Ziegler, Andreas; Arminger, Gerhard

Analyzing the Employment Status with Panel Data from the GSOEP: A Comparison of the MECOSA and the GEE1 Approach for Marginal Models

Vierteljahrshefte zur Wirtschaftsforschung

Provided in Cooperation with:
German Institute for Economic Research (DIW Berlin)

Suggested Citation: Ziegler, Andreas; Arminger, Gerhard (1995) : Analyzing the Employment Status with Panel Data from the GSOEP: A Comparison of the MECOSA and the GEE1 Approach for Marginal Models, Vierteljahrshefte zur Wirtschaftsforschung, ISSN 0340-1707, Duncker & Humblot, Berlin, Vol. 64, Iss. 1, pp. 72-80

Analyzing the Employment Status with Panel Data from the GSOEP

A Comparison of the MECOSA and the GEE1 Approach for Marginal Models*

by Andreas Ziegler** and Gerhard Arminger***

1. Introduction

Two different approaches for the analysis of dichotomous panel data have been well developed in the last years, that is the GEE approach of Liang and Zeger (1986) and the approach of estimating the models of Heckman (1981a).

First, a special case of the model Heckman proposed in 1981 is embedded into the Mean and Covariance Structure Model for non-metric dependent variables which has been introduced by Muthén (1984) and extended by Küsters (1987) and Schepers and Arminger (1992). The model parameters can be estimated with the assumption of multivariate normality of the error terms. Special problems of non-metric panel data such as time dependent variances, unobserved heterogeneity and serial correlation can be directly solved by models using the polychoric covariance matrix (Arminger 1992).

Second, the approach of the generalized estimating equations (GEE) proposed in a series of papers by Liang and Zeger (Liang and Zeger, 1986; Zeger and Liang, 1986; Zeger, 1988) is discussed. Here, the original model of Liang and Zeger (1986) is considered and termed as GEE1. Like in the Mean and Covariance Structure Analysis approach, the mean structure for the dependent dichotomous variable is formulated with a distributional assumption such as in the probit or the logit model. If the mean structure is correctly specified and first order identifiable Pseudo Maximum Likelihood (PML) estimation developed by Gourieroux et al. (1984) is used a consistent estimate of the parameter vector may be obtained.

Third, both estimation strategies are illustrated in an empirical example: The employment status of 1246 men of the GSOEP from 1985 to 1988 dependent on explanatory variables such as age, professional education, family status and history of unemployment at the first wave is considered. The microeconomic specification of models for the employment status is found in Flaig et al. (1993). However, these authors concentrate on models for state dependence, while we consider models without state dependence.

2. The Heckman Model for Dichotomous Variables

Heckman[1] has proposed a general model for the analysis of dichotomous panel data. He considers the unobserved variable y_{it}^*, $i = 1,...,K$, $t=1,...,T$, where i denotes the individual and t denotes a sequence of (fixed) equispaced time points:

$$y_{it}^* = \mu_{it} + \epsilon_i^* = (\epsilon_{i1}^*,...,\epsilon_{iT}^*)^T,\ \epsilon_i^* \sim N(0, \Sigma) \quad (1)$$

The latent metric variable y_{it}^* may be considered as a propensity or utility. In the context of models for unemployment it denotes the propensity for individual i to be unemployed in period t. It is decomposed into a systematic part μ_{it} and an error term ϵ_{it}^*. Observable is a dependent dummy variable y_{it} which is connected to y_{it}^* through a dichotomous threshold model:

$$y_{it} = \begin{cases} 1 \text{ if } y_{it}^* > 0 \\ \\ 0 \text{ if } y_{it}^* \leq 0 \end{cases} \quad (2)$$

Note, that the thresholds are 0 for every single wave. This restriction is only necessary to avoid identification problems, if the waves are considered separately. But the thresholds should be set equal across panel waves, otherwise the meaning of the categories varies over time.

The systematic part μ_{it} follows a linear model:

$$\mu_{it} = x_{it}^T \beta \quad (3)$$

$x_{it}^T \beta$ represents the effect of possibly time varying explanatory variables x_{it}. In this model the parameter vector β is time constant. The random variables (y_i, x_i) are assumed to be independent and identically distributed which corresponds to simple random sampling from a population. In addition, it is assumed that strong exogeneity of the error terms ϵ_{it} holds, i.e. the ϵ_{it} are uncorrelated with past, present and future explanatory variables. No assumption about stationarity of the disturbances is imposed.

Equation (3) can be extended by including duration dependence (inclusion of $\prod_{t-s,s\geq 1} y_{is}$), state dependence (inclusion of $y_{i,t-s,}\ s \geq 1$) and habit persistance (inclusion of y_{it}^* in the regression)[2]. However, these models are not considered here, but see Flaig et al. (1993).

---

* This paper was prepared for The 1993 Conference of German Socio-Economic Panel Study Users: Using Panel Data to Answer Policy Questions, June 7 and 8, 1993 in Berlin. The authors are greatful to Michael Lechner for comments on an earlier draft of this paper.

** Philipps-Universität Marburg.

*** Bergische Universität — GH Wuppertal.

[1] Heckman (1981a), ch. 3.3.

[2] Heckman (1981a).

Two problems that arise are connected to the inclusion of state dependence or duration dependence. First, the initial states have to be known. Either they have to be fixed outside the model or they can be taken into account as discussed in Heckman (1981b) or Arminger (1992). Second, the assumption of strong exogeneity is usually violated if lagged dependent variables are included into equation (3). Models without state dependence are called marginal models.

The values are collected in the following vectors:

$$y_i^* = (y_{i1}^*,\ldots,y_{iT}^*)^T, \quad (T\,x\,1),$$
$$x_i = (x_{i1}^T,\ldots,x_{iT}^T)^T, \quad (T \cdot p\,x\,1) \tag{4}$$

The model mentioned above is extended by allowing time varying parameters β_t.

Four waves are considered in the data from the GSOEP. Equations (1) and (3) can be written as:

$$y_{it}^* = \varkappa_t + x_{it}^T\beta_\tau + \epsilon_{it}^*, \quad t = 1, 2, 3, 4 \tag{5}$$

Using matrix notation equation (5) can be collected to

$$y_i^* = \varkappa + \Gamma x_i + \epsilon_i^* \tag{6}$$

with vectors

$$y_i^* = \begin{pmatrix} y_{i1}^* \\ y_{i2}^* \\ y_{i3}^* \\ y_{i4}^* \end{pmatrix}, \quad x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ x_{i4} \end{pmatrix}, \quad \varkappa = \begin{pmatrix} \varkappa_1 \\ \varkappa_2 \\ \varkappa_3 \\ \varkappa_4 \end{pmatrix} \text{ and } \epsilon_i^* = \begin{pmatrix} \epsilon_{i1}^* \\ \epsilon_{i2}^* \\ \epsilon_{i3}^* \\ \epsilon_{i4}^* \end{pmatrix} \tag{7}$$

and a parameter matrix

$$\Gamma = \begin{pmatrix} \beta_1^T & 0 & 0 & 0 \\ 0 & \beta_2^T & 0 & 0 \\ 0 & 0 & \beta_3^T & 0 \\ 0 & 0 & 0 & \beta_4^T \end{pmatrix}. \tag{8}$$

After the formulation of the simultaneous equation model, a model for the error term is specified. Usually ϵ_{it}^* is separated into two terms

$$\epsilon_{it}^* = \alpha_i + \epsilon_{it}, \tag{9}$$

where α_i denotes the subject specific error term[3] which does not vary over time and may be interpreted as unobserved heterogeneity. The values of α_i are considered as random effects such that $\alpha_i \sim N(0,\sigma_\alpha^2)$. If the α_is are treated as fixed effects it is possible to eliminate the effect in linear models by taking the first differences. In logit models the α_is may be eliminated by conditioning on a sufficient statistic as shown in Hamerle and Ronning (1995). In probit models it is not possible to eliminate them, hence they are not treated as fixed.

Note that restrictions about the variances must be set up in dichotomous panel analysis to avoid identification problems[4].

More generally, ϵ_{it} has a serial structure — as in the $AR(1)$ — or the structure of a factor model. Heckman (1981a) discusses the interpretation of general one factor schemes. The most general form is to assume that there is no specific structure of $V(\epsilon_i^*)$. Under the assumption of normal distributed ϵ_i^*these models can be estimated by using Maximum Likelihood Estimation[5]. For the data of the GSOEP a covariance matrix with no specific structure is assumed, but for the comparison with the GEE1 all variances are set to 1, so that

$$V(\epsilon_{it}^*) = 1, \text{ and } Cov(\epsilon_{it}^*, \epsilon_{it'}^*) = \sigma_{tt'}^2,\ t,t' = 1, 2, 3, 4\ t \neq t'. \tag{17}$$

3. Embedding Heckman's Model into the Mean and Covariance Structure Approach

For general mean and covariance structures it is assumed that a r x 1 vector of y_i^* latent dependent variables follows a multivariate normal distribution with conditional mean and covariance:

$$E(y_i^*|x_i) = \gamma(\theta) + \Pi(\theta)x_i$$
$$V(y_i^*|x_i) = \Sigma(\theta) \tag{11}$$

Here, panel data are analysed, so r equals T. $\gamma(\theta)$ is a r x 1 vector of regression constants and $\Pi(\theta)$ is a r x p matrix of reduced form regression coefficients. x_i is a p x 1 vector of explanatory variables. $\Sigma(\theta)$ is the r x r covariance matrix of the errors of the reduced form. θ is the $\tilde{q}$ x 1 vector of structural parmeters to be estimated. The reduced form parameters $\gamma(\theta)$, $\Pi(\theta)$, $\Sigma(\theta)$ are continuously differentiable functions of a common vector θ. This model can be extended so that $\Sigma(\theta)$ is a function of the explanatory variables as well.

One typical example is the simultaneous equation system described in equation (6) with the reduced form parameters

$$\gamma(\theta) = \varkappa, \quad \Pi(\theta) = \Gamma \quad \text{and} \quad \Sigma(\theta) = \Omega.$$

The estimation of the structural parameter vector from the observed data vector y_i proceeds in three stages. Algorithmic details are found in Schepers (1991). Computation of the estimates with the MECOSA program is described in Schepers and Arminger (1992).

In the first stage, the reduced form coefficients γ_t, Π_t and the reduced form error variance σ_t^2 of the t^{th} equation are estimated using marginal maximum likelihood. This first stage is the estimation of the parameters of the mean structure without restrictions of y_{it}^*, *given* x_{it}.

In the second stage, the covariances of the error terms in the reduced form equations are estimated. In this stage, the covariances are estimated without parametric restrictions.

All estimates of the first two stages are collected in a vector $\hat{\xi}_K$ which depends on the number of individuals (sample size). The asymptotic covariance of $\hat{\xi}_K$ is denoted by Λ. For the third stage the vector $\hat{\xi}_K(\theta)$ is written as a function of the structural parameters of interest, collected in a parameter vector θ. Küsters (1987) shows that

$$\hat{\xi}_K(\theta) \overset{a}{\sim} N(\xi(\theta), \Lambda), \tag{13}$$

---

[3] Liang et al. (1988).

[4] Arminger (1992), Heckman (1981a).

[5] Heckman (1981a).

where $\overset{a}{\sim}$ denotes „asymptotically distributed as". The various elements of the asymptotic covariance matrix can be found there. It is difficult to derive θ because the estimates of the second stage depend on the estimates of the first stage.

The parameter vector θ is estimated by using minimum distance estimation (MDE) with weight $\hat{\Lambda}_K$ after computing a strongly consistent estimate $\hat{\Lambda}_K$ of Λ:

$$Q_K(\theta) = (\hat{\xi}_K(\theta) - \xi_K(\theta))^T \hat{\Lambda}_K^{-1} (\hat{\xi}_K(\theta) - \xi(\theta)) \tag{14}$$

If the model is correctly specified

$$Q_K(\theta) \overset{a}{\sim} X^2_{r-\tilde{q}}, \tag{15}$$

since $\hat{\Lambda}_K$ is strongly consistent for Λ.

To analyse the data from the GSOEP the program MECOSA is used which follows these three estimation steps. MECOSA is implemented in GAUSS[6] and many matrix features of this programming language are used to estimate the parameters.

The MECOSA approach is not restricted to dichotomous dependent variables. Within MECOSA, the models discussed in this section may be extended to deal with metric, censored metric and / or ordered categorical dependent variables.

4. The GEE1 Approach

The general model of Heckman is specified in terms of a latent variable vector y_i^* with a linear model $E(y_i^* | x_i) = \gamma(\theta) + \Pi(\theta) x_i$ for the expected value of y_i^* given x_i and the conditional covariance $V(y_i^* \mid x_i) + \Sigma(\theta)$ of y_i^* given x_i. An essential assumption is the multivariate normality of y_i^* given x_i with correctly specified mean and covariance.

A different approach to estimate the parameters of a general mean structure has been proposed by Liang and Zeger (Zeger and Liang, 1986; Liang and Zeger, 1986; Zeger, 1988). This approach has primarily been used in biometrics. There are two important differences to the Heckman model. First, the notion of multivariate normality of y_i^*, given x_i, is given up. It is replaced by the weaker condition that y_{it}^* given x_{it}, follows a univariate normal or logistic distribution. Second, only the parameters of the mean structure are of interest. The assumptions about the covariance matrix of y_{it}^*, given x_i are replaced by assumptions about the covariance matrix of the observed vector y_i, given x_i. Parameters that correspond to the assumption about the covariance matrix of y_i, given x_i, are collected in the so called „working" covariance matrix.

Here, the original model of Zeger and Liang (1986) is considered:

$$y_{it} = \mu_{it}(\theta) + \epsilon_{it}, \quad t = 1,...,T_i, \quad i = 1,...,K \tag{16}$$

$$E(y_{it} | x_{it}) = \mu_{it}(\theta), \quad \text{and} \quad \mu_{it}(\theta) = \mu_{it}(x_{it}, \theta), \tag{17}$$

$$V(y_{it} | x_{it}) = g(\mu_{it}) \cdot \phi \tag{18}$$

$$E(\epsilon_{it} | x_{it'}) = 0, \quad t' = 1,...,T_i \text{ (strong exogeneity)} \tag{19}$$

In this model, only the first two moments of the marginal density are specified. For our application it is assumed that y_i is dichotomous with values 1 and 0.

Equation (19) implies that the mean structure is correctly specified. The function g from equation (18) is known as the variance function in Generalized Linear Models. It is also assumed that the variances of the model are correctly specified. However, the correlations of y_{it} and $y_{it'}$ are not yet specified.

Typical examples for mean structures in dichotomous models are the logit model

$$\mu(x_{it}, \theta) = \frac{\exp(x_{it}^T \theta)}{1 + \exp(x_{it}^T \theta)} \tag{20}$$

and the probit model

$$\mu(x_{it}, \theta) = \phi(x_{it}^T \theta), \tag{21}$$

where ϕ denotes the standard normal distribution function. In these cases the variance function is given by

$$g(\mu(x_{it}, \theta)) = \mu(x_{it}, \theta) \cdot [1 - \mu(x_{it}, \theta)] \tag{22}$$

and the dispersion parameter $\phi = 1$. For the following considerations it is assumed that the mean structure is first order identifiable, that is $\mu(\theta_1) = \mu(\theta_2) \overset{(a.s.)}{\rightarrow} \theta_1 = \theta_2$.

Usually, the values are collected in vectors and matrices: $y_i = (y_{i1},...,y_{iT_i})^T$, $x_i = (x_{i1},...,x_{iT_i})^T$, and so forth. The data (y_i, x_i) are assumed to be independent and identically distributed. The $T_i \times T_i$ covariance matrix of ϵ_i is denoted by Ω_i. No assumptions about the structure of ω_i need to be established.

Now it is assumed that the vector y_i is multivariate normal distributed with identity covariance matrix: $y_i \sim N(\mu_i, I_{T_i \times T_i})$. This assumption is certainly wrong for dichotomous variables, but it is only used to derive the estimating equations Liang and Zeger introduced. One may use the theory of Pseudo Maximum Likelihood (PML) estimation developed by Gourieroux et al. (1984) to show, that the maximization of the normal distribution pseudo log-likelihood

$$L(\theta) = \sum_{i=1}^{K} \ln \varphi(y_i | \mu_i(\theta), I_{T_i \times T_i}) \tag{23}$$

yields a consistent estimator of θ which is denoted by $\hat{\theta}$. $\varphi(y | \mu, \Sigma)$ is the density of a normal distribution with $E(y) = \mu$ and $V(y) = \Sigma$. Here, the maximization of $L(\theta)$ is equivalent to the minimization of a quadratic form

$$Q(\theta) = \sum_{i=1}^{K} (y_i - \mu_i(\theta))^T I_{T_i \times T_i} (y_i - \mu_i(\theta)). \tag{24}$$

This is a minimum distance estimation for $\mu_i(\theta)$ with weight $I_{T_i \times T}$ because the euclidian distance between y_i and $\mu_i(\theta)$ has to be minimized. The asymptotic covariance matrix $V(\hat{\theta})$ of the PML estimator θ depends on the first and second order derivatives of $\ln \varphi(y_i | \mu_i(\theta), I)$ with respect to θ and the true unknown covariance matrix Ω_i, which has not been specified.

[6] Gauss (1982).

A strongly consistent estimator of $V(\hat{\theta})$ has the typical sandwich form[7]:

$$\hat{V}(\hat{\theta}) = \hat{C}(\hat{\theta})^{-1}\hat{B}(\hat{\theta})\hat{C}(\hat{\theta})^{-1} \tag{25}$$

with elements

$$\hat{C}(\hat{\theta}) = \sum_{i=1}^{K} D_i^T(\hat{\theta})D_i(\hat{\theta}) \tag{26}$$

and

$$\hat{B}(\hat{\theta}) = \sum_{i=1}^{K} D_i^T(\hat{\theta})\left(y_i - \mu_i(\hat{\theta})\right)\left(y_i - \mu_i(\hat{\theta})\right)^T D_i(\hat{\theta}) \tag{27}$$

where $D_i(\hat{\theta}) = \left.\frac{\partial \mu_i(\theta)}{\partial \theta^T}\right|_{\theta=\hat{\theta}}$. The advantage of this estimator is that it does not depend on the correct assumption of multivariate normality and on the assumption of correct specification of the covariance structure Ω_i. Only the correct specification of the mean structure and the i.i.d. distribution of the observations (y_i, x_i) is necessary.

To overcome the inefficiency of the estimator $\hat{\theta}$, Zeger and Liang (1986) introduced (a „working" correlation matrix R_i of ϵ_i which may be thought of as an approximation of the true correlation matrix. In the dichotomous case, one may assume, that the variance function has the form of equation (22), so that

$$V(\xi_{it}|x_{it}) = \mu_{it}(\theta)[1 - \mu_{it}(\theta)]. \tag{28}$$

If A_i is $\text{diag}(V(y_{it} \mid x_{it}))_{t=1,\ldots,T_i}$, then the „working" covariance matrix of Σ_i *of* ϵ_i is given by

$$\Sigma_i = A_i^{1/2}R_iA_i^{1/2}. \tag{29}$$

The „working" correlation matrix R_i may depend on an additional parameter vector α that is treated as nuisance. The number of nuisance parameters and the estimator of α depend on the choice of R_i. Therefore: $R_i = R_i(\alpha)$ and $\Sigma_i = \Sigma_i(\theta, \alpha)$. In most applications the „working correlation" is set to be the same for all individuals i, that is $R_i(\alpha) = R(\alpha)$. The parameter vector α may be computed using simple least squares methods on the empirical correlation matrix which may be computed from y_i, x_i and the PML estimator $\hat{\theta}$. Note that the „working" covariance need not to be equal to the true covariance matrix Ω_i. But if $\Sigma_i(\alpha, \theta)$ is equal or very close to Ω_i the estimator $\tilde{\theta}$ is more efficient than $\hat{\theta}$. This has been demonstrated by Zeger and Liang (1986) in Monte Carlo simulations and has been proven by Gourieroux et al. (1984) in the context of PML estimation. The estimator $\tilde{\theta}$ is found by using the PML estimation with the possibly wrong density function $\varphi(\mu_i, \Sigma_i(\hat{\alpha}, \hat{\theta}))$ with fixed $\Sigma_i(\hat{\alpha}, \hat{\theta})$ or — equivalently — minimizing the sum of Mahalanobis distances:

$$U(\theta) = \sum_{i=1}^{K} (y_i - \mu_i(\hat{\theta}))^T \Sigma_i(\hat{\alpha}, \hat{\theta})^{-1}(y_i - \mu_i(\hat{\theta})) \tag{30}$$

Derivatives of $U(\theta)$ with respect to θ yields the Generalized Estimating Equations:

$$\frac{\partial U(\theta)}{\partial \theta} = \sum_{i=1}^{K} D_i^T(\theta)\,\Sigma_i(\hat{\alpha}, \hat{\theta})^{-1}(y_i - \mu_i(\theta)) = 0 \tag{31}$$

Note that the „working" covariance matrix is treated as fixed in eq. (31) because the values have been estimated by using and $\hat{\alpha}$ and $\hat{\theta}$.

A consistent estimator of the asymptotic covariance matrix of the GEE1 estimator $\tilde{\theta}$ is given by:

$$\hat{V}(\tilde{\theta}) = \hat{C}(\tilde{\theta})^{-1}\hat{B}(\tilde{\theta})\hat{C}(\tilde{\theta})^{-1} \tag{32}$$

If $\hat{\Sigma}_i$ denotes $\Sigma_i(\hat{\alpha}, \hat{\theta})$, then the matrices $\hat{C}(\tilde{\theta})$ and $\hat{B}(\tilde{\theta})$ are defined by:

$$\hat{C}(\tilde{\theta}) = \sum_{i=1}^{K} D_i^T(\tilde{\theta})\hat{\Sigma}_i^{-1} D_i(\tilde{\theta}) \tag{33}$$

and

$$\hat{B}(\tilde{\theta}) = \sum_{i=1}^{K} D_i^T(\tilde{\theta})\,\hat{\Sigma}_i^{-1}\left((y_i - \mu_i(\tilde{\theta}))\right)\left(y_i - y_i(\tilde{\theta})\right)^T \hat{\Sigma}_i^{-1} D_i(\tilde{\theta}) \tag{34}$$

If the first two moments are correctly specified, then asymptotically $C(\tilde{\theta}) \stackrel{\alpha}{=} B(\tilde{\theta})$ and equation (32), called „robust variance estimation", reduces to

$$V(\tilde{\theta}) = C(\tilde{\theta})^{-1} \tag{35}$$

which is termed „model based variance estimation".

There is a second — more intuitive — way to derive the generalized estimation equations: Consider the normal equations for θ of a generalized linear model:

$$D^T V(y)^{-1}(y - \mu) = 0 \tag{36}$$

Usually, in the iterative estimating procedure $V(y)$ is substituted by $diag(\hat{\sigma}_{it}^2)$ where $\hat{\sigma}_{it}^2$ is calculated using the variance function. To increase efficiency a „working" variance matrix of block diagonal form should be used instead of a diagonal variance matrix in the case of panel data. But it is difficult to derive the asymptotic properties of this estimate.

There are several specific choices of the „working" correlation matrix $R(\alpha)$. Each leads to a different analysis. Examples are described in detail in Liang and Zeger (1986).

A generalization of the GEE1, referred to GEE2, has been developed by Zhao and Prentice (1990). There the PML approach is used to estimate the parameter vector and the parameters of the correlation matrix simultaneously. The estimation is based on quadratic exponential families rather than linear exponential families. This approach has only been used for dichotomous data because the computational effort is high. An overview of generalizations concerning the generalized estimating equations approach is given in Davis (1991) and Ziegler (1994a).

The algorithm to solve the GEE1 (eq. 31) is an iterated two step procedure: In the first step a modified FISHER-SCORING procedure is used for the estimation of θ. In the second step moment estimation is used for the estimation of α. Given the current estimates $\hat{R}_{(j)}$, and $\hat{\alpha}_{(j)}$, the following iterative procedure is used for estimating θ:

[7] White (1992).

$$\hat{\theta}_{j+1} = \left(\sum_{i=1}^{K} D_i^T(\hat{\theta}_{(j)})\, \hat{\Sigma}_{i(j)}^{-1}\, D_i(\hat{\theta}_{(j)})\right)^{-1} \left(\sum_{i=1}^{K} D_i^T(\hat{\theta}_{(j)})\, \hat{\Sigma}_{i(j)}\, \{ D_i(\hat{\theta}_{(j)}) + S_i(\hat{\theta}_{(j)})\}\right) \quad (37)$$

where $\hat{\Sigma}_{i(j)} = \Sigma_i(\hat{\theta}_{(j)}, \hat{\alpha}_{(j)})$, $S_i(\hat{\theta}_{(j)}) = y_i - \hat{\mu}(\hat{\theta}_{(j)})$.

A general difference in the choice of the structure of R must be pointed out concerning the computational effort: Either a fixed „working" covariance matrix, say R_0 is used or a correlation matrix that is updated while iterating. The theorems of Gourieroux and Monfort (1993) only show the asymptotic consistence and normality of $\hat{\theta}$ given a consistent estimate for R_i. But it is more efficient to use the above two step procedure for a stable variance estimation as shown for linear models by Carroll et al. (1988). Problems arise in estimating R, if the number of observations varies across individuals. In this case, may $\hat{R}$ occur to be not positive definite[8]. This problem can be avoided by using the EM-Algorithmus[9] or the approach described in Ziegler (1994a).

McDonald (1993) criticizes the approach of Liang and Zeger (1986) because of the small sample properties of the GEE1 estimator, the high computational effort and the boundary constraints on the additional parameter α: For example, a $2x2$ table is considered with entries π_{11}, π_{12}, π_{21} and π_{22}. Therefore, $\mu_A = \pi_{21} + \pi_{22}$, $\mu_B = \pi_{12} + \pi_{22}$. The correlation between factors A and B is:

$$\varrho = \frac{\pi_{22} - \mu_A \mu_B}{\sqrt{\mu_A(1-\mu_A) \cdot \mu_B(1-\mu_B)}} \quad (38)$$

A resulting constraint is: $max(0,\ \mu_A + \mu_B - 1) \leq \pi_{22} \leq min(\mu_A,\ \mu_B)$. For instance, if $\mu_A = \mu_B = 0.3$, then the correlation ϱ is bounded through the inequality: $-0.43 \leq \varrho \leq 1$.

5. A Marginal Model for Employment Status

The employment status of a cohort of 1246 men of the GSOEP from 1985 to 1988 is considered as dependent variable over four waves. The codification is: 0=employed, 1=unemployed. A simple model for the disposition to become unemployed is used. The model does not include interactions. Flaig et al. (1993) specify different models for the employment status and discuss the relevant variables in the GSOEP data set.

The following variables are used as exogenous variables for waves 1 to 4: ALD=duration of unemployment between 1974 and 1984 in months, ALDSQ=AALD squared and divided by 100, ALH=frequency of unemployment between 1974 and 1984, ALT=age in years, ALTSQ=ALT squared and divided by 100, GZ=1 if a person is severely handicapped and 0 otherwise, BB1=1 if a person has finished some professional education and 0 otherwise, BB2=1 if a person has a university degree and 0 otherwise, BST2=1 if a person is a white collar employee and 0 otherwise, FST (Family status)=1 if a person is married and 0 otherwise.

Effects like economic situation are taken into account by including a constant for every single wave. The variable ALH could be updated for the years after 1984 and other variables like length of unemployment 1.5 years before and 0.5 years after the last interview could also be included, but in this case the assumption of strong exogeneity may be violated. Therefore, these variables are not included in the marginal model.

Note, that equation (5) allows time varying coefficients which can be restricted later. Therefore, separate probit models — for each wave — are estimated in the first two estimation stages. Marginal (univariate) probit models are as well computed using the program GEE1[10] for the comparison with the MECOSA-results. Indeed, this can be considered as a first step like in the Mean and Covariance Structure Analysis approach because it is possible to compute a „working" correlation matrix using the marginal — probably time varying — estimates for θ_t[11]. For this, the empirical correlation matrix has to be calculated using Pearson residuals after finishing the marginal estimation.

The estimations of the univariate probit models using MECOSA are given in Table 1 and the GEE1 results in Table 2. Note, that the parameter estimations are the same for MECOSA and the GEE1. Only the z-values differ slightly because the variances (and standard errors) are calculated using different algorithms.

The variables duration of unemployment, age, age squared, being severely handicapped and being a white collar employee are significant at the 5% test level over all four waves. The signs of the parameter coefficients are all in the expected direction.

The coefficients seem to vary across the waves. Only one difference arises: Usually, the z-values calculated by MECOSA are smaller than the values calculated by GEE1. The MECOSA and GEE1 results were compared with univariate probit models calculated by GLIM 3.77 (1985). The estimated GEE1 standard deviations lie between the MECOSA and the GLIM results. The differences arise because of numerical inaccuracy.

However, there are great differences in the correlations between the Mean and Covariance Structure Model and the GEE1 model. All correlations calculated by GEE1 are smaller than the MECOSA correlations by factors that vary between $\frac{1}{1.8}$ and $\frac{1}{3.7}$. This result is plausible: In MECOSA the correlations of the lagged endogenous variables are considered while GEE1 computes the correlations between the observed values. And it was shown (cf. eq. 38) that the correlations for binary outcomes are bounded.

The results for the partially restricted probit model for unemployment status calculated by MECOSA are presen-

[8] Davis (1991).

[9] Dempster et al. (1977).

[10] Ziegler (1994b).

[11] Stram et al. (1998).

Table 1

Marginal probit models for unemployment status (and z-values for MECOSA in parenthesis)

Explanatory variables	Wave 1	Wave 2	Wave 3	Wave 4
CONST	3.088 (1.599)	3.249 (1.758)	3.499 (1.923)	2.947 (1.485)
ALD	0.108 (5.173)	0.091 (4.306)	0.065 (3.686)	0.084 (5.548)
ALDSQ	−0.086 (−1.785)	−0.074 (−1.409)	−0.057 (−1.477)	−0.057 (−1.670)
ALH	−0.045 (−0.731)	0.050 (0.976)	0.071 (1.386)	0.009 (0.178)
ALT	−0.246 (−2.237)	−0.266 (−2.656)	−0.259 (−2.701)	−0.240 (−2.351)
ALTSQ	0.306 (2.125)	0.350 (2.780)	0.328 (2.755)	0.318 (2.571)
GZ	0.713 (3.493)	0.652 (3.339)	0.776 (4.264)	0.724 (3.950)
BB1	−0.208 (−1.214)	−0.113 (−0.764)	−0.101 (−0.691)	−0.330 (−2.236)
BB2	−0.010 (−0.043)	−0.299 (−1.154)	−0.120 (−0.477)	−0.194 (−0.850)
BST2	−0.397 (−2.048)	−0.391 (−2.251)	−0.396 (−2.131)	−0.374 (−2.200)
FST	−0.230 (−1.187)	−0.115 (−0.625)	−0.309 (−1.835)	−0.253 (−1.475)
Loglikelihood	−193.691	−247.436	−252.445	−254.662
Correlations				
Wave 1	1.000			
Wave 2	0.824	1.000		
Wave 3	0.640	0.762	1.000	
Wave 4	0.545	0.645	0.888	1.000

ted in Table 3. There a model with equal coefficients ALD to FST was estimated. The results for the GEE1 are shown in Table 3, too, classified by unspecified and user specified variances. The user specified „working" correlation matrix is presented in Table 2, the correlation matrices calculated by MECOSA and by GEE1 using the option „unspecified working" correlation matrix can be found in Table 4.

Only one difference appears between MECOSA and GEE1 outcomes: In MECOSA the variable family status turned out to be significant while in GEE1 the family status turned out to be significant at the 5% test level.

The X^2 statistic of step 3 in MECOSA has a value of 35.83 with 30 degrees of freedom. Therefore the null hypothesis of proportionality of the parameters cannot be rejected at the 5% test level. Note that standard errors and z-values for the correlations can also be computed in MECOSA, while this is not possible using the GEE1 approach.

For the estimations a 486–50 MHz IBM computer with 8 MB RAM, MS-DOS 6.0 and GAUSS-386i-VM 3.0.1 (Rev 25) has been used. MECOSA is written for the GAUSS 2.x version requiring an XT with minimum 640 KB RAM. The program GEE1 requires a 386 or 486 processor and GAUSS

Table 2

Marginal probit models for unemployment status (and z-values for GEE1 in parenthesis)

Explanatory variables	Wave 1	Wave 2	Wave 3	Wave 4
CONST	3.088 (1.764)	3.249 (1.923)	3.499 (2.003)	2.947 (1.548)
ALD	0.108 (6.259)	0.091 (5.088)	0.065 (3.835)	0.084 (4.910)
ALDSQ	−0.086 (−2.168)	−0.074 (−1.797)	−0.057 (−1.579)	−0.057 (−1.463)
ALH	−0.045 (−0.831)	0.050 (0.892)	0.071 (1.320)	0.009 (0.168)
ALT	−0.246 (−2.526)	−0.266 (−2.948)	−0.259 (−2.850)	−0.240 (−2.517)
ALTSQ	0.306 (2.384)	0.350 (3.049)	0.328 (2.901)	0.318 (2.777)
GZ	0.713 (3.408)	0.652 (3.407)	0.776 (4.193)	0.724 (3.860)
BB1	−0.208 (−1.323)	−0.113 (−0.813)	−0.101 (−0.733)	−0.330 (−2.435)
BB2	−0.010 (−0.038)	−0.299 (−1.076)	−0.120 (−0.455)	−0.194 (−0.761)
BST2	−0.397 (−2.026)	−0.391 (−2.268)	−0.396 (−2.275)	−0.374 (−2.112)
FST	−0.230 (−1.319)	−0.115 (−0.692)	−0.309 (−1.993)	−0.253 (−1.522)
Correlations				
Wave 1	1.000			
Wave 2	0.354	1.000		
Wave 3	0.215	0.230	1.000	
Wave 4	0.146	0.216	0.495	1.000

Table 3

Partially restricted probit model for unemployment status (z-values in parenthesis)

Explanatory variables	MECOSA estimations[1]		GEE1 estimations[2]		GEE1 estimations[3]	
CONST (Wave 1)	3.695	(3.003)	3.726	(3.162)	3.731	(3.180)
CONST (Wave 2)	3.660	(2.937)	3.623	(3.130)	3.625	(3.143)
CONST (Wave 3)	3.838	(3.091)	3.347	(2.916)	3.354	(2.933)
CONST (Wave 4)	3.708	(2.975)	3.421	(2.907)	3.425	(2.920)
ALD	0.084	(7.224)	0.092	(7.999)	0.091	(7.942)
ALDSQ	−0.066	(−2.981)	−0.079	(−3.445)	−0.078	(−3.441)
ALH	0.020	(0.632)	0.014	(0.375)	0.016	(0.406)
ALT	−0.279	(−4.069)	−0.273	(−4.642)	−0.273	(−4.663)
ALTSQ	0.364	(4.057)	0.356	(4.834)	0.356	(4.860)
GZ	0.726	(4.565)	0.683	(4.080)	0.683	(4.081)
BB1	−0.220	(−1.976)	−0.268	(−2.427)	−0.267	(−2.414)
BB2	−0.111	(−0.610)	−0.213	(−0.972)	−0.210	(−0.960)
BST2	−0.367	(−1.717)	−0.439	(−3.386)	−0.441	(−3.390)
FST	−0.289	(−2.138)	−0.188	(−1.519)	−0.188	(−1.525)

[1] Estimations of MECOSA (third stage) using the equality restriction and unspecified correlation matrix (Table 4). — [2] Estimations of GEE1 using the option unspecified correlation matrix (Table 4). — [3] Estimations of GEE1 using the correlation matrix presented in Table 2.

3.0 or above. Large data sets need the virtual memory manager. The great advantage of GAUSS is the ability of reading and processing the data blockwise, not only row by row as in other program systems. Therefore, the runtime for the GEE1 estimations was less than $3\frac{1}{2}$ min.

Finally, the correlation matrices for the MECOSA and the GEE1 are presented:

Table 4

Correlation matrices estimated

	In the third step be MECOSA			By GEE1 using the "unspecified" option			
Correlations				Correlations			
Wave 1	1.000			Wave 1	1.000		
Wave 2	0.812	1.000		Wave 2	0.319	1.000	
Wave 3	0.626	0.772	1.000	Wave 3	0.182	0.300	1.000
Wave 4	0.568	0.637	0.903	Wave 4	0.116	0.217	0.519

References

Arminger, Gerhard (1992): Analyzing Panel Data with Non-Metric Dependent Variables: Probit Models, Generalized Estimating Equations, Missing Data and Absorbing States, DIW Discussion Paper, No. 59, Berlin.

Arminger, Gerhard, Clifford C. *Clogg*, Michael E. *Sobel* (1995): Handbook of Statistical Modeling for the Social and Behavioral Sciences, New York: Plenum.

Carroll, Raymond J., C.F. Jeff *Wu*, David *Ruppert* (1988): The Effect of Estimating Weights in Weighted Least Squares, Journal of the American Statistical Association, Vol. 83, pp. 1045-1054.

Davis, Charles S. (1991): Semi-Parametric and Non-Parametric Methods for the Analysis of Repeated Measurements With Applications to Clinical Trials, Statistics in Medicine, Vol. 10, pp. 1959-1980.

Dempster, Arthur P., Nan M. *Laird*, Donald B. *Rubin* (1977): Maximum Likelihood from Incomplete Data via the EM Algorithm (with Discussion), Journal of the Royal Statistical Society, Series B, Vol. 39, pp. 1-38.

Flaig, Gebhard, Georg *Licht*, Victor *Steiner* (1993): Testing for State Dependence Effects in a Dynamic Model of Male Unemployment Behaviour, ZEW Discussion paper No. 93-07.

Gauss, Version 3.1.4 (1992): System and Graphics Manual, Aptech Systems, Inc., Maple Valley, Washington, USA.

Glim, release 3.77 (1985): The GLIM System Release 3.77 Manual, Numerical Algorithms Group Ltd, Mayfield House, Oxford, GB.

Gourieroux, Christian, Alain *Monfort*, Alain *Trognon* (1984): Pseudo Maximum Likelihood Methods: Theory, Econometrica, Vol. 52, pp. 682-700.

Gourieroux, Christian, Alain *Monfort* (1993): Pseudo-Likelihood Methods, in Maddala et al. (1993), pp. 335-362.

Hamerle, Alfred, Gerd *Ronning* (1995): Analysis of Discrete Panel Data, in Arminger et al. (1995).

Heckman, James J. (1981a): Statistical Models for Discrete Panel Data, in Manski and McFadden (1981), pp. 114-178.

Heckman, James J. (1981b): The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Stochastic Process, in Manski and McFadden (1981), pp. 179-195.

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

Liang, Kung-Yee, Scott L. *Zeger* (1986): Longitudinal Data Analysis Using Generalized Linear Models, Biometrika, Vol. 73, pp. 13-22.

Maddala, Gangadharrao S., Calyampudi R. *Rao*, Hrishikesh D. *Vinod* (1993): Handbook of Statistics, Vol. 11, Amsterdam: Elvesier.

Manski, Charles F., Daniel *McFadden* (1990): Structural Analysis of Discrete Data with Econometric Applications, London: Harvard University Press.

McDonald, Barry W. (1993): Estimating Logistic Regression Parameters for Bivariate Binary Data, Journal of the Royal Statistical Society, Ser. B, Vol. 55, pp. 391-397.

Muthén, Bengt (1984): A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators, Psychometrika, Vol. 49, pp. 115-132.

Paik, Myunghee C. (1988): Repeated Measurement Analysis for Nonnormal Data in Small Samples, Communications in Statistics and Simulation, Ser. B, Vol. 17, pp. 1155-1171.

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

Schepers, Andreas, Gerhard *Arminger* (1992): MECOSA: A Program for the Analysis of General Mean- and Covariance Structures with Non-Metric Variables, User Guide, Frauenfeld (Schweiz): SLI-AG.

Stram, Daniel O., L.J. *Wei*, James H. *Ware* (1988): Analysis if Repeated Ordered Categorical Outcomes With Possiby Missing Observations and Time-Dependent Covariates, Journal of the American Statistical Association, Vol. 83, pp. 631-637.

White, Halbert (1982): Maximum Likelihood Estimation of Misspecified Models, Econometrica, Vol. 50, 1-25.

Zeger, Scott L. (1988): Commentary, Statistics in Medicine, Vol. 7, pp. 161-168.

Zeger, Scott L., Kung-Yee *Liang* (1986): Longitudinal Data Analysis for Discrete and Continuous Outcomes, Biometrics, Vol. 42, pp. 121-130.

Ziegler, Andreas (1994a): Verallgemeinerte Schätzgleichungen zur Analyse korrelierter Daten, Dissertation, Universität Dortmund.

Ziegler, Andreas (1994b): GEE1: Ein Programmsystem zur Schätzung von Parameterstrukturen in multivariaten verallgemeinerten linearen Modellen mit Generalized Estimating Equations, Arbeitspapier des FB Wirtschaftswissenschaft der BUGH Wuppertal, Nr. 167, Wuppertal.

Zhao, Lue P., Ross L. *Prentice* (1990): Correlated Binary Regression Using a Quadratic Exponential Model, Biometrika, Vol. 77, pp. 642-648.

Summary

Analyzing the Employment Status with Panel Data from the GSOEP

Two different approaches for the analysis of dichotomous panel data have been well developed in the last years, that is the GEE approach of Liang and Zeger (1986) and the approach of estimating the models of Heckman (1981a).

First, a special case of the model Heckman proposed in 1981 is embedded into the Mean and Covariance Structure Model for non-metric dependent variables (Muthén 1984; Küsters 1987; Schepers and Arminger 1992).

Second, the approach of the generalized estimating equations (GEE) proposed in a series of papers by Liang and Zeger is discussed. Here, the original model of Liang and Zeger (1986) is considered. If the mean structure is correctly specified and first order identifiable Pseudo Maximum Likelihood (PML) estimation developed by Gourieroux et al. (1984) is used to compute a consistent estimate of the parameter vector.

Both models and the corresponding estimation methods are illustrated by analyzing panel data from the GSOEP.

Zusammenfassung

Analyse des Erwerbsstatus anhand von Paneldaten aus dem SOEP

Der Beitrag enthält zwei verschiedene Ansätze zur Analyse dichotomer Paneldaten, die in den letzten Jahren entwickelt wurden.

Einmal wird ein Spezialfall des Heckman Modells in ein Mittelwert und Kovarianz Strukturmodell für nicht metrische abhängige Variablen integriert.

Zum zweiten wird ein Generalized Estimating Equations (GEE) Ansatz von Liang und Zeger diskutiert. Eine korrekt spezifizierte Mittelwertstruktur und eine identifizierbare Pseudo Maximum Likelihood Schätzung erster Ordnung werden benutzt, um konsistente Schätzungen für den Parameter Vektor zu erhalten.

Beide Modelle und die entsprechenden Schätzmethoden werden schließlich durch die Verwendung von Panel Daten (GSOEP) illustriert.