Spiess, Nagl, Hamerle:

Probit models: Regression parameter estimation using the ML principle despite misspecification of the correlation structure

Sonderforschungsbereich 386, Paper 67 (1997)

Online unter: http://epub.ub.uni-muenchen.de/

Probit models: Regression parameter estimation using the ML principle despite misspecification of the correlation structure

Martin Spiess*, Willi Nagl† and Alfred Hamerle*

Abstract

In this paper it is shown that using the maximum likelihood (ML) principle for the estimation of multivariate probit models leads to consistent and normally distributed pseudo maximum likelihood regression parameter estimators (PML estimators) even if the 'true' correlation structure of the responses is misspecified. As a consequence, e.g. the PML estimator of the random effects probit model may be used to estimate the regression parameters of a model with any 'true' correlation structure. This result is independent of the kind of covariates included in the model. The results of a Monte Carlo experiment show that the PML estimator of the independent binary probit model is inefficient relative to the PML estimator of the random effects binary panel probit model and two alternative estimators using the 'generalized estimating equations' approach proposed by Liang and Zeger (1986), if the 'true' correlations are high. If the 'true' correlations are low, the differences between the estimators are small in finite samples and for the models used. Generally, the PML estimator of the random effects probit panel model is very efficient relative to the GEE estimators. Therefore, if correlations between binary responses cannot be ruled out and the 'true' structure of association is unknown or infeasible to estimate, the PML estimator of the random effects probit model is recommended.

Key words: Multivariate probit model; Random effects probit model; Categorical responses; Maximum likelihood; Misspecification; Simulation study

*Department of Economics, University of Regensburg, Germany. M. Spiess is supported by the Deutsche Forschungsgemeinschaft, SFB 386

†Department of Economics, University of Konstanz, Germany

1 Introduction

The use of panel data for the estimation of regression models has become increasingly popular in econometric, biometric and psychometric applications. One important advantage using panel data sets compared to using cross section or single time-series data sets is the possibility to estimate more realistic behavioral models (see e.g. Hsiao, 1986). If the correlated response variables are continuous, then efficient estimation of the parameters in regression models generally poses no problems. In contrast, efficient estimation of categorical response probit models, which are the focus of the present paper, generally is quite expensive if at all possible.

There are, however, some special cases where maximum likelihood (ML) estimation is feasible. For example, Fitzmaurice and Laird (1993) proposed the ML based estimation of mixed parameter models for binary data, where the association of the observable responses is modelled in terms of conditional log odds-ratios. The regression parameter estimator can then be shown to be consistent even if the association structure is misspecified. In this case, the so-called 'sandwich' estimate of the covariance matrix of the regression parameter estimator has to be used. If, however, a threshold model, i.e. a latent variable model, is assumed, then generally not the association structure of the observable responses but of the latent errors is modelled. Furthermore, the interpretation of the association parameters based upon log odds-ratios is at least unusual if e.g. a probit link function is used. Another drawback is that these mixed parameter models are not appropriate for the analysis of models with an unequal number of observations within each block or cluster. In applications, calculation of the ML estimates is only feasible for less than five or six observations within each block.

As an example of a model for which ML estimation is feasible, consider the random effects probit models, where unobservable random effects — often termed heterogeneity — which are assumed to be independently distributed from the covariates are introduced in the model. A well known and often used model is the random effects probit model with binary responses which can be estimated using Gauss-Hermite quadrature (see e.g. Bock and Aitkin, 1981; Bock and Liebermann, 1970; Butler and Moffit, 1982; Heckman, 1981; Maddala, 1987). However, the formulation of the random effects probit model implies the assumption of a certain correlation structure. As a consequence of this restrictive assumption three research topics are of special interest.

The first topic is the research on simulation methods for the approximative calculation of the ML estimator in general probit models (e.g. Lerman and Manski, 1981; McFadden, 1989). Although promising progress is made within this approach (e.g. Börsch-Supan and Hajivassiliou, 1993; Hajivassiliou, McFadden and Ruud, 1996), not much is known about the required computing time and the finite sample properties of the estimators using these methods compared to the ML estimators even in the simple case of the binary random effects probit model

with a correctly specified correlation structure.

The second important direction of research is the work on derivation of non-ML estimators for probit models with any correlation structure which are consistent and asymptotically normal. Starting with the implicite or explicite premise that the ML regression parameter estimator of a multivariate probit model generally is not robust with respect to misspecification of the true correlation structure, this second topic is addressed e.g. by Avery, Hansen and Hotz (1983) or Liang and Zeger (1986). Both approaches allow the estimation of probit models with general serial correlations, however, at the cost of lower efficiency relative to the ML estimator even if the assumed equicorrelation structure is correct (Maddala, 1987; Spiess and Hamerle, 1995).

The third topic concerns the question whether ML regression parameter estimators are still consistent and normally distributed even if the correlation structure is misspecified. Within the more general framework of research on estimation in models where e.g. the link function or error distribution is misspecified (e.g. Fahrmeir, 1990; Gourieroux, Montfort and Trognon, 1984a, 1984b; Huber, 1967; Li and Duan, 1989; Robinson, 1982; Ruud, 1983, 1986; White, 1981, 1982), Gourieroux, Montfort and Trognon (1984a) (see also Robinson, 1982) showed that the ML estimator of the independent probit model with binary responses remains consistent — although inefficient (Maddala, 1987) — even if the responses are not independent. Under certain conditions concerning the covariates it can be shown that the use of the ML principle still leads to consistent estimators even if the distributional assumptions are violated (Li and Duan, 1989; Ruud, 1983, 1986). However, it is not clear whether the ML regression parameter estimator of the general probit model is still consistent in the presence of any kind of covariates if the correlation structure is misspecified. In particular, within the proposed approaches on estimation of misspecified models, the ML estimator of the random effects binary probit model cannot be shown to be consistent in this case.

This paper is organized as follows. In section 2 the general model is described and in section 3 it is shown that the ML regression parameter estimator of a probit model is consistent without any assumption on the distribution of the covariates even if the correlation structure is misspecified. In section 4 as a special case of the general probit model the binary model for panel data is considered. Section 4.1 presents the statistical specification of the ML estimator of the random effects binary probit model and section 4.2 shortly describes as an alternative approach for the estimation of binary models with correlated responses the 'generalized estimating equations' approach suggested by Liang and Zeger (1986). The format and results of a Monte Carlo experiment designed to evaluate the efficiency of the above estimators in finite samples are presented in section 5 and 6, respectively. Conclusions can be found in section 7.

2 Model and Notation

For the models considered we have N blocks ($n = 1, \ldots, N$) and T observations ($t = 1, \ldots, T$) in every block (generally individuals). Although not considered, the results discussed in this paper may be generalized allowing the number of time-series observations, T_n, to vary between blocks. Only for simplicity, we assume $T_1 = \cdots = T_N$. Let $y_n = (y_{n1}, \ldots, y_{nT})'$ denote the $(T \times 1)$ vector of observable responses for the nth block. Furthermore, let $x_{nt} = (x_{nt1}, \ldots, x_{ntP})'$ denote the $(P \times 1)$ vector of covariates associated with the ntth observation, X_n the $(T \times P)$ matrix of covariates associated with the nth block and X the $(NT \times P)$ matrix having full column rank associated with all NT observations. Since the essential arguments that follow do not change whether the covariates are known constants or considered as random, no assumption will be made concerning the kind of covariates. However, using panel data sets both kinds of covariates are plausible: for example modelling time points often leads to a design matrix with known constants whereas e.g. the expense for certain goods at different time points may be considered as a realization of a random variable.

Throughout we assume a latent linear regression model

$$y_n^* = X_n\theta + \upsilon_n, \tag{1}$$

where y_n^* is the latent, generally not observable, $(T \times 1)$ response variable and υ_n is a not observable $(T \times 1)$ random error distributed independently of X_n as $\mathrm{N}(0, \Sigma)$, where Σ is any positive definite covariance matrix. θ is an unknown $(P \times 1)$ parameter vector which for simplicity is assumed to be identical for all t, $t = 1, \ldots, T$.

The observable $(T \times 1)$ vector of responses y_n is a known transformation of the latent y_n^*:

$$y_n = \tau(y_n^*) = (\tau_1(y_{n1}^*), \ldots, \tau_T(y_{nT}^*))', \tag{2}$$

where the transformations $\tau_t(\cdot)$ are functions from $\mathcal{R}^1$ onto $\mathcal{R}^1$. Observations from different blocks are assumed to be independent.

For example, in multivariate binary probit models, $\tau_t(y_{nt}^*) = 1(y_{nt}^* > c)\ \forall t$, where $1(\cdot)$ is the indicator function and is one if the argument is true and zero otherwise. Note that the formulation (1) and (2) also includes the representation of models with continuous, ordinal or multinomial responses as well as models with mixed continuous and discrete responses.

For the arguments that follow, define $\Sigma = A + b^2\,\mathrm{I}_T$ (e.g. Stern, 1992), where A is positive (semi-) definite, I_T is the $(T \times T)$ identity matrix, and L is a $(T \times T)$ matrix which satisfies $A = LL'$. Note, that there always exists a $b^2 > 0$ and a matrix A positive (semi-) definite for $\Sigma = A + b^2\,\mathrm{I}_T$ to hold (see Appendix). The error υ_n may then be written as $\upsilon_n = L\nu_n + b\,\varepsilon_n$, where ν_n and ε_n are both unobservable $(T \times 1)$ random variables, $\nu_n = (\nu_{n_1}, \ldots, \nu_{n_T})'$ and $\varepsilon_n = (\varepsilon_{n_1}, \ldots, \varepsilon_{n_T})'$.

Let $\nu_n \sim \mathrm{N}(0, \mathrm{I}_T)$, $\varepsilon_n \sim \mathrm{N}(0, \mathrm{I}_T)$ and $\mathrm{E}(\nu_n, \varepsilon_n) = 0$, then $L\nu_n + b\,\varepsilon_n$ has covariance matrix $\mathrm{Cov}(L\nu_n + b\,\varepsilon_n) = \mathrm{Cov}(\upsilon_n) = \Sigma$ and is normally distributed with expectation $\mathrm{E}(L\nu_{n1} + b\,\varepsilon_n) = 0$.

Therefore (1) may also be written as

$$y_n^* = X_n\theta + L\nu_n + b\,\varepsilon_n \tag{3}$$

and the complete general model, i.e. (2) and (3), may be written as

$$y_n = \tau(h(X_n\theta + L\nu_n, b\,\varepsilon_n)), \quad \nu_n \sim \mathrm{N}(0, \mathrm{I}_T), \quad \varepsilon_n \sim \mathrm{N}(0, \mathrm{I}_T), \tag{4}$$

where τ is again a function which relates the elements of y_n to the elements of y_n^*. Note that although $\mathrm{Cov}(\varepsilon_n) = \mathrm{I}_T$, this does not restrict model (4) to the homoscedastic case: heteroscedasticity may be absorbed into matrix A.

3 Consistent regression parameter estimation

In this section it will be shown that maximum likelihood estimation of a probit model yields estimators which are consistent for the identifiable regression parameters regardless of the true correlation structure of the error terms and without imposing any restriction on the distribution of the covariates.

The log likelihood function for the maximum likelihood estimation of the general model (4) is

$$\begin{aligned} l_N &= \sum_n \ln \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(X_n\theta + L\nu_n, y_n)\, \varphi(L\nu_n)\, |L|\, d\nu_n \\ &= \sum_n \ln \mathrm{E}_{L\nu_n}(g(X_n\theta + L\nu_n, y_n)), \end{aligned} \tag{5}$$

where $g(\cdot,\cdot)$ is the joint probability density function (pdf) or joint probability of y_n given $X_n\theta$ and $L\nu_n$, $\varphi(\cdot)$ denotes the density function of the normal distribution and $A = LL'$ is positive definite. Note that generally L is not unique.

Let $A_0 = L_0 L_0'$ be an unknown 'true' positive definite matrix satisfying $\Sigma_0 = A_0 + b_0^2\,\mathrm{I}_T$, where Σ_0 is the 'true' covariance matrix and b_0 is the 'true' scalar parameter value. The structure of the 'true' matrix L_0 may not coincide with the assumed structure of L used for estimation. Let γ_0 be the 'true' vector which includes θ_0 and possibly one or more parameters characterizing the distribution of $L_0\nu_n$ and $b_0\,\varepsilon_n$. Note that often γ_0 is identifiable only up to a multiplicative scalar.

Now the assumption $\nu_n \sim \mathrm{N}(0, \mathrm{I}_T)$ implies $\mathrm{E}(L\nu_n | L_0\nu_n) = ML_0\nu_n$, where M is a $(T \times T)$ matrix, and since L_0 is of rank T, $\mathrm{Cov}(L\nu_n | L_0\nu_n) = 0$, i.e. $L\nu_n = ML_0\nu_n$ with probability 1. Therefore, Jensens inequality becomes an equality and we have, where conditioning in taking the expectation will be denoted by a

superscript over the expectation sign E,

$$\begin{aligned}
&\mathrm{E}_{L\nu_n}(g(X_n\theta + L\nu_n, \tau(h(X_n\theta + L_0\nu_n, b\,\varepsilon_n))))\\
&\quad = \mathrm{E}_{L_0\nu_n}\mathrm{E}^{L_0\nu_n}_{L\nu_n}(g(X_n\theta + L\nu_n, \tau(h(X_n\theta + L_0\nu_n, b\,\varepsilon_n))))\\
&\quad = \mathrm{E}_{L_0\nu_n}(g(X_n\theta + ML_0\nu_n, y_n))
\end{aligned}$$

and

$$\begin{aligned}
l_N &= \sum_n \ln \mathrm{E}_{L\nu_n}(g(X_n\theta + L\nu_n, \tau(h(X_n\theta + L_0\nu_n, b\,\varepsilon_n))))\\
&= \sum_n \ln \mathrm{E}_{L_0\nu_n}(g(X_n\theta + ML_0\nu_n, y_n)).
\end{aligned}$$

The same arguments hold if A is positive semi-definite. In this case, the expectation $\mathrm{E}_{L\nu_n}(\cdot)$ has to be taken with respect to a singular density (see Appendix).

Let $l_N = \sum_{n=1}^{N} l_n$ be the log likelihood function (5), then under standard regularity conditions (e.g., Amemiya, 1973), and the assumption that $\mathrm{E}(l_n)$ has a unique maximizer in Θ, $N^{-1}l_N \xrightarrow{a.s.} \mathrm{E}(l_n)$ as $N \to \infty$, where convergence is uniform, and $\hat{\gamma} \xrightarrow{a.s.} \gamma^*$ as $N \to \infty$, where γ^* is the unique solution to $\max_{\gamma\in\Theta} \mathrm{E}(l_n(\gamma))$ ($l_n(\gamma) \equiv l_n$). Maximizing (5) with respect to γ then leads to estimators $\hat{\theta}$ which are consistent for θ_0. Using again standard assumptions, asymptotic normality and the asymptotic covariance matrix of $\hat{\gamma}$ can be derived, where the asymptotic covariance matrix may be estimated using the robust covariance matrix estimator (e.g. Li and Duan, 1989).

Note that no assumption was necessary concerning the kind of covariates — beside the standard assumptions needed to derive the asymptotic properties of a ML estimator.

A special binary model for which the rank of L is T follows from the assumption $LL' = l^2\,\mathrm{I}_T$, which leads to an independent model for estimation. In this case $L = l\,\mathrm{I}_T$ and the results from above apply. The consistency of the identifiable regression parameter estimator regardless of the 'true' association structure using this model is a well known result, already shown by Gourieroux, Montfort and Trognon (1984a) or Robinson (1982).

As another interesting special case, consider the random effects binary probit model, leading to a covariance matrix Σ having identical off-diagonal elements. To see that this model is a special case of the general model (4), let $L = \Gamma\Lambda^{1/2}$, where Λ is the diagonal matrix of eigenvalues of the positive semidefinite matrix $A = 1_T 1_T' \sigma_\pi^2$, 1_T be a $(T \times 1)$ vector with all elements equal to unity, and Γ be an orthogonal matrix whose columns are the corresponding standardized eigenvectors. Since for $A = 1_T 1_T' \sigma_\pi^2$ there is only one eigenvalue greater than zero, which is $\lambda = T\sigma_\pi^2$, and the corresponding standardized eigenvector is $T^{-1/2}1_T$, we have $L = \sigma_\pi(1_T\, 0_{T\times(T-1)})$, where $0_{T\times(T-1)}$ is a $(T \times (T-1))$ matrix whose elements are all equal to zero, and $L\nu_n = \sigma_\pi 1_T \nu_{n_1}$. Defining $\tilde{\pi}_n = \nu_{n_1}$ we may write

$$y_n = \tau(h(X_n\theta + \sigma_\pi 1_T \tilde{\pi}_n, b\,\varepsilon_n)), \quad \tilde{\pi}_n \sim \mathrm{N}(0,1), \quad \varepsilon_n \sim \mathrm{N}(0, \mathrm{I}_T). \tag{6}$$

The expectation $\mathrm{E}_{L\nu_n}(g(X_n\theta + L\nu_n, y_n))$ then has to be taken with respect to a singular density (see Appendix) and the log likelihood function (5) reduces to the log likelihood function for the estimation of the random effects model (6),

$$\begin{aligned} l_N &= \sum_n \ln \int_{-\infty}^{\infty} g(X_n\theta + \sigma_\pi 1_T \tilde{\pi}_n, \tau(h(X_n\theta + L_0\nu_n, \varepsilon_n)))\, \varphi(\tilde{\pi}_n) d\tilde{\pi} \\ &= \sum_n \ln \mathrm{E}_{\tilde{\pi}_n}(g(X_n\theta + \sigma_\pi 1_T \tilde{\pi}_n, \tau(h(X_n\theta + L_0\nu_n, \varepsilon_n)))) \\ &= \sum_n \ln \mathrm{E}_{L_0\nu_n}(g(X_n\theta + ML_0\nu_n, y_n)), \end{aligned}$$

where $\sigma = \sigma_\pi b^{-1}$ is identifiable up to sign. Again, the estimator $\hat{\theta}$ is consistent for the identifiable regression parameter, regardless of the 'true' covariance matrix Σ_0. Note that the above representation is not unique. For example using $L = \Gamma\Lambda^{1/2}\Gamma'$, the symmetric square root of $A = \sigma_\pi^2 1_T 1_T'$, would lead to an equivalent random effects model.

Since the regression parameter estimators of both, the independent binary probit model and the random effects binary probit model are consistent, it seems to be worthwhile to compare these estimators, henceforth called pseudo maximum likelihood (PML) estimators (cf. Gourieroux, Montfort and Trognon, 1984a), with each other and with an alternative non-ML estimator with respect to efficiency. Therefore, in the next section the estimation models and the approaches used are described. The question of relative efficiency of the different estimators in finite samples will be addressed in section 6 via a Monte Carlo experiment.

4 Estimation of the binary probit model

4.1 Maximum Likelihood

The log likelihood function of the independent binary probit model is

$$l_N = \sum_{n=1}^{N} \sum_{t=1}^{T} \ln \Phi(\psi_{nt}),$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function, $\psi_{nt} = (2y_{nt} - 1)x_{nt}'\theta$, $\theta = \sigma_v^{-1}\theta^\star$, and $\theta^\star$ is a regression parameter vector which is not identifiable. Only the parameter $\gamma = \theta$ is identified in this model. The asymptotic covariance matrix of $\hat{\gamma}$ is estimated using the robust estimator $\widehat{\mathrm{Var}}(\hat{\gamma}) = \hat{\mathrm{A}}^{-1}\hat{\mathrm{B}}\hat{\mathrm{A}}^{-1}$, where $\hat{\mathrm{A}}$ ist the matrix of second derivatives evaluated at $\hat{\gamma}$ and $\hat{\mathrm{B}}$ is the matrix of the sum over $n = 1, \ldots, N$ outer products of first derivatives of individual log likelihoods evaluated at $\hat{\gamma}$. The estimator $\hat{\gamma}$ will be denoted as PML_I estimator.

The log likelihood function of the random effects binary probit model is

$$l_N = \sum_{n=1}^{N} \ln \int_{-\infty}^{\infty} \prod_{t=1}^{T} \Phi(\psi_{nt})\, \varphi(\tilde{\pi}_n)\, d\tilde{\pi}_n,$$

where $\psi_{nt} = (2y_{nt} - 1)(x'_{nt}\theta_A + \tilde{\pi}_n\sigma_A)$, $\theta_A = \sigma_\varepsilon^{-1}\theta^\star$, $\sigma_A = \sigma_\varepsilon^{-1}\sigma_\pi$ and $\theta^\star$ and σ_π are not identifiable parameters. Only the parameter $\gamma_A = (\theta'_A\ \sigma_A)'$ is identified.

This log-likelihood function and their derivatives can approximately be calculated using Gauss-Hermite quadrature (Bock and Lieberman, 1970; Butler and Moffit, 1982). Let $\tilde{\pi}_n = \sqrt{2}\,m$ and therefore $d\tilde{\pi}_n = \sqrt{2}\,dm$, we have

$$l_N(\gamma_A) = \ln L_N(\gamma_A) \approx -\frac{N}{2}\ln\pi + \sum_{n=1}^{N}\ln\sum_{k=1}^{K}\exp\left\{\sum_{t=1}^{T}\ln\Phi(\psi_{nt}(m_k))\right\}w_k,$$

where K is the number of evaluation points m_k $(k = 1, \ldots, K)$, w_k is the weight given to the kth evaluation point and $\psi_{nt}(m_k) = (2y_{nt} - 1)(x'_{nt}\theta_A + \sqrt{2}\,m_k\sigma_A)$ is evaluated at the kth point. Evaluation points and corresponding weights can be found in Stroud and Secrest (1966).

For the estimates of the different models to be comparable, we compute $\hat{\gamma} = (\theta'\ \sigma)' = (1 + \hat{\sigma}_A^2)^{-1/2}\hat{\gamma}_A$ for the estimates of the random effects binary probit model. The estimated covariance matrix calculated using again the robust covariance matrix estimator has to be transformed correspondingly. Provided that enough evaluation points are used (see e.g. Butler, 1985) this estimator, henceforth denoted as PML_E estimator, is consistent and asymptotically normal. To avoid the bias in the parameter and variance estimates caused by the use of too few evaluation points (see e.g. Guilkey and Murphy, 1993, Spiess and Hamerle, 1995), we calculated the estimates over all s replications at a time, successively increasing the number of evalution points by one until the results remained stable.

To compute the PML estimate in both models, the Newton-Raphson method was used. Furthermore, a line search method for global convergence (Dennis and Schnabel, 1983) was added for the estimation of the the random effects binary probit model.

4.2 GEE approach

The generalized estimation equations (Liang and Zeger, 1986) for the estimation of the regression parameter $\gamma = \theta = \sigma_v^{-1}\theta^\star$ using the binary probit model considered above are

$$\sum_{n=1}^{N} X'_n \mathrm{D}_n \Omega_n^{-1}(y_n - \Phi(X_n\theta)) = 0$$

where $\mathrm{D}_n = \mathrm{diag}(\varphi(x'_{n1}\theta), \ldots, \varphi(x'_{nT}\theta))$ is a diagonal matrix and $\Phi(X_n\theta) = (\Phi(x'_{n1}\theta), \ldots, \Phi(x'_{nT}\theta))'$ is a $(T \times 1)$ vector. Furthermore, $\Omega_n = \mathrm{A}_n^{1/2}\mathrm{R}(\alpha)\mathrm{A}_n^{1/2}$ and $\mathrm{A}_n = \mathrm{diag}(\mathrm{Var}(y_{n1}), \ldots, \mathrm{Var}(y_{nT}))$ where $\mathrm{Var}(y_{nt}) = \Phi(x'_{nt}\theta)(1 - \Phi(x'_{nt}\theta))$. $\mathrm{R}(\alpha)$ is a 'working correlation matrix' whose structure reflects the assumed correlation structure in the observable response variables and α is a vector that fully characterizes this structure. If $\mathrm{R}(\alpha)$ is the 'true' correlation matrix and $\alpha = \alpha_0$,

the 'true' value, then Ω_n will be equal to the 'true' correlation matrix of the observable response variables.

The estimator $\hat{\gamma}$, denoted as GEE estimator, is consistent and asymptotically normally distributed (Liang and Zeger, 1986). The asymptotic covariance matrix can consistently be estimated by $N^{-1}\widehat{\mathrm{G}}^{-1}\widehat{\mathrm{W}}\widehat{\mathrm{G}}^{-1}$, where

$$\widehat{\mathrm{G}} = -N^{-1}\sum_{n=1}^{N}\left(X_n'\mathrm{D}_n\Omega_n^{-1}\mathrm{D}_nX_n\right)_{\theta=\hat{\theta},\,\alpha=\hat{\alpha}},$$

$$\widehat{\mathrm{W}} = N^{-1}\sum_{n=1}^{N}\left(X_n'\mathrm{D}_n\Omega_n^{-1}\widehat{\mathrm{COV}}(y_n)\Omega_n^{-1}\mathrm{D}_nX_n\right)_{\theta=\hat{\theta},\,\alpha=\hat{\alpha}}$$

and

$$\widehat{\mathrm{COV}}(y_n) = (y_n - \Phi(X_n\hat{\theta}))(y_n - \Phi(X_n\hat{\theta}))'.$$

The above properties do not depend on the assumed correlation structure, that is, they hold — beside some regularity conditions — as long as $\hat{\alpha}$ is consistent (Liang and Zeger, 1986).

Following Liang and Zeger (1986) $\hat{\gamma} = \hat{\theta}$ is iteratively computed switching between a modified Fisher scoring for γ_0 and the estimation of α_0. Given current estimates $\hat{\alpha}_j$ and $\hat{\gamma}_j$ $(j = 1, 2, \ldots)$, $\hat{\gamma}_{j+1}$ is calculated by

$$\hat{\gamma}_{j+1} = \hat{\gamma}_j + (\sum_{n=1}^{N} X_n'\mathrm{D}_n\Omega_n^{-1}\mathrm{D}_nX_n)^{-1}\sum_{n=1}^{N} X_n'\mathrm{D}_n\Omega_n^{-1}(y_n - \Phi(X_n\theta)).$$

Unlike e.g. Liang and Zeger (1986) we estimate α_0 starting with the Pearson correlation matrix of the residuals $(y - \Phi(X\hat{\theta}))$, denoted as $\widehat{\mathrm{R}}$, which is — under the usual assumptions — guaranteed to be positive definite.

The off-diagonal elements of the matrix $\widehat{\mathrm{R}}$ are then Z-transformed (Fisher, 1963) to get unbiased estimates $\hat{\alpha}$. If all off-digonal elements are restricted to the same value (i.e. α is (1×1), $0 < |\alpha| < 1$ and $\alpha \neq -1/(T-1)$), the resulting correlation structure is an equicorrelation structure in the observable response variables. In this case $\hat{\alpha}$ is calculated as $\hat{\alpha} = (\exp(2\bar{z}) - 1)/(\exp(2\bar{z}) + 1)$ where $\bar{z}$ is the arithmetic mean of the Z-transformed off-diagonal elements of the matrix $\widehat{\mathrm{R}}$. The corresponding GEE estimator will be denoted GEE_E estimator.

Another specification leads to an estimator which will be denoted GEE_{AE} estimator. For the calculation of this estimator we estimate $\hat{\alpha} = (\hat{\sigma}, \hat{\varrho})'$, the parameter of a mixed AR(1)- and equicorrelation structure in the observable response variables, where the correlation between the observations at t and t' is $r_{t,t'} = \sigma^2 + (1 - \sigma^2)\varrho^{|t-t'|}$, $|\varrho| < 1$ and $\sigma^2 < 1$. In this case the estimator $\hat{\alpha}$ is iteratively calculated using the Newton-Raphson method (for details see Spiess, 1995).

5 Simulation Study: Description

The main program and the modules for simulation and estimation were written in SAS/IML (SAS Institute Inc., 1989). Random numbers were generated using the random number generators RANNOR, RANUNI and RANGAM provided by the SAS system (SAS Institute Inc., 1990).

In generating the datasets the following factors were varied: (1) the sample size, i.e. the number of blocks ($N = 50$, $N = 100$, $N = 500$ and $N = 1000$), (2) the number of observations within each block ($T = 5$ and $T = 10$), (3) the structure of the correlation matrix of the error terms in the latent model, i.e. (i) AR(1) with correlations $\rho_{t,t'} = \varrho^{|t-t'|}$ ($|\varrho| < 1$) and $\varrho = .4$ and $\varrho = .8$, (ii) a 'free' correlation structure with correlations varying between .95 and .40 and (iii) a Toeplitz correlation matrix with correlations ranging from .8 to .2, (4) the kind of covariates (discrete, i.e. dichotomous and trichotomous variates, and continuous variates from an uniform as well as two skewed distributions) and (5) the values of the parameter vector θ. Other factors such as the starting value of the random number generators or number of replications were also variied, but had only marginal effects on the results.

The values of the error term υ_n were independently drawn from the standard normal distribution using RANNOR. The different correlation structures were simulated multiplying these errors by the Cholesky root of the corresponding correlation matrix. As 'free' correlation matrix we used

$$\mathrm{Corr}(\upsilon_n) = \begin{pmatrix} 1 & .95 & .80 & .64 & .40 \\ .95 & 1 & .90 & .73 & .56 \\ .80 & .90 & 1 & .93 & .84 \\ .64 & .73 & .93 & .1 & .89 \\ .40 & .56 & .84 & .89 & 1 \end{pmatrix}$$

and as Toeplitz correlation matrix a matrix with entries $\rho_{t,t-1} = .8$, $\rho_{t,t-2} = .6$, $\rho_{t,t-3} = .4$ and $\rho_{t,t-4} = .2$.

The dichotomous, trichotomous and the uniformly distributed variates were generated using the uniformly distributed random number generator RANUNI. The uniformly distributed covariate, weighted by the parameter θ_u, has expected value zero and will be denoted as x_u. The dichotomous covariates were generated to have value 1 with probability .3 (.4 or .5) and will be denoted as x_{d_1}, x_{d_2} and x_{d_3}, respectively, with corresponding parameters θ_{d_1}, θ_{d_2} and θ_{d_3}. The trichotomous variate (x_t) was generated to have value -1 with probability .2 and 0 with probability .4. The parameter which weights this variate will be denoted as θ_t. As one of the continuous variates from a skewed distribution we generated an exponentially distributed variate (x_e), weighted by the parameter θ_e, using the RANGAM function with shape parameter $\alpha = 1$ and scale parameter $\beta = .5$. We also generated a variate from a gamma distribution with shape parameter $\alpha = .5$ and scale parameter $\beta = 1$. This variate was then standardized (x_s).

The corresponding regression parameter will be denoted as θ_s. The parameter weighting the constant term will be denoted as θ_c.

As starting values for the calculation of the GEE estimates the 'true' parameter values of θ were used. Calculating the GEE_{AE} estimates, a starting value for α is needed within each iteration step for the estimation of θ. This value was calculated using arithmetic means of the Z-transformed off-diagonal elements of the matrix $\widehat{\text{R}}$. Since the log likelihood function of the random effects probit model is not globally concave (see Spiess and Hamerle, 1995), the ML estimates of the independent probit model were used as starting values for the regression parameter θ. As starting value for σ we used values which led to converging sequences of estimates $\{\hat{\gamma}_j\}$, where $j = 1, 2, \ldots$ denotes the jth iteration. However, starting with different values we found for some randomly choosen datasets all sequences $\{\hat{\gamma}_j\}$ to converge — if they converged — to the same values using the ML as well as the GEE approach.

The iterations stopped in all cases if all elements of the vector of first derivatives or estimation equations and all elements of the vector of increments of the last iteration were smaller in absolute value than 10^{-6}.

6 Simulation Study: Results

To compare the estimation results, the mean (m) and the standard deviation (sd) of the estimates over s replications or data generations, the root mean squared error, defined as $\text{rmse} = \left(s^{-1} \sum_{j=1}^{s} (\hat{\gamma}_{lj} - \gamma_{l0})^2\right)^{1/2}$, where $\hat{\gamma}_{lj}$ is the lth element of $\hat{\gamma}$ in the jth simulated sample, were calculated. For the PML estimators we calculated the 'robust' estimated standard deviation of the estimates, defined as $\widehat{\text{sd}} = \left(s^{-1} \sum_{j=1}^{s} \widehat{\text{Var}}(\hat{\gamma}_{lj})\right)^{1/2}$, where $\widehat{\text{Var}}(\hat{\gamma}_{ln})$ is the diagonal entry of the 'robust' covariance matrix estimate as described e.g. in Li and Duan (1989). For the GEE estimates $\widehat{\text{sd}}$ denotes the corresponding estimated standard deviation calculated as described above for the PML estimates and using the diagonal entries of the estimated covariance matrices as described in Section 4.2. To compare the four different estimators 500 datasets were generated according to every model used to calculate the different estimates. However, the different estimates did not converge in all datasets.

The distributions of the estimates $\hat{\gamma}_l$ were graphically displayed using the procedure CHART and a test of normality (Shapiro-Wilk test) was performed using the procedure UNIVARIATE (SAS Institute Inc., 1990b). For the latter test the 1% level of significance will be used throughout.

The number of evaluation points for the PML_E estimates was defined to be sufficient if the measures m, sd and $\widehat{\text{sd}}$ did not change in the first four non-zero digits. Using this strategy up to 92 evaluation points were needed.

Because of the limited space, not all simulation results can be presented.

However, the results reported in this section are consistent with the findings using different models.

Another point that should be mentioned is the use of the GEE_{AE} estimator when in fact the correlation structure of the latent error term is AR(1). This is because the correlation structure in the observable response variables which is no more an AR(1) structure especially for high values of ϱ is better modelled by the GEE_{AE} estimator than by a GEE estimator modelling an AR(1) structure in the correlation matrix of the observable response variables (Spiess, 1995).

Although the results of Section 3 are independent of the type of covariates included into the model, it is well known that the properties of estimators are affected by the type of covariates used (e.g. Li and Duan, 1989, Mancl and Leroux, 1996, McDonald, 1993, Spiess and Hamerle, 1996). Therefore, in a first step datasets according to six models with only one of the following covariates at a time were generated: x_{d1}, x_{d2}, x_{d3} and x_t. The covariates were weighted by the 'true' slope parameter value $\theta_0 = -1$. Two different sample sizes ($N = 50$ and $N = 1000$), 'true' value $\theta_{0c} = .3$, $T = 5$ and an AR(1) structure with 'true' value $\varrho_0 = .8$ were used. The estimation results for the slope parameter in terms of the mean, root mean squared error, the standard deviation and estimated standard deviation can be found in Table 1.

As can be seen from the results in Table 1, there are no systematic or significant deviations of the means from the 'true' value. As expected, the root mean squared errors are larger in the small sample cases than in the large sample cases. Although the differences between root mean squared errors and standard deviations are larger for $N = 50$ than for $N = 1000$, root mean squared errors are essentially standard deviations for all four estimators. A comparison of rmse and $\widehat{\text{sd}}$ reveals that the root mean squared errors are generally underestimated if $N = 50$ for all models with but one exception, whereas they are slightly overestimated if $N = 1000$.

In terms of smallest values rmse, sd and $\widehat{\text{sd}}$, the most inefficient estimator is the PML_I estimator. On the other hand, the differences between the PML_E, GEE_E and GEE_{AE} estimators are not large. However, for almost all models considered in this first step, the PML_E, and the GEE_{AE} estimator are both slightly more efficient than the GEE_E estimator in terms of smaller rmse, sd and $\widehat{\text{sd}}$. The differences in rmse, sd and $\widehat{\text{sd}}$ between the PML_E estimator and the GEE_{AE} estimator depend upon sample size and the covariate included into the model. For $N = 50$ the GEE_{AE} estimator is slightly more efficient in three of the four models than the PML_E estimator. On the other hand, for $N = 1000$ the relative efficiency of the two estimators clearly depend upon the type of covariate used. Generally, the PML_E estimator is very efficient relative to the GEE_{AE} estimator although using the latter it is possible to model the structure of dependence more properly.

The distributions of the estimates $\hat{\theta}$ deviated only in three cases significantly from the normal distribution, i.e. using the PML_I and the GEE_{AE} estimator

Table 1: *Mean* (m), *root mean squared error* (rmse), *standard deviation* (sd) *and estimated standard deviation* ($\widehat{\text{sd}}$) *of* $\hat{\theta}$ *for four models with covariate* x_{d1}, x_{d2}, x_{d3} *or* x_t, $T = 5$, $\theta_{0c} = .3$, $\theta_0 = -1$ *and an AR(1) correlation structure with* $\varrho_0 = .8$ *over* $s = 500$ *replications*

	Model with covariate			
	x_{d1}	x_{d2}	x_{d3}	x_t
m(rmse)	$N = 50$			
PML_I	−1.009 (.2018)	−1.008 (.1940)	−1.014 (.1852)	−1.017 (.1436)
PML_E	−1.008 (.1814)	−1.007 (.1691)	−1.010 (.1657)	−1.014 (.1293)
GEE_E	−1.006 (.1841)	−1.006 (.1715)	−1.010 (.1677)	−1.014 (.1332)
GEE_{AE}	−1.006 (.1805)	−1.005 (.1683)	−1.007 (.1619)	−1.014 (.1327)
$\text{sd}(\widehat{\text{sd}})$				
PML_I	.2018 (.1987)	.1940 (.1929)	.1849 (.1818)	.1427 (.1450)
PML_E	.1814 (.1711)	.1691 (.1664)	.1656 (.1608)	.1287 (.1335)
GEE_E	.1841 (.1750)	.1715 (.1690)	.1676 (.1622)	.1327 (.1363)
GEE_{AE}	.1805 (.1722)	.1683 (.1662)	.1619 (.1588)	.1320 (.1361)
m(rmse)	$N = 1000$			
PML_I	−1.002 (.0434)	−1.000 (.0413)	−1.001 (.0406)	−1.001 (.0322)
PML_E	−0.999 (.0370)	−0.999 (.0347)	−0.999 (.0341)	−1.000 (.0281)
GEE_E	−1.003 (.0379)	−1.002 (.0355)	−1.002 (.0345)	−1.002 (.0281)
GEE_{AE}	−1.002 (.0375)	−1.001 (.0349)	−1.001 (.0338)	−1.002 (.0279)
$\text{sd}(\widehat{\text{sd}})$				
PML_I	.0434 (.0446)	.0413 (.0421)	.0406 (.0409)	.0322 (.0325)
PML_E	.0370 (.0381)	.0347 (.0359)	.0341 (.0350)	.0281 (.0291)
GEE_E	.0378 (.0395)	.0355 (.0370)	.0345 (.0360)	.0281 (.0300)
GEE_{AE}	.0375 (.0389)	.0349 (.0364)	.0338 (.0354)	.0279 (.0299)

to estimate the model with covariate x_{d1} and $N = 50$ and using the GEE_{AE} estimator to estimate the model with covariate x_t and $N = 50$. In all these cases, the distributions are negatively skewed.

If the PML_E estimator is used, one also gets an estimate of σ_0. For $N = 50$, the mean of the estimates $\hat{\sigma}$ for the different models ranged from $\text{m}(\hat{\sigma}) = .8056$ to $\text{m}(\hat{\sigma}) = .8093$. For $N = 1000$ they ranged from $\text{m}(\hat{\sigma}) = .8097$ to $\text{m}(\hat{\sigma}) = .8126$. Comparing these values with the square root of the mean of the 'true' correlations, it seems as if $\hat{\sigma}$ is an estimator for something like the square root of the average 'true' correlation, which is .8095. The distributions of $\hat{\sigma}$ are all normal for $N = 1000$. For $N = 50$ all the distributions are negatively skewed. However, this is not surprising because of the large variances of the estimates on one hand and the restriction $\sigma^2 < 1$ on the other.

We also considered two more models with the same correlation structure but

Table 2: *Mean* (m), *root mean squared error* (rmse), *standard deviation* (sd) *and estimated standard deviation* ($\widehat{\text{sd}}$) *of* $\hat{\theta}_s$ *for a model with* $T = 5$, $\theta_{0c} = -.3$, $\theta_{0d1} = .5$, $\theta_{0s} = -1$ *and* $\theta_{0t} = 1.5$ *and an AR(1) correlation structure with* $\varrho_0 = .8$

m(rmse)	$N = 50$	$N = 100$	$N = 500$	$N = 1000$
PML_I	-1.042 (.1937)	-1.026 (.1271)	-1.002 (.0547)	-1.003 (.0382)
PML_E	-1.036 (.1773)[a]	-1.024 (.1189)	-1.000 (.0506)	-1.002 (.0356)
GEE_E	-1.039 (.1845)[b]	-1.026 (.1238)	-1.002 (.0521)	-1.003 (.0363)
GEE_{AE}	-1.040 (.1854)[b]	-1.025 (.1232)	-1.002 (.0519)	-1.003 (.0363)
$\text{sd}(\widehat{\text{sd}})$				
PML_I	.1893 (.1701)	.1246 (.1192)	.0547 (.0529)	.0382 (.0382)
PML_E	.1738 (.1599)[a]	.1166 (.1140)	.0506 (.0490)	.0355 (.0350)
GEE_E	.1805 (.1672)[b]	.1212 (.1147)	.0521 (.0504)	.0362 (.0362)
GEE_{AE}	.1812 (.1652)[b]	.1208 (.1139)	.0519 (.0503)	.0363 (.0361)

Note: If not otherwise stated, all results are based upon $s = 500$ replications.
[a]Results are based upon $s = 498$ replications.
[b]Results are based upon $s = 499$ replications.

covariates x_u or x_s. Furthermore, instead of an AR(1) correlation structure a correlation structure according to the Toeplitz correlation matrix described in Section 5 was simulated for all the above models. However, the basic results remained unchanged.

In a second step four different sample sizes ($N = 50$, $N = 100$, $N = 500$ and $N = 1000$) were used to compare the different estimators for a model with covariates x_{d1}, x_s and x_t, 'true' regression parameter values $\theta_{0c} = -.3$, $\theta_{0d1} = .5$, $\theta_{0s} = -1$, $\theta_{0t} = 1.5$, $T = 5$ and an AR(1) correlation structure with $\varrho_0 = .8$. The estimation results for one of the regression parameters (θ_{0s}) in terms of the above described measures can be found in Table 2.

The results reported in Table 2 are generally based upon $s = 500$ replications. However, in the small sample case only the results for the PML_I estimator are based upon $s = 500$ replications. For $N = 50$, the PML_E estimator converged only in $s = 498$, the GEE_E and the GEE_{AE} estimators only in $s = 499$ cases.

Again, the differences between the mean of the estimates over simulations and the 'true' value are small. As above, the PML_I estimator is the most inefficient estimator in terms of rmse, sd and $\widehat{\text{sd}}$. Although the differences in rmse, sd and $\widehat{\text{sd}}$ between the PML_E, the GEE_E and the GEE_{AE} estimator are small, the PML_E estimator tends to be the most efficient estimator followed by the GEE_{AE} and by the GEE_E estimator.

For large sample sizes the differences between rmse and sd are negligible and the differences between rmse and $\widehat{\text{sd}}$ are small. These differences become larger if the sample sizes are reduced. For small sample sizes there is a tendency to

underestimate the rmse.

With but few exceptions, the distributions of the estimates are not significantly different from the normal distribution. Furthermore, there is no systematic pattern in the deviations from normality. If the estimation results for the other regression parameters are consiederd (not shown), the general results remain unchanged.

The distributions of the estimates $\hat{\sigma}$ are significantly different from normality, i.e. negatively skewed, only if $N = 50$, which again ist not surprising because of the large variances and the restriction $\sigma^2 < 1$. The means of the estimates $\hat{\sigma}$ ranged from $\mathrm{m}(\hat{\sigma}) = .8050$ ($N = 1000$) to $\mathrm{m}(\hat{\sigma}) = .8125$ ($N = 50$).

The same model was also estimated using the Toeplitz correlation structure. Again, the basic results remain unchanged. Using this correlation matrix, the square root of the mean of the 'true' correlations is .7746. The means of the estimates $\hat{\sigma}$ for that model ranged from .7664 ($N = 100$) to .7715 ($N = 50$).

From these results it may be concluded that the PML_E, GEE_E and GEE_{AE} estimators are more efficient than the PML_I estimator and that the PML_E estimator is very efficient relative to the GEE_{AE} estimator, which in turn tends to be more efficient than the GEE_E estimator for the regression parameters and in the models considered. Furthermore, in general there are no systematic deviations of the distributions of the estimates from normality. Only the distributions of the estimates $\hat{\sigma}$ tend to be negatively skewed for small samples.

To see whether the results are still the same if a different correlation structure and different covariates are used, a model with $T = 5$, covariates x_{d1}, x_t, x_{d2}, $\theta_0 = (-.3\ .5\ -1\ 1.5)'$ but with a 'free' correlation structure was simulated using $N = 50$ and $N = 1000$ (see Table 3).

The question whether the general results found above are also valid if the 'true' correlation is rather low, a model with $T = 5$, x_{d1}, x_t, x_{d2}, $\theta_0 = (-.3\ .5\ -1\ 1.5)'$, an AR(1) structure with $\varrho_0 = .4$, $N = 50$ and $N = 1000$ was simulated. Since $\hat{\sigma}$ using the PML_E estimator converged for $N = 50$ in 22 datasets to zero, the results in Table 4 for the PML_E estimator are based on $s = 478$ simulated datasets only. However, if the results based upon all $s = 500$ replications are used, the general results concerning the regression parameters are the same. Only the results concerning the estimation of σ_0 change ($\mathrm{m}(\hat{\sigma}) = .4466$ for $s = 500$).

Table 5 contains the results of the estimation of a model with more extreme 'true' values of the regression parameters, i.e. $\theta_0 = (-.3\ .05\ 4)'$, x_{d3}, x_u, $T = 10$, an AR(1) structure with $\varrho_0 = .8$, $N = 50$ and $N = 1000$.

From the results in Table 3 to Table 5 it may be concluded that the PML_I estimator generally is the most inefficient estimator in terms of largest values rmse, sd and $\widehat{\mathrm{sd}}$. The differences between the PML_I estimator on one hand and the PML_E, GEE_E and GEE_{AE} estimators on the other is larger for higher 'true' correlations (Table 3 and Table 5) and smaller for lower 'true' correlations (Table 4).

The differences between the PML_E, GEE_E and GEE_{AE} estimator are not as

Table 3: *Mean* (m), *root mean squared error* (rmse), *standard deviation* (sd) *and estimated standard deviation* ($\widehat{\text{sd}}$) *for a model with* $N = 50$ *and* $N = 1000$, $T = 5$, $\theta_{0c} = -.3$, $\theta_{0d1} = .5$, $\theta_{0t} = -1$ *and* $\theta_{0d2} = 1.5$ *and a 'free' correlation structure (see text)*

	$N = 50$				$N = 1000$			
m rmse sd $\widehat{\text{sd}}$	PML_I	PML_E	GEE_E	GEE_{AE}	PML_I	PML_E	GEE_E	GEE_{AE}
$\hat{\theta}_c$	$-.2950$	$-.2919^a$	$-.2979^b$	$-.3016^c$	$-.2992$	$-.2982$	$-.3001$	$-.3002$
	.1966	.1828	.1929	.1919	.0406	.0384	.0396	.0398
	.1967	.1828	.1929	.1919	.0406	.0384	.0397	.0399
	.1800	.1689	.1785	.1776	.0393	.0368	.0387	.0386
$\hat{\theta}_{d1}$	.5375	5283^a	$.5309^b$	$.5301^c$	.4967	.4968	.4980	.4984
	.2370	.1913	.2004	.2008	.0474	.0387	.0418	.0407
	.2343	.1894	.1982	.1988	.0473	.0386	.0418	.0408
	.2099	.1729	.1881	.1854	.0475	.0394	.0417	.0413
$\hat{\theta}_t$	-1.038	-1.026^a	-1.029^b	-1.027^c	-1.001	$-.9976$	-1.002	-1.001
	.1779	.1546	.1585	.1600	.0372	.0320	.0321	.0322
	.1739	.1525	.1559	.1579	.0372	.0319	.0321	.0322
	.1592	.1455	.1499	.1490	.0364	.0325	.0336	.0335
$\hat{\theta}_{d_2}$	1.550	1.534^a	1.546^b	1.548^c	1.503	1.497	1.504	1.503
	.2543	.2247	.2385	.2375	.0521	.0451	.0460	.0467
	.2496	.2224	.2342	.2378	.0521	.0450	.0458	.0466
	.2375	.2146	.2249	.2235	.0532	.0474	.0488	.0487
$\hat{\sigma}$		$.8805^a$				.8695		
		$--$				$--$		
		.0539				.0117		
		.0493				.0118		

Note: If not otherwise stated, all results are based upon $s = 500$ replications.
[a]Results are based upon $s = 499$ replications.
[b]Results are based upon $s = 497$ replications.
[c]Results are based upon $s = 495$ replications.

large. If the 'free' correlation structure is used, the PML_E estimator is the most efficient estimator in terms of these statistics, followed by the GEE estimators (see Table 3). This result shows that the PML_E estimator may even be more efficient than a GEE estimator if the correlation structure is not modelled correctly by the PML_E as well as the GEE estimator. For lower 'true' correlations (AR(1) with $\varrho_0 = .4$), the differences between the PML_E, GEE_E and the GEE_{AE} estimator are small (see Table 4). The results of the estimation of a model with $T = 10$, x_{d3}, x_u,

Table 4: *Mean* (m), *root mean squared error* (rmse), *standard deviation* (sd) *and estimated standard deviation* ($\widehat{\text{sd}}$) *for a model with* $N = 50$ *and* $N = 1000$, $T = 5$, $\theta_{0c} = -.3$, $\theta_{0d1} = .5$, $\theta_{0t} = -1$ *and* $\theta_{0d2} = 1.5$ *and an AR(1) structure with* $\varrho_0 = .4$

m rmse sd $\widehat{\text{sd}}$	$N = 50$				$N = 1000$			
	PML_I	PML_E	GEE_E	GEE_{AE}	PML_I	PML_E	GEE_E	GEE_{AE}
$\hat{\theta}_c$	.3017	$-.3031^a$	$-.3024$	$-.3018$	$-.2984$	$-.2983$	$-.2984$	$-.2982$
	.1578	.1572	.1576	.1557	.0345	.0341	.0341	.0337
	.1578	.1573	.1578	.1558	.0345	.0341	.0341	.0337
	.1504	.1501	.1500	.1489	.0327	.0324	.0325	.0323
$\hat{\theta}_{d_3}$	.5214	$.5215^a$	.5196	.5198	.4974	.4972	.4975	.4975
	.2217	.2197	.2195	.2191	.0476	.0467	.0467	.0457
	.2209	.2189	.2188	.2185	.0476	.0467	.0466	.0456
	.2097	.2062	.2082	.2060	.0471	.0463	.0464	.0560
$\hat{\theta}_t$	-1.025	-1.019^a	-1.022	-1.023	$-.9997$	$-.9997$	$-.9995$	$-.9994$
	.1593	.1564	.1579	.1572	.0311	.0303	.0303	.0302
	.1575	.1554	.1566	.1557	.0311	.0303	.0304	.0303
	.1477	.1455	.1472	.1462	.0329	.0325	.0326	.0323
$\hat{\theta}_{d_2}$	1.538	1.537^a	1.539	1.539	1.499	1.499	1.499	1.498
	.2368	.2309	.2346	.2312	.0465	.0459	.0461	.0453
	.2339	.2282	.2315	.2281	.0465	.0459	.0461	.0453
	.2134	.2097	.2121	.2108	.0476	.0468	.0470	.0466
$\hat{\sigma}$		$.4671^a$				.4696		
		$--$				$--$		
		.1232				.0288		
		.1463				.0285		

Note: If not otherwise stated, all results are based upon $s = 500$ replications.
[a] Results are based upon $s = 478$ replications.

$\theta_0 = (-.3\ .05\ 4)'$, $N = 50$ and $N = 1000$ show that, again, for an AR(1) structure and high 'true' correlations the differences between the PML_E, the GEE_E and the GEE_{AE} estimator concerning the efficiency are small. As above, however, the PML_E and the GEE_{AE} estimators tend to be more efficient than the GEE_E estimator (see Table 5).

Following the general results found in Table 1 and Table 2, again, there are no systematic and significant deviations of the distributions of the regression parameter estimates from normality. Only the distributions of the estimates $\hat{\sigma}$ are negatively skewed for $N = 50$ in all cases.

If the sample sizes are large, root mean squared errors are essentially standard deviations and the differences between rmse and $\widehat{\text{sd}}$ are rather small. For $N = 50$,

Table 5: *Mean* (m), *root mean squared error* (rmse), *standard deviation* (sd) *and estimated standard deviation* ($\widehat{\text{sd}}$) *for a model with* $N = 50$ *and* $N = 1000$, $T = 10$, $\theta_{0c} = -.3$, $\theta_{0d3} = .05$ *and* $\theta_{0u} = 4$ *and an AR(1) structure with* $\varrho = .8$ *over* $s = 500$ *replications*

m rmse sd $\widehat{\text{sd}}$	$N = 50$				$N = 1000$			
	PML_I	PML_E	GEE_E	GEE_{AE}	PML_I	PML_E	GEE_E	GEE_{AE}
$\hat{\theta}_c$	−.3078	−.3067	−.3066	−.3073	−.3001	−.3005	−.3008	−.3011
	.1481	.1449	.1534	.1510	.0286	.0297	.0303	.0300
	.1481	.1449	.1534	.1510	.0286	.0297	.0303	.0300
	.1415	.1364	.1429	.1404	.0285	.0298	.0308	.0306
$\hat{\theta}_{d3}$	.0446	.0458	.0472	.0486	.0487	.0500	.0500	.0501
	.1426	.1267	.1299	.1236	.0332	.0286	.0292	.0283
	.1427	.1267	.1300	.1236	.0332	.0286	.0292	.0283
	.1436	.1264	.1306	.1252	.0339	.0278	.0285	.0278
$\hat{\theta}_u$	4.080	4.067	4.071	4.061	4.008	3.997	4.001	4.001
	.3893	.3763	.3798	.3745	.0793	.0761	.0771	.0768
	.3831	.3707	.3735	.3698	.0793	.0761	.0771	.0768
	.3615	.3512	.3572	.3544	.0814	.0779	.0786	.0784
$\hat{\sigma}$		.6932				.6997		
		−−				−−		
		.0643				.0140		
		.0628				.0140		

the small sample cases, the root mean squared errors are generally underestimated with but a few exceptions (see Table 5) and the differences between root mean squared errors and standard deviations are larger.

The mean of the 'true' correlations using the 'free' correlation matrix is .8741, using $T = 5$ and an AR(1) structure with $\varrho_0 = .4$ the mean is .4767 and using $T = 10$ and an AR(1) structure with $\varrho_0 = .8$ the mean is .7015. Comparing the means of the estimates $\hat{\sigma}$ with these values, the results again suggest that $\hat{\sigma}$ is an estimator of the averaged 'true' correlations.

7 Discussion

In this paper it is shown that the ML regression parameter estimator using a multivariate probit model remains consistent and normally distributed even if the 'true' correlation structure is misspecified. This result holds regardless of the kind of covariates included in the model. As special but often used estimators in

applications, the ML estimator of the random effects panel probit model as well as the ML estimator of the probit model assuming independent responses are considered. Both estimators are consistent even if the responses are correlated or the correlation structure of the responses is misspecified.

A Monte Carlo experiment designed to compare the two ML regression parameter estimators (denoted as PML_I and PML_E estimator, respectively) and two GEE estimators (GEE_E and GEE_{AE} estimator, respectively) showed the PML_I estimator to be the most inefficient estimator for high 'true' correlations. If the 'true' correlations are low, there is not much difference between the four estimators in terms of root mean squared errors, standard deviations and estimated standard deviations. In applications, therefore, if possible dependencies between the responses can assumed to be low, the estimator of the independent probit model may be used without loss of too much efficieny.

If no reasonable assumption concerning the 'true' correlation structure can be made and the dependencies must assumed to be stronger, then the estimator of the random effects probit model may be used. The PML_E estimator was found to be at least as efficient as the GEE_E estimator in the models used and as very efficient relative to the GEE_{AE} estimator which allowed the modelling of the 'true' correlation structure more properly in most of the models considered.

For small samples, generally, the root mean squared errors were found to be underestimated by the estimated standard deviations. However, with larger sample sizes there was no systematic over- or underestimation of the root mean squared errors.

Although no systematic deviations of the distributions of the regression parameter estimates was found, in small samples the distributions of $\hat{\sigma}$ using the PML_E estimator were negatively skewed. This is not surprising, since the estimates $\hat{\sigma}$ are restricted to be below one and the variance of the estimates is large in small samples. The above restriction, however, was necessary to make the different estimators comparable. In practice the restriction $\sigma^2 < 1$ could be replaced by the restriction $\sigma_\epsilon^2 = 1$ (see Section 4.1) to avoid a skewed distribution of the estimator. A look at the means of the estimates $\hat{\sigma}$ using the PML_E estimator suggest that the estimator $\hat{\sigma}$ is an estimator of something like the averaged 'true' correlations.

As the results show that the PML_E estimator is very efficient relative to the GEE estimators in applications, the PML_E estimator implemented in some statistical software packages may be calculated if no reasonable assumption about the 'true' correlation can be made and the main interest lies on the estimation of regression parameters. Furthermore, a possible superiority of alternative estimators of multivariate probit models concerning their efficiency relative to the PML_I and PML_E estimators should be checked, e.g. in finite samples by running Monte Carlo experiments.

Appendix

Lemma 1 *Let Σ be a positive definite covariance matrix, A be a positive (semi-) definite matrix and I the identity matrix. Then there always exists a $b^2 > 0$, such that*

$$\Sigma = A + b^2\, I. \tag{A.1}$$

Proof: Since A is symmetric, there always exists an orthogonal matrix P such that $P'AP = D$, where D is a diagonal matrix with diagonal elements being the eigenvalues of A. Multiplying (A.1) with P' and P leads to

$$P'\Sigma P = D + b^2\, \mathrm{I}.$$

Since P is a non-singular matrix both matrices, Σ and $D + b^2\, \mathrm{I}$, are similar and have the same set of eigenvalues. The eigenvalues of $D + b^2\, \mathrm{I}$ are $d_{ii} + b^2$, where d_{ii} are the diagonal entries of D. For A positive semi-definite there is at least one $d_{ii} = 0$. Therefore, $b^2 > 0$. If A is positive definite, then $d_{ii} > 0\ \forall i$. However, we may always choose a $b^2 > 0$ and a matrix A such that A has positive eigenvalues.

Lemma 2 *Let $A = LL'$ be a symmetric $(T \times T)$ matrix of rank K $(K < T)$ and $\nu_n \sim N(0, I_T)$. Then the expectation $E_{L\nu_n}(g(X_n\theta + L\nu_n, y_n))$ reduces to an expectation with respect to a K-variate normal density.*

Proof: Define $u_n = L\nu_n$ and write $\Gamma = (B\, C)$, where B is a $(T \times K)$ matrix of orthonormal column vectors belonging to the linear space spanned by A and C is a $(T \times (T-K))$ matrix of rank $T-K$, satisfying the conditions $C'A = 0$ and $C'C = \mathrm{I}_{(T-K)}$. It can then be shown (see Rao, 1973, pp. 527–528), that the distribution of u_n is specified by

$$\frac{(2\pi)^{-K/2}}{(\lambda_1, \cdots, \lambda_K)^{1/2}} \exp\left\{-\frac{1}{2}\, u_n'\, A^{-} u_n\right\}, \tag{A.2}$$

and $C'u_n = 0$ with probability 1, where $\lambda_1, \cdots, \lambda_K$ is the product of the nonzero eigenvalues of A and A^{-} is a generalized inverse of A. Thus, u_n lies on the hyperplane $C'u_n = 0$ and the result follows.

Remark: For a random effects binary probit model let $u_n = L\nu_n$, where $L = \sigma_\pi(1_T\, 0_{T\times(T-1)})$, $B = T^{-1/2}1_T$ and C as defined above. Define $z_n = (z_{n_1}\, z_{n_2}')'$, where $z_{n_1} = B'u_n$ is a scalar and $z_{n_2} = C'u_n$ is a $(T-1)$ column vector. Then $z_{n_1} \sim \mathrm{N}(0, T\sigma_\pi^2)$ with density $(2\pi T\sigma_\pi^2)^{-1/2} \exp\{-.5\, u_n' A^{-} u_n\}$, where $u_n' A^{-} u_n = z_{n_1}^2 (T\sigma_\pi^2)^{-1}$. Now, let $\tilde{\pi}_n = T^{-1/2}\sigma_\pi^{-1} z_{n_1}$, then $f(\tilde{\pi}_n) = (2\pi)^{-1/2} \exp\{-.5\, \tilde{\pi}_n^2\}$ and $\tilde{\pi}_n = T^{-1/2}\sigma_\pi^{-1} B'u_n = \nu_{n_1}$. Therefore,

$$\mathrm{E}_{L\nu_n}(g(X_n\theta + L\nu_n, y_n)) = \mathrm{E}_{\tilde{\pi}_n}(g(X_n\theta + \sigma_\pi 1_T \tilde{\pi}_n, y_n)).$$

References

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica, 41(6)*, 997–1016.

Avery, R.B., Hansen, L.P. & Hotz, V.J. (1983). Multiperiod probit models and orthogonality condition estimation. *International Economic Review, 24(1)*, 21–35.

Bock, R.D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46(4*, 443–459.

Bock, R.D. & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika, 35(2)*, 179–197.

Börsch-Supan, A. & Hajivassiliou, V.A. (1993). Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of Econometrics, 58*, 347–368.

Butler, J.S. (1985). The statistical bias of numerically integrated statistical procedures. *Computer and Mathematics with Applications, 11(6)*, 587–593.

Butler, J.S. & Moffit, R. (1982). Notes and comments: A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica, 50(3)*, 761–764.

Dennis, J.E. Jr. & Schnabel, R.B. (1983). *Numerical methods for unconstrained optimization and nonlinear equations.* Englewood Cliffs, New Jersey: Prentice-Hall.

Fahrmeir, L. (1990). Maximum likelihood estimation in misspecified generalized linear models. *Statistics, 21(4)*, 487–502.

Fitzmaurice, G.M. & Laird, N.M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika, 80(1)*, 141–151.

Gourieroux, C., Monfort, A. & Trognon, A. (1984a). Estimation and test in probit models with serial correlation. In J.P. Florens, M. Mouchart, J.P. Raoult & L. Simar (Eds.), *Alternative approaches to time series analysis, Proceedings of the 3rd Franco-Belgian Meeting of Statisticians, november 1982* (pp. 169–209). Bruxelles: Publications des Facultès universitaires Saint-Louis, Boulevard de Jardin botanique 43.

Gourieroux, C., Monfort, A. & Trognon, A. (1984b). Pseudo maximum likelihood methods: Theory. *Econometrica, 52(3)*, 681–700.

Guilkey D.K. & Murphy, J.L. (1993). Estimation and testing in the random effects probit model. *Journal of Econometrics, 59*, 301–317.

Hajivassiliou, V., McFadden, D. & Ruud, P. (1996). Simulation of multivariate normal rectangle probabilites and their derivatives. Theoretical and computational results. *Journal of Econometrics, 72*, 85–134.

Heckman J.J. (1981). Statistical models for discrete panel data. In C.F. Manski, & D. McFadden, (eds.), *Structural analysis of discrete data with econometric applications* (pp. 114–178). Cambridge: The MIT Press.

Hsiao, C. (1986). *Analysis of panel data.* Cambridge: University Press.

Huber, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In L.M. LeCam & J. Neyman, (eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. I, pp. 221–233). Berkeley: University of California Press.

Lerman, S. & Manski, C. (1981). On the use of simulated frequencies to approximate choice probabilites. In C. Manski & D. McFadden, (eds.), *Structural analysis of discrete data with econometric applications* (pp. 305–319). Cambridge: MIT Press.

Li, K.-C. & Duan, N. (1989). Regression analysis under link violation. *The Annals of Statistics, 17(3)*, 1009–1052.

Liang, K.-Y. & Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73(1)*, 13–22.

Maddala, G.S. (1987). Limited dependent variable models using panel data. *The Journal of Human Resources, 22(3)*, 307–338.

Mancl, L.A. & Leroux, B.G. (1996). Efficiency of regression estimates for clustered data. *Biometrics, 52*, 500–511.

McDonald, B.W. (1993). Estimating logistic regression parameters for bivariate binary data. *Journal of the Royal Statistical Society, Series B, 55(2)*, 391–397.

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

Rao, C.R. (1973). *Linear statistical inference and its applications* (2nd ed.). New York: John Wiley & Sons.

Robinson, P.M. (1982). On the asymptotic properties of estimators of models containing limited dependent variables. *Econometrica, 50(1)*, 27–41.

Ruud, P.A. (1983). Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete choice models. *Econometrica, 51(1)*, 225–228.

Ruud, P.A. (1986). Consistent estimation of limited dependent variable models despite misspecification of distribution. *Journal of Econometrics, 32*, 157–187.

SAS Institute Inc. (1989). *SAS/IML software: Usage and reference, version 6.* Cary, NC: SAS Institute.

SAS Institute Inc. (1990). *SAS language: Reference, version 6.* Cary, NC: SAS Institute.

Spiess, M. & Hamerle, A. (1995). *A comparison of different methods for the estimation of regression models with correlated binary responses* (Regensburger Beiträge zur Statistik und Ökonometrie, Nr. 37). Regensburg: Lehrstuhl für Statistik.

Spiess, M. & Hamerle, A. (1996). On the properties of GEE estimators in the presence of invariant covariates. *Biometrical Journal, 38(8)*, 931–940.

Stern, S. (1992). A method for smoothing simulated moments of discrete probabilities in multinomial probit models. *Econometrica, 60*, 943–952.

Stroud, A.H. & Secrest, D. (1966). *Gaussian quadrature formulas.* Englewood Cliffs N.J.: Prentice-Hall.

White, H. (1981). Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association, 76(374)*, 419–433.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica, 50(1)*, 1–25.