



UPPSALA  
UNIVERSITET

Working Paper 2011:4

Department of Statistics

# Pairwise Likelihood Estimation for factor analysis models with ordinal data

Myrsini Katsikatsou

Irini Moustaki

Fan Yang-Wallentin

Karl G. Jöreskog



Working Paper 2011:3  
November 2011  
Department of Statistics  
Uppsala University  
Box 513  
SE-751 20 UPPSALA  
SWEDEN

Working papers can be downloaded from [www.statistics.uu.se](http://www.statistics.uu.se)

Title: Pairwise Likelihood Estimation for factor analysis models with ordinal data.  
Authors: Myrsini Katsikatsou, Irini Moustaki, Fan Yang-Wallentin, Karl G.Jöreskog  
E-mail: myrsini.katsikatsou@statistics.uu.se



# Pairwise Likelihood Estimation for factor analysis models with ordinal data

Myrsini Katsikatsou<sup>1</sup>      Irimi Moustaki<sup>2</sup>      Fan Yang-Wallentin<sup>1</sup>  
Karl G. Jöreskog<sup>1,3</sup>

<sup>1</sup>Department of Statistics, University of Uppsala, Sweden

<sup>2</sup>Department of Statistics, London School of Economics, United Kingdom

<sup>3</sup>Department of Economics, Norwegian School of Management, Oslo, Norway

## Abstract

Pairwise maximum likelihood (PML) estimation is developed for factor analysis models with ordinal data fitted both in an exploratory and confirmatory set-up, and its performance is studied and compared with full information maximum likelihood (FIML) and a three-stage limited information estimation method. More specifically, estimates and standard errors obtained from PML are compared with those obtained from FIML and those from robust unweighted least squares (3S-RULS). All three methods provide very close estimates and standard errors. However, the PML estimates and standard errors are on average slightly closer to FIML than the 3S-RULS are. The advantage of PML over FIML is mainly computational. The computational complexity of FIML increases with the number of factors or observed variables depending on the model formulation, while that of PML is affected by neither of them. Contrary to 3S-RULS, in PML, all model parameters are simultaneously estimated and therefore the final estimates reflect all the sampling variability. In the 3S-RULS method the standard errors of the parameter estimates in stage three do not incorporate the variability of the estimates obtained in step one. Furthermore, PML does not require the estimation of a weight matrix for computing correct standard errors. The performance of PML estimates and their estimated asymptotic standard errors are investigated through a simulation study where the effect of different models and sample sizes are studied. The bias and mean squared error of PML estimators and their standard errors are found to be small in all experimental conditions and decreasing with the sample size.

**Keywords:** composite maximum likelihood; factor analysis; ordinal data; maximum likelihood; 3-stage estimation; item response theory approach.

## 1 Introduction

Factor analysis is frequently employed in social sciences where the main interest lies in measuring and relating unobserved constructs such as emotions, attitudes, beliefs and behaviour. The main idea of the analysis is that the latent variables, referred to also as factors, account for the dependencies among the observed variables, referred to also as items or indicators, in the sense that

if the factors are held fixed, the observed variables would be independent. Theoretically, factor analysis can be distinguished between exploratory and confirmatory analysis, but practice most of the time lies between the two. In exploratory factor analysis the goal is the following: for a given set of observed variables  $x_1, \dots, x_p$  one wants to find a set of latent factors  $\xi_1, \dots, \xi_k$ , fewer in number than the observed variables ( $k < p$ ), that contain essentially the same information. In confirmatory factor analysis, the objective is to verify a social theory. Hence, a factor model is specified in advance and its fit to the empirical data is tested.

The data usually encountered in social sciences is of categorical nature (ordinal or nominal). In the literature, there are two main approaches for analysing ordinal variables with factor models. The Underlying Response Variable (URV) approach (e.g. Jöreskog 1990, 1994; Lee et al. 1990, 1992; Muthén, 1984), and the Item Response Theory (IRT) approach (e.g. Bartholomew et al. 2011; Muraki, 1990; Muraki & Carlson, 1995; Samejima, 1969). In the URV approach, the ordinal variables are generated by underlying continuous variables partially observed through their ordinal counterparts. In the IRT approach, ordinal indicators are treated as they are. In both approaches, one must specify the probability of each response pattern as a function of  $\xi_1, \xi_2, \dots, \xi_k$ :

$$Pr(x_1 = c_1, x_2 = c_2, \dots, x_p = c_p \mid \xi_1, \dots, \xi_k) = f(\xi_1, \xi_2, \dots, \xi_k), \quad (1)$$

where  $c_1, c_2, \dots, c_p$  represent the different response categories of  $x_1, x_2, \dots, x_p$ , respectively.

In this study we consider these two different ways of specifying the model for ordinal variables, namely the URV and IRT approach, and three estimation methods, the full information maximum likelihood<sup>1</sup> (FIML), the composite pairwise maximum likelihood (PML), and the three-stage limited information robust unweighted least squares (3S-RULS). We mainly focus on the computational issues arising at each estimation method within each approach which seriously affect their applicability. More specifically, FIML involves high-dimensional integrations under the URV approach as the number of items increases. When the number of items is greater than five ( $p > 5$ ), FIML is only feasible when the IRT framework is used. However, even in IRT, FIML becomes very computationally heavy as the number of factors increases. To tackle the practical restrictions of FIML in the case of URV approach limited information estimation methods have been developed such as the three-stage estimation methods: robust unweighted least squares (3S-RULS), robust diagonally weighted least squares (3S-RDWLS), and weighted least squares (3S-WLS) (see e.g. Yang-Wallentin et. al., 2010). The restricted applicability of FIML in both URV and IRT approaches along with the theoretical developments of PML estimation (Lindsay, 1988) motivated us to consider PML as an alternative estimation method. PML estimators have the desired properties of being asymptotically unbiased, consistent, and normally distributed (Varin, 2008; Varin et. al., 2011). PML can be applied to any of the two aforementioned model formulations although the computational gain is only when the URV approach is used; its computational complexity can be kept low regardless of the number of observed variables or factors. Based on the results of Jöreskog & Moustaki (2001), de Leon (2005), and Liu (2007), we propose PML estimation within the URV approach. In particular, Jöreskog & Moustaki (2001) suggest the *underlying bivariate normal* (UBN) method within the URV approach which is found to yield estimates close to FIML ones. Although it has not been presented as such, UBN can be seen as a pseudo maximum likelihood method involving the univariate and bivariate marginal distributions. De Leon (2005) suggests using only the bivariate marginal distributions, to estimate thresholds and polychoric correlations of grouped continuous data. His simulation study indicates that PML estimates are quite accurate, yielding minimal bias and small root mean squared errors. Liu (2007) proposes a new multistage

---

<sup>1</sup>The term “full” is used here to emphasize that we refer to the joint likelihood of all observed variables.

estimation method for factor models and structural equation models, alternative to the commonly used three-stage methods. In particular, in a first stage, thresholds, polychoric, and polyserial correlations are estimated simultaneously by using PML, while structural parameters, like loadings and factor correlations are estimated in a second stage given the estimates of the first stage using generalised least squares. The simulation studies Liu (2007) conducted show that the proposed methodology performs equally well and is slightly more robust than the conventional three-stage methods.

The structure of the paper is as follows: Section 2 provides a brief presentation of the URV and IRT approaches with a focus on the estimation. We discuss the computational issues arising in the case of FIML and the advantages and disadvantages of the three-stage limited information estimators. Section 3 provides the definition of composite maximum likelihood and its properties. A discussion of the UBN method suggested by Jöreskog & Moustaki (2001) which is closely related to our proposed methodology is also included. Section 4 presents the proposed methodology, namely the composite pairwise maximum likelihood estimation (PML), and Section 5 demonstrates it with some examples of real data both in the case of exploratory and confirmatory factor analysis. Furthermore, PML estimates and their estimated asymptotic standard errors are compared to the FIML and 3S-RULS ones. Finally, Section 6 reports the results of our simulation study which aims to investigate the effect of model and sample size on the performance of PML estimator and its estimated asymptotic standard error as measured in terms of bias and mean squared error. Discussion and conclusions are provided in the end.

## 2 Factor analysis models with ordinal observed variables

### 2.1 Basic framework and notation

Let  $\mathbf{x}' = (x_1, x_2, \dots, x_p)$  denote the vector of  $p$  ordinal observed variables<sup>2</sup>, where  $x_i$  has  $m_i$  ordered categories,  $i = 1, \dots, p$ . Thus, there are  $R = \prod_{i=1}^p m_i$  possible response patterns of the form  $\mathbf{x}_r = (c_1, c_2, \dots, c_p)$ , where  $c_i = 1, \dots, m_i$ . For a random sample of size  $n$  the log-likelihood is:

$$\ln L(\boldsymbol{\theta}; \mathbf{x}) = \sum_{r=1}^R n_r \ln \pi_r(\boldsymbol{\theta}), \quad (2)$$

where  $\boldsymbol{\theta}$  is a parameter vector,  $n_r$  and  $\pi_r(\boldsymbol{\theta})$  are the observed frequency and the probability under the model respectively for the response pattern  $r$ ,  $\pi_r(\boldsymbol{\theta}) > 0$ ,  $\sum_{r=1}^R n_r = n$ , and  $\sum_{r=1}^R \pi_r(\boldsymbol{\theta}) = 1$ . Each approach imposes a different model on the probability  $\pi_r(\boldsymbol{\theta})$  but both URV and IRT methods assume the presence of a  $k$ -dimensional vector of continuous latent variables  $\boldsymbol{\xi}' = (\xi_1, \dots, \xi_k)$ , where  $k < p$ .

### 2.2 Underlying Response Variable (URV) approach

Under the URV approach, the observed ordinal variables are taken to be manifestations of some underlying continuous variables partially observed through their ordinal counterparts. The connection between an observed ordinal variable  $x_i$  and the underlying continuous variable  $x_i^*$  is

$$x_i = c_i \iff \tau_{c_i-1}^{(x_i)} < x_i^* < \tau_{c_i}^{(x_i)} \quad (3)$$

---

<sup>2</sup>Small letters are used to denote both the variables and their values.

where  $\tau_{c_i}^{(x_i)}$  is the  $c_i^{\text{th}}$  threshold of variable  $x_i$  and  $-\infty = \tau_0^{(x_i)} < \tau_1^{(x_i)} < \dots < \tau_{m_i-1}^{(x_i)} < \tau_{m_i}^{(x_i)} = +\infty$ . Since only ordinal information is available, the distribution of  $x_i^*$  is determined only up to a monotonic transformation. As Jöreskog (2002) points out, one can choose any continuous distribution for  $x_i^*$ , but any continuous variable can be transformed by a monotonic transformation to a normally distributed variable. Thus, it is convenient to choose the standard normal distribution for each  $x_i^*$ . In the case that the mean and the variance of  $x_i^*$  are of interest, Jöreskog (2002) discusses an alternative parametrization<sup>3</sup>.

The factor model is of the form

$$\mathbf{x}^* = \Lambda \boldsymbol{\xi} + \boldsymbol{\delta}, \quad (4)$$

where  $\mathbf{x}^*$  is the  $p$ -dimensional vector of the underlying variables,  $\Lambda$  is the  $p \times k$  matrix of loadings, and  $\boldsymbol{\delta}$  is the  $p$ -dimensional vector of unique variables. The matrix  $\Lambda$  is determined only up to an orthogonal transformation. Furthermore, it is assumed that  $\boldsymbol{\xi} \sim N_k(\mathbf{0}, \Phi)$  where  $\Phi$  has ones on its main diagonal and is the correlation matrix of latent factors<sup>4</sup>. In addition, it is assumed that  $\boldsymbol{\delta} \sim N_p(\mathbf{0}, \Theta)$  with  $\Theta$  a diagonal matrix. In particular,  $\Theta$  follows the structure  $\Theta = I - \text{diag}(\Lambda \Phi \Lambda')$ , which is a direct consequence of fixing the variance of  $x_i^*$ 's equal to 1. Finally,  $\text{Cov}(\boldsymbol{\xi}, \boldsymbol{\delta}) = \mathbf{0}$ . The parameter vector  $\boldsymbol{\theta}' = (\boldsymbol{\lambda}, \boldsymbol{\varphi}, \boldsymbol{\tau})$  contains  $\boldsymbol{\lambda}$  and  $\boldsymbol{\varphi}$  which are the free non-redundant parameters in matrices  $\Lambda$  and  $\Phi$ , respectively, and  $\boldsymbol{\tau}$  is a vector of all free thresholds.

The model along with the aforementioned assumptions implies that the correlation matrix of the underlying variables vector  $\mathbf{x}^*$  follows the structure

$$\Sigma_{\mathbf{x}^*} = \Lambda \Phi \Lambda' + I - \text{diag}(\Lambda \Phi \Lambda')$$

and the expected under the model probability of a response pattern  $r$  is

$$\pi_r(\boldsymbol{\theta}) = \pi(x_1 = c_1, x_2 = c_2, \dots, x_p = c_p; \boldsymbol{\theta}) = \underbrace{\int_{\tau_{c_1-1}^{(x_1)}}^{\tau_{c_1}^{(x_1)}} \dots \int_{\tau_{c_p-1}^{(x_p)}}^{\tau_{c_p}^{(x_p)}}}_{p\text{-dimensional}} \phi_p(\mathbf{x}^*; \Sigma_{\mathbf{x}^*}) d\mathbf{x}, \quad (5)$$

where  $\phi_p(\mathbf{x}^*; \Sigma_{\mathbf{x}^*})$  is a  $p$ -dimensional normal density with zero mean and correlation matrix equal to  $\Sigma_{\mathbf{x}^*}$ .

The maximization of log-likelihood defined in (2) over the parameter  $\boldsymbol{\theta}$  requires the evaluation of the  $p$ -dimensional integral given in (5) which cannot be written in a closed form. A version of the algorithm proposed by Schervish (1984) can be used but the computational time increases rapidly with the number of observed variables  $p$ , rendering FIML estimation impractical for general factor analysis problems. Lee et al. (1990) discuss FIML estimation in the case of URV approach but restrict their example to the case of four ordinal observed variables. As a consequence, limited

---

<sup>3</sup>When the number of response categories is at least 3,  $m_i \geq 3$ , the mean and the variance of an underlying variable  $x_i^*$  can be identified. The assumption of normal distribution is retained but instead of fixing the mean and the variance, the first and the second threshold,  $\tilde{\tau}_1^{(x_i)}$  and  $\tilde{\tau}_2^{(x_i)}$ , are set equal to 0 and 1, respectively. The two sets of parameters are connected as follows:  $\tilde{\mu}_{x_i} = -\tau_1^{(x_i)} / (\tau_2^{(x_i)} - \tau_1^{(x_i)})$ ,  $\tilde{\sigma}_{x_i} = 1 / (\tau_2^{(x_i)} - \tau_1^{(x_i)})$ , and  $\tilde{\tau}_{c_i}^{(x_i)} = (\tau_{c_i}^{(x_i)} - \tau_1^{(x_i)}) / (\tau_2^{(x_i)} - \tau_1^{(x_i)})$ , where  $\{\tau_{c_i}^{(x_i)}\}$  is the set of thresholds when it is set  $\mu_{x_i} = 0$  and  $\sigma_{x_i} = 1$ , and  $\{\tilde{\tau}_{c_i}^{(x_i)}\}$  is the set of thresholds when  $\tilde{\tau}_1^{(x_i)} = 0$  and  $\tilde{\tau}_2^{(x_i)} = 1$ . It becomes obvious that having estimated the one set of parameters, the estimates of the other set can be obtained directly using the above relationships.

<sup>4</sup>By setting the mean and the variance of a factor  $\xi_i$  equal to 0 and 1 the origin and the unit of its scale are defined, respectively. Instead of fixing the variance, another way to define the scale unit is to set one of the loadings referring to factor  $\xi_i$  equal to 1.

information estimation methods have been proposed and implemented in commercial software, with the most widely used the three-stage estimation methods (Jöreskog, 1990, 1994; Muthén, 1984). In these methods, thresholds are estimated by maximizing the univariate marginal distributions separately. Given the estimated thresholds, polychoric correlations are estimated by maximizing the bivariate marginal distributions separately. In the third stage, the factor analysis model given in (4) is fitted on the estimated polychoric correlation matrix using a version of generalized least squares (GLS), such as unweighted least squares (ULS), diagonally weighted least squares (DWLS), and weighted least squares (WLS) estimators (e.g. Jöreskog 1990, 1994; Jöreskog & Sörbom 1996, pp. 23-24; Muthén 1984; Muthén et al. 1997). In WLS, the weight matrix is an estimate of the inverse of the asymptotic covariance matrix of polychoric correlations, while DWLS involves only the diagonal elements of that weight matrix. Several studies have been carried out to compare these three least square methods (e.g. Forero et. al., 2009; Yang-Wallentin et al., 2010) and all lead to similar conclusions. The WLS estimator converges very slowly to its asymptotic properties and therefore does not perform well in small sample sizes. DWLS and ULS are preferable to WLS and they seem to perform similarly well in finite samples. However, in order to compute correct standard errors, the full weight matrix is needed and then the methods are called robust ULS (RULS), robust DWLS (RDWLS), and robust WLS (RWLS). The advantages of the three-stage RDWLS and RULS estimators are that they are computationally less demanding. On the downside, they are of stepwise nature and the standard errors of the model parameters from step three do not account for the sampling variability induced in step one. Besides, the estimate of the weight matrix is quite unstable for small sample sizes.

### 2.3 Item Response Theory (IRT) approach

Under the assumption of conditional independence, the factors account for all the dependencies among the ordinal variables. Therefore, one can write

$$\pi_r(\boldsymbol{\theta}|\boldsymbol{\xi}) = \pi_r(x_1 = c_1, \dots, x_p = c_p; \boldsymbol{\theta}|\boldsymbol{\xi}) = \prod_{i=1}^p \pi(x_i = c_i; \boldsymbol{\theta}|\boldsymbol{\xi}). \quad (6)$$

where the conditional response category probability is

$$\pi(x_i = c_i; \boldsymbol{\theta}|\boldsymbol{\xi}) = \gamma(x_i = c_i; \boldsymbol{\theta}|\boldsymbol{\xi}) - \gamma(x_i = c_i - 1; \boldsymbol{\theta}|\boldsymbol{\xi}), \quad (7)$$

and  $\gamma(x_i = c_i; \boldsymbol{\theta}|\boldsymbol{\xi})$  is the cumulative probability of a response into category  $c_i$  for variable  $x_i$ . The cumulative probability is modeled as follows:

$$\text{link}(\gamma(x_i = c_i; \boldsymbol{\theta}|\boldsymbol{\xi})) = \alpha_{c_i}^{(x_i)} - \sum_{j=1}^k \beta_{ij} \xi_j \quad (8)$$

or equivalently

$$\gamma(x_i = c_i; \boldsymbol{\theta}|\boldsymbol{\xi}) = F \left( \alpha_{c_i}^{(x_i)} - \sum_{j=1}^k \beta_{ij} \xi_j \right), \quad (9)$$

where the  $\alpha_{c_i}^{(x_i)}$ 's are thresholds ( $-\infty = \alpha_0^{(x_i)} < \alpha_1^{(x_i)} < \dots < \alpha_{m_i-1}^{(x_i)} < \alpha_{m_i}^{(x_i)} = +\infty$ ), and the  $\beta_{ij}$ 's are factors loadings. The link function can be any monotonically increasing function mapping  $(0, 1)$  onto  $(-\infty, \infty)$  such as the logit, the inverse normal (also called probit), the log-log, the

complementary log, and the inverse Cauchy function. When the logit link  $\Psi(x)$  is used, the model is known as the proportional odds model (Samejima, 1969). The probit function is denoted by  $\Phi(x)$  and then the model is referred to as the normal ogive response model (Jöreskog & Moustaki, 2001). The two functions,  $\Phi(x)$  and  $\Psi(x)$ , are of similar shapes. Lord and Novick (1968, p. 299) noted that  $|\Phi(x) - \Psi(1.7x)| < 0.01$  for all  $x$ .

The probability  $\pi_r(\boldsymbol{\theta})$  is now written as:

$$\pi_r(\boldsymbol{\theta}) = \int_{R_{\boldsymbol{\xi}}} \pi_r(\boldsymbol{\theta}|\boldsymbol{\xi})f(\boldsymbol{\xi})d\boldsymbol{\xi}, \quad (10)$$

where  $f(\boldsymbol{\xi})$  is the joint distribution of latent variables, usually assumed to be the  $k$ -dimensional standard<sup>5</sup> normal density function since the normal distribution has rotational advantages.

The log-likelihood is maximized using the E-M algorithm (Bartholomew et al., 2011; Muraki, 1990; Muraki & Carlson, 1995). FIML requires the evaluation of the  $k$ -dimensional integral. This integral cannot be written in a closed form but there are several numerical methods that can be used such as Gauss-Hermite quadrature, adaptive quadrature, Monte Carlo, and Laplace approximation. However, all of these methods have their limitations. Gauss-Hermite quadrature (Bock & Aitkin, 1981) is usually adequate for small models with small number of factors (e.g. two factors). Otherwise, it leads to a poor approximation and it is practically unfeasible when the number of factors is large. Adaptive quadrature (Schilling & Bock, 2005) is an improved version of Gauss-Hermite where accuracy and efficiency is increased by taking into account the mode and the curvature of the function integrated at each iterate. This way, the quadrature points are appropriately centered and rescaled. This adjustment not only speeds the convergence of the E-M algorithm but also “facilitates integration to the extent that fewer quadrature points can be used while still obtaining a convergent solution” (Schilling & Bock, 2005). However, the procedure can be quite slow in the case of a large number of factors. For higher dimensional integrals, Monte Carlo (Sammel et al. 1997) seems more feasible where randomly sampled quadrature points are used. A fairly large sample of points is required to obtain precise estimates. As a result, the convergence of the algorithm slows down as dimensionality increases. Laplace approximation (e.g. Huber et al. 2004) uses Taylor’s expansion to approximate the function to be integrated. Therefore, in all methods the computational burden increases rapidly with the number of factors  $k$  rendering FIML estimation in the case of IRT approach quite impractical or even unfeasible beyond a number of factors.

### 3 Composite Maximum Likelihood (CML) estimation

In situations where the full likelihood function either cannot be specified or is impractical to work with due to high computational complexity, alternative approaches based on modifications of the full information maximum likelihood method have been adopted with composite maximum likelihood methods being a competitive option. Composite likelihood methods can simplify computations substantially and at the same time yield estimators with the desired asymptotic properties of unbiasedness, consistency, and normality (Varin, 2008; Varin et al., 2011). Varin (2008) and Varin et al. (2011) give an extensive overview of these methods, present a wide range of application areas, and provide the following definition. Consider a  $p$ -dimensional random vector  $Y$  with probability density  $f(y; \theta)$  for some unknown  $d$ -dimensional vector parameter  $\theta \in \Theta$ . Also,

---

<sup>5</sup>Correlated latent variables can be fitted as well.

denote by  $\{\mathcal{A}_1, \dots, \mathcal{A}_K\}$  a set of measurable marginal or conditional events with associated likelihoods  $\mathcal{L}_k(\theta; y) \propto f(y \in \mathcal{A}_k; \theta)$ . Then, a composite likelihood (CL) is the weighted product of the likelihoods corresponding to each single event,

$$CL(\theta; y) = \prod_{k=1}^K \mathcal{L}_k(\theta; y)^{w_k},$$

where  $w_k$  are non-negative weights to be chosen. The composite maximum likelihood estimator  $\hat{\theta}_{CML}$  is obtained by maximizing the composite likelihood  $CL(\theta; y)$  over the parameter  $\theta$ . Under regularity conditions on the component likelihoods, the central limit theorem for the composite likelihood score statistic can be applied leading to the result

$$\sqrt{n} \left( \hat{\theta}_{CML} - \theta \right) \xrightarrow{d} N_d \left( 0, G^{-1}(\theta) \right),$$

where  $G(\theta)$  is the Godambe information matrix of a single observation (also referred to as Sandwich Estimator). In particular,

$$G(\theta) = H(\theta)J^{-1}(\theta)H(\theta),$$

where  $H(\theta)$  is the sensitivity matrix,  $H(\theta) = E \{-\nabla^2 \ln CL(\theta; y)\}$ , and  $J(\theta)$  is the variability matrix,  $J(\theta) = Var \{\nabla \ln CL(\theta; y)\}$ . In general, the identity  $H(\theta) = -J(\theta)$  does not hold in the case of composite likelihoods. The assumed independence among the likelihood terms forming the composite likelihood is usually not valid when the full likelihood is considered. The sample estimates of  $H(\theta)$  and  $J(\theta)$  are

$$\hat{H}(\theta) = -\frac{1}{n} \sum_{i=1}^n \nabla^2 \ln CL(\hat{\theta}_{CL}; y_i), \text{ and} \quad (11)$$

$$\hat{J}(\theta) = \frac{1}{n} \sum_{i=1}^n \left( \nabla \ln CL(\hat{\theta}_{CL}; y_i) \right) \left( \nabla \ln CL(\hat{\theta}_{CL}; y_i) \right)^T \quad (12)$$

respectively.

Varin et al. (2011) discuss some other qualities of composite likelihood as well. Composite likelihood can be seen as a robust alternative in terms of modeling. In some cases it is easier and more straightforward to model lower order dimensional distributions while modeling uncertainty increases with dimensionality. By applying composite likelihood, possible misspecification of the higher order dimensional distributions can be avoided. In addition, a model assumed for lower order distributions can be compatible with more than one possible modeling options available for higher dimensional distributions. Moreover, in some settings, there are no obvious high dimensional distributions. Also, the composite likelihood function surface is usually smoother than that of full likelihood and therefore, easier to maximize.

When it comes to goodness-of-fit and model selection criteria, related theory applicable to the case of composite likelihood methods has been developed. In general, the fit of a model with ordinal data is usually assessed by the Pearson's test statistic

$$X^2 = n \sum_r (p_r - \pi_r(\theta))^2 / \pi_r(\theta)$$

or the likelihood ratio test statistics

$$G^2 = 2n \sum_{p_r > 0} p_r \ln (p_r / \pi_r(\theta)),$$

which are asymptotically equivalent under the null hypothesis. However, as it is well known, these two test statistics are strongly affected by the sparseness of the observed  $p$ -dimensional contingency table. Besides, in the case of rejection of the null hypothesis,  $X^2$  and  $G^2$  statistics do not offer much insight where the source of misfit is. To overcome these difficulties, Maydeu-Olivares & Joe (2005; 2006) propose a family of goodness-of-fit statistics where the statistics are quadratic forms of marginal residuals up to order  $l$ ,  $l = 1, \dots, p$ . This way, Pearson's test statistic  $X^2$  is a special case of this family (when  $l = p$ ). These statistics are asymptotically chi-square distributed under the null hypothesis when the model parameter  $\theta$  is estimated using any asymptotically normal consistent estimator. Hence, their theory applies to the case of composite likelihood estimation as well. An even more general theory was presented by Joe & Maydeu-Olivares (2010) where the statistics need not be based on marginal residuals but are quadratic forms of linear functions of cell residuals. About model selection criteria, Varin & Vidoni (2005) introduced a criterion to be used with composite likelihood estimation methods.

### 3.1 *CML versus Partition Maximum Likelihood estimation*

At this point, it might be helpful to briefly discuss the similarities and differences between CML and partition maximum likelihood estimation methods (e.g. Chan & Bentler, 1998; Poon & Lee, 1987). Partition ML is also a method based on a modification of full ML in order to simplify computations and yield an estimator with the desired asymptotic properties of unbiasedness, consistency, and normality (e.g. Chan & Bentler, 1998). The full joint likelihood is broken down into lower order likelihoods in a similar fashion as in CML. However, in partition ML, maximization over the parameter is carried out at each individual lower-order likelihood separately. This way, we get a set of estimates for the parameter which have to be combined in one final estimate, usually done by taking their average. Hence, the direct advantage of CML over partition ML is that, in CML, there is only one objective function to be maximized and all parameters are estimated in one step simultaneously.

### 3.2 *Application of CML methods in factor analysis with ordinal data*

To our knowledge, there are very few studies investigating the applicability and performance of CML approaches within the context of factor analysis with ordinal data. So far, Jöreskog & Moustaki (2001) has proposed the underlying bivariate normal (UBN) approach to estimate the model parameters, thresholds, loadings, and factor correlations. In this approach, the log-likelihood is composed of the sum of all univariate and bivariate marginal log-likelihoods. Although it can be seen as a composite likelihood method, it has not been studied as such. We briefly present the UBN approach below as it was the starting point of our research. De Leon (2005) suggests pairwise maximum likelihood (PML) to estimate the thresholds and polychoric correlations of ordinal data. PML is similar to the UBN likelihood, except only the sum of the bivariate distributions is involved in the likelihood. De Leon (2005) studies the performance of the PML estimator in terms of bias and root mean squared errors with a small scale simulation study where three ordinal variables and sample sizes of 50 and 100 are considered. The main conclusion is that the PML estimates are very close to the true values with very small bias and root mean squared errors. Liu (2007) proposes a new multistage estimation method for factor models and structural equation models with data of mixed type (both ordinal and continuous). Instead of maximizing each univariate and bivariate marginal distribution separately to estimate thresholds and polychoric/ polyserial correlations respectively, as in the common three-stage estimation methods, he suggests that all

these parameters can be estimated simultaneously by using PML; and then, in a second stage, similarly to the third stage of the conventional three-stage methods, the structural parameters, such as the loadings and factor correlations, are estimated given the estimates of the first stage and taking into account their asymptotic distribution. In this second stage, generalized least squares is employed with the asymptotic covariance matrix of the first stage estimates being used as a weight matrix. Based on the results of simulation studies, Liu (2007) concludes that the proposed method performs equally well and is slightly more robust than the usual three-stage estimation approaches.

### *The Underlying Bivariate Normal (UBN) approach*

The UBN approach has been developed within the more general framework of URV approach. Based on the expression of the expected probability of a response pattern  $\pi_r(\boldsymbol{\theta})$  given in (5), the probability of a response pattern to a pair of indicators  $(x_i, x_j)$  in terms of the parameter  $\boldsymbol{\theta}' = (\boldsymbol{\lambda}^{(x)}, \boldsymbol{\varphi}, \boldsymbol{\tau}^{(x)})$  can be written as

$$\begin{aligned} \pi_{c_i c_j}^{(x_i x_j)}(\boldsymbol{\theta}) &= \pi(x_i = c_i, x_j = c_j; \boldsymbol{\theta}) = \\ &= \Phi_2\left(\tau_{c_i}^{(x_i)}, \tau_{c_j}^{(x_j)}; \rho_{x_i x_j}\right) - \Phi_2\left(\tau_{c_i}^{(x_i)}, \tau_{c_j-1}^{(x_j)}; \rho_{x_i x_j}\right) - \end{aligned} \quad (13)$$

$$- \Phi_2\left(\tau_{c_i-1}^{(x_i)}, \tau_{c_j}^{(x_j)}; \rho_{x_i x_j}\right) + \Phi_2\left(\tau_{c_i-1}^{(x_i)}, \tau_{c_j-1}^{(x_j)}; \rho_{x_i x_j}\right), \quad (14)$$

and the probability of a certain response to a single indicator  $x_i$  as

$$\pi_{c_i}^{(x_i)}(\boldsymbol{\theta}) = \pi(x_i = c_i; \boldsymbol{\theta}) = \Phi\left(\tau_{c_i}^{(x_i)}\right) - \Phi\left(\tau_{c_i-1}^{(x_i)}\right),$$

where  $\Phi_2(a, b; \rho)$  is the bivariate cumulative normal distribution with correlation coefficient equal to  $\rho$  evaluated at point  $(a, b)$ ,  $\Phi(a)$  is the univariate cumulative normal distribution evaluated at point  $a$ , and  $\rho_{x_i x_j}$  is the polychoric correlation of observed variables  $x_i$  and  $x_j$ . The latter is expressed in terms of the model parameter as

$$\rho_{x_i x_j}(\boldsymbol{\theta}) = \boldsymbol{\lambda}_i \cdot \Phi \boldsymbol{\lambda}_j',$$

where  $\boldsymbol{\lambda}_i$  is an  $1 \times k$  row vector containing the elements of the  $i^{\text{th}}$  row of matrix  $\Lambda_x$ . The UBN approach suggests that, instead of maximizing the full log-likelihood in (2) over the parameter  $\boldsymbol{\theta}$ , one could maximize the log-likelihood composed of the sum of all univariate and the sum of all bivariate marginal log-likelihoods, i.e.

$$\ln L_{UBN}(\boldsymbol{\theta}; \mathbf{x}) = \sum_{i < j} \ln L(\boldsymbol{\theta}; (x_i, x_j)) + \sum_i \ln L(\boldsymbol{\theta}; x_i),$$

where the bivariate log-likelihoods are of the form

$$\ln L(\boldsymbol{\theta}; (x_i, x_j)) = \sum_{c_i=1}^{m_i} \sum_{c_j=1}^{m_j} n_{c_i c_j}^{(x_i x_j)} \ln \pi_{c_i c_j}^{(x_i x_j)}(\boldsymbol{\theta}),$$

and the univariate log-likelihoods are of the form

$$\ln L(\boldsymbol{\theta}; x_i) = \sum_{c_i=1}^{m_i} n_{c_i}^{(x_i)} \ln \pi_{c_i}^{(x_i)}(\boldsymbol{\theta}).$$

The UBN log-likelihood  $\ln L_{UBN}(\boldsymbol{\theta}; \mathbf{x})$ , despite including only the first and second order marginal distributions, contains information for all model parameters to be estimated. Its obvious advantage over FIML estimation is that it only requires the evaluation of up to two-dimensional normal probabilities, regardless of the number of observed or latent variables. This way, it is always computationally feasible. Comparing the UBN approach with the three-stage limited information estimation methods, the advantage of the former is that the estimation of all parameters is carried out simultaneously. Moreover, the standard errors of the estimates can be obtained without the usage of any weight matrix.

Jöreskog & Moustaki (2001) demonstrate UBN method using a real data set and two generated data sets. Keeping the number of factors up to two, they were able to compare the UBN estimates with the FIML estimates derived under the IRT approach when the link function is logit or probit. The main conclusion is that the estimates were similar across the three approaches and fairly close to the true values in the case of the simulated data. However, Jöreskog & Moustaki (2001) do not incorporate their approach within the CML framework or any other general framework. Thus, they do not provide any discussion about the standard errors and the properties of the UBN estimator.

## 4 Proposed methodology

The promising empirical results of the aforementioned studies motivated us to develop the UBN approach further. First of all, the UBN estimation method can be incorporated in the larger framework of composite likelihood methods. This way, we can take advantage of the related theory. The computation of standard errors become straightforward and test statistics used for inference are available. Secondly, one can question the role of the sum of univariate distributions in the accuracy and efficiency of UBN method. The univariate distributions contain information only of the thresholds, while the bivariate ones contain information of all URV model parameters  $\boldsymbol{\lambda}$ ,  $\boldsymbol{\varphi}$ , and  $\boldsymbol{\tau}$ . Can this information overlap be balanced out by an “optimal” weighting scheme which improves the efficiency of UBN method? Cox & Reid (2004) provide a general framework that can be used in order to deal with such a question. Their proposed log-likelihood is very similar to the UBN log-likelihood with the only difference that they put a weight on the sum of the univariate distributions. Following their suggestion, the UBN log-likelihood should be modified as follows:

$$l(\boldsymbol{\theta}; \mathbf{x}) = \sum_{i < j} \ln L(\boldsymbol{\theta}; (x_i, x_j)) - ap \sum_i \ln L(\boldsymbol{\theta}; x_i) ,$$

where  $a$  is a constant to be chosen for optimal efficiency and  $p$  is the number of observed variables. Cox & Reid (2004) point out that if the univariate likelihoods are independent of  $\boldsymbol{\theta}$  then the choice of  $a = 0$  is appropriate; taking  $a = \frac{1}{2}$  corresponds to the situation where all possible conditional distributions of one variable given another are considered. In general, they suggest that a non-negative value of  $a$  is appropriate.

Trying different values of  $a$  so that the value of  $ap$  ranges from 0 to 1 and conducting some small scale simulation studies, our results indicated that practically the sum of univariate distributions affect neither the estimate accuracy nor the efficiency. Therefore, we concluded that the most appropriate choice of  $a$  is zero in our case. Subsequently, instead of UBN log-likelihood, we suggest that one could consider the composite pairwise log-likelihood to estimate the URV parameter  $\boldsymbol{\theta}' = (\boldsymbol{\lambda}, \boldsymbol{\varphi}, \boldsymbol{\tau})$ . That is of the form:

$$pl(\boldsymbol{\theta}; \mathbf{x}) = \sum_{i < j} \ln L(\boldsymbol{\theta}; (x_i, x_j)) = \sum_{i < j} \sum_{c_i=1}^{m_i} \sum_{c_j=1}^{m_j} n_{c_i c_j}^{(x_i x_j)} \ln \pi_{c_i c_j}^{(x_i x_j)}(\boldsymbol{\theta}) , \quad (15)$$

where  $\pi_{c_i c_j}^{(x_i x_j)}$  is as defined in expression (14),  $i = 1, \dots, p - 1$ , and  $j = i + 1, \dots, p$ . Note that in the case of cross-sectional data, which is of interest in the current study, an equal weighting of all pairwise likelihood components seems to be the most appropriate option<sup>6</sup>. Maximizing the function over the parameter  $\boldsymbol{\theta}$  we get the composite pairwise maximum likelihood (PML) estimator  $\hat{\boldsymbol{\theta}}_{PML}$ . The gradient of the pairwise log-likelihood  $\nabla pl(\boldsymbol{\theta}; \mathbf{x})$  is equal to the sum of the gradients of the bivariate log-likelihood components  $\nabla \ln L(\boldsymbol{\theta}; (x_i, x_j))$ . The explicit form of the latter is given in Appendix I. To estimate the asymptotic standard error we apply expressions (11) and (12) to our case and get

$$\hat{H}(\hat{\boldsymbol{\theta}}_{PML}) = \nabla^2 pl(\hat{\boldsymbol{\theta}}_{PML}; \mathbf{x}), \text{ and}$$

$$\hat{J}(\hat{\boldsymbol{\theta}}_{PML}) = \frac{1}{n} \sum_{h=1}^n \left( \sum_{i < j} \nabla \ln \pi_{c_i c_j}^{(x_i^{(h)} x_j^{(h)})}(\hat{\boldsymbol{\theta}}_{PML}) \right) \left( \sum_{i < j} \nabla \ln \pi_{c_i c_j}^{(x_i^{(h)} x_j^{(h)})}(\hat{\boldsymbol{\theta}}_{PML}) \right)^T.$$

Our research interest is twofold. The first objective is to compare the PML estimates and their standard errors with those obtained by FIML and 3S-RULS using real data examples where either exploratory or confirmatory factor analysis is carried out. We choose RULS among the three-stage limited information estimators because it performs well in small samples. Moreover, between RULS and RDWLS, RULS is recommended by Forero et. al. (2009) who performed a detailed simulation study comparing the two methods under numerous experimental conditions. Section 5 presents the related results. Secondly, we aim to study the performance of the PML estimator  $\boldsymbol{\theta}_{PML}$  and its estimated asymptotic standard error in terms of bias and mean squared error under different sample and model sizes. Our simulation study and its results are reported in Section 6.

#### 4.1 *Technical Issues*

In the current study, the maximization of  $pl(\boldsymbol{\theta}; \mathbf{x})$  has been carried out by using the “maxLik” command of “maxLik” package of R software. As an input to the command we have given two “function objects” (according to R language) which we wrote. One function is for the composite pairwise log-likelihood as expressed in (15) and the other one for its gradient given in the appendix. The “maxLik” command offers a few options about the maximization algorithm. We have used the Broyden-Fletcher-Goldfarb-Shanno method denoted as BFGS in R, which is a member of the larger group of quasi-newton methods. Regarding the estimate  $\hat{H}(\hat{\boldsymbol{\theta}}_{PML})$  that is part of the output of “maxLik” command. About  $\hat{J}(\hat{\boldsymbol{\theta}}_{PML})$  we wrote our own function object. The 2.10.1 version of R software has been used.

## 5 Comparison of PML, FIML, and 3S-RULS estimates and standard errors

In this section, the PML estimation is demonstrated by applying the method in exploratory and confirmatory factor analysis of some empirical data. Moreover, the PML estimates and standard

---

<sup>6</sup>Unequal weighting schemes arise naturally in the case of time series or spatial data where larger weights are usually put on observations being close in time or space. Also, in cluster data using weights inversely related to cluster sizes is almost always the case.

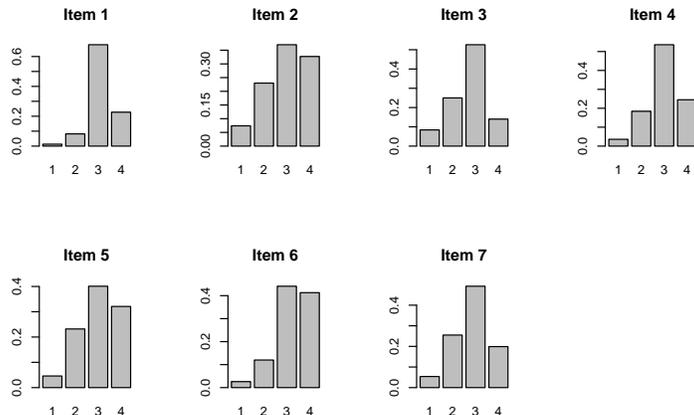


Figure 1: Observed distribution of each item in S&T data

errors are compared with those obtained by FIML and 3S-RULS. The FIML estimates and standard errors are derived under the IRT approach with probit as the link function. Although the parameters in this case are the  $\alpha_{c_i}^{(x_i)}$  and  $\beta_{ij}$ , the corresponding estimates of  $\lambda_{ij}$  and  $\tau_{c_i}^{(x_i)}$  can be obtained easily, as the two sets of parameters are related to each other. The relationship between them were noted by Takane & de Leeuw (1987) and Bartholomew et al. (2011), and the related formulas can be found in Jöreskog & Moustaki (2001). All statistical packages provide both sets of parameters. In our examples, the package Mplus 6.1 has been used to get the FIML estimates and standard errors. In the examples of one- and two-factor models, the numerical algorithm used to evaluate the integral in (10) was the default of Mplus, i.e. adaptive quadrature with 15 integration points for each dimension. In the example of the four-factor model, the default method is remarkably slow and, as the program suggests, Monte Carlo with the default settings of Mplus has been used. To get the 3S-RULS estimates and their standard errors LISREL 8.80 has been used. However, similar results can be obtained by Mplus as well using the same estimation method. In all the examples, we focus on the estimates of loadings  $\lambda_{ij}$  and factor correlations  $\phi_{jh}$ , since these are the parameters that inference is based on and practitioners are interested in.

### 5.1 Exploratory factor analysis - Science & Technology (S&T) data

The data used in this example come from the Consumer Protection and Perceptions of Science and Technology section of the 1992 Eurobarometer Survey (Karlheinz & Melich, 1992) and particularly, it is based on a sample from Great Britain. Seven items are used in the current analysis, all presented in Appendix II. All the items were measured on a four-point scale with response categories “strongly disagree”, “disagree to some extent”, “agree to some extent”, and “strongly agree”. The sample size is 392 after eliminating the cases with missing values in any of the items (listwise deletion). Almost all items present a considerable amount of skewness as it can be seen by Figure 1 which shows the observed distribution of each item.

Exploratory factor analysis with one and two factors have been carried out where the factors are assumed to follow standard normal distribution. For the two-factor exploratory analysis the loading of the first item on the second factor, namely  $\lambda_{12}$ , has been fixed to 0, for identification reasons.

The estimates of factor loadings obtained by the three estimation methods in the case of one

factor are given in Table 1 and plotted in Figure 2. In the figure the parameters are denoted with an index on the horizontal axis and presented in the same order as in the table. As it can be seen, all three methods give very close estimates with the PML estimates being on average slightly closer to the FIML ones. The average of absolute differences between PML and FIML estimates is 0.049, while that between 3S-RULS and FIML is 0.079. For the standard errors, the picture is opposite. The 3S-RULS standard errors are on average slightly closer to the FIML ones with 0.031 average absolute difference, while that between PML and FIML is 0.055. An interesting pattern though, that can be seen in Figure 2, is that 3S-RULS gives the smaller standard errors for all parameters, while PML the larger.

Table 2 and Figure 3 report the results of the exploratory analysis with two factors. As before, all three estimation methods give very similar estimates with PML ones being slightly closer to FIML than 3S-RULS are. The average of absolute differences between FIML and PML estimates is 0.012, while that between FIML and 3S-RULS is 0.020. About standard errors the picture is similar; the PML standard errors are a bit closer to FIML ones than 3S-RULS are. The average of absolute differences between FIML and PML standard errors is 0.0027, while that between FIML and 3S-RULS is 0.0081. As expected, the 3S-RULS standard errors are the smaller ones among the three methods for almost all the parameter estimates, since 3S-RULS does not incorporate the sampling variability of the first stage.

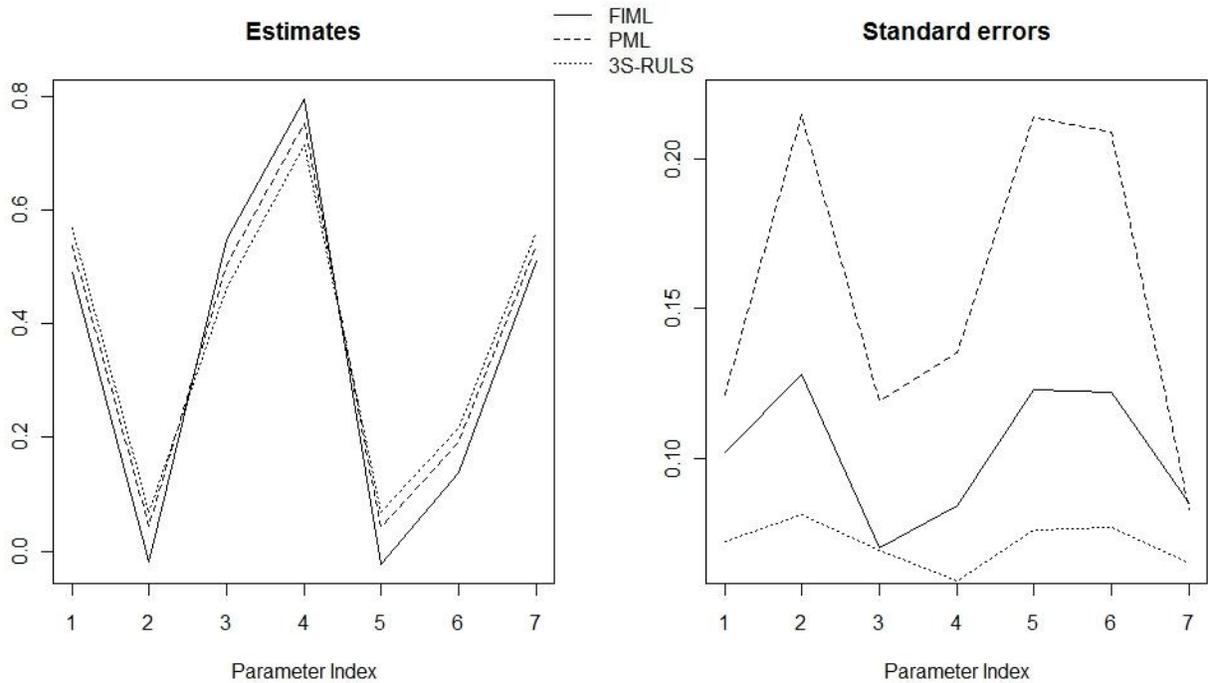


Figure 2: The loading estimates and standard errors obtained by PML, FIML, and 3S-RULS, S&T data, one-factor exploratory analysis

Parameter	PML	FIML	3S-RULS
$\lambda_1$	0.536 (0.121)	0.491 (0.102)	0.568 (0.072)
$\lambda_2$	0.044 (0.215)	-0.018 (0.128)	0.069 (0.081)
$\lambda_3$	0.503 (0.119)	0.548 (0.070)	0.464 (0.069)
$\lambda_4$	0.752 (0.135)	0.795 (0.084)	0.713 (0.059)
$\lambda_5$	0.042 (0.214)	-0.023 (0.123)	0.068 (0.076)
$\lambda_6$	0.192 (0.209)	0.139 (0.122)	0.217 (0.077)
$\lambda_7$	0.538 (0.083)	0.511 (0.085)	0.562 (0.065)

Table 1: Estimated loadings and standard errors in brackets obtained by PML, FIML, and 3S-RULS, S&T data, one-factor exploratory analysis

Parameter	PML	FIML	3S-RULS	Parameter	PML	FIML	3S-RULS
$\lambda_{11}$	0.545 (0.076)	0.529 (0.081)	0.556 (0.069)	$\lambda_{12}$	0* (fixed) —	0* (fixed) —	0* (fixed) —
$\lambda_{21}$	0.183 (0.116)	0.197 (0.121)	0.174 (0.111)	$\lambda_{22}$	0.633 (0.074)	0.637 (0.074)	0.634 (0.074)
$\lambda_{31}$	0.464 (0.080)	0.479 (0.079)	0.449 (0.079)	$\lambda_{32}$	-0.297 (0.100)	-0.312 (0.101)	-0.283 (0.087)
$\lambda_{41}$	0.719 (0.069)	0.731 (0.070)	0.707 (0.063)	$\lambda_{42}$	-0.279 (0.127)	-0.285 (0.131)	-0.261 (0.112)
$\lambda_{51}$	0.189 (0.116)	0.206 (0.120)	0.179 (0.111)	$\lambda_{52}$	0.674 (0.071)	0.675 (0.073)	0.675 (0.070)
$\lambda_{61}$	0.353 (0.100)	0.369 (0.102)	0.345 (0.098)	$\lambda_{62}$	0.549 (0.083)	0.554 (0.085)	0.552 (0.082)
$\lambda_{71}$	0.510 (0.077)	0.493 (0.083)	0.532 (0.066)	$\lambda_{72}$	-0.135 (0.077)	-0.153 (0.080)	-0.140 (0.087)

Table 2: Estimated loadings and their standard errors in brackets obtained by PML, FIML, and 3S-RULS, S&T data, two-factor exploratory analysis

## 5.2 Confirmatory factor analysis - Relationship Learning (RL) data

In this example we use part of the data gathered by Selnes & Sallis (2003) who aimed to study whether specific factors affect the learning capabilities of targeted customer-supplier relationships. We focus on the data coming from suppliers and referring to 18 specific items which measure four factors; collaborative commitment ( $\xi_1$ ), internal complexity ( $\xi_2$ ), relational trust ( $\xi_3$ ), and environmental uncertainty ( $\xi_4$ ) as named by Selnes & Sallis (2003). The items used to measure each factor are presented in Appendix III. All items were measured on a seven-point scale with 1 referring to “strongly disagree” or “low” and 7 to “strongly agree” or “high” depending on the form

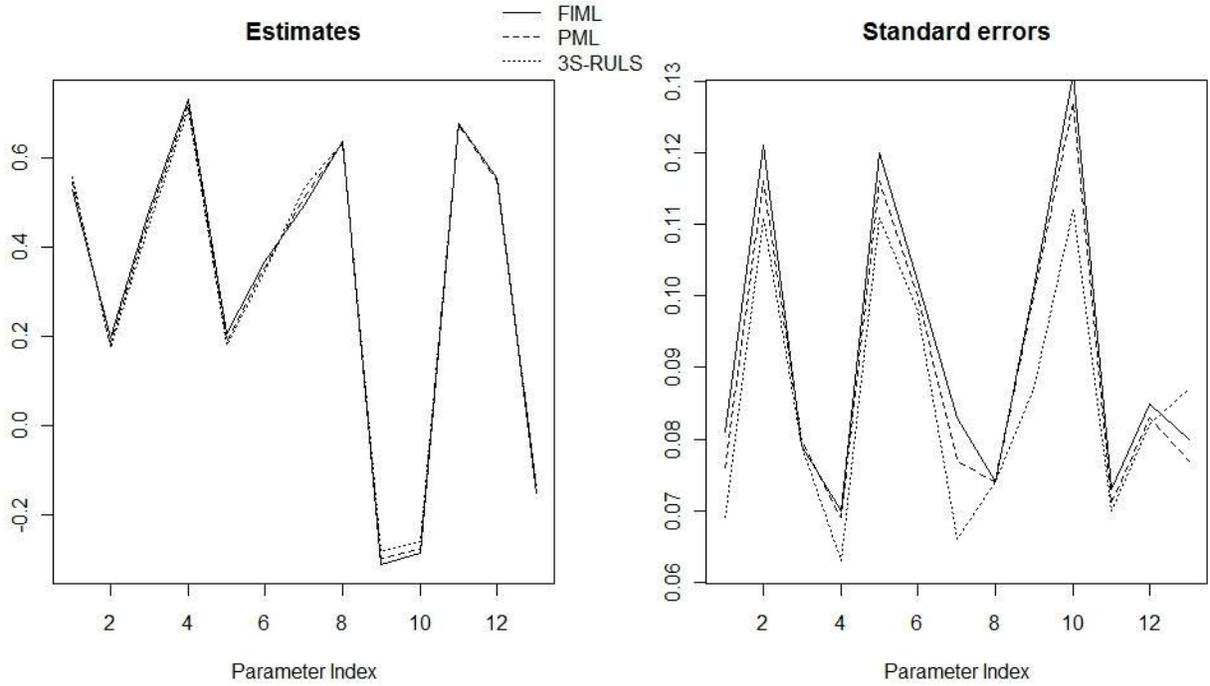


Figure 3: Estimated loadings and their standard errors obtained by PML, FIML, and 3S-RULS, S&T data, two-factor exploratory analysis

of the question. The sample size is 286 after listwise deletion. The observed distribution of each item is shown in Figure 5.2. As it can be seen, quite many of them are rather skewed while the rest somehow closer to a more symmetric distribution.

The structure of matrices  $\Lambda$  and  $\Phi$  to be estimated are

$$\Lambda = \begin{pmatrix} \lambda_{1,1} \\ \lambda_{2,1} \\ \lambda_{3,1} \\ \lambda_{4,1} \\ \lambda_{5,1} \\ & \lambda_{6,2} \\ & \lambda_{7,2} \\ & \lambda_{8,2} \\ & & \lambda_{9,3} \\ & & \lambda_{10,3} \\ & & \lambda_{11,3} \\ & & \lambda_{12,3} \\ & & \lambda_{13,3} \\ & & & \lambda_{14,4} \\ & & & \lambda_{15,4} \\ & & & \lambda_{16,4} \\ & & & \lambda_{17,4} \\ & & & \lambda_{18,4} \end{pmatrix}, \quad \Phi = \begin{pmatrix} 1 & & & \\ \phi_{21} & 1 & & \\ \phi_{31} & \phi_{32} & 1 & \\ \phi_{41} & \phi_{42} & \phi_{43} & 1 \end{pmatrix}. \quad (16)$$

In total, along with the thresholds which are six for each item, there are 132 free parameters to be

estimated.

The estimates of factor loadings and correlations obtained by the three estimation methods are reported in Table 3 and are depicted in Figure 5. As it can be seen, all three methods give very similar estimates for all parameters with PML estimates being slightly closer to FIML ones than 3S-RULS are. The average of absolute differences between the FIML and PML estimates is 0.022, while that between the FIML and 3S-RULS is 0.047. For the standard errors, the picture is similar. The average of absolute differences between the FIML and PML standard errors is 0.0036, while that between the FIML and 3S-RULS is 0.0090. It is interesting to note that the 3S-RULS standard errors are the larger ones for more than half of the estimates.

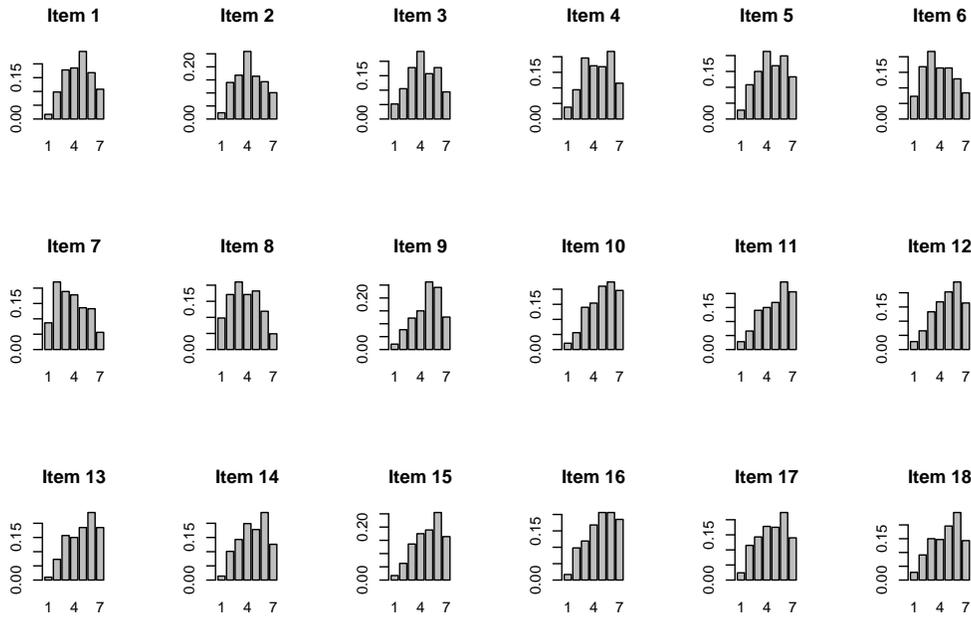


Figure 4: Observed distribution of each item in RL data

Parameter	PML	FIML	3S-RULS	Parameter	PML	FIML	3S-RULS
$\lambda_{1,1}$	0.882 (0.023)	0.880 (0.022)	0.905 (0.027)	$\lambda_{14,4}$	0.767 (0.033)	0.775 (0.031)	0.774 (0.035)
$\lambda_{2,1}$	0.891 (0.017)	0.894 (0.016)	0.897 (0.023)	$\lambda_{15,4}$	0.854 (0.027)	0.802 (0.032)	0.873 (0.028)
$\lambda_{3,1}$	0.884 (0.018)	0.900 (0.015)	0.867 (0.023)	$\lambda_{16,4}$	0.752 (0.040)	0.783 (0.035)	0.740 (0.042)
$\lambda_{4,1}$	0.897 (0.016)	0.905 (0.015)	0.907 (0.018)	$\lambda_{17,4}$	0.697 (0.044)	0.743 (0.044)	0.697 (0.043)
$\lambda_{5,1}$	0.875 (0.020)	0.886 (0.018)	0.850 (0.025)	$\lambda_{18,4}$	0.705 (0.042)	0.724 (0.042)	0.710 (0.042)
$\lambda_{6,2}$	0.622 (0.079)	0.520 (0.064)	0.775 (0.113)	$\phi_{21}$	0.255 (0.083)	0.208 (0.084)	0.285 (0.080)
$\lambda_{7,2}$	0.821 (0.065)	0.849 (0.051)	0.711 (0.090)	$\phi_{31}$	0.627 (0.044)	0.627 (0.042)	0.630 (0.043)
$\lambda_{8,2}$	0.784 (0.069)	0.834 (0.050)	0.669 (0.093)	$\phi_{41}$	0.658 (0.047)	0.659 (0.049)	0.658 (0.047)
$\lambda_{9,3}$	0.808 (0.027)	0.801 (0.027)	0.826 (0.031)	$\phi_{32}$	0.125 (0.073)	0.113 (0.076)	0.135 (0.074)
$\lambda_{10,3}$	0.866 (0.023)	0.873 (0.022)	0.857 (0.026)	$\phi_{42}$	0.197 (0.079)	0.147 (0.079)	0.220 (0.077)
$\lambda_{11,3}$	0.867 (0.024)	0.884 (0.019)	0.842 (0.028)	$\phi_{43}$	0.651 (0.048)	0.641 (0.052)	0.638 (0.047)
$\lambda_{12,3}$	0.908 (0.016)	0.913 (0.015)	0.903 (0.019)				
$\lambda_{13,3}$	0.871 (0.020)	0.865 (0.020)	0.888 (0.023)				

Table 3: Estimated loadings, correlations and their standard errors in brackets obtained by PML, FIML, and 3S-RULS, RL data, confirmatory analysis

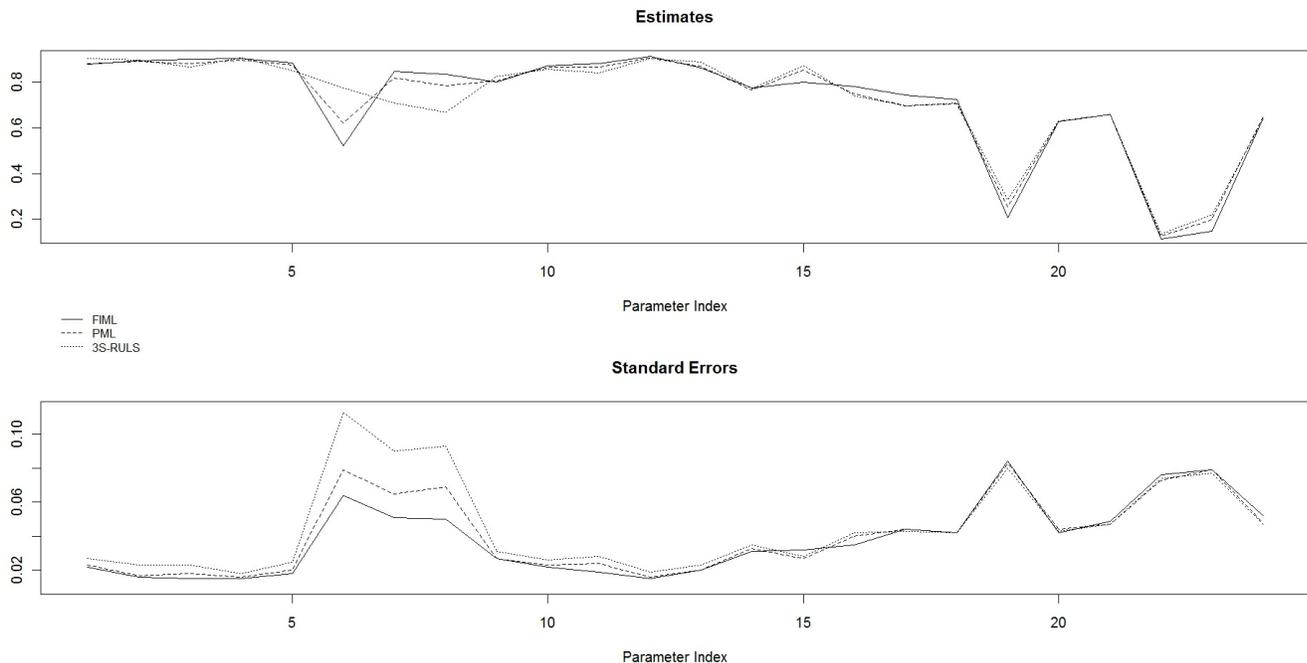


Figure 5: Estimated loadings, correlations and their standard errors obtained by FIML, PML, and 3S-RULS, RL data, confirmatory analysis

## 6 Simulation study on the performance of PML estimator

### 6.1 Set-up of the simulation study

A simulation study has been conducted to evaluate the performance of the PML estimator within the framework of confirmatory factor analysis with ordinal data. Six experimental conditions have been investigated by combining three different sample sizes, 200, 500, and 1000, with two different models, referred to as Model I and Model II. For each condition 1000 replicates have been carried out.

Model I is regarded as a small size model with 6 observed variables ( $p = 6$ ) and 2 factors ( $k = 2$ ). The number of indicators for each factor is relatively small, 3 or 4 indicators for each factor. The true values of matrices  $\Lambda$  and  $\Phi$  are

$$\Lambda = \begin{pmatrix} 0.9 \\ 0.8 \\ 0.7 \\ 0.5 & 0.6 \\ & 0.7 \\ & & 0.8 \end{pmatrix}, \quad \Phi = \begin{pmatrix} 1 & \\ 0.5 & 1 \end{pmatrix}. \quad (17)$$

As it can be seen the loadings range from high, 0.9, to relatively low, 0.5, and there is one multidimensional indicator whose loadings are relatively small and close in value, 0.5 and 0.6. The factor correlation is of moderate size, 0.5. All observed variables are assumed to have 4 response categories, a case quite often met in applications, and same thresholds, namely  $\tau_1^{(x_i)} = -1.2$ ,

$\tau_2^{(x_i)} = 0$ ,  $\tau_3^{(x_i)} = 1.2$ ,  $i = 1, \dots, 6$ . Therefore, there are 26 free parameters to be estimated for Model I.

Model II is of larger size with 15 observed variables ( $p = 15$ ) and 3 factors ( $k = 3$ ). The number of indicators per factor is moderate, 5 or 6 indicators per factor. The true values of matrices  $\Lambda$  and  $\Phi$  are

$$\Lambda = \begin{pmatrix} 0.4 \\ 0.5 \\ 0.6 \\ 0.7 \\ 0.8 \\ 0.3 & 0.8 \\ & 0.7 \\ & 0.6 \\ & 0.5 \\ & 0.4 & 0.5 \\ & & 0.6 \\ & & 0.7 \\ & & 0.8 \\ & & 0.9 \\ & & 0.4 \end{pmatrix}, \quad \Phi = \begin{pmatrix} 1 & & \\ 0.2 & 1 & \\ 0.5 & 0.8 & 1 \end{pmatrix}. \quad (18)$$

Again the loadings range from high, 0.9, to low, 0.3. In this model, there are two multidimensional indicators. The first one has one high loading, 0.8, and one low, 0.3, while the second one have both loadings low and close in value, 0.4 and 0.5, i.e. a similar pattern as in Model I. The factor correlations range from low to high, 0.2, 0.5, and 0.8. Again all observed variables are assumed to have 4 response categories and the same thresholds, namely the same thresholds as in Model I. There are 65 free parameters to be estimated in the larger model.

## 6.2 Data generation

The following steps have been taken to generate data within each model and sample size:

1. A random vector  $\boldsymbol{\xi}$  and a random vector  $\boldsymbol{\delta}$  are generated from  $N_k(\mathbf{0}, \Phi)$  and  $N_p(\mathbf{0}, \Theta)$ , respectively. (Recall that  $\Theta = I - \text{diag}(\Lambda\Phi\Lambda')$ .)
2. A random vector of underlying variables  $\mathbf{x}^*$  is generated by applying the assumed model  $\mathbf{x}^* = \Lambda\boldsymbol{\xi} + \boldsymbol{\delta}$ .
3. A random vector of ordinal variables  $\mathbf{x}$  is obtained from  $\mathbf{x}^*$  by applying the relationship in (3) which connects the continuous underlying variables with the ordinal observed ones. Thus, the values of thresholds  $\tau_{c_i}^{(x_i)}$  are used in this step.
4. Steps 1-3 are repeated  $n$  times to get a sample of size  $n$ .

## 6.3 Performance Criteria

Bias and mean squared error (MSE) are defined as follows:

$$\text{Bias} = \frac{1}{R} \sum_{i=1}^R (\hat{\theta}_i - \theta), \quad \text{and}$$

$$MSE = \frac{1}{R} \sum_{i=1}^R (\hat{\theta}_i - \theta)^2 ,$$

where  $R$  here is the number of replicates, 1000 in this study,  $\hat{\theta}_i$  is the estimate of a parameter or of its asymptotic standard error at the  $i^{th}$  replication, and  $\theta$  is the corresponding true value. In the case of standard errors, where the true value  $\theta$  is unknown, the standard deviation of the corresponding 1000 estimates of the parameter is used.

## 6.4 Results

The first thing to study is the proportion of valid replications per condition which, in our case, is defined as the percentage of replications per condition providing proper solutions, i.e. estimated loadings and factor correlations have values between -1 and 1. We consider valid replications as the only valid observations that can be included in the analysis, as in Forero & Maydeu-Olivares (2009) and Forero et. al. (2011). Table 4 reports these percentages. It can be seen that, for the smaller model, we got 100% valid replications with all sample sizes. However, for Model II, the percentage ranges from 94.6% to 99.3%, increasing as the sample size increases. It is worthwhile to note that all invalid replications gave improper solutions only for the loadings of the second multidimensional item,  $\lambda_{10,2}$  and  $\lambda_{10,3}$ , and a few of them for the correlation between the 2nd and 3rd factor,  $\phi_{23}$ . Removing the invalid replications from the analysis, that has affected only the results referring to these three aforementioned parameters, while the results for the rest parameters have remained practically unchanged. Interestingly, Forero et. al. (2011) comment that “convergence and estimation accuracy problems are aggravated in the presence of... multidimensional indicators”. In our case, the related indicator has besides relatively low and close loadings.

	Model Size	
Sample size $n$	Model I	Model II
200	100%	94.6%
500	100%	98.2%
1000	100%	99.3%

Table 4: Percentage of valid replications per condition

The results with respect to the bias and MSE of estimates and their estimated asymptotic standard errors are presented in Tables 5 - 8. Each table reports for each parameter the following: the true value, the mean value of the PML estimates, the corresponding bias and MSE, the standard deviation of these PML estimates, the mean of the estimated standard errors, and the corresponding bias and MSE. Figures 6.4 and 7 depict the bias and MSE of both estimates and standard errors for all parameters and sample sizes in the case of Model I and Model II, respectively. On the horizontal axis of the graphs the parameters are denoted with an index and are presented in the same order as in Tables 5 - 8. As explained earlier, our analysis focuses on loadings and factor correlations, but the results regarding the thresholds are very similar.

Inspecting both the tables and the figures the main conclusion is that, for both models, the PML estimates and their estimated asymptotic standard errors have very small estimated bias and MSE, both decreasing as the sample size increases. Some secondary comments are that, in the case of Model I and sample size 200, the bias of standard errors is mostly negative. Also, in the case of Model I, the MSE of both estimates and standard errors are relatively higher for the

loadings of the multidimensional item,  $\lambda_{41}$  and  $\lambda_{42}$  (parameter indices 4 and 5 respectively). A similar pattern occurs in Model II regarding the loadings of the second multidimensional indicator,  $\lambda_{10,2}$  and  $\lambda_{10,3}$ , (parameter indices 11 and 12 respectively); the bias and MSE of their estimates and standard errors are relatively higher in absolute values than these of the other loadings.

	True	Estimate			Standard Deviation	Estimated standard error		
		Mean	Bias	MSE		Mean	Bias	MSE
<i>n</i> = 200								
$\lambda_{11}$	0.9	0.9014	0.0014	0.0012	0.0348	0.0344	-0.0004	0.000026
$\lambda_{21}$	0.8	0.8026	0.0026	0.0019	0.0436	0.0410	-0.0026	0.000041
$\lambda_{31}$	0.7	0.7014	0.0014	0.0026	0.0509	0.0498	-0.0011	0.000042
$\lambda_{41}$	0.5	0.4998	-0.0002	0.0057	0.0758	0.0715	-0.0043	0.000244
$\lambda_{42}$	0.6	0.6001	0.0001	0.0058	0.0760	0.0726	-0.0034	0.000225
$\lambda_{52}$	0.7	0.7040	0.0040	0.0035	0.0592	0.0587	-0.0005	0.000059
$\lambda_{62}$	0.8	0.8027	0.0027	0.0035	0.0594	0.0574	-0.0020	0.000079
$\phi_{12}$	0.5	0.5038	0.0038	0.0068	0.0825	0.0804	-0.0022	0.000073
<i>n</i> = 500								
$\lambda_{11}$	0.9	0.9002	0.0002	0.0005	0.0219	0.0220	0.0001	0.000004
$\lambda_{21}$	0.8	0.8007	0.0007	0.0007	0.0262	0.0263	0.0001	0.000005
$\lambda_{31}$	0.7	0.7006	0.0006	0.0010	0.0312	0.0318	0.0006	0.000007
$\lambda_{41}$	0.5	0.4967	-0.0033	0.0020	0.0441	0.0446	0.0005	0.000023
$\lambda_{42}$	0.6	0.6023	0.0023	0.0021	0.0460	0.0450	-0.0010	0.000022
$\lambda_{52}$	0.7	0.6988	-0.0012	0.0015	0.0381	0.0371	-0.0010	0.000010
$\lambda_{62}$	0.8	0.8011	0.0011	0.0013	0.0358	0.0360	0.0002	0.000011
$\phi_{12}$	0.5	0.5034	0.0034	0.0027	0.0521	0.0513	0.0008	0.000012
<i>n</i> = 1000								
$\lambda_{11}$	0.9	0.9005	0.0005	0.0003	0.0161	0.0155	-0.0006	0.000001
$\lambda_{21}$	0.8	0.8003	0.0003	0.0003	0.0182	0.0186	0.0004	0.000001
$\lambda_{31}$	0.7	0.6999	-0.0001	0.0005	0.0218	0.0225	0.0007	0.000002
$\lambda_{41}$	0.5	0.4989	-0.0011	0.0010	0.0317	0.0312	-0.0004	0.000006
$\lambda_{42}$	0.6	0.6007	0.0007	0.0010	0.0322	0.0314	-0.0008	0.000006
$\lambda_{52}$	0.7	0.6988	-0.0012	0.0007	0.0260	0.0263	0.0003	0.000002
$\lambda_{62}$	0.8	0.8012	0.0012	0.0007	0.0267	0.0255	-0.0012	0.000004
$\phi_{12}$	0.5	0.5002	0.0002	0.0014	0.0378	0.0364	-0.0013	0.000005

Table 5: Simulation results for Model I

	True	Estimate			Standard Deviation	Estimated standard error		
		Mean	Bias	MSE		Mean	Bias	MSE
$\lambda_{1,1}$	0.4	0.3987	-0.0013	0.0071	0.0842	0.0849	0.0007	0.000074
$\lambda_{2,1}$	0.5	0.4958	-0.0042	0.0065	0.0807	0.0791	-0.0016	0.000080
$\lambda_{3,1}$	0.6	0.5992	-0.0008	0.0052	0.0718	0.0722	0.0004	0.000079
$\lambda_{4,1}$	0.7	0.6972	-0.0028	0.0043	0.0653	0.0662	0.0009	0.000077
$\lambda_{5,1}$	0.8	0.8012	0.0012	0.0038	0.0614	0.0616	0.0002	0.000077
$\lambda_{6,1}$	0.3	0.3041	0.0041	0.0058	0.0759	0.0752	-0.0007	0.000094
$\lambda_{6,2}$	0.8	0.7994	-0.0006	0.0024	0.0486	0.0470	-0.0016	0.000040
$\lambda_{7,2}$	0.7	0.6998	-0.0002	0.0028	0.0526	0.0537	0.0011	0.000051
$\lambda_{8,2}$	0.6	0.5996	-0.0004	0.0040	0.0631	0.0619	-0.0012	0.000054
$\lambda_{9,2}$	0.5	0.4994	-0.0006	0.0048	0.0691	0.0694	0.0003	0.000055
$\lambda_{10,2}$	0.4	0.4090	0.0090	0.0179	0.1337	0.1373	0.0036	0.002527
$\lambda_{10,3}$	0.5	0.4904	-0.0096	0.0172	0.1309	0.1340	0.0031	0.002486
$\lambda_{11,3}$	0.6	0.6018	0.0018	0.0031	0.0556	0.0561	0.0005	0.000040
$\lambda_{12,3}$	0.7	0.7006	0.0006	0.0024	0.0486	0.0470	-0.0016	0.000039
$\lambda_{13,3}$	0.8	0.7987	-0.0013	0.0015	0.0384	0.0374	-0.0009	0.000029
$\lambda_{14,3}$	0.9	0.8983	-0.0017	0.0008	0.0289	0.0279	-0.0010	0.000019
$\lambda_{15,3}$	0.4	0.3967	-0.0033	0.0046	0.0681	0.0702	0.0021	0.000038
$\phi_{12}$	0.2	0.1994	-0.0006	0.0106	0.1032	0.1036	0.0004	0.000091
$\phi_{13}$	0.5	0.5036	0.0036	0.0060	0.0773	0.0734	-0.0040	0.000077
$\phi_{23}$	0.8	0.7993	-0.0007	0.0028	0.0525	0.0521	-0.0004	0.000066

Table 6: Simulation results for Model II with  $n = 200$

	True	Estimate			Standard Deviation	Estimated standard error		
		Mean	Bias	MSE		Mean	Bias	MSE
$\lambda_{1,1}$	0.4	0.4007	0.0007	0.0029	0.0535	0.0535	0.0000	0.000011
$\lambda_{2,1}$	0.5	0.5011	0.0011	0.0026	0.0510	0.0497	-0.0013	0.000014
$\lambda_{3,1}$	0.6	0.6029	0.0029	0.0020	0.0442	0.0453	0.0011	0.000014
$\lambda_{4,1}$	0.7	0.7006	0.0006	0.0017	0.0417	0.0414	-0.0003	0.000012
$\lambda_{5,1}$	0.8	0.7975	-0.0025	0.0016	0.0401	0.0387	-0.0014	0.000014
$\lambda_{6,1}$	0.3	0.2984	-0.0016	0.0023	0.0483	0.0472	-0.0011	0.000014
$\lambda_{6,2}$	0.8	0.7990	-0.0010	0.0009	0.0294	0.0292	-0.0003	0.000005
$\lambda_{7,2}$	0.7	0.6991	-0.0009	0.0012	0.0345	0.0339	-0.0006	0.000009
$\lambda_{8,2}$	0.6	0.5998	-0.0002	0.0016	0.0406	0.0391	-0.0015	0.000011
$\lambda_{9,2}$	0.5	0.4986	-0.0014	0.0020	0.0452	0.0439	-0.0013	0.000011
$\lambda_{10,2}$	0.4	0.4016	0.0016	0.0069	0.0829	0.0815	-0.0014	0.000149
$\lambda_{10,3}$	0.5	0.4988	-0.0012	0.0067	0.0818	0.0796	-0.0023	0.000157
$\lambda_{11,3}$	0.6	0.5994	-0.0006	0.0013	0.0357	0.0356	-0.0001	0.000006
$\lambda_{12,3}$	0.7	0.7007	0.0007	0.0009	0.0303	0.0298	-0.0005	0.000006
$\lambda_{13,3}$	0.8	0.8006	0.0006	0.0006	0.0236	0.0235	-0.0001	0.000004
$\lambda_{14,3}$	0.9	0.8996	-0.0004	0.0003	0.0177	0.0175	-0.0002	0.000003
$\lambda_{15,3}$	0.4	0.3988	-0.0012	0.0018	0.0425	0.0444	0.0019	0.000009
$\phi_{12}$	0.2	0.2050	0.0050	0.0042	0.0647	0.0653	0.0006	0.000014
$\phi_{13}$	0.5	0.5038	0.0038	0.0021	0.0462	0.0463	0.0001	0.000009
$\phi_{23}$	0.8	0.8023	0.0023	0.0011	0.0331	0.0325	-0.0006	0.000010

Table 7: Simulation results for Model II with  $n = 500$

	True	Estimate			Standard Deviation	Estimated standard error		
		Mean	Bias	MSE		Mean	Bias	MSE
$\lambda_{1,1}$	0.4	0.3998	-0.0002	0.0015	0.0382	0.0379	-0.0002	0.000003
$\lambda_{2,1}$	0.5	0.4992	-0.0008	0.0012	0.0353	0.0352	-0.0001	0.000003
$\lambda_{3,1}$	0.6	0.5985	-0.0015	0.0010	0.0321	0.0321	0.0000	0.000003
$\lambda_{4,1}$	0.7	0.7015	0.0015	0.0008	0.0289	0.0291	0.0002	0.000003
$\lambda_{5,1}$	0.8	0.7995	-0.0005	0.0007	0.0271	0.0273	0.0002	0.000003
$\lambda_{6,1}$	0.3	0.2994	-0.0006	0.0011	0.0329	0.0332	0.0003	0.000003
$\lambda_{6,2}$	0.8	0.8008	0.0008	0.0004	0.0205	0.0204	-0.0001	0.000001
$\lambda_{7,2}$	0.7	0.7002	0.0002	0.0006	0.0253	0.0239	-0.0013	0.000004
$\lambda_{8,2}$	0.6	0.6003	0.0003	0.0008	0.0283	0.0276	-0.0007	0.000003
$\lambda_{9,2}$	0.5	0.4991	-0.0009	0.0010	0.0322	0.0310	-0.0012	0.000003
$\lambda_{10,2}$	0.4	0.4034	0.0034	0.0030	0.0551	0.0559	0.0008	0.000028
$\lambda_{10,3}$	0.5	0.4974	-0.0026	0.0028	0.0533	0.0547	0.0014	0.000030
$\lambda_{11,3}$	0.6	0.6012	0.0012	0.0007	0.0259	0.0251	-0.0008	0.000002
$\lambda_{12,3}$	0.7	0.6999	-0.0001	0.0004	0.0209	0.0211	0.0002	0.000001
$\lambda_{13,3}$	0.8	0.8009	0.0009	0.0003	0.0168	0.0166	-0.0002	0.000001
$\lambda_{14,3}$	0.9	0.8996	-0.0004	0.0002	0.0124	0.0124	0.0000	0.000001
$\lambda_{15,3}$	0.4	0.4015	0.0015	0.0010	0.0320	0.0314	-0.0005	0.000002
$\phi_{12}$	0.2	0.2002	0.0002	0.0021	0.0457	0.0461	0.0004	0.000004
$\phi_{13}$	0.5	0.5007	0.0007	0.0011	0.0329	0.0328	-0.0001	0.000002
$\phi_{23}$	0.8	0.8004	0.0004	0.0005	0.0225	0.0229	0.0004	0.000003

Table 8: Simulation results for Model II with  $n = 1000$

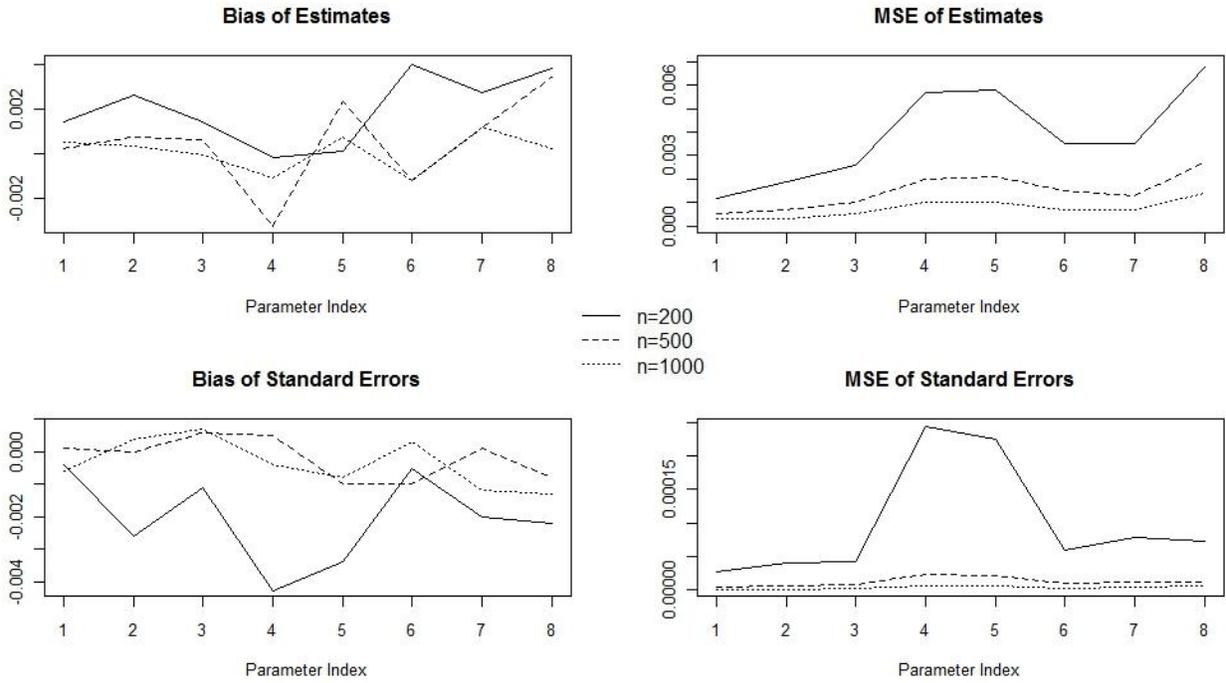


Figure 6: Bias and MSE of estimates and standard errors for all parameters and sample sizes in the case of Model I

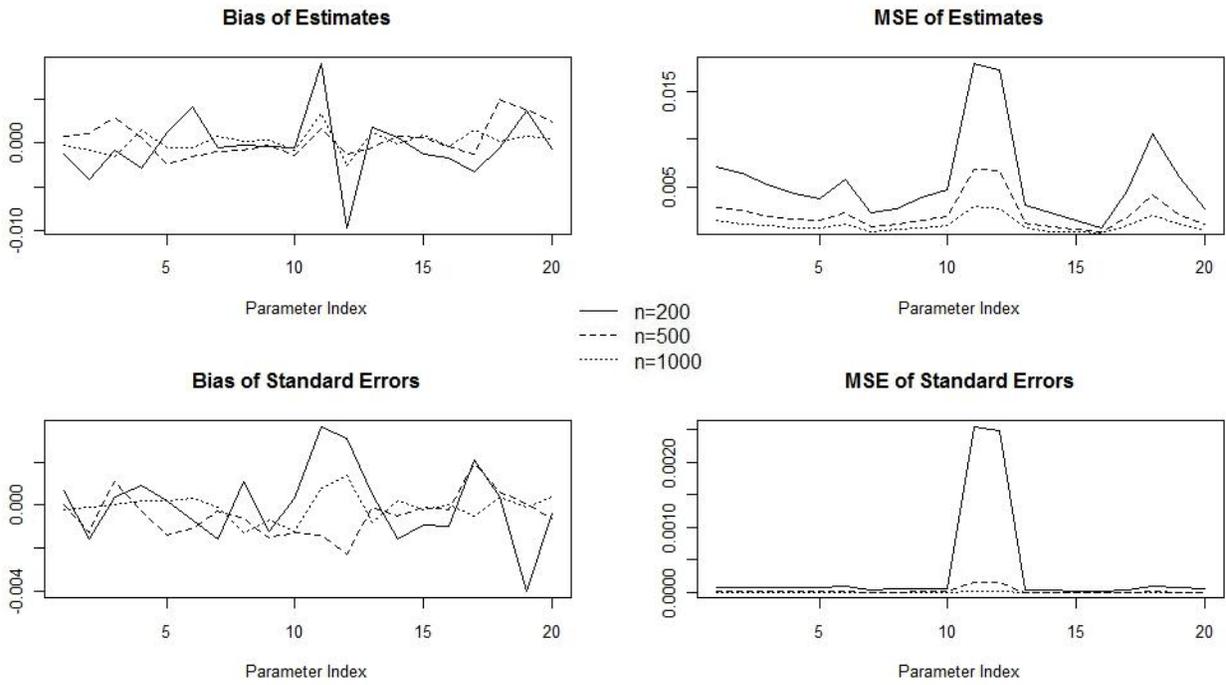


Figure 7: Bias and MSE of estimates and standard errors for all parameters and sample sizes in the case of Model II

## 7 Discussion and Conclusions

There are two main approaches for analysing ordinal variables within the context of factor analysis models, the item response theory (IRT) approach and the underlying response variable (URV) approach. In both approaches, full information maximum likelihood (FIML) estimation cannot be considered as a general method practically since its computational complexity increases rapidly with the model size. In particular, it requires the numerical evaluation of multidimensional integrals whose dimensionality either depends on the number of factors, the case of IRT approach, or on the number of observed variables, the case of URV approach. In the latter case, as the number of observed variables is often large, FIML is not used at all in practice. Instead, three-stage limited information estimation methods have been developed, such as robust unweighted least squares (RULS), robust diagonally weighted least squares (RDWLS), and weighted least squares (WLS). In all these methods, the model parameters are estimated in stage three given the estimates of thresholds and polychoric correlations being estimated in the first and second stages, respectively. Besides, to get the right standard errors, an estimate of the inverse of the asymptotic covariance matrix of polychoric correlations should be used as a weight matrix.

In this paper, we propose a limited information maximum likelihood estimation, namely a pairwise maximum likelihood (PML) method, within the URV approach. That is a computationally general method since it involves the evaluation of up to two-dimensional integrals written in a closed form, regardless of the number of observed variables or factors. Moreover, the PML estimator is asymptotically unbiased, consistent, and normally distributed. The main advantages of our proposed method over the commonly used three-stage limited information estimators (3S-RULS, 3S-RDWLS, 3S-WLS) are that all model parameters are estimated in one single step and there is no need of estimating a weight matrix to obtain the right standard errors.

To compare the estimates and standard errors provided by FIML, PML, and 3S-RULS we use some real data examples. The comparisons are made in an exploratory and confirmatory set-up, with one and two factors in the case of exploratory analysis, and four factors in the case of confirmatory analysis. The main conclusion is that the estimates and standard errors of all methods are fairly close to each other. However, these obtained by PML are on average slightly closer to FIML ones than 3S-RULS are.

To investigate the performance of PML estimator and its standard error we carry out a simulation study where the effect of model size and sample size are studied. In particular, we examine 6 experimental conditions derived by the combination of two models, one smaller with 26 free parameters and one larger with 65 free parameters, and three sample sizes, 200, 500, and 1000. The number of replications with each condition was 1000. The main result of the study is that PML estimates and their standard errors are found to have small bias and mean squared error both decreasing with the sample size.

Our study indicates that PML can be considered as a competitive method to FIML for estimating factor analysis models with ordinal data. Subsequently, further study might needed to examine the efficiency of the proposed method relatively to FIML. Also, PML estimation can be readily extended to the case of full structural equation model with mixed type data.

## Appendix I - The gradient of the bivariate log-likelihood $\ln L(\boldsymbol{\theta}; (x_i, x_j))$

The gradient  $\nabla \ln L(\boldsymbol{\theta}; (x_i, x_j))$ ,  $i = 1, \dots, p$ ,  $j \neq i$ , can be distinguished in three main blocks as follows:

$$\nabla \ln L(\boldsymbol{\theta}; (x_i, x_j)) = \begin{pmatrix} \frac{\partial \ln L(\boldsymbol{\theta}; (x_i, x_j))}{\partial \boldsymbol{\lambda}} \\ \frac{\partial \ln L(\boldsymbol{\theta}; (x_i, x_j))}{\partial \boldsymbol{\varphi}} \\ \frac{\partial \ln L(\boldsymbol{\theta}; (x_i, x_j))}{\partial \boldsymbol{\tau}} \end{pmatrix}.$$

The elements of the subvector  $\frac{\partial \ln L(\boldsymbol{\theta}; (x_i, x_j))}{\partial \boldsymbol{\tau}}$  are the first derivatives with respect to thresholds  $\frac{\partial \ln L(\boldsymbol{\theta}; (x_i^*, x_j^*))}{\partial \tau_{c_i}^{(x_i)}}$  and are given in Olsson (1979, eq. (13)). In particular,

$$\frac{\partial \ln L(\boldsymbol{\theta}; (x_i, x_j))}{\partial \tau_{c_i}^{(x_i)}} = \sum_{c_j=1}^{m_j} \left( \frac{n_{c_i c_j}^{(x_i x_j)}}{\pi_{c_i c_j}^{(x_i x_j)}} - \frac{n_{(c_i+1) c_j}^{(x_i x_j)}}{\pi_{(c_i+1) c_j}^{(x_i x_j)}} \right) \frac{\partial \pi_{c_i c_j}^{(x_i x_j)}}{\partial \tau_{c_i}^{(x_i)}},$$

where

$$\frac{\partial \pi_{c_i c_j}^{(x_i x_j)}}{\partial \tau_{c_i}^{(x_i)}} = \phi_1(\tau_{c_i}^{(x_i)}) \left[ \Phi_1 \left( \frac{\tau_{c_j}^{(x_j)} - \rho_{x_i x_j} \tau_{c_i}^{(x_i)}}{\sqrt{1 - \rho_{x_i x_j}^2}} \right) - \Phi_1 \left( \frac{\tau_{c_j-1}^{(x_j)} - \rho_{x_i x_j} \tau_{c_i}^{(x_i)}}{\sqrt{1 - \rho_{x_i x_j}^2}} \right) \right],$$

and  $\phi_1$  and  $\Phi_1$  are the standard univariate normal density and distribution respectively. To find the partial derivatives with respect to  $\boldsymbol{\lambda}$  and  $\boldsymbol{\varphi}$  we use the chain rule, i.e.

$$\begin{aligned} \frac{\partial \ln L(\boldsymbol{\theta}; (x_i, x_j))}{\partial \boldsymbol{\lambda}} &= \frac{\partial \ln L(\boldsymbol{\theta}; (x_i, x_j))}{\partial \rho_{x_i x_j}} \frac{\partial \rho_{x_i x_j}}{\partial \boldsymbol{\lambda}} \text{ and} \\ \frac{\partial \ln L(\boldsymbol{\theta}; (x_i, x_j))}{\partial \boldsymbol{\varphi}} &= \frac{\partial \ln L(\boldsymbol{\theta}; (x_i, x_j))}{\partial \rho_{x_i x_j}} \frac{\partial \rho_{x_i x_j}}{\partial \boldsymbol{\varphi}}. \end{aligned}$$

The partial derivative with respect to  $\rho_{x_i x_j}$  is

$$\frac{\partial \ln L(\boldsymbol{\theta}; (x_i, x_j))}{\partial \rho_{x_i x_j}} = \sum_{c_i=1}^{m_i} \sum_{c_j=1}^{m_j} \frac{n_{c_i c_j}^{(x_i x_j)}}{\pi_{c_i c_j}^{(x_i x_j)}} \frac{\partial \pi_{c_i c_j}^{(x_i x_j)}}{\partial \rho_{x_i x_j}},$$

where  $\frac{\partial \pi_{c_i c_j}^{(x_i x_j)}}{\partial \rho_{x_i x_j}}$  is given in Olsson (1979, eqs (8)) and it is

$$\frac{\partial \pi_{c_i c_j}^{(x_i x_j)}}{\partial \rho_{x_i x_j}} = \phi \left( \tau_{c_i}^{(x_i)}, \tau_{c_j}^{(x_j)}; \rho_{x_i x_j} \right) - \phi \left( \tau_{c_i}^{(x_i)}, \tau_{c_j-1}^{(x_j)}; \rho_{x_i x_j} \right) - \phi \left( \tau_{c_i-1}^{(x_i)}, \tau_{c_j}^{(x_j)}; \rho_{x_i x_j} \right) + \phi \left( \tau_{c_i-1}^{(x_i)}, \tau_{c_j-1}^{(x_j)}; \rho_{x_i x_j} \right).$$

The partial derivative of  $\rho_{x_i x_j}$  with respect to  $\boldsymbol{\lambda}$  in vector - matrix form is as follows

$$\frac{\partial \rho_{x_i x_j}}{\partial \boldsymbol{\lambda}} = \frac{\partial \rho_{x_i x_j}}{\partial \boldsymbol{\lambda}_i} \frac{\partial \boldsymbol{\lambda}_i}{\partial \boldsymbol{\lambda}} + \frac{\partial \rho_{x_i x_j}}{\partial \boldsymbol{\lambda}_j} \frac{\partial \boldsymbol{\lambda}_j}{\partial \boldsymbol{\lambda}} = \boldsymbol{\lambda}_j \cdot \Phi \frac{\partial \boldsymbol{\lambda}_i}{\partial \boldsymbol{\lambda}} + \boldsymbol{\lambda}_i^{(x)} \Phi \frac{\partial \boldsymbol{\lambda}_j}{\partial \boldsymbol{\lambda}},$$

where the terms of type  $\frac{\partial \boldsymbol{\lambda}_i}{\partial \boldsymbol{\lambda}}$  are matrices of zeroes and ones with  $k$  rows (as many as the number of columns of matrix  $\Lambda$ ) and as many columns as the size of  $\boldsymbol{\lambda}$  (the number of free parameters in matrix  $\Lambda$ ).

The partial derivative of  $\rho_{x_i x_j}$  with respect to  $\boldsymbol{\varphi}$  in vector - matrix form is

$$\frac{\partial \rho_{x_i x_j}}{\partial \varphi} = \frac{\partial \rho_{x_i x_j}}{\partial \Phi} \frac{\partial \Phi}{\partial \varphi} = \left( \lambda_{i \cdot}^{(x)} \right)' \lambda_{j \cdot}^{(x)} \frac{\partial \Phi}{\partial \varphi},$$

where  $\frac{\partial \Phi}{\partial \varphi}$  is a matrix of zeroes and ones with  $r^2$  rows (as many as the total number of elements of matrix  $\Phi$ ) and as many columns as the size of vector  $\varphi$  (the number of free non-redundant parameters in  $\Phi$ ). Note that  $\frac{\partial \Phi}{\partial \varphi}$  is sometimes (more appropriately) denoted as  $\frac{\partial \text{vec}(\Phi)}{\partial \varphi'}$ , where *vec* is the function transforming a  $m \times n$  matrix to a  $mn \times 1$  vector by stacking its columns one underneath the other.

## Appendix II - The items of Science & Technology data

1. Science and technology are making our lives healthier, easier and more comfortable.
2. Scientific and technological research cannot play an important role in protecting the environment and repairing it.
3. The application of science and new technology will make work more interesting.
4. Thanks to science and technology, there will be more opportunities for the future generations.
5. New technology does not depend on basic scientific research.
6. Scientific and technological research do not play an important role in industrial development.
7. The benefits of science are greater than any harmful effects it may have.

## Appendix III - The items of Relationship Learning data

### Collaborative Commitment

1. To what degree do you discuss company goals with the other party in this relationship?
2. To what degree are these goals developed through joint analysis of potentials?
3. To what degree are these goals formalized in a joint agreement or contract?
4. To what degree are these goals implemented in day-to-day work?
5. To what degree have you developed measures that capture performance related to these goals?

### Internal Complexity

6. The products we exchange are generally very complex.
7. There are many operating units involved from both organizations.
8. There are many contract points between different departments or professions between the two organizations.

### Relational Trust

9. I believe the other organization will respond with understanding in the event of problems.
10. I trust that the other organization is able to fulfill contractual agreements.
11. We trust that the other organization is competent at what they are doing.
12. There is a general agreement in my organization that the other organization is trustworthy.
13. There is a general agreement in my organization that the contact people on the other organization are trustworthy.

### Environmental Uncertainty

14. End-users needs and preferences change rapidly in our industry.
15. The competitors in our industry frequently make aggressive moves to capture market share.
16. Crises have caused some of our competitors to shut down or radically change the way they operate.
17. It is very difficult to forecast where the technology will be in the next 2-3 years in our industry.
18. In recent years, a large number of new product ideas have been made possible through technological breakthroughs in our industry.

## References

- Bartholomew, D., Knott, M., & Moustaki, I. (2011). *Latent Variable Models and Factor Analysis: A Unified Approach*. John Wiley series in Probability and Statistics, 3rd ed.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an em algorithm. *Psychometrika*, *46*(4), 443–459.
- Chan, W., & Bentler, P. (1998). Covariance structure analysis of ordinal ipsative data. *Psychometrika*, *63*, 369–399.
- Cox, D., & Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, *91*(3), 729–737.
- de Leon A.R. (2005). Pairwise likelihood approach to grouped continuous model and its extension. *Statistics & Probability Letters*, *75*, 49–57.
- Forero, A., C. G. Maydeu-Olivares, & Gallardo-Pujol, D. (2011). Factor analysis with ordinal indicators: A monte carlo study comparing dwls and uls estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 625–641.
- Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of irt graded response models: Limited versus full information methods. *Psychological Methods*, *14*, 275–299.
- Huber, P., Ronchetti, E., & Victoria-Feser, M.-P. (2004). Estimation of generalized latent linear variable models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *66*(4), 893–908.
- Joe, H., & Maydeu-Olivares, A. (2010). A general family of limited information goodness-of-fit statistics for multinomial data. *Psychometrika*, *75*(3), 393–419.
- Jöreskog, K. (1990). New developments in lisrel: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity*, *24*, 387–404.
- Jöreskog, K. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, *59*, 381–389.
- Jöreskog, K. (2002). Structural equation modeling with ordinal variables using lisrel. <http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf>.
- Jöreskog, K., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, *36*(3), 347–387.
- Jöreskog, K., & Sörbom, D. (1996). *LISREL 8 User's Reference Guide*. Chicago, IL: Scientific Software International.
- Karlheinz, & Melich (1992). Euro-barometer 38.1: Consumer protection and perceptions of science and technology.
- Lee, S., Poon, W., & Bentler, P. (1990). Full maximum likelihood analysis of structural equation models with polytomous variables. *Statistics and Probability Letters*, *9*, 91–97.

- Lee, S., Poon, W., & Bentler, P. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika*, *57*, 89–105.
- Lindsay, B. (1988). Composite likelihood methods. *Contemporary Mathematics*, *80*, 221–239.
- Liu, J. (2007). *Multivariate ordinal data analysis with pairwise likelihood and its extension to SEM*. Ph.D. thesis, University of California, Los Angeles, <http://theses.stat.ucla.edu/72/Thesis>
- Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores*. Addison-Wesley Publishing Co.
- Maydeu-Olivares, A., & Joe, H. (2005). Limited- and full-information estimation and goodness-of-fit testing in  $2^n$  contingency tables: A unified approach. *Journal of the American Statistical Association*, *100*, 1009–1020.
- Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, *71*(4), 713–732.
- Muraki, E. (1990). Fitting a polytomous item response model to likert-type data. *Applied Psychological Measurement*, *14*, 59–71.
- Muraki, E., & Carlson, E. (1995). Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement*, *19*, 73–90.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered, categorical, and continuous latent variables indicators. *Psychometrika*, *49*, 115–132.
- Muthén, B., du Toit, S., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. [http://gseis.ucla.edu/faculty/muthen/articles/Article\\_075.pdf](http://gseis.ucla.edu/faculty/muthen/articles/Article_075.pdf).
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus 6 [Computer Software]*. Muthén and Muthén, Los Angeles.
- Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*(4), 443–460.
- Poon, W. Y., & Lee, S. Y. (1987). Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients. *Psychometrika*, *52*, 409–430.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, No. 17.
- Sammel, M., Ryan, L., & Legler, J. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *59*(3), 667–678.
- Schervish, M. (1984). Algorithm as 195: Multivariate normal probabilities with errors bound. *Applied Statistics*, *3*, 81–94.
- Schilling, S., & Bock, R. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*, *70*(3), 533–555.

- Selnes, F., & Sallis, J. (2003). Promoting relationship learning. *Journal of Marketing*, *67*, 80–95.
- Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*, 393–408.
- Team, R. D. C. (2005). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org>.
- Varin, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis*, *92*(1), 1–28.
- Varin, C., Reid, N., & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, *21*(1), 1–41.
- Varin, C., & Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, *92*(3), 519–528.
- Yang-Wallentin, F., Jöreskog, K., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*, 392–423.