Spiess, Hamerle:

Regression Models with Correlated Binary Response Variables: A Comparison of Different Methods in Finite Samples

Projektpartner

gsf

MAX-PLANCK-GESELLSCHAFT

TUM

Regression Models with Correlated Binary Response Variables: A Comparison of Different Methods in Finite Samples

Martin Spiess and Alfred Hamerle[*]

Abstract

The present paper deals with the comparison of the performance of different estimation methods for regression models with correlated binary responses. Throughout, we consider probit models where an underlying latent continous random variable crosses a threshold. The error variables in the unobservable latent model are assumed to be normally distributed. The estimation procedures considered are (1) marginal maximum likelihood estimation using Gauss-Hermite quadrature, (2) generalized estimation equations (GEE) techniques with an extension to estimate tetrachoric correlations in a second step, and, (3) the MECOSA approach proposed by Schepers, Arminger and Küsters (1991) using hierarchical mean and covariance structure models. We present the results of a simulation study designed to evaluate the small sample properties of the different estimators and to make some comparisons with respect to technical aspects of the estimation procedures and to bias and mean squared error of the estimators. The results show that the calculation of the ML estimator requires the most computing time, followed by the MECOSA estimator. For small and moderate sample sizes the calculation of the MECOSA estimator is problematic because of problems of convergence as well as a tendency of underestimating the variances. In large samples with moderate or high correlations of the errors in the latent model, the MECOSA estimators are not as efficient as ML or GEE estimators. The higher the 'true' value of an equicorrelation structure in the latent model and the larger the sample sizes are, the more is the efficiency gain of the ML estimator compared to the GEE and MECOSA estimators. Using the GEE approach, the ML estimates of tetrachoric correlations calculated in a second step are biased to a smaller extent than using the MECOSA approach.

Key words: maximum likelihood; Gauss-Hermite quadrature, generalized estimation equations; mean and covariance structure analysis; tetrachoric correlations; simulation study

[*]Lehrstuhl für Statistik, Wirtschaftswissenschaftliche Fakultät, Universität Regensburg

1   Introduction

The subject of the present paper is to discuss and compare three different methods for the estimation of regression models applied to data sets with correlated binary response variables. This kind of data sets often arise in applied sciences, for example in studies with $(t = 1, \ldots, T)$ repeated measures on $(n = 1, \ldots, N)$ individuals or with T_n measures on different individuals within the same families or blocks. The selection of a statistical model as well as an estimation method then hinges e.g. upon the number of observations per realization of the vector of covariates or whether the structure of association between the response variables is of scientific interest or not. In general, serious computational difficulties arise in the application of the method of maximum likelihood (ML) to these models because of the lack of a rich class of distributions such as the multivariate Gaussian in the case of continuous response variables. Hence likelihood methods are only available in a few cases.

An example is the random effects probit model (e.g. Bock and Lieberman, 1970; Heckman, 1981). Starting with a latent linear regression model, with the observable response variable taking on the value 1 if (and only if) the latent, not observable response variable crosses a threshold, and 0 otherwise (Pearson, 1900), it is often assumed that the error term of the latent linear model has a components of variance structure which in turn implies an equicorrelation structure in the correlation matrix of the latent errors. Assuming this association structure the computation of the log-likelihood function and their derivatives becomes feasible because only one-fold integrals have to be evaluated. This can approximately be done using Gauss-Hermite quadrature (Anderson and Aitkin, 1985; Bock and Lieberman, 1970; Butler and Moffit, 1982). Provided that enough evaluation points are used the ML estimators and their estimated variances are unbiased (Butler, 1985).

To estimate a broader class of models alternative approaches have been suggested. One approach is the 'generalized estimation equations' approach (Liang and Zeger, 1986; Zeger and Liang, 1986) which leads to consistent and asymptotically normally distributed estimates for the vector of regression parameters given only the correct specification of the first moments of the response variables (GEE1 approach). In addition, consistent estimates for the variances of the regression parameter estimators are available. Although the dependencies between the observable or manifest response variables are taken into account to increase efficiency, the association is treated as a nuisance. In contrast, Zhao and Prentice (1990; see also Prentice (1988) and Liang, Zeger and Qaqish (1992)) define a 'generalized estimation equations' approach for simultaneous inference on regression and association parameters (GEE2 approach). The consistency of these parameters depends upon the correct specification of the first moments of the response variables and the correct modelling of the association structure. Note that the regression parameter estimates using GEE1 are consistent whether or

not the association structure is correctly specified, while this not necessarily holds
for the GEE2 approach. Liang, Zeger and Qaqish (1992) found the regression
parameter estimates using both approaches to be similar efficient if they are esti-
mated under correct specifications. On the other hand, the GEE1 approach may
lead to inefficient estimation of the association parameters.

The results of simulation studies using the GEE1 approach have been pre-
sented by Hamerle and Nagl (1987) or Sharples and Breslow (1992) showing that
in general the relative efficiency of the GEE1 estimators calculated under different
assumptions on the association structure between the observable response vari-
ables depends upon e.g. the true structure and strength of the association, N and
T_n. In both studies the GEE1 estimator calculated under the assumption of an
equicorrelation structure was in general found to be very efficient relative to ML
estimators. Hamerle and Nagl (1987) considered models with one time invariant
and one free varying covariate. In contrast, Sharples and Breslow (1992) used
dichotomous covariates. In the present paper we consider models with a dichoto-
mous, a normal and an uniform distributed covariate, because in most practical
applications the covariates do not belong to the same scale or distribution. The
three covariates used are varying over all NT observations. Different effects of
time invariant, block invariant and free varying covariates on the properties of
the GEE1 estimators are discussed in Spiess and Hamerle (in preparation). Un-
like Hamerle and Nagl (1987) or Sharples and Breslow (1992), who used small
to medium sample sizes ($N = 50$ to $N = 200$) we present estimation results for
small ($N = 100$) to large ($N = 1000$) sample sizes.

Using a threshold probit model not only the regression parameters but also the
pairwise correlations of the error components of the latent regression model may
be of scientific interest. In the present paper an extension of the GEE1 approach
(henceforth called 'GEE' approach) is suggested which — in an additional ML-
estimation step — allows the estimation of these tetrachoric correlations.

Another approach was suggested by Küsters (1987, 1990), Schepers (1991)
and Schepers, Arminger and Küsters (1991). They propose the estimation of
complex hierarchical mean and covariance structure models e.g. for metric, or-
dinal or binary response variables — again using a threshold model — in three
stages ('mean and covariance structure analysis' henceforth called 'MECOSA'
approach). This estimation procedure leads to consistent and asymptotically
normally distributed estimates (Küsters, 1987). Again, consistent estimators for
the asymptotic variances of the parameter estimates are available.

Although some special cases of the general model have been shown to work
using 'real' datasets (e.g. Sobel and Arminger, 1992) the properties of these
'MECOSA'-estimators in finite samples have not yet been investigated.

One simple submodel of the general model is the probit regression model
with dependent binary response variables which can also be estimated using the
GEE approach. In the present article results of a simulation study will be pre-
sented comparing the properties of the estimators in finite samples using both

approaches[1]. For special correlation matrices in the error component of the latent regression model the appropriate ML-estimator is computed and compared to GEE- and MECOSA-estimators. In all cases estimation of the regression parameters is the main interest.

In section 2 the general model (section 2.1) and the different estimation approaches are described, i.e. the ML approach (section 2.2), the GEE approach (section 2.3) and the MECOSA approach (section 2.4). Section 3 gives a description of the simulated models, e.g. the used combination of sample sizes and correlation structures of the latent error terms. The results can be found in section 4. In section 4.1 the technical results, e.g. required computing time or convergence problems, and in section 4.2 the results with regard to bias and efficiency are presented. A discussion of the results and concluding remarks can be found in section 5.

2   Estimation Procedures

2.1   Model

For the models considered we have N blocks ($n = 1, \ldots, N$) and T observations ($t = 1, \ldots, T$) in every block[2]. Let $Y_n = (Y_{n1}, \ldots, Y_{nT})'$ denote the ($T \times 1$) vector of binary responses for the nth block and Y the ($NT \times 1$) vector of binary responses for all NT observations. 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.

Throughout we assume a threshold model (Pearson, 1900) with

$$Y_{nt}^* = X_{nt}' \beta^* + v_{nt},$$

a linear latent regression model where Y_{nt}^* is the latent, i.e. not observable, continuous response variable, v_{nt} is the error term, and β^* is the unknown regression parameter vector ($n = 1, \ldots, N$ and $t = 1, \ldots, T$). For the binary probit model considered in this article let $v_n = (v_{n1}, \ldots, v_{nT})'$, $v_n \sim N(0, \Sigma)$ — observations from different blocks are assumed to be independent — and

$$Y_{nt} = \begin{cases} 1 & \text{if } Y_{nt}^* > 0, \\ 0 & \text{otherwise.} \end{cases}$$

In the sequel let $\Phi(\cdot)$ denote the standard normal cumulative distribution function (cdf) and $\varphi(\cdot)$ denote the standard normal density function (df).

---

[1] A third possibility would be to compute the ML-estimator using the simulation method (e.g. Börsch-Supan and Hajivassiliou, 1993). The comparison of this method with the ML, GEE and MECOSA approach is left for further studies.

[2] The results discussed in this paper may easily be generalized allowing the number of time-series observations, T_n, to vary between blocks. Only for simplicity, we assume $T_1 = \cdots = T_N$.

2.2   Maximum Likelihood

Assuming $v_{nt} = \pi_n + \epsilon_{nt}$, where $\pi_n \sim N(0, \sigma_\pi^2)$, $\epsilon_{nt} \sim N(0, \sigma_\epsilon^2)$ and $E(\pi_n \epsilon_{nt}) = 0$, leads to the random effects probit model with an equicorrelation structure in the errors of the latent model. Because the observations Y_{nt} and $Y_{nt'}$ are conditionally independent, the probability $p(Y | X)$ is

$$p(Y | X) = \prod_{n=1}^{N} p(Y_n | X_n)$$

$$p(Y_n | X_n) = \int_{-\infty}^{\infty} \prod_{t=1}^{T} \Phi(\psi_{nt}) \, \varphi(\tilde{\pi}_n) \, d\tilde{\pi}_n,$$

where $\tilde{\pi}_n = \sigma_\pi^{-1} \pi_n$, $\psi_{nt} = (2y_{nt} - 1)(X'_{nt}\beta_A + \tilde{\pi}_n \sigma_A)$, $\beta_A = \sigma_\epsilon^{-1} \beta^\star$ and $\sigma_A = \sigma_\epsilon^{-1} \sigma_\pi$. In this model only the parameter $\theta_A = (\beta'_A \ \sigma_A)'$ is identified.

The log-likelihood function $l_N(\theta_A) = \ln L_N(\theta_A) = \sum_{n=1}^{N} \ln p(Y_n | X_n, \theta_A)$ 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

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

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 $\psi_{nt}(m_k) = (2y_{nt} - 1)(X'_{nt}\beta_A + \sqrt{2}\,m_k \sigma_A)$ is evaluated at the kth point. Evaluation points and corresponding weights can be found in Stroud and Secrest (1966).

To compute the ML estimate $\hat{\theta}_A$ the Newton-Raphson method together with a line search method for global convergence is used (Dennis and Schnabel, 1983). The diagonal elements of $-\mathrm{H}^{-1}(\hat{\theta}_A)$, where $\mathrm{H}(\theta)$ is the matrix of partial second derivatives, are used as estimates for the variances of $\hat{\theta}_A$.

For the estimates to be comparable across the different approaches we compute $\hat{\theta} = (1 + \hat{\sigma}_A^2)^{-1/2} \hat{\theta}_A$. The estimated variances have to be transformed correspondingly. Provided that enough evaluation points are used (see e.g. Butler, 1985) the ML-estimators are consistent and asymptotically normal. The number of evaluation points for an unbiased estimation of parameters and their variances is affected by several factors (Spiess, 1995). Above all the value of σ plays a significant role: the higher the value of σ the more evaluation points are needed.

2.3   GEE Approach

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

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

where $\mathrm{D}_n = \mathrm{DIAG}(\varphi(X'_{n1}\beta), \ldots, \varphi(X'_{nT}\beta))$ is a diagonal matrix and $\Phi(X_n\beta) = (\Phi(X'_{n1}\beta), \ldots, \Phi(X'_{nT}\beta))'$ 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 $\mathrm{VAR}(Y_{nt}) = \Phi(X'_{nt}\beta)(1 - \Phi(X'_{nt}\beta))$. $\mathrm{R}(\alpha)$ 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.

Given a consistent estimator $\hat{\alpha}$, Liang and Zeger (1986) have shown that the GEE-estimator $\hat{\beta}$ is consistent and asymptotically normal with covariance matrix $N^{-1}\mathrm{G}_0^{-1}\mathrm{W}_0\mathrm{G}_0^{-1}$, where

$$\mathrm{G}_0 = -\lim_{N\to\infty} N^{-1} \sum_{n=1}^{N} \left(X'_n \mathrm{D}_n \Omega_n^{-1} \mathrm{D}_n X_n \right)_{\beta=\beta_0,\ \alpha=\hat{\alpha}}$$

and

$$\mathrm{W}_0 = \lim_{N\to\infty} N^{-1} \sum_{n=1}^{N} \left(X'_n \mathrm{D}_n \Omega_n^{-1} \mathrm{COV}(Y_n) \Omega_n^{-1} \mathrm{D}_n X_n \right)_{\beta=\beta_0,\ \alpha=\alpha_0}.$$

A consistent estimator for this covariance matrix is $N^{-1}\hat{\mathrm{G}}_N^{-1}\widehat{\mathrm{W}}_N\hat{\mathrm{G}}_N^{-1}$, where

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

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

and

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

Note that these properties do not depend on the assumed correlation structure, that is, they hold — beside some regularity conditions — as long as $\hat{\alpha}$ is consistent.

In the case of time or block invariant covariates some special results can be derived (see Spiess and Hamerle, in preparation). Therefore in the present paper we only consider free varying covariates, i.e. covariates that vary freely over all NT observations.

Following Liang and Zeger (1986) $\hat{\beta}$ is iteratively computed switching between a modified Fisher scoring for β_0 and a moment estimation for α_0. Given current estimates $\hat{\alpha}_j$ and $\hat{\beta}_j$ ($j = 1, 2, \ldots$), $\hat{\beta}_{j+1}$ is estimated by

$$\hat{\beta}_{j+1} = \hat{\beta}_j + (X'\mathrm{D}\Omega^{-1}\mathrm{D}X)^{-1}X'\mathrm{D}\Omega^{-1}(Y - \Phi(X\beta)),$$

where $D = DIAG(D_1, \dots, D_N)$ and $\Omega = DIAG(\Omega_1, \dots, \Omega_N)$ are both block diagonal matrices and $\Phi(X\beta) = (\Phi(X_1\beta)', \dots, \Phi(X_N\beta)')'$ is a $(NT \times 1)$ vector consisting of the $(T \times 1)$ vectors $\Phi(X_n\beta)$, $n = 1, \dots, N$.

Unlike Liang and Zeger (1986) or Sharples and Breslow (1992) we estimate α_0 starting with the Pearson correlation matrix of the residuals[3] $(Y - \Phi(X\hat{\beta}))$,

$$\hat{R} = DIAG(\hat{S}_d)^{1/2} \, \hat{S} \, DIAG(\hat{S}_d)^{1/2},$$

where $\hat{S}_d$ is the vector of diagonal entries of the covariance matrix $\hat{S}$ computed as

$$\hat{S} = N^{-1}\hat{U}' \left(I_N - N^{-1}1_N 1_N'\right) \hat{U}$$

where $\hat{U} = (\hat{U}_1, \dots, \hat{U}_T)$ is a $(N \times T)$ matrix with elements $\hat{U}_t = ((Y_{1t} - \Phi(X_{1t}'\hat{\beta})), \dots, (Y_{Nt} - \Phi(X_{Nt}'\hat{\beta})))'$, I_N is the $(N \times N)$-identity matrix and $1_N = (1_1, \dots, 1_N)'$ is a $(N \times 1)$ vector. For $\hat{U}$ having full column rank this correlation matrix is guaranteed to be positive definite.

The off-diagonal elements of the matrix $\hat{R}$ are then Z-transformed (Fisher, 1963) for all but one choice of correlation structure to get unbiased estimates $\hat{\alpha}$, the exception being a 'free' correlation structure where $\hat{\alpha}$ is a vector whose elements are the off-diagonal elememts of $\hat{R}$. The corresponding GEE-estimator will be denoted GEE_F-estimator.

If all off-digonal elements are restricted to the same value (i.e. α is (1×1) and $0 < |\alpha| < 1$), 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 $\hat{R}$. The corresponding GEE-estimator will be denoted GEE_E-estimator. The restriction $\alpha = 0$ implies an GEE-estimator calculated under the assumption of independence.

Two other specifications lead to estimators which will be denoted GEE_{AR1}- and GEE_{ARH}-estimator, respectively. Under both specifications the estimates $\hat{\alpha}$ are calculated iteratively using the Newton-Raphson method (for details see Spiess (1995)). The calculation of the GEE_{AR1}-estimator is based upon the assumption of a stationary stochastic AR(1) process in the observable response variables. In this case the off-diagonal elements of the matrix $\hat{R}$, $r_{tt'}$, were $t \neq t'$, are defined as $r_{tt'} = \rho^{|t-t'|}$ $(|\rho| < 1)$ and $\hat{\alpha} = \hat{\rho}$. For the GEE_{ARH}-estimator we estimate $\hat{\alpha} = (\hat{\sigma}, \hat{\rho})'$, the parameter of a mixed AR(1)- and equicorrelation structure $r_{tt'} = \sigma^2 + (1 - \sigma^2)\rho^{|t-t'|}$, $|\rho| < 1$ and $\sigma^2 < 1$.

Although Prentice (1988) pointed out the restrictions on the values of the correlations of binary variates, it is not clear if and in which way or to which extent the estimators themselves or the calculation of the estimates are influenced negatively in some sense if these restrictions are violated. Only Sharples and Breslow (1992) reported some problems, noting that for cases in which these

---

[3]We also used the standardized residuals but found no advantage over the not standardized residuals with regard to the properties of the GEE estimators.

constraints were violated there were multiple solutions to the generalized estimation equations. However, in cases with covariates varying over all observations it is to be expected that these restrictions are violated at least for some pairs of observations. Therefore in our simulation study we calculated the percentage of violations in each dataset to search for a connection between these violations and possible problems in the calculation of the estimates or defective parameter or variance estimates.

If a latent regression model is assumed one may be interested in the correlations $\rho_{tt'}$ between the errors of that latent model. To estimate these tetrachoric correlations we use the ML-method in a second step after the solution to the generalized estimation equations is found (for a different approach see Qu, Williams, Beck and Medendorp, 1992), that is we maximize the log-likelihood function

$$l_N(\rho_{tt'}) = \sum_{n=1}^{N} \ln p(Y_{nt} = y_{nt}, Y_{nt'} = y_{nt'} | \hat{\eta}_{nt}, \hat{\eta}_{nt'}, \rho_{tt'}),$$

where $\hat{\eta}_{nt} = X'_{nt}\hat{\beta}$ and $\hat{\eta}_{nt'} = X'_{nt'}\hat{\beta}$ and $\hat{\beta}$ is the GEE-estimator from the first step. The probability $p(Y_{nt} = y_{nt}, Y_{nt'} = y_{nt'} | \hat{\eta}_{nt}, \hat{\eta}_{nt'}, \rho_{tt'})$ is a function of Y_{nt}, $Y_{nt'}$, $\Phi(\hat{\eta}_{nt})$, $\Phi(\hat{\eta}_{nt'})$ and

$$p(Y_{nt} = 1, Y_{nt'} = 1 | \hat{\eta}_{nt}, \hat{\eta}_{nt'}, \rho_{tt'}) = \int_{-\infty}^{\hat{\eta}_{nt}} \int_{-\infty}^{\hat{\eta}_{nt'}} \varphi_2(x, y, \rho_{tt'}) \, dx \, dy$$

where $\varphi_2(x, y, \rho_{tt'})$ is the df of a bivariate standard normal distribution evaluated at the points $\hat{\eta}_{nt}$, $\hat{\eta}_{nt'}$ and $\rho_{tt'}$. This two dimensional problem may be reduced to a one dimensional problem (Owen, 1956) and the resulting integral may approximately be calculated using Gauss-Legendre quadrature. We estimated tetrachoric correlations using different numbers of evaluation points, different true values for $\rho_{0tt'}$ and different values for $\hat{\beta}_t$ and $\hat{\beta}_{t'}$ and found sixteen points in all cases to be enough to get stable results[4].

Although following Küsters (1990) this 'second stage' estimator, $\hat{\rho}_{tt'}$, is consistent its asymptotic normality has still to be shown.

2.4   MECOSA Approach

The probit model described in section 2.1 may also be derived from a more general 'mean and covariance structure' model as discussed by Küsters (1987) (see also Sobel and Arminger (1992) for a special application). The latent linear regression model simplifies in the case considered in the present paper to

$$Y_n^* = \Gamma X_n + v_n,$$

---

[4]For the computation of the estimates $\hat{\rho}_{tt'}$ we again used the Newton-Raphson method together with a line search method for global convergence (Dennis and Schnabel, 1983; for the derivatives see Spiess, 1995).

where $\Gamma = \text{DIAG}(\beta_1^{*\prime}, \ldots, \beta_T^{*\prime})$ is a $(T \times TP)$ matrix, $X_n = (X_{n1}^\prime, \ldots, X_{nT}^\prime)^\prime$ is a $(TP \times 1)$ vector, $v_n \sim N(0, \Sigma)$ and the elements of the $(T \times 1)$ vector of latent response variables Y_n^*, are connected to the observable response variables Y_{nt} by means of the threshold relation described in section 2.1.

The estimation of the parameters, henceforth denoted MEC estimators, is carried out in three steps. In the first step the ML estimates of $\beta_{0t} = \sigma_{v_t}^{-1}\beta_{0t}^*$ for T independent probit models are calculated using the Newton Raphson method. In the second step the ML estimates for pairwise tetrachoric correlations are calculated using bivariate marginal models and the estimated values for the regression parameters from step one (see section 2.3). The techniques used to calculate the estimates are the same as described in section 2.3. Still in the second step, an estimator of the asymptotic covariance matrix of all estimators of the first two steps is calculated.

In the third step a weighted least squares estimator for ϑ_0, a vector of fundamental parameters, is — in the most cases — iteratively calculated again using the Newton Raphson method (for the derivatives see Spiess, 1995), where the quadratic function

$$q_N(\vartheta) = (\hat{\delta} - \text{g}(\vartheta))^\prime \, \widehat{W}^{-1} \, (\hat{\delta} - \text{g}(\vartheta)),$$

is minimized for ϑ, where $\hat{\delta}$ is the vector containing all the parameter estimates of the first two steps and $\widehat{W}$ is the estimate of the asymptotic covariance matrix of $\hat{\delta}$. The restrictions imposed on the elements of δ, namely $\beta_t = \beta_{t^\prime} \; \forall t, t^\prime = 1, \ldots, T$ and $t \neq t^\prime$, to make the estimates comparable using the different approaches, are realized through the function $\text{g}(\vartheta)$. In the models considered here $\theta = \vartheta = (\beta^\prime, \vartheta_C^\prime)^\prime$, were β is the $(P \times 1)$ vector of regression parameters and ϑ_C is a scalar or vector depending on the assumed correlation structure, e.g. if an equicorrelation structure is assumed $\vartheta_C = \rho_{tt^\prime} \; \forall t, t^\prime$.

It can be shown (see Küsters, 1987; Shapiro, 1986) that the estimators, $\hat{\theta}_N$, are consistent and asymptotically normal with asymptotic covariance matrix $N^{-1}(G_0 W_0^{-1} G_0^\prime)^{-1}$, where W_0 is the asymptotic covariance matrix of $\hat{\delta}_N$, $G_0 = (\partial \text{g}(\vartheta)/\partial \vartheta)_{\vartheta = \vartheta_0}$ and W_0 can consistently be estimated using $\widehat{W}$.

As mentioned above, the dimension of the vector θ in the models considered here depends upon the dimension of β and the assumed correlation structure of the latent error terms. For the sake of comparability we restrict the matrix Σ to be a correlation matrix. This correlation matrix is assumed to have one of the following structures: an equicorrelation structure (the estimator will be denoted MEC_E), an AR(1) structure (MEC_{AR}), a mixed equicorrelation and AR(1) structure (MEC_{ARH}) and no structure at all, that is, the off-diagonal elements of Σ are allowed to vary freely (MEC_F).

Since the estimators of the parameters determining the correlation structure were usually biased, the estimated tetrachoric correlations in step two were Z-transformed (see Fisher, 1963) to get unbiased estimates in the third step. The

estimated variances and covariances of the Z-transformed correlations were transformed correspondingly.

3 Simulation Study: Description

All three approaches lead to consistent and asymptotically normally distributed estimators. The question that motivated this study then is: Which of the approaches is preferable in which situation not only in terms of bias and relative efficiency of the estimators in finite samples, but also with respect to computing time or robustness of the estimation method. To answer this question we used the three approaches to estimate simulated datasets where the following factors were varied: (1) the sample size ($N = 100$, $N = 500$ and $N = 1000$) and (2) the structure of the correlation matrix of the error terms in the latent model and the values of the corresponding parameter values, i.e. (i) equicorrelation: $\rho_{tt'}$ identical $\forall t, t' = 1, \ldots, T$ ($t \neq t'$) with values 0.2, 0.5 and 0.8, (ii) AR(1): $\rho_{tt'} = \varrho^{|t-t'|}$, $|\varrho| < 1$, whith values $\varrho = 0.2$ and $\varrho = 0.8$ and (iii) mixed equicorrelation and AR(1): $\rho_{tt'} = \sigma^2 + (1 - \sigma^2)\varrho^{|t-t'|}$, $|\varrho| < 1$ and $\sigma^2 \equiv \sigma_\pi^2/\sigma_v^2$, with $\sigma^2 = 0.8$ and $\varrho = 0.2$

The main program and the modules for simulation and estimation were written in SAS/IML, the 'interactice matrix language' included in the SAS system ('statistical analysis system'), version 6 (SAS Institute Inc., 1989). Random numbers were generated using the random number generators RANNOR and RANUNI provided by the SAS system (SAS Institute Inc., 1990).

We generated dichotomous, normal and uniform distributed covariates — the latter two having mean zero — which were held constant over the s simulated samples. The dichotomous variables were generated via the uniform distributed random number generator RANUNI with the value 1 if the generated random number exceeds .5 and 0 otherwise. The corresponding regression coefficients are β_2, β_3, β_4 and β_1 denotes the intercept.

The values of the error term were drawn from the standard normal distribution using RANNOR. For the simulation of an equicorrelation structure we generated $\{\pi_n\} \sim N(0, \sigma_\pi^2)$ and $\{v_{nt}\} \sim N(0, \sigma_v^2)$ independently from each other, where $\sigma_\epsilon^2 = \sigma_\pi^2 + \sigma_v^2 = 1$. The AR(1) structure was simulated multiplying $\epsilon = (\epsilon_{n1}, \ldots, \epsilon_{nT})'$ by the Cholesky root of the corresponding Toeplitz matrix, where $\{\epsilon_{nt}\} \sim N(0, 1)$. The mixed AR(1) and equicorrelation structure was generated mixing both of these techniques, again with $\sigma_\epsilon^2 = \sigma_\pi^2 + \sigma_v^2 = 1$.

If possible the parameter values that were used to simulate the datasets were also used as starting values for the calculation of the estimates. However, in some cases this was either not possible or some other strategy was superior.

As mentioned in section 2.2 for the ML estimator to be unbiased a sufficient number of evaluation points is needed. Because this number is a priori unknown a predetermined number of evaluation points has to be increased successively by

one. If the estimates and the estimated variances are constant within a predetermined range of at least three such trials, a sufficient number of evaluation points is found.

Because the ML estimators may be biased, the parameter values used to simulate the datasets were not always optimal as starting values for the estimation procedure. We therefore used the ML estimates of the independent probit model as starting values for the regression parameters and the 'true' value for $\sigma = \sigma_A/(1 + \sigma_A^2)^{1/2}$. Even if the number of evaluation points were sufficient, this strategy in general led to lesser computing time required.

Calculating the GEE estimates under the assumption of an AR(1) or a mixed AR(1) and equicorrelation structure in the correlation matrix of the observable response variables the corresponding correlation structure parameters have to be calculated iteratively within each iteration step for the GEE estimates. As starting values we calculated arithmetic means of the Z-transformed Pearson correlations of the residuals at the first iteration step in the calculation of the GEE estimates. At any further call of the corresponding module, the estimates calculated within the previous iteration were used as starting values.

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 1×10^{-6}.

4 Simulation Study: Results

4.1 Convergence

The estimation of the simulated datasets showed that although the log likelihood function from section 2.2 is not globally concave, and few random starting values led to diverging sequences of estimates, in all cases the sequence of estimates $\{\hat{\beta}_j\}$, where j denotes the jth iteration ($j = 1, 2, \ldots$), converged — if they converged — to the same solution. Furthermore we compared the iterations and the time needed for convergence using the matrix of second derivatives vs. using the sum of the outer product of first derivatives in the calculation of the estimates. In the examples we considered using the matrix of second derivatives for the calculation of the estimates about only half as much iterations were needed than using the first derivatives only. Accordingly, using the matrix of second derivatives led to considerably less required computing time.

Calculating the GEE estimates we encountered convergence problems only with two out of thousands of datasets, the estimation results of most of them being not reported here because of the limited space. In both cases — GEE$_E$ and GEE$_{AR1}$ estimators, respectively, for two different datasets both with a 'true' equicorrelation structure of the latent error terms with $\rho_{tt'} = .8$ — no solution was found despite trying different starting values and the implementation of a

global strategy (see Dennis and Schnabel, 1983). On the other hand we found
violations of the constraints on the estimated correlations of the observable re-
sponse variables in most of the cases (see section 2.3). When we calculated the
portions of violations over s samples in each situation we found the highest por-
tion of violations of the upper bound to be .68 and of the lower bound to be
.09.

We also calculated the portions of response variables having value one at any
point t over s simulations at a time — the portions varied depending on the
covariates used between .57 and .44. In both cases there were no convergence
problems[5].

Defective parameter or covariance matrix estimates (e.g. not positive definite)
did in no case emerge. Estimating the tetrachoric correlations as described in
section 2.3 was problematic in cases with small N and high true values. In these
cases the estimator often converged to the boundary point unity.

This was also true for the MECOSA estimators. Whereas the GEE regres-
sion parameter estimates were calculated before the tetrachoric correlations, the
calculation of the MECOSA estimates depend heavily on the estimation of the
tetrachoric correlations. With small sample sizes ($N = 100$) and moderate to
high true correlations of the latent error terms not only the estimates of the
tetrachoric correlations often converged to boundary points but also singular ma-
trices $\widehat{W}$ occured. This is not surprising since for example with $T = 5$ and four
regression parameters to be estimated $\widehat{W}$ is a (30×30) matrix.

We also found considerable problems in the calculation of the MEC_{ARH} es-
timator with low ($N = 100$) to moderate sample sizes ($N = 500$). For example
with $N = 500$, $T = 5$, $\beta_0 = (-.3, .8, .8, -.8)'$ and $\sigma_0^2 = .2$ and $\rho_0 = .8$ the esti-
mate for σ_0 converged in 79 out of $s = 200$ simulations to zero and the estimated
variance to infinity. In those cases — for sample sizes $N = 500$ but also, although
to a smaller extent, for sample sizes $N = 1000$ — in which all the elements of the
MEC_{ARH} estimate converged to values inside the paramater space the estimators
$\hat{\sigma}$ and $\hat{\varrho}$ turned out to be highly correlated in this nonlinear model.

The computing time required for the calculation of the ML estimates depends
on different factors. As mentioned in section 2.2 a sufficient number of evalua-
tion points is needed to get unbiased estimates. The number of evaluation points
depends above all upon the value of the true intraclass correlation. The higher
this value the more evaluation points are needed and the more computing time is
required. By comparing the mean and the estimated standard deviations of the
ML estimates over s simulations (see section 4.2) with different numbers of eval-
uation points we ensured the estimation results to be unbiased. In problematic
cases, e.g. for a model with $N = 100$ blocks and a high value of $\rho_{0tt'}$, up to 64
evaluation points were needed.

---

[5] This did also hold using block invariant covariates with portions of response variables having
value one varying between .69 and .14.

The time required to calculate the GEE estimates depends upon the number of observations or the number of parameters to be estimated, i.e. factors that influence the time required in the calculation of the ML and the MECOSA estimates as well. In fact, we found the calculation of the different GEE estimates to be very similar in required computing time and independent of factors such as the true values of the correlations.

A main factor that influences the time required to calculate the MECOSA estimates is the number of observations T in every block n. The value of T determines the number of independent probit models to be estimated in the first step and the number of correlations to be estimated in the second step.

The calculation of the ML estimates as well as the calculation of the MECOSA estimates required definitely more computing time in all our simulations than the calculation of the GEE estimates. In the simulated models used in this paper the calculation of the ML estimates in general required more computing time than the calculation of the MECOSA estimates, but these differences depend heavily on the true correlations and on the number of observations within each block.

As an example with $N = 500$ blocks, $T = 5$ observations within each block, four regression parameters to estimate and a true correlation of the latent errors of $\rho = .5$ over $s = 200$ simulations the ML estimation required about 348 minutes with 21 evaluation points needed for unbiased estimation, the GEE_E estimation required about 17 minutes and the MEC_E estimation required about 149 minutes of computing time. Although these values — beside the above mentioned factors — are also subject to programming techniques, they nevertheless give rough hints on the differences in required computing time.

4.2   Bias and efficiency

To compare the results of the estimation of the simulated datasets using the three approaches the mean of the estimates (m), the root mean squared error, defined as $\text{rmse} = \left(s^{-1} \sum_{r=1}^{s} (\hat{\theta}_r - \theta_0)^2 \right)^{1/2}$, where s is the number of simulations, $\hat{\theta}_r$ is the estimate for the 'true' value θ_0 in the rth simulated sample, and the estimated standard deviation of the estimates, defined as $\widehat{\text{sd}}(\hat{\theta}) = \left(s^{-1} \sum_{r=1}^{s} \widehat{\text{var}}(\hat{\theta}_r) \right)^{1/2}$, where $\widehat{\text{var}}(\hat{\theta}_r)$ is the estimated asymptotic variance of $\hat{\theta}_r$, are calculated. In all cases reported in the following section the values of rmse were virtually the same as the standard deviations of the estimates over the simulations.

Using $N = 500$ blocks and $T = 5$ observations in every block for small values of $\rho_{0tt'}$ of an equicorrelation structure the ML and the GEE_E estimators have virtually the same values $\widehat{\text{sd}}$ and rmse (see Table 1). The higher the value of the true correlation is, the larger the difference between the values $\widehat{\text{sd}}$ and rmse of the ML and the GEE_E estimator, with the ML estimator being more efficient than the GEE_E estimator in terms of these measures.

The picture becomes slightly different if one compares the values $\widehat{\text{sd}}$ and rmse

Table 1: *Mean (m), estimated standard deviation $(\widehat{sd})$ and root mean squared error (rmse) of different estimators for a model with $N = 500$, $T = 5$, $\beta_{01} = -.3$, $\beta_{02} = .8$, $\beta_{03} = .8$ and $\beta_{04} = -.8$ and different values for an equicorrelation structure over $s = 200$ simulations*

m $\widehat{sd}$ rmse	$\rho_{0tt'} = .2$			$\rho_{0tt'} = .5$			$\rho_{0tt'} = .8$		
	ML	GEE_E	MEC_E	ML	GEE_E	MEC_E	ML	GEE_E	MEC_E
$\hat{\beta}_{N1}$	$-.3003$	$-.3000$	$-.3001$	$-.3006$	$-.3006$	$-.3021$	$-.2989$	$-.2991$	$-.2989$
	0.0438	0.0439	0.0433	0.0479	0.0485	0.0485	0.0512	0.0542	0.0535
	0.0425	0.0428	0.0447	0.0440	0.0449	0.0484	0.0464	0.0502	0.0524
$\hat{\beta}_{N2}$	0.8019	0.8016	0.8020	0.8031	0.8028	0.8040	0.8009	0.8019	0.7997
	0.0575	0.0574	0.0570	0.0544	0.0548	0.0577	0.0500	0.0523	0.0592
	0.0552	0.0553	0.0578	0.0494	0.0503	0.0565	0.0471	0.0500	0.0625
$\hat{\beta}_{N3}$	0.8038	0.8036	0.8017	0.8024	0.8016	0.7993	0.8020	0.8007	0.7953
	0.0346	0.0347	0.0341	0.0345	0.0348	0.0356	0.0353	0.0364	0.0385
	0.0336	0.0337	0.0355	0.0318	0.0318	0.0358	0.0338	0.0351	0.0422
$\hat{\beta}_{N4}$	$-.8019$	$-.8022$	$-.8022$	$-.7999$	$-.7998$	$-.7991$	$-.8033$	$-.8031$	$-.7976$
	0.0980	0.0977	0.0972	0.0915	0.0924	0.0979	0.0801	0.0858	0.0992
	0.0990	0.0981	0.1016	0.0978	0.0972	0.1085	0.0836	0.0865	0.1102
$\hat{\sigma}_N$	0.4491		0.4535	0.7081		0.7100	0.8942		0.8942
	0.0388		0.0371	0.0246		0.0237	0.0128		0.0124
	0.0380		0.0391	0.0231		0.0240	0.0138		0.0147

of the MEC_E estimator on one hand and of the ML and GEE_E estimator on the other. With a small value of $\rho_{0tt'}$ the values $\widehat{sd}$ are systematically lower and the values of the rmse are systematically higher for the MEC_E estimator than for the ML and the GEE_E estimator. In terms of higher values $\widehat{sd}$ and rmse the MEC_E becomes more inefficient relativ to the other two estimators the higher the true correlation is.

A look at Table 2 reveals that essentially the same results hold for the GEE and the MECOSA estimators using an AR(1) and a mixed correlation structure in the correlation structure of the latent error term. In this Table only the estimation results of a model with a mixed correlation structure with $\sigma_0^2 = .8$ and $\varrho_0 = .2$ are shown because — as described in section 4.1 — in the case of a model with $\sigma_0^2 = .2$ and $\varrho_0 = .8$ only 121 estimation results out of 200 were valid using the MEC_{ARH} estimator (there were no problems with the use of the GEE_{ARH} estimator).

Another point that should be mentioned is the use of the GEE_{ARH} estimator when in fact the true correlation structure of the latent error term is AR(1). This is because the correlation structure in the observable response variables which is

Table 2: *Mean (m), estimated standard deviation $(\widehat{sd})$ and root mean squared error (rmse) of different estimators for a model with $N = 500$, $T = 5$, $\beta_{01} = -.3$, $\beta_{02} = .8$, $\beta_{03} = .8$ and $\beta_{04} = -.8$ and different 'true' correlation structures of the latent errors $\mathbf{C}_0$ over $s = 200$ simulations*

m $\widehat{sd}$ rmse	$\mathbf{C}_0$					
	$\varrho_0 = .2$		$\varrho_0 = .8$		$\sigma_0^2 = .8,\ \varrho_0 = .2$	
	GEE_{ARH}	MEC_{AR1}	GEE_{ARH}	MEC_{AR1}	GEE_{ARH}	MEC_{ARH}
$\hat{\beta}_{N1}$	$-.2996$	$-.3000$	$-.2985$	$-.2961$	$-.2976$	$-.2975$
	0.0419	0.0412	0.0509	0.0511	0.0541	0.0539
	0.0418	0.0424	0.0539	0.0568	0.0501	0.0519
$\hat{\beta}_{N2}$	0.7986	0.7986	0.8010	0.7981	0.7995	0.7971
	0.0578	0.0569	0.0529	0.0584	0.0520	0.0592
	0.0574	0.0601	0.0559	0.0646	0.0485	0.0598
$\hat{\beta}_{N3}$	0.8038	0.8025	0.8022	0.7995	0.8001	0.7944
	0.0346	0.0338	0.0356	0.0372	0.0366	0.0388
	0.0356	0.0371	0.0351	0.0394	0.0339	0.0410
$\hat{\beta}_{N4}$	$-.8054$	$-.8032$	$-.8070$	$-.8047$	$-.8031$	$-.7950$
	0.0984	0.0971	0.0877	0.0985	0.0850	0.0992
	0.1053	0.1047	0.0893	0.0995	0.0864	0.1089
$\hat{\sigma}_N$						0.8889
						0.0296
						0.0288
$\hat{\varrho}_N$	0.2025		0.8013		0.2129	
	0.0416		0.0197		0.1378	
	0.0481		0.0220		0.1798	

no more an AR(1) structure especially for high values of ϱ_0 is better modeled by the GEE_{ARH} estimator than by the GEE_{AR1} estimator (see Spiess, 1995). For a low value of ϱ_0 ($\varrho_0 = .2$) both estimators turned out to lead to the same mean, $\widehat{sd}$ and rmse of the estimates.

To see whether these results depend upon s, we increased the number of simulations up to $s = 500$, using the same 'true' models and the same estimators. With $s = 500$ simulations the numerical results were similar and the overall results did not change at all.

We also increased and decreased the number of observation blocks to $N = 1000$ and $N = 100$, respectively. Because of the problems calculating the ME-COSA estimators for small sample sizes (see section 4.1) the results for the ME-COSA estimator were not valid and are not reported for $N = 100$ blocks.

For $N = 1000$ blocks and $\rho_{0tt'} = .2$ for the true equicorrelation structure the ML estimates and the GEE_E estimates have nearly the same means and values

$\widehat{sd}$ and rmse. With increasing values of $\rho_{0tt'}$ the GEE_E estimator becomes more inefficient relativ to the ML estimator (see Table 3). For low $\rho_{0tt'}$ the means and values of $\widehat{sd}$ and rmse of the MEC_E estimates are similar to the means and values $\widehat{sd}$ and rmse of the ML and the GEE_E estimates, but are significantly higher for moderate and high values of $\rho_{0tt'}$ than for the GEE_E estimates. The same systematic difference between the GEE and the MECOSA estimates is also true for other 'true' correlation structures, that is for an AR(1) and a mixed equicorrelation and AR(1) structure (See Table 4).

Table 3: *Mean (m), estimated standard deviation* $(\widehat{sd})$ *and root mean squared error (rmse) of different estimators for a model with* $N = 1000$, $T = 5$, $\beta_{01} = -.3$, $\beta_{02} = .8$, $\beta_{03} = .8$ *and* $\beta_{04} = -.8$ *and different values for an equicorrelation structure over* $s = 200$ *simulations*

m $\widehat{sd}$ rmse	$\rho_{0tt'} = .2$			$\rho_{0tt'} = .5$			$\rho_{0tt'} = .8$		
	ML	GEE_E	MEC_E	ML	GEE_E	MEC_E	ML	GEE_E	MEC_E
$\hat{\beta}_{N1}$	$-.3009$	$-.3009$	$-.3003$	$-.3011$	$-.3014$	$-.3005$	$-.3043$	$-.3050$	$-.3037$
	0.0310	0.0312	0.0310	0.0339	0.0345	0.0348	0.0363	0.0381	0.0384
	0.0318	0.0316	0.0325	0.0349	0.0346	0.0359	0.0355	0.0370	0.0385
$\hat{\beta}_{N2}$	0.8010	0.8010	0.8002	0.8009	0.8004	0.7988	0.8034	0.8024	0.8000
	0.0406	0.0406	0.0408	0.0385	0.0390	0.0414	0.0354	0.0371	0.0425
	0.0425	0.0424	0.0439	0.0394	0.0391	0.0445	0.0371	0.0386	0.0467
$\hat{\beta}_{N3}$	0.7989	0.7988	0.7977	0.7988	0.7984	0.7958	0.7992	0.7991	0.7958
	0.0244	0.0244	0.0243	0.0243	0.0245	0.0255	0.0249	0.0258	0.0278
	0.0220	0.0221	0.0224	0.0227	0.0233	0.0243	0.0222	0.0230	0.0263
$\hat{\beta}_{N4}$	$-.7995$	$-.7994$	$-.8004$	$-.8010$	$-.8017$	$-.8013$	$-.7971$	$-.7988$	0.7978
	0.0692	0.0695	0.0697	0.0645	0.0655	0.0700	0.0564	0.0601	0.0704
	0.0644	0.0650	0.0674	0.0644	0.0674	0.0715	0.0593	0.0660	0.0777
$\hat{\sigma}_N$	0.4464		0.4481	0.7068		0.7077	0.8938		0.8939
	0.0276		0.0271	0.0175		0.0173	0.0091		0.0091
	0.0253		0.0264	0.0174		0.0182	0.0094		0.0103

With a small sample size ($N = 100$) and a true correlation structure with low and moderate values for $\rho_{0tt'}$ there is no systematic difference in the $\widehat{sd}$ and rmse between the ML and GEE_E estimators (see Table 5). Only for a high correlation ($\rho_{0tt'} = .8$) the GEE_E estimator seems to become inefficient relative to the ML estimator.

For moderate sample sizes ($N = 500$) there is a general tendency to underestimate the variances of the parameters using the MECOSA approach. For $N = 100$ blocks the calculation of the MECOSA estimates may lead to singular matrices $\widehat{W}_N$ or to convergence of the estimates of the tetrachoric correlations to boundary points. In the case of large sample sizes ($N = 1000$) the MECOSA estimators are

Table 4: *Mean (m), estimated standard deviation $(\widehat{sd})$ and root mean squared error (rmse) of different estimators for a model with $N = 1000$, $T = 5$, $\beta_{01} = -.3$, $\beta_{02} = .8$, $\beta_{03} = .8$ and $\beta_{04} = -.8$ and different 'true' correlation structures of the latent errors $\mathbf{C}_0$ over $s = 200$ simulations*

m $\widehat{sd}$ rmse	$\mathbf{C}_0$					
	$\varrho_0 = .2$		$\varrho_0 = .8$		$\sigma_0^2 = .8, \varrho_0 = .2$	
	GEE_{ARH}	MEC_{AR1}	GEE_{ARH}	MEC_{AR1}	GEE_{ARH}	MEC_{ARH}
$\hat{\beta}_{N1}$	$-.3013$	$-.3004$	$-.2995$	$-.2979$	$-.3046$	$-.3026$
	0.0298	0.0296	0.0361	0.0365	0.0384	0.0387
	0.0297	0.0299	0.0346	0.0359	0.0374	0.0385
$\hat{\beta}_{N2}$	0.8006	0.7989	0.7996	0.7957	0.8026	0.7998
	0.0410	0.0408	0.0377	0.0419	0.0370	0.0426
	0.0432	0.0442	0.0348	0.0406	0.0394	0.0475
$\hat{\beta}_{N3}$	0.8009	0.7994	0.7997	0.7957	0.7991	0.7959
	0.0245	0.0243	0.0252	0.0267	0.0259	0.0280
	0.0244	0.0251	0.0236	0.0271	0.0244	0.0276
$\hat{\beta}_{N4}$	$-.8000$	$-.7999$	$-.7978$	$-.7979$	$-.7975$	$-.7965$
	0.0697	0.0695	0.0617	0.0698	0.0596	0.0704
	0.0668	0.0678	0.0619	0.0757	0.0648	0.0771
$\hat{\sigma}_N$						0.8925
						0.0126
						0.0145
$\hat{\varrho}_N$		0.2008		0.8008		0.1997
		0.0304		0.0144		0.1015
		0.0300		0.0146		0.1149

as efficient as the GEE estimators or, in the case of an equicorrelation structure, as the ML estimator, for low true correlations of the latent error terms. As the values of the true correlations increase, the MECOSA estimators become more and more inefficient relative to the GEE and ML estimators, respectively.

The calculation of the GEE and the ML estimates in general were not problematic with $N = 100$, $N = 500$ and $N = 1000$ blocks. In equicorrelation models with low true correlations the ML and the GEE_E estimators are nearly equally efficient, whereas with increasing true correlations the GEE_E estimator becomes inefficient relative to the ML estimator in terms of the used measures. This difference is clearer in large samples ($N = 1000$) than in small samples ($N = 100$).

To compare the estimation of the tetrachoric correlations using the GEE approach as described in section 2.3 and the MECOSA approach we simulated a model with $N = 500$ and $N = 1000$ blocks, respectively, $T = 5$, $\beta = (-.3, .8, .8, -.8)'$ and a true equicorrelation matrix with $\rho_{0tt'} = .2$ and an AR(1)

Table 5: *Mean (m), estimated standard deviation $(\widehat{sd})$ and root mean quared error (rmse) of different estimators for a model with $N = 100$, $T = 5$, $\beta_{01} = -.3$, $\beta_{02} = .8$, $\beta_{03} = .8$ and $\beta_{04} = -.8$ and different values for an equicorrelation structure over $s = 200$ simulations*

m $\widehat{sd}$ rmse	$\rho_{0tt'} = .2$		$\rho_{0tt'} = .5$		$\rho_{0tt'} = .8$	
	ML	GEE_E	ML	GEE_E	ML	GEE_E
$\hat{\beta}_{N1}$	$-.2906$	$-.2900$	$-.2974$	$-.2962$	$-.2981$	$-.2983$
	0.1019	0.1007	0.1107	0.1110	0.1175	0.1227
	0.1115	0.1119	0.1242	0.1251	0.1279	0.1315
$\hat{\beta}_{N2}$	0.7944	0.7945	0.7991	0.8001	0.8019	0.8027
	0.1312	0.1300	0.1251	0.1254	0.1151	0.1193
	0.1404	0.1411	0.1332	0.1353	0.1170	0.1246
$\hat{\beta}_{N3}$	0.8122	0.8118	0.8098	0.8101	0.8082	0.8078
	0.0787	0.0777	0.0781	0.0776	0.0799	0.0807
	0.0776	0.0776	0.0801	0.0800	0.0785	0.0796
$\hat{\beta}_{N4}$	$-.8121$	$-.8112$	$-.8188$	$-.8177$	$-.8211$	$-.8269$
	0.2198	0.2175	0.2059	0.2074	0.1797	0.1908
	0.2253	0.2267	0.2139	0.2164	0.1921	0.2034
$\hat{\sigma}_N$	0.4350		0.7056		0.8963	
	0.1034		0.0560		0.0286	
	0.0944		0.0560		0.0302	

structure in the correlation matrix of the latent error term with $\varrho_0 = .8$ (see Table 6). As estimators we used the GEE_F and the MEC_F estimator.

Looking at Table 6 there seems to be no significant difference between the means of the estimated tetrachoric correlations using the GEE and the MECOSA approach, respectively. Altogether, if the values of the l_2-norm of the vectors of differences between the means of the estimated tetrachoric and the realized underlying correlations are considered, the GEE approach leads to a slightly better fit.

As can be seen from the means of the estimates in Table 1 to Table 5 there is no systematic and significant different bias in the mean values of the ML, GEE and MECOSA estimates. This holds for all our simulation results.

For the different estimates over the s simulations, in general, there were no systematic significant deviations from the normal distribution, the exceptions beeing distributions of the estimates $\hat{\sigma}_N$ with $\sigma_0^2 = .8$ either of the MEC_{ARH} estimator and true mixed correlations structures (Tables 2 and 4) or of the ML estimate for σ_0^2 in the model with $N = 100$ blocks and an equicorrelation structure with $\sigma_0^2 = .8$ (see Table 5). In all three cases the distributions of the estimates

were negatively skewed. This is not surprising because of the high mean and the large variance of the estimates.

Table 6: *True correlations ($\rho_{0tt'}$), mean of simulated correlations ($m(\rho_{Ntt'})$) and mean of estimated tetrachoric correlations ($m(\hat{\rho}_{Ntt'})$), $t < t'$, over $s = 200$ using GEE_F- and MEC_F estimator respectively for a model with $T = 5$, $\beta_{01} = -.3$, $\beta_{02} = .8$, $\beta_{03} = .8$, $\beta_{04} = -.8$ and $N = 500$ and $N = 1000$, respectively*

		$\mathbf{C_0} : \rho_{0tt'} = .2$							
			$N = 500$	GEE_F	MEC_F		$N = 1000$	GEE_F	MEC_F
		$\rho_{0tt'}$	$m(\rho_{Ntt'})$	$m(\hat{\rho}_{Ntt'})$	$m(\hat{\rho}_{Ntt'})$	$\rho_{0tt'}$	$m(\rho_{Ntt'})$	$m(\hat{\rho}_{Ntt'})$	$m(\hat{\rho}_{Ntt'})$
	21	.2	.2052	.2133	.2145	.2	.2022	.2126	.2130
	31	.2	.2044	.2127	.2147	.2	.1979	.1958	.1964
	32	.2	.2064	.2063	.2074	.2	.2051	.2111	.2102
	41	.2	.1997	.2143	.2150	.2	.2001	.2022	.2078
tt'	42	.2	.1974	.2071	.2060	.2	.2004	.2052	.2046
	43	.2	.1950	.1932	.1953	.2	.1957	.1886	.1886
	51	.2	.1981	.1962	.1975	.2	.1987	.1952	.1961
	52	.2	.1990	.2040	.2075	.2	.2012	.2021	.2031
	53	.2	.2046	.2048	.2036	.2	.2012	.1964	.1958
	54	.2	.1991	.1920	.1939	.2	.1996	.1978	.1979
		$\mathbf{C_0} : \varrho_0 = .8$							
			$N = 500$	GEE_F	MEC_F		$N = 1000$	GEE_F	MEC_F
		$\rho_{0tt'}$	$m(\rho_{Ntt'})$	$m(\hat{\rho}_{Ntt'})$	$m(\hat{\rho}_{Ntt'})$	$\rho_{0tt'}$	$m(\rho_{Ntt'})$	$m(\hat{\rho}_{Ntt'})$	$m(\hat{\rho}_{Ntt'})$
	21	.8000	.8015	.8076	.8116	.8000	.8010	.8039	.8052
	31	.6400	.6416	.6521	.6538	.6400	.6390	.6413	.6420
	32	.8000	.8020	.8085	.8117	.8000	.8005	.8014	.8035
	41	.5120	.5154	.5270	.5292	.5120	.5121	.5106	.5123
tt'	42	.6400	.6433	.6492	.6513	.6400	.6417	.6454	.6473
	43	.8000	.8007	.8043	.8059	.8000	.7998	.8012	.8031
	51	.4096	.4098	.4136	.4159	.4096	.4081	.4146	.4149
	52	.5120	.5120	.5091	.5117	.5120	.5122	.5133	.5157
	53	.6400	.6398	.6372	.6397	.6400	.6391	.6466	.6477
	54	.8000	.7991	.8041	.8068	.8000	.7995	.8042	.8054

Note: The means of the correlations are calculated using Fisher's Z-transformation.

A further — although not surprising — result indicates the essential but often overlooked effect of the distribution, or more generally, of the type of the covariates on the properties of the estimators: the values $\widehat{sd}$ and rmse are systematically highest for the estimated regression parameter which weights the uniform distributed covariate, whereas those for the parameter estimator which

weights the normal distributed covariates are systematically lowest. The values $\widehat{\text{sd}}$ and rmse for the regression parameter estimator that weights the dichotomous covariates are in between.

5 Discussion

The GEE estimates required substantially lesser computing time than the other estimates. Furthermore, in contrast to Sharples and Breslow (1992) in the calculation of the GEE estimates we found no connection between convergence problems and features of datasets or true values of some parameters. However, in further studies the calculation of the GEE estimates may be found to be more problematic for datasets with very low or very high portions of response variables having value one or for datasets were the portions of violations of the restrictions on the correlations (Prentice, 1988) are higher than in our study.

Again, from a technical point of view the calculation of the MECOSA estimates as well as the calculation of the tetrachoric correlations using the GEE approach are not recommended in small samples because of the possibility of considerable convergence problems.

Although the calculation of the ML estimates required the most computing time, this approach was the one that did not cause any problems, provided suitable starting values were used. Whereas for the regression parameters the ML estimates from the independent probit model seemed to be a good choice in practice, the starting value for the standard deviation of the heterogenity component has to be chosen by theoretical considerations.

Another point worth mentioning is the use of the matrix of second derivatives in the calculation of the ML estimator. Although the second derivative of the log likelihood function is costly to derive, its use leads to lesser computing time required than the use of the first derivatives only and, furthermore, a robust variance estimate may be calculated in practical applications (see White, 1982).

The ML estimator in general seems to be the most efficient estimator[6]. Therefore, if a latent variance component model with an equicorrelation structure can be assumed and computing time is no issue the ML approach is prefered over the GEE and MECOSA approach.

On the other hand, if only small correlations between the error terms can be assumed, the GEE estimator may be used in practical applications with only a negligible loss of efficiency. The same seems to be true in small samples and for low to moderate true correlations, where we found no significant and systematic difference in the efficiency of the ML and GEE estimators. Clearly, if no equicorrelation structure of the error terms of a latent model can be assumed,

---

[6]Although it is clear that the results reported should not be overgeneralized, we expect them to be valid not only for the examples considered in this article, since we found the same general results simulating and estimating a lot of more models than reported here.

the adoption of the ML approach as described in section 2.2 leads to a model misspecification. Using the GEE approach in this case it is possible to model the structure of dependence in the observable response variables more properly.

In small samples the use of the MECOSA approach led to results which are not reliable, mainly caused by estimates converging to boundary points in the second step and by nearly singular weight matrices used in the third step. In moderate samples we observed a tendency of underestimating the variances of the estimators. This tendency vanished with the use of a large number of observation blocks but in this case the MECOSA estimators were found to be inefficient relative to the GEE estimators for moderate to high true correlations. Hence the MECOSA estimators cannot be recommended for small or moderate sample sizes.

Obviously, an advantage of the MECOSA approach is its generality and the possibility to estimate complicated models including the estimation of parameters determining different correlation structures in the latent model. As was shown in section 2.3 and in section 4.2, the GEE approach may be extended to estimate the tetrachoric correlations using the ML method in a second step. Beyond the estimation of the pairwise correlations of the latent errors it should be possible to estimate functions of the correlations, dependent upon the assumed correlation structure in the latent errors. The advantage of this approach over the MECOSA approach is that the properties of the estimators of the regression parameters would not depend on the properties of the estimators of the tetrachoric correlations. Further theoretical and practical work is needed to derive those estimators and their asymptotic properties as well as to investigate their properties in finite samples.

In this paper we included only free varying covariates in our models. The results for the different regression parameter estimates illustrate the fact, that although overlooked in many cases, the distribution, or more generally spoken the kind of covariates, play an important role regarding the properties of the estimators (see also Li and Duan, 1989). Therefore, the results presented in this paper are strictly speaking only valid for estimators of models which include free varying covariates. In a different paper we address the question of the effect of time and block invariant covariates on the properties of the GEE estimators (Spiess and Hamerle, in preparation).

References

Anderson, D.A. & Aitkin, M. (1985). Variance component models with binary responses: Interviewer variability. *Jornal of the Royal Statistical Society, Ser. B, 47(2)*, 203–210.

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.

Fisher, R.A. (1963). *Statistical methods for research workers* (13th ed). Edinburgh: Oliver and Boyd.

Hamerle, A. & Nagl, W. (1987). *Misspecification in models for discrete panel data: Applications and comparisons of some estimators* (Diskussionsbeiträge Nr. 105/s). Konstanz: Fakultät für Wirtschaftswissenschaften und Statistik.

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.

Küsters, U. (1987). *Hierarchische Mittelwert- und Kovarianzstrukturmodelle mit nichtmetrischen endogenen Variablen*. Heidelberg: Physica-Verlag.

Küsters, U. (1990). A note on sequential ML estimates and their asymptotic covariances. *Statistical Papers, 31*, 131–145.

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

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

Liang, K-Y., Zeger, S.. & Qaqish, B. (1992). Multivariate regression analysis for categorical data. *Journal of the Royal Statistical Society, Ser. B., 54(1)*, 3–40.

Owen, D.B. (1956). Tables for computing bivariate normal probabilities. *The Annals of Mathematical Statistics, 27(2)*, 1075–1090.

Pearson, K. (1900). Mathematical contributions to the theory of evolution. – VIII. On the inheritance of characters not capable of exact quantitative measurement. *Philosophical Transactions Of The Royal Society Of London, Series A, 195*, 79–150.

Prentice, R.L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics, 44*, 1033–1048.

Qu, Y., Williams, G.W., Beck, G.J. & Medendorp, S.V. (1992). Latent variable models for clustered dichotomous data with multiple subclusters. *Biometrics, 48*, 1095–1102.

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.

Schepers, A. (1991). *Numerische Verfahren und Implementation der Schätzung von Mittelwert- und Kovarianzstrukturmodellen mit nichtmetrischen Variablen*. Ahaus: Verlag Frank Hartmann.

Schepers, A., Arminger, G. & Küsters, U. (1991). The analysis of non-metric endogenous variables in latent variable models: The MECOSA approach. In J. Gruber (ed.), *Econometric Decision Models: New Methods of Modelling and Applications. Lecture Notes in Economics and Mathematical Systems: 366. (pp. 459–472)*. Berlin: Springer.

Shapiro, A. (1986). Asymptotic theory of overparameterized structural models. *Journal of the American Statistical Association, 81*, 142–149.

Sharples, K. & Breslow, N. (1992). Regression analysis of correlated binary data: Some small sample results for the estimating equation approach. *Journal of Statistical Computation and Simulation, 42*, 1–20.

Sobel, M.E. & Arminger, G. (1992). Modeling household fertility decisions: A nonlinear simultaneous probit model. *Journal of the American Statistical Association, 87(417)*, 38–47.

Spiess, M. (1995). *Parameterschätzung in Regressionsmodellen mit abhängigen diskreten endogenen Variablen*. Konstanz: Hartung-Gorre Verlag.

Spiess, M. & Hamerle, A. (in preparation). On the properties of GEE estimators in the presence of invariant covariates.

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

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

Zeger, S. & Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics, 42*, 121–130.

Zhao, L.P. & Prentice, R.L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika, 77*, 642–648.