

A comparison of a pseudo-likelihood estimation and full information maximum likelihood estimation for fitting multivariate longitudinal ordinal data

Vassilis Vasdekis and Silvia Cagnone and Irini Moustaki

Abstract We consider a latent variable model for multivariate ordinal responses accounting for dependencies among items. Time-dependent latent variables and random effects account for the inter-dependencies of the multivariate ordinal items. Model estimation is usually obtained using the full maximum likelihood via the EM algorithm. However, computationally problems can arise due to the calculation of multiple integrals involved in the likelihood. The paper proposes a pseudo-likelihood approach which involves only bivariate marginal probabilities. The proposed estimation method is evaluated by means of a little simulation study. A real data example illustrates the performance of both the full and the limited information estimation method.

Key words: latent variables, ordinal longitudinal data, pseudo-likelihoods, EM algorithm

1 Introduction

Latent variable models are widely used in Social Sciences for analyzing multivariate outcomes. Social science data are mainly of categorical nature. It is also common to collect data on different time points. When fitting a latent variable model to longitudinal data one needs to account for the interrelationships of the observed variables or indicators within each time point as well as the interrelationships of the same indicator across time in order to measure changes in the latent variables across time. In

Vassilis Vasdekis

Department of Statistics, Athens University of Economics and Business, e-mail: vasdekis@aueb.gr

Silvia Cagnone

Department of Statistics, University of Bologna, e-mail: silvia.cagnone@unibo.it

Irini Moustaki

Department of Statistics, London School of Economics e-mail: i.moustaki@lse.ac.uk

the classical normal linear factor model the total variance of an item is decomposed into two components: the common factor variance and the specific factor variance that confounds measurement error variance and residual variance. However, when measurements are made across time it is possible to decompose the measurement error from the residual error variance.

[9] discusses a generalized factor model for different type of responses that accounts for dependencies within time and across time using time specific factors and a heteroscedastic item-dependent random effect term. An autoregressive structure that allows for covariates is used to model the structural part of the model. The model is estimated by using MCMC. [7] propose a Full Information ML (FIML) approach implemented with the E-M algorithm for a special case of the family of models presented in [9] for ordinal variables. Time-dependent latent variables account for item dependencies within and across time where item-specific random effects account for item dependencies across time. The time-dependent latent variables have an autoregressive structure. The E-step of the algorithm is computationally intensive due to high-order integrations. The number of integrations depend on the number of items (random effects) and the number of latent variables fitted at each time point.

Fitting latent variable models to ordinal longitudinal data using full maximum likelihood estimation methods is unattractive due to the increase of computational complexity as the number of observed and latent variables increases. FIML estimation can only be feasible for small scale problems (no more than six latent variables and/or random effects). Therefore, alternative estimation methods that reduce the complexity of the estimation and produce estimators with desirable statistical properties are needed. Such complexity can be easily reduced by using pseudo-likelihood approaches. [14] divide the pseudo-likelihood approaches into two categories. The first category contains the ‘subsetting methods’ ([8]) which are based on likelihoods that use lower-order densities also known as composite marginal likelihoods ([12]) and the second category contains the ‘omission methods’ (see [5], [6], [3]) which omit parts of the likelihood that do not contain important information about the model parameters.

For the latent variable model with longitudinal ordinal responses, instead of maximizing the marginal likelihood of the whole response pattern, we propose a pairwise likelihood approach, that is we maximize products of bivariate likelihoods. Bivariate likelihoods contain much of the information about the model parameters. That will immediately reduces the number of fitted random effects. The pairwise likelihood approach is a special case of composite likelihoods approach and have been studied by different authors. Among others [4] applied this approach to the generalized linear models with crossed random effects; [13] considered spatial generalized linear mixed models and used an EM algorithm for estimating the parameters. [10] refer to the Rasch model for longitudinal data. [14] also introduced a model selection criterion based on composite likelihoods.

In this paper we use the ideas on composite marginal likelihoods suggested by [8] and [16]. We develop an EM algorithm for parameter estimation. The paper is organized as follows: Section 2 discusses the methodology for multivariate longitudinal ordinal responses; Section 3 describes the composite likelihood estimation.

In section 4 we report a small simulation study to evaluate the performance of the estimation method proposed. Finally in Section 5 we present a real data analysis where both FIML and the pairwise likelihood approach are used.

2 Proposed methodology

Suppose that p items are measured for n individuals for T time points providing measurements y_{it} for a randomly selected individual. Let $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_t, \dots, \mathbf{y}'_T)'$ represent the whole response pattern where each component of that vector is also a $p \times 1$ vector $\mathbf{y}_t = (y_{1t}, \dots, y_{pt})'$. We assume that the same ordinal items are measured across time. The m_i ordered categories of variable i measured at time t have probabilities $\pi_{it(1)}(\mathbf{z}, u_i), \pi_{it(2)}(\mathbf{z}, u_i), \dots, \pi_{it(m_i)}(\mathbf{z}, u_i)$, which are functions of the time-dependent latent variables $\mathbf{z} = (z_1, \dots, z_T)'$ and the item-specific random effects u_i . At the moment we assume that items within each time point are uni-dimensional for keeping the notation simpler. Models with multidimensional items can also be estimated using composite likelihoods.

The random component of the model is such that the pT random response variables, y_{it} , conditional on the latent variables z_t and the random effects u_i , are independent observations with a distribution from the exponential family. The systematic component for the i, t combination is one in which the latent variable z_t , and the random effect term u_i produce a linear predictor $\eta_{it}^{(s)}$ corresponding to each of the m_i categories of y_{it} :

$$\eta_{it}^{(s)} = \tau_{it}^{(s)} - \alpha_{it} z_t - u_i, \quad i = 1, \dots, p; \quad s = 1, \dots, m_i; \quad t = 1, \dots, T. \quad (1)$$

where $\tau_{it}^{(s)}$ are item, time, category specific fixed intercepts and α_{it} are factor loadings different for each item and time. Finally, the link between the systematic components and the conditional means of the random component distributions is $\eta_{it}^{(s)} = v_{it}^{(s)}(\gamma_{it}^{(s)})$ where $\gamma_{it}^{(s)} = P(y_{it} \leq s \mid z_t, u_i)$ with $v_{it}^{(s)}(\cdot)$ any monotonic differentiable function.

Defining the linear predictor in this way, the associations among the items measured at time t are explained by the variance of the latent variable z_t . The associations among the same item measured across time are explained by both the item-specific random effect u_i and the covariances between z_t 's. A reasonable structure for these covariances is necessary for the adequate modelling of these associations.

In particular, we assume that the time dependent latent variables are linked with a first order autoregressive structure

$$z_t = \phi z_{t-1} + \delta_t \quad (2)$$

where for identification purposes $\delta_t \sim N(0, 1)$ and $z_1 \sim N(0, \sigma_1^2)$. We also assume that the random effects u_i are independent of z_t and their common distribution function is $N_p(\mathbf{0}, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = \text{diag}_{i=1, \dots, p}(\sigma_{ui}^2)$. Equation (2) accounts for serial correla-

tion in the latent variables and expresses the dynamic nature of our latent constructs. As a consequence, the covariances between cumulative probabilities of response seeing under the scale provided by link v , are given by

$$Cov(\eta_{it}, \eta_{i't'}) = \alpha_{it} \alpha_{i't'} Var(z_t), \quad i \neq i' \quad (3)$$

$$Cov(\eta_{it}, \eta_{i't'}) = \alpha_{it} \alpha_{i't'} Cov(z_t, z_{t'}) + \sigma_{ui}^2, \quad t < t' \quad (4)$$

$$Cov(\eta_{it}, \eta_{i't'}) = \alpha_{it} \alpha_{i't'} Cov(z_t, z_{t'}), \quad i \neq i', \quad t < t' \quad (5)$$

with $Var(z_t) = \phi^{2(t-1)} \sigma_1^2 + I(t \geq 2) \sum_{k=1}^{t-1} \phi^{2(k-1)}$ and $Cov(z_t, z_{t'}) = \phi^{t+t'-2} \sigma_1^2 + I(t \geq 2) \sum_{k=0}^{t'-t+2k} \phi^{t'-t+2k}$, where $I(\cdot)$ is the indicator function.

The model applied to the ordinal items is the *proportional odds model*:

$$\log \left[\frac{\gamma_{it}^{(s)}(z_t, u_i)}{1 - \gamma_{it}^{(s)}(z_t, u_i)} \right] = \tau_{it}^{(s)} - \alpha_{it} z_t - u_i, \quad s = 1, \dots, m_i - 1, \quad t = 1, \dots, T \quad (6)$$

The π 's are obtained from the γ 's by the relationships $\pi_{it}^{(1)}(z_t, u_i) = \gamma_{it}^{(1)}(z_t, u_i)$ and $\pi_{it}^{(s)}(z_t, u_i) = \gamma_{it}^{(s)}(z_t, u_i) - \gamma_{it}^{(s-1)}(z_t, u_i)$, $s = 2, 3, \dots, m_i$.

Let $\mathbf{y}_r = (y_{11} = a_{11}, y_{21} = a_{21}, \dots, y_{pT} = a_{pT})$ represent a full response pattern for $r = 1, \dots, (\prod_{i=1}^p m_i)^T$. The vectors of the time dependent latent variables and the item random effects will be represented by $\mathbf{z} = (z_1, z_2, \dots, z_T)$ and by $\mathbf{u} = (u_1, u_2, \dots, u_p)$ respectively. We also denote by $\boldsymbol{\tau} = (\{\tau_{it}^{(s)}\}, i = 1, \dots, p, t = 1, \dots, T, s = 1, \dots, m_i)$, the vector of unknown fixed intercepts and by $\boldsymbol{\theta} = (\phi, \sigma_1^2, \sigma_{u1}^2, \dots, \sigma_{up}^2)$ the vector of variance and covariance parameters. Under the assumption of conditional independence, the conditional probability, for given $\mathbf{z}$ and $\mathbf{u}$, of the response pattern $\mathbf{y}_r$ is

$$\pi_r(\mathbf{z}, \mathbf{u}; \boldsymbol{\tau}) = \prod_{i=1}^p \prod_{t=1}^T \pi_{it}^{a_{it}}(z_t, u_i) = \prod_{i=1}^p \prod_{t=1}^T [\gamma_{it}^{a_{it}}(z_t, u_i) - \gamma_{it}^{(a_{it}-1)}(z_t, u_i)] .$$

The unconditional probability π_r of the response pattern $\mathbf{y}_r$ is obtained by integrating the joint density of $(\mathbf{y}_r, \mathbf{z}, \mathbf{u})$, over the $(T + p)$ -dimensional factor space:

$$\pi_r(\boldsymbol{\tau}, \boldsymbol{\theta}) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \pi_r(\mathbf{z}, \mathbf{u}; \boldsymbol{\tau}) h(\mathbf{z}, \mathbf{u}; \boldsymbol{\theta}) d\mathbf{z} d\mathbf{u} ,$$

where $h(\mathbf{z}, \mathbf{u}; \boldsymbol{\theta})$ is the common density function of $\mathbf{z}, \mathbf{u}$. The latent variables $\mathbf{z}$ are dependent through the autoregressive structure but they are independent from the random effects $\mathbf{u}$ according to the model specification given above (2). A convenient assumption is that h is multivariate normal with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Phi}$

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega} \end{bmatrix} \quad (7)$$

where $\mathbf{\Omega} = \text{diag}_{i=1,\dots,p}\{\sigma_{ii}^2\}$ and the inverse of $\mathbf{\Gamma}$ has a well known special pattern as

$$\mathbf{\Gamma}^{-1} = \begin{bmatrix} \frac{1}{\sigma_{12}^2} + \phi^2 & -\phi & 0 & \dots & 0 & 0 & 0 \\ -\phi & 1 + \phi^2 & -\phi & \dots & 0 & 0 & 0 \\ 0 & -\phi & 1 + \phi^2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\phi & 1 + \phi^2 & -\phi \\ 0 & 0 & 0 & \dots & 0 & -\phi & 1 \end{bmatrix}$$

Therefore, for a random sample of size n , the log-likelihood is written as:

$$L = \sum_{l=1}^n \log \pi_l(\boldsymbol{\tau}, \boldsymbol{\theta}) \quad (8)$$

The maximization of the marginal log-likelihood given in (8) is achieved via an E-M algorithm. At this regard a FIML has been developed by [7]. The marginal distribution of the data requires integration over the sample space of all latent variables and random effects but for a large number of items and time-points FIML becomes intractable. Therefore, we propose here a pseudo-likelihood estimation that depends only on the bivariate marginal distributions.

3 Composite likelihood estimation

The pseudo likelihood is obtained by replacing the joint likelihood of the data $\mathbf{y}$ by a suitable product of conditional or marginal densities ([5], [12], [11]). Here, we use the bivariate marginal distributions. In order to identify all possible marginal pairwise probabilities, we shall consider the sets of indices $S_1 = \{(i, j, t); 1 \leq i < j \leq p; t = 1, \dots, T\}$ which defines all pairs of observations from different items at the same time, and $S_2 = \{(i, i', t, t'); 1 \leq i \leq p; 1 \leq i' \leq p; 1 \leq t < t' \leq T\}$ which defines pairs of observations from either different or same items but at different time points. We denote by $\pi_{it, i't', z_t, z_{t'}, u_i, u_{i'}}^{a_{it}, a_{i't'}}(\boldsymbol{\tau}, \boldsymbol{\theta})$ the joint density of $(y_{it}, y_{i't'}, z_t, z_{t'}, u_i, u_{i'})$ and by $\pi_{it, i't'}^{a_{it}, a_{i't'}}(\boldsymbol{\tau}, \boldsymbol{\theta}) = P(y_{it} = a_{it}, y_{i't'} = a_{i't'}; \boldsymbol{\tau}, \boldsymbol{\theta})$ the marginal probability density of any pair of observations. In the following, we shall distinguish between the $Tp(p-1)/2$ pairs of items within the same time and the $p^2T(T-1)/2$ pairs of items across time, points mainly for computational reasons. The bivariate probability for a pair of response $(y_{it} = a_{it}, y_{i't} = a_{i't})$ is:

$$\pi_{it, i't}^{a_{it}, a_{i't}}(\boldsymbol{\tau}, \boldsymbol{\theta}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \pi_{it}^{a_{it}}(z_t, u_i) \pi_{i't}^{a_{i't}}(z_t, u_{i'}) h(z_t, u_i, u_{i'}) dz_t du_i du_{i'} \quad (9)$$

and for a pair of response $(y_{it} = a_{it}, y_{i't'} = a_{i't'})$ is:

$$\pi_{it,i't'}^{a_{it},a_{i't'}}(\boldsymbol{\tau}, \boldsymbol{\theta}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \pi_{it}^{a_{it}}(z_t, u_i) \pi_{i't'}^{a_{i't'}}(z_{t'}, u_{i'}) h(z_t, z_{t'}, u_i, u_{i'}) dz_t dz_{t'} du_i du_{i'} \quad (10)$$

The latter can be three dimensional if it happens that $i = i'$. In any case, the maximum number of integrations we have to deal with, is four.

The contribution of any given individual to the log pairwise likelihood is therefore:

$$pl(\boldsymbol{\tau}, \boldsymbol{\theta}; \mathbf{y}) = \left[\sum_{S_1} \log \pi_{it,i't'}^{a_{it},a_{i't'}}(\boldsymbol{\tau}, \boldsymbol{\theta}) + \sum_{S_2} \log \pi_{it,i't'}^{a_{it},a_{i't'}}(\boldsymbol{\tau}, \boldsymbol{\theta}) \right] \quad (11)$$

Since each component of equation (11) is a likelihood object we can claim that $\nabla pl(\boldsymbol{\theta}) = 0$ is unbiased under regularity conditions. For the pairwise likelihood defined in (11), the maximum pairwise likelihood estimator $\hat{\boldsymbol{\theta}}_{PL} = (\hat{\boldsymbol{\tau}}_{PL}, \hat{\boldsymbol{\psi}}_{PL})$ converges in probability to the true parameter value $\boldsymbol{\theta}_0$ and

$$\hat{\boldsymbol{\theta}}_{PL} \rightarrow N_q(\boldsymbol{\theta}_0, J^{-1}(\boldsymbol{\theta}_0)H(\boldsymbol{\theta}_0)J^{-1}(\boldsymbol{\theta}_0))$$

where $J(\boldsymbol{\theta}) = E_y \{ \nabla^2 pl(\boldsymbol{\theta}; \mathbf{y}) \}$ and $H(\boldsymbol{\theta}) = \text{var}_y(\nabla pl(\boldsymbol{\theta}; \mathbf{y}))$ ([2], [8]).

The maximization of the likelihood in (11) is done with the E-M algorithm. The E-step of the algorithm requires the expected score function of the joint log-likelihood of the data, the latent variables and the random effects. In the M step we get not-explicit solutions for all the parameters of the model. A Newton Raphson algorithm is used to solve non linear maximum likelihood equations. Since the integrals cannot be solved analytically they are approximated using the Gauss Hermite quadrature points.

The asymptotic covariance matrix of the pairwise likelihood (PL) estimators assumes a typical sandwich form and it is given by the inverse of the pairwise information

$$I(\boldsymbol{\tau}, \boldsymbol{\theta}) = J(\boldsymbol{\theta})H^{-1}(\boldsymbol{\theta})J(\boldsymbol{\theta})$$

A convenient estimator of $H(\boldsymbol{\theta})$, which does not require the evaluation of second-order derivatives, is given by

$$\hat{H} = \sum_{l=1}^n \nabla pl_l(\hat{\boldsymbol{\tau}}_{PL}, \hat{\boldsymbol{\theta}}_{PL}; \mathbf{y}) \nabla pl_l(\hat{\boldsymbol{\tau}}_{PL}, \hat{\boldsymbol{\theta}}_{PL}; \mathbf{y})'$$

The empirical estimate of $J(\boldsymbol{\theta})$ is

$$\hat{J} = - \sum_{l=1}^n \left[\sum_{S_{l,1}} \nabla \log \pi_{it,i't'}^{(a_{it},a_{i't'})}(z_t, z_{t'}, u_i, u_{i'}) + \sum_{S_{l,2}} \nabla \log \pi_{it,i't'}^{(a_{it},a_{i't'})}(z_t, z_{t'}, u_i, u_{i'}) \right]$$

Model selection criteria have been derived to be used with composite likelihood estimation methods. [14] derived the following predictive model criterion that is a generalization of the AIC criterion ([1]):

$$CLIC = -2(pl(\hat{\boldsymbol{\tau}}_{PL}, \hat{\boldsymbol{\theta}}_{PL}; \mathbf{y}) - \text{tr}(\hat{H}\hat{J}^{-1})) \quad (12)$$

4 Simulation study

In this section, we evaluate the performance of the bivariate likelihood estimation for our model by means of a small simulation study. The design of study includes number of items $p = 3$, number of time points $T = 4$ and $\phi = 0.4$. The number of quadrature points in the Gauss Hermite approximation are fixed to 8. Different sample sizes are considered ($n = 200, 500, 1000$). Notice that a FIML approach is computationally unfeasible under these conditions, even using 5 quadrature points.

In Table 1 we report the true values used to simulate the data, the mean values of the estimated parameters across simulations together with their corresponding variances and mean square error (MSE). The loadings were constrained to be equal across time points. Although the bias is similar for all the three cases, the MSE reduces for all parameters as the sample size increases from 200 to 1000. The degree of bias is similar to the bias one will find if FIML estimation is used since even in such case an approximate solution is obtained with the Gauss-Hermite approximation of the integrals.

Table 1 Simulation results for $p = 3$, $T = 4$, $\phi = 0.4$, quad=8, 500 replications

Parameters	True	$n=200$			$n=500$			$n=1000$		
		Mean	Var	MSE	Mean	Var	MSE	Mean	Var	MSE
α_1	1.00	1.00			1.00			1.00		
α_2	1.20	1.36	0.08	0.11	1.41	0.03	0.08	1.43	0.01	0.07
α_3	0.66	0.70	0.04	0.04	0.70	0.01	0.02	0.70	0.01	0.01
ϕ	0.40	0.37	0.01	0.01	0.37	0.00	0.01	0.36	0.00	0.00
σ_{u1}^2	1.00	1.18	0.08	0.11	1.17	0.04	0.07	1.19	0.02	0.05
σ_{u2}^2	1.50	1.71	0.18	0.22	1.77	0.06	0.14	1.79	0.03	0.11
σ_{u3}^2	2.50	2.58	0.31	0.31	2.56	0.11	0.12	2.55	0.04	0.05
σ^2	2.00	1.95	0.42	0.42	1.80	0.12	0.16	1.78	0.06	0.11

5 British Household Panel Survey

The data set is extracted from the British Household Panel Survey (BHPS). Three waves (1994, 1996 and 1998) and three ordinal items are analyzed. The three items measure social and political attitudes. The original sample size was 3784. Since we consider several specifications of the latent variable model in our analyses, we reduce computational time by analysing a random subsample of 500 individuals. The items are:

1. Private enterprise solves economic problems [Enterp]

2. Government has obligation to provide jobs [Govern]
3. Strong trade unions protect employees [TrUnion]

Permitted responses to these questions were *agree strongly (AS)*, *agree (A)*, *not agree/disagree (Neither A nor D)*, *disagree (D)*, *disagree strongly (DS)*. Due to the small proportion of people that responded into the first and the last categories of the items, the first two categories and the last two categories have been collapsed as to obtain with three categories for each item. Finally, the response categories of item ‘Enterp’ have been reversed.

The aim of the analysis is to estimate changes that have occurred in the variance of the attitudinal latent variable from 1994 to 1998. We started the analysis by fitting an one-factor model at each time point separately. The one-factor model fitted very well in all three instances. That implied that the three items being analyzed were uni-dimensional in all three occasions. To be able to estimate differences in the latent variable across time one needs to establish that the latent variable is on the same metric scale for those three different occasions (origin and scale needs to be the same). To achieve the common metric for the latent variable we need to constrain the thresholds and the factor loadings to be equal across time for the same item.

We started the analysis by allowing all model item-parameters associated with the measurement model (thresholds and slopes) to vary across time points (Model 1). We then proceeded fitting models with only equal thresholds (Model 2), only equal loadings (Model 3) and finally equal thresholds and equal loadings (Model 4). Model 4 sets the same metric (origin and scale) to the latent variable across the five time points and allows to estimate differences in the latent variable across time (strictly measurement invariance).

We estimated the model using both the FIML approach discussed by [7] and the pairwise approach presented above. Notice that the former approach is computationally feasible only with 5 quadrature points whereas with the latter we used 8 quadrature points.

The four models with different constraints were compared using the AIC criterion in the FIML approach and the CLIC criterion described above in the pairwise approach (Table 2). We can observe that in both cases Model 3 is selected.

Table 2 Model selection criteria

Models	FIML	Pairwise
	AIC	CLIC
Model 1	8123.694	73114.35
Model 2	8154.829	73322.34
Model 3	8123.385	73113.41
Model 4	8153.579	73321.95

Table 3 gives the estimated model parameters for Model 3 with the two estimation methods. All the loadings are positive and close to each other (similar discrimination

power). They are similar between the two approaches, even if slightly higher with the pairwise approach. The large estimated variances for the random effects indicate large homogeneity in the responses within each item across time. Homogeneity that obviously could not all be accounted for by the autoregressive model fitted on the time-dependent latent variables.

Table 3 Estimated factor loadings with standard errors in brackets for the non-stationary model with time specific latent variables and estimated variances for the item-specific random effects, FIML and Pairwise approach, BHPS

Items	FIML		Pairwise	
	$\hat{\alpha}_i$	$\hat{\sigma}_{ui}^2$	$\hat{\alpha}_i$	$\hat{\sigma}_{ui}^2$
Enterp	1.00	1.71 (0.73)	1.00	2.46(0.52)
Govern	0.86 (0.12)	5.09 (3.41)	1.05(0.09)	5.54(1.98)
TrUnion	1.01 (0.18)	5.45 (3.91)	1.28(0.15)	5.63(1.43)

The estimated covariance matrices of the time-dependent attitudinal latent variables are close with the two estimation methods:

$$\hat{\mathbf{\Gamma}}_{FIML} = \begin{bmatrix} 4.22 & 3.99 & 3.78 \\ 3.99 & 4.78 & 4.52 \\ 3.78 & 4.52 & 5.28 \end{bmatrix} \quad \hat{\mathbf{\Gamma}}_{Pair} = \begin{bmatrix} 4.25 & 3.64 & 3.11 \\ 3.64 & 4.11 & 3.52 \\ 3.11 & 3.52 & 4.01 \end{bmatrix}$$

There is also a strong correlation between the time-dependent factor from one occasion to the other. Indeed we obtained a large estimated autoregressive parameter with both methods. With FIML it is $\hat{\phi} = 0.95$ and estimated standard error equal to 0.01, with the pairwise approach it is $\hat{\phi} = 0.86$ and estimated standard error equal to 0.04. The estimated variances of the time-dependent latent variable do not change very much across the three waves. This implies that variability of opinions measured by the identified factor remains similar across time.

References

1. Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Petrov, B.B., Caski, F. (eds.) Proc. Second International Symposium on Information Theory, 267–281 (1973).
2. Arnold, D.A., Strauss, D. Pseudolikelihood estimation: some examples. *sankhya Ser. B* **53**, 233–243 (1991).
3. Azzalini, A.: Maximil likelihood of order m for stationary stochastic pocesses. *Biometrika*, **70**, 381–387 (1983).
4. Bellio R. Varin C.: A pairwise likelihood approach to generalized linear models with crossed random effects, *Statistical modelling*, **5**, 217–227. (2005)
5. Besag, J.E.: Spatial interaction and the statistical analysis of lattice system (with discussion). *Jornal of the royal statistical society B*, **34**, 192-236 (1974).

6. Cox, D.R.: Partial likelihood. *Biometrika* **62**, 269–276 (1975).
7. Cagnone, S., Moustaki, I., Vasdekis, V.: Latent variable models for multivariate longitudinal ordinal responses. *British Journal of Mathematical and Statistical Psychology*, **62**, 401–415 (2009).
8. Cox, D. R., Reid, N.: A note on pseudolikelihood constructed from marginal densities. *Biometrika*, **91**, 729–737 (2004).
9. Dunson, D.: Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association*, **463**, 555–563 (2003).
10. Feddag, M.L., Bacci, S.: Pairwise likelihood for the longitudinal mixed Rasch model. *Computational statistics and data analysis*, **53**, 1027–1037 (2009).
11. Jöreskog, K. G. and Moustaki, I.: Factor analysis of ordinal variables: A comparison of three approaches, *Multivariate Behavioral Research*, **36**, 347–387 (2001)
12. Lindsay, B.G.: Composite likelihood methods. *Contemporary Mathematics*, **80**, 221–239 (1988).
13. Varin C., Gudmund H., Øivind S.: Pairwise likelihood inference in spatial generalized linear mixed models. *Computational Statistics and Data analysis*, **49**, 1173–1191 (2005).
14. Varin, C., Vidoni, P.: A note on composite likelihood inference and model selection. *Biometrika*, **92**, 519–528 (2005).
15. Varin, C., Vidoni, P.: Pairwise Likelihood Inference for Ordinal Categorical Time Series., *Computational statistics and data analysis*, **51**, 2365–2373. (2006)
16. Varin, C.: On composite marginal likelihoods. *AstA* **92**, 1–28 (2008) doi: 10.1007/s10182-008-0060-7.