Revealing Additional Dimensions of Preference Heterogeneity in a Latent Class Mixed Multinomial Logit Model

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

David A. Hensher
Institute of Transport and Logistics Studies
Faculty of Economics and Business
The University of Sydney
NSW 2006 Australia
David.Hensher@sydney.edu.au

16 September 2010

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 (i.e., fixed parameters), 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 paper sets out the random parameter latent class model, building on the fixed 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
1. Introduction

The focus of the majority of discrete choice modelling over the last ten years has been centred on the mixed (or random parameter) logit model, and more recently on extensions of mixed logit to incorporate scale heterogeneity and estimation in willingness to pay space (referred to as generalised mixed logit) (see for example Train and Weeks 2005, Fiebig et al. 2009, 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 that compares latent class and mixed logit models, and a recent revisit of the Greene-Hensher contribution by Shen 2009). Within each class, preference homogeneity is usually assumed (i.e., fixed parameters), 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 paper sets out the random parameter latent class model, known as the latent class mixed multinomial logit model, building on the fixed parameter latent class model. As far as we are aware, this is the first paper to present this extension and application of this method, although Bujosa et al. (2010) have recently also 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.

2. 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 and Uebersax 1999). The natural approach assumes that parameter vectors, $\beta_i$, are distributed among individuals with a discrete distribution, rather than the continuous distribution that lies behind the mixed logit model. Thus, it is assumed that the population consists of a finite number, $Q$, of groups of individuals. The groups are heterogeneous, with common parameters, $\beta_q$, 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.

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.
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):

\[ f(y_i|x_i, \text{class } = q) = g(y_i | x_i, \beta_q) \]  

(1)

\[ \text{Prob(class } = q) = \pi_q(\theta), q = 1, \ldots, Q. \]  

(2)

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

\[ f(y_i|x_i) = \sum_q \pi_q(\theta) g(y_i | x_i, \beta_q). \]  

(3)

In this paper, we extend the latent class model to allow for heterogeneity both within and across groups. That is, we allow for variation of the parameter vector within classes as well as between classes. The extended model is a straightforward combination of the mixed logit and latent class models. 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 as (4) and (5).

\[ f(y_i|x_i, \text{class } = q) = g(y_i | x_i, \beta_{iq}) \]  

(4)

\[ \text{Prob(class } = q) = \pi_q(\theta), q = 1, \ldots, Q. \]  

(5)

The within-class heterogeneity is structured as

\[ \beta_{iq} = \beta_q + w_{iq} \]  

(6)

\[ w_{iq} \sim E[w_{iq}|X] = 0, \text{ Var}[w_{iq}|X] = \Sigma_q \]  

(7)

where the use of \( X \) indicates that \( w_{iq} \) 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 0 and covariance matrix \( \Sigma \). In a given application, it may be appropriate to further assume that certain rows and corresponding columns of \( \Sigma_q \) 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 will allow for a panel data, or stated preference data setting, hence the observed vector of outcomes is denoted \( y_i \) and the observed data on exogenous variables are collected in \( X_i = [X_{i1}, \ldots, X_{iT}] \). The individual is assumed to engage in \( T_i \) choice situations, where \( T_i \geq 1 \). The generic model is given in (8).

\[ f(y_i|X_i, \beta_1, \ldots, \beta_Q, \theta, \Sigma_1, \ldots, \Sigma_Q) = \sum_{q=1}^{Q} \pi_q(\theta) \int_{w} \prod_{t=1}^{T_i} f(y_{it} | (\beta_q + w_{t}), X_{it}) h(w_{t}|\Sigma_q) dw_t. \]  

(8)
We parameterize the class probabilities using a multinomial logit formulation to impose the adding up and positivity restrictions on \( \pi_q(\theta) \). Thus,

\[
\pi_q(\theta) = \frac{\exp(\theta_q)}{\sum_{q=1}^Q \exp(\theta_q)}, \quad q = 1, \ldots, Q; \; \theta_Q = 0.
\]

(9)

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

\[
\pi_q(z_i, \theta) = \frac{\exp(\theta'_q z_i)}{\sum_{q=1}^Q \exp(\theta'_q z_i)}, \quad q = 1, \ldots, Q; \; \theta_Q = 0.
\]

(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 as (11).

\[
f[y_{it} | (\beta_q + w_{it}), X_{it}] = \frac{\exp[\Sigma_{j=1}^J y_{it,j} (\beta_q + w_{it})' x_{it,j}]}{\sum_{j=1}^J \exp[\Sigma_{j=1}^J y_{it,j} (\beta_q + w_{it})' x_{it,j}]}, \quad j = 1, \ldots, J,
\]

(11)

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

The integrals cannot be evaluated analytically. We use maximum simulated likelihood (along the same lines as mixed logit) 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).

\[
f^{S_i}(y_{it}|X_{it}, \beta_1, \ldots, \beta_Q, \Sigma_1, \ldots, \Sigma_Q) = \sum_{q=1}^Q \pi_q(\theta) \frac{1}{R} \sum_{r=1}^R \prod_{t=1}^T f[y_{it} | (\beta_q + w_{it,r}), X_{it}]
\]

(12)

\( w_{it,r} \) is the \( r \)th of \( R \) random draws (Halton draws in our implementation) on the random vector \( w_i \). Collecting all terms, the simulated log likelihood is given as (13).

\[
\log L^S = \sum_{i=1}^N \log \left[ \sum_{q=1}^Q \pi_q(\theta) \frac{1}{R} \sum_{r=1}^R \prod_{t=1}^T f[y_{it} | (\beta_q + w_{it,r}), X_{it}] \right]
\]

(13)

The functional forms for \( \pi_q(\theta) \) and \( f[y_{it} | (\beta_q + w_{it,r}), X_{it}] \) are given in (9) and (10), respectively.

### 3. 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 (detailed below), 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, Sandor and Wedel 2001) was developed to vary the levels 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. Whilst in-depth interviews and literature reviews revealed myriad attributes that influence freight decision making (see Puckett et al. 2006, Cullinane and Toy 2000, Danielas et al. 2005, Fowkes et al. 2004), we focussed on the subset of these attributes that were most likely to be directly affected by congestion charges. Hence, the attributes that reside within the choice sets 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 (i.e., changes in fuel taxes, road user charges), 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 five percent (e.g., a reference value of 75% reduced by 50% would yield a raw figure of 37.5%, which would be rounded to 40%). If the resulting value is 100%, the value is expressed as 99%. If the reference level is greater than 92%, the pivot base is set to 92%. If the pivot base is greater than 66 percent (i.e., if 1.5 times the base would be greater than 100%) let the pivot base equal X, and let the difference between 99% and X equal Y. The range of attribute levels for on-time arrival when X > 66% are (in percentage terms): X-Y, X-.5*Y, X, X+.5*Y, X+Y. This yields five equally-spaced attribute levels between X-Y and 99%.
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 .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 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 Figure 1. Freight
transporters were faced with four choice sets. They were asked to assume that, for each of the choice sets given, the same goods need to be carried for the same client, subject to the same constraints faced when the reference trip was undertaken. Respondents are then informed that the choice sets involve three alternative methods of making the trip: their stated trip and two SC alternatives that involve VUCs.

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. Initially a sample of transporters were selected and screened for participation by a telephone call. Eligibility to participate in the CAPI survey required a respondent having (i) input into the routing or scheduling of freight vehicles used by their organisation (ii) input into the business arrangements made with their organisation’s customers, and (iii) their organisation carry truckloads that contain cargo either sent by, or intended for, one single company. The resulting estimation sample is 108 transporters yielding 432 choice sets.

![Figure 1 Main Choice Set Screen](image)

### 4. Results

In this illustrative 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 the fuel costs and variable user charge. From supplementary questions, over 80 percent 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 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 small estimates of the class
probabilities, wild values of the structural parameters, and huge estimated standard errors.

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, \( \pi_1 \) and \( \pi_2 \). If the parameter vectors are the same, then regardless of the values of \( \pi_1 \) and \( \pi_2 \), 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 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\(^2\), 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, M5) and (M5, M6) are distinguished by the inclusion of the freight rate as a source of systematic variation in the class membership probabilities.

<table>
<thead>
<tr>
<th>Table 1 Summary of Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>432 observations. For mixed logit we used constrained t-distributions(^3) and 500 Halton draws</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attributes:</th>
<th>MNL (Model M1)</th>
<th>Mixed Logit (Model M2)</th>
<th>Latent Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Class 1</td>
<td>Class 2</td>
<td>Class 1</td>
</tr>
<tr>
<td>Total time</td>
<td>-0.0056 (-3.44)</td>
<td>-0.0053 (-2.97)</td>
<td>-0.00177 (-3.82)</td>
</tr>
<tr>
<td>On time delivery</td>
<td>0.03161 (4.75)</td>
<td>0.0360 (4.51)(^a)</td>
<td>0.0296 (3.35)</td>
</tr>
<tr>
<td>Total cost</td>
<td>-0.0030 (-4.15)</td>
<td>-0.0039 (-3.74)(^a)</td>
<td>-0.0051 (-3.95)</td>
</tr>
<tr>
<td>No variable charge dummy</td>
<td>0.9406 (5.71)</td>
<td>0.8798 (4.84)</td>
<td>0.7716 (2.35)</td>
</tr>
<tr>
<td>Class membership probability:</td>
<td>0.7137 (14.2)</td>
<td>0.2853 (4.04)</td>
<td>0.574 (6.9)</td>
</tr>
<tr>
<td>BIC</td>
<td>1.8777</td>
<td>1.8651</td>
<td>1.8857</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-393.44</td>
<td>-390.71</td>
<td>-380.01</td>
</tr>
</tbody>
</table>

\(^2\) All models are estimated using Nlogit 5 (Pre-release).

\(^3\) The triangular distribution ensures that the sign of the parameter estimate is behaviourally plausible throughout the entire distribution. The triangular distribution was first used for random coefficients by Train and Revel (2000), later incorporated into Train (2003). Hensher and Greene (2003) also used it and it is increasingly being used in empirical studies. Let \( c \) be the centre and \( s \) the spread. The density starts at \( c-s \), rises linearly to \( c \), and then drops linearly to \( c+s \). It is zero below \( c-s \) and above \( c+s \). The mean and mode are \( c \). The standard deviation is the spread divided by \( \sqrt{6} \); hence the spread is the standard deviation times \( \sqrt{6} \). The height of the tent at \( c \) is \( 1/s \) (such that each side of the tent has area \( s \times (1/s) \times (1/2) = 1/2 \), and both sides have area \( 1/2 + 1/2 = 1 \), as required for a density). The slope is \( 1/s \).
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 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. 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. Furthermore, when we account for within-class preference heterogeneity at the attribute level, we see a significant change in the mix of membership probability between the two classes; namely 42.64 and 57.4 percent for the LC-MMNL model 4 (and 42.5 and 57.5 percent for model 6) and 71.37 and 28.53 percent for the fixed parameter latent class model 3 (and 61.6 and 38.4 percent 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.

The only random parameter is on time delivery, which is statistically significant in one class only for all four latent class models, and together with the fixed parameter estimate for total time, 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 Remaining parameter estimates in one class are statistically non-significant. 

4 Bujosa et al. (2010) also found a number of statistically non-significant parameter estimates in one class.
transporters who focus on the trade-off between on time delivery and total cost up to a probability. The marginal utility distribution for on time delivery in class 2 for LC-MMNL Model 6 and the Mixed Logit model (both sources of random parameters) are given in Figures 2 and 3 respectively. The on time delivery parameter is not significant for class 1, suggesting that the mixed logit form confounds the mixture of respondents, up to a probability of latent class membership, with both a significant and a non-significant on time delivery parameter estimate which converges to statistical significance when combined into one class. The most informative evidence in Figures 2 and 3 is the range of estimates, since the two distributions are not directly comparable across the models. Some additional amount of preference heterogeneity is added in when we distinguish between latent classes in contrast to a ‘single class’ treatment in mixed logit.

Figure 2 LC-MMNL Marginal Utility Distribution for On-Time Delivery
mean = 0.045, std = 0.0178, min =0.0009, max = 0.0895

Figure 3 Mixed Logit Marginal Utility Distribution for On-Time Delivery
mean = 0.035, std = 0.0143, min =0.0003, max = 0.0707

Willingness to pay (WTP) estimates for the value of total travel time savings ($/hour) and on time delivery ($/percent point on time) are summarised in Table 2. 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 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.

### Table 2 Willingness to Pay Estimates for Total Time and On Time Delivery

Values in brackets are standard deviations for random parameter specifications (based on 500 draws)

<table>
<thead>
<tr>
<th>WTP</th>
<th>MNL</th>
<th>Mixed Logit</th>
<th>Latent Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M1 M2 M3 M5 M4 M6</td>
<td>Fixed Parameters</td>
<td>Random Parameters</td>
</tr>
<tr>
<td>Total time ($/hr)</td>
<td>112</td>
<td>81.5 (0.05)</td>
<td>20.8 ns 37.2</td>
</tr>
<tr>
<td>On time delivery ($/percent point)</td>
<td>10.53</td>
<td>17.13 (31.2)</td>
<td>4.76 6.49</td>
</tr>
</tbody>
</table>

### Conclusions

This paper 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.

The evidence herein shows that the introduction of random parameters within class is straightforward, and that this LC-MMNL model, in the context of one data set, delivers the best overall fit in contrast to MNL, mixed logit and fixed parameter latent class models. In the specific application however, this comes at a price, in that researchers interested in willingness to pay estimates, a major output of choice analysis, may find that the cost parameter is not statistically significant in all classes, as is the case herein (as well as Bujosa et al. (2010) and also found in other studies that use fixed latent class models, although not so in the current study), denying calculation of overall weighted average WTP estimates, which are necessary to obtain meaning to the notion of WTP being weighted by the probability of class membership.

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.

### References


