Diskussionspapiere
Discussion Papers

Discussion Paper No. 59

Analyzing Panel Data with Non-Metric Dependent Variables: Probit Models, Generalized Estimating Equations, Missing Data and Absorbing States

by
*Gerhard Arminger**

Deutsches Institut für Wirtschaftsforschung, Berlin
German Institute for Economic Research, Berlin

Deutsches Institut für Wirtschaftsforschung

Discussion Paper No. 59

Analyzing Panel Data with Non-Metric Dependent Variables: Probit Models, Generalized Estimating Equations, Missing Data and Absorbing States

by

*Gerhard Arminger**

*) Prof. Dr. Gerhard Arminger (Bergische Universität Wuppertal) was at the DIW as a Visiting Scholar in November 1991.

Berlin, October 1992

Deutsches Institut für Wirtschaftsforschung, Berlin
Königin-Luise-Str. 5, 1000 Berlin 33
Telefon: 49-30 - 82 991-0
Telefax: 49-30 - 82 991-200

Analyzing Panel Data with Non-Metric Dependent Variables: Probit Models, Generalized Estimating Equations, Missing Data and Absorbing States

Gerhard Arminger
Bergische Universität Wuppertal

*Paper prepared for the International Conference in Social Science Methodology, Trento, Italy, June 22-26, 1992, session on Household Panels and Longitudinal Data Analysis. Comments and criticisms should be adressed to Gerhard Arminger, Department of Economics (FB 6), Bergische Universität, Gaußstr. 20, D-5600 Wuppertal, Germany. I am grateful to Georg Licht and Viktor Steiner at the *Zentrum für Europäische Wirtschaftsforschung (ZEW)*, Mannheim, for the data set on employment status and to Rainer Pischner and Gert Wagner at the *Deutsches Institut für Wirtschaftsforschung (DIW)*, Berlin for the data set on the need for nursing care. Andreas Schepers of the Bergische Universität Wuppertal has been enormously helpful by providing technical expertise for setting up the Mecosa runs and writing the Gauss program for the analysis in section 6. I am obliged to Ulrich Rendtel of the *DIW* and to Olaf Hübler at the University of Hannover for a critical reading of a first draft of the manuscript. Special thanks are due to the DIW which gave me the opportunity to work with the German Socio Economic Panel (GSOEP) as a visiting scholar in November 1991.

1 Introduction

In the first part of the paper we focus on the construction of models for non-metric panel data. The construction principles of the seminal work of Heckman (1981a, 1981b) on the specification and estimation of dichotomous outcomes in panel data are extended to include ordered categorial and censored dependent variables as well as simultaneous equation systems of non-metric dependent variables. In the third section we discuss different strategies to estimate the models discussed before. First it is shown that the model parameters can be estimated in a straightforward manner by assuming multivariate normality of the error terms and using conditional polychoric and polyserial correlation and covariance coefficients in the framework of mean- and covariance structure models for non-metric dependent variables. These models and the corresponding estimation technique has been introduced by Muthén (1984) and extended by Küsters (1987) and Schepers, Arminger and Küsters (1991). Special attention is given to the problem of initial states. Second, the approach of general estimating equations (GEE) proposed by Liang and Zeger in a series of papers is briefly dicussed. A special case of their approach coincides with the estimation strategy of Chamberlain (1984). The GEE approach is considerably easier to implement numerically than the polychoric correlation coefficients but cannot capture some of the relevant parameters in the Heckman type models directly. In the fourth section we illustrate the first estimation method by reanalyzing data from the German Socio Economic Panel (GSOEP, Wagner, Schupp and Rendtel 1991, Projektgruppe Panel 1991) on employment status of males previously analyzed by Flaig, Licht and Steiner (1992). The fifth section extends the results of Verbeek (1991) for the inclusion of sample selection bias in random effect models for metric dependent variables to limited dependent variables. Finally, the special problem of the estimation of random effect models for ordered categorical dependent variables in which one category is an absorbing state is treated. For this case, the likelihood function is derived. The last model is illustrated by a brief analysis of people who need nursing care. The data come again from the GSOEP.

2 Extension of Heckman's model

2.1 Heckman's model for dichotomous variables

Heckman (1981a, ch. 3.3) considers the following model for a latent variable y_{it}^*, $i = 1, \ldots, n$, $t = 1, \ldots, T$ where i denotes the individual and t denotes a sequence of equispaced time points:

$$y_{it}^* = \mu_{it} + \epsilon_{it}^* \tag{1}$$

$$\mu_{it} = x_{it}\beta + \sum_{j=1}^{\infty} \gamma_{t-j,t} y_{i,t-j} + \sum_{j=1}^{\infty} \lambda_{j,t-j} \prod_{l-1}^{j} y_{i,t-l} + \sum_{k=1}^{K} \delta_k y_{i,t-k}^* \tag{2}$$

$$\epsilon_i^* = (\epsilon_{i1}^*, \ldots, \epsilon_{iT}^*)', \quad \epsilon_i^* \sim \mathcal{N}(0, \Sigma), \tag{3}$$

$$y_{it} = \begin{cases} 1 & \text{if} \quad y_{it}^* > 0 \\ 0 & \text{if} \quad y_{it}^* \leq 0 \end{cases} \tag{4}$$

The latent variable y_{it}^* is considered as a disposition or utility which is connected to the observed dependent variable y_{it} through a dichotomous threshold model. The values y_{it}, y_{it}^* and x_{it} are collected in $T \times 1$ vectors y_i, y_i^* and the $R \times 1$ vector $x_i = (x_{i1}, \ldots, x_{iT})$. The random variables $\{y_i, x_i\}$ are identically and independently distributed which corresponds to a simple random sample from a population. The model specification consists in determining the structure of the systematic part μ_{it} and of the stochastic part ϵ^*_{it} of the model. A discussion of the different parts of the model specification for y^*_{it} is found in Heckman (1981a) and in Hamerle and Ronning (1993). We repeat here only the most important elements of the specification.

The first component of μ_{it} is the variation induced by possibly time varying explanatory variables x_{it}. The parameter vector β is time constant in this model which may changed to parameter vectors $\beta_t, t = 1, \ldots, T$ that vary over time.

The second component of μ_{it} captures the influence of former states of the observed dependent variable $y_{i,l-j}, j \geq 1$ which is called true state dependence. If $\gamma_{t-1,t} = \gamma_1 \neq 0$ and $\gamma_{t-j,t} = 0$ for all $j > 1$ and $t = 1, \ldots, T$ we have a simple Markov model. Note that the inclusion of former states requires the knowledge of initial states $y_{i0}, y_{i,-1}, \ldots$ depending on the specification of the parameters $\gamma_{t-j,t}$. If the initial states are known and non-stochastic they can be included in the variables x_{it} as additional explanatory variables. If the initial states are themselves outcomes of the process that generates y^*_{it}, the distribution of the initial states must be taken into account as discussed by Heckman (1981b) and in section 3 of this paper. Note that the effects of the former states $y_{i,t-j}$ may change for each time point. This is captured by $\gamma_{t-j,t}$. In most applications, $\gamma_{t-j,t}$ is set to γ_{t-j} and almost all parameters are set to 0.

The third component of μ_{it} models the dependence of y^*_{it} on the duration of the state $y_{i,t} = 1$. Again, the effects of duration may be different for each time length. These different effects are parameterized in $\lambda_{j,t-j}$. If $\lambda_{j,t-j} = \lambda$ we have the simple case of a linear duration effect.

The fourth component of μ_{it} models the dependence of y^*_{it} on former values $y^*_{i,t-j}$ of the latent endogenous variables. Structures that incorporate this component are called models with habit formation or habit persistence. The idea behind such models is that y^*_{it} is not dependent on the actual former state of the observed variable but rather on the former disposition or habit identified rather with $y^*_{i,t-j}$ than $y_{i,t-j}$. If the initial variables $y^*_{i,0}, y^*_{i,-1}, \ldots, y^*_{i,-(K-1)}$ are known and non-stochastic they may be included in the list of explanatory variables. Otherwise, assumptions about the distribution of the initial variables $y^*_{i,0}, y^*_{i,-1}, \ldots, y^*_{i,-(K-1)}$ must be made. This is especially the case, if $y^*_{i,0}, y^*_{i,-1}, \ldots, y^*_{i,-(K-1)}$ are not observed and only $y_{i,0}, y_{i,-1}, \ldots, y_{i,-(K-1)}$ are observed. This point is taken up in section 3.

Now we turn to the specification of the error term ϵ^*_{it}. Throughout this paper it is assumed that the error terms are uncorrelated with past, present and future explanatory variables (strong exogeneity). The method of Keane and Runkle (1992) to deal with weak exogeneity cannot be extended to limited dependent variables in a straightforward manner. The usual decomposition of ϵ^*_{it} is

$$\epsilon^*_{it} = \alpha_i + \epsilon_{it} \tag{5}$$

where α_i denotes the error term which varies across individuals but not over time and may be considered as unobserved heterogeneity just as in the metric case (cf. Hsiao 1986). The values of α_i may be considered as fixed effects for each i or may be considered as random effects such that $\alpha_i \sim \mathcal{N}(0, \sigma^2_\alpha)$. In the first case, α_i is an individual specific parameter. If y^*_{it} is modelled by $y^*_{it} = x_{it}\beta + \alpha_i + \epsilon^*_{it}$ and is actually observed as in the metric case then α_i may be eliminated by taking the first differences $y^*_{it} - y^*_{i,t-1} = (x_{it} - x_{i,t-1})\beta + \epsilon^*_{i,t} - \epsilon^*_{i,t-1}$. This technique does not work for non-metric models such as the probit model. In the case of a dichotomous logit model, the α_i's may be eliminated by conditioning on a sufficient statistic as shown in Hsiao (1986) and Hamerle and Ronning (1993). If α_i is a random variable, then it is assumed to be uncorrelated with z_{it} and ϵ_{it}. If ϵ_{it} has constant variance σ^2_ϵ and is serially uncorrelated, then ϵ^*_i has the typical covariance structure:

$$V(\epsilon_i^*) = \begin{pmatrix} \sigma_\alpha^2 + \sigma_\epsilon^2 & & & \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\epsilon^2 & & \\ \vdots & \vdots & \ddots & \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 + \sigma_\epsilon^2 \end{pmatrix} = \sigma_\alpha^2 11' + \sigma_\epsilon^2 I$$

where 1 is a $T \times 1$ vector of ones and I is the $T \times T$ identity matrix.

More generally, one can assume that ϵ_{it} has a serial structure, for instance a first order autoregressive process

$$\epsilon_{it} = \rho\epsilon_{i,t-1} + \nu_{it} \tag{6}$$

where the ν_{it} are *iid* normal with $\nu_{it} \sim \mathcal{N}(0, \sigma_\nu^2)$ and $|\rho| < 1$. The elements σ_{ts} of the covariance matrix Σ of ϵ_i^* are given by (cf. Hamerle and Ronning 1993):

$$\sigma_{ts} = \begin{cases} \sigma_\alpha^2 + \dfrac{\sigma_\nu^2}{1-\rho^2} & t = s \\ \sigma_\alpha^2 + \dfrac{\sigma_\nu^2}{1-\rho^2}\rho^{|t-s|} & t \neq s \end{cases} \tag{7}$$

The unobserved heterogeneity α_i may have different effects for each panel wave t. In this case ϵ_{it}^* is written as a one factor model

$$\epsilon_{it}^* = \lambda_t\alpha_i + \epsilon_{it} \tag{8}$$

which may be considered as a special case of a multifactor model. In a multifactor model, there is not only one type of unobserved heterogeneity but there are K factors α_{ik} which are collected in a $K \times 1$ vector α_i. In this case we have

$$\epsilon_{it}^* = \lambda_{t1}\alpha_{i1} + \lambda_{t2}\alpha_{i2} \ldots + \lambda_{tK}\alpha_{iK} + \epsilon_{it} \tag{9}$$

which may be written in matrix notation as

$$\epsilon_i^* = \Lambda\alpha_i + \epsilon_i \tag{10}$$

where Λ is the $T \times K$ matrix of factor loadings, α_i is the $K \times 1$ vector of factor scores with covariance matrix Φ and ϵ_i is uncorrelated with α_i and has covariance matrix Θ. Consequently, $V(\epsilon_i^*)$ may be written as

$$V(\epsilon_i^*) = \Lambda\Phi\Lambda^T + \Theta \tag{11}$$

Note that identification restrictions must be placed on Λ, Φ and Θ. The interpretation of general one factor schemes is discussed in Heckman (1981a, ch. 3.18).

Finally, we can assume that there is no specific structure of Σ. In this case, all parameters $\sigma_{ts}, s \leq t$ are estimated from the data without any constraints.

Heckman (1981a) discusses the ML estimation of these models under the assumption that $\epsilon_i^\star$ is normally distributed. Furthermore, he assumes that only dichotomous outcomes are observed. Since $\epsilon_i^\star \sim \mathcal{N}(0, \Sigma)$ the parameter vector of the explanatory variables can be estimated from a univariate probit model for each y_{it} with the usual identification restriction $\sigma_{tt} = 1, t = 1, \ldots, T$ which implies that β may only be estimated up to a scale factor. Note that this identification restriction is not required for all panel waves. If β is constant for all waves then the variance of $\epsilon_{it}^\star$ need only be restricted for one wave, usually the first (cf. Arminger 1987).

2.2 Extension to general threshold models

A thorough discussion and some extensions of Heckman's models to fixed effects logit and tobit models as well as to random effects probit models are found in Madalla (1987). We extend now Heckman's (1981a) dichotomous models in a systematic way to censored, metrically classified and ordered categorical dependent variables and simultaneous equation systems that allow as dependent variables any mixture of metric and/or limited dependent variables. Only random effect models are considered. Like Heckman (1981a) we consider the case that the initial states are known and non stochastic or that information about the distribution of the initial states is known so that the initial states can either be modelled simultaneously with the dependent variables or can be included in the explanatory variables. In this case the $T \times 1$ vector $y_i^\star$ may be written as

$$y_i^\star = \gamma + \Pi x_i + \epsilon_i^\star \tag{12}$$

where $\epsilon_i^\star$ follows a normal distribution with $\epsilon_i^\star \sim \mathcal{N}(0, \Sigma)$. The $R \times 1$ vector x_i contains the explanatory variables $x_{it}, t = 1, \ldots, T$. The $T \times 1$ vector γ is the vector of regression constants. The $T \times R$ matrix Π is the matrix of regression coefficients. If $\mu_{it} = x_{it}\beta_t$, then Π is structured as follows:

$$\Pi = \begin{pmatrix} \beta_1 & 0 & \ldots & 0 \\ 0 & \beta_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \beta_T \end{pmatrix} \tag{13}$$

In the case of time constant parameters, the vector β_t is replaced by β.

Together with the specification of Σ through a model for unobserved heterogeneity and for serial correlation, the above specification of $y_i^\star = \gamma + \Pi x_i + \epsilon_i^\star$ yields a conditional mean and covariance structure in the latent variable vector $y_i^\star$ with $y_i^\star \sim \mathcal{N}(\gamma + \Pi x_i, \Sigma)$.

The model is now extended by allowing not only the dichotomous threshold model of equation (4), but any one of the following threshold models that maps $y_{it}^\star$ onto the observed variable y_{it} (cf. Schepers, Arminger and Küsters, 1991). For convenience, the case index $i = 1, \ldots, n$ is omitted.

- y_t is metric (identity relation).

$$y_t = y_t^\star \tag{14}$$

- y_t is ordered categorical with unknown thresholds $\tau_{t,1} < \tau_{t,2} < \ldots < \tau_{t,K_t}$ and categories $y_t = 1, \ldots, K_t + 1$ (ordinal probit relation, McKelvey and Zavoina 1975).

$$y_t = k \Longleftrightarrow y_t^\star \in [\tau_{t,k-1}, \tau_{t,k}) \tag{15}$$

$$\text{with}[\tau_{t,0}, \tau_{t,1}) = (-\infty, \tau_{t,1}) \quad \text{and} \quad \tau_{t,K_t+1} = +\infty \quad .$$

Note that for reasons of identification the threshold $\tau_{t,1}$ is set to 0 and the variance of the reduced form error term σ_t^2 is set to 1. The parameters in β_t are only identified up to scale.

- Classified metric variables may be treated analogously to the ordinal probit case with the difference that the class limits are now used as known thresholds (Stewart 1983). No identification restrictions are necessary.
- y_t is one-sided censored with a threshold value $\tau_{t,1}$ known a priori (tobit relation, Tobin 1958).

$$y_t = \begin{cases} y_t^\star & \text{if } y_t^\star > \tau_{t,1} \\ \tau_{t,1} & \text{if } y_t^\star \leq \tau_{t,1} \end{cases} \tag{16}$$

- y_t is double-sided censored with threshold values $\tau_{t,1} < \tau_{t,2}$ known a priori (two-limit-probit relation, Rosett and Nelson 1975).

$$y_t = \begin{cases} \tau_{t,1} & \text{if } y_t^\star \leq \tau_{t,1} \\ y_t^\star & \text{if } \tau_{t,1} < y_t^\star < \tau_{t,2} \\ \tau_{t,2} & \text{if } \tau_{t,2} \leq y_t^\star \end{cases} \tag{17}$$

In the case of general threshold models several modifications of the dichotomous models for panel data have to considered. First, the state dependence must be modified for non-dichotomous dependent variables. In the case of a censored dependent variable there should be a dummy variable $d_{i,t-j}$ which takes on the value of 1 if the variable $y_{i,t-j}^\star$ has been observed and a value of 0 if $y_{i,t-j}$ takes on the threshold value. For metrically classified variables and ordered categorical variables a $K \times 1$ vector of dummy variables $d_{i,t-j}^{(k)}$ must be defined. The dummy variable $d_{i,t-j}^{(k)}$ equals 1 if $y_{i,t-j}^\star$ falls in category k, $k = 2, \ldots, K+1$ and is 0 otherwise.

Second, the identification restrictions for the variances σ_t^2, $t = 1, \ldots, T$ of $\epsilon_{it}^\star$ and for the thresholds collected in the vector $\tau^{(t)}$ for each panel wave t must be chosen with great care. In my opinion, the threshold vectors $\tau^{(t)}$ should be set equal across panels waves, otherwise the meaning of the categories of an ordered categorical variable is assumed to vary across time. This restriction immediately implies that, except for the first wave, the variances σ_t^2 need not be restricted but can vary across the panel waves. For censored dependent variables, no restrictions for σ_t^2 are necessary.

A further extension of Heckman's model concerns the specification of models for a system of variables. Instead of considering only one variable over T waves one may consider a vector $y_{it}^\star$ of H dependent variables over time. Each component of $y_{it}^\star$ is denoted by $y_{ith}^\star$, $h = 1, \ldots, H$. The vector of dependent variables $y_i^\star$ of the ith observation is then a $H \cdot T \times 1$ vector of dependent variables observed at T time points. Each latent variable $y_{ith}^\star$ at each time point is then mapped onto the observation y_{ith} through a threshold model of the form given above. The covariance matrix Σ of $\epsilon_i^\star$ contains in this case not only the conditional serial covariance structure for each variable $y_{th}^\star$ but also the conditional covariance structure between the variables across all time points. An example for such a model is found in Arminger and Ronning (1991).

3 Estimation strategies

3.1 Direct estimation of mean and covariance structures with conditional polychoric correlation coefficients

For general mean and covariance structures we assume that a $P \times 1$ vector $y_i^\star$ of latent dependent variables follows a multivariate normal distribution with conditional mean and covariance:

$$\begin{aligned} E(y_i^\star|x_i) &= \gamma(\vartheta) + \Pi(\vartheta)x_i \\ V(y_i^\star|x_i) &= \Sigma(\vartheta) \end{aligned} \tag{18}$$

In the analysis of panel data, P equals T if a univariate dependent variable is analyzed or P equals $H \cdot T$ if a multivariate dependent variable is analyzed. $\gamma(\vartheta)$ is a $P \times 1$ vector of regression constants and $\Pi(\vartheta)$ is a $P \times R$ matrix of reduced form regression coefficients. x_i is a $R \times 1$ vector of explanatory variables. $\Sigma(\vartheta)$ is the $P \times P$ covariance matrix of the errors of the reduced form. ϑ is the $\tilde{q} \times 1$ vector of structural parameters to be estimated. The reduced form parameters $\gamma(\vartheta)$, $\Pi(\vartheta)$ and $\Sigma(\vartheta)$ are continously differentiable functions of a common vector ϑ.

In the MECOSA program most parametrizations useful in applications can be implemented using the matrix features of GAUSS. Typical examples are simultaneous equation systems

$$y_i^\star = By_i^\star + \Gamma x_i + \epsilon_i \quad \text{with} \quad \epsilon_i \sim N(0, \Omega) \tag{19}$$

with the reduced form parameters

$$\Pi(\vartheta) = (I - B)^{-1}\Gamma \quad \text{and} \quad \Sigma(\vartheta) = (I - B)^{-1}\Omega(I - B)^{-1\prime} \tag{20}$$

and confirmatory factor analysis

$$y_i^\star = \Lambda\eta_i + \epsilon_i \quad \text{with} \quad \eta_i \sim N(0, \Phi) \quad \text{and} \quad \epsilon_i \sim N(0, \Theta) \tag{21}$$

and the reduced form parameters

$$\Pi(\vartheta) = 0 \quad , \quad \Sigma(\vartheta) = \Lambda\Phi\Lambda' + \Theta \tag{22}$$

In the first example ϑ consists of the structural parameter matrices B, Γ and Ω. In the second example, ϑ consists of Λ, Φ and Θ.

Note that the typical structure of panel models, that is $\mu_{it} = x_{it}\beta$ cannot be embedded in the reduced form of equation (19) directly. If one uses the model of (19) with $x_i = (x_{i1}, \ldots, x_{iT})'$ then the dependent variables $y_{it}^\star$ are regressed not only on x_{it} but also on all other variables x_{is}, $s \neq t$. The parameter restrictions that apply to the reduced form have to be introduced at the third stage of the estimation procedure.

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

1. In the first stage the threshold parameters τ, the reduced form coefficients γ and Π of the regression equation, and the reduced form error variance σ_t^2 of the tth equation are estimated using marginal maximum likelihood. Note that this first stage is the estimation of the mean structure without restrictions as in equation (19). The parameters to be estimated in the tth equation are the thresholds denoted by the vector τ_t, the regression constant denoted by γ_t, the regression coefficients, i.e. the tth row of Π denoted by $\Pi_{t\cdot}$ and the variance denoted by σ_t^2. The loglikelihood function that is maximized with respect to $\{\tau_t, \gamma_t, \Pi_{t\cdot}, \sigma_t^2\}$ of the tth equation is given by:

$$l_t(\tau_t, \gamma_t, \Pi_{t\cdot}, \sigma_t^2) = \sum_{i=1}^{n} \ln P(y_{it}|x_i) \tag{23}$$

The formulation of $P(y_{it}|x_i)$ depends on the measurement level of the observed variable y_t. If y_t is metric then $P(y_{it}|x_i)$ is the univariate normal density with expected value $\gamma_t + \Pi_{t\cdot}x_i$ and variance σ_t^2, i.e.,

$$P(y_{it}|x_i) = \varphi(y_{it}|\gamma_t + \Pi_{t\cdot}x_i, \sigma_t^2) \quad , \tag{24}$$

in which $\varphi(y|\mu, \sigma^2)$ denotes the univariate normal density. Another example is the ordinal probit model that is applied if y_t is an ordinal variable. In this case the probability that $y_{it} = k$ must be computed as:

$$P(y_{it} = k|x_i) = \int_{\tau_{t,(k-1)}}^{\tau_{t,(k)}} \varphi(y^\star|\gamma_t + \Pi_{t\cdot}x_i, 1)dy^\star \tag{25}$$

The probability of the last equation is the basis of the likelihood used in the ordinal probit model (McKelvey and Zavoina 1975). If other models are used such as the tobit model (Tobin 1958) for dependent variables censored on one side only or the two limit probit model (Rosett and Nelson 1975) for dependent variables censored on two sides then the probabilities $P(y_{it}|x_i)$ have to modified accordingly.

A solution to the likelihood equations obtained by setting the first derivatives of $l_t(\tau_t, \gamma_t, \Pi_{t\cdot}, \sigma_t^2)$ to 0 is computed in MECOSA by applying a Quasi-Newton algorithm using analytical first derivatives. The second order derivatives are approximated by computing the crossproduct of the first order derivatives. The first estimation stage yields strongly consistent estimates of $\tau_t, \gamma_t, \Pi_{t\cdot}$ and σ_t^2 for each equation t. Regularity conditions and the proof of strong consistency are given in Küsters (1987).

2. In the second stage the problem is to estimate the covariances of the error terms in the reduced form equations. Note that in this stage the covariances are estimated without parametric restrictions. Since the errors are assumed to be normally distributed and strongly consistent estimators of the reduced form coefficients have already been obtained in the first stage the estimation problem reduces to maximizing the loglikelihood function

$$l_{tj}(\sigma_{tj}) = \sum_{i=1}^{n} \ln P(y_{it}, y_{ij}|x_i, \hat{\tau}_t, \hat{\gamma}_t, \hat{\Pi}_{t\cdot}, \hat{\sigma}_t^2, \hat{\tau}_j, \hat{\gamma}_j, \hat{\Pi}_{j\cdot}, \hat{\sigma}_j^2, \sigma_{tj}) \quad , \tag{26}$$

in which $P(y_{it}, y_{ij}|x_i, \hat{\tau}_t, \hat{\gamma}_t, \hat{\Pi}_{t\cdot}, \hat{\sigma}_t^2, \hat{\tau}_j, \hat{\gamma}_j, \hat{\Pi}_{j\cdot}, \hat{\sigma}_j^2, \sigma_{tj})$ is the bivariate probability of y_{it} and y_{ij} given x_i and the reduced form coefficients. A typical example of this bivariate probability is the case when y_t and y_j are both ordinal. Then the probability that $y_{it} = k$ and $y_{ij} = l$ is given by:

$$P(y_{it} = k, y_{ij} = l|x_i) = \int_{\hat{\tau}_{t,(k-1)}}^{\hat{\tau}_{t,(k)}} \int_{\hat{\tau}_{j,(l-1)}}^{\hat{\tau}_{j,(l)}} \varphi(y_t^*, y_j^*|\hat{\mu}_{it}, \hat{\sigma}_t^2, \hat{\mu}_{ij}, \hat{\sigma}_j^2, \sigma_{tj}) dy_j^* dy_t^* \tag{27}$$

in which $\hat{\mu}_{it} = \hat{\gamma}_t + \hat{\Pi}_t.x_i, \hat{\mu}_{ij} = \hat{\gamma}_j + \hat{\Pi}_j.x_i$ and $\varphi(y_t^*, y_j^*|\mu_t, \sigma_t^2, \mu_j, \sigma_j^2, \sigma_{tj})$ is the bivariate normal density function. Note that in the ordinal case $\hat{\sigma}_t^2 = \hat{\sigma}_j^2 = 1$. Hence, σ_{tj} is a correlation coefficient which is called the polychoric correlation coefficient. The loglikelihood function $l_{tj}(\sigma_{tj})$ has to be modified accordingly if variables with other measurement levels are used.

The objective function $l_{tj}(\sigma_{tj})$ is maximized using the regula falsi algorithm with analytical first derivatives. The resulting estimates $\hat{\sigma}_{tj}$ are strongly consistent estimators of σ_{tj}. Note that these covariances are the covariances of the error terms in the equations for y_t^*, $t = 1, \ldots, P$ conditional on x_i. In contrast to LISREL 7, we do not assume that the variables y_{it}^* and y_{ij}^*, which depend on x_i, are normal. We only assume that the errors are normal.

The estimated thresholds $\hat{\tau}_t$, the reduced form coefficients $\hat{\gamma}_t$ and $\hat{\Pi}_{t\cdot}$, the variances $\hat{\sigma}_t^2$ and the covariances $\hat{\sigma}_{tj}$ from all equations are then collected in a vector $\hat{\kappa}_n$ which depends on the sample size n. For the final estimation stage, a strongly consistent estimate of the asymptotic covariance matrix W of $\hat{\kappa}_n$ is computed. This estimate is denoted by $\hat{W}_n$. The asymptotic covariance matrix W is difficult to derive since the estimates of $\hat{\sigma}_{tj}$ of the second stage depend on the estimated coefficients $\hat{\tau}_f, \hat{\gamma}_f, \hat{\Pi}_{f\cdot}, \hat{\sigma}_f^2, f = t, j$ of the first stage. The various elements of the asymptotic covariance matrix W are given in Küsters (1987). The estimate $\hat{W}_n$ is computed in MECOSA by using analytical first order and numerical second order derivatives of the first and second stage loglikelihood function.

3. In the third stage the vector κ of thresholds, the reduced form regression coefficients and the reduced form covariance matrix is written as a function of the structural parameters of interest, collected in the parameter vector ϑ. The parameter vector ϑ is then estimated by minimizing the quadratic form

$$Q_n(\vartheta) = (\hat{\kappa}_n - \kappa(\vartheta))'\hat{W}_n^{-1}(\hat{\kappa}_n - \kappa(\vartheta)) \tag{28}$$

which is a minimum distance approach based on the asymptotic normality of the estimators of the reduced form coefficients. The vector $\hat{\kappa}_n$ is asymptotically normal with expected value $\kappa(\vartheta)$ and covariance matrix W. Since $\hat{W}_n$ is a strongly consistent estimate of W the quadratic form $Q_n(\vartheta)$ is centrally χ^2 distributed with $p - \tilde{q}$ degrees of freedom if the model is specified correctly and the sample size is sufficiently large. The number p indicates the number of elements in $\hat{\kappa}_n$ while $\tilde{q}$ is the number of elements in ϑ. The computation of $\hat{W}_n$ is quite cumbersome for models with many parameters. As an alternative, one can use the weight matrix of the GLS approach in LISREL which may be based on the estimated covariance matrix of y_i^* and x_i. This covariance matrix may be computed from the estimated matrix $\hat{\Sigma}$ of the error terms, the matrix $\hat{\Pi}$ of the reduced form and the estimated covariance matrix of the explanatory variables. The function $Q_n(\vartheta)$ is minimized using the Davidon-Fletcher-Powell algorithm (Luenberger 1984) with numerical first derivatives.

The program MECOSA follows these three estimation stages. In the third stage, the facilities of GAUSS are fully exploited to be able to estimate parameters under arbitrary restrictions. The parameter vector $\kappa(\vartheta)$ may be defined using the matrix language and the procedure facility of GAUSS. Consequently, arbitrary restrictions can be placed on $\tau(\vartheta)$, $\gamma(\vartheta)$, $\Pi(\vartheta)$ and $\Sigma(\vartheta)$ including the restrictions placed on $V(\epsilon_i^*) = \Sigma(\vartheta)$ in the analysis of panel data.

As a technical note on the MECOSA program we remark that MECOSA estimates the reduced form parameters $\Pi(\vartheta)$ in the first stage by regressing all dependent variables in $y_i^* = (y_{i1}^*, \ldots, y_{iT}^*)'$ on all regressors collected in x_i without the usual restriction that the effect of x_{it} on $y_{i,t-j}$, $j \geq 1$ is zero. This restriction has to be put into the program in the third stage. If the model includes lagged dependent variables, it may be necessary to restrict the parameter estimates already in the first stage. In this case, one can trick the program into restricted estimation already in the first stage by running the first stage of MECOSA T times separately for each y^*_{it} which is then regressed on x_{it}. In the second stage, the standard batch input of MECOSA has to corrected to deal with the T estimation results from the first stage. The polychoric covariances and/or correlation coefficients of the reduced form are then computed under the restriction of the first stage.

3.2 The problem of initial states

For the estimation of the general Heckman model it was assumed that the initial states for a model with state dependence or with habit persistence are known and non-stochastic or that information about the initial states could be used in the list of explanatory variables. If this is not the case, one must assume that either a new process begins at the first wave of the panel or that the error terms of the process for the initial states are uncorrelated with the error terms of the T panel waves. (cf. Heckman 1981a, b). Note that the first assumption holds for the attrition process in a panel where the dependent variable is the participation of a person in panel wave t given that this person has participated in wave 1. At the first wave, all persons participate. Hence, the initial states are known and a new process starts at time 1.

Heckman (1981b) considers the following special case of the initial states problem:

$$y^*_{it} = \beta + \gamma y_{i,t-1} + \alpha_i + \epsilon_{it} \quad \text{with} \tag{29}$$

$$y_{it} = \begin{cases} 1 & \text{if} \quad y^*_{it} > 0 \\ 0 & \text{if} \quad y^*_{it} \leq 0 \end{cases}$$

The random effects are $\alpha_i \sim \mathcal{N}(0, \sigma^2_\alpha)$, $E(\alpha_i \epsilon_{it}) = 0$, $\epsilon_{it} \sim \mathcal{N}(0, \sigma^2_\epsilon)$ with $\sigma^2_\epsilon = 1$.

In this model, the only initial value is y_{i0}. What happens if y_{i0} is not known and non-stochastic, but is determined by the same process as y_{it} for $t = 1, \ldots, T$? In this case, y_{i0} is dependent on α_i and the sample conditional likelihood function for $y_{it}, t = 1, \ldots, T$ given y_{i0} is given by integration over the unobserved variable α which may be written as $\alpha = \sigma_\alpha \eta$ with $\eta \sim \mathcal{N}(0, 1)$.

$$\begin{aligned} L(\beta_0, \gamma, \sigma^2_\alpha) \quad &= \quad \prod_{i=1}^{n} \int_{-\infty}^{\infty} \prod_{t=1}^{T} (\Phi(\beta + \gamma y_{i,t-1} + \sigma_\alpha \eta)^{y_{it}} (1 - \Phi(\beta + \gamma y_{i,t-1} + \sigma_\alpha \eta))^{1-y_{it}} \\ &\times \quad P(y_{i0} | \sigma_\alpha \eta) \varphi(\eta) d\eta \end{aligned} \tag{30}$$

where $\varphi(.)$ is the standard normal density. The term $P(y_{i0}|\alpha)$ is the probability that the random variable y^*_{i0} is either > 0 or ≤ 0. Possible specifications of $P(y_{i0}|\alpha)$ by introducing restrictions on the process generating the exogenous variables x_{it} and application of the estimation method of Kiefer and Wolfowitz (1956) are discussed by Heckman (1981b). However, for all practical purposes this estimation method is too cumbersome.

As a first simple solution for this estimation problem, Heckman proposes the estimation of α_i as a parameter by ML methods. This method, however, yields only consistent estimates of the structural parameters β and γ if $T \to \infty$ (cf. Neyman and Scott 1948 and Andersen 1973) which cannot be assumed for panel data. The limited Monte Carlo evidence given by Heckman (1981b) for $T = 8$ panel waves in the model of equation (30) is certainly not sufficient to recommend the ML estimation of the α_i's in conjunction with β and γ as a general solution.

The second simple solution proposed by Heckman (1981b) is the substitution of $P(y_{i0}|\alpha)$ by $F(x_{i0}\delta)$, where x_{i0} is a vector of explanatory variables for y_{i0}^* and $F(.)$ is the distribution function of the random variable

$$y_{i0}^* = x_{i0}\delta + \epsilon_{i0}^* \tag{31}$$

This ad hoc solution is simply the attempt to replace the unobserved heterogeneity in y_{i0}^* by observed heterogeneity in x_{i0}. The error term is allowed to be correlated with α_i and ϵ_{it} for $t = 1, \ldots, T$ without restrictions to capture the serial correlation between ϵ_{i0}^* and $\epsilon_{it}^* = \alpha_i + \epsilon_{it}$ induced by α_i. In practice, this procedure amounts to the inclusion of y_{i0}^* in the vector of dependent variables and additional parameters in the augmented covariance matrix for $\bar{\epsilon}_i^* = (\epsilon_{i0}^*, \epsilon_{i1}^*, \ldots, \epsilon_{iT}^*)'$. If $\epsilon_{it}^* = \alpha_i + \epsilon_{it}$ with $V(\alpha_i) = \sigma_\alpha^2$ and $V(\epsilon_{it}) = \sigma_\epsilon^2$, then the augmented covariance matrix is given by

$$V(\bar{\epsilon}_i^*) = \begin{pmatrix} \sigma_{\epsilon 0}^2 & & & & \\ \sigma_{10} & \sigma_\alpha^2 + \sigma_\epsilon^2 & & & \\ \sigma_{20} & \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\epsilon^2 & & \\ \vdots & \vdots & \vdots & \ddots & \\ \sigma_{T0} & \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 + \sigma_\epsilon^2 \end{pmatrix} \tag{32}$$

The limited Monte Carlo simulation of Heckman (1981b) shows that this approach works better than the simultaneous ML method to yield consistent estimates of β and γ. If $F(.) = \Phi(.)$, the above method can immediately be incorporated in the MECOSA framework by specifying

$$y_{i0}^* = x_{i0}\delta + \epsilon_{i0}^* \tag{33}$$

$$y_{it}^* = x_{it}\beta + \epsilon_{it}^* \tag{34}$$

The joint covariance matrix of ϵ_{i0}^* and ϵ_i^* may be structured as in equation (32).

This solution for the initial states problem depends crucially on the specification of a model for the initial states. The key idea is to specify a model for the initial states that gives as good predictions as possible. The model for the initial states need not be the same model as for $y_{it}, t = 1, \ldots, T$. The errors $\epsilon_{i0}^*, \epsilon_{i,-1}^*, \epsilon_{i,-K}^*$ are allowed to correlate freely with ϵ_{it}^*. If the model for the initial states is misspecified, the simple solution will work badly. If the model for the initial states is correct, the ML estimators of B, Γ and $V(\epsilon_i^*)$ will be consistent for fixed T as $n \to \infty$.

A problem that is easier to handle occurs if habit persistence instead of state dependence is taken into account. Assume that the following model is formulated with $\epsilon_{it}^* = \alpha_i + \epsilon_{it}$, $V(\alpha_i) = \sigma_\alpha^2$, $V(\epsilon_{it}) = \sigma_\epsilon^2$ and $E(\epsilon_{it}\epsilon_{is}) = 0$ for $t \neq s$:

$$y_{it}^* = x_{it}\beta + \gamma y_{i,t-1}^* + \epsilon_{it}^*, \quad t = 1, \ldots, T \tag{35}$$

Furthermore, we assume like Bhargava and Sargan (1983) for the metric case that the start value y_{i0}^* has been generated by the same process as $y_{i,t}^*$, that is:

$$y_{i0}^{\star} = x_{i0}\beta + \gamma y_{i,-1}^{\star} + \epsilon_{i0}^{\star} \tag{36}$$

This implies that $y_{i0}^{\star}$ is determined by

$$y_{i0}^{\star} = \sum_{k=0}^{\infty} \gamma^k (x_{i,-k}\beta + \epsilon_{i,-k}^{\star}) = \bar{y}_{i0} + \delta_{i,-k} \tag{37}$$

where $\delta_{i,-k} = \sum_{k=0}^{\infty} \gamma^k \epsilon_{i,-k}^{\star}$. Assume that the optimal predictor of $\bar{y}_{i0}$ given all observed values x_{it}, $t = 1, \ldots, T$ of the exogenous variables is given by

$$\tilde{y}_{i0} = x_i \delta \tag{38}$$

were δ is the vector of regression coefficients for the vector $x_i = (x_{i0}, x_{i1}, \ldots, x_{iT})$. Then $\bar{y}_{i0}$ may be written as

$$\bar{y}_{i0} = \tilde{y}_{i0} + u_{i0} \tag{39}$$

where $u_{i0} \sim \mathcal{N}(0, \sigma_u^2)$ and u_{i0} is independent of α_i and ϵ_{it}. The model for the augmented vector $\tilde{y}_i^{\star} = (y_{i0}^{\star}, y_i^{\star\prime})'$ can then be written in the form of a simultaneous equation system

$$\tilde{y}_i^{\star} = B\tilde{y}_i^{\star} + \Gamma \tilde{x}_i^{\star} + \tilde{\epsilon}_i^{\star} \tag{40}$$

with the $(T+1) \times (T+1)$ parameter matrix B and the $(T+1) \times q$ parameter matrix Γ:

$$B = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \gamma & 0 & \cdots & 0 & 0 \\ 0 & \gamma & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \gamma & 0 \end{pmatrix} \quad , \quad \Gamma = \begin{pmatrix} \delta & 0 & 0 & \cdots & 0 \\ 0 & \beta & 0 & \cdots & 0 \\ 0 & 0 & \beta & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \beta \end{pmatrix} \tag{41}$$

The vector of explanatory variables $\tilde{x}_i$ is $\tilde{x}_i = (x_{i0}, x_{i1}, \ldots, x_{iT})'$. The covariance matrix $V(\tilde{\epsilon}_i^{\star})$ of the augmented vector $\tilde{\epsilon}_i^{\star} = (\epsilon_{i0}, \epsilon_i^{\star})'$ is given by the elements (cf. Bhargava and Sargan 1983):

$$\begin{aligned} v_{00} &= \frac{\sigma_\alpha^2}{1-\gamma^2} + \frac{\sigma_\epsilon^2}{1-\gamma^2} + \sigma_u^2 \\ v_{0t} &= \frac{\sigma_\alpha^2}{1-\gamma} \\ v_{tt} &= \sigma_\alpha^2 + \sigma_\epsilon^2, \quad t = 1, \ldots, T \\ v_{ts} &= \sigma_\alpha^2, \quad t \neq s \end{aligned}$$

The reduced form parameters estimated in the first two stages of MECOSA are given by:

$$\Pi = (I - B)^{-1}\Gamma \quad \text{and} \quad \Sigma = (I - B)^{-1} V(\tilde{\epsilon}_i^{\star})(I - B)^{-1\prime} \tag{42}$$

3.3 Estimation of mean structures with generalized estimating equations

In section 3.1 we have discussed the direct estimation of the parameters of mean and covariance structures for non–metric dependent variables. The model was specified in terms of a latent variable vector y_i^* with a linear model $E(y_i^*|x_i) = \gamma(\vartheta) + \Pi(\vartheta)x_i$ for the expected value of y_i^* given x_i and the conditional covariance $V(y_i^*|x_i) = \Sigma(\vartheta)$ of y_i^* given x_i: The elements y_{it}^* of y_i^* are connected to the observations y_i through a threshold model. An essential assumption of the model is conditional multivariate normality given x_i. This allows one to use dichotomous and ordinal probit models and conditional polychoric and polyserial correlation coefficients to estimate the parameters ϑ of the joint mean and covariance structure of y_i^* given x_i directly. The same parameter vector applies to the mean as well as to the covariance structure.

A different approach to estimate the parameters of a general mean structure

$$y_{it} = \mu_{it}(\vartheta) + \epsilon_{it}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T, \tag{43}$$

has been proposed in a series of paper by Zeger and Liang (Zeger and Liang 1986, Zeger, Liang and Albert 1988, Zeger, Liang and Qaqish 1992). Here, we explain briefly the main ideas of the generalized estimating equation (GEE) approach and discuss some of the differences and similarities of the GEE approach and the approach in section 3.1.

We consider the original model of Zeger and Liang (1986). Assume that $y_{it}, i = 1, \ldots, n, t = 1, \ldots, T_i$ is dichotomous with $y_{it} = 1$ if an event occurs and 0 if it does not occur. Let $y_{it} = \mu_{it}(\vartheta) + \epsilon_{it}$ where $\mu_{it}(\vartheta) = \mu(x_{it}, \vartheta)$ is the mean structure of y_{it} and ϵ_{it} is the error term. The mean structure is assumed to be correctly specified, that is $E(\epsilon_{it}|x_{it}) = 0$ for $t = 1, \ldots, T_i$ (strong exogeneity), and first order identifiable, that is $\mu(\vartheta_1) = \mu(\vartheta_2) \Rightarrow \vartheta_1 = \vartheta_2$ almost surely.

Typical examples for such mean structures are the logit model

$$\mu(x_{it}, \vartheta) = \exp(x_{it}\vartheta)/(1 + \exp(x_{it}\vartheta)) \tag{44}$$

and the probit model

$$\mu(x_{it}, \vartheta) = \Phi(x_{it}\vartheta) \tag{45}$$

where $\Phi(.)$ is the standard normal distribution function. Note that these models may also be derived as threshold models.

Consider now the vector y_i of length T_i with expected value $\mu_i(\vartheta)$ and error term ϵ_i with unknown $T_i \times T_i$ covariance matrix Ω_i which may have a serial structure. The vectors y_i are assumed to be independent. Assume now that the vector y_i follows a normal distribution with identity covariance matrix, that is $y_i \sim \mathcal{N}(\mu_i, I)$. Although this assumption is certainly wrong for dichotomous variables $y_{it}, t = 1, \ldots, T_i$, one may use the theory of Pseudo Maximum Likelihood (PML) estimation developed by Gourieroux, Monfort and Trognon (1984) to show that the maximization of the normal pseudo loglikelihood function

$$l(\vartheta) = \sum_{i=1}^{n} \ln \varphi(y_i|\mu_i(\vartheta), I) \tag{46}$$

yields a consistent estimator of ϑ which is denoted by $\hat{\vartheta}$. $\varphi(y|\mu, \Sigma)$ is the multivariate normal density with expected value μ and covariance matrix Σ. The maximization of $l(\vartheta)$ is equivalent to the minimization of a sum of quadratic forms in I that is

$$Q(\vartheta) = \sum_{i=1}^{n}(y_i - \mu_i(\vartheta))' I (y_i - \mu_i(\vartheta)) \tag{47}$$

This is equivalent to minimizing the sum of squared Euclidian distances and is therefore a minimum distance estimator (MDE). The asymptotic covariance matrix $V(\hat{\vartheta})$ matrix of the PML estimator $\hat{\vartheta}$ depends on the first and second derivatives of $[\ln \varphi(y_i|\mu_i(\vartheta), I)]$ w.r.t. ϑ and the true but unknown covariance matrices Ω_i which have not been specified.

A consistent estimator of $V(\hat{\vartheta})$ is given by the "sandwich" form

$$\hat{V}(\hat{\vartheta}) = \hat{A}(\hat{\vartheta})^{-1}\hat{B}(\hat{\vartheta})\hat{A}(\hat{\vartheta})^{-1} \tag{48}$$

with the elements

$$\hat{A}(\hat{\vartheta}) = \sum_{i=1}^{n} \frac{\partial \mu_i'(\hat{\vartheta})}{\partial \vartheta} \frac{\partial \mu_i(\hat{\vartheta})}{\partial \vartheta'} \tag{49}$$

$$\hat{B}(\hat{\vartheta}) = \sum_{i=1}^{n} \frac{\partial \mu_i'(\hat{\vartheta})}{\partial \vartheta}(y_i - \mu_i(\hat{\vartheta}))(y_i - \mu_i(\hat{\vartheta}))' \frac{\partial \mu_i(\hat{\vartheta})}{\partial \vartheta'} \tag{50}$$

The advantage of this estimator is that it does not depend on the 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 is needed. Note that this is the approach of Chamberlain (1984) for estimating a multivariate probit model for dichotomous outcomes in panel data.

To overcome the inefficiency of the estimator $\hat{\vartheta}$, Zeger and Liang (1986) introduce a "working" covariance matrix Σ_i of ϵ_i which may be thought of as an approximation of Ω_i. In the dichotomous case, one may assume that the variances of y_{it} given x_{it} are given by $\mu_{it}(\vartheta)(1 - \mu_{it}(\vartheta))$. Hence, only a working correlation matrix R_i needs to be specified as a function of a small number of parameters collected in α , that is $R_i = R_i(\alpha)$ and $\Sigma_i = \Sigma_i(\vartheta, \alpha)$.

The parameter vector α may be estimated consistently using simple least squares methods on the empirical covariance matrix $\hat{R}_i$ which may be computed from y_i, x_i and the PML estimate $\hat{\vartheta}$. A simple example is equicorrelation where it is assumed that $\rho_{t,s} = \alpha$ for $t \neq s$. For $\alpha > 0$, this model corresponds to the random effect model for unobserved heterogeneity in the analysis of panel data. Note that the "working" covariance matrix $\Sigma_i(\hat{\alpha}, \hat{\vartheta})$ need not to be equal to the true covariance matrix Ω_i which is still unspecified. But if $\Sigma_i(\hat{\alpha}, \hat{\vartheta})$ is equal or close to Ω_i, the estimator $\tilde{\vartheta}$ that takes $\Sigma_i(\hat{\alpha}, \hat{\vartheta})$ into account will be more efficient than $\hat{\vartheta}$. 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 $\tilde{\vartheta}$ is found by using the PML estimation with the possibly wrong density function $\varphi(y_i|\mu_i, \Sigma_i(\hat{\alpha}, \hat{\vartheta}))$ or – equivalently – minimizing the sum of the Mahalanobis distances:

$$M(\vartheta) = \sum_{i=1}^{n}(y_i - \mu_i(\vartheta))' \Sigma_i(\hat{\alpha}, \hat{\vartheta})^{-1}(y_i - \mu_i(\vartheta)) \tag{51}$$

Setting the first derivative of $M(\vartheta)$ w.r.t. ϑ to 0 yields the generalized estimating equations:

$$\frac{\partial M(\vartheta)}{\partial \vartheta} = \sum_{i=1}^{n} \frac{\partial \mu_i(\vartheta))'}{\partial \vartheta} \Sigma_i(\hat{\alpha}, \hat{\vartheta})^{-1}(y_i - \mu_i(\vartheta)) = 0 \tag{52}$$

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

$$\tilde{V}(\tilde{\vartheta}) = \tilde{A}(\tilde{\vartheta})^{-1}\tilde{B}(\tilde{\vartheta})\tilde{A}(\tilde{\vartheta})^{-1} \tag{53}$$

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

$$\tilde{A}(\tilde{\vartheta}) = \sum_{i=1}^{n} \frac{\partial \mu_i'(\tilde{\vartheta})}{\partial \vartheta} \hat{\Sigma}_i^{-1} \frac{\partial \mu_i(\tilde{\vartheta})}{\partial \vartheta'} \tag{54}$$

$$\tilde{B}(\tilde{\vartheta}) = \sum_{i=1}^{n} \frac{\partial \mu_i'(\tilde{\vartheta})}{\partial \vartheta} \hat{\Sigma}_i^{-1} (y_i - \mu_i(\tilde{\vartheta}))(y_i - \mu_i(\tilde{\vartheta}))' \hat{\Sigma}_i^{-1} \frac{\partial \mu_i(\tilde{\vartheta})}{\partial \vartheta'} \tag{55}$$

The GEE approach for mean structures can be further refined by estimating not only the parameters vector ϑ of the mean structure with generalized estimating equations derived from the normal distribution but by estimating the parameter vector α of the covariance structure with generalized estimating equations as well. This strategy has been proposed by Prentice and Zhao (1991). Again, it is based on the PML approach of Gourieroux et al (1984) using the quadratic rather than the linear exponential family for constructing the pseudo loglikelihood function.

What are the similarities and the differences of the GEE and the MECOSA approach? In the GEE approach the focus is on the estimation of the parameter ϑ of the mean structure. There is no need to assume that y_i given x_i is normal. The normality assumption is only used to specify the marginal probit model for $E(y_{it}|x_{it})$ and to derive the generalized estimating equations. The mean structure formulation $\mu_i(\vartheta)$ is quite general, it can easily be extended to deal with unordered categorical outcomes. Important advantages of the GEE method are comparatively easy implementation and the possibility to let T_i vary over i.

The working covariance matrix Σ_i is primarily introduced to increase efficiency. The model of Σ_i is not formulated in the covariance matrix of the latent variable vector y_i^* given x_i but rather in the vector y_i given x_i. This makes it difficult to formulate and test hypotheses about specific covariance structures found in section 2 that may be introduced by specific assumption about the error process or by a simultaneous equation system. There is no possibility to include habit persistence effects and to solve the initial states problem in a manner analogous to Bhargava and Sargan (1983). At least at present there seems to be no possibility to include factor analytic measurement models jointly with regression models in the GEE approach which precludes statistical analysis if the regressors have been measurerd with error. Measurements with error occur often in economics as well as in epidemiology (for instance in dietary assessment).

The MECOSA approach on the other hand is more restricted by its assumption of multivariate normality of y_i given x_i. Computationally, it is much more cumbersome because the polychoric coefficients and an estimate of their asymptotic covariance matrix have to be computed. It cannot be generalized to unordered categorical outcomes. The consideration of varying T_i is in theory possible but has not been practically implemented. The strength of this approach lies in the possibility to include tobit models and to specify and test hypotheses not only about the mean structure but also about the covariance structure of the latent variables y_i^* given x_i in an almost unlimited way including the formulation of factor analytic measurement models. Consequently, it will depend on the specific model that a researcher is interested in, which model specification and estimation strategy is used.

4 Testing for state dependence in employment status

To illustrate the model in section 2 and the estimation strategies in section 3.1 and 3.2 we consider as dependent variable the employment status (1 = employed, 2 = unemployed) of a cohort of 1246 men of the GSOEP from 1985 to 1988, that is over four waves. The microeconomic specification of models for the employment status and a discussion of the relevant variables in the GSOEP data set is found in Flaig, Licht and Steiner (1992). We focus on a simple model for the disposition to become unemployed:

$$y_{i0}^{\star} = \kappa_0 + x_{i0}\beta_0 + \epsilon_{i0}^{\star} \tag{56}$$

$$y_{it}^{\star} = \kappa_t + \gamma_t y_{i,t-1} + x_{it}\beta_t + \epsilon_{it}^{\star}, \quad t = 1,2,3 \tag{57}$$

The model includes state dependence. The error term $\epsilon_{it}^{\star} = \alpha_i + \epsilon_{it}$ is decomposed into a random effects component α_i with $V(\alpha_i) = \sigma_{\alpha}^2$ and a component ϵ_{it} with $V(\epsilon_i) = \sigma_{\epsilon}^2$ which are uncorrelated. The initial state is modelled according to the proposal of Heckman discussed in subsection 3.2 as a function of explanatory variables x_{i0} and an error term $\epsilon_{i0}^{\star}$. This error term is allowed to correlate freely with $\epsilon_{it}^{\star}$, $t = 1,2,3$. The variance $V(\epsilon_{it}^{\star})$ may be written as $V(\epsilon_{it}^{\star}) = \sigma_{\alpha}^2 + \sigma_{\epsilon}^2$ under the model. Since the dependent variables are dichotomous, the variances $V(\epsilon_{i0}^{\star})$ and $V(\epsilon_{it}^{\star})$, $t = 1,2,3$ are set to 1. The covariance between ϵ_{it} and $\epsilon_{is}, t \neq s, t, s = 1,2,3$ is σ_{α}^2 under the model.

The explanatory variables in x_{i0} are ALD = duration of unemployment between 1974 and 1984 in months, ALDSQ = ALD squared and devided by 100, ALH = frequency of unemployed between 1974 and 1984, ALT = age in years, ALTSQ = ALT squared devided by 100, GZ = 0 if a person is not severely handicapped and 1 if he is, BB1 = 1 if the person has finished some professional education and 0 otherwise, BB2 = 1 if the person has a university degree and 0 otherwise, and BST2 = 1 if a person is a white collar employee and 0 otherwise. The explanatory variables in $x_{it}, t = 1,2,3$ are ESTL = employed status in the year before with 0 denoting employment and 1 denoting enemployment, NBDL = length of unemployment 1.5 years before and 0.5 years after the last interview, NBDLS = NBDL squared divided by 100, and FST = family status with 0 if a person is not married and 1 if the person is married. Additional variables are ALT, ALTSQ and BST2 which have been explained before. More complex models have been formulated by Flaig et al. (1992).

Note that the formulation in equation (57) assumes time varying coefficients which can later be restricted to be equal. In the first two stages of the MECOSA-estimation we estimate separate probit models for y_{i0} and $y_{it}, t = 1,2,3$ and the conditional polychoric correlation matrix without restriction. The results are given in table 1.

Table 1. Unrestricted probit models for unemployment status
(t - values in parentheses)

Explanatory variables	Wave 0	Wave 1	Wave 2	Wave 3
CONSTANT	3.409 (1.911)	-1.292 (-0.524)	-0.107 (-0.044)	-2.762 (-1.028)
ALD	0.108 (5.201)	—	—	—
ALDSQ	-0.081 (-1.702)	—	—	—
ALH	-0.049 (-0.833)	—	—	—
ESTL	—	0.715 (2.878)	0.467 (2.257)	1.071 (4.509)
NBDL	—	0.287 (6.852)	0.296 (6.257)	0.314 (6.499)
NBDLS	—	-0.667 (-3.433)	-0.736 (-3.836)	-0.847 (-4.271)
ALT	-0.272 (-2.711)	-0.071 (-0.536)	-0.121 (-0.912)	-0.005 (-0.036)
ALTSQ	0.336 (2.524)	0.120 (0.688)	0.159 (0.908)	0.051 (0.268)
FST	—	-0.192 (-0.746)	-0.252 (-1.034)	-0.451 (-1.682)
GZ	0.710 (3.443)	—	—	—
BB1	-0.149 (-1.170)	—	—	—
BB2	0.000 (0.001)	—	—	—
BST2	-0.375 (-1.996)	-0.749 (-3.560)	-0.656 (-2.514)	-0.665 (-2.348)
Loglikelihood	-194.530	-126.946	-124.865	-111.840
Correlations				
Wave 0	1.000			
Wave 1	-0.092	1.000		
Wave 2	-0.046	0.282	1.000	
Wave 3	-0.062	0.098	0.097	1.000

The equation for employment in wave 0 is not of substantive interest, although, taking over the model of Flaig et al (1992), we have tried to predict the employment status in wave 0 as well as possible with variables like duration and frequency of unemployment during the last 10 years before wave 0. Of substantive interest are the equations for the first, second and third panel wave. The variables age, age squared and being married turn out to be insignificant at the 0.05 test level as soon as the lagged employment status is taken into account. The lagged employment status, length of unemployment and its square as well as being a white collar employee have significant effects on employment status. The signs of the coefficients are all in the expected direction. The hypothesis that there exists no state dependence is rejected in this model. Note however, that the coefficients seem vary greatly across waves 1, 2 and 3. The effect of previous unemployment doubles between wave 2 and wave 3. Also the polychoric correlation coefficients beween waves 1, 2 and 3 which should be equal to the variance of the unobserved heterogeneity term α_i vary considerably. Hence, a model with equal coefficients for all three waves may seem to be misspecified.

This impression may however be misleading since the variance of the error terms has been set to 1 in each equation for identification in the univariate probit model. Hence, the regression coefficients are identified only up to scale. Assume that the variance of the error in wave 1 is equal to 1 and that the variance in wave 2 and wave 3 are ω_2^2 and ω_3^2. If $\beta_j^{(1)}, \beta_j^{(2)}$ and $\beta_j^{(3)}$ are the true regression coefficients of variable j in waves 1, 2 and 3, in univariate probit models only the parameters $\beta_j^{(1)}, \beta_j^{(2)*} = \beta_j^{(2)}/\omega_2$ and $\beta_j^{(3)*} = \beta_j^{(3)}/\omega_3$ are identified and estimable. Hence, the hypothesis of equality: $\beta_j^{(1)} = \beta_j^{(2)} = \beta_j^{(3)}$ cannot be tested, only the hypothesis of proportionality, that is: $\beta_j^{(2)*} = \beta_j^{(1)}/\omega_2$ and $\beta_j^{(3)*} = \beta_j^{(1)}/\omega_3$ can be tested. Setting $\lambda_2 = 1/\omega_2$ and $\lambda_3 = 1/\omega_3$ allows one to write the hypothesis of proportionality as: $\beta_j^{(2)*} = \lambda_2\beta_j^{(1)}$ and $\beta_j^{(3)*} = \lambda_3\beta_j^{(1)}$.

Under the model of unobserved heterogenity the covariance between the errors $\varepsilon_t^*, t = 1,2,3$ is given by σ_α^2. However, since the variance of ε_t^* is not identified, only the correlations of ε_t^* under the assumptions of proportionality can be written as: $\rho_{12} = \sigma_\alpha^2\lambda_2$, $\rho_{13} = \sigma_\alpha^2\lambda_3$ and $\rho_{23} = \sigma_\alpha^2\lambda_2\lambda_3$. The parameter estimates under the hypothesis of proportionality for waves 1, 2 and 3 are given in table 2.

Table 2. Partially restricted probit model for unemployment status (t - values in parentheses)

Explanatory variables	Wave 1	Wave 2	Wave 3
CONSTANT	-1.395 (-0.601)	-2.406 (-0.933)	-1.149 (-0.523)
ALT	-0.069 (-0.582)	—	—
ALTSQ	0.111 (0.711)	—	—
ESTL	0.826 (4.126)	—	—
NBDL	0.304 (5.303)	—	—
NBDLS	-0.791 (-3.853)	—	—
FST	-0.412 (-1.153	—	—
BST2	-0.641 (-1.580)	—	—
λ	1	0.682	1.062
McKelvey-Zavoina Pseudo R_{MZ}^2	0.517	0.352	0.559
σ_α^2	0.194 (0.875)		
χ^2 statistic	7.440	df	14

The χ^2 statistic of the minimum distance estimation of step 3 in MECOSA has a value of 7.44 with 14 degrees of freedom. Therefore the null hypothesis of proportionality cannot be rejected at the $\alpha = 0.05$ test level. The regression constants have not been restricted since the general level of unemployment can vary over the years. Only the variables employment status at time $t-1$, length of unemployment and its square are significant at the $\alpha = 0.05$ test level. Being married and being a white colour employee lower the probability of becoming unemployed but are not significant at the $\alpha = 0.05$ test level. To compare the regression coefficients of table 2 with the regression coefficients of table 1, one has to take the proportionality constants of wave 2 and 3 into account. $\lambda_2 = 0.682$ shows that the error variance of wave 2 is greater than the error variances of wave 1 and wave 3. Accordingly, one finds a smaller pseudo R^2. The pseudo R^2_{MZ} of McKelvey and Zavoina is discussed in Sobel and Arminger (1992). The unobserved heterogeneity parameter σ^2_α is not significant.

The estimation of the parameters of a panel model under the restrictions discussed here is not entirely straightforward in MECOSA. The difficulties are caused by the fact that the polychoric and polyserial covariances of the second estimation stage have to be computed under restrictions in the first estimation stage. In addition, the rows and columns of the estimated asymptotic covariance matrix of the estimated parameters have to be eliminated if they correspond to restricted parameters. Otherwise the t-values of the parameters will be incorrect. Hence, the MECOSA programs for the second and the third estimation step are given in the appendix.

5 Random effect models under sample selection bias

5.1 Metric response variables

This subsection is based on Ridder (1990) and Verbeek (1991, section 8.4). These authors analyse the estimation of a random effects model for a metric dependent variable

$$y_{it} = x_{it}\beta + \alpha_i + \epsilon_{it} \tag{58}$$

where y_{it} is only observed if the person i participates in wave t of the panel. Otherwise y_{it} is missing. The explanatory variables may or may not be missing. The participation is modelled with the unobserved variable

$$r^*_{it} = z_{it}\gamma + \xi_i + \eta_{it} \tag{59}$$

which is mapped onto $r_{it} = 1$ if $r^*_{it} > 0$ and onto $r_{it} = 0$ if $r^*_{it} \leq 0$. The value $r_{it} = 0$ indicates that y_{it} cannot be observed. Note that the explanatory variables for r^*_{it} are never missing.

The model for r^*_{it} is written by Ridder (1990) as follows

$$r^*_{it} = w_{it}\gamma_0 + y_{it}\gamma_1 + r_{i,t-1}\gamma_2 + \cdots + r_{i,1}\gamma_t + d_{it}\gamma_{t+1} + (1 - r_{i,t-1})\gamma_{t+2} + \tilde{\xi}_i + \tilde{\eta}_{it} \tag{60}$$

The explanatory variables for r^*_{it} are collected in the vector w_{it}. The disposition variables r^*_{it} may depend on the present value of y_{it} from the response equation. It may also depend on the former participation pattern collected in the sequence $\{r_{i,t-1}, \cdots, r_{i1}\}$ and the duration variable d_{it}:

$$d_{it} = \sum_{k=1}^{t-1} \prod_{s=1}^{k} r_{i,t-s} \tag{61}$$

If $\gamma_{t+2} = -\infty$, then non-participation is an absorbing state. $\tilde{\xi}_i$ denotes unobserved heterogeneity in the participation process and $\tilde{\eta}_{it}$ is the serially uncorrelated error term. Replacing y_{it} by $x_{it}\beta + \alpha_i + \epsilon_{it}$ yields the following model for the disposition to participate:

$$r_{it}^{*} = w_{it}\gamma_0 + x_{it}\beta\gamma_1 + \sum_{j=1}^{T-1} r_{i,t-j}\gamma_{j+1} + d_{it}\gamma_{t+1} + (1 - r_{i,t-1})\gamma_{t+2} + \alpha_i\gamma_1 + \tilde{\xi}_i + \epsilon_{it}\gamma_1 + \tilde{\eta}_{it} \qquad (62)$$

Collecting the explanatory variables $w_{it}, x_{it}, r_{i,t-j}, d_{it}$ and $(1 - r_{i,t-1})$ in a vector z_{it} and setting $\xi_i = \alpha_i\gamma + \tilde{\xi}_i$ and $\eta_{it} = \epsilon_{it}\gamma + \tilde{\eta}_{it}$ gives equation (59). The errors in y_{it} and r_{it}^{*} are obviously correlated if r_{it}^{*} depends on y_{it}.

The joint distribution of the error terms collected in the $T \times 1$ vectors ϵ_i and η_i and the scalars ξ_i and α_i has the structure:

$$\begin{pmatrix} \epsilon_i \\ \eta_i \\ \xi_i \\ \alpha_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\epsilon^2 I & & & \\ \sigma_{\epsilon\eta} I & \sigma_\eta^2 I & & \\ 0 & 0 & \sigma_\xi^2 & \\ 0 & 0 & \sigma_{\alpha\xi} & \sigma_\alpha^2 \end{pmatrix} \right) \qquad (63)$$

In the spirit of Heckman (1976, 1979) a correction term for the specification of the response variable may be derived. Here, we follow Verbeek (1991, ch.8.4) and derive some of his results explicitly. Some minor errors are corrected.

The history of panel participation is denoted by the vector $r_i = (r_{i1}, \ldots, r_{iT})$. The expected value of y_{it} conditional on r_i is given by:

$$E(y_{it}|r_i) = x_{it}\beta + E(\alpha_i|r_i) + E(\epsilon_{it}|r_i) \qquad (64)$$

$E(\alpha_i|r_i)$ may be written as an iterated expected value

$$E_\alpha(\alpha_i|r_i) = E_{(\xi+\eta)} E_\alpha(\alpha_i|(\xi_i + \eta_{it}|r_i)) \qquad (65)$$

The joint distribution of $(\xi_i + \eta_i, \alpha_i)$ is immediately found from equation (63):

$$\begin{aligned} \begin{pmatrix} \xi_i + \eta_i \\ \alpha_i \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\xi^2 11' + \sigma_\eta^2 I & \sigma_{\xi\alpha} 1 \\ \sigma_{\xi\alpha} 1' & \sigma_\alpha^2 \end{pmatrix} \right) \\ &= \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \right) \end{aligned} \qquad (66)$$

From standard results on the conditional normal distribution one finds:

$$E(\alpha_i|(\xi_i + \eta_i|r_i)) = \Omega_{21}\Omega_{11}^{-1} E(\xi_i + \eta_i|r_i) \qquad (67)$$

Since the matrix Ω_{11} is defined as $\sigma_\xi^2 11' + \sigma_\eta^2 I$, the updating formula for the inverse of a matrix

$$[A + bb'] = A^{-1} - [\frac{1}{1 + b'A^{-1}b}]A^{-1}bb'A^{-1} \tag{68}$$

with $A = \sigma_\eta^2 I$ and $b = \sigma_\xi 1$ may be applied (cf. Greene, 1990 ch. 2.5). With $\Omega_{21} = \sigma_{\xi\alpha}1'$ one finds that $E(\alpha_i|r_i) = E_{(\xi+\eta)}E_\alpha(\alpha_i|(\xi_i + \eta_i|r_i)) = \sigma_{\xi\alpha}A_{1i}$ with

$$A_{1i} = \frac{1}{\sigma_\eta^2 + T\sigma_\xi^2}\sum_{t=1}^{T} E(\xi_i + \eta_{it}|r_i) \tag{69}$$

The same technique may be used to write $E(\epsilon_{it}|r_i)$ as

$$E(\epsilon_{it}|r_i) = E_{(\xi+\eta)}E_\epsilon(\epsilon_{it}|(\xi_i + \eta_{it}|r_i)) \tag{70}$$

and to derive the joint distribution of $(\xi_i + \eta_i, \epsilon_i)$. As a result, one finds that $E(\epsilon_{it}|r_i) = \sigma_{\epsilon\eta}A_{2it}$ with

$$A_{2it} = \frac{1}{\sigma_\eta^2}[E(\xi_i + \eta_{it}|r_i) - \frac{\sigma_\xi^2}{\sigma_\eta^2 + T\sigma_\xi^2}\sum_{t=1}^{T} E(\xi_i + \eta_{it}|r_i)] \tag{71}$$

The expected values $E(\alpha_i|r_i) = \sigma_{\xi\alpha}A_{1i}$ and $E(\epsilon_{it}|r_i) = \sigma_{\epsilon\eta}A_{2it}$ can be used as correction terms in equation (58) with $\sigma_{\xi\alpha}$ and $\sigma_{\epsilon\eta}$ as parameters that must be estimated. Before this can be done, the expected value $E(\xi_i + \eta_{it}|r_i)$ must be derived.

First, one notes that $E(\xi_i + \eta_{it}|r_i) = E_\xi(\xi_i|r_i) + E_\xi E_\eta(\eta_{it}|\xi_i, r_i))$ Hence, we find

$$E(\xi_i + \eta_{it}|r_i) = \int_{-\infty}^{\infty}(\xi_i + E_\eta(\eta_{it}|\xi_i, r_i))f(\xi_i|r_i)d\xi_i \tag{72}$$

The expected value inside the integral is given by

$$E(\eta_{it}|r_i, \xi_i) = (2r_{it} - 1)\sigma_\eta \frac{\varphi((z_{it}\gamma + \xi_i)/\sigma_\eta)}{\Phi((2r_{it} - 1)(z_{it}\gamma + \xi_i)/\sigma_\eta)} \tag{73}$$

This result may be derived by setting r_{it} first to 1 and then to 0. For $r_{it} = 1$ we have

$$\begin{aligned} E(\eta_{it}|r_{it} = 1, \xi_i) &= E(\eta_{it}|\eta_{it} > -z_{it}\gamma - \xi_i, \xi_i) \\ &= \left[\Phi(\frac{z_{it}\gamma + \xi_i}{\sigma_\eta})\right]^{-1} \int_{(-z_{it}\gamma-\xi_i)}^{\infty} \eta_{it}\varphi(\eta_{it}|0, \sigma_\eta^2)d\eta_{it} \end{aligned} \tag{74}$$

where $\varphi(.|0, \sigma_\eta^2)$ is the normal density of η_{it}. $\Phi\ ((z_{it}\gamma + \xi_i)/\sigma_\eta)$ is the probability $P(\eta_{it} > -z_{it}\gamma - \xi_i)$ which is used because the variable η_{it} conditional on $r_{it} = 1$ and ξ_i is truncated. With the variable transformation $z = \eta/\sigma_\eta$ and the well known result (cf. Ronning 1991) that

$$\int_a^\infty z\varphi(z)dz = \varphi(a)$$

where $\varphi(.)$ is the standard normal density one gets the result of equation (73) for $r_{it} = 1$. The same exercise for $r_{it} = 0$ gives the desired result.

The density $f(\xi_i|r_i)$, that is used for the integral of equation (72) may be derived by noting that

$$f(\xi_i|r_i) = f(\xi_i, r_{i1}, \cdots r_{iT}) / \int_{-\infty}^{\infty} f(\xi, r_{i1}, \cdots, r_{iT}) d\xi \tag{75}$$

The common density $f(\xi_i, r_{i1}, \cdots, r_{iT})$ may be written as

$$f(r_{i1}, r_{i2}, \cdots, r_{iT}|\xi_i) \cdot f(\xi_i) = \prod_{t=1}^{T} \Phi\left((2r_{it} - 1)\frac{z_{it}\gamma + \xi_i}{\sigma_\eta}\right) \frac{1}{\sigma_\xi}\varphi(\frac{\xi_i}{\sigma_\xi}) \tag{76}$$

where $\varphi(.)$ is again the standard normal density.

Note that the computation of the correction terms in a panel is, as opposed to the case of the Heckman correction for sample selection in a cross-sectional sample, quite cumbersome. First, the parameters of γ have to be estimated using a multivariate probit model for ML estimation or MECOSA for the three stage marginal ML estimation. Second, the denominator of equation (75) has to be calculated using numerical integration. Third the expected value $E(\xi_i + \eta_{it}|r_i)$ of equation (72) has to be computed using again numerical integration. Finally A_{1i} and A_{2it} can be computed and inserted into equation (58) as additional explanatory variables that correct for sample selection in the panel. Since the length of the vector of observed variables y_{it} may differ for each person, the dependent variable in the multivariate regression to estimate the parameters $\beta, \sigma_{\xi\alpha}, \sigma_{\eta\epsilon}, \sigma_\alpha^2$ and σ_ϵ^2 cannot be computed with standard regression methods and programs. Here, procedures such as the multiple group option in LISREL or the methods discussed in Arminger and Sobel (1990) have to be applied.

Note that the true values of γ and σ_ξ^2 are not known except for the restriction that $\sigma_\xi^2 + \sigma_\eta^2 = 1$. Hence, the estimates of the asymptotic standard errors for $\beta, \sigma_{\xi\alpha}$ and $\sigma_{\eta\epsilon}$ would have be corrected like in the cross-sectional case (cf. Heckman 1976). Because of the cumbersome computation of the correction terms, Verbeek (1991, ch.9) proposes a simple to compute test procedure based on a Hausman test statstic to check for non-ignorable sample selection bias before computation of the correction terms. This Hausman test is based on differences between the estimates of β in fixed and random effects models.

5.2 Non–metric response variables

Let now the dependent variables y_{it} be censored, dichotomous or ordered categorical, but the structure of equation (58) is retained for the unobserved variable.

$$y_{it}^\star = x_{it}\beta + \alpha_i + \epsilon_{it} \tag{77}$$

The unobserved variable is connected to the observed variable through one of the observation rules discussed in section 2.2. The response process is structured in the same way as before and z_{it} is observed for all time points $t = 1, \ldots, T$. The argument of the last section for the derivation of $E(\alpha_i|r_i)$ and $E(\epsilon_{it}|r_i)$ is exactly the same. Hence, for all observed data points y_{it}, the misspecification induced by panel sample selection bias can be corrected by estimating the parameters $\beta, \sigma_{\xi\alpha}$ and $\sigma_{\epsilon\eta}$ of the equation

$$y_{it}^\star = x_{it}\beta + \sigma_{\xi\alpha}A_{1i} + \sigma_{\epsilon\eta}A_{2it} + \alpha_i + \epsilon_{it} \tag{78}$$

Since y_{it} is only observed if $r_{it} = 1$, the sequence of observations $\{y_{it}, \cdots, y_{it}\}$ will again be different for each person. Sample elements, that are only observed in y_{it} up to a time point $T' < T$ are dropped from the data set after T'. For practical estimation of the parameters in the last estimation, one must first compute consistent estimates of A_{1i} and A_{2it} and then augment the list of explanatory variables correspondingly. Second, the missing data patterns in y_{it} must be identified and the individuals, that have the same number of panel waves in which y_{it} has been observed, are grouped together. This can be done since the same model holds for all waves. For the first wave, there is no need to put in the Heckman correction variables A_{1i} and A_{2it}. There are at most T different groups. For each group, MECOSA may be used to estimate the reduced form parameters of $y^{\star}_{it}$. Finally, these estimates are pooled together in the multiple group option of the third stage of MECOSA to yield the final estimates of $\beta, \sigma_{\xi\alpha}$ and $\sigma_{\alpha\eta}$ of equation (78).

6 Modeling absorbing states

We consider now a version of the model

$$y^{\star}_{it} = x_{it}\beta + \alpha_i + \epsilon_{it}$$

with dichotomous or ordered categorical observation rules for $y^{\star}_{it}$ where the last state of the observed dependent variable y_{it} is an absorbing state. The variable y_{it} with categories $\{1, 2, \ldots, K\}$ remains in the state $y_{it} = K$ for all waves $t > T'$ if it has switched to state K at time T'. The unobserved heterogeneity is captured in the random effects α_i with $V(\alpha_i) = \sigma^2_\alpha$ which are uncorrelated with the error terms ϵ_{it} with $V(\epsilon_{it}) = 1$.

Such processes occur quite often in panel data as shown by examples from the GSOEP. A first example is given by Helberger, Rendtel and Schwarze (1992) for the transition of adolescents from high school into vocational training or participation in the labor force where the dependent variable is dichotomous and only two time points are considered. A second example is the ordinal variable *need for nursing care* with the states {no nursing care necessary, nursing care necessary, dead} which plays currently a great role in the discussion about compulsory insurance for nursing care in Germany (Pischner and Wagner 1991). In both examples it is necessary to take the unobserved heterogeneity into account. If this is not deemed necessary one can immediately estimate β by pooling the data and computing a standard ordered categorical probit.

To derive the likelihood function for the case of unobserved heterogeneity the approach of Helberger, Rendtel and Schwarze (1992, sec. 5) is generalized to deal with absorbing states not only for dichotomous, but also for ordinal variables. Let the dummy variable $d_{itk} = 1$ if y_{it} falls in category $k \in \{1, \ldots, K\}$ of a dichotomous or ordered categorical variable with the restriction that $\sum_{k=1}^{K} d_{itk} = 1$.

The probability that $y_{it} = k$, conditional on x_{it} and α_i is given the ordinal probit model with thresholds $\tau_0 = -\infty, \tau_1 = 0, \tau_2, \ldots, \tau_{K-1}, \tau_K = \infty$, by:

$$P(y_{it} = k | x_{it}, \alpha_i) = \begin{cases} \Phi(-x_{it}\beta - \sigma_\alpha\eta) & ,k = 1 \\ \Phi(\tau_k - x_{it}\beta - \sigma_\alpha\eta) - \Phi(\tau_{k-1} - x_{it}\beta - \sigma_\alpha\eta) & ,k = 2, \ldots, K \\ 1 - \Phi(\tau_{K-1} - x_{it}\beta - \sigma_\alpha\eta) & ,k = K \end{cases} \tag{79}$$

Note that the variance of ϵ_{it} has been set to 1 for all $t = 1, \ldots, T$ in the conditional model. The thresholds $\tau_k, k = 2, \ldots, K - 1$ are assumed to be equal for all waves and must be estimated from the data. They are collected in a vector $\tau \sim (K-1) \times 1$. The value of α_i is written as $\sigma_\alpha \eta_i$ where η_i follows a standard normal distribution. For each wave t the conditional likelihood function of individual i is given by:

$$L_{it}(\tau, \beta, \sigma_\alpha | \eta) = \prod_{k=1}^{K} P(y_{it} = k | x_{it}, \alpha_i = \sigma_\alpha \eta_i)^{d_{itk}} \tag{80}$$

The tth wave contributes to the conditional likelihood function of all waves only if an individual i is still in the risk set at time t which is determined by the value of the dummy variable $c_{it} = \prod_{s=1}^{t-1}(1 - d_{isK})$ for $t = 2, \ldots, T$. If $c_t = 0$ then the absorbing state has been reached before wave t. If $c_{it} = 1$, then wave t still contributes to the likelihood function. Since every individual is in the risk set at wave 1, c_{i1} is set to 1. Consequently, the marginal likelihood function of individual i for all waves is given by:

$$L_i(\tau, \beta, \sigma_\alpha) = \int_{-\infty}^{\infty} \prod_{t=1}^{T} \left[L_{it}(\tau, \beta, \sigma_\alpha | \eta)\right]^{c_{it}} \varphi(\eta) d\eta \tag{81}$$

where $\varphi(.)$ is the univariate normal density. Note that in this way of writing the likelihood function, the number of waves does not differ across persons. For the missing data y_{it} and x_{it} of individuals that have dropped out of the risk set, any numerical values can be used. Hence, the standard Gauss-Hermite numerical integration procedure (cf. Butler and Moffit 1982, Davis and Rabinowitz 1984) can be used for each individual to compute the integral in the last equation. Unfortunately, this model cannot be embedded in MECOSA. Hence, no standard software is available yet to estimate the parameters of such models. To estimate the model discussed below, a GAUSS program has been written by Andreas Schepers.

To illustrate the model with an absorbing state we analyze a data set of 388 persons that have been in need of nursing care in the first wave of the GSOEP in 1984. Persons under 21 have been dropped from this subset because of their special characteristics. The dependent variable is *nursing care* defined in the beginning of this section. We consider nursing care as an ordered categorical variable with {dead} as an absorbing state. The model is a simple random effects model

$$y_{it}^{*} = x_{it}\beta + \alpha_i + \epsilon_{it}, \quad t = 1, \ldots, 6 \quad , \tag{82}$$

where $V(\alpha_i) = \sigma_\alpha^2$ and $V(\epsilon_{it}) = \sigma_\epsilon^2$ and α_i and ϵ_{it} are uncorrelated with x_{it} and with each other over all $t = 1, \ldots, T$. The variance $\sigma_\epsilon^2 = 1$ because the model for y_{it} given α_i is assumed to be a univariate probit model. Hence the size of σ_α^2 can be compared directly with σ_ϵ^2. The regression constant is set to 0. Instead, both threshold values, τ_1 and τ_2 are estimated. The first time point is 1985. The explanatory variables are WEIBL = 0 if person is male and 1 if person is female (time constant), ALTERD100 = age divided by 100 (time constant), ZUFR = satisfaction with health measured on a scale between 0 (complete dissatisfaction) and 10 (complete satisfaction) at time $t-1$ (time varying), PARTNER = 0 if there is no husband or wife and 1 if there is a husband or wife at time $t-1$ (time varying), EINK = average monthly income in DM divided by 10000 (time varying), OLD = 0 if a person is younger than 70 years in 1984 and 1 if a person is older than 70 years (time constant). The results of the analysis are given in table 3.

Table 3. Parameter estimates of a model for the stability of the need for nursing care in the GSOEP		
Variable	Parameter estimate	t-value
Threshold τ_1	-0.435	-3.404
Threshold τ_2	1.325	64.517
WEIBL	0.092	0.998
ALTERD100	0.724	5.307
ZUFR	-1.247	-10.755
PARTNER	-0.453	-5.947
EINK	0.264	1.543
OLD	-0.414	-3.839
σ_α^2	0.220	10.149

The parameter estimates show the expected effects. Increasing age, dissatisfaction with health status at time $t-1$ and no partner at time $t-1$ generate higher values of y_{it}^* with a greater likelihood that nursing care is necessary or that a person dies. The best predictor seems to be satisfaction with one's health status. The variable WEIBL is not significant at the $\alpha = 0.05$ test level. The variable OLD has an unexpected sign that we cannot explain since the interaction between WEIBL and OLD turned out to be not significant at the $\alpha = 0.05$ test level. In a more elaborate model, the distance to one's life expectancy should be taken into account. The estimate of σ_α^2 turns out to be significant, that is we observe a small but significant effect of unobserved heterogeneity.

References

Andersen, E. B. (1973), *Conditional Inference and Models for Measuring*, Copenhagen: Mentalhygiejnisk Forsknings Institut.

Arminger, G. (1987), Misspecification, Asymptotic Stability and Ordinal Measurements in Models for the Analysis of Panel Data, *Sociological Methods and Research*, Vol.15,No.3: 336-348.

Arminger, G. and Sobel, M. (1990), Pseudo maximum likelihood estimation of mean- and covariance structures with missing data. *Journal of the American Statistical Association (JASA)*, 85, 195–203.

Arminger, G. and Ronning, G. (1991), Ein Strukturmodell für Preis-, Produktions- und Lagerhaltungsentscheidungen von Firmen, *IFO-STUDIEN, Zeitschrift für empirische Wirtschaftsforschung*, 37, 229-254.

Bhargava, A. and Sargan, J.D. (1983). Estimating Dynamic Random Effect Models from Panel Data Covering Short Time Periods, *Econometrica*, 51, 1635-1659.

Butler, J. and Moffitt, R. (1982), A Computationally Efficient Quadrature Procedure of the One-Factor Multinomial Probit Model, *Econometrica*, 50, 761-764.

Chamberlain, G. (1984). Panel Data, in Griliches, Z. and Intriligator, M.D. (eds.), *Handbook of Econometrics*, Vol. II, Amsterdam: North Holland, 1247–1318.

Davis, P. J. and Rabinovitz, P. (1984), *Methods of Numerical Integration*, Orlando: Academic Press.

Flaig, G., Licht, G. and Steiner, V. (1992). Testing for State Dependence Effects in a Dynamic Model of Male Unemployment Behaviour, paper prepared for the workshop "Arbeitslosigkeits-, Beschäftigungs- und Lohndynamik auf der Basis von Mikrodaten", Zentrum für Europäische Wirtschaftsforschung, February 27–28, 1992.

Gourieroux, C., Monfort, A. and Trognon, A. (1984). Pseudo Maximum Likelihood Methods: Theory, *Econometrica*, 52, 681–700.

Greene, W. H. (1990), *Econometric Analysis*, New York: MacMillan.

Hamerle, A. and Ronning, G. (1993), Analysis of Discrete Panel Data, forthcoming in G. Arminger, C. C. Clogg and M. E. Sobel (eds.), *Handbook of Statistical Modeling for the Behavioral Sciences*, New York: Plenum.

Heckman, J. J. (1976), The Common Structure of Staistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models, *Annals of Economic and Social Measurement*, 475-492.

Heckman, J. J. (1979), Sample Selection Bias as a Specification Error, *Econometrica*, 47, 153-161.

Heckman, J. J. (1981a), Statistical Models for Discrete Panel Data, in C. F. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*, 114-178.

Heckman, J. J. (1981b), The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Stochastic Process, in C. F. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*, 179-195.

Helberger, C., Rendtel, U. and Schwarze, J. (1992), Der Weg in die betriebliche Berufsausbildung und in die Erwerbstätigkeit, *Discussion paper No. 54*, Deutsches Institut für Wirtschaftsforschung, Berlin.

Hsiao, C. (1986), *Analysis of Panel Data*,Cambridge, Mass.: Cambridge University Press.

Keane, M.P. and Runkle, D.E. (1992). On the Estimation of Panel-Data Models With Serial Correlation When Instruments Are Not Strictly Exogenous, *Journal of Business & Economic Statistics*, 10, No. 1, 1–29.

Kiefer, J. and Wolfowitz, J. (1956), Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters, *Annals of Mathematical Statistics*, 27, 887-906.

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

Liang, K-Y., Zeger, S.L. and Qaqish, B. (1992). Multivariate Regression Analysis for Categorical Data (with discussion), *Journal of the Royal Statistical Society*, B, 42, 109–142.

Maddala, G.S. (1987). Limited Dependent Variable Models Using Panel Data, *Journal of Human Resources*, XXII, 3, 307–336.

McKelvey, R.D. and Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology 4*, 103–120.

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

Muthén, B. (1988). *LISCOMP — Analysis of Linear Equations Using a Comprehensive Measurement Model.* Scientific Software, Inc., Mooresville.

Nelson, F.D. (1976). On a general computer algorithm for the analysis of models with limited dependent variables. *Annals of Econometrics and Social Measurement 5*, 493–509.

Neyman, J. and Scott, E. (1948). Consistent Estimates Based on Partially Consistent Observations, *Econometrica*, 16, 1-32.

Pischner, R. and Wagner, G. (1991), Pflegebedürftigkeit im Zeitverlauf — Ergebnisse des Sozio-ökonomischen Panels 1985 – 1990, in *Hilfsbedarf und Behinderungen bei Alltagsaktivitäten in ausgewählten Bevölkerungsgruppen*, Forschungsprojekt im Auftrag des Bundesministers für Familie und Senioren, München, 85–103.

Prentice, R.L. and Zhao, L.P. (1991). Estimating Equations for Parameters in Means and Covariances of Multivariate Discrete and Continuos Responses, *Biometrics*, 47, 825–839.

Projektgruppe Panel (1991), Das Sozio-ökonomische Panel (SOEP) im Jahr 1990/91, *Vierteljahreshefte zur Wirschaftsforschung*, 3-4, 146–155.

Ridder, G. (1990), Attrition in Multi-wave Panel Data, in J. Hartog, G. Ridder and J. Theeuwes (eds.), *Panel Data and Labor Market Studies*, Amsterdam: North Holland, 45-68.

Ronning, G. (1991), *Mikroökonometrie*, Heidelberg: Springer Verlag.

Rosett, R.N. and Nelson, F.D. (1975). Estimation of the two-limit probit regression model. *Econometrica 43*, 141–146.

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

Schepers A., Arminger G. and Küsters U. (1991), The Analysis of Non-Metric Endogenous Variables in Latent Variable Models: The MECOSA Approach, in P. Gruber (ed.), *Econometric Decision Models: New Methods of Modeling and Applications*, Springer Verlag, Heidelberg, 1991, 459-472.

Schepers, A. and Arminger, G. (1992), *MECOSA: A Program for the Analysis of General Mean- and Covariance Structures with Non-Metric Variables, User Guide*, SLI-AG, Züricher Str. 300, CH-8500 Frauenfeld, Switzerland.

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

Sobel, M. and Arminger, G. (1992), Modeling Household Fertility Decisions: A Nonlinear Simultaneous Probit Model, *Journal of the American Statistical Association*, 87, 38-47.

Stewart, M.B. (1983). On least squares estimation when the dependent variable is grouped. *Review of Economic Studies L*, 737–753.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica 26*, 24–36.

Verbeek, M. J. (1991), *The Design of Panel Surveys and the Treatment of Missing Observations*, Dissertation at the Catholic University Brabant.

Wagner, G. (1992), Personal Communication, Deutsches Institut für Wirtschaftsforschung, Berlin.

Zeger, S.L. and Liang, K-Y (1986). Longitudinal Data Analysis for Discrete and Continuous Outcomes, *Biometrics*, 42, 121–130.

Zeger, S.L., Liang, K-Y and Albert, P.S. (1988). A Generalized Estimating Equation Approach, *Biometrics*, 44, 1049–1060.

Appendix

In this appendix, the MECOSA program to estimate the restricted model in section 4, is briefly described. The meaning of the names of the variables have been explained in section 4. The suffixes A, B, C and D in the list of names denote the panel waves 0, 1, 2 and 3. Since the program uses only the indices of the vector of variables, the correspondence between variables and index numbers is given in table A1.

Table A1. Correspondence of index numbers and variable names

wave 0		wave 1		wave2		wave3	
1	EST_A	11	EST_B	17	EST_C	23	EST_D
2	ALD_A	12	ESTL_B	18	ESTL_C	24	ESTL_D
3	ALDSQ_A	13	NBDL_B	19	NBDL_C	25	NBDL_D
4	ALH_A	14	NBDLS_B	20	NBDLS_C	26	NBDLS_D
5	ALT_A	15	FST_B	21	FST_C	27	FST_D
6	ALTSQ_A	16	BST2_B	22	BST2_C	28	BST2_D
7	GZ_A						
8	BB1_A						
9	BB2_A						
10	BST2_A						

The estimates shown in table 1 of section 4 have been obtained by regressing the dependent variable employment status on the regressors for each wave in four separate runs of the first step of the MECOSA program. These estimates have been augmented by zeros to yield a restricted parameter matrix for all four dependent variables jointly according to section 4.6 in the MECOSA manual.

The polychoric correlations of the reduced form and the estimate of the asymptotic covariance matrix of the reduced form parameter estimates are computed as the second estimation step with the following MECOSA program:

```
/**   M E C O S A  -  Command File  **/

let __srf= 0.3 -0.3;    @ starting values for regula falsi procedure @

let NAME    = HECK;      @ input data file @
let END_VEC= 1 11 17 23;  @ vector of endogenous variables @
let END_TYP= OR OR OR OR    ;
let EX_VEC =
 2  3  4  5  6  7  8  9  10
12 13 14 15 16
18 19 20 21 22  @ vector of exogenous variables @
24 25 26 27 28;
let GAM     =Y;
let WEIGHT =.        ;
let TAU_N=
 1 1 1 1 ;
let TAU=
 0.000000
 0.000000
 0.000000
 0.000000 ;
let SIGMA2 =
 1.000000
 1.000000
 1.000000
 1.000000 ;

let BETA[ 4, 25]=
3.40884
 0.108498  -0.0811215 -0.0494523 -0.271570   0.335869   0.709946
-0.198878   0.000299084 -0.375160
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0

-1.292089
0 0 0  -0.071350  0.120203  0 0 0 0
0.715143   0.287117  -0.667141  -0.192329  -0.749421
0 0 0 0 0
0 0 0 0 0

-0.106953
0 0 0  -0.120715  0.159380  0 0 0 0
0 0 0 0 0
0.467492   0.295880  -0.735902  -0.251616  -0.656112
0 0 0 0 0

-2.762146
0 0 0  -0.005224  0.050875  0 0 0 0
0 0 0 0 0
0 0 0 0 0
1.070616   0.313798  -0.846792  -0.451427  -0.665418
;

RUN _STEP2A.PRG;
```

The program treats the parameters that have been restricted to zeros as free parameters and computes therefore the asymptotic variances and covariance for these parameters. Since only the covariance matrix for the non-restricted parameters of the reduced form is needed in the third estimation stage, the covariances for the restricted parameters must be taken out. This is done in the next MECOSA command file before the actual estimation is carried out.

```
/***************************************************************************/
/**                    M E C O S A  -  Command File                       **/
/**           (STEP 3: Estimation of Fundamental Parameters)              **/
/***************************************************************************/

COMMENT= "Estimation with weight matrix  of unrestricted parameters";
#lineson;
/** Output File **/
  let NAME = HECK3;

/** Number of Groups **/
  N_GROUPS=1;

/**************   Loading Reduced Form Estimators   *******************/

/** GROUP 1 **/

{NAME_1,CASES_1,TAU_H_1,GAM_H_1,PI_H_1,SIG_H_1,W_MAT_1}
= LOADRP(inpath $+ "HECK2A1    ");

/*
The following commands produce the correct weight matrix for
the third estimation step for the case that the first two steps
were estimated under restrictions. Here, the restrictions are
characterized by the fact, that parameters are set to zero.
The following typical pattern is formed:

A_11   0    A_13
 0     1     0
A_31   0    A_33

*/

screen off;
output file = heckw.out on;
ro = rows(w_mat_1);       @ ro = number of rows of weight matrix @
w = eye(ro);              @ w = identity matrix @

/*
The reduced form parameters are computed either from
KAP_H_1 if it is available, otherwise from the *.FMT file
of parameters in the LOADRP procedure.
*/

/*
```

```
kap = kap_h_1;            @ vector of reduced form parameters @
*/

load _x = heck2a1;
rom = ro.*(ro+1)./2;      @ rom = number of elements in weight matrix @
rox = rows(_x)-rom;
"Number of reduced form parameters ";; rox;
kap = packr(_x[8:rox]);  @ extract the reduced form parameter vector @
rok=rows(kap);
"Number of reduced form parameters without missings";; rok;

w1 = 1 - (kap .== 0);
exvec = seqa(1,1,ro);    @ exvec = 1 2 3 ... ro @
exvec = exvec.*w1;       @ exvec = 1 2 3 ... ro  if param. is not restricted@

?;
"The indices of the parameters NOT restricted to zero are";
?;
format /rd 6,0;
exvec';
?;
no = miss(exvec,0);
no = packr(no);
no = rows(no);

"The number of reduced form parameters is:";; no;

format /rd 10,7;

w = diagrv(w, (1-exvec .> 0));
                       @ The diagonal values of w are set to zero @
                       @ for the unrestricted parameters          @


w = w+(w1.*w1').*w_mat_1;@ w contains the matrix w_hat for the @
                         @ unrestricted parameters in the above form @

output file = heckw.out off;

w_mat_1 = w;
screen on;




/*****************************************************************/
/**************            Model  Specification       ***********/
/*****************************************************************/

/*---------------------  Help Procedures  ---------------------*/

proc (1) = _ALPHA(theta);
        local a,t;
        t=theta;
        a=t[1]|t[2]|t[3]|t[4];
retp(a);
```

```
endp;

proc (1)=_GA(theta);
     local ga,g0,g1,g2,g3,g4,t;
     t=theta;
     g0=zeros(1,3) ~ t[14] ~ t[15] ~ zeros(1,4);
     g1=t[5:13]';
     g2=t[16:20]';
     g3=t[16:20]';
     g4=t[16:20]';
     g1=g1~zeros(1,15);
     g2=g0~g2~zeros(1,10);
     g3=g0~zeros(1,5)~g3~zeros(1,5);
     g4=g0~zeros(1,10)~g4;
     ga=g1|g2|g3|g4;
retp(ga);
endp;


proc (1)=_S1(theta);
        local s1,t;
        t=theta;
        s1 = (miss(0,0)~t[21]~t[22]~t[23])|
             (t[21]~miss(0,0)~t[24]~t[24])|
             (t[22]~t[24]~miss(0,0)~t[24])|
             (t[23]~t[24]~t[24]~miss(0,0));
        s1 = omres(_LA(theta),s1, sig_h_1);
retp(s1);
endp;

proc (1)=_LA(theta);
        local la,t;
        t=theta;
        la = (1  ~   0   ~    0    ~   0)|
             (0  ~   1   ~    0    ~   0)|
             (0  ~   0   ~t[25]    ~   0)|
             (0  ~   0   ~    0    ~ t[26]);
retp(la);
endp;


/*---------------------------------------------------------------*/
/*------------------   MECOSA Parameterisation   ---------------*/
/*---------------------------------------------------------------*/

/** GROUP 1 **/

  proc (1)=TAU_1(theta);
  retp(miss(0,0));
  endp;


  proc (1)=GAM_1(theta);
       local g;
       g = _LA(theta)*_ALPHA(theta);
  retp(g);
```

```
  endp;

  proc (1)=PI_1(theta);
       local p;
       p = _LA(theta)*_GA(theta);
  retp(p);
  endp;

  proc (1)=SIG_1(theta);
       local s;
       s=_LA(theta)*_S1(theta)*_LA(theta)';
  retp(s);
  endp;


/*----------------------   Starting Values  -----------------------*/

  let theta_s =
@ alpha @
@ ga1   @
@ ga2 - ga6  @

@ s1   @
@ la   @

 2.96
-1.41
-2.39
-1.1
0.108  -0.0811   -0.54   -0.244    0.3359   0.7099 -0.199 0.0003 -0.43
-0.0658    0.1102
0.85    0.2989   -0.77   -0.39   -0.6903
-0.0917   -0.0457   -0.0618   0.196
0.69 1.102
;

/* Output of Help Procedures */
  u_procs =("ALPHA" ~ &_ALPHA)
           |("GAMMA" ~ &_GA )
           |("S1"    ~ &_S1)
           |("LAMBDA" ~ &_LA);
/***********************  START of Step 3  ************************/
  RUN _STEP3.PRG;
```