A comparison of generalized multinomial logit (GMNL) and latent class approaches to studying consumer heterogeneity with some extensions of the GMNL model by Peter J. Lenk

1. Introduction

I wish to congratulate the authors on their comparison of the newly proposed generalized multinomial logit (GMNL) of Fiebig *et al.* [1], henceforth 'FKLW,' with the widely-used, latent class or finite mixture model of Kamakura and Russell [2]. Both models use the same logistic regression likelihood for product choice but differ in their distributions for individual parameter heterogeneity. I am agnostic about the 'best' model for parameter heterogeneity. Ultimately, it is an empirical issue. When models are nested, the smaller model may be preferred because of its simplicity. Conversely, even when the smaller model has better AIC or BIC, a practitioner may choose the larger model if it provides useful information for marketing strategy. When models are not nested, as in this paper, comparisons beyond fit are more nuanced because the models have different foundations and implications. Academics who propose new models tend to be more focused on improved fit, while practitioners tend to look for new models that extend the set of functional problems that can be addressed.

Aitchison and Bennet [3] introduced the probit model for product choice based on maximizing random 'enjoyment,' which has normal distributions. Partly because of limitations in computing power at the time and partly because of lack of visibility among economist, this work was superseded by the logistic random utility model of McFadden [4]. The random utility for subject i, product j, and choice occasion t is

$$\widetilde{U}_{i,j,t} = \theta_i x_{i,j,t} + \sigma_i^{-1}\varepsilon_{i,j,t}, \tag{1}$$

where θ_i is subject i's 'taste' or 'preference' parameters; σ_i is subject i's scale parameter, and $\{\varepsilon_{i,j,t}\}$ have independent and identically distributed (i.i.d.) Type I extreme value distributions. Subjects pick the product that maximizes their random utility. McFadden's derivation of the logistic model for discrete choice data remains a major contribution to econometrics. The scale and preference parameters are confounded in the likelihood function based on observed choices. A common identifying constraint sets the scale parameter to one.

FKLW's main motivation is to separate the estimation of scale heterogeneity from preference heterogeneity. Multiplying the random utilities in Equation (1) by σ_i does not alter their ordering within subject i, and results in

$$U_{i,j,t} = \beta_i x_{i,j,t} + \varepsilon_{i,j,t}, \tag{2}$$

where $\beta_i = \sigma_i \theta_i$. FKLW extends this parameterization of β_i

$$\begin{aligned} \beta_i &= \sigma_i \bar{\beta} + \gamma \eta_i + (1-\gamma)\sigma_i \eta_i \\ \sigma_i &= \exp(-\frac{\tau^2}{2} + \tau \varepsilon_{0,i}) \end{aligned}, \tag{3}$$

where $\{\eta_i\}$ have i.i.d. multivariate normal distributions with zero mean, and $\{\varepsilon_{0,i}\}$ have i.i.d. standard normal distributions. $\bar{\beta}$ is the population mean for $\{\beta_i\}$, and $0 \leq \gamma \leq 1$. The scale parameter σ_i has a log-normal distribution with mean 1, which helps with model identification. Lenk *et al.* [5] also allows for scale and taste heterogeneity in the context of metric (ratings) conjoint and detected considerable amounts of both.

FKLW nests several model specifications. When $\tau = 0$ and $\gamma = 1$, preference heterogeneity is multivariate normal, $\beta_i = \bar{\beta} + \eta_i$, the workhorse for hierarchical Bayes, discrete choice models. A more interesting case obtains when $\gamma = 0$

and $\beta_i = \sigma_i(\bar{\beta} + \eta_i)$, which is the main motivation for FKLW. This model for heterogeneity is similar to the multivariate t-distribution: $\beta_i = \bar{\beta} + \xi_i \eta_i$ where ξ_i has an inverse gamma distribution. Multivariate t-distributions are popular candidates for heterogeneity distributions [6] and are easily implemented in WINBUGS (MRC Biostatistics Unit, Cambridge, UK) or SAS's MCMC procedure (SAS Institute, Cary, NC). We shall see that multiplying the population mean by the scale parameter has a huge impact on the heterogeneity distribution.

If the variances of η_i are close to zero, then the heterogeneity in $\{\beta_i\}$ is mostly due to heterogeneity in the scale parameter and not tastes. That is, idiosyncratic factors during the purchase occasion (e.g., the shopper is rushed and does not have time to compare prices) or response tendencies (e.g., pick the product at eye-level) or store environment (e.g., the favorite brand is out of stock) not captured in $x_{i,j,t}$ drive the apparent heterogeneity in preferences.

Figure 1 plots the marginal distribution of a univariate $\beta_i = \sigma_i(\bar{\beta} + \eta_i)$ when $\bar{\beta} = 0$ and var(η_i). Figure 1 shows two cases: $\sigma_i = 1$(or $\tau = 0$), so the marginal distribution is a standard normal; and $\sigma_i \sim$ log normal with $\tau = 1$. The second case results in a long-tailed distribution that is symmetric about the mean of $\bar{\beta} = 0$. Figure 1 compares these distributions to a t-distribution with 2 DOFs. The marginal distribution for the log-normal scale is more peaked at 0 than the t-distribution and has shorter tails. Log-normal distributions, hence the marginal distribution of β_i, have finite moments of all orders, while the t-distribution only has moments of order less than the DOFs.

When $\bar{\beta}$ is not zero, the marginal distributions are not symmetric about the mean $\bar{\beta}$, as illustrated in Figure 2. The distribution is right skewed if $\bar{\beta}$ is positive and left skewed if it is negative. $\beta_i = \sigma_i(\bar{\beta} + \eta_i)$ is the convolution of a symmetric random variable $\sigma_i \eta_i$ with a skewed random variables $\sigma_i \bar{\beta}$, thus resulting in a left or right skewed distribution depending on the sign of $\bar{\beta}$. FKLW comments on the long-tailed property as a way of capturing extreme preferences, which is similar to multivariate t-distributions, but does not comment on its skewness related to $\bar{\beta}$. For some attributes, such as price, it may be reasonable to assume negative skewness; it is less clear for other attributes that the mean and skewness should
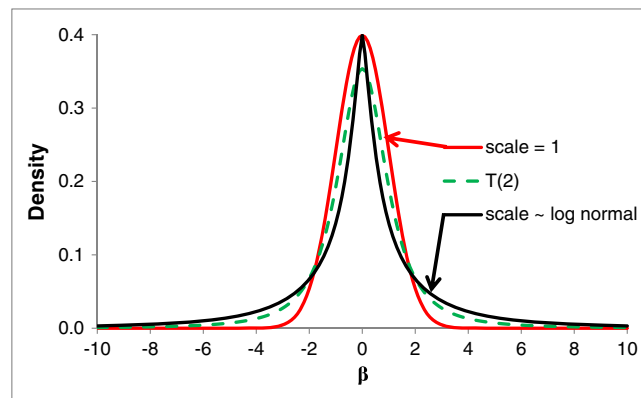

Figure 1. Marginal distribution of $\beta_i = \sigma_i(\bar{\beta} + \eta_i)$ for $\sigma_i = 1$ and $\sigma_i \sim$ log normal with $\tau = 1$; $\bar{\beta} = 0$, and var$(\eta_i) = 1$. A t-distribution (dashed line) with 2 DOFs is included as a reference.
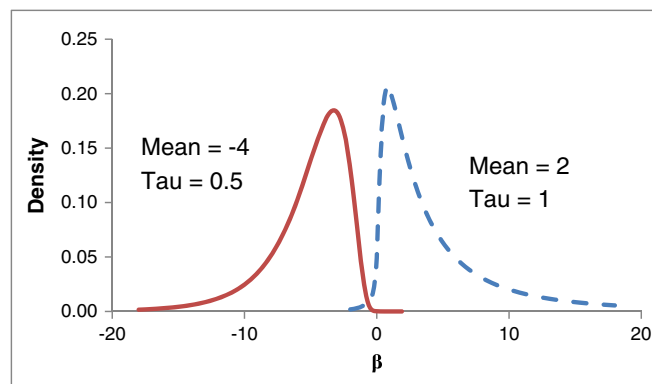

Figure 2. Marginal distribution of $\beta_i = \sigma_i(\bar{\beta} + \eta_i)$ with negative (solid) and positive (dashed) means for $\sigma_i \sim$ log normal with parameter τ, and var$(\eta_i) = 1$.

Appl. Stochastic Models Bus. Ind. **2011,** 27 580–583

581

be related. One could imagine the situation where the heterogeneity for the price coefficient is skewed, but the brand preferences are symmetric, which is the case for the empirical examples in FKLW and the current paper.

The constraint $E(\sigma_i) = 1$ identifies the population parameters of the FKLW model, but may not identify or only weakly identifies subject-level parameters. For example, one could swap scale parameters for pairs of subjects and adjust the taste parameters, thus preserving the scale heterogeneity distribution, without changing the likelihood function. Similarly, multiplying and dividing σ_i and θ_i by ψ_i, which has a log normal distribution with mean 1 and small variance, does not change the individual-level likelihood function, and $\sigma_i \psi i$ also has a log-normal distribution with mean one. It will change the distribution of heterogeneity for θ_i / ψ_i, but the extent of deformation may not be sufficient to force identification. The paper uses simulated maximum likelihood (SML), which maximizes the marginal likelihood function after integrating over heterogeneity distributions to estimate the population level parameters. SML does not require identification at the individual level. Also, FKLW pursue two remedies. On p. 399 in their paper, they set the average draw to one, not just the expectations, and in Equation (11) they do not apply the scale parameter to the intercepts (Equation (15) of the paper). This identifies the model, but loses the motivating interpretation of scale heterogeneity: σ_i no longer is the scale parameter for the error terms and merely becomes a mathematical device for creating interesting distributions. In the empirical example, there is a skewed distribution for price, and normal distributions for intercepts. A simpler model that is consistent with economic theory by forcing the price coefficient to be negative is

$$U_{i,j,k} = \beta_{i,(p)} x_{i,j,t,(p)} - \exp(\beta_{i,p}) x_{i,j,t,p} + \varepsilon_{i,j,t}, \tag{4}$$

where $x_{i,j,t,(p)}$ is the design vector without prices; $x_{i,j,t,(p)}$ is price; and $\beta_i = (\beta'_{i,(p)}, \beta_{i,p})'$ has a multivariate normal distribution.

The paper compares FKLW to latent class or finite mixtures distributions [2] where the heterogeneity distribution is

$$P(\beta_i = \beta_k) = \lambda_k \text{ for } k = 1, \ldots, K. \tag{5}$$

The coefficients for each subject are equal to one of K support points β^k. The paper assumes that the prior probabilities $\{\lambda_k\}$ of class membership are equal. One of the strong points of latent class heterogeneity is that it can approximate any distribution if the number of support points is sufficiently large. However, the number of population parameters $\{\beta^k, \lambda_k\}$ grows linearly with K. If the true heterogeneity distribution is continuous, one needs a large number of subjects to accurately approximate it with finite mixtures using large K. Conversely, if the true distribution is discrete with small or moderate K, then continuous distributions, such as FKLW, will never provide a good description. Finite mixtures of normal distributions include latent class models as a limiting case [7] if the interclass variances shrink to zero, but it is much easier to fit latent classes model with the EM (Expectation-Maximization) algorithm than finite mixtures of normal distributions with MCMC because of label switching. Gilbride and Lenk [8] noted that solutions for mixtures of normal distributions depend on the prior distribution of the interclass covariance: tight priors result in more mixture components, while less informative ones produce fewer components.

In the empirical application, the authors find that FKLW fits better than the latent class model. One interpretation of this result is that the heterogeneity is continuous and there is not enough information to fit enough support points to accurately approximate continuous heterogeneity with latent classes. Despite similar findings in other studies, practitioners find latent class models to be very useful for their purposes for a very good reason: they provide a ready-made segmentation scheme. In a perfect world, marketing researchers would combine parameter heterogeneity with profitability functions to identify segments. In the real world, this analysis is challenging because profitability functions are not well specified or only partially known, especially for competitors. Lacking profitability functions, the support points from finite mixture models are frequently interpreted as market segments. Kamakura and Russell [2] identified segments of brand loyal and brand switching customers. Segmentation is less straightforward with continuous heterogeneity. An easy approach is to classify customers who are in the lower or upper percentiles for particular attributes; however, this approach may miss segments that are defined in multiple dimensions.

The paper uses SML to estimate the population parameters then applies importance sampling (Equations (13) to (18) of the paper) to estimate subject level parameters. Bayesian statisticians used similar procedures in pre-MCMC days, but with prior distributions. The heterogeneity distribution may not be a particularly good importance sampling distribution for many subjects. It is not unusual for most of the importance sampling weights w_d^i (Equation (17)) to be near zero, which provides a diagnostic for the performance of the importance sampler. In higher dimensions, the draws of $\{\hat{\beta}_d^i\}$ may be in the tails of the posterior distribution of β_i, in which case many of the sampling weights may be non-negligible, but the importance sampling estimator $\hat{\beta}_i$ of Equation (18) is inaccurate. It seems that MCMC methods would be a more accurate method to estimate both individual and population parameters, if the researchers are willing to use prior distributions for the hyper parameters. Of course, supporters of SML will throw stones at MCMC's glass house. However, MCMC methods have been well studied for hierarchical Bayes logistic regression, and commercial software, such as SAWTOOTH

Software (Sawtooth Software, Orem, UT) and SAS, has made the glass house bullet proof. That being said, the paper's main contribution is not the estimation method. However, I am left wondering, 'Why not MCMC?'

The paper extends FKLW by incorporating state dependence (previous brand purchased) as a covariate in the utility function. In a repeat purchase setting, adding state dependence frequently improves model fit. I believe that I am not taking a big risk by saying that consumers do not literally follow the random utility model in Equation (1). Adding state dependence allows for dynamic utilities. Including previous purchases introduce long-term effects for marketing activities. Today's promotion not only effects today's purchase, but also temporarily shifts customers' utility functions, thus effecting future purchases. Estimating the long-term consequences requires simulating future buying behavior, which increases uncertainty, when using lagged purchases. Allenby and Lenk [9] pursue a different approach through autocorrelated error terms. If *post hoc* utility exceeds *ex ante* expectations, then the residual positively impacts future purchases of the brand, and the effect attenuates at a known rate on the logit scale.

The paper ends by incorporating purchase history as a covariate in the heterogeneity distribution of the scale parameter. The finding that purchase history decreases the scaling parameter in the FKLW model may be consistent with the story that more frequent purchasers become more knowledgeable about the product category. When a new shopper to a category (or an experienced shopper in a new store) is faced with a long grocery aisle of offerings (or zillions of web pages), his or her choices may be driven more by the random component of the utility model, while a seasoned shopper quickly zeros-in on the utility maximizing choice and may have a smaller random component.

New models of heterogeneity need to be compared and contrasted with existing models across a number of data sets. Their foundations and implications need careful consideration to guide practitioners in their use. The paper is a needed step in the critical comparison of FKLW to a popular and successful latent class model.

PETER LENK
Ross School of Business
University of Michigan
Ann Arbor, MI 48109, USA
E-mail: plenk@umich.edu

References

1. Fiebig DG, Kean MP, Louviere JJ, Wasi N. The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. *Marketing Science* 2010; **29**:393–421.
2. Kamakura WA, Russell GJ. A Probabilistic Choice Model for Market Segmentation and Elasticity Structure. *Journal of Marketing Research* 1989; **26**(4):389–390.
3. Aitchison J, Bennet J A. Polychotomous quantal response by maximum indicant. *Biometrika* 1970; **57**:253–262.
4. McFadden D. Conditional logit analysis of quantitative choice behavior. *Frontiers in Econometrics*, Vol. 105–142, Zarembka P (ed.). Academic Press: New York, 1974.
5. Lenk P, DeSarbo W, Green P, Young M. Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs. *Marketing Science* 1996; **15**(2):173–191.
6. Berger JO, Roberts C. Subjective Hierarchical Bayes Estimation of a Multivariate Normal Mean: on the Frequentist Interface. *The Annals of Statistics* 1990; **18**(2):617–651.
7. Lenk P, DeSarbo W. Bayesian Inference for Finite Mixtures of Generalized Linear Models with Random Effects. *Psychometrika* 2000; **65**(1):93–119.
8. Glibride T, Lenk P. Posterior Predictive Model Checking: An Application to Multivariate Normal Heterogeneity. *Journal of Marketing Research* 2010; **47**(5):896–909.
9. Allenby G, Lenk P. Modeling Household Purchase Behavior with Logistic Normal Regression. *Journal of the American Statistical Association* 1994; **83**(428):1218–1231.