

Vassilis G.S. Vasdekis, Dimitris Rizopoulos and [Irimi Moustaki](#)

Swarm intelligence: when uncertainty meets conflict

Article (Accepted version)
(Refereed)

Original citation:

Vasdekis, Vassilis G. S., Rizopoulos, Dimitris and Moustaki, Irimi (2014) Weighted pairwise likelihood estimation for a general class of random effects models. [Biostatistics](#), Online . ISSN 1465-4644 (In Press)

DOI: [10.1093/biostatistics/kxu018](https://doi.org/10.1093/biostatistics/kxu018)

© 2014 The Author

This version available at: <http://eprints.lse.ac.uk/56733/>

Available in LSE Research Online: August 2014

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

Weighted pairwise likelihood estimation for a general class of random effects models

Vassilis G.S Vasdekis¹

Athens University of Economics and Business

Dimitris Rizopoulos

Erasmus University Medical Center

Irini Moustaki

London School of Economics & Political Science

¹Correspondence Author: Department of Statistics, Greece, email: vasdekis@aueb.gr

Abstract

Models with random effects/latent variables are widely used for capturing unobserved heterogeneity in multilevel/hierarchical data and account for associations in multivariate data. The estimation of those models becomes cumbersome as the number of latent variables increases due to high-dimensional integrations involved. Composite likelihood is a pseudo-likelihood that combines lower-order marginal or conditional densities such as univariate and/or bivariate; it has been proposed in the literature as an alternative to full maximum likelihood estimation. We propose a weighted pairwise likelihood estimator based on estimates obtained from separate maximizations of marginal pairwise likelihoods. The derived weights minimize the total variance of the estimated parameters. The proposed weighted estimator is found to be more efficient than the one that assumes all weights to be equal. The methodology is applied to a multivariate growth model for binary outcomes in the analysis of four indicators of schistosomiasis before and after drug administration.

Keywords composite likelihood; generalized linear latent variable models; longitudinal data; categorical data.

1. INTRODUCTION

Models with random effects known as mixed effects models or multilevel models, as well as factor analysis models and structural equation models (SEM) are widely used in Social Sciences, Health Sciences and Economics for analyzing associations among variables in cross-sectional and longitudinal studies. Random effects are unobserved random variables employed to capture associations and heterogeneity above the one explained by explanatory variables. In cross-sectional studies, random effects are often used with nested data (e.g students (low level) nested within schools (high level)) to allow for higher-level heterogeneity as well as higher-level covariates. In multivariate longitudinal studies such as the one that will be examined here, four indicators/items/variables of schistosomiasis are measured on children in Tanzania at three occasions before and after the administration of drugs. In this set up, item-specific correlated random effects are used to account for the serial correlation of the same variables across time and correlations of the four indicators within time.

Estimation of random effects and factor analysis models entail heavy integrations that make the use of full information maximum likelihood (FIML) infeasible in practice. Composite likelihood estimation provides a feasible alternative to FIML. It simplifies the likelihood to be maximised and provides estimates with desirable statistical properties. For an excellent review of recent methodological developments and applications of composite

likelihood methods see Varin, Reid, and Firth (2011) and the special issue of *Statistica Sinica* published in 2011. Composite likelihood estimation is based upon lower-order densities - marginal or conditional likelihoods - (Lindsay, 1988; Arnold & Strauss, 1991; Geys, Molenberghs, & Ryan, 1999; Cox & Reid, 2004; Varin, 2008).

In particular, composite likelihood estimation has been shown to work satisfactorily and be computationally attractive over FIML for SEM for binary, ordinal and ranking variables when the underlying variable approach (each categorical variable is assumed to be a manifestation of a normally distributed variable) is adopted (Jöreskog & Moustaki, 2001; Liu, 2007; Katsikatsou, Moustaki, Yang-Wallentin, & Jöreskog, 2012; Katsikatsou, 2013) and for factor analysis models for longitudinal data where both latent variables and random effects are used to account for dependencies (Vasdekis, Cagnone, & Moustaki, 2012). In all aforementioned papers, composite likelihood is defined as the sum of all log pairwise likelihoods. Furthermore, Chan and Bentler (1998) and Fieuws and Verbeke (2006) used the composite likelihood for a covariance structure analysis for ranking data and for estimating mixed effects models for multivariate longitudinal outcomes respectively. In their implementation of the composite likelihood, each pairwise likelihood is maximized separately and the final parameter estimates are obtained as a simple average of the estimates produced by the separate bivariate maximizations.

However, in some cases, lower-order margins provide no information for some of the model parameters and the amount of information available for estimating a single parameter may vary according to data availability or data characteristics. That provided the motivation to propose a weighted estimator for a general class of random effects and factor analysis models under a pairwise likelihood estimation. The proposed estimator is shown through simulations and the data application to have greater efficiency.

The paper is organized as follows. Section 2 provides a description of the data, Section 3 presents a general model framework that includes both random effects and latent variables; Section 4 discusses composite likelihood estimation and the weighted estimator and Section 5 presents results from simulations that show the effectiveness of the proposed methodology. The results and discussion of the data obtained from the Schistosomiasis Control Initiative based at Imperial College London are given in Section 6 and the paper concludes in Section 7.

2. EXAMPLE: SCHISTOSOMIASIS DATA

Schistosomiasis remains one of the most prevalent parasitic diseases in developing countries. After malaria, schistosomiasis is the most important tropical disease in terms of human morbidity with significant economic and public health consequences. In fact, Schistosomiasis Control Initiative (SCI) implements and evaluates control of schistosomi-

asis and thus invests in process monitoring, drug evaluation and morbidity measurements throughout the programmes in each supported country. The data analyzed here are obtained from SCI (Fenwick et al., 2009) and they are part of longitudinal morbidity surveys on children over 4 years in Tanzania during 2005 (n=2157, no intervention), 2006 (n=1048, where 1 mass drug administration (MDA) is evaluated) and 2008 (n=717, 2 MDA are evaluated). Two of the four variables being analyzed in the present study are blood in urine and pain when urinating which are self-reported symptoms by children when they were asked whether they had felt any of these, during the last 2 weeks when surveys took place. The other two variables are: ‘Do you know what schistosomiasis is?’ and blood urine as detected by reagent test strips. Recent epidemiological studies (Clements, Brooker, Nyandindi, Fenwick, & Blair, 2008; Koukounari et al., 2006) have suggested the reagent strips to be a good indicator for urinary schistosomiasis in endemic settings. All responses were binary and coded ‘1’ for a Yes response and ‘0’ for a No response. Between baseline and follow-ups, there were children missing either because they were lost to follow up or some schools were not revisited and so data collection did not happen in these, due to logistical constraints. The aim of our analysis is twofold, first to study the measurement properties of the four indicators and second to study simultaneously changes in the symptoms, presence and knowledge of schistosomiasis before and after the drug administration. The model used is a multivariate growth model that accounts for the associations of the four items within and across time.

3. GENERALIZED LINEAR LATENT VARIABLE MODEL

We adopt the generalized linear latent variable model (GLLVM) specification as described in Skrondal and Rabe-Hesketh (2004). The model framework accommodates models with random effects and latent variables. Let y_{jit} be the observation of individual j on item i at time t . To accommodate different types of responses in a unified framework, we postulate a multivariate generalized linear mixed effects model. In particular, the conditional distribution of y_{jit} given a vector of latent variables $\boldsymbol{\eta}_{jit}$ and covariates \boldsymbol{x}_{jit} is assumed to be a member of the exponential family, with linear predictor ν_{jit} given by

$$\nu_{jit} = \boldsymbol{x}_{jit}^{\top} \boldsymbol{\beta}_{it} + \boldsymbol{z}_{jit}^{\top} \boldsymbol{\eta}_{jit}, j = 1, \dots, n, i = 1, \dots, p, t = 1, \dots, T,$$

where \boldsymbol{x}_{jit} denotes a design vector of possibly time-dependent covariates with a corresponding vector of fixed effects parameters $\boldsymbol{\beta}_{it}$. The design vector \boldsymbol{z}_{jit} is associated with the vector of latent variables $\boldsymbol{\eta}_{jit}$ which are assumed to be independent between sample units. The latent variable vector can include latent variables that depend on individuals and items, on individuals and times or on individuals only but not on all three since those

would not be identified. The conditional expectation of the response y given covariates \mathbf{x} and latent variables $\boldsymbol{\eta}$ is linked to the linear predictor ν via a link function $g_{it}(\cdot)$:

$$g_{it}\{E(y_{jit} \mid \mathbf{x}_{jit}, \boldsymbol{\eta}_{jit})\}.$$

For the remaining of the paper, we will use the term latent variable to indicate either a random effect or a latent variable.

A special case of the GLLVM framework is the multivariate growth model for binary variables that will be used for analysing the schistosomiasis data. The model has been proposed by Fieuws and Verbeke (2006) for multivariate continuous data and it is extended here to multivariate binary outcomes. Assuming that y_{jit} is now binary, related to time and a set of L covariates, the model is:

$$\begin{aligned} \text{logit}\{P(y_{jit} = 1 \mid \mathbf{x}_{jit}, \eta_{ji,0}, \eta_{ji,1})\} &= \beta_{i,0} + \beta_{i,1}x_{jit,1} + \sum_{\ell=2}^L \beta_{i,\ell} x_{jit,\ell} \\ &+ \eta_{ji,0} + \eta_{ji,1}x_{jit,1}, \end{aligned} \quad (1)$$

where $\text{logit}(x) = \log(x/(1-x))$, $x_{jit,1}$ indicates time and the latent variables $\eta_{ji,0}$ and $\eta_{ji,1}$ are item-individual, multivariate normals with mean zero and unrestricted covariance matrix, representing item-individual variation at the intercept and at the slope or growth level respectively. The covariances among the latent variables capture the associations of items within and across time. The η s are assumed to be independent from the observed covariates x . This modeling approach can deal with both balanced and unbalanced data and also with unequal time spaced measurements.

Maximum likelihood is often employed to estimate the parameters of the GLLVM. To define the observed data likelihood we typically assume that the vector of latent variables, $\boldsymbol{\eta}_{j^{**}}$, follows a distribution $\mathcal{F}_{\theta_\eta}$ parameterized by $\boldsymbol{\theta}_\eta$. A standard choice is the normal distribution with mean zero and covariance matrix D , in which case $\boldsymbol{\theta}_\eta = \text{vech}(D)$ denotes the unique elements of D . Under the assumption of conditional independence, the log-likelihood contribution of the j th sample unit is:

$$\begin{aligned} \ell_j(\boldsymbol{\theta}) = \log p(\mathbf{y}_{j^{**}} \mid \mathbf{x}_{j^{**}}; \boldsymbol{\theta}) &= \log \int p(\mathbf{y}_{j^{**}} \mid \mathbf{x}_{j^{**}}, \boldsymbol{\eta}_{j^{**}}; \boldsymbol{\theta}_y) p(\boldsymbol{\eta}_{j^{**}}; \boldsymbol{\theta}_\eta) d\boldsymbol{\eta}_{j^{**}} \\ &= \log \int \prod_t^T \prod_i^p p(y_{jit} \mid \mathbf{x}_{jit}, \boldsymbol{\eta}_{jit}; \boldsymbol{\theta}_y) p(\boldsymbol{\eta}_{j^{**}}; \boldsymbol{\theta}_\eta) d\boldsymbol{\eta}_{j^{**}}, \end{aligned} \quad (2)$$

where $\mathbf{y}_{j^{**}}$ and $\mathbf{x}_{j^{**}}$ as in $\boldsymbol{\eta}_{j^{**}}$ denote the vector of responses and covariates respectively with indices defining all possible values of i and t respectively. Expression (2) defines a general latent variable model for the y items given the latent variables $\boldsymbol{\eta}$ and the covariates \mathbf{x} . The model consists of the measurement model $p(\mathbf{y}_{j^{**}} \mid \mathbf{x}_{j^{**}}, \boldsymbol{\eta}_{j^{**}}; \boldsymbol{\theta}_y)$ which

describes how the items y measure η and the structural model $p(\boldsymbol{\eta}_{j^{**}}; \boldsymbol{\theta}_\eta)$ which specifies the distribution of the latent variables. The q -dimensional parameter vector $\boldsymbol{\theta}$ is written as $\boldsymbol{\theta}^\top = (\boldsymbol{\theta}_y^\top, \boldsymbol{\theta}_\eta^\top)$ where $\boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_\eta$ denote the parameters of the measurement part and the structural part of the model respectively. The assumption of conditional independence implies that the latent variables and covariates account for the interdependencies among the observed variables. This greatly facilitates the computation of the likelihood because, each density $p(y_{jit} | \mathbf{x}_{jit}, \boldsymbol{\eta}_{jit}; \boldsymbol{\theta}_y)$ is a member of the exponential family or of the extended exponential family in the case of ordinal responses.

Under the normality assumption for the latent variables, the integrals in the definition of the log-likelihood in (2) do not have, in general, a closed-form solution and as a result the location of the MLEs of GLLVMs requires a combination of numerical integration and optimization. For the maximization of the log-likelihood function standard algorithms can be utilized, such as the Expectation-Maximization algorithm (E-M) (Dempster, Laird, & Rubin, 1977) or the Newton-Raphson algorithm (Lange, 2004). For the numerical approximation of the integrals in (2) various simulation techniques provide powerful tools. Standard choices are Gaussian quadrature rules (Bock & Aitkin, 1981; Schilling & Bock, 2005; Press, Teukolsky, Vetterling, & Flannery, 2007), simple Monte Carlo methods (Sammel, Ryan, & Legler, 1997) but also advanced sampling algorithms such as the MCMC (importance sampling, rejection sampling), or Laplace approximations as described in Huber, Ronchetti, and Victoria-Feser (2004) and the work related with the ADMB project that combines Laplace approximation and MCMC sampling algorithms (Fournier et al., 2012). Those techniques work efficiently for solving high dimensional problems.

The computational complexity of a GLLVM increases exponentially with the increase of the latent variables and therefore composite likelihood estimation can reduce the dimensionality of the problem at the expense of loss of some information. Under composite likelihood estimation, and more specifically pairwise likelihood, an approximated method from the ones mentioned above still needs to be applied but to a lower dimensional problem.

4. PROPOSED WEIGHTED PAIRWISE ESTIMATOR

A pairwise likelihood replaces the full likelihood in (1) by a set of bivariate likelihoods. Let us denote by \mathbf{y}_{ji^*} the T -dimensional vector of all observations on time for

subject j and item i . The pairwise log-likelihood for a random sample of size n is

$$p\ell(\boldsymbol{\theta}) = \sum_{j=1}^n p\ell_j(\boldsymbol{\theta}) = \sum_{j=1}^n \sum_{i=1}^{p-1} \sum_{k=i+1}^p \log p(\mathbf{y}_{ji*}, \mathbf{y}_{jk*}, \boldsymbol{\theta}) \quad (3)$$

where the pairwise likelihood estimator $\hat{\boldsymbol{\theta}}_{PL}$ is a consistent estimator of $\boldsymbol{\theta}$ under suitable regularity conditions (Arnold & Strauss, 1991).

We show that $\hat{\boldsymbol{\theta}}_{PL}$ is asymptotically a weighted estimator of estimates obtained from separate maximizations of pairwise log-likelihoods each one summed across subjects. Let us denote by r , one of N possible combinations of item i and item k above. The pairwise likelihood $p\ell(\boldsymbol{\theta})$ becomes $p\ell(\boldsymbol{\theta}) = \sum_{r=1}^N f_r(\boldsymbol{\theta})$ where $f_r(\boldsymbol{\theta}) = \sum_{j=1}^n \log p(\mathbf{y}_{ji*}, \mathbf{y}_{jk*}, \boldsymbol{\theta})$. Let us also denote by $\hat{\boldsymbol{\theta}}_r$ the estimator resulting from maximizing $f_r(\boldsymbol{\theta})$. For a large sample size n , a Taylor series expansion of $\frac{\partial f_r(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ around $\hat{\boldsymbol{\theta}}_r$ gives

$$\frac{\partial f_r(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial f_r(\hat{\boldsymbol{\theta}}_r)}{\partial \boldsymbol{\theta}} + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_r)^\top \frac{\partial^2 f_r(\hat{\boldsymbol{\theta}}_r)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top},$$

where the first component on the right hand side is zero.

By definition of $\hat{\boldsymbol{\theta}}_{PL}$, $\partial p\ell(\hat{\boldsymbol{\theta}}_{PL})/\partial \boldsymbol{\theta} = \sum_{r=1}^N \partial f_r(\hat{\boldsymbol{\theta}}_{PL})/\partial \boldsymbol{\theta} = 0$. Therefore, $\sum_{r=1}^N (\hat{\boldsymbol{\theta}}_{PL} - \hat{\boldsymbol{\theta}}_r)^\top \partial^2 f_r(\hat{\boldsymbol{\theta}}_r)/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top = 0$, from which we obtain,

$$\hat{\boldsymbol{\theta}}_{PL} = \left(\sum_{s=1}^N \frac{\partial^2 f_s(\hat{\boldsymbol{\theta}}_s)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right)^{-1} \sum_{r=1}^N \frac{\partial^2 f_r(\hat{\boldsymbol{\theta}}_r)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \hat{\boldsymbol{\theta}}_r = \sum_{r=1}^N A_r \hat{\boldsymbol{\theta}}_r,$$

where

$$A_r = \left(\sum_{s=1}^N \frac{\partial^2 f_s(\hat{\boldsymbol{\theta}}_s)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right)^{-1} \frac{\partial^2 f_r(\hat{\boldsymbol{\theta}}_r)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}, \quad r = 1, \dots, N.$$

The aim of the paper is to find new weights A_r that satisfy some optimality criterion.

To maximize each $f_r(\boldsymbol{\theta})$ separately, we rewrite the pairwise loglikelihood for pairs of items given in (3) as

$$p\ell_j(\boldsymbol{\theta}) = \sum_{i=1}^{p-1} \sum_{k=i+1}^p \log p(\mathbf{y}_{ji*}, \mathbf{y}_{jk*}; \boldsymbol{\theta}_{i,k}), \quad (4)$$

where $\boldsymbol{\theta}_{i,k}$ denotes the q -dimensional model parameter vector indexed by the pair of items $\{i, k\}$. Similarly, for pairs of time points, the pairwise log-likelihood function is:

$$p\ell_j(\boldsymbol{\theta}) = \sum_{t=1}^{T-1} \sum_{s=t+1}^T \log p(\mathbf{y}_{j*t}, \mathbf{y}_{j*s}; \boldsymbol{\theta}_{t,s}), \quad (5)$$

where $\boldsymbol{\theta}_{t,s}$ denotes the q -dimensional parameter vector for the pair of time points $\{t, s\}$.

The choice between (4) or (5) is according to which of the two representations reduces the computational complexity. The separate maximizations give as many consistent estimates for a single parameter as the total number of possible pairs. Fieuw and Verbeke (2006) suggested to take the simple average as the final estimate. Although this is the simplest solution, it may not lead to the most efficient estimator.

Let us denote by $\hat{\boldsymbol{\Theta}}^\top = (\hat{\boldsymbol{\theta}}_1^\top, \hat{\boldsymbol{\theta}}_2^\top, \dots, \hat{\boldsymbol{\theta}}_N^\top)$, the Nq -dimensional vector with elements the estimates $\hat{\boldsymbol{\theta}}_r$, $r = 1, \dots, N$ obtained from separate maximizations of (4) or (5). The asymptotic covariance matrix of $\hat{\boldsymbol{\Theta}}$ has the form (Fieuw & Verbeke, 2006)

$$V = \frac{1}{n} J^{-1} K J^{-1}, \quad (6)$$

where J is a $Nq \times Nq$ block diagonal matrix and K a matrix of the same dimensions with each element of J and K is of dimension $q \times q$

$$J_{rr} = -\frac{1}{n} \sum_{j=1}^n E \left(\frac{\partial^2 p l_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}_r^\top} \right), \quad K_{ru} = \frac{1}{n} \sum_{j=1}^n E \left(\frac{\partial p l_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r} \frac{\partial p l_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_u^\top} \right),$$

$r = 1, \dots, N$, $u = 1, \dots, N$.

The weighted estimator is of the form $\hat{\boldsymbol{\theta}}_{WPL} = A \hat{\boldsymbol{\Theta}}$ where A is a $q \times Nq$ block matrix of weights. A is obtained by minimizing the total variance of the weighted estimator, $\hat{\boldsymbol{\theta}}_{WPL}$, given by the trace of $AV A^\top$ where V is the asymptotic covariance matrix given in (6). Matrix A is of the form $A = (A_1, A_2, \dots, A_N)$ where each element A_r of dimension $q \times q$ gives the weight for the estimate $\hat{\boldsymbol{\theta}}_r$. By denoting with $A_{r,ij}$ and $\hat{\theta}_{r,j}$ the i, j element of matrix A_r and the j th component of $\hat{\boldsymbol{\theta}}_r$ respectively, each parameter θ_i , $i = 1, \dots, q$, will be estimated by $\hat{\boldsymbol{\theta}}_{WPL,i} = \sum_{r=1}^N \sum_{j=1}^q A_{r,ij} \hat{\theta}_{r,j}$. To guarantee consistency, the minimization is done under the constraint that the weights for each i sum to 1, that is $\sum_{r=1}^N \sum_{j=1}^q A_{r,ij} = 1$.

Let us denote with $\boldsymbol{\alpha}$ the q_A -dimensional vector of unique weights contained in A . Vector $\boldsymbol{\alpha}$ is linked with $vec(A)$ through a matrix Ψ which determines the position of each element of $\boldsymbol{\alpha}$ in $vec(A)$, as $vec(A) = \Psi \boldsymbol{\alpha}$. The sum-to-one constraints are now applied to the elements of $\boldsymbol{\alpha}$ by defining a $q \times q_A$ matrix M such that $M \boldsymbol{\alpha} = \mathbf{1}_q$. To understand the nature of $\boldsymbol{\alpha}$, Ψ and M , we give two simple examples. Let us consider two estimates of $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$ denoted by $\hat{\boldsymbol{\theta}}_1 = (\hat{\theta}_{11}, \hat{\theta}_{12})^\top$ and $\hat{\boldsymbol{\theta}}_2 = (\hat{\theta}_{21}, \hat{\theta}_{22})^\top$. The weight matrix A has the form $A = (A_1, A_2)$ where A_1 and A_2 are 2×2 weight matrices for $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ respectively. In the first example, we assume that all elements of $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ can be used to estimate θ_1 and θ_2 and that all elements of A are unique and $\boldsymbol{\alpha} = vec(A)$. In this case $\Psi = I_8$. The elements of the first row of A should add up to one and the same is true for the elements of the second row of A . Therefore, the form of matrix M will be

$M = (I_2, I_2, I_2, I_2)$. As a second example, let us assume that θ_1 is estimated using $\hat{\theta}_{11}$ and $\hat{\theta}_{21}$ and θ_2 is estimated using $\hat{\theta}_{12}$ and $\hat{\theta}_{22}$. Then matrices A_1 and A_2 are diagonal and $\boldsymbol{\alpha} = (A_{1,11}, A_{1,22}, A_{2,11}, A_{2,22})^\top$ and $\Psi = I_2 \otimes \text{diag}(\mathbf{e}_1, \mathbf{e}_2)$ where \mathbf{e}_k is a 2×1 vector with 1 in the k -th place and 0 elsewhere. Matrix M has the form $M = (I_2, I_2)$.

The solution of this minimization problem is given in Theorem 1 and the proof is in the supplementary material.

Theorem 1 *The minimizer of $\text{tr}AVA^\top$ under the constraint $M\boldsymbol{\alpha} = \mathbf{1}_q$ is given by*

$$\boldsymbol{\alpha}^* = \Omega^{-1}M^\top(M\Omega^{-1}M^\top)^{-1}\mathbf{1}_q, \quad (7)$$

where $\Omega = \Psi^\top(V \otimes I_q)\Psi$ and Ψ is such that $\text{vec}(A) = \Psi\boldsymbol{\alpha}$.

If we assume that all weight matrices A_r are diagonal, then each θ_i is estimated by the corresponding components $\hat{\theta}_{r,i}$, $r = 1, \dots, N$, only. In this case, $q_A = Nq$ and Ψ is a $Nq^2 \times Nq$ matrix having the form $\Psi = I_N \otimes \text{diag}(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_q)$ where \mathbf{e}_i is a vector with 1 in the i th place and zeros everywhere else. In this case, it can be shown that $\Omega = V \odot (J_N \otimes I_q)$, where J_N is a $N \times N$ matrix of ones and \odot is the Hadamard product between two matrices. This is equivalent with getting as Ω , a block matrix for which each block Ω_{rj} is a $q \times q$ diagonal matrix with elements, the corresponding elements of V but with zeros everywhere else. The optimal vector of weights $\boldsymbol{\alpha}$ takes the form

$$\boldsymbol{\alpha} = \Omega^{-1}(1_N \otimes I_q) \left[\left(1_N^\top \otimes I_q \right) \Omega^{-1} (1_N \otimes I_q) \right]^{-1} \mathbf{1}_q.$$

Since all Ω_{rj} matrices are diagonal, it can be shown that Ω^{-1} is again a block matrix and each block Ω^{rj} is again a $q \times q$ diagonal matrix. The i, i element of the A_r matrix is

$$A_{r,ii} = \frac{\sum_{j=1}^N \Omega_{ii}^{rj}}{\sum_{r=1}^N \sum_{j=1}^N \Omega_{ii}^{rj}}. \quad (8)$$

This means that the weights of $\hat{\theta}_{r,i}$, $r = 1, \dots, N$, the linear combination of which will estimate θ_i , are given by,

$$\frac{\sum_{j=1}^N \Omega_{ii}^{1j}}{\sum_{r=1}^N \sum_{j=1}^N \Omega_{ii}^{rj}}, \frac{\sum_{j=1}^N \Omega_{ii}^{2j}}{\sum_{r=1}^N \sum_{j=1}^N \Omega_{ii}^{rj}}, \dots, \frac{\sum_{j=1}^N \Omega_{ii}^{Nj}}{\sum_{r=1}^N \sum_{j=1}^N \Omega_{ii}^{rj}}. \quad (9)$$

5. SIMULATIONS

In this section, we present results from three simulation scenarios using the multivariate growth model for binary outcomes both with a linear and a quadratic time latent variable. In all simulations, we analysed $p = 4$ variables and ran 200 simulations. The

other parameters were selected as follows: under simulation scenario 1, the sample size was chosen as $n = 100$ and the number of time points $T = 10$, under scenario 2, $n = 500$ and $T = 10$ and under scenario 3, $n = 500$ and $T = 10$. The model used for simulating data under scenarios 1 and 2 is:

$$\text{logit}\{P(y_{jit} = 1)\} = \beta_{i,0} + \beta_{i,1}x_j + \beta_{i,2}t + \beta_{i,3}x_j \times t_j + \eta_{ji,0} + \eta_{ji,1} \times t_j, \quad (10)$$

and the model used for simulating data under scenario 3 is:

$$\text{logit}\{P(y_{jit} = 1)\} = \beta_{i,0} + \beta_{i,1}x_j + \beta_{i,2}t_j + \beta_{i,3}t_j^2 + \beta_{i,4}x_j \times t_j + \beta_{i,5}x_j \times t_j^2 + \eta_{ji,0} + \eta_{ji,1} \times t_j + \eta_{ji,2} \times t_j^2, \quad (11)$$

where x_j and t_j denote a group effect and the linear effect of time for individual j respectively and i denotes the item. In model (10), we used as parameter values for the fixed part of the model ($\beta_{i,0} = -2, \beta_{i,1} = -0.2, \beta_{i,2} = 0.5, \beta_{i,3} = 0.5$) and in model (11), ($\beta_{i,0} = -1, \beta_{i,1} = 0.5, \beta_{i,2} = 0.1, \beta_{i,3} = -0.2, \beta_{i,4} = 0.5, \beta_{i,5} = -0.1$). We assumed that the parameters of the fixed effects are the same across time for all items. The values of the parameters for the simulation study were chosen such that there is sufficient information (i.e., proportions of the binary responses not tending to zero or one) in each item and time point.

Furthermore, under simulation scenarios 1 and 2, the distribution of the latent variables $(\eta_{j1,0}, \eta_{j1,1}, \dots, \eta_{j4,0}, \eta_{j4,1})$ was assumed to be 8-dimensional normal with mean zero and covariance matrix

$$\Sigma = I_4 \otimes \begin{pmatrix} 0.3 & 0.1 \\ 0.2 & 0.3 \end{pmatrix} + 1_4 1_4^\top \otimes \begin{pmatrix} 0.7 & 0.7 \\ 0.6 & 0.6 \end{pmatrix}$$

For simulation scenario 3, the latent variables $(\eta_{j1,0}, \eta_{j1,1}, \eta_{j1,2}, \dots, \eta_{j4,0}, \eta_{j4,1}, \eta_{j4,2})$ were assumed to be 12-dimensional normal with mean zero and covariance matrix

$$\Sigma = I_4 \otimes \begin{pmatrix} 0.8 & 0.6 & 0.3 \\ 0.6 & 0.7 & 0.5 \\ 0.3 & 0.5 & 0.6 \end{pmatrix} + 1_4 1_4^\top \otimes 0.2 1_3 1_3^\top$$

Table 1 summarizes the results of the three simulation scenarios in terms of bias, efficiency and coverage of the 95% asymptotic confidence intervals of the unweighted and weighted estimator. For saving space, the results of the first two outcomes are presented. By comparing the columns of ‘S.E. unweighted’ and ‘S.E. weighted’ in the two tables, one can immediately see that the weighted estimator has a higher efficiency compared to the unweighted one. The difference between the asymptotic standard errors computed under the unweighted and the weighted estimator becomes smaller as the sample size increases

from 100 to 500 but still the difference is noticeable. The coverage of asymptotic 95% confidence intervals in most cases exceeds 95% and reaches one.

Although the weighted estimators have much smaller standard errors, these are still large compared to optimal ones and this could be a reason for empirical coverages reaching values near to one. No differences are found in the bias between the two estimators.

6. APPLICATION

For the data introduced in Section 2 on reported symptoms of schistosomiasis we used the multivariate growth model already discussed in Section 3 with a logistic link. The baseline and follow up of one and two mass drug administration (MDA) are modeled through a time covariate with values 0, 1 and 3 to allow for the two years gap between the second and the third measurements and also to make the interpretation of the intercept parameter meaningful. We also attempted to fit a model with two latent variables one to allow for a random intercept and random slope but we experienced numerical difficulties with the estimated variance of the random slopes practically zero, suggesting that this extension was not supported by the data. The latent variable at the intercept level represents the combined effect of all omitted children and item specific covariates that cause some children to be more prone to report symptoms of schistosomiasis or detected with blood in their urine through the test than other children. The model is:

$$\text{logit}\{P(y_{jit} = 1 \mid t, \eta_{ji,0})\} = \beta_{i,0} + \beta_{i,1}t + \eta_{ji,0} \quad (12)$$

where $\eta_{ji,0}$ are item-individual zero mean multivariate normals with an unrestricted covariance matrix representing item-individual variation at the intercept level. Table 2 gives the simple average estimates and the corresponding weighted ones for the intercept and the coefficient for time for all four items together with their standard errors. As it was expected, the weighted estimates have smaller standard errors. All four coefficients for time were found to be highly statistically significant. Obtaining the exponentiated coefficients of the time variable to bring their interpretation to odds ratios we find that after the MDAs, the chance of self-reporting and detected blood in the urine by the reagent strips decreased by 15% and 53% respectively. At the same time, the chances of knowing about the disease and self-reporting of pain when urinating have increased with time. Although we might have expected that MDA treatments to reduce not only the blood in urine but also the pain, it is still reasonable to expect that pain when urinating could be attributed to many other characteristics and therefore this does not necessarily reflect failure of MDA. The increase of knowledge and awareness is the positive effect of the SCI programme.

Table 1: Scenarios 1, 2 and 3: True values, unweighted and weighted estimates along with asymptotic standard errors and coverages, $n = 100$ or 500 , $p = 4$, $T = 10, 200$ simulations

	Outcome	True	Unweighted	Weighted	S.E. unweighted	S.E. weighted	Coverage unweighted	Coverage weighted
Scn. 1 ($n = 100$)	1	-2.00	-2.06	-2.06	1.28	0.93	1.00	0.99
		-0.20	-0.25	-0.25	1.69	1.24	1.00	1.00
		0.50	0.54	0.54	2.00	1.51	1.00	1.00
	2	0.50	0.54	0.54	2.60	2.02	1.00	1.00
		-2.00	-2.01	-2.01	1.31	0.98	1.00	1.00
		-0.20	-0.27	-0.27	1.62	1.19	1.00	1.00
Scn. 2 ($n = 500$)	1	0.50	0.52	0.52	2.04	1.59	1.00	1.00
		0.50	0.57	0.57	2.53	1.99	1.00	1.00
		-2.00	-1.98	-1.98	0.33	0.30	0.99	0.99
	2	-0.20	-0.25	-0.25	0.42	0.37	1.00	1.00
		0.50	0.51	0.51	0.53	0.47	1.00	1.00
		0.50	0.53	0.53	0.70	0.61	1.00	1.00
Scn. 3 ($n = 500$)	1	-2.00	-1.99	-1.99	0.34	0.30	1.00	1.00
		-0.20	-0.24	-0.24	0.41	0.36	1.00	1.00
		0.50	0.50	0.50	0.54	0.46	1.00	1.00
	2	0.50	0.52	0.52	0.70	0.60	1.00	1.00
		-1.00	-0.90	-0.88	1.79	1.10	1.00	0.99
		0.50	0.41	0.40	2.55	1.67	1.00	1.00
	1	0.10	-0.04	-0.05	1.56	1.01	1.00	1.00
		-0.20	-0.21	-0.21	5.69	3.57	1.00	1.00
		0.50	0.63	0.64	2.30	1.57	1.00	1.00
	2	-0.10	-0.10	-0.10	8.13	5.42	1.00	1.00
		-1.00	-0.91	-0.90	1.78	1.11	1.00	0.99
		0.50	0.44	0.43	2.56	1.71	1.00	1.00
	0.10	-0.04	-0.04	1.58	1.00	1.00	1.00	
	-0.20	-0.22	-0.22	5.52	3.32	1.00	1.00	
	0.50	0.62	0.62	2.28	1.52	1.00	1.00	
		-0.10	-0.08	-0.08	7.98	5.27	1.00	1.00

Table 2: Schistosomiasis data: unweighted and weighted estimated parameter values and standard errors for the fixed effects

Parameter	Unweighted estimate	Unweighted S.E.	Weighted estimate	Weighted S.E.
$\hat{\beta}_{1,0}$	-2.52	0.122	-2.51	0.107
$\hat{\beta}_{2,0}$	-1.69	0.072	-1.66	0.063
$\hat{\beta}_{3,0}$	-0.79	0.046	-0.79	0.045
$\hat{\beta}_{4,0}$	-1.54	0.066	-1.52	0.063
$\hat{\beta}_{1,1}$	-0.17	0.064	-0.17	0.064
$\hat{\beta}_{2,1}$	0.27	0.036	0.27	0.035
$\hat{\beta}_{3,1}$	0.75	0.035	0.75	0.035
$\hat{\beta}_{4,1}$	-0.75	0.084	-0.75	0.084

It is clear from Table 3 that gives the estimated covariance and correlation matrix for the random intercept term, that MDA is not enough to account for the dependencies among the items and children’s characteristics. Variances and correlations seem high enough to justify the use of the random intercept in the model. The highest correlations are between the random intercepts for the two self-reporting symptoms and between the self-reporting symptom of blood in urine and the detection through the reagent strips which implies that at baseline, children with low probability of reporting a blood urine symptom will also have a low probability of reporting pain and also low probability of detecting blood in their urine via the reagent strips. In addition to that, at baseline, there is a much smaller correlation between the two self-reporting symptoms, the reagent strips and how much children seem to know about the disease. The intra-class coefficient controlling for the effect of the explanatory variable time (intervention) is 0.493, 0.322, 0.160 and 0.281 for items 1 to 4 respectively. The intra-class correlation measures the dependencies among the dichotomous responses on the same children for each item. The largest homogeneity within children responses is detected for items 1 and 2 which are the self-reporting items. Children will continue reporting the presence of the symptoms even after the intervention (i.e. MDA) where in the case of the reagent strips the correlation of the measurements reduces to 0.281. We should note that in the data analysis our primary interest is in the parameters of the fixed effects of the model. The parameters associated with the item-specific latent variables are more seen as nuisance parameters and therefore we do not produce standard errors for those. However, inference on those parameters can be based on likelihood ratio tests already developed under the framework of composite likelihood estimation (Varin, 2008) or on bootstrapping methodology.

Table 3: Schistosomiasis data: weighted estimated covariance and correlation matrix for the random intercept

Item	1	2	3	4
1	3.194	2.178	0.668	1.687
2	2.178	1.563	0.422	0.980
3	0.668	0.422	0.628	0.518
4	1.687	0.980	0.518	1.286
1	1.000	0.975	0.472	0.832
2	0.975	1.000	0.426	0.691
3	0.472	0.426	1.000	0.576
4	0.832	0.691	0.576	1.000

In Table 4, we computed some fitted probabilities for various values of the latent variable and for all three time points and items. It is quite evident that the estimated probabilities for all four items vary significantly at different values of the latent variable indicating large variability in children responses. There is less of that variability for item 4 (reagent strips) at wave 3 in which the estimated probabilities of detecting blood in urine through the reagent strips is relatively low at all values of the latent variable. There is a clear effect of MDA which is reflected in the large reduction of the probability of detecting the disease through the reagent strips. This leads us to the conclusion that one needs to treat the self-reporting items (excluding the blood in urine) with caution and if possible avoid them when one wants to measure intervention effectiveness.

Table 4: Schistosomiasis data: fitted probabilities of a positive response for different values of the random intercept

	$-2 \times \hat{\sigma}_{\eta_{0ji}}$	$-1 \times \hat{\sigma}_{\eta_{0ji}}$	0	$1 \times \hat{\sigma}_{\eta_{0ji}}$	$2 \times \hat{\sigma}_{\eta_{0ji}}$
Item 1					
t	-3.574	-1.787	0	1.787	3.574
0	0.002	0.013	0.075	0.327	0.742
1	0.002	0.012	0.069	0.293	0.712
3	0.001	0.008	0.047	0.231	0.642
Item 2					
t	-2.500	-1.250	0	1.250	2.500
0	0.016	0.053	0.162	0.404	0.703
1	0.020	0.067	0.201	0.468	0.754
3	0.034	0.108	0.297	0.596	0.838
Item 3					
t	-1.585	-0.792	0	0.792	1.585
0	0.085	0.170	0.312	0.501	0.689
1	0.165	0.303	0.490	0.680	0.824
3	0.469	0.661	0.812	0.905	0.955
Item 4					
t	-2.268	-1.134	0	1.134	2.268
0	0.022	0.066	0.179	0.405	0.679
1	0.011	0.032	0.094	0.243	0.500
3	0.002	0.007	0.023	0.067	0.182

We do not report here the weights for each pair but we should note that for our data

application, weights took values close to 0 or 1 for pairs of items indicating that only a percentage of all pairs contributed to estimate the model parameters. Unfortunately that information becomes available after the estimation is complete and therefore it cannot be used to reduce the computational time but rather to increase the efficiency of the estimates.

7. CONCLUSION

The paper studies the use of weights in pairwise likelihood estimation for a family of models with random effects / latent variables. It is shown that the pairwise estimator obtained from maximizing the sum of all pairwise log-likelihoods can be written as a weighted sum of estimates obtained from separate maximizations of each pairwise log-likelihood. We propose a new set of weights that improve the efficiency of the pairwise likelihood estimator and apply them to a data set collected as part of the SCI programme. The four binary indicators of schistosomiasis are analyzed with a multivariate growth model. Simulations show that the proposed weights improve the efficiency of the estimators obtained from the pairwise maximum likelihood estimation. Finally, in our developments we have assumed that we work with complete data or in the presence of incomplete response patterns that the missing data mechanism is missing completely at random. However, due to the fact that missing data and complex missing data mechanisms are the norm rather than the exception in applied research, we are currently extending our proposed weights to also account for missing at random mechanisms. This is in line with recent developments in pseudo-likelihood methodology for incomplete data proposed by Molenberghs, Kenward, Verbeke, and Birhanu (2011).

SUPPLEMENTARY MATERIAL

Proof of Theorem 1

The variance of $A\hat{\Theta}$ is AVA^\top . The choice of A is based on the minimization of the total variance of the weighted estimator which is the trace of AVA^\top under the sum-to-one constraints. Let us define vector α which contains all q_A unique elements of A and assume that there is a design matrix M which imposes the sum-to-one restrictions to these elements. This matrix consists of zeros and ones at appropriate places. Each row of this matrix imposes a sum-to-one restriction to the elements of α suitable for estimating each one of the q model parameters. Here these restrictions are such that $M\alpha = \mathbf{1}_q$ where M is a $q \times q_A$ matrix and $\mathbf{1}_q$ is a q -dimensional vector of ones. Each of these q restrictions corresponds to a different set of estimated parameters therefore, matrix M is of full row rank. Let us define a $Nq^2 \times q_A$ matrix Ψ such that $\Psi\alpha = \text{vec}(A)$. Such a matrix defines

the position of each unique element of α in $\text{vec}(A)$ vector, therefore Ψ is of full column rank. Examples of vector α and matrices M and Ψ are found in the main document.

Consider the Lagrangian:

$$\phi = \frac{1}{2} \text{tr} AV A^\top - \ell^\top (M\alpha - \mathbf{1}_q) \quad (13)$$

where ℓ is a $q \times 1$ vector of Lagrange coefficients. Then,

$$\phi = \frac{1}{2} \{\text{vec}(A)\}^\top (V \otimes I_q) \{\text{vec}(A)\} - \ell^\top (M\alpha - \mathbf{1}_q).$$

Since $\Psi\alpha = \text{vec}(A)$,

$$\phi = \frac{1}{2} \alpha^\top \Psi^\top (V \otimes I_q) \Psi \alpha - \ell^\top (M\alpha - \mathbf{1}_q).$$

The differential is

$$d\phi = \left\{ \Psi^\top (V \otimes I_q) \Psi \alpha \right\}^\top d\alpha - \ell^\top M d\alpha.$$

Hence, the first order equations are

$$\Psi^\top (V \otimes I_q) \Psi \alpha = M^\top \ell \text{ and } M\alpha = \mathbf{1}_q. \quad (14)$$

Since Ψ is of full column rank, matrix $\Psi^\top (V \otimes I_q) \Psi$ is invertible and since M is of full row rank, matrix $M \left\{ \Psi^\top (V \otimes I_q) \Psi \right\}^{-1} M^\top$ is invertible too. Therefore, the solution with respect to ℓ is

$$\ell = \left[M \left\{ \Psi^\top (V \otimes I_q) \Psi \right\}^{-1} M^\top \right]^{-1} \mathbf{1}_q.$$

If we define $\Omega = \Psi^\top (V \otimes I_q) \Psi$ and substitute ℓ into (14), we get

$$\Omega\alpha = M^\top \left(M\Omega^{-1}M^\top \right)^{-1} \mathbf{1}_q,$$

from which the solution

$$\alpha^* = \Omega^{-1} M^\top (M\Omega^{-1}M^\top)^{-1} \mathbf{1}_q$$

emerges.

ACKNOWLEDGMENTS

We would like to thank Professor Alan Fenwick, Director of Schistosomiasis Control Initiative (SCI) organization based at the Department of Infectious Disease Epidemiology at Imperial College for providing us with the data and Dr Artemis Koukounari for her comments on the data analysis results. We also wish to thank the Referees for their constructive comments and suggestions. This research was funded by the Basic Research Funding Program 2010-2011 of the Athens University of Economics and Business.

References

- Arnold, B. C., & Strauss, D. (1991). Pseudolikelihood estimation: some examples. *Sankya: The Indian Journal of Statistics*, *53*, 233–243.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*, *46*, 443–459.
- Chan, W., & Bentler, P. (1998). Covariance structure analysis of ordinal ipsative data. *Psychometrika*, *63*, 369–399.
- Clements, A. C., Brooker, S., Nyandindi, U., Fenwick, A., & Blair, L. (2008). Bayesian spatial analysis of a national urinary schistosomiasis questionnaire to assist geographic targeting of schistosomiasis control in Tanzania, East Africa. *International Journal of Parasitology*, *38*, 401–415.
- Cox, D. R., & Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, *91*, 729–737.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of The Royal Statistical Society, Series B*, *39*, 1–38.
- Fenwick, A., Webster, J. P., Bosque-Oliva, E., Blair, L., Fleming, F. M., Zhang, Y., et al. (2009). The Schistosomiasis Control Initiative (SCI): rationale, development and implementation from 2002-2008. *Parasitology*, *136*(13), 1719–1730.
- Fieuws, S., & Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*, *62*, 424–431.
- Fournier, D., Skaug, H., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M., et al. (2012). Ad model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods Software*, *27*, 233–249.
- Geys, H., Molenberghs, G., & Ryan, L. M. (1999). Pseudolikelihood modeling of multivariate outcomes in developmental toxicology. *Journal of the American Statistical Association*, *94*, 734–745.
- Huber, P., Ronchetti, E., & Victoria-Feser, M.-P. (2004). Estimation of generalized linear latent variable models. *Journal of the Royal Statistical Society, Series B*, *66*, 893–908.
- Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: a comparison of three approaches. *Multivariate Behavioral Research*, *36*, 347–387.
- Katsikatsou, M. (2013). *Composite likelihood estimation for latent variable models with ordinal and continuous or ranking variables*. Unpublished doctoral dissertation, Uppsala Universitet.
- Katsikatsou, M., Moustaki, I., Yang-Wallentin, F., & Jöreskog, K. G. (2012). Pairwise likelihood estimation for factor analysis models with ordinal data. *Computational Statistics and Data Analysis*, *56*, 4243–4258.
- Koukounari, A., Sacko, M., Keita, A. D., Gabrielli, A. F., Landoure, A., Dembele, R., et al. (2006). Assessment of ultrasound morbidity indicators of schistosomiasis in the context of large-scale programs illustrated with experiences from Malian children. *American Journal of Tropical Medicine and Hygiene*, *75*(6), 1042–1052.

- Lange, K. (2004). *Optimization*. New York: Springer-Verlag.
- Lindsay, B. (1988). Composite likelihood methods. In N. U. Prabhu (Ed.), *Statistical Inference from Stochastic Processes* (pp. 221–239). Providence, RI: American Mathematical Society.
- Liu, J. (2007). *Multivariate ordinal data analysis with pairwise likelihood and its extension to SEM*. Unpublished doctoral dissertation, University of California.
- Molenberghs, G., Kenward, M., Verbeke, G., & Birhanu, T. (2011). Pseudo-likelihood estimation for incomplete data. *Statistica Sinica*, *21*, 187–206.
- Press, W., Teukolsky, S., Vetterling, W., & Flannery, B. (2007). *Numerical recipes: The art of scientific computing* (3rd ed.). New York: Cambridge University Press.
- Sammel, R. D., Ryan, L. M., & Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society, B*, *59*, 667–678.
- Schilling, S., & Bock, R. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*, *70*, 533–555.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multi-level, Longitudinal and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC.
- Varin, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis*, *92*, 1–28.
- Varin, C., Reid, N., & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, *21*, 5–42.
- Vasdekis, V., Cagnone, S., & Moustaki, I. (2012). A pairwise likelihood inference in latent variable models for ordinal longitudinal responses. *Psychometrika*, *77*, 425–441.