Discussion Paper No. 01-53

Simulated z-Tests in Multinomial Probit Models

Andreas Ziegler

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 01-53

Simulated z-Tests in Multinomial Probit Models

Andreas Ziegler

Non-technical summary

In economic studies, it is often necessary to use discrete choice models with more than two alternatives of a qualitative variable. Examples are the analysis of the choice of living arrangements, of the brand choice of consumers, of the choice of modes for the journey to work, of the household portfolio choice, or the analysis of employment reactions of firms. Multinomial probit models (in contrast e.g. to the restrictive multinomial logit models) are particularly suitable for the examination of such economic problems because of the flexible structure. For a long time, the application of flexible one- or multiperiod multinomial probit models was restricted because of the appearing multidimensional integrals. The use of such approaches has become numerically feasible, however, since the investigation of simulated estimation methods. With regard to the empirical application of flexible one- and multiperiod multinomial probit models, the simulated maximum likelihood method (SMLM), e.g. the simulated counterpart of the maximum likelihood method (MLM) (including the so-called GHK simulator), seems to be preferable. The asymptotic properties of the SMLM estimator as well as the properties with finite numbers of observations and with finite numbers of random draws in the GHK simulator have been investigated in the past. Such studies are essential to handle estimation results in the empirical work.

Often, the absolute estimated values are not the focus of empirical applications, but it is more interesting to know if the choice of an alternative depends on certain explanatory variables or on certain contemporary and/or intertemporal relationships. Based on the classical MLM, such problems can be examined with z-tests as special cases of classical test procedures. In a flexible one- or multiperiod multinomial probit model, however, the MLM, and thus the z-test, can be computationally infeasible because of the underlying multidimensional integrals. According to the inclusion of simulators in the MLM, classical test procedures can also be associated with simulation methods. By embedding a simulator in the z-test, one gets the simulated z-test. The asymptotic properties of simulated classical tests in general have been investigated in the past, too. But, in view of the empirical application of simulated z-tests, the properties with finite sample sizes and with finite amounts of random draws in the GHK simulator are again more important than the asymptotic properties.

Hence, within the framework of Monte Carlo experiments, this paper systematically compares different versions of the simulated z-test (using the GHK simulator) in one- and multiperiod multinomial probit models. In this context, deviations between the shares of type I errors and the basic significance levels are examined as well as the number of type II errors. In view of empirical applications, the number of observations and the number of random draws in the GHK simulator are varied. One important finding is that, in the flexible probit models, the tests on parameters of explanatory variables mostly provide robust results in contrast to the tests on variance-covariance parameters. Overall, neither the amount of random draws in the GHK simulator nor the choice of a certain version of the simulated z-test have a strong influence on the results. This finding refers to the conformity between the shares of type I errors and the basic significance levels as well as to the number of type II errors. In contrast, the number of type II errors in the simulated z-tests on variance-covariance parameters is reduced by increasing the sample size. Effects of misspecifications on simulated z-tests only appear in the multiperiod multinomial probit model. In this special case, the inclusion of the concept of the quasi maximum likelihood theory in the simulated z-test provides comparatively more favourable results.

Simulated z-Tests in Multinomial Probit Models

Andreas Ziegler*
Centre for European Economic Research (ZEW)
Department of Environmental and Resource Economics,
Environmental Management
P.O. Box 103443, D-68034 Mannheim, Germany
Phone: +49/621/1235-219, E-Mail: ziegler@zew.de

September 3, 2001

Abstract

Within the framework of Monte Carlo experiments, this paper systematically compares different versions of the simulated z-test (using the GHK simulator) in one- and multiperiod multinomial probit models. One important finding is that, in the flexible probit models, the tests on parameters of explanatory variables mostly provide robust results in contrast to the tests on variance-covariance parameters. Overall, neither the amount of random draws in the GHK simulator nor the choice of a certain version of the simulated z-test have a strong influence on the test results. This finding refers to the conformity between the shares of type I errors and the basic significance levels as well as to the number of type II errors. In contrast, the number of type II errors in the simulated z-tests on variance-covariance parameters is reduced by increasing the sample size. Effects of misspecifications on simulated z-tests only appear in the multiperiod multinomial probit model. In this case, the inclusion of the concept of the quasi maximum likelihood theory in the simulated z-test provides comparatively more favourable results.

*I would like to thank Axel Börsch-Supan, Elke Eberts, Angelika Eymann, Klaus Winckler and Thomas Zwick as well as the discussion partners Bernd Fitzenberger, Olaf Hübler, Winfried Pohlmeier, Gerd Ronning and the participants of the "DStatG-Nachwuchs-Workshop: Mikroökonometrie" in Blaubeuren for their useful comments.

1 Introduction

In economic examinations, it is often necessary to study discrete choice models with more than two alternatives of a qualitative variable. Examples are the analysis of the choice of living arrangements, the brand choice of consumers, the choice of modes for the journey to work, the choice of the practice location of general practitioners, or the household portfolio choice. Due to the flexible structure, multinomial probit models (in contrast e.g. to the restrictive multinomial logit models) are particularly suitable for the investigation of such economic problems. In particular, any intertemporal relationship which has an important influence on economic decisions can be modelled in this approach. Due to the increasing availability of panel data that include qualitative variables in several periods, such multiperiod multinomial probit models (MMPM) are likely to be applied more frequently in the future.

For a long time, the application of flexible one- and multiperiod multinomial probit models was restricted because of the underlying multidimensional integrals. The use of such approaches has become numerically feasible, however, since the investigation of simulated estimation methods (see e.g. Lerman and Manski, 1981, McFadden, 1989, Börsch-Supan and Hajivassiliou, 1993, Keane, 1994, Hajivassiliou and McFadden, 1998). In fact, such simulated (classical) estimations were already used in empirical applications of multinomial probit models (for the economic problems mentioned above see Börsch-Supan, 1992, Chintagunta, 1992, Bolduc, 1994, Bolduc et al., 1997, Asea and Turnovsky, 1998).

By combining classical estimation methods and simulators, several approaches are possible. With regard to the empirical use of multinomial probit models, the simulated maximum likelihood method (SMLM), e.g. the simulated counterpart of the maximum likelihood method (MLM), including the so called GHK simulator, seems to be preferable. On the one hand, this can be explained by the favourable numerical properties of the SMLM and the high precision of the GHK simulator. This simulated estimator is extremely practicable, too, since the common software packages can be used for the implementation, merely supplemented by the GHK simulation of the multidimensional integrals in the choice probabilities. In particular, this simulated estimation method was recently implemented directly in some software packages (e.g. GAUSSX and LIMDEP). More examples of the empirical use of the SMLM, including the GHK simulator, in multinomial probit models are Börsch-Supan and Pfeiffer, 1992, Börsch-Supan et al., 1992, and Bolduc et al., 1996.

The asymptotic properties of simulated classical estimation methods in general and of the SMLM in particular have been well known for a long time. Furthermore, within the framework of Monte Carlo experiments, properties of the SMLM have been investigated with finite numbers of observations and with finite numbers of random draws in the GHK simulator (see e.g. Keane, 1994, Lee, 1995, 1997a, Hyslop, 1999, Inkmann, 2000, on binary multiperiod probit models, Börsch-Supan and Hajivassiliou, 1993, Geweke et al., 1994, Mühleisen, 1994, Stern, 2000, on one-period multinomial probit models, Geweke et al., 1997, on the MMPM, Ziegler and Eymann, 2001, on one- and multiperiod multinomial probit models). Such systematic analyses are essential to handle estimation results in the empirical work.

But it is not only the absolute estimated value that is the focus of many empirical applications of probit models. Often, it is more interesting to know if the choice of an alternative depends on certain explanatory variables as well as on certain contemporary and/or intertemporal relationships. Based on the classical MLM, such problems can be examined with z-tests as special cases of classical test procedures. In a flexible multinomial probit model, however, the MLM, and thus the z-test, can be computationally infeasible because of the underlying multidimensional integrals. According to the inclusion of simulators in the MLM, classical test procedures can also be associated with simulation methods. By embedding a simulator in the z-test, one gets the simulated z-test. Simulated classical tests in general are discussed, and their asymptotic properties are derived, in the basic work of Lee (1999). But, in view of empirical applications of simulated z-tests, the properties with finite sample sizes and with finite numbers of random draws in the GHK simulator are again more important than the asymptotic properties.

Such simple simulated counterparts of the z-test were calculated regularly in the previous empirical SMLM estimations of probit models (see e.g. the literature cited above). The problem of the inclusion of simulators is left entirely unconsidered in these applications, however. In particular, it is often unclear which specific version of the simulated z-test is used. Despite the application in the empirical work, simulated z-tests in probit models (to my knowledge) have not been systematically examined yet in the literature. Even analyses of z-tests based on the MLM estimation are rare (so e.g. Guilkey and Murphy, 1993, on the binary multiperiod probit model). The only Monte Carlo experiments about simulated classical test procedures are available in Lee (1997b, 1999). Indeed, only selected probit models are tested in these articles, but tests on single parameters are ignored.

Hence, within the framework of Monte Carlo experiments, the goal of the present paper is to provide a systematic comparison of different versions of the simulated z-test in multinomial probit models representative of the applications. In this context, the deviations between the shares of the type I errors and the basic significance levels are considered as well as the number of the type II errors. The versions of the simulated z-test differ in the various simulated estimations of the information matrix. The three most important approaches are examined (see Lee, 1999). The first version uses the simulated Hessian matrix of the simulated log-likelihood function. The second estimation of the information matrix is constructed by the simulated counterpart of an outer product of gradient vectors of the simulated loglikelihood function. The third version is derived from the quasi maximum likelihood theory according to White (1982). This estimation of the information matrix includes the Hessian matrix as well as the gradients of the simulated loglikelihood function.

Due to the favourable properties, the GHK simulator is included in the SMLM estimation as well as in the various estimations of the information matrix. Thus, the Monte Carlo experiments in this paper try to give practical evidence about the handling of simulated z-tests (including the GHK simulator) in multinomial probit models. One important purpose of the present paper is the examination if the third version of the simulated z-test provides comparatively more preferable results (particularly when misspecified probit models are used) because the inclusion of the concept of the quasi maximum likelihood theory in simulated classical tests seems to outperform the other examined simulated classical tests in the experiments of Lee (1999).

Furthermore, in view of empirical applications, the number of observations and the number of random draws in the GHK simulator are varied in this paper. In contrast, Lee (1999) only examines one sample size. This makes statements about the effect of different numbers of observations on the results of simulated classical tests impossible. In particular, Lee only examines binary multiperiod probit models, but he does not analyze the empirically important multinomial probit models. In contrast, the present paper compares for the first time test results in one- and multiperiod multinomial probit models. The experimental design allows the inclusion of contemporary and (in the MMPM) intertemporal relationships. Due to the different results (in the flexible models), the comparative examination distinguishes between hypotheses about the coefficients of the explanatory variables and hypotheses about the variance-covariance parameters.

The organization of this paper is as follows. In the second section, the SMLM estimation in a flexible MMPM is explained. In the third section, simulated z-tests are described. In the fourth section, the design of the Monte Carlo experiments is illustrated. The Monte Carlo results are discussed in the fifth section, and in the last section, some conclusions are drawn.

2 Simulated maximum likelihoood estimation in multiperiod multinomial probit models

The origin of the microeconomic derivation of the MMPM (as well as of other discrete choice models) is that an agent chooses among a finite number of mutually exclusive alternatives of a qualitative variable in each of the considered time periods. In this paper, the following hypothetical utility function of observation i for alternative j in period t is examined:

$$\upsilon_{ijt} = \gamma' z_{ijt} + \varepsilon_{ijt} \qquad i = 1, \ldots, N; j = 1, \ldots, J; t = 1, \ldots, T$$

In this function, $z_{ijt} = (z_{ijt1}, \ldots, z_{ijtK})'$ is a (K-dimensional) vector with alternative specific attributes weighted by the parameter vector $\gamma = (\gamma_1, \ldots, \gamma_K)'$. Below, the z_{ijt} are summarized in the ($J \cdot K$-dimensional) vector $z_{it} = (z'_{i1t}, \ldots, z'_{iJt})'$, and then the z_{it} are subsumed in the ($T \cdot J \cdot K$-dimensional) vector $X_i = (z'_{i1}, \ldots, z'_{iT})'$. One arrives at the MMPM if the stochastic components ε_{ijt} are:

$$\varepsilon_i = (\varepsilon_{i11}, \ldots, \varepsilon_{iJ1}, \ldots\ldots, \varepsilon_{i1T}, \ldots, \varepsilon_{iJT})' \sim NV(0; \Sigma)$$

The ($J \cdot T$-dimensional) random vectors ε_i ($i = 1, \ldots, N$) are independent among each other and are independent of all X_i. Different versions of the MMPM result from various restrictions about Σ. If Σ is the identity matrix, one arrives at the specific multiperiod multinomial independent probit model. But in particular, a flexible structure of Σ, and thus, a flexible approach of the MMPM according to Börsch-Supan et al. (1992) is considered in the following.

Here, the stochastic utility components ε_{ijt} permit any contemporary correlation between the alternatives $j = 1, \ldots, J$ as well as time invariant stochastic effects and intertemporal autoregressive relationships (see also Ziegler and Eymann, 2001):

$$\varepsilon_{ijt} = \alpha_{ij} + \nu_{ijt} \qquad i = 1, \ldots, N; j = 1, \ldots, J; t = 1, \ldots, T$$

with

$$\nu_{ijt} = \rho_j \nu_{i,j,t-1} + \sqrt{1 - {\rho_j}^2}\eta_{ijt}$$

or with the assumption $\nu_{ij0} = \eta_{ij0}$

$$\nu_{ijt} = \sqrt{1 - {\rho_j}^2} \sum_{m=0}^{t-1} {\rho_j}^m \eta_{i,j,t-m} + {\rho_j}^t \eta_{ij0}$$

For $t = 0, 1, \ldots, T$ it is $\eta_{ijt} \sim NV(0; \sigma^2_{\eta_j})$, whereby the η_{ijt} are uncorrelated over all periods. For $t = 1, \ldots, T$ it is $(\forall j, j')$ $cov(\eta_{ijt}, \eta_{ij't}) = \sigma_{\eta_{jj'}}$. The ρ_j (whereby $|\rho_j| < 1$) are termed autocorrelation coefficients for category j. Further on, it is $\alpha_{ij} \sim NV(0; \sigma^2_{\alpha_j})$ with $cov(\alpha_{ij}, \alpha_{ij'}) = \sigma_{\alpha_{jj'}}$, whereby the α_{ij} and the ν_{ijt} are uncorrelated with each other. Finally, the components of the variance-covariance matrix Σ of ε_i ($i = 1, \ldots, N; j, j' = 1, \ldots, J; t, t' = 1, \ldots, T$ and $t \geq t'$) are:

$$cov(\varepsilon_{ijt}, \varepsilon_{ij't'}) = \sigma_{\alpha_{jj'}} + {\rho_j}^{(t-t')} \frac{\sqrt{1 - {\rho_j}^2}\sqrt{1 - {\rho_{j'}}^2}}{1 - \rho_j \rho_{j'}} \sigma_{\eta_{jj'}}$$

With respect to the formal model identification, in the Monte Carlo experiments of this paper the coefficients $\sigma^2_{\eta_J}$ and $\sigma^2_{\eta_{J-1}}$ are restricted to the value one and the coefficients $\sigma_{\eta_{jJ}}$ $(\forall j \neq J)$ are restricted to the value zero. Upon consideration of the multiperiod approach, the variance-covariance parameters of the stochastic effects $\sigma^2_{\alpha_J}$ and $\sigma_{\alpha_{jj'}}$ $(\forall j \neq j')$

as well as the autocorrelation coefficient ρ_J are restricted to the value zero (for the formal identification of multinomial probit models, see also Bolduc, 1992, Bunch, 1991, Dansie, 1985). Notice that in the basic SMLM estimations of this paper, instead of the variances $\sigma^2_{\eta_j}$ $(j = 1, \ldots, J-2)$ and $\sigma^2_{\alpha_j}$ $(j = 1, \ldots, J-1)$, the corresponding standard deviations σ_{η_j} and σ_{α_j} are included. Furthermore, instead of the covariances $\sigma_{\eta_{jj'}}$ $(j, j' = 1, \ldots, J-1; j \neq j')$, the corresponding correlation coefficients $corr(\eta_{ijt}, \eta_{ij't}) = \sigma_{\eta_{jj'}}/\sigma_{\eta_j}\sigma_{\eta_{j'}}$ are included. Accordingly, in the Monte Carlo experiments, the formulation of the null hypotheses refer to these transformed parameters.

Below, all free and estimating coefficients (i.e. the parameters of the explanatory variables and, in the flexible approach, the variance-covariance parameters) of the examined probit models are summarized in the vector $\theta = (\theta^{(1)}, \theta^{(2)}, \ldots)$. According to the stochastic utility maximization hypothesis (see e.g. Börsch-Supan, 1987: 12 ff), the observation i chooses category j in period t if j maximizes the utility under all J alternatives of the qualitative variable. Over time, every observation can choose J^T different category sequences in a multiperiod consideration. Thus, in regard to a chosen category sequence s, an observation has to choose a particular alternative in every period. In the flexible MMPM, the resulting probability $P_{is}(\theta)$ that observation i chooses category sequence s is a $(J-1) \cdot T$-dimensional integral.

As J and/or T grow, the computation of these multidimensional integrals is not feasible with deterministic numerical integration methods. Instead, the choice probabilities $P_{is}(\theta)$ can be approximated quickly and well with (unbiased) stochastic simulation methods, i.e. with R repeated transformed draws of pseudo random numbers (see e.g. the overviews in Hajivassiliou et al., 1996, Vijverberg, 1997). A simulated choice probability $\widetilde{P}_{is}(\theta)$ can be obtained by including such a simulator. In comparative Monte Carlo experiments, it has been shown that the GHK (Geweke-Hajivassiliou-Keane) simulator (see Börsch-Supan and Hajivassiliou, 1993, Geweke et al., 1994, Keane, 1994) outperforms other simulation methods for approximating the true probability (see also Mühleisen, 1994). Hence, in this paper, only the GHK simulator is considered.

By connecting an (unbiased) simulation method and the MLM, one obtains the SMLM (see e.g. Gouriéroux and Monfort, 1996: 41 ff). Below, all explanatory variables are subsumed in the vector X_i, and the (J^T-dimensional) vector $Y_i = (Y_{i1}, Y_{i2}, \ldots)'$ contains the observable endogenous variables

$$Y_{is} = \begin{cases} 1 & \text{if } i \text{ chooses category sequence } s \\ 0 & \text{else} \end{cases}$$

whereby $s \in S$, and S represents the set of all potential J^T category sequences. By embedding the simulator $\widetilde{P}_{is}(\theta)$ in the MLM approach and considering N independent observations

(Y_i, X_i) in the MMPM, one obtains the particular SMLM estimator:

$$\hat{\theta}_{SMLM} = (\hat{\theta}^{(1)}_{SMLM}, \hat{\theta}^{(2)}_{SMLM}, \ldots) = \arg\max_{\theta} \left[\sum_{i=1}^{N} \sum_{s \in S} Y_{is} ln \widetilde{P}_{is}(\theta) \right]$$

Below, the true unknown and estimating parameter vector is termed $\dot{\theta} = \left(\dot{\theta}^{(1)}, \dot{\theta}^{(2)}, \ldots\right)$.

3 Simulated z-tests

Statistical hypotheses about several components $\dot{\theta}^{(q)}$ ($q = 1, \ldots, dim\,\theta$) of $\dot{\theta}$ are tested in this paper. Thus, the null hypotheses have the appearance:

$$H_0 : \dot{\theta}^{(q)} = a$$

To examine such problems, the z-test as a special case of the Wald test (see e.g. Gouriéroux and Monfort, 1995: 84 ff) is the classical approach. The basis for this test procedure is the classical MLM estimation. But due to the existence of multidimensional integrals in the flexible MMPM, the MLM, and thus the z-test, is computationally not feasible if J and/or T are sizable. Corresponding to the inclusion of simulators in the MLM, such simulation methods can also be connected with classical test procedures (see Lee, 1999). Specifically, by embedding an (unbiased) simulator in the z-test, one obtains the simulated z-test. The test statistic is:

$$SZT = \frac{\hat{\theta}^{(q)}_{SMLM} - a}{\sqrt{\widehat{var}\left(\hat{\theta}^{(q)}_{SMLM}\right)}}$$

Note that the computation of SZT depends on the SMLM estimator $\hat{\theta}_{SMLM}$. In addition, in the flexible MMPM, more simulations have to be performed in the context of $\widehat{var}\left(\hat{\theta}^{(q)}_{SMLM}\right)$ and thus in the context of the estimation of the information matrix .

The GHK simulated estimation of the information matrix takes place in different ways in this paper. The first approach uses the Hessian matrix, the second approach uses the outer product of gradient vectors of the simulated loglikelihood function. Corresponding to the quasi maximum likelihood theory (see White, 1982), the third approach includes the gradients as well as the Hessian matrix of the simulated loglikelihood function. The three versions of the test statistic of the simulated z-test derived in these ways are referred to as SZT_1, SZT_2 and SZT_3. Note that the gradients of the simulated loglikelihood function are calculated analytically in the computation of the various simulated z-test statistics. Indeed, these gradients are differentiated numerically (by using the GAUSS module OPTMUM).

Neither the inclusion of a specific (unbiased) simulator nor the inclusion of a specific (consistently simulated) estimation of the information matrix in the simulated Wald test in general and in the simulated z-test in particular have any influence on the asymptotic properties (see Lee, 1999). The asymptotic properties of the simulated classical test procedures differ from the asymptotic properties of the unsimulated classical test procedures, however. This can particularly be ascribed to the different asymptotic properties between the underlying MLM and SMLM estimators (see also Gouriéroux and Monfort, 1991, Hajivassiliou and Ruud, 1994, Lee, 1995). The analyses of Lee (1999) implicate that under H_0, the test statistic SZT asymptotically has a normal distribution with mean $\sqrt{\lambda}$ and variance 1, if $\lim_{N\to\infty} \frac{\sqrt{N}}{R} = c$, in which R is the number of pseudo random draws in the considered simulator, c is a finite constant, and λ is a noncentrality parameter that arises from the noncentral χ^2 distribution of the test statistics of simulated classical test procedures. If $c = 0$, then $\lambda = 0$, and thus the asymptotic properties of the unsimulated test procedures are reached so that under H_0, the test statistic SZT asymptotically has a standard normal distribution.

When the simulated classical test procedures are applied to the empirical work, the asymptotic properties become less interesting again. Thus, the properties with finite numbers N of observations and with finite numbers R of random draws in the included simulator are much more important. Remember that in the following Monte Carlo experiments about simulated z-tests, the GHK simulator is considered exclusively. This choice refers to the underlying SMLM estimation as well as to the simulated estimation of the information matrix. Furthermore, it refers to the analysis of independent probit models, too, although the problem of multidimensional integrals does not appear in that approach, even if J and/or T are high. This strategy ensures that the influences of the model specification on the test results can be exclusively examined.

4 Design of the Monte Carlo studies

The following Monte Carlo experiments try to give practical evidence about the handling of simulated z-tests in one- and multiperiod multinomial probit models. As an example of the one-period multinomial probit model, the one-period four-alternative probit model is examined (this model is e.g. applied in Bolduc et al., 1996). In contrast to the consideration of a simpler one-period three-alternative probit model, in such a model, a more complex analysis of simulated z-tests on variance-covariance parameters is possible since in the one-period three-alternative approach, only two coefficients of the contemporary correlations can maximally be estimated. As an example of the MMPM, the five-period three-alternative probit model is considered (this model is e.g. applied in Börsch-Supan and Pfeiffer, 1992, and Börsch-Supan et al., 1992). Besides the common empirical application of such a MMPM, this choice results from the acceptable computing time, too. A strong increase of the number T of periods and/or the number J of alternatives would lead to calculating time problems.

In all experiments, 200 replications of the data generating process (DGP) are considered. It should be mentioned that this number is rather small for a systematic examination of test procedures. However, due to the long computing time, it was not possible to use many more replications of a DGP, even for the considered probit models. Furthermore, the strict investigation of the conformity with the underlying significance levels is not focussed in this paper, but the comparative analysis of the various versions of the simulated z-test, of several multinomial probit model specifications, and of different test problems are. In addition, the influence of the sample size N and of the amount R of random draws in the GHK simulator is studied. In this respect, 200 replications of the DGP are sufficient to draw many conclusions. The tested null hypotheses are

$$H_0 : \dot{\theta}^{(q)} = 0$$

or (if the parameter $\theta^{(q)}$ refers to a standard deviation σ_{η_j} or σ_{α_j})

$$H_0 : \dot{\theta}^{(q)} = 1$$

Based on the DGP explained below, these formulations of the null hypotheses guarantee that deviations between the shares of type I errors and the basic significance levels as well as numbers of type II errors can be investigated. According to the 5% and 10% quantiles of the standard normal distribution, the relative frequencies of the rejected null hypotheses in all 200 replications of the DGP are examined. The outcomes are derived from the three versions SZT_1, SZT_2 and SZT_3 of the simulated z-test statistic which refer to the particular simulated variance estimations.

The DGP in the considered flexible multinomial probit models are the same as in Ziegler and Eymann (2001). In this article the SMLM estimations of these specific probit models are exclusively investigated. By considering the same DGP in the present paper, relations between the basic SMLM estimates and the simulated z-tests on these parameters can be examined. In the experiments, the same (pseudo) random generated explanatory variables are used in all replications of the DGP (even for different numbers R of random draws in the GHK simulator). The explanatory variables generated at lower N are included in the SMLM estimation when the number N of observations is increased. By comparison, the random draws for deriving the GHK simulator are modified for any observation over the 200 replications of the DGP. But when N or R are increased successively, the random draws generated at lower N or R are included correspondingly.

4.1 Experiment one: One-period four-alternative probit model

Firstly, the following one-period multinomial probit model is examined ($i = 1, \ldots, N$; $j = 1, \ldots, 4$):

$$\upsilon_{ij1} = \gamma_1 z_{ij11} + \gamma_2 z_{ij12} + \varepsilon_{ij1}$$

The two alternative specific attributes are ($i = 1, \ldots, N; j = 1, \ldots, 4$):

$$z_{ij11} \sim NV(0;2) \quad z_{ij21} \sim NV(0;2)$$

In the DGP, the values of the parameters of these attributes are:

$$\dot{\gamma}_1 = 1 \quad \dot{\gamma}_2 = 0$$

In regard to the variance-covariance parameters of the DGP, on the one hand, the independent probit model is examined and, on the other hand, contemporary correlations are considered (since $T = 1$, it follows $\alpha_{ij} = \rho_j = 0; \forall j$), so that

$$\dot{\sigma}_{\eta_j} = 1 \quad (j = 1, \ldots, 4)$$

$$c\dot{o}rr(\eta_{ij1}, \eta_{ij'1}) = 0 \quad (j, j' = 1, \ldots, 4; j \neq j')$$

and

$$\dot{\sigma}_{\eta_1} = 1.5 \quad \dot{\sigma}_{\eta_2} = 0.5$$

$$c\dot{o}rr(\eta_{i11}, \eta_{i21}) = c\dot{o}rr(\eta_{i11}, \eta_{i31}) = c\dot{o}rr(\eta_{i21}, \eta_{i31}) = 0.5$$

The SMLM estimation is undertaken either in the independent probit model or in the flexible one-period four-alternative probit model. Here, five variance-covariance parameters are estimated in the general case. By including the last DGP and by estimating in the independent probit, a model misspecification occurs. The number of observations varies between $N = 1000$ and $N = 2000$, and the number of random draws in the GHK simulator varies between $R = 10$, $R = 50$ and $R = 200$.

4.2 Experiment two: Five-period three-alternative probit model

Concerning the analysis of panel data, the following MMPM is examined ($i = 1, \ldots, N$; $j = 1, \ldots, 3$; $t = 1, \ldots, 5$):

$$\upsilon_{ijt} = \gamma_1 z_{ijt1} + \gamma_2 z_{ijt2} + \varepsilon_{ijt}$$

In view of intertemporal relationships (see also the examinations in Geweke et al., 1997), the two alternative specific attributes are ($i = 1, \ldots, N; j = 1, \ldots, 3; t = 1, \ldots, 5$):

$$z_{ijt1} = z_{ij1}^{(1)} + z_{ijt1}^{(2)} \quad \text{whereby} \quad z_{ij1}^{(1)} \sim NV(0;1) \quad \text{and} \quad z_{ijt1}^{(2)} \sim NV(0;1)$$

$$z_{ijt2} = z_{ij2}^{(1)} + z_{ijt2}^{(2)} \quad \text{whereby} \quad z_{ij2}^{(1)} \sim NV(0;1) \quad \text{and} \quad z_{ijt2}^{(2)} \sim NV(0;1)$$

In the DGP, the values of the parameters of these attributes are:

$$\dot{\gamma}_1 = 1 \quad \dot{\gamma}_2 = 0$$

In regard to the variance-covariance parameters of the DGP, on the one hand, the independent probit model is examined, and on the other hand, various contemporary and intertemporal correlations are considered, so that

$$\dot{\sigma}_{\eta_j} = 1 \quad (j = 1, \ldots, 3) \qquad c\dot{o}rr(\eta_{ijt}, \eta_{ij't}) = 0 \quad (j, j' = 1, \ldots, 3; j \neq j')$$

$$\dot{\sigma}_{\alpha_j} = 0 \quad (j = 1, \ldots, 3)$$

$$\dot{\rho}_j = 0 \quad (j = 1, \ldots, 3)$$

and

$$\dot{\sigma}_{\eta_1} = 1.5 \qquad c\dot{o}rr(\eta_{i1t}, \eta_{i2t}) = 0.5$$

$$\dot{\sigma}_{\alpha_1} = 1.5 \qquad \dot{\sigma}_{\alpha_2} = 0.5$$

$$\dot{\rho}_1 = 0.8 \qquad \dot{\rho}_2 = 0.5$$

The SMLM estimation is undertaken either in the independent probit model or in the flexible MMPM. Here, six variance-covariance parameters are estimated in the general case. By including the last DGP and by estimating in the independent probit model, a misspecification occurs. The number of observations varies between $N = 250$ and $N = 500$, and the number of random draws in the GHK simulator varies between $R = 10$, $R = 50$ and $R = 200$.

5 Results

5.1 Experiment one: One-period four-alternative probit models

5.1.1 Simulated z-tests on the parameters in the independent probit model

Table 1 reports the results of the simulated z-tests in the one-period four-alternative independent probit model. The outcomes in the upper part of the table refer to the DGP characterized by the independent probit model. The results in the lower part of the table are based on the DGP that consists of contemporary correlations. Consequently, simulated z-tests are analyzed in a misspecified probit model in this part. Overall, the relative frequencies of the rejection of the null hypotheses $H_0 : \dot{\gamma}_1 = 0$ and $H_0 : \dot{\gamma}_2 = 0$ are illustrated in the table based on the significance levels 5% and 10%. The use of the three test statistics SZT_1, SZT_2 and SZT_3 is considered first.

In the table, the analysis of the tested null hypothesis $H_0 : \dot{\gamma}_1 = 0$, and thus the analysis of the number of the type II errors, is clear (since $\dot{\gamma}_1 = 1$ in the DGP, the validity of the alternative hypothesis H_1 is considered here). Independent from the number N of observations and the number R of random draws in the GHK simulator, as well as independent from the various versions of the simulated z-test statistic, the null hypothesis is correctly rejected without

Table 1: Share of the rejection of H_0 (simulated z-tests on the parameters in the one-period four-alternative independent probit model)

DGP: Independent probit model							
		5%			10%		
	H_0	SZT_1	SZT_2	SZT_3	SZT_1	SZT_2	SZT_3
$N = 1000$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 10$	$\dot{\gamma}_2 = 0$	0.030	0.030	0.030	0.075	0.080	0.065
$N = 1000$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 50$	$\dot{\gamma}_2 = 0$	0.030	0.030	0.030	0.080	0.080	0.080
$N = 1000$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 200$	$\dot{\gamma}_2 = 0$	0.030	0.030	0.030	0.075	0.080	0.075
$N = 2000$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 10$	$\dot{\gamma}_2 = 0$	0.060	0.060	0.055	0.110	0.110	0.110
$N = 2000$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 50$	$\dot{\gamma}_2 = 0$	0.050	0.055	0.050	0.105	0.100	0.105

DGP: Contemporary correlations							
		5%			10%		
	H_0	SZT_1	SZT_2	SZT_3	SZT_1	SZT_2	SZT_3
$N = 1000$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 10$	$\dot{\gamma}_2 = 0$	0.020	0.020	0.020	0.045	0.040	0.045
$N = 1000$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 50$	$\dot{\gamma}_2 = 0$	0.020	0.020	0.020	0.040	0.040	0.045
$N = 1000$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 200$	$\dot{\gamma}_2 = 0$	0.020	0.020	0.020	0.050	0.045	0.050
$N = 2000$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 10$	$\dot{\gamma}_2 = 0$	0.005	0.010	0.005	0.035	0.040	0.035
$N = 2000$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 50$	$\dot{\gamma}_2 = 0$	0.005	0.010	0.005	0.035	0.035	0.035

exception. Thus, there is no occurrence of a single type II error. Note that this result holds if the DGP is characterized by contemporary correlations.

In contrast to the aforementioned analysis, the testing of $H_0 : \dot{\gamma}_2 = 0$, and thus the analysis of the deviations between the shares of the type I errors and the basic significance levels, is slightly more sophisticated (since $\dot{\gamma}_2 = 0$ in the DGP, the validity of the null hypothesis H_0 is considered here). If the DGP is characterized by the one-period four-alternative independent probit model (see the upper part of table 1), the shares of the rejected H_0 have an excellent conformity with the underlying significance levels, in particular for $N = 2000$. Note that this result holds despite the small number of 200 replications of the DGP. Neither R nor the various simulated z-test statistics have a systematic influence on the relative frequencies.

If the DGP consists of contemporary correlations, the shares of the incorrectly rejected null hypotheses $H_0 : \dot{\gamma}_2 = 0$ are below the underlying significance levels (see the lower part of table 1). This finding holds for all variations of the sample size N and the amount R of random draws in the GHK simulator. Whereas an increase of N mostly effects a repeated decrease of the relative frequencies, the various test statistics SZT_1, SZT_2 and SZT_3 as well as R, again, have no systematic influence. Particularly, the number of the type I errors in this misspecified probit model is always lower than the corresponding number in the correctly specified independent probit model (see the upper part of table 1).

5.1.2 Simulated z-tests on the parameters of the explanatory variables

The test results in table 1 refer to the independent probit model. But the examination of simulated z-tests in the flexible one-period four-alternative probit model is much more interesting because of its avoidance of a model misspecification. Indeed, contrary to the analysis in section 5.1.1, the (simulated) estimation of the information matrix is hereby numerically problematic with the aid of the Hessian matrix of the simulated loglikelihood function. Repeatedly, negative estimated values of the variances of the SMLM estimates occur. In these cases, complex values of the test statistic SZT_1 appear. Obviously, these problems are related to the numerical (and not analytical) differentiation of the gradients of the simulated loglikelihood function. In this respect, the calculation of the two other versions SZT_2 and SZT_3 of the simulated z-test statistic are not problematical. Below, SZT_1 is not considered if negative simulated variance estimates occur.

The results of the simulated z-tests on the parameters of the explanatory variables in the flexible one-period four-alternative probit model are reported in table 2. As in table 1, this table contains the relative frequencies of the rejected null hypotheses $H_0 : \dot{\gamma}_1 = 0$ and $H_0 : \dot{\gamma}_2 = 0$ over all 200 replications of both considered DGP. According to the remarks above, only the test statistics SZT_2 and SZT_3 are analyzed. The findings on the left side of table 2 refer to the DGP characterized by the one-period four-alternative independent probit model. The test results on the right side of the table are based on the DGP that consists of contemporary correlations.

According to table 2, independently from the sample size N and independently from the amount R of random draws in the GHK simulator, $H_0 : \dot{\gamma}_1 = 0$ is correctly rejected in every case if the DGP consists of the independent probit model. Thus, neither in these simulated z-tests nor in the corresponding tests in the independent probit model (see table 1), does a

Table 2: Share of the rejection of H_0 (simulated z-tests on the parameters of the explanatory variables in the flexible one-period four-alternative probit model)

		DGP: Independent probit model				DGP: Contemporary correlations			
		5%		10%		5%		10%	
	H_0	SZT_2	SZT_3	SZT_2	SZT_3	SZT_2	SZT_3	SZT_2	SZT_3
$N = 1000$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$R = 10$	$\dot{\gamma}_2 = 0$	0.030	0.030	0.080	0.075	0.030	0.035	0.060	0.050
$N = 1000$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	0.995	1.000	0.995
$R = 50$	$\dot{\gamma}_2 = 0$	0.030	0.025	0.065	0.095	0.025	0.020	0.055	0.065
$N = 1000$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	0.990	1.000	0.995
$R = 200$	$\dot{\gamma}_2 = 0$	0.030	0.025	0.075	0.080	0.030	0.030	0.055	0.055
$N = 2000$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$R = 10$	$\dot{\gamma}_2 = 0$	0.050	0.050	0.110	0.110	0.015	0.010	0.025	0.035
$N = 2000$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	0.980	1.000	0.980
$R = 50$	$\dot{\gamma}_1 = 0$	0.050	0.050	0.100	0.105	0.015	0.010	0.035	0.040

type II error ever occur (since $\dot{\gamma}_1 = 1$ in the DGP, the validity of the alternative hypothesis H_1 is considered here again). But also if the DGP consists of contemporary correlations, the null hypothesis is only sporadically incorrectly maintained (applying the test statistic SZT_3). Consequently, independent from the underlying DGP, type II errors rarely occur in these simulated z-tests.

In contrast, the testing of $H_0 : \dot{\gamma}_2 = 0$ is again affected by the underlying DGP. If the DGP is characterized by the independent probit model, there are good conformities between the relative frequencies of the incorrectly rejected null hypotheses and the basic significance levels (see the left side of table 2) (since $\dot{\gamma}_2 = 0$ in the DGP, the validity of the null hypothesis H_0 is considered here again). When the number of observations increases to $N = 2000$, almost an identity arises between the shares of the type I errors and the underlying significance levels. In contrast, if the DGP consists of contemporary correlations, most of the shares of the incorrectly rejected $H_0 : \dot{\gamma}_2 = 0$ are comparatively lower (see the right side of table 2). The relative frequencies are particularly below the basic significance levels (this is noticeable for $N = 2000$). Again, R and the test statistics SZT_2 and SZT_3 have no systematic influence on the frequency of the rejected $H_0 : \dot{\gamma}_2 = 0$.

Overall, it should be emphasized that in the considered one-period four-alternative probit model, the simulated z-test on γ_2 is affected by the underlying DGP. If the DGP is characterized by the independent probit model, the relative frequencies of the incorrectly rejected

$H_0 : \dot{\gamma}_2 = 0$ are close to the basic significance levels. In contrast, if the DGP consists of contemporary correlations, the corresponding frequencies are below the basic significance levels. The last result holds in the misspecified independent probit model, too. Thus, no specific effects of the model misspecification on the testing of $H_0 : \dot{\gamma}_2 = 0$ arise here. Note that by using the various test statistics SZT_1 (in the independent probit model) as well as SZT_2 and SZT_3, the shares of the type I errors (independent of N, R and the underlying DGP) are very similar.

5.1.3 Simulated z-tests on the variance-covariance parameters

The simulated z-tests on the variance-covariance parameters in the flexible one-period four-alternative probit model give more mixed results. Table 3 reports the shares of the rejected null hypotheses about the coefficients of the contemporary correlations based on the significance levels 5% and 10%. According to the remarks at the beginning of section 5.1.2, only the test statistics SZT_2 and SZT_3 are analyzed again. On the left side of the table, the validity of the various null hypotheses H_0 is considered. Thus, the DGP is characterized by the one-period four-alternative independent probit model with $\dot{\sigma}_{\eta_1} = \dot{\sigma}_{\eta_2} = 1$ and $co\dot{r}r(\eta_{i11}, \eta_{i21}) = co\dot{r}r(\eta_{i11}, \eta_{i31}) = co\dot{r}r(\eta_{i21}, \eta_{i31}) = 0$. This allows the examination of the conformity between the shares of type I errors and the basic significance levels. In contrast, the validity of the various alternative hypotheses H_1 is considered on the right side of the table. Here, the DGP consists of contemporary correlations with $\dot{\sigma}_{\eta_1} = 1.5$, $\dot{\sigma}_{\eta_2} = 0.5$ and $co\dot{r}r(\eta_{i11}, \eta_{i21}) = co\dot{r}r(\eta_{i11}, \eta_{i31}) = co\dot{r}r(\eta_{i21}, \eta_{i31}) = 0.5$. Hence, the number of type II errors can be analyzed.

According to the left side of table 3, the testing of $H_0 : co\dot{r}r(\eta_{i11}, \eta_{i31}) = 0$ leads to the best conformities between the shares of the type I errors and the underlying significance levels, but the corresponding shares are also close to the basic significance levels if $H_0 : co\dot{r}r(\eta_{i21}, \eta_{i31}) = 0$ is tested (applying SZT_3). The relative frequencies of the incorrect rejection of the null hypotheses about the correlation coefficients are above as well as below the underlying significance levels. In contrast, if $H_0 : \dot{\sigma}_{\eta_1} = 1$ and $H_0 : \dot{\sigma}_{\eta_2} = 1$ are tested, the corresponding frequencies are never above the theoretical values of 5% and 10%. This finding is valid for all variations of the sample size N and the amount R of random draws in the GHK simulator as well as for both used simulated z-test statistics.

Indeed, the simulated z-tests on these variance parameters (in contrast to the simulated z-tests on the coefficients of the explanatory variables) show differences between the use of the test statistics SZT_2 and SZT_3. When SZT_2 is used, the share of the incorrectly rejected null hypotheses, in particular for $N = 1000$, is noticeably below the basic significance levels. The application of SZT_3 has a better conformity, however. But also the testing of $H_0 : co\dot{r}r(\eta_{i11}, \eta_{i21}) = 0$ shows for $N = 1000$ different shares of the type I errors when

Table 3: Share of the rejection of H_0 (simulated z-tests on the variance-covariance parameters in the flexible one-period four-alternative probit model)

		DGP: Independent probit model (Validity of H_0)				DGP: Contemporary correlations (Validity of H_1)			
		5%		10%		5%		10%	
	H_0	SZT_2	SZT_3	SZT_2	SZT_3	SZT_2	SZT_3	SZT_2	SZT_3
	$\dot{\sigma}_{\eta_1} = 1$	0.015	0.050	0.035	0.080	0.665	0.670	0.775	0.750
$N = 1000$	$\dot{\sigma}_{\eta_2} = 1$	0.010	0.050	0.030	0.080	0.090	0.235	0.305	0.365
	$c\dot{o}rr(\eta_{i11}, \eta_{i21}) = 0$	0.045	0.095	0.060	0.115	0.420	0.475	0.475	0.530
$R = 10$	$c\dot{o}rr(\eta_{i11}, \eta_{i31}) = 0$	0.075	0.065	0.090	0.110	0.860	0.860	0.910	0.870
	$c\dot{o}rr(\eta_{i21}, \eta_{i31}) = 0$	0.045	0.050	0.065	0.090	0.600	0.615	0.635	0.665
	$\dot{\sigma}_{\eta_1} = 1$	0.010	0.040	0.015	0.090	0.540	0.635	0.735	0.745
$N = 1000$	$\dot{\sigma}_{\eta_2} = 1$	0.010	0.050	0.035	0.080	0.060	0.350	0.235	0.500
	$c\dot{o}rr(\eta_{i11}, \eta_{i21}) = 0$	0.055	0.095	0.065	0.130	0.315	0.440	0.365	0.470
$R = 50$	$c\dot{o}rr(\eta_{i11}, \eta_{i31}) = 0$	0.060	0.055	0.085	0.095	0.890	0.865	0.935	0.915
	$c\dot{o}rr(\eta_{i21}, \eta_{i31}) = 0$	0.050	0.060	0.090	0.100	0.545	0.640	0.580	0.665
	$\dot{\sigma}_{\eta_1} = 1$	0.010	0.045	0.020	0.095	0.530	0.630	0.700	0.745
$N = 1000$	$\dot{\sigma}_{\eta_2} = 1$	0.015	0.040	0.030	0.085	0.045	0.370	0.265	0.475
	$c\dot{o}rr(\eta_{i11}, \eta_{i21}) = 0$	0.050	0.095	0.065	0.120	0.275	0.415	0.335	0.490
$R = 200$	$c\dot{o}rr(\eta_{i11}, \eta_{i31}) = 0$	0.060	0.055	0.085	0.105	0.870	0.865	0.930	0.920
	$c\dot{o}rr(\eta_{i21}, \eta_{i31}) = 0$	0.055	0.050	0.080	0.100	0.485	0.600	0.535	0.655
	$\dot{\sigma}_{\eta_1} = 1$	0.020	0.040	0.055	0.080	0.930	0.890	0.965	0.940
$N = 2000$	$\dot{\sigma}_{\eta_2} = 1$	0.025	0.045	0.065	0.090	0.535	0.625	0.685	0.815
	$c\dot{o}rr(\eta_{i11}, \eta_{i21}) = 0$	0.035	0.040	0.050	0.070	0.615	0.625	0.635	0.670
$R = 10$	$c\dot{o}rr(\eta_{i11}, \eta_{i31}) = 0$	0.035	0.045	0.100	0.115	0.990	0.975	0.995	0.985
	$c\dot{o}rr(\eta_{i21}, \eta_{i31}) = 0$	0.020	0.030	0.065	0.070	0.725	0.740	0.765	0.785
	$\dot{\sigma}_{\eta_1} = 1$	0.015	0.025	0.060	0.070	0.920	0.825	0.960	0.915
$N = 2000$	$\dot{\sigma}_{\eta_2} = 1$	0.025	0.035	0.065	0.075	0.615	0.650	0.790	0.795
	$c\dot{o}rr(\eta_{i11}, \eta_{i21}) = 0$	0.035	0.040	0.045	0.075	0.445	0.575	0.515	0.615
$R = 50$	$c\dot{o}rr(\eta_{i11}, \eta_{i31}) = 0$	0.040	0.050	0.100	0.100	0.990	0.965	0.995	0.970
	$c\dot{o}rr(\eta_{i21}, \eta_{i31}) = 0$	0.030	0.040	0.080	0.080	0.680	0.710	0.720	0.780

the various simulated z-test statistics are applied. Obviously, the use of SZT_3, and thus the inclusion of the concept of the quasi maximum likelihood theory in the simulated z-test, provides hereby slightly more robust results concerning the conformity with the basic significance levels. In contrast, R has again no systematic effects.

Differences occur in particular between the simulated z-tests on the coefficients of the explanatory variables (see table 2) and the simulated z-tests on the variance-covariance parameters when the number of the type II errors is analyzed. On the right side of table 3, incorrectly maintained null hypotheses about the parameters of the contemporary correlations arise repeatedly. The smallest number of type II errors, independent of N and R, occurs in the testing of $H_0 : \dot{corr}(\eta_{i11}, \eta_{i31}) = 0$. $H_0 : \dot{\sigma}_{\eta_1} = 1$ is also frequently correctly rejected, in particular for $N = 2000$. In contrast, for $N = 1000$, the testing of $H_0 : \dot{corr}(\eta_{i11}, \eta_{i21}) = 0$, and particularly the testing of $H_0 : \dot{\sigma}_{\eta_2} = 1$ lead to many type II errors. Again, R has no systematic influence, whereas the increase of N always generates an increase of the correctly rejected null hypotheses.

By comparison, the test statistic SZT_2 provides a higher number of correct rejections of $H_0 : \dot{corr}(\eta_{i11}, \eta_{i31}) = 0$ than SZT_3. This finding holds for all variations of N and R. Indeed, the values differ little, and are on a high level. In contrast, if $H_0 : \dot{corr}(\eta_{i11}, \eta_{i21}) = 0$ and $H_0 : \dot{corr}(\eta_{i21}, \eta_{i31}) = 0$ are tested, and in particular if $H_0 : \dot{\sigma}_{\eta_2} = 1$ is tested, much more type II errors arise. The relative frequencies of the correct rejection of $H_0 : \dot{\sigma}_{\eta_2} = 1$ are very low, especially for $N = 1000$. In this case, based on a significance level of 5% (and for $R = 50$ and $R = 200$), the share is merely close to the underlying significance level. By using SZT_3, the numbers of the type II errors can be decreased, even if there are still high values. Overall, it is shown that (also for high $N = 2000$ and $R = 50$) even the use of SZT_3 repeatedly leads to type II errors in the one-period four-alternative probit model if hypotheses about variance-covariance parameters are tested.

5.2 Experiment two: Five-period three-alternative probit model

5.2.1 Simulated z-tests on the parameters in the independent probit model

With regard to the analysis of panel data, first of all, table 4 reports the results of simulated z-tests in the five-period three-alternative independent probit model. Based on the significance levels 5% and 10%, the shares of the rejection of the null hypotheses $H_0 : \dot{\gamma}_1 = 0$ and $H_0 : \dot{\gamma}_2 = 0$ are examined. The results in the upper part of the table refer to the DGP characterized by the corresponding independent probit model. In contrast, the test results in the lower part of the table are based on the DGP that consists of contemporary and intertemporal correlations. Consequently, simulated z-tests are analyzed in a misspecified MMPM in this part. In order to test the null hypotheses, all versions SZT_1, SZT_2 and SZT_3 of the simulated z-test statistic are considered first.

Just as the analysis of $H_0 : \dot{\gamma}_1 = 0$ in the one-period four-alternative independent probit model (see table 1), according to table 4, this null hypothesis is also correctly rejected in the five-period three-alternative independent probit model in all 200 replications of the DGP.

Table 4: Share of the rejection of H_0 (simulated z-tests on the parameters in the five-period three-alternative independent probit model)

DGP: Independent probit model							
		5%			10%		
	H_0	SZT_1	SZT_2	SZT_3	SZT_1	SZT_2	SZT_3
$N = 250$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 10$	$\dot{\gamma}_2 = 0$	0.050	0.060	0.060	0.085	0.085	0.090
$N = 250$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 50$	$\dot{\gamma}_2 = 0$	0.055	0.060	0.055	0.080	0.080	0.090
$N = 250$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 200$	$\dot{\gamma}_2 = 0$	0.055	0.055	0.055	0.085	0.085	0.090
$N = 500$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 10$	$\dot{\gamma}_2 = 0$	0.040	0.040	0.040	0.095	0.105	0.085
$N = 500$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 50$	$\dot{\gamma}_2 = 0$	0.040	0.040	0.045	0.100	0.095	0.085
DGP: Contemporary and intertemporal correlations							
		5%			10%		
	H_0	SZT_1	SZT_2	SZT_3	SZT_1	SZT_2	SZT_3
$N = 250$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 10$	$\dot{\gamma}_2 = 0$	0.170	0.230	0.090	0.220	0.280	0.140
$N = 250$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 50$	$\dot{\gamma}_2 = 0$	0.165	0.225	0.085	0.220	0.295	0.145
$N = 250$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 200$	$\dot{\gamma}_2 = 0$	0.170	0.230	0.085	0.220	0.295	0.150
$N = 500$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 10$	$\dot{\gamma}_2 = 0$	0.140	0.210	0.085	0.185	0.270	0.120
$N = 500$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	1.000
$R = 50$	$\dot{\gamma}_2 = 0$	0.140	0.205	0.085	0.195	0.275	0.120

Hence, independent of the number N of observations and of the number R of random draws in the GHK simulator as well as independent of the various versions of the simulated z-test statistic, there is no single occurrence of a type II error (since $\dot{\gamma}_1 = 1$ in the DGP, the validity of the alternative hypothesis H_1 is considered here). Note that this test result is also valid in the misspecified MMPM, i.e. if the DGP is characterized by contemporary and intertemporal correlations.

Again, the analysis of $H_0 : \dot{\gamma}_2 = 0$ is more sophisticated. According to the upper part of table 4, the shares of the type I errors are extremely close to the basic significance levels (since $\dot{\gamma}_2 = 0$ in the DGP, the validity of the null hypothesis H_0 is considered here). This result holds for all variations of N and R as well as for all test statistics SZT_1, SZT_2 and SZT_3 again. Note that for a small number $N = 250$ of observations, in this correctly specified five-period three-alternative independent probit model, the conformity between the relative frequencies and the underlying significance levels is more accurate than the (good) conformity in the correctly specified one-period four-alternative independent probit model for moderate $N = 1000$ (see the upper part of table 1).

In contrast, if the DGP consists of contemporary and intertemporal correlations, the shares of the incorrect rejection of $H_0 : \dot{\gamma}_2 = 0$ are (partially noticeably) higher than the basic significance levels. Thus, strong impacts on the test results occur in the misspecified five-period three-alternative independent probit model. Whereas R, again, has no systematic influence on the frequencies of the type I errors, an increase of N mostly decreases the number of the incorrect rejection of $H_0 : \dot{\gamma}_2 = 0$, but to a minor degree. In contrast, the choice of the version of the simulated z-test statistics has stronger effects. When SZT_3 is applied, the shares of the incorrect rejection of $H_0 : \dot{\gamma}_2 = 0$ are closer to the basic significance levels in comparison to the use of SZT_1, and in particular to the use of SZT_2. Hence, in this misspecified independent probit model, the inclusion of the concept of the quasi maximum likelihood theory in the simulated z-test is preferable. But note that even by applying SZT_3, the relative frequencies of the type I errors are, without exception, all higher than the underlying significance levels.

5.2.2 Simulated z-tests on the parameters of the explanatory variables

In order to avoid misspecifications, contemporary and intertemporal correlations should be taken into account in the five-period three-alternative probit model. In this context, simulated z-tests on the parameter of the explanatory variables are examined now. Table 5 (like table 4) reports the shares of the rejection of $H_0 : \dot{\gamma}_1 = 0$ and $H_0 : \dot{\gamma}_2 = 0$. In accordance to the remarks in section 5.1.2, only the test statistics SZT_2 and SZT_3 are analyzed. The test results on the left side of the table refer to the DGP characterized by the corresponding independent probit model. The outcomes on the right side of the table are based on the DGP that consists of contemporary and intertemporal correlations.

According to table 5, $H_0 : \dot{\gamma}_1 = 0$ is correctly rejected in all 200 replications of both DGP if the test statistic SZT_2 is used. This finding holds for all variations of the number N of observations and the number R of random draws in the GHK simulator. The application of SZT_3 leads to very few type II errors (since $\dot{\gamma}_1 = 1$ in the DGP, the validity of the alternative hypothesis H_1 is considered here again). Thus, like in the flexible one-period four-alternative

Table 5: Share of the rejection of H_0 (simulated z-tests on the parameters of the explanatory variables in the flexible five-period three-alternative probit model)

		DGP: Independent probit model				DGP: Cont. and intert. correlations			
		5%		10%		5%		10%	
	H_0	SZT_2	SZT_3	SZT_2	SZT_3	SZT_2	SZT_3	SZT_2	SZT_3
$N = 250$	$\dot{\gamma}_1 = 0$	1.000	0.995	1.000	1.000	1.000	0.990	1.000	0.990
$R = 10$	$\dot{\gamma}_2 = 0$	0.055	0.065	0.080	0.085	0.075	0.050	0.125	0.115
$N = 250$	$\dot{\gamma}_1 = 0$	1.000	0.995	1.000	0.995	1.000	0.990	1.000	0.990
$R = 50$	$\dot{\gamma}_2 = 0$	0.055	0.065	0.075	0.085	0.060	0.060	0.115	0.120
$N = 250$	$\dot{\gamma}_1 = 0$	1.000	1.000	1.000	1.000	1.000	0.975	1.000	0.975
$R = 200$	$\dot{\gamma}_2 = 0$	0.055	0.060	0.075	0.090	0.055	0.070	0.110	0.110
$N = 500$	$\dot{\gamma}_1 = 0$	1.000	0.995	1.000	0.995	1.000	0.985	1.000	0.985
$R = 10$	$\dot{\gamma}_2 = 0$	0.035	0.035	0.095	0.090	0.075	0.060	0.110	0.095
$N = 500$	$\dot{\gamma}_1 = 0$	1.000	0.990	1.000	0.990	1.000	0.980	1.000	0.990
$R = 50$	$\dot{\gamma}_2 = 0$	0.040	0.040	0.090	0.095	0.070	0.065	0.105	0.110

probit model (see table 2), $H_0 : \dot{\gamma}_1 = 0$ is only sporadically incorrectly maintained. Overall, the application of SZT_2 seems to provide slightly more favourable results in this test problem in comparison to the test statistic SZT_3.

In view of the conformity between the shares of the type I errors and the basic significance levels, the testing of $H_0 : \dot{\gamma}_2 = 0$ also provides robust outcomes (since $\dot{\gamma}_2 = 0$ in the DGP, the validity of the null hypothesis H_0 is considered here again). Contrary to the flexible one-period four-alternative probit model (see table 2), the shares of the incorrectly rejected null hypotheses are all close to the underlying significance levels. According to the right side of table 5, this finding also arises if the DGP consists of contemporary and intertemporal correlations. In this case, $H_0 : \dot{\gamma}_2 = 0$ is, for the most part, only slightly more frequently rejected than on the basis of the DGP characterized by the independent probit model (see the left side of table 5). It should be emphasized that neither N nor the test statistics SZT_2 and SZT_3 have a systematic influence. Furthermore, R still has no specific effects.

5.2.3 Simulated z-tests on the variance-covariance parameters

The simulated z-tests on the variance-covariance parameters in the flexible five-period three-alternative probit model lead again to substantially more mixed results. Based on the significance levels 5% and 10%, table 6 reports the shares of the rejected null hypotheses about the coefficients of the contemporary, time invariant and autoregressive correlations.

According to the remarks in section 5.1.2, only the test statistics SZT_2 and SZT_3 are examined again. On the left side of the table, the validity of the various null hypotheses H_0 is considered. Thus, the DGP is characterized by the five-period three-alternative independent probit model with $\dot{\sigma}_{\eta_1} = 1$, $\dot{corr}(\eta_{i1t}, \eta_{i2t}) = \dot{\rho}_1 = \dot{\rho}_2 = 0$. This allows the analysis of the conformity between the shares of type I errors and the basic significance levels. It should be taken into account that under the consideration of the null hypotheses $H_0 : \dot{\sigma}_{\alpha_1} = 1$ and $H_0 : \dot{\sigma}_{\alpha_2} = 1$, this analysis is not possible for the two parameters of the stochastic effects, since it is $\dot{\sigma}_{\alpha_1} = \dot{\sigma}_{\alpha_2} = 0$ in the DGP. On the right side of table 6, the validity of the various alternative hypotheses H_1 is considered. Here, the DGP consists of contemporary and intertemporal correlations with $\dot{\sigma}_{\eta_1} = 1.5$, $\dot{corr}(\eta_{i1t}, \eta_{i2t}) = 0.5$, $\dot{\sigma}_{\alpha_1} = 1.5$, $\dot{\sigma}_{\alpha_2} = 0.5$, $\dot{\rho}_1 = 0.8$, $\dot{\rho}_2 = 0.5$. Thus, the number of type II errors can be examined.

According to the left side of table 6, in regard to the conformity with the basic significance levels, more instabilities occur in comparison with the simulated z-tests on the parameters of the explanatory variables (see table 5). Independent of N and R, the relative frequencies of the incorrect rejection of $H_0 : \dot{\sigma}_{\eta_1} = 1$ are always below the underlying significance levels, in particular when SZT_2 is used. In contrast, when $H_0 : \dot{corr}(\eta_{i1t}, \eta_{i2t}) = 0$ is tested, the corresponding shares never fall below the theoretical values 5% and 10%. With regard to the shares of the incorrect rejection of $H_0 : \dot{\rho}_1 = 0$ and $H_0 : \dot{\rho}_2 = 0$, values above as well as below the basic significance levels arise. When the last two null hypotheses are tested, the application of SZT_3 leads to noticeably higher shares of the type I errors than the application of SZT_2. Overall, in respect to the conformity with the basic significance levels, no general advantage of one version of the simulated z-test statistic can be derived. Furthermore, N and R have no systematic influence on the numbers of the type I errors.

In contrast, the increase of the sample size N (for the same amount R of random draws in the GHK simulator) leads again (as expected) always to a decrease of the numbers of the type II errors (see the right side of table 6). Indeed, like in the simulated z-tests on the variance-covariance parameters in the flexible one-period four-alternative probit model (see the right side of table 3), incorrectly maintained null hypotheses repeatedly occur. Again, these results contradict the outcomes when hypotheses about the coefficients of the explanatory variables are tested (see table 5). It can be recognized that the testing of $H_0 : \dot{\rho}_2 = 0$ causes, without exception, more type II errors than the testing of $H_0 : \dot{\rho}_1 = 0$. This finding is not surprising, however, since in the DGP it is $\dot{\rho}_1 = 0.8$, but $\dot{\rho}_2 = 0.5$

Furthermore, the increase from $R = 10$ to $R = 50$ leads (for the same N) to an increase in the number of the correct rejections of the null hypotheses mentioned at last. It should be noted, however, that using SZT_2 (for $N = 250$) the share of the incorrectly maintained $H_0 : \dot{\rho}_1 = 0$ rises if the number of random draws in the GHK simulator increases from $R = 50$ to $R = 200$. Generally, the increase of R often causes an increase of the number of the type

Table 6: Share of the rejection of H_0 (simulated z-tests on the variance-covariance parameters in the flexible five-period three-alternative probit model)

		DGP: Independent probit model (Validity of H_0)				DGP: Cont. and intert. corr. (Validity of H_1)			
		5%		10%		5%		10%	
	H_0	SZT_2	SZT_3	SZT_2	SZT_3	SZT_2	SZT_3	SZT_2	SZT_3
	$\dot{\sigma}_{\eta_1} = 1$	0.015	0.030	0.035	0.070	0.145	0.190	0.240	0.250
$N = 250$	$co\dot{r}r(\eta_{i1t}, \eta_{i2t}) = 0$	0.090	0.105	0.130	0.140	0.730	0.715	0.805	0.755
	$\dot{\sigma}_{\alpha_1} = 1$	–	–	–	–	0.685	0.625	0.700	0.665
$R = 10$	$\dot{\sigma}_{\alpha_2} = 1$	–	–	–	–	0.330	0.445	0.475	0.580
	$\dot{\rho}_1 = 0$	0.050	0.075	0.070	0.140	0.750	0.635	0.800	0.710
	$\dot{\rho}_2 = 0$	0.035	0.095	0.065	0.135	0.260	0.350	0.450	0.465
	$\dot{\sigma}_{\eta_1} = 1$	0.020	0.025	0.040	0.070	0.065	0.225	0.115	0.310
$N = 250$	$co\dot{r}r(\eta_{i1t}, \eta_{i2t}) = 0$	0.085	0.085	0.100	0.120	0.700	0.750	0.785	0.805
	$\dot{\sigma}_{\alpha_1} = 1$	–	–	–	–	0.425	0.455	0.480	0.525
$R = 50$	$\dot{\sigma}_{\alpha_2} = 1$	–	–	–	–	0.060	0.440	0.180	0.550
	$\dot{\rho}_1 = 0$	0.055	0.070	0.070	0.115	0.870	0.800	0.910	0.860
	$\dot{\rho}_2 = 0$	0.050	0.095	0.075	0.175	0.470	0.605	0.635	0.665
	$\dot{\sigma}_{\eta_1} = 1$	0.020	0.030	0.040	0.060	0.060	0.265	0.120	0.335
$N = 250$	$co\dot{r}r(\eta_{i1t}, \eta_{i2t}) = 0$	0.080	0.080	0.105	0.100	0.650	0.755	0.750	0.805
	$\dot{\sigma}_{\alpha_1} = 1$	–	–	–	–	0.320	0.465	0.395	0.520
$R = 200$	$\dot{\sigma}_{\alpha_2} = 1$	–	–	–	–	0.055	0.460	0.125	0.585
	$\dot{\rho}_1 = 0$	0.040	0.060	0.060	0.115	0.855	0.845	0.885	0.880
	$\dot{\rho}_2 = 0$	0.040	0.105	0.065	0.165	0.505	0.650	0.655	0.680
	$\dot{\sigma}_{\eta_1} = 1$	0.015	0.025	0.065	0.085	0.310	0.235	0.405	0.345
$N = 500$	$co\dot{r}r(\eta_{i1t}, \eta_{i2t}) = 0$	0.080	0.075	0.135	0.125	0.920	0.845	0.960	0.880
	$\dot{\sigma}_{\alpha_1} = 1$	–	–	–	–	0.800	0.715	0.835	0.755
$R = 10$	$\dot{\sigma}_{\alpha_2} = 1$	–	–	–	–	0.715	0.705	0.785	0.800
	$\dot{\rho}_1 = 0$	0.035	0.090	0.060	0.105	0.935	0.825	0.945	0.870
	$\dot{\rho}_2 = 0$	0.065	0.090	0.100	0.145	0.640	0.490	0.770	0.595
	$\dot{\sigma}_{\eta_1} = 1$	0.015	0.025	0.030	0.065	0.145	0.305	0.270	0.400
$N = 500$	$co\dot{r}r(\eta_{i1t}, \eta_{i2t}) = 0$	0.065	0.070	0.115	0.120	0.835	0.785	0.895	0.835
	$\dot{\sigma}_{\alpha_1} = 1$	–	–	–	–	0.515	0.520	0.545	0.585
$R = 50$	$\dot{\sigma}_{\alpha_2} = 1$	–	–	–	–	0.330	0.580	0.490	0.680
	$\dot{\rho}_1 = 0$	0.055	0.080	0.075	0.115	0.985	0.930	0.995	0.955
	$\dot{\rho}_2 = 0$	0.065	0.065	0.090	0.135	0.845	0.650	0.915	0.740

II errors. Using SZT_2, this finding holds in particular when $H_0 : \dot{\sigma}_{\alpha_2} = 1$ is tested and when R increases from 10 to 50. The choice between the test statistics SZT_2 and SZT_3 again provides no systematic advantage of one version. Overall, it should be emphasized that the testing of the hypotheses about the variance-covariance parameters in the five-period three-alternative probit model leads to many type II errors. This important outcome holds when both versions of the simulated z-test statistic are used, even for high $N = 500$ and $R = 50$.

6 Conclusions

Within the framework of Monte Carlo experiments, this paper examines various versions of the simulated z-test (using the GHK simulator) in one-period four-alternative probit models and in five-period three-alternative probit models. One important finding is that the tests on the parameter of the explanatory variables in the correctly specified probit models mostly provide robust results. This outcome refers to the precise conformity between the shares of the type I errors and the basic significance levels as well as to the rare type II errors. Only in the flexible one-period four-alternative probit model, the relative frequencies of the type I errors are noticeably lower than the basic significance levels. This result only holds if the DGP consists of contemporary and intertemporal correlations, however. It should be emphasized that the numbers of the incorrectly rejected null hypotheses about the coefficients of the explanatory variable are surprisingly not influenced by the sample size N or by the amount R of random draws in the GHK simulator.

In the context of the simulated z-tests on the coefficients of the explanatory variables, N and R have no effects on the numbers of the type II errors, too. In the one-period four-alternative probit model as well as in the five-period three-alternative probit model, the null hypothesis $H_0 : \dot{\gamma}_1 = 0$ is almost always correctly rejected. This result holds in the correctly specified probit models as well as in the misspecified independent probit models. It can not be proved that these test results can be transferred to other probit models, however. It is plausible that the outcomes are strongly influenced by the parameter formations in the DGP, in particular by $\dot{\gamma}_1 = 1$. In the future, more investigations on this test problem are desirable.

Furthermore, the results refer to multinomial probit models which exclusively include alternative specific explanatory variables. Own unpublished studies have shown that the added SMLM estimation of coefficients of individual specific explanatory variables (which do not vary between the alternatives) can be less precise and less stable than the SMLM estimation of coefficients of alternative specific variables (for the problem of the identification of the estimation of one-period multinomial probit models which exclusively include individual specific explanatory variables, see also Keane, 1992, or the application in Rennings et al., 2001). Thus, it is not clear if simulated z-tests on parameters of individual specific explanatory variables lead to similar results as simulated z-tests on parameters of alternative specific explanatory variables. Since the inclusion of individual specific explanatory variables in probit models occurs in empirical applications, a systematic examination of simulated z-tests on such parameters would be very desirable in the future.

In the misspecified five-period three-alternative independent probit model, the shares of the incorrect rejection of $H_0 : \dot{\gamma}_2 = 0$ are noticeably higher than the underlying significance levels. In this case, the inclusion of the concept of the quasi maximum likelihood theory (according to White, 1982) in the test statistic SZT_3 of the simulated z-tests is preferable. Hence, in comparison to the use of the test statistics SZT_1 and SZT_2, the conformity between the relative frequencies of the type I errors and the basic significance levels can be noticeably improved. In contrast, in the misspecified one-period four-alternative independent probit model, such differences between the various versions of the simulated z-test do not exist.

Indeed, in order to avoid misspecifications in the empirical application, flexible models should be used, e.g. the flexible one-period four-alternative probit model or the flexible five-period three-alternative probit model. With regard to the conformity between the shares of the type I errors and the basic significance levels in simulated z-tests on the coefficients of the (alternative specific) explanatory variables, the use of SZT_3 (surprisingly) does not yield substantial advantages in these approaches. In contrast, the application of SZT_1 can not be suggested because of the repeatedly occurring numerical problems. It should be remarked that such computational problems do not appear in the independent probit models, but in the empirically more important flexible multinomial probit models.

The results of the tests on variance-covariance parameters permit no clear recommendations of a specific version of the simulated z-test statistic, too. On the one hand, the use of SZT_3 seems to provide (as expected) more robust results in the flexible one-period four-alternative probit model. But, with regard to the deviations between the shares of the type I errors and the basic significance levels, as well as with regard to the numbers of the type II errors, this finding does not hold for all formulated hypotheses. In particular, compared to the use of SZT_2, the use of SZT_3 is not systematically preferable in the flexible five-period three-alternative probit model.

In contrast, a clear result is the substantially less stable testing of the hypotheses about the parameters of the contemporary and (in the five-period three-alternative probit model) intertemporal correlations compared with the testing of the hypotheses about the coefficients of the (alternative specific) explanatory variables. This finding corresponds to the likewise substantially less stable SMLM estimation of the variance-covariance parameters compared with the SMLM estimation of the parameters of the explanatory variables (see Ziegler and Eymann, 2001). Consequently, a strong relation between the underlying SMLM estimations of the coefficients and the corresponding simulated z-tests on these coefficients exists here.

Concerning the number of the type I errors in the testing of the hypotheses about the variance-covariance parameters, very uneven values occur when the various simulated z-test statistics are applied. It should be emphasized that besides the choice of SZT_2 and SZT_3, the number N of observations as well as the number R of random draws in the GHK simulator have surprisingly no systematic influence on the conformities between the shares of the incorrectly rejected null hypotheses and the basic significance levels. In contrast, the underlying SMLM estimation of the variance-covariance parameters with lower R and N provides stronger biases than the corresponding SMLM estimation with higher R and N (see also Ziegler and Eymann, 2001).

The instability of the simulated z-tests on the variance-covariance parameters particularly arises for the amounts of the type II errors. For all variations of N and R, hypotheses about the parameters of the contemporary and intertemporal relationships are very often incorrectly maintained in the examined flexible multinomial probit models. These outcomes contradict the results in the simulated z-tests on the coefficients of the (alternative specific) explanatory variables. Note that in the simulated z-tests on the variance-covariance parameters, R again has no systematic effect on the numbers of the type II errors. Only if hypotheses about the autocorrelation coefficients are tested, an increase of R (for the same N) often (but not without exception) leads to an increase in the numbers of the correctly rejected null hypotheses.

In contrast, an increase of the number N of observations always has a positive effect on the numbers of the correct rejections of the null hypotheses. Thus, in the examined flexible one- and multiperiod multinomial probit models, a rise of N leads to a partly substantial decrease in the numbers of the type II errors when the simulated z-tests on the variance-covariance parameters are undertaken. According to the remarks above, this finding indicates that an increasing number N of observations could also reduce the amount of the incorrectly maintained null hypotheses about the coefficients of the explanatory variables if the DGP is characterized by other parameter constellations. Hence, a systematic analysis of this problem would be desirable in the future, too.

References

Asea, P.K. and S.J. Turnovsky (1998), Capital Income Taxation and Risk-Taking in a Small Open Economy, *Journal of Public Economics* 68, 55-90.

Börsch-Supan, A. (1987), *Econometric Analysis of Discrete Choice. With Applications on the Demand for Housing in the U.S. and West-Germany*, Berlin u.a.

Börsch-Supan, A. (1992), Der Wohnungskonsum älterer Mitbürger: Wie lange selbständig? Wie oft in Mehrpersonenhaushalten?, in: Hujer R. et al. (Ed.), *Herausforderungen an den Wohlfahrtsstaat im strukturellen Wandel*, Frankfurt/New York, 143-172.

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

Börsch-Supan, A. and F. Pfeiffer (1992), Determinanten der Selbständigkeit in der Bundesrepublik Deutschland, in: Hujer R. et al. (Ed.), *Herausforderungen an den Wohlfahrtsstaat im strukturellen Wandel*, Frankfurt/New York, 257-287.

Börsch-Supan, A., V.A. Hajivassiliou, L.J. Kotlikoff and J.N. Morris (1992), Health, Children, and Elderly Living Arrangements. A Multiperiod-Multinomial Probit Model with Unobserved Heterogeneity and Autocorrelated Errors, in: Wise D.A. (Ed.), *Topics in the Economics of Aging*, Chicago, 79-104.

Bolduc, D. (1992), Generalized Autoregressive Errors in the Multinomial Probit Model, *Transportation Research B*, 26B (2), 155-170.

Bolduc, D. (1994), *A Practical Technique to Estimate Multinomial Probit Models in Transportation: Computational Details and an Application to a Disaggregate Mode Choice Problem*, Série des Documents de Travail du CREST No. 9421, Institut National de la Statistique et des Etudes Economique.

Bolduc, D., B. Fortin and S. Gordon (1997), Multinomial Probit Estimation of Spatially Interdependent Choices: An Empirical Comparison of Two New Techniques, *International Regional Science Review* 20 (1+2), 77-101.

Bolduc, D., G. Lacroix and C. Muller (1996), The Choice of Medical Providers in Rural Bénin: A Comparison of Discrete Choice Models, *Journal of Health Economics* 15, 477-498.

Bunch, D.S. (1991), Estimability in the Multinomial Probit Model, *Transportation Research B*, 25B (1), 1-12.

Chintagunta, P.K. (1992), Estimating a Multinomial Probit Model of Brand Choice Using the Method of Simulated Moments, *Marketing Science*, 11, 386-407.

Dansie, B.R. (1985), Parameter Estimability in the Multinomial Probit Model, *Transportation Research B*, 19B (6), 526-528.

Geweke, J., M. Keane and D. Runkle (1994), Alternative Computational Approaches to Inference in the Multinomial Probit Model, *The Review of Economics and Statistics* LXXVI (4), 609-632.

Geweke, J., M. Keane and D. Runkle (1997), Statistical Inference in the Multinomial Multiperiod Probit Model, *Journal of Econometrics* 80, 125-165.

Gouriéroux, C. and A. Monfort (1991), Simulation Based Inference in Models with Heterogeneity, *Annales d'économie et de statistique* 20/21, 69-107.

Gouriéroux, C. and A. Monfort (1995), *Statistics and Econometric Models, Vol. 2. Testing, Confidence Regions, Model Selection, and Asymptotic Theory*, Cambridge.

Gouriéroux, C. and A. Monfort (1996), *Simulation-Based Econometric Methods*, Oxford.

Guilkey, D.K. and J.L. Murphy (1993), Estimation and Testing in the Random Effects Probit Model, *Journal of Econometrics* 59, 301-317.

Hajivassiliou, V.A. and D. McFadden (1998), The Method of Simulated Scores for the Estimation of LDV Models, *Econometrica* 66 (4), 863-896.

Hajivassiliou, V.A. and P. Ruud (1994), Classical Estimation Methods for LDV Models Using Simulation, in: Engle, R.F. and D. McFadden (Eds.), *Handbook of Econometrics*, Vol IV, 383-2441.

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

Hyslop, D.R. (1999), State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women, *Econometrica* 67 (6), 1255-1294.

Inkmann, J. (2000), Misspecified Heteroscedasticity in the Panel Probit Model: A Small Sample Comparison of GMM and SML Estimators, *Journal of Econometrics* 97, 227-259.

Keane, M. (1992), A Note on Identification in the Multinomial Probit Model, *Journal of Business+Economic Statistics*, 10 (2), 193-200.

Keane, M. (1994), A Computationally Practical Simulation Estimator for Panel Data, *Econometrica* 62 (1), 95-116.

Lee, L.-F. (1995), Asymptotic Bias in Simulated Maximum Likelihood Estimation of Discrete Choice Models, *Econometric Theory* 11, 437-483.

Lee, L.-F. (1997a), Simulated Maximum Likelihood Estimation of Dynamic Discrete Choice Statistical Models. Some Monte Carlo Results, *Journal of Econometrics* 82, 1-35.

Lee, L.-F. (1997b), *Some Common Structures of Simulated Specification Tests in Multinormal Discrete and Limited Dependent Variables Models*, Working Paper No. 97-4, The Hong Kong University of Science and Technology, Department of Economics.

Lee, L.-F. (1999), Statistical Inference with Simulated Likelihood Functions, *Econometric Theory* 15, 337-360.

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

McFadden, D. (1989), A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration, *Econometrica* 57 (5), 995-1026.

Mühleisen, M. (1994), *Human Capital Decay and Persistence. A Simulation Approach to German Unemployment*, Frankfurt am Main.

Rennings, K., A. Ziegler and T. Zwick (2001), *Employment Changes in Environmentally Innovative Firms*, Discussion Paper No. 01-46, Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim.

Stern, S. (2000), Simulation-Based Inference in Econometrics: Motivation and Methods, in: Mariano R. et al. (Eds.), *Simulation-Based Inference in Econometrics: Methods and Applications*, Cambridge, 9-37.

Vijverberg, W.P.M. (1997), Monte Carlo Evaluation of Multivariate Normal Probabilities, *Journal of Econometrics* 76, 281-307.

White, H. (1982), Maximum Likelihood Estimation of Misspecified Models, *Econometrica* 50 (1), 1-25.

Ziegler, A. and A. Eymann (2001), Zur Simulated Maximum-Likelihood-Schätzung von Mehrperioden-Mehralternativen-Probitmodellen, *Allgemeines Statistisches Archiv* 85 (3), 319-342.