ELSEVIER Computational Statistics & Data Analysis 34 (2000) 33–50

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

A Monte Carlo EM method for estimating multinomial probit models

Ranjini Natarajan[a,*], Charles E. McCulloch[b], Nicholas M. Kiefer[c]

[a] *Department of Statistics and Division of Biostatistics, University of Florida, Gainesville, FL 32610, USA*

[b] *Department of Statistical Science and Biometrics Unit, Cornell University, Ithaca, NY 14853, USA*

[c] *Department of Economics, Cornell University, Ithaca, NY 14853, USA*

Received 1 January 1998; received in revised form 1 August 1999

Abstract

We develop a framework to perform maximum likelihood estimation in the multinomial probit model using a Monte Carlo EM algorithm. Our method includes a Gibbs step. This approach is different from likelihood procedures currently in use for this class of models, in that it does not involve direct evaluation and maximization of the observed data likelihood. Instead, we take advantage of the underlying continuum to simplify calculations. We also develop extensions of this Monte Carlo EM method for analyzing multi-period data. The computations are illustrated through real and simulated data. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Gibbs sampling; Maximum likelihood estimation; Menu pricing; Monte Carlo EM; Observed information; Panel data

1. Introduction

The multinomial probit (MNP) model for unordered categorical data has its roots in the biometrics, econometrics and transportation modeling literature (e.g., Ashford and Sowden, 1970; Hausman and Wise, 1978; Daganzo, 1980). The appeal of this model is that it can accommodate arbitrarily complicated relationships among the probabilities of the various categories quite naturally. However, despite potential broad applications, its usage has been limited by the computational burden associated

* Corresponding author.

PII: S 0167-9473(99)00073-0

with estimating the parameters. Evaluation of the likelihood function for this model requires a method to compute multi-Normal orthant probabilities. This calculation is difficult unless the dimension of the multi-Normal distribution is less than five, or the variance matrix has a special structure (Hajivassiliou et al., 1992).

Current approaches to calculate maximum likelihood estimates (MLEs) for the MNP model involve numerically maximizing a simulation-based estimate of the likelihood function (Lerman and Manski, 1981; Geweke, 1989; Stern, 1992; Börsch-Supan and Hajivassiliou, 1994). While this strategy works reasonably in many problems, it appears to place a misplaced focus on the accuracy and precision of the simulator, rather than efficient maximization using such a simulator. Some work has been done on developing alternative estimation methods. McFadden (1989) proposed the method of simulated moments (MSM) which involves substitution of simulated orthant probabilities into moment conditions. The computational effort required for this method grows quite rapidly relative to the size of the problem, especially for multi-period (panel) data. Keane (1994) suggested a computationally feasible variant of the MSM approach by factoring the orthant probabilities into transition probabilities. Albert and Chib (1993) and McCulloch and Rossi (1994) have described a Bayesian analysis of the MNP model using the Gibbs sampler. An important and worrisome issue that arises in this context is prior selection in the absence of well-formulated subjective information about the parameters. Diebolt and Ip (1998, p. 259) discuss the use of a stochastic EM algorithm to estimate the parameters of a binary probit model. While their approach is related to ours, it is different in that the resulting point estimates do not in general produce maximum likelihood estimates.

We describe maximum likelihood estimation in the MNP model via a Monte Carlo EM (MCEM) algorithm (Tanner, 1993, p. 52; McCulloch, 1994). Our motivation for using this approach is three-fold: (i) it circumvents direct evaluation of the likelihood function; (ii) the iterates are automatically confined to lie in the parameter space; and (iii) it can be readily adapted for other limited dependent variable models, such as the multinomial logit or Tobit regression models. This paper is organized as follows. The MNP model is formulated in Section 2. In Section 3 we describe the MCEM algorithm for calculating MLEs, while standard error estimation is discussed in Section 4. We compare our MCEM approach with the smooth simulated maximum likelihood (SSML) approach advocated by Börsch-Supan and Hajivassiliou (1994) through a simulated example in Section 5. Our results show that MCEM approaches the neighborhood of the MLE more quickly than SSML. On the other hand, given good starting values, the SSML method converges more accurately. This may suggest a hybrid algorithm, beginning with a few iterations of MCEM to get "close" to the MLE, followed by SSML to refine the estimate. We illustrate the use of our MCEM algorithm in Section 6 by analyzing actual consumer data on the quantity demanded of several menu items at a medium-priced family restaurant (Kiefer et al., 1994). Extensions to panel data are described in Section 7, and demonstrated on the household purchase of peanut butter (ERIM panels, A. C. Nielsen) in Section 8. Finally, Section 9 offers some concluding remarks.

2. The model

Assume that N agents/individuals choose among a set of c choices. The observed data for agent i is a multinomial vector $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{ic})^{\mathrm{T}}$. Each component of $\boldsymbol{w}_i$ is binary and specifically $w_{ij} = 1$ if agent i makes the choice j and zero otherwise. Further, it is assumed that the set of c choices are discrete so that $\sum_{j=1}^{c} w_{ij} = 1$. The MNP model arises by postulating the existence of a latent continuum $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{ic})^{\mathrm{T}}$ which generates the observed $\boldsymbol{w}_i$ in the following manner:

$$w_{ij} = I\left(u_{ij} = \max_k \{u_{ik}\}\right), \tag{1}$$

and

$$\boldsymbol{u}_i \sim N(\boldsymbol{X}_i^* \boldsymbol{\beta}^*, \boldsymbol{\Omega}^*), \tag{2}$$

where $I\ (.)$ denotes the indicator function, $\boldsymbol{X}_i^*$ is a known $c \times p$ design matrix of exogenous variables, $\boldsymbol{\beta}^*$ is an unknown $p \times 1$ vector of fixed effects, and $\boldsymbol{\Omega}^*$ is a $c \times c$ variance–covariance matrix.

It is typical to regard the $\boldsymbol{u}_i$ as the unmeasured utility or value of the c choices to the ith individual. Thus, model (1) suggests that an agent picks the choice that has the largest utility to them. It is unimportant whether we actually believe in these underlying $\boldsymbol{u}_i$ or simply use it as a mechanism for estimation. However, in practice there are numerous applications where the existence of the $\boldsymbol{u}_i$ can be easily justified.

The choice model as stated in (1) and (2) is not identified (Dansie, 1985). In order to achieve identification, it is conventional to reformulate it in terms of the $(c-1)$ relative differences $y_{ij} = u_{ij} - u_{ic}$. It can be seen that $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ic-1})^{\mathrm{T}}$ satisfies

$$\boldsymbol{y}_i \sim N(\boldsymbol{X}_i \boldsymbol{\beta}, \boldsymbol{\Omega}),$$

where $\boldsymbol{X}_i$, $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$ are the appropriate transformations of $\boldsymbol{X}_i^*$, $\boldsymbol{\beta}^*$ and $\boldsymbol{\Omega}^*$ respectively. Further, Eq. (1) may be re-expressed as $w_{ij} = I\ (y_{ij} > 0,\ y_{ij} = \max_k\{y_{ik}\})$ when $j = 1, \ldots, c-1$ and $w_{ic} = I\ (y_{ik} < 0,\ \forall k)$. Since the scale of the relative utilities y_{ij} is indeterminate, we set the first diagonal element of $\boldsymbol{\Omega}$ equal to 1 (e.g., Geweke et al., 1994).

2.1. The estimation problem

The focus of this paper is on ML estimation of $\boldsymbol{\beta}$ and Ω. The MNP likelihood is given by

$$L(\boldsymbol{\beta}, \boldsymbol{\Omega} | \boldsymbol{w}_1, \ldots, \boldsymbol{w}_N) = \prod_{i=1}^{N} \prod_{j=1}^{c} \pi_{ij}^{w_{ij}},$$

where π_{ij} is the probability that individual i makes choice j. For each i, these π_{ij} satisfy the constraints:

$$\pi_{ij} > 0, \quad \sum_{j=1}^{c} \pi_{ij} = 1,$$

and are given by the multi-Normal orthant probabilities

$$\pi_{ij} \propto \int \exp\left(-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right) I\left(y_{ij} > 0,\ y_{ij} = \max_k \{y_{ik}\}\right) \mathrm{d}\boldsymbol{y}_i,$$

when $j = 1,\ldots,c-1$. The above integral does not have a closed form except for special forms of $\boldsymbol{\Omega}$. We develop a Monte Carlo EM approach to estimate the MLEs of $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$. This approach circumvents direct evaluation and maximization of the observed likelihood L by taking advantage of the latent structure of the model.

3. Maximum likelihood estimation

The Expectation Maximization (EM) algorithm (Dempster et al., 1977) is a powerful numerical tool used for computing ML estimates in standard incomplete data problems. The basic premise is that the maximization based on the observed (incomplete) data is computationally intractable. However, by augmenting the observed data, the hard maximization can be reduced into a sequence of easier problems. We first review the EM algorithm in general and then describe the specifics for the MNP model.

3.1. Review of the EM algorithm

Following the usual EM terminology, we let $\boldsymbol{z} = (\boldsymbol{y}, \boldsymbol{w})$ denote the complete data, where $\boldsymbol{w}$ is observed and $\boldsymbol{y}$ is missing data. We assume that $\boldsymbol{z}$ is indexed by a d-dimensional parameter $\boldsymbol{\theta}$, and the goal is to find the MLE of $\boldsymbol{\theta}$. If $\boldsymbol{z}$ were observed, the objective would be to maximize $\ell(\boldsymbol{\theta}|\boldsymbol{z}) = \ln L(\boldsymbol{\theta}|\boldsymbol{z})$. However, since only the $\boldsymbol{w}$ are observed, we need to maximize:

$$\begin{aligned}\ell(\boldsymbol{\theta}|\boldsymbol{w}) &= \ln L(\boldsymbol{\theta}|\boldsymbol{w}),\\ &= \ln \int L(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{w})\, \mathrm{d}\boldsymbol{y}.\end{aligned}$$

It is the integration which can make the maximization of the observed data log-likelihood tedious, even when maximizing the complete data log-likelihood is trivial.

The EM algorithm maximizes $\ell(\boldsymbol{\theta}|\boldsymbol{w})$ by iteratively maximizing $E(\ell(\boldsymbol{\theta}|\boldsymbol{z}))$. Each iteration has 2 steps: an Expectation-step and a Maximization-step. The $(m+1)$st E-step computes

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) = E(\ell(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{w})),$$

where the expectation is with respect to the conditional density $f(\boldsymbol{y}|\boldsymbol{w}, \boldsymbol{\theta}^{(m)})$. The $(m+1)$st M-step then finds $\boldsymbol{\theta}^{(m+1)}$ to maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$. Although this algorithm works quite generally for any model, it is particularly useful when the complete data are from an exponential family, since then the E-step merely reduces to finding the complete-data sufficient statistics.

Sometimes the computations required for the E-step are hefty. In such cases a Monte Carlo estimate can be obtained by estimating $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$ with

$$\frac{\sum_{r=1}^{R} \ell(\boldsymbol{\theta}|\boldsymbol{y}^{(r)}, \boldsymbol{w})}{R},$$

where $\boldsymbol{y}^{(r)} \sim f(\boldsymbol{y}|\boldsymbol{w}, \boldsymbol{\theta}^{(m)})$, $r = 1, \ldots, R$. This leads to a Monte Carlo EM method.

3.2. MCEM for the MNP model

We use the EM algorithm with the following definitions. We regard the vector of relative utilities $\boldsymbol{y}_i$ as the complete data. The complete data log-likelihood is therefore a product of multi-Normal densities given by

$$\ell(\boldsymbol{\beta}, \boldsymbol{\Omega}|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N) \propto -\frac{N}{2} \ln \det(\boldsymbol{\Omega}) - \frac{1}{2} \operatorname{tr}\left(\boldsymbol{\Omega}^{-1} \sum_{i=1}^{N} \boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{T}}\right),$$

where $\boldsymbol{e}_i = \boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$. The E-step is conceptually simple and requires the calculation of

$$\begin{aligned} Q_i(\boldsymbol{\beta}|\boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)}) &= E(\boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{T}}|\boldsymbol{w}_i, \boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)}), \\ &= \sigma_i^{2(m)} + (\mu_i^{(m)} - \boldsymbol{X}_i\boldsymbol{\beta})(\mu_i^{(m)} - \boldsymbol{X}_i\boldsymbol{\beta})^{\mathrm{T}}, \end{aligned}$$

where $\sigma_i^{2(m)} = \mathrm{Var}(\boldsymbol{y}_i|\boldsymbol{w}_i, \boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)})$ and $\mu_i^{(m)} = E(\boldsymbol{y}_i|\boldsymbol{w}_i, \boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)})$. The M-step is also straightforward. We maximize

$$-\frac{N}{2} \ln \det(\boldsymbol{\Omega}) - \frac{1}{2} \operatorname{tr}\left(\boldsymbol{\Omega}^{-1} \sum_{i=1}^{N} Q_i(\boldsymbol{\beta}|\boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)})\right)$$

with respect to $\boldsymbol{\beta}$ and the $(c(c-1)//2 - 1)$ parameters in $\boldsymbol{\Omega}$. It is well known that the joint maximizing values of $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$ are not available in closed form. Thus, we adopt Meng and Rubin's (1993) suggestion of replacing this M-step with 2 conditional M-steps: the first is a maximization with respect to $\boldsymbol{\beta}$ conditional on the elements of $\boldsymbol{\Omega}$, and the second is a maximization over the unknown elements of $\boldsymbol{\Omega}$ conditional on the updated value of $\boldsymbol{\beta}$. This results in a Monte Carlo Expectation Conditional Maximization (MCECM) algorithm. However, in the rest of this paper, we will continue to refer to our approach as a MCEM method. The conditional ML estimate of $\boldsymbol{\beta}$ is simply the generalized least squares estimator:

$$\boldsymbol{\beta}^{(m+1)} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\Omega}^{(m)-1} \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\Omega}^{(m)-1} \mu_i^{(m)}\right),$$

while the conditional ML estimate of $\boldsymbol{\Omega}$ is obtained by maximizing

$$-\frac{N}{2} \ln(\det(\boldsymbol{\Omega})) - \frac{1}{2} \operatorname{tr}\left(\boldsymbol{\Omega}^{-1} \sum_{i=1}^{N} Q_i(\boldsymbol{\beta}^{(m+1)}|\boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)})\right) \tag{3}$$

with respect to the parameters in $\boldsymbol{\Omega}$. Although the maximization of (3) requires an iterative procedure, it is a standard calculation and would need to be performed even if the $\boldsymbol{y}_i$ were observed. Non-linear functions such as (3) are relatively well-studied and there are several extremely efficient algorithms to perform the maximization (Jennrich and Schluchter, 1986). For the examples described in this paper, we used the non-linear optimization routine OPTMUM (Aptech Systems, 1994) to maximize (3).

Thus, we have shown that the only additional computations required for ML estimation of discrete choice data is the calculation of $\mu_i^{(m)}$ and $\sigma_i^{2(m)}$. We are now ready to make a formal statement of the MCEM algorithm.

The Algorithm.

Step 0: Obtain starting values $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\Omega}^{(0)}$. Set counter $m = 0$.

Step 1: (E-step) For each i, calculate $\mu_i^{(m)}$ and $\sigma_i^{2(m)}$

Step 2a: (Conditional M-step 1) Set:

$$\boldsymbol{\beta}^{(m+1)} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{\Omega}^{(m)-1} \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{\Omega}^{(m)-1} \mu_i^{(m)} \right)$$

Step 2b: (Conditional M-step 2) Maximize:

$$-\frac{N}{2} \ln(\det(\boldsymbol{\Omega})) - \frac{1}{2} \operatorname{tr} \left(\boldsymbol{\Omega}^{-1} \sum_{i=1}^{N} Q_i(\boldsymbol{\beta}^{(m+1)} | \boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)}) \right)$$

over the unknown elements of $\boldsymbol{\Omega}$ to obtain $\boldsymbol{\Omega}^{(m+1)}$.

Step 3: If convergence is reached set $\hat{\boldsymbol{\beta}}_{\text{MCEM}} = \boldsymbol{\beta}^{(m+1)}$ and $\hat{\boldsymbol{\Omega}}_{\text{MCEM}} = \boldsymbol{\Omega}^{(m+1)}$; else increment counter m by one and return to Step 1).

Most of the computational effort is expended in computing the conditional means and variances of $\boldsymbol{y}_i$ given the observed data $\boldsymbol{w}_i$. For small problems, and simple covariance structures $\boldsymbol{\Omega}$ these may be computed using direct numerical integration. However, for more complicated models we propose the use of the Gibbs sampler (Geman and Geman, 1984) to estimate them. More details on this will be discussed in Section 3.4. An important consideration in implementing Monte Carlo EM is the assessment of convergence. We will now discuss this issue.

3.3. *Convergence of MCEM*

The convergence of MCEM can be monitored by plotting the parameter value at each iteration versus the iteration number. After a certain number of iterations, the plot will reveal random fluctuation about the MLE, due to the randomness introduced by the Monte Carlo E step (see Figs. 1–3).At this point one may either terminate the algorithm, or continue with a large number of Gibbs samples to decrease the Monte Carlo variability. Chan and Ledolter (1995) provide a stopping criterion as well as rules for selecting the appropriate Monte Carlo sample size. However, their method requires repeated evaluation of the likelihood function, which proves too computationally intensive and impractical for the models considered here. More recently, Booth and Hobert (1999) have proposed an automated MCEM algorithm for generalized linear mixed models that allows the Monte Carlo error at each iteration to be assessed using central limit theory and Taylor series approximations. However,

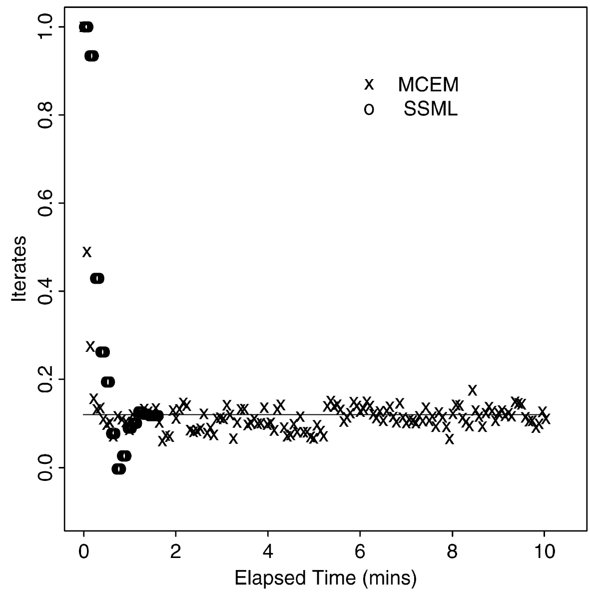

Fig. 1. Convergence of MCEM and SSML for β when $c=3$ and $\boldsymbol{\Omega}=\boldsymbol{\Omega}_1$. The solid line represents the true MLE.

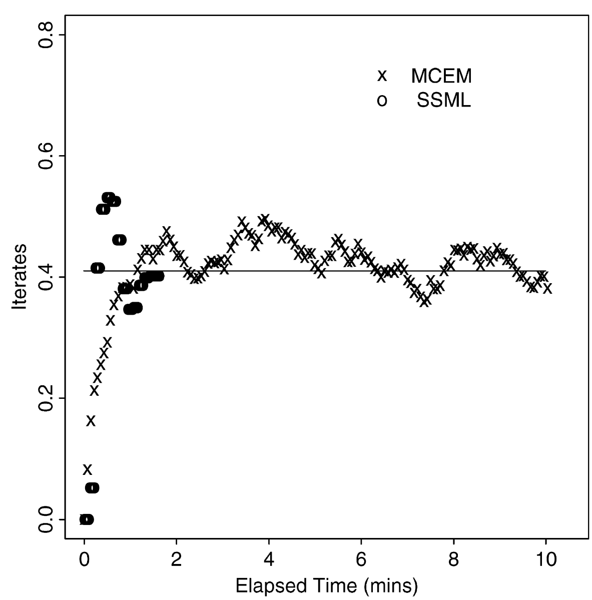

Fig. 2. Convergence of MCEM and SSML for ω_{12} when $c=3$ and $\boldsymbol{\Omega}=\boldsymbol{\Omega}_1$. The solid line represents the true MLE.

the method they propose may break down when the intractable integrals in the likelihood function are of high dimension. For the examples described in this paper, we monitored convergence by investigating plots of the parameter iterates and terminated the algorithm when the iterates appeared to fluctuate randomly.

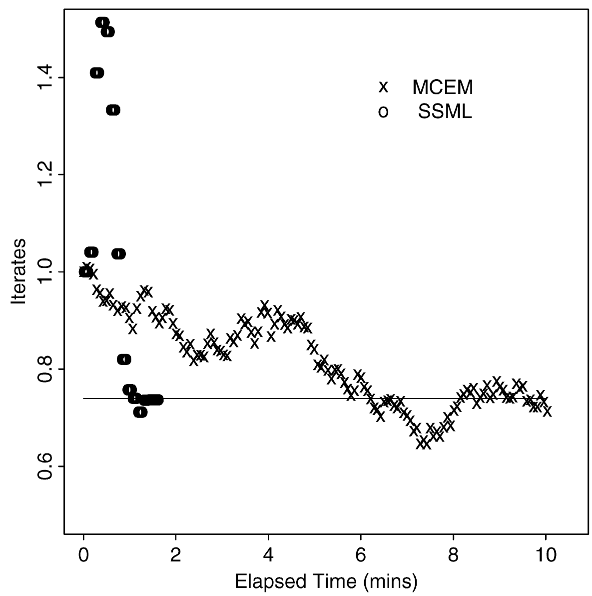

Fig. 3. Convergence of MCEM and SSML for ω_{22} when $c = 3$ and $\boldsymbol{\Omega} = \boldsymbol{\Omega}_1$. The solid line represents the true MLE.

We will now discuss the estimation of the expectations involved in the E-step.

3.4. The Gibbs sampler

The computational difficulties associated with calculating the moments of the conditional distribution $f(\boldsymbol{y}_i|\boldsymbol{w}_i)$ can be burdensome, since it involves integrals similar to those that appear in the likelihood function. Thus, we propose the use of the Gibbs sampler to provide a simulation-based estimate of these moments. McCulloch (1994) developed a similar Monte Carlo EM algorithm for binary probit models. The efficient implementation of the Gibbs sampler rests on the fact that fast acceptance-rejection algorithms exist to generate from truncated univariate Normal distributions (e.g., inverse transform method, Devroye, 1986).

In order to generate a sample from the density $f(\boldsymbol{y}_i|\boldsymbol{w}_i)$ using the Gibbs sampler, we need to cycle through the full conditional specifications $f(y_{ij}|\boldsymbol{y}_{i(-j)}, \boldsymbol{w}_i)$, where $\boldsymbol{y}_{i(-j)} = (y_{i1}, \ldots, y_{ij-1}, y_{ij+1}, \ldots, y_{ic-1})^{\mathrm{T}}$. Using standard results on Normal theory, it can be shown that these conditional densities are simply univariate truncated normal distributions (Searle et al., 1992, p. 464). The following is an outline of how the Gibbs sampler is used to generate a sample of $\boldsymbol{y}_i$ from the conditional distribution of $f(\boldsymbol{y}_i \mid \boldsymbol{w}_i)$ for a fixed i.

(a) Obtain starting values for $\boldsymbol{y}_i$. For the examples described in this paper, we intialized the chain at the zero vector.

(b) For each $j = \{1, \ldots, (c-1)\}$ calculate

$$\beta_{j|-j} = \mathrm{Cov}(y_{ij}, \boldsymbol{y}_{i(-j)}),$$

$$\sigma^2_{j|-j} = \mathrm{Var}(y_{ij}|\boldsymbol{y}_{i(-j)}),$$

$$= (\mathrm{Var}(y_{ij}) - \beta_{j|-j}\,\mathrm{Var}^{-1}(\boldsymbol{y}_{i(-j)})\beta^{\mathrm{T}}_{j|-j})^{-1}.$$

These are standard calculations for normally distributed variates.

(c) For each $j = \{1, \ldots, (c-1)\}$ calculate

$$\mu_{ij|i(-j)} = E(y_{ij}|\boldsymbol{y}_{i(-j)}),$$

$$= \boldsymbol{x}^{\mathrm{T}}_{ij}\boldsymbol{\beta} + \beta_{j|-j}\,\mathrm{Var}^{-1}(\boldsymbol{y}_{i(-j)})(\boldsymbol{y}_{i(-j)} - \boldsymbol{X}_{i(j)}\boldsymbol{\beta}),$$

where $\boldsymbol{X}_{i(j)} = \boldsymbol{X}_i$ with row j deleted and $\boldsymbol{x}^{\mathrm{T}}_{ij}$ is the jth row of $\boldsymbol{X}_i$.

(d) Simulate y_{ij} from a normal distribution with mean $\mu_{ij|i(-j)}$ and variance $\sigma^2_{j|-j}$ truncated in the following manner: if $w_{ij} = 1$, simulate y_{ij} truncated above $\max\{0, \max\{\boldsymbol{y}_{i(-j)}\}\}$; else simulate y_{ij} truncated below $\max\{0, \max\{\boldsymbol{y}_{i(-j)}\}\}$.

Repeat Steps (c) and (d) a large number of times, say M, to obtain $\boldsymbol{y}_i^{(1)}, \boldsymbol{y}_i^{(2)}, \ldots, \boldsymbol{y}_i^{(M)}$. Discard a suitable number from the beginning of the sequence, *nburn*, and then accept every *nskip*th one to form a sample of size *nrep*. This sample is then used to estimate $E(\boldsymbol{y}_i|\boldsymbol{w}_i)$ and $\mathrm{Var}(\boldsymbol{y}_i|\boldsymbol{w}_i)$. There are no hard and fast rules on the choice of *nburn*, *nrep* and *nskip*. On account of the iterative nature of EM and the desire to take as few Gibbs samples as possible at the beginning of EM, it is usually efficient let these numbers depend on the iteration of EM; i.e., later iterations do more Gibbs sampling.

It is evident that this approach can accommodate arbitrarily complicated covariance structures since they only affect Step (a), which is performed only once before initiating the Gibbs chain. We will now discuss the calculation of standard errors for the ML estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$.

4. Standard error calculations

In this section we describe a Monte Carlo approach to Louis' method (1982) to estimate the asymptotic standard errors of the MLE of the MNP model. Guo and Thompson (1992) outline a similar method for genetic models.

4.1. Louis's method

Louis developed a technique to compute the observed information matrix within the EM framework. It requires computation of the complete-data gradient vector and second derivative matrix and can be embedded quite simply in the EM iterations. In order to describe his method we use the notation developed in the review of EM in Section 3.1. Louis proved that the observed information matrix $I_w(\boldsymbol{\theta})$ satisfies the following identity:

$$I_w(\boldsymbol{\theta}) = -E(H(\boldsymbol{\theta}|\boldsymbol{y})|\boldsymbol{w}) - \mathrm{Var}(S(\boldsymbol{\theta}|\boldsymbol{y})|\boldsymbol{w}), \tag{4}$$

where S and H are the complete data score vector and Hessian matrix respectively. Expressions for the elements of S and H are given in Appendix A. Of course, they need be evaluated only on the last iteration of EM at the MLE.

Despite the conceptual simplicity and elegance of Louis' identity, this method has not been used extensively due to computational difficulties in evaluating the expectations involved. A computationally feasible variant of Louis's identity can be obtained by replacing all the expectations in (4) by their Monte Carlo estimates, in the following manner:

(1) Generate $\boldsymbol{y}^{(r)} \sim f(\boldsymbol{y}|\boldsymbol{w},\boldsymbol{\theta}), r = 1,\ldots,R$.

(2) Replace each term in (4) by its Monte Carlo estimate, e.g., replace the first term by $-(1/R)\sum_{r=1}^{R} H(\boldsymbol{\theta}|\boldsymbol{y}^{(r)},\boldsymbol{w})$.

5. Simulation study

This section reports results of simulations run to compare the performance of the MCEM algorithm with the SSML approach of Börsch-Supan and Hajivassiliou (1994). We considered a MNP model with $c=\{3,5\}$ choices and a single fixed effect. More formally, the latent variables $\boldsymbol{y}_i$ were assumed to arise from a multi-normal distribution with mean $\boldsymbol{X}_i\beta$ and an unrestricted error variance matrix $\boldsymbol{\Omega}$, where $\boldsymbol{X}_i = (x_{i1},\ldots,x_{ic})^{\mathrm{T}}$ and $x_{i1},\ldots,x_{ic}$ were drawn independently from a uniform distribution on $(-0.5, 0.5)$. The true value of β was taken to be zero and two choices were considered for $\boldsymbol{\Omega}$: $\boldsymbol{\Omega}_1 = 0.5\boldsymbol{I} + 0.5\boldsymbol{J}$ and $\boldsymbol{\Omega}_2 = 0.3\boldsymbol{I} + 0.7\boldsymbol{J}$, where $\boldsymbol{I}(c \times c)$ is the identity matrix and $\boldsymbol{J}(c \times c)$ is a matrix of ones. The first choice of $\boldsymbol{\Omega}$ results in a correlation of 0.5 among the latent utilities while the second results in a correlation of 0.7. For each of the four combinations of c and $\boldsymbol{\Omega}$, a 500-observation data set was generated. The observed data likelihood function is not too difficult to calculate for this example and may be evaluated using numerical integration. It is therefore a good situation to compare MCEM and SSML.

Table 1 compares the performance of these methods in terms of proximity to the true MLE (obtained by direct numerical maximization of the likelihood) after 30 iterations. The numbers displayed for MCEM and SSML (with a simulation sample of size 10 to estimate the likelihood) are the average estimate over 25 independent runs of each of these methods. In order to make the MCEM algorithm comparable with SSML, we did not allow *nrep* to vary with iteration but instead set it equal to 10. Both methods used the same starting values, namely, $\beta = 1$ and $\boldsymbol{\Omega}$ equal to the identity matrix.

Two overall conclusions may be drawn from our simulation study and these are described below:

1. MCEM and SSML are competitive with one another for the 3-choice models, with the SSML approach appearing to have a slight edge over MCEM both in terms of computational time and accuracy. Inspection of the MCEM iterates (not reported in the table) shows that they approach a neighborhood of the MLE relatively quickly, but continue to bounce around the MLE even after a fairly long. Figs.

Table 1
MCEM and SSML iterates after 30 iterations

Method	β	ω_{21}	ω_{22}	ω_{31}	ω_{32}	ω_{33}	ω_{41}	ω_{42}	ω_{43}	ω_{44}	Time (min)
					$c = 3,\ \boldsymbol{\Omega} = \boldsymbol{\Omega}_1$						
SSML	0.119	0.400	0.727								1.71
MCEM	0.126	0.412	0.819								2.36
MLE	0.119	0.406	0.737								
					$c = 3,\ \boldsymbol{\Omega} = \boldsymbol{\Omega}_2$						
SSML	0.056	0.751	1.234								1.90
MCEM	0.063	0.716	1.318								2.31
MLE	0.055	0.748	1.225								
					$c = 5,\ \boldsymbol{\Omega} = \boldsymbol{\Omega}_1$						
SSML	0.036	1.232	2.240	0.066	1.463	3.288	−0.126	0.098	1.053	0.832	32.79
MCEM	0.004	0.526	1.139	0.517	0.631	1.216	0.533	0.678	0.642	1.338	7.38
MLE	−0.041	0.896	1.699	0.307	1.008	1.160	0.466	0.756	0.451	1.370	
					$c = 5,\ \boldsymbol{\Omega} = \boldsymbol{\Omega}_2$						
SSML	0.019	1.210	2.500	0.463	1.890	2.853	−0.247	0.585	1.500	1.303	31.45
MCEM	0.111	0.679	1.123	0.711	0.841	1.405	0.700	0.730	0.869	1.168	7.39
MLE	0.060	0.919	1.630	0.732	1.386	1.818	0.507	0.778	0.924	1.222	

Note: The numbers reported are the average of 25 independent runs of each method.

1–3 display this behavior for β and the elements of $\boldsymbol{\Omega}$ when $c = 3$ and $\boldsymbol{\Omega} = \boldsymbol{\Omega}_1$. The number of Gibbs samples would have to increase drastically to achieve higher accuracy. Note that this is not done in the results displayed.

2. The comparative performance of SSML dramatically worsens for the 5-choice problem. While estimates of β may still be accurately obtained using either method, the same is not true for estimates of $\boldsymbol{\Omega}$. For example, the average squared-error distance of the SSML estimate of $\boldsymbol{\Omega}$ (based on the 25 independent runs) from the true MLE is 15.51 under $\boldsymbol{\Omega}_1$ and 7.53 under $\boldsymbol{\Omega}_2$, while the corresponding numbers for MCEM are 5.72 and 1.00 respectively (not shown in the table). Further, the speed of the SSML algorithm is significantly reduced as evidenced by the four-fold increase in computational time compared with MCEM.

We re-ran the SSML algorithm for the 5-choice model using starting values which were closer to the true MLE, and noted an improvement in performance. This suggests a hybrid MCEM and SSML implementation, with a few initial rounds of MCEM to identify good starting values followed by SSML for more accurate convergence.

6. Menu pricing data

We now illustrate our MCEM method on data gathered on the quantity demanded of several menu items at a medium-priced family restaurant offering a range of

Table 2
Purchase frequencies of items from the menu-pricing data

Menu item	Frequency of demand
Steak and other seafood	214
Fish fry	214
Pinesburger	419
Tullyburger	127

entrées including steak, seafood as well as specialty sandwiches (Kiefer et al., 1994). We are interested in studying the demand of four menu items: fish fry, steak and other seafood, Pinesburger and the Tullyburger. Together these items account for more than 65% of the total items ordered. Kiefer et al. (1994) present a subset of this data on the demand for fish fry, a popular item priced at \$8.95. Of special interest in their study was the effect of changes in price on the demand for fish fry. Four price levels were experimented with (\$8.95, \$9.50, \$9.95, \$10.95). However, actual checks were presented at original price levels so that the customer did not suffer any monetary loss. The data were collected over four winter weekends (Friday/Saturday).

Table 2 displays the menu items under study along with the sample frequency of demand over the four weekends, for a total of 974 orders.

We fit an MNP model to these data using fish fry as the reference item and with nine fixed effects parameters: three item-specific intercepts, three day-specific intercepts and the effect of price of fish fry on demand. Thus, the latent model for the relative utilities is (dropping the subscript for individual) given by

$$\begin{aligned}
y_{\text{steak}} &= \beta_{\text{steak}} + x_1\gamma_{\text{steak}} + x_2\delta_{\text{steak}} + \varepsilon_{\text{steak}},\\
y_{\text{pines}} &= \beta_{\text{pines}} + x_1\gamma_{\text{pines}} + x_2\delta_{\text{pines}} + \varepsilon_{\text{pines}},\\
y_{\text{tully}} &= \beta_{\text{tully}} + x_1\gamma_{\text{tully}} + x_2\delta_{\text{tully}} + \varepsilon_{\text{tully}},
\end{aligned}$$

where the β are item-specific intercepts, x_1 is an indicator for day (1 for Friday and 0 for Saturday), γ measures the effect of Friday on the different items, and x_2 is the price of fish fry with its corresponding effect δ. The error variance matrix $\boldsymbol{\Omega}$ is assumed to be a general 3×3 matrix. Thus, there are 14 identified parameters in this model. Fig. 4 graphs the convergence of the MCEM iterates for the fixed effect parameters.

Table 3 displays the MLEs and asymptotic standard errors of these parameters.

The Friday effect on the demand of fish fry is clearly seen. The price coefficients are small but not statistically significant. Thus, it appears that even with a fairly large data set, the response of customers to a price change in fish fry is small and insignificant. Our conclusions are consistent with the results of Kiefer et al. (1994) who analyzed the data using an independent logit model. Hence a price increase could be supported without serious substitution effects. In fact, the price increase has occurred (\$10.95) without any appreciable change in the menu mix or the overall sales.

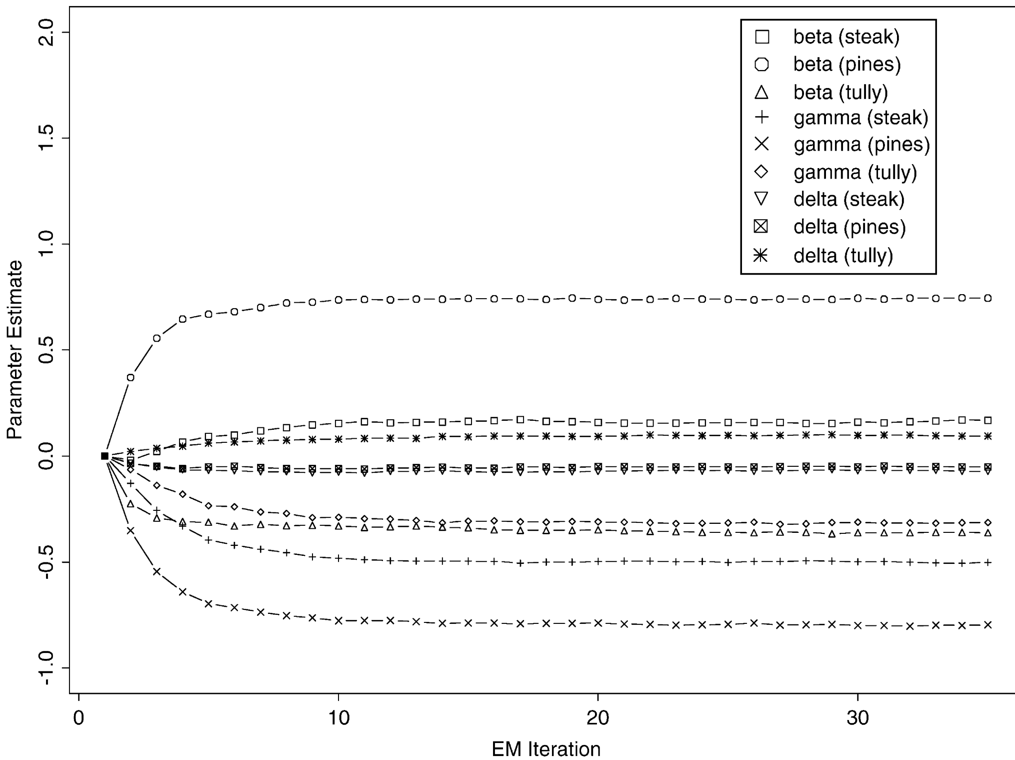

Fig. 4. Convergence of MCEM iterates for fixed effects parameters for the menu-pricing data.

Table 3
MLEs and asymptotic standard errors for 4-choice model for the menu-pricing data

	MLE	Std. Error
Item Intercept		
β_{steak}	0.168	0.390
β_{pines}	0.745[a]	0.261
β_{tully}	−0.362	0.340
Friday Effect		
γ_{steak}	−0.502[a]	0.178
γ_{pines}	−0.797[a]	0.170
γ_{tully}	−0.313[a]	0.156
Price Coefficient		
δ_{steak}	−0.072	0.072
δ_{pines}	−0.051	0.070
δ_{tully}	0.095	0.076

[a] Indicates statistical significance at the 5% level.

7. Extensions to panel data

Several studies investigate discrete choices made by individuals or households over time (Allenby and Lenk, 1994; Geweke et al., 1994). These data are known as panel data. In modeling this data, it is typical to allow the repeated observations on a household to be correlated. We describe how MCEM can be easily adapted to handle panel data, when the correlation among the relative utilities (for each time period) is modeled through the use of random effects. All the ideas extend directly for other general covariance structures (e.g., autoregressive patterns, general covariance patterns, etc.), as well as random coefficient models.

Assume that for each individual i, we observe T_i correlated multinomial vectors $\boldsymbol{w}_{it} = (w_{it1}, \ldots, w_{itc})^{\mathrm{T}}$. Further, suppose the latent utilities $\boldsymbol{y}_{it}$ follow a mixed model:

$$\boldsymbol{y}_{it}|\boldsymbol{b}_i \sim N(\boldsymbol{X}_{it}\boldsymbol{\beta} + \boldsymbol{Z}_{it}\boldsymbol{b}_i, \boldsymbol{\Omega}),$$
$$\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D}),$$

where $\boldsymbol{Z}_{it}$ is the incidence matrix. The parameters of interest are the fixed effects $\boldsymbol{\beta}$, the cross-correlation parameters in $\boldsymbol{\Omega}$ and the serial correlation parameters in $\boldsymbol{D}$. To implement the MCEM approach, we simply treat the vector of unobserved random effects $\boldsymbol{b}_i$ as part of the complete data, in addition to the latent utilities $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}^{\mathrm{T}}, \ldots, \boldsymbol{y}_{iT_i}^{\mathrm{T}})^{\mathrm{T}}$.

The MCEM algorithm of Section 3.2 may be adapted for the panel data setting by replacing Step 1 with

Step 1′ (E-step): For each i and purchase instance t, calculate $E(\boldsymbol{y}_i|\boldsymbol{w}_i, \boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)}, \boldsymbol{D}^{(m)})$, $\mathrm{Var}(\boldsymbol{y}_i|\boldsymbol{w}_i, \boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)}, \boldsymbol{D}^{(m)})$, $E(\boldsymbol{b}_i|\boldsymbol{w}_i, \boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)}, \boldsymbol{D}^{(m)})$, and $E(\boldsymbol{b}_i\boldsymbol{b}_i^{\mathrm{T}}|\boldsymbol{w}_i, \boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)}, \boldsymbol{D}^{(m)})$, and Step 2 with

Step 2a′: Set

$$\boldsymbol{D}^{(m+1)} = N^{-1}\sum_{i=1}^{N} E(\boldsymbol{b}_i\boldsymbol{b}_i^{\mathrm{T}}|\boldsymbol{w}_i, \boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)}, \boldsymbol{D}^{(m)}).$$

Step 2b′: Set

$$\beta^{(m+1)} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T_i} \boldsymbol{X}_{it}^{\mathrm{T}}\boldsymbol{\Omega}^{(m)-1}\boldsymbol{X}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T_i} \boldsymbol{X}_{it}^{\mathrm{T}}\boldsymbol{\Omega}^{(m)-1}\boldsymbol{y}_{it}^{*}\right).$$

where $\boldsymbol{y}_{it}^{*} = E(\boldsymbol{y}_{it}|\boldsymbol{w}_i, \boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)}, \boldsymbol{D}^{(m)}) - \boldsymbol{Z}_{it}E(\boldsymbol{b}_i|\boldsymbol{w}_i, \boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)}, \boldsymbol{D}^{(m)})$.

Step 2c′: Maximize

$$-\frac{\sum_{i=1}^{N} T_i}{2}\ln(\det(\boldsymbol{\Omega})) - \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Omega}^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T_i} Q_{it}(\boldsymbol{\beta}^{(m+1)}|\boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)})\right)$$

over the unknown elements of $\boldsymbol{\Omega}$ to obtain $\boldsymbol{\Omega}^{(m+1)}$. Note that the expression $Q_{it}(\boldsymbol{\beta}|\boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)}) = E(\boldsymbol{e}_{it}^{*}\boldsymbol{e}_{it}^{*\mathrm{T}}|\boldsymbol{w}_i, \boldsymbol{\beta}^{(m)}, \boldsymbol{\Omega}^{(m)}, \boldsymbol{D}^{(m)})$ where $\boldsymbol{e}_{it}^{*} = \boldsymbol{y}_{it}^{*} - \boldsymbol{X}_{it}\boldsymbol{\beta}$.

In order to estimate the expectations involved in the E-step, we use the Gibbs sampler to generate from the conditional specifications $f(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{w}_i)$ and $f(\boldsymbol{b}_i|\boldsymbol{y}_i, \boldsymbol{w}_i)$. This involves generation from truncated Normal and Normal distributions respectively.

Table 4
Purchase frequencies and costs for the peanut butter data

Brand	Frequency of purchase	Unit price (in cents)
Jiff (creamy)	476	9.81
Jiff (chunky)	141	10.00
Peter Pan (creamy)	123	9.61
Peter Pan (chunky)	53	9.96
Skippy (creamy)	581	9.32
Skippy (chunky)	387	9.18

8. Peanut butter data

This example illustrates the feasibility of MCEM calculations for a panel data choice model. We study consumer purchases of six different brands of peanut butter (ERIM panels, A.C. Nielsen, 1985–1988). The six brands under consideration are: Jiff (creamy & chunky), Peter Pan (creamy & chunky) and Skippy (creamy & chunky). Together these brands account for more than 75% of all purchases in the peanut butter category. Our sample of 105 households have 1761 purchase records for these six top brands. Table 4 displays the brands, sample frequency of purchase and the unit price averaged over the 1761 records.

The reference brand is Skippy (creamy). We fit the following random effects model for the relative utilities:

$$\begin{aligned}&\boldsymbol{y}_{it} = \boldsymbol{X}_{it}\boldsymbol{\beta} + \boldsymbol{I}\boldsymbol{b}_i + \boldsymbol{\varepsilon}_{it},\\&\boldsymbol{b}_i \sim N(\mathbf{0}, \theta I_5),\\&\boldsymbol{\varepsilon}_{it} \sim N(\mathbf{0}, \boldsymbol{I}),\end{aligned}$$

where $\boldsymbol{X}_{it}$ is a 5×6 design matrix corresponding to five brand-specific intercepts and the logarithm of price, and $\boldsymbol{I}$ is a 5×5 identity matrix. The above model allows for household heterogeneity in the intercept. The component of variance, θ captures the variability among household preferences. Table 5 reports the MLEs (along with asymptotic standard errors) of the parameters $\boldsymbol{\beta}$ and θ.

As might be expected in a choice situation with close substitutes, the price coefficient is large and negative (-6.33). Peter Pan (chunky) has the largest negative intercept, indicating that households accord a much lower utility to this brand. This fact is further corroborated by the large negative intercept (-2.00) of the Peter Pan (creamy) brand as well. The estimate of θ is large and significantly different from zero, supporting the belief that households are extremely heterogeneous in their choice behavior.

9. Summary

The goal of the work reported here was to calculate maximum likelihood estimates in the MNP model. By resorting to an EM approach, we circumvented direct evaluation of the intractable likelihood function. We showed through a simulated example

Table 5
MLEs and asymptotic standard errors for 6-choice hierarchical model for the peanut butter data

	MLE	Std. Error
Brand Intercept		
Jiff (creamy)	−0.721[a]	0.174
Jiff (chunky)	−1.718[a]	0.193
Peter Pan (creamy)	−1.981[a]	0.197
Peter Pan (chunky)	−2.423[a]	0.210
Skippy (chunky)	−0.968[a]	0.172
Price coefficient	−6.335[a]	0.595
Variance component	2.451[a]	0.071

[a] Indicates statistical significance at the 5% level.

that our MCEM method approaches the neighborhood of the MLE rapidly. Further, it may be readily adapted to handle more complicated models, such as models for panel data. The calculations were illustrated through two real data examples with four and six categories.

Acknowledgements

This work was supported by NSF grants DMS-9625476 and DMS-9709091. We thank two anonymous referees for valuable suggestions that greatly improved this manuscript.

Appendix

Expressions for the elements of S and H are

$$S_{\boldsymbol{\beta}} = \sum_{i=1}^{N} \frac{\partial}{\partial \boldsymbol{\beta}} \ln f(\boldsymbol{y}_i | \boldsymbol{\beta}, \boldsymbol{\Omega}),$$

$$= \sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{\Omega}^{-1} \boldsymbol{e}_i,$$

$$S_{\omega_{rs}} = \sum_{i=1}^{N} \frac{\partial}{\partial \omega_{rs}} \ln f(\boldsymbol{y}_i | \boldsymbol{\beta}, \Omega),$$

$$= -\frac{N}{2} \operatorname{tr}(\boldsymbol{\Omega}^{-1} \dot{\boldsymbol{\Omega}}_{rs}) + \frac{1}{2} \sum_{i=1}^{N} \boldsymbol{e}_i^{\mathrm{T}} \boldsymbol{\Omega}^{-1} \dot{\boldsymbol{\Omega}}_{rs} \boldsymbol{\Omega}^{-1} \boldsymbol{e}_i,$$

$$H_{\boldsymbol{\beta\beta}} = \sum_{i=1}^{N} \frac{\partial^2}{\partial \boldsymbol{\beta} \boldsymbol{\beta}^{\mathrm{T}}} \ln f(\boldsymbol{y}_i | \boldsymbol{\beta}, \boldsymbol{\Omega})$$

$$= -\sum_{i=1}^{N} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\Omega}^{-1} \boldsymbol{X}_i,$$

$$H_{\boldsymbol{\beta}\omega_{rs}} = \sum_{i=1}^{N} \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \omega_{rs}} \ln f(\boldsymbol{y}_i | \boldsymbol{\beta}, \boldsymbol{\Omega})$$

$$= -\sum_{i=1}^{N} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\Omega}^{-1} \dot{\boldsymbol{\Omega}}_{rs} \boldsymbol{\Omega}^{-1} \boldsymbol{e}_i,$$

$$H_{\omega_{rs}\omega_{tu}} = \sum_{i=1}^{N} \frac{\partial^2}{\partial \omega_{tu} \partial \omega_{rs}} \ln f(\boldsymbol{y}_i | \boldsymbol{\beta}, \boldsymbol{\Omega})$$

$$= \frac{N}{2} \operatorname{tr}(\boldsymbol{\Omega}^{-1} \dot{\boldsymbol{\Omega}}_{tu} \boldsymbol{\Omega}^{-1} \dot{\boldsymbol{\Omega}}_{rs}) - \sum_{i=1}^{N} \boldsymbol{e}_i^{\mathrm{T}} \boldsymbol{\Omega}^{-1} (\dot{\boldsymbol{\Omega}}_{tu} \boldsymbol{\Omega}^{-1} \dot{\boldsymbol{\Omega}}_{rs} + \dot{\boldsymbol{\Omega}}_{rs} \boldsymbol{\Omega}^{-1} \dot{\boldsymbol{\Omega}}_{tu}) \boldsymbol{\Omega}^{-1} \boldsymbol{e}_i$$

where ω_{rs} is the (r,s)th element of $\boldsymbol{\Omega}, \dot{\boldsymbol{\Omega}}_{rs} = (\partial/\partial\omega_{rs})\boldsymbol{\Omega}$ and $\boldsymbol{e}_i = \boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$.

References

Albert, J.H., Chib, S., 1993. Bayesian analysis of polychotomous response data. J. Amer. Statist. Assoc. 88, 669–679.

Allenby, G.M., Lenk, P.J., 1994. Modeling household purchase behavior with logistic normal regression. J. Amer. Statist. Assoc. 89, 1218–1231.

Aptech Systems, 1994. The GAUSS System, Version 3.0. Maple Valley, WA.

Ashford, J.R., Sowden, R.R., 1970. Multi-variate probit analysis. Biometrics 26, 535–546.

Booth, J.G., Hobert, J.P., 1999. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. J. Roy. Statist. Soc. B 61, 265–285.

Börsch-Supan, A., Hajivassiliou, V., 1994. Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. J. Econometrics 58, 347–368.

Chan, K.S., Ledolter, J., 1995. Monte carlo EM estimation for time series models involving counts. J. Amer. Statist. Assoc. 90, 242–252.

Daganzo, C., 1980. Multinomial Probit. Acedemic Press, New York, 1980.

Dansie, B., 1985. Parameter estimability in the multinomial probit model. Transportation Res. 19B, 526–528.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood estimation from incomplete observations. J. Roy. Statist. Soc. B 39, 1–38.

Devroye, L., 1986. Non-Uniform Random Variate Generation. Springer, New York.

Diebolt, J., Ip, E.H.S., 1998. Stochastic EM: Method and Application. in: Markov Chain Monte Carlo in Practice, Eds: Gilks, Richardson, Spiegelhalter. Chapman & Hall, New York, NY.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. 6, 721–741.

Geweke, J.F., 1989. Efficient simulation from the multivariate normal distribution subject to linear inequality constraints and the evaluation of constraint probabilities. Mimeo. Department of Economics, Duke University, Durham, NC.

Geweke, J.F., Keane, M.P., Runkle, D.E., 1994. Statistical inference in the multinomial multiperiod probit model. Research Department Staff Report 177. Federal Reserve Bank of Minneapolis, MN.

Guo, S.W., Thompson, E.A., 1992. A Monte Carlo method for combined segregation and linkage analysis. Amer. J. Hum. Genet. 51, 1111–1126.

Hajivassiliou, V., McFadden, D., Ruud, D., 1992. Simulation of multivariate normal orthant probabilities: methods and programs. Discussion Paper 1021. Cowles Foundation, Yale University, CT.

Hausman, J., Wise, D., 1978. A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences. Econometrica 45, 319–339.

Jennrich, R.I., Schluchter, M.D., 1986. Unbalanced repeated measures models with structured covariance matrices. Biometrics 42, 805–820.

Keane, M.P., 1994. A computationally practical simulation estimator for panel data. Econometrica 62, 95–116.

Kiefer, N.M., Kelly, T.J., Burdett, K., 1994. Menu pricing: an experimental approach. J. Bus. Econom. Statist. 3, 329–337.

Lerman, S., Manski, C., 1981. On the use of simulated frequencies to approximate choice probabilities. In: Manski, C., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA, pp. 305–319.

Louis, T.A., 1982. Finding the observed information matrix when using the EM algorithm. J. Roy. Statist. Soc. B 44, 226–233.

McCulloch, C.E., 1994. Maximum likelihood variance components estimation for binary data. J. Amer. Statist. Assoc. 89, 330–335.

McCulloch, R., Rossi, P.E., 1994. An exact likelihood analysis of the multinomial probit model. J. Econometrics 64, 207–240.

McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica 57, 995–1027.

Meng, X.L., Rubin, D.B., 1993. Maximum likelihood estimation via the ECM algorithm: a general framework. Biometrika 80, 267–278.

Searle, S., Casella, G., McCulloch, C.E., 1992. Variance Components. Wiley, New York.

Stern, S., 1992. A method for smoothing simulated moments of discrete probabilities in multinomial probit models. Econometrica 60, 943–952.

Tanner, M.A., 1993. Tools for Statistical Inference. Methods for the Exploration of Posterior Distributions and Likelihood Functions. Springer, New York.