DIW Berlin
German Institute
for Economic Research

Discussion Papers

663

Markus Pannenberg
Martin Spiess

GEE Estimation of a Two-Equation Panel Data Model: An Analysis of Wage Dynamics and the Incidence of Profit-Sharing in West Germany

Berlin, January 2007

GEE Estimation of a Two-Equation Panel Data Model: An Analysis of Wage Dynamics and the Incidence of Profit-Sharing in West Germany

Markus Pannenberg*
FH Bielefeld & DIW Berlin, IZA Bonn

Martin Spiess†
DIW Berlin & University of Flensburg

Abstract

We propose a generalized estimating equations approach to the analysis of the mean and the covariance structure of a bivariate time series process of panel data with mixed continuous and discrete dependent variables. The approach is used to jointly analyze wage dynamics and the incidence of profit-sharing in West Germany. Our findings reveal a significantly positive conditional correlation of wages and the incidence of profit-sharing. Furthermore, they indicate that permanent unobserved individual ability is comparatively more important in the profit-sharing than in the wage equation and show that shocks have a long-lasting effect on transitory wages but not on the incidence of profit-sharing. Hence, the results support theoretical predictions that selection into profit-sharing is mostly due to unobservable ability and that profit-sharing ties wages more closely to productivity.

JEL Classification: C33, C35, D31, J31, J33

Keywords: generalized estimating equations, covariance structure, longitudinal data, real wages, variable pay

* markus.pannenberg@fh-bielefeld.de

† mspiess@diw.de, SOEP, DIW Berlin, Königin-Luise-Str. 5, 14195 Berlin, Germany, phone +49-30-897 89 602 (corresponding author).

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
http://www.diw.de

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

Available for free downloading from the DIW Berlin website.

1. INTRODUCTION

Profit-sharing as a means of increasing productivity, employment, individual earnings and labor market flexibility has been discussed among politicians as well as economists for a long time (e.g. OECD 1995). Yet, the empirical evidence on the economic effects of profit sharing is still mixed. First, while there is overwhelming evidence from studies in the 1990s based on (cross-sectional) firm data suggesting that profit-sharing increases productivity (e.g. OECD 1995), recent studies point out that the results might be plagued by selectivity issues as well as unobserved heterogeneity (e.g. Kraft/Ugarkovic 2005a). Second, if profit-sharing is beneficial for firms, the question arises why still only a minority of firms in most industrialized countries has implemented profit-sharing schemes (eg. Poutsma 2001). To give an example, a current study for Germany finds that in 2005 roughly 9% of all firms use profit-sharing schemes (Bellmann/Möller 2006). Third, empirical evidence on the effects of profit-sharing on wages is scarce (e.g. Hart/Hübler 1990, Kraft/Ugarkovic 2005b). Fourth, even though theoretical models show that the major determinant of whether or not an employee participates in a profit-sharing scheme is unobservable individual ability (e.g. Booth/Frank 1999, p. 449), empirical studies that explicitly control for unobserved individual heterogeneity by means of using representative individual panel data are rare (Booth/Frank 1999).

Our study contributes to the empirical literature on the effects of profit-sharing by jointly analyzing individual wage dynamics and the incidence of profit-sharing in West Germany. West Germany is of interest, since our data indicate that the prevalence of profit-sharing among workers has increased during the 1990s. Taking the result of the theoretical literature seriously that the major determinant of the incidence of profit-sharing schemes is unobservable individual ability, our focus is on the joint analysis of the mean and the covariance structure of individual wages and the incidence of profit-sharing. We allow for unobserved individual heterogeneity, whose impact might vary across wages and the incidence of profit-sharing, as well as for different dynamics in the two individual time series. This enables us to test whether unobserved heterogeneity is indeed that important for the probability of receiving variable pay as suggested by the theoretical literature. Moreover, the estimated correlation coefficient of the two-equation system indicates whether there is a link between profit-sharing and wages conditional on selection on observables. This might give support to the premise in the theoretical literature that profit-sharing ties wages more closely to productivity.

From an econometric point of view our study contributes to the literature on the covariance structure of individual earnings (e.g. Alvarez 2004, Baker and Solon 2003, Biewen 2005, Cappellari 2004, MaCurdy 1982) as well as on the covariance structure of bivariate time serie processes (e.g. Abowd/Card 1989, Hall/Mishkin 1982). Most of these studies share the use of a two-step estimation procedure: in a first step, estimated residuals are obtained from (pooled) regression models typically including some observable individual characteristics as well as time dummies to control for aggregate common effects. These 'mean-adjusted' raw earnings are then used to test models of the covariance structure of earnings applying generalized methods of moments (GMM), maximum likelihood (ML) or pseudo-ML estimators. Arellano (2003, chap. 5) gives a comprehensive survey of covariance structures for dynamic error component models. In our study, however — and like Alvarez (2004) with respect to the univariate time series process of wages with seasonality estimated by GMM or ML — we choose a one-step approach to analyzing the joint dynamics of wages and the incidence of profit-sharing in a panel data framework. In particular, we propose a generalized estimating equations (GEE-) type two-equation panel data model with mixed continuous and binary dependent variables. Compared to the GEE approach proposed by Liang and Zeger (1986) our approach is more general, since it allows us to simultaneously estimate the parameters of the systematic and the covariance part of the model. Yet, like standard GEE estimation procedures, the estimators of the systematic part are robust with respect to potential misspecification of the covariance structure.

We proceed as follows: In Section 2, we start by describing our panel data set from West Germany, present evidence on the evolution of real monthly wages, and describe the prevalence of profit-sharing. In Section 3 we set out the econometric model, describe the estimating equations for our empirical example, outline the estimation of our model and discuss the relationship of our GEE estimator to better known GMM estimators in applied economics. Section 4 contains the empirical results. Section 5 concludes.

2. DATA AND DESCRIPTIVE EVIDENCE

Our empirical analysis is based on the German Socio-Economic Panel (GSOEP) which is a nationally representative longitudinal data set for Germany (Wagner et al. 1993, SOEP Group 2001). We use data for the years 1991 to 2000 for West Germany. The analysis is restricted to part- and full-time workers in the private sector aged 18 to 65 in the relevant years. The econometric model is estimated on a balanced panel data set with 7200 observations. To compensate for unit nonresponse up to 2000, we use longitudinal attrition factors provided with the GSOEP to weight individual contributions to the estimating equations (cf. Wooldridge 2002, pp. 577). These longitudinal attrition factors are the product of inverse conditional estimated response probabilities and design weights of the first wave (cf. Pannenberg et al. 2004).

The wage measure used is the monthly gross real labor earnings in the month preceding the interview including overtime payments. Nominal wages are deflated by the national consumer price index (base year 1995). Information on extra pay such as a 13th or 14th month salary, holiday pay as well as profit-sharing-schemes are drawn from the subsequent wave, divided by 12 and added to the monthly wage measure. The profit-sharing dummy equals one if the respondent answers that he or she received extra pay from profit-sharing schemes or premiums or bonuses in the respective year. Like Hart/Hübler (1990) we interpret the collected information from the GSOEP as incidence of profit-sharing. This is in line with evidence provided by Kruse (1993) that premiums and bonuses are mostly used in a similar way as profit-sharing schemes by firms. Moreover, following the empirical literature (e.g. Booth/Frank 1999, Hart/Hübler 1990) we include the following covariates in the profit-sharing equation: experience (in years), experience squared, years of schooling, the amount of overtime work (in hours per month), dummy variables for gender, German nationality, part-time work, occupational status (worker) as well as full sets of firm size dummies, industry dummies and time dummies. With respect to the wage equation we add tenure (in years), tenure squared and an interaction term of experience and gender. Table 4 in the appendix shows the summary statistics of the regressors for our subsample.

[Table 1 about here.]

Regarding the evolution of wages, the figures in Table 1 reveal a remarkable increase in real monthly wages over the period 1991 – 2000 in West Germany. On average, a worker earns 22% more in real terms in 2000 than in 1991. The standard deviation of the monthly real wage is also steadily increasing over the years and the percentage change adds up to 36%. The incidence of profit-sharing schemes over time in West Germany is documented in Column 3 of Table 1. In

1991, 16% of all workers in our subsample received some type of profit-sharing. This share falls to a minimum of 14% in 1994, increases to a maximum of 25% in 1998 and amounts to 23% in 2000. The OECD report (OECD 1995) gives a slightly lower figure of the incidence of profit-sharing schemes of about 10% based on information from 1994. Avalaible evidence from establishment data for the 1990s reveals that roughly 12% of all firms have implemented profit-sharing schemes in Germany (Poutsma 2001). With respect to the average real monthly amount of profit-sharing, we observe both a clear upward trend as well as substantial cyclical variation. Also, the standard deviation of the amount of profit-sharing is notably increasing over time. If we calculate the ratio of profit-sharing to the basic fixed wage without any type of incentive pay, the ratio has its minimum of 6% in 1995 and its maximum in 2000 with 8%. This is in line with the results of 5%-10% from the OECD report.

[Table 2 about here.]

Regarding the evolution of real wages for workers with and without profit-sharing (Table 2), we observe a remarkably higher level of real wages as well as a stronger increase over time for workers with profit-sharing schemes (23% vs. 16%). Moreover, the increase of the coefficient of variation is more pronounced for the group with profit-sharing (14%) than for the group without profit-sharing (5%). The descriptive evidence is in line with the argument that profit-sharing schemes tie wages more closely to productivity and hence the variation of real wages is higher in case of the existence of profit-sharing schemes.

3. ECONOMETRIC MODEL AND ESTIMATION ISSUES

In our econometric study we focus on the analysis of the joint covariance structure of wages and profit-sharing. This allows us to test whether unobservable individual ability is indeed a major determinant of an individual's participation in a profit-sharing scheme as suggested by the theoretical literature. In preliminary regressions we did start with a flexible structure of the joint covariance matrix including unobserved person-specific invariant variables, whose impact might vary across equations and time as well as AR(1) processes of the disturbances for every equation.

It turned out, however, that — at least in the data set at hand — there is a trade-off between the use of an extensive set of regressors and a complex structure of the covariance matrix. To give an example, we did observe a trade-off between including a full set of time dummies and allowing equation-specific unobserved heterogeneity to vary over time. We therefore decided to implement some restrictions on the covariance matrix as described below while including full sets of indudstry, firm size and time dummies as well as set of covariates usually employed in the relevant literature.

3.1 The Model

Consider measurements on a continuous and a binary dependent variable obtained for each of N units at each of T points in time ($n = 1, \ldots, N$; $t = 1, \ldots, T$). Let $j = 1$ denote the equation with the continuous and $j = 2$ the equation with the binary dependent variable. Our econometric model is a two-equation panel data model with a continuous dependent variable log(wage), denoted as y_{nt1}, and a binary dependent variable, incidence of profit-sharing, denoted as y_{nt2}.

For each dependent variable, we assume the following latent model

$$y^*_{ntj} = \mathbf{x}^T_{nt}\boldsymbol{\beta}_j + \epsilon_{ntj} \qquad \mathrm{var}(\epsilon_{ntj}) = \sigma^2_{\epsilon_j},$$

where $\mathbf{x}_{nt}$ is a $(K \times 1)$ vector of covariables including the element one as the constant, and the random error ϵ_{ntj} is independent of $\mathbf{x}_{ntj}$ for all n, t, j. We assume that $(\epsilon_{nt2}, \epsilon_{nt'2})$ $(t \neq t')$ are bivariate normally distributed for all t, t' with mean zero, each y^*_{nt2} conditional on covariates and $y_{n11}, \ldots, y_{nT1}$ is univariate normally distributed and each ϵ_{nt2} does depend on all ϵ_{nt1}, $t = 1, \ldots, T$, only through a linear function. We do not need the assumption of multivariate normality. The $(K \times 1)$ parameter vector $\boldsymbol{\beta}_j$ is equation-specific and may contain parameters restricted to zero for one of the two equations.

The latent dependent variables, y^*_{ntj}, are related to the observable dependent variables, y_{ntj} as follows. For the continuous variable equation, $y_{nt1} = y^*_{nt1}$ holds, and for the equation with the binary variable we have

$$y_{nt2} = \begin{cases} 0 & \text{if} \quad y^*_{nt2} \leq \kappa \\ 1 & \text{else} \end{cases}$$

where we impose the restriction $\kappa = 0$.

With respect to the joint covariance matrix of wages and profit-sharing we assume unobserved person-specific time- and equation-invariant random variables, which might have a different impact on the continuous and the binary dependent variable. This assumption implies a time-invariant biserial correlation of the error terms of the two equations. The respective correlation coefficient is denoted as ζ. Furthermore, we consider a stationary AR(1) process of the remainder disturbances in each equation to allow for additional serial dependence. The corresponding model in the error terms is

$$\epsilon_{ntj} = \vartheta_j \pi_n + \nu_{ntj} \qquad \text{and} \qquad \nu_{ntj} = \varrho_j \nu_{n(t-1)j} + \sigma_j w_{ntj},$$

where $\pi_n \sim N(0,1)$, $E(w_{nt1}) = 0$, $\text{var}(w_{nt1}) = 1$, $w_{nt2} \sim N(0,1)$, $E(\nu_{ntj}) = \mu_{\nu,j}$, $\text{var}(\nu_{ntj}) = \sigma^2_{\nu,j}$, $\text{cov}(\nu_{ntj}, \nu_{nt'j}) = \gamma_{j,tt'}$, $\nu_{n02} \sim N(\mu_{\nu,2}, \sigma^2_{\nu,2})$, $|\varrho_j| < 1$ and $E(\pi_n \nu_{n0j}) = E(\nu_{n01}\nu_{n02}) = E(\pi_n w_{ntj}) = E(\nu_{n0j} w_{ntj}) = E(w_{ntj} w_{nt'j'}) = 0$ for all j, j', t, t'. From these assumptions, $\mu_{\nu,j} = 0$ and $\sigma^2_{\nu,2} = \sigma^2_j/(1-\varrho^2_j)$ follows. With respect to the wage equation, we estimate ϑ_1, ϱ_1 and σ^2_1. However, in the profit-sharing equation, we cannot identify all parameters. Therefore, we impose the restriction $\sigma^2_2 = 1 - \varrho^2_2$ and estimate ϑ_2 and ϱ_2, respectively. The elements of

$\boldsymbol{\Sigma}$, the covariance matrix of the latent model, are

$$\begin{aligned}
\mathrm{var}(\epsilon_{nt1}) &= \vartheta_1^2 + \frac{\sigma_1^2}{1-\varrho_1^2}, \qquad \mathrm{var}(\epsilon_{nt2}) = \vartheta_2^2 + 1,\\
\mathrm{cov}(\epsilon_{nt1}\epsilon_{nt'1}) &= \vartheta_1^2 + \frac{\sigma_1^2 \varrho_1^{|t-t'|}}{1-\varrho_1^2} \qquad \text{if} \quad t \neq t',\\
\mathrm{cov}(\epsilon_{nt2}\epsilon_{nt'2}) &= \vartheta_2^2 + \varrho_2^{|t-t'|} \qquad \text{if} \quad t \neq t' \text{ and}\\
\mathrm{cov}(\epsilon_{nt1}\epsilon_{nt'2}) &= \vartheta_1\vartheta_2.
\end{aligned}$$

3.2 The Generalized Estimating Equations Approach

The approach adopted in this paper is based on the generalized estimating equations (GEE) approach proposed by Liang and Zeger (1986) which has become a standard tool in statistics and biometrics. However, as discussed below, under the assumption that the mean and the covariance structure are correctly specified, our particular GEE estimator can be interpreted as a generalized methods of moments (GMM) estimator (e.g. Hansen, 1982), although this does not hold in general. The GEE approach was introduced as an extension of univariate generalized linear models (cf. Fahrmeir und Tutz, 2001; McCullagh and Nelder, 1990) and has its roots in the methods of moments advocated by Karl Pearson as well as in the theory of optimal unbiased estimating functions (e.g. Godambe, 1960).

Let $\boldsymbol{\beta}$ be the (identifiable) parameter vector of the mean structure with possibly vector valued elements $\boldsymbol{\beta}_1$ and $\tilde{\boldsymbol{\beta}}_2$, let $\boldsymbol{\omega}$ be the vector of all (identifiable) parameters of the covariance structure of the observed dependent variables, let $\mathbf{X}_n$ be the fixed matrix collecting all vectors $\mathbf{x}_{ntj}^T$ and $\mathbf{y}_n = (y_{n11}, \ldots, y_{nTJ})^T$. It will be assumed throughout that the interpretation of $\boldsymbol{\beta}$ does not depend on the value of $\boldsymbol{\omega}$.

The starting point is the assumption of the existence of a set of unbiased estimating functions for the parameters of the mean structure, denoted as $\mathbf{g}_n \equiv \mathbf{g}_n(\mathbf{y}_n, \mathbf{X}_n, \boldsymbol{\beta})$, such that $E(\mathbf{g}_n; \boldsymbol{\beta}, \boldsymbol{\omega}|\mathbf{X}_n) = 0$ for all possible $\boldsymbol{\beta}$, $\boldsymbol{\omega}$, which are uncorrelated with each other. Optimal estimating functions in a variance minimizing sense with respect to $\mathbf{g}$ are given by

$$\mathbf{g} = \sum_{n=1}^{N} E\left(\frac{\partial \mathbf{g}_n}{\partial \boldsymbol{\beta}}\right)^T \mathrm{Cov}^{-1}(\mathbf{g}_n)\, \mathbf{g}_n,$$

where $\mathrm{Cov}(\mathbf{g}_n)$ is the covariance of $\mathbf{g}_n$, conditional on $\mathbf{X}_n$ (Godambe, 1960, 1995, Liang and Zeger, 1995). The use of $\mathbf{g}_n = (\mathbf{y}_n - \boldsymbol{\mu}_n)$, where $\boldsymbol{\mu}_n$ is a correctly specified model of the conditional mean $E(\mathbf{y}_n|\mathbf{X}_n)$ and is a function of $\boldsymbol{\beta}$ but not of $\boldsymbol{\omega}$, leads to estimating functions which have been referred to as the generalized estimating equations (GEEs) by Liang and Zeger (1986). A GEE estimator of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$, is obtained as the root of the unbiased estimating functions

$$\mathbf{0} = \sum_{n=1}^{N} \left(\frac{\partial \boldsymbol{\mu}_n}{\partial \boldsymbol{\beta}}\right)^T \mathrm{Cov}^{-1}(\mathbf{y}_n)\, (\mathbf{y}_n - \boldsymbol{\mu}_n), \tag{1}$$

where $\mathrm{Cov}(\mathbf{y}_n)$ is the covariance of $\mathbf{y}_n$, conditional on $\mathbf{X}_n$, and depends on $\boldsymbol{\omega}$.

Usually $\boldsymbol{\omega}$ is unknown and must be estimated. However, it can be shown that the nuisance parameter, $\boldsymbol{\omega}$, has only little impact on $\mathbf{g}$ and on the solution of $\mathbf{g} = \mathbf{0}$ at least for large N (Liang and Zeger, 1986, 1995). Thus, replacing $\boldsymbol{\omega}$ by any consistent estimator $\hat{\boldsymbol{\omega}}$ of $\boldsymbol{\omega}$, e.g. the classical minimum distance estimator, the asymptotic variance of $\hat{\boldsymbol{\beta}}$ is not affected. Further, if $\mathrm{Cov}(\mathbf{y}_n)$ is correctly specified $\hat{\boldsymbol{\beta}}$ has minimum asymptotic variance within the class of asymptotically linear estimators (McCullagh, 1983). But even if $\mathrm{Cov}(\mathbf{y}_n)$ is misspecified, $\boldsymbol{\beta}$ is consistently estimated by the root of (1), although efficiency is lost in this case.

Liang and Zeger (1986) assume that each y_{ntj} follows a simple univariate exponential distribution (e.g. Fahrmeir and Tutz, 2001; McCullagh and Nelder, 1990). The advantage of this assumption is that it implies not only a model for the theoretical (conditional) mean of y_{ntj}, but also of the theoretical (conditional) variance, which is a known function of $\mathbf{x}_{ntj}$, $\boldsymbol{\beta}_1$, $\tilde{\boldsymbol{\beta}}_2$ and a dispersion parameter, ϕ_j, which might be known in some models but must be estimated in others. For example, for continuous dependent variables and assuming normality $\phi = \sigma^2$ and for binary dependent variables $\phi = 1$ (Fahrmeir and Tutz, 2001; McCullagh and Nelder, 1990). To complete the estimating equations for $\boldsymbol{\beta}$, Liang and Zeger (1986) propose a 'working' correlation matrix, $\mathbf{R}(\boldsymbol{\alpha})$, which is common to all units and is a 'working' model of the correlation structure in the observed dependent variables, where $\boldsymbol{\alpha}$ is a possibly vector valued parameter. It can be shown, that if the y_{ntj} are independent and follow an exponential distribution, then choosing $\alpha = 0$ leads to estimating equations which correspond to score equations in many cases (McCullagh and Nelder, 1990).

3.3 Estimation of Mean and Covariance Parameters

Unlike standard GEE approaches we are interested in estimating both mean and covariance structure parameters of a latent model based on the GEE approach. Such estimators have been proposed by Prentice (1988), Zhao and Prentice (1990), Qu et al (1992, 1994) or Reboussin and Liang (1998) among others (for GMM estimators see, e.g., Breitung and Lechner 1999). In contrast to Prentice (1988), Zhao and Prentice (1990) and Qu et al. (1992, 1994), our estimating equations for the covariance parameters of the latent model, $\boldsymbol{\theta}$, are equal to score equations under the assumption that subsets of the dependent variables are independent, thereby generalizing Spiess (1998) and Spiess and Keller (1999). Unlike Reboussin and Liang (1998), we estimate the mean and covariance parameters as if they were orthogonal. Thus at the price of lower efficiency, the parameters of the mean structure can consistently be estimated even if the models for the covariances structure are misspecified, given a correct specification of the mean model only.

To estimate the parameters of the mean structure, we adopt estimating equations (1), where for each binary dependent variable, y_{nt2}, we assume the corresponding element of $\boldsymbol{\mu}$, μ_{nt2}, to be equal to $\mu_{nt2} = \Phi(\eta_{nt2})$, where $\Phi(\cdot)$ is the cumulative function of the standard normal distribution and $\eta_{nt2} = \mathbf{x}_{nt}^T \tilde{\boldsymbol{\beta}}_2$, $\tilde{\boldsymbol{\beta}}_2 = \sigma_{\epsilon_2}^{-1} \boldsymbol{\beta}_2$, $\sigma_{\epsilon_2}^2 = \vartheta_2^2 + 1$. For each continuous dependent variable, y_{nt1}, we assume a linear model, i.e. the corresponding elements of $\boldsymbol{\mu}$, μ_{nt1}, reduce to $\mu_{nt1} = \eta_{nt1}$, where

$\eta_{nt1} = \mathbf{x}_{nt}^T \boldsymbol{\beta}_1$. However, unlike e.g. Liang and Zeger (1986), we do not adopt a 'working' correlation matrix common to all individuals, but use individual specific covariance matrices instead which follow from the assumed covariance structure and depend on $\boldsymbol{\theta}$, the vector of all covariance structure parameters of the latent model (see Appendix A.1).

For simplicity, let the elements of the vector of all identifiable covariance parameters of the latent model, $\boldsymbol{\delta}$, be arranged in subvectors, so that $\boldsymbol{\delta}_{c,v}$ denotes the vector of variances, $\boldsymbol{\delta}_{c,c}$ the vector of all correlations of the error terms $\epsilon_{n11}, \ldots, \epsilon_{nT1}$ of the linear equations, $\boldsymbol{\delta}_b$ the vector of all (tetrachoric) correlations of $\epsilon_{n12}, \ldots, \epsilon_{nT2}$, and $\boldsymbol{\delta}_{cb}$ denotes the vector of (biserial) correlations of all pairs $(\epsilon_{nt1}, \epsilon_{nt'2})$. Further, $\boldsymbol{\Sigma}_c$ is the covariance matrix part of $\boldsymbol{\Sigma}$ corresponding to all continuous dependent variables and $\boldsymbol{S}_n = (\mathbf{y}_{n1} - \boldsymbol{\mu}_{n1})(\mathbf{y}_{n1} - \boldsymbol{\mu}_{n1})^T$. Accordingly, denote by $\mathbf{R}_c$ that part of the correlation matrix of the latent errors that correspond to all continuous dependent variables. Note, however, that the estimating equations for $\boldsymbol{\delta}$ are usually not equal to those for the parameter of interest, $\boldsymbol{\theta}$, which are in general of lower dimensionality. In fact, it would be not a good idea to really estimate all possible covariance parameters, as this would in our case imply the estimation of $2T^2 = 200$ covariance parameters.

The individual contributions to the estimating equations for $\boldsymbol{\delta}_{c,v}$ and $\boldsymbol{\delta}_{c,c}$ are given by

$$\mathbf{u}_{n,c} = \left(\frac{\partial \boldsymbol{\Sigma}_c}{\partial \boldsymbol{\delta}_c}\right)^T (\boldsymbol{\Sigma}_c^{-1} \otimes \boldsymbol{\Sigma}_c^{-1})\, \mathbf{e}_{n,c}, \tag{2}$$

where $\mathbf{e}_{n,c} = \mathrm{vec}(\mathbf{S}_n - \boldsymbol{\Sigma}_c)$, vec$(\cdot)$ is the usual vec operator, $\otimes$ is the Kronecker product and $\boldsymbol{\delta}_c = (\boldsymbol{\delta}_{c,v}^T, \boldsymbol{\delta}_{c,c}^T)^T$. Note that $E(\mathbf{u}_{n,c}) = \mathbf{0}$ at the true parameter values if $\boldsymbol{\Sigma}_c$ is correctly specified and that for consistent estimation of $\boldsymbol{\delta}_{c,v}$ and $\boldsymbol{\delta}_{c,c}$ no distributional assumption is necessary. It is easy to see that (2) are equal to the score equations derived from the log likelihood under multivariate normality if all outcomes were continuous.

The estimating equations for the tetrachoric correlations consider each possible pair of binary dependent variables as a three-dimensional polytomous variable and equate this variable with its theoretical mean. Denote the (3×1)-vector representing a pair of binary variables as $\mathbf{v}_{ntt',b}$ and its theoretical mean, which is equal to a vector of probabilities, as $\boldsymbol{\mu}_{ntt',b}$. Note that the latter can, under bivariate normality of the corresponding errors in the latent model, easily be evaluated using the bivariate cumulative standard normal distribution function. Then the individual contributions to the estimating equations for the tt'th element $(t = 2, \ldots, T,\ t' = 1, \ldots, t-1)$ of $\boldsymbol{\delta}_b$ are

$$u_{n,tt',b} = \left(\frac{\partial \boldsymbol{\mu}_{ntt',b}}{\partial \delta_{tt',b}}\right)^T \mathbf{W}_{ntt',b}^{-1}(\mathbf{v}_{ntt',b} - \boldsymbol{\mu}_{ntt',b}). \tag{3}$$

where $\mathbf{W}_{ntt',b} = (\mathrm{diag}(\boldsymbol{\mu}_{ntt',b}) - \boldsymbol{\mu}_{ntt',b}\boldsymbol{\mu}_{ntt',b}^T)$ and diag$(\mathbf{a})$ denotes a diagonal matrix with diagonal elements equal to $\mathbf{a}$. It is easy to show (cf. Amemiya, 1985, sec. 9.3) that the estimating equations (3) are equal to individual pseudo-score equations derived from the pseudo-log likelihood for $\delta_{tt',b}$ based on observations y_{nt2} and $y_{nt'2}$ under the assumption of bivariate normality of the errors and mutual independence of all possible pairs. All $(T(T-1)/2 \times 1)$ vectors $(\mathbf{v}_{ntt'b} - \boldsymbol{\mu}_{ntt'b})$ are collected in $\mathbf{e}_{n,b}$ and all matrices $\mathbf{W}^{-1}_{ntt',b}$ in the block diagonal matrix $\mathbf{W}^{-1}_{n,b}$.

The estimating equations for the biserial correlations are obtained by equating the binary dependent variables and their theoretical means given the observed continuous dependent variables. Thus the individual contributions to the estimating equations for $\boldsymbol{\delta}_{t,cb}$ are

$$\mathbf{u}_{n,t,cb} = \Big(\frac{\partial \Phi(\psi_{nt})}{\partial \boldsymbol{\delta}_{t,cb}}\Big)^T (\Phi(\psi_{nt})(1 - \Phi(\psi_{nt})))^{-1}(y_{nt2} - \Phi(\psi_{nt})), \tag{4}$$

where

$$\psi_{nt} = (1 - \boldsymbol{\delta}^T_{t,cb}\mathbf{R}^{-1}_c\boldsymbol{\delta}_{t,cb})^{-1/2}(\eta_{nt2} + \boldsymbol{\delta}^T_{t,cb}\mathbf{R}^{-1}_c\mathbf{V}^{-1/2}_c(\mathbf{y}_{n1} - \boldsymbol{\mu}_{n1})), \tag{5}$$

and $\mathbf{V}_c = \mathrm{Diag}(\boldsymbol{\Sigma}_c)$. See Appendix A.2 for a derivation of (4) from the corresponding pseudo-log likelihood. All T scalar terms $(y_{nt2} - \Phi(\psi_{nt}))$ are collected in $\mathbf{e}_{n,cb}$ and all scalar terms $(\Phi(\psi_{nt})(1 - \Phi(\psi_{nt})))^{-1}$ are collected in the diagonal matrix $\mathbf{W}^{-1}_{n,cb}$.

As the above discussion shows, only the correct specification of uni- and bivariate distributions of subsets of the dependent variables is necessary for a consistent estimation of all parameters of interest. Hence, only one- and two-dimensional integrals have to be evaluated. This is a clear advantage of our approach compared to maximum likelihood estimators, where the joint multivariate distribution of all error terms must be specified and high-dimensional integrals must be evaluated. However, the advantageous properties of our GEE-type approach come at the price of a loss of efficiency. Simulation results for a simpler one-equation panel model with binary dependent variables suggest that the efficiency loss relative to the maximum likelihood estimator is rather small (Spiess 1998). This is in line with Liang and Zeger (1995), who state that according to their experience, the gain in robusteness is far greater than the loss in efficiency. Further, the results in Spiess (1998) imply that using individual covariance matrices leads to more efficient estimators as compared to adopting a 'working' correlation matrix, common to all units, as proposed by Liang and Zeger (1986).

3.4 Estimation of the Model

Collect all vectors $\mathbf{e}_{n,c}$, $\mathbf{e}_{n,b}$ and $\mathbf{e}_{n,cb}$ in $\mathbf{e}_{n,2}$ and all matrices $(\mathbf{\Sigma}_c^{-1} \otimes \mathbf{\Sigma}_c^{-1})$, $\mathbf{W}_{n,b}^{-1}$ and $\mathbf{W}_{n,cb}^{-1}$ in the block diagonal matrix $\mathbf{W}_{n,2}^{-1}$. Further let $\mathbf{D}_{n,2}$ include the terms $\partial\mathbf{\Sigma}_c/\partial\boldsymbol{\delta}_c$, $\partial\boldsymbol{\mu}_{ntt',b}/\partial\delta_{tt',b}$ $(t = 2, \ldots, T, \quad t' = 1, \ldots, t-1)$, $\partial\Phi(\psi_{nt})/\partial\boldsymbol{\delta}_{t,cb}$ and $\partial\Phi(\psi_{nt})/\partial\boldsymbol{\delta}_c$ $(t = 1, \ldots, T)$.

Then, to estimate the parameters of the assumed covariance structure, we use

$$\mathbf{u}_2 = \left(\frac{\partial\,\boldsymbol{\delta}}{\partial\,\boldsymbol{\theta}}\right)^T \sum_{n=1}^{N} \mathbf{D}_{n,2}\mathbf{W}_{n,2}^{-1}\mathbf{e}_{n,2} = \mathbf{0},$$

where $\boldsymbol{\delta} = (\boldsymbol{\delta}_c^T, \boldsymbol{\delta}_b^T, \boldsymbol{\delta}_{cb}^T)^T$ and, in this paper, $\boldsymbol{\theta} = (\vartheta_1, \vartheta_2, \sigma_1^2, \varrho_1, \varrho_2)^T$. Let $\mathbf{u}_1 = \sum_{n=1}^N \mathbf{u}_{n,1}$. All parameters of interest are estimated simultaneously by stacking the estimating equations and solving $(\mathbf{u}_1^T, \mathbf{u}_2^T)^T = \mathbf{0}$ for $\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the vector of elements of $\boldsymbol{\beta}_1$ and $\tilde{\boldsymbol{\beta}}_2$ not restricted to zero, and $\boldsymbol{\theta}$.

Let $\mathbf{\Gamma}_{11} = \partial\mathbf{u}_1/\partial\boldsymbol{\beta}$ and $\mathbf{\Gamma}_{22} = \partial\mathbf{u}_2/\partial\boldsymbol{\theta}$. The vector of estimates, $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\theta}}^T)^T$, is iteratively calculated with updated value in the $(j+1)$th iteration given by

$$\hat{\boldsymbol{\gamma}}_{j+1} = \hat{\boldsymbol{\gamma}}_j + \begin{pmatrix} \boldsymbol{\Gamma}_{11}^T & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{22}^T \end{pmatrix}^{-1}_{\gamma=\hat{\gamma}_j} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix}_{\gamma=\hat{\gamma}_j}.$$

The unknown parameters in $\boldsymbol{\Omega}_n$ are replaced with their estimates.

The asymptotic covariance matrix is estimated by

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\gamma}}) = \begin{pmatrix} \boldsymbol{\Gamma}_{11}^T & \mathbf{0} \\ \boldsymbol{\Gamma}_{12}^T & \boldsymbol{\Gamma}_{22}^T \end{pmatrix}^{-1}_{\gamma=\hat{\gamma}} \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix}_{\gamma=\hat{\gamma}} \begin{pmatrix} \boldsymbol{\Gamma}_{11} & \boldsymbol{\Gamma}_{12} \\ \mathbf{0} & \boldsymbol{\Gamma}_{22} \end{pmatrix}^{-1}_{\gamma=\hat{\gamma}}$$

where

$$\boldsymbol{\Gamma}_{12} = \frac{\partial \mathbf{u}_2}{\partial \boldsymbol{\beta}}$$

which accounts for the fact that $\mathbf{u}_2$ is a function of $\boldsymbol{\beta}$, and

$$\begin{aligned} \mathbf{W}_{11} &= \sum_{n=1}^{N} \mathbf{u}_{n,1}\mathbf{u}_{n,1}^T \quad \mathbf{W}_{12} = \sum_{n=1}^{N} \mathbf{u}_{n,1}\mathbf{u}_{n,2}^T \Big(\frac{\partial \boldsymbol{\delta}}{\partial \boldsymbol{\theta}}\Big)^T \\ \mathbf{W}_{22} &= \Big(\frac{\partial \boldsymbol{\delta}}{\partial \boldsymbol{\theta}}\Big) \sum_{n=1}^{N} \mathbf{u}_{n,2}\mathbf{u}_{n,2}^T \Big(\frac{\partial \boldsymbol{\delta}}{\partial \boldsymbol{\theta}}\Big)^T \quad \text{and} \quad \mathbf{W}_{21} = \mathbf{W}_{12}^T. \end{aligned}$$

Note that weighting factors to compensate for missing observations can easily be incorporated in our estimation framework following, e.g., Wooldridge (2002).

3.5 GEE vs. GMM

To compare our GEE approach with the GMM approach, which is more widely adopted in economic applications, first note that the number of equations in (1) is equal to the dimension of $\boldsymbol{\beta}$. Hence, the problem is just identified. Further, (1) times N^{-1} is the sample counterpart of the set of (conditional) moment conditions

$$E(\mathbf{H}(\mathbf{X}_n, \boldsymbol{\beta}, \boldsymbol{\delta})(\mathbf{y}_n - \boldsymbol{\mu}_i)) = \mathbf{0}$$

which is implied by the assumption $E(\mathbf{g}_n; \boldsymbol{\beta}, \boldsymbol{\omega}|\mathbf{X}_n) = \mathbf{0}$ for all possible $\boldsymbol{\beta}$, $\boldsymbol{\omega}$ and $n = 1, \ldots, N$ (cf. section 3.2) and fixed matrix of instruments $\mathbf{H}(\mathbf{X}_n, \boldsymbol{\beta}, \boldsymbol{\delta})$. An optimal matrix of instruments is given by

$$\mathbf{H}(\mathbf{X}_n, \boldsymbol{\beta}, \boldsymbol{\delta}) = E(\partial \boldsymbol{\mu}_n / \partial \boldsymbol{\beta}) \mathrm{Cov}^{-1}(\mathbf{y}_n)$$

(Newey and McFadden, 1994, p. 2170) leading to the estimating equations described in section 3.2 (cf. Godambe, 1995), which depend on $\boldsymbol{\beta}$ and the nuisance parameter $\boldsymbol{\delta}$. However, $\boldsymbol{\delta}$ is unknown and must be estimated. The plug-in approach, using some consistent estimator for $\boldsymbol{\delta}$ leads to a feasible GMM estimator for $\boldsymbol{\beta}$ using (1) (Newey and McFadden, 1994, p. 2171). By construction of $\mathbf{g}_n$, estimation of $\boldsymbol{\omega}$ or $\boldsymbol{\delta}$ does not affect the asymptotic variance of $\hat{\boldsymbol{\beta}}$ (see section 3.2).

In general, the GEE approach does not require that the estimator of $\boldsymbol{\omega}$ is a GEE (or GMM) estimator. It may be any estimator converging in probability to some $\boldsymbol{\omega}$, which not even need to be the 'true' value. In the latter case, where the instruments are not optimal, efficiency of $\hat{\boldsymbol{\beta}}$ is lost. Thus, stacking the estimating functions for $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ does not in general lead to a two-step GMM estimator, although $\hat{\boldsymbol{\beta}}$ can always be interpreted as a „plug-in" GMM estimator.

However, section 3.3 describes estimating equations for $\boldsymbol{\delta}$, the vector of all covariance parameters of the latent model which are derived from pseudo log-likelihood functions assuming independence of certain subsets of variables. Again, this is a just identified problem, but the instruments chosen now to simplify calculations are not optimal. The estimating equations are equivalent to conditional moment conditions and thus, the resulting estimator $\hat{\gamma}$ could be denoted as a „multi-step plug-in“ GMM estimator if the covariance structure is correctly specified.

An interesting general difference between the GEE and the GMM approach is that in the former optimality results are ascribed to the estimating functions whereas in the latter they are ascribed to the estimator. This difference has been an issue in the statistical literature (e.g. Crowder, 1989) although in general it may not make a big difference in large samples (cf. Liang and Zeger, 1995, Godambe, 1995). However, given the equality of the GMM and the GEE estimator described in this paper, optimality results hold for both, the estimating functions and the estimator.

4. RESULTS

Table 3 provides the parameter estimates of our two-equation panel data model.

[Table 3 about here.]

Starting with the estimated parameters of the joint covariance matrix, Table 3 reveals that the estimated correlation coefficient between the two equations, ζ, equals 0.30 and is significantly different from zero. Hence, we do observe a notable positive co-movement between variations in the wage residuals and variations in the profit-sharing residuals conditional on the sets of observed variables in both equations. This indicates that there is a positive link between shocks in the wage equation and the incidence of profit-sharing. Hence, we find supportive evidence for the premise in the theoretical literature that profit-sharing ties wages more closely to productivity. For example, Booth/Frank (1999) illustrate in their theoretical model that conditional on observed individual characteristics average wage differentials across different payment schemes are a good measure of average productivity differences. Our result also implies that the variation of real wages is higher under profit-sharing regimes as indicated by our descriptive evidence.

The variance components ϑ_1 and ϑ_2, capturing the impact of time invariant unobservable individual ability on wages respectively the incidence of profit-sharing, are significantly different from zero. The share of the variance due to the permanent component relative to the overall variance amounts to 51% for the profit-sharing equation while it is only 17% for the wage equation. This clearly shows that unobservable individual ability is a major determinant of whether someone participates in a profit-sharing scheme or not as suggested by the theoretical model of Booth/Frank (1999).

The estimated parameters of the AR(1) process in both equations are positive and significantly different from zero. The estimated parameter of $\varrho_1 = 0.87$ for the wage equation implies that after five years, 50% of a shock is still present in transitory wages. This indicates that shocks have a long-lasting effect on transitory real wages in the 1990s in West Germany conditional on observed characteristics as well as on the time-invariant permanent earnings component of our specification. Regarding the estimated AR(1) parameter for the profit-sharing equation, we find that the effect of a shock on the transitory component of the likelihood of receiving variable pay is less important than in the wage equation. After five years, only 7% of a shock is still present in the transitory component of the probability of profit-sharing incidence. Hence, transitory shocks concerning the use of profit-sharing schemes by firms do not exhibit a long memory.

With respect to the observed characteristics of the wage equation, we find that the estimated parameters for male workers, German workers, years of schooling and firm size are significantly positive, the estimated parameter for part-time work is significantly negative and the estimated experience profile is concave. These results are generally in line with findings reported in the literature (e.g. Wolf 2002, Fitzenberger/Kunze 2006). This also holds for the significantly positive estimate of the amount of overtime worked in the month before the interview, which indicates that paid as well as other types of overtime work exert a positive impact on current wages (Pannenberg 2005). Note however, that the estimated gender wage differential in our data is at the upper bound estimated in the literature. Reasons might be that our balanced panel requires continuous labor market participation over 10 years and covers all age cohorts. Moreover, we cannot identify significant effects of tenure on wages. This might also be explained by the balanced panel data structure of our data at hand, since we can only separately identify experience and tenure effects if we observe a remarkable amount of workers who switch firms.

Considering the results of the profit-sharing equation, the estimated parameters indicate that the probability of receiving variable pay increases with firm size. This is in line with evidence based on establishment data for Germany (Kraft/Ugarkovic 2005a). Moreover, German workers as well as part-time workers exhibit a higher probability of receiving profit-sharing. None of the other estimated parameters is significantly different from zero supporting again and in line with our findings with respect to the variance components the theoretical model of Booth/Frank (1999) that suggests that profit-sharing mainly rewards unobservable ability.

The asymptotic Wald test indicates that the joint impact of all regressors is significantly different from zero. The Pseudo-R^2 measure given in Table 3 is identical to $\hat{\rho}_2^2$, proposed by Spiess and Tutz (2004, p. 138) as a measure of the explanatory power of the model. The intuition underlying $\hat{\rho}_2^2$ is that a measure of the explanatory power should take into account all components which are explicitly modeled when a regression model is estimated, i.e. in our case also the covariance model. This is based on ideas presented in Glahn (1969) and Carter and Nagar (1977). Its value suggests that the model possesses substantial explanatory power, i.e. both its systematic and its covariance part.

5. CONCLUSIONS

We analyze the joint covariance structure of monthly wages and the incidence of profit-sharing in West Germany. We show that (a) the ratio of the permanent variance component and the overall variance is greater for the profit-sharing equation than for the wage equation, that (b) there is a significantly positive co-movement of the wage and profit-sharing residuals and that (c) shocks in the wage equation have a long-lasting effect on transitory labor earnings, while they exhibit no enduring impact on the likelihood of receiving variable pay. Our findings therefore give supportive evidence for theoretical models stressing the impact of unobservable individual attributes on the probability of participating in profit-sharing schemes. The results also demonstrate that profit-sharing indeed ties individual wages more closely to productivity as suggested in the theoretical literature. Furthermore, the variation of real wages is higher under profit-sharing regimes. Combining all these findings with the descriptive evidence of an increase in the incidence of profit-sharing, one might conclude that one reason for the increasing wage inequality in Germany in the 1990s (e.g. Riphahn 2002) is the increasing prevalence of profit-sharing among employees. Future research investigating more explicitly the link between profit-sharing and the increase in wage inequality in Germany therefore seems promising.

Our proposed GEE-type approach for the analysis of a two-equation panel data model with a continuous and a discrete dependent variable as well as a joint covariance matrix, which is equivalent to a multi-step plug-in GMM estimator, can in principal be extended to systems of multiple equations for panel data with mixed continuous, discrete and ordered dependent variables. If sufficient panel data is at hand to identify all parameters of the specified variance covariance matrix, the approach is quite flexible, since in principle no restriction on the covariance model — beyond being positive (semi-) definite — is required. Essentially, the covariance structure parameters are formulated as functions of all identifiable correlations and variances. Yet, the approach remains quite simple technically as there are no integrals to be calculated of a dimension higher than two.

ACKNOWLEDGMENTS

We would like to thank two anonymous referees, an associate editor and Uwe Jirjahn for helpful comments and suggestions on a previous version of the paper. Responsibility for any remaining errors is entirely our own.

APPENDIX A.1: THE COVARIANCE MATRIX $\mathbf{\Omega}_n$

The matrix $\mathbf{\Omega}_n$ is a 'working' covariance matrix and is considered to be fixed in the first set of estimating equations (1). For each continuous variable, the corresponding element on the diagonal is equal to the corresponding entry in $\mathbf{\Sigma}$. For each binary dependent variable, the corresponding element is $\mu_{nt2}(1-\mu_{nt2})$. For all pairs of continuous dependent variables, the corresponding off-diagonal entries in $\mathbf{\Omega}_n$ are identical to those of $\mathbf{\Sigma}$. For all pairs of binary dependent variables, we have

$$\text{cov}(y_{nt2}, y_{nt'2}|\mathbf{x}_{nt}, \mathbf{x}_{nt'}) = \Phi(\eta_{nt2}, \eta_{nt'2}, \rho_{b,tt'}) - \Phi(\eta_{nt2})\Phi(\eta_{nt'2}),$$

where $\Phi(\cdot, \cdot, \rho_{tt',b})$ is the cumulative function of the bivariate standard normal distribution and $\rho_{tt',b}$ is the correlation of ϵ_{nt2} and $\epsilon_{nt'2}$. The entries in $\mathbf{\Omega}_n$ corresponding to the covariances of each binary with all continuous dependent variables are easily derived from the corresponding moments representation and are given by

$$\text{cov}(\mathbf{y}_{n1}, y_{nt2}|\mathbf{X}_n) = -\sigma_{\epsilon_2}^{-1}\mathbf{\Sigma}_{t,cb}\phi(\eta_{nt2}),$$

where $\mathbf{\Sigma}_{t,cb}$ is that part of $\mathbf{\Sigma}$ that denotes the covariances between all continuous dependent variables and y^*_{nt2} given the covariates, and $\phi(\cdot)$ is the density function of the standard normal distribution.

APPENDIX A.2: ESTIMATING EQUATIONS FOR BISERIAL CORRELATIONS

The likelihood for $\boldsymbol{\delta}_{t,cb}$ based on one binary dependent variable, y_{nt2}, and the $(T \times 1)$-vector of continuous variables $\mathbf{y}_{n1}$, can be written as

$$\begin{aligned} L(\boldsymbol{\delta}_{t,cb}) &= \prod_{n=1}^{N} \Big(\Pr(y_{nt2} = 1|\mathbf{y}_{n1}, \mathbf{x}_{nt2})^{y_{nt2}} (1 - \Pr(y_{nt2} = 1|\mathbf{y}_{n1}, \mathbf{x}_{nt2}))^{1-y_{nt2}} \\ &\qquad \times f(\mathbf{y}_{n1}|\mathbf{x}_{n11}, \ldots \mathbf{x}_{nT1})\Big). \end{aligned}$$

The log likelihood is

$$\begin{aligned} &l(\boldsymbol{\delta}_{t,cb}) = \text{const}+ \\ &\quad \sum_{n=1}^{N} \Big(y_{nt2} \Pr(y_{nt2} = 1|\mathbf{y}_{n1}, \mathbf{x}_{nt2}) + (1 - y_{nt2})(1 - \Pr(y_{nt2}|\mathbf{y}_{n1}, \mathbf{x}_{nt2}))\Big) \end{aligned}$$

where const is a term not involving $\boldsymbol{\delta}_{t,cb}$. Assuming that ϵ_{nt2} is normally distributed given $\mathbf{y}_{n1}$, does only linearly depend on $\boldsymbol{\epsilon}_{n1}$ and that y_{nt2} does not depend on covariates $\mathbf{x}_{n11}, \ldots, \mathbf{x}_{nT1}$ given $\mathbf{x}_{nt2}$, this is

$$l(\boldsymbol{\delta}_{t,cb}) = \text{const} + \sum_{n=1}^{N} (y_{nt2} \log \Phi(\psi_{nt}) + (1 - y_{nt2}) \log(1 - \Phi(\psi_{nt}))),$$

where ψ_{nt} is given by (5).

APPENDIX B: COVARIATES IN THE REGRESSION MODEL

[Table 4 about here.]

REFERENCES

Abowd J, Card D (1989) On the Covariance Structure of Earnings and Hours Changes. Econometrica 57: 411–445.

Alvarez J (2004) Dynamics and Seasonality in Quarterly Panel Data: An Analysis of Earnings Mobility in Spain. Journal of Business and Economics Statistics 22: 443–456.

Amemiya T (1985) Advanced Econometrics. Harvard University Press, Cambridge, Massachusetts.

Arellano M (2003) Panel Data Econometrics. Oxford University Press, Oxford, U.K.

Baker M, Solon G (2003) Earnings Dynamics and Inequality among Canadian Men, 1976 - 1992: Evidence from Longitudinal Income Tax Records. Journal of Labor Economics 21: 289–321.

Bellmann L, Möller I. (2006) Gewinn- und Kapitalbeteiligung der Mitarbeiter. IAB Kurzbericht 13.

Biewen, M. (2005) The Covariance Structure of East and West German Incomes and its Implications for the Persistence of Poverty and Inequality, German Economic Review 6: 445–469

Booth A, Frank J (1999) Earnings, Productivity, and Performance-Related Pay. Journal of Labor Economics 17: 447–463.

Breitung J, Lechner M (1999) Alternative GMM Methods for Nonlinear Panel Data Models, in: Mátyás, L. (ed.), Generalised Methods of Moments Estimation, Cambridge University Press, 1999, 248–274.

Cappellari L (2004) The Dynamics and Inequality of Italian Men's Earnings: Long-term Changes or Transitory Fluctuations ? The Journal of Human Resources 39: 475–499.

Carter RAL, Nagar AL (1977) Coefficients of Correlation for Simultaneous Equation Systems. Journal of Econometrics: 39–50.

Crowder, M.J. (1989). Comment on „An Extension of quasi-likelihood estimation“ by V.P. Godambe and M.E. Thompson, J. Statist. Plann. Inference, 22, 167–168.

Fahrmeir, L. & Tutz, G. (2001, 2nd ed.). Multivariate Statistical Modelling Based on Generalized Linear Models. New York: Springer.

Fitzenberger, B. and Kuntze, A. (2006) Vocational training and gender: Wages and occupational mobility among young workers. Oxford Review of Economic Policy (forthcoming).

Glahn HR (1969) Some Relationships Derived From Canonical Correlation Theory. Econometrica: 252–256.

Godambe VP (1960) An optimum property of regular maximum likelihood estimation. Annals of Mathematical Statistics 31: 1208–1212.

Godambe VP (1995) Comment on „Inference Based on Estimating Functions in the Presence of Nuisance Parameters“ by K-Y Liang and SL Zeger. Statistical Science 10: 173–174.

Hall RE, Mishkin FS (1982) The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households. Econometrica 50: 461–481.

Hansen LP (1982) Large Sample Properties of Generalized Method of Moments Estimators. Econometrica 50: 1029–1054.

Hart RA, Hübler O. (1990) Wage, Labour Mobility and Working Time Effects of Profit Sharing. Empirica 17: 115–130.

Kraft K., Urgarkovic, M. (2005a) The Output, Employment and Productivity Effects of Profit Sharing: A Matching Approach, DP Dortmund/Mannheim.

Kraft K., Urgarkovic, M. (2005b) Profit-Sharing: Supplement or Substitute, DP Dortmund/Mannheim.

Kruse D. (1993) Profit-sharing:Does it make a difference? W.E. Upjohn Institute for Employment Research, Kalamazoo, MI USA.

Liang K-Y, Zeger SL (1986) Longitudinal Data Analysis using Generalized Linear Models. Biometrika 73: 13–22.

Liang K-Y, Zeger SL (1995) Inference Based on Estimating Functions in the Presence of Nuisance Parameters. Statistical Science 10: 158–173.

MaCurdy TE (1982) The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis. Journal of Econometrics

18: 83–114.

McCullagh, P. (1983). Quasi-Likelihood Functions. The Annals of Statistics, 11(1), 59–67.

McCullagh P, Nelder JA (1990) Generalized Linear Models (2nd Ed.). London: Chapman and Hall.

Meghir C, Pistaferri L (2004) Income Variance Dynamics and Heterogeneity. Econometrica 72: 1–32.

Newey WK, McFadden D (1994) Large Sample Estimation and Hypothesis Testing. In RF Engle and DL McFadden (Eds.) Handbook of Econometrics, Volume IV, (pp. 2111–2245). Elsevier: Amsterdam.

OECD (1995) Profit-Sharing in OECD Countries. Employment Outlook 139-169.

Pannenberg, M. et al. (2004) Sampling and Weighting. In: Haisken-DeNew, J., Frick, J. (eds.), Desktop Companion to the GSOEP, Berlin.

Pannenberg M (2005) Long-Term Effects of Unpaid Overtime. Evidence for West Germany. Scottish Journal of Political Economy 52: 177–193.

Poutsma E (2001) Recent Trends in Employee and Financial Participation in the European Union, European Commission.

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

Qu Y, Piedmonte MR, Williams GW (1994) Small sample validity of latent variable models for correlated binary data. Communications in Statistics — Simulation and Computation 23: 243–269.

Qu Y, Williams GW, Beck GJ, Medendorp SV (1992) Latent variable models for clustered dichotomous data with multiple subclusters. Biometrics 48: 1095–1102.

Reboussin BA, Liang K-Y (1998) An Estimating Equations Approach for the LISCOMP Model. Psychometrika 63: 165–182.

Riphahn, R. (2002) Bruttoeinkommensverteilung in Deutschland 1984 - 1999 und Ungleichheit unter auslaendischen Erwerbstaetigen DIW Discussion Paper, Berlin.

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, 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.

Spiess M, Tutz G. (2004) Alternative measures of the explanatory power of general multivariate regression models. Journal of Mathematical Sociology 28: 125–146.

Wagner G, Burkhauser R, Behringer F (1993) The English Language Public Use File of the German Socio-Economic Panel. The Journal of Human Resources 28: 429–433.

Wolf E (2002) Lower Wage Rates for Lesser Hours? A simultaneous wage- hours model for Germany. Labour Economics 9: 643–663.

Wooldridge JM (2002) Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge.

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

Table 1: *Real Wages and Profit-Sharing in West Germany 1991–2000*

Year	Real wages Mean	Real wages Std.Dev.	Incidence profit-sharing Mean	Amount of profit-sharing Mean	Amount of profit-sharing Std.Dev.	Ratio profit-sharing/Fixed real wage Mean
1991	4855.57	2295.63	0.16	463.84	673.92	0.08
1992	5044.09	2321.57	0.18	428.01	676.64	0.07
1993	5161.24	2494.86	0.16	398.81	603.95	0.06
1994	5222.89	2497.59	0.14	555.69	742.75	0.08
1995	5443.35	2667.42	0.18	407.74	566.59	0.06
1996	5532.62	2725.87	0.19	480.68	678.01	0.06
1997	5591.18	2772.91	0.18	490.73	920.46	0.07
1998	5716.53	2935.16	0.25	601.35	1024.50	0.07
1999	5837.46	3056.37	0.21	717.55	1071.90	0.08
2000	5904.03	3132.08	0.23	762.82	1097.92	0.08

Source: GSOEP. Sample weights used.

Table 2: *Real Wages for Groups with and without Profit-Sharing (DM)*

Year	Without profit-sharing			With profit-sharing		
	Mean	Standard deviation	Coefficient of Variation	Mean	Standard deviation	Coefficient of Variation
1991	4486.32	1932.07	0.43	6733.26	2989.99	0.44
1992	4699.97	1965.52	0.42	6585.27	3063.83	0.47
1993	4857.77	2120.83	0.44	6708.97	3502.66	0.52
1994	4845.74	2076.34	0.43	7612.24	3456.62	0.45
1995	5005.30	2200.85	0.44	7429.71	3563.69	0.48
1996	5063.59	2201.65	0.44	7579.52	3695.67	0.49
1997	5250.36	2466.32	0.47	7123.95	3483.37	0.49
1998	5012.27	2190.90	0.44	7829.91	3766.17	0.48
1999	5254.80	2442.04	0.47	8079.34	4014.14	0.50
2000	5203.76	2350.62	0.45	8282.59	4139.68	0.50

Source: GSOEP. Sample weights used.

Table 3: *Estimation results*

Variable	Profit-Sharing Equation Estimate	Profit-Sharing Equation Std.Dev	Wage Equation Estimate	Wage Equation Std.Dev
Male	0.298	0.188	0.413**	0.080
German	1.041**	0.200	0.097**	0.027
Part-time	0.463*	0.202	-.261**	0.036
Tenure (\10)			-.019	0.022
Tenure^2 (\1000)			0.122	0.096
Experience (\10)	0.623^{+}	0.351	0.177**	0.056
Experience^2 (\1000)	-1.56*	0.661	-.295**	0.083
Years of schooling			2.345**	0.544
Amount Overtime (\100)			0.148**	0.021
Firm size: $20 \leq X < 200$	0.121	0.118	0.031^{+}	0.017
Firm size: $200 \leq X < 2000$	0.320*	0.151	0.052**	0.019
Firm size: $X \geq 2000$	0.665**	0.150	0.075**	0.023
Chemistry, mining	-.170	0.258	0.012	0.014
Construction	0.098	0.333	0.016	0.020
Finance, insurance, wholesale trade	0.126	0.245	-.011	0.014
Manufacturing	0.204	0.239	0.002	0.014
Transportation, warehousing	0.202	0.307	0.014	0.024
Male*Experience (\10)			-.004	0.028
Worker	-.032	0.118	-.061**	0.014
Constant	3.616**	0.603	7.515**	0.135
Covariance Matrix				
Parameter	Estimate	Std.Dev		
Correlation Coefficient (ζ)	0.295**	0.055		
ϑ_1	0.122**	0.037		
ϑ_2	1.021**	0.131		
σ_1^2	0.018**	0.006		
ϱ_1	0.867**	0.032		
ϱ_2	0.591**	0.072		
Wald Test	667.8**	(df=34)		

Source: GSOEP 1991-2000 (weighted estimation).
Significance Level: ** 0.01; * 0.05; $^{+}$ 0.10.
Pseudo R^2: $\hat{\rho}_2^2$=0.596; $NT = 7200$.

Table 4: *Variables in Regression Analysis*

Variable	Mean	Std.Dev
Log(real wage)	8.48	0.48
Profit-sharing	0.18	0.39
Male	0.70	0.46
German	0.89	0.31
Part-time	0.10	0.30
Tenure	13.36	9.15
Experience	24.53	9.59
Years of schooling	11.53	2.49
Worker	0.42	0.49
Amount overtime (hours)	11.33	16.80
Firm size: $20 \leq X < 200$	0.26	0.44
Firm size: $200 \leq X < 2000$	0.27	0.44
Firm size: $X \geq 2000$	0.31	0.46
Chemistry, mining	0.16	0.37
Construction	0.08	0.27
Finance, insurance, wholesale trade	0.23	0.42
Manufacturing	0.40	0.49
Transportation, warehousing	0.05	0.21

Source: GSOEP. Sample weights used.