The Stability of Simulation Based Estimation of the Multiperiod Multinominal Probit Model with Individual specific Covariates

Ulrich Rendtel[*†], **Ulrich Kaltenborn**

Abstract

The multi-period multinomial Probit model (MMPM) is seen as a flexible tool to explain individual choices among several alternatives over time. There are two versions of this model: a) for each individual the covariates for all alternatives are known and b) for each individual only the parameters of the alternative which was chosen is known. The main difficulty with the MMPM was the calculation of the probability for the individual sequence of chosen alternatives, which requires the computation of the integral over a high dimensional multivariate Normal density. This remedy was removed by the Smooth Recursive Conditional (SRC) simulator. Several simulation studies have investigated the stability of the MMPM estimates with special emphasis to the number of replications of the SRC routine. In contrast to these studies, which use the case of alternative specific covariates, we use the case of the individual specific covariates.

We conclude that the MMPM with individual specific covariates is only weakly identified, generalizing Keane's (1992) result for the one period case. As a consequence the maximization of the simulated likelihood often converges to a singular covariance structure so that the SRC-routine stops iterating. This feature cannot be avoided by increasing the number of replications in the SRC-routine. The percentage of these failures rapidly increases with the number of alternatives.

Key words: Discrete Choice Models, Multi-period Multinomial Probit Models, Simulated Maximum Likelihood Method, Smooth Recursive Conditional Simulator, Panel Data.

[*]I thank Andreas Ziegler for the discussion of a previous version of this paper.

[†]Correspondence address: Institut für Statistik und Ökonometrie, FB Wirtschaftswissenschaft, Freie Universität Berlin, Garystr.21, D-14195 Berlin, Germany, Email: Rendtel@wiwiss.Fu-berlin.de

1 Introduction

In panel surveys on individual behaviour one is interested to explain how covariates explain the individual choice among a set of alternatives. The set of alternatives may be given by preference for a political party, the choice of living arrangements, the brand choice of consumers or the choice of traffic mode. We have two types of covariates: alternative specific covariates and individual specific covariates. For example, we may know the fairs and the duration for a trip to work for different traffic means. Here we have one coefficient for the whole set of variables for all alternatives, for example one coefficient for trip duration.

Individual specific covariates like income induce for each alternative a specific coefficient, for example the impact of income for the preference of party A, party B or party C. Unless a survey is especially designed for a certain purpose, like a panel on travel behaviour, the covariates for the alternatives which are not chosen remain unknown. For example, we know only the travel duration of the means chosen. Therefore we can only estimate the impact of travel duration by a special coefficient for car, bike or public transport. In panel surveys with general scope we will have almost always only individual specific covariates.

In the statistical analysis both cases are treated quite similar: the covariates enter a linear model for utility which is augmented by a stochastic component and the alternative with the maximum utility is chosen, see for example Manski/McFadden (1981), Börsch-Supan (1987), McFadden (1976). If we choose the multivariate Normal distribution for the stochastic component, we obtain the multiperiod multinominal Probit model (MMPM). Although this model class is very flexible it has not been used in empirical work because the likelihood involves multiple integrals over the normal distribution. This is a purely numerical problem, for which an efficient simulation based routine has been developed, the Geweke- Hajivassiliou-Keane (GHK) or the Smooth-Recursive-Conditioning (SRC) simulator, see Hajivassiliou/McFadden (1990), Börsch-Supan/Hajivassiliou (1993), Börsch-Supan (1994) or Keane (1994). If we maximize the SRC simulator approximation of the likelihood instead of the true likelihood we end up with the Simulated Maximum Likelihood Method (SMLM). Thus numerical feasibility is achieved now. Asymptotic results on the number of replications of the SRC are given by Hajivassiliou et al. (1996). Most examples deal with the case of two alternatives observed over some panel waves, the panel Probit model, see for example Keane (1994), Lee (1995, 1997), Inkmann (2000). Only very few authors used the model in the multi-alternative case, see for example Geweke et al. (1997), Kaltenborn (1997).

The starting point of this paper is the distinction between models with individual covariates and models with alternative specific covariates. Ziegler/Eymann (2001) use only models that are entirely based on alternative specific covariates. On the other hand, there is a result by Keane (1992), which seems not been widely known. Keane states that the estimation of the one period multinominal Probit model with only individual covariates and sufficient restrictions for formal identification is extremely unstable unless further model restrictions were made.

In this paper we will investigate this problem in a panel context. We use restrictions on the inter-temporal covariances of the random utility term that have an easy interpretation as conditional independence. Furthermore we restrict the parameters to be time-homogeneous. However, we did not impose additional restrictions on the contemporaneous covariance parameters between the altenatives. The question arises whether these restriction are sufficient to stabilize the likelihood maximization. Our analysis bases on a series of simulation experiments. The parameters of the simulation design were derived from empirical data on political party preference in the German socio-economic panel (GSOEP), see Kaltenborn (1997). With these empirical data there occurred severe numerical problems in the maximization of the simulated likelihood.

The article is organized as follows: section 2 sketches the MMPM and its likelihood. The formal identification restrictions and the conditional independence restrictions follow in section 2. The SRC simulator for the approximation of likelihood is displayed in section 4. In section 5 we introduce our simulation design. The results of the simulation runs are displayed in section 6. Section 7 concludes.

2 The multiperiod multinominal Probit model

Each individual n out of a sample of N persons has the opportunity at time t to make a choice i out of I alternatives. Let $u_{i,n,t}$ be the utility of choice i of individual n at time t. The standard linear utility model is:

$$(1) \quad u_{int} = X_{nt}^{'}\beta_i + Z_{int}^{'}\gamma + \varepsilon_{int} \quad i = 1, ..., I; n = 1, ..., N; t = 1, ..., T$$

Here X_{nt} is the $(K \times 1)$ vector of the individual specific covariates while Z_{int} is the $(L \times 1)$ vector of the alternative specific covariates. β_i is the $(K \times 1)$ parameter vector for the explanation of the deterministic effect of X_{nt} on utility for alternative i. γ is the $(L \times 1)$ parameter vector for the alternative specific covariates.

The ϵ_{int} are the stochastic utility components. We may stack the ϵ_{int} for each n and t to an $(I \times 1)$ vector ϵ_{nt}, which may be stacked again for each n into a $(IT \times 1)$ vector ϵ_n. It is assumed that the ϵ_n are independent and identical distributed following a $I \times T$ multivariate Normal distribution with mean 0 and covariance matrix Σ. We put our emphasis on models with only the term $X_{nt}^{'}\beta_i$, which is also called a discrete choice model.

The utility u_{int} in (1) cannot be observed. For each individual n we observe a sequence of selected alternatives $s_n = (i_{n1}, ..., i_{nT})$ from period 1 to period T. This selection sequence may be also characterized by the decision indicator variables:

$$D_{int} = \begin{cases} 1 & \text{if} \quad i = i_{nt} \\ 0 & \text{if} \quad i \neq i_{nt} \end{cases}$$

Utility maximization implies that for each person n and period t the alternative with the maximum utility u_{int} is chosen. This gives:

$$\begin{aligned} (2) \quad \mathrm{P}(D_{int} = 1) &= \mathrm{P}(u_{int} > u_{jnt} \quad \forall j \neq i) \\ &= \mathrm{P}(X'_{nt}\beta_i + \varepsilon_{int} > X'_{nt}\beta_j + \varepsilon_{jnt} \quad \forall j \neq i) \\ &= \mathrm{P}(X'_{nt}\beta_i - X'_{nt}\beta_j > \varepsilon_{jnt} - \varepsilon_{int} \quad \forall j \neq i) \end{aligned}$$

In order to ease the notation we use the shorthand $w_{jnt}^{(i)} = X'_{nt}\beta_i - X'_{nt}\beta_j$ and $\eta_{jnt}^{(i)} = \varepsilon_{jnt} - \varepsilon_{int}$. By

$$\begin{aligned} \boldsymbol{\eta}_{nt}^{(i)} &= \left(\eta_{1nt}^{(i)}, \ldots, [\eta_{int}^{(i)}], \ldots, \eta_{Int}^{(i)}\right)' \\ \mathbf{w}_{nt}^{(i)} &= \left(w_{1nt}^{(i)} \ldots, [w_{int}^{(i)}], \ldots, w_{Int}^{(i)}\right)' \end{aligned}$$

we denote the vectors of the corresponding elements where the element in brackets is omitted.

Thus in the one period case we have:

$$\begin{aligned} (3) \quad \mathrm{P}(D_{int} = 1) &= \mathrm{P}\left(\eta_{jnt}^{(i)} < w_{jnt}^{(i)} \quad \forall j \neq i\right) \\ &= F_{\eta_{nt}^{(i)}}\left(\mathbf{w}_{nt}^{(i)}\right) \end{aligned}$$

where $F_{\eta_{nt}^{(i)}}$ is the cumulative distribution function of $\eta_{nt}^{(i)}$.

Thus the computation of $P(D_{int} = 1)$ involves the evaluation of a $(I-1)$ dimensional integral over the density function of $w_{nt}^{(i)}$. For the computation

of the likelihood function we need the probability for the whole sequence s_n of selected alternatives:

$$\begin{aligned} P_{n,s_n} &= \mathrm{P}(D_{i_{n1},n,1}=1, D_{i_{n2},n,2}=1, \ldots, D_{i_{nT},n,T}=1) \\ &= \mathrm{P}(u_{i_{nt},n,t} > u_{jnt} \quad \forall t, \forall j \neq i_{nt}) \\ &= \mathrm{P}(X'_{nt}\beta_{i_{nt}} - X'_{nt}\beta_j > \varepsilon_{jnt} - \varepsilon_{i_{nt},n,t} \quad \forall t, \forall j \neq i_{nt}) \end{aligned} \tag{4}$$

For the $T \times (I-1)$ vector

$$\eta_n = \left(\{\eta_{n1}^{(i_{n1})}\}', \ldots, \{\eta_{nT}^{(i_{nT})}\}'\right)'$$

with the corresponding cumulative function F_{η_n} we have:

$$P_{n,s_n} = F_{\eta_n}(\mathbf{w}_{n1}^{(i_{n1})}, \ldots, \mathbf{w}_{nT}^{(i_{nT})}) \tag{5}$$

Therefore the loglikelihood l is given by:

$$l = \sum_{n=1}^{N} \ln F_{\boldsymbol{\eta}_n}\left(\mathbf{w}_{n1}^{(i_{n1})}, \ldots, \mathbf{w}_{nT}^{(i_{nT})}\right) \tag{6}$$

3 Identification of parameters

The MMPM is formulated by the slope vectors $\beta_i \quad (i = 1, \ldots, I)$ and by the $T \times I$ elements of the covariance matrix Σ. We have to distinguish restrictions to reach a formal identification of the model (see Bunch 1991) and restrictions which imply a certain interpretation of the model parameters. It is an inherent feature of the utility maximization model that:

1. only contemporaneous differences of the utilities are relevant
2. only the sign of these differences are relevant.

A standard identification rule with respect to 1) is to use a reference choice, say alternative I as the zero of the utility scale. For that reason we have $\beta_I = 0$ and $Cov(\epsilon_{Jnt}, \epsilon_{int}) = 0$ and $Var(\epsilon_{Int}) = 1$ for all $i \neq I$ and all n and t, see for example Keane (1992). The observation of only the sign of the utility differences is resolved by one free positive scale parameter, which can be fixed to achieve formal identification. Thus the most general formally identified MMPM has $1/2[(I-1)T][(I-1)T+1]-1$ free parameters for the covariance matrix Σ of the ϵ_{int}.

As Keane (1992) demonstrated for the one period case with $T = 1$ and $I = 3$ there is a remarkable tradeoff between the estimation of the slope

parameters and the covariance parameters if there are no further restrictions of Σ. This instability resulted in severe numerical problems during the likelihood maximization[1]:

- the Hessian was not invertable
- the algorithm did stop after the first iteration at different starting points that were obtained by restricting one covariance parameter at a fixed value
- there were very high covariances for the estimates obtained.

Usually one has only unspecific ideas about the contemporaneous covariances $Cov(\epsilon_{int}, \epsilon_{jnt})$ hence one would like to leave this part of the MMPM, apart from formal identification, unrestricted. However, in the multi-period case one can define restrictions with respect for intertemporal covariances and/or with respect to time homogenity. Hence one may hope to overcome the numerical instability of the MMPM estimates by using some adequate set of the above restrictions without imposing restriction on the contemporaneous covariances where no intuitive ideas exist in general. Kaltenberg (1997) used a conditional independence model:

$$(7) \quad \epsilon_{ins} \perp \epsilon_{jnt} \mid \epsilon_{int} \quad j \neq i, s \neq t$$

Conditional independence means that ϵ_{jnt} is influenced by ϵ_{ins} only by the contemporaneous value ϵ_{int} of ϵ_{ins}. For example, in the context of political party preference, where Kaltenberg (1997) introduced his model, conditional independence means: the impact of unexplained preference for party i at time s on preference for party j at time t is transmitted through the contemporaneous preference for party i at time t. In the context of political party preference this is a reasonable model.

Under the Normal distribution the conditional independence reduces to:

$$(8) \quad Cov(\epsilon_{ins}, \epsilon_{jnt} \mid \epsilon_{int}) = 0 \quad j \neq i, s \neq t$$

The above condition is equivalent under Normality with:

$$(9) \quad Cov(\epsilon_{ins}, \epsilon_{jnt}) = \frac{Cov(\epsilon_{int}, \epsilon_{jnt})\, Cov(\epsilon_{ins}, \epsilon_{int})}{Var(\epsilon_{int})}$$

see Tong (1990).

[1]For the integration Keane did not use the simulation approach described below but numerical integration.

We introduce the following shorthand notation: $\sigma_{ij}^{(t)} = Cov(\epsilon_{int}, \epsilon_{jnt})$ for the contemporaneous covariances at time t and $\sigma_{st}^{(i)} = Cov(\varepsilon_{ins}, \varepsilon_{int})$ for the intertemporal covariances for alternative i. With this shorthand (9) reads as:

$$(10) \quad \text{Cov}(\varepsilon_{ins}, \varepsilon_{jnt}) = \frac{\sigma_{ij}^{(t)} \sigma_{st}^{(i)}}{\sigma_{ii}^{(t)}}$$

In the case of temporal homogenity, i.e. $\sigma_{ij}^{(t)} = \sigma_{ij}$, and uniformity of intertemporal stability across alternatives (10) reduces to:

$$(11) \quad \text{Cov}(\varepsilon_{ins}, \varepsilon_{jnt}) = \sigma_{ij}\rho_{st}$$

where ρ_{st} is the intertemporal correlation between the alternatives.

This model is different from the model that Ziegler/Eymann (2001) use for their simulation study. They use a decomposition of ϵ_{int} into a constant individual preference for alternative i, α_{in}, and a time dependent term ν_{int} that follows an autoregressive first order process. It is apparent that the additive time-independent component α_{in} does not fit the multiplicative conditional independence scheme of equation (10). However, if the time-independent component is skipped and for the remaining term an autoregressive model is used, the Ziegler/Eymann model fits well into the conditional independence approach and we obtain:

$$\rho_{st}^{(i)} = \rho_i^{|s-t|}$$

where ρ_i is the serial correlation for alternative $(i = 1, ..., I)$.

The model restrictions have to fulfill the identification restrictions. In general this may be a cumbersome work. However, in the conditional independence model of equation (11) we may simply restrict $\sigma_{iI} = 0 \quad (i = 1, ..., I-1)$ and use $\sigma_{II} = 1$. For scale identification we have to restrict one further element of the diagonal.

4 The SRC simulator

Even for a small number of alternatives, for example $I = 4$, and a reasonable choice of panel waves, for example $T = 3$, one has to compute a $(4-1) \times 3 = 9$ dimensional integral over the multivariate Normal distribution, which is infeasible by numerical integration. For this reason Monte-Carlo integration methods have been developed. These methods approximate the value of

$$Q = \int g(\mathbf{x})\phi(\mathbf{x})d\mathbf{x}$$

where $\phi(x)$ is the joint density of a multivariate Normal density and $g(x)$ is an indicator of the multivariate integration interval. In the present case ϕ is the joint density of $\eta_n = (\epsilon_{1n1} - \epsilon_{i_{n1}n1}, \ldots, \epsilon_{InT} - \epsilon_{i_{nT}nT})$ and $g(x)$ is the indicator function of $A = (-\infty, X'_{n1}(\beta_{i_{n1}} - \beta_1)] \times \ldots \times (-\infty, X'_{nT}(\beta_{i_{nT}} - \beta_I)]$ The simplest method to simulate the value Q is to draw random vectors $\mathbf{x}$ according to the density ϕ and to count the relative frequency of random vectors falling into A. This is the so-called Frequency simulator, see Lerman/Manski (1981). The Frequency simulator has two drawbacks: a) it is inefficient for the computation of small probabilities and b) it is not smooth in the parameters $\beta_1, ..., \beta_I$.

The Smooth Recursive Conditioning (SRC) simulator prevents these drawbacks: it is an Importance simulator, that generate its random vectors in such a manner that many of them fall into region A, see Ripley (1987), Hajivassiliou et al. (1992,1996) or Stern (1992) for Importance simulators. And the SRC simulator is smooth in the parameters of the MMPM. Hence, the simulated integrals may be used instead of the exact integral in the Maximum Likelihood estimation of the MMPM. This generates the Simulated Maximum Likelihood Method (SMLM).

The SRC simulator is also called GHK simulator, named after Geweke, Hajivassiliou and Keane who independently developed this simulator, see Geweke (1991), Hajivassiliou/Mc Fadden (1990) and Keane (1990). A comparison of the SRC simulator with other simulators can be found in Hajivassiliou et al. (1992,1996), Börsch-Supan (1994) Börsch-Supan/Hajivassiliou (1993) and Kaltenborn (1997). One feature of the attractiveness of the SRC simulator is its ability to approximate the corresponding probabilities even with a small number of replications R. As a role of thumb R may be taken proportional to the number of dimensions of the multivariate normal distribution, see Börsch-Supan (1994).

The construction of the SRC simulator bases entirely on the Cholesky decomposition of the covariance matrix $\Omega = \Omega(s_n)$ of $\eta = \eta_n$. For the simulation of the probability $P(\mathbf{a} \leq \boldsymbol{\eta} \leq \mathbf{b})$ the SRC routine proceeds as follows. Let $\boldsymbol{\upsilon} \sim N(\mathbf{0}, \mathbf{I})$. If μ is the expected value of η and if Γ, with $\Gamma\Gamma' = \Omega$, is the Cholesky decomposition of the variance of η, we have:

$$\boldsymbol{\eta} - \boldsymbol{\mu} = \boldsymbol{\Gamma\upsilon} \sim N(\mathbf{0}, \boldsymbol{\Gamma\Gamma'})$$

Then the event $\mathbf{a} \leq \boldsymbol{\eta} \leq \mathbf{b}$ is equivalent to:

$$\mathbf{a}^* \leq \boldsymbol{\Gamma\upsilon} \leq \mathbf{b}^* \quad \text{with } \mathbf{a} - \boldsymbol{\mu} = \mathbf{a}^*, \quad \mathbf{b} - \boldsymbol{\mu} = \mathbf{b}^*$$

$$(12) \quad \widehat{=} \quad \begin{pmatrix} a_1^* \\ a_2^* \\ \vdots \\ a_J^* \end{pmatrix} \leq \begin{pmatrix} \gamma_{11} & 0 & \cdots & 0 \\ \gamma_{21} & \gamma_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \gamma_{J1} & \gamma_{J2} & \cdots & \gamma_{JJ} \end{pmatrix} \begin{pmatrix} \upsilon_1 \\ \upsilon_2 \\ \vdots \\ \upsilon_J \end{pmatrix} \leq \begin{pmatrix} b_1^* \\ b_2^* \\ \vdots \\ b_J^* \end{pmatrix}$$

where $J = (I-1)T$ is the dimension of η. If we solve the $i-$th row of the above equation system for υ_i we get for $i > 1$:

$$c_i \leq \upsilon_i \leq d_i \quad \text{with} \qquad c_i = \frac{a_i^* - \sum_{j=1}^{i-1} \gamma_{ij}\upsilon_j}{\gamma_{ii}} \quad \text{and}$$
$$d_i = \frac{b_i^* - \sum_{j=1}^{i-1} \gamma_{ij}\upsilon_j}{\gamma_{ii}}$$

For $i > 1$ the interval boundaries depend on the υ-values of the preceding rows. For $i = 1$ we have $c_1 = a_1^*/\gamma_{11}$ and $d_1 = b_1^*/\gamma_{11}$. The joint probability Q is obtained by the following factorization:

$$\begin{aligned} (13) \quad Q &= P(c_1 \leq \upsilon_1 \leq d_1) \\ &\quad \times P(c_2 \leq \upsilon_2 \leq d_2 \mid c_1 \leq \upsilon_1 \leq d_1) \\ &\quad \times \vdots \\ &\quad \times P(c_J \leq \upsilon_J \leq d_J \mid c_1 \leq \upsilon_1 \leq d_1, \ldots, c_{J-1} \leq \upsilon_{J-1} \leq d_{J-1}) \\ &= \Upsilon_1 \Upsilon_2(\upsilon_1) \cdots \Upsilon_J(\upsilon_1, \ldots, \upsilon_{J-1}) \end{aligned}$$

with

$$\begin{aligned} \Upsilon_i(\upsilon_1, \ldots, \upsilon_{i-1}) &= \Phi\left(\frac{b_i^* - \sum_{j=1}^{i-1} \gamma_{ij}\upsilon_j}{\gamma_{ii}}\right) - \Phi\left(\frac{a_i^* - \sum_{j=1}^{i-1} \gamma_{ij}\upsilon_j}{\gamma_{ii}}\right) \quad , \quad i > 1 \\ \Upsilon_1 &= \Phi\left(\frac{b_1}{\gamma_{11}}\right) - \Phi\left(\frac{a_1}{\gamma_{11}}\right) \end{aligned}$$

and Φ for the cumulative of the standard Normal distribution.

If we denote the truncated Normal distribution with bounds $a < b$ by $N(\mu, \sigma^2, a, b)$, the SRC simulator with R replications is given by:

$$\Xi_{src} = \frac{1}{R} \sum_{r=1}^{R} \prod_{i=1}^{I} \Upsilon_i(\xi_{1r} \cdots \xi_{i-1,r}) = \frac{1}{R} \sum_{r=1}^{R} \prod_{i=1}^{I} (\Phi(d_{ir}) - \Phi(c_{ir}))$$

where $\xi_{jr} \in N(0, 1, c_{jr}, d_{jr})$ with:

$$c_{ir} = \frac{a_i^* - \sum_{j=1}^{i-1} \gamma_{ij}\xi_{jr}}{\gamma_{ii}} \quad \text{and} \quad d_{ir} = \frac{b_i^* - \sum_{j=1}^{i-1} \gamma_{ij}\xi_{jr}}{\gamma_{ii}}$$

For each r the ξ_{ir} have to be generated recursively from the preceding ξ_{jr} for the alternatives with $j < i$. Although Ξ_{src} is unbiased for P_{n,s_n}, the simulated loglikelihood where P_{n,s_n} is replaced by Ξ_{src}, is not unbiased because of the nonlinear logarithm transformation.

The parameters of the MMPM enter not only the boundaries of the truncated Normal distribution but also the Cholesky decomposition. During the maximization of the simulated likelihood all trial values for new parameters have to deliver a positive definite Ω-matrix for all persons, otherwise the Cholesky decomposition cannot be performed and in consequence the simulated likelihood cannot be computed. In the Appendix we list the numerical problems we encountered and our strategies to overcome them.

5 The design of the simulation experiments

The motivation of these simulation experiments arose from an empirical analysis of political party preferences with data from the German socio-economic panel (GSOEP), see Kaltenborn (1997). We encountered severe numerical problems even with such a restrictive model for the intertemporal correlation like $\rho_{t,s} = \rho^{|t-s|}$, i.e. a simple serial correlation without differences with respect to alternatives. However, the contemporaneous covariances σ_{ij} were not restricted. We used only two metric covariates (age and income) for a model with 4 alternatives and two panel waves with about $N = 4000$ individuals. Despite $R = 500$ replications the obtained solution for the ML estimate for the Σ-matrix was quasi degenerate. Neither different starting values nor different sets of random numbers did remove the problem. It was not possible to find a global maximum within the range of non-degenerate covariance matrices, with the consequence that the SRC routine was not able to calculate the likelihood function. The only exceptions were models where we restricted the ρ-parameter. For $\rho = 0$ and $\rho = 0.5$ the SRC-routine reached convergence. However, for $\rho = 0.85$ the same numerical problem occurred again. In order to check also the algorithmic accuracy of our SRC-simulation module we did run the following simulation experiments[2].

[2]The results of the simulation runs indicated no apparent programming error for the cases with no convergence problems as appropriate estimates were obtained. Due to the bias of the SRC-routine one cannot expect that the true values are met in the mean.

Simulation studies have been presented so far only for the multiperiod Probit model with only two alternatives, see for example Keane (1994), Lee (1995,1997), Hyslop (1999) and Inkmann (2000). The multiperiod multinomial model was investigated so far only for the case of models with alternative specific covariates, see Ziegler/Eymann (2001).

The emphasis of our simulation design is the stability of the SRC-simulator, i.e. whether the routine converges at all. Furthermore, the extra variance induced by the random character of the simulated likelihood and the extra bias with respect to the ML solution are of interest[3]. As the ML solution is infeasible in this case we have to use a non-standard simulation design.

Let θ represent the parameter vector of the MMPM that is used for the simulation of the data. For each set of simulated data there exists an infeasible ML estimate $\hat{\theta}_{ML}$. In order to approximate $\hat{\theta}_{ML}$ the SRC-routine is used which bases itself on a random sample. The estimation result $\hat{\theta}_{SRC}$ will be in general different from $\hat{\theta}_{ML}$ and depends on the random likelihood.

If we use different $\hat{\theta}_{SRC}$ estimates for the same data and calculate their variance, this is a reasonable estimate of the additional variance σ^2_{SRC} due to the randomness of the SRC routine. On the other hand, if we keep the SRC random numbers fixed and vary the data, the corresponding variance σ^2_{Data} reflects the statistical precision of $\hat{\theta}_{ML}$.

In order to estimate the bias due to the SRC-routine we cannot compare $\hat{\theta}_{SRC}$ with $\hat{\theta}_{ML}$. However, if we use for each SRC-random sample also a new data set, the mean of the $\hat{\theta}_{ML}$ will be near[4] θ. Therefore $\hat{\theta}_{SRC} - \theta$ may be taken as an approximation for the simulation bias.

The covariates different from the constant were taken as multivariate normal. In order to mimic the empirical data from Kaltenberg (1997) we used high correlations ($\rho = 0.85$) between these variables and also with respect to time. The variances we normed to 1. The choice of the MMPM parameters was taken similar to those obtained in the empirical example of Kaltenborn (1997). However, the values were modified somehow to keep them distant from a singularity of the covariance matrix.

6 The results of the simulation experiments

Table 1 displays the results for the variance terms σ_{Data} and σ_{SRC}. There were $I = 3$ alternatives and $T = 2$ waves. The covariates contained the constant

[3]In non-linear models like the MMPM the ML estimate is in general not unbiased for a fixed number of observations.

[4]If the number of observations is reasonably high the asymptotic unbiasedness of the ML estimates applies.

and one Normal distributed metric variable. The slope coefficients for the metric variable were taken as $\beta_{21} = 0.198$ and $\beta_{22} = -3.785$. Thus negative values of this variable strongly favor the choice of the second alternative. The covariance between the unobserved utilities for alternative 1 and 2 was taken as $\sigma_{12} = -0.7$. As all variances were taken to be 1 the covariances represent correlations[5]. The temporal stability of the unobserved utilities was set quite high ($\rho = 0.85$). This may be seen as typical for individual habits, like political party preference.

The row "Fail" reports the percentage of simulations, which end up in a degenerate covariance matrix[6]. Due to time restrictions the number of simulation runs, reported in row "Sim", was only moderate. If the number of observations N is large for this simple model, like $N = 2000$, degenerate cases did not occur. However, for a lower sample size, N=500, degenerate solutions occur in a reasonable number of cases. Interestingly the number of these cases increases with the number of replications R, a fact that will be discussed later. It happened that once an estimation failed for a new data set then almost all other SRC-estimates failed for the same data set. This behaviour suggests that the poor convergence behavior is due to the underlying data set and does not depend on the randomness of the SRC-simulator.

The additional variance σ_{SRC} that is due to the simulation of the likelihood is substantial in the case of a low number of replications as $R = 10$. Therefore the asymptotic covariance expression derived from the Hessian will grossly under-estimate the variability of the SRC-estimates. If we increase R or N the additional variance decreases in a $1/R$ or $1/N$ fashion, as expected from asymptotic results, see Lee (1995).

The bias of the simulated ML estimation is displayed in table 2. For $R = 10$ and $N = 500$ the bias amounts 10 to 25 percent of the true value. If R is increased to 100, the bias decreases substantially. However, if we increase instead the number of the observations from $N = 500$ to $N = 2000$ the bias remains almost constant. σ^2_{total} gives the total variance with respect to the two sources: data generation and likelihood simulation. It is roughly identical with $\sigma^2_{Data} + \sigma^2_{SRC}$. In this experiment for each simulation run a new data set and a new set of random numbers for the SRC-routine is generated.

[5]The necessary identification restrictions request only two of these variances to be set to a fixed value. So there remains one free variance parameter, that could be estimated. However, in order to reduce the numerical problems with the positivity restrictions this parameter was regarded as known.

[6]However, the results of these simulation runs were taken into account for the computation of the variance terms. In the cases the value of the last iteration before singularity was reached was used.

Table 1: The additional variance of the SRC-routine. Number of alternatives $I = 3$, Number of covariates $K = 2$, Number of waves $T = 2$

R		**10**		**100**		**10**	
N		**500**		**500**		**2000**	
Sims		150		50		100	
Fail		10%		18%		0%	
	θ	σ_{data}	σ_{SRC}	σ_{data}	σ_{SRC}	σ_{data}	σ_{SRC}
beta11	-0.637	0.128	0.066	0.110	0.029	0.057	0.035
beta12	0.103	0.105	0.037	0.115	0.017	0.048	0.020
beta21	0.198	0.151	0.050	0.107	0.021	0.054	0.028
beta22	-3.785	0.221	0.096	0.275	0.036	0.139	0.046
rho	0.850	0.039	0.022	0.039	0.005	0.019	0.011
sig21	-0.700	0.210	0.162	0.221	0.073	0.146	0.093

Hence the simulation runs are independent from each other and the standard deviation of the mean is computed by dividing σ_{total} by $\sqrt{100} = 10$. The last row of Table 2 gives the p-value of a test whether the bias is different from zero. This hypothesis is to be rejected for a moderate replication size as $R = 10$. Only for $R = 100$ the significance of the bias disappears for three out of six parameters. Also in Table 2 the number of fails increases with R and decreases with N.

In the next tables we increase the number of alternatives and covariates. In Table 3 we display the simulation results for $I = 4$ alternatives and $K = 3$ covariates. Due to the increased complexity of the model we used an increased number of observations, $N = 2000$. However, as the calculations for one simulation run are very time consuming the number of simulations runs was only 40 for $R = 10$ and 13 for $R = 100$. For this large sample size there occurred no degenerated cases, like in the previous Tables 1 and 2 for $N = 2000$. Also the size of the bias remained approximately at the same level.

For $I = 5$ alternatives and $K = 4$ covariates there appeared massive numerical problems. In order to obtain at least some convergent cases all correlations terms in the Σ-matrix of the underlying model were set to zero. Still 3 out of 10 simulation runs ended with a degenerate covariance matrix. The average bias seems to be of the same magnitude as in the preceding simulation experiments.

The basic result of Keane (1992) for the one period multinomial Probit model was the trade–off between the slope parameters and the covariance

Table 2: The bias of the SRC-routine. Number of alternatives $I = 3$, Number of covariates $K = 2$, Number of waves $T = 2$

R		**10**				**100**			
N		**500**				**500**			
Sims		100				100			
Fail		8%				26%			
	θ	mean	σ_{total}	$\|bias\|$	p-val	mean	σ_{total}	$\|bias\|$	p-val
beta11	-0.637	-0.548	0.140	0.089	0.000	-0.606	0.152	0.031	0.046
beta12	0.103	0.130	0.124	0.027	0.031	0.119	0.120	0.017	0.168
beta21	0.198	0.145	0.144	0.052	0.000	0.161	0.142	0.036	0.012
beta22	-3.785	-3.669	0.309	0.117	0.000	-3.804	0.277	0.018	0.511
rho	0.850	0.786	0.042	0.064	0.000	0.845	0.038	0.005	0.216
sig21	-0.700	-0.527	0.297	0.173	0.000	-0.619	0.323	0.081	0.013

R		**10**			
N		**2000**			
Sims		100			
Fail		0%			
	θ	mean	σ_{total}	$\|bias\|$	p-val
beta11	-0.637	-0.539	0.084	0.098	0.000
beta12	0.103	0.146	0.059	0.043	0.000
beta21	0.198	0.146	0.072	0.051	0.000
beta22	-3.785	-3.615	0.141	0.171	0.000
rho	0.850	0.788	0.021	0.062	0.000
sig21	-0.700	-0.502	0.156	0.198	0.000

Table 3: The bias of the SRC-routine. Number of alternatives $I = 4$, Number of covariates $K = 3$, Number of waves $T = 2$

R		**10**				**100**			
N		**2000**				**2000**			
Sims		40				13			
Fail		0%				0%			
	θ	mean	σ_{total}	$\|bias\|$	p-val	mean	σ_{total}	$\|bias\|$	p-val
beta11	-0.637	-0.485	0.142	0.152	0.000	-0.564	0.159	0.073	0.124
beta12	0.103	0.249	0.117	0.146	0.000	0.165	0.095	0.062	0.035
beta13	0.048	0.156	0.123	0.109	0.000	0.102	0.114	0.055	0.108
beta21	0.198	0.022	0.155	0.176	0.000	0.131	0.185	0.067	0.214
beta22	-3.785	-3.440	0.239	0.345	0.000	-3.684	0.200	0.102	0.090
beta23	0.084	-0.145	0.241	0.229	0.000	0.064	0.224	0.020	0.758
beta31	-0.878	-0.799	0.175	0.079	0.007	-0.859	0.191	0.019	0.721
beta32	0.363	0.080	0.243	0.283	0.000	0.264	0.111	0.099	0.007
beta33	-3.732	-3.436	0.267	0.296	0.000	-3.725	0.291	0.007	0.929
rho	0.700	0.612	0.033	0.088	0.000	0.687	0.024	0.013	0.076
sig21	-0.500	-0.088	0.190	0.412	0.000	-0.301	0.168	0.199	0.001
sig31	0.200	0.199	0.157	0.001	0.959	0.189	0.229	0.011	0.871
sig32	0.100	0.339	0.174	0.239	0.000	0.143	0.129	0.043	0.256

Table 4: The bias of the SRC-routine. Number of alternatives $I = 5$, Number of covariates $K = 3$, Number of waves $T = 2$

R		**100**			
N		**2000**			
Sims		10			
Fail		30%			
	θ	mean	σ_{total}	$\|bias\|$	p-val
beta11	-0.637	-0.662	0.171	0.025	0.656
beta12	0.103	0.126	0.094	0.024	0.440
beta13	0.048	0.039	0.124	0.008	0.834
beta14	-1.594	-1.374	0.233	0.220	0.014
beta21	0.198	0.246	0.210	0.049	0.478
beta22	-3.785	-3.808	0.240	0.023	0.771
beta23	0.084	0.054	0.177	0.030	0.601
beta24	0.481	0.432	0.185	0.050	0.415
beta31	-0.878	-0.718	0.223	0.160	0.047
beta32	0.363	0.371	0.234	0.008	0.920
beta33	-3.732	-3.721	0.246	0.011	0.892
beta34	0.047	-0.003	0.174	0.049	0.391
rho	0.000	-0.001	0.065	0.001	0.969
sig21	0.000	-0.074	0.205	0.074	0.278
sig31	0.000	-0.067	0.357	0.067	0.565
sig32	0.000	0.003	0.151	0.003	0.943
sig41	0.000	0.366	0.441	0.366	0.025
sig42	0.000	0.031	0.165	0.031	0.569
sig43	0.000	-0.029	0.180	0.029	0.617

parameters that resulted in extremely unstable ML estimates of the model. However, Keane did not use the SRC-routine but numerical integration for the evaluation of the integrals over the bivariate Normal distribution. In the two-period model of our analysis we used a restrictive model for the contemporaneous and the intertemporal covariance terms. This stabilizes the model estimates to some degree. However, there still remains some trade-off between the covariate estimates: if we restrict the serial correlation between the unobserved utilities to zero this is compensated by high contemporaneous correlations between all alternatives, see Kaltenborn (1997) for an empirical example with data for political party preferences. This means that the likelihood function is flat along these trade-off parameter values. The MMPM itself does not assure that the covariance matrix of the random utilities is non-degenerate per se. However, the positive definiteness of the covariance matrix is a necessary prerequisite of the SRC-routine to evaluate the likelihood.

The SRC-routine uses a random approximation of the likelihood function. The lower R the more imprecise is the approximation of the likelihood function. As the approximation depends on random numbers the systematic dependencies of the parameters will be obscured. The higher we take R the smaller will be the random approximation error and the more apparent become the systematic parameter dependencies. For this reason one should expect that the percentage of degenerate cases increases with R, what is confirmed by the simulation experiments.

If we increase the number of alternatives to $I = 5$ or $I = 6$ the degenerate cases of the SRC-routine become a severe problem. The bias displayed in Table 4 is not larger than for smaller values of I. However, the variance – especially for parameter σ_{41} in Table 4 – becomes larger. So there will be cases where the estimated correlation between ϵ_{1nt} and ϵ_{4nt} is near 1. The resulting Σ-matrix is degenerate in this case.

7 Conclusions

The identification of the MMPM yields quite different results for models with alternative specific covariates and models with individual specific covariates. In the one period case Ziegler/Eymann (2001) state for the alternative specific case a high stability of the SRC-routine and a low bias even for moderate values of N (=1000, 2000) and R (=10,50). For the individual specific model Keane (1994) reported that the unrestricted covariance model is "quasi unidentified".

More panel waves offer the possibility to stabilize the estimation of the

MMPM with individual specific covariates by setting restrictions with respect to the intertemporal covariances and with respect to homogenity of the contemporaneous covariances across periods. We used a model derived from conditional independence which stabilizes the estimation of the MMPM to some extent. For high sample sizes like $N = 2000$ we encountered no numerical problems for models with only two periods and $I \leq 4$ alternatives. However, we observed substantial biases of the SRC-routine for small values of $R = 10$ which are reduced by increasing R to 100. This finding is different from the results of Ziegler/Eymann for the alternative specific MMPM with $I = 3, 4$. Even for small values of $N = 250, 500$ and $R = 10, 50$ there occurred no convergence problems and only small biases.

In contrast to the alternative specific case, there appeared numerical problems for smaller sample size like $N = 500$. In about 10 to 20 percent of the cases the SRC-routine converged to a degenerate covariance structure that stops the SRC-routine. There is a relationship between R and the percentage where the SRC-routine fails: for a given set of data the percentage of fails increases with R. This indicates that the convergence to a degenerate solution is a matter of insufficient model identification and not a matter of lacking accuracy of the simulated likelihood. This finding is also supported by the fact that a different random number choice for the SRC-routine for the same data does not remove the converge problems. Thus the problem of insufficient model identification cannot be solved by using high values for R.

If we increase the number of alternatives to $I = 5, 6$ the convergence problems become the main pattern even for large sample sizes like $N = 2000$. This happens also for data that come from a data generating process with no serial and no contemporaneous dependencies in the random utilities. the fact that an increase of the number of alternatives and/or the number of panel waves reduces the numerical stability of the MMPM is also reported for the case of alternative specific covariates, see Ziegler/Eymann (2001). However, for models with only individual specific covariates the instability is considerably more pronounced.

The source of the insufficient model identification are trade-offs between the contemporaneous covariances and the parameters for the serial correlation of the random utilities. Similar trade–offs were also obtained for the MMPM with only alternative specific covariates, see Ziegler/Eymann (2001). Even in those cases, where we have a data set without convergence problems in the likelihood maximization, there are two caveats for the obtained estimates. First,for small R the extra-variance due to the randomization of the likelihood function may be substantial in relationship to the standard asymptotic variance expressions. If we ignore this additional variance term we will over-estimate the precision of our estimates. However, this effect may

be reduced. by increasing the number of replications R.

A more serious problem is the potential bias for the slope parameters of the covariates. This problem is closely linked to the identification problem. In order to stabilize the estimates of the MMPM we have to restrict some parameters of the covariance matrix. It is an unpleasant feature of the MMPM that these restrictions effect also the estimation of the slope parameters. This was the result of Keane's (1994) analysis of the one period model. If we restrict the serial correlation parameters in a multiperiod model to different fixed values, we obtain quite different results for the slope parameters for political party preference, see Kaltenborn (1997). However, also in the case of models with alternative specific covariates Ziegler/Eymann (2001) report in their conclusions serious biases of the slope parameters if the covariance structure is mis-specified.

Thus we conclude that the MMPM with individual specific covariates is only weakly identified even in the presence of restrictive assumptions on the intertemporal error structure. Additionally, there appear numerical problems that force the SRC-routine to stop the likelihood maximization. These problems cannot be solved by increasing the number of replications in the SRC-routine.

Appendix

There are a number of numerical problems that have to be solved to run the SRC-routine which will be described in this appendix.

Random number generation

It is essential for the SRC-algorithm that the random numbers for the simulation of the P_{n,s_n} are kept constant during the likelihood maximization. Otherwise the simulated likelihood would not be smooth in the parameters. This can be achieved by starting the random number generator for each evaluation with the same seed value.

Due to the recursive nature of the SRC-algorithm it is necessary to use a random number generator with a long period and no systematic structures even in high dimensional spaces, see Ripley (1987, Chapter 3). For our simulation experiments we used the random number generator[7] of Matsumoto/Nishimura (1998) with an extreme long period.

[7] Code and information available at http://www.math.keio.ac.jp/matsumoto/emt.html

Computation of the likelihood derivatives

The analytical computation of the simulated likelihood is tedious as the model parameters enter the Cholesky decomposition. However, the analytical computation of the gradient is still feasible, see Kaltenborn (1997). If the gradient is computed numerically – in order to check the algebraic correctness of the analytical solution – we have to face the fact that due to the recursive scheme the numerical accuracy of the numerical gradient is only 5 to 7 digits. This is not enough to compute the second derivative numerically. So one should either use the analytical gradient and compute numerically the second derivative matrix from the analytical gradient or one should compute the Hessian matrix from the outer product of the gradient.

Choice of the maximization strategy

Because of the poor identification of the model parameters there are combinations of parameters with a flat profile of the corresponding likelihood values. Besides there are positivity and full rank constrains for the resulting covariance matrix. For such profiles the Newton-Raphson algorithm often fails to find the maximum. An alternative is to use a stepwise search technique along the gradient by the BFGS algorithm or the Fletcher-Reeves algorithm, see Press et al. (1992). These strategies avoid the computation of the second derivative and are therefore rather fast. They are known to be quite robust against the above maximization problems. Even more robust and stable but slower are Downhill-Simplex methods that do not use the gradient, see Press et al. (1992). In our simulation experiments we used a combination that switches between the Fletcher–Reeves and the Downhill–Simplex algorithm.

During the maximization we have to cope with situations where the Cholesky decomposition does not exist for the present parameter. Note that the MMPM does not exclude such parameters and the gradient may also be different from zero. In these situations we did step somewhat back along the gradient and used the conjugate gradient to reach a parameter region somewhat distant from the degenerate solutions. This procedure is repeated if the new parameter is also degenerate. After three unsuccessful trials the routine stops and returns the last valid parameter value. We made the experience that this routine is helpful to restart the optimization process but after a few iterations almost all experiments returned to a final degenerate solution. This indicates that the degenerate solution is not a matter of a poor optimization routine.

References

Börsch-Supan, A. (1987): Econometric Analysis of Discrete Choice; *Springer*

Börsch-Supan, A. (1994): Simulationsmethoden für die Analyse qualitativer Daten; *Allgemeines Statistisches Archiv 78, S. 20-39*

Börsch-Supan, A. / Hajivassiliou, V. (1993): Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models; *Journal of Econometrics 58, S. 347-368*

Bunch, D. (1991): Estimability in the Multinomial Probit Model; *Transportation Research 25B No 1, S. 1-12*

Geweke, J. (1991): Efficient simulation from the multivariate normal and student-t distributions subject to linear constaints; *in Keramidas, E. R. (Ed.); Computing Science and Statistics: Proc. 23rd Sympos. on the Interface; Interface Foundation Inc.; S. 571-578*

Geweke, J. / Keane, M. / Runkle, D. (1994): Statistical Inference in the multinomial multiperiod Probit Model; *Journal of Econometrics, 80, P. 125-165*

Hajivassiliou, V. A. (1993): Simulation Estimation Methods for Limited Dependent Variable Models; *in Maddala, G. S / Rao, C. R. / Vinod, H. D.; Handbook of Statistics Volume 11, Econometrics; North Holland; S. 519-543*

Hajivassiliou, V. A. / McFadden, D. (1990): The Method of Simulated Scores for the Estimation of LDV-Models with an Application to External Debt Crises; *Cowles Foundation Discussion Paper No.967; Yale University*

Hajivassiliou, V. A. / McFadden, D. / Ruud, P. (1992): Simulation of Multivariate Normal Orthant Probabilities: Theoretical and Computational Results; *Crowles Foundation Discussion paper No. 1021; Yale University*

Hajivassiliou, V. A. / McFadden, D. / Ruud, P. (1996): Simulation of Multivariate Normal rectangle Probabilities and their derivatives: Theoretical and Computational Results; *Journal of Econometrics, 72, 85-134*

Inkmann, J. (2000): Misspecified heteroskedasticity in the panel Probit model: *Journal of Econometrics 97, P. 227-259*

Kaltenborn, U. (1997): Die Anwendung simulativer Schätzverfahren für Discrete-Choice-Modelle am Beispiel von Daten des sozioökonomischen Panels; *Dissertation FU Berlin*

Keane, M. (1990): Four Essays in Empirical Macro and Labor Economics; *Unpublished PhD Dissertation, Brown University*

Keane, M. (1992): A Note on Identification in the Multinomial Probit Model; *Journal of Business and Economic Statistics 10, S. 193-201*

Keane, M. (1994): A Computationally Practical Simulation Estimator for Panel Data; *Econometrica 62, S. 95-116*

Lee, L.-F. (1995): Asyptotic bias in simulated maximum likelihood estimation of discrete choice models; *Econometric Theory 11, P. 437-483*

Lee, L.-F. (1997): Simulated maximum likelihood estimation of dynamic discrete choice models. Some Monte Carlo results; *Journal of Econometrics 82, P. 1-35*

Lerman, R. L. / Manski, C. F. (1981): On the Use of Simulated Frequencies to Approximate Choice Probabilities; *in Manski , C. F. / McFadden, D. (Eds.); Structural Analysis of Discrete Choice with Econometric Applications; MIT Press; S. 305-319*

Manski, C. F. / McFadden, D.;Eds. (1981): Structural Analysis of Discrete Choice with Econometric Applications; *MIT Press*

Matsumoto, M. / Nishimura, T. (1998): Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator; *ACM Transactions on Modeling and Computer Simulation*, 8, 3-30.

McCormick, G. P. (1983): Nonlinear Programming. Theory, Algorithms and Applications; *John Wiley & Sons*

McFadden, D. (1976): Quantal Choice Analysis: A Survey; *Annals of Economic and Social Measurement 5, S. 363-390*

McFadden, D. (1989): A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration; *Econometrica 57 No 5, S. 995-1026*

Press, W. H. / Teukolski, S. A. / Vettering, W. T. / Flannery, B. P. (1992): Numerical Recipes in C, The Art of Scientific Programming (Second Edition); *Cambridge University Press*

Ripley, B. D. (1987): Stochastic Simulation; *John Wiley & Sons*

Stern, S. (1992): A Method for Smoothing Simulated Moments of Discrete Probabilities in Multinomial Probit Models; *Econometrica 60, S. 943-952*

Tong, Y. L. (1990): The Multivariate Normal Distribution; *Springer*

Ziegler, A. / Eymann, A. (2001): Zur Simulated Maximum Likelihood-Schätzung von Mehrperioden-Mehralternativen-Modellen; *Allgemeines Statistisches Archiv 85, p. 319-342*