A Note on Identification in the Multinomial Probit Model in the Presence of Weak Correlation

Murat K. Munkin
Department of Economics
4202 E Fowler Avenue, CMC 207E
University of South Florida
Tampa, FL, 33620, U.S.A.
Email: mmunkin@usf.edu
Phone: +1 (813) 974 6517

June 27, 2023

Abstract

An identification problem with the Multinomial Probit Model arises when correlations among choice alternatives are weak. Then even if formal conditions for model identification are satisfied, practical estimation of the unrestricted variance parameters is tenuous. An estimable specification of the model, in which all variance parameters are fixed, resolves the problem. This paper demonstrates the identification problem and presents an MCMC algorithm to estimate this restricted specification of the Multinomial Probit model.

Key words: Multinomial Probit Model, Tenuous Identification, Estimable Specification, Bayesian Estimation.

JEL Codes: C11, C14

1. Introduction

Parameter identification in the Multinomial Probit (MNP) model can be tenuous even if formal identification conditions are satisfied as shown by Keane (1992). Specifically, the log-likelihood function in the MNP model is flat in the absence of exclusion restrictions. This paper presents yet another identification problem which occurs in the presence of weak correlation. Then identification of the unrestricted variance parameters is difficult in the same sense that the log-likelihood function does not change much for a wide range of variance values. A possible solution to this problem is to present an estimable specification of the MNP model which restricts all variance parameters. This paper develops an MCMC estimation method for the restricted MNP model and shows that it produces Markov chains with good mixing properties uniform across different parameter values and numbers of alternatives.

The MNP model was introduced by Aitchison and Bennett (1970) to model choices among several alternatives, which cannot be ranked uniformly for all individuals from most preferred to least. Since the observed outcomes are defined by differences in levels of unobservable random utility, not all parameters may be identified. Several alternative identification restrictions are possible, as presented, for example, by Bunch (1991). It is customary for formal identification to normalize the utility of one alternative to 0 and restrict one variance parameter of the covariance matrix to 1. Heckman and Sedlacek (1985) find that for formal identification in the trinomial probit model (TPM) it is also necessary that the linear-in-parameters latent utility contains a single regressor that varies over individuals.

Keane (1992) raises a very important question of practical estimability of the MNP model in the cases when formal identification restrictions are satisfied, and shows that in the absence of exclusion restrictions, identification in the TPM model is fragile in the sense that the objective function is flat with respect to the parameters of the covariance matrix. Practical identification, however, has no problem in the presence of exclusion restrictions. This restriction might be just a single case among several identification issues in practice caused by different reasons. For example, Bunch and Kitamura (1991) point out that the flexible covariance structure of the MNP model requires "huge datasets" as the number of alternatives increases, but the issue of overparametrization is an open research question. McCulloch et al. (2000) present an MCMC estimation method for the MNP model, but find that in some cases of high dimensions and situations, when the likelihood is not very informative, their Markov chains fail to converge. They do not provide additional details on such cases. This paper shows that even in the case of the TPM model of low dimension when correlation is weak ($\rho \leqslant 0.2$), the algorithm of McCulloch et. al (2000) fails to converge with only one covariance and one variance parameter. This convergence problem is caused by fragile identification occurring when the log-likelihood function is flat with respect to the variance parameter in the presence of weak correlation.

Geweke et al. (1994) compare several approaches to inference in the MNP model including simulated maximum likelihood, method of simulated moments and Gibbs sampling with data augmentation and find that the Gibbs sampling algorithm performs relatively better. Bayesian MCMC estimation algorithms of the MNP model include Albert and Chib (1993), McCulloch and Rossi (1994), Nobile (1998), McCulloch et al. (2000), Nobile (2000), Imai and van Dyk (2005) and Burgette and Nordheim (2012). McCulloch and Rossi (1994) estimate the MNP model without placing any identification restrictions on the covariance matrix, specifying a single Wishart prior for both identifiable and unidentifiable parameters, which means that a direct improper prior on the identifiable parameters is not possible. McCulloch et al. (2000) use a reparametrization that allows to make an identifying restriction and specify improper priors on the identifiable parameters. The produced Markov chains, however, are somewhat tenuous to converge. As reported by McCulloch et al. (2000) and Nobile (2000), the

MCMC algorithm can be much slower to converge than either the procedure of McCulloch and Rossi (1994) or Nobile (1998). Imai and van Dyk (2005) introduce a set of MCMC algorithms based on the method of marginal data augmentation to estimate the MNP model. The produced algorithm has better computational properties in terms of autocorrelations and sensitivity to starting values. Burgette and Nordheim (2012) offer a different identification strategy of the MNP model, in which instead of fixing a diagonal element of the covariance matrix they restrict its trace. The mixing properties of the algorithm are reported to be similar to those of Imai and van Dyk (2005).

The rest of the paper is organized as follows. Section 2 designs a numerical example to show that the unrestricted variance parameter is difficult to identify in the presence of weak correlation in the trinomial specification of the MNP model. Section 3 restricts the variance parameter, develops an MCMC procedure for the restricted TPM model and compares it with the algorithm of McCulloch et al. (2000). Section 4 introduces an estimable specification of the MNP model with any number of alternatives, restricting all variance parameters, and develops an MCMC procedure for the general case. Section 5 concludes.

2. Weak Correlation in the TPM model

This section describes an identification problem in the TPM model that arises when the correlation parameter in the covariance matrix is small in value ($\rho \leqslant 0.2$). To provide evidence of fragile identification I construct a numerical example designed in a similar way to Keane (1992), who specifies latent utilities Z_{1i} and Z_{2i} as

$$\begin{aligned} Z_{1i} &= X_{1i}\alpha_1 + u_{1i}, \\ Z_{2i} &= X_{2i}\alpha_2 + u_{2i}, \end{aligned} \tag{2.1}$$

where X_{1i} and X_{2i} are different vectors of covariates, α_1 and α_2 are conformable parameter vectors and N observations are independent over i ($i = 1, ..., N$). Latent utility Z_3

is restricted for identification to zero $(Z_3 \equiv 0)$. The errors (u_{1i}, u_{2i}) are independently and identically distributed as $N(0, \Sigma)$ where

$$\Sigma = \begin{pmatrix} 1 & \rho\sigma_2 \\ \rho\sigma_2 & \sigma_2^2 \end{pmatrix}$$

and variance parameter σ_1 is restricted to 1 for identification since Z_{1i} and Z_{2i} are latent. The observed outcome variables d_{1i}, d_{2i}, d_{3i} are defined as

$$\begin{aligned} d_{1i} &= 1 \text{ if and only if } Z_{1i} \geqslant Z_{2i},\ Z_{1i} \geqslant 0, \\ d_{2i} &= 1 \text{ if and only if } Z_{2i} > Z_{1i},\ Z_{2i} \geqslant 0, \\ d_{3i} &= 1 \text{ if and only if } Z_{1i} < 0,\ Z_{2i} < 0. \end{aligned}$$

The probabilities of these outcomes can be expressed as bivariate integrals. Denote bivariate probability

$$\Pr(X \leqslant x, Y \leqslant y) = \Phi(x, y, \rho)$$

where random variables X and Y are distributed bivariate normal

$$(X, Y) \sim N\left((0,0), \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

Then

$$\begin{aligned} \Pr(d_{1i} = 1) &= \Pr(Z_{1i} > Z_{2i}; Z_{1i} > 0) = \Phi\left(\frac{X_{1i}\alpha_1 - X_{2i}\alpha_2}{\sqrt{\sigma_2^2 + 1 - 2\sigma_{12}}}, X_{1i}\alpha_1, \frac{1 - \sigma_{12}}{\sqrt{\sigma_2^2 + 1 - 2\sigma_{12}}}\right), \\ \Pr(d_{2i} = 1) &= \Pr(Z_{2i} > Z_{1i}; Z_{2i} > 0) = \Phi\left(\frac{X_{2i}\alpha_2 - X_{1i}\alpha_1}{\sqrt{\sigma_2^2 + 1 - 2\sigma_{12}}}, \frac{X_{2i}\alpha_2}{\sigma_2}, \frac{\sigma_2^2 - \sigma_{12}}{\sigma_2\sqrt{\sigma_2^2 + 1 - 2\sigma_{12}}}\right), \\ \Pr(d_{3i} = 1) &= \Pr(Z_{1i} < 0; Z_{2i} < 0) = \Phi\left(-X_{1i}\alpha_1, \frac{-X_{2i}\alpha_2}{\sigma_2}, \frac{\sigma_{12}}{\sigma_2}\right). \end{aligned}$$

Most statistical packages have very efficient procedures to numerically calculate bivariate integrals $\Phi(x, y, \rho)$, which do not have a closed form solution. These expressions are used in the maximum likelihood (ML) programs for the probabilities of the observed outcomes in the TPM model.

The ML programs estimating the TPM model, used in this paper, have been tested with data, generated according to Keane (1992) such that $N = 8000$, $X_{1i} \sim N(6,5)$, $X_{2i} \sim N(6,5)$, $\alpha_1 = (-0.8, 0.2)$, $\alpha_2 = (-2.0, 0.4)$, $\rho = 0.6$, $\sigma_2 = 1.5$. However, in my model specification parameter $\sigma_{12} = \rho\sigma_2$ is used instead of ρ. The obtained results produce a similar conclusion that the log-likelihood function does not change much for a range of different values of covariance parameters σ_{12} and σ_2^2 when X_{1i} and X_{2i} are restricted to be the same. Next, I generate zero correlation data with the following specification: $N = 8000$, $X_{1i} \sim N(4,5)$, $X_{2i} \sim N(4,5)$, $\alpha_1 = (-0.8, 0.2)$, $\alpha_2 = (-2.0, 0.4)$, $\rho = 0$, $\sigma_2 = 1.5$. The true value $\rho = 0.6$ in Keane (1992) produces outcomes with approximately equal relative frequencies. However, when ρ is restricted to zero the generated proportions are far from being equal. The adjustment in drawing X_1 and X_2 as $N(4,5)$ instead of $N(6,5)$ produces approximately equal proportions of observations in each of the three choice groups.

Table 1 presents estimation results for several specifications. The true values of all parameters in the data generating process are given in the column titled "True Value". The column titled "ML" presents maximum likelihood estimates of the TPM model with no restrictions imposed on the log-likelihood function. The numbers in the first rows are the ML estimates, and the second rows are the corresponding standard errors. Additionally, restricted ML estimates of the TPM model, fixing parameter σ_2 to a range of values from 1.3 to 2.9, are given in columns (1) to (5). The estimated log-likelihood values are presented in the row titled "$\log \widehat{L}$". To test these restrictions the last row of the table reports the likelihood ratio (LR) test statistic $\log \widehat{L}_U - \log \widehat{L}_R$ (U and R stand for unrestricted and restricted). The critical value of the test statistic at 10% level of significance is $\chi^2_{0.1}(1) = 2.706$, so that the null hypothesis that the restriction of σ_2 is valid cannot be rejected for all cases presented in Table 1.

Figure 1 shows how the restricted log-likelihood (the red line) deteriorates from the unrestricted maximum when variance parameter σ_2 is fixed at values changing from 1 to 3 in increments of 0.1. The same log-likelihood values are also given in Table 2 in the column titled "$\rho = 0$". The unrestricted model produces $\log \widehat{L}_{ML} = -7764.17$ with the ML estimate $\widehat{\sigma}_{2,ML} = 1.85$ (0.30), where the red line is maximized. The lower bound for the log-likelihood values, above which validity of σ_2 restriction cannot be rejected, is $\log \widehat{L}_{ML} - \chi^2_{0.1}(1) = -7766.87$ (the blue line). The corresponding "no rejection" interval in this case is $[1.3, 2.9]$. It is defined with respect to σ_2 and referred to as "LR interval" in Table 2 with the lower and upper bounds rounded up to 0.1. The log-likelihood values corresponding to this interval in Table 2 are in boldface. The left tail of the log-likelihood function (the red line) around the ML estimate $\widehat{\sigma}_{2,ML}$ is relatively shorter with values decreasing fast, however, the right tail persists being above the critical value much longer. Since parameter σ_2 is formally identified it produces some slight deteriorations of the objective function, however, it remains relatively "flat" for a very large range of values, $[1.3, 2.9]$. Given that $\widehat{\sigma}_{2,ML} = 1.85$ (0.30) when the true value is $\sigma_2 = 1.5$, the fact that the restriction $\sigma_2 = 2.9$ cannot be rejected is a sign that identification of σ_2 in practice is difficult.

The shapes of the log-likelihood functions change with larger values of ρ. Figures 2 and 3 present them for the cases when $\rho = 0.6$ and $\rho = 0.9$ respectively. Table 2 also provides estimates of the corresponding log-likelihoods. The unrestricted maximum likelihood estimates $\widehat{\sigma}_{2,ML} = 1.65$ (0.17) and $\widehat{\sigma}_{2,ML} = 1.61$ (0.11) move closer to the true value of 1.5 with standard deviations becoming markedly smaller. Thus, ML estimates of the variance parameter become more precise. The "no rejection" LR intervals narrow to $[1.4, 2.1]$ and $[1.4, 1.8]$ respectively. The right tails shorten and the shapes become more symmetric with a distinct maximum.

The null hypothesis for validity of the restrictions for the variance parameter σ_2 can also be tested based on the asymptotic normality of the maximum likelihood estimator $\widehat{\sigma}_{2,ML}$, centered at 1.61 with standard deviation 0.11 when $\rho = 0.9$. Then, a confidence interval at 10% level of significance can be estimated as $1.61 \pm z_{0.05} \times 0.11$. Table

2 refers to such intervals estimated based on asymptotic normality as "ML intervals". Rounding the lower and upper bounds up to 0.1 results in $[1.4, 1.8]$. Thus, in the case when $\rho = 0.9$ both likelihood ratio (LR) and asymptotic normality (ML) approaches give practically equal estimates. The objective function does not have any signs of being "flat" and identification of σ_2 has no problems.

Due to computational errors and different asymptotic properties of the test statistics the estimated LR and ML intervals are not expected to be identical. However, if they differ by a large margin and inference based on them leads to completely different conclusions, this could be interpreted as a sign of potential identification problems. In the case when $\rho = 0$, the no rejection ML interval at 10% level of significance is $[1.4, 2.3]$, estimated as $1.852 \pm z_{0.05} \times 0.305$, which is substantially narrower than the LR interval $[1.3, 2.9]$. A possible explanation for this is that ML optimization and estimated standard errors are based on the inverse of the Hessian matrix, which produces imprecise estimates when the objective function flattens.

The next results in Table 2 show that the identification problem is present in other cases of weak correlation. The no rejection LR intervals $[1.4, 2.9]$ and $[1.4, 2.8]$ for the cases when $\rho = 0.1$ and $\rho = 0.2$, respectively, are not much different from that of zero correlation. The no rejection ML intervals are substantially smaller, $[1.3, 2.4]$ and $[1.4, 2.3]$, respectively. The standard errors of $\widehat{\sigma}_{2,ML}$ are as high as in the case of zero correlation, 0.32 ($\rho = 0.1$) and 0.30 ($\rho = 0.2$), indicating that the objective function remains "flat". Thus, it can be stated that identification of the variance parameter σ_2 is tenuous at least for the cases when $\rho \leqslant 0.2$, however, identification of σ_2 is not a problem when $\rho \geqslant 0.8$. The cases in between, when $0.2 < \rho < 0.8$, produce mixed results. What is important is the conclusion that for some parameter values consistent with weak correlation ($\rho \leqslant 0.2$) identification is difficult. This paper offers a practical solution to this identification problem by presenting an estimable specification of the TPM model, which fixes all variance parameters including σ_2, and extends this specification to the

MNP model of any number of alternatives.

3. Estimable TPM Specification: MCMC Estimation

First, before presenting an estimable specification of the TPM model, define Σ without placing any identification restrictions

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}.$$

The joint distribution of latent variables in system (2.1) can be written as the conditional distribution $Z_{1i}|Z_{2i}$ and marginal distribution of Z_{2i}, where $Z_{1i}|Z_{2i} \sim N(X_{1i}\alpha_1 + \sigma_{12}\sigma_{22}^{-1}(Z_{2i} - X_{2i}\alpha_2), \sigma_{11} - \sigma_{12}^2\sigma_{22}^{-1})$ and $Z_{2i} \sim N(X_{2i}\alpha_2, \sigma_{22})$. Denote $\phi_{11} = \sigma_{11} - \sigma_{12}^2\sigma_{22}^{-1}$, $\phi_{12} = \sigma_{12}\sigma_{22}^{-1}$, $\phi_{22} = \sigma_{22}$. There is a one-to-one correspondence between parameters $(\sigma_{11}, \sigma_{21}, \sigma_{22})$ and $(\phi_{11}, \phi_{12}, \phi_{22})$. Then the model can be presented as

$$\begin{aligned} Z_{1i} &= X_{1i}\alpha_1 + (Z_{2i} - X_{2i}\alpha_2)\phi_{12} + \varepsilon_{1i}, \\ Z_{2i} &= X_{2i}\alpha_2 + \varepsilon_{2i}, \end{aligned}$$

where

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \overset{i.i.d.}{\sim} N\left(0, \begin{pmatrix} \phi_{11} & 0 \\ 0 & \phi_{22} \end{pmatrix}\right).$$

Now constraints can be imposed on the original variance parameters σ_{11} and σ_{22} by setting $\phi_{11} = 1$ and $\phi_{22} = 1$. Then $\sigma_{11} = 1 + \sigma_{12}^2\sigma_{22}^{-1}$ and $\sigma_{22} = 1$ are the additional identifying restrictions. This is the estimable specification of the TPM model.

Next I develop an MCMC algorithm to estimate this restricted model. Denote $d_i = (d_{1i}, d_{2i}, d_{3i})$ and $\Delta_i = (X_{1i}, X_{2i}, \alpha_1, \alpha_2, \phi_{12})$. For each observation i the joint density of the observed data and latent variables is

$$\begin{aligned} \Pr(Z_{1i}, Z_{2i}, d_i|\Delta_i) &= (2\pi)^{-1} \exp\left[-0.5\left(Z_{1i} - X_{1i}\alpha_1 - (Z_{2i} - X_{2i}\alpha_2)\phi_{12}\right)^2\right] \\ &\quad \times \exp\left[-0.5(Z_{2i} - X_{2i}\alpha_2)^2\right] \\ &\quad \times \left[\sum_{j=1}^{3} d_{ji} \prod_{k=1}^{3} I_{[0,+\infty)}(Z_{ji} - Z_{ki})\right], \end{aligned}$$

where $I_{[0,+\infty)}$ is the indicator function for the set $[0,+\infty)$. The joint distribution of the observed and latent variables for all observations is the product of N such independent terms over i $(i = 1, ...N)$. The posterior density is proportional to the product of the prior density of the parameters and the joint distribution of the observed and included latent variables. Proper priors are selected for all parameters. The prior distributions for parameters α_1 $(k_1 \times 1)$, α_2 $(k_2 \times 1)$ and ϕ_{12} (1×1) are normal $N(\underline{\alpha}_1, \underline{H}_{\alpha_1}^{-1})$, $N(\underline{\alpha}_2, \underline{H}_{\alpha_2}^{-1})$ and $\phi_{12} \sim N\left(\underline{\phi}, \underline{H}_{\phi}^{-1}\right)$ respectively, centered at zero with relatively large variance

$$\alpha_1 \sim N\left(0, 10^2 I_{k_1}\right),\ \alpha_2 \sim N\left(0, 10^2 I_{k_2}\right),\ \phi_{12} \sim N\left(0, 10^2\right).$$

The parameter set is blocked as $[Z_{1i}, Z_{2i}]$, $[\alpha_1, \phi_{12}]$ and α_2. The details of the MCMC algorithm are given in Appendix A1.

In order to assess performance of the developed MCMC algorithm I utilize the same zero correlation data set which was generated to produce ML results presented in Table 1. However, for comparison I also estimate the model with unrestricted variance parameter σ_2, using the MCMC procedure developed by McCulloch et al. (2000) referred to as the MPR (McCulloch, Polson and Rossi) method. The priors for the MPR method are specified to be

$$\begin{aligned} \alpha_1 &\sim N\left(0, 10^2 I_{k_1}\right),\ \alpha_2 \sim N\left(0, 10^2 I_{k_2}\right), \\ \phi_{12} &\sim N\left(0, 10^2\right),\ \sigma_2^{-1} \sim W(5, 1). \end{aligned}$$

The MCMC procedure for the restricted model, presented in this section, is referred to as the M (Munkin) method. The results are given in Table 3 and based on 200,000 replications following 1000 replications of the burn-in phase. The reason for such a large number of replications is that the Markov chains produced by the MPR method display very high serial correlations. Since the true value of σ_2 in the data generating process is 1.5 but in the M model it is restricted to 1 the corresponding posterior of α_2 should be centered around the true values divided with $\sqrt{1.5}$. Figures 4 and 5 present Markov chains and autocorrelation functions up to lag 20 for the slowest to converge parameters in the M and MPR models respectively. For example, for the MPR model the levels of autocorrelation function at lag 20 for parameter σ_2 is $\gamma_{20} = 0.97$, and relative numerical efficiency, $RNE = 0.0008$. The estimated posterior mean is 3.627 and posterior standard deviation is 1.470. However, other runs of the MPR algorithm produce Markov chains that center around different posterior means with different standard deviations. The Markov chain for parameter σ_2 fails to converge. McCulloch et al. (2000) use a tighter prior on parameter $\phi_{12} \sim N(0, 1/8)$, however, the MPR algorithm fails to converge in that case as well. This result is consistent with the conclusion based on the ML estimates, that identification of the unrestricted variance parameter is tenuous. Inference based on Markov chains with such poor mixing properties would be unreliable.

It is interesting to notice that Markov chains of the MPR algorithm for parameter vector α_1 display reasonable convergence properties, however, for parameter α_2 the autocorrelation function is similar to that of σ_2 with the intercept being more serially correlated than the slope parameter. This is likely caused by poor identification of the variance parameter σ_2 of latent variable Z_2 which affects identification of the mean parameters in Z_2. Overall, the Markov chains of the M output have much better convergence properties than those of the MPR algorithm. The covariance parameter σ_{12} has the largest lag 20 autocorrelation among all parameters, however, it is at a reasonable level of $\gamma_{20} = 0.64$ and $RNE = 0.012$. It is smaller than that of the MPR output for σ_{12} with $\gamma_{20} = 0.77$ and $RNE = 0.0019$.

Next I examine how convergence properties of the Markov chains change as the true value of ρ is moved away from 0. Data sets are generated for a range of ρ values and the estimation results are presented in Tables 4 and 5 together with estimates of γ_{20} and RNE for the M and MPR outputs respectively. Overall, the produced output of the M method is consistently similar in its good mixing properties across all ρ values. The MPR method produces Markov chains with better convergence when $\rho > 0.2$, however, the autocorrelations for parameter σ_2 are still high. Placing the tighter prior $\phi_{12} \sim N(0, 1/8)$ improves the mixing properties of the chains, but only slightly with γ_{20} still being above 0.9.

4. Estimable MNP Specification: MCMC Estimation

This section presents an estimable specification of the MNP model with any number of alternatives and develops an MCMC algorithm to estimate the posterior distribution of model parameters. Assume that we observe N independent individuals ($i = 1, ..., N$) and each individual i makes a choice among J alternatives based on latent utility Z_{ji} ($j = 1, ..., J$) defined as

$$Z_{ji} = X_{ji}\alpha_j + u_{ji}, \tag{4.1}$$

where X_{ji} is a vector of exogenous regressors specific to individual i and alternative j, α_j is a conformable vector of parameters and u_{ji} is the error. Let d_{1i}, $d_{2i}, ...,$ d_{Ji} be a set of binary random variables representing this choice defined as

$$d_{ji} = \prod_{k=1}^{J} I_{[0,+\infty)}\left(Z_{ji} - Z_{ki}\right).$$

Individual i chooses alternative j if utility level Z_{ji} exceeds those of the alternatives in which case $d_{ji} = 1$. Otherwise $d_{ji} = 0$. Even before a formal specification of the distribution for the errors u_{ji} it can be noticed that the probabilities of outcomes d_{ji} depend on pairwise differences $Z_{ji} - Z_{ki}$ or $(\alpha_j - \alpha_k)$, such that all parameters α_j ($j = 1, ..., J$) are not identifiable. For identification it is necessary to restrict one of the latent utilities to a constant, say $Z_J = 0$.

The MNP model assumes that the distribution of the error term $u_i = \left(u_{1i}, u_{2i}, ..., u_{(J-1)i}\right)'$ is $(J-1)$-variate normal $N(0, \Sigma)$. Once again just like in the TPM case before imposing any additional identifying restrictions specify symmetric matrix Σ without any constraints

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1(J-2)} & \sigma_{1(J-1)} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \cdots & \sigma_{2(J-2)} & \sigma_{2(J-1)} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \cdots & \sigma_{3(J-3)} & \sigma_{3(J-1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{1(J-2)} & \sigma_{2(J-2)} & \sigma_{3(J-2)} & \cdots & \sigma_{(J-2)(J-2)} & \sigma_{(J-2)(J-1)} \\ \sigma_{1(J-1)} & \sigma_{2(J-1)} & \sigma_{3(J-1)} & \cdots & \sigma_{(J-2)(J-1)} & \sigma_{(J-1)(J-1)} \end{pmatrix}.$$

Define an $1 \times (J-j)$ vector

$$\Sigma_{j,12} = \begin{pmatrix} \sigma_{(j-1)j} & \sigma_{(j-1)(j+1)} & \cdots & \sigma_{(j-1)(J-1)} \end{pmatrix},$$

and an $(J-j) \times (J-j)$ matrix

$$\Sigma_{j,22} = \begin{pmatrix} \sigma_{jj} & \sigma_{j(j+1)} & \cdots & \sigma_{j(J-1)} \\ \sigma_{j(j+1)} & \sigma_{(j+1)(j+1)} & & \sigma_{(j+1)(J-1)} \\ \cdots & & \cdots & \cdots \\ \sigma_{j(J-1)} & \sigma_{(j+1)(J-1)} & \cdots & \sigma_{(J-1)(J-1)} \end{pmatrix},$$

subsets of matrix Σ $(j = 2, ..., J-1)$. Following Munkin and Trivedi (2008) the joint distribution of $u_{1i}, u_{2i}, u_{3i}, ..., u_{(J-1)i}$ can be written as

$$f\left(u_{1i}, u_{2i}, u_{3i}, ..., u_{(J-1)i}\right) = f\left(u_{(J-1)i}\right) \prod_{j=2}^{J-1} f\left(u_{(j-1)i} | u_{ji}, u_{(j+1)i}, ..., u_{(J-1)i}\right)$$

where the conditional distributions of $u_{(j-1)i} | u_{ji}, u_{(j+1)i}, ..., u_{(J-1)i}$ $(j = 2, ..., J-1)$ are

$$N\left(\Sigma_{j,12}\Sigma_{j,22}^{-1}\left(u_{ji}, ..., u_{(J-1)i}\right)', \sigma_{(j-1)(j-1)} - \Sigma_{j,12}\Sigma_{j,22}^{-1}\Sigma_{j,12}'\right),$$

and $u_{(J-1)i} \sim N\left(0, \sigma_{(J-1)(J-1)}\right)$.

Introduce Φ to be a $(J-1) \times (J-1)$ matrix and define its elements as

$$\phi_{(j-1)(j-1)} = \sigma_{(j-1)(j-1)} - \Sigma_{j,12}\Sigma_{j,22}^{-1}\Sigma_{j,12}',$$

$$\begin{pmatrix} \phi_{(j-1)j} & \phi_{(j-1)(j+1)} & \cdots & \phi_{(j-1)(J-1)} \end{pmatrix} = \Sigma_{j,12}\Sigma_{j,22}^{-1}$$

and $\phi_{(J-1)(J-1)} = \sigma_{(J-1)(J-1)}$ $(j = 2, ..., J-1)$. There is a one-to-one correspondence between parameters in Σ and Φ. Then the equations in (4.1) can be written as

$$
\begin{aligned}
Z_{1i} &= X_{1i}\alpha_1 + (Z_{2i} - X_{2i}\alpha_2)\phi_{12} + ... + \left(Z_{(J-1)i} - X_{(J-1)i}\alpha_{(J-1)}\right)\phi_{1(J-1)} + \varepsilon_{1i}, \\
Z_{2i} &= X_{2i}\alpha_2 + (Z_{i3} - X_{3i}\alpha_3)\phi_{23} + ... + \left(Z_{(J-1)i} - X_{(J-1)i}\alpha_{(J-1)}\right)\phi_{2(J-1)} + \varepsilon_{2i}, \\
Z_{3i} &= X_{3i}\alpha_3 + (Z_{4i} - X_{4i}\alpha_4)\phi_{34} + ... + \left(Z_{(J-1)i} - X_{(J-1)i}\alpha_{(J-1)}\right)\phi_{3(J-1)} + \varepsilon_{3i}, \\
&\quad ... \\
Z_{(J-2)i} &= X_{(J-2)i}\alpha_{(J-2)} + \left(Z_{(J-1)i} - X_{(J-1)i}\alpha_{(J-1)}\right)\phi_{(J-2)(J-1)} + \varepsilon_{(J-2)i}, \\
Z_{(J-1)i} &= X_{(J-1)i}\alpha_{(J-1)} + \varepsilon_{(J-1)i}.
\end{aligned}
$$

where the errors are distributed as

$$
\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ ... \\ \varepsilon_{(J-2)i} \\ \varepsilon_{(J-1)i} \end{pmatrix} \overset{i.i.d.}{\sim} N\left(0, \begin{pmatrix} \phi_{11} & 0 & 0 & ... & 0 & 0 \\ 0 & \phi_{22} & 0 & ... & 0 & 0 \\ 0 & 0 & \phi_{33} & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & \phi_{(J-2)(J-2)} & 0 \\ 0 & 0 & 0 & ... & 0 & \phi_{(J-1)(J-1)} \end{pmatrix}\right).
$$

To define the estimable specification of the MNP model I choose the following variance restrictions: $\phi_{(j-1)(j-1)} = 1$ $(j = 2, ..., J)$. These impose $\sigma_{(j-1)(j-1)} = 1 + \Sigma_{j,12}\Sigma_{j,22}^{-1}\Sigma_{j,12}'$ $(j = 2, ..., J-1)$ and $\sigma_{(J-1)(J-1)} = 1$ restrictions on the diagonal elements of matrix Σ.

Denote $Z_i = (Z_{1i}, ..., Z_{Ji})$, $d_i = (d_{1i}, ..., d_{Ji})$, $X_i = \left(X_{1i}, ..., X_{(J-1)i}\right)$, $\alpha = (\alpha_1, ..., \alpha_{J-1})$, $\mathbf{\Delta}_i = (X_i, \alpha, \Phi)$. For each observation i the joint density of the observed data and latent variables is

$$
\begin{aligned}
&\Pr\left(Z_i, d_i|\mathbf{\Delta}_i\right) = (2\pi)^{-(J-1)/2} \\
&\times \exp\left[-0.5\left(Z_{1i} - X_{1i}\alpha_1 - (Z_{2i} - X_{2i}\alpha_2)\phi_{12} - ... - \left(Z_{(J-1)i} - X_{(J-1)i}\alpha_{(J-1)}\right)\phi_{1(J-1)}\right)^2\right] \\
&\times \exp\left[-0.5\left(Z_{2i} - X_{2i}\alpha_2 - (Z_{3i} - X_{3i}\alpha_3)\phi_{23} - ... - \left(Z_{(J-1)i} - X_{(J-1)i}\alpha_{(J-1)}\right)\phi_{2(J-1)}\right)^2\right] \\
&\times \exp\left[-0.5(Z_{3i} - X_{3i}\alpha_3 - (Z_{4i} - X_{4i}\alpha_4)\phi_{34} - ... - \left(Z_{(J-1)i} - X_{(J-1)i}\alpha_{(J-1)}\right)\phi_{3(J-1)})^2\right] \\
&... \\
&\times \exp\left[-0.5(Z_{(J-2)i} - X_{(J-2)i}\alpha_{(J-2)} - \left(Z_{(J-1)i} - X_{(J-1)i}\alpha_{(J-1)}\right)\phi_{(J-2)(J-1)})^2\right] \\
&\times \exp\left[-0.5(Z_{(J-1)i} - X_{(J-1)i}\alpha_{(J-1)})^2\right] \\
&\times \left[\sum_{j=1}^{J} d_{ji} \prod_{k=1}^{J} I_{[0,+\infty)}\left(Z_{ji} - Z_{ki}\right)\right].
\end{aligned}
$$

The parameters in the model are blocked as Z_{1i}, Z_{2i}, Z_{3i}, ..., $Z_{(J-2)i}$, $Z_{(J-1)i}$, $\left[\alpha_1, \phi_{12}, ..., \phi_{1(J-1)}\right]$, α_2, α_3, ..., α_{J-2}, α_{J-1}, $\left[\phi_{23}, ..., \phi_{2(J-1)}\right]$, $\left[\phi_{34}, ..., \phi_{3(J-1)}\right]$, ..., $\left[\phi_{(J-2)(J-1)}\right]$. The details of the MCMC algorithm including formal specification of the prior distributions are given in Appendix A2. However, the priors are chosen to be similar to those in the TPM model, normally distributed centered at zero and with large variance 10^2.

Next I examine properties of Markov chains produced by the MCMC algorithm developed in this section generating data according to the MNP model for the cases when $J = 4$ and $J = 5$ respectively. Once again the numerical examples are chosen to have data generating specifications similar to those of Keane (1992). Specifically I select $N = 8000$, $X_{ji} \overset{i.i.d.}{\sim} N\left(0, 5\right)$, $\alpha_j = (-0.5, 1)$ for $i = 1, ..., N$ and $j = 1, ..., J-1$. The true values of the variance and covariance parameters are chosen to be

$$
\begin{aligned}
\sigma_{12} &= \sigma_{13} = \sigma_{23} = 0, \\
\sigma_1 &= 1, \sigma_2 = 1.3, \sigma_3 = 1.6
\end{aligned}
$$

and

$$\begin{aligned}\sigma_{12} &= \sigma_{13} = \sigma_{14} = \sigma_{23} = \sigma_{24} = \sigma_{34} = 0, \\ \sigma_1 &= 1, \sigma_2 = 1.2, \sigma_3 = 1.4, \sigma_4 = 1.6\end{aligned}$$

when $J = 4$ and $J = 5$ respectively. The selected specifications and parameter values generate approximately equal numbers of observations in each choice category. The estimation results are presented in Table 6 which includes the posterior means and standard deviations, autocorrelation at lag 20 denoted as γ_{20} and RNE of the parameters. Posterior means of all covariance parameters are indistinguishable from zeros given the estimated posterior standard deviations. It must be noted that the posterior means of parameters α_j $(j = 1, ..., J-1)$ do not have to be centered at $(-0.5, 1)$ since variances σ_j in the restricted model are fixed at 1, different from the true values in the data generating specification.

Assessment of the convergence properties of the produced Markov chains are based on the values of the autocorrelation function and relative numerical efficiency. In the case of the TPM model (Table 3) there is a single covariance parameter σ_{12} for which $\gamma_{20} = 0.64$ and $RNE = 0.012$ in the case when $\rho = 0$. Table 4 calculates the same convergence statistics for a range of ρ values and shows that γ_{20} for σ_{12} overall fluctuates between 0.59 and 0.69. The Markov chains for the cases of $J = 4$ and $J = 5$ have similar convergence properties as in the case of $J = 3$ with the values of γ_{20} and RNE being in the same range, although Table 6 presents results only for a single set of parameter values when all covariances are restricted to 0. I do not report convergence statistics similar to those in Table 4 since the number of covariance parameters increases with J and the quantity of sets of results is large. However, choosing various values for the covariance parameters produces Markov chains with similar convergence characteristics as in Table 6. As the number of alternatives in the MNP model increases the mixing properties of the Markov chains do not deteriorate.

5. Conclusion

An important conclusion from the work of Keane (1992) is that even though formal conditions for identification in the MNP model are satisfied, additional restriction might be necessary for practical identification. This paper finds a practical identification problem with the MNP model which arises in the presence of weak correlation among alternatives. This might be a reason why some proposed in the literature Bayesian estimation methods produce Markov chains that display slow convergence and in some cases fail to converge as reported, for example, by McCulloch et al. (2000). This paper presents an estimable specification of the MNP model, restricting all variance parameters, and develops an MCMC procedure to estimate this restricted specification. The Markov chain output produced by the estimation procedure shows good mixing properties consistent with respect to the numbers of alternatives and different parameter values.

Appendix A1

The steps of the MCMC algorithm are the following:

1. The latent variable Z_{1i} $(i = 1, ...N)$ is conditionally independent with normal distribution $Z_{1i} \overset{iid}{\sim} N(\overline{Z}_{1i}, 1)$ where

$$\overline{Z}_{1i} = X_{1i}\alpha_1 + (Z_{2i} - X_{2i}\alpha_2)\,\phi_{12},$$

   and subject to

$$\begin{aligned} Z_{1i} &\geqslant \max\{Z_{2i}, 0\} \text{ if } d_{1i} = 1, \\ Z_{1i} &< \max\{Z_{2i}, 0\} \text{ if } d_{1i} = 0. \end{aligned}$$

   The latent vectors Z_{2i} $(i = 1, ...N)$ are conditionally independent with the normal distribution $Z_{2i} \overset{iid}{\sim} N(\overline{Z}_{2i}, \overline{H}^{-1})$ where

$$\begin{aligned} \overline{H} &= 1 + \phi_{12}^2, \\ \overline{Z}_{2i} &= X_{2i}\alpha_2 + \overline{H}^{-1}\phi_{12}\,(Z_{1i} - X_{1i}\alpha_1)\,, \end{aligned}$$

   and truncated such that

$$\begin{aligned} Z_{2i} &\geqslant \max\{Z_{1i}, 0\} \text{ if } d_{2i} = 1, \\ Z_{2i} &< \max\{Z_{1i}, 0\} \text{ if } d_{2i} = 0. \end{aligned}$$

2. Let $W_i = (X_{1i}, (Z_{2i} - X_{2i}\alpha_2))$, $\theta' = (\alpha_1', \phi_{12})$. Given the prior distribution of $\alpha_1 \sim N(\underline{\alpha}_1, \underline{H}_{\alpha_1}^{-1})$ and $\phi_{12} \sim N\left(\underline{\phi}, \underline{H}_{\phi}^{-1}\right)$ form the prior for $\theta \sim \mathcal{N}(\underline{\theta}, \underline{H}_{\theta}^{-1})$ such that

$$\begin{aligned} \underline{\theta} &= \begin{pmatrix} \underline{\alpha}_1 \\ \underline{\phi} \end{pmatrix}, \\ \underline{H}_{\theta} &= \begin{pmatrix} \underline{H}_{\alpha_1} & 0 \\ 0 & \underline{H}_{\phi} \end{pmatrix}. \end{aligned}$$

Then the full conditional distribution of θ is $\theta \sim N(\bar{\theta}, \overline{H}_\theta^{-1})$ where

$$\begin{aligned}
\overline{H}_\theta &= \underline{H}_\theta + \sum_{i=1}^{N} W_i' W_i \\
\bar{\theta} &= \overline{H}_\theta^{-1} \left[\underline{H}_\theta \underline{\theta} + \sum_{i=1}^{N} W_i' Z_{1i} \right].
\end{aligned}$$

3. Given the prior distribution of α_2, $N(\underline{\alpha}_2, \underline{H}_{\alpha_2}^{-1})$, the full conditional distribution of α_2 is $\alpha_2 \sim \mathcal{N}(\overline{\alpha}_2, \overline{H}_\alpha^{-1})$ where

$$\begin{aligned}
\overline{H}_{\alpha_2} &= \underline{H}_\alpha + \sum_{i=1}^{N} X_{2i}'(1+\phi_{12}^2) X_{2i} \\
\overline{\alpha}_2 &= \overline{H}_{\alpha_2}^{-1} [\underline{H}_{\alpha_2} \underline{\alpha}_2 + \sum_{i=1}^{N} X_{2i}'(1+\phi_{12}^2) Z_{2i} - X_{2i}\phi_{12} \left(Z_{1i} - X_{1i}\alpha_1 \right).
\end{aligned}$$

This concludes the MCMC algorithm.

Appendix A2

The steps of the MCMC algorithm are the following:

1. The latent variable Z_{1i} $(i = 1, ...N)$ is conditionally independent with normal distribution $Z_{1i} \overset{iid}{\sim} \mathcal{N}(\overline{Z}_{1i}, 1)$ where

$$\overline{Z}_{i1} = X_{1i}\alpha_1 + (Z_{2i} - X_{2i}\alpha_2)\phi_{12} + (Z_{3i} - X_{3i}\alpha_3)\phi_{13} + ... + \left(Z_{(J-1)i} - X_{(J-1)i}\alpha_{(J-1)} \right) \phi_{1(J-1)}.$$

The latent vectors Z_{2i} $(i = 1, ...N)$ are conditionally independent with the normal distribution $Z_{2i} \overset{iid}{\sim} \mathcal{N}(\overline{Z}_{2i}, \overline{H}_2^{-1})$ where

$$\begin{aligned}
\overline{H}_2 &= 1 + \phi_{12}^2, \\
\overline{Z}_{i2} &= X_{i2}\alpha_2 \\
&\quad + \overline{H}_2^{-1} [\phi_{12} \left(Z_{1i} - X_{1i}\alpha_1 - (Z_{3i} - X_{3i}\alpha_3)\phi_{13} - ... - \left(Z_{(J-1)i} - X_{(J-1)i}\alpha_{(J-1)} \right) \phi_{1(J-1)} \right) \\
&\quad + (Z_{3i} - X_{3i}\alpha_3)\phi_{23} + ... + \left(Z_{(J-1)i} - X_{(J-1)i}\alpha_{(J-1)} \right) \phi_{2(J-1)}].
\end{aligned}$$

The latent vectors Z_{ji} $(i=1,...N;\ j=3,...,J-2)$ are conditionally independent with the normal distribution $Z_{ji} \overset{iid}{\sim} \mathcal{N}(\overline{Z}_{ji}, \overline{H}_j^{-1})$ where

$$\begin{aligned}
\overline{H}_j &= 1 + \sum_{k=1}^{j-1} \phi_{kj}^2, \\
\overline{Z}_{ji} &= X_{ji}\alpha_j + \overline{H}_j^{-1} \left[\sum_{s=1}^{j-1} \phi_{sj} \left\{ Z_{si} - X_{si}\alpha_s - \sum_{t=s+1, t\neq j}^{J-1} (Z_{ti} - X_{ti}\alpha_t)\, \phi_{st} \right\} \right. \\
&\quad \left. + \sum_{t=j+1}^{J-1} (Z_{ti} - X_{ti}\alpha_t)\, \phi_{jt} \right].
\end{aligned}$$

The latent vectors $Z_{(J-1)i}$ $(i=1,...N)$ are conditionally independent with the normal distribution $Z_{(J-1)i} \overset{iid}{\sim} \mathcal{N}(\overline{Z}_{(J-1)i}, \overline{H}_{J-1}^{-1})$ where

$$\begin{aligned}
\overline{H}_{J-1} &= 1 + \phi_{(J-2)(J-1)}^2 + ... + \phi_{2(J-1)}^2 + \phi_{1(J-1)}^2, \\
\overline{Z}_{(J-1)i} &= X_{(J-1)i}\alpha_{(J-1)} + \overline{H}_{J-1}^{-1} \\
&\quad \times \Big[\phi_{1(J-1)} \{ Z_{1i} - X_{1i}\alpha_1 - (Z_{2i} - X_{2i}\alpha_2)\, \phi_{12} - ... - (Z_{(J-2)i} - X_{(J-2)i}\alpha_{(J-2)}) \phi_{1(J-2)} \} \\
&\quad + \phi_{2(J-1)} \{ Z_{2i} - X_{2i}\alpha_2 - (Z_{3i} - X_{3i}\alpha_3)\, \phi_{23} - ... - \big(Z_{(J-2)i} - X_{(J-2)i}\alpha_{(J-2)} \big)\, \phi_{2(J-2)} \} \\
&\quad + ... \\
&\quad + \phi_{(J-3)(J-1)} \{ Z_{(J-3)i} - X_{(J-3)i}\alpha_{(J-3)} - \big(Z_{(J-2)i} - X_{(J-2)i}\alpha_{(J-2)} \big)\, \phi_{(J-3)(J-2)} \} \\
&\quad + \phi_{(J-2)(J-1)} \{ Z_{(J-2)i} - X_{(J-2)i}\alpha_{(J-2)} \} \Big]
\end{aligned}$$

Each latent variable Z_{ji} $(j=1,...,J-1)$ is truncated such that

$$\begin{aligned}
Z_{ji} &\geqslant \max\{Z_{ki} |\ k=1,...,J,\ k\neq j\} \text{ if } d_{ji}=1, \\
Z_{ji} &< \max\{Z_{ki} |\ k=1,...,J,\ k\neq j\} \text{ if } d_{ji}=0.
\end{aligned}$$

2. Let $C_{1i} = (X_{1i}, (Z_{2i} - X_{2i}\alpha_2), ..., (Z_{(J-1)i} - X_{(J-1)i}\alpha_{(J-1)}))$, $\beta_1' = \Big(\alpha_1', \phi_{12}, ..., \phi_{1(J-1)} \Big)$. Given the prior distribution for $\beta_1 \sim \mathcal{N}(\underline{\beta}_1, \underline{H}_{\beta_1}^{-1})$ the full conditional distribution

of β_1 is $\beta_1 \sim \mathcal{N}(\overline{\beta}_1, \overline{H}_{\beta_1}^{-1})$ where

$$\begin{aligned}
\overline{H}_{\beta_1} &= \underline{H}_{\beta_1} + \sum_{i=1}^{N} C_{1i}' C_{1i} \\
\overline{\beta}_1 &= \overline{H}_{\beta_1}^{-1} \left[\underline{H}_{\beta_1} \underline{\beta}_1 + \sum_{i=1}^{N} C_{1i}' Z_{1i} \right].
\end{aligned}$$

3. Given the prior distribution of α_j $(j = 2, ..., J-2)$, $\mathcal{N}(\underline{\alpha}_j, \underline{H}_{\alpha_j}^{-1})$, the full conditional distribution of α_j is $\alpha_j \sim \mathcal{N}(\overline{\alpha}_j, \overline{H}_{\alpha_j}^{-1})$ where

$$\begin{aligned}
\overline{H}_{\alpha_j} &= \underline{H}_{\alpha_j} + \sum_{i=1}^{N} X_{ji}' \left(1 + \sum_{k=1}^{j-1} \phi_{kj}^2 \right) X_{ji} \\
\overline{\alpha}_j &= \overline{H}_{\alpha_j}^{-1} [\underline{H}_{\alpha_j} \underline{\alpha}_j + \sum_{i=1}^{N} X_{ji}' \left(1 + \sum_{k=1}^{j-1} \phi_{kj}^2 \right) Z_{ji} \\
&\quad - X_{ji}' \sum_{s=1}^{j-1} \phi_{sj} \left(Z_{si} - X_{si}\alpha_s - \sum_{t=s+1, t\neq j}^{J-1} \left(Z_{ti} - X_{ti}\alpha_t \right) \phi_{st} \right) \\
&\quad - X_{ji}' \sum_{t=j+1}^{J-1} \left(Z_{ti} - X_{ti}\alpha_t \right) \phi_{jt}].
\end{aligned}$$

4. Given the prior distribution of α_{J-1}, $\mathcal{N}(\underline{\alpha}_{J-1}, \underline{H}_{\alpha_{J-1}}^{-1})$, the full conditional distribution of α_j is $\alpha_{J-1} \sim \mathcal{N}(\overline{\alpha}_{J-1}, \overline{H}_{\alpha_{J-1}}^{-1})$ where

$$\begin{aligned}
\overline{H}_{\alpha_{J-1}} &= \underline{H}_{\alpha_{J-1}} + \sum_{i=1}^{N} X_{(J-1)i}' \left(1 + \sum_{k=1}^{J-2} \phi_{k(J-1)}^2 \right) X_{(J-1)i} \\
\overline{\alpha}_{J-1} &= \overline{H}_{\alpha_{J-1}}^{-1} [\underline{H}_{\alpha_{J-1}} \underline{\alpha}_{J-1} + \sum_{i=1}^{N} X_{(J-1)i}' \left(1 + \sum_{k=1}^{J-2} \phi_{k(J-1)}^2 \right) Z_{(J-1)i} \\
&\quad - X_{(J-1)i}' \sum_{s=1}^{J-3} \phi_{s(J-1)} \left(Z_{si} - X_{si}\alpha_s - \sum_{t=s+1}^{J-2} \left(Z_{ti} - X_{ti}\alpha_t \right) \phi_{st} \right) \\
&\quad - X_{(J-1)i}' \phi_{(J-2)(J-1)} \left(Z_{(J-2)i} - X_{(J-2)i}\alpha_{(J-2)} \right)]
\end{aligned}$$

5. Let $C_{ji} = (Z_{(j+1)i} - X_{(j+1)i}\alpha_{(j+1)}), ..., (Z_{(J-1)i} - X_{(J-1)i}\alpha_{(J-1)})), \beta_j' = \left(\phi_{j(j+1)}, ..., \phi_{j(J-1)}\right)$ for $j = 2, ..., J-2$. Given the prior distribution for $\beta_j \sim \mathcal{N}(\underline{\beta}_j, \underline{H}_{\beta_j}^{-1})$ the full conditional distribution of β_j is $\beta_j \sim \mathcal{N}(\overline{\beta}_j, \overline{H}_{\beta_j}^{-1})$ where

$$
\begin{aligned}
\overline{H}_{\beta_j} &= \underline{H}_{\beta_j} + \sum_{i=1}^{N} C_{ji}' C_{ji} \\
\overline{\beta}_j &= \overline{H}_{\beta_j}^{-1} \left[\underline{H}_{\beta_j} \underline{\beta}_j + \sum_{i=1}^{N} C_{ji}' \left(Z_{ji} - X_{ji}\alpha_j \right) \right].
\end{aligned}
$$

This concludes the MCMC algorithm.

References

Aitchison, J., Bennett, J.A., 1970. Polychotomous quantal response by maximum indicant. Biometrika 57, 253-262.

Albert, J.H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 88, 669-679.

Bunch, D.S., 1991. Estimability in the multinomial probit Model. Transportation Research B 25, 1-12.

Bunch, D.S., Kitamura, R., 1991. Probit model estimation revisited: trinomial models of household car ownership. Working paper 70, University of California, Berkeley, https://escholarship.org/content/qt2hr8d4bs/qt2hr8d4bs.pdf

Burgette, L.F., Nordheim, E.V., 2012. The trace restriction: An alternative identification strategy for the Bayesian multinomial probit model. Journal of Business & Economic Statistics 30, 404–410.

Geweke, J., Keane, M.P., Runkle, D., 1994. Alternative computational approaches to inference in the multinomial probit model. The Review of Economics and Statistics 76, 609-632.

Heckman, J., Sedlacek, G., 1985. Heterogeneity, aggregation, and market wage functions: an empirical model of self- selection in the labor market. Journal of Political Economy 93, 1077-1125.

Imai, K., Van Dyk, D.A., 2005. A Bayesian analysis of the multinomial probit model using marginal data augmentation. Journal of Econometrics 124, 311–334.

Keane, M.P., 1992. A note on identification in the multinomial probit model. Journal of Business and Economic Statistics 10, 193-200.

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

McCulloch, R.E., Polson, N.G., Rossi, P.E., 2000. A Bayesian analysis of the multinomial probit model with fully identified parameters. Journal of Econometrics 99, 173-193.

Munkin, M.K., Trivedi, P.K., 2008. Bayesian analysis of the ordered probit model with endogenous selection. Journal of Econometrics 143, 334-348.

Nobile, A., 1998. A hybrid Markov chain for the Bayesian analysis of the multinomial probit model. Statistics and Computing 8, 229-242.

Nobile, A., 2000. Comment: Bayesian multinomial probit models with normalization constraint. Journal of Econometrics 99, 335–345.

Table 1. Trinomial Probit ($\sigma_{12} = 0$): ML Estimates with σ_2 Restrictions.

Parameter	True Value	ML	(1)	(2)	(3)	(4)	(5)
α_{11}	-0.8	-0.892	-0.893	-0.892	-0.892	-0.895	-0.897
		0.045	0.046	0.046	0.046	0.045	0.045
α_{12}	0.2	0.212	0.206	0.208	0.213	0.215	0.216
		0.008	0.008	0.008	0.008	0.008	0.008
α_{21}	-2.0	-2.571	-1.763	-2.053	-2.790	-3.539	-4.141
		0.459	0.062	0.069	0.088	0.105	0.118
α_{22}	0.4	0.494	0.354	0.405	0.531	0.657	0.758
		0.078	0.011	0.012	0.016	0.020	0.023
σ_{12}	0	-0.205	-0.102	-0.135	-0.239	-0.368	-0.484
		0.144	0.091	0.104	0.140	0.172	0.198
σ_2	1.5	1.852	1.3	1.5	2	2.5	2.9
		0.305					
$\log \widehat{L}$		-7764.17	-7766.53	-7764.96	-7764.26	-7765.37	-7766.61
$\log \widehat{L}_U - \log \widehat{L}_R$		0	2.36	0.79	0.09	1.2	2.44

(ML estimates are in the first rows, standard errors are in the second rows)

Table 2. The Estimated Restricted Log-likelihoods and "No rejection" Intervals for Various ρ Values.

Parameters	$\rho = 0$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 0.9$
$\sigma_2 = 1$	-7771.93	-7768.05	-7745.57	-7712.11	-7642.87	-7415.713	-7079.98	-6750.80
$\sigma_2 = 1.1$	-7769.57	-7765.44	-7742.68	-7709.03	-7639.52	-7411.03	-7072.93	-6740.48
$\sigma_2 = 1.2$	-7767.81	-7763.49	-7740.53	-7706.77	-7637.10	-7407.71	-7068.05	-6733.12
$\sigma_2 = 1.3$	**-7766.53**	-7762.06	-7738.96	-7705.13	-7635.39	-7405.42	-7064.83	-6728.03
$\sigma_2 = 1.4$	**-7765.60**	**-7761.03**	**-7737.83**	**-7703.97**	**-7634.23**	**-7403.94**	**-7062.90**	**-6724.72**
$\sigma_2 = 1.5$	**-7764.96**	**-7760.30**	**-7737.05**	**-7703.20**	**-7633.50**	**-7403.09**	**-7061.97**	**-6722.87**
$\sigma_2 = 1.6$	**-7764.54**	**-7759.83**	**-7736.55**	**-7702.71**	**-7633.09**	**-7402.71**	**-7061.82**	**-6722.21**
$\sigma_2 = 1.7$	**-7764.30**	**-7759.54**	**-7736.25**	**-7702.45**	**-7632.94**	**-7402.72**	**-7062.29**	**-6722.53**
$\sigma_2 = 1.8$	**-7764.19**	**-7759.40**	**-7736.12**	**-7702.37**	**-7632.99**	**-7403.01**	**-7063.24**	**-6723.62**
$\sigma_2 = 1.9$	**-7764.18**	**-7759.38**	**-7736.12**	**-7702.43**	**-7633.19**	**-7403.53**	-7064.54	-6725.32
$\sigma_2 = 2$	**-7764.26**	**-7759.45**	**-7736.23**	**-7702.59**	**-7633.51**	**-7404.21**	-7066.11	-6727.47
$\sigma_2 = 2.1$	**-7764.41**	**-7759.59**	**-7736.41**	**-7702.84**	**-7633.91**	**-7405.01**	-7067.87	-6729.94
$\sigma_2 = 2.2$	**-7764.60**	**-7759.79**	**-7736.65**	**-7703.15**	**-7634.38**	-7405.91	-7069.77	-6732.64
$\sigma_2 = 2.3$	**-7764.83**	**-7760.03**	**-7736.94**	**-7703.51**	**-7634.90**	-7406.86	-7071.74	-6735.47
$\sigma_2 = 2.4$	**-7765.09**	**-7760.30**	**-7737.26**	**-7703.90**	-7635.46	-7407.86	-7073.77	-6738.38
$\sigma_2 = 2.5$	**-7765.37**	**-7760.60**	**-7737.61**	**-7704.32**	-7636.03	-7408.89	-7075.81	-6741.32
$\sigma_2 = 2.6$	**-7765.67**	**-7760.91**	**-7737.98**	**-7704.76**	-7636.62	-7409.92	-7077.85	-6744.26
$\sigma_2 = 2.7$	**-7765.98**	**-7761.24**	**-7738.35**	-7705.21	-7637.22	-7410.96	-7079.87	-6747.16
$\sigma_2 = 2.8$	**-7766.29**	**-7761.57**	**-7738.74**	-7705.66	-7637.83	-7411.99	-7081.86	-6750.01
$\sigma_2 = 2.9$	**-7766.61**	**-7761.91**	-7739.13	-7706.12	-7638.43	-7413.01	-7083.81	-6752.79
$\sigma_2 = 3$	-7766.93	-7762.25	-7739.53	-7706.58	-7639.02	-7414.02	-7085.71	-6755.50
LR Intervals	[1.3,2.9]	[1.4,2.9]	[1.4,2.8]	[1.4,2.6]	[1.4,2.3]	[1.4,2.1]	[1.4,1.8]	[1.4,1.8]
ML Intervals	[1.4,2.3]	[1.3,2.4]	[1.4,2.3]	[1.4,2.2]	[1.3,2.1]	[1.4,1.9]	[1.4,1.8]	[1.4,1.8]
$\log \widehat{L}_{ML}$	-7764.17	-7759.37	-7736.11	-7702.37	-7632.94	-7402.68	-7061.79	-6722.20
$\widehat{\rho}_{ML}$	1.85	1.87	1.85	1.80	1.72	1.65	1.57	1.61
$(std.err)$	0.30	0.32	0.30	0.26	0.24	0.17	0.12	0.11
$\log \widehat{L}_{ML} - \chi^2_{0.1}(1)$	-7766.87	-7762.07	-7738.81	-7705.07	-7635.64	-7405.38	-7064.49	-6724.90

Table 3. Trinomial Probit: MPR and M Estimates when $\sigma_{12} = 0$.

Parameter	True Value	M			MPR		
			Lag 20	RNE		Lag 20	RNE
α_{11}	-0.8	-0.900	0.31	0.025	-0.894	0.27	0.026
		0.048			0.046		
α_{12}	0.2	0.201	0.06	0.131	0.211	0.10	0.016
		0.008			0.009		
α_{21}	-2.0	-1.332	0.25	0.030	-2.602	0.96	0.0007
		0.049			0.545		
α_{22}	0.4	0.277	0.02	0.192	0.498	0.95	0.0007
		0.009			0.093		
σ_{12}	0	-0.060	0.64	0.012	-0.227	0.77	0.0019
		0.068			0.164		
σ_2	1.5				3.627	0.97	0.0008
					1.470		

(Posterior means are in the first rows, posterior standard deviations are in the second rows)

Table 4. Trinomial Probit: RNE and Lag 20 Values of M Markov Chains with Respect to ρ.

Parameter	$\rho = 0.1$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$		$\rho = 0.9$	
	Lag 20	RNE	Lag 20	RNE	Lag 20	RNE	Lag 20	RNE	Lag 20	RNE
α_{11}	0.25	0.029	0.20	0.042	<0.01	0.229	0.12	0.057	0.13	0.052
α_{21}	0.27	0.028	0.29	0.028	0.51	0.012	0.45	0.015	0.35	0.022
σ_{12}	0.63	0.012	0.61	0.013	0.69	0.009	0.66	0.010	0.59	0.013

Table 5. Trinomial Probit: RNE and Lag 20 Values of MPR Markov Chains with Respect to ρ.

Parameter	$\rho = 0.1$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$		$\rho = 0.9$	
	Lag 20	RNE	Lag 20	RNE	Lag 20	RNE	Lag 20	RNE	Lag 20	RNE
α_{11}	0.28	0.027	0.31	0.024	0.29	0.027	0.40	0.018	0.52	0.012
α_{21}	0.95	0.001	0.95	0.008	0.93	0.001	0.90	0.002	0.92	0.001
σ_{12}	0.69	0.001	0.64	0.010	0.69	0.003	0.85	0.002	0.93	0.001
σ_2	0.96	0.001	0.97	0.0008	0.94	0.001	0.91	0.002	0.94	0.001

Table 6. Multinomial Probit: M Estimates when $J = 4$ and $J = 5$.

Parameter	True Value	$J = 4$			$J = 5$		
			Lag 20	RNE		Lag 20	RNE
α_{11}	-0.5	-0.582	0.38	0.021	-0.570	0.50	0.013
		0.038			0.044		
α_{12}	1	0.984	0.47	0.017	0.931	0.53	0.013
		0.023			0.024		
α_{21}	-0.5	-0.486	0.43	0.014	-0.539	0.62	0.009
		0.038			0.051		
α_{22}	1	0.929	0.47	0.014	0.938	0.62	0.009
		0.023			0.026		
α_{31}	-0.5	-0.356	0.32	0.025	-0.451	0.54	0.011
		0.034			0.045		
α_{32}	1	0.845	0.37	0.025	0.887	0.52	0.013
		0.020			0.023		
α_{41}	-0.5				-0.423	0.48	0.014
					0.042		
α_{42}	1				0.863	0.46	0.016
					0.022		
σ_{12}	0	-0.021	0.56	0.013	0.112	0.64	0.011
		0.062			0.073		
σ_{13}	0	0.015	0.50	0.017	0.034	0.62	0.012
		0.058			0.069		
σ_{14}	0				-0.014	0.61	0.012
					0.070		
σ_{23}	0	-0.046	0.56	0.012	-0.068	0.68	0.009
		0.061			0.077		
σ_{24}	0				-0.117	0.66	0.009
					0.074		
σ_{34}	0				-0.070	0.63	0.010
					0.070		

(Posterior means are in the first rows, posterior standard deviations are in the second rows)

Figure 1. The Shape of the Log-likelihood Function when $\rho = 0$.

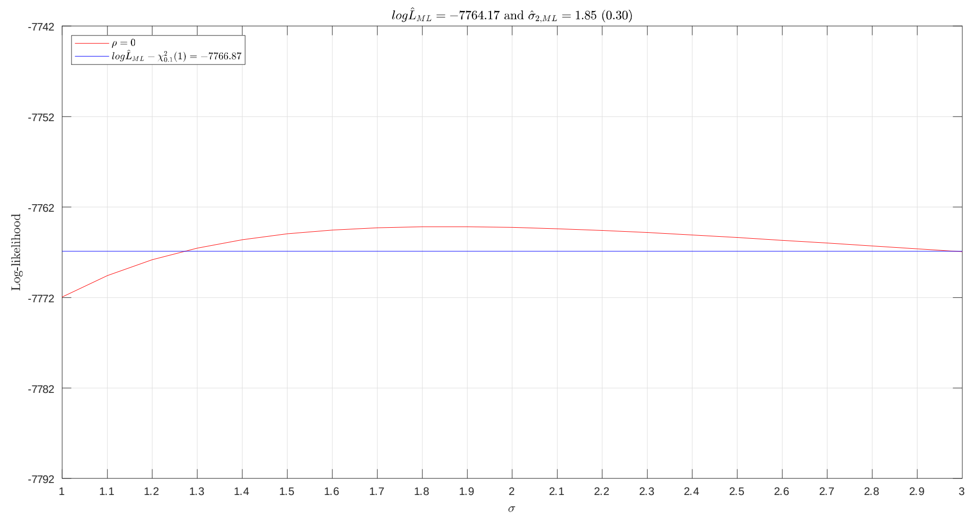

Figure 2. The Shape of the Log-likelihood Function when $\rho = 0.6$.

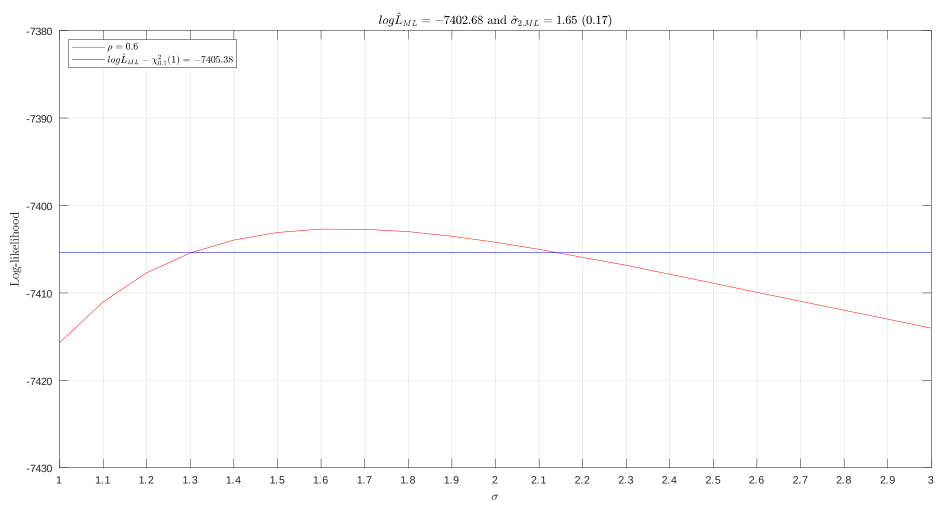

Figure 3. The Shape of the Log-likelihood Function when $\rho = 0.9$.

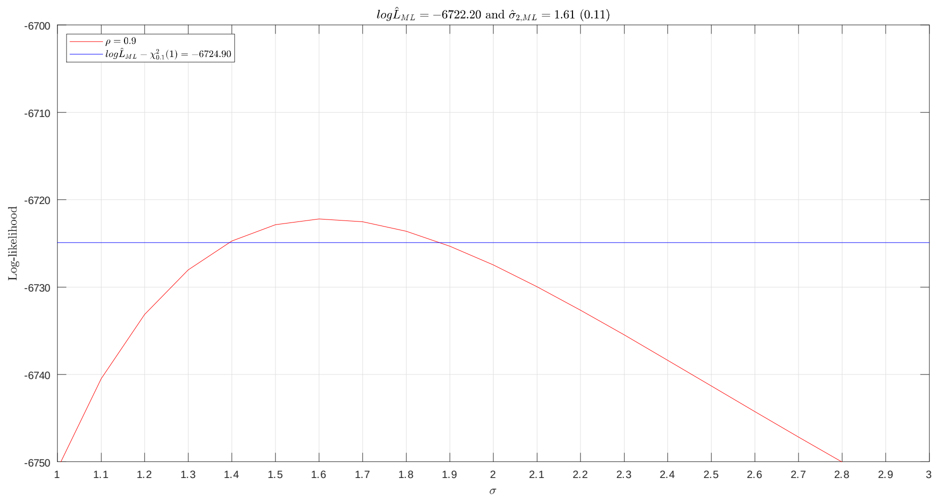

Figure 4. Time Series and Autocorrelation Function Plots: M Estimates.

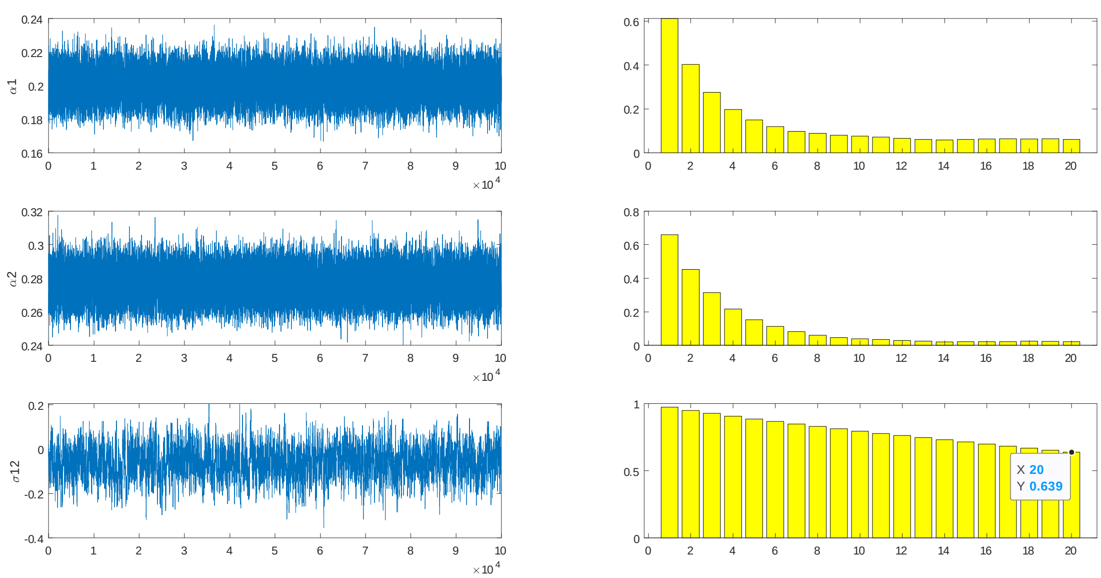

Figure 5. Time Series and Autocorrelation Function Plots: MPR Estimates.