 (-0.86)	-0.1023 (-0.19)	0.4522 (0.44)	0.6772 (1.78)
Probability of early arrival	-0.0063 (-0.71)	-0.0093 (-1.62)	-0.0246 (-5.16)	-0.0073 (-0.90)	-0.0024 (-0.35)	0.0054 (0.37)
Probability of late arrival	-0.0396 (-4.66)	-0.0205 (-2.99)	-0.0390 (-7.23)	-0.0008 (-0.08)	-0.0109 (-1.42)	0.0009 (0.06)
Stated choice alt 1 (1,0)	-0.2523 (-2.75)	-0.0220 (-0.18)	-0.2523 (-2.75)	-0.1743 (-1.38)	0.3169 (2.55)	0.2824 (0.86)
Class membership probability	0.243 (10.8)	0.154 (7.08)	0.174 (7.66)	0.056 (3.95)	0.051 (4.95)	0.322 (14.3)
Log-likelihood			-4711.59			
AIC/N			1.013			
McFadden pseudo-R^2			0.5441			

Latent Class, Random Parameters, No ANA, No ACMA (Model 3)

Attributes	Class 1	Class 2	Class 3	Class 4
Random parameters (constrained triangular distribution)				
Free flow time	-0.0216 (-0.89)	-0.0237 (-1.59)	-0.0779 (-8.64)	-0.0366 (-6.03)
Slowed down and Stop/start/crawling time	0.00055 (0.03)	-0.0514 (-4.80)	-0.1116 (-17.9)	-0.0641 (-10.9)
Running cost	-0.7729 (-5.22)	-0.5957 (-7.99)	-0.5065 (-10.4)	-0.2229 (-6.24)

10

Toll cost	-1.3314 (-12.2)	-0.4316 (-5.52)	-0.2060 (-4.57)	-0.0879 (-2.61)
Fixed parameters				
Reference alt (1,0)	1.2552 (5.95)	0.3044 (1.61)	0.2439 (1.52)	-0.4626 (-3.37)
Commuter trip purpose (1,0)	-0.2665 (-0.77)	-0.5026 (-2.67)	-0.4992 (-3.90)	-0.5344 (-4.15)
Probability of early arrival	-0.0228 (-1.55)	-0.0038 (-0.62)	-0.0261 (-6.04)	-0.0096 (-2.51)
Probability of late arrival	0.0042 (0.24)	-0.0227 (-3.59)	-0.0269 (-5.54)	-0.0052 (-1.24)
Stated choice alt 1 (1,0)	-0.2283 (0.69)	-0.0231 (-0.18)	-0.0301 (-0.35)	0.0710 (1.16)
Class membership probability	0.367 (14.57)	0.202 (7.38)	0.211 (8.11)	0.221 (8.19)
Log-likelihood	-4803.2			
AIC/N	1.033			
McFadden pseudo-R^2	0.5352			

Latent Class, Random Parameters, FAA, ANA, ACMA (Model 4)

Attributes	FAA 1	ANA 1	ANA 2	ACMA
Random parameters				
(constrained triangular distribution)				
Free flow time	-0.0495 (-5.39)	fixed	-0.0845 (-8.58)	-0.3561 (-36.8)
Slowed down and	-0.0608 (-8.82)	-0.0839 (-14.6)	fixed	-0.3561 (-36.8)
Stop/start/crawling time				
Running cost	-0.6603 (-12.9)	-0.3114 (-6.25)	-0.0799 (-1.74)	-0.4662 (-12.9)
Toll cost	-0.5858 (-29.7)	-0.3352 (-10.1)	-0.2462 (-8.15)	-0.4662 (-12.9)
Fixed parameters				
Reference alt (1,0)	1.8524 (22.6)	0.1051 (0.84)	-1.1228 (-8.76)	0.1678 (1.21)
Commuter trip purpose (1,0)	0.1945 (2.49)	0.2533 (2.45)	-0.7462 (-4.17)	0.0635 (0.56)
Probability of early arrival	-0.0019 (-0.41)	-0.0327 (-6.91)	-0.0053 (-1.11)	-0.0135 (-1.59)
Probability of late arrival	-0.0231 (-4.23)	-0.0451 (-9.33)	-0.0118 (-2.29)	-0.0103 (-1.33)
Stated choice alt 1 (1,0)	-0.0952 (-0.82)	-0.1442 (-1.32)	0.3708 (5.79)	0.7924 (5.98)
Class membership probability	0.656 (22.5)	0.167 (7.48)	0.0611 (4.41)	0.116 (5.39)
Log-likelihood	-4705.2			
AIC/N	1.010			
McFadden pseudo-R^2	0.5447			

The incorporation of the ANA and ACMA attribute processing rules (Model 2 and 4) increases the overall goodness-of-fit compared to the model without allowance for attribute processing. The overall log-likelihood improves from -4817.71 (Model 1) to -4711.59 (Model 2) under fixed parameters, and from -4803.2 (Model 3)[7] to -4705.2 (Model 4) under random parameters. On the AIC test, adjusted for sample size, it improves from 1.035 to 1.013 from Model 1 to Model 2 and 1.033 to 1.010 respectively for Models 3 and 4. Model 4 is only a marginal improvement over Model 2 on overall fit, and after extensive investigation of possible reasons (including changing the number of Halton draws from 250 up to 1500 with 250 increments), we could not find any circumstance under which Model 4 performs considerably better than Model 2. What this suggests to us is that the added layer of behavioural complexity to allow for taste heterogeneity may indeed be adding little once attribute processing is accounted for. This finding is reinforced by the worse fit for Model 3 with random parameters in the absence of attribute processing (-4803.2) in contrast to Model 2 with attribute processing under fixed parameters (-4711.59). It is noteworthy that including three FAA classes in Model 2 is also a way of capturing (discrete) random preference heterogeneity in a probabilistic decision process model, which

---

[7] We also ran Model 3 as a standard mixed logit model. Three models were estimate – an unconstrained triangular distribution with and without correlated random parameters, and a constrained triangular distribution that does not permit correlated parameters. The respective log-likelihoods (and AIC) at convergence were -5512.22 (1.176), -5568.57 (1.187) and -6158.89 (1.311). In all cases these models are inferior, statistically, to Models 1-4 in Table 2, although an expected improvement over multinomial logit (-6729.90, 1.433). Given that Model 4 outperforms the standard mixed logit model, then Model 4 is additionally an improvement over a model with preference heterogeneity accommodated with continuous random taste heterogeneity (as opposed to the Model 1 with discrete preference heterogeneity).

11

introduces useful behavioural information that aligns to a comparison of two alternative specifications of random preference heterogeneity. What this indicates is an expectation that Model 2 may be capturing some amount of the random preference heterogeneity that Model 4 is seeking to reveal, which was not obtained in Model 4 through multiple classes of FAA.

When we consider the class allocation, up to a probability, some interesting findings emerge. Class membership probabilities are statistically significant in all models, with a good spread of membership in all models. A comparison of Models 2 and 4 is especially interesting, given the difference in the treatment of the parameters (fixed or random) under a common set of attribute processing rules. Introducing random parameters to account for taste heterogeneity with a class of probabilistic decision rules reduces the probability of membership of the ACMA class from 0.322 to 0.116. A closer look at the classes representing ANA shows an increase in a move back to full attribute attendance under a random parameter specification from 0.571 (the sum of three FAA classes) to 0.656, and an increase in membership probabilities for the ANA classes. This might suggest that an amount of attribute processing (both ANA and ACMA) is being accommodated through random parameters embedded in FAA. Specifically, one implication is that a low marginal disutility associated with attributes that are otherwise assigned a zero value in ANA are instead associated with a very low marginal disutility and appear in FAA; and small differences in marginal disutility are revealed under ACMA in contrast to equal marginal disutility. There does however remain a sizeable (but smaller) incidence of ANA and ACMA.

Willingness to pay (WTP) estimates for the value of total travel time savings ($/hour) obtained for all four models are summarised in Table 3 and in Figure 2. The averaging is undertaken for the random parameters as per formula (13). We find that the mean estimates increase as we account for attribute processing. On a test of statistical differences of VTTS estimates, the z values are greater than 1.96 (ranging from 13.72 for M1 vs. M3 to 31.7 for M3 vs. M4), except for the comparison of Models 2 and 4 (z=0.85). Thus we can conclude that adding a layer of random parameters to the model that accounts for FAA, ANA and ACMA does not result in a statistically significant difference in the mean estimate of VTTS (M2 vs. M4); however this is not the situation for comparisons between the fixed parameter models M1 and M2, or between the random parameters models M3 and M4, where attribute processing clearly influences VTTS in an upward direction.

Table 3 Willingness to Pay Estimates

Weighted Average VTTS (2008 AUD$/person hour) using weights for components of time and components of cost (standard deviations in brackets)

	Model 1	Model 2	Model 3	Model 4
Free flow time	6.86 (2.25)	8.91 (2.01)	5.90 (3.17)	12.78 (0.79)
Slowed down and stop start time	12.38 (3.12)	14.78 (1.38)	8.72 (3.91)	12.30 (1.14)
Total time	10.17 (2.77)	12.42 (1.63)	7.59 (3.61)	12.49 (1.00)
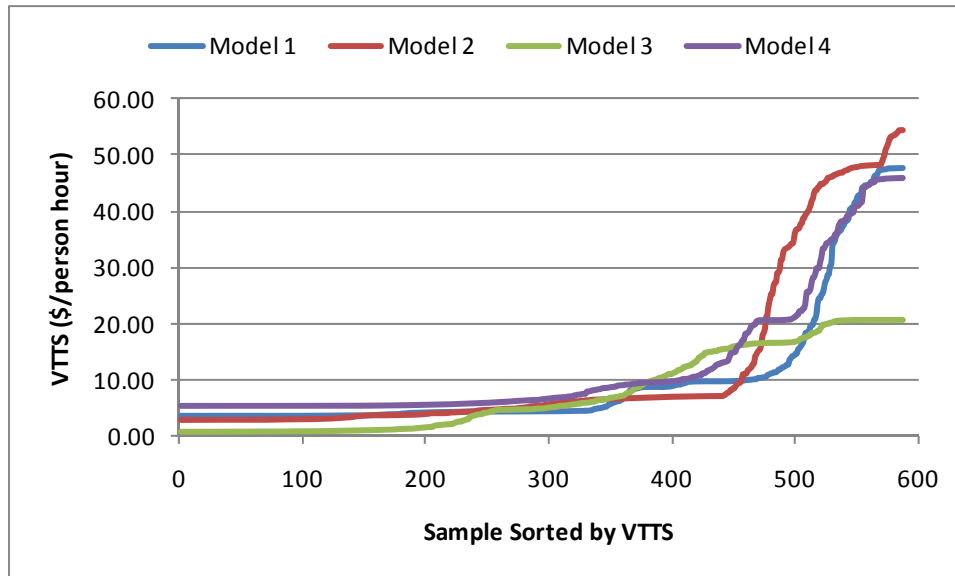
Figure 2 Distribution of VTTS for all Models

We also observe that the incorporation of attribute processing reduces the standard deviation of the VTTS quite considerably for both the fixed and random parameter models, as well as increasing the mean estimate of VTTS. What this suggests is that it is the allowance for attribute processing, and not the allowance for preference heterogeneity within classes through random parameters, that is the key influence on the higher mean estimate of VTTS and accompanying lower standard deviation. Model 3 is of particular interest, since it suggests that in the absence of allowance for FAA, ANA and ACMA, the mean estimate of VTTS is significantly deflated but with an inflated standard deviation when preference heterogeneity through random parameters is accommodated.

Conclusions

This paper has introduced a generalisation of the fixed parameter latent class model through a layering of random parameters within each class, and the redefinition of classes as probabilistic decision rules associated with two specific attribute processing rules. We implement this extended model structure in the context of a toll verses free road choice setting and estimate four models as a way of seeking an understanding of the role of attribute processing in the presence of fixed or random parameters within each probabilistic decision rule class.

What we find, for the data set analysed, is that if attribute processing is handled through discrete distributions defined in a sufficiently flexible way, adding an additional layer of taste heterogeneity through random parameters within a latent class delivers only a very small improvement in the statistical and behavioural contribution of the model. The flexibility is achieved by not equality constraining coefficients

13

across classes, and, crucially, allowing multiple FAA classes to be specified. Compared to the random parameter approach, this model is simple, fast to estimate, and in the empirical setting presented herein, very close in model fit to the model that included continuously distributed random parameters.

One implication is that a random parameter treatment in this setting may be confounding with attribute processing; and that including attribute processing in the absence of continuously distributed random parameters is preferred to including continuously distributed random parameters in the absence of attribute processing. This is an important finding that might suggest the role that attribute processing rules play in accommodating attribute heterogeneity, and that random parameters within class are essentially a potential confounding effect. We offer this finding as a potential concern, conditioned on evidence from one data set, about the possible presence of an identification problem when attribute processing and random parameters within-class are simultaneously considered.

Despite the marginal influence of preference heterogeneity[8] in the overall fit of the models, we find potentially important behavioural evidence to suggest that inclusion of random parameters may be a way of accommodating small marginal disutilities (in contrast to ANA set equal to zero marginal disutility), and small differences in marginal disutilities (in contrast to equal marginal disutilities under ACMA), as observed by a 'move back' to FAA when fixed parameters become random parameters under attribute processing. If this argument has merit, we may have identified one way of recognising what the broader literature (e.g., Hess *et al.* 2011, Campbell *et al.* 2011) refers to as low sensitivity in contrast to zero sensitivity.

The findings in this paper are specific to the data set being analysed[9]; however like any empirical study there is a need to assess the findings and conclusions on a number of data sets. We encourage the research community to undertake this task, not only for the attribute processing strategies we have assessed, but a broader set of heuristics on how attributes are processed. There is the possibility that our findings might be different for different data sets; this is however not a concern about our evidence, but rather a reminder that behavioural processes are often context dependent. If additional studies support the evidence herein on many occasions, then there is a case for recognising the practical value of selecting a latent class framework with fixed parameters for attribute processing, given that inclusion of random parameters adds

---

[8] Noting that in all of the estimated models, we have preference heterogeneity of some kind, whether discrete or continuous.

[9] We are aware of only two studies that have estimated random parameter latent class models allowing for ANA (Hess et al. 2011, Collins et al. 2012). They are not directly comparable with the current study because they do not allow for ACMA and multiple FAA classes, using instead a single FAA class. The main finding, however, of both of these studies, is that the inclusion of random parameters and ANA does improve the model fit. There is thus a consistent message under different assumptions re the role of random parameters. Studies that introduce random parameters into latent class models without attribute processing are Greene and Hensher (2012) who found clear improvement with random parameters added to a LCM; Bujosi et al (2010) who also found an improvement, albeit with the random parameters making only a small contribution (in line with only a small improvement of a RPL model over the MNL), and Vij et al (2012) who does not report the LCM without random parameters, so a comparison cannot be made.

very little in terms of predictive performance while adding significant complexity in estimation.

Other authors have recently used the latent class structure to compare processing heterogeneity with regard to several types of behavioural processes, with other types of heterogeneity (e.g., scale see Thiene *et al.* 2012 and taste, see Hess *et al.* 2012). Although they deal with different decision processes and use different model specifications, they offer general findings on the confounding issue that is discussed in this paper. They propose like us, a latent class (or probabilistic decision process) approach with some conditions imposed on classes to reflect a decision process. They then layer additional heterogeneity on top (random taste or scale) to establish the robustness of both the specifications of heterogeneity, and the alternative model specifications that represent the different decision processes. They conclude that the LC approach has great merit as a framework within which to represent multiple decision processes with and without a random parameter treatment.

References

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

Campbell, D., Hensher, D.A. and 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., Hutchinson, W. and Scarpa, R. (2008) Incorporating discontinuous preferences into the analysis of discrete choice experiments, *Environmental and Resource Economics,* 41 (3), 401–417.

Carlsson, F., Kataria, M. and Lampi, E. (2010) Dealing with ignored attributes in choice experiments on valuation of Sweden, *Environmental and Resource Economics,* 47 (1), 65–89.

Collins, A.T., Rose, J.M., and Hensher, D.A. (2012) The random parameters attribute nonattendance model. Paper presented at The 13th International Conference on Travel Behavior Research, Toronto, July.

Econometric Software (2012), *Nlogit5 Reference Manual*, Econometric Software, New York and Sydney.

Everitt, B. (1988) A finite mixture model for the clustering of mixed-mode data, *Statistics and Probability Letters*, 6, 305-309.

Fiebig, D., Keane, M., Louviere, J., and Wasi, N. (2010) The generalized multinomial logit: accounting for scale and coefficient heterogeneity, *Marketing Science*, 29 (3), 393-421.

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

Greene, W.H. and Hensher, D.A. (2010) Does scale heterogeneity matter? a comparative assessment of logit models, *Transportation*, 37 (3), 413-428.

Greene, W.H. and Hensher, D.A. (2012) Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model, *Applied Economics*, DOI:10.1080/00036846.2011.650325.

Heckman, J. and B. Singer (1984) A method for minimizing the impact of distributional assumptions in econometric models, *Econometrica*, 52, 271-320.

Hensher, D.A. (2006) How do respondents process stated choice experiments? Attribute consideration under varying information load. *Journal of Applied Econometrics*, 21 (6), 861-878.

Hensher, D.A. and Greene, W.H. (2010) Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification, *Empirical Economics* 39 (2), 413-426.

Hensher, D.A. and Rose, J.M. (2009) Simplifying choice through attribute preservation or non-attendance: Implications for willingness to pay. *Transportation Research Part E: Logistics and Transportation Review*, 45 (4), 583–590.

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

Hensher, D.A., Li, Z., and Rose, J.M. (2012a) Accommodating Risk in the Valuation of Expected Travel Time Savings. *Journal of Advanced Transportation*, online 16 January 2011, DOI: 10.1002/atr.160

Hensher, D.A., Rose, J.M. and Greene, W.H. (2012b) 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), 235-246.

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

Hess, S. and Rose, J.M. (2007), A latent class approach to modelling heterogeneous information processing strategies in SP studies, paper presented at the Oslo Workshop on Valuation Methods in Transport Planning, Oslo.

Hess, S., Stathopoulos, A., Campbell, D., O'Neill, V. and Caussade, S. (2011) It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity, Institute of Transport Studies, University of Leeds Working Paper, December.

Hess, S., Stathopoulos, A. and Daly, A, (2012) Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies, *Transportation*, 39 (3) 565-591.

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

Layton, D. and Hensher, D.A. (2010) Aggregation of common-metric attributes in preference revelation in choice experiments and implications for willingness to pay, *Transportation Research Part D*, 15, 394-404.

Li, Z., Hensher, D.A. and Rose, J.M. (2010) Willingness to pay for travel time reliability for passenger transport: a review and some new empirical evidence *Transportation Research Part E,* 46 (3), 384-403

McNair, B., Hensher, D.A. and Bennett, J. (2012) Modelling heterogeneity in response behaviour towards a sequence of discrete choice questions: a probabilistic decision process model, *Environment and Resource Economics*, 51, 599–616.

Puckett, S.M. and Hensher, D.A. (2008) The role of attribute processing strategies in estimating the preferences of road freight stakeholders under variable road user charges, *Transportation Research* E, 44, 379-395.

Rose, J.M., Bliemer, M.C., Hensher and Collins, A. T. (2008) Designing efficient stated choice experiments in the presence of reference alternatives, *Transportation Research B* 42 (4), 395-406

Scarpa, R., Gilbride, T.J., Campbell, D., and 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. and 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.

Shen, J. (2009) Latent class model or mixed logit model? A comparison by transport mode choice data, *Applied Economics*, 41, 2915–2924.

Thiene, M., Meyerhoff, J. and De Salvo, M. (2012) Scale and taste heterogeneity forforest biodiversity: models of serial nonparticipation and their effects, *Journal of Forest Economics,* 18 (4), 355-369.

Train, K. (2008) EM algorithms for nonparametric estimation of mixing distributions, *Journal of Choice Modeling,* 1(1), 40-69.

Train, K. and Weeks, M. (2005) Discrete choice models in preference space and willing to-pay space, *in* R. Scarpa and A. Alberini, eds., *Applications of Simulation Methods in Environmental and Resource Economics,* Springer Publisher, Dordrecht, chapter 1, 1–16.

Wasi, N. and Richard T. Carson (2011) The Influence of Rebate Programs on the Demand for Water Heaters The Case of New South Wales, UC Center for Energy and Environmental Economics, Berkeley, California.

Vij, A., André Carrel1, and Joan L. Walker (2012) Latent modal preferences: behavioral mixture models with longitudinal data, E3 WP-025, UC Berkeley, Department of Civil and Environmental Engineering, Berkeley California.