NISS

A Random Effects Multinomial Probit Model of Car Ownership Choice

Agostino Nobile, Chandra R. Bhat and Eric I. Pas

Technical Report Number 41
November, 1995

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

A Random Effects Multinomial Probit Model

of

Car Ownership Choice[1]

Agostino Nobile[2]

Chandra R. Bhat[3]

and

Eric I. Pas[4]

November 21, 1995

[1]Research supported by NSF grants DMS9208758 and DMS9313013 to NISS.

[2]National Institute of Statistical Sciences and Institute of Statistics and Decision Sciences, Duke University

[3]Department of Civil Engineering, University of Massachusetts at Amherst

[4]Department of Civil and Environmental Engineering, Duke University

Abstract

The number of cars in a household has an important effect on its travel behavior (e.g., choice of number of trips, mode to work, and non-work destinations), hence car ownership modeling is an essential component of any travel demand forecasting effort. In this paper we report on a random effects multinomial probit model of car ownership level, estimated using longitudinal data collected in the Netherlands.

A Bayesian approach is taken and the model is estimated by means of a modification of the Gibbs sampling with data augmentation algorithm considered by McCulloch and Rossi (1994). The modification consists in performing, after each Gibbs sampling cycle, a Metropolis step along a direction of constant likelihood. An examination of the simulation output illustrates the improved performance of the resulting sampler.

Keywords: Multinomial probit model, Panel data, Gibbs sampling, Metropolis algorithm, Bayesian analysis.

1 Introduction

The number of cars in a household plays an important role in travel demand forecasting since it is a major determinant of the trip-making behavior of individuals and households. Among other things, household car ownership affects the number of trips made by a household (Meurs, 1990), choice of destination for non-work activity participation (Wrigley, 1990), mode choice to work and to other non-work activities (Bhat, 1995; Uncles, 1987), and chaining of activities in a tour (Hamed and Mannering, 1993; Strathman et al., 1994).

The critical need for accurate car ownership forecasting as a precursor to reliable travel demand forecasting has led, in the past few years, to a number of important developments in car ownership modeling. One such development is the use of longitudinal analysis methods. Longitudinal panel data methods enable the incorporation of the intertemporal dimension present in car ownership choices (for example, resistance or delays to change in car ownership levels due to search costs and uncertainty of financial position in the future; see Goodwin and Mogridge, 1981).

Longitudinal analysis also enables the control of unobserved intertemporal factors, such as acquired tastes for a certain lifestyle, that are likely to affect household car ownership decisions. These unobserved factors are also likely to make some car ownership alternatives closer substitutes than others (Small, 1987), thus inducing an intratemporal correlation among alternatives. From a modeling perspective, this implies that the independence of irrelevant alternatives (IIA) assumption (Ben-Akiva and Lerman, 1985), often maintained in discrete choice models, is likely to be violated in the context of car ownership choices.

Ideally, then, car ownership choices should be modeled accounting for both intertemporal correlation of unobserved determinants over time and intratemporal correlation of unobserved determinants across choice alternatives. While researchers have recognized the need to accommodate both these types of correlation, previous efforts in car ownership modeling have either a) ignored both types of correlation (Lerman and Ben-Akiva, 1975) or b) focused only on the intratemporal dimension of correlation (Bunch and Kitamura, 1990; Small, 1987), or c) addressed only the intertemporal dimension of correlation (Hensher et al., 1992).

In this paper we present a multinomial probit (MNP) model which accommodates both intratemporal and intertemporal correlation. The intratemporal correlation is accommodated by allowing a general form for the error term covariance matrix. The intertemporal correlation is captured by unobserved individual-specific time-invariant attributes (also referred to as household heterogeneity). For other ways of allowing for intratemporal and intertemporal correlations see Börsch-Supan et al. (1992).

The approach is Bayesian: a prior distribution for the parameters of the longitudinal MNP model is specified and the posterior is examined using Markov chain Monte Carlo methods. The Markov chain is similar to the Gibbs sampler with data augmentation proposed by McCulloch and Rossi (1994), with the additional Metropolis step introduced by Nobile (1995).

The rest of the paper is organized as follows. The structure of the longitudinal MNP model is presented in Section 2. Details on the Markov chain sampler are reported in Section 3. Section 4 describes the data set used in our modeling effort and presents empirical results. Conclusions are presented in Section 5. The Appendix provides a comparison between the performances of the samplers with and without the additional Metropolis step, by exhibiting simulated sample paths for some parameters in the model.

2 The model

Suppose that the choices of n agents over a set of p alternatives are observed at different times and denote by $\boldsymbol{d}_i = (d_{it}, t \in S_i)$ the choices made by agent i. Thus, $d_{it} = j$ means that at time t the i-th agent has chosen the j-th alternative. S_i is the set of times in which the choice of agent i is observed; this allows, for instance, for the usual turn-over of households occurring in panel surveys. We denote by T_i the number of elements in S_i.

Denote by $\boldsymbol{w}_{it}$ the $(p-1)$-vector containing the utility differentials of agent i at time t between the first $p-1$ alternatives and the last alternative, used as reference. Assume that these utility differentials are linear combinations of q characteristics Y_{itm}, $m = 1, \ldots, q$, of the agent, plus error. Then, one can write the MNP model under consideration as

$$d_{it} = \begin{cases} j & \text{if } w_{itj} \geq \max_b w_{itb} \geq 0 \\ p & \text{if } \max_b w_{itb} < 0 \end{cases} \tag{1}$$

$$w_{itj} = \alpha_{ij} + \sum_{m=1}^{q} \delta_{jm} Y_{itm} + \epsilon_{itj} \tag{2}$$

where $i = 1, \ldots, n$; $j = 1, \ldots, p-1$ and $t \in S_i$.

Equation (2) can be rewritten in matrix notation as

$$\boldsymbol{w}_{it} = \boldsymbol{\alpha}_i + \boldsymbol{\Delta}\mathbf{Y}_{it} + \boldsymbol{\epsilon}_{it}. \tag{3}$$

Agents' heterogeneity is modeled by assuming that the intercepts $\boldsymbol{\alpha}_i$ are normal random effects:

$$\boldsymbol{\alpha}_i \overset{i.i.d.}{\sim} \mathrm{N}_{p-1}(\boldsymbol{\lambda}, \mathbf{V}), \qquad i = 1, \ldots, n. \tag{4}$$

The other component $\boldsymbol{\epsilon}_{it}$ of the error term is normally distributed, independently of $\boldsymbol{\alpha}_i$:

$$\boldsymbol{\epsilon}_{it} \overset{i.i.d.}{\sim} \mathrm{N}_{p-1}(\mathbf{0}, \boldsymbol{\Sigma}) \qquad i = 1, \ldots, n;\ t \in S_i \tag{5}$$

The specification of the model is completed by assigning a prior distribution on the parameters. The covariance matrices $\boldsymbol{\Sigma}$ and $\mathbf{V}$ are assumed to have an inverse Wishart, while $\boldsymbol{\Delta}$ is matrix-variate normal and $\boldsymbol{\lambda}$ multivariate normal:

$$\boldsymbol{\Sigma}^{-1} \sim \mathrm{W}_{p-1}(\nu, \mathbf{P}) \tag{6}$$

$$\boldsymbol{\Delta} \sim \mathrm{N}_{(p-1)\times q}(\mathbf{D}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}^{-1}) \tag{7}$$

$$\mathbf{V}^{-1} \sim \mathrm{W}_{p-1}(\eta, \mathbf{Q}) \tag{8}$$

$$\boldsymbol{\lambda} \sim \mathrm{N}_{p-1}(\boldsymbol{\ell}, \varphi^{-1}\mathbf{V}). \tag{9}$$

The symbol θ will be used to denote the parameter vector, $\theta = \{\boldsymbol{\Sigma}, \boldsymbol{\Delta}, \mathbf{V}, \boldsymbol{\lambda}\}$ and $\pi(\theta)$ will denote the prior distribution (6)–(9).

3 The hybrid Markov chain sampler

Inference about the parameters in the model of the previous section follows the Bayesian approach described in McCulloch and Rossi (1994). Values of the hyperparameters in (6)–(9) are chosen so that the prior distribution is proper, yet rather diffuse, so that the posterior distribution of θ will mostly reflect the likelihood. A sample from the posterior of θ is obtained using the Gibbs sampler with data augmentation algorithm. The constraint $\sigma_{11} = 1$, usually imposed to obtain identifiability in a MNP model, is enforced "a posteriori" through the consideration of "identified" functionals of the sampled parameter values, such as $\delta_{jm}/\sqrt{\sigma_{11}}$, σ_{jj}/σ_{11}, $\sigma_{jk}/\sqrt{\sigma_{jj}\sigma_{kk}}$. After each Gibbs cycle through the full conditional distributions, the Metropolis step proposed by Nobile (1995) is performed, to ameliorate the convergence properties of the sampling Markov chain.

The next subsection is devoted to the full conditional distributions necessary to implement the Gibbs sampler with data augmentation, while the following one deals with the specification of the Metropolis step for the model at hand.

3.1 The Gibbs sampler

When the parameter vector θ can be partitioned as $\theta = \{\theta_1, \ldots, \theta_m\}$ in a way that the full conditional distributions

$$[\theta_i | \{\theta_j, j \neq i\}, Data] \tag{10}$$

can be sampled from, then successively drawing from these distributions will asymptotically give a sample from the posterior distribution of θ. This is the Gibbs sampler (Gelfand and Smith, 1990).

In models where the distributions (10) are available only conditional on some latent variable W, sampling from the posterior of θ is accomplished by successively drawing from the following full conditional distributions:

$$[\theta_i | \{\theta_j, j \neq i\}, W, Data] \tag{11}$$

$$[W | \theta, Data]. \tag{12}$$

This technique is called data augmentation (Tanner and Wong, 1987).

Papers dealing with the application of the Gibbs sampler with data augmentation to the MNP model include Albert and Chib (1993), McCulloch and Rossi (1994), and Geweke et al. (1994). The last cited paper compares the performance of the Gibbs sampler to that of other estimation methods for the MNP model.

Below we report the full conditional distributions necessary to implement the Gibbs sampler with data augmentation for the model (1)–(9). We will denote by *Data* the sequence of choice vectors $\boldsymbol{d}_i$ and by $\boldsymbol{\alpha}$ the sequence of random effects $\boldsymbol{\alpha}_i$.

1. $\left[w_{itj}|\boldsymbol{\Sigma}, \boldsymbol{\Delta}, \mathbf{V}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathbf{W}_{-(itj)}, Data\right]$.

The distribution of w_{itj} conditional on the parameters, the other utility differentials $\mathbf{W}_{-(itj)}$ and the observed data is normal $\mathrm{N}(m_{ijt}, \tau^2_{ijt})$ truncated either to $[\max(w_{it,-j}, 0), \infty)$ if $d_{it} = j$, or to $(-\infty, \max(w_{it,-j}, 0)]$ if $d_{it} \neq j$. The mean m_{ijt} and variance τ^2_{ijt} are

$$m_{itj} = \alpha_{ij} + \boldsymbol{\delta}_j^T \mathbf{Y}_{it} - g_{jj}^{-1} \boldsymbol{g}_{j,-j}^T (\boldsymbol{w}_{it,-j} - \boldsymbol{\alpha}_{i,-j} - \boldsymbol{\Delta}_{-j} \mathbf{Y}_{it})$$

$$\tau^2_{ijt} = g_{jj}^{-1}$$

where

$$g_{jj} = \left[\sigma_{jj} - \boldsymbol{\sigma}_{j,-j}^T \boldsymbol{\Sigma}_{-j,-j}^{-1} \boldsymbol{\sigma}_{j,-j}\right]^{-1}$$

$$\boldsymbol{g}_{j,-j} = -\boldsymbol{\Sigma}_{-j,-j}^{-1} \boldsymbol{\sigma}_{j,-j} g_{jj}$$

and $\boldsymbol{w}_{it,-j}$ is the vector $\boldsymbol{w}_{it}$ with w_{itj} deleted; $\boldsymbol{\alpha}_{i,-j}$ is $\boldsymbol{\alpha}_i$ with α_{ij} deleted; $\boldsymbol{\Delta}_{-j}$ is $\boldsymbol{\Delta}$ with the j-th column $\boldsymbol{\delta}_j$ deleted; $\boldsymbol{\Sigma}_{-j,-j}$ is $\boldsymbol{\Sigma}$ with the j-th row and column deleted; $\boldsymbol{\sigma}_{j,-j}$ is the j-th column of $\boldsymbol{\Sigma}$ after the deletion of its j-th element σ_{jj}.

2.$[\boldsymbol{\Sigma}|\boldsymbol{\Delta}, \mathbf{V}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathbf{W}, Data]$.

This is an inverse Wishart distribution:

$$\boldsymbol{\Sigma}^{-1} \sim \mathrm{W}_{p-1}\left(\nu + \sum_{i=1}^{n} T_i, \mathbf{P} + \sum_{i=1}^{n} \sum_{t \in S_i} \boldsymbol{\epsilon}_{it} \boldsymbol{\epsilon}_{it}^T\right).$$

3.$[\boldsymbol{\Delta}|\boldsymbol{\Sigma}, \mathbf{V}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathbf{W}, Data]$.

This is a matrix-variate distribution:

$$\boldsymbol{\Delta}^T \sim \mathrm{N}_{q\times(p-1)}\left(\widehat{\boldsymbol{\Delta}}^T, \boldsymbol{\Omega}_1^{-1} \otimes \boldsymbol{\Sigma}\right)$$

where

$$\widehat{\boldsymbol{\Delta}}^T = \boldsymbol{\Omega}_1^{-1}\left[\boldsymbol{\Omega}\mathbf{D}^T + \sum_{i=1}^{n}\sum_{t\in S_i}\mathbf{Y}_{it}(\boldsymbol{w}_{it} - \boldsymbol{\alpha}_i)^T\right]$$

and

$$\boldsymbol{\Omega}_1 = \boldsymbol{\Omega} + \sum_{i=1}^{n}\sum_{t\in S_i}\mathbf{Y}_{it}\mathbf{Y}_{it}^T.$$

4.$[\boldsymbol{\alpha}|\boldsymbol{\Sigma}, \boldsymbol{\Delta}, \mathbf{V}, \boldsymbol{\lambda}, \mathbf{W}, Data]$.

Here we have a set of n conditionally independent multivariate normals:

$$\boldsymbol{\alpha}_i \sim \mathrm{N}_{p-1}(\widehat{\boldsymbol{\alpha}}_i, \mathbf{V}_1)$$

where

$$\mathbf{V}_1^{-1} = \mathbf{V}^{-1} + T_i\boldsymbol{\Sigma}^{-1}$$

and

$$\widehat{\boldsymbol{\alpha}}_i = \mathbf{V}_1\left[\mathbf{V}^{-1}\boldsymbol{\lambda} + \boldsymbol{\Sigma}^{-1}\sum_{t\in S_i}(\boldsymbol{w}_{it} - \boldsymbol{\Delta}\mathbf{Y}_{it})\right].$$

5.$[\boldsymbol{\lambda}|\boldsymbol{\Sigma}, \boldsymbol{\Delta}, \mathbf{V}, \boldsymbol{\alpha}, \mathbf{W}, Data]$.

The conditional distribution of $\boldsymbol{\lambda}$ is multivariate normal:

$$\boldsymbol{\lambda} \sim \mathrm{N}_{p-1}(\widehat{\boldsymbol{\lambda}}, (\varphi + n)^{-1}\mathbf{V})$$

with

$$\widehat{\boldsymbol{\lambda}} = (\varphi\boldsymbol{\ell} + n\overline{\boldsymbol{\alpha}})/(\varphi + n)$$

and

$$\overline{\boldsymbol{\alpha}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\alpha}_i.$$

6.$[\mathbf{V}|\Sigma, \boldsymbol{\Delta}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathbf{W}, Data]$.

This is also inverse Wishart:

$$\mathbf{V}^{-1} \sim \mathrm{W}_{p-1}\left(\eta + n, \mathbf{Q} + \sum_{i=1}^{n}(\boldsymbol{\alpha}_i - \boldsymbol{\lambda})(\boldsymbol{\alpha}_i - \boldsymbol{\lambda})^T\right).$$

3.2 The Metropolis step

For the sake of simplicity, in this section we reparameterize $\theta = \{\Sigma, \boldsymbol{\Delta}, \mathbf{V}, \boldsymbol{\lambda}\}$ to $\vartheta = \{\mathbf{L}_{\Sigma}, \boldsymbol{\Delta}, \mathbf{L}_{\mathbf{V}}, \boldsymbol{\lambda}\}$, where $\mathbf{L}_{\Sigma}$ and $\mathbf{L}_{\mathbf{V}}$ are the lower triangular Cholesky factors of Σ and $\mathbf{V}$, respectively. $\pi(\vartheta)$ will denote the prior distribution on ϑ induced by (6)–(9).

As already mentioned, the model (1)–(5) is not identified, since rescaling the parameters and the utility differentials leaves unchanged the distribution of the *Data*. More precisely, for any $c > 0$,

$$k(Data, \mathbf{W}, \boldsymbol{\alpha}|\vartheta) = k(Data, c\mathbf{W}, c\boldsymbol{\alpha}|c\vartheta) \tag{13}$$

where k denotes the joint probability density of observed choices, unobserved utility differentials and random effects, given the parameters. Identification is usually achieved by imposing $\sigma_{11} = 1$, thus implicitly fixing the arbitrary constant c.

In the Bayesian approach, the model (1)–(9) is, strictly speaking, identified, because of the proper prior distribution. This allows to run the Markov chain sampler without imposing $\sigma_{11} = 1$. However, on sets of the form $\{c\vartheta, c > 0\}$, variation in the posterior distribution only reflects variation in the prior, since the likelihood is constant. Therefore, as the sample size increases, the posterior mass will concentrate rather than on a single point $\tilde{\vartheta}$, on the set $\{c\tilde{\vartheta}, c > 0\}$. While this explains the need to look at "identified" functionals of the sampled parameter values, it also suggests a technique to improve the speed with which the sampling Markov chain explores the parameter/latent variables space.

Suppose a Gibbs cycle has yielded $\psi_0 = \{\vartheta, \boldsymbol{\alpha}, \mathbf{W}\}$. Then, before performing the next Gibbs cycle, one may change the scale of ψ_0, having care to respect the equilibrium distribution of the sampling chain. One way this can be accomplished is by means of a Metropolis step: a candidate

$\psi^* = c\psi_0$ is drawn from some distribution $h(\psi|\psi_0)$ and it is accepted with probability

$$\min\left[\frac{f(\psi^*)}{f(\psi_0)}\frac{h(\psi_0|\psi^*)}{h(\psi^*|\psi_0)}, 1\right],$$

where $f(\psi) \propto k(Data, \mathbf{W}, \boldsymbol{\alpha}|\vartheta)\pi(\vartheta)$ is the equilibrium distribution of the Markov chain. Because of (13), $f(\psi^*)/f(\psi_0)$ depends only on the ratio between the prior distribution at $\vartheta_0 = \{\mathbf{L}_{\boldsymbol{\Sigma}}, \boldsymbol{\Delta}, \mathbf{L}_{\mathbf{V}}, \boldsymbol{\lambda}\}$ and $\vartheta^* = c\vartheta_0$:

$$\begin{aligned}\frac{f(\psi^*)}{f(\psi_0)} = \frac{\pi(\vartheta^*)}{\pi(\vartheta_0)} &= c^{-(p-1)(\nu+\eta+p+q+1)} \exp\left\{-\frac{1}{2}(c^{-1}-1)\left[(c^{-1}+1)\left(\mathrm{tr}\mathbf{P}\boldsymbol{\Sigma}^{-1} + \mathrm{tr}\mathbf{Q}\mathbf{V}^{-1}\right)\right.\right.\\ &\left.\left. + \mathrm{tr}\boldsymbol{\Omega}\mathbf{D}^T\boldsymbol{\Sigma}^{-1}\left((c^{-1}+1)\mathbf{D} - 2\boldsymbol{\Delta}\right) + \varphi\boldsymbol{\ell}^T\mathbf{V}^{-1}\left((c^{-1}+1)\boldsymbol{\ell} - 2\boldsymbol{\lambda}\right)\right]\right\}.\end{aligned} \tag{14}$$

Also, the candidate ψ^* is best obtained by sampling a rescaling factor c. One can show that if c is drawn from an $Exp(1)$ distribution (other viable alternatives are considered in Nobile, 1995), the acceptance probability of the Metropolis step is

$$\min\left[\frac{\pi(\vartheta^*)}{\pi(\vartheta_0)}\frac{1}{c}\exp\left\{\frac{c^2-1}{c}\right\}, 1\right].$$

Considering that the limiting distributions of both Markov chains, with and without the rescaling Metropolis step, coincide with the posterior distribution on the parameter/latent variables space, an argument needs to be made to support the slightly more expensive use of the hybrid chain. Our still limited experience suggests that, at least in problems involving binary exogenous variables, and within the constraints imposed by limited computing resources, the hybrid chain seems to provide more representative samples. Further discussion of this issue is deferred to the Appendix.

4 Data and Results

4.1 Data Source

The data source used in the present study is the Dutch National Mobility Panel. This panel was instituted in 1984 and involves a one-week travel diary and household and personal questionnaires collected at biannual and annual intervals. Ten waves (a wave refers to cross-sectional data at one time point) were collected between March 1984 and March 1989. See van Wissen and Meurs (1989) for additional information. Data for our analysis is obtained from waves 3, 5, 7 and 9 of the panel, collected during the spring of the years 1985 through 1988 (data from wave 1 were omitted because there was considerable sample attrition between waves 1 and 3). The four waves comprise (after deletion of some cases due to missing values) 2731 households for a total of 6882 observed choices.

Since fewer than 1% of the choices corresponded to three or more automobiles, we decided to consider only $p = 3$ alternatives: 0, 1, 2 or more cars. The alternative "0 cars" was used as the reference alternative.

We considered several model specifications, some of them were quite comparable in terms of goodness of fit. We present here the results for one such specification.

The set of $q = 15$ exogenous variables comprises a) wave dummy variables accounting for fixed temporal effects relative to wave 3, b) residential location indicators for city and suburb dwelling as opposed to rural dwelling, and c) socio-demographic variables. The socio-demographic variables include three license-dummy variables for 1, 2 and 3 or more license holders, respectively, in the household (the base is households with no licensed members), annual household income (in tens of thousands of Dutch Guilders), number of workers (full time and part-time) in the household, two dummy variables to represent the effect of number of adults (the base is household with one adult), and the number of children in the household classified by age-group.

4.2 Results

We estimated the model of Section 2 using the sampling procedure described in Section 3. The following hyperparameter values were used: $\nu = 4$, $\mathbf{P} = 4I_{p-1}$, $\mathbf{D} = \mathbf{0}_{(p-1)\times q}$, $\mathbf{\Omega} = 0.01 I_{q\times q}$, $\eta = 4$, $\mathbf{Q} = 4I_{p-1}$, $\boldsymbol{\ell} = \mathbf{0}_{p-1}$, $\varphi = 1$.

The sampling Markov chain was run for 50000 cycles (each consisting of a Gibbs/data augmentation cycle and a rescaling Metropolis step). The first 5000 draws were discarded. Due to the relatively large number of variables involved, the remaining 45000 draws were sampled at the rate of 1 every 4 and this subsample was output for analysis. An examination of time series plots of the sampled values of the parameters supported the presumption that the Markov chain had stabilized within the initial burn-in of 5000 draws.

Table 1 contains the means of the (marginal) posterior distributions of the parameters obtained from the simulation sample. All values are standardized so that $\sigma_{11} = 1$. Along with each mean, an interval accounting for 95% of the (marginal) posterior probability is also reported (the extremes of the intervals are the 0.025 and 0.975 quantiles of the empirical cdf's of the simulated parameter values). Estimates of other quantiles of the marginal posterior distributions, density estimates of the marginal posterior densities or joint posterior probability regions for more than one parameter could also be computed from the simulation output.

The estimates of the coefficients for the wave dummies are all negative and most of the probability is assigned to negative values by the posterior distributions, especially for the coefficients in the "2 or more cars" equation. This seems to point to the presence of generic temporal effects. Taking wave 3 as reference point, the effect is negative in all the remaining waves, it is strongest in wave 7 and rather mild in wave 9. This seems to agree somewhat with the Dutch business cycle in the years 1985-88, during which the data was collected.

Having as reference rural location, residence location in a suburban area tends to lower the utility of owning 1 and 2 or more cars approximately in the same amount, with respect to 0 cars. The same holds true for urban location, with the effect being even stronger, as would be expected. Thus the model seems to capture a clear progression of the utility of owning cars, increasing from urban to suburban to rural location. This agrees with the usually greater availability of public transportation in more densely populated areas and with their characteristically higher costs of parking.

For each level of the number of licenses in the household, there is a positive effect on the utility of 1 and 2 or more cars, with respect to no cars. Moreover, as one would expect, one license favors 1 car over 2 or more, while two licenses correspond to a slight preference for 2 or more cars and three or more licenses yield a definite increase in the utility for 2 or more cars. The estimates of the parameters in the equation for 1 car suggest a non-linear relationship, so that it would be inappropriate to replace the three dummy variables with a single quantitative variable "number of licenses".

Income enters with a positive coefficient in both the utility functions and the coefficient is larger, as one would expect, in the utility function for 2 or more cars. The number of workers in the household seems to matter especially in the utility function for 2 or more cars. The coefficients of 2 ADULTS and 3+ ADULTS indicate that the presence of more than one adult in the household is associated with a larger number of automobiles. While the two dummy variables enter in the equation for 1 car with approximately the same coefficient, there seems to be a much larger effect of three or more adults, with respect to that of two, on the utility of two or more cars.

The variables accounting for the number of young people in the household have coefficients very close to 0 and of varied sign, in the first equation. In the equation for 2 or more cars, however, they show a progression from negative to positive, as age increases, as would expected.

The ratio between the variances of the two components of the error term ϵ_{it} is about equal to 2 and the components are approximately uncorrelated. The variability of the random intercepts is much larger: v_{11} and v_{22} are about 5 and 12 times larger than σ_{11}, respectively. Thus, much of the variability in the choice behavior has to be attributed to between-household differences, rather than shocks occurring within the households. Moreover, behavior relating to the alternative "2 or more cars" is more volatile. The components of the random intercepts are positively correlated, so that larger utilities of the alternative "1 car" in a certain household are usually associated with larger utilities, in the same household, for "2 or more cars".

Using the simulation output from the Markov chain sampler, and with little additional computation, one can estimate, for each observed choice, $\int \Pr[d_{it}|\theta]\pi(\theta|Data)d\,\theta$ and from this compute a quantity analogous to the usually reported maximum log-likelihood. The estimate of this quantity for the present model is -3748. A quantity usually reported for comparison is the log-likelihood in a model where the choice probabilities are equal to the relative frequencies of choices in the sample. This quantity equals -6018.

5 Summary and Conclusions

We have presented a random effects multinomial probit model of household car ownership level. This structure allows one to capture both intratemporal and intertemporal correlations. The model was estimated using longitudinal data from the Dutch National Mobility Panel. Residential location, number of license holders in the household, household income, number of adults and number of employed people in the household appear to be important factors in car ownership decisions. Most of the variability in the observed choices can be attributed to between-household differences rather than to within-household random disturbances.

We have also described a Markov chain Monte Carlo method which combines the Gibbs sampler and the Metropolis algorithm. This technique seems to provide, at a small additional computational cost relative to Gibbs sampling, a more reliable tool in the exploration of the posterior distribution of the parameters in the model.

Acknowledgments

The authors would like to thank Mike West and Michael Lavine for useful discussions at various stages of this work. Clearly, the authors are solely responsible for any errors.

The present research was supported by National Science Foundation grants DMS9208758 and DMS9313013 to the National Institute of Statistical Sciences.

References

Albert, J. H. and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data, *Journal of the American Statistical Association*, 88, 669-679.

Bhat, C. R. (1995). A joint unordered-ordered discrete model of work mode choice and number of stops during the work commute. Accepted (subject to editorial revisions) for publication in *Transportation Research.*

Ben-Akiva, M. and Lerman, S. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*, The MIT Press.

Börsch-Supan, A., Hajivassiliou, V., Kotlikoff, L. J. and Morris, J. N. (1992). Health, Children and Elderly Living Arrangements: A Multiperiod-Multinomial Probit Model with Unobserved Heterogeneity and Autocorrelated Errors, in *Topics in the Economics of Aging*, edited by D. A. Wise, NBER Project Report.

Bunch, D. S. and Kitamura, R. (1990). Multinomial probit model estimation revisited: testing estimable model specifications, maximum likelihood algorithms and probit integral approximations for trinomial models of car ownership. Institute of Transportation Studies Technical Report, University of California, Davis.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85, 398-409.

Geweke, J., Keane, M. and Runkle, D. (1994). Alternative Computational Approaches to Inference in the Multinomial Probit Model. *The Review of Economics and Statistics*, 76, 609-632.

Goodwin, P. B. and Mogridge, M. J. H. (1981). Hypotheses for a fully dynamic model of car ownership. *International Journal of Transport Economics*, 8, 313-326.

Hamed, M. M. and Mannering, F. L. (1993). Modeling travelers' postwork activity involvement: Toward a new methodology. *Transportation Science*, 27, 381-394.

Hensher, D. A., Smith, N. C., Milthorpe, F. M., and Barnard, P. O. (1992). *Dimensions of Automobile Demand: A Longitudinal Study of Automobile Ownership and Use*, Amsterdam, North Holland.

Lerman, S. and Ben-Akiva, M. (1975). Disaggregate Behavioral Model of Automobile Ownership, *Transportation Research Record*, 569, 43-51.

McCulloch, R. E. and Rossi, P. E. (1994). An Exact Likelihood Analysis of the Multinomial Probit Model, *Journal of Econometrics*, 64, 207-240.

Meurs, H. (1990). Trip generation models with permanent unobserved effects, *Transportation Research*, 24B, 145-158.

Nobile, A. (1995). A Hybrid Markov Chain for the Bayesian Analysis of the Multinomial Probit Model. National Institute of Statistical Sciences, Technical Report # 36.

Small, K.A. (1987). A discrete choice model for ordered alternatives, *Econometrica*, 55, 409-424.

Strathman, J. G., Dueker, K. J. and Davis, J. S. (1994). Effects of household structure and selected travel characteristics on trip chaining. *Transportation*, 21, 23-45.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82, 528-540.

Uncles, M. D. (1987). A beta-logistic model of mode choice: goodness of fit and intertemporal dependence. *Transportation Research*, 21B, 195-205.

Wissen, L. van and Meurs, H. (1989). The Dutch mobility panel: Experiences and evaluation. *Transportation*, 16, 99-119.

Wrigley, N. (1990). Dirichlet-Logistic Models of Spatial Choice, in *Spatial Choices and Processes*, edited by M. M. Fischer, P. Nijkamp and Y. Papageorgiou, North-Holland, Amsterdam.

Appendix

Here we provide some evidence in support of our contention that the little extra computation necessary to implement the hybrid chain, basically the evaluation of (14), is well worth its cost.

The results reported in Section 4 were obtained using the hybrid Markov chain sampler. We repeated the simulation using the Gibbs sampler with data augmentation, with the same stream of random numbers. Figure 1 displays time series plots of simulated values for some parameters in the model, obtained using the hybrid chain (darker dots) and Gibbs sampling (lighter dots). Both samplers were run for 50000 cycles, the plots display a subsample of 1 every 4.

For about a half of the identified parameters in the model, time series plots of the simulated values obtained with the two methods exhibited quite similar characteristics, even though, typically, the Gibbs sampler simulated paths were more strongly autocorrelated. An example is in panel (b) of Figure 1, which refers to σ_{22}/σ_{11}. For these cases, inferences based on the two sampling methods did not differ much. However, for other parameters, there were clear discrepancies. For instance, in panel (a) of Figure 1 the simulated values of $\delta_{1,CITY}/\sqrt{\sigma_{11}}$ from the Gibbs sampler are consistently larger than those from the hybrid chain. Similar differences were present for the other location variables, for the number of licenses variables and for the means of the random effects. Differences were really marked for the correlation coefficients in $\mathbf{\Sigma}$ and $\mathbf{V}$. Part (c) of Figure 1 presents time series plots for $\sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$. Notice how the properties of time series from the Gibbs sampler change as the simulation progresses. A similar pattern was exhibited by the values of $v_{12}/\sqrt{v_{11}v_{22}}$.

An examination of the absolute scale of the sampled parameter values may help understand how these differences might arise. This is provided in panels (d) and (e) of Figure 1, both containing the simulated values of $\sqrt{\sigma_{11}}$ produced by the hybrid chain and the Gibbs sampler, respectively. From these plots one evinces that, even though the two Markov chains yield overall similar results in terms of identified parameters, they are actually exploring different regions of the "unidentified" parameter space. The hybrid chain stabilizes rather quickly on values of $\sqrt{\sigma_{11}}$ close to 0.35, while the Gibbs sampler at first oscillates around values near 1, then shows a tendency towards larger

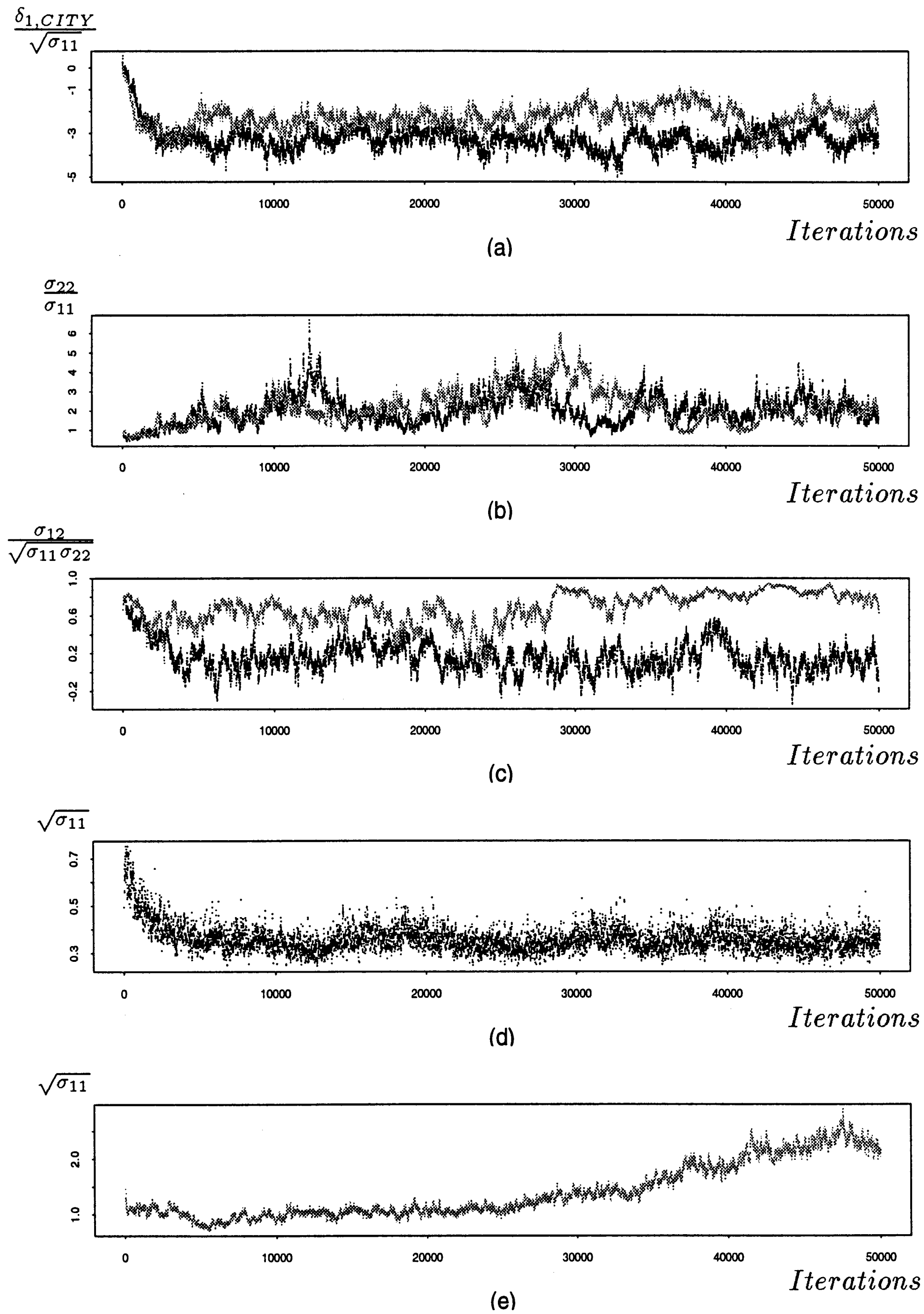

Figure 1: Time series plots of simulated values for some identified parameters: (a) coefficient of CITY in equation for 1 car; (b) ratio of the variances of ϵ_{it2} and ϵ_{it1}; (c) correlation coefficient of ϵ_{it1} and ϵ_{it2}; (d) and (e) standard deviation of ϵ_{it1}. Darker dots represent the output of the hybrid Markov chain, lighter dots the output of the Gibbs sampler with data augmentation.

values.

Now, as we have shown in Section 3.2, the likelihood contains no information about the overall scale of the parameter vector, which depends on the prior distribution only. Therefore, one can ascertain whether the sampler has wondered in a region of low posterior probability by comparing the prior density at the current state of the Markov chain with the prior density at rescaled versions of the current state. We considered the last parameter vector ϑ_H obtained from the hybrid chain and computed the scale factor c_H which maximize $r_H = \pi(c_H \vartheta_H)/\pi(\vartheta_H)$. This yielded $c_H = 1.077$ and $r_H = 1.369$. The corresponding values for the Gibbs sampler were $c_G = 0.231$ and $r_G = 1.678 \cdot 10^{23}$, suggesting that it spent the last part of the simulation in the tails of the posterior distribution. Therefore, in the present situation, unless one feels as confident about the tail behavior of a high dimensional prior distribution as one does about its behavior near the mode, it is difficult to imagine preferring the Gibbs sampler over the hybrid chain, when making inferences about the "identified" parameters in the model.

Variable	1 car		2 or more cars	
λ	-1.50	(-2.24 , -0.75)	-6.89	(-9.86 , -4.23)
WAVE 5	-0.13	(-0.37 , 0.09)	-0.53	(-0.98 , -0.12)
WAVE 7	-0.17	(-0.41 , 0.08)	-0.62	(-1.10 , -0.20)
WAVE 9	-0.02	(-0.27 , 0.23)	-0.16	(-0.60 , 0.27)
CITY	-3.34	(-4.21 , -2.57)	-3.84	(-5.17 , -2.56)
SUBURB	-2.34	(-3.20 , -1.55)	-2.30	(-3.54 , -1.03)
1 LICENSE	4.79	(4.26 , 5.37)	3.39	(2.25 , 4.82)
2 LICENSES	5.97	(5.33 , 6.70)	6.81	(5.28 , 8.82)
3+ LICENSES	6.42	(5.15 , 7.85)	9.37	(7.05 , 12.12)
INCOME	0.39	(0.26 , 0.53)	0.69	(0.43 , 0.97)
# WORKERS	0.05	(-0.15 , 0.25)	0.45	(0.12 , 0.82)
2 ADULTS	0.91	(0.56 , 1.27)	0.53	(-0.30 , 1.38)
3+ ADULTS	1.07	(0.34 , 1.82)	2.96	(1.49 , 4.68)
# KIDS < 12	-0.11	(-0.26 , 0.04)	-0.14	(-0.40 , 0.11)
# KIDS 12-17	0.12	(-0.11 , 0.36)	0.25	(-0.11 , 0.62)
# KIDS > 17	-0.03	(-0.31 , 0.27)	1.04	(0.58 , 1.60)
σ_{22}/σ_{11}	2.12	(1.04 , 3.76)		
$\sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$	0.14	(-0.12 , 0.42)		
v_{11}/σ_{11}	5.41	(4.23 , 6.88)		
v_{22}/σ_{11}	12.18	(6.58 , 21.99)		
$v_{12}/\sqrt{v_{11}v_{22}}$	0.41	(0.22 , 0.62)		

Table 1: Means and 95% probability intervals of the marginal posterior distributions of the parameters.