A Note on the Estimation of the Multinomial Logit Model with Random Effects

Zhen CHEN and Lynn KUO *

Abstract

The multinomial logit model with random effects is often used in modeling correlated nominal polytomous data. Given that there is no standard software of fitting it, we advocate using either a Poisson log-linear model or a Poisson nonlinear model, both with random effects. Their implementations can be carried out easily by many existing commercial statistical packages including SAS. A brand choice data set is used to illustrate the proposed methods.

KEY WORDS: Discrete choice model; Multinomial logit model; Poisson log-linear model; Poisson nonlinear model; Polytomous data; Unobserved heterogeneity.

*Zhen Chen is graduate student, Department of Statistics, University of Connecticut, Storrs, CT 06269 (Email: zhen@stat.uconn.edu). Lynn Kuo is Professor, Department of Statistics, University of Connecticut, Storrs, CT 06269 (Email: lynn@stat.uconn.edu). The authors thank the editor, associated editor, and the referee for their helpful suggestions that have improved this article.

1. INTRODUCTION

In the analysis of nominal polytomous data, correlated observations are usually present. This may arise from repeated measurements taken on the same experimental unit, or from observations of experimental units which can be grouped into clusters. For example, in the panel data for a consumer behavior study, the brands chosen by a consumer are recorded over a period of time along with a set of brand attributes and possibly a set of consumer characteristics for each purchase (Chintagunta, Jain, and Vilcassim 1991; Gonul and Srinivasan 1993; Jain, Vilcassim, and Chintagunta 1994). Appropriately handling heterogeneity in the data due to clustering (repeated purchases of the same household) is essential in these studies.

We consider the multinomial logit model (cf. McFadden 1974) that is often used in nominal polytomous data. To account for heterogeneity, a random coefficient formulation has been proposed (Longford 1993; Jain, Vilcassim and Chintagunta 1994; Revelt and Train 1998). The dependence structure among the multiple observations in a cluster is modeled by using random intercepts and/or random slopes in the logit expression.

Although many applications of the multinomial logit model with random effects appear in the literature (Chintagunta, Jain, and Vilcassim 1991; Gonul and Srinivasan 1993; Jain, Vilcassim, and Chintagunta 1994; and Revelt and Train 1998), few suggest an easy way of implementing the estimation through existing software. Since a stratified proportional hazards model is routinely used to estimate the multinomial logit model (So and Kuhfeld 1995; SAS Institute 1995, pp.131-144), one might attempt to incorporate random effects into this survival model. However, the requirement of using a stratified proportional hazards model complicates the extension.

In this note, we propose two models which yield likelihoods that are equivalent (identical or proportional) to that of the multinomial logit model. Both models are constructed utilizing the well-known connection between multinomial variates and Poisson variates (e.g., McCullagh and Nelder 1989, Chapter 6; Agresti 1990, Chapter 6; Lindsey 1995, Chapter 2). As a consequence, both models can be fitted by existing software such as SAS. More importantly, SAS can also handle the estimation if random effects are introduced into the models. The first model of the two, a Poisson log-linear model, can be estimated by any generalized linear model procedures, such as SAS GENMOD (SAS Institute 2000, Chapter 29). The Poisson log-linear model with random effects can be estimated by the generalized linear mixed procedures, such as the SAS macro GLIMMIX (Littell, Milliken, Stroup, and Wolfinger, 1996). The second model is a Poisson nonlinear model. The SAS procedure NLMIXED (SAS Institute 2000, Chapter 46) can be used with or without random effects.

We organize the rest of the note as follows. Section 2 reviews the multinomial logit model with random effects. In Section 3 we first review the survival model method of estimating the multinomial logit model. Then we discuss the proposed models and extend them to incorporate random effects. We provide a yogurt brand choice example in Section 4 to illustrate our proposed methods, and conclude in Section 5. Proofs of the likelihood-equivalence among the four models (the multinomial logit model, the stratified proportional hazards model, the Poisson log-linear model, and the Poisson nonlinear model) are provided in Appendix A. SAS programs for the example are given in Appendix B.

2. THE MULTINOMIAL LOGIT MODEL WITH RANDOM EFFECTS

Consider the nominal polytomous data with multiple measurements on a subject (household). Suppose the data consist of observations of n subjects, and the i^{th} subject has T_i observations. Denote the t^{th} observation of the i^{th} subject by $(y_{it}, \boldsymbol{X}_{it})$, where y_{it} takes on any integer value from 1 to J, and $\boldsymbol{X}_{it} = (\boldsymbol{x}_{it1}, \ldots, \boldsymbol{x}_{itJ})'$ is the covariate matrix with $\boldsymbol{x}_{itj}$ a column vector associated with value j of y_{it}. In the context of consumer behavior study mentioned at the beginning of the introduction, y_{it} denotes the brand chosen by the i^{th} household at the t^{th} purchase, J is usually the total number of brands, and $\boldsymbol{x}_{itj}$ could be, say, the price and feature advertisement of the j^{th} brand at the time that the i^{th} household makes the t^{th} purchase. We assume that each household purchases only one brand at each occasion.

If heterogeneity is not a concern, then the y_{it} can be modeled as independent across all subjects and across all repeated observations, where the probability that y_{it} takes on the value j is

$$p_{itj} = \Pr(y_{it} = j) = \frac{\exp(\alpha_j + \boldsymbol{x}_{itj}^{'}\boldsymbol{\beta})}{\sum_{j=1}^{J}\exp(\alpha_j + \boldsymbol{x}_{itj}^{'}\boldsymbol{\beta})},\ j = 1, \ldots, J, \tag{1}$$

where α_j's and $\boldsymbol{\beta}$ are the parameters to be estimated from the data. The multinomial logit model with the j^{th} cell probability defined in (1) is also called the McFadden model in econometrics, because McFadden (1974) derived it from a stochastic random utility framework. The log-likelihood contribution from the i^{th} subject can be easily shown to be

$$l_i = \sum_{t=1}^{T_i}\left\{\alpha_{j^*} + \boldsymbol{x}_{itj^*}'\boldsymbol{\beta} - \log\left(\sum_{j=1}^{J}\exp(\alpha_j + \boldsymbol{x}_{itj}'\boldsymbol{\beta})\right)\right\}, \tag{2}$$

where j^* is the value that y_{it} takes on. The full log-likelihood function l is then simply the summation of (2) over i.

The above formulation of the model depends heavily on the assumption that the data within subjects are independent. This is hardly realistic in real data. To account for correlated errors caused by multiple observations for a subject, we allow some or all of the parameters in the multinomial logit model to vary randomly across subjects. This idea leads to the multinomial logit model with random effects (Jain, Vilcassim, and Chintagunta 1994; Revelt and Train 1998) where the choice probabilities for repeated purchases of a household share the same unobserved random effects. So we modify (1) by conditioning on the same household's random effects $\boldsymbol{u}_i$:

$$p_{itj|\boldsymbol{u}_i} = \Pr(y_{it} = j|\boldsymbol{u}_i) = \frac{\exp(\alpha_j + \boldsymbol{x}_{itj}^{'}\boldsymbol{\beta} + \boldsymbol{z}_{itj}^{'}\boldsymbol{u}_i)}{\sum_{j=1}^{J}\exp(\alpha_j + \boldsymbol{x}_{itj}^{'}\boldsymbol{\beta} + \boldsymbol{z}_{itj}^{'}\boldsymbol{u}_i)}, \tag{3}$$

Note we still consider the multinomial logit model except for the addition of the known design vector $\boldsymbol{z}_{itj}$, and the vector of unknown random effects, $\boldsymbol{u}_i$ in the cell probabilities. Conditional on $\boldsymbol{u}_i$, the observations from the i^{th} subject are independent. Several special cases of (3) include the random intercepts model ($\boldsymbol{z}_{itj} = 1$), the random slopes model ($\boldsymbol{z}_{itj} = \boldsymbol{x}_{itj}$), and the random intercepts and slopes model ($\boldsymbol{z}_{itj} = (1, \boldsymbol{x}_{itj}^{'})'$). It is also possible to have only a subset of slopes to be random. To finish the specification of the multinomial logit model with random effects, the $\boldsymbol{u}_i$ are assumed to be independent and identically distributed according to a distribution F with density f. Usually the multivariate normal distribution $\mathcal{N}(\boldsymbol{u}_i|\boldsymbol{0}, \boldsymbol{G})$ is used for F.

The conditional likelihood contribution of the subject i, $L_i(\cdot|\boldsymbol{u}_i)$, is the exponentiation of (2) with $\boldsymbol{x}_{itj}^{'}\boldsymbol{\beta}$ replaced by $\boldsymbol{x}_{itj}^{'}\boldsymbol{\beta} + \boldsymbol{z}_{itj}^{'}\boldsymbol{u}_i$ and $\boldsymbol{x}_{itj^*}^{'}\boldsymbol{\beta}$ replaced by $\boldsymbol{x}_{itj^*}^{'}\boldsymbol{\beta} + \boldsymbol{z}_{itj^*}^{'}\boldsymbol{u}_i$. The unconditional likelihood from the i^{th} subject is then

$$\int L_i(\cdot|\boldsymbol{u}_i) f(\boldsymbol{u}_i|\boldsymbol{G}) d\boldsymbol{u}_i. \tag{4}$$

The full unconditional likelihood function is simply the product of the above term over all subjects indexed by i.

3. ESTIMATION

3.1 Using the Stratified Proportional Hazards Model

For the multinomial logit model with repeated independent observations, a readily available method for fitting it is to use a stratified proportional hazards model. This can be done because the log-likelihood function from an appropriately designed stratified proportional hazards model can be shown to be identical to the log-likelihood function under the multinomial logit model. For completeness, we sketch a proof in Appendix A. Indeed, this method has been advocated in the literature since most of the statistical software packages have the capability of fitting the stratified proportional hazards model. In this method, the original nominal polytomous dataset is first converted to an artificial survival dataset. Specifically, for a fixed (i, t) pair, create the J observed survival times (censored or uncensored) $s_{itj} = I(y_{it} \neq j)$ for $j = 1, \ldots, J$; also construct the censoring indicators $\delta_{itj} = I(y_{it} = j)$ for each "patient" where $I(\cdot)$ is the indicator function. By convention, we let $\delta_{itj} = 0$ to indicate that the j^{th} survival time s_{itj} is right censored. To include the $J-1$ intercept terms α_j in the survival model, we need to create $J-1$ binary variables as well. It is done by stacking the $J \times (J-1)$ matrix $(\mathrm{I}_{J-1}, \mathbf{0}_{J-1})'$ for $\sum_{i=1}^{N} T_i$ times, where I_{J-1} is the $(J-1) \times (J-1)$ identity matrix and $\mathbf{0}_{J-1}$ is a $(J-1) \times 1$ vector of 0's. Now we apply a stratified proportional hazards model to the survival data (s_{itj}, δ_{itj}) with covariates $\boldsymbol{x}_{itj}$ and the newly created $J-1$ variables, using the observation number in the original polytomous dataset as the stratification variable. This estimation method can be implemented with SAS procedure PHREG (SAS Institute 2000, Chapter 49). For more details, see So and Kuhfeld (1995) and SAS Institute (1995, pp.131-144).

Although the stratified proportional hazards model can be used for the multinomial logit model with fixed effects, it is not clear how to extend it for the multinomial logit model with random effects. The usual techniques for multivariate survival models, such as the frailty model (Klein and Moeschberger, 1997, Chapter 9), do not extend to stratified survival models readily.

3.2 Using the Poisson Log-Linear Model

It is well known that a multinomial model is likelihood-equivalent to a Poisson log-linear model. See Palmgren (1981), McCullagh and Nelder (1989, Chapter 6), and Lindsey (1995, Chapter 2). To be more specific, for i from 1 to n, t from 1 to T_i, and j from 1 to J, define $w_{itj} = I(y_{it} = j)$. Suppose w_{itj} are independent Poisson variates with mean u_{itj}, $w_{itj} \sim \mathcal{P}(u_{itj})$ with u_{itj} specified by

$$u_{itj} = \exp(\theta_{it} + \alpha_j + \boldsymbol{x}_{itj}'\boldsymbol{\beta}), \tag{5}$$

where θ_{it} are the incidental parameters. The log-likelihood function for the data w_{itj} is the log-likelihood function under the multinomial logit model plus a constant. This can be shown in the same spirit as in Whitehead (1980). We provide a proof in Appendix A.

Many existing software fit the Poisson log-linear model, since it belongs to the generalized linear models family. SAS procedure GENMOD has this capability.

Before we show that the above method can be extended to handle the multinomial logit model with random effects, we pause to compare it to the stratified proportional hazards model approach.

First, the Poisson log-linear model is easier to fit. While four groups of new variables (the survival time, the censoring indicator, the strata, and the $J-1$ intercepts) are created in the survival model approach, only two (the Poisson counts and the auxiliary θ_{it} effects) need to be created in the Poisson log-linear approach. The stratified proportional hazards model approach also has the complexity of relating survival times to censoring variables and assigning larger values of survival times to right censored observations, an issue that may be confusing to users. Moreover, many empirical researchers are more familiar with generalized linear models than survival models.

Of course, the biggest advantage of the Poisson log-linear model is that it can be extended to incorporate random effects yet still enjoy the easy implementation of model fitting. Within the generalized linear model framework, it is straightforward to introduce random effects to model (5). Suppose $\boldsymbol{z}_{itj}$ is formed and $\boldsymbol{u}_i$ is a vector of random effects having multivariate normal distribution $\mathcal{N}(\boldsymbol{u}_i|\mathbf{0}, \boldsymbol{G})$ with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{G}$. Then the model with random effects indexed by $\boldsymbol{u}_i$ is

$$w_{itj} \overset{ind}{\sim} \mathcal{P}(u_{itj})$$

with

$$u_{itj} = \exp(\theta_{it} + \alpha_j + \boldsymbol{x}'_{itj}\boldsymbol{\beta} + \boldsymbol{z}'_{itj}\boldsymbol{u}_i). \tag{6}$$

This model belongs to the class of generalized linear mixed models (e.g., Wolfinger and O'Connell 1993) which can be handled by many standard statistical packages. In particular, the SAS macro GLIMMIX is designed specifically for fitting this type of models. SAS procedure NLMIXED can also fit the Poisson log-linear model, though with a different estimation method than GLIMMIX (See Section 4).

3.3 Using the Poisson Nonlinear Model

An alternative to the Poisson log-linear model exists. If $w_{itj} \overset{ind}{\sim} \mathcal{P}(\lambda_{itj})$ and the Poisson mean takes the form

$$\lambda_{itj} = \frac{\exp(\alpha_j + \boldsymbol{x}'_{itj}\boldsymbol{\beta})}{\sum_{j=1}^{J} \exp(\alpha_j + \boldsymbol{x}'_{itj}\boldsymbol{\beta})},$$

then it is easy to show the log likelihood is again equivalent to that of the multinomial logit model (a proof is provided in Appendix A). This model differs from the above log-linear model in that it is a Poisson nonlinear model. Random effects enter into this model in the same way as in the log-linear model, where the Poisson parameter takes the form

$$\lambda_{itj} = \frac{\exp(\alpha_j + \boldsymbol{x}'_{itj}\boldsymbol{\beta} + \boldsymbol{z}'_{itj}\boldsymbol{u}_i)}{\sum_{j=1}^{J} \exp(\alpha_j + \boldsymbol{x}'_{itj}\boldsymbol{\beta} + \boldsymbol{z}'_{itj}\boldsymbol{u}_i)}.$$

SAS procedure NLMIXED is designed for these types of problems. Notice that the Poisson nonlinear method is much more efficient than the Poisson log-linear model because it does not require the specification of the nuisance parameters, θ_{it}. Consequently, it has $\sum_{i=1}^{n} T_i$ fewer parameters to estimate.

4. AN EXAMPLE

We use a brand choice data set previously analyzed by Jain *et al.* (1994) to demonstrate the advocated estimation approaches for fitting the multinomial logit model with random effects by means of the Poisson log-linear model or by the Poisson nonlinear model. The original version of the data can be found at *ftp://www.amstat.org/JBES_View/94-3-JUL/jain_chinta/*, while the data used in this paper is located at *http://www.stat.uconn.edu/~zhen/estmlm_data/*. The later rearranges the variables in a different order for clarity in programming.

This data set consists of purchases of yogurt by a panel of 100 households in Springfield, Missouri over a period of about two years. The data collected by optical scanners contain information on the brand purchased, store environment variables (e.g., prices of all brands in the product category), and the marketing environment (e.g., presence or absence of newspaper feature advertisements) for each purchase made by each household in the panel. There are four brands of yogurt: Yoplait, Dannon, Weight Watchers, and Hiland, with market shares of 34%, 40%, 23% and 3%, respectively. A total of 2412 purchases are recorded for this panel during the study time. The maximum and minimum number of purchases of a household is 185 and 4, respectively, with the average at about 24.

One may use this data to study the effects of price and feature advertisement on the choice probabilities of the four brands. Let $BRAND_{it}$ denote the brand chosen by household i at the t^{th} purchase. Let $PRICE_{itj}$ and $FEATURE_{itj}$ denote the price and feature advertisement for brand j household i faces at the t^{th} purchase. We use 1, 2, 3, 4 to denote the brand Yoplait, Dannon, Weight Watchers, and Hiland, respectively. Define $CHOICE_{itj} = I(BRAND_{it} = j)$. Suppose heterogeneity is not a concern at the moment. Then the multinomial logit model can be similarly established as in Subsection 3.2:

$$\begin{aligned} &CHOICE_{itj} \sim \mathcal{P}(u_{itj}), \\ &\log(u_{itj}) = \theta_{it} + \alpha_j + \beta_1\ FEATURE_{itj} + \beta_2\ PRICE_{itj}, \qquad (7) \\ &i = 1, \ldots, 100,\ t = 1, \ldots, T_i,\ j = 1, \ldots, 4, \end{aligned}$$

where θ_{it} are the (auxiliary) incidental parameters, α_j is the intercept term corresponding to the preference effect for brand j, β's are the response coefficients for price and feature advertisement effects. Note it's necessary to have one of α_j to be 0 for the identifiability issue. We put Hiland as the reference brand in the analyses (hence $\alpha_4 = 0$). SAS procedures PHREG and GENMOD can be used to estimate these parameters as well as SAS macro GLIMMIX and SAS procedure NLMIXED. But we really do not need the latter two if GENMOD is already available. Indeed all four procedures give identical point estimates for α_j, β_1, and β_2 (θ_{it} are nuisance parameters). All procedures but NLMIXED give identical estimates for the standard errors. The standard errors from NLMIXED are slightly larger. We report the results of PHREG, GENMOD, and GLIMMIX in the second column of Table 1, and provide the SAS programs for PHREG and GENMOD in Appendices B.2 and B.3, respectively.

The nonheterogeneity assumption in the above analysis hardly reflects the reality. In brand choice studies, it is usually the case that the observations from the same household are correlated. Failing to model the dependence structure of the data will likely produce biased estimates. The multinomial logit model with random effects extends the multinomial logit models to incorporate unobserved heterogeneity as random effects. On one hand, we can build a multinomial logit model with random intercepts by letting the intercept terms be random effects. On the other hand, we can have both the intercepts and the slopes be random. The former is more parsimonious and is usually preferred in applications over the latter. For the Yogurt data, Jain *et al.* show that the random intercepts model performs as well as the model with both intercepts and slopes random. So we only consider the random intercepts model here.

Suppose the intercepts α_j are random effects and are distributed according to a trivariate normal distribution; *PRICE* and *FEATURE* are still assumed to be fixed effects. Then (7) becomes

$$\begin{aligned} \log(u_{itj}) &= \theta_{it} + \alpha_{ij} + \beta_1\ FEATURE_{itj} + \beta_2\ PRICE_{itj}, \qquad (8)\\ \boldsymbol{\alpha}_i &\sim \mathcal{N}_3\left(\overline{\boldsymbol{\alpha}}, \Sigma\right), \end{aligned}$$

where $\boldsymbol{\alpha}_i = (\alpha_{i1},\alpha_{i2},\alpha_{i3})'$ and $\overline{\boldsymbol{\alpha}} = (\overline{\alpha}_1,\overline{\alpha}_2,\overline{\alpha}_3)'$. Note (8) is a special case of (6) with $\mathbf{x}_{itj} = (FEATURE_{itj},\ PRICE_{itj})'$, $\boldsymbol{\beta} = (\beta_1, \beta_2)$, $\mathbf{z}_{itj}$ a one-dimensional column vector of value 1, and $\mathbf{u}_i$ a scalar (hence $\alpha_{ij} = \alpha_j + \mathrm{u}_i$). The SAS macro GLIMMIX can be used to estimate the mean vector $\overline{\boldsymbol{\alpha}}$, the response parameters β_1 and β_2, the variance-covariance matrix Σ, and the subject-specific preference effects α_{ij}. We provide the estimates of β_1, β_2, and $\overline{\boldsymbol{\alpha}}$ (population averages) in column 3 of Table 1. The SAS program is also provided in the Appendix B.4.

If a nonlinear Poisson model is used, then the mean of the Poisson random variable, λ_{itj}, is no longer related to the coefficient parameters linearly in the log scale; rather, it takes the following form:

$$\lambda_{itj} = \frac{\exp(\alpha_{ij} + \beta_1 \; FEATURE_{itj} + \beta_2 \; PRICE_{itj})}{\sum_{j=1}^{J} \exp(\alpha_{ij} + \beta_1 \; FEATURE_{itj} + \beta_2 \; PRICE_{itj})}.$$

Suppose $\boldsymbol{\alpha}_i = (\alpha_{i1},\alpha_{i2},\alpha_{i3})'$ is still distributed as a trivariate normal with mean vector $\overline{\boldsymbol{\alpha}} = (\overline{\alpha}_1,\overline{\alpha}_2,\overline{\alpha}_3)'$ and variance matrix Σ. Using NLMIXED, we obtain the results as listed in column 4 of Table 1 (Appendix B.5 provides the SAS program). For comparison, we also list in the last column of Table 1 the random intercepts model results of Jain *et al.* (1994).

(insert Table 1 around here)

In all four sets of results (column 2 through column 5 in Table 1), the estimates of the slope parameters have the expected signs. *PRICE* effect is negative, implying that if the price of a particular brand goes up, the chance that a household purchases it goes down. *FEATURE* effect is positive, indicating that feature advertisement of a brand tends to encourage its sale. The preference ordering of the brands in the four sets of results are the same. Yoplait is the most preferred brand by the households, followed by Dannon, Weight Watchers, and Hiland.

We see the results of NLMIXED are slightly different from those of GLIMMIX. Particularly, estimates of standard errors from NLMIXED are uniformly larger. The point estimates from NLMIXED are also larger except for the intercept term associated with the third brand (Weight Watchers). The difference between these two procedures is due to different estimation techniques used in the two algorithms. Roughly speaking, GLIMMIX iteratively fits a set of generalized estimating equations while NLMIXED directly maximizes an approximate integrated likelihood. In other words, GLIMMIX uses quasi-likelihood estimation techniques and NLMIXED uses marginal maximum likelihood estimation methods. For more on these estimation techniques, see, for example, Lindstrom and Bates (1990), Breslow and Clayton (1993) and Wolfinger and O'Connell (1993).

In comparing the above results to those of Jain *et al.*, we note that the latter tend to give larger point estimates and larger standard errors for the intercept parameters but smaller point estimates and smaller standard errors for the slope parameters, except for the variable *FEATURE* in GLIMMIX. These differences are primarily due to two different distribution assumptions of the random effects. In SAS mixed-model procedures, the vector $\boldsymbol{u}_i$ is assumed to be a realization from a multivariate normal distribution with mean vector $\mathbf{0}$ and variance covariance matrix $\boldsymbol{G}$, i.e., $\boldsymbol{u}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{G})$; In Jain *et al.*, the distribution of $\boldsymbol{u}_i$ is approximated by a discrete distribution

$\pi(\boldsymbol{u}_s)$ at $s = 1, \ldots, S$, with the support located at vectors $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_S)$ and the probability masses associated with the supports $(\pi(\boldsymbol{u}_1), \ldots, \pi(\boldsymbol{u}_S))$ estimated empirically from the data. Wedel *et al.* (1999) provide further discussions on the merits of each approach.

As has been illustrated by Jain *et al.*, the preferred model specification for this Yogurt dataset is the multinomial logit model with random intercepts. Their results show that the estimate for FEATURE variable is biased downward and the estimate for PRICE variable is biased upward in the fixed-effects model, with the magnitude of the bias in the FEATURE variable being more pronounced than that for the PRICE variable. Our results confirm their findings. For comparing the various ways of fitting the multinomial logit model with random intercepts, we provided minus log-likelihood value and the number of parameters estimated in each approach in Table 1. The minus log-likelihood values for NLMIXED is computed as the minus log likelihood reported in NLMIXED output minus the constant $\sum T_i = 2412$ (See Appendix A.3). The minus log-likelihood values for GLIMMIX is computed as the half of the deviance reported in GLIMMIX output (The log likelihood reported in GLIMMIX output is not useful in reconstructing the log likelihood for the multinomial logit model). We note that the parametric approach that GLIMMIX and NLMIXED are employing produces smaller minus log-likelihood values than the nonparametric approach of Jain *et al.* Moreover, the Poisson nonlinear model (NLMIXED) has few number of parameters to estimate. This suggests the Poisson nonlinear model is preferred to the model of Jain et al. Between the two SAS estimation methods (GLIMMIX and NLMIXED), NLMIXED is preferred since it has fewer number of parameters to estimate and with much less execution time. In terms of programming efforts, both GLIMMIX and NLMIXED statements are straightforward to write.

5. SUMMARY

In this note, we first review three methods of fitting the multinomial logit model with fixed effects. They can be carried out by fitting a stratified proportional hazards model, a Poisson log-linear model, or a Poisson nonlinear model. We conclude that both Poisson log-linear model and Poisson nonlinear model are preferred to the stratified proportional hazards model due to the simplicity of fitting them and the easiness of extending them to incorporate random effects. Consequently, the two methods provide attractive tools for empirical researchers and applied statisticians studying nominal polytomous data with clustered observations.

APPENDIX A: LIKELIHOOD-EQUIVALENCE

In this appendix, we provide proofs on the likelihood-equivalence between a stratified proportional hazards model and a multinomial logit model, between a Poisson log-linear model and a multinomial logit model, and between a Poisson nonlinear model and a multinomial logit model.

A.1 Stratified proportional hazards model

Let s_{itj} and δ_{itj} be the survival time and censoring indicator variables defined in Subsection 3.1. Consider a proportional hazards model for the data $\{s_{itj},\ \delta_{itj},\ \boldsymbol{x}^*_{itj},\ j = 1, \ldots, J\}$, where $\boldsymbol{x}^*_{itj}$ includes both $\boldsymbol{x}_{itj}$ and the $J-1$ newly created binary variables (corresponding to α_j in the multinomial logit model with one of them fixed). It is clear from the construction of s_{itj} and δ_{itj} that the total number of deaths is 1 and the risk set at this death time is the set of all J observations, i.e., $\{1, \ldots, J\}$. Hence the partial log-likelihood function is

$$\boldsymbol{x}^{*\prime}_{itj^*}\boldsymbol{\beta} - \log\left(\sum_{j=1}^{J}\exp(\boldsymbol{x}^{*\prime}_{itj}\boldsymbol{\beta})\right). \tag{9}$$

Now apply a stratified proportional hazards model to the full survival data constructed from all the (i,t) pairs and use the subject index as the stratification variable. Following Kalbfleisch and Prentice (1980), the overall log-likelihood function for this stratified proportional hazards model is simply the summation of (9) over all i and t. Thus the result is proved.

A.2 Poisson log-linear model

Household $i's$ log likelihood for the model with $w_{itj} \overset{ind}{\sim} \mathcal{P}(u_{itj})$ and $u_{itj} = \exp(\theta_{it} + \alpha_j + \boldsymbol{x}'_{itj}\boldsymbol{\beta})$ is

$$\sum_{t=1}^{T_i}\sum_{j=1}^{J}\log\left(u_{itj}^{w_{itj}}\exp(-u_{itj})\right) = \sum_{t=1}^{T_i}\left\{(\theta_{it} + \alpha_{j^*} + \boldsymbol{x}'_{itj^*}\boldsymbol{\beta}) - \sum_{j=1}^{J}\exp(\theta_{it} + \alpha_j + \boldsymbol{x}'_{itj}\boldsymbol{\beta})\right\}, \tag{10}$$

where j^* is the j such that $w_{itj} = 1$. The essence of the proof is to use the profile likelihood idea, and express the maximum likelihood estimator of θ_{it} in terms of $\boldsymbol{\beta}$ and α_j. To maximize (10) as a function of θ_{it}, we differentiate it with respect to θ_{it} and set the result equal to 0, yielding

$$1 - \sum_{j=1}^{J}\exp(\theta_{it} + \alpha_j + \boldsymbol{x}'_{itj}\boldsymbol{\beta}) = 0.$$

This implies $\sum_{j=1}^{J}\exp(\theta_{it} + \alpha_j + \boldsymbol{x}'_{itj}\boldsymbol{\beta}) = 1$. Therefore,

$$\exp(\theta_{it}) = \frac{1}{\sum_{j=1}^{J}\exp(\alpha_j + \boldsymbol{x}'_{itj}\boldsymbol{\beta})},$$

and

$$\theta_{it} = -\log\left(\sum_{j=1}^{J}\exp(\alpha_j + \boldsymbol{x}'_{itj}\boldsymbol{\beta})\right).$$

Substituting the above expressions for both $\exp(\theta_{it})$ and θ_{it} into (10) gives the log likelihood for the i^{th} household:

$$\sum_{t=1}^{T_i}\left\{\alpha_{j^*} + \boldsymbol{x}'_{itj^*}\boldsymbol{\beta} - \log\left(\sum_{j=1}^{J}\exp(\alpha_j + \boldsymbol{x}'_{itj}\boldsymbol{\beta})\right) - 1\right\}.$$

Thus household $i's$ log likelihood for the Poisson log-linear model equals to that for the multinomial logit model minus T_i, and thus they are equivalent.

A.3 Poisson nonlinear model

If $w_{itj} \overset{ind}{\sim} \mathcal{P}(\lambda_{itj})$ and

$$\lambda_{itj} = \frac{\exp(\alpha_j + \boldsymbol{x}'_{itj}\boldsymbol{\beta})}{\sum_{j=1}^{J}\exp(\alpha_j + \boldsymbol{x}'_{itj}\boldsymbol{\beta})}, \tag{11}$$

then the log-likelihood contribution from household i is

$$\sum_{t=1}^{T_i}\sum_{j=1}^{J}\{w_{itj}\log\lambda_{itj} - \lambda_{itj}\} = \sum_{t=1}^{T_i}\left\{\sum_{j=1}^{J}w_{itj}\log\lambda_{itj} - \sum_{j=1}^{J}\lambda_{itj}\right\} = \sum_{t=1}^{T_i}\{\log\lambda_{itj^*} - 1\}, \tag{12}$$

where j^* is defined as before. Upon substituting (11) with j replaced by j^* into (12), the log-likelihood function becomes

$$\sum_{t=1}^{T_i}\left\{\alpha_{j^*} + \boldsymbol{x}'_{itj^*}\boldsymbol{\beta} - \log\left(\sum_{j=1}^{J}\exp(\alpha_j + \boldsymbol{x}'_{itj}\boldsymbol{\beta})\right) - 1\right\}.$$

Therefore the log-likelihood function of household i for a Poisson nonlinear model equals again to that for a multinomial logit model minus T_i. This concludes the proof.

APPENDIX B: SAS PROGRAMMING STATEMENTS FOR THE YOGURT DATA

The original yogurt data is purchase level data where each purchase record constitutes an observation. The purchase level data contain the following variables: id, yoplait, dannon, weight, hiland, fy, fd, fw, fh, py, pd, pw, and ph. The variable id is the subject id number. The next four variables are binary (0/1) valued, indicating whether the brand (weight is for Weight Watchers) is purchased (if value 1) or not (if value 0). The variables fy to fh are features of the brands, in the order of Yoplait, Dannon, Weight Watchers, and Hiland. The final four variables, py through ph, are prices of the four brands, with the same order as that of the feature.

In the following programming statements, B.1 first constructs a brand level dataset from the purchase level data. Four new brand level observations, each with variables choice, feature, price, and brand, are created from each observation in the original purchase level data. The brand variable represents the intercepts. The data construction proceeds as follows. The new variables (in the order of choice, feature, price, and brand) are assigned the values of yoplait, fy, py, and integer 1 for the first brand level observation, the value of Dannon, fd, pd, and integer 2 for the second brand level observation, and so on for the third and fourth brand level observations. This process is repeated for each purchase level observation. The second part of B.1 creates the survival time variable, the censoring indicator variable, and the three new covariates corresponding to the three intercepts for the stratified proportional hazards model approach.

The next two parts (B.2 and B.3) fit multinomial logit model by PHREG and GENMOD. The last two parts (B.4 and B.5) fit the random intercepts multinomial logit model using GLIMMIX and NLMIXED.

B.1 Data preparation

```
data yogurt;
  infile ''s:\zhen\estmlm\yogurt.dat'';
  input id yoplait dannon weight hiland fy fd fw fh py pd pw ph;
  occasion=_n_;
  choice=yoplait;feature=fy;price=py;brand=1;output;
  choice=dannon;feature=fd;price=pd;brand=2;output;
  choice=weight;feature=fw;price=pw;brand=3;output;
  choice=hiland;feature=fh;price=ph;brand=4;output;
run;

/* the following data step is necessary only for the survival model approach */
data yogsurv;
  set yogurt;
```

```
  brandy=(brand=1); brandd=(brand=2); brandw=(brand=3);
  time=1-choice; censor=choice;
run;
```

B.2 Fitting the multinomial logit model by the stratified proportional hazards model

```
proc phreg data=yogsurv nosummary;
  strata occasion;
  model time*censor(0)=feature price brandy brandd brandw;
run;
```

B.3 Fitting the multinomial logit model by the Poisson log-linear model

```
proc genmod data=yogurt;
  class occasion brand;
  model choice=occasion feature price brand
            / noint d=poisson link=log;
run;
```

B.4 Fitting the multinomial logit model with random intercepts by the Poisson log-linear model with random intercepts

```
%glimmix (data=yogurt,
  procopt=order=data,
  stmts=%str(
    class id occasion brand;
    model choice=occasion feature price brand / noint s;
    random brand / type=un subject=id;),
  error=poisson,
  link=log)
run;
```

B.5 Fitting the multinomial logit model with random intercepts by the Poisson non-linear model with random intercepts

```
proc nlmixed data=yogurt;
  parms beta1 0 beta2 -36;
  array p[4] py pd pw ph;
  array f[4] fy fd fw fh;
  array alpha[4] alpha1-alpha4;
  alpha4=0;
  bot=0;
  do j=1 to 4;
    bot=bot+exp(beta1*f[j]+beta2*p[j]+alpha[j]);
```

```
  end;
  top=exp(beta1*f[brand]+beta2*p[brand]+alpha[brand]);
  lambda=top/bot;
  model choice ~ poisson(lambda);
  random alpha1 alpha2 alpha3
         ~normal([a1,a2,a3],[sig11,sig21,sig31,sig22,sig32,sig33]) subject=id;
run;
```

References

Agresti, A. (1990), *Categorical Data Analysis*, New York: Wiley.

Breslow, N. E., and Clayton, D. G. (1993), "Approximate Inference in Generalized Linear Mixed Models," *Journal of the American Statistical Association,* 88, 9-25.

Chintagunta, P. K., Jain, D. C., and Vilcassim, N. J. (1991), "Investigation Heterogeneity in Brand Preferences in Logit Models for Panel Data," *Journal of Marketing Research*, 28, 417-428.

Gonul, F., and Srinivasan, K. (1993), "Modeling Unobserved Heterogeneity in Multinomial Logit Models: Methodological and Managerial Implications," *Marketing Science*, 12, 213-229.

Jain, D., Vilcassim, N., and Chintagunta, P. (1994), "A Random-Coefficients Logit Brand-Choice Model Applied to Panel Data," *Journal of Business and Economic Statistics,* 12, 317-328.

Kalbfleisch, J. D., and Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data,* New York: John Wiley and Sons.

Klein, J. P., and Moeschberger, M. L. (1997), *Survival Analysis: Techniques for Censored and Truncated Data,* New York: Springer-Verlag.

Lindsey, J. K. (1995), *Modelling Frequency and Count Data*, Oxford: Clarendon Press.

Lindstrom, M. J., and Bates, D. M. (1990), "Nonlinear Mixed Effects Models for Repeated Measurements Data," *Biometrics*, 46, 673-687.

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (1996), *SAS System for Mixed Models,* Cary, NC: SAS Institute Inc.

Longford, N. T. (1993), *Random Coefficient Models,* Oxford: Clarendon Press.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models*, London: Chapman & Hall.

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

Palmgren, J. (1981), "The Fisher Information Matrix for Log Linear Models Arguing Conditionally on Observed Explanatory Variables," *Biometrika*, 68, 2, 563-566.

Revelt, D., and Train, K. (1998), "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level," *The Review of Economics and Statistics*, 80, 647-657.

SAS Institute Inc. (1995), *Logistic Regression Examples Using the SAS System, Version 6, First Edition,* Cary, NC: SAS Institute Inc.

SAS Institute Inc. (2000), *SAS/STAT User's Guide, Version 8*, Cary, NC: SAS Institute Inc.

So, Y., and Kuhfeld, W. F. (1995), "Multinomial Logit Models," in *SUGI 20 Conference Proceedings*, Cary, NC: SAS Institute Inc.

Wedel, M., Kamakura, W., Arora, N., Bemmaor, A., Chiang, J., Elrod, T., Johnson, R., Lenk, P., Neslin, S., and Poulsen, C. S. (1999), "Discrete and Continuous Representations of Unobserved Heterogeneity in Choice Modeling," *Marketing Letters,* 10, 219-232.

Whitehead, J. (1980), "Fitting Cox's Regression Model to Survival Data using GLIM," *Applied Statistics*, 29, 268-275.

Wolfinger, R., and O'Connell, M. (1993), "Generalized Linear Mixed Models: a Pseudo-Likelihood Approach," *Journal of Statistical and Computational Simulation,* 48, 233-243.

Table 1: Multinomial Logit Models for the Yogurt Data: Parameter Estimates and Their Standard Errors

Variables	*Fixed Effects Model*[1]	*Random Intercepts Model*		
		GLIMMIX[2]	*NLMIXED*[3]	*Jain et al.*[4]
Feature	0.491	0.730	0.956	0.851
	(0.120)	(0.121)	(0.185)	(0.169)
Price	-36.658	-40.012	-36.686	-43.243
	(2.437)	(2.562)	(3.725)	(2.346)
Yoplait	4.450	4.626	4.805	11.419
	(0.187)	(0.261)	(0.699)	(0.856)
Dannon	3.716	3.838	4.130	10.618
	(0.145)	(0.231)	(0.648)	(8.10)
Weight Watchers	3.074	2.242	1.046	8.108
	(0.145)	(0.241)	(0.671)	(1.075)
-Log likelihood	2,656.89	956.52	1,242.9	1,318.74
# Parameters	5	2423	11	25

[1] The fixed effects model was fit by SAS procedures PHREG, GENMOD, and SAS macro GLIMMIX, all of which give identical results.

[2] SAS macro GLIMMIX uses quasi-likelihood estimation method.

[3] SAS procedure NLMIXED uses marginal maximum likelihood estimation method.

[4] Jain *et al.* (1994) use a nonparametric procedure programmed in Gauss.