ISSN: 0003-6846 (Print) 1466-4283 (Online) Journal homepage: http://www.tandfonline.com/loi/raec20

Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model

William H. Greene & David A. Hensher

Applied Economics, 2013, **45**, 1897–1902

Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model

William H. Greene[a] and David A. Hensher[b,*]

[a]*Department of Economics, Stern School of Business, New York University, New York 10012, USA*
[b]*The Business School, Institute of Transport and Logistics Studies, The University of Sydney, NSW 2006, Australia*

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. This article sets out the random parameter latent class model and illustrates its applications using a stated choice data set on alternative freight distribution attribute packages pivoted around a recent trip in Australia.

Keywords: latent class mixed multinomial logit; random parameters; preference heterogeneity; stated choice experiment; freight distribution

JEL Classification: C10; C25; C51; C90; L92

I. Introduction

The focus of the majority of discrete choice modelling over the last 10 years has been centred on the 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, which may be systematically assigned to specific observations through an interaction with observed covariates.

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; 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.

This article sets out the random parameter latent class model, known as the latent class mixed multinomial logit model. As far as we are aware, this is the first article to present this extension and application of this method, although Bujosa *et al*. (2010)[1] have also recently undertaken a very similar extension. We illustrate its applications using an unlabelled stated choice data set on alternative freight distribution trip attribute packages pivoted around a recent trip in Australia.

*Corresponding author. E-mail: David.Hensher@sydney.edu.au

[1] Chronologically, we developed our model in November 2009 and were made aware of Bujosa *et al*. in early 2010. Like Bujosa *et al*., we have a 2-class model; however they use a single observation per respondent revealed preference data, whereas we use stated choice data and allow for correlation amongst the observations common to each respondent.

Applied Economics ISSN 0003–6846 print/ISSN 1466–4283 online © 2013 Taylor & Francis
http://www.tandfonline.com
http://dx.doi.org/10.1080/00036846.2011.650325

II. Random Parameter Latent Class Model

Latent class modelling provides an alternative approach to accommodating heterogeneity to models such as multinomial logit and mixed logit (see Everitt, 1988). The natural approach assumes that parameter vectors, $\boldsymbol{\beta}_i$ are distributed among individuals with a discrete distribution, rather than the continuous distribution that lies behind the mixed logit model. It is assumed that the population consists of a finite number, Q, of groups of individuals. The groups are heterogeneous, with common parameters, $\boldsymbol{\beta}_q$, for the members of the group, but the groups themselves are different from one another. We assume that the classes are distinguished by the different parameter vectors, though the fundamental data generating process, the probability density for the interesting variable under study, is the same.

The analyst does not know from the data which observation is in which 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).

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

$$\text{Prob}(\text{class} = q) = \pi_q(\theta), \quad q = 1, \ldots, Q \tag{2}$$

The unconditional probability attached to an observation is obtained by integrating out the heterogeneity due to the distribution across classes, given in Equation 3,

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

We extend the latent class model to allow for heterogeneity both within and across groups. To accommodate the two layers of heterogeneity, we allow for continuous variation of the parameters within classes. The latent class aspect of the model is given by Equations 4 and 5, as follows:

$$f(y_i|\mathbf{x}_i, \text{class} = q) = g(y_i|\mathbf{x}_i, \boldsymbol{\beta}_i|q) \tag{4}$$

$$\text{Prob}(\text{class} = q) = \pi_q(\theta), \quad q = 1, \ldots, Q \tag{5}$$

The within-class heterogeneity is structured as

$$\boldsymbol{\beta}_i|q = \boldsymbol{\beta}_q + \mathbf{w}_{i|q} \tag{6}$$

$$\mathbf{w}_{i|q} \sim E\left[\mathbf{w}_{i|q}|\mathbf{X}\right] = \mathbf{0}, \quad \text{Var}\left[\mathbf{w}_{i|q}|\mathbf{X}\right] = \boldsymbol{\Sigma}_q \tag{7}$$

where the $\mathbf{X}$ indicates that $\mathbf{w}_{i|q}$ 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 $\boldsymbol{\Sigma}_q$ equal zero, indicating that the variation of the corresponding parameter is entirely across classes.

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 and then the class heterogeneity. We allow for a panel data setting, hence the observed vector of outcomes is denoted as $\mathbf{y}_i$ and the observed data on exogenous variables are collected in $\mathbf{X}_i = [\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iT_i}]$. An individual is assumed to engage in T_i choice situations, where $T_i \geq 1$. The generic model is given in Equation 8,

$$f(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_Q, \theta, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_Q) = \sum_{q=1}^{Q} \pi_q(\theta) \int_{\mathbf{w}_i} \prod_{t=1}^{T_i} f[\mathbf{y}_{it}|(\boldsymbol{\beta}_q + \mathbf{w}_i), \mathbf{X}_{it}]\, h(\mathbf{w}_i|\boldsymbol{\Sigma}_q) \mathrm{d}\mathbf{w}_i \tag{8}$$

We parameterize the class probabilities using a multinomial logit formulation to impose the adding up and positivity restrictions on $\pi_q(\theta)$. Thus,

$$\pi_q(\theta) = \frac{\exp(\theta_q)}{\Sigma_{q=1}^{Q} \exp(\theta_q)}, \quad q = 1, \ldots, Q; \; \theta_Q = 0 \tag{9}$$

A useful refinement of the class probabilities model is to allow the probabilities to be dependent on individual data, such as demographics. The class probability model becomes

$$\pi_{iq}(\mathbf{z}_i, \theta) = \frac{\exp(\theta'_q \mathbf{z}_i)}{\Sigma_{q=1}^{Q} \exp(\theta'_q \mathbf{z}_i)}, \quad q = 1, \ldots, Q; \; \theta_Q = 0 \tag{10}$$

The model employed in this application is a Latent Class, Mixed Multinomial Logit (LC_MMNL) model. Individual i chooses among J alternatives with conditional probabilities given by Equation 11,

$$f[\mathbf{y}_{it}|(\boldsymbol{\beta}_q + \mathbf{w}_i), \mathbf{X}_{it}] = \frac{\exp[\sum_{j=1}^{J} \mathbf{y}_{it,j}(\boldsymbol{\beta}_q + \mathbf{w}_i)'\mathbf{x}_{it,j}]}{\sum_{j=1}^{J} \exp[\sum_{j=1}^{J} \mathbf{y}_{it,j}(\boldsymbol{\beta}_q + \mathbf{w}_i)'\mathbf{x}_{it,j}]}, \quad j = 1, \ldots J \tag{11}$$

$\mathbf{y}_{it,j} = 1$ 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 12,

$$f^S(\mathbf{y}_i|\mathbf{X}_i, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_Q, \theta, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_Q) = \sum_{q=1}^{Q} \pi_q(\theta) \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} f[\mathbf{y}_{it}|(\boldsymbol{\beta}_q + \mathbf{w}_{i,r}), \mathbf{X}_{it}] \tag{12}$$

$\mathbf{w}_{i,r}$ is the rth of R random draws on the random vector $\mathbf{w}_i$. Collecting all terms, the simulated log-likelihood is given by Equation 13,

$$\log L^S = \sum_{i=1}^{N} \log\left[\sum_{q=1}^{Q} \pi_q(\theta) \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} f[\mathbf{y}_{it}|(\boldsymbol{\beta}_q + \mathbf{w}_{i,r}), \mathbf{X}_{it}]\right] \tag{13}$$

III. Illustrative Empirical Context

To illustrate the application of the model, we have selected a Stated Choice (SC) framework (Louviere *et al.*, 2000) within which a freight transporter defined a recent reference trip in terms of its time and cost attributes, treating fuel as a separate cost item to the Variable User Charge (VUC), with the latter being zero at present. An SC design of two alternatives using principles of D-optimality in experimental design (see Hensher *et al.*, 2007; Rose *et al.*, 2008) was developed to vary the levels

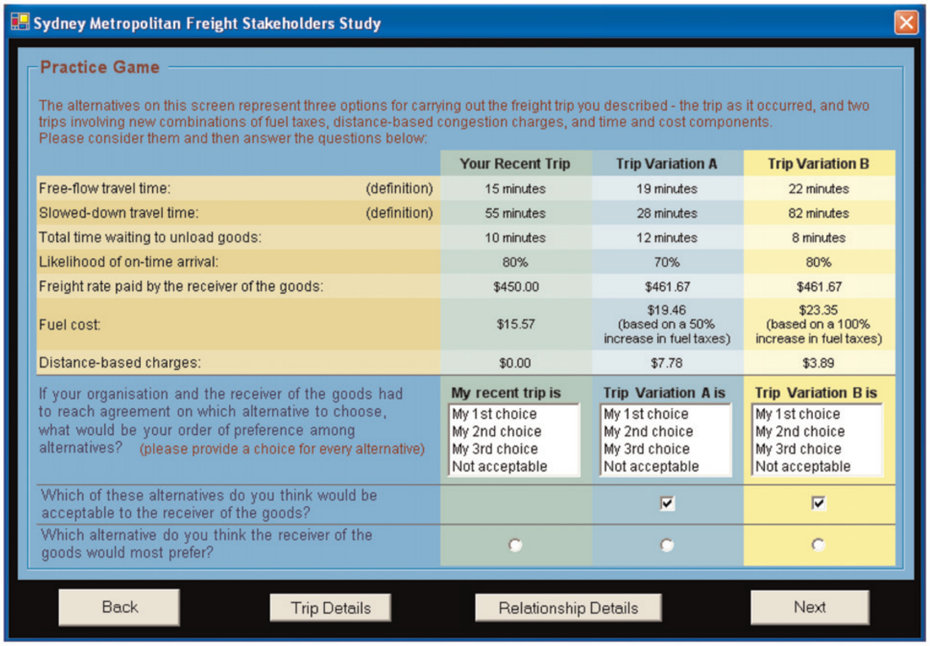

Fig. 1. Main choice set screen

of existing attributes around the recent reference trip levels plus introduce a VUC based on distance travelled but with varying rates per kilometre. Puckett *et al.* (2007) provide extended details of how the empirical study was designed and the data collection strategy.

Selecting the set of attributes for the choice sets involved an iterative process of finding candidate attributes and determining how they could fit intuitively into the choice sets. The attributes are: free-flow travel time, slowed-down travel time, time spent waiting to unload at the final destination, likelihood of on-time arrival, fuel cost and distance-based road user charges. These attributes are either an input into a congestion-charging policy or direct functions of such a policy. The levels of the attributes are expressed as deviations from the reference level, which is the exact value specified in the corresponding non-SC questions, unless noted:

(1) Free-flow time: −50%, −25%, 0, +25%, +50%.
(2) Congested time: −50%, −25%, 0, +25%, +50%.
(3) Waiting time at destination: −50%, −25%, 0, +25%, +50%.
(4) Probability of on-time arrival: −50%, −25%, 0, +25%, +50%, with the resulting value rounded to the nearest 5%.
(5) Fuel cost: −50%, −25%, 0, +25%, +50% (representing changes in fuel taxes of −100%, −50%, 0, +50%, +100%).
(6) Distance-based charges: Pivot base equals $0.5 *$(reference fuel cost), to reflect the amount of fuel taxes paid in the reference alternative. Variations around the pivot base are: −50%, −25%, 0, +25%, +50%.

The survey was conducted via a Computer-Aided Personal Interview (CAPI) in which an interviewer asked the questions and the respondents provided a response. In many situations the respondent was the logistics manager, although a driver was present to provide details of a specific reference trip that were used to establish the attribute levels for the reference trip. An example of an SC screen is given in Fig. 1. Freight transporters were faced with four choice sets. The survey was undertaken in 2005, sampling transporters who were delivering goods on behalf of a single shipper to and/or from the Sydney Metropolitan area. The resulting estimation sample is 108 transporters yielding 432 choice sets.

IV. Results

In this application of the LC_MMNL model, we have simplified the model and aggregated free-flow time, congested time and time waiting at the destination to unload goods, as well as aggregating the fuel costs and variable user charge. From supplementary questions, over 80% of respondents indicated that they added up the times and costs respectively before making a choice. A dummy variable was included in the model to indicate the absence of a variable user charge in the total cost.

We have assumed to this point that the number of classes, Q, is known. This will rarely be the case, so a question naturally arises, how can the analyst determine 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 Akaike Information Criterion (AIC), to guide them toward the appropriate value. Heckman and Singer (1984) note a practical guidepost. If the model is fit with too many classes, then estimates will become imprecise, even varying wildly. Signature features of a model that has been overfit will be exceedingly

Table 1. Summary of models

	MNL (Model M1)	Mixed logit (Model M2)	Latent class: Fixed parameters (M3)		Latent class: Random parameters (M4)	
Attributes			Class 1	Class 2	Class 1	Class 2
Total time (min)	−0.0056	−0.0053	−0.00177	−0.0451	−0.0082	0.0008
	(−3.44)	(−2.97)^{rp}	(−3.82)	(−1.73)	(1.95)	(0.25)
On time delivery (%)	0.03161	0.0360	0.0296	−0.0296	−0.0005^{rp}	0.02713^{rp}
	(4.75)	(4.51)^{rp}	(3.35)	(−0.55)	(−0.03)	(2.53)
Total cost ($)	−0.0030	−0.0039	−0.0051	−0.0134	−0.0135	−0.0052
	(−4.15)	(−3.74)^{rp}	(−3.95)	(−2.09)	(−2.13)	(−3.33)
No variable charge dummy (1,0)	0.9406	0.8798	0.7716	4.2819	1.5933	−1.2945
	(5.71)	(4.84)	(2.35)	(2.18)	(4.48)	(−2.1)
Class membership probability			0.7137	0.2853	0.574	0.4264
			(14.2)	(4.04)	(6.9)	(5.14)
BIC	1.8777	1.8651	1.8857		1.6650	
Log-likelihood	−393.44	−390.71	−380.01		−350.63	

	Latent class with decomposition of class membership probability: Fixed parameters (M5)		Random parameters (M6)	
Attributes	Class 1	Class 2	Class 1	Class 2
Total time (min)	0.0056	−0.00015	−0.0069	0.0015
	(0.69)	(−0.04)	(−1.96)	(0.56)
On time delivery (%)	−0.0062	0.0279	−0.00056^{rp}	0.0276^{rp}
	(−0.21)	(2.34)	(−0.040	(2.86)
Total cost ($)	−0.0474	−0.0043	−0.0157	−0.0055
	(−3.30)	(−1.91)	(−3.14)	(−4.26)
No variable charge dummy (1,0)	1.4220	−2.1567	1.4109	−1.3472
	(2.84)	(−1.34)	(4.25)	(−2.95)
Class membership probability	0.616	0.384	0.575	0.425
Constant	1.1741 (3.30)	0	1.8200 (2.79)	0
Freight rate ($)	−0.00076 (−2.32)	0	−0.0016 (−2.54)	0
BIC	1.7724		1.6498	
Log-likelihood	−372.85		−346.36	

Notes: Elasticities are not informative for this model given that the alternatives are unlabelled. BIC – Bayesian Information Criterion. There are 432 observations given. For mixed logit we used constrained *t*-distributions and 500 Halton draws. *t*-ratios are given within parentheses. ^{rp} Denotes random parameter.

small estimates of the class probabilities, wild values of the structural parameters, and huge estimated SEs.

Statistical inference about the parameters can be made in the familiar fashion. The Wald test or likelihood ratio tests will probably be more convenient. Hypothesis tests across classes are unlikely to be meaningful. For example, suppose we fit a three class model. Tests about the equality of some of the coefficients in one class to those in another would probably be ambiguous, because the classes themselves are indeterminate. It is rare that one can even put a name on the classes, other than, '1', '2', etc. Likewise, testing about the number of classes is an uncertain exercise.

If the parameters of the two classes are identical, it would seem that there is a single class. The number of restrictions would seem to be the number of model parameters. However, there remain two class probabilities, π_1 and π_2. If the parameter vectors are the same, then regardless of the values of π_1 and π_2, there is only one class. Thus, the degrees of freedom for this test are ambiguous. The same log-likelihood will emerge for any pair of probabilities that sum to one. We found that two classes delivered statistical significance for each parameter estimate in at least one class, in contrast three and four classes had at least one variable that was statistically insignificant in all classes. The findings are presented in Table 1 for the two latent class-mixed MNL model, with random parameters (rp) defined by a constrained triangular distribution with SD equal to the mean estimate,[2] together with an MNL model, a mixed logit model, and two fixed parameter latent class models. The two versions of the latent class models (M3, M4) and (M5, M6) are distinguished by the inclusion of the freight rate as a source of systematic variation in the class membership probabilities.

The overall fit of the LC-MMNL model is a statistically significant improvement over the other models on the BIC index, given the additional two parameters, compared to the fixed parameter latent class model, and three extra parameters

[2] All models are estimated using Nlogit 5 (Pre-release).

Table 2. WTP Estimates for total time and on time delivery

	MNL	Mixed logit	Latent class			
			Fixed parameters		Random parameters	
WTP	M1	M2	M3	M5	M4	M6
Total time ($/hr)	112	81.5 (0.05)	20.8 (Class 1)	–	36.4 (Class 1)	26.37 (Class 1)
On time delivery ($/percent point)	10.53	17.13 (31.2)	4.76 (Classes 1 and 2)	6.49 (Class 2)	5.08 (2.15) (Class 2)	5.17 (2.01) (Class 2)

Notes: Values in parentheses are SDs for random parameter specifications (based on 500 draws). The time parameter is not statistically significant (see Table 1). '–' denotes not significant.

compared to the mixed logit model (excluding the membership class probabilities). The addition of the freight rate in M6 as a systematic conditioning source on the probability of class provides the best fit model. The improvement of the LC-MMNL model over all other models is impressive, and supports the presence of preference heterogeneity in on time delivery in one class.

Whereas all parameters are statistically significant for the MNL and mixed logit models, this is not the case for the latent class models, especially the random parameter specifications and the fixed parameter version of latent class when we allow for the decomposition of the class membership probability by the freight rate.[3] This suggests that, up to a class membership probability, that some attributes have statistical merit for only one latent class. In particular, the LC-MMNL models 4 and 6 have two latent classes in which there is a clear preference for total time and total cost in one class, and on-time delivery and total cost in the other class. This is also the case under the fixed parameter latent class form. When we account for within-class preference heterogeneity at the attribute level, we see significant change in the mix of membership probability between the two classes; namely 42.64% and 57.4% for the LC-MMNL model 4 (and 42.5% and 57.5% for model 6) and 71.37% and 28.53% for the fixed parameter latent class model 3 (and 61.6% and 38.4% for model 5). It should be noted that class 1 and class 2 in the latent class models are not equivalent in any sense across the models, and should not be directly compared *per se*.

In the latent class models we investigated the possibility of random parameters for all attributes, but we could only establish statistical significance for one random parameter, namely on time delivery, which is statistically significant in one class only for all four latent class models. Together with the fixed parameter estimate for total time, this suggests that we have one class of transporters who focus on a trade-off between total time and total cost, up to a probability, and another class of transporters who focus on the trade-off between on time delivery and total cost, up to a probability.

Willingness To Pay (WTP) estimates for the value of total travel time savings ($/hour) and on time delivery ($/percent point on time) are summarized in Table 2. Only estimates based on statistically significant parameters are included. The WTP estimates where the numerator is a random parameter distribution, involve an estimate for each respondent drawn off the distribution, divided by the fixed mean beta for cost, and then averaged across the sample. The mean estimates for both time dimensions are considerably lower for the latent class models compared to MNL and mixed logit, for reasons that are unclear other than the recognition that the MNL and mixed logit results ignore the possibility of the two classes of utility representation (up to a probability). The presence of statistically significant effects in only one of the two classes on both time attributes might be reason for concern; however, what this tells us is that the WTP estimate is driven by membership of one latent class only. Given the overall fit, under BIC, is considerably better for the latent class models, the WTP estimates associated with MNL and mixed logit might be queried, suggesting a significant upward bias in the mean estimates for both total time and on time delivery. Whether this evidence is transferable to other situations can only be confirmed by additional empirical studies, although we note that Bujosa *et al*. (2010) in a totally different context of recreational trips to forest sites in Mallorca, find significantly higher mean estimates for mixed logit and LC models for attributes with random parameters compared to MNL and fixed parameter LC.

V. Conclusions

This article extends the latent class model to accommodate preference heterogeneity within each class through the use of random parameters. Although a subtle extension in many ways, in line with the evolution of choice models specified with continuous mixtures, it was only a matter of time before the interest in a latent class model with random parameters surfaced.

Ongoing research using a number of data sets will ensure that we gain a greater understanding of the gains to be made from the inclusion of random parameters in a latent class model, which we suspect will be data-set specific.

Acknowledgements

The comments of two referees are appreciated.

[3] Bujosa *et al*. (2010) also found a number of statistically nonsignificant parameter estimates in one class.

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**, 477–93.

Everitt, B. (1988) A finite mixture model for the clustering of mixed-mode data, *Statistics and Probability Letters*, **6**, 305–9.

Fiebig, D., Keane, M., Louviere, J. and Wasi, N. (2010) The generalized multinomial logit: accounting for scale and coefficient heterogeneity, *Marketing Science*, **29**, 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**, 681–98.

Greene, W. H. and Hensher, D. A. (2010) Does scale heterogeneity matter? A comparative assessment of logit models, *Transportation*, **37**, 413–28.

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

Hensher, D. A. and Greene, W. H. (2003) Mixed logit models: state of practice, *Transportation*, **30**, 133–76.

Hensher, D. A., Puckett, S. and Rose, J. (2007) Agency decision making in freight distribution chains: revealing a parsimonious empirical strategy from alternative behavioural structures, *Transportation Research Part B*, **41**, 924–49.

Louviere, J. J., Hensher, D. A. and Swait, J. (2000) *Stated Choice Methods: Analysis and Applications in Marketing, Transportation and Environmental Valuation*, Cambridge University Press, Cambridge.

Puckett, S. M., Hensher, D. A., Rose, J. and Collins, A. (2007) Design and development of a stated choice experiment in a two-agent setting: interactions between buyers and sellers of urban freight services, *Transportation*, **34**, 429–51.

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

Shen, J. (2009) Latent class model or mixed logit model? A comparison by transport mode choice data, *Applied Economics*, **41**, 2915–24.

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