Comparing Alternative Models of Heterogeneity in Consumer Choice Behavior

by

Michael Keane and Nada Wasi

University of Technology Sydney

May 2009

Abstract

While there is general agreement that consumer taste heterogeneity is crucially important in marketing, there is no consensus on a preferred approach to modeling heterogeneity. In this paper, we assess the performance of five alternative choice models, using ten empirical data sets. We include the popular latent class (LC) model, and the mixed logit (MIXL) model where utility weights are assumed to be multivariate normal. The new generalized multinomial logit (G-MNL) and scale heterogeneity (S-MNL) models are also included. G-MNL generalizes MIXL by allowing for heterogeneity in the scale coefficient. S-MNL is a special case of G-MNL where only scale heterogeneity is present. Finally, we consider the potentially more flexible mixture-of-normals logit or "mixed-mixed" logit (MM-MNL) model. We find that according to the Bayes information criteria, G-MNL is preferred in 4 datasets while MM-MNL and S-MNL are preferred in 3 datasets each. By further investigating what behavioural patterns each model can capture better than others, we find that: (i) the more flexible heterogeneity distributions of G-MNL and MM-MNL allow them to better capture "extreme" (i.e., lexicographic) as well as "random" behaviour; and (ii) which model is preferred depends on the structure of heterogeneity, which differs across datasets.

Keywords: Choice models, Mixture models, Consumer heterogeneity, Choice experiments

Acknowledgements: Keane's work on this project was supported by ARC grant FF0561843.

Corresponding author: Michael P. Keane, Faculty of Business, UTS.
Phone: +61 2 9514 9742. Fax: +61 2 9514 9743 Email: michael.keane@uts.edu.au

I. Introduction

For at least 25 years, there has been a large ongoing research program in marketing on modeling consumer heterogeneity. Much of this work was motivated by the classic Guadagni and Little (1983) paper on modeling choice behavior using scanner panel data. The issue of how best to model heterogeneity is important for many reasons. Most obviously, estimates of own and cross price elasticities of demand may be severely biased if one does not properly account for taste heterogeneity. But understanding taste heterogeneity is also critical for a host of other problems, such as new product development (NPD), product positioning and advertising, optimal price discrimination strategies, the development of menus of product offerings, considerations of product image and/or brand equity, etc..

Most researchers would now agree that the simple multinomial logit (MNL) model of McFadden (1974), which assumes homogeneous tastes for observed product attributes, is inadequate to model choice behaviour in many contexts. Many popular models extend MNL to allow for unobserved heterogeneous tastes over the observed product attributes. However, there is no general consensus within marketing on a preferred approach to modeling heterogeneity.

In this paper, we consider several alternative approaches to modeling consumer taste heterogeneity. The models we consider include two that are currently quite popular – the latent class (LC) model and the mixed logit (MIXL) model. We also consider two new models recently developed by Fiebig, Keane, Louviere and Wasi (2009) – the scale heterogeneity logit (S-MNL) and the generalized multinomial logit (G-MNL). G-MNL nests the scale heterogeneity model and the MIXL model. Finally, we also consider a model that is present in the literature but has rarely been applied, the mixture of normals logit model or "mixed mixed" logit (MM-MNL) model. MM-MNL specifies that the mixing distribution in MIXL is a discrete *mixture*-of-multivariate normals.

Fiebig, Keane, Louviere and Wasi (2009) report a series of experiments where G-MNL generally fits consumer choice behavior better than MIXL. However, the G-MNL and MM-MNL models are closely related, in that G-MNL can be interpreted as letting heterogeneity take the form of a continuous mixture of scaled normals. This is typically more parsimonious than the MM-MNL specification (i.e., discrete mixture of normals), but still quite flexible. It is thus of considerable interest to compare the performance of MM-MNL and G-MNL.

Here, we compare the performance of all 5 of these alternative models of heterogeneity on 10 empirical data sets. Using the Bayes information criteria (BIC), we find that G-MNL is the preferred model of heterogeneity in 4 data sets, while MM-MNL and S-MNL are preferred in 3 each. Strikingly, the MIXL and latent class models, which are arguably the most popular models of heterogeneity in use in marketing today, are never preferred (and rarely come even close to matching the fit of the preferred models). In the paper, we explain what features of the data the G-MNL, MM-MNL and S-MNL models capture that MIXL and LC models fail to capture.

Essentially, what we find is that G-MNL, MM-MNL and S-MNL all do a better job of capturing behaviour of consumers who exhibit "extreme" or lexicographic behaviour, in the sense that they base choice largely on a single attribute (i.e., thus choose the alternative that has the lowest price, highest quality, etc.). At the same time, these models are also better able to capture the behaviour of "random" consumers, whose choices are only slightly influenced by observed product attributes. What determines which model is preferred among G-MNL, MM-MNL and S-MNL is more subtle.

G-MNL and S-MNL can never dominate MM-MNL in terms of the likelihood function, because the mixture-of-normals can approximate any heterogeneity distribution arbitrarily well. But it may require large number of parameters to do so. Thus, G-MNL and S-MNL can potentially dominate MM-MNL according to information criteria like BIC that favor more parsimonious models. Where G-MNL dominates MM-MNL, the two models actually capture similar observed choice patterns, but G-MNL is preferred because it achieves this in a more parsimonious way. In some cases, S-MNL is preferred because almost all the likelihood improvement that can be achieved by introducing a flexible heterogeneity distribution is achieved by introducing scale heterogeneity alone. As S-MNL is a very parsimonious model, it is preferred by BIC in such cases.

MM-MNL outperforms G-MNL when there is a small, but not trivial, fraction of respondents who exhibit "extreme" behaviour, but whose choices are largely determined by some attributes which are not viewed as important by the majority of the respondents. For example, regarding pizza delivery services, we find there are large segments of consumers who place great weight on either price or ingredient quality. In the text we will refer to these as "major" attributes. But there are also small fractions who care greatly about other attributes, like gourmet pizza, woodfire cooking, etc.. In the text we will refer to these as "minor" attributes. In some (but not all) cases, MM-MNL provides a much better fit to consumers who have a strong preference for one of these minor attributes.

Finally, we conduct counterfactual simulations to examine differences in the demand predictions from G-MNL and MM-MNL. In most cases these models give very similar predictions for the effects of changing product attributes. In particular, they always give very similar predictions for the effects of changing "major" attributes. In some (but not all cases) the two models give quite different predictions for the effects of changing "minor" attributes.

II. Alternative Models of Consumer Choice Behaviour

In the traditional multinomial logit (MNL) model (McFadden, 1974), the utility to person n from choosing alternative j on purchase occasion (or in choice scenario) t is given by:

$$U_{njt} = \beta x_{njt} + \varepsilon_{njt} \qquad n = 1,...,N; \quad j = 1,...,J; \quad t = 1,...,T, \tag{1}$$

where x_{njt} is a K-vector of observed attributes of alternative j, β is a vector of utility weights (assumed homogenous across consumers) and ε_{njt} ~ iid extreme value is the "idiosyncratic" error component. The x_{njt} for $j = 1,...,J$ may include alternative specific constants (ASCs) to capture persistence in the unobserved attributes associated with each option j over choice occasions. The iid extreme value assumption leads to a closed form expression for the choice probabilities:

$$P(j \mid X_{nt}) = \exp(\beta x_{njt}) \Big/ \sum_{k=1}^{J} \exp(\beta x_{nkt})$$

where X_{nt} is the vector of attributes of all alternatives j=1,…J. However, due to the restrictive assumptions that (i) the ε_{njt} are iid and (ii) tastes for observed attributes are homogenous, MNL imposes a very special structure on how changes in elements of x_{njt} can affect choice probabilities. For instance, the IIA property is implied by the iid assumption. And, with panel data, the basic MNL model does not incorporate individual-specific unobserved taste heterogeneity.

Several alternative models which avoid IIA and/or allow for unobserved heterogeneous tastes over the observed product attributes have been proposed. These included the nested logit model (McFadden, 1978), the generalized extreme value (GEV) model (McFadden, 1978), the multinomial probit (MNP) model (Thurston, 1927), the mixed MNL or mixed logit (MIXL) model (Ben-Akiva and McFadden *et al.*, 1997; McFadden and Train, 2000), the latent class (LC) model (Kamakura and Russell, 1989), and their variants.

The mixed logit and latent class models appear to be the most popular models of heterogeneity in use in marketing and other fields today. Because of its ability to approximate some other existing models (e.g., nested logit and MNP) together with its ease of use, MIXL is appealing. For instance, one may try to approximate MNP by specifying the mixing distribution of the alternative specific constants to be multivariate normal (MVN). Most applications of the MIXL model assume the whole vector of preference weights has a MVN distribution in the population.[1]

The latent class model is thought to be robust to non-normal heterogeneity distributions as its semi-parametric formulation allows for a more flexible shape of the taste distribution (*e.g.*, skewed or multi-modal). Of course, if the true heterogeneity distribution departs substantially from normality, LC maybe better able to capture that than MIXL. However, by using a finite number of homogenous segments (typically 3-5), LC is usually found to understate the extent of heterogeneity in the data (Elrod and Keane, 1995; Allenby and Rossi, 1998). Applied researchers are often aware of the disadvantages of MIXL and LC, and sometimes estimate both models in their empirical applications (see, e.g., Greene and Hensher (2003), Hole (2008)).

Recent studies have started to question whether the specification of MIXL with a MVN mixing distribution is adequate for explaining key features of choice data. In particular, Louviere and colleagues (1999, 2002) have argued that the major source of heterogeneity in choice data comes from "scale heterogeneity" – i.e., a generally scaling up or down of the entire vector of attribute weights – as opposed to the random coefficients specification of MIXL.

Recently, Fiebig, Keane, Louviere and Wasi (2009) developed a generalized multinomial logit model (G-MNL) which extends MIXL by incorporating both scale heterogeneity and a random coefficient vector. G-MNL nests both MIXL and the "scale heterogeneity" model (S-MNL). The latter, as its name implies, includes only scale heterogeneity. They found that G-MNL outperform MIXL with MVN in eight out of the 10 datasets that we examine in this paper.

Another flexible model that is present in the literature but has rarely been applied is the mixture of normals logit model or "mixed mixed" logit (MM-MNL) model. MM-MNL specifies that the mixing distribution in MIXL is mixture-of-multivariate normals. Under the Hierarchical Bayes approach, a literature is also moving from using normal weak priors to using mixtures-of-

[1] There seems to be a misconception among practitioners that mixed logit with a MVN mixing distribution can approximate any random utility model well. In fact, McFadden and Train (2000) showed that the MIXL model can approximate any random utility model arbitrarily well *if the researcher specifies the correct mixing distribution*.

normal priors for the individual level parameters as it accommodates a larger range of non-normal posteriors. This approach has been extended to probit by Geweke and Keane (1999, 2001) and to MIXL by Rossi et al (2005) and Burda, Harding and Hausman (2008).

Next, Section II.A reviews the MIXL and LC models. Section II.B reviews the S-MNL and G-MNL models. And Section III.C discusses the MIXL model with a discrete mixture-of-normals as the mixing distribution (that is, the MM-MNL model).

II.A. Models with Unobserved Taste Heterogeneity: Mixed Logit and Latent Class models

The MIXL and LC models simply extend the MNL to allow for random coefficients on the observed attributes, while continuing to assume the "idiosyncratic" error component ε_{njt} is iid extreme value.

$$U_{njt} = \beta_n X_{njt} + \varepsilon_{njt} \qquad n = 1,...,N;\ j = 1,...,J;\ t = 1,...,T\ . \tag{2}$$

The difference is that MIXL specifies a continuous distribution for β_n while LC assumes that the underlying distribution is discrete. For MIXL, the model is often written as

$$U_{njt} = (\beta + \eta_n)x_{njt} + \varepsilon_{njt} \qquad n = 1,...,N; \quad j = 1,...,J; \quad t = 1,...,T, \tag{3}$$

Here, β is the vector of <u>mean</u> attribute utility weights in the population, while η_n is the person n specific deviation from the mean. The investigator is free to specify any distribution for the η vector, but in most applications it is assumed to be multivariate normal, MVN(0, Σ). Equivalently, we can write that $\beta_n \sim MVN(\beta, \Sigma)$.

The LC model assumes that consumers belong to one of several classes (also called "segments" or "types"). Classes are latent and the number of classes is not known a priori. The β_n's differ across classes but are identical within classes, *i.e.*,

$$\beta_n = \beta_s \text{ with probability } w_{n,s}$$

$$\sum_s w_{n,s} = 1 \qquad \text{and} \qquad w_{n,s} > 0\ \forall\ s$$

$$\text{for } s = 1, 2, ..., S;\ n = 1,...,N;\ j = 1,...,J;\ t = 1,...,T \tag{4}$$

where S is number of classes and $w_{n,s}$ is the probability of person n to be a member of class s. $w_{n,s}$ may depend on characteristics of person n or maybe assumed identical across consumers, $w_{n,s} = w_s$. In typical applications, the researcher estimates models with different numbers of classes, and the best model is chosen using BIC or AIC.

In both the MIXL and LC models, the choice probabilities *conditional* on β_n still have the closed form logit expression:

$$P(j \mid X_{nt}, \beta_n) = \exp(\beta_n x_{njt}) \Big/ \sum_{i=1}^{J} \exp(\beta_n x_{nit}) \, .$$

With panel data, the probability that we observe a sequence of choices $\{y_{njt}\}_{t=1}^{T}$ from period 1 to T for person n is just the product of the period-by-period logit expressions. Denote $y_{njt} = 1$ if choice j is chosen and 0 otherwise. Because β_n is unobserved, the unconditional choice probabilities have to be evaluated over all possible values of β_n. The choice probabilities of MIXL and LC are given by (5) and (6), respectively.

$$prob(\{y_{njt}\}_{t=1}^{T}) = \int \left[\prod_t \prod_j \left(\frac{e^{\beta_n x_{njt}}}{\sum_i e^{\beta_n x_{nit}}} \right)^{y_{njt}} \right] f(\beta) d\beta \qquad (5)$$

$$prob(\{y_{njt}\}_{t=1}^{T}) = \sum_{s=1}^{S} w_{n,s} \left[\prod_t \prod_j \left(\frac{e^{\beta_s x_{njt}}}{\sum_i e^{\beta_s x_{nit}}} \right)^{y_{njt}} \right] \qquad (6)$$

The only difference between (5) and (6) is whether the possible values of β_n are generated from a continuous distribution or a discrete distribution.

II.B. The Scale Heterogeneity and Generalized Multinomial Logit Models

Louviere and colleagues (1999, 2002) have argued that "scale heterogeneity" is a major source of taste heterogeneity in choice models. They have also argued that the MIXL model is seriously mis-specified because it ignores scale heterogeneity. Their argument led Fiebig, Keane, Louviere and Wasi (2009) to develop the scale heterogeneity (S-MNL) model, as well as the generalized multinomial logit (G-MNL) model that nests MIXL and S-MNL.

To understand what scale heterogeneity means, one must first recognize that the variance of the extreme value idiosyncratic error of MNL model is $\sigma^2\pi^2/6$ where σ, the scale parameter, has been implicitly normalized to one to achieve identification. The simple logit model can be written with the scale of the error made explicit:

$$U_{njt} = \beta x_{njt} + \varepsilon_{njt} / \sigma \qquad n = 1, \ldots, N; \quad j = 1, \ldots, J; \quad t = 1, \ldots, T,$$

The scale heterogeneity model assumes that σ is heterogeneous in the population, and hence denotes its value for person n by σ_n:

$$U_{njt} = \beta x_{njt} + \varepsilon_{njt} / \sigma_n \qquad n = 1,\ldots,N; \quad j = 1,\ldots,J; \; t = 1,\ldots,T \tag{7}$$

The expression in (7) suggests that the source of heterogeneity is the variance of the idiosyncratic error, not the taste parameters, β. Heterogeneity in scale, however, is observationally equivalent to a particular type of heterogeneity in the utility weights. Multiplying (7) through by σ_n we obtain:

$$U_{njt} = (\sigma_n \beta) x_{njt} + \varepsilon_{njt} \qquad n = 1,\ldots,N; \quad j = 1,\ldots,J; \; t = 1,\ldots,T \tag{8}$$

That is, equation (8) can be considered a random coefficient model but with $\beta_n = \sigma_n \beta$, a restriction that the vector of utility weights β is scaled up or down proportionately across consumers by the scaling factor σ_n.

The G-MNL model incorporates two ways to nest the S-MNL and MIXL models. The first (called G-MNL-I) combines (3) with (8):

$$U_{njt} = (\sigma_n \beta + \eta_n) x_{njt} + \varepsilon_{njt} \,. \tag{9}$$

The other one (called G-MNL-II) starts with MIXL and multiplies through by σ_n,

$$U_{njt} = \sigma_n (\beta + \eta_n) x_{njt} + \varepsilon_{njt} \tag{10}$$

Both (9) and (10) incorporate MIXL and S-MNL as special cases. G-MNL adds parameter γ, varying between 0 and 1, to nest G-MNL-I and G-MNL-II as well as the hybrid case. The difference between the two cases is that in G-MNL-I, η_n is independent of the scaling of β while in G-MNL-II, η_n and β are both scaled by σ_n. The utility function of G-MNL model is given by:

$$U_{njt} = [\sigma_n \beta + \gamma \eta_n + (1-\gamma)\sigma_n \eta_n] x_{njt} + \varepsilon_{njt} \tag{11}$$

The following table lists the special cases of G-MNL model:

$\sigma_n = \sigma = 1$	Var(η_n) = 0		$\beta_n = \beta$	MNL
$\sigma_n \neq \sigma$	Var(η_n) = 0		$\beta_n = \sigma_n \beta$	S-MNL
$\sigma_n = \sigma = 1$	Var(η_n) ≠ 0		$\beta_n = \beta + \eta_n$	MIXL
$\sigma_n \neq \sigma$	Var(η_n) ≠ 0	$\gamma = 1$	$\beta_n = \sigma_n \beta + \eta_n$	G-MNL-I
$\sigma_n \neq \sigma$	Var(η_n) ≠ 0	$\gamma = 0$	$\beta_n = \sigma_n (\beta + \eta_n)$	G-MNL-II

To complete the specification of the G-MNL model the distribution of σ_n must be specified. Because it is the "scale" parameter, its distribution should have positive support. We have used the lognormal distribution, $\ln(\sigma_n) \sim N(\overline{\sigma}, \tau^2)$.

Note that the parameters $\bar{\sigma}, \tau$ and β cannot be separately identified. To achieve identification we normalize E[σ_n] to one. Also, in order to constrain γ to lie between 0 and 1, we use a logistic transformation.

Finally, it is important to note that this model differs from the model of Sonnier et al. (2007) or Train and Weeks (2005). Theirs are still the mixed logit model in (3), but with an alternative normalization used to achieve identification. Instead of setting the scale parameter to one as in the standard procedure, they set the coefficient of one attribute (price) to minus one (for further discussion see Fiebig, Keane, Louviere and Wasi (2009)).

II.C. The Mixed-Mixed Multinomial Logit Model

The use of mixture-of-multivariate normals as an alternative flexible distribution is present in the literature. Geweke and Keane (1999, 2001) develop the mixture-of-normals probit model, and Rossi et al (2005) develop the mixture-of-normals logit model. Figure 5.7 in Rossi et al (2005) provides a nice illustration of how flexible the distribution of household posterior means can be in a mixture-of-normals model compared to standard MIXL. Burda et al (2008) specified a subset of coefficients in MIXL model to follow mixture-of-normal distributions while some other still follow a simple MVN distribution. Train (2008) and Bajari et al (2007) also consider specifying mixture-of-normals for MIXL but both studies focus on alternative algorithms rather than the performance of MM-MNL compared to existing models.

The MM-MNL essentially nests the MIXL with LC models, and minimizes the disadvantages of each. Specifying the mixing distribution of MIXL to be mixture-of-MVN is actually equivalent to extending LC models to incorporate unobserved heterogeneity <u>within</u> class. In the MM-MNL model, the utility of person n in period t conditional on choice of alternative j is specified as:

$$U_{njt} = \beta_n X_{njt} + \varepsilon_{njt}$$

where $\beta_n \sim MVN(\beta_s, \Sigma_s)$ with probability $w_{n,s}$

$$\sum_s w_{n,s} = 1 \qquad \text{and} \qquad w_{n,s} > 0 \; \forall \, s$$

$$\text{for s} = 1,2,\ldots,\text{S}; \; n = 1,\ldots,N; \; j = 1,\ldots,J; \; t = 1,\ldots,T \tag{12}$$

As we can see if $w_{n,s} \to 0$ for all classes except one, (12) becomes the mixed logit model in (3). If $\Sigma_s \to 0 \; \forall \, s$, (12) becomes the latent class model in (4). The choice probabilities are given by

$$prob(\{y_{njt}\}_{t=1}^{T}) = \sum_{s=1}^{S} w_{n,s} \left\{ \int \left[\prod_t \prod_j \left(\frac{e^{\beta_{n|s} x_{njt}}}{\sum_k e^{\beta_{n|s} x_{nkt}}} \right)^{y_{njt}} \right] f\left(\beta^s\right) d\beta^s \right\}. \tag{13}$$

where $f(\beta^s)$ refers to $MVN(\beta_s, \Sigma_s)$.

III. Some Notes on the Estimation Procedures

The LC models are estimated by maximum likelihood, and we estimate them using many alternative values for S, the number of classes. It is well known that estimation results for LC models are sensitive to starting values. Thus, for each number of classes, we use 50 different random values and the solution to the model with one fewer class as starting values. For each data set, we kept adding one more class until that model yielded a smaller AIC than the model with one fewer class. We report the results for the LC model that is preferred by BIC.

For MIXL and G-MNL, we consider both the case where the covariance matrix of η, denoted Σ, is a full covariance matrix (correlated errors) and the case where Σ is a diagonal matrix (uncorrelated errors). We again report results from the version of each model that is preferred by BIC.

One key detail about estimation of the S-MNL model is worth noting. For datasets in which choices are labeled (e.g., buy or don't buy), our models include alternative specific constants, or 'ASCs.' Of course, ASCs are not needed in datasets where choices are generic (e.g., pizza A or pizza B). Fiebig, Keane, Louviere and Wasi (2009) found that scaling the ASCs in the S-MNL model leads to non-sensical results. So instead we assume the ASCs are normally distributed random coefficients.

For the MM-MNL model, the choice probabilities in (13) can be simulated as follows. First, <u>conditional</u> on being in class *s*, the simulated probability of observing person *n* choose a sequence of choices is given by:

$$\hat{P}_n \mid s = \frac{1}{D} \sum_{d=1}^{D} \prod_t \prod_j \left(P(j \mid X_{nt}, \eta^{s,d}, s) \right)^{y_{njt}} = \frac{1}{D} \sum_{d=1}^{D} \prod_t \prod_j \left(\frac{\exp(\beta^s + \eta^{s,d}) X_{njt}}{\sum_{k=1}^{J} \exp(\beta^s + \eta^{s,d}) X_{nkt}} \right)^{y_{njt}}$$

where $\eta^{s,d}$ is a K-vector distributed MVN$(0, \Sigma^s)$. The simulation involves drawing $\{\eta^{s,d}\}$ for *d* =1,…,D; and *s*=1,…,S. To obtain the <u>unconditional</u> probability of a person's choice sequence, we take a weighted average of the simulated probabilities for each class, where the weights are the probability of being in that class: $\hat{P}_n = \sum_s w_{n,s}(\hat{P}_n \mid s)$. The simulated log likelihood for the sample is the sum of the simulated likelihood contributions for all individuals: $\ln \hat{L} = \sum_n \ln \hat{P}_n$. If there is only one class, this is exactly the simulated likelihood of the mixed logit model.

Because few personal characteristics are available in these data sets, we specify $w_{n,s} = w_s$ for both LC and MM-MNL models. To impose $\sum_s w_s = 1$, we use the logistic transformation, $w_s = \exp(w_s^*) \Big/ 1 + \sum_{s=1}^{S-1} \exp(w_s^*)$, and set w_s^* for the last class to zero. To avoid cases where w_s^* may run off to infinity as we iterate, we also set upper and lower bounds of 5 and -5, implying the membership probability for each class is at least 0.01. We also impose upper and lower bounds on taste parameters because, when a large fraction of respondents chooses based on one or two attributes, LC and MM-MNL are likely to generate one class to capture that behaviour. The estimates of utility weights on those attributes can then run off to infinity.

The number of parameters of MM-MNL proliferates with the number of classes, especially if one specifies Σ^s to be a full variance-covariance matrix for all classes. We adopt two alternative restrictions here. In the first case, we assume that Σ_s is a diagonal covariance matrix for all *s,* and Σ_s differ across class s. The second case specifies the variance-covariance matrices for all classes to be proportional: $\Sigma_s = k_s \Sigma$ where Σ is a full variance-covariance matrix. Note that the first case nests (i) MIXL with uncorrelated errors and (ii) LC models. The second case nests (i) MIXL with correlated errors and (ii) LC models.

The second case is closely related to the G-MNL model. If one specifies Σ_s for each class to be $\gamma\Sigma + (1-\gamma)k_s\Sigma$, restricts the mean vector to vary proportionally across classes, $\beta_s = k_s \beta$, and lets the number of classes goes to infinity, then one obtains the G-MNL model. G-MNL as a continuous mixture of normal can be written: $\beta_n \mid \sigma_s \sim MVN(\beta_s, \Sigma_s)$ and $\ln(\sigma_s) \sim N(\overline{\sigma}, \tau^2)$ where β_s and Σ_s follow the restrictions above.

IV. Empirical results

The five models of consumer heterogeneity (S-MNL, MIXL, G-MNL, LC and MM-MNL) are evaluated based on ten stated preference choice experiment data sets. Three of the data sets concern medical decision making, specifically, preferences for genetic and cervical cancer test options. Seven of the data sets concern choice of various consumer products, ranging from pizza delivery services to holiday packages, mobile phones and charge cards. Table 1 describes the general characteristics of each dataset (i.e., number of attributes, number of choices, number of choice occasions). Table 2 lists the attributes and their levels.

Tables 3-12 report estimation results for the MNL, S-MNL, MIXL, G-MNL, LC and MM-MNL models on each of the ten datasets. As there are so many models, we only present a subset of the parameter estimates. Also, within each type of model, we have generally estimated several different versions (e.g., different numbers of latent classes, errors correlated or uncorrelated, etc.). For each class of model, we only report results for member of the class that was preferred by the Bayes information criterion (BIC). [2] This is because in Monte Carlo work reported in Fiebig, Keane, Louviere and Wasi (2009) found that BIC was the most reliable criteria for choosing the correct model in this type of data. Still, we report how each model performed on 3 information criteria, Akaike (AIC), Bayes (BIC) and consistent Akaike (CAIC).

IV.A. Estimation Results for the Ten Data Sets

Table 3 presents the result of the first data set. Here, subjects were asked whether they would chose to receive diagnostic tests for Tay Sachs disease, cystic fibrosis, both or neither, giving four alternatives. Covariates include cost of the tests, whether the person's doctor recommends it, risk factors, and alternative specific constants. The sample members are all Ashkenazi Jews, who have a relatively high probability of carrying Tay Sachs.

For this data set, Fiebig, Keane, Louviere and Wasi (2009) found that G-MNL with correlated errors dominated S-MNL and MIXL (and G-MNL with uncorrelated errors) by all three model selection criteria (AIC, BIC and CAIC). The 4th column reports the estimates of this G-MNL model. Note that G-MNL achieves log-likelihood of -2480 using 79 parameters, giving a BIC value of 5601. Estimates of the mean preference weights have expected signs and most are statistically significant. The estimated mean ASCs are not statistically significant different from

[2] Given N people and T choices per person we have that BIC = -2LL + (#parameters) * log(NT).

zero. But their estimated variances are large and significant, suggesting a high degree of heterogeneity in how people value each test option, holding other observed attributes constant.

Note that the G-MNL estimate of the scale parameter τ is 0.45 with a standard error of 0.08, which implies substantial scale heterogeneity in the data. As $\sigma_n = \exp(-\tau^2/2 + \tau\varepsilon_{0n})$, the estimates imply a person at the 90th percentile of the scale parameter would have his/her vector of utility weights scaled up by 57%, while a person at the 10th percentile would have his/her vector of utility weights scaled down by 46%.[3]

The 5th column reports the latent class (LC) model. We estimated LC models with various numbers of classes, and found that a model with 5 classes is preferred by both BIC and AIC. Given the large number of parameters, we only report the attribute coefficient vectors for the 3 largest classes, which account for 76% of the population. Note that the largest class (class 1) places much greater weight on risk factors than do classes 2 and 3. The second largest class (class 2) places a much greater weight on cost than do classes 1 and 3. Class 3 has a very high intercept for the "both" option, implying they are very likely to get both tests regardless of attribute settings. Regarding the two smaller classes not reported, class 4 is characterized by placing a very great weight on price. Class 5 has a configuration of parameters such that they will usually choose to get either both tests or neither, and they have low sensitivity to attributes in making these decisions. The 5-segment LC model achieves a log-likelihood of -2701 using 59 parameters, giving a BIC of 5882. Thus, it is dominated by the G-MNL model according to all three model selection criteria. In particular, G-MNL is superior on BIC by 281 points.

The 6th column presents the MM-MNL model. We estimated several versions of this model, using different numbers of classes, and assuming either independent or correlated normal coefficient vectors. In this case BIC preferred a model with a mixture of two independent normal coefficient vectors. The 6th column is MM-MNL with 2-independent-normals, which achieves a log-likelihood of -2620 using 45 parameters, giving a BIC of 5605.[4] Note that the larger class

[3] The estimate of γ is 0.11, which implies the data is closer to the G-MNL-II model (see equation (10)), where the variance of residual taste heterogeneity increases with scale, than the G-MNL-I model (see equation (9)), where it is invariant to scale.

[4] MM-MNL with 3-independent-normals (not reported) achieves the log-likelihood of -2555 but is beaten by MM-MNL with 2-independent-normal using BIC. MM-MNL with 2 correlated normal coefficient vectors (with the covariance proportionality constraint discussion in the text imposed) achieves a log-likelihood of -2455, but this model has 90 parameters, and gives a BIC of 5640. Thus, it is also beaten by MM-MNL with 2 independent normal coefficient vectors.

(Class 1) places much greater weight on risk factors, while the smaller class (Class 2) places much greater weight on costs.

In summary, BIC prefers G-MNL over all other models, the closest competitor being MM-MNL (5601 vs. 5605) followed by MIXL (5626). There is then a rather wide gap before we get to S-MNL (5777) and another wide gap before we get to LC (5882).

The comparison between G-MNL and MM-MNL is more complex if we also consider AIC and CAIC. This is because G-MNL (with correlated errors) has 79 parameters while MM-MNL with 2 independent normals has only 45.[5] Hence, the advantage of G-MNL is much greater if we look at AIC, which imposes a smaller penalty for extra parameters (5118 vs. 5330). Indeed, MIXL is also preferred to MM-MNL according to AIC. On the other hand, MM-MNL is slightly preferred by CAIC, which imposes a larger penalty for extra parameters (5680 vs. 5650).

The comparison between G-MNL and MIXL is unaffected by the criterion used, since these models have a similar number of parameters (79 vs. 77). Furthermore, G-MNL and MM-MNL remain heavily preferred to LC regardless of the criteria used.

As we will see below, this pattern of the LC model performing poorly relative to the other 4 models that include heterogeneity holds consistently across all 10 data sets. According to BIC, the LC model performs worst in 5 datasets, and next to worst in the other 5. Thus, we would not advise using the LC model for demand prediction. However, as we will also see, the LC model estimates are very useful for gaining an intuitive understanding of the nature of consumer segmentation in each category.

For instance, in Table 3, the LC results indicate that the largest segment of consumers place great emphasis on risk factors, the next largest cares a lot about cost, and the third largest pretty much chooses to get the tests regardless of attribute settings (i.e., they have very large intercepts). Segment 4 (not reported) cares *extremely* much about cost, and segment 5 (not reported) tends to behave fairly randomly, choosing to get either both tests or nether test with little effect of attribute settings. These patterns are basically born out when we look at posterior distributions of attribute weights derived from the better fitting G-MNL and MM-MNL models.

[5] It is notable that this situation is a bit unusual, as G-MNL only ends up with more parameters than MM-MNL because BIC prefers the G-MNL model with a correlated random coefficient vector. In most cases we consider, the G-MNL model with *uncorrelated* heterogeneity is preferred, and as a result G-MNL has fewer parameters than MM-MNL.

It is interesting to note that the two classes identified by the MM-MNL model basically correspond to the two largest classes identified by the LC model (i.e., class 1 cares a lot about risk, class 2 cares a lot about costs). MM-MNL is able to capture the behavior of members of smaller segments by relying on the randomness of its coefficient vectors. In contrast, G-MNL is able to capture the behavior of various segments via the interaction of the random coefficients with the scaling parameter. For example, there will be some cases where the random draws for price coefficients are large, and in addition the random draw for the scale parameter is also large. This generates behavior where consumers care very much about price.

Table 4 reports results for an identical Tay Sachs/Cystic Fibrosis screening test choice experiment, except now the sample is chosen from the general population. In this dataset, according to BIC, the MM-MNL model (with a mixture of two independent normals) is preferred over G-MNL (6420 vs. 6487). The ordering of the other models is the same as before – i.e., MIXL (6535), S-MNL (6591) and LC (6723).

Interestingly, in the general population data the nature of heterogeneity appears to be more complex than in sample of Ashkenazi Jews. The LC model identifies 7 consumer segments, compared to 5 in the previous example, and here the three largest segments capture only 57% of the population, compared to 76% in Table 3. Below we will see that this pattern holds generally. That is, MM-MNL will be preferred by BIC in the 3 data sets where the number of segments in the LC model is 6+, G-MNL will be preferred in 4 data sets where the number of segments is 4 or 5, and S-MNL will be preferred in 3 datasets where the number of segments is 4.

Substantively, the difference in the structure of heterogeneity between the Ashkenazi and general population samples is quite interesting. In the general population, there is a segment that will almost never choose to get the tests, and it is actually the largest segment (22%). It basically replaces segment 3 in the Ashkenazi population (20%), which almost always gets the tests. This pattern would not be surprising unconditionally (i.e., we would expect the Ashkenazis to be more interested in getting the tests as they know they are at higher risk). But the fact this holds even conditional on risk factors is consistent with a view that the experimental subjects are behaving as Bayesians – i.e., updating their own priors on risks with the information given in the experiment. This highlights the fact that even in a choice experiment it is not possible to fully control subjects' perceptions of the attribute levels. This of course has more general implications for social advertising (e.g., how we convey information about risks to the population).

The two extra segments that appear in the general population are #6, who rely heavily on doctor's recommendation (7%), and #7, who care a lot about *both* risk factors and price (7%). As we will see below, this constitutes a general pattern for cases where MM-MNL is preferred to G-MNL. That is, MM-MNL tends to be preferred in cases where there are some small but non-trivial segments of consumers who care a great deal about particular attributes that are not weighed so heavily by members of the larger segments. Of course, this statement is essentially equivalent to our earlier observation that MM-MNL is preferred in cases where the number of segments identified by the LC model is large (i.e., 6 or more). In such cases there will almost inevitably be a few small segments (i.e., segment proportions in the single digits).

Table 5 reports estimates from the mobile phone choice experiment. Here, the choice is simply whether or not to buy a mobile phone with the specified attributes. In this dataset the structure of heterogeneity is fairly simple. The LC model identifies only 4 segments. The largest (32%) is not very sensitive to any particular attribute (i.e., they exhibit fairly "random" behavior). Segment 2 (28%) is sensitive to price but not other attributes. Segment 3 (22%) is very sensitive to price. And segment #4 (not reported, 18%) is modestly sensitive to price. Thus, one could think of the consumers as being segmented into 4 levels of price sensitivity (very sensitive to not at all), with other attributes being fairly unimportant. This apparent lack sensitivity to extra features of cell phones (beyond the basic features all phones have) is consistent with information we were given by industry executives.

Given the simple structure of heterogeneity, it is not surprising that the very parsimonious S-MNL model is actually the preferred model in this dataset, with a BIC of 8121. There is little to choose between MIXL and G-MNL, which have BIC values of 8197 and 8190, respectively. Finally, MM-MNL and LC lag far behind, with BIC values of 8359 and 8426, respectively.

Even though MIXL beats G-MNL for the mobile phone data (albeit by only 7 points), we would argue that the results are still supportive of the use of the G-MNL specification over MIXL. It is important to note that S-MNL is a special case of G-MNL, and a researcher who started with G-MNL would have tested down to the more parsimonious S-MNL specification. As emphasized by Fiebig, Keane, Louviere and Wasi (2009), either G-MNL or its S-MNL special case is preferred over MIXL in all 10 datasets we examine.

We have now discussed in detail results from three data sets that illustrate contexts where G-MNL, MM-MNL or S-MNL is the preferred model. We will now discuss the other datasets in less detail. Table 6 reports results from Pizza delivery service choice experiment A. The two services are only labeled by A and B, and hence the model does not contain ASCs. This data set is a bit exceptional, in that the LC model identifies only 4 segments, yet the G-MNL model is preferred. It appears that S-MNL does relatively poorly because the four segments are very different (i.e., type 1 cares greatly about freshness, type 2 exhibits random behavior, type 3 cares greatly about the pizza being hot and type 4 (not reported) cares a lot about price).

Table 7 reports results from the Holiday A dataset. Participants choose between two Holiday packages labeled A and B, so there is no ASC. Here, the LC model identifies 5 segments.[6] Given this intermediate level of heterogeneity we are not surprised that G-MNL is preferred, as per our earlier discussion. Table 8 reports the results from the Papsmear test choice experiment. Here the LC model identifies 5 segments, and G-MNL is preferred by BIC as we would expect.

Substantively, it is interesting that type 1's chose to get the test as needed (i.e., when due and doctor recommended), type 2's almost always chose to get the test, and type 3's are sensitive to a range of factors (test due, doctor characteristics, and doctor recommendation). Types 4 and 5 (not reported) make up 24% of the population and they are, respectively, either very averse or extremely averse to male doctors. No type cares much about price.

Table 9 reports the results from Pizza deliver service choice experiment B. This differs from experiment A in that the number of attributes of the pizza (and the delivery service) is increased from 8 to 16. Not surprisingly, this increases the number of classes identified by the LC model from 4 to 6. Hence, it is not surprising that that MM-MNL is the preferred model in this dataset according to BIC (11527 vs. 11693 for G-MNL). There is quite a large gap before we come to MIXL (12081), followed by LC (12118) and then S-MNL (13372).

The structure of heterogeneity in this data set is quite interesting. The first segment identified by the LC model makes up 51% of the population, and is it shows only very modest sensitivity to the attributes (i.e., close to random choice behavior). Members of the 2nd segment (14%) care greatly about price, the 3rd (12%) cares greatly about quality (fresh ingredients), the 4th (10%) cares greatly about crust type, the 5th (9%) cares greatly about hot delivery, and the 6th

[6] The first segment cares a lot about price, the second about quality of accommodation and the third cares only modestly about price and accommodation. The 4th and 5th either like or do not like overseas destinations, respectively.

(4%) wants a vegetarian option.[7] Thus, we have several small segments that care greatly about different attributes (as oppose to just few large segments). Contrast this to Pizza A, where, e.g., there was one large segment (36%) that cared greatly about price. The more complex heterogeneity structure in Pizza B falls neatly into our characterization of when the MM-MNL model would be preferred. Also, this is the first dataset we have examined where the structure of heterogeneity is complex enough that MM-MNL supports a 3-class model.

Table 10 reports results from the Holiday B dataset. This differs from experiment A in that the number of attributes of the holiday packages is increased from 8 to 16. This increases the number of classes identified by the LC model from 5 to 9. This is more segments than for any other dataset.[8] Hence, it is not surprising that that MM-MNL is preferred over G-MNL according to BIC (23002 vs. 23291). Next comes MIXL (23519), followed by LC (23981) and then S-MNL (26224). This is only other dataset (besides Pizza B) where a MM-MNL model with a mixture of 3 normals is preferred over models with only two.

Table 11 reports results from an experiment where a bank offers a credit card and a debit card (which, along with "neither", giving 3 alternatives). The structure of heterogeneity in this data set is quite simple. According to the LC model, there are only 4 segments. The largest segment (48%) basically doesn't want either card (large negative intercepts). Segment #2 (27%) prefers a debit card while segment #3 (19%) prefers a credit card. Segment #4 (not reported, 7%) is fairly indifferent between the two. All types dislike interest and fees. Given this simple structure, it is not surprising that the S-MNL model is preferred by BIC (5707). It is followed by MIXL and G-MNL, whose BIC values are fairly close (5883 and 5898, respectively). Then comes MM-MNL (5988) and, finally, LC (6039).

Table 12 reports results from the 2nd charge card experiment, which is identical except that a transaction card option is added. Again there are only 4 segments according to the LC model. The largest segment (44%) basically doesn't want either card (large negative intercepts).

[7] The crust type and vegetarian segments were not identified by LC in the Pizza A dataset (the latter because it was not one of the listed attributes).

[8] Segment 1 (27%) cares modestly about price, but has small attribute coefficients in general (i.e., close to "random" choice behaviour. Segment 2 (15%) cares moderately about quality accommodation, price, meals and length of stay. Segment #3 (14%) cares intensely about quality accommodation. Segment #4 (14%) wants overseas travel. Segment #5 cares about price and meal inclusion. Segment #6 (7%) cares mostly about having a beach or swimming pool. Segment #7 (6%) intensely dislikes overseas travel. Segment #8 (5%) likes personal tours. And segment #9 (4%) cares about price, quality accommodation and length of stay (i.e., "value"). Sadly, not enough respondents value cultural activities to form a segment.

Segment #2 (21%) is indifferent among type of card, but cares about interest, fees and access. Segment #3 (19%) prefers a transaction card and cares about interest rates. Segment #4 (not reported, 16%) is indifferent among type of card, but cares about interest, fees and access. All types dislike interest. Given this simple structure, it is not surprising that the S-MNL model is again preferred by BIC (7007). It is followed by MIXL and G-MNL, whose BIC values are fairly close (7166 and 7182, respectively). Then comes MM-MNL (7258) and, finally, LC (7391).

IV.B. Comparing Model Fit across Data Sets.

Table 13 summarizes the fit of the 6 models (MNL, S-MNL, G-MNL, LC and MM-MNL) across the ten datasets. For MIXL, G-MNL, LC and MM-MNL we estimated various versions of each model, which differ in number of segments and whether random coefficients are allowed to be correlated. Within each type of model, we report results for the version preferred by BIC. Nevertheless, we will compare these models based on AIC, BIC and CAIC.

According to AIC, MM-MNL is preferred in 7 out of 10 datasets. G-MNL is preferred only in the two Tay Sachs datasets. In the Papsmear dataset there is a virtual tie between MM-MNL and G-MNL. However, in Monte Carlo work reported in Fiebig, Keane, Louviere and Wasi (2009), we found that BIC was much more accurate than AIC in selecting the correct model for data of this type. AIC tends to choose over-parameterized models (as it imposes a smaller penalty for additional parameters).

According to BIC, G-MNL is the preferred model in 4 datasets, MM-MNL is preferred in 3 and S-MNL is preferred in 3. Another way to look at this is that models that allow for scale heterogeneity – that is, G-MNL or its S-MNL special case – are preferred in 7 out of 10 datasets. Furthermore, G-MNL is preferred to MM-MNL in 7 out of 10 data sets (the 4 where it is preferred overall plus the 3 where S-MNL is preferred overall). Thus, the results clearly support a conclusion that it is important to consider models with scale heterogeneity.

It is also notable that in 4 of the 7 datasets where MM-MNL loses to a scale heterogeneity model (mobile phones, papsmear test, card cards A and B) it losses rather soundly (i.e., by 3% to 5% on BIC). But in the three datasets where MM-MNL wins (Tay Sachs general, Pizza B, Holiday B), the G-MNL model is a closer competitor, losing by only 1% to 1.4% on BIC. Thus, G-MNL appears to be rather robust to alternative data structures in the following sense: in data where the structure of heterogeneity is "very complex" MM-MNL is preferred, but G-MNL proves to be a close competitor. In data where heterogeneity is "moderately complex" these roles are reversed. But in data sets where the structure of heterogeneity is "simple" the S-MNL special case of G-MNL is preferred by a wide margin over MM-MNL. Thus, the use of G-MNL (including the possibility of testing down to the S-MNL special case) will produce reasonably good results in all three cases.

Recall that we have defined heterogeneity as "very complex," "moderately complex" or "simple" in a specific way in this paper. We have based it on the number of segments of consumers identified by the LC model. In the 3 "very complex" datasets where MM-MNL is preferred (Tay Sachs general, Pizza B, Holiday B) there are 6+ segments. In the 4 "moderately complex" datasets where G-MNL wins (Tay Sachs Jewish, Pizza A, Holiday A, Papsmear) there are 4 to 5 segments. And in the 3 "simple" datasets where S-MNL wins (mobile phones, charge cards A and B) there are only 4 segments.

More subtly, "simplicity" is also defined by how the segments differ. For instance, in mobile phones the four segments differ only in the sense of having different levels of price sensitivity. This is quite easy for the S-MNL model to capture. Pizza A also has only 4 segments, but their behavior differs quite substantially. Thus, the more general G-MNL model is needed.

The results for CAIC are almost identical to those for BIC. In 9 datasets the preferred model is unchanged. The only change is Tay Sachs Jewish data set, where the G-MNL model beats MM-MNL by a small margin. Using CAIC this result is reversed. The reason for the reversal is that CAIC imposes a larger penalty for additional parameters than BIC, and in the Tay Sachs Jewish dataset the G-MNL model has more parameters than MM-MNL. (This is actually a bit unusual as in most cases G-MNL has fewer parameters).

Strikingly, the MIXL and LC models, which are arguably the most popular models of heterogeneity in use in marketing today, are never preferred by any model selection criteria. The performance of LC is particularly weak: MIXL is preferred over LC by BIC in all 10 datasets. Nevertheless, we found that the LC model is very useful for gaining an intuitive understanding of the structure of heterogeneity in each dataset, and for understanding why G-MNL, MM-MNL or S-MNL is preferred in each case. Of course, ease of interpretation has always been the strength of the LC model. Thus, we would advocate estimating the LC model in conjunction with the G-MNL and MM-MNL models. The better fitting of the latter can be used for actual demand prediction, while the LC model can be used for intuition.

IV.C. Understanding the Behavioral differences between the Models

Simply knowing that one model fits better than another is not in itself particularly interesting. We would all also like to understand the behavioral differences between the models. What aspect(s) of behavior can one model capture better than another? In this section we examine the behavioral differences between our 6 models (MNL, S-MNL, LC, MIXL, G-MNL and MM-MNL).

First, we examine the degree of flexibility of each model in fitting distributions of taste heterogeneity. To do this, we use the estimated model to calculate person-specific parameters. Adopting what Allenby and Rossi (1998) call an "approximate Bayesian" approach, the estimated heterogeneity distribution is taken as the prior, and the posterior means of each person-specific vector of preference weights are then calculated conditional on his/her observed choices (see Train (2003), chapter 11.)

For the Pizza B dataset, Figures 1-2 plot the posterior distributions of the person level coefficients on "price" and "ingredient freshness," respectively. Consider first figure 1, which plots the posterior distribution of the price coefficient. Of course, the MNL posterior puts all mass on a point, as there is no heterogeneity in this model. Notice next that the MIXL posterior has a distinctly normal shape. As Allenby and Rossi (1998) pointed out, the normal prior in the MIXL model has a strong tendency to draw in outliers, so this model has a hard time capturing "extreme" consumers – e.g., a mass of consumers who place great weight on price. Similarly, the S-MNL posterior departs only slightly from its log-normal prior.

And, as Allenby and Rossi (1998) also pointed out, the LC posterior is less dispersed than that for MIXL, as it is constrained to lie within the convex hull of β_s. Thus, as Elrod and Keane (1995) found, LC tends to understate the degree of heterogeneity in the data.

In contrast to MIXL, the posterior distributions of G-MNL and MM-MNL depart quite substantially from normality, with more mass in the tail. Notice that both models generate a mass of consumers in the left tail who care intensely about price (e.g., G-MNL generates a local mode at a price coefficient of -4.1). Both models also generate excess kurtosis relative the normal (i.e., a mass of consumers with price coefficients near zero).

Notice that the G-MNL and MM-MNL posteriors for price look fairly similar (especially relative to the other models). This is not really surprising, because, as noted earlier, these two models are actually closely related. In the case of G-MNL, the posterior is a continuous mixture of scaled normals, while for MM-MNL the posterior is a discrete mixture of normals (see Fiebig, Keane, Louviere and Wasi (2009) for further discussion). Both posteriors are quite flexible, which lets the data have more impact on the shape of the posterior.

Figure 2 reports the posterior distribution for the ingredient quality coefficient. The story here is very similar to Figure 1. Both G-MNL and MM-MNL are able to capture that there is a segment of consumers who put great positive weight on fresh ingredients. MIXL is again unable to capture this as these outliers are pulled in by the normal prior. LC does capture that there is a segment that cares a lot about freshness, but it puts almost all the mass of the heterogeneity distribution on a few points, understating the true extent of heterogeneity in the data.

Figure 3 reports an experiment where we look at the predictions of the various models for how changes in product attributes would affect consumer demand. Specifically, we start from a baseline where pizza delivery services A and B both offer identical attributes. Of course, in that case people are indifferent between the two services, and all models predict that 100% of consumers choose service A exactly 50% of the time. In the experiment, service A improves ingredient quality (i.e., fresh ingredients) while also increasing price by $4.

G-MNL predicts that, after the policy change, 16% of consumers still have roughly a 50% chance of choosing A. Strikingly, 8% of consumers have a near 100% chance of choosing A (these are the types who put great weight on fresh ingredients) while 5% have a near 0% chance of choosing A (these are the types who care primarily about price).[9] The predictions for MM-MNL are quite similar. It predicts that 16% of consumers remain near 50%, while 9% have a near 100% chance of choosing A and 7% have a near 0% chance of choosing A.

As we would expect based on the coefficient distributions in Figures 1-2, MIXL predicts fewer people stay indifferent, but also that fewer people have extreme reactions. Specifically, MIXL predicts that only 8% of consumers stay at roughly a 50% chance of choosing A, while almost no consumers have their choice probabilities move all the way to 100% or 0%.

In the actual Pizza B data, 24/328 = 7.3% of subjects choose the fresh ingredient Pizza on all choice occasions regardless of other attribute settings, while 27/328 = 8.2% always choose the less expensive Pizza. *The Figure 3 results show that G-MNL and MM-MNL can both generate such extreme (lexicographic) behavior, while MIXL cannot.*

[9] In these calculations, we define "roughly 50%" as between .475 and .525, while we define "essentially 100%" as greater than .95 and "essentially 0%" as less than .05.

In Table 13 we see that G-MNL fits better than MIXL in 7 out of 10 datasets, often by wide margins. MIXL is only marginally preferred to G-MNL in the 3 data sets with "simple" heterogeneity structures (i.e., mobile phones and charge cards A and B). In order to understand why G-MNL generally fits better than MIXL, Fiebig, Keane, Louviere and Wasi (2009) ordered people by their likelihood contribution in the MIXL model (from best to worst). They then plotted the G-MNL likelihood contributions against the MIXL contributions and found two patterns of improvement:

First, G-MNL is better able to capture the behavior of consumers who base choice largely on a single attribute (i.e., people who have nearly lexicographic preferences). Examples are the people who put great weight on price or freshness in choosing pizza. As we see in Figures 1-2, G-MNL can generate very large coefficients on price or freshness, while MIXL cannot.

Second, G-MNL is better ability to capture the behavior of "random" consumers, whose choices are only slightly influenced by observed product attributes. This is again because G-MNL is better able to generate excess kurtosis (i.e., more mass near zero than in the normal).

We performed the same type of analysis to compare MM-MNL to MIXL, and we found identical results. Thus, MM-MNL has the same advantages over MIXL as does G-MNL. This is quite apparent from the policy experiment results in Figure 3.

We performed the same type of analysis to compare MM-MNL vs. G-MNL. Recall from Table 13 that G-MNL is preferred to MM-MNL in 7 out of 10 datasets. Thus, we looked at both types of datasets to see if we could discern, via this approach, the types of people that each model tended to fit better. Unfortunately, we could not discern any clear patterns. In particular, MM-MNL and G-MNL are both able to capture the behavior of consumers who exhibit either nearly lexicographic or nearly random behavior, so no differences emerged there. Indeed, for the most part, MM-MNL and G-MNL predict similar patterns in choice behavior, and thus provide a similar fit to the data. Which model is preferred is therefore often determined not by differences in ability to fit various data patterns, but rather by issues of parsimony. BIC often tends to choose G-MNL because in most cases it provides a similar fit to MM-MNL but with fewer parameters.

Thus, we sought to find more subtle behavioral patterns that might distinguish the two models. A clue is provided by the fact that we already found that MM-MNL is preferred only in the datasets where the structure of heterogeneity is complex. For example, in Pizza B we found that there are five "major" attributes that are important enough for the LC model to devote a segment to consumers who place great weight on them: price, fresh ingredients, crust, hot delivery and vegetarian. But an examination of the posterior distribution of attribute weights in this data set indicated that there are some additional attributes that small but non-trivial segments of consumers also care a lot about. We will call these "minor attributes."

For instance, the top panel of Figure 4 shows the posterior distribution of the coefficient on baking method (i.e., woodfire) for the G-MNL and MM-MNL models. In contrast to the posteriors for the "major attributes," which are very similar between the two models (see Figures 1 and 2), the posteriors for this "minor attribute" are quite different. For G-MNL, most of the mass is near zero, and only 9% is in the .30 to .50 range. In contrast, for MM-MNL, 29% of the mass is in the .30-.50 range. Thus, the MM-MNL model implies that a non-negligible fraction of the population has a modest preference for woodfire cooking.

This is illustrated in the bottom panel of Figure 4. Here we consider an experiment where, starting from identical offerings, firm A offers a woodfire cooked pizza. G-MNL predicts that about 35% of consumers remain essentially indifferent between A and B. An additional 28% have their probability of choosing A increase to only about 55%. But MM-MNL predicts a bigger effect. For about 30% of consumers, the probability of choosing A jumps to about 65%.

Thus, while G-MNL and MM-MNL give very similar predictions for what happens when firms change "major" attributes, they do predict different responses when firms change "minor" attributes. Based on this, we decided to classify respondents into types based on how they respond to both major and minor attributes. Details of how this classification is done are provided in the Appendix, but here we just give an overview.

Consider again the Pizza B dataset. Some consumers have an extremely strong preference weight on only one attribute. These consumers appear in the top panel of Table 14. Within this group, consumers can be further divided into whether that attribute is one of the "major" attributes (i.e., price, quality, crust, hot, vegetarian) or whether it is a "minor" attribute. These people are reported separately in the first two rows of Table 14. Furthermore, these groups of consumers can be further divided into those who (i) have negligible preference weights for all other attributes, (ii) have modest preference weights for one or a few other attributes, or (iii) have modest preference weights for several other attributes. These sub-groups are reported in the 3 columns of Table 14. For example, in the top row of Table 14, we see there are 39 consumers who have a large preference weight on one major attribute, and do not care about other attributes.

Similarly, there are consumers who have an extremely strong preference weight on two attributes. This group can be further divided based on whether those attributes are (i) both "major," (ii) one "major" and one "minor, or (iii) both minor. These groups are reported in rows 3 to 5 of Table 14. As above, we can split each of these groups up into subgroups based on whether they have modest preference weights on any other attributes.

Next, we have consumers with large preference weights on three or more attributes. These are reported in rows 6 to 8 of Table 14. They are further divided into sub-groups as above.

Finally, we have people who do not exhibit extreme behavior. These are people who have modest preference weights on a few attributes. These people are reported in rows 9 to 10 of table 14, where they are differentiated by how many of those attributes are "major." Additionally, these people may (or may not) have small but not completely negligible weights on a few other attributes, as indicated in the three columns of the table.

For the Pizza B dataset, Table 14 lists the number of people in each of the above groups. It also lists the BIC difference between the MM-MNL and the G-MNL model for each group. We have highlighted in yellow the cases where MM-MNL has a large advantage, and in blue the cases where G-MNL has a large advantage. Recall that, for the dataset as a whole, the MM-MNL model has a BIC advantage of 11527 vs. 11693, which is 166 points or 1.4%. Rather strikingly, as we see in row (2) column (1), MM-MNL achieves an advantage of 196 points on just 17 consumers who exhibit an "extreme" preference for one of the "minor" attributes (and have negligible preferences for all other attributes). Furthermore, in row (5) column (1), we see that MM-MNL achieves a substantial BIC advantage of 52 points for just 6 consumers who exhibit a strong preference for two "minor" attributes (and have negligible preferences for all other attributes). And in row (5) column (2), we see that MM-MNL achieves a substantial BIC advantage of 34 points for just 3 consumers who exhibit a strong preference for two "minor" attributes (and have modest preference weights on a few other attributes).

There are also groups where G-MNL is favored over MM-MNL. In particular, in rows (9) and (10), we see that G-MNL has a BIC advantage in fitting the behavior of "non extreme" consumers who show modest preference weights on multiple attributes.

We have performed this same analysis on other datasets and come to the same general conclusion. G-MNL performs better than MM-MNL for consumers who exhibit moderate preference weights on multiple attributes, as well as for consumers who exhibit fairly random choice behavior (i.e., their choices are not strongly influenced by any attribute settings). MM-MNL is better at fitting the behavior of consumers who have a very strong preference for one or more minor attributes (i.e., attributes that are not weighted heavily by very many consumers).

This analysis is consistent with our discussion in Section IV.A. There we found that MM-MNL is preferred over G-MNL in datasets that exhibit very complex patterns of heterogeneity, in the sense that there are a large number of consumer segments. It is precisely in these types of datasets that there tend to be relatively small subsets of consumers who have strong preferences for "minor" attributes that the majority of consumers are relatively uninterested in.

V. Conclusion

In a recent paper, Fiebig, Keane, Louviere and Wasi (2009) evaluated the performance of a new choice model called the "generalized multinomial logit" or G-MNL model. The G-MNL model generalizes the popular MIXL model by allowing for heterogeneity in the scale coefficient (in addition to normally distributed random coefficients). G-MNL also has an important special case – the scale heterogeneity or "S-MNL" model – in which only scale heterogeneity is present. Using ten empirical datasets, Fiebig, Keane, Louviere and Wasi (2009) found that either G-MNL or its S-MNL special case is always preferred to MIXL according to BIC. They showed that the reason for the superior performance of G-MNL is its more flexible specification of the heterogeneity distribution, which allows it to accommodate highly non-normal posterior distributions for individual level coefficients.

Note that, in G-MNL, the normal coefficient vector is multiplied by a continuously distributed scale coefficient. Thus, G-MNL can be interpreted as a model where the coefficient vector is assumed to be distributed as a continuous mixture of scaled normals. There is a rapidly growing literature in statistics and econometrics that uses discrete mixtures-of-normals as a flexible modeling device. A key reference is Ferguson (1973), who used this approach for density estimation. It has been extended to probit by Geweke and Keane (1999, 2001) and to MIXL by Rossi et al (2005) and Burda et al (2008).[10] The appeal of this approach is that the discrete mixture-of-normals can approximate any heterogeneity distribution arbitrarily well.

[10] In addition Geweke and Keane (2007) introduced the "smoothly mixing regression" (SMR) model, in which the class probabilities in a mixture-of-normals model are determined by a multinomial probit. SMR is closely related to what are known as "mixture of experts" models in statistics (see Jiang and Tanner (1999), Villani, M., R. Kohn and P. Giordani (2007))

Given this important property of the mixture-of-normals, it would not be surprising if a mixture-of-normals generalization of mixed logit were to outperform G-MNL, by providing a yet more flexible specification of the heterogeneity distribution. Indeed that was our prior when we began this study. Here, using the same 10 datasets as in Fiebig, Keane, Louviere and Wasi (2009), we have compared the performance of G-MNL to that of the mixture-of-normals generalization of MIXL. We refer to the latter as the "mixed-mixed logit or "MM-MNL" model.

We found the results of the study somewhat surprising. Based on the BIC criterion, G-MNL outperformed MM-MNL in 4 of the 10 data sets. Even more surprising, the S-MNL special case of G-MNL, which only allows for scale heterogeneity, was preferred in 3 out of 10. MM-MNL was only the preferred model in 3 datasets. Viewed another way, G-MNL or its S-MNL special case are preferred in 7 out of 10 datasets. We also noted that when G-MNL loses to MM-MNL it is always by a rather small margin.

These results suggest that the G-MNL model is in fact quite competitive with the MM-MNL approach. It also reaffirms the conclusion of Fiebig, Keane, Louviere and Wasi (2009) that scale heterogeneity may account for much of the heterogeneity in consumer choice behavior, and that it is important for researchers to consider models that accommodate scale heterogeneity.

We also carefully investigated why MM-MNL fits better than G-MNL in some cases and not in others. That is, what behavioral patterns does each model have an advantage in fitting? We found that the MM-MNL model only outperforms G-MNL in datasets with very complex patterns of heterogeneity, by which we mean that there are several attributes that a non-trivial fraction of consumers treat as very important when making decisions. In that case one can divide these attributes into "major" attributes (i.e., ones that large segments of consumers treat as extremely important) and "minor" attributes (i.e., ones that small but non-trivial segments of consumers treat as important). As an example, for pizza, we found that price and ingredient quality are major attributes, while woodfire cooking and gourmet are minor attributes. The MM-MNL model provides a clearly better fit for the small groups of consumers who place a great deal of weight on "minor" attributes. Conversely, the G-MNL model provides a better fit to consumers who exhibit "non-extreme" behavior, meaning that they don't place great weight on just one or two attributes (i.e., they may put modest weight on several attributes).

Aside from these fairly small differences, G-MNL and MM-MNL predict very similar behavioral patterns for most consumers. For instance, in most cases the two models make very similar predictions for how changes in attribute levels affect consumer demand.

We also included MIXL and latent class (LC) models among the set of models we compared. Neither of these models was preferred in any of the 10 data sets. The comparison with LC is particularly interesting, as this is also a method that is intended to relax the normality assumption often invoked for heterogeneity distributions. However, the LC model performed quite poorly in our comparisons. Indeed, it ranked last in 5 datasets and next to last in the other 5, and never came within 160 points of the preferred model on BIC.

Despite this, we found that the LC model results were very useful in order to gain an intuitive understanding of the patterns of heterogeneity in the datasets. For example, we found that MM-MNL is the preferred model when LC identified 6+ segments, that G-MNL is preferred when LC identified 5 segments, and that S-MNL is preferred when LC identified 4 segments (except in one case where G-MNL was still preferred). It is precisely in those cases where LC identifies a large number of segments that there tend to be some small segments made up of consumers who value some “minor” attribute very highly. These are the cases where MM-MNL outperforms G-MNL on the BIC criterion. Thus, we would advocate estimating the LC model as an aid to understanding the nature of heterogeneity in a market, while using the better fitting of the G-MNL, S-MNL or MM-MNL models for actual demand prediction.

Appendix: Classification of Consumers into Types

Let's consider the binary choice case where both options have binary attributes (say, dummy coded with 1 or 0). Define A_{njk} as a measure of strength of preference of person n for attribute k of option j, $A_{njk} = \sum_{t=1}^{T} y_{njt} I(x_{jkt} = 1, x_{ikt} = 0) \Big/ \sum_{t=1}^{T} I(x_{jkt} = 1, x_{ikt} = 0)$ for j, i = 1,2. The denominator is the sum of number of choice occasions where the attributes of option j takes the value 1 and that of option i takes the value 0. The numerator is the number of times option j is chosen out of those choice occasions. When choices are unlabelled, which is the case of four pizza and holiday data sets, we can further sum these measures across two choices, weighted by their denominator: $A_{nk} = w_1 A_{n1k} + w_2 A_{n2k}$. For example, one pizza attribute is "steaming hot" vs. "warm"(say, hot = 1, warm = 0). A_{nk} equaling one implies that consumer n extremely likes "hot" pizza. A_{nk} equaling zero implies the opposite – that they extremely like "warm". A_{nk} being around .5 means consumer n is quite indifferent for this attribute. We will use "extremely prefer" to refer to the case where $A_{nk} \leq .2$ or $A_{nk} \geq .8$; "like" refer to the case where $.2 < A_{nk} \leq .4$ or $.6 \leq A_{nk} < .8$; and "indifferent" refer to the case where $.4 < A_{nk} < .6$.

Consumer n's observed choice behavior is then described by the vector $\{A_{n1}, A_{n2},\ldots, A_{nK}\}$, where K is the number of attributes. Many possibilities arise. For example, consider the Pizza B data set with 16 attributes. If a consumer "extremely prefers" fresh ingredients (the 3rd attribute) but is totally indifferent with other attributes, his/her vector would be {.5, .5, 1, .5, .5, .5, .5, .5, .5, .5, .5, .5, .5, .5, .5, .5}. If a consumer "extremely prefers" fresh ingredients (the 3rd attribute) and also "likes" gourmet (1st), low price (2nd) and steaming hot (7th), but is indifferent to other attributes, his/her vector might be {.75, .3, 1, .5, .5, .5, .7, .5, .5, .5, .5, .5, .5, .5, .5, .5}. The third possibility is similar to the second but the consumer "extremely prefers" fresh ingredients and "likes" many attributes. Then, for example, we might have: {.75, .3, 1, .5, .5, .25, .7, .5, .3, .5, .65, .7, .75, .5, .5, .5}.[11]

[11] The "extreme" preference is defined in a less extreme sense here than in the Fiebig, Keane, Louviere and Wasi (2009) paper. They only counted only when A_{nk} equals one or zero, and did not look at consumer's preferences on other attributes.

References

Allenby and P. Rossi (1998), “Marketing Models of Consumer Heterogeneity,” *Journal of Econometrics*, 89(1-2), p. 57-78.

Bajari, P., J.T. Fox and S. Ryan (2007), “Linear Regression Estimation of Discrete Choice Models with Nonparametric Distributions of Random Coefficients," *American Economic Review,* 97, 459-463.

Ben-Akiva, M., D. McFadden, et al. (1997), “Modelling Methods for Discrete Choice Analysis,” *Marketing Letters*, 8:3, 273-286.

Burda, M., Harding, M. and Hausman, J. (2008), “A Bayesian mixed logit-probit model f or multinomial choice”, *Journal of Econometrics*, 147:2, 232-246.

Elrod, Terry and Michael P. Keane (1995), “A Factor Analytic Probit Model for Representing the Market Structure in Panel Data,” *Journal of Marketing Research*, 32, 1-16.

Ferguson, T.S. (1973), “A Bayesian analysis of some nonparametric problems,” *The Annals of Statistics* 1, 209-230.

Fiebig, D.G and J.P. Hall (2005), “Discrete choice experiments in the analysis of health policy,” *Productivity Commission Conference, November 2004: Quantitative Tools for Microeconomic Policy Analysis,* 119-136.

Fiebig, D.G, M.P. Keane, J.J. Louviere and N. Wasi (2009) “The Generalized Multinomial Logit Model: Accounting for scale and coefficient heterogeneity,” forthcoming *Marketing Science*

Geweke, J. and M. Keane (1999), “Mixture of Normals Probit Models,” in *Analysis of Panels and Limited Dependent Variable Models*, Hsiao, Lahiri, Lee and Pesaran (eds.), Cambridge University Press, 49-78.

Geweke, J. and M. Keane (2001), “Computationally Intensive Methods for Integration in Econometrics,” in *Handbook of Econometrics: Vol. 5*, J.J. Heckman and E.E. Leamer (eds.), Elsevier Science B.V., 3463-3568.

Geweke, J. and M. Keane (2007), “Smoothly Mixing Regressions,” *Journal of Econometrics*, 138:1 (May), p. 291-311.

Greene, W.H. and D.A. Hensher (2003), “A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit,” *Transportation Research Part B*, 37, 681-698.

Guadagni, P. and J. Little (1983), “A Logit Model of Brand Choice Calibrated on Scanner Data,” *Marketing Science*, 2, 203:238.

Hall, J.P., D.G. Fiebig, M. 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, 520-537.

Hole, A.R. (2008), "Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment," *Journal of Health Economics,* 27:4, 1078-1094.

Jiang, W. and M. Tanner (1999), "Hierachical mixtures-of-experts for exponential family regression models: approximation and maximum likelihood estimation," *Annals of Statistics*, 27, 987-1011.

Kamakura, W. and G. Russell (1989), "A Probabilistic Choice Model for market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 25, 379-390.

Louviere, Jordan J., Robert J. Meyer, David S. Bunch, Richard Carson, Benedict Dellaert, W. Michael Hanemann, David Hensher and Julie Irwin (1999), "Combining Sources of Preference Data for Modelling Complex Decision Processes," *Marketing Letters*, 10:3, 205-217.

Louviere, J.J., R.T. Carson, A. Ainslie, T. A. Cameron, J. R. DeShazo, D. Hensher, R. Kohn, T. Marley and D.J. Street (2002), "Dissecting the random component of utility," *Marketing Letters*, 13, 177-193.

Louviere, J.J., T. Islam, N. Wasi, D.J. Street and L. Burgess (2008), "Designing discrete choice experiments: Do optimal designs come at a price?," *Journal of Consumer Research,* 35, 360-375.

McFadden, D. (1974), Conditional Logit Analysis of Qualitative Choice Behavior, in *Frontiers in Econometrics*, in P. Zarembka (ed.), New York: Academic Press, 105-42.

McFadden, D. (1978), Modeling the choice of residential location', in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds., *Spatial Interaction Theory and Planning Models*, North-Holland, Amsterdam, 75–96.

McFadden, D. and K. Train. (2000), "Mixed MNL models for discrete response," *Journal of Applied Econometrics,* 15, 447-470.

Rossi, P., Allenby, G. and R. McCulloch (2005), *Bayesian Statistics and Marketing*, John Wiley and Sons, Hoboken, N.J..

Sonnier, G., A. Ainslie, and Thomas Otter (2007), "Heterogeneity Distributions of Willingness-to-pay in Choice Models," *Quantitative Marketing and Economics*, 5, 313-331.

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press.

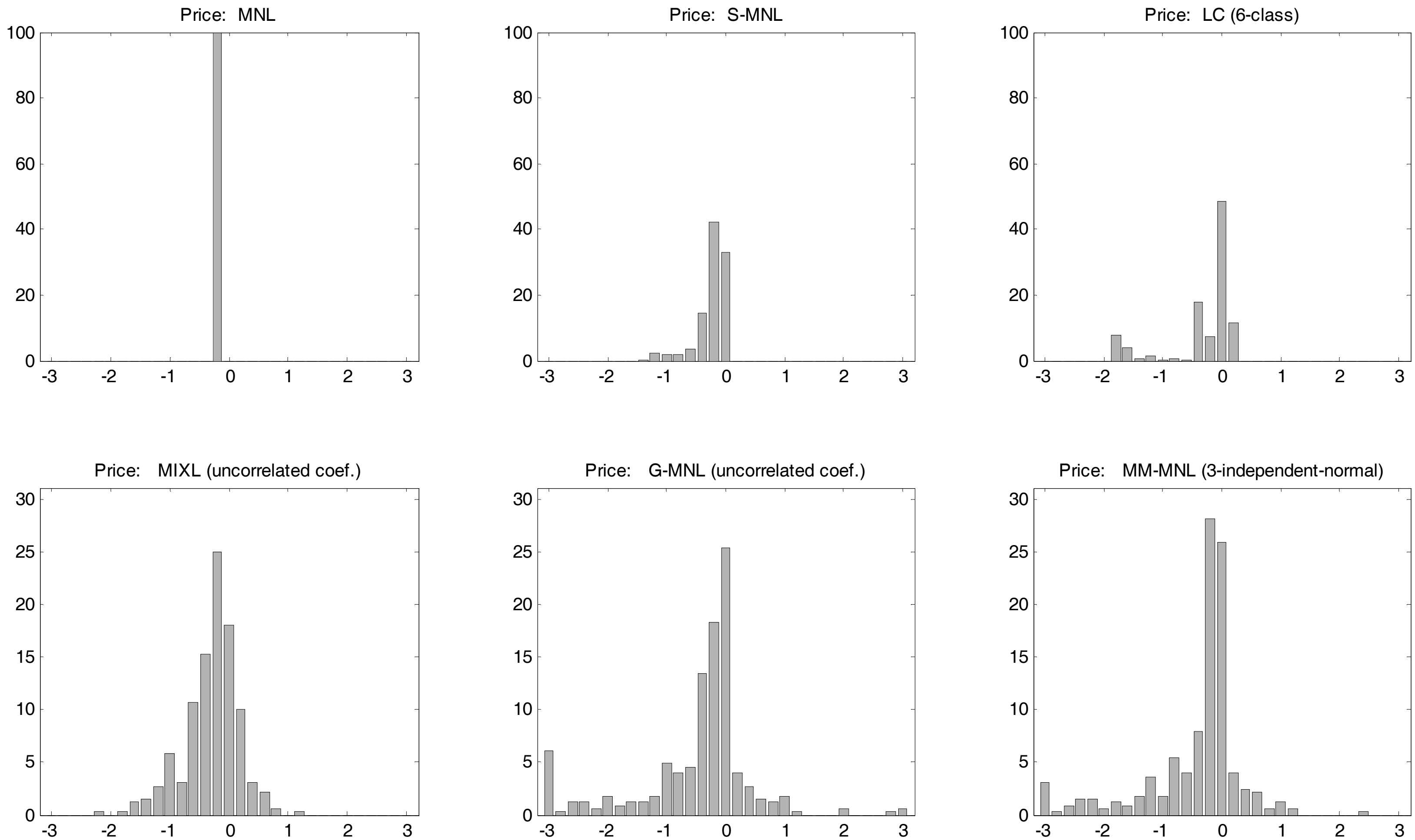
Train, K. (2008), "EM Algorithms for Nonparametric Estimation of Mixing Distributions," *Journal of Choice Modeling,* 1:1, 40-69.

Train, K. and M.Weeks (2005), Discrete Choice Models in Preference Space and Willingness-to-Pay Space, in *Applications of Simulation Methods in Environmental Resource Economics*, in A. Alberini and R. Scarpa, eds., Springer Publisher: Dordrecht, The Netherlands

Thurstone, L. (1927), "A Law of Comparative Judgment," *Psychological Review,* 34, 273-86.

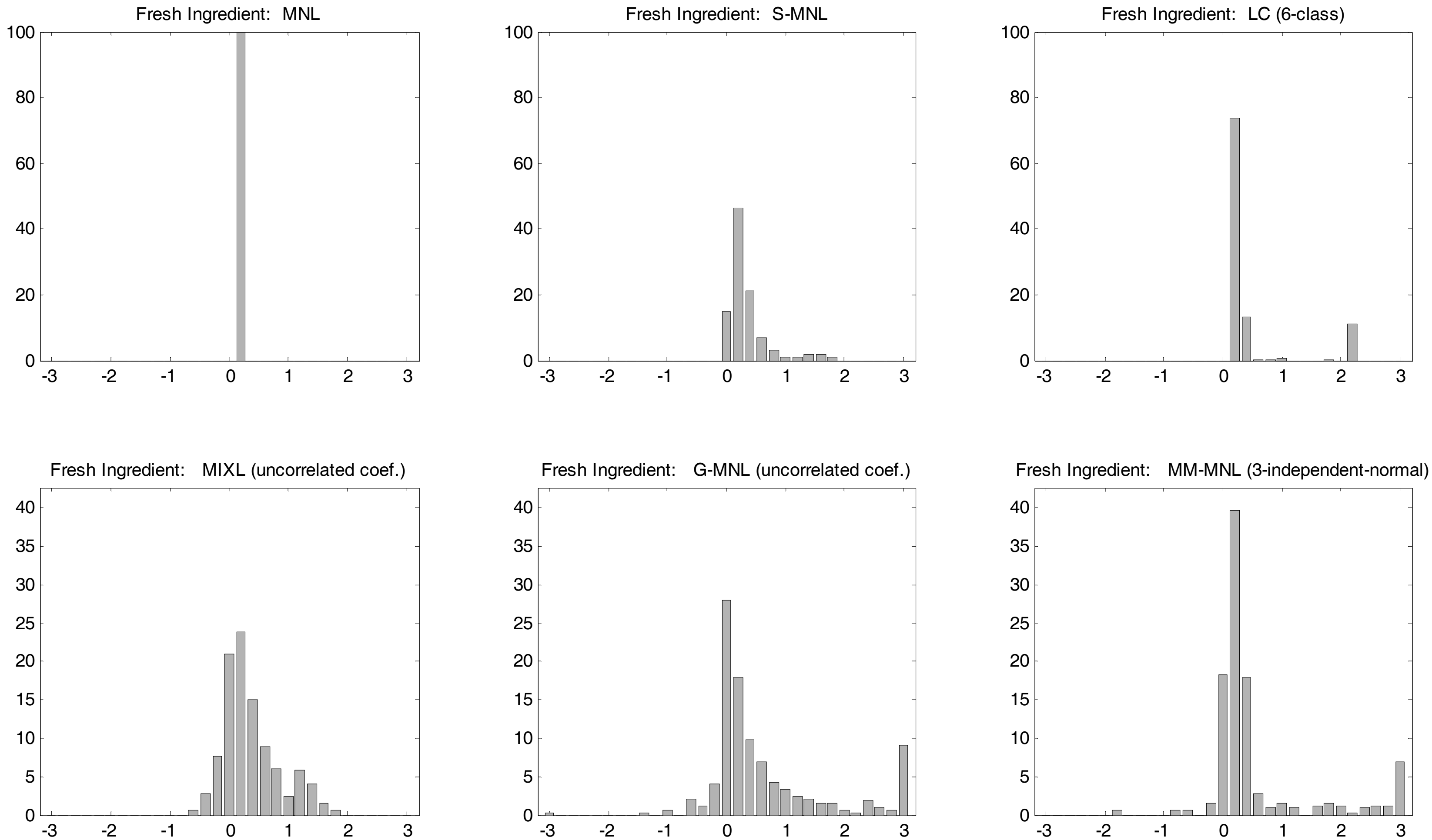
Villani, M., R. Kohn and P. Giordani (2007), "Nonparametric Regression Density Estimation Using Smoothly Varying Normal Mixtures," Sveriges Riksbank Research Paper Series #211.

Figure 1: Posterior distribution of individual-level PRICE coefficient from Pizza B dataset



Note: The first bin includes data between –infinity and the first center (-3) and the last bin includes data between the last bin center (3) and infinity. For G-MNL, the left tail span to -15.9 and there is a small mode at -4.1.

Figure 2: Posterior distribution of individual-level FRESH INGREDIENT coefficient from Pizza B dataset



Note: The first bin includes data between –infinity and the first center (-3) and the last bin includes data between the last bin center (3) and infinity. The maximum values of the right tails of G-MNL and MM-MNL are 12.4 and 5.2, respectively. MM-MNL also has a small mode at 3.

Figure 3: Predicted distribution of probability of choosing firm A from MIXL, LC, G-MNL and MM-MNL models when firm A <u>improves ingredient quality</u> and <u>increases price $4</u>

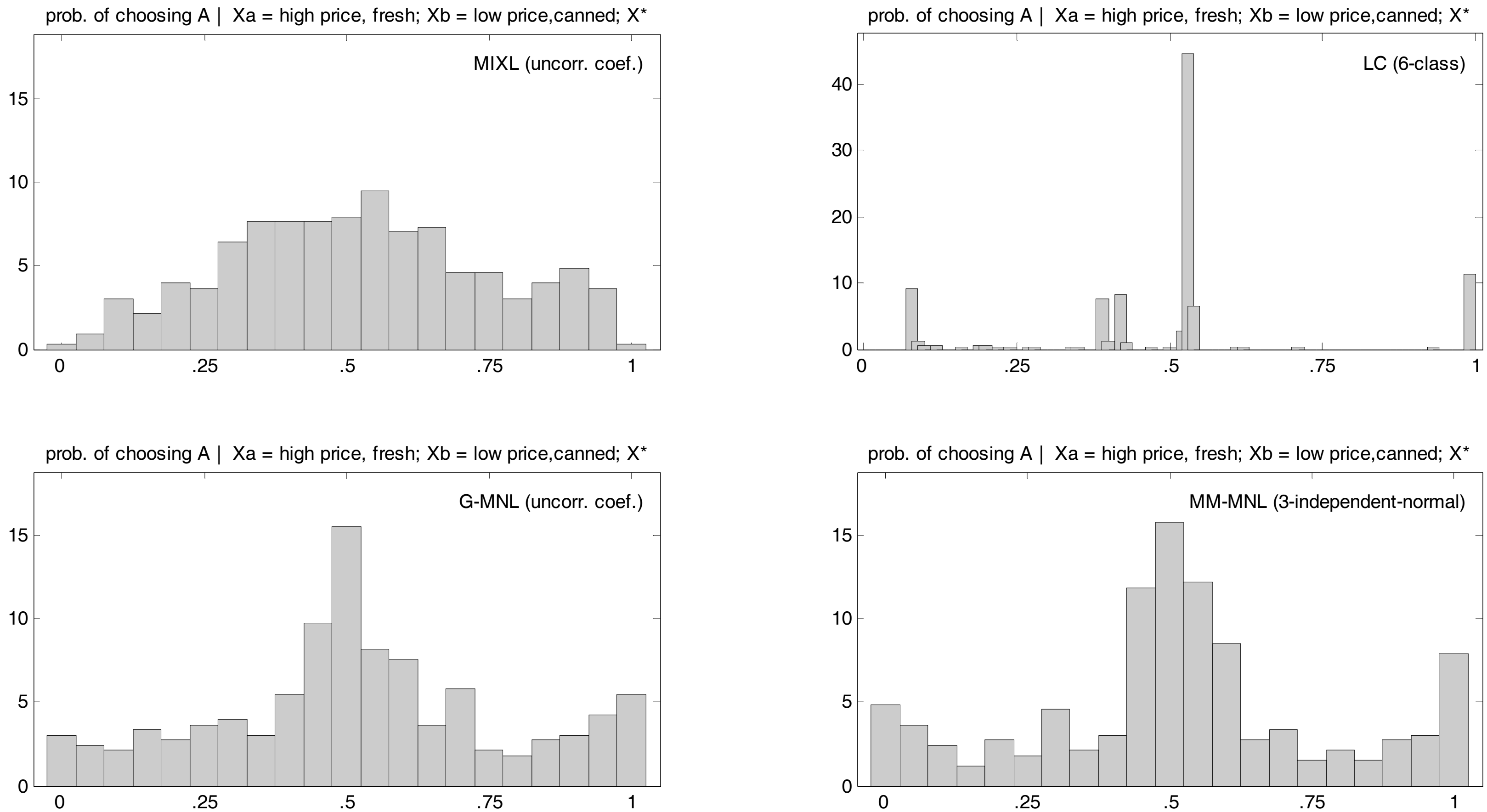


Figure 4: Posterior distribution of individual-level BAKING METHOD coefficient and predicted probability of choosing firm A from G-MNL and MM-MNL models when firm A uses woodfire baking method but does not increase price

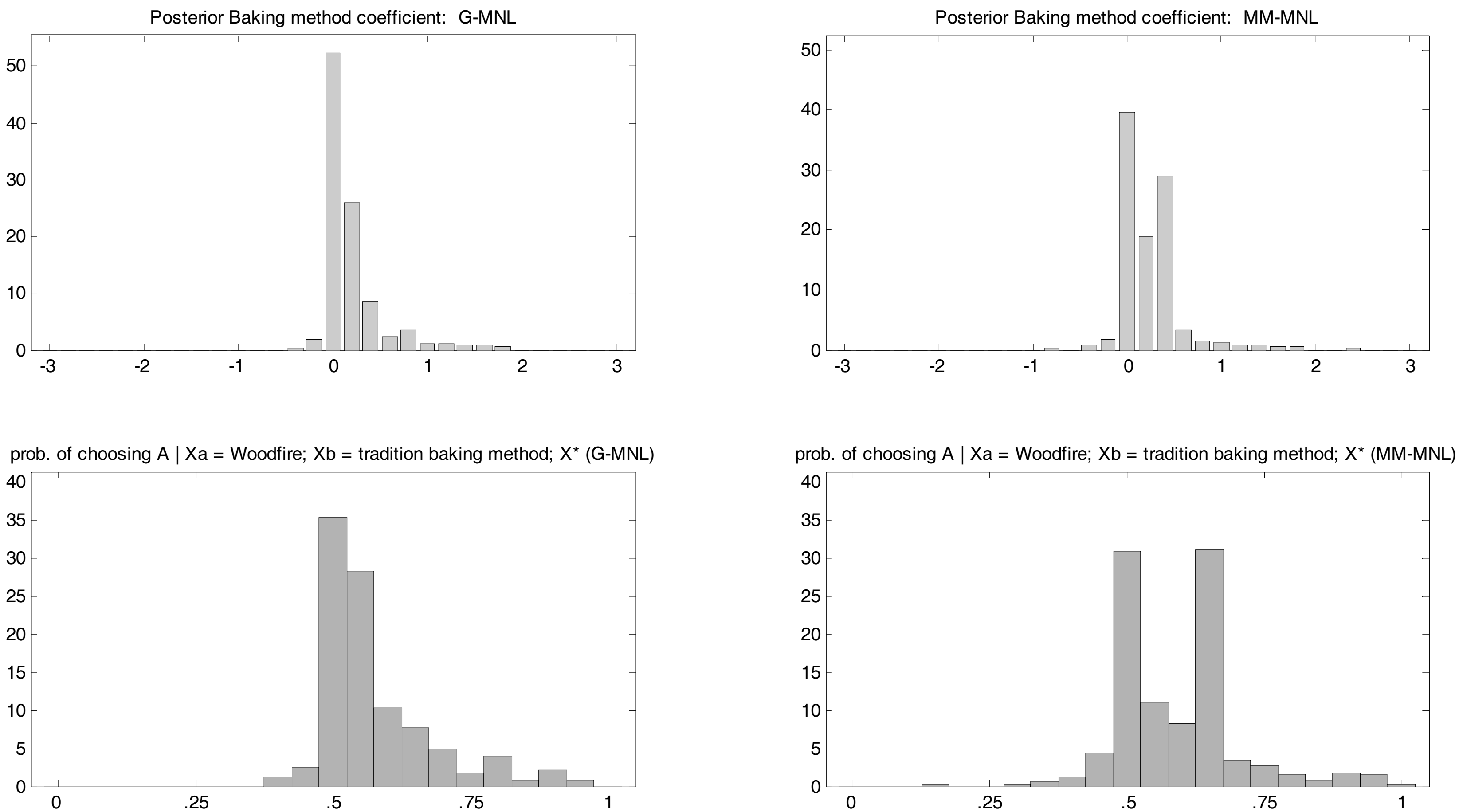


Table 1: Empirical Data Sets

		No. of choices	No. of choice occasions	No. of respondents	No. of observations	No. of attributes
1	Tay Sachs Disease & Cystic Fibrosis test Jewish sample (3 ASCs)	4	16	210	3360	11
2	Tay Sachs Disease & Cystic Fibrosis test General population sample (3 ASCs)	4	16	261	4176	11
3	Mobile phone (1 ASC)	4	8	493	3944	15
4	Pizza A (no ASC)	2	16	178	2848	8
5	Holiday A (no ASC)	2	16	331	5296	8
6	Papsmear test (1 ASC)	2	32	79	2528	6
7	Pizza B (no ASC)	2	32	328	10496	16
8	Holiday B (no ASC)	2	32	683	21856	16
9	Charge card A (2 ASCs)	3	4	827*	3308	17
	Charge card B (3 ASCs)	4	4	827*	3308	18

Note: * The respondents in the two credit card data sets are the same. They first complete 4 tasks with 3 options and then answer 4 tasks with 4 options. Some data sets were used in previous research (see Hall et al (2006) for data sets 1 and 2, Fiebig and Hall (2005) for data set 6, and Louviere et al (2008) for data sets 4, 5, 7 and 8).

Table 2: Attributes and Levels

Tay Sachs disease (TS) & Cystic Fibrocis (CF) test: Jewish and General population

	Attributes	Levels
1	ASC for TS test	0,1
2	ASC for CF test	0,1
3	ASC for both tests	0,1
4	Cost to you of being tested for TS	(0,150,300,600)/1000
5	Cost to you of being tested for CF	(0,375,750,1500)/1000
6	Cost to you of being tested both TS and CF	(0,150,…,1800,2100)/1000
7	Whether your doctor recommends you have a test	-1(no),1(yes)
8	The chance that you are a carrier even if the test is negative	(15,30,45,60)/10
9	Whether you are told your carrier status as an individual or as a couple	-1(individual), 1(couple)
10	Risk of being a carrier for TS	log base 10 of (.004,.04,.4,4) x 10^3
11	Risk of being a carrier for CF	log base 10 of (.004,.04,.4,4) x 10^3

Papsmear test

	Attributes	Levels
1	ASC for test	0(no),1(yes)
2	Whether you know doctor	0(no),1(yes)
3	Whether doctor is male	0(no),1(yes)
4	Whether test is due	0(no),1(yes)
5	Whether doctor recommends	0(no),1(yes)
6	Test cost	{0,10,20,30}/10

Pizza A: attributes 1-8; Pizza B: attributes 1-16 (No ASC)

	Attributes	Levels
1	Gourmet	-1 (Traditional),1(Gourmet)
2	Price	-1 ($13),1 ($17)
3	Ingredient freshness	-1 (some canned),1(all fresh ingredients)
4	Delivery time	-1(45 mins),1(30 min)
5	Crust	-1(thin),1(thick)
6	Sizes	-1(single size),1(3 sizes)
7	Steaming hot	-1(warm),1(steaming hot)
8	Late open hours	-1(till 10 pm.), 1 (till 1 am.)
9	Free delivery charge	-1($2),1 (free)
10	Local store	-1(chain),1(local)
11	Baking Method	-1(traditional),1(wood fire)
12	Manners	-1(friendly),1(polite & friendly)
13	Vegetarian availability	-1(no),1(yes)
14	Delivery time guaranteed	-1(no),1(yes)
15	Distance to the outlet	-1(in other suburb),1(in own suburb)
16	Range/variety availability	-1(restricted menu),1(large menu)

Mobile phone

	Attributes	Levels
1	ASC for purchase (phone 1, phone 2 or phone 3)	0,1
	Voice Commands (omitted Text to voice or voice to text converter)	
2	(1) No	-1,0,1
3	(2) Voice dialling by number or name	-1,0,1
4	(3) Voice operating commands	-1,0,1
	Push to Communicate (omitted to share video)	
5	(1) No	-1,0,1
6	(2) to talk	-1,0,1
7	(3) to share pictures or video	-1,0,1
	Email Access (omitted email with attachments)	
8	(1) personal emails	-1,0,1
9	(2) corporate emails (VPN, RIM)	-1,0,1
10	(3) both personal & corporate emails on multiple accounts	-1,0,1
11	WiFi	-1(No), 1(Yes)
12	USB Cable or Cradle connection	-1(No), 1(Yes)
13	Thermometer	-1(No), 1(Yes)
14	Flashlight	-1(No), 1(Yes)
15	Price	(0,11.7,19.5,…,497.25, 563.55)/100 (36 unique values)

Holiday A: attributes 1-8; Holiday B: attributes 1-16 (No ASC)

	Attributes	Levels
1	Price	-1($999), 1 ($1200)
2	Overseas destination	-1(Australia), 1(Overseas)
3	Airline	-1(Qantas), 1(Virgin)
4	Length of stay	-1(7 days), 1(12 days)
5	Meal inclusion	-1(no), 1(yes)
6	Local tours availability	-1(no), 1(yes)
7	Peak season	-1(off-peak), 1(peak)
8	4-star Accommodation	-1(2-star), 1 (4-star)
9	Length of Trip	-1(3 hours), 1 (5 hours)
10	Cultural activities	-1(Historical sites), 1 (museum)
11	Distance from hotel to attractions	-1 (200m), 1 (5km)
12	Swimming pool avail.	-1(no), 1(yes)
13	Helpfulness	-1(helpful), 1(very helpful)
14	Individual tour	-1 (organized tour), 1 (individual)
15	Beach availability	-1(no), 1(yes)
16	Brand	-1(Jetset), 1 (Creative Holidays)

Table 2 (continued)

Charge card A & B (no transaction option for Card A)

	Attributes	Levels
1	ASC for credit card	0,1
2	ASC for debit card	0,1
3	ASC for transaction card	0,1
4	Annual fee	(-70,-30,10,70)/10
5	Transaction fee	(-.5, -.3, .1, .5)*10
6	Permanent overdraft facility	
	credit:	0 (N/A)
	debit/trans:	-1(Available), 1(Not available)
7	overdraft interest free days (up to)	(-30, 5, 15, 30)/10
8	Interest charged on outstanding credit/overdraft	(-.075, -.035, .015, .075)*100
9	Interest earned on positive balance	
	credit:	(-.025, .025)*100
	debit/trans:	0.015*100
10	Cash advance interest rate	
	credit:	(-.035, -.005, .015, .035)*100
	debit/trans:	0.015*100
	Location and shop access (omited EFTPOS + telephone + internet + mail, use world wide)	
11	(1) Nowhere else, use Australia wide	-1,0,1
12	(2) EFTPOS + telephone + internet + mail, use Australia wide	-1,0,1
13	(3) Nowhere else, use world wide	-1,0,1
14	Loyalty scheme	0(None), 1(Frequent Flyer/Fly Buys and other rewards)
15	Loyalty scheme annual fees	(-40,40)/10 if Loyalty scheme = 1; 0 if Loyalty scheme = 0
16	Loyalty scheme points earning	-1(points on outstanding balance interest paid on), 1(points on purchases only)
17	Merchant surcharge for using card	(-.03, -.01, .01, .03)*100
18	Surcharge for transactions at other banks ATM	
	credit:	-1.5
	debit/trans:	(-1.5, -.5, .5, 1.5)

Table 3: Tay Sachs Disease (TS) and Cystic Fibrosis (CF) test: Jewish sample (3 ASCs)

	MNL		S-MNL (with R.E.)		MIXL[a]		G-MNL[a]		Latent class[b]						MM-MNL[c]			
									class 1		class 2		class 3		class 1		class 2	
	est	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
ASC for TS test	**-0.57**	0.14	**-0.57**	0.20	-0.67	0.47	-0.17	0.41	-0.33	0.36	**-0.64**	0.28	2.64	2.16	**-0.74**	0.31	**-0.69**	0.33
ASC for CF test	**-0.82**	0.15	**-0.88**	0.22	-0.74	0.42	-0.27	0.36	-0.39	0.30	**-0.92**	0.35	3.18	2.35	**-0.69**	0.34	**-0.86**	0.34
ASC for both tests	-0.08	0.15	0.01	0.27	-0.38	0.52	0.01	0.45	**-0.98**	0.31	-0.42	0.29	**6.26**	2.63	-0.42	0.36	**0.64**	0.36
TS cost	**-2.51**	0.24	**-3.45**	0.34	**-4.75**	0.63	**-5.62**	0.78	**-2.03**	0.40	**-5.20**	0.68	-1.78	1.21	**-3.00**	0.42	**-9.38**	1.64
CF cost	**-1.43**	0.13	**-1.96**	0.20	**-3.24**	0.38	**-3.57**	0.42	**-1.46**	0.22	**-3.03**	0.30	-2.50	1.29	**-2.19**	0.27	**-6.30**	1.01
Both cost	**-1.20**	0.07	**-2.70**	0.17	**-3.65**	0.26	**-4.25**	0.37	**-1.88**	0.14	**-4.77**	0.32	**-1.94**	0.44	**-2.59**	0.26	**-4.86**	0.51
Recommend	**0.33**	0.04	**0.56**	0.06	**0.95**	0.13	**1.00**	0.19	**0.43**	0.10	**0.64**	0.08	0.72	0.89	**0.56**	0.12	**0.92**	0.16
Inaccuracy	**-0.12**	0.02	**-0.15**	0.03	-0.14	0.09	**-0.36**	0.10	**-0.29**	0.06	**-0.12**	0.04	**-0.55**	0.26	**-0.27**	0.08	**-0.19**	0.09
Form	0.07	0.04	**0.12**	0.05	0.28	0.16	0.15	0.19	**-0.21**	0.10	**0.19**	0.08	-0.43	0.41	-0.05	0.15	**0.45**	0.20
Own risk of TS	**0.50**	0.03	**1.05**	0.08	**1.39**	0.12	**1.67**	0.18	**1.43**	0.12	**0.81**	0.08	**0.96**	0.45	**1.64**	0.13	**0.46**	0.10
Own risk of CF	**0.47**	0.04	**1.02**	0.07	**1.26**	0.12	**1.50**	0.18	**1.30**	0.09	**0.77**	0.12	**0.66**	0.22	**1.54**	0.11	**0.37**	0.10
τ			**0.64**	0.06			**0.45**	0.08										
γ							0.11	0.15										
Class probability									**0.29**	0.03	**0.27**	0.03	**0.10**	0.02	**0.62**	0.04	**0.38**	0.04
No. of parameters	11		18		77		79		59						45			
LL	-3717		-2815		-2500		-2480		-2701						-2620			
AIC	7455		5666		5154		5118		5521						5330			
BIC	7523		5777		5626		**5601**		5882						5605			
CAIC	7534		5795		5703		5680		5941						**5650**			

Notes: [a] estimates from correlated coefficient specification; [b] estimates from LC with 5 classes; [c] estimates from MM-MNL with 2 independent normals. Bold estimates are statistically significant at 5%.

Table 4: Tay Sachs Disease (TS) and Cystic Fibrosis (CF) test: General population sample (3 ASCs)

	MNL		S-MNL (with R.E.)		MIXL[a]		G-MNL[a]		Latent class[b]						MM-MNL[c]			
									class 1		class 2		class 3		class 1		class 2	
	est	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
ASC for TS test	**-2.18**	0.13	**-3.17**	**0.20**	**-3.24**	0.32	**-3.29**	0.31	-5.89	3.53	**-2.57**	0.53	**-1.28**	0.37	**-2.74**	0.46	**-3.02**	0.26
ASC for CF test	**-1.92**	0.12	**-2.92**	**0.21**	**-2.61**	0.33	**-2.64**	0.29	-4.86	3.98	**-1.90**	0.52	-0.19	0.36	**-1.27**	0.57	**-2.57**	0.24
ASC for both tests	**-1.49**	0.13	**-3.63**	**0.27**	**-3.13**	0.44	**-3.73**	0.40	-4.42	3.17	**-3.73**	0.51	0.19	0.40	0.80	0.62	**-3.65**	0.26
TS cost	**-1.12**	0.25	**-1.39**	0.25	**-2.71**	0.50	**-2.99**	0.52	1.28	5.43	**-1.42**	0.66	-0.96	0.51	**-9.89**	1.35	**-1.13**	0.43
CF cost	**-0.73**	0.10	**-0.87**	0.11	**-2.17**	0.30	**-2.56**	0.32	-0.76	4.33	**-0.75**	0.28	**-1.27**	0.20	**-10.55**	1.79	**-0.98**	0.18
Both cost	**-0.51**	0.06	**-1.11**	0.10	**-2.13**	0.23	**-2.27**	0.22	-2.82	6.98	**-1.08**	0.19	**-1.34**	0.17	**-4.39**	0.59	**-1.26**	0.14
Recommend	**0.35**	0.03	**0.61**	0.05	**0.95**	0.12	**0.94**	0.12	0.14	1.03	**0.45**	0.15	**0.27**	0.11	**1.08**	0.17	**0.68**	0.12
Inaccuracy	0.02	0.02	**0.05**	0.02	0.10	0.07	0.02	0.06	-0.32	0.52	-0.18	0.12	**0.10**	0.05	**-0.53**	0.15	0.07	0.06
Form	**0.06**	0.03	**0.08**	0.04	**0.25**	0.10	0.21	0.13	0.29	0.69	-0.12	0.18	-0.15	0.10	-0.16	0.15	0.21	0.12
Own risk of TS	**0.39**	0.03	**0.91**	0.07	**1.06**	0.11	**1.26**	0.13	0.56	1.03	**1.94**	0.15	**0.30**	0.09	0.15	0.13	**1.27**	0.07
Own risk of CF	**0.37**	0.03	**0.88**	0.06	**0.99**	0.10	**1.16**	0.10	0.40	0.54	**1.65**	0.14	**0.20**	0.08	-0.13	0.17	**1.23**	0.07
τ			**0.89**	0.07			**0.56**	0.07										
γ							**0.64**	0.08										
Class probability									**0.22**	0.03	**0.18**	0.03	**0.17**	0.03	**0.505**	0.03	**0.495**	0.03
No. of parameters	11		18		77		79		83						45			
LL	-4649		-3221		-2946		-2914		-3016						-3022			
AIC	9320		6477		6047		5986		6197						6134			
BIC	9390		6591		6535		6487		6723						**6420**			
CAIC	9401		6610		6612		6566		6806						**6465**			

Notes: [a] estimates from correlated coefficient specification; [b] estimates from LC with 7 classes; [c] estimates from MM-MNL with 2 independent normals. Bold estimates are statistically significant at 5%.

Table 5: Mobile phones (1 ASC)

	MNL		S-MNL (with R.E.)		MIXL[a]		G-MNL[a]		Latent class[b]						MM-MNL[c]			
									class 1		class 2		class 3		class 1		class 2	
	est	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
ASC for purchase	**-0.80**	0.05	**-0.35**	0.12	**-0.50**	0.11	**-0.46**	0.12	**-1.15**	0.14	**-2.96**	0.27	**1.49**	0.26	0.22	0.25	**-1.31**	0.20
No voice comm.	0.04	0.04	0.06	0.05	0.04	0.05	0.04	0.06	0.02	0.08	0.03	0.31	0.07	0.11	0.07	0.08	-0.02	0.12
Voice dialing	**0.08**	0.04	0.05	0.06	**0.10**	0.05	0.09	0.06	0.08	0.09	0.23	0.26	-0.12	0.12	0.12	0.08	0.10	0.12
Voice operation	**-0.12**	0.04	-0.11	0.06	**-0.13**	0.05	**-0.12**	0.06	**-0.21**	0.10	-0.37	0.39	0.07	0.11	-0.08	0.08	-0.21	0.14
No push to com.	0.06	0.04	**0.12**	0.06	0.05	0.05	0.06	0.06	0.05	0.10	-0.22	0.32	0.18	0.12	0.07	0.08	0.05	0.12
Push to talk	0.03	0.04	0.03	0.07	0.05	0.05	0.07	0.06	0.17	0.09	-0.21	0.39	0.05	0.14	0.00	0.09	0.12	0.11
Push to share pics/video	-0.02	0.04	-0.08	0.07	-0.02	0.05	-0.04	0.06	**-0.23**	0.11	0.51	0.28	-0.06	0.13	0.05	0.09	-0.18	0.13
Personal e-mail	-0.07	0.04	-0.04	0.06	-0.08	0.05	-0.07	0.06	-0.15	0.10	0.32	0.27	0.03	0.13	-0.03	0.09	-0.13	0.13
Corporate e-mail	**0.09**	0.04	0.08	0.07	0.08	0.05	0.08	0.06	0.10	0.08	0.00	0.31	-0.01	0.14	0.06	0.09	0.09	0.11
both e-mails	-0.05	0.04	-0.08	0.06	-0.03	0.05	-0.04	0.06	0.08	0.09	-0.41	0.39	-0.05	0.13	-0.11	0.09	0.08	0.12
WiFi	0.001	0.02	-0.02	0.03	-0.002	0.03	-0.01	0.03	0.08	0.06	0.05	0.17	-0.08	0.07	-0.07	0.05	0.09	0.07
USB Cable/Cradle	**0.06**	0.03	**0.08**	0.04	**0.07**	0.03	**0.08**	0.03	0.05	0.06	-0.01	0.18	**0.20**	0.08	0.08	0.05	0.07	0.07
Themometer	**0.07**	0.03	0.05	0.03	**0.07**	0.03	**0.08**	0.03	0.05	0.05	0.00	0.18	0.10	0.06	**0.11**	0.05	0.02	0.07
Flashlight	0.05	0.03	0.01	0.03	0.05	0.03	0.04	0.03	**0.16**	0.06	-0.10	0.17	-0.03	0.08	-0.02	0.05	**0.18**	0.07
Price/100	**-0.32**	0.02	**-1.02**	0.16	**-0.76**	0.06	**-0.88**	0.10	-0.04	0.05	**-0.64**	0.20	**-2.06**	0.21	**-1.57**	0.20	-0.05	0.08
τ			**1.45**	0.15			**0.66**	0.18										
γ							0.01	0.49										
Class probability									**0.32**	0.03	**0.28**	0.03	**0.22**	0.03	**0.67**	0.05	**0.33**	0.05
No. of parameters	15		17		30		32		63						61			
LL	-4475		-3990		-3971		-3966		-3952						-3927			
AIC	8980		8014		8002		7996		8030						**7976**			
BIC	9074		**8121**		8190		8197		8426						8359			
CAIC	9089		**8138**		8220		8229		8489						8420			

Notes: [a] estimates from uncorrelated coefficient specification; [b] estimates from LC with 4 classes; [c] estimates from MM-MNL with 2 independent normals. Bold estimates are statistically significant at 5%.

Table 6: Pizza A (No ASC)

	MNL		S-MNL		MIXL[a]		G-MNL[a]		Latent class[b]						MM-MNL[c]			
									class 1		class 2		class 3		class 1		class 2	
	est	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
Gourmet	0.02	0.02	0.03	0.04	0.03	0.05	**0.45**	0.22	-0.01	0.05	0.02	0.02	0.08	0.10	0.02	0.07	0.14	0.47
Price	**-0.16**	0.02	**-0.19**	0.05	**-0.35**	0.06	**-1.67**	0.65	**-0.20**	0.06	**-0.16**	0.03	**-0.39**	0.09	**-0.18**	0.06	-4.63	2.71
Ingredient freshness	**0.48**	0.03	**1.45**	0.29	**0.96**	0.08	**4.65**	1.69	**1.57**	0.09	**0.12**	0.06	0.30	0.16	**0.59**	0.08	13.47	7.73
Delivery time	**0.09**	0.03	**0.16**	0.08	**0.16**	0.05	**0.74**	0.35	0.10	0.09	**0.10**	0.04	**0.32**	0.09	0.06	0.05	3.95	2.36
Crust	0.02	0.03	0.01	0.04	0.02	0.06	0.42	0.26	**-0.12**	0.06	0.01	0.05	**-0.30**	0.09	-0.06	0.08	1.18	1.05
Sizes	**0.09**	0.03	**0.12**	0.06	**0.20**	0.05	**0.81**	0.37	**0.15**	0.07	0.06	0.04	**0.23**	0.11	**0.23**	0.07	0.92	0.81
Steaming hot	**0.38**	0.03	**1.02**	0.24	**0.87**	0.08	**4.46**	1.64	**0.50**	0.08	**0.12**	0.06	**1.60**	0.18	**0.50**	0.08	9.85	5.76
Late open hours	**0.04**	0.02	0.08	0.06	0.07	0.05	0.29	0.17	0.09	0.08	**0.06**	0.03	0.02	0.07	**0.12**	0.06	-0.97	0.72
τ			**1.69**	0.18			**1.79**	0.24										
γ							0.01	0.01										
Class probability									**0.36**	0.04	**0.32**	0.04	**0.23**	0.04	**0.57**	0.04	**0.43**	0.04
No. of parameters	8		9		16		18		35						33			
LL	-1657		-1581		-1403		-1373		-1418						-1328			
AIC	3330		3179		2838		2782		2907						2722			
BIC	3378		3233		2933		**2889**		3115						2919			
CAIC	3386		3242		2949		**2907**		3150						2952			

Note: [a] estimates from uncorrelated coefficient specification; [b] estimates from LC with 4 classes; [c] estimates from MM-MNL with 2 independent normals. Bold estimates are statistically significant at 5%.

Table 7: Holiday A (No ASC)

	MNL		S-MNL		MIXL[a]		G-MNL[a]		Latent class[b]						MM-MNL[c]			
									class 1		class 2		class 3		class 1		class 2	
	est	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
Price	**-0.16**	0.02	**-0.17**	0.03	**-0.33**	0.04	**-0.74**	0.12	**-0.35**	0.05	**-0.17**	0.06	**-0.16**	0.03	**-0.35**	0.06	**-0.30**	0.07
Overseas destination	**0.09**	0.02	**0.17**	0.02	**0.23**	0.06	**0.32**	0.11	**0.12**	0.04	**0.29**	0.06	0.01	0.05	**0.27**	0.06	0.21	0.17
Airline	-0.01	0.02	**-0.05**	0.02	-0.02	0.03	-0.1	0.06	-0.03	0.04	-0.01	0.08	-0.03	0.04	0.00	0.06	-0.04	0.07
Length of stay	**0.26**	0.02	**0.35**	0.04	**0.52**	0.04	**1.24**	0.19	**0.56**	0.05	**0.36**	0.10	0.05	0.04	**0.57**	0.07	**0.61**	0.10
Meal inclusion	**0.27**	0.02	**0.31**	0.03	**0.56**	0.04	**1.29**	0.2	**0.73**	0.06	**0.28**	0.07	0.07	0.08	**0.71**	0.09	**0.38**	0.08
Local tours availability	**0.09**	0.02	**0.09**	0.03	**0.19**	0.03	**0.45**	0.09	**0.24**	0.05	**0.23**	0.07	-0.01	0.04	**0.32**	0.07	0.02	0.08
Peak season	0.03	0.02	-0.004	0.03	**0.06**	0.03	**0.14**	0.07	0.05	0.05	0.03	0.10	0.02	0.03	0.06	0.06	0.06	0.07
4-star Accommodation	**0.44**	0.02	**0.65**	0.05	**0.86**	0.06	**1.99**	0.29	**0.49**	0.04	**1.50**	0.11	**0.13**	0.06	**1.23**	0.12	**0.41**	0.06
τ			**0.97**	0.08			**1.19**	0.10										
γ							0.00	0.18										
Class probability									**0.34**	0.04	**0.26**	0.03	**0.18**	0.03	**0.58**	0.04	**0.42**	0.04
No. of parameters	8		9		16		18		44						33			
LL	-3066		-2967		-2553		-2519		-2502						-2464			
AIC	6149		5952		5139		5074		5092						4994			
BIC	6201		6011		5244		**5192**		5354						5211			
CAIC	6209		6020		5260		**5210**		5398						5244			

Note: [a] estimates from uncorrelated coefficient specification; [b] estimates from LC with 5 classes; [c] estimates from MM-MNL with 2 independent normals.
Bold estimates are statistically significant at 5%.

Table 8: Papsmear test (1 ASC)

	MNL		S-MNL		MIXL[a]		G-MNL[a]		Latent class[b]						MM-MNL[c]			
									class 1		class 2		class 3		class 1		class 2	
	est	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
ASC for test	**-0.40**	0.14	-0.60	0.37	**-1.26**	0.30	**-0.80**	0.31	**-1.59**	0.22	4.31	9.57	**-1.37**	0.34	-0.16	0.43	-1.35	1.10
If know doctor	**0.32**	0.09	**0.63**	0.14	**0.78**	0.18	**0.68**	0.21	0.02	0.27	-1.34	9.51	**1.27**	0.13	0.20	0.28	2.15	1.21
If doctor is male	**-0.70**	0.09	**-1.24**	0.16	**-1.39**	0.30	**-1.99**	0.32	-0.18	0.25	0.90	4.55	**-0.75**	0.27	-0.40	0.23	**-6.14**	1.46
If test is due	**1.23**	0.10	**2.74**	0.29	**3.26**	0.31	**3.35**	0.42	**3.15**	0.16	2.67	12.80	**0.88**	0.22	**3.20**	0.41	**3.82**	0.65
If doctor recommends	**0.51**	0.10	**0.74**	0.17	**1.33**	0.23	**1.65**	0.31	**1.57**	0.18	0.62	15.60	**0.52**	0.27	**1.31**	0.38	**1.53**	0.69
Test cost	**-0.08**	0.04	**-0.17**	0.07	**-0.22**	0.09	**-0.28**	0.09	**-0.18**	0.09	-0.50	1.85	-0.23	0.14	-0.16	0.12	-0.45	0.34
τ			**0.81**	0.11			**1.00**	0.11										
γ							0.01	0.38										
Class probability									**0.37**	0.05	**0.20**	0.04	**0.19**	0.04	**0.70**	0.07	**0.30**	0.07
No. of parameters	6		8		12		14		34						25			
LL	-1528		-1063		-945		-935		-958						-923			
AIC	3069		2143		1914		1897		1985						1896			
BIC	3104		2189		1984		**1979**		2183						2042			
CAIC	3110		2197		1996		**1993**		2217						2067			

Notes: [a] estimates from uncorrelated coefficient specification; [b] estimates from LC with 5 classes; [c] estimates from MM-MNL with 2 independent normals. Bold estimates are statistically significant at 5%.

Table 9: Pizza B (No ASC)

	MNL		S-MNL		MIXL[a]		G-MNL[a]		Latent class[b]						MM-MNL[c]					
									class 1		class 2		class 3		class 1		class 2		class 3	
	est	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
Gourmet	0.01	0.01	**0.05**	0.01	0.01	0.02	0.03	0.03	0.01	0.02	0.02	0.07	0.09	0.05	-0.03	0.04	-0.12	0.07	**0.37**	0.08
Price	**-0.17**	0.01	**-0.25**	0.02	**-0.30**	0.03	**-0.79**	0.07	-0.04	0.02	**-1.71**	0.09	0.24	0.11	-0.10	0.04	**-0.86**	0.10	-0.17	0.13
Ingredient freshness	**0.21**	0.01	**0.36**	0.03	**0.34**	0.03	**1.05**	0.08	**0.10**	0.02	**0.46**	0.06	**2.17**	0.19	**0.12**	0.03	**0.29**	0.07	**1.02**	0.13
Delivery time	**0.03**	0.01	0.04	0.02	0.05	0.02	**0.15**	0.04	0.02	0.02	0.14	0.10	-0.03	0.16	0.02	0.03	0.19	0.07	0.14	0.08
Crust	**0.08**	0.01	**0.09**	0.01	**0.08**	0.03	**0.59**	0.06	**-0.04**	0.01	-0.05	0.04	**0.31**	0.08	-0.03	0.03	**0.62**	0.09	0.15	0.07
Sizes	**0.07**	0.01	**0.08**	0.02	**0.11**	0.02	**0.23**	0.03	**0.05**	0.02	**0.19**	0.07	**0.28**	0.07	0.06	0.03	**0.31**	0.07	**0.26**	0.09
Steaming hot	**0.20**	0.01	**0.35**	0.03	**0.34**	0.02	**1.15**	0.09	**0.10**	0.02	**0.22**	0.07	**0.67**	0.07	**0.11**	0.03	**0.37**	0.06	**1.43**	0.17
Late open hours	**0.04**	0.01	0.02	0.02	**0.08**	0.02	0.08	0.04	**0.04**	0.01	0.06	0.06	0.07	0.10	0.01	0.02	**0.29**	0.07	**0.19**	0.06
Free delivery charge	**0.12**	0.01	**0.15**	0.02	**0.20**	0.02	**0.56**	0.06	**0.11**	0.01	**0.56**	0.04	0.15	0.08	**0.22**	0.05	**0.26**	0.06	**0.28**	0.07
Local store	**0.08**	0.01	**0.06**	0.02	**0.15**	0.02	**0.42**	0.05	**0.14**	0.01	-0.01	0.07	0.10	0.12	**0.09**	0.03	**0.43**	0.07	0.08	0.08
Baking Method	**0.07**	0.01	**0.07**	0.02	**0.11**	0.02	**0.25**	0.04	**0.06**	0.01	0.16	0.07	**0.29**	0.07	0.01	0.03	**0.32**	0.06	**0.35**	0.11
Manners	0.01	0.01	-0.004	0.02	0.02	0.02	0.01	0.04	0.03	0.02	0.03	0.08	-0.06	0.11	0.03	0.03	-0.06	0.08	0.11	0.11
Vegetarian availability	**0.09**	0.01	**0.06**	0.01	**0.13**	0.03	**0.34**	0.06	0.02	0.02	**0.15**	0.04	0.04	0.11	0.04	0.03	**0.35**	0.09	0.04	0.07
Delivery time guaranteed	**0.07**	0.01	**0.07**	0.02	**0.11**	0.02	**0.15**	0.04	**0.08**	0.02	**0.17**	0.05	0.12	0.12	**0.14**	0.04	0.07	0.08	**0.19**	0.07
Distance to the outlet	**0.06**	0.01	0.04	0.02	**0.09**	0.02	0.10	0.04	**0.09**	0.02	0.11	0.07	-0.12	0.10	**0.11**	0.04	0.09	0.07	0.06	0.07
Range/variety availability	**0.06**	0.02	0.04	0.02	**0.09**	0.02	**0.14**	0.05	**0.07**	0.03	0.03	0.07	0.07	0.10	**0.10**	0.03	0.03	0.07	0.19	0.08
τ			**1.22**	0.08			**1.26**	0.06												
γ							0.01	0.01												
Class probability									**0.51**	0.03	**0.14**	0.02	**0.12**	0.02	**0.41**	0.03	**0.31**	0.03	**0.28**	0.03
No. of parameters	16		17		32		34		101						98					
LL	-6747		-6607		-5892		-5689		-5591						-5310					
AIC	13525		13249		11849		11446		11385						10815					
BIC	13641		13372		12081		11693		12118						**11527**					
CAIC	13657		13389		12113		11727		12219						**11625**					

Notes: [a] estimates from uncorrelated coefficient specification; [b] estimates from LC with 6 classes; [c] estimates from MM-MNL with 3 independent normals. Bold estimates are statistically significant at 1%.

Table 10: Holiday B (No ASC)

	MNL		S-MNL		MIXL[a]		G-MNL[a]		Latent class[b]						MM-MNL[c]					
									class 1		class 2		class 3		class 1		class 2		class 3	
	est	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
Price	**-0.16**	0.01	**-0.16**	0.01	**-0.25**	0.02	**-0.34**	0.02	**-0.15**	0.02	**-0.18**	0.04	**-0.27**	0.04	**-0.45**	0.04	**-0.12**	0.05	**-0.16**	0.03
Overseas destination	**0.08**	0.01	**0.12**	0.01	**0.12**	0.02	**0.24**	0.03	-0.01	0.02	0.02	0.03	**0.25**	0.04	**0.42**	0.05	-0.10	0.07	0.06	0.03
Airline	-0.02	0.01	-0.02	0.01	**-0.03**	0.01	-0.03	0.02	-0.03	0.02	-0.09	0.04	0.00	0.07	-0.04	0.03	0.03	0.04	**-0.08**	0.03
Length of stay	**0.18**	0.01	**0.19**	0.01	**0.29**	0.02	**0.40**	0.02	**0.09**	0.02	**0.41**	0.03	**0.19**	0.05	**0.65**	0.05	**0.28**	0.06	**0.11**	0.03
Meal inclusion	**0.20**	0.01	**0.24**	0.02	**0.34**	0.02	**0.46**	0.03	**0.09**	0.02	**0.54**	0.04	**0.23**	0.08	**0.74**	0.04	**0.25**	0.04	**0.12**	0.02
Local tours availability	**0.07**	0.01	**0.08**	0.01	**0.11**	0.01	**0.17**	0.02	0.02	0.02	**0.21**	0.03	**0.21**	0.05	**0.21**	0.03	**0.20**	0.05	0.02	0.02
Peak season	0.003	0.01	0.02	0.01	0.001	0.01	-0.01	0.02	-0.02	0.01	0.04	0.03	0.07	0.05	0.02	0.03	-0.05	0.04	0.01	0.03
4-star Accommodation	**0.34**	0.01	**0.54**	0.03	**0.50**	0.02	**0.69**	0.03	**0.12**	0.02	**0.64**	0.04	**1.75**	0.06	**0.58**	0.04	**1.14**	0.07	**0.18**	0.03
Length of Trip	-0.02	0.01	**-0.03**	0.01	**-0.03**	0.01	-0.03	0.02	-0.03	0.02	**-0.11**	0.04	0.03	0.06	**-0.07**	0.03	0.02	0.03	-0.04	0.02
Cultural activities	**-0.05**	0.01	**-0.05**	0.01	**-0.09**	0.01	**-0.12**	0.01	**-0.07**	0.01	**-0.10**	0.03	-0.11	0.06	-0.06	0.03	**-0.14**	0.04	**-0.10**	0.03
Distance to attractions	**-0.08**	0.01	**-0.07**	0.01	**-0.12**	0.01	**-0.17**	0.02	**-0.06**	0.01	**-0.18**	0.03	-0.11	0.06	**-0.21**	0.03	**-0.12**	0.04	**-0.11**	0.03
Swimming pool avail.	**0.09**	0.01	**0.09**	0.01	**0.15**	0.01	**0.23**	0.02	0.00	0.02	**0.19**	0.03	**0.15**	0.04	**0.43**	0.04	0.08	0.04	0.02	0.02
Helpfulness	**0.04**	0.01	**0.03**	0.01	**0.06**	0.01	**0.07**	0.02	0.03	0.02	**0.08**	0.04	-0.07	0.08	**0.10**	0.03	0.07	0.04	0.03	0.02
Individual tour	**0.07**	0.01	**0.07**	0.01	**0.13**	0.02	**0.20**	0.02	-0.02	0.02	0.01	0.03	**0.20**	0.04	0.00	0.03	**0.61**	0.06	0.01	0.03
Beach availability	**0.11**	0.01	**0.10**	0.01	**0.18**	0.01	**0.22**	0.02	**0.06**	0.02	**0.20**	0.04	0.10	0.06	**0.33**	0.04	**0.12**	0.04	**0.13**	0.03
Brand	0.001	0.01	-0.01	0.02	0.003	0.02	0.004	0.02	0.00	0.03	0.01	0.04	0.02	0.08	-0.01	0.03	-0.01	0.05	0.01	0.03
τ	-		**1.13**	0.05	-		**0.72**	0.04												
γ							0.01	0.02												
Class probability									**0.27**	0.02	**0.15**	0.02	**0.14**	0.01	**0.39**	0.02	**0.33**	0.02	**0.28**	0.02
No. of parameters	16		17		32		34		152						98					
LL	-13478		-13027		-11600		-11476		-11231						-11012					
AIC	26988		26088		23263		23019		22766						22219					
BIC	27116		26224		23519		23291		23981						**23002**					
CAIC	27132		26241		23551		23325		24133						**23100**					

Notes: [a] estimates from uncorrelated coefficient specification; [b] estimates from LC with 9 classes; [c] estimates from MM-MNL with 3 independent normals. Bold estimates are statistically significant at 1%.

Table 11: Charge Card A (2 ASCs)

	MNL		S-MNL		MIXL[a]		G-MNL[a]		Latent class[b]						MM-MNL[c]			
									class 1		class 2		class 3		class 1		class 2	
	est	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
ASC for credit	**-0.85**	0.08	**-0.90**	0.18	**-1.31**	0.27	**-1.31**	0.27	**-2.83**	0.24	0.50	0.23	**1.98**	0.38	-19.99	13.72	**2.16**	0.38
ASC for debit	**-0.99**	0.08	**-1.22**	0.18	**-2.07**	0.31	**-2.05**	0.32	**-3.54**	0.24	**1.64**	0.21	0.36	0.44	-30.00	20.73	**2.34**	0.38
annual fee	**-0.08**	0.01	**-0.13**	0.01	**-0.18**	0.02	**-0.19**	0.02	**-0.11**	0.02	**-0.11**	0.02	**-0.08**	0.02	-1.16	0.81	**-0.18**	0.04
trans fee	**-0.53**	0.07	**-0.82**	0.11	**-1.34**	0.20	**-1.37**	0.21	**-1.25**	0.21	**-0.51**	0.17	**-0.75**	0.32	-8.72	6.15	**-0.71**	0.27
overdraft facility	**0.28**	0.06	**0.43**	0.09	**0.70**	0.15	**0.75**	0.16	**0.79**	0.17	-0.10	0.15	**0.88**	0.30	5.59	4.09	0.32	0.21
overdraft free days	0.04	0.02	**0.06**	0.02	0.07	0.03	0.07	0.03	0.04	0.04	0.04	0.04	0.11	0.06	0.45	0.43	0.06	0.05
interest charged	**-0.43**	0.06	**-0.67**	0.09	**-1.00**	0.15	**-1.01**	0.16	**-0.70**	0.15	**-0.54**	0.13	**-0.50**	0.20	-7.11	5.10	**-0.68**	0.22
interest earned	**0.04**	0.01	0.04	0.02	0.06	0.03	0.06	0.03	0.03	0.03	0.07	0.03	-0.02	0.05	0.23	0.28	0.05	0.04
access_1	-0.05	0.02	**-0.08**	0.02	-0.05	0.03	-0.06	0.03	-0.01	0.03	**-0.10**	0.04	-0.12	0.07	-0.12	0.25	**-0.17**	0.06
access_2	**-0.21**	0.05	**-0.31**	0.08	**-0.42**	0.13	**-0.39**	0.12	-0.33	0.14	-0.20	0.13	-0.31	0.21	-3.08	2.25	**-0.34**	0.16
access_3	0.06	0.05	0.11	0.07	0.22	0.11	0.23	0.11	0.15	0.12	-0.11	0.14	0.20	0.18	3.60	2.58	0.01	0.16
cash advance interest	-0.06	0.05	-0.12	0.08	-0.29	0.13	-0.34	0.14	-0.29	0.14	0.01	0.12	-0.08	0.23	-2.24	1.83	-0.02	0.17
loyal scheme	**0.26**	0.06	**0.33**	0.08	**0.44**	0.14	**0.47**	0.15	**0.43**	0.14	**0.29**	0.13	0.33	0.17	1.11	1.35	0.32	0.20
loyal fee	**-0.03**	0.01	**-0.05**	0.01	**-0.06**	0.02	**-0.06**	0.02	-0.05	0.02	-0.02	0.03	-0.06	0.04	-0.55	0.41	**-0.09**	0.04
loyal point	-0.04	0.04	0.04	0.06	0.07	0.09	0.07	0.09	0.02	0.10	0.02	0.10	0.19	0.18	1.41	1.12	-0.02	0.15
merchant surcharge	-0.02	0.01	**-0.07**	0.02	**-0.08**	0.03	**-0.08**	0.03	**-0.09**	0.03	-0.06	0.03	-0.04	0.06	-0.45	0.37	-0.09	0.05
surcharge at other ATM	-0.10	0.04	**-0.17**	0.06	-0.20	0.11	-0.19	0.11	-0.12	0.12	**-0.21**	0.10	-0.12	0.18	-0.17	0.75	-0.29	0.14
τ			**0.40**	0.17			0.21	0.24										
γ							0.50	0.56										
Class probability									**0.48**	0.02	**0.27**	0.03	**0.19**	0.03	**0.62**	0.02	**0.38**	0.02
No. of parameters	17		21		51		53		71						69			
LL	-3354		-2768		-2735		-2734		-2732						-2714			
AIC	6742		5579		5572		5574		5606						**5567**			
BIC	6846		**5707**		5883		5898		6039						5988			
CAIC	6863		**5728**		5934		5951		6110						6057			

Notes: [a] estimates from correlated coefficient specification; [b] estimates from LC with 4 classes; [c] estimates from MM-MNL with 2 independent normals. Bold estimates are statistically significant at 1%.

Table 12: Charge Card B (3 ASCs)

	MNL		S-MNL		MIXL[a]		G-MNL[a]		Latent class[b]						MM-MNL[c]			
									class 1		class 2		class 3		class 1		class 2	
	est	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
ASC for credit	**-0.97**	0.07	**-0.83**	0.18	**-1.29**	0.24	**-1.29**	0.24	**-2.98**	0.22	0.19	0.33	-0.12	0.35	**-1.75**	0.25	9.19	15.05
ASC for debit	**-1.29**	0.08	**-1.47**	0.20	**-1.99**	0.27	**-1.99**	0.27	**-3.76**	0.23	0.58	0.30	0.53	0.32	**-2.29**	0.28	-2.93	7.07
ASC for transaction	**-1.32**	0.08	**-1.59**	0.21	**-2.12**	0.29	**-2.12**	0.29	**-3.63**	0.23	0.10	0.32	**1.23**	0.30	**-2.54**	0.29	2.72	6.72
annual fee	**-0.10**	0.01	**-0.16**	0.01	**-0.22**	0.02	**-0.22**	0.02	**-0.17**	0.02	**-0.28**	0.03	-0.04	0.01	**-0.23**	0.02	-1.56	2.34
trans fee	**-0.61**	0.07	**-0.94**	0.10	**-1.32**	0.17	**-1.32**	0.17	**-1.10**	0.20	**-1.66**	0.32	-0.15	0.21	**-1.30**	0.19	-12.43	19.31
overdraft facility	**0.30**	0.06	**0.42**	0.08	**0.48**	0.11	**0.48**	0.11	0.20	0.14	**0.53**	0.19	0.16	0.17	**0.40**	0.12	6.30	9.71
overdraft free days	**0.06**	0.02	**0.09**	0.02	**0.10**	0.03	**0.10**	0.03	0.06	0.04	**0.28**	0.06	**0.10**	0.05	0.05	0.03	2.34	3.59
interest charged	**-0.56**	0.06	**-0.80**	0.08	**-0.90**	0.12	**-0.90**	0.13	**-0.45**	0.16	**-0.90**	0.21	**-0.91**	0.20	**-0.57**	0.13	-20.00	30.41
interest earned	0.02	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.03	**0.13**	0.05	-0.02	0.03	0.03	0.03	0.38	0.76
access_1	-0.01	0.02	0.00	0.02	-0.01	0.03	-0.01	0.03	-0.05	0.04	0.11	0.06	-0.08	0.07	0.03	0.04	-1.91	3.07
access_2	**-0.21**	0.05	**-0.35**	0.07	**-0.44**	0.10	**-0.44**	0.10	-0.25	0.13	**-0.67**	0.17	0.02	0.14	**-0.65**	0.12	7.59	11.84
access_3	**0.13**	0.05	**0.19**	0.06	**0.32**	0.09	**0.32**	0.09	0.28	0.12	**0.62**	0.17	-0.05	0.13	**0.59**	0.11	-11.97	18.43
cash advance interest	**-0.19**	0.05	**-0.32**	0.06	**-0.45**	0.11	**-0.45**	0.11	**-0.42**	0.14	**-0.90**	0.22	-0.07	0.15	**-0.68**	0.12	5.34	8.55
loyal scheme	**0.24**	0.05	**0.37**	0.07	**0.46**	0.11	**0.46**	0.11	**0.37**	0.14	0.36	0.17	**0.51**	0.15	**0.37**	0.13	10.73	16.01
loyal fee	-0.02	0.01	**-0.04**	0.01	-0.04	0.02	-0.04	0.02	-0.05	0.02	-0.05	0.03	-0.03	0.03	**-0.07**	0.02	0.54	0.97
loyal point	-0.03	0.04	-0.06	0.06	-0.06	0.08	-0.06	0.08	-0.02	0.09	-0.20	0.16	-0.08	0.12	0.01	0.09	-2.02	3.43
merchant surcharge	**-0.06**	0.01	**-0.08**	0.02	**-0.13**	0.03	**-0.13**	0.03	**-0.15**	0.03	-0.01	0.05	-0.07	0.04	**-0.13**	0.03	-0.86	1.49
surcharge at other ATM	-0.07	0.03	-0.11	0.04	**-0.19**	0.07	**-0.19**	0.07	-0.24	0.10	0.07	0.11	-0.08	0.08	-0.18	0.08	-3.94	6.22
τ			**0.38**	0.12			0.00	0.19										
γ							0.99	171										
Class probability									**0.44**	0.02	**0.21**	0.02	**0.19**	0.02	**0.72**	0.04	**0.28**	0.04
No. of parameters	18		25		54		56		75						74			
LL	-4100		-3402		-3364		-3364		-3391						-3329			
AIC	8236		6854		6836		6840		6933						**6806**			
BIC	8346		**7007**		7166		7182		7391						7258			
CAIC	8364		**7032**		7220		7238		7466						7332			

Notes: [a] estimates from correlated coefficient specification; [b] estimates from LC with 4 classes; [c] estimates from MM-MNL with 2 proportional covariance normal. Bold estimates are statistically significant at 1%.

Table 13: Comparing Model Fit Across Data Sets

		MNL	S-MNL	MIXL	G-MNL	LC	MM-MNL
Tay Sachs Disease & Cystic Fibrosis test Jewish sample (3 ASCs) T = 16; N = 210	# parameters	11	18	77	79	59	45
	LL	**-3717**	**-2815**	**-2500**	**-2480**	**-2701**	**-2620**
	AIC	7455	5666	5154	5118	5521	5330
	BIC	7523	5777	5626	**5601**	5882	5605
	CAIC	7534	5795	5703	5680	5941	**5650**
Tay Sachs Disease & Cystic Fibrosis test General population (3 ASCs) T = 16, N = 261	# parameters	11	18	77	79	83	45
	LL	**-4649**	**-3221**	**-2946**	**-2914**	**-3016**	**-3022**
	AIC	9320	6477	6047	5986	6197	6134
	BIC	9390	6591	6535	6487	6723	**6420**
	CAIC	9401	6610	6612	6566	6806	**6465**
Mobile phone (1 ASC) T = 8; N = 493	# parameters	15	17	30	32	63	61
	LL	**-4475**	**-3990**	**-3971**	**-3966**	**-3952**	**-3927**
	AIC	8980	8014	8002	7996	8030	**7976**
	BIC	9074	**8121**	8190	8197	8426	8359
	CAIC	9089	**8138**	8220	8229	8489	8420
Pizza A (No ASC) T=16; N = 178	# parameters	8	9	16	18	35	33
	LL	**-1657**	**-1581**	**-1403**	**-1373**	**-1418**	**-1328**
	AIC	3330	3179	2838	2782	2907	2722
	BIC	3378	3233	2933	**2889**	3115	2919
	CAIC	3386	3242	2949	**2907**	3150	2952
Holiday A (No ASC) T=16; N = 331	# parameters	8	9	16	18	44	33
	LL	**-3066**	**-2967**	**-2553**	**-2519**	**-2502**	**-2464**
	AIC	6149	5952	5139	5074	5092	4994
	BIC	6201	6011	5244	**5192**	5354	5211
	CAIC	6209	6020	5260	**5210**	5398	5244
Papsmear test (1 ASC) T = 32; N = 79	# parameters	6	8	12	14	34	25
	LL	**-1528**	**-1063**	**-945**	**-935**	**-958**	**-923**
	AIC	3069	2143	1914	1897	1985	1896
	BIC	3104	2189	1984	**1979**	2183	2042
	CAIC	3110	2197	1996	**1993**	2217	2067
Pizza B (No ASC) T = 32; N = 328	# parameters	16	17	32	34	101	98
	LL	**-6747**	**-6607**	**-5892**	**-5689**	**-5591**	**-5310**
	AIC	13525	13249	11849	11446	11385	10815
	BIC	13641	13372	12081	11693	12118	**11527**
	CAIC	13657	13389	12113	11727	12219	**11625**
Holiday B (No ASC) T = 32; N = 683	# parameters	16	17	32	34	152	98
	LL	**-13478**	**-13027**	**-11600**	**-11476**	**-11231**	**-11012**
	AIC	26988	26088	23263	23019	22766	22219
	BIC	27116	26224	23519	23291	23981	**23002**
	CAIC	27132	26241	23551	23325	24133	**23100**
Credit card A (2 ASCs) T = 4; N = 827	# parameters	17	21	51	53	71	69
	LL	**-3354**	**-2768**	**-2735**	**-2734**	**-2732**	**-2714**
	AIC	6742	5579	5572	5574	5606	**5567**
	BIC	6846	**5707**	5883	5898	6039	5988
	CAIC	6863	**5728**	5934	5951	6110	6057
Credit card B (3 ASCs) T = 4; N = 827	# parameters	18	25	54	56	75	74
	LL	**-4100**	**-3402**	**-3364**	**-3364**	**-3391**	**-3329**
	AIC	8236	6854	6836	6840	6933	**6806**
	BIC	8346	**7007**	7166	7182	7391	7258
	CAIC	8364	**7032**	7220	7238	7466	7332

Table 14: BIC gain of MM-MNL over G-MNL from different types of observed choice pattern from Pizza B dataset

	Attribute preferences	Totally indifferent for 'other' attributes Freq	BIC gain of MM-MNL over G-MNL	Also like some 'other' attributes Freq	BIC gain of MM-MNL over G-MNL	Also like many 'other' attributes Freq	BIC gain of MM-MNL over G-MNL
	Extremely prefer one attribute						
(1)	One of major attributes (price, fresh ingredient, crust, hot or vegetarian)	39	74	42	-31	17	-29
(2)	One of minor attributes (other attributes)	17	196	12	40	16	15
	Extremely prefer 2 attributes						
(3)	both in major attributes	9	-33	14	-7	5	-26
(4)	1 in major and 1 in minor attributes	5	29	21	10	10	-10
(5)	both in minor attributes	6	52	3	34	7	15
	Extremely prefer 3 or more attributes						
(6)	2 in major attributes			12	-32	2	-1
(7)	1 in major and 1 in minor attributes			7	6	3	-5
(8))	2 in minor attributes			1	12		
	Not extreme						
(9)	like at least 3 of major attributes	2	-11	17	-48	6	-27
(10)	like 2 of major attributes	1	-4	44	-50	10	0