Approximation and Simulation of the Multinomial Probit Model: An Analysis of Covariance Matrix Estimation

Paul A. Ruud*
Department of Economics, University of California, Berkeley

April, 1996

1 Introduction

The multinomial probit (MNP) model is a primary application for combining simulation with estimation. Indeed, McFadden (1989) featured the MNP model in his seminal paper. As random utility model, the MNP model offers a highly desirable flexibility in substitution among alternatives that its chief rival, the multinomial logit model, fails to possess. The unrestricted character of the variance matrix in the multivariate normal distribution that underlies probit cannot be produced by logit, even in its generalized extreme value forms.

As experience with the MNP model has developed recently, researchers have developed an appreciation for the practical difficulties that estimation with simulation presents. McFadden & Ruud (1994) give a general description, analyzing some of the generic problems with the methods of simulated moments and maximum simulated likelihood. In this paper, we draw on their analysis to develop a new estimation strategy for the MNP model based on the method of simulated moments.

*I am indebted to Kenneth Train for discussions about many topics related to the material in this paper. He helped me particularly with the history of the random parameters logit model. He has not even seen this paper yet, so he cannot be held responsible for its contents.

In addition, we focus on a problem particular to the MNP model, the estimation of covariance parameters. Keane (1992) for example finds in his applications that the covariance parameter estimators are difficult to compute and imprecisely estimated when the attributes of the alternatives are similar.

2 Multinomial Probit

Most of the research into the estimation of the MNP model using simulation has focussed on developing new simulators for the multivariate normal c.d.f. See, for example, Hajivassiliou, McFadden & Ruud (forthcoming). This research has provided immediate and significant practical improvements in estimation. These simulators are used in two principal ways. First of all, following McFadden (1989), the method of simulated moments builds on a set of moment equations constructed with residuals that are differences in observed choice frequencies and their simulated probabilities. Alternatively, one computes a maximum simulated likelihood estimator by placing the simulated probabilities into the log-likelihood function as though the simulators contained no simulation error.

Both methods have drawbacks, regardless which probabilities simulators are used. The method of simulated moments estimator is often difficult to compute. In our experience, search algorithms frequently wander toward the boundary of the parameter space. This seems to occur because the instrumental variables introduce undesirable small sample idiosyncracies to the moment equations. The instrumental variables are usually chosen so that the moment equations approximate the normal equations of maximum likelihood. A popular approximation method substitutes, like maximum simulated likelihood, simulators for intractable terms and the simulation noise surely causes some of the computational difficulties.

The method of maximum simulated likelihood appears to be much easier to apply. Again, our own experience suggests that the optimization of the simulated likelihood is usually straightforward. But the method produces estimators that are generally inconsistent, unless the number of simulations are increased with the sample size. New methods for correcting such estimators for their asymptotic bias remains an interesting research question.

In this paper, we return to the method of simulated moments and focus on the specification of the moment equations. We offer a new method for deriving instrumental variables based upon approximating the log likelihood function without simulation. Instead, we derive analytical instrumental variables based on an analysis of extreme values. The tractable character of these instrumental variables makes the estimation of the covariance parameters of the MNP model more straightforward.

2.1 Notation

We will use the following notation for the MNP model. Let $\mathfrak{N}(\mu, \Omega)$ denote the multivariate normal distribution with expectation vector μ and variance-covariance matrix Ω. Each choice among alternatives indexed $j = 1, \ldots, J$ is generated by the latent random utility model $y^* \sim \mathfrak{N}(X^*\beta, \Omega^*)$ where y^* is a vector of J random utility indexes, X^* is a $J \times K$ matrix of attribute K variables, x_{jk}, for each of the J choices, and $\theta = (\beta, \Omega^*)$ contains the unknown population parameters. The vector β contains K elements and $\Omega^* = [\omega_{ij}]$ is a $J \times J$ positive, semi-definite matrix. The specification is completed by the observation rule

$$y = \left[\mathbf{1}\left\{y_j^* = \max_{m=1,\ldots,J} y_m^*\right\};\ j = 1, \ldots, J\right]. \tag{1}$$

Each element of y is an indicator for whether the alternative with the corresponding index was chosen according to whether its latent utility index was the greatest.

Because only differences in the elements of y^* are necessary to determine y, we will normalize as follows. We reduce y^* to the $J-1$ dimensional vector

$$y^{\Delta} = \left[y_j^* - y_1^*;\ j = 2, \ldots, J\right]$$

and X correspondingly to the $(J-1) \times K$ matrix

$$X = [x_{jk} - x_{1k};\ j = 2, \ldots, J,\ k = 1, \ldots, K].$$

Without loss of generality, we set the first row and column of Ω^* to zeros and refer to the remaining submatrix as Ω:

$$\Omega = [\omega_{ij};\ i, j = 2, \ldots, J].$$

We also note that Ω or β must be normalized in some way because the scale of the distribution of y^{Δ} is not identifiable. For example, if Ω is constant across observations, one can constrain $\omega_{22} = 1$.

We denote the $\mathfrak{N}(\mu, \Omega)$ p.d.f. by

$$\phi(z-\mu,\,\Omega) = \frac{1}{\sqrt{\det\left(2\pi\Omega\right)}} \exp\left(-\frac{1}{2}\left(x-\mu\right)\Omega^{-1}\left(x-\mu\right)\right) \tag{2}$$

and the corresponding c.d.f. by

$$\Phi(Z-\mu,\,\Omega) = \int_{-\infty}^{Z} \phi(z-\mu,\,\Omega)\,dz. \tag{3}$$

Using this notation, the log-likelihood function of the MNP model is

$$L(\theta; y, X) = \sum_{j=1}^{J} y_j \log \Phi(\Delta_j X\beta,\, \Delta_j \Omega \Delta_j')$$

where Δ_j is a differencing matrix, except for $\Delta_1 = I_{J-1}$ which is an identity matrix. The Δ_j matrix is constructed by replacing the elements of j^{th} column of I_{J-1} with -1. The score can be written in the general form

$$\frac{\partial L(\theta; y, X)}{\partial\theta} = \sum_{j=1}^{J} \frac{\partial \log \Phi(\Delta_j X\beta,\, \Delta_j \Omega \Delta_j)}{\partial\theta} \left[y_j - \Phi(\Delta_j X\beta,\, \Delta_j \Omega \Delta_j')\right].$$

In MSM, it is the simulation of the derivatives of the log probability term that appears to cause small sample difficulties. See McFadden & Ruud (1994) for further discussion and examples.

2.2 A Special MSM Estimator

Our basic goal is to construct simpler and more tractable MSM estimators for the MNP model by focussing on the instrumental variables, rather than the probability simulators. We made an initial step toward this goal in Ruud (1991), where we noted that one of the simplest simulated moment equations for multinomial choice models can be integrated back to a globally concave objective function. Given N observations (indexed $n = 1, \ldots, N$), the moment equations for estimation of β are

$$\sum_{n=1}^{N} \sum_{j=1}^{J} x_{nj} \left(y_{nj} - \tilde{y}_{nj}(\hat{\theta})\right) = 0 \tag{4}$$

where

$$\tilde{y}_{nj}(\theta) = \mathbf{1}\left\{x'_{nj}\beta + \varepsilon_{nj} = \max_k \left(x'_{nk}\beta + \varepsilon_{nk}\right)\right\} \tag{5}$$

is the crude frequency simulator of y_{nj} based on simulations ε_n drawn from the $\mathfrak{N}(0, \Omega^*)$ distribution. These moment equations have the appealing form of standard orthogonality conditions in linear models. They are also extremely easy to compute and suffer no simulation noise. But their most interesting feature is that these moment equations are the gradient of the objective function

$$\sum_{n=1}^{N}\left[\sum_{j=1}^{J} y_{nj}\cdot x'_{nj}\beta - \tilde{y}_{nj}(\theta)\left(x'_{nj}\beta + \varepsilon_{nj}\right)\right] = \sum_{n=1}^{N}\left[\sum_{j=1}^{J} y_{nj}\cdot x'_{nj}\beta\right] - \max_k \left(x'_{nk}\beta + \varepsilon_{nk}\right). \tag{6}$$

This function of β is piece-wise linear and concave: it can be maximized quickly with standard linear programming techniques. The calculations are quite similar to those of the least absolute deviations estimator of the linear regression model.

One can actually specify any distribution for the disturbance terms, not only the multivariate normal distribution. The MSM estimator is fully characterized and easy to compute. Unfortunately, we have never been able to include covariance parameters in an equally attractive manner and so the estimator has remained a theoretical curiosity.

3 The Random-Parameters Logit Model

There is a reason for this failure that relates this special result to the random-parameters logit (RPL) model, an alternative to MNP. This reason motivates the approach that we describe in the next section.

The RPL model is a mixture of logit and probit specifications:

$$\Pr\left\{y_j = 1 \mid X\right\} = \mathrm{E}\left[\frac{\exp\left(y_j^*\right)}{\sum_{m=1}^{J}\exp\left(y_m^*\right)}\right].$$

This model, and its simulation, was described by Boyd & Mellman (1980) and Cardell & Dunbar (1980). A variant, with the log-normal distribution, is mentioned in Ben-Akiva & Lerman (1985). Bolduc & Ben-Akiva (1989) actually describe MSM estimation for this model. Bolduc, Fortin & Fournier (1993), McFadden & Train (1996) and Revelt & Train (1996) are examples of actual application.

Because the RPL model is a mixture of normal and logistic distributions, estimates of this model will not be consistent for the parameters of the MNP specification. Nevertheless, this is an attractive specification for (at least) two reasons: First of all, these probabilities are easy to simulate and they always sum over the alternatives to one and secondly, they can probably approximate the MNP probabilities well.

Indeed the MNP model is an extreme special case. If we increase the scales of β and Ω appropriately, the RPL and MNP probabilities are identical. Let σ be a positive scale parameter. Note that

$$\lim_{\sigma \to \infty} \frac{\exp\left(\sigma\, y_j^*\right)}{\sum_{m=1}^{J} \exp\left(\sigma\, y_m^*\right)} = \mathbf{1}\left\{y_j^* = \max_m y_m^*\right\}.$$

This limit is, of course, the component of (1) and analogous to the crude frequency simulator in (5) and

$$\mathrm{E}\left[\mathbf{1}\left\{y_j^* = \max_m y_m^*\right\}\right] = \Phi(\Delta_j X\beta,\, \Delta_j \Omega \Delta_j'),$$

exhibiting the MNP case.

This feature also shows that the RPL model effectively has an inherent scaling problem. The RPL specification requires an allocation of random variation between the normal and logistic components that seems difficult to discern with discrete choice data.[1] Amemiya (1981) discusses the close similarity of binomial probit and logit probability functions. Stern (1989) makes a similar observation for multinomial models. Given the difficulty of telling these models apart, their convolution is unlikely to yield much information about their individual contributions.

This leads us to the connection between the simple MSM estimator for choice models and the RPL model. We take the normalized limit of the simulated RPL log likelihood function when each observation is simulated once, as the scale grows. We find that the elements of the objective function

[1]Only the differences of i.i.d. extreme value random variables, which have logistic distributions, enter into the random utility specification.

approach

$$\lim_{\sigma\to\infty} \frac{1}{\sigma} \sum_{j=1}^{J} y_{nj} \log\left(\frac{\exp\left(\sigma\, y_j^*\right)}{\sum_{m=1}^{J} \exp\left(\sigma\, y_m^*\right)} \right) = \left[\sum_{j=1}^{J} y_{nj} \cdot \left(x'_{nj}\beta + \varepsilon_{nj} \right) \right] - \max_k \left(x'_{nk}\beta + \varepsilon_{nk} \right).$$

This is an element of the objective function (6) of the simulated moments method described in the previous section above, except that an additional term

$$\sum_{n=1}^{N} \sum_{j=1}^{J} y_{nj} \cdot \varepsilon_{nj}$$

that is irrelevant to β appears. This limit is useful for analyzing the simulated RPL model, because it enables us to examine the influence function of the most influential observations on the maximum simulated likelihood estimator, that is the observations with the smallest fitted probabilities.

This connection explains two characteristics of the MSM estimator. First of all, we see the concavity of the MSM objective function as consistent with the well-known concavity of the multinomial logit log-likelihood function. Secondly, note that the normal equations for parameters in the covariance matrix Ω are (like 4)

$$\sum_{n=1}^{N} \sum_{j=1}^{J} \frac{\partial \varepsilon_{nj}}{\partial \theta} \left(y_{nj} - \tilde{y}_{nj}(\hat{\theta}) \right) = 0$$

which appear nonsensical. The key terms in this normal equation involve the data y_{nj} and they all have marginal expectation equal to zero. Because the simulations that enter into the ε_{nj} are i.i.d. standard normal random variables drawn independently of the y_{nj}, the expectation of the $\frac{\partial \varepsilon_{nj}}{\partial \theta} y_{nj}$ is zero. Therefore, the moment equations provide no information at all about the covariance parameters these equations serve to estimate. Similar phenomena occur for replicated simulations, except that the problem is restricted to a local one.

We conclude that there is a region of the parameter space of the simulated random parameters logit model where the likelihood is quite flat with respect to all the covariance parameters.

4 Extreme Normal Approximation

The weakness of the logit specification leads us to investigate whether another simple MSM estimator could be derived from the limiting behavior of the MNP log-likelihood function. That is, replace the multinomial logit kernel with a multinomial probit kernel. Such specifications have been suggested before by Bolduc & Ben-Akiva (1989) and Train (1995), but only for the restrictive case where the kernel is the product of *univariate* c.d.f.s. Our goal is to produce tractable instrumental variables that approximate the efficient instrumental variables possessed by the maximum likelihood normal equations. Therefore, we require the kernel c.d.f. to exhibit the same covariance parameters as the data generating process itself.

Our basic result gives the limiting behavior of the MNP log probability function as a quadratic spline function.

Theorem 1 *Let $\Phi(\cdot)$ be defined by (2) and (3). Then*

$$\begin{aligned}\lim_{\sigma\to\infty}\frac{1}{\sigma^2}\log\Phi(\sigma\cdot x,\,\Omega) &= \lim_{\sigma\to 0^+}\sigma\,\log\Phi(x,\sigma\,\cdot\Omega)\\ &= \max_{z\leq x}-\frac{1}{2}z'\Omega^{-1}z.\end{aligned}$$

Proof. See appendix.

For expositional convenience, let us denote this limiting function by

$$Q(x,\Omega)=\max_{z\leq x}-\frac{1}{2}z'\Omega^{-1}z$$

and let us describe some of its important features. First of all, this function is nondecreasing and it reaches a maximum of zero over the entire positive orthant. Secondly, $Q(x,\Omega)$ is a quadratic spline. For example, over the region where

$$\frac{\partial}{\partial x}\left(-\frac{1}{2}x'\Omega^{-1}x\right)=-\Omega^{-1}x>0,$$

the function is given by

$$Q(x,\Omega)=-\frac{1}{2}x'\Omega^{-1}x.$$

At a boundary of this region, an element of $\Omega^{-1}x$ becomes zero. Let us say it is the last element and partition

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}.$$

Then,

$$x_2 = \Omega_{21}\Omega_{11}^{-1}x_1$$

at the boundary and on the other side of this boundary

$$Q\left(x, \Omega\right) = -\frac{1}{2}x_1'\Omega_{11}^{-1}x_1,$$

so that $Q(x, \Omega)$ becomes flat in the element x_2. One can see that such boundaries continue to reduce the number of terms in the quadratic until, in the positive orthant, the function is constant and equal to zero. The loss of terms preserves the nondecreasing (in x) character of $Q(x, \Omega)$.

Finally, note that $Q(x, \Omega)$ possesses a continuous gradient. The boundaries just described are locations of knots for the continuous linear spline that comprises the gradient function.

We propose an MSM estimator that uses the derivatives of $Q(x, \Omega)$ as the instrumental variables for simulated residuals, in analogy with the moment equations in (4),

$$\hat{\theta} : \sum_{n=1}^{N}\sum_{j=1}^{J} \left. \frac{\partial Q(\Delta_j X_n \beta, \Delta_j \Omega \Delta_j')}{\partial \theta} \right|_{\theta=\hat{\theta}} \left(y_{nj} - \tilde{y}_{nj}(\hat{\theta}) \right) = 0,$$

where

$$\frac{\partial Q(\Delta_j X \beta, \Delta_j \Omega \Delta_j')}{\partial \theta} = \begin{pmatrix} -Z_j' \left(\Gamma_j\right)^{-1} Z_j \beta \\ -\Delta_j' S_j' \left(\Gamma_j\right)^{-1} Z_j \beta \otimes \Delta_j' S_j' \left(\Gamma_j\right)^{-1} Z_j \beta \end{pmatrix},$$

$$\begin{aligned} Z_j &= S_j \Delta_j X, \\ \Gamma_j &= S_j \Delta_j \Omega \Delta_j' S_j', \end{aligned}$$

and S_j is a selection matrix for the active terms in the quadratic spline $Q(\Delta_j X \beta, \Delta_j \Omega \Delta_j')$. One can use any simulator for the $\tilde{y}_j(\theta)$ replacing the

MNP probabilities, although a smooth one will provide a computationally easier estimator. Thus, we construct a set of moment equations with simple, tractable instrumental variables that contain no simulation noise. Furthermore, these instrumental variables provide local identification of *all* the parameters of the MNP model, if there is sufficient variation in the attributes of the alternatives in X across observations.

This caveat, that the variation in X_n may fail to identify the parameters, is apparently generic to MNP estimation. Obviously, the X_n must not be linearly dependent across observations. In addition, there must be adequate variation across alternatives to identify the parameters of the variance matrix Ω. This has been noted by Keane (1992). We view our specification for the instrumental variables as having the analytical advantage that one can easily confirm failures of local identification. Nevertheless, it is also the case that our instrumental variables will fail to identify the parameters in situations where they are formally identifiable. Our preliminary computation experience suggests that this will not be an issue in practice.

It is helpful to look at the simplest case of a binomial probit model to understand the nature of our proposed MSM estimator. In that case, the classical (without simulation) method-of-moments estimator that simulation approximates is the solution to the moment equations

$$\sum_{n=1}^{N} x_n \; |x_n' \beta| \; [y_n - \Phi(x_n' \beta)] \, ,$$

where $\Phi(x_n'\beta)$ is the univariate standard normal c.d.f.. The $|x_n'\beta|$ term replaces the efficient term

$$\frac{\phi(x_n'\beta)}{\Phi(x_n'\beta)\Phi(-x_n'\beta)}.$$

The difference between these two influence functions is plotted in Figure 4. One can see that these functions are essentially equivalent at extreme values and that the overall convex character of efficient weighting function is captured by the extreme normal approximation.

The efficiency loss of the approximation depends on the distribution of $X_n\beta$ in the sample. If the estimated slopes are near zero or the attributes exhibit little variation, then the loss will be greatest. One should compare this efficiency loss with that introduced by simulating the efficient weights with error. In many cases, the latter will be more damaging to efficiency.

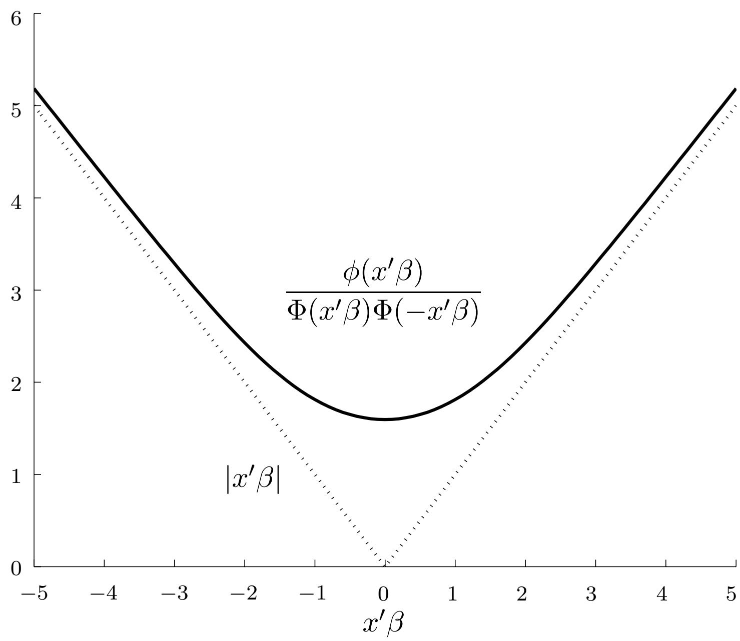

Figure 1: Comparison of Influence Functions

5 Computational Strategy

We plan to use a combination of indirect inference (Gourieroux, Monfort & Renault (1993), Gallant & Tauchen (1992)) and the method of simulated moments. We can construct a probability choice model from the extreme normal approximation. If we think of the limiting log likelihood function as the log likelihood of a probability choice model, then $\exp\left[Q(\Delta_j X\beta,\, \Delta_j \Omega \Delta_j')\right]$ would be interpreted as $\Pr\{y_j = 1 \mid X\}$. But these "probabilities" do not sum to one. This is easily corrected using the logit form

$$\begin{aligned}\Pr\{y_j = 1 \mid X\} &= \frac{\exp\left[Q_j(\theta)\right]}{\sum_{m=1}^{J} \exp\left[Q_m(\theta)\right]} \\ &\equiv p_j(\theta),\end{aligned}$$

where

$$Q_j(\theta) \equiv Q(\Delta_j X\beta,\, \Delta_j \Omega \Delta_j').$$

Such a specification has the form of a so-called mother logit model, in which the characeristics of other alternatives enter into the index $Q(\Delta_j X\beta,\, \Delta_j \Omega \Delta_j')$ for the j^{th} alternative.

As a first step in estimation, we use indirect inference. This step has two steps itself. First of all, we compute the quasi-maximum likelihood estimator for θ:

$$\tilde{\theta} = \arg\max_{\theta} \sum_{n=1}^{N} \left(\sum_{j=1}^{J} y_{nj}\, Q_{nj}(\theta) \right) - \log \left(\sum_{m=1}^{J} \exp\left[Q_{nm}(\theta)\right] \right).$$

We have found this extremely easy to do. This yields the relationship

$$\sum_{n=1}^{N} \sum_{j=1}^{J} \left. \frac{\partial Q_{nj}(\theta)}{\partial \theta} \right|_{\theta=\tilde{\theta}} \left(y_{nj} - p_{nj}(\tilde{\theta}) \right) = 0.$$

In the second step of the method of indirect inference, we obtain a consistent estimator $\hat{\theta}$ by minimizing the length of

$$\sum_{n=1}^{N} \sum_{j=1}^{J} \left. \frac{\partial Q_{nj}(\theta)}{\partial \theta} \right|_{\theta=\tilde{\theta}} \left(\tilde{y}_{nj}(\theta) - p_{nj}(\tilde{\theta}) \right)$$

over θ, where θ appears only in the unbiased simulations for y_{nj}, the $\tilde{y}_{nj}(\theta)$ that replace the y_{nj} in the moment equations. We expect this to be straightforward; indeed, we anticipate that most solutions will yield the zero vector. In that case, we will have found a $\hat{\theta}$ which solves a form of MSM estimation:

$$\sum_{n=1}^{N}\sum_{j=1}^{J} \left.\frac{\partial Q_{nj}(\theta)}{\partial \theta}\right|_{\theta=\tilde{\theta}} \left(y_{nj} - \tilde{y}_{nj}(\hat{\theta})\right) = 0.$$

We will also compute a second-round estimator on the basis of $\hat{\theta}$. In the second round, we improve the instrumental variables by replacing $\tilde{\theta}$ in the instrumental variables with the value of θ that minimizes a measure of distance between the matrices

$$\sum_{n=1}^{N}\sum_{j=1}^{J} \frac{\partial Q_{nj}(\theta)}{\partial \theta}\frac{\partial Q_{nj}(\theta)}{\partial \theta'}\left(1 - \tilde{y}_{nj}(\hat{\theta})\right)\tilde{y}_{nj}(\hat{\theta})$$

and

$$\sum_{n=1}^{N}\sum_{j=1}^{J} \frac{\partial Q_{nj}(\theta)}{\partial \theta}\frac{\partial \tilde{y}_{nj}(\hat{\theta})}{\partial \theta'}.$$

If these two matrices have the same probability limit, then we obtain the optimal generalized MSM estimator, following the classical generalized method of moments (GMM) theory developed by Hansen (1982). This process will lower the asymptotic variance of certain functions of the estimator based on the resultant instrumental variables.

6 Conclusion

We will report on our computational experience in the future. Our limited experience has confirmed that the instruments do identify the variance matrix parameters and that computation of the MSM estimator is quite straightforward. In addition, we are able to accomodate to singularities in the variance matrix easily. The maximum likelihood estimator of the MNP model can easily be singular, so that this is also an important feature for the approximation to allow.

In future research, we plan to investigate whether the extreme normal approximation leads to a new, tractable, random utility model (RUM) with a flexible covariance structure.[2] It is an open question whether such probabilities, or similar ones, can be generated by a RUM.

7 Appendix

In this appendix we prove Theorem 1.

If $x \geq 0$, then $\lim_{\sigma\to 0^+} \Phi(x,\sigma\Omega) > 0$ and $\lim_{\sigma\to 0^+} \sigma \log \Phi(x,\sigma\Omega) = 0$.

Suppose that $x \ngeq 0$ so that $\lim_{\sigma\to 0^+} \Phi(x,\sigma\Omega) = 0$. Consider the multivariate normal p.d.f.

$$\phi(x,\Omega) = \frac{1}{\sqrt{\det(2\pi\Omega)}} \exp\left(-\frac{1}{2}x'\Omega^{-1}x\right).$$

Then

$$\frac{\partial \log\phi(x,\Omega)}{\partial \operatorname{vec}\Omega} = -\frac{1}{2}\operatorname{vec}\left(\Omega^{-1} - \Omega^{-1}xx'\Omega^{-1}\right)$$

and

$$\frac{\partial \operatorname{vec}(\sigma\Omega)}{\partial\sigma} = \operatorname{vec}(\Omega)$$

so that

$$\begin{aligned}
\frac{\partial\,\phi(x,\sigma\Omega)}{\partial\sigma} &= -\frac{1}{2}\phi(x,\sigma\Omega)\operatorname{vec}(\Omega)'\operatorname{vec}\left(\sigma^{-1}\Omega^{-1} - \sigma^{-2}\Omega^{-1}xx'\Omega^{-1}\right) \\
&= -\frac{1}{2\sigma^2}\phi(x,\sigma\Omega)\ \operatorname{tr}\left(\sigma I - xx'\Omega^{-1}\right) \\
&= -\frac{1}{2\sigma^2}\phi(x,\sigma\Omega)\ \left(\sigma J - x'\Omega^{-1}x\right)
\end{aligned}$$

and therefore,

$$\begin{aligned}
\lim_{\sigma\to 0^+} \sigma\log\Phi(x,\sigma\Omega) &= \lim_{\sigma\to 0^+} \frac{1}{-\sigma^{-2}} \frac{-\frac{1}{2\sigma^2}\int \phi(x,\sigma\Omega)\ (\sigma N - x'\Omega^{-1}x)\,dx}{\Phi(x,\sigma\Omega)} \\
&= \lim_{\sigma\to 0^+} -\frac{1}{2}\frac{\int x'\Omega^{-1}x\,\phi(x,\sigma\Omega)\,dx}{\Phi(x,\sigma\Omega)} \\
&= \lim_{\sigma\to 0^+} \mathrm{E}\left[-\frac{1}{2}z'\Omega^{-1}z \mid z \leq x\right].
\end{aligned}$$

[2]See McFadden (1981).

Consider the distribution of the truncated multivariate normal random variable z generated by taking random draws from the $N(0,\Omega)$ distribution and retaining only those less than or equal x. The p.d.f. of this random variable,

$$f(z;x,\Omega) = \left\{ \begin{array}{ccc} \frac{\phi(z,\Omega)}{\Phi(x,\Omega)} & if & z \leq x \\ 0 & if & z \nleq x \end{array} \right\}$$

is proportional to the normal p.d.f.. Let x^* denote the location of the unique maximum of this truncated p.d.f.. We can easily see that

$$\lim_{\sigma \to 0^+} f(z;x,\sigma\Omega) = 0, \quad z \neq x^*.$$

Proof. By definition,

$$\begin{aligned} x^* &= \arg\max_{z \leq x} \phi(z,\Omega) \\ &= \arg\max_{z \leq x} -z'\Omega^{-1}z \\ &= \arg\min_{z \leq x} z'\Omega^{-1}z. \end{aligned}$$

It is unique, because the quadratic form is globally concave and the feasible set is convex. Thus, for every $z \neq x^*$, there is a nonempty, closed region $\mathbf{A}(z)$ containing z and x^* given by

$$\mathbf{A}(z) = \left\{ w \mid w \leq x,\ w'\Omega^{-1}w \leq z'\Omega^{-1}z \right\}.$$

Therefore,

$$\begin{aligned} f(z;x,\sigma\Omega) &= \frac{\phi(z,\sigma\Omega)}{\Phi(x,\sigma\Omega)} \\ &< \left[\int_{\mathbf{A}(z)} \exp\left(\frac{1}{2\sigma} \left(z'\Omega^{-1}z - w'\Omega^{-1}w \right) \right) dw \right]^{-1} \\ &\underset{\sigma \to 0^+}{\longrightarrow} 0 \end{aligned}$$

because the integral is unbounded. □

On the other hand, $f(x^*;x,\sigma\Omega)$ is unbounded in σ:

$$f(z;x,\sigma\Omega) = \left[\int \exp\left(\frac{1}{2\sigma} \left(x^{*\prime}\Omega^{-1}x^* - w'\Omega^{-1}w \right) \right) dw \right]^{-1}$$

can be made arbitrarily large because $x^{*\prime}\Omega^{-1}x^* - w'\Omega^{-1}w < 0$ for all $w \neq x^*$.

Therefore,

$$\begin{aligned}\lim_{\sigma\to 0^+} \sigma \log \Phi(x, \sigma\Omega) &= \lim_{\sigma\to 0^+} \mathrm{E}\left[-\frac{1}{2}z'\Omega^{-1}z \mid z \leq x\right]. \\ &= \max_{z\leq x} -\frac{1}{2}z'\Omega^{-1}z.\end{aligned}$$

References

Amemiya, T. (1981), 'Qualitative response models: A survey', *Journal of Economics Literature* **19**(4), 1483–1536.

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

Bolduc, D. & Ben-Akiva, M. (1989), Multinomial probit with taste variation and autoregressive alternatives, mimeograph.

Bolduc, D., Fortin, B. & Fournier, M.-A. (1993), The impact of incentive policies on the practice location of doctors: A multinomial probit analysis, working paper 9308, Université Laval, Groupe de Recherche en Politique Economique, Département d'économique.

Boyd, J. H. & Mellman, R. E. (1980), 'The effect of fuel economy standards on the u. s. automotive market: An hedonic demand analysis', *Transportation Research–A* **14A**, 367–378.

Cardell, S. & Dunbar, F. (1980), 'Not available at this time', *Transportation Research–A* **14A**.

Gallant, R. & Tauchen, G. (1992), Which moments to match?, mimeograph.

Gourieroux, C., Monfort, A. & Renault, E. (1993), 'Indirect inference', *Journal of Applied Econometrics* **8**, S85–S118. Supplement.

Hajivassiliou, V. A., McFadden, D. L. & Ruud, P. A. (forthcoming), 'Simulation of multivariate normal orthant probabilities', *Journal of Econometrics* .

Hansen, L. P. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica* **50**(4), 1029–1054.

Keane, M. P. (1992), 'A note on identification in the multinomial probit model', *Journal of Business and Economic Statistics* **10**(2), 193–200.

McFadden, D. L. (1981), Structural discrete probability models derived from theories of choice, *in* C. F. Manksi & D. L. McFadden, eds, 'Structural Analysis of Discrete Data with Econometric Applications', MIT, chapter 5, pp. 198–272.

McFadden, D. L. (1989), 'A method of simulated moments for estimation of discrete response models without numerical integration', *Econometrica* **57**(5), 995–1026.

McFadden, D. L. & Ruud, P. A. (1994), 'Estimation with simulation', *Review of Economics and Statistics* **76**(4), 591–608.

McFadden, D. L. & Train, K. E. (1996), Simulation variance in parameter estimation for a random-parameters logit model of car class choice, mimeograph, University of California, Berkeley, Department of Economics.

Revelt, D. & Train, K. E. (1996), Incentives for appliance efficiency: Random parameters logit models of households' choices, mimeograph, University of California, Berkeley, Department of Economics.

Ruud, P. A. (1991), Estimating multinomial choice models with simulation, mimeograph, Massachusetts Institute of Technology, Department of Economics.

Stern, S. (1989), 'Rules of thumb for comparing multinomial logit and multinomial probit coefficients', *Economics Letters* **31**(3), 235–238.

Train, K. E. (1995), Simulation methods for probit and related models based on convenient error partitioning, working paper 95-237, University of California, Berkeley, Department of Economics.