Spiess, Hamerle:

Estimation of multivariate probit models: A mixed generalized estimating/pseudo-score equations approach and some finite sample results

Projektpartner

Estimation of multivariate probit models: A mixed generalized estimating/pseudo-score equations approach and some finite sample results

Martin Spiess and Alfred Hamerle
Lehrstuhl für Statistik
Wirtschaftswissenschaftliche Fakultät
Universität Regensburg

Abstract

In the present paper a mixed approach is proposed for the simultaneously estimation of regression and correlation structure parameters in multivariate probit models using generalized estimating equations for the former and pseudo-score equations for the latter. The finite sample properties of the corresponding estimators are compared to estimators proposed by Qu, Williams, Beck and Medendorp (1992) and Qu, Piedmonte and Williams (1994) using generalized estimating equations for both sets of parameters via a Monte Carlo experiment. As a 'reference' estimator for an equicorrelation model, the maximum likelihood (ML) estimator of the random effects probit model is calculated. The results show the mixed approach to be the most robust approach in the sense that the number of datasets for which the corresponding estimates converged was largest relative to the other two approaches. Furthermore, the mixed approach led to the most efficient non-ML estimators and to very efficient estimators for regression and correlation structure parameters relative to the ML estimator if individual covariance matrices were used.

Key words: Generalized estimating equations; Pseudo-score equations; Multivariate binary data; Panel data; Simulation study

1   Introduction

In the last few years estimation of regression models with correlated binary response variables has received considerable attention. A prominent and flexible

subclass within this class of regression models consists of binary probit models with correlated responses which are the focus of the present paper. Since high-dimensional integrals have to be evaluated, maximum likelihood (ML) estimation of the general binary probit models is quite expensive if at all possible.

One possibility avoiding the evaluation of high-dimensional integrals is to use the 'generalized estimating equations' (GEE) approach proposed by Liang and Zeger (1986) which leads to consistent and asymptotically normally distributed estimators if only the expectation of the responses given the covariates are correctly specified. Originally, this approach was not developed for the estimation of probit models solely, but was introduced by Liang and Zeger (1986) and Zeger and Liang (1986) using a linear and two logit models as examples. Furthermore, using the GEE approach, the structure of dependence within blocks (e.g. subjects or families) are modelled in the observable response variables and not in the reponse variables of an underlying model given exogenous variables. The assumption of a latent regression model, however, is crucial e.g. in structural equation models and is a plausible assumption in many contexts (e.g. Ashford and Sowden, 1970; Bartholomew, 1987; Muthén 1984).

In their work Qu, Williams, Beck and Medendorp (1992) and Qu, Piedmonte and Williams (1994) incorporate the assumption of a latent regression model with multivariate normally distributed errors into the GEE approach. They propose the simultaneously estimation of regression parameters and of functions of the tetrachoric correlations, i.e. functions of the correlations of the latent errors, using an extension of the GEE approach which is similar to an approach introduced by Prentice (1988). Calculating the estimates, the regression parameters and the correlation structure parameters are assumed to be orthogonal to one another. The advantage of this approach is that the regression parameter estimator remains consistent if only the expectation of the response given the covariates is correctly specified even if the correlation structure is misspecified. The results of a simulation study designed to '...examine the small sample performance of GEE estimates for both regression and covariance parameters ...' (pp. 244–245) are presented in Qu et al. (1994). Unfortunately, no results for the covariance parameters, i.e. correlation structure parameters, are given allowing the evaluation of the finite sample properties.

Zhao and Prentice (1990) proposed the non-orthogonal simultaneously estimation of regression parameters and correlation structure parameters, again as in Prentice (1988) of the correlation structure of the observable responses, leading to more efficient estimators. In this case, however, the consistency of all estimators depends not only on the correct specification of the expectation of the response given the covariates but also on the correct specification of the structure of dependence (Liang, Zeger and Qaqish, 1992). Furthermore, the efficiency gain for the regression parameter estimators seems to be rather small compared to the above mentioned GEE approach (Liang, Zeger and Qaqish, 1992). On the other hand, estimation of the parameters of the dependence structure (Liang et

al., 1992, used odds-ratios) were more efficient using the method of Zhao and Prentice (1992) than using the approach proposed by Liang and Zeger (1986).

To increase efficiency of the parameters modelling the correlation structure of the latent errors, in the present paper a mixed generalized estimating/pseudo-score equations (GEPSE) approach is proposed, where similar to Qu et al. (1992) and Qu et al. (1994) both sets of parameters are estimated simultaneously. However, whereas the regression parameter estimators are calculated using the GEE approach, the estimators of the correlation structure parameters are defined to be the solution of so-called pseudo-score equations (Gourieroux, Montfort and Trognon, 1984a, 1984b). This approach combines the efficiency of estimators calculated using the maximum likelihood principle and the robustness of the regression parameters estimators against misspecification of the correlation structure by treating both sets of parameters as if they were orthogonal.

The finite sample properties of these estimators are compared with those of the GEE estimators as proposed by Qu et al. (1992) and Qu et al. (1994) and, for an equicorrelations structure in the latent errors, with the properties of the ML estimator of a random effects probit model via a Monte Carlo experiment.

This paper is organized as follows. In section 2 the general model is introduced and in section 3 the estimation procedures are described. Section 3.1 gives the approach proposed by Qu et al. (1992) and Qu et al. (1994), whereas in section 3.2 the mixed approach is described. The format and results of a Monte Carlo experiment designed to evaluate the finite sample efficiency of the estimators are presented in section 4.1 and section 4.2, respectively. Conclusions can be found in section 5.

2   Model and Notation

Let N $(n = 1, \ldots, N)$ be the number of blocks and T $(t = 1, \ldots, T)$ be the number of observations within every block. The $(T \times 1)$ vector of observable binary responses for the nth block will be denoted as $y_n = (y_{n1}, \ldots, y_{nT})'$. Let $x_{nt} = (x_{nt1}, \ldots, x_{ntP})'$ denote the $(P \times 1)$ vector of covariates associated with the ntth observation and X_n the $(T \times P)$ matrix of covariates associated with the nth block. The $(NT \times P)$ data matrix is assumed to have full column rank.

Throughout a threshold model (Pearson, 1900)

$$y_{nt}^* = x_{nt}'\beta^\star + v_{nt} \quad \text{and} \quad y_{nt} = \begin{cases} 1 & \text{if } y_{nt}^* > 0, \\ 0 & \text{otherwise,} \end{cases}$$

is assumed, where y_{nt}^* is the latent, i.e. not observable, continuous response variable, $\beta^\star$ is the unknown regression parameter vector and v_{nt} is an unobservable error term independently distributed from the covariates. For the multivariate binary probit model considered, let $v_n = (v_{n1}, \ldots, v_{nT})'$, $v_n \sim N(0, \Sigma)$ and $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_T) \, \text{R} \, \text{diag}(\sigma_1, \ldots, \sigma_T)$, where $\text{diag}(\sigma_1, \ldots, \sigma_T)$ denotes a

diagonal matrix with the diagonal elements being the standard deviations of v_{nt}, σ_t, and R denotes the correlation matrix of v_n with elements $\rho_{tt'}$ $(t \neq t')$. Throughout, ρ denotes the vector of all $T(T-1)/2$ off diagonal elements of R, i.e. $\rho = (\rho_{21}, \rho_{31}, \rho_{32}, \ldots, \rho_{T(T-1)})'$, where $\rho_{tt'}$ is the element in the tth row and t'th column of R. The structure of R is determined by a $K \times 1$ parameter vector denoted by ϑ. For example, if an equicorrelation structure is assumed in the correlation matrix of the latent error terms $\rho = 1_{T(T-1)/2} \vartheta$, where $1_{T(T-1)/2} = (1, \ldots, 1)'$ is a $(T(T-1)/2 \times 1)$-vector and ϑ is a scalar $(|\vartheta| < 1$ and $\vartheta \neq -1/(T-1))$. If a stationary first-order autoregressive process (AR(1) process) is assumed in the latent error terms, then the corresponding structure of the assumed correlation matrix is an AR(1) structure, where the elements of ρ are $\rho_{tt'} = \vartheta^{|t-t'|}$ and ϑ again is a scalar $(|\vartheta| < 1)$. Observations from different blocks are assumed to be independent.

In the sequel let $\Phi(\cdot)$ denote the standard normal cumulative distribution function, $\varphi(\cdot)$ the standard normal density function, $\Phi(\cdot, \cdot, \rho_{tt'})$ the standard bivariate normal cumulative distribution function and $\varphi(\cdot, \cdot, \rho_{tt'})$ the standard bivariate normal density function.

3 Estimation approaches

In the model of section 2 not all parameters are identified. Therefore, the restriction $\sigma_1 = \sigma_t = \cdots = \sigma$ will be adopted for the remainder of this paper[1]. The identifiable parameters then are $\beta = \sigma^{-1}\beta^\star$, the regression parameter, and ϑ, the correlation structure parameter determining the structure of R.

3.1 The GEE approach

Following Prentice (1988), Qu et al. (1992) and Qu et al. (1994) define the estimator of the regression parameter β to be the solution to the generalized estimating equations

$$\sum_n A_n' \, \Omega_n^{-1} \, e_n = 0, \tag{1}$$

where $e_n = (y_n - \mu_n)$. For the multivariate probit model considered in this paper, $\mu_n = (\mu_{n1}, \ldots, \mu_{nT})'$, where $\mu_{nt} = \Phi(x_{nt}'\beta)$, $\Omega_n = \text{Cov}(y_n)$, with diagonal elements $\mu_{nt}(1 - \mu_{nt})$ and covariance, i.e. off diagonal element, $\Phi(x_{nt}'\beta, x_{nt'}'\beta, \rho_{tt'}) - \mu_{nt}\mu_{nt'}$ in the tth row and t'th column $(t \neq t')$ and $A_n' = X_n'\text{diag}(\varphi(x_{n1}'\beta), \ldots, \varphi(x_{nT}'\beta))$.

The generalized estimating equations for the estimation of the correlation structure parameter ϑ are given by

$$\sum_n E_n' V_n^{-1} w_n = 0, \tag{2}$$

---

[1]This constraint is more restrictive than necessary and could be relaxed in what follows.

where $w_n = (s_n - \omega_n)$, $\omega_n = (\omega_{n21}, \omega_{n31}, \ldots, \omega_{nT(T-1)})'$ and $\omega_{ntt'}$ $(t' < t)$ are the elements below the diagonal of the covariance matrix Ω_n. Correspondingly, $s_n = (s_{n21}, s_{n31}, \ldots, s_{nT(T-1)})'$ with elements $s_{ntt'} = (y_{nt} - \mu_{nt})(y_{nt'} - \mu_{nt'})$ $(t' < t)$. The matrix V_n is a covariance matrix for s_n. Qu et al. (1992) and Qu et al. (1994) take $V_n = \text{diag}(v_{n21}, v_{n31}, \ldots, v_{nT(T-1)})'$, where

$$v_{ntt'} = \text{Var}(s_{ntt'}) = (1 - 2\mu_{nt})(1 - 2\mu_{nt'})\omega_{ntt'} + \omega_{ntt}\omega_{nt't'} - \omega_{ntt'}^2$$

and ω_{ntt} is the tth diagonal element of Ω_n. The matrix E_n' is defined as

$$\frac{\partial \omega_n}{\partial \vartheta} = \frac{\partial \rho}{\partial \vartheta} \text{diag}(\varphi(\mu_{n2}, \mu_{n1}, \rho_{21}), \ldots, \varphi(\mu_{nT}, \mu_{n(T-1)}, \rho_{T(T-1)})),$$

where ρ is considered as a function of the structural parameter ϑ which may be a vector or a scalar depending upon the assumed correlation strucure. The GEE estimator $\hat{\theta} = (\hat{\beta}', \hat{\vartheta}')'$ is calculated iteratively with updated value in the $(j+1)$th iteration given by

$$\hat{\theta}_{j+1} = \hat{\theta}_j +$$
$$\left(\begin{array}{cc} \sum_n A_n' \Omega_n^{-1} A_n & 0 \\ 0 & \sum_n E_n' V_n^{-1} E_n \end{array} \right)_{\theta = \hat{\theta}_j}^{-1} \left(\begin{array}{c} \sum_n A_n' \Omega_n^{-1}(y_n - \mu_n) \\ \sum_n E_n' V_n^{-1}(s_n - \omega_n) \end{array} \right)_{\theta = \hat{\theta}_j}.$$

The asymptotic covariance matrix of $\sqrt{N}(\hat{\theta} - \theta)$ can consistently be estimated by

$$\widehat{\text{Cov}} = N \left(\begin{array}{cc} L & 0 \\ M & Q \end{array} \right)^{-1} \left(\begin{array}{cc} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12}' & \Lambda_{22} \end{array} \right) \left(\begin{array}{cc} L & M' \\ 0 & Q \end{array} \right)^{-1},$$

where

$$L = \left(\sum_n A_n' \Omega_n^{-1} A_n \right)_{\theta = \hat{\theta}},$$

and

$$\Lambda_{11} = \left(\sum_n A_n' \Omega_n^{-1} e_n e_n' \Omega_n^{-1} A_n \right)_{\theta = \hat{\theta}}$$

(see Qu et al., 1992; Qu et al., 1994; Prentice, 1988). For M, Q, Λ_{12} and Λ_{22} Qu et al. (1992) and Qu et al. (1994) refer to Prentice (1988). According to Prentice (1988)

$$Q = \left(\sum_n E_n' V_n^{-1} E_n \right)_{\theta = \hat{\theta}},$$
$$\Lambda_{12} = \left(\sum_n A_n' \Omega_n^{-1} e_n w_n' V_n^{-1} E_n \right)_{\theta = \hat{\theta}}$$

and

$$\Lambda_{22} = \left(\sum_n \mathrm{E}'_n \mathrm{V}_n^{-1} \mathrm{w}_n \mathrm{w}'_n \mathrm{V}_n^{-1} \mathrm{E}_n \right)_{\theta = \hat{\theta}}.$$

However, the matrix M is not identical to the corresponding matrix given by Prentice (1988, p. 1039). Given that $\mathrm{E}(s_n)$ is correctly specified, then under mild regularity conditions (Prentice, 1988) for $N \to \infty$

$$N^{-1} \frac{\partial}{\partial \beta'} \sum_n \mathrm{E}'_n \mathrm{V}_n^{-1} (s_n - \omega_n) \xrightarrow{p} N^{-1} \sum_n \mathrm{E}'_n \mathrm{V}_n^{-1} \frac{\partial (s_n - \omega_n)}{\partial \beta'},$$

since both s_n and ω_n are functions of β. Therefore,

$$\mathrm{M} = - \left(\sum_n \mathrm{E}'_n \mathrm{V}_n^{-1} \frac{\partial (s_n - \omega_n)}{\partial \beta'} \right)_{\theta = \hat{\theta}}.$$

In section 4 two correlation structures in the correlation matrix of the latent error terms are considered: an equicorrelation and an AR(1) structure. The corresponding estimators will be denoted as GEE_E and GEE_A estimator, respectively.

Following the idea in Liang and Zeger (1986) of using a correlation matrix which is identical for all N observation blocks, instead of using the 'working' covariance matrix $\Omega_n = \mathrm{Cov}(y_n)$ as defined above, the 'working' covariance matrix Ω_n^* with off-diagonal elements

$$\omega_{ntt'}^* = \left(\mu_{nt}(1 - \mu_{nt}) \right)^{1/2} r_{tt'} \left(\mu_{nt'}(1 - \mu_{nt'}) \right)^{1/2},$$

where

$$r_{tt'} = \left(\sum_n \mu_{nt}(1 - \mu_{nt}) \right)^{-1/2} \left(\sum_n \omega_{ntt'} \right) \left(\sum_n \mu_{nt'}(1 - \mu_{nt'}) \right)^{-1/2}$$

and diagonal elements

$$\omega_{ntt}^* = \mu_{nt}(1 - \mu_{nt})$$

will be calculated. The corresponding estimators will be denoted as GEE_E^* and GEE_A^* estimator, respectively.

3.2 The GEPSE approach

As another approach consider again the generalized estimating equations (1) for the estimation of β as defined in section 3.1 and the pseudo-score equations for the estimation of the correlation structure parameter

$$\sum_n \mathrm{B}'_n \mathrm{W}_n^{-1} \mathrm{v}_n = 0. \tag{3}$$

The elements of the $(T(T-1)/2 \times 1)$ vector v_n are $(2y_{nt}-1)(2y_{nt'}-1)$,

$$W_n = \text{diag}(P_{n(2,1)}, \ldots, P_{n(T,T-1)}),$$

where $P_{n(t,t')} = \Pr(y_{nt}, y_{nt'} \mid x'_{nt}\hat{\beta}, x'_{nt'}\hat{\beta}, \rho_{tt'})$ is the probability of the variables y_{nt} and $y_{nt'}$ assuming specific values, given the covariates, the regression parameter and the correlation, and

$$B'_n = \frac{\partial \rho}{\partial \vartheta} \, \text{diag}(\varphi(x'_{n2}\hat{\beta}, x'_{n1}\hat{\beta}, \rho_{21}), \ldots, \varphi(x'_{nT}\hat{\beta}, x'_{n(T-1)}\hat{\beta}, \rho_{T(T-1)})),$$

where ρ is again considered as a function of the structural parameter ϑ. Note that (2) is just the vector of first derivatives of the pseudo-maximum likelihood functions

$$l(\vartheta) = \sum_n l_n(\vartheta) = \sum_n \sum_{\substack{t,t' \\ (t'<t)}} \log P_{n(t,t')}$$

with respect to ϑ, where $\sum_{\substack{t,t' \\ (t'<t)}}$ means summation over all $T(T-1)/2$ probabilities $P_{n(2,1)}, P_{n(3,1)}, P_{n(3,2)}, \ldots, P_{n(T,T-1)}$. Note that $P_{n(t,t')}$ is also a function of β, so if necessary the function $l(\vartheta)$ will also be written as $l(\vartheta, \beta)$.

The corresponding estimator $\hat{\vartheta}$ is similar to the pseudo ML (PML) estimator described in Gourieroux, Monfort and Trognon (1984a), since both estimators are calculated as if the $y_t y_{t'}$ were independent. However, contrary to Gourieroux et al. (1984a) who used PML estimators for β calculated under the assumption of an independent probit model, in the approach proposed above both sets of parameters are estimated simultaneously and, furthermore, in estimating the regression parameters the assumed structure of association is taken into account.

The vector of estimates $\hat{\theta} = (\hat{\beta}', \hat{\vartheta}')'$ is iteratively calculated with updated value in the $(j+1)$th iteration given by

$$\hat{\theta}_{j+1} = \hat{\theta}_j -$$
$$\begin{pmatrix} -\sum_n A'_n \Omega_n^{-1} A_n & 0 \\ 0 & \sum_n \frac{\partial^2 l_n(\vartheta)}{\partial \vartheta \, \partial \vartheta'} \end{pmatrix}_{\theta=\hat{\theta}_j}^{-1} \begin{pmatrix} \sum_n A'_n \Omega_n^{-1} e_n \\ \sum_n B'_n W_n^{-1} v_n \end{pmatrix}_{\theta=\hat{\theta}_j}.$$

Since the estimator $\hat{\theta}$ is defined to be the solution to two sets of equations, namely the generalized estimating equations (1) and the pseudo-score equations (3), it will be denoted as GEPSE estimator.

Following closely the theory in Prentice (1988) $\sqrt{N}(\hat{\theta} - \theta_0)$, where θ_0 is the true value, can be shown to be asymptotically normally distributed with zero mean and covariance matrix consistenly estimated by

$$\widehat{\text{Cov}} = N \begin{pmatrix} L & 0 \\ M & Q \end{pmatrix}^{-1} \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda'_{12} & \Lambda_{22} \end{pmatrix} \begin{pmatrix} L & M' \\ 0 & Q \end{pmatrix}^{-1},$$

7

where L and Λ_{11} as defined in section 3.1,

$$
\begin{aligned}
\text{M} &= \left(\sum_n \frac{\partial^2 l_n(\vartheta, \beta)}{\partial \vartheta \, \partial \beta'} \right)_{\theta = \hat{\theta}}, \\
\text{Q} &= \left(\sum_n \frac{\partial^2 l_n(\vartheta)}{\partial \vartheta \, \partial \vartheta'} \right)_{\theta = \hat{\theta}}, \\
\Lambda_{12} &= \left(\sum_n \text{A}_n' \Omega_n^{-1} \text{e}_n \text{v}_n' \text{W}_n^{-1} \text{B}_n \right)_{\theta = \hat{\theta}}
\end{aligned}
$$

and

$$
\Lambda_{22} = \left(\sum_n \text{B}_n' \text{W}_n^{-1} \text{v}_n \text{v}_n' \text{W}_n^{-1} \text{B}_n \right)_{\theta = \hat{\theta}}
$$

(see Appendix).

Again, two different correlation structures will be estimated, i.e. an equicorrelation structure leading to an estimator denoted as GEPSE_E estimator and an AR(1) structure leading to an estimator denoted as GEPSE_A estimator. Furthermore, as in section 3.1 an alternative 'working' covariance matrix Ω_n^* will be used, which is calculated as described in section 3.1. The corresponding estimators will be denoted as GEPSE_E^* and GEPSE_A^* estimator, respectively.

4 Simulation Study

4.1 Description

To compare the finite sample properties of the estimators described above, a simulation study was conducted using the 'interactice matrix language' (IML) included in the SAS system ('statistical analysis system'; SAS Institute Inc., 1989). To save space, in subsection 4.2 not all simulation results are presented. Therefore, in this subsection only those factors of the Monte Carlo experiment relevant to the main results reported in the next subsection are described.

According to a panel regression model with $T = 5$ observations within each block, for each of $s = 500$ replications samples of $N = 50$, $N = 100$ and $N = 500$ blocks were generated. Using the random number generators RANNOR and RANUNI provided by the SAS system (SAS Institute Inc., 1990), one dichotomous, one normally and one uniformly distributed covariate was generated for every observation. These covariates variied freely over all NT observations but were held constant over the s replications. The values of the regression parameters were chosen to be $\beta_2 = .1$ weighting the dichotomous, $\beta_3 = .8$ weighting the normally distributed, $\beta_4 = -1.5$ weighting the uniformly distributed covariate and $\beta_1 = -.3$ weighting the constant term in the data matrix. The error terms were generated as standard normally distributed variates according to one of

two correlation structures: (1) an equicorrelation structure with $\rho_{tt'} = \vartheta = .2$ or $\rho_{tt'} = \vartheta = .8$ for all t, t' and (2) an AR(1) structure with $\rho_{tt'} = \vartheta^{|t-t'|}$ and $\vartheta = .2$ or $\vartheta = .8$ for all t, t'. The standard deviations were restricted to $\sigma = 1$, hence $\beta = \beta^\star$. The 'latent' responses and the 'observable' responses y_{nt} were then generated according to the model described in section 2.

As a 'reference' estimator in the case of the equicorrelation model, we calculated the maximum likelihood estimator of the random effects probit model (e.g. Butler and Moffit, 1982), restricting the error variance to unity. For the ML estimator, $\hat{\theta} = (\hat{\beta}_1, \ldots, \hat{\beta}_4, \hat{\vartheta})'$, to be unbiased in this model, a necessary condition is a sufficient number of points used for the approximative evaluation of the integrals in the log likelihood function and their derivatives. To ensure this, we calculated the ML estimates for the models over $s = 500$ replications, successively increasing the number of evaluation points by one until the results remained stable. For the computation of the estimates and for the estimation of the variances, the matrix of analytical second derivatives was used (for details see Spiess and Hamerle, 1995).

As starting values for the calculation of the estimates the 'true' parameter values used to create the datasets were used. Since the log likelihood function for the computation of the ML estimate 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 β in this case.

The iterations stopped 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}.

To compare the results, we used the following measures, where for simplicity $\hat{\theta}$ denotes a scalar estimate and θ denotes the 'true' value:

1. the arithmetic mean of the estimates over s replications (m),

2. the relative bias (bias), defined as $|(\text{m}(\hat{\theta}) - \theta)/\theta|$,

3. the standard deviation (sd) of the estimates over s replications,

4. the estimated standard deviation defined as $\widehat{\text{sd}} = \big(s^{-1} \sum_{r=1}^{s} \widehat{\text{Var}}(\hat{\theta}_r)\big)^{1/2}$, where $\widehat{\text{Var}}(\hat{\theta}_r)$ is the estimated asymptotic variance of $\hat{\theta}_r$ $(r = 1, \ldots, s)$,

5. $\text{sd}/\widehat{\text{sd}}$, indicating an over- or underestimation of the variance,

6. the root mean squared error (rmse) of the estimates and

7. the proportion of rejections (rej) at the 5% level of significance of the null hypothesis that the parameter is identical to the 'true' value against a two-sided alternative (df $= N - P - K$).

9

4.2 Simulation Study: Results

Because of the limited space, not all simulation results are presented. However, the results reported in this section are consistent with the findings using different models.

In small samples ($N = 50$), the calculation of the different estimators turned out to be problematic. For example, for an equicorrelation structure with $\vartheta = .2$, the ML estimator of ϑ converged to the boundary point zero in 9 out of 500 datasets. For $\vartheta = .8$, again for a 'true' equicorrelation matrix, the GEE_E estimates converged only in 450 out of 500 datasets to a solution although a global strategy (see Dennis and Schnabel, 1983) was implemented and different starting values were used. Therefore, the estimation results are hardly comparable over estimation methods and simulated models. However, it is possible to draw some conclusions about the robustness of the estimation procedures in small samples. Considering the results in Table 1, the different estimation procedures may be ranked according to the number of datasets for which the estimates converged: (1) $\text{GEPSE}^*_{E/A}$, (2) $\text{GEPSE}_{E/A}$, (3) $\text{GEE}^*_{E/A}$ and (4) $\text{GEE}_{E/A}$. Clearly, the most robust estimation procedure is the GEPSE^* procedure followed by the GEPSE procedure. If only the equicorrelation models are considered, the ML estimation procedure lies somewhere between the GEPSE^*/GEPSE and GEE^*/GEE estimation procedures.

Table 1: *Number of datasets for which the estimates converged for a model with $N = 50$, $T = 5$, $\beta_1 = -.3$, $\beta_2 = .1$, $\beta_3 = .8$, $\beta_4 = -1.5$ and an equicorrelation (Equi) and an AR(1) (AR(1)) structure with $\vartheta = .2$ and $\vartheta = .8$ over $s = 500$ replications*

		GEE^*_E	GEE_E	GEPSE^*_E	GEPSE_E	ML
Equi	$\vartheta = .2$	500	500	500	500	491
	$\vartheta = .8$	475	450	500	498	497
		GEE^*_A	GEE_A	GEPSE^*_A	GEPSE_A	ML
AR(1)	$\vartheta = .2$	496	495	500	499	—
	$\vartheta = .8$	500	500	500	500	—

Generally, for all models used, the relative bias tended to be larger for $N = 100$ than for $N = 500$, independently of the estimation approach. However, there were no systematic and significant deviations of the averaged estimates from the 'true' values for all estimators. Since the overall results therefore do not change whether standard deviations or root mean squared errors are considered, only the root mean squared errors are reported. Furthermore, given a simulated model, there was no systematic difference in the measure bias using the different estimation approaches. Since the results are similar for the different regressions parameters,

only the estimation results for $\beta_4 = -1.5$ and ϑ are reported.

Using the different approaches, the estimation results for an equicorrelation model with $N = 100$ and $N = 500$ blocks are given in Table 2 and 3, respectively. Note, that for the low 'true' correlation ($\vartheta = .2$) and $N = 100$ the ML estimates converged only in $s = 499$ cases. For the high 'true' correlation model ($\vartheta = .8$) and $N = 100$ the GEE_E^* estimates converged only in 499 and the GEE_E estimates converged only in 498 out of $s = 500$ cases.

Table 2: m, $\widehat{\text{sd}}$, rmse, sd/$\widehat{\text{sd}}$ and rej *for* $\hat{\beta}_4$ *and* $\hat{\vartheta}$ *for a model with* $N = 100$, $T = 5$, $\beta_4 = -1.5$ *and an equicorrelation structure with* $\vartheta = .2$ *and* $\vartheta = .8$ *over s replications*

m $\widehat{\text{sd}}$ rmse sd/$\widehat{\text{sd}}$ rej	GEE_E^*	GEE_E	GEPSE_E^*	GEPSE_E	ML
			$\vartheta = .2$		
	$s = 500$	$s = 500$	$s = 500$	$s = 500$	$s = 499$
	-1.511	-1.512	-1.511	-1.512	-1.511
	0.2130	0.2128	0.2130	0.2128	0.2141
$\hat{\beta}_4$	0.2124	0.2129	0.2124	0.2129	0.2132
	1.00	1.00	1.00	1.00	1.00
	0.04	0.042	0.04	0.042	0.05
	0.1946	0.1949	0.1944	0.1945	0.1952
	0.0787	0.0787	0.0796	0.0795	0.0791
$\hat{\vartheta}$	0.0821	0.0825	0.0816	0.0817	0.0807
	1.04	1.05	1.02	1.07	1.02
	0.066	0.07	0.06	0.06	0.05
			$\vartheta = .8$		
	$s = 499$	$s = 498$	$s = 500$	$s = 500$	$s = 500$
	-1.522	-1.522	-1.521	-1.522	-1.522
	0.1982	0.1901	0.1981	0.1901	0.1910
$\hat{\beta}_4$	0.2011	0.1905	0.2003	0.1897	0.1871
	1.01	1.00	1.01	0.99	0.97
	0.052	0.04	0.05	0.036	0.032
	0.8023	0.8040	0.7980	0.7997	0.7997
	0.0730	0.0691	0.0531	0.0523	0.0529
$\hat{\vartheta}$	0.0755	0.0730	0.0570	0.0572	0.0550
	1.04	1.06	1.07	1.10	1.04
	0.10	0.11	0.078	0.088	0.066

For $N = 100$ as well as for $N = 500$ blocks if $\vartheta = .2$, the differences in the measures m, $\widehat{\text{sd}}$ and rmse for the different estimators are negligible. However,

Table 3: m, $\widehat{\text{sd}}$, rmse sd/$\widehat{\text{sd}}$ *and* rej *for* $\hat{\beta}_4$ *and* $\hat{\vartheta}$ *for a model with* $N = 500$, $T = 5$, $\beta_4 = -1.5$ *and an equicorrelation structure with* $\vartheta = .2$ *and* $\vartheta = .8$ *over* $s = 500$ *replications*

m $\widehat{\text{sd}}$ rmse sd/$\widehat{\text{sd}}$ rej	GEE$^*_E$	GEE$_E$	GEPSE$^*_E$	GEPSE$_E$	ML
			$\vartheta = .2$		
$\hat{\beta}_4$	-1.505	-1.505	-1.505	-1.505	-1.505
	0.1026	0.1025	0.1026	0.1025	0.1026
	0.1005	0.1004	0.1005	0.1004	0.1006
	0.98	0.98	0.98	0.98	0.98
	0.044	0.044	0.044	0.044	0.04
$\hat{\vartheta}$	0.1985	0.1985	0.1985	0.1985	0.1986
	0.0350	0.0350	0.0350	0.0350	0.0349
	0.0345	0.0346	0.0344	0.0344	0.0344
	0.99	0.99	0.98	0.98	0.98
	0.052	0.054	0.05	0.05	0.048
			$\vartheta = .8$		
$\hat{\beta}_4$	-1.506	-1.505	-1.506	-1.505	-1.506
	0.0947	0.0912	0.0947	0.0912	0.0905
	0.0933	0.0875	0.0933	0.0875	0.0871
	0.98	0.96	0.98	0.96	0.96
	0.062	0.048	0.062	0.05	0.046
$\hat{\vartheta}$	0.8000	0.8003	0.8001	0.8004	0.8005
	0.0326	0.0309	0.0226	0.0225	0.0229
	0.0323	0.0312	0.0237	0.0237	0.0233
	0.99	1.01	1.05	1.05	1.02
	0.064	0.068	0.064	0.062	0.052

the picture is quite different for the models with high 'true' correlation $\vartheta = .8$. For the regression parameter β_4 the GEE$^*_E$ and GEPSE$^*_E$ estimators are the most inefficient estimators if the values of rmse and $\widehat{\text{sd}}$ are considered. Clearly, the GEE$_E$, GEPSE$_E$ and ML estimators for β_4 are more efficient. On the other hand, the most efficient estimators for the correlation structure parameter ϑ are the GEPSE$^*_E$, the GEPSE$_E$ and the ML estimator. The least efficient estimators are the GEE$^*_E$ and the GEE$_E$ estimators, with the GEE$_E$ being more efficient than the GEE$^*_E$ estimator. If both parameters are taken into consideration, the GEPSE estimator is the most efficient of the non-ML estimators and is very efficient relative to the ML estimator. This statement also holds if the other regression

parameters (not reported) are considered. Although to a smaller degree, the same pattern in the measures rmse and $\widehat{\text{sd}}$ also resulted if a moderate 'true' correlation ($\vartheta = .5$) was used.

Note, that there is no significant or systematic difference in the measures sd/$\widehat{\text{sd}}$ from unity for all estimators, indicating that there is no systematic under- or overestimation of the variances. Since all experiments were conducted with approximately 500 replications, the critical values for a test of the hypothesis of the proportions of rejections being 0.05 are approximately 0.05 ± 0.02 ($\alpha = 0.05$). The statistic rej lies outside this interval for all non-ML estimators only in the model with $N = 100$ and $\vartheta = .8$ for the parameter ϑ (see Table 2).

Guilkey and Murphy (1993) identified situations in which the ML estimator performed poorly on the measure sd/$\widehat{\text{sd}}$ for the regression parameter estimator. For an equicorrelation model with $N = 100$, $\vartheta = .25$ and $T = 2$ they found sd/$\widehat{\text{sd}} = .4$ for the ML regression parameter estimator using the DFP algorithm. To see whether the results for the non-ML estimators are better in this case, additional experiments were run using an equicorrelation model with $N = 100$, $\vartheta = .25$, $T = 2$ and exogenous variables and β as in the models described above.

For the non-ML estimators the values sd/$\widehat{\text{sd}}$ for all parameters were between 1.07 and 1.02. However, the ML estimates converged only in 460 out of $s = 500$ cases. The values sd/$\widehat{\text{sd}}$ for the regression parameters were between 0.99 and 1.04 for these $s = 460$ cases. On the other hand, the results for ϑ were not as good, i.e. sd/$\widehat{\text{sd}} = .92$ and m $= .2838$. Considering the estimation results for the non-ML estimators, this value for sd/$\widehat{\text{sd}}$ and the large bias can easily be explained: Whereas the ML estimator forces $\hat{\vartheta}$ to be positive — which is a consequence of the random effects model —, the non-ML estimators described in section 3 also allow negative values for $\hat{\vartheta}$. For this particular model, for example the GEPSE approach leads to negative values for $\hat{\vartheta}$ in 62 cases ($s = 499$). In 23 out of these 62 simulated datasets the ML estimates for ϑ converged to a positive value near zero. In 39 of these datasets and one dataset for which also the GEPSE estimates did not converge, the Newton-Raphson algorithm used together with a global strategy (Dennis and Schnabel, 1983) ended up with a non-negative definite matrix of second derivatives. Therefore, the variance of the ML estimates $\hat{\vartheta}$ is forced to be too small and the mean value too large.

In a different experiment we used exactly the same model as Guilkey and Murphy (1993), i.e. $N = 100$, $\vartheta = .25$, $T = 2$, $\beta_1 = 0$, $\beta_2 = 0$, weighting a normally distributed covariate, and $s = 500$, but again found the same general results as above.

To summarize, contrary to Guilkey and Murphy (1993) we found no systematic or significant overestimation of the variances of the ML regression parameter estimators for a model with $N = 100$, $\vartheta = .25$ and $T = 2$. However, the ML estimator of the random effects probit model forces $\hat{\vartheta}$ to be positive even if $\hat{\vartheta}$ should be negative, which is clearly possible in simulated datasets if the

'true' value of ϑ is low, the number of blocks is small and $T = 2$. This in turn leads a large value of sd/$\widehat{\text{sd}}$ for $\hat\vartheta$. Since the results of Guilkey and Murphy (1993) could not be replicated, it may only be speculated about the cause of their results for the regression parameter estimates. Clearly, it is not, as they suppose, that there is not sufficient information available for ϑ to be estimated accurately.

Overall, concerning the relative efficiency, the same pattern of results as in Table 2 and 3 can be derived for the non-ML estimators from Table 4 were the estimation results for a model with an AR(1) structure with $\vartheta = .2$ and $\vartheta = .8$ for $N = 100$ and $N = 500$ are given. When the 'true' value of ϑ is low ($\vartheta = .2$) the differences between the estimators in the measures rmse and $\widehat{\text{sd}}$ are negligible. If $\vartheta = .8$, the most efficient estimator (rmse and $\widehat{\text{sd}}$) is the GEPSE$_A$ estimator, whereas the least efficient estimator is the GEE$^*_A$ estimator if both parameters are considered. Again, this statement is true if the other regression parameters are considered. If a moderate value for ϑ was used ($\vartheta = .5$), the same pattern in the measures rmse and $\widehat{\text{sd}}$ resulted as for $\vartheta = .8$, although the differences were smaller.

Again, there are no significant and systematic differences in the measure sd/$\widehat{\text{sd}}$ for the different estimators. The statistic rej lies inside the interval 0.05 ± 0.02 for all estimators and all models.

From the results in Table 2 to Table 4 for moderate to high 'true' correlations it may be concluded for the non ML-estimators that — at least for the different models considered in this paper — using an individual covariance matrix in calculating the regression parameter estimates leads to more efficient estimators for β than using a covariance matrix composed of a correlation matrix identical for all blocks and individual variances — if the association structure is correctly specified. On the other hand, the use of pseudo-score equations lead to more efficient estimators for the correlation structure parameter than using the generalized estimating equations. The conclusion clearly is, that the GEPSE estimator is the most efficient non-ML estimator and is very efficient relative to the ML estimator. Beside these two 'main' effects of 'type of covariance matrix' and 'estimation method for the correlation structure parameter' there also seems to be an 'interaction' effect on the efficiency of $\hat\vartheta$: Whereas there is no difference in efficiency ($\widehat{\text{sd}}$ and rmse) of $\hat\vartheta$ using the GEPSE* and GEPSE approach, there is a clear difference using the GEE* and GEE approach, with the $\hat\vartheta$ estimator being more efficient using the GEE approach than using the GEE* approach. A possible explanation for this 'interaction' effect is that the pseudo-score equations for the estimation of ϑ are more robust with respect to inefficient estimators of β than the generalized estimating equations.

Overall, it may be concluded that the different estimators do not differ significantly or systematically in the statistics bias, sd/$\widehat{\text{sd}}$ and rej. They are, however, systematically different for moderate to high 'true' correlations if the statistics rmse and $\widehat{\text{sd}}$ are considered: The GEPSE$_{E/A}$ estimator is the most efficient non-

Table 4: m, $\widehat{\text{sd}}$, rmse, sd/$\widehat{\text{sd}}$ *and* rej *for* $\hat{\beta}_4$ *and* $\hat{\vartheta}$ *for a model with* $N = 100$ *and* $N = 500$, $T = 5$, $\beta_4 = -1.5$ *and an AR(1) structure with* $\vartheta = .2$ *and* $\vartheta = .8$ *over* $s = 500$ *replications*

m $\widehat{\text{sd}}$ rmse sd/$\widehat{\text{sd}}$ rej	$N = 100$							
	GEE^*_A	GEE_A	GEPSE^*_A	GEPSE_A	GEE^*_A	GEE_A	GEPSE^*_A	GEPSE_A
			$\vartheta = .2$				$\vartheta = .8$	
	-1.522	-1.522	-1.522	-1.522	-1.517	-1.519	-1.517	-1.519
	0.2150	0.2149	0.2150	0.2149	0.1979	0.1925	0.1979	0.1925
$\hat{\beta}_4$	0.2153	0.2150	0.2154	0.2151	0.2012	0.2003	0.2011	0.2001
	1.00	1.00	1.00	1.00	1.01	1.04	1.01	1.04
	0.044	0.04	0.044	0.04	0.048	0.06	0.048	0.058
	0.1995	0.1997	0.2001	0.2002	0.7951	0.7960	0.7955	0.7964
	0.0994	0.0993	0.1026	0.1025	0.0571	0.0547	0.0488	0.0484
$\hat{\vartheta}$	0.1016	0.1017	0.1018	0.1019	0.0569	0.0555	0.0480	0.0480
	1.02	1.03	0.99	1.00	0.99	1.01	0.98	0.99
	0.046	0.044	0.048	0.046	0.052	0.058	0.066	0.07

m $\widehat{\text{sd}}$ rmse sd/$\widehat{\text{sd}}$ rej	$N = 500$							
	GEE^*_A	GEE_A	GEPSE^*_A	GEPSE_A	GEE^*_A	GEE_A	GEPSE^*_A	GEPSE_A
			$\vartheta = .2$				$\vartheta = .8$	
	-1.505	-1.505	-1.505	-1.505	-1.501	-1.499	-1.501	-1.499
	0.1033	0.1033	0.1033	0.1032	0.0950	0.0925	0.0950	0.0925
$\hat{\beta}_4$	0.1011	0.1011	0.1011	0.1011	0.0940	0.0914	0.0940	0.0914
	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.99
	0.044	0.042	0.044	0.042	0.046	0.038	0.046	0.036
	0.1985	0.1985	0.1985	0.1985	0.7985	0.7985	0.7985	0.7986
	0.0443	0.0443	0.0445	0.0444	0.0244	0.0236	0.0206	0.0205
$\hat{\vartheta}$	0.0437	0.0437	0.0436	0.0436	0.0246	0.0241	0.0215	0.0215
	0.99	0.99	0.098	0.098	1.01	1.02	1.04	1.05
	0.04	0.042	0.05	0.05	0.048	0.05	0.068	0.068

ML estimator if compared to the $GEE^*_{E/A}$, $GEE_{E/A}$ and $GEPSE^*_{E/A}$ estimators. In the equicorrelation model it is a very efficient estimator relative to the ML estimator.

5   Discussion

In the present paper a mixed approach for the estimation of regression and correlation structure parameters in multivariate probit models is proposed, where the regression and correlation structure parameters are estimated using generalized estimating equations for the former and pseudo-score equations for the latter (GEPSE approach). Via a Monte Carlo experiment the corresponding estimator is compared with an estimator proposed by Qu et al. (1992) and Qu et al. (1994) (GEE approach) in finite samples. For an equicorrelation model both estimators are compared with the ML estimator of a random effects probit model. For both non-ML approaches two different types of 'working' covariance matrices for the estimation of the regression parameters are used: GEPSE and GEE use individual covariance matrices, whereas $GEPSE^*$ and GEE^* use individual variances and correlation matrices which are identical for all blocks.

The results of the Monte Carlo experiment show the $GEPSE^*$ estimator to be the most robust estimator, in the sense that the number of datasets for which the corresponding estimates converged was largest, closely followed by the GEPSE and the ML approach. In any case, convergence problems only occured in small samples. In general, the GEPSE approach led to very efficient regression and correlation structure parameters relative to the ML estimator, and is therefore recommended in applications if an equicorrelation cannot be assumed. Furthermore, since the ML estimator of the random effects probit model forces the correlation between the responses to be positive, the GEPSE approach may also be preferred to the ML estimator of the random effects probit model in small samples with a small number of observations within blocks and if a low 'true' correlation has to be assumed.

The results of the present study are, however, restricted to models as used in the present paper. If for example, special types of covariates are included into the regression model, then the results are expected to change accordingly (see e.g. Mancl and Leroux, 1996, or Spiess and Hamerle, forthcoming, for the GEE estimators as proposed by Liang and Zeger, 1986). Clearly, further research is needed to examine the finite sample properties of the GEPSE estimator under various standard and non-standard conditions not realized in the above Monte Carlo experiment.

Appendix

For the following see also Prentice (1988) and Küsters (1990). Let $\theta = (\beta', \vartheta')'$,

$$\mathbf{u}_1(\beta, \vartheta) = \sum_n \mathbf{A}_n' \Omega_n^{-1} \mathbf{e}_n$$

and

$$\mathbf{u}_2(\beta, \vartheta) = \sum_n \mathbf{B}_n' \mathbf{W}_n^{-1} \mathbf{v}_n,$$

then, using a Taylor expansion, $\sqrt{N}(\hat{\theta} - \theta_0)$ can — under some regularity conditions — be approximated by

$$-\left[\begin{array}{cc} N^{-1}\frac{\partial\, \mathbf{u}_1(\beta,\vartheta)}{\partial\, \beta} & N^{-1}\left(\frac{\partial\, \mathbf{u}_1(\beta,\vartheta)}{\partial\, \vartheta}\right)' \\ N^{-1}\left(\frac{\partial\, \mathbf{u}_2(\beta,\vartheta)}{\partial\, \beta}\right)' & N^{-1}\frac{\partial\, \mathbf{u}_2(\beta,\vartheta)}{\partial\, \vartheta} \end{array} \right]_{\theta=\theta_0}^{-1} \left[\begin{array}{c} N^{-1/2}\mathbf{u}_1(\beta,\vartheta) \\ N^{-1/2}\mathbf{u}_2(\beta,\vartheta) \end{array} \right]_{\theta=\theta_0}. \quad \text{(A.1)}$$

It can then be shown, that the functions $(N^{-1/2}\mathbf{u}_1'(\beta,\vartheta),\ N^{-1/2}\mathbf{u}_2'(\beta,\vartheta))'_{\theta=\theta_0}$ have an asymptotic normal distribution as $N \to \infty$ with mean zero and covariance matrix

$$\lim_{N\to\infty} N^{-1}$$
$$\left[\begin{array}{cc} \sum_n \mathbf{A}_n' \Omega_n^{-1} \mathrm{Cov}(y_n)\Omega_n^{-1}\mathbf{A}_n & \sum_n \mathbf{A}_n'\Omega_n^{-1}\mathrm{E}(\mathbf{e}_n, \mathbf{v}_n'\mathbf{W}_n^{-1})\mathbf{B}_n \\ \sum_n \mathbf{B}_n'\mathrm{E}(\mathbf{W}_n^{-1}\mathbf{v}_n, \mathbf{e}_n')\Omega_n^{-1}\mathbf{A}_n & \sum_n \mathbf{B}_n'\mathrm{E}(\mathbf{W}_n^{-1}\mathbf{v}_n, \mathbf{v}_n'\mathbf{W}_n^{-1})\mathbf{B}_n \end{array} \right]_{\theta=\theta_0}. \quad \text{(A.2)}$$

Again, under mild regularity conditions, it can be shown that as $N \to \infty$

$$N^{-1}\frac{\partial\, \mathbf{u}_1(\beta, \vartheta)}{\partial\, \beta'} \quad \xrightarrow{p} \quad -N^{-1}\left(\sum_n \mathbf{A}_n'\Omega_n^{-1}\mathbf{A}_n\right)_{\theta=\theta_0},$$

$$N^{-1}\frac{\partial\, \mathbf{u}_1(\beta, \vartheta)}{\partial\, \vartheta'} \quad \xrightarrow{p} \quad 0,$$

$$N^{-1}\frac{\partial\, \mathbf{u}_2(\beta, \vartheta)}{\partial\, \beta'} \quad \xrightarrow{p} \quad N^{-1}\left(\sum_n \frac{\partial^2\, l_n(\vartheta, \beta)}{\partial\, \vartheta\, \partial\, \beta'}\right)_{\theta=\theta_0}$$

and

$$N^{-1}\frac{\partial\, \mathbf{u}_2(\beta, \vartheta)}{\partial\, \vartheta'} \quad \xrightarrow{p} \quad N^{-1}\left(\sum_n \frac{\partial^2\, l_n(\vartheta)}{\partial\, \vartheta\, \partial\, \vartheta'}\right)_{\theta=\theta_0}.$$

Therefore, the matrix in (A.1) converges as $N \to \infty$ to

$$\lim_{N\to\infty} N^{-1}\left[\begin{array}{cc} -\sum_n \mathbf{A}_n'\Omega_n^{-1}\mathbf{A}_n & 0 \\ \sum_n \frac{\partial^2\, l_n(\vartheta,\beta)}{\partial\, \vartheta\, \partial\, \beta'} & \sum_n \frac{\partial^2\, l_n(\vartheta)}{\partial\, \vartheta\, \partial\, \vartheta'} \end{array} \right]_{\theta=\theta_0}. \quad \text{(A.3)}$$

Combining (A.3) with (A.2) and inserting estimates $\hat{\theta}$ for θ leads to the covariance matrix estimator described in section 3.

References

Ashford, J.R. & Sowden, R.R. (1970). Multi-variate probit analysis. *Biometrics, 26,* 535–546.

Bartholomew, D.J. (1987). Latent variable models and factor analysis. London: Charles Griffin & Co.

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.

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.

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

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, B, 54(1),* 3–40.

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

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49(1),* 115–132.

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. *Biomet-
rics, 48*, 1095–1102.

Qu, Y., Piedmonte, M.R. & Williams, G.W. (1994). Small sample validity of la-
tent variable models for correlated binary data. *Communications in Statis-
tics — Simulation and Computation, 23(1)*, 243–269.

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). *Regression models with correlated binary
response variables: A comparison of different methods in finite samples*
(Discussion Paper No. 10). München: SFB 386 der Ludwig-Maximilians-
Universität-München.

Spiess, M. & Hamerle, A. (forthcoming). On the properties of GEE estimators
in the presence of invariant covariates. *Biometrical Journal.*

Zeger, S.L. & 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 quad-
ratic exponential model. *Biometrika, 77*, 642–648.