# Hierarchical Bayes of Mixed Logit: an application to genetic testing

Dean A. Regier[1], Mandy Ryan[2], and Euan Phimister[3]

[1] PhD Candidate, Health Economics Research Unit (HERU), University of Aberdeen, Scotland
[2]Professor, Health Economics Research Unit (HERU), University of Aberdeen, Scotland
[3] Professor, Business School, Univeristy of Aberdeen Scotland.

Submitted for the 71st Health Economists' Study Group Meeting.
September 5-7th, 2007.

## 1. Introduction

Demand models for health technologies can estimate the value that individuals place on different factors of the good that affect their choice. Under a discrete choice experiment (DCE) framework, respondents choose between alternatives that differ on several attributes. Including a cost attribute allows for an empirical measure of willingness to pay (WTP) for incremental changes between the attributes of the novel technology and the current standard, or for the technology itself. The multinomial logit and nested logit models are well-established behavioural specifications used to model DCE data. These models, however, can be limiting because they assume (1) the parameter coefficient is the same for each person (or the same for each person within a group); (2) proportional substitution patterns[1]; and (3) independence over choices of individuals' unobserved factors (McFadden 1974).

The mixed logit (MXL) behavioural model is increasingly being used in health economics because it allows for realistic substitution patterns and correlation of unobserved factors across choice questions. The MXL can also incorporate random taste variation that permits the estimation of individual partworths, identification of outliers, and calculation of more accurate choice probabilities (Train and Sonnier 2005). Unlike the random effects probit model, the MXL can incorporate non-normal distributions for the parameters. McFadden and Train show that any discrete choice random utility model can be approximated to any degree of accuracy by a MXL (McFadden and Train 2000).

In health economics, the MXL has been estimated using a classical, or frequentist, approach (Hall et al. 2006; King et al. 2007; Lancsar et al. 2007). Classical estimation of the MXL employs maximum simulated likelihood estimation (MSLE). MSLE combines maximum likelihood estimates of the mean and standard deviation of the asymptotic sampling distribution with individuals' choices (Huber and Train 2001). MSLE works well when normal distributions are specified for the parameters. There may, however, be difficulties when using MSLE such as reaching convergence if the starting values are far from their maximum, or when bounded distributions such as the log-normal are specified. Furthermore, issues of local versus global maxima can complicate the maximization procedures since convergence does not guarantee a global maximum has been attained (Train 2003).

[1] This is a result of the independence from irrelevant alternatives assumption (IIA). For logit models, IIA holds for all alternatives, and for nested logit models IIA holds within each nest.

An alternative estimation approach employs Bayesian procedures. Hierarchical Bayes (HB) of MXL uses Markov Chain Monte Carlo (MCMC) techniques to obtain the joint posterior of part-worths. HB models estimate a distribution of coefficients and combine information of each individual's choice to derive person-specific posterior estimates (Huber and Train 2001). Unlike MSLE, HB does not require the maximization of a likelihood function, which alleviates problems of convergence due to poor starting values or non-quadratic likelihood functions (Train 2003). Under certain conditions, MCMC may result in faster estimation time than classical procedures (Train 2003). Bayesian procedures, however, can also be complicated to estimate. To simulate the relevant statistics, HB estimation uses an iterative process that converges with a sufficient number of iterations to draws from the posterior distribution. Knowing if the MCMC algorithm is drawing from the distribution is not easily ascertained (Kass et al. 1998).

Despite coming from different estimation algorithms and interpretive philosophies of probability, HB and MSLE procedures are related and for large enough samples the estimates from the two procedures may converge asymptotically (Huber and Train 2001).[2] In small samples, the two procedures are likely to provide different results because they handle uncertainty differently (Train 2003). The aim of this paper is to investigate HB procedures as an alternative to classical estimation, and to compare the numerical results of the two approaches. The ability of the MXL model to approximate a random utility model depends on the type of distributions assumed for the parameters. For example, if the researcher wants to calculate WTP estimates, a normal distribution on the price coefficient is not ideal given it has support on both the negative and positive side of the distribution, which necessarily straddles zero. It is possible, therefore, that for some individuals the estimation results may provide an infinite WTP for some incremental good. If a respondent's estimate is positive for the cost attribute, they necessarily will prefer a higher price to a lower price. Thus, this paper also explores the use of bounded distributions under the HB and classical procedures.

The application under consideration is structured around a DCE eliciting preferences for a novel genetic technology that aims to identify genetic causes of developmental delay/mental retardation (MR) in individuals whose aetiology is unknown. MR is a frequently occurring disorder that has a major impact not only on the person with MR, but also on his/her family and society. Establishing aetiology can be difficult and costly, but having a diagnosis provides

[2] The Bernstein-von Mises theorem outlines the classical properties of the mean of the Bayesian posterior.

information on recurrence risk, prognosis, and treatment options (Rauch et al. 2006). Having a diagnosis may also help families with acceptance of the disability (Lenhard et al. 2005). In health care systems funded by the public, policy makers are not only interested to know how families value genetic testing, but also the preferences of society. The choice data are derived from respondents of the public from British Columbia, Canada.

This paper is divided into six sections. Section 2 outlines the MXL using a random coefficients interpretation, and provides an overview of the classical and Bayesian approaches to estimation. Section 3 reviews the DCE application and describes the questionnaire design. In Section 4, the econometric modeling is discussed, and section 5 presents the results of the DCE under the alternate econometric specifications. Section 6 concludes with a discussion of the merits and limitations of the two procedures.

## 2. The mixed logit model

### *2.1 Behavioural specification*

Suppose the utility of person n for profile j=1….J in each choice set s=1….S can be specified as:

(1) $$U_{njs} = \beta_n' X_{njs} + \varepsilon_{njs},$$

where $X_{njs}$ is non-stochastic vector of the attributes for alternative j, n=1,…,N identifies each respondent, and $\varepsilon_{njs}$ is an error term following an *iid extreme value* distribution. $\beta_n$ is the vector of coefficients for person n: indicating that different individuals have different tastes. The researcher specifies a distribution for $\beta_n$, which has a mean vector, b, and covariance matrix W. The estimates of these parameters characterize the distribution of the random coefficients. Under random utility theory (RUT), each customer chooses the alternative that provides him or her with the greatest utility (McFadden 1974; McFadden and Train 2000).

Conditional on $\beta_n$, the MXL probability of individual n choosing alternative i where $j \neq i$ can be expressed as:

(2) $$L_{nis} = \frac{e^{\beta_{ni} x_{si}}}{\sum_j e^{\beta_{nj} x_{sj}}}.$$

The researcher does not know $\beta_n$. The unconditional probability is the integral of $L_{nis}$ over all possibilities of $\beta_n$:

(3) $$P_{nis} = \int \prod_s L_{nis} f(\beta \mid \theta) db .$$

The distribution of the parameters is denoted as $f(\beta|\theta)$, where $\theta$ is the vector of parameters (b and W) characterizing the distribution. In the health economics literature, $f(\beta|\theta)$ has been typically specified as a Normal: $\beta \sim N(b,W)$ (Hall et al. 2006; King et al. 2007; Lancsar et al. 2007). Other investigators in economics and marketing have used the log-normal distribution, $\ln\beta \sim N(b,W)$ (Train 1998) , uniform distribution (Hensher and Greene 2003; Revelt and Train 2000), triangular distribution (Hensher and Greene 2003; Revelt and Train 2000), Johnson's $S_B$ distribution (Train and Sonnier 2005), and a normal distribution censored from below at zero(Bhat 2000). Coefficients can also be specified to be fixed; that is, they do not vary in the population.

*2.2. Maximum simulated likelihood estimation*

The classical procedures specify b and W as fixed, representing the true mean and covariance of $\beta_n$ in the population(Huber and Train 2001). Their estimators are stochastic due to variation in sampling. The integral in equation 3 does not have a closed-form solution. That is, $P_{nis}$ needs to be approximated by taking draws of $\beta_n$ from the density $f(\beta|\theta)$. Taking draws from the density is achieved through simulation: take a draw of $\beta_n$ from $f(\beta|\theta)$, calculate the conditional logit probability given in equation 1; repeat many times and average over the results. For panel data, the probability of each person's choice is the integral in equation 3 over a product of conditional logit probabilities (Train 2003). The resulting simulated probability is an unbiased estimator of $P_{nis}$ and is used to approximate the maximum likelihood function (Train 2003). The simulated likelihood estimators are the estimates of b and W that maximize the simulated likelihood function.

*2.3 Bayesian estimation*

The parameters b and W are considered stochastic from a Bayesian perspective (Train 2003). The Bayesian posterior distribution, $K(\cdot)$, is proportional to the prior distribution on b

and W, p(b,W), multiplied by the likelihood function. Under the HB specification, the joint posterior of the parameters is defined as:

$$K(\beta_n \forall n, b, W \mid Y) \propto \prod_s L_{nis}(\beta_n) f(\beta \mid \theta) p(b, W) . \quad (4)$$

Draws from the joint posterior are obtained using the Gibbs sampling and the Metropolis-Hasting (MH) algorithm. Taken together, these procedures are sometimes called MCMC. Gibbs sampling entails estimating a sequence of draws where each new draw for a parameter is estimated conditional on the other parameters in the model (in a hierarchical form). For example, begin with initial values of $b^0$, $W^0$, and $\beta^0_n$. The $t^{th}$ iteration of the Gibbs sampler is estimated using the three steps: (1) take a draw of $b^t$ from f($\beta$ ,W), where $\beta$ is the mean, conditional on $W^0$ and $\beta^0_n$; (2) take a draw of $W^t$ from an inverted Wishart distribution conditional on $b^t$ and $\beta^0_n$; and (3) take a draw of $\beta^t_n \ \forall$ n conditional on $b^t$ and $W^t$ using the MH algorithm (Train 2003). This sequence of steps is repeated for many iterations until the values have converged to draws from the posterior. The MH algorithm is necessary for $\beta_n$ because the posterior for each individual does not follow a convenient form, whereas for b and W the distribution is specified by the investigator and is easy to draw from.[3] Unlike the classical approach, HB provides individual-specific posterior estimates of $\beta_n$.

Iterations occurring prior to convergence are discarded; these iterations are termed the 'burn-in period'(Cowles and Carlin 1996). Following convergence, a number of draws from the posterior are used to calculate the relevant statistics (means and standard errors); only every $k^{th}$ iteration is retained in this process to reduce correlation across the Markov Chain (Kass et al. 1998). Of interest is determining if the MCMC procedures have indeed reached convergence. There are a number of diagnostic tests for this purpose; however, these tests are somewhat difficult to apply and may not be able to detect convergence failure. Gelman and Rubin suggest estimating the model starting from several different starting points and formally testing if the statistics are the same (Gelman and Rubin 1992). Train takes a pragmatic approach and suggests that for HB models, convergence can be ascertained by examining if the draws after burn-in are trending. If convergence has been attained, the draws move around the posterior (Train 2003).

[3] Interested readers are referred to Train 2001 for a more detailed description of MCMC methods.

## 3. **DCE Application: Genetic Testing**

The application involves the introduction of a novel health technology that identifies chromosomal abnormalities in children or adults with MR. MR affects 2 to 3% of the Canadian population. Individuals with MR have limitations in their mental functioning and in skills such as communicating and personal care. These limitations will cause the individual to learn and develop more slowly than a typical child. Children with MR may take longer to learn how to speak, walk, and attend to their personal needs. These individuals are also likely to have trouble learning in school. Whilst there are many possible reasons why a person has MR, the cause is often a genetic condition, a problem during pregnancy or at birth, or childhood disease. Knowing the cause of MR can result in more accurate genetic counselling and prognosis, availability of prenatal diagnosis, elimination of numerous diagnostic tests, and improved quality of life for the affected children and their families.

When the paediatrician suspects an individual has MR, she/he is referred to a geneticist to search for possible genetic causes. The geneticist typically takes a sample of DNA and uses cytogenetic analysis to identify chromosomal abnormalities. The clinical standard for cytogenetic testing has been to employ a karyotype. Karyotyping, in its various forms, has been the standard of care for identifying chromosomal abnormalities for 40 years (Friedman et al. 2006). Karyotyping has evolved since its introduction, but it is a lengthy procedure that requires highly skilled interpretation. Although exact rates of positive diagnosis are not possible, most case series report that standard cytogenetic screening can identify a genetic cause in approximately 10% of those who are tested (van Karnebeek et al. 2005). Unfortunately, for approximately one-third to one-half of all children with MR, the cause will remain unknown. Genome British Columbia has recently developed a technology called array genomic hybridization (AGH) that uses the entire genome to identify chromosomal imbalances that are very small (<5-10 Mb). Compared to Karyotyping, AGH can detect chromosomal abnormalities up to 100 times smaller, and may result in faster test results because it is an automated process. AGH is also expected to be significantly more expensive at approximately CDN$1,547 versus CDN$444 per test for a standard karyotype.[4]

Although AGH is a technology that is continuing to be developed and refined, when 100 children whose MR was idiopathic were tested with AGH, 11 children were found to have a

[4] Costs do not include the use of confirmatory tests such as fluorescent in-situ hybridization (FISH).

genetic cause of their MR (Friedman et al. 2006). Given that AGH can also detect chromosome aberrations found by karyotype methods, these initial findings suggest that AGH may detect chromosomal abnormalities in more than twice as many cases.

*3.1 Identification of attributes and questionnaire design*

The identification of the attributes and levels of the DCE began with expert interviews with geneticists, genetic counselors, and health outcome researchers from the University of British Columbia and the University of Aberdeen; the process also included a literature review. After a pilot study, three attributes were included in the DCE design: (1) number of children test whose genetic condition is identified with this test (levels: 10 in 100, 14 in 100, 20 in 100, 25 in 100); (2) time waiting for results (levels: 1, 3, 6, 12); and (3) cost to you (levels: 750, 1100, 1750, 2500). Individuals were to choose between two generic alternatives and a 'neither test' alternative to allow for non-demanders. This specification led to 64 possible profiles or treatment combinations for the experimental design. An example of a choice question for the final questionnaire is presented in Figure 1.

**Figure 1. Choice question example**

**Choice question 1: Would you prefer postnatal test A, postnatal test B or neither?**

| | **Test A** | **Test B** | **Neither Test** |
|---|---|---|---|
| Number of children tested whose genetic condition is identified with this test | 10 children in 100 with DD who are tested | 14 children in 100 with DD who are tested | In this scenario, you would prefer neither of the tests to be conducted |
| Time waiting for the results of the test | 1 week | 3 weeks | |
| Cost to you | $750 | $1100 | |

**In this situation would you choose (please select only one of the following options):**
**___ 1) Test A**
**___ 2) Test B**
**___ 3) Neither**

D-error heterogeneous procedures (Sandor and Wedel 2005) for mixed logit models were employed to construct the final designs. A D-error design seeks to minimize the variability in the parameters of the model by minimizing a scalar measure of the variance-covariance matrix called the $D_M$-error ; the resulting designs allow the parameter estimates to be measured with greater accuracy (Sandor and Wedel 2005; Sandor and Wedel 2002). The base design was constructed using orthogonal arrays, and 16 choice sets were generated using the foldover technique to ensure level balance and minimum level overlap. To search for more efficient

designs, choice sets were generated using two design algorithms—swapping and cycling. Prior coefficients obtained from the pilot study were used to estimate the information matrix. Because the prior coefficients were based on limited information, heterogeneous designs were constructed to mitigate the potential effects of misspecified priors. The heterogeneous design approach produces several subdesigns to be administered across study participants. This allows for greater variability in the attribute levels, which may result in more efficient designs if the priors are misspecified (Sandor and Wedel 2005). For each design, relative efficiency was derived as the ratio of the D-errors. The base design resulted in a $D_M$-error of 0.147; the $D_M$-error for the heterogeneous design was 0.047. The expected efficiency gain of using the D-error design was 34%.

**4. Econometric specification**

Two econometric models were examined: M1 was an all parameters random specification, and M2 specified coefficients that were both fixed and random. Under the M1 specification, partwoths are estimated for number of children tested whose genetic condition is identified, time waiting for results, cost to you, and the neither test alternative. The neither alternative was dummy coded with 'neither test' as the base so that the value of testing is estimated relative to no-test. If individuals have an overall preference to have genetic testing conducted, the neither parameter will be positive.

In the M1 specification the neither alternative is specified to follow the normal distribution. Defining the neither alternative in this way has the benefit of nesting the two testing options, which allows for correlation amongst these alternatives (Hall et al. 2006; Train 2003). This random-errors specification is analogous to the nested logit. For both M1 and M2, the variables time and cost were designated to follow a log-normal distribution. The parameters on these attributes were *a priori* expected to be negative because individuals do not prefer higher prices or longer waiting times. The log-normal specification, however, has support strictly on the positive side of the distribution; as such, the negative of the data for these variables was entered into the model so that the coefficients are expected to be positive. Cost was scaled to be hundreds of Canadian dollars. Finally, number of children diagnosed was given a normal distribution to account for the possibility that respondents may dislike having a higher rate of detected genetic abnormalities. In the M2 model, some coefficients are specified as fixed. Including fixed coefficients may be desirable in several circumstances. For example, it has been argued that the

MXL with all random coefficients is nearly unidentified empirically (Ruud 1996). This is dually true when alternative specific constants are modeled and specified to follow distributions that are similar to the extreme value distribution, such as the normal (Train 2003). In these circumstances, it is recommended that at least one coefficient is held constant. The M2 model specifies the neither coefficient as fixed. Given that different segments of the population may feel differently about the overall value of testing, this constant is interacted with a number of demographic variables.

*4.1 MSLE simulation*

The classical models used 2,000 Halton draws to approximate the integral. A fewer number of draws can be specified. Using a smaller number draws, however, may mask problems related to model identification (Walker 2007). For example, if a model is unidentified empirically and a small number of draws are used, the model may still converge (Walker 2007). If a large number of draws are used, the information matrix (inverse Hessian) will likely be singular if the model is unidentified. Under both the Bayesian and classical estimation procedures, the information matrix is specified to allow for correlation between the coefficients. The MSLE MXL model is estimated in Gauss7.0 using code from Kenneth Train (Train 2004).

*4.2 Hierarchical Bayes*

In the Bayesian models, specifying a prior distribution for the parameters is necessary (equation 4). Under both specifications, the priors on the random parameters were specified to be uninformative. The prior on the covariance matrix was specified to be inverted Wishart. All prior distributions in the HB model were conjugate priors.[5] The Bayesian models employed 10,000 iterations for the burn-in period. Further to this, 2,000 iterations were retained to calculate the parameters and standard errors of the posterior. Every $10^{th}$ iteration was kept after convergence to reduce correlation between draws from the posterior; this meant that after convergence, a total of 20,000 draws were estimated. The iterations were inspected visually to assess if convergence had successfully been reached by examining if the draws traversed the posterior. The model was estimated using Gauss7.0 and Matlab using codes obtained from Kenneth Train (Train 2006).

[5] The prior distribution is considered to be conjugate if it belongs to a class of likelihood functions where the resulting posterior distribution is in the same family as the prior.

## 5. Results

Ethics to conduct the DCE was approved by the University of British Columbia, Vancouver, Canada. A market research company recruited individuals greater than 18 years of age who were residents of British Columbia and could read and understand English. 785 individuals of 1057 who choose to participate in the survey completed all 16 choice questions over the internet. 29 of 785 responses were discarded because of evidence these individuals 'clicked-through' the survey. Table 1 provides an overview of the sociodemographic variables. The average age of respondents in this population was 50 years. Income was defined as total family income; in our sample, 18, 44 and 27% of individuals were in the low, middle, and high-income groups, respectively. 11% of the sample choose not to state their income. 12% of individuals were single, and 29% did not have any children. 13% of individuals had obtained a university degree, 25% had completed a vocational or trades degree, and 29% completed at least some post-secondary education.

Table 1.
Demographic Characteristics of the Sample

| Variable and coding | Average or Percentage of Sample |
|---|---|
| **Continuous Variables** | |
| Age | 50 |
| **Categorical Variables** | |
| **Income** | |
| Income <=$20,000, low income | 18% |
| Income >20,001 and <=80,000, mid income | 44% |
| Income >=80,001, high income | 27% |
| Income not reported, NR | 11% |
| **Family** | |
| Marital Status, single | 12% |
| Children | 29% |
| Gender (=1 if female) | 51% |
| **Education** | |
| High school diploma, HS | 31% |
| Some post-secondary, PS | 30% |
| Completed vocational or trade diploma | 25% |
| Completed university degree, Uni | 13% |
| Education not available, ENA | 1% |

### *5.1. M1, all parameters random*

The results of M1 using both classical and Bayesian procedures are given in Table 2. The starting values used in the MSLE procedures were obtained from the parameters of a standard main-effects multinomial logit model. Using these initial values, convergence to the maximum

was not obtained using the maximum likelihood approach. Problems of convergence are not unusual when log-normal distributions are specified because this distribution can produce log-likelihood functions that are highly non-quadratic. To obtain better starting values, a MXL estimated using normal distributions for each parameter was necessary. The parameters of the log-normal were then calculated using standard transformations. Under these starting values, the classical model reached convergence in 14.50 minutes. The starting values used in the HB model were set to be zero. The HB model took 8.6 minutes to complete using a burn-in of 10,000 iterations and 2,000 retained draws to approximate the posterior. For these data, the Bayesian procedure was 1.68 times quicker than the classical procedure. The mean and heterogeneity parameters in both models were highly significant ($p<0.00$), which indicates that the hypothesis of no variance is rejected.

Table 2 Expected partwoths under classical and Bayesian approaches

| Variables | Classical | | | | Bayesian | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean Coeffic ient | SE | Standard Deviation | SE | Mean Posterior | SE | Variance/Std. Deviation | SE of Variance |
| Testing | 5.09 | 1.24 | 4.45 | 1.05 | 5.49 | 0.28 | 24.33/4.93 | 2.64 |
| NC | 17.67 | 1.96 | 19.13 | 2.58 | 17.98 | 0.89 | 356/18.87 | 35.92 |
| Time | -2.64 | 0.18 | 1.19 | 0.22 | -2.65 | 0.10 | 1.56/1.25 | 0.25 |
| Cost/100 | -1.48 | 0.10 | 1.22 | 0.38 | -1.52 | 0.06 | 1.80/1.34 | 0.16 |
| Number of draws: 2,000<br>Log-likelihood=-7021<br>Minutes to Convergence: 14.50<br>Percentage of correct predictions: 10,809 of 12,096 (89.36%) | | | | | Burn-in:10,000;2,000 to approximate posterior<br>Simulated log-likelihood:-7025<br>Minutes to Convergence:8.6<br>Percentage of correct predictions: 10,811 (89.37%) | | | |

Examination of the mean and standard deviations arising from the two approaches shows the classical and Bayesian procedures yield similar results. For the testing parameter, there was an overall preference to have a genetic test conducted. The exact percentage of respondents who were estimated to have a negative preference for testing can be calculated using the Z distribution. Under the classical procedure, 12.6% of respondents had negative coefficients. This number was 13% in the Bayesian procedure. If a 95% confidence level approach is taken, the numbers are 7.6% and 8.3% for the classical and HB approaches, respectively. For the variable NC, the overall mean was positive, which indicated that as the number of children diagnosed increased, so did the average preference for that attribute. Again using the Z table, MSLE resulted in 17.7% of draws below 0 for NC, and HB resulted in 17% draws below zero. The number of draws below zero are somewhat larger than expected. A possible interpretation of this

result is that a percentage of individuals prefer less children to have a diagnosis because they do not approve of a label being put on these children, especially in the case of mild MR.

The parameters on time and cost were restricted to be all the same sign, which is a product of specifying that these parameters follow a log-normal distribution. If the estimates on time and cost are intended to be used for WTP calculations, the coefficients for these parameters need to be transformed from log-normal to normal. Transforming the mean partworth for time resulted in classical estimate of the mean and mode of 0.14 and 0.07, respectively; the cost partworth had a mean of 0.48 and a mode of 0.23 using MSLE. In the Bayesian procedures, time has a mode of 0.07 and a mean of 0.15, whilst cost had a mode of 0.22 and a mean of 0.54. The large difference between the mean and mode is because the log-normal distribution is skewed, which can result in draws from the distribution being extreme, increasing the mean of the distribution.

The ability of the estimated model to predict the observed choices of respondents was also tested under the two estimation procedures. For this exercise, the probability of each alternative for every respondent was calculated using parameters from both estimation procedures. Under RUT, the alternative with the highest probability should be the alternative that the respondent chooses. For the classical model, the MXL was able to correctly predict 10,809 of 12,096 choices (89.36%), whilst for the Bayesian model the number of correctly predicted choices was 10,811 (89.37%).

*5.2 M2, estimation with fixed and random coefficients.*

The inclusion of parameters that do not vary in the population may be necessary when, for example, the researcher encounters problems with model identification, or the marginal utility of the parameter is required to be constant across individuals. In HB, the inclusion of a fixed coefficient may increase estimation time because an additional level of the MH algorithm is necessary; drawing from the posterior using MH is slower than Gibbs sampling (Train 2003). Classical likelihood procedures may have the advantage in this instance because including a fixed coefficient does not require any alteration to the estimation process.

An HB model with random parameters for NC, time and cost, and fixed parameters for testing and the interaction between testing and: gender, income, education, age/100, and if the respondent was single and did not have children was estimated. All categorical variables were

effects coded. Gender had a base category for being male; for income, the base category was low income; education's base category was high school (HS) or if data on education was not available (ENA); and single had a base category of having children or being married. The results from this model are presented in Table 3.

Table 3. HB with fixed coefficients

| **Variables** | **Mean Coefficient** | **Standard Error** | **Variance/Standard Deviation** | **Standard Error of Variance** |
|---|---|---|---|---|
| **Random Parameters** | | | | |
| NC | 14.94* | 0.78 | 236.31*/15.38 | 25.93 |
| Time | -2.80* | 0.09 | 1.34*/1.15 | 0.21 |
| Cost | -1.49* | 0.04 | 1.20*/1.09 | 0.10 |
| | | | | |
| **Fixed Parameters** | | | | |
| Testing | 3.16* | 0.28 | | |
| Gender ⊗ Testing | -0.16** | 0.08 | | |
| Age ⊗ Testing | 0.15 | 0.54 | | |
| Mid income ⊗ Testing | 0.33* | 0.11 | | |
| High income ⊗ Testing | -0.68* | 0.12 | | |
| Voc ⊗ Testing | -0.023 | 0.18 | | |
| Uni ⊗ Testing | 0.07 | 0.14 | | |
| Single ⊗ Testing | -0.42* | 0.11 | | |
| Burn-in:10,000; 2,000 approximated posterior<br>Simulated log-likelihood: -7397<br>*Significant at 0.00% level<br>**Significant at 10% | | | | |

The simulated log-likelihood for this HB model was -7397, which indicated that M2 loses explanatory power when compared to M1 (log-likelihood -7025). The interaction between testing demographic characteristics yielded interesting results. Each interaction parameter adjusts the mean testing parameter according to the demographics categories. The interaction between the variables was modeled because only differences in utility are important, and sociodemographic variables do not differ by alternative. Women, on average, are expected to have a slightly smaller overall preference for testing compared to men. Individuals who have family income between CDN$20,001 and CDN$80,000 per year preferred having testing more than those in the high or low-income categories. Individuals who were single and did not have children preferred testing to a lesser degree than those who have families. The coefficients on the interaction variables were all small in magnitude compared to the overall testing parameter. This indicates that the heterogeneity in the testing variable observed in the M1 specification does not, to a large degree, depend on the sociodemographic variables collected in the DCE. This result is not surprising

because demographic variables are often defined too broad to adequately describe latent unobservable characteristics.

The results between the HB and classical procedures were again close for M2, except for the parameters on number of children diagnosed and cost. The classical parameters for the mean and standard deviation of the coefficient on number of children diagnosed was 15.73 and 15.07, respectively, whilst these parameter using HB were 14.94 and 15.38, respectively. For cost, the mean and standard deviation was -1.60 and 1.13 using MSLE, and -1.49 and 1.20 under the Bayesian analysis.

The HB model for M2 took 31 minutes to run, whilst the same model under MSLE procedures took 28 minutes to reach convergence. This meant that the run time for the Bayesian model was approximately increased by a factor of three with the inclusion of fixed coefficients, whilst for the classical procedures, run time was increased by a factor of less than 2. Still, the relative similarity between the HB and classical models was somewhat surprising as other investigations report that the classical approach is substantially quicker than HB when fixed coefficients are included (Train 2003).

## 6. Discussion

The use of the MXL behavioural specification to model DCE data in health economics will become more common because it avoids some significant shortcomings inherent in the multinomial logit and nested logit models. This paper applied two estimation approaches, hierarchical Bayes and maximum simulated likelihood, to estimate a MXL with bounded distributions for the parameters. Previous work in health economics have solely used MSLE and normal distributions specified for the parameters. Despite the substantial differences in the estimation algorithms, the Bayesian and classical estimates were observed to be very close.

Under a specification of all normal distributions for the coefficients, both models are straightforward to implement if a sufficient number of draws are used in the classical approach, or a sufficient burn-in period is specified so that the MCMC algorithm is drawing from the posterior. The use of bounded distributions, however, is sometimes necessary, especially when a good deal of heterogeneity has been observed in the data. Bounded distributions ensure that the interpretation of the individual level partworths is consistent with economic theory; improvements in model fit are also observed as the specified distribution more accurately

approximates *a priori* economic interpretation. As exampled using the M1 specification, the use of bounded distributions will complicate convergence in the classical estimation approach because the log-normal distribution can create a log-likelihood function that is non-quadratic. The maximization procedures developed for MSLE works best when the function is close to quadratic; if this is not the case, the likelihood estimation procedures may fail to find an increase even when a maximum has not been obtained.

Bayesian procedures do not require the maximization of a likelihood function or the simulation of an integral, which gives HB an advantage over classical techniques under circumstances where convergence is difficult, or if it is uncertain if a global maximum has been attained. In these situatations, HB is recommended regardless of the researcher's philosophical interpretative beliefs on probability because the mean of the Bayesian posterior will likely be close to the classical mean estimate. The use of the mean of HB posterior in this way is supported by the Bernstein von-Mises theorem (Train 2003), which states that the parameters share the same information matrix and are sampled from the same asymptotic distribution.

The MXL behavioural specification requires substantially more computational estimation time compared to the logit and nested logit specifications. This can make the process of choosing the final model tedious. Under the M1 specification, it was demonstrated that HB takes substantially less estimation time when compared to classical procedures. Thus, using the Bayesian approach may result in modeling taking weeks instead of months.

Researchers can specify a smaller number of draws so that estimation time decreases; however, with a smaller number of draws, numerical unidentification may be masked, which may mislead the researcher and result in biased and inefficient parameters. If the model is unidentified, the priors in the Bayesian approach may be used to provide information for the identification of the model; such an exercise is beyond the scope of this paper, but is an avenue for future research in HB modeling.

Incorporating fixed coefficients is desirable if the model is unidentified, or if, from a theoretical perspective, the use of random coefficients is not desirable or unjustified. A loss of fit compared to M1 was observed in Model M2. Although useful information was obtained in terms of how individuals grouped by socio demographic data view the overall value of any genetic test, the degree of their deviation form the overall preference from testing was not particularly

illuminating. Incorporating fixed coefficients under the Bayesian procedures slowed estimation time considerable, but not to a degree that greatly differed from the classical procedures.

The similarity between the estimates of the HB and classical estimates does not mitigate the interpretive differences between the approaches (Huber and Train 2001). Furthermore, these results are not likely to hold in smaller sample sizes because the two procedures handle uncertainty in different ways (Huber and Train 2001). For example, under Bayesian analysis, model uncertainty is characterized by the posterior distribution, whilst in classical procedures uncertainty arises because different samples produce different results and the resulting sampling distribution of point estimates is the distribution obtained if many different samples were taken. Nonetheless, the similarity between the two approaches provides a useful tool to examine issues of convergence to the posterior (in HB), or finding a global versus local maxima under classical approaches. Researchers that wish to maintain a classical approach to statistics can use HB procedures to find good starting points if convergence to the maximum is difficult; Bayesian statisticians can used MSLE to examine if the MCMC algorithm is indeed sampling from the posterior distribution.

References

Bhat, C. R. (2000), "Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling," Transportation Science, 34 (2), 228-38.

Cowles, M. K. and B. P. Carlin (1996), "Markov chain Monte Carlo convergence diagnostics: A comparative review," Journal of the American Statistical Association, 91 (434), 883-904.

Friedman, J. M., A. Baross, A. D. Delaney, A. Ally, L. Arbour, J. Asano, D. K. Bailey, S. Barber, P. Birch, M. Brown-John, M. Q. Cao, S. Chan, D. L. Charest, N. Farnoud, N. Fernandes, S. Flibotte, A. Go, W. T. Gibson, R. A. Holt, S. J. M. Jones, G. C. Kennedy, M. Krzywinski, S. Langlois, H. Y. I. Li, B. C. McGillivray, T. Nayar, T. J. Pugh, E. Rajcan-Separovic, J. E. Schein, A. Schnerch, A. Siddiqui, M. I. Van Allen, G. Wilson, S. L. Yong, F. Zahir, P. Eydoux, and M. A. Marra (2006), "Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation," American Journal of Human Genetics, 79 (3), 500-13.

Gelman, A. and R. Rubin (1992), "Inference from iterative simulation using multiple sequences," Statistical Science, 7 (4), 457-72.

Hall, J., D. G. Fiebig, M. T. King, I. Hossain, and J. J. Louviere (2006), "What influences participation in genetic carrier testing? Results from a discrete choice experiment," Journal of Health Economics, 25 (3), 520-37.

Hensher, D. and W. Greene (2003), "The mixed logit model: The state of practice," Transportation, 30, 132-76.

Huber, J. and K. Train (2001), "On the similarity of classical and bayesian estimates of individual mean partworths," Marketing Letters, 12 (2), 259-69.

Kass, RE., P. Carlin, A. Gelman, and R. Neal (1998), "Markov chain monte carlo in practice: a roundtable discussion," The American Statistician, 52 (2), 93-100.

King, M. T., J. Hall, E. Lancsar, D. Fiebig, I. Hossain, J. Louviere, H. K. Reddel, and C. R. Jenkins (2007), "Patient preferences for managing asthma: results from a discrete choice experiment," Health Econ, 16 (7), 703-17.

Lancsar, E. J., J. P. Hall, M. King, P. Kenny, J. J. Louviere, D. G. Fiebig, I. Hossain, F. C. Thien, H. K. Reddel, and C. R. Jenkins (2007), "Using discrete choice experiments to investigate subject preferences for preventive asthma medication," Respirology, 12 (1), 127-36.

Lenhard, W., E. Breitenbach, H. Ebert, H. J. Schindelhauer-Deutscher, and W. Henn (2005), "Psychological benefit of diagnostic certainty for mothers of children with disabilities: lessons from Down syndrome," Am J Med Genet, 133 (2), 170-5.

McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in Frontiers in Econometrics, Zarembka P, Ed. New York: Academic Press.

McFadden, D. and K. Train (2000), "Mixed MNL models for discrete response," Journal of Applied Econometrics, 15 (5), 447-70.

Rauch, A., J. Hoyer, S. Guth, C. Zweier, C. Kraus, C. Becker, M. Zenker, U. Huffmeier, C. Thiel, F. Ruschendorf, P. Nurnberg, A. Reis, and U. Trautmann (2006), "Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation," Am J Med Genet A, 140 (19), 2063-74.

Revelt, D. and K. Train (2000), "Specific tastes parameters and mixed logit," working paper, Department of Economics, University of California, Berkeley.

Ruud, P. (1996), "Simulation of the multinomial probit model: an analysis of covariance matrix estimation," working paper, Department of Economics, University of California, Berkeley.

Sandor, Z. and M. Wedel (2005), "Heterogeneous Conjoint Choice Designs," Journal of Marketing Research, XLII, 210-18.

---- (2002), "Profile construction in experimental choice designs for mixed logit models," Marketing Science, 21 (4), 455-75.

Train, K. (2003), Discrete Choice Methods with Simulation (1st ed.). Cambridge: Cambridge University Press.

---- (2004), "Mixed logit estimation for panel data using maximum simulated likelihood," Retrieved December 17 2005.From http://elsa.berkeley.edu/software/abstracts/train0296.html.

---- (2006), "Mixed logit with bounded distributions of correlated partwoths," Retrieved August 19 2006.From http://www.elsa.berkeley.edu/software/abstracts/0016.html.

---- (1998), "Recreation demand models with taste differences over people," Land Economics, 74 (2), 230-39.

Train, K. and G. Sonnier (2005), "Mixed logit with bounded distributions of correlated partworths," in Applications of simulation mehtods in environmental and resource economics, A. Alberini and R. Scarpa, Eds. Vol. 1. Dordrecht: Springer.

van Karnebeek, C. D., M. C. Jansweijer, A. G. Leenders, M. Offringa, and R. C. Hennekam (2005), "Diagnostic investigations in individuals with mental retardation: a systematic literature review of their usefulness," Eur J Hum Genet, 13 (1), 6-25.

Walker, J. (2007), "Mixed logit (or logit kernel) model: dispelling misconception of identification," Transportation research record, 1805, 86-98.