DIW Berlin

German Institute
for Economic Research

Discussion Papers

564

Martin Spiess

On the Returns to Occupational Qualification in Terms of Subjective and Objective Variables: A GEE-type Approach to the Estimation of Two-Equation Panel Models

Berlin, March 2006

Opinions expressed in this paper are those of the author and do not necessarily reflect views of the Institute.

IMPRESSUM

DIW Berlin
German Institute for Economic Research
Königin-Luise-Str. 5
14195 Berlin
Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
www.diw.de

ISSN print edition 1433-0210
ISSN electronic edition 1619-4535

On the Returns to Occupational Qualification in Terms of Subjective and Objective Variables: A GEE-type Approach to the Estimation of Two-Equation Panel Models

Martin Spiess*
SOEP, DIW Berlin, Germany

Abstract

This article proposes an estimation approach for panel models with mixed continuous and ordered categorical outcomes based on generalized estimating equations for the mean and pseudo-score equations for the covariance parameters. A numerical study suggests that efficiency can be gained as concerns the mean parameter estimators by using individual covariance matrices in the estimating equations for the mean parameters. The approach is applied to estimate the returns to occupational qualification in terms of income and perceived job security in a nine-year period based on the German Socio-Economic Panel (SOEP). To compensate for missing data, a combined multiple imputation/weighting approach is adopted.

Key words: generalized estimating equations, mean and covariance model, multiple imputation, pseudo-score equations, status inconsistency, weighting

*Martin Spiess, SOEP, DIW Berlin, Königin-Luise-Str. 5, 14195 Berlin, Germany, mspiess@diw.de.

1 Introduction

This paper describes the estimation of a panel model with mixed continuous and ordered categorical outcomes. The proposed estimation approach was designed to achieve two ends: first to study the returns to occupational qualification (university, apprenticeship, other completed training; reference category: none) in terms of objective and subjective gratification variables, i.e., in terms of the dependent variables income (log of monthly gross real labor income) and perceived job security (very concerned about job security, somewhat concerned, not concerned at all). Second, it was designed to answer the question of whether both outcomes depend on common unobserved individual and time-invariant variables, given the covariates explicitly controlled for in the regression model.

A growing body of evidence in economics and related fields supports the view that in addition to routinely used objective variables, such as income, subjective variables can be understood as gratification variables as well (e.g., Easterlin, 2002; Diener and Seligman, 2004). According to this view and following the suggestion, e.g., of Zimmermann (1985) to consider subjective in addition to objective variables in research on status inconsistency, this paper understands both gratification variables and occupational qualification as factors that define social positions. This means, for example, that when occupational qualification shows a declining effect on income over time, unusual combinations of income and occupational qualification become more likely to emerge. An increasing proportion of such status inconsistencies may help explain deviating behavior, voting decisions and social change (e.g., Geschwender, 1967a, 1967b). Furthermore, given the importance of subjective variables, represented by perceived job security in this paper, it is even more important to examine their relation to established objective gratification variables like income.

The analysis is based on data from the German Socio-Economic Panel (SOEP; www.diw.de/soep). The SOEP is a longitudinal data set of individuals aged 16 and older living in private households in Germany and surveyed on a yearly basis (SOEP Group, 2001). It consists of several subsamples, the first two of which started in 1984. Information is collected about the household as a whole and additionally about each household member. Topics covered by the SOEP include, among others, occupation, employment, earnings, household composition and housing, socio-demographic variables and health, as well as subjective variables such as worries about certain aspects of life. Hence the SOEP is particularly suitable for studying income dynamics, the dynamics of subjective gratification variables, as well as their possible interrelations.

The estimation of models with both continuous and categorical, in most cases binary, outcomes has in recent years attracted increasing interest in various areas of research. Since the estimation of these models based on likelihood approaches (e.g., Fitzmaurice and Laird, 1995; Regan and Catalano, 1999; Gueorguieva and Agresti, 2001) and Bayesian approaches (e.g., Dunson, 2000) is rather cumbersome due to excessive computational burdens (Sammel et al., 1997), alternative approaches have been proposed (e.g., Muthén, 1984; Faes et al., 2004).

If not only the parameters of the mean but also of the covariance structure are of interest, then approaches are attractive that draw on the assumption of a latent linear model where each observable outcome is related to a continuous latent outcome. Each latent outcome is in turn a linear function of covariates and, given the covariates, is generally assumed to be normally distributed. To estimate the parameters in a model known as the LISCOMP (Linear Structural Equations with a Comprehensive Measurement) model, e.g., Muthén (1984), Muthén and Satorra (1995) or Arminger and Küsters (1988), proposed a three-stage estimation approach. The first step estimates the parameters of the mean structure and, if they are identifiable, variances under the independence assumption. The second step estimates the correlations of the errors of the latent model, based on estimators from the first step and under independence of pairs of outcomes. The third step estimates the parameters of interest, i.e., functions of the parameters from the first two steps, based on a weighted least squares approach. However, this approach turned out to perform poorly with respect to bias, efficiency and convergence (Reboussin and Liang, 1998; Spiess and Hamerle, 2000). Hence, Reboussin and Liang (1998) proposed that the latent model parameters be estimated simultaneously using a quadratic estimating equations approach based on the correct specification of the means of the outcomes and the covariances of pairs of outcomes (cf. Zhao and Prentice, 1990).

In the LISCOMP model, the parameters of interest are usually functions of both the parameters of the mean and of the covariance structure. Hence, in these models it makes sense to use all information available in the mean and the covariance structure and explicitly consider all dependencies to estimate the parameters of interest. The present paper focuses not on functions of both sets of parameters but separately on mean and covariance structure parameters. This leads to a more robust approach, by estimating both sets of parameters as if they were orthogonal (Prentice, 1988; Prentice and Zhao, 1991). Thus at the price of lower efficiency, the parameters of the mean can be estimated consistently even if the covariance structure is misspecified, necessitating a correct specification of the mean model only (cf. Zhao and Prentice, 1990).

Following Prentice (1988) and Zhao and Prentice (1990), Qu et al. (1992, 1994) adopted this approach to estimate probit models with correlated binary outcomes based on two sets of generalized estimating equations. Spiess (1998) and Spiess and Keller (1999) proposed a similar approach, where pseudo-score equations based on pairs of outcomes replace the estimating equations for the correlation structure parameters. In a simulation study, Spiess (1998) compared this mixed approach with the one adopted by Qu et al. (1992, 1994). The results suggest that the parameter estimators of the mean structure are equally efficient, but the mixed approach leads to estimators of the correlation structure which are substantially more efficient. Hence, the present paper generalizes this approach to estimate probit models with correlated continuous and ordered categorical outcomes.

The paper is organized as follows. Section 2 describes the panel model with mixed continuous and ordered categorical outcomes, and Section 3 outlines its estimation. One version of estimating equations adopts a working correlation matrix which is common to all units, whereas another version takes advantage of individual covariance matrices that follow from the covariance structure model. Section 4 presents the results of a numerical study comparing these two versions of the estimator with respect to their efficiency. Section 5 describes the mixed imputation/weighting strategy adopted to compensate for missing data and gives the estimation results with respect to the returns to occupational qualification. Section 6 concludes.

2 The Model

Consider measurements on a continuous outcome, y_{it1}, and an ordered categorical outcome, z_{it2}, obtained on each of N units at each of T points in time ($i = 1, \dots, N$; $t = 1, \dots, T$). In addition, there is a vector of fixed covariates $\mathbf{x}_{it1}$ thought to be related to y_{it1} and a vector of fixed covariates $\mathbf{x}_{it2}$ thought to be related to z_{it2}. In the model to be estimated, y_{it1} is log(income), where income is the monthly gross real labor income, deflated by the national consumer price index (base year 1995), and z_{it2} is perceived job security (0: very concerned about job security, 1: somewhat concerned and 2: not concerned at all). The covariates assumed to have an effect on both outcomes are age, number of children under 17 living in the same household, marital status (married: yes, no), nationality (German nationality: yes, no), industrial sector (chemicals industry, building trade, commerce, metalworking industry; reference category: other), occupational qualification (university, apprenticeship, other completed training; reference category: none) and tenure (in years) as well as tenure squared.

The model assumes that each observable outcome is related to a continuous latent variable. In particular, the observable continuous outcome is identical to the latent outcome. The ordered categorical outcome, z_{it2} with $K+1$ possible values $0,\ldots,K$ (in our application $K=2$), is represented by a $(K\times 1)$ vector of binary indicators, $\mathbf{y}_{it2} = (y_{it21},\ldots,y_{it2K})^T$. The binary indicators relate to the continuous latent variable, y^*_{it2}, via the threshold relation

$$y_{it2k} = \begin{cases} 1 & \text{if} \quad \kappa_{tk} < y^*_{it2} \leq \kappa_{t(k+1)} \\ 0 & \text{else} \end{cases} \qquad \text{for} \quad k = 1,\ldots,K,$$

where κ_{tk} and $\kappa_{t(k+1)}$ are unknown thresholds and $\kappa_{t(K+1)} = \infty$.

The latent model is

$$y^*_{itj} = \eta_{itj} + \epsilon_{itj} \qquad \text{and} \qquad \eta_{itj} = \mathbf{x}^T_{itj}\boldsymbol{\beta}_{tj},$$

where $j=1$ denotes the equation with the continuous outcome, $j=2$ denotes the equation with the ordered categorical outcome, $\boldsymbol{\beta}_{tj}$ is an unknown time and equation-specific vector of parameters of the mean structure. The random error ϵ_{itj} is independent of η_{itj} for all i,t,j.

Let $\boldsymbol{\epsilon}_i$ be the $(2T\times 1)$ vector with elements $\epsilon_{i11},\ldots,\epsilon_{iT2}$. Since the estimation approach discussed in the next section does not involve higher-order moments specifications, only conditional first and second moments need to be correctly specified. That is, the underlying assumptions are that all possible pairs of ϵ_{it2}'s are bivariate normally distributed, each y^*_{it2} conditional on $y_{i11},\ldots,y_{iT1}$ is univariate normally distributed and each ϵ_{it2} depends on all ϵ_{it1}, $t=1,\ldots,T$, only through a linear function. Note that for a valid inference with respect to the parameters of the mean structure, only the assumption of univariate normality of the ϵ_{it2}'s is necessary. The covariance matrix of $\boldsymbol{\epsilon}_i$ will be denoted as $\boldsymbol{\Sigma}$. The units i are assumed to be independent throughout.

In the general model not all parameters are identifiable. Hence, the errors in the regression equations with the observable continuous outcome, i.e., in the linear part of the model, are restricted to have mean zero. In the nonlinear part, i.e., in the regression equations corresponding to the ordered categorical outcomes, constants and means are set equal to zero. Furthermore, in the simulation section, unit variance of the errors and in the application section, unit variance of a component of the errors is assumed in the equations with ordered categorical variables.

Depending on the covariance structure, $\boldsymbol{\Sigma}$ is a function of at most $2T^2$ parameters. For example, since the population considered can — for fixed covariates — be assumed to be rather stable with respect to income but there is no indication of the same stability with respect to perceived job security, the model estimated assumes unobserved subject-specific time-invariant random variables with equation-specific effects and a stationary AR(1) process over time in the equation with the ordered categorical outcome. More general covariance structures than the one described above were considered as well but were not identified. Together with the assumption of constant variances over time, this amounts to the estimation of four covariance structure parameters. The corresponding model in the errors is

$$
\begin{aligned}
\epsilon_{nt1} &= \theta_{11}\pi_n + \theta_{12}w_{nt1}, \\
\epsilon_{nt2} &= \theta_{21}\pi_n + \nu_{nt2}, \quad \nu_{nt2} = \theta_{22}\nu_{n(t-1)2} + w_{nt2},
\end{aligned}
$$

$\pi_n \sim N(0,1)$, $E(w_{nt1}) = 0$, $\mathrm{var}(w_{nt1}) = 1$, $E(\nu_{nt2}) = \mu_{\nu,2}$, $\mathrm{var}(\nu_{nt2}) = \sigma^2_{\nu,2}$, $\mathrm{cov}(\nu_{nt2}, \nu_{nt'2}) = \gamma_{t,t'}$, $|\theta_{22}| < 1$, $\nu_{n02} \sim N(\mu_{\nu,2}, \sigma^2_{\nu,2})$, $w_{nt2} \sim N(0, 1-\theta^2_{22})$ and $E(\pi_n\nu_{n02}) = E(\pi_n w_{ntj}) = E(\nu_{n02}w_{ntj}) = E(w_{ntj}w_{nt'j'}) = 0$ for all j, j', t, t'. From these assumptions, $\mu_{\nu,2} = 0$, $\sigma^2_{\nu,2} = 1$ and $\mathrm{cov}(\nu_{nt2}, \nu_{nt'2}) = \theta_{22}^{|t-t'|}$. The elements of $\boldsymbol{\Sigma}$ are

$$
\begin{aligned}
\mathrm{var}(\epsilon_{nt1}) &= \theta^2_{11} + \theta^2_{12}, & \mathrm{cov}(\epsilon_{nt1}\epsilon_{nt'1}) &= \theta^2_{11} \quad \text{if} \quad t \neq t', \\
\mathrm{cov}(\epsilon_{nt2}\epsilon_{nt'2}) &= \theta^2_{21} + \theta_{22}^{|t-t'|} \qquad \text{and} & \mathrm{cov}(\epsilon_{nt1}\epsilon_{nt'2}) &= \theta_{11}\theta_{21}.
\end{aligned}
$$

3 Estimation of the Model

Generalizing the work of Spiess (1998) and Spiess and Keller (1999), who consider panel models with binary outcomes, this section describes an approach for estimating the parameters of the mean and covariance structure simultaneously as if they were orthogonal. Hence, the mean parameter estimator which is the root of generalized estimating equations (Liang and Zeger, 1986) is robust against misspecification of the covariance structure.

Parameters of the covariance structure are estimated using generalized estimating equations which are equal to pseudo-score equations derived from pseudo-log-likelihood functions based on subsets of outcomes. This is in contrast to Prentice (1988), Qu et al. (1992, 1994) and Reboussin and Liang (1998), who consider estimating equations equating empirical and theoretical centered second moments in models with binary outcomes (Prentice, 1988, Qu et al., 1992, 1994) and in a model with mixed continuous and binary outcomes, respectively (Reboussin and Liang, 1998). To avoid a heavy computational burden, they adopt the identity matrix as a working correlation matrix in the estimating equations for the covariance parameters. In a simulation study, Spiess (1998) compares the approach of Qu et al. (1992, 1994) with an approach based on pseudo-score equations in longitudinal models with binary outcomes. The results suggest that both approaches lead to equally efficient estimators in finite samples for the mean structure but the estimators of the covariance structure turn out to be substantially more efficient based on the latter approach. Furthermore, in a simple model with an exchangeable correlation structure, the loss of efficiency relative to the marginal maximum likelihood estimator is negligible.

3.1 Estimating Equations for the Mean Parameters

Let $\mathbf{y}_i$ be the vector of all continuous outcomes and binary indicators representing the ordered categorical outcomes of unit i, and $\boldsymbol{\mu}_i$ be the vector of the theoretical first conditional moments, $E(\mathbf{y}_{itj}|\mathbf{x}_{itj})$. For continuous outcomes, $E(y_{it1}|\mathbf{x}_{it1})$ is equal to η_{it1}. For binary indicators, $E(y_{it2k}|\mathbf{x}_{it2}) = \Phi(\eta_{it2} - \kappa_{tk}) - \Phi(\eta_{it2} - \kappa_{t(k+1)})$, where $\Phi(\cdot)$ is the cumulative standard normal distribution. Further, $\mathbf{y}_i$ and $\boldsymbol{\mu}_i$ are partitioned into vectors $\mathbf{y}_{i1}$, $\mathbf{y}_{i2}$ and $\boldsymbol{\mu}_{i1}$, $\boldsymbol{\mu}_{i2}$, respectively, where $\mathbf{y}_{i1} = (y_{i11}, \ldots, y_{iT1})^T$, $\boldsymbol{\mu}_{i1} = \{E(y_{i11}|\mathbf{x}_{i11}), \ldots, E(y_{iT1}|\mathbf{x}_{iT1})\}^T$, $\mathbf{y}_{i2} = (\mathbf{y}_{i12}^T, \ldots, \mathbf{y}_{iT2}^T)^T$,

$\boldsymbol{\mu}_{i2} = (\boldsymbol{\mu}_{i12}^T, \ldots, \boldsymbol{\mu}_{iT2}^T)^T$ and $\boldsymbol{\mu}_{it2} = E(\mathbf{y}_{it2}|\mathbf{x}_{it2})$.

Collecting all $\boldsymbol{\beta}_{tj}$ and thresholds in $\boldsymbol{\beta}$, this parameter is estimated by solving the estimating equations

$$\mathbf{0} \;=\; \sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right)^T \boldsymbol{\Omega}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i), \tag{1}$$

(Liang and Zeger, 1986), where $\boldsymbol{\Omega}_i$ is a unit-specific covariance matrix. Usually $\boldsymbol{\Omega}_i$ is of the form $\boldsymbol{\Omega}_i = \mathbf{V}_i^{1/2}\mathbf{R}(\boldsymbol{\alpha})\mathbf{V}_i^{1/2}$, where $\mathbf{V}_i$ is a block-diagonal matrix with diagonal entries equal to $\mathrm{var}(y_{it1}|\mathbf{x}_{it1})$ and $\mathrm{Cov}(\mathbf{y}_{it2}|\mathbf{x}_{it2})$, respectively, and $\mathbf{R}(\boldsymbol{\alpha})$ is a suitable 'working' correlation matrix common to all units. However, here $\boldsymbol{\Omega}_i$ follows directly from the assumed latent covariance structure and is a function of the covariance parameters.

3.2 Estimating Equations for the Covariance Parameters

Let $\boldsymbol{\Sigma}_{11}$ denote the part of $\boldsymbol{\Sigma}$ corresponding to variances and covariances of the linear part of the model, $\mathbf{R}_{11}$ the corresponding correlation matrix and $\mathbf{V}_{11}$ a diagonal matrix with diagonal entries equal to the diagonal elements of $\boldsymbol{\Sigma}_{11}$. Let $\boldsymbol{\delta}_{11}$ be the vector of all elements on and below the diagonal of $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\delta}_{12}$ the vector of all T^2 polyserial correlations between errors of the T nonlinear and T linear equations and $\boldsymbol{\delta}_{22}$ the vector of all $T(T-1)/2$ polychoric correlations corresponding to the nonlinear part of the model, and $\boldsymbol{\delta} = (\boldsymbol{\delta}_{11}^T, \boldsymbol{\delta}_{12}^T, \boldsymbol{\delta}_{22}^T)^T$.

Estimation of all $2T^2$ covariance parameters is usually prohibitive as it may lead to unstable results or convergence problems in small data sets. Furthermore, the research question may imply a covariance structure parameter of lower dimensionality. Hence, usually one would be interested in a function of all $2T^2$ parameters of lower dimensionality. Let $\boldsymbol{\delta} = \boldsymbol{\delta}(\boldsymbol{\theta})$ be a differentiable vector-valued function, where $\boldsymbol{\theta}$ is the vector of covariance structure parameters of interest. The vector $\boldsymbol{\theta}$ fully determines $\boldsymbol{\Sigma}$.

The ith individual contribution to the estimating equations for $\boldsymbol{\delta}_{11}$ is

$$\mathbf{u}_{i,11} = \left(\frac{\partial \boldsymbol{\Sigma}_{11}}{\partial \boldsymbol{\delta}_{11}}\right)^T (\boldsymbol{\Sigma}_{11}^{-1} \otimes \boldsymbol{\Sigma}_{11}^{-1})\, \mathrm{vec}(\mathbf{S}_i - \boldsymbol{\Sigma}_{11}), \tag{2}$$

where $\mathbf{S}_i = (\mathbf{y}_{i1} - \boldsymbol{\mu}_{i1})(\mathbf{y}_{i1} - \boldsymbol{\mu}_{i1})'$, $\otimes$ is the Kronecker product and $\mathrm{vec}(\cdot)$ is the vec operator. Note that if all outcomes were continuous, then (2) would be equal to the individual score equations derived from the log-likelihood under multivariate normality.

While the estimating equations (2) require no distributional assumption, this is different for the estimation of polyserial and polychoric correlations. Let $\boldsymbol{\delta}_{t,12}$ denote the vector of correlations of the errors of the linear part of the model with the error of the tth latent equation corresponding to an ordered categorical outcome. Let Diag($\mathbf{a}$) denote a diagonal matrix with diagonal elements equal to $\mathbf{a}$. Then the ith individual contribution to the estimating equations for $\boldsymbol{\delta}_{t,12}$ is

$$\mathbf{u}_{it,12} = \Big(\frac{\partial \boldsymbol{\mu}_{it,\mathbf{b}|\mathbf{c}}}{\partial \boldsymbol{\delta}_{t,12}}\Big)^T \mathbf{W}_{it,12}^{-1}(\mathbf{y}_{it2} - \boldsymbol{\mu}_{it,\mathbf{b}|\mathbf{c}}), \tag{3}$$

where $\boldsymbol{\mu}_{it,\mathbf{b}|\mathbf{c}}$ is the conditional mean of $\mathbf{y}_{it2}$ given $\mathbf{y}_{i1}$ for fixed $\mathbf{x}_{it2}, \mathbf{x}_{i11}, \ldots, \mathbf{x}_{iT1}$, and $\mathbf{W}_{it,12} = (\text{Diag}(\boldsymbol{\mu}_{it,\mathbf{b}|\mathbf{c}}) - \boldsymbol{\mu}_{it,\mathbf{b}|\mathbf{c}}\boldsymbol{\mu}_{it,\mathbf{b}|\mathbf{c}}^T)$. The estimating equations (3) are equal to the pseudo-score equations derived from the pseudo-log-likelihood function of $\boldsymbol{\delta}_{t,12}$. Details are given in Appendix A.1.

The estimating equations (3) are generalizations of those given in Catalano and Ryan (1992), who consider a model with mixed continuous and binary outcomes and an exchangeable correlation matrix. In contrast to Catalano and Ryan (1992), the above estimating equations are not solved to estimate both mean and covariance structure parameters. Hence estimation of the mean structure parameters via (1) remains robust with respect to a misspecification of the covariance structure.

The estimating equations for the polychoric correlations consider each possible pair of ordered categorical outcomes as one polytomous variable and equate this variable with its theoretical expectation. Let $\mathbf{v}_{itt'2}$ be a vector with binary indicators coding the observed non-redundant combination of values of the two ordered categorical outcomes for the ith unit at time points t and t'. Let $\boldsymbol{\mu}_{itt'2}$ denote its mean for fixed $\mathbf{x}_{it2}$ and $\mathbf{x}_{it'2}$, which is easily evaluated using the bivariate cumulative standard normal distribution function. Then the individual contribution to the estimating equations for the tt'th element ($t = 2, \ldots, T$, $t' = 1, \ldots, t$) of $\boldsymbol{\delta}_{22}$, $\delta_{tt',22}$, is

$$u_{itt',22} = \Big(\frac{\partial \boldsymbol{\mu}_{itt'2}}{\partial \delta_{tt',22}}\Big)^T \mathbf{W}_{itt',22}^{-1}(\mathbf{v}_{itt'} - \boldsymbol{\mu}_{itt'2}). \tag{4}$$

where $\mathbf{W}_{itt',22} = (\text{Diag}(\boldsymbol{\mu}_{itt'2}) - \boldsymbol{\mu}_{itt'2}\boldsymbol{\mu}_{itt'2}^T)$. Again, the above estimating equations are equal to pseudo-score equations derived from the pseudo-log-likelihood for $\delta_{tt',22}$ based on observations $\mathbf{y}_{it2}$ and $\mathbf{y}_{it'2}$ under the assumption of bivariate normality of the errors (Appendix A.1). All $u_{i,22,tt'}$ are collected in the $(T(T-1)/2 \times 1)$ vector of individual estimating equations $\mathbf{u}_{i,22}$.

3.3 The Stacked Estimating Equations

Let $\boldsymbol{\vartheta} = (\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T$ be the parameter to be estimated. For ease of presentation, let $\mathbf{B}_{i,1}$ be equal to $\partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}$ and $\mathbf{B}_{i,2}$ a partitioned matrix with block diagonal elements equal to $\partial \boldsymbol{\Sigma}_{11} / \partial \boldsymbol{\delta}_{11}$, $\partial \boldsymbol{\mu}_{it,\mathbf{b}|\mathbf{c}} / \partial \boldsymbol{\delta}_{t,12}$ $(t = 1, \ldots, T)$, $\partial \boldsymbol{\mu}_{itt'2} / \partial \delta_{22,tt'}$ $(t = 2, \ldots, T,\ t' = 1, \ldots, t)$ and off-diagonal blocks equal to null matrices and $\partial \boldsymbol{\mu}_{it,\mathbf{b}|\mathbf{c}} / \partial \boldsymbol{\delta}_{11}$ $(t = 1, \ldots, T)$. Let $\mathbf{B}_i$ be a block diagonal matrix with diagonal blocks equal to $\mathbf{B}_{i,1}$ and $\partial \boldsymbol{\delta} / \partial \boldsymbol{\theta}\ \mathbf{B}_{i,2}$. Let $\boldsymbol{\Gamma}_i$ be a block diagonal matrix with diagonal blocks equal to $\boldsymbol{\Omega}_i$, $\boldsymbol{\Sigma}_{11} \otimes \boldsymbol{\Sigma}_{11}$, $\mathbf{W}_{it,12}$ $(t = 1, \ldots, T)$ and $\mathbf{W}_{itt',22}$ $(t = 2, \ldots, T,$ $t' = 1, \ldots, t)$, where $\boldsymbol{\Omega}_i$ is given in Appendix A.2. Further, let $\mathbf{e}_i$ be a vector with subvectors equal to $(\mathbf{y}_i - \boldsymbol{\mu}_i)$, $(\mathrm{vec}(\mathbf{S}_i) - \mathrm{vec}(\boldsymbol{\Sigma}_i))$, $(\mathbf{y}_{it2} - \boldsymbol{\mu}_{it,\mathbf{b}|\mathbf{c}})$ $(t = 1, \ldots, T)$ and $(\mathbf{v}_{itt'} - \boldsymbol{\mu}_{itt'2})$ $(t = 2, \ldots, T,\ t' = 1, \ldots, t)$.

The estimating equations for mean and covariance structure parameters are stacked to yield the unbiased estimating equations

$$\mathbf{0} = \sum_{i=1}^{N} \mathbf{B}_i^T \boldsymbol{\Gamma}_i^{-1} \mathbf{e}_i$$

for the simultaneous estimation of mean and covariance structure parameters.

The vector of estimates, $\hat{\boldsymbol{\vartheta}}$, is iteratively calculated with updated value in the $(q+1)$th iteration given by

$$\hat{\boldsymbol{\vartheta}}_{q+1} = \hat{\boldsymbol{\vartheta}}_q + \left(\sum_{i=1}^{N} \hat{\mathbf{B}}_i^T \hat{\boldsymbol{\Gamma}}_i^{-1} \hat{\mathbf{B}}_i \right)^{-1} \sum_{i=1}^{N} \hat{\mathbf{B}}_i^T \hat{\boldsymbol{\Gamma}}_i^{-1} \hat{\mathbf{e}}_i.$$

Adapting results given in Prentice and Zhao (1991), $\sqrt{N}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0)$ is asymptotically normally distributed with mean zero and asymptotic covariance matrix, $\mathbf{V}_{\hat{\boldsymbol{\vartheta}}}$, consistently estimated by

$$\hat{\mathbf{V}}_{\hat{\boldsymbol{\vartheta}}} = \left(\sum_{i=1}^{N} \hat{\mathbf{B}}_i^T \hat{\boldsymbol{\Gamma}}_i^{-1} \hat{\mathbf{B}}_i \right)^{-1} \left\{ \sum_{i=1}^{N} \hat{\mathbf{B}}_i^T \hat{\boldsymbol{\Gamma}}_i^{-1} \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^T \hat{\boldsymbol{\Gamma}}_i^{-1} \hat{\mathbf{B}}_i \right\} \left(\sum_{i=1}^{N} \hat{\mathbf{B}}_i^T \hat{\boldsymbol{\Gamma}}_i^{-1} \hat{\mathbf{B}}_i \right)^{-1},$$

where all unknowns are replaced by their sample counterparts and estimates, respectively.

The following points should be noted. First, solving the estimating equations requires evaluation of one- and two-dimensional integrals only. Second, the estimate $\hat{\mathbf{\Omega}}_i$ needed to calculate $\hat{\boldsymbol{\vartheta}}$ and $\hat{\mathbf{V}}_{\hat{\boldsymbol{\vartheta}}}$ is not guaranteed to be positive definite in general. Hence, one strategy is to consider the simple function $\bar{\mathbf{\Omega}} = N^{-1} \sum_{i=1}^{N} \hat{\mathbf{\Omega}}_i$. Let $\bar{\mathbf{V}} = \mathrm{Diag}(\bar{\mathbf{\Omega}})$. Then a working correlation matrix is $\hat{\mathbf{R}} = \bar{\mathbf{V}}^{-1/2}\bar{\mathbf{\Omega}}\bar{\mathbf{V}}^{-1/2}$ and individual covariance matrices can be calculated by $\tilde{\mathbf{\Omega}}_i = \hat{\mathbf{V}}_i^{1/2}\hat{\mathbf{R}}\hat{\mathbf{V}}_i^{1/2}$, where $\hat{\mathbf{V}}_i$ is a diagonal matrix with diagonal elements equal to the estimated variances of the corresponding outcomes. This closely resembles the strategy of Liang and Zeger (1986), who adopt a working correlation matrix identical for all N. However, note that the consistent estimation of $\hat{\boldsymbol{\vartheta}}$ and $\hat{\mathbf{V}}_{\hat{\boldsymbol{\vartheta}}}$ does not depend on the correlation matrix implied by $\hat{\mathbf{\Omega}}_i$. A second strategy is to first try to invert $\hat{\mathbf{\Omega}}_i$ for each unit and replace this individual matrix by $\tilde{\mathbf{\Omega}}_i$ only if the former is not positive definite.

4 A Numerical Study

This section compares the estimator based on a correlation matrix common to all units, denoted as GEE$_*$ with the estimator based on covariance matrices not depending on a common correlation matrix, denoted as GEE. The data sets were simulated according to the model described in Section 2 with $T = 5$ with four covariates generated independently of each other. The covariates followed a uniform, a standard normal, a Bernoulli, and a mixture of a gamma and a uniform distribution. The former two were generated independently over time and equations, the third was held fixed over equations, and the fourth was correlated over time with a correlation of 0.5. The parameters weighting the covariates in the linear equations were $\beta_{c,1} = -1$, $\beta_{c,2} = .8$, $\beta_{c,3} = 0$ and $\beta_{c,4} = -.1$, those weighting the covariates in the equations with ordered categorical outcomes were $\beta_{o,1} = -1$, $\beta_{o,2} = .8$, $\beta_{o,3} = -.8$ and $\beta_{o,4} = 0$. The constant in the linear equations was $-.3$ and the thresholds in the equations with ordered categorical outcomes were set equal to $\kappa_1 = -.4$ and $\kappa_2 = .7$, respectively.

The error term followed a multivariate normal distribution with unit variances and $\mathbf{\Sigma} = (\mathbf{Q}_{\boldsymbol{\theta}} \otimes \mathbf{1}_2\mathbf{1}_2^T) + (1 - \theta_1)\mathbf{I}_{10}$, where $\mathbf{Q}_{\boldsymbol{\theta}}$ is a symmetric Toeplitz matrix with diagonal elements θ_1 and off-diagonal elements $\theta_2, \ldots, \theta_5$, arranged in such a way that the correlations decrease with increasing distance in time. Two versions of true correlation matrices were considered. For the high correlation condition, Model I, $\theta_1 = .8$, $\theta_2 = .68$, $\theta_3 = .584$, $\theta_4 = .507$ and $\theta_5 = .4468$. For the low correlation condition, Model II, $\theta_1 = .4$, $\theta_2 = .25$, $\theta_3 = .175$, $\theta_4 = .138$ and $\theta_5 = .119$, respectively.

According to both models, data sets were generated with $N = 200$, $N = 500$ and $N = 1000$ units. Statistics calculated over 500 simulations under each condition are the mean (m) and standard deviation (sd) of the estimates, the square root of the mean of estimated variances of the estimators, denoted as estimated standard deviation ($\widehat{sd}$), and the portion of rejections of the hypothesis $H_0 : \vartheta_s = \vartheta_{s,0}$ for $\alpha = .05$, where s denotes the sth element of $\boldsymbol{\vartheta}$.

The general pattern of results does not differ with increasing sample sizes, the only obvious differences being decreasing standard deviations and decreasing differences of means of estimates and true values. The portions of rejections of the null are in an acceptable range of approximately $.05 \pm .02$. Further, since there are nearly no differences with respect to the estimators of the covariance structure parameters under both types of covariance matrices, Table 1 gives the results for $N = 500$ and the parameters $\beta_{c,1}$, $\beta_{c,4}$, $\beta_{o,1}$ and $\beta_{o,2}$ only.

Table 1: *Mean (m), estimated standard deviation ($\widehat{sd}$) and standard deviation (sd) over* 500 *simulations.*

	Model I				Model II			
	GEE$_*$	GEE	GEE$_*$	GEE	GEE$_*$	GEE	GEE$_*$	GEE
	$\beta_{c,1} = -1$		$\beta_{o,1} = -1$		$\beta_{c,1} = -1$		$\beta_{o,1} = -1$	
m	$-.9997$	$-.9997$	-1.005	-1.005	$-.9979$	$-.9981$	-1.005	-1.005
$\widehat{sd}$	0.0289	0.0284	0.0541	0.0488	0.0412	0.0410	0.0597	0.0564
sd	0.0294	0.0286	0.0555	0.0491	0.0413	0.0408	0.0617	0.0574
	$\beta_{c,4} = -.1$		$\beta_{o,2} = .8$		$\beta_{c,4} = -.1$		$\beta_{o,2} = .8$	
m	$-.0988$	$-.0987$	0.8037	0.8031	$-.0988$	$-.0988$	0.8083	0.8029
$\widehat{sd}$	0.0151	0.0141	0.0296	0.0266	0.0200	0.0196	0.0300	0.0278
sd	0.0154	0.0140	0.0293	0.0258	0.0206	0.0201	0.0312	0.0277

The results in Table 1 suggest that there is a gain in efficiency if individual correlation matrices are used as compared to using a common correlation matrix. This efficiency gain seems to be largest for the parameters in the equations with ordered categorical outcomes under the high correlation condition. They are negligible for the mean structure parameters of the linear part under Model I and Model II. In fact, the relative variances of the GEE and the GEE$_*$ estimator range from approximately 0.98 for $\beta_{c,1}$ under Model II to 0.78 for $\beta_{o,1}$ and $\beta_{o,2}$ under Model I. Put differently, a gain in efficiency up to 22% is possible by using the GEE instead of the GEE$_*$ estimator, which is a substantial improvement given that using individual correlation matrices does not necessitate additional assumptions. Hence, the GEE approach is used to estimate the model described in Section 2.

5 Application

5.1 The Data

The analysis involves panel data on full-time employed males from nine panel waves beginning in 1991 of the West German subsamples of the German Socio-Economic Panel (SOEP). As is typical for survey data, the SOEP suffers from missing information. Not all households sampled in 1984 were actually observed in the first wave (62%) and not all individuals interviewed in 1984 were also observed in 1991. That is, in 1984 the number of individuals aged 16 years and older observed is 12,245 living in 5,921 households. The same subsamples cover 9,467 individuals living in 4,669 households in 1991, the eighth wave of SOEP. Additionally, there is item nonresponse, up to approximately 20% (e.g., monthly gross income 1985), varying depending on the question being asked.

5.2 Handling of Missing Data

To compensate for attrition from 1991 up to 1999 and for missing items, the application draws on multiple imputations. Basically, multiple imputations should be draws from the joint posterior (predictive) distribution of the variables whose values are unobserved given the observed values of all other variables and should reflect the entire uncertainty inherent in these predictions (Rubin, 1987, 1996). However, a general problem with complicated patterns of missing values is that it is hard to specify this joint predictive distribution. Therefore, simpler and less formally rigorous methods that approximate draws from this distribution have been proposed.

One such method is implemented in the program IVEware (Raghunathan et al., 2001, 2002), where imputations are generated based on the repeated estimation of regression models for the variables to be imputed conditional on all other variables with observed or already imputed values, and assuming non-informative prior distributions for the parameters of these models. For a continuous response variable this amounts to estimating a simple linear regression model, for a binary response variable a logit model, for categorical response variables a polytomous logit model, and for count variables a Poisson loglinear model. For a detailed description see Raghunathan et al. (2001, 2002). The whole procedure was repeated ten times to create $M = 10$ completed data sets.

The imputations were generated based on males selected into the imputation data set if they were observed in 1991 and unless their year of death was known to lie between 1991 and 1999 or if they were in the army or doing civilian service. Finally, $N = 4043$ males entered the imputation data set. Among the variables considered to be important with respect to the imputation models, aside from those included in the final model of interest, are variables indicating a separation from or the death of a partner within the last year, schooling, working experience, size of the firm or institute, number of overtime hours worked in the month before the interview, employment status and a dummy variable coding whether the individual is employed in the public sector. Additionally, tenure squared and experience squared but also the estimated probability of observing the corresponding household in 1991 entered the imputation models.

The assumption underlying the generation of imputations was that the missing-data mechanism is ignorable, which is largely equivalent to assuming that the missing values are missing at random (MAR; Little and Rubin, 2002). Unfortunately it is not possible to test the MAR assumption against the assumptions that the missing data are not missing at random (NMAR; Little and Rubin, 2002). Furthermore, with survey data and a complex missing pattern it is hard to justify any hypothesis about the exact missing mechanism, which, if misspecified, would usually lead to improper imputations. On the other hand, by including as many variables as possible that are thought to be relevant, the MAR assumption becomes more likely to hold (e.g., Rubin, 1996). Furthermore, if the missing values are NMAR, then proper imputation methods based on the MAR assumption are still preferable to procedures that rely on the missing data being missing completely at random, such as simply ignoring the missing data (Schafer, 1997, Little and Rubin, 2002).

Although the validity of Rubin's (1987) variance estimator based on multiple imputations has been questioned in the context of frequentist inference if the estimator is not fully efficient (e.g., Nielsen, 2003), simulation results in Paik (1997), Xie and Paik (1997) and Spiess and Keller (1999) do not suggest invalid variance estimation for GEE estimators. In contrast, the variance estimation seems to be quite robust even with respect to moderate misspecifications of the imputation model.

To compensate for first wave nonresponse and attrition up to 1991 in the data analyzed, again under the MAR assumption, each individual contribution to the estimating equations is divided by the estimated probability of observing that unit in 1991 (e.g., Robins et al., 1995, or, Wooldridge, 2002). These weights can be derived from information delivered with the SOEP. There is, however, no information available that allows one to take into account the uncertainty in the estimated probabilities, leading to conservative inferences (Robins et al., 1995).

To estimate the model, individuals were selected from the multiply imputed data set into the final samples if they lived in former West Germany and were full-time employed in the private sector in each year from 1991 to 1999. Observations with high leverage were excluded from the analysis. Thus, only those individuals with weights lower than or equal to the 99%-quantile were included in the final analysis. Since selection is based on variables with missings replaced by multiple imputations, the size of the ten final samples varies from $N = 702$ to $N = 781$. The weighted standard analysis was then carried out $M = 10$ times and the estimation results were combined according to the rules given, e.g., in Little and Rubin (2002).

5.3 Results

Table 2 displays the test statistic $\tilde{D}$ and the corresponding p-values to test the hypothesis that the effects of the covariates are identical over time based on the weighted analyses of the $M = 10$ imputed data sets (Little and Rubin, 2002, see also Li et al., 1991). In the case of the variables economic sector and occupational qualification, the corresponding test is a test that the differences of the effects of the corresponding dummy variables over time are, separately for both variables, simultaneously zero. The test for equal effects of tenure amounts to testing that differences of linear and quadratic effects over time are simultaneously zero.

Table 2: *Test statistic* ($\tilde{D}$) *and p-values to test* H_0 : *'Effects are identical over time'.*

Response: log(Income)			Response: Perceived job security		
Variable	$\tilde{D}_5$	p	Variable	$\tilde{D}_5$	p
Constant	1.17	0.31	Threshold 1	0.70	0.69
			Threshold 2	0.54	0.83
Age	0.70	0.69	Age	1.14	0.33
Children	0.31	0.96	Children	0.66	0.72
Married	0.37	0.94	Married	1.38	0.20
Nationality	0.16	0.99	Nationality	1.61	0.12
Econ. Sector	0.21	1.00	Econ. Sector	0.73	0.82
Occup. Qual.	0.56	0.92	Occup. Qual.	1.04	0.41
Tenure	0.54	0.93	Tenure	0.48	0.96

Obviously, the hypothesis of time-invariant effects cannot be rejected for the covariates ($\alpha = .05$). In particular, the results in Table 2 do not support the hypothesis of a changing effect of the social investment variable occupational qualification on the objective and subjective gratification variables income and perceived job security. This is also supported if one considers the unrestricted parameter estimates weighting the three dummies university, apprenticeship and other graduation separately (not shown in form of a table): for all three dummies, the estimates do not change in a systematic way over time.

Table 3 provides the estimation results with the parameters of the systematic part of the model restricted to be equal over time.

According to the upper first part of Table 3, the hypothesis of no effect on gross income can be rejected at the .05-level for the covariates age, children, university, apprenticeship, tenure and tenure squared, the former covariates having a positive effect, whereas tenure squared seems to have a negative effect.

With respect to perceived job security, the results in the upper part of Table 3 suggest that only the covariates university, tenure and tenure squared seem to have an effect at the .05 level. Interestingly, the nonlinear effect of job tenure

Table 3: *Estimates ($\hat{\theta}$), standard errors ($\widehat{sd}$) and p-values; mean structure parameters and thresholds restricted to the same value over time.*

	Mean Structure					
	log(Income)			Perceived job security		
Variable	$\hat{\theta}$	$\widehat{sd}$	p	$\hat{\theta}$	$\widehat{sd}$	p
Constant	7.84	0.08	0.00			
Threshold 1				−1.58	0.23	0.00
Threshold 2				−0.19	0.23	0.41
Age[a]	1.15	0.16	0.00	−0.58	0.47	0.21
Tenure[a]	0.64	0.30	0.03	−3.27	1.06	0.00
Tenure Squared[a]	−0.02	0.01	0.03	0.08	0.03	0.01
Children[b]	0.23	0.11	0.04	−0.47	0.29	0.11
Married[b]	−0.03	0.29	0.91	−0.14	0.77	0.85
Nationality[b]	0.77	0.46	0.11	0.97	1.10	0.38
Building Trade[b]	−0.35	0.55	0.53	1.66	1.20	0.17
Chemicals Ind.[b]	−0.30	0.46	0.51	−0.91	1.17	0.43
Commerce[b]	−0.55	0.52	0.29	1.16	1.21	0.34
Metalworking Ind.[b]	−0.41	0.46	0.38	−1.47	1.15	0.20
University	0.56	0.08	0.00	0.59	0.17	0.00
Apprenticeship	0.15	0.06	0.02	0.15	0.12	0.22
Other Graduation	−0.00	0.08	0.96	0.19	0.13	0.15
Covariance Structure						
θ_{11}	0.24	0.005	0.00			
θ_{12}^2	0.03	0.004	0.00			
θ_{21}	0.10	0.017	0.18			
θ_{22}	0.72	0.013	0.00			

[a] Estimate and standard deviation multiplied by 10^2,

[b] Estimate and standard deviation multiplied by 10^1.

on perceived job security is contrary to its effect on log(income): the longer the employee belongs to a company, the higher the probability of reporting concerns about job security, although with a diminishing effect over time. Not surprisingly, having a university degree seems to have a positive effect on perceived job security.

Estimation results with respect to the covariance structure are presented in the lower part of Table 3. In the light of the assumption of a possible dependence of the responses given the covariates, a surprising result is that the estimated correlation of the corresponding error terms is close to zero, i.e., θ_{21} is not significantly different from zero. This does not support the hypothesis of unobserved individual effects that are important with respect to both outcomes simultaneously. On the other hand, the estimates of the correlation structure parameters θ_{11} and θ_{22} for each equation over time are significantly different from zero. This implies that there is substantive dependence over time within this nine-year interval.

6 Discussion

The results of the last section do not suggest noticeable changes in the returns to social investments. As might be expected, holding a university degree or having finished training seems to have a rather stable positive effect on income. Holding a university degree also seems to have a positive effect on perceived job security. Interestingly, the results also suggest that the effect of job tenure on log(income) is contrary to its effect on perceived job security. In the former case, job tenure has a positive although decreasing effect, whereas its effect on perceived job security is, although diminishing, negative. However, given the covariates, there seems to be no additional dependence between the objective and the subjective gratification variable, suggesting that, given that the underlying model assumptions are correct, both gratification variables are not interchangeable and should be treated as conditionally independent returns to the social investment variable occupational qualification. This result reveals, ex post, that fitting two separate regression models for the two outcomes would lead to the same results.

The general approach of estimating the covariance parameters in the above model based on pseudo-score equations was justified with simulation results presented in Spiess (1998), which suggest that this approach leads to more efficient covariance parameter estimators than estimators based on equating empirical and theoretical centered second moments under a working independence assumption. As suggested by the simulation results of the present paper, additional efficiency can be gained for the estimators of mean parameters by using unit-specific covariance matrices that directly follow from the assumed covariance structure and do not depend on a correlation matrix common to all units. Having the most efficient estimators possible with a given set of assumptions is particularly important if estimation is based on survey data with missing values. For example, if the weights used to compensate for missing units are based on estimated response probabilities, as delivered with some public-use data sets, then one usually has no information available to properly account for the uncertainty in these estimates. As a consequence, the resulting inference tends to be conservative (Robins et al., 1995). Similarly, mild misspecifications of models to generate imputations typically lead to an overestimation of standard errors and to conservative inferences (e.g., Little and Rubin, 2002; Rubin, 2003).

It should be noted that although the estimation approach proposed in this paper is applied to a model with one continuous and one ordered categorical outcome, it can easily be generalized to models with more than two outcomes and unequal numbers of possible values of ordered categorical outcomes. Further, to avoid problems with estimating too many unrestricted mean parameters, the estimation approach can easily be supplemented to estimate lower dimensional functions of the mean parameters. However, it should also be noted that the approach proposed in this paper is appropriate only for a balanced panel design.

A Appendix

A.1 Pseudo-log-likelihood functions for $\boldsymbol{\delta}_{t,12}$ and $\delta_{tt',22}$

To derive the estimating equations for $\boldsymbol{\delta}_{t,12}$ note that the pseudo-log-likelihood of $\boldsymbol{\delta}_{t,12}$ can be written as

$$l^*(\boldsymbol{\delta}_{t,12}) = \text{const} + \sum_{i=1}^{N} \{\mathbf{y}_{it2}^T \log \boldsymbol{\mu}_{it,\mathbf{b}|\mathbf{c}} + (1 - \mathbf{1}_K^T \mathbf{y}_{it2}) \log(1 - \mathbf{1}_K^T \boldsymbol{\mu}_{it,\mathbf{b}|\mathbf{c}})\}$$

where $\mathbf{1}_K$ is a $(K \times 1)$ vector of ones, const is a term not involving $\boldsymbol{\delta}_{t,12}$ and $\boldsymbol{\mu}_{it,\mathbf{b}|\mathbf{c}}$ is the vector of conditional means $E(\mathbf{y}_{it2}|\mathbf{y}_{i1}, \mathbf{x}_{it2}, \mathbf{x}_{i11}, \ldots, \mathbf{x}_{iT1})$, with elements $\Pr(y_{it2k} = 1|\mathbf{y}_{it1}, \mathbf{x}_{it2}, \mathbf{x}_{i11}, \ldots, \mathbf{x}_{iT1}) = \Phi(\psi_{it2k}) - \Phi(\psi_{it2(k+1)})$, $k = 1, \ldots, K$, and

$$\psi_{it2r} = \{\eta_{it2} - \kappa_{tr} + \boldsymbol{\delta}_{t,12}^T \mathbf{R}_{11}^{-1} \mathbf{V}_{11}^{-1/2} (\mathbf{y}_{i1} - \boldsymbol{\mu}_{i1})\} / (1 - \boldsymbol{\delta}_{t,12}^T \mathbf{R}_{11}^{-1} \boldsymbol{\delta}_{t,12})^{1/2}$$

if $r \leq K$ and $\psi_{it2r} = -\infty$ if $r = K + 1$. The derivative of $l^*(\boldsymbol{\delta}_{t,12})$ with respect to $\boldsymbol{\delta}_{t,12}$ leads to the estimating equations (3) given in Section 3.

To derive the estimating equations (4) note that each pair of binary indicator vectors representing a pair of ordered categorical outcomes is represented by a polytomous variable. Thus, let $\mathbf{y}_{it2}^{+} = (y_{it20}, \mathbf{y}_{it2}^T)^T$, where $y_{it20} = 1 - \mathbf{1}_K^T \mathbf{y}_{it2}$, let $\mathbf{v}_{itt'2}$ be a $S = \{(K+1)^2 - 1\}$-dimensional column vector of elements $\mathrm{vec}(\mathbf{y}_{it2}^{+} \mathbf{y}_{it'2}^{+,T})$, $t \neq t'$ but not including the element $y_{it20} y_{it'20}$, and $\boldsymbol{\mu}_{itt'2} = E(\mathbf{v}_{itt'2} | \mathbf{x}_{it2}, \mathbf{x}_{it'2})$, with elements $\Pr(y_{it2l} = 1, y_{it'2l'} = 1 | \mathbf{x}_{it2}, \mathbf{x}_{it'2})$, $l = 0, \ldots, K$, not including the elements $\Pr(y_{it20} = 1, y_{it'20} = 1 | \mathbf{x}_{it2}, \mathbf{x}_{it'2})$. Then the derivative of the pseudo-log-likelihood function

$$l^*(\delta_{tt',22}) = \sum_{i=1}^{N} \{\mathbf{v}_{itt'2}^T \log \boldsymbol{\mu}_{itt'2} + (1 - \mathbf{1}_S^T \mathbf{v}_{itt'2}) \log(1 - \mathbf{1}_S^T \boldsymbol{\mu}_{itt'2})\}.$$

with respect to $\delta_{tt',22}$ leads to the estimating equations (4) given in Section 3.

A.2 The Matrix $\boldsymbol{\Omega}_i$

The assumed structure of $\boldsymbol{\Sigma}$ implies $\boldsymbol{\Omega}_i$ as follows. The partition of $\boldsymbol{\Omega}_i$ corresponding to variances and covariances of the linear part of the model is identical to $\boldsymbol{\Sigma}_{11}$. Let $\zeta_{it2k}^* = \eta_{it2} - \kappa_{tk}$ and $\zeta_{it2(k+1)}^* = \eta_{it2} - \kappa_{t(k+1)}$, respectively. The partition of $\boldsymbol{\Omega}_i$ corresponding to covariances of the linear and the non-linear part of the model, $\boldsymbol{\Omega}_{12}$, is a matrix made up of vectors

$$\mathrm{Cov}(\mathbf{y}_{i1}, y_{it2k}) = -\mathbf{V}_{11}^{1/2} \boldsymbol{\delta}_{t,12} \{\varphi(\zeta_{it2k}^*) - \varphi(\zeta_{it2(k+1)}^*)\},$$

where $\varphi(\cdot)$ is the standard normal density function. The elements of the partition of $\boldsymbol{\Omega}_i$ corresponding to the nonlinear part of the model represented by the binary indicators are

$$\mathrm{Cov}(\mathbf{y}_{it2}) = \mathrm{Diag}(\boldsymbol{\mu}_{it2}) - \boldsymbol{\mu}_{it2} \boldsymbol{\mu}_{it2}^T \qquad t = 1, \ldots, T$$

on the diagonal, and, as off-diagonal elements,

$$\mathrm{Cov}(y_{it2k}, y_{it'2k'}) = \mu_{itt'2,kk'} - \mu_{it2k} \mu_{it'2k'},$$

where $\mu_{itt'2,kk'}$ and $\mu_{it2,k}$ are the corresponding elements from $\boldsymbol{\mu}_{itt'2}$ and $\boldsymbol{\mu}_{it2}$, respectively.

References

Arminger, G. and Küsters, U. (1988) Latent Trait Models with Indicators of Mixed Measurement Levels. In *Latent Trait and Latent Class Models* (eds. R. Langeheine and J. Rost), pp. 51–73. New York: Plenum Press.

Catalano, P.J. and Ryan, L.M.(1992) Bivariate Latent Variable Models for Clustered Discrete and Continuous Outcomes. *J. Am. Statist. Ass.,* 87, 651–658.

Diener, E. & Seligman, M.E.P. (2004) Beyond Money. Toward an Economy of Well-Being. *Psychological Science in the Public Interest,* 5, 1–31.

Dunson, D.B. (2000) Bayesian latent variable models for clustered mixed outcomes. *J. R. Statist. Soc. B,* 355–366.

Easterlin, R.A. (2002) *Happiness in Economics.* Cheltenham: Edward Elgar.

Faes, C., Geys, H., Aerts, M., Molenberghs, G. and Catalano, P.J. (2004) Modeling combined continuous and Ordinal Outcomes in a Clustered Setting. *Journal of Agricultural, Biological, and Environmental Statistics,* 9, 515–530.

Fitzmaurice, G.M. and Laird, N.M. (1995) Regression Models for Bivariate Discrete and Continuous Outcome With Clustering. *J. Am. Statist. Ass.,* 90, 845–852.

Geschwender, J.A. (1967a) Continuities in Theories of Status Consistency and Cognitive Dissonance. *Social Forces, 46*, 160–171.

Geschwender, J.A. (1967b) Status Inconsistency, Social Isolation, and Individual Unrest. *Social Forces, 46*, 477–483.

Gueorguieva, R.V. and Agresti, A. (2001) A Correlated Probit Model for Joint Modeling of Clustered Binary and Continuous Responses. *J. Am. Statist. Ass.,* 96, 1102–1112.

Li, K.H., Raghunathan, T.E. and Rubin, D.B. (1991) Large-Sample Significance Levels From Multiply Imputed Data Using Moment-Based Statistics and an F Reference Distribution. *J. Am. Statist. Ass.,* 86, 1065–1073.

Liang, K.-Y. and Zeger, S.L. (1986) Longitudinal Data Analysis using Generalized Linear Models, *Biometrika, 73*, 13–22.

Little, R.J.A. and Rubin, D.B. (2002) *Statistical Analysis with Missing Data*, 2nd edn. New York: John Wiley & Sons.

Muthén, B. (1984) A General Structural Equation Model With Dichotomous, Ordered Categorical, And Continuous Latent Variable Indicators. *Psychometrika 49,* 115-132.

Muthén, B.O. and Satorra, A. (1995) Technical Aspects of Muthéns LISCOMP Approach to Estimation of Latent Variable Relations with a Comprehensive Measurement Model. *Psychometrika, 60,* 489–503.

Nielsen, S.F. (2003) Proper and Improper Multiple Imputation. *International Statistical Review, 71*, 593–627.

Paik, M.C. (1997) The Generalized Estimating Equation Approach When Data Are Not Missing Completely at Random. *J. Am. Statist. Ass.,* 92, 1320–1329.

Prentice, R.L. (1988) Correlated Binary Regression with Covariates Specific to Each Binary Observation. *Biometrics, 44*, 1033–1048.

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

Qu, Y., Piedmonte, M.R. and Williams, G.W. (1994) Small sample validity of latent variable models for correlated binary data. *Communs Statist. — Simulation and Computation, 23*, 243–269.

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

Raghunathan, T.E., Lepkowski, J.M., Van Hoewyk, J. and Solenberger, P. (2001) A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models. *Survey Methodology, 27,* 85–95.

Raghunathan, T.E., Solenberger, P. and Van Hoewyk, J. (2002) *IVEware:* Imputation and Variance Estimation Software. Michigan: Survey Methodology Program, Survey Research Center, Institute for Social Research, University of Michigan. http://www.isr.umich.edu/src/smp/ive/.

Reboussin, B.A. and Liang, K.-Y. (1998) An Estimating Equations Approach for the LISCOMP Model. *Psychometrika*, 63, 165–182.

Regan, M.M. and Catalano, P.J. (1999) Likelihood Models for Clustered Binary and Continuous Outcomes: Application to Developmental Toxicology. *Biometrics,* 55, 760–768.

Robins, J.M., Rotnitzky, A. and Zhao, L.P. (1995) Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. *J. Am. Statist. Ass.,* 90, 106–121.

Rubin, D.B. (1987) *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley & Sons.

Rubin, D.B. (1996) Multiple Imputation After 18+ Years. *J. Am. Statist. Ass.,* 91, 473–489.

Rubin, D.B. (2003) Discussion on Multiple Imputation. *International Statistical Review, 71(3),* 619–625.

Sammel, M.D., Ryan, L.M. and Legler, J.M. (1997) Latent Variable Models for Mixed Discrete and Continuous Outcomes. *J. R. Statist. Soc. B,* 667–678.

Schafer, J.L. (1997) *Analysis of Incomplete Multivariate Data.* London: Chapman & Hall.

SOEP Group (2001) The German Socio-Economic Panel After More Than 15 Years – Overview. *Vierteljahrshefte zur Wirtschaftsforschung, 70,* 7–14.

Spiess, M. (1998) A Mixed Approach for the Estimation of Probit Models with Correlated Responses: Some Finite Sample Results. *Journal of Statistical Computation and Simulation, 61,* 39–59.

Spiess, M. and Hamerle, A. (2000) A Comparison of Different Methods for the Estimation of Regression Models with Correlated Binary Responses. *Computational Statistics & Data Analysis, 33,* 439–455.

Spiess, M. and Keller, F. (1999) A Mixed Approach and a Distribution Free Multiple Imputation Technique for the Estimation of a Multivariate Probit Model with Missing Values, *British Journal of Mathematical and Statistical Psychology, 52,* 1–17.

Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data.* Cambridge, Massachusetts: The MIT Press.

Xie, F. and Paik, M.C. (1997) Multiple Imputation Methods for the Missing Covariates in Generalized Estimating Equation. *Biometrics, 53*, 1538–1546.

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

Zimmermann, E. (1985) Almost All You Wanted to Know About Status Inconsistency But Never Dared to Measure: Theoretical Deficits in Empirical Research on Status Inconsistency. *Social Behavior and Personality, 13(2),* 195–214.