A Comparison of Estimation Methods in Latent Variable Models for Binary Panel Data

Published Version:
Guastadisegni, L., Bianconcini, S., Cagnone, S. (2025). A Comparison of Estimation Methods in Latent Variable Models for Binary Panel Data. Cham : Springer [10.1007/978-3-032-03042-9_23].

Availability:
This version is available at: https://hdl.handle.net/11585/1022978 since: 2025-09-11

Published:
DOI: http://doi.org/10.1007/978-3-032-03042-9_23

Metadata of the chapter that will be visualized in SpringerLink

Book Title	Supervised and Unsupervised Statistical Data Analysis	
Series Title		
Chapter Title	A Comparison of Estimation Methods in Latent Variable Models for Binary Panel Data	
Copyright Year	2025	
Copyright HolderName	The Author(s), under exclusive license to Springer Nature Switzerland AG	

Corresponding Author	Family Name	**Guastadisegni**
	Particle	
	Given Name	**Lucia**
	Prefix	
	Suffix	
	Role	
	Division	Department of Statistical Sciences “Paolo Fortunati”
	Organization	University of Bologna
	Address	Bologna, Italy
	Email	lucia.guastadisegni2@unibo.it
Author	Family Name	**Bianconcini**
	Particle	
	Given Name	**Silvia**
	Prefix	
	Suffix	
	Role	
	Division	Department of Statistical Sciences “Paolo Fortunati”
	Organization	University of Bologna
	Address	Bologna, Italy
	Email	silvia.bianconcini@unibo.it
Author	Family Name	**Cagnone**
	Particle	
	Given Name	**Silvia**
	Prefix	
	Suffix	
	Role	
	Division	Department of Statistical Sciences “Paolo Fortunati”
	Organization	University of Bologna
	Address	Bologna, Italy
	Email	silvia.cagnone@unibo.it

Abstract

This paper examines and compares various estimation methods for generalized linear latent variable models for multidimensional longitudinal binary data. In such cases, likelihood-based methods are problematic due to the high dimensional integrals involved, which lack analytical solution. Among the methods proposed in the literature to address this issue, we focus on approximate likelihood and composite likelihood methods. Specifically, within the first class, we examine the dimension-wise quadrature, while within the second class, we consider composite likelihood methods based on bivariate densities. The properties of these methods are evaluated through a comprehensive simulation study.

Keywords (separated by '-')	Approximate likelihood methods - pairwise likelihood methods - high dimensional integrals

A Comparison of Estimation Methods in Latent Variable Models for Binary Panel Data

Lucia Guastadisegni(✉), Silvia Bianconcini, and Silvia Cagnone

Department of Statistical Sciences "Paolo Fortunati", University of Bologna, Bologna, Italy
{lucia.guastadisegni2,silvia.bianconcini,silvia.cagnone}@unibo.it

AQ1

Abstract. This paper examines and compares various estimation methods for generalized linear latent variable models for multidimensional longitudinal binary data. In such cases, likelihood-based methods are problematic due to the high dimensional integrals involved, which lack analytical solution. Among the methods proposed in the literature to address this issue, we focus on approximate likelihood and composite likelihood methods. Specifically, within the first class, we examine the dimension-wise quadrature, while within the second class, we consider composite likelihood methods based on bivariate densities. The properties of these methods are evaluated through a comprehensive simulation study.

AQ2

Keywords: Approximate likelihood methods · pairwise likelihood methods · high dimensional integrals

1 Introduction

Multivariate longitudinal data are common in many fields of research, such as biostatistics and psychology, and can be used to measure changes in outcomes over time or to identify the determinants of such changes in a set of individuals. Among the different models proposed in the literature to analyze multivariate longitudinal data (see [15] for a review), we focus on those that includes random effects and latent variables. We adopt the generalized linear latent variable model (GLLVM) described by [6], to handle various response types in a multivariate longitudinal setting. This model is convenient due to its high flexibility, as it incorporates time-specific latent variables to capture associations among different response variables at each time point, random effects to account for the association of each response over time, and an autoregressive structure to represent the dynamic nature of the time-dependent latent variables, allowing for the inclusion of covariates. When likelihood-based methods are applied to GLLVM for multidimensional discrete data, a challenge arises because the high dimensional integrals involved in the likelihood lack analytical solutions. Among

A. D'Ambrosio et al. (Eds.): CLADAG-VOC 2025, SCDAKO, pp. 1–12, 2025.
https://doi.org/10.1007/978-3-032-03042-9_23

the discrete outcomes, binary data are often regarded as the most complex case, as the limited information provided by binary responses further complicates the process of likelihood maximization.

Classic solutions to this problem include approximate likelihood methods, which are based on Gauss-Hermite quadrature approximations [5]. A popular method within this class is the Adaptive Gauss-Hermite (AGH) quadrature, originally introduced in a Bayesian framework to efficiently approximate posterior densities [12]. However, it becomes infeasible as the number of latent variables and random effects increases [3]. A simpler way to estimate the model is to use the first-order Laplace approximation of the likelihood [8], which is feasible even with a large number of latent variables and random effects, but it yields inaccurate results with discrete data [9]. A recent method for approximating the likelihood, is the dimension-wise (DW) quadrature, proposed by [4]. This method reduces the dimension of the multidimensional integrals by truncating the Taylor series expansion of the integrand as the Laplace approximation, but without requiring any derivative computation. The corresponding estimators are asymptotically as accurate as the AGH estimators when using the same number of quadrature points [4]. Moreover, DW has lower computational complexity compared to AGH, making it feasible even when the latter is not [4].

An alternative way to address the problem of the integrals involved in the likelihood of GLLVMs for multidimensional longitudinal discrete data is to use composite likelihood methods [10], that combine likelihood components, associated with small subsets of data. Within this class, we consider the pairwise likelihood (PL) approach [11], which is based on marginal densities for pairs of observations. The estimator resulting from the maximization of the PL is asymptotically unbiased but less efficient than the full maximum likelihood (ML) estimator [13]. The PL method can also be derived using the method of separate maximizations [7]. According to this approach, the composite likelihood is defined as the sum of all log pairwise likelihoods across pairs of items or times, with each pairwise likelihood maximized separately. The obtained parameter estimates are then combined to produce a single parameter estimate through different weighting matrices. We consider the average (AVE) method, where the weight matrix is designed to compute the arithmetic mean of the estimates [7] and the weighted mean (WM) method [14], where the weight matrix produces a combined estimator that is asymptotically equivalent to the classic PL estimator.

In the context of GLLVM for multidimensional longitudinal data, some comparisons have been made between the classical PL method and the DW quadrature method [2,3]. Overall, the DW quadrature method outperforms the PL method in terms of efficiency, but it has a higher computational burden.

In this work, we extend the comparisons by also considering PL methods based on separate maximizations. Specifically, we evaluate the performance of the DW quadrature method, the classic PL method, and the AVE and WM methods, for a GLLVM for multidimensional longitudinal binary data which, as outlined before, is the most challenging case. The performance of these methods is assessed through a simulation study, along with a comparison of different weighting approaches used in the separate maximizations methods.

The paper is organized as follows. Section 2 introduces the GLLVM for multidimensional longitudinal binary data. Section 3 describes the DW method. Section 4 reviews the classic PL method, while Sect. 5 discusses the PL methods based on separate maximizations. Section 6 presents a simulation study. Finally, Sect. 7 offers discussion.

2 The GLLVM for Longitudinal Binary Data

Let $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}, \cdots, \boldsymbol{y}_{iT})$ be a $p \times T$ matrix, where $\boldsymbol{y}_{it} = (y_{i1t}, y_{i2t}, \cdots, y_{ipt})'$ is a vector of p binary variables measured at time t for individual i, with $i = 1, \ldots, n$. For each individual i, let $\boldsymbol{z}_i = (z_{i1}, z_{i2}, \ldots, z_{iT})^\top$ be a vector representing a latent variable measured at T time points, that accounts for the association among the p items at each time point, and $\boldsymbol{u}_i = (u_{i1}, u_{i2}, \ldots, u_{ip})'$ be a p-dimensional vector of random effects that accounts for the association of each item over time.

According to the GLLVM framework [1], we can express the joint probability of the individual response pattern $f(\boldsymbol{y}_i)$ as

$$f(\mathbf{y}_i) = \int_{\mathbb{R}^K} g(\mathbf{y}_i|\mathbf{u}_i, \mathbf{z}_i) h(\mathbf{u}_i, \mathbf{z}_i) \mathrm{d}\mathbf{u}_i \mathrm{d}\mathbf{z}_i, \tag{1}$$

where $g(\boldsymbol{y}_i|\boldsymbol{u}_i, \boldsymbol{z}_i)$ is referred to as the measurement part of the model and $h(\boldsymbol{u}_i, \boldsymbol{z}_i)$ as the structural part of the model. The dimension of the integral K is equal to the total number of latent variables and random effects.

For the measurement part of the model, the variables $\boldsymbol{y}_i$ are assumed to be conditionally independent given $\boldsymbol{u}_i$ and $\boldsymbol{z}_i$. This implies that

$$g(\mathbf{y}_i|\mathbf{u}_i, \mathbf{z}_i) = \prod_{j=1}^{p} \prod_{t=1}^{T} g(y_{ijt}|u_{ij}, z_{it}) = \pi_{ijt}(u_{ij}, z_{it})^{y_{ijt}} (1 - \pi_{ijt}(u_{ij}, z_{it}))^{1-y_{ijt}} \tag{2}$$

being $g(y_{ijt}|u_{ij}, z_{it})$ the random component and $\pi_{ijt}(u_{ij}, z_{it}) = P(y_{ijt} = 1|u_{ij}, z_{it})$. The latent variable z_{it}, $t = 1, \cdots, T$ and the random effects u_{ij}, $j = 1, \cdots, p$, determine the linear predictor that represents the systematic component of the model, that is $\eta_{ijt} = \alpha_{0jt} + \alpha_{1jt} z_{it} + u_{ij}$. The parameters α_{0jt} and α_{1jt} are the item- and time-specific intercept and loading, respectively. Under the assumption of measurement invariance, which must be tested on real data, the intercepts and factor loadings do not vary over time. This implies that $\alpha_{0jt} = \alpha_{0j}$ and $\alpha_{1jt} = \alpha_{1j}$ $\forall t$. The link between the systematic and the random component is the logit link function, $\eta_{ijt} = \ln\left(\frac{\pi_{ijt}(u_{ij}, z_{it})}{1-\pi_{ijt}(u_{ij}, z_{it})}\right)$ such that $\pi_{ijt} = \frac{\exp \eta_{ijt}}{1+\exp \eta_{ijt}}$.

The structural part of the model describes the dynamic nature of the model determined by the time-dependent latent variables z_{it}, $t = 1, \cdots, T$, and their relationship with the item-specific random effects u_{ij}, $j = 1, \cdots, p$. To account for the serial correlation in the latent variables, a time-homogeneous nonstationary autoregressive process of order one is specified. That is,

$$z_{it} = \phi z_{i(t-1)} + \delta_{it}, \qquad t = 2, \cdots, T, \tag{3}$$

with $z_{i1} \sim N(0, \sigma_1^2)$ and $\delta_{it} \sim N(0,1)$ for identification purposes. The latent variables and the random effects are assumed to be independent and normally distributed with $\mathbf{0}$ mean vector and a block diagonal covariance matrix $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Xi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma} \end{bmatrix}$, where $\boldsymbol{\Xi} = \mathrm{diag}_{j=1,\ldots,p}(\sigma_{uj}^2)$, and $\boldsymbol{\Gamma}$ is the autocovariance matrix of the latent variables.

Full maximum likelihood is performed by maximizing the following log-likelihood function

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(\boldsymbol{y}_i; \boldsymbol{\theta}) = \sum_{i=1}^{n} \log \int_{\mathbb{R}^K} g(\boldsymbol{y}_i | \boldsymbol{u}_i, \boldsymbol{z}_i) h(\boldsymbol{u}_i, \boldsymbol{z}_i) d\boldsymbol{u}_i d\boldsymbol{z}_i, \tag{4}$$

where $\boldsymbol{\theta}$ is the q-dimensional vector of model parameters. For binary data, the integrals cannot be solved analytically and K is equal to the number of random effects p plus the number of time-dependent latent variables T, that is $K = p+T$.

3 Dimension-Wise Quadrature Method

Let $\boldsymbol{b}_i = (\boldsymbol{u}_i, \boldsymbol{z}_i)'$ denotes the vector of latent variables and random effects. Similar to the AGH approximation, the DW quadrature method starts from the following representation of the marginal density function of the observed variables where, for simplicity, the individual subscript i has been omitted

$$f(\boldsymbol{y}; \boldsymbol{\theta}) = \int_{R^K} \frac{\prod_{j=1}^{p} \prod_{t=1}^{T} g(y_{jt} | \boldsymbol{b}) h(\boldsymbol{b})}{\phi(\boldsymbol{b}; \boldsymbol{b}_{mo}, \boldsymbol{\Sigma}_{mo})} \phi(\boldsymbol{b}; \boldsymbol{b}_{mo}, \boldsymbol{\Sigma}_{mo}) d\boldsymbol{b}, \tag{5}$$

where $\boldsymbol{b}_{mo} = \max_b g(\boldsymbol{y}|\boldsymbol{b})h(\boldsymbol{b})$ and $\boldsymbol{\Sigma}_{mo}$ is minus the inverse of the corresponding Hessian matrix evaluated in the mode $\boldsymbol{b}_{mo}$. $\phi(\cdot)$ is the multivariate normal distribution, that is the posterior distribution of the latent vector $\boldsymbol{b}$. We can rewrite function (5) in terms of standardized variables

$$\begin{aligned} f(\boldsymbol{y}; \boldsymbol{\theta}) &= |\boldsymbol{C}_{mo}| \int_{R^K} \frac{\prod_{j=1}^{p} \prod_{t=1}^{T} g(y_{jt} | \boldsymbol{C}_{mo}\boldsymbol{b}^* + \boldsymbol{b}_{mo}) h(\boldsymbol{C}_{mo}\boldsymbol{b}^* + \boldsymbol{b}_{mo})}{\phi(\boldsymbol{b}^*; \mathbf{0}, \boldsymbol{I})} \phi(\boldsymbol{b}^*; \mathbf{0}, \boldsymbol{I}) d\boldsymbol{b}^* = \\ &= |\boldsymbol{C}_{mo}| \int_{R^K} m(\boldsymbol{b}^*) \phi(\boldsymbol{b}^*; \mathbf{0}, \boldsymbol{I}) d\boldsymbol{b}^* = |\boldsymbol{C}_{mo}| E_{\phi}[m(\boldsymbol{b}^*)], \end{aligned} \tag{6}$$

where $m(\boldsymbol{b}^*) = \frac{\prod_{j=1}^{p} \prod_{t=1}^{T} g(y_{jt} | \boldsymbol{C}_{mo}\boldsymbol{b}^* + \boldsymbol{b}_{mo}) h(\boldsymbol{C}_{mo}\boldsymbol{b}^* + \boldsymbol{b}_{mo})}{\phi(\boldsymbol{b}^*; \mathbf{0}, \boldsymbol{I})}$ is a positive function and $\boldsymbol{C}_{mo}$ is derived by the Cholesky decomposition of $\boldsymbol{\Sigma}_{mo} = \boldsymbol{C}_{mo}\boldsymbol{C}'_{mo}$. The DW method is applied to $E_{\phi}[m(\boldsymbol{b}^*)]$, and it considers the Taylor expansion of $m(\mathbf{b}^*)$ around 0 up to the s term. The function $m(\mathbf{b}^*)$ can be written as [4]

$$\hat{m}(\mathbf{b}^*) = \sum_{l=0}^{s} (-1)^l \binom{K-s+l-1}{l} m_{s-l}(\mathbf{b}^*), \tag{7}$$

where $m_{s-l}(\mathbf{b}^*) = m(0, \cdots, b^*_{k_1}, 0, \cdots, 0, b^*_{k_{s-l}}, 0, \cdots, 0)$. Thus, $m_{s-l}(\mathbf{b}^*)$ is a function of just $s-l$ variables, with all the remaining variables fixed to 0. Substituting (7) to (6), we obtain the approximated marginal density function

$$f_a(\mathbf{y};\boldsymbol{\theta}) = f_L + |\boldsymbol{C}_{mo}| \left[\sum_{l=0}^{s-1}(-1)^l \binom{K-s+l-1}{l} \int_{R^{s-l}} \sum_{k_1<\ldots<k_{s-l}} m_{s-l}(\boldsymbol{b}^*)\phi(b^*_{k_1})\cdots\phi(b^*_{k_{s-l}})db^*_{k_1}..db^*_{k_{s-l}}\right], \quad (8)$$

where f_L is the Laplace approximation of the integral. The dimension of the integrals in (8) is determined by the choice of s. When s is small, these integrals can be easily approximated using the Gauss-Hermite quadrature.

When $s = 0$, DW is equivalent to the classical Laplace approximation. As s increases, the DW method consists of approximating the K-dimensional integral with the sum of one-, two-, up to s-dimensional integrals. When $s = K$, no dimension reduction is performed, and the method reduces to the AGH method. The DW estimators are consistent, as long as the number of observed variables and sample size increase, and asymptotically normally distributed, with a sandwich-type covariance matrix. The DW quadrature method is more accurate than the Laplace approximation and, for small values of s, at most 3, it can achieve the same accuracy as the AGH quadrature [4].

4 The Pairwise Likelihood Approach

The PL replaces the likelihood in (4) with bivariate marginal densities as follows [10]:

$$pl(\boldsymbol{\theta};\boldsymbol{y}) = \sum_{i=1}^{n} pl(\boldsymbol{\theta};\boldsymbol{y}_i) = \sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{j'=j}^{p}\sum_{t,t'} \log f(y_{ijt}, y_{ij't'};\boldsymbol{\theta}), \quad (9)$$

where $t, t' = 1, \cdots, T$. Moreover, $t = t' \Rightarrow j < j'$ and $j = j' \Rightarrow t < t'$. Formula (9) can be rewritten as the sum of bivariate densities over all possible pairs of items or time points. We consider the one for pairs of items. The number of pairs of items in this model is given by $N_p = \binom{p}{2} + p$, where $\binom{p}{2}$ accounts for the bivariate densities formed by different items over time, and p accounts for the bivariate densities of the same item at different time points. Thus, formula (9) can be rewritten as

$$pl(\boldsymbol{\theta};\boldsymbol{y}) = \sum_{r=1}^{N_p} f_r(\boldsymbol{\theta}) \quad (10)$$

where $\sum_{r=1}^{N_p} = \sum_{j=1}^{p}\ \sum_{j'=j}^{p}$ and

$$f_r(\boldsymbol{\theta}) = \sum_{i=1}^{n} f_{ri}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{t,t'} \log f(y_{ijt}, y_{ij't'};\boldsymbol{\theta}), \quad (11)$$

where $t, t' = 1, \cdots, T$. Moreover, $t = t' \Rightarrow j < j'$ and $j = j' \Rightarrow t < t'$. As clarified in Sect. 5, rewriting formula (9) as (10) is convenient for the formulation of the PL method based on separate maximization.

The PL estimator is obtained as

$$\hat{\boldsymbol{\theta}}_{PL} = \arg\max_{\boldsymbol{\theta}} pl(\boldsymbol{\theta}; \boldsymbol{y}). \tag{12}$$

Under standard regularity conditions [11], $\hat{\boldsymbol{\theta}}_{PL}$ is asymptotically [13]

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{PL} - \boldsymbol{\theta}) \xrightarrow{d} N(0, H(\boldsymbol{\theta})^{-1} J(\boldsymbol{\theta}) H(\boldsymbol{\theta})^{-1}), \tag{13}$$

where the matrices $H(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$[1] can be estimated by their observed versions evaluated at $\hat{\boldsymbol{\theta}}_{PL}$, denoted as $\mathcal{H}(\hat{\boldsymbol{\theta}}_{PL})$ and $\mathcal{J}(\hat{\boldsymbol{\theta}}_{PL})$[2].

5 Pairwise Likelihood Methods Based on Separate Maximizations

[7] proposed an alternative formulation of the PL method, based on separate maximizations. Consider formula (10), where $pl(\boldsymbol{y}; \boldsymbol{\theta})$ is expressed as the sum of $f_r(\boldsymbol{\theta})$ across all possible pairs of items N_p. Instead of maximizing the entire $pl(\boldsymbol{y}; \boldsymbol{\theta})$, each $f_r(\boldsymbol{\theta})$ is maximized separately. Let $\boldsymbol{\theta}_r$ represent the vector of parameters of $f_r(\boldsymbol{\theta}_r)$. The ML estimator $\hat{\boldsymbol{\theta}}_r$ is defined as

$$\hat{\boldsymbol{\theta}}_r = \arg\max_{\boldsymbol{\theta}_r} \ f_r(\boldsymbol{\theta}_r). \tag{14}$$

Let us denote $\hat{\boldsymbol{\Theta}}' = (\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_{N_p})'$, and let $\boldsymbol{\Theta}$ represent the corresponding true parameter vector. Both $\hat{\boldsymbol{\Theta}}$ and $\boldsymbol{\Theta}$ are Q_p-dimensional vectors, where $Q_p = N_p q$, and q is the dimension of $\boldsymbol{\theta}$. It is important to note that the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\Theta}$ are not equivalent. Specifically, certain parameters in $\boldsymbol{\theta}$ have more counterparts in $\boldsymbol{\Theta}$ than others [7]. For instance, the parameters ϕ and σ^2 can be estimated from all $f_r(\boldsymbol{\theta}_r)$, meaning they have N_p counterparts in $\boldsymbol{\Theta}$. In contrast, the item-specific intercepts and slopes can only be estimated from the functions $f_r(\boldsymbol{\theta}_r)$ that involve the respective items and therefore have fewer counterparts in $\boldsymbol{\Theta}$.

The estimator $\hat{\boldsymbol{\Theta}}$ shares the same asymptotic properties as its constituting elements, which are ML estimators. For this reason, asymptotically [7]

$$\sqrt{n}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}) \xrightarrow{d} N(\mathbf{0}, H(\boldsymbol{\Theta})^{-1} J(\boldsymbol{\Theta}) H(\boldsymbol{\Theta})^{-1}). \tag{15}$$

The asymptotic covariance matrix of $\hat{\boldsymbol{\Theta}}$, of dimension $Q_p \times Q_p$, has the form

$$V(\boldsymbol{\Theta}) = H(\boldsymbol{\Theta})^{-1} J(\boldsymbol{\Theta}) H(\boldsymbol{\Theta})^{-1}, \tag{16}$$

where $H(\boldsymbol{\Theta})$ and $J(\boldsymbol{\Theta})$ are $Q_p \times Q_p$ block matrices, where $H(\boldsymbol{\Theta})$ is block diagonal. Each block is denoted as H_{rr} and J_{ru}[3], for $r, u = 1, \ldots, N_p$. These matrices

[1] $H(\boldsymbol{\theta}) = -\frac{1}{n} E\left[\frac{\partial^2 pl(\boldsymbol{\theta};\boldsymbol{y})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}\right]$, $J(\boldsymbol{\theta}) = \frac{1}{n} E\left[\left(\frac{\partial pl(\boldsymbol{\theta};\boldsymbol{y})}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial pl(\boldsymbol{\theta};\boldsymbol{y}}{\partial\boldsymbol{\theta}}\right)'\right]$.

[2] $\mathcal{H}(\hat{\boldsymbol{\theta}}_{PL}) = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 pl(\mathbf{y}_i,\hat{\boldsymbol{\theta}}_{PL})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}$, $\mathcal{J}(\hat{\boldsymbol{\theta}}_{PL}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial pl(\boldsymbol{y}_i,\hat{\boldsymbol{\theta}}_{PL})}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial pl(\boldsymbol{y}_i,\hat{\boldsymbol{\theta}}_{PL})}{\partial\boldsymbol{\theta}}\right)'$.

[3] $H_{rr} = -\frac{1}{n}\sum_{i=1}^{n} E\left[\frac{\partial^2 f_{ri}(\boldsymbol{\theta}_r)}{\partial\boldsymbol{\theta}_r \partial\boldsymbol{\theta}_r'}\right]$, $J_{ru} = \frac{1}{n}\sum_{i=1}^{n} E\left[\left(\frac{\partial f_{ri}(\boldsymbol{\theta}_r)}{\partial\boldsymbol{\theta}_r}\right)\left(\frac{\partial f_{ui}(\boldsymbol{\theta}_u)}{\partial\boldsymbol{\theta}_u}\right)'\right]$.

can be estimated by their observed versions, evaluated at $\hat{\boldsymbol{\Theta}}$, denoted as $\mathcal{H}(\hat{\boldsymbol{\Theta}})$ and $\mathcal{J}(\hat{\boldsymbol{\Theta}})$. The matrix $\mathcal{H}(\hat{\boldsymbol{\Theta}})$ is block diagonal, and $\mathcal{J}(\hat{\boldsymbol{\Theta}})$ is a block matrix, both of size $Q_p \times Q_p$, with each block denoted as $\mathcal{H}_{rr}(\hat{\boldsymbol{\theta}}_r)$ and $\mathcal{J}_{ru}(\hat{\boldsymbol{\theta}}_r, \hat{\boldsymbol{\theta}}_u)$[4] for $r, u = 1, \ldots, N_p$.

The estimators in $\hat{\boldsymbol{\Theta}}$ should be combined to obtain a single estimator for the model parameter vector $\boldsymbol{\theta}$. This can be achieved by pre-multiplying $\hat{\boldsymbol{\Theta}}$ by a matrix of weights $\boldsymbol{A}$ [7]. Consequently,

$$\sqrt{n}(\boldsymbol{A}\hat{\boldsymbol{\Theta}} - \boldsymbol{A}\boldsymbol{\Theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{A}V(\boldsymbol{\Theta})\boldsymbol{A}'). \tag{17}$$

The asymptotic covariance matrix of the combined estimator takes the form $\boldsymbol{A}V(\boldsymbol{\Theta})\boldsymbol{A}'$ and can be estimated as $\boldsymbol{A}\mathcal{V}(\hat{\boldsymbol{\Theta}})\boldsymbol{A}'$, where $\mathcal{V}(\hat{\boldsymbol{\Theta}}) = \mathcal{H}(\hat{\boldsymbol{\Theta}})^{-1}\mathcal{J}(\hat{\boldsymbol{\Theta}})\mathcal{H}(\hat{\boldsymbol{\Theta}})^{-1}$. Among the various weighting matrices proposed in the literature, we consider those based on the AVE and WM methods.

5.1 The AVE Method

The estimators in $\hat{\boldsymbol{\Theta}}$ are averaged to obtain one single estimator for the model parameter in $\boldsymbol{\theta}$. The estimator $\hat{\boldsymbol{\theta}}_{AVE}$ is obtained as

$$\hat{\boldsymbol{\theta}}_{AVE} = \boldsymbol{A}\hat{\boldsymbol{\Theta}}. \tag{18}$$

In this case, $\boldsymbol{A} = [A_1 \cdots A_{N_p}]$ has dimension $q \times qN_p$. Each matrix A_r is diagonal, with the weights on the main diagonal denoted as $\alpha_{hh}^{(r)}$, for $r = 1, \ldots, N_p$ and $h = 1, \ldots, q$. If θ_h is estimated in the r-th pair, the corresponding weight in A_r can be computed as

$$\alpha_{hh}^{(r)} = \frac{1}{n_h}, \tag{19}$$

where n_h is the number of times θ_h is estimated in the N_p separate maximizations. Moreover, it holds that $\sum_{r=1}^{N_p} \sum_{h=1}^{q} \alpha_{hh}^{(r)} = 1$.

5.2 The Weighted Mean Method

[14] demonstrated that $\hat{\boldsymbol{\theta}}_{PL}$ is asymptotically a weighted estimator of estimates obtained from separate maximizations. The detailed proof is provided in [14]. We refer to this approach as the weighted mean (WM) method. The weight matrix A_r, for $r = 1, \ldots, N_p$, that ensures the asymptotic equivalence between the WM and the PL methods is defined as

$$A_r = \left(\sum_{g=1}^{N_p} \frac{\partial^2 f_g(\hat{\boldsymbol{\theta}}_g)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)^{-1} \frac{\partial^2 f_r(\hat{\boldsymbol{\theta}}_r)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}, \tag{20}$$

where the index g denotes the summation across item pairs, with r held fixed.

---

4 $\mathcal{H}_{rr}(\hat{\boldsymbol{\theta}}_r) = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 f_{ri}(\hat{\boldsymbol{\theta}}_r)}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}_r'}$, $\mathcal{J}_{ru}(\hat{\boldsymbol{\theta}}_r, \hat{\boldsymbol{\theta}}_u) = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{\partial f_{ri}(\hat{\boldsymbol{\theta}}_r)}{\partial \boldsymbol{\theta}_r}\right)\left(\frac{\partial f_{ui}(\hat{\boldsymbol{\theta}}_u)}{\partial \boldsymbol{\theta}_u}\right)'$.

The final estimator $\hat{\boldsymbol{\theta}}_{WM}$ is obtained as

$$\hat{\boldsymbol{\theta}}_{WM} = \boldsymbol{A}\hat{\boldsymbol{\Theta}}, \tag{21}$$

where $\boldsymbol{A} = [A_1 \cdots A_{N_p}]$ is a matrix of dimension $q \times N_p q$.

6 Simulation Study

We conduct a simulation study to evaluate the performance of DW, PL, WM, and AVE methods for a GLLVM applied to multidimensional longitudinal binary data. Some preliminary results on the performance of the classical PL and DW methods for this type of model have been presented in [2], for different values of T and p fixed to 3. That study shows that the performance of DW and PL remains consistent for $T = 6, 15$. In this work, we also explore the WM and AVE methods and compare their respective weighting schemes. We focus on the case $T = 6$ time points, which represents a plausible setting for panel data, and consider sample sizes $n = 200$ and $n = 1000$. For consistency with the study of [2], we set the number of items to $p = 3$.

The same data-generating values used in the study of [2] are considered: $\phi = 0.5$, $\sigma_1^2 = 2$, and $\sigma_{u_j}^2$ (for $j = 1, 2, 3$) ranging from 1 to 2. The item intercepts are generated from a log-normal distribution with a mean of 0 and a standard deviation (SD) of 0.1, while the item slopes are generated from a log-normal distribution with a mean of 0 and an SD of 0.5. We consider the classic PL method and the AVE and WM methods. For DW, we set $s = 0, 1, 2, 3$. The values of s are not increased further because, already for $s = 2$ or 3, a high level of accuracy can be obtained [2,3]. Moreover, increasing s would also lead to higher computational time.

The optimization for all methods is performed in R using the *optim* function, which employs a Quasi-Newton algorithm, combined with Fortran. For all methods, the number of quadrature points is fixed at eight, as this represents a good trade-off between computation time and estimator accuracy. We consider 500 replications for each condition of the study.

6.1 Results

The following figures display the boxplots of parameter estimates for ϕ, σ_1^2, $\sigma_{u_3}^2$, and α_{1_2}. The remaining parameter estimates exhibit similar behavior and are omitted for brevity.

The boxplots of parameter estimates for $p = 3$, $T = 6$, $n = 200$ are reported in Fig. 1.

Starting from the composite likelihood methods we observe several differences between the classic PL and WM, with the latter producing less accurate results, particularly for the parameters ϕ and σ_1^2. The AVE method exhibits noticeable bias for $\sigma_{u_3}^2$ and α_{1_2}. DW demonstrates accurate results for $s = 2$ and $s = 3$, and shows lower variability than the composite likelihood methods for most parameters.

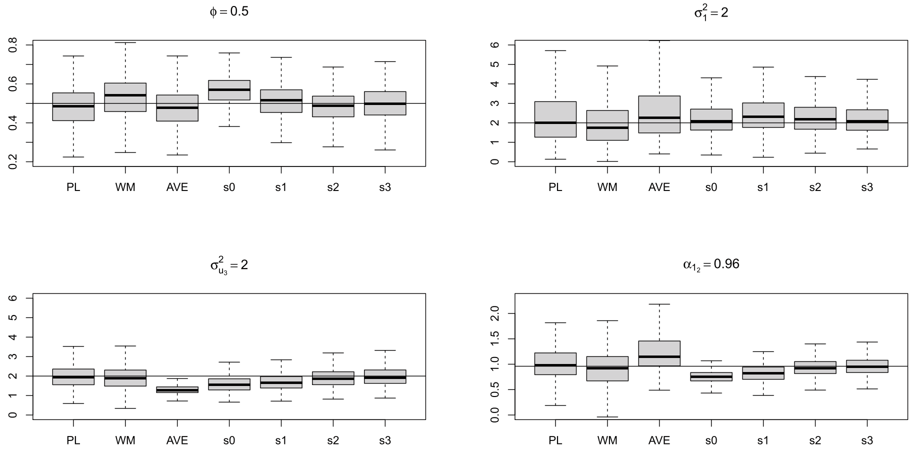


Fig. 1. Boxplots of parameter estimates, $n = 200$.

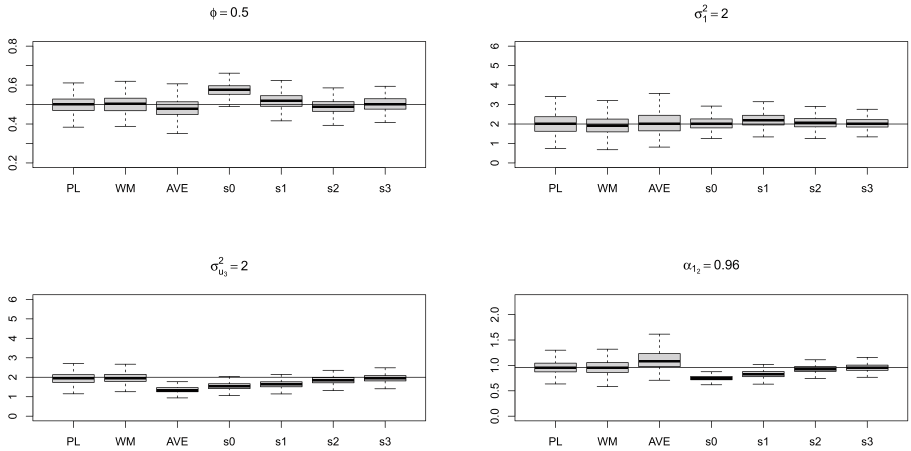


Fig. 2. Boxplots of parameter estimates, $n = 1000$.

For $p = 3$, $T = 6$, $n = 1000$ (Fig. 2), all methods show reduced variability compared to $n = 200$. Notably, also for large sample sizes, DW demonstrates lower variability than the composite likelihood methods for most parameters and produces accurate results for $s = 2$ and $s = 3$. Concerning the composite likelihood methods, as the sample size increases, WM and PL exhibit a similar performances. However, with AVE, the bias in the estimates of $\sigma^2_{u_3}$ and α_{1_2} does not reduce as the sample size increases.

Further insights into the results can be obtained by examining the behaviour of the weights.

Figure 3 presents scatter plots of the WM and AVE weights corresponding to the estimates of ϕ, σ^2_1, $\sigma^2_{u_3}$ and α_{1_2} for $p = 3$, $T = 6$, and $n = 1000$. On the x-

axis, the WM weights are plotted, while the corresponding parameter estimates are shown on the y-axis. The red vertical line identifies the AVE weights, which are the same for all estimates, while the black horizontal line represents the true parameter value.

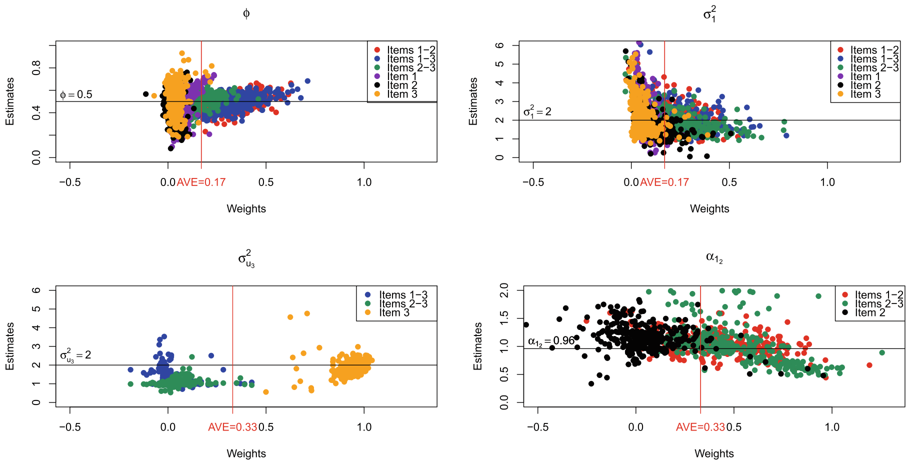


Fig. 3. Scatterplots for parameter estimates and weights for $n = 1000$.

We can notice that the weights of the WM method are very different from those provided by the AVE method for all model parameters. Interestingly, we observe a clear trend for $\sigma^2_{u_3}$: the WM weights concentrate around 1 for the estimates obtained from $f_r(\boldsymbol{\theta}_r)$ formed by bivariate densities of the same item over time, and around 0 for the estimates obtained from $f_r(\boldsymbol{\theta}_r)$ formed by bivariate densities of different items over time. A similar behavior is also observed for the variances of the remaining random effects. This reflects the fact that the random effects account for the variability of the same item over time. In contrast, for α_{1_2}, higher WM weights are observed for the estimates obtained from $f_r(\boldsymbol{\theta}_r)$ formed by bivariate densities of different items over time. This is also true for the loading of the third item, even though it is not reported in the Figure. Indeed, the loadings account for the correlations between different items. The AVE weights are equal to 0.33 for all variances of the random effect and loading estimates, without accounting for any differences between them. Interestingly, for ϕ and σ^2_1, we observe that the WM weights, which concentrate around 0, correspond to the less accurate estimates, which are those farther from the black horizontal line. The AVE weights are equal to 0.17 for all estimates, regardless of their distance from the true values.

Even though the AVE method is simpler and faster, it provides rough weights for the estimates and does not capture the specific characteristics of the parameters as effectively as the WM method.

7 Conclusion

In this work, we compared several composite likelihood methods and the DW method to address the high-dimensional integrals involved in the likelihood of GLLVM models for multidimensional longitudinal binary data, which do not have analytical solutions. Among the composite likelihood methods, we evaluated the classic PL method as well as some PL methods based on separate maximizations, namely the WM and AVE methods.

Generally, DW is more efficient than the composite likelihood methods considered for $s = 2$ and $s = 3$, but it involves a high computational cost. Similar findings are reported by [2] for longitudinal binary data and [3] for longitudinal ordinal data. Among the separate maximizations methods, WM outperforms AVE, which, despite its simplicity, produces inconsistent results, particularly for the variances of the random effects and the loadings. WM is asymptotically equivalent to the classic PL, but it does not offer a significant practical advantage over the latter. Therefore, it is valuable to explore weights that minimize the variance of the combined estimator, such as those proposed by [14]. Regarding the DW method, future work should focus on improving the computational time of the method and studying its asymptotic properties in the presence of longitudinal data.

Acknowledgements. This study was funded by the European Union - NextGenerationEU, in the framework of the "GRINS -Growing Resilient, INclusive and Sustainable project" (PNRR - M4C2 - I1.3 - PE00000018 – CUP J33C22002910001). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

References

1. Bartholomew, D.J., Knott, M., Moustaki, I.: Latent Variable Models and Factor Analysis: A Unified Approach. Wiley (2011). https://doi.org/10.1002/9781119970583
2. Bianconcini, S., Cagnone, S.: Estimation issues in multivariate panel data. In: Studies in Classification, Data Analysis, and Knowledge Organization, Springer (Forthcoming)
3. Bianconcini, S., Cagnone, S.: Comparison between different estimation methods of factor models for longitudinal ordinal data. Wiberg M.; Molenaar D.; González J.; Böckenholt U.; Kim J.S., Quantitative Psychology: The 85th Annual Meeting of the Psychometric Society, Virtual, Springer Proceedings in Mathematics & Statistics, 9–21 (2021). https://doi.org/10.1007/978-3-030-74772-5_2
4. Bianconcini, S., Cagnone, S., Rizopoulos, D.: Approximate likelihood inference in generalized linear latent variable models based on the dimension-wise quadrature. Electron. J. Stat. **11**(2), 4404–4423 (2017). https://doi.org/10.1214/17-EJS1360
5. Bock, R.D., Aitkin, M.: Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. Psychometrika **46**(4), 443–459 (1981). https://doi.org/10.1007/BF02293801

6. Dunson, D.B.: Dynamic latent trait models for multidimensional longitudinal data. J. Am. Stat. Assoc. **98**(463), 555–563 (2003). https://doi.org/10.1198/016214503000000387
7. Fieuws, S., Verbeke, G.: Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. Biometrics **62**(2), 424–431 (2006). https://doi.org/10.1111/j.1541-0420.2006.00507.x
8. Huber, P., Ronchetti, E., Victoria-Feser, M.-P.: Estimation of generalized linear latent variable models. J. R. Stat. Soc. Ser. B Stat. Methodol. **66**(4), 893–908 (2004). https://doi.org/10.1111/j.1467-9868.2004.05627.x
9. Joe, H.: Accuracy of Laplace approximation for discrete response mixed models. Comput. Stat. Data Anal. **52**(12), 5066–5074 (2008). https://doi.org/10.1016/j.csda.2008.05.002
10. Lindsay, B.G.: Composite likelihood methods. Contemp. Math. **80**, 221–239 (1988). https://doi.org/10.1090/conm/080/999014
11. Molenberghs, G., Verbeke, G.: Models for Discrete Longitudinal Data. Springer, Heidelberg (2005). https://doi.org/10.1007/0-387-28980-1
12. Naylor, J.C., Smith, A.F.M.: Applications of a method for the efficient computation of posterior distributions. J. R. Stat. Soc.: Ser. C: Appl. Stat. **31**(3), 214–225 (1982). https://doi.org/10.2307/2347995
13. Varin, C., Vidoni, P.: A note on composite likelihood inference and model selection. Biometrika **92**(3), 519–528 (2005). https://doi.org/10.1093/biomet/92.3.519
14. Vasdekis, V., Rizopoulos, D., Moustaki, I.: Weighted pairwise likelihood estimation for a general class of random effects models. Biostatistics **15**(4), 677–689 (2014). https://doi.org/10.1093/biostatistics/kxu018
15. Verbeke, G., Fieuws, S., Molenberghs, G., Davidian, M.: The analysis of multivariate longitudinal data: a review. Stat. Methods Med. Res. **23**(1), 42–59 (2014). https://doi.org/10.1177/0962280212445834

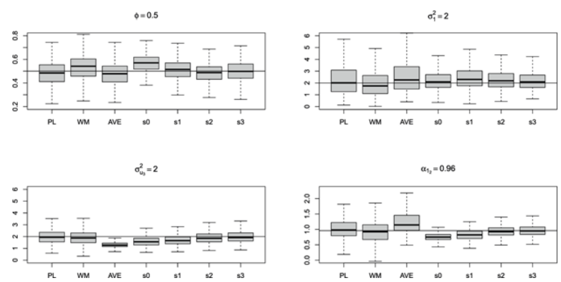
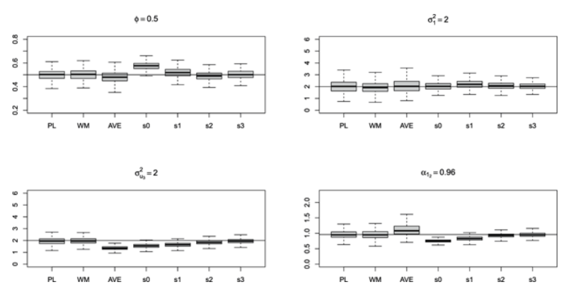
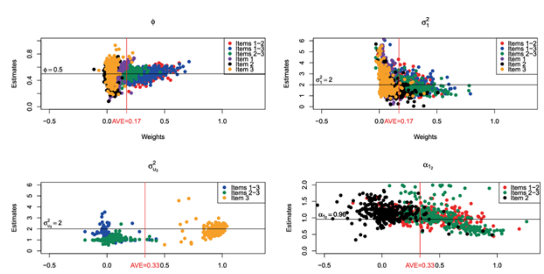
Author Proof

Author Queries

Chapter 23

Query Refs.	Details Required	Author's response
AQ1	This is to inform you that corresponding author has been identified as per the information available in the Copyright form.	
AQ2	Per Springer style, both city and country names must be present in the affiliations. Accordingly, we have inserted the city and country names in affiliation. Please check and confirm if the inserted city and country names is correct. If not, please provide us with the correct city and country names.	

Alternative Texts for Your Images, Please Check and Correct them if Required

Page no	Fig/Photo	Thumbnail	Alt-text Description
9	Fig1		The image consists of four box plots arranged in a 2x2 grid. Each plot displays data distributions for categories labeled PL, IW, AW, and AL. The top left plot is labeled with a parameter phi equals 0.6, the top right with sigma squared equals 2, the bottom left with epsilon squared equals 2, and the bottom right with alpha equals 0.98. Each box plot shows median, quartiles, and potential outliers. The plots are used to compare variations across different conditions.
9	Fig2		Box plots showing data distributions across different categories labeled PL, WM, AVE, s0, s1, s2, and s3. Each plot is annotated with parameters: top left with δ=0.5, top right with σ_1^2=2, bottom left with σ_5^2=2, and bottom right with α_{12}=0.96. The plots illustrate variations in data spread and central tendency for each category.
10	Fig3		Four-panel figure showing X-Y charts with scatter plots. Each chart displays "Estimates" on the Y-axis and "Weights" on the X-axis. The top left chart is labeled with the Greek letter phi (φ) and includes a vertical red line at 0.0 with "AVE=0.17" in red. The top right chart is labeled with sigma squared subscript 1 ($\sigma^2{}_1$) and has a similar red line and label. The bottom left chart is labeled with sigma squared subscript u3 ($\sigma^2{}_{u3}$) and "AVE=0.33." The bottom right chart is labeled with alpha subscript 12 (α_{12}) and "AVE=0.33." Each chart

Page no	Fig/Photo	Thumbnail	Alt-text Description
			includes a legend indicating different item groupings with various colored dots.