Received: 25 April 2023 | Accepted: 18 January 2024

DOI: 10.1111/bmsp.12337

ARTICLE

the british psychological society
promoting excellence in psychology

Fast estimation of generalized linear latent variable models for performance and process data with ordinal, continuous, and count observed variables

Maoxin Zhang[1] | **Björn Andersson**[1,2] | **Shaobo Jin**[3]

[1]Center for Educational Measurement, University of Oslo, Oslo, Norway

[2]Centre for Research on Equality in Education (CREATE), University of Oslo, Oslo, Norway

[3]Department of Mathematics, Uppsala University, Uppsala, Sweden

Correspondence
Björn Andersson, Center for Educational Measurement, University of Oslo, Oslo, Norway.
Email: bjorn.andersson@cemo.uio.no

Funding information
This work was partially supported by the Research Council of Norway through its Centres of Excellence scheme, Grant/Award Number: 331640

Different data types often occur in psychological and educational measurement such as computer-based assessments that record performance and process data (e.g., response times and the number of actions). Modelling such data requires specific models for each data type and accommodating complex dependencies between multiple variables. Generalized linear latent variable models are suitable for modelling mixed data simultaneously, but estimation can be computationally demanding. A fast solution is to use Laplace approximations, but existing implementations of joint modelling of mixed data types are limited to ordinal and continuous data. To address this limitation, we derive an efficient estimation method that uses first- or second-order Laplace approximations to simultaneously model ordinal data, continuous data, and count data. We illustrate the approach with an example and conduct simulations to evaluate the performance of the method in terms of estimation efficiency, convergence, and parameter recovery. The results suggest that the second-order Laplace approximation achieves a higher convergence rate and produces accurate yet fast parameter estimates compared to the first-order Laplace approximation, while the time cost increases with higher model complexity. Additionally, models that consider the dependence of variables from the same stimulus fit the empirical data substantially better than models that disregarded the dependence.

KEYWORDS
estimation efficiency, high dimensionality, Laplace approximation, mixed data types

1 | INTRODUCTION

Due to technological advances in data collection via digital devices, mixtures of continuous and discrete data (e.g., binary, categorical, and count) frequently occur in assessment contexts (De Leon & Chough, 2013, Chapter 1). Compared to traditional paper-and-pencil tests that only record the final answer to the items, computer-based assessments can track the entire human–computer interaction sequence (e.g., mouse clicks and keyboard input with time-stamps) and compile such information in log files. From log files, researchers can extract different types of indicators for further analysis, such as scored responses (categorical), response time spent on single items (continuous), response time until the first action (continuous), and the number of actions for each item (count). Such information is widely available in large-scale assessments, providing abundant research material for understanding participants' task-taking behaviours (Costa & Chen, 2023; De Boeck & Scalise, 2019; Ulitzsch et al., 2020b). These data also commonly exist in game-based assessments (Landers et al., 2022), which routinely collect the number of correct or incorrect trials, the number of mouse clicks, and the performance scores. In addition, sophisticated measurement tools such as eye-tracking devices also produce mixed data such as fixation count and fixation duration (Man et al., 2022; Man & Harring, 2023; Steinfeld, 2016). Hence, a combination of continuous and discrete data widely exists in educational and psychological assessments, providing researchers and practitioners with valuable information on diverse aspects of the respondents.

However, a mixture of different types of data poses challenges for conventional statistical methods because of the complex dependence structures that often exist (De Leon & Chough, 2013, Chapter 1). To be specific, dependence can stem from the same type of indicator, such as the responses to a number of items or tasks, and from different types of indicators based on the same stimulus, such as the item response and the response time from the same task. The former type of dependence is typically handled by introducing latent variables, while the latter type is often ignored. However, ignoring the dependence of indicators from the same task can lead to biased parameter estimation (De Boeck & Scalise, 2019; Meng et al., 2015). Additionally, the inherent non-normality of categorical and count data means that traditional analysis methods that assume continuous and normally distributed observed variables are less suitable to use.

Despite the above-mentioned challenges, multiple approaches to handling the issue of a mixture of different types of data exist. Among them, drawing inferences for each type of measure via separate models is the simplest approach. For example, researchers can analyse ordinal performance data via item response theory models, continuous response time via factor analysis, and the number of actions via count data models. However, a multiple-testing issue arises (De Leon & Chough, 2013, Chapter 2) and the approach cannot capture relationships between the measures because they are modelled separately. Therefore, a single multivariate model is regarded as a more appealing approach. To estimate such models with traditional methods, data may be converted into the same type by recoding continuous data as categorical data according to certain cut-off values or treating discrete data as continuous. The former method causes a loss of information while the latter violates a model assumption. Neither of these approaches is ideal and it is instead recommended to treat the observed variables as they are (Huber et al., 2004). For joint analysis of ordinal and continuous data, limited-information estimation with polyserial correlations may be used (Olsson et al., 1982). However, such a method cannot handle count data and the existence of missing data poses an issue in estimation.

An alternative approach is therefore to model mixed data jointly under the framework of generalized linear latent variable models (GLLVMs; Bartholomew et al., 2011; Huber et al., 2004; Rabe-Hesketh et al., 2004). A complicating factor for GLLVMs concerns the estimation of model parameters. Typically, full-information maximum likelihood or Bayesian estimation has been proposed. Bayesian inference is based on the posterior distribution of the freely estimated parameters given the data and priors of the parameters (Bartholomew et al., 2011, p. 30). When the dimension is high or models are very complex, Markov chain Monte Carlo (MCMC) methods are often used. For example, Man and Harring (2023) and Qiao et al. (2022) jointly modelled ordinal, continuous, and count data with Bayesian methods. MCMC methods are computationally demanding with many latent variables, and residual dependence between multiple observed variables related to the same stimulus or task has therefore commonly been ignored when using Bayesian methods (Man & Harring, 2023; Qiao et al., 2022; Ulitzsch et al., 2020b).

In contrast to Bayesian estimation, full-information maximum likelihood integrates out the latent variables from the likelihood function and maximizes the marginal likelihood. However, the integrals do not have closed-form solutions for GLLVMs, and approximation methods are required to compute them. One approach is Gauss–Hermite quadrature (GHQ), which has been implemented for GLLVMs with a collection of data from different distributions in the exponential family (Moustaki, 1996; Moustaki & Knott, 2000). GHQ works well for simple models but becomes unfeasible with more than three latent variables because the computational cost grows exponentially as the latent variable dimension increases (Andersson & Xin, 2021; Huber et al., 2004). Adaptive Gauss–Hermite quadrature (AGHQ) identifies integration intervals with rapid changes and reduces the required number of quadrature points (Rabe-Hesketh et al., 2002). AGHQ methods for generalized linear latent and mixed models are available in the Stata package *gllamm* (Rabe-Hesketh et al., 2004) and both quadrature methods are available in Mplus (Muthén & Muthén, 2017). Although AGHQ is faster than GHQ, it is still computationally demanding when the dimension is high. Instead, methods using Laplace approximations are showing promise for approximating the required integrals accurately and fast (Andersson & Xin, 2021; Huber et al., 2004; Niku et al., 2017). First-order Laplace approximations have been proposed to estimate GLLVMs for mixed data with distributions in the exponential family (Huber et al., 2004). Estimation with first-order Laplace approximations has been implemented in the R package *gllvm* (Niku et al., 2017) but supports only one type of indicator at a time. It is worth noting that first-order Laplace approximations (Lap1) are equivalent to AGHQ with one quadrature point per dimension when using the posterior mode and Hessian, and the method thus works highly efficiently with complex, high-dimensional models. However, this comes at the cost of non-convergence and inaccuracy with binary data and few observed variables (Andersson & Xin, 2021; Joe, 2008). To handle issues regarding convergence and accuracy in parameter recovery, a second-order Laplace approximation (Lap2) can be used (Shun, 1997). Lap2 requires higher-order derivatives to obtain a more accurate approximation by including more information, but with the downside that it needs more time in estimation (Andersson et al., 2023b).

In this paper we propose to apply both first- and second-order Laplace approximations to GLLVMs with mixed observed variables. Our main interest is in applying Laplace approximations (both Lap1 and Lap2) to enable joint modelling of ordinal, continuous, and count variables based on process data and performance data in educational and psychological measurement, where the residual dependence between observed variables related to the same stimulus is accounted for. This paper makes three main contributions beyond the existing literature. First, we implement estimation of joint models for count data, continuous data, and ordinal data using Laplace approximations, extending the papers by Huber et al. (2004) and Niku et al. (2017). Second, compared to Huber et al. (2004) and Niku et al. (2017) who only implemented the first-order Laplace approximation, we further implement a second-order Laplace approximation. Third, we provide a comparison between Lap1 and Lap2 in estimating GLLVMs with a mixture of different observed variables, extending the comparison in Andersson and Xin (2021) and Andersson et al. (2023b) from only categorical data to also include continuous and count data.

The remainder of the paper is organized as follows. In Section 2 we introduce GLLVMs and derive the estimation algorithm. In Section 3 we then describe a motivating example from the Programme for International Student Assessment (PISA) to guide our simulation design. In Section 4 simulations are conducted to evaluate estimation of joint models for ordinal, continuous, and count data with or without considering the dependence of indicators from the same task. Section 5 concludes.

2 | METHODS

2.1 | Generalized linear latent variable models

GLLVMs are extensions of generalized linear models (Nelder & Wedderburn, 1972), which are a class of regression models for discrete or continuous outcomes. GLMs consist of three components (Nelder & Wedderburn, 1972): a linear combination of predictors,

$$\nu = b + \boldsymbol{\beta}'\mathbf{w}, \tag{1}$$

where b and $\boldsymbol{\beta}$ are the intercept and regression coefficients, and $\mathbf{w}$ represents D-dimensional predictors; the outcome variable belonging to an exponential dispersion family; and a monotone and differentiable link function g, such as the identity, logit, or probit function, which relates the expected value of the outcome variable to the linear combination of predictors ν. In GLMs, there is only one outcome variable and all the variables are observable. When there are multiple correlated indicators that are developed to measure the same construct, such as responses from several cognitive tasks, we can incorporate latent variables to account for the dependence between the indicators. In social science, it is common to develop a battery of tests to measure theoretical constructs since they cannot be directly observed.

GLMs are extended to GLLVMs (Bartholomew et al., 2011) by introducing latent variables. Let y_i denote the ith observed outcome variable. Following Rabe-Hesketh et al. (2004), a general formula for GLLVMs can be written as

$$g_i(E[y_i \mid \mathbf{w}, \mathbf{z}]) = b_i + \boldsymbol{\beta}_i'\mathbf{w} + \mathbf{a}_i'\mathbf{z}, \tag{2}$$

where $\mathbf{a}_i$ is a vector of slope parameters or factor loadings of variable i, $\boldsymbol{\beta}_i$ is a D-dimensional vector of regression coefficients, and $\mathbf{z}$ is a P-dimensional vector of latent variables. To link the linear combination and the expected value of the observed variables, the link function g_i must be defined for each observed variable. For the distribution of the latent variables, we assume a multivariate normal distribution. For identification purposes, the means and variances of the latent variables are constrained to zeros and ones, respectively. The observed outcome variables are assumed to be independent conditional on the latent variables (Huber et al., 2004).

Let $\mathbf{y}$ be the $I \times 1$ vector of observed outcome variables. The marginal log-likelihood for a response vector $\mathbf{y}$ is then

$$l(\boldsymbol{\theta}|\mathbf{y}, \mathbf{w}) = \log \int \prod_{i=1}^{I} P_i(y_i|\mathbf{w}, \mathbf{z})\psi(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma})d\mathbf{z}, \tag{3}$$

where $\boldsymbol{\theta}$ represents the unknown parameters, P_i defines the measurement model for variable i, and $\psi(\cdot)$ is the multivariate normal density function with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The latent variables $\mathbf{z}$ are unknown and need to be integrated out, which requires approximation methods.

2.2 | Measurement models

Equation (3) provides a general form of the marginal log-likelihood function for GLLVMs. As mentioned above, the measurement models P_i and link functions g_i need to be defined according to the specified distribution of the observed variable. Recall that $\mathbf{z}$ is the P-dimensional vector of latent variables and let b_i be the intercept parameter, $\mathbf{a}_i$ be a vector of slope parameters, and ϕ_i be the scale parameter, all for the observed variable i. Three types of observed data, namely ordinal, continuous, and count data, are considered in this paper and the associated measurement models are given below.

1. *Ordinal responses*. One model for ordinal data is the generalized partial credit model (GPCM; Muraki, 1992) where the probability of observing each category $c \in 1, \ldots, m_i$ given the latent variables is

$$P_i(y_i = c|\mathbf{z}, \mathbf{w}) = \frac{\exp\left[\sum_{v=1}^{c} (\mathbf{a}_i'\mathbf{z} + b_{iv} + \boldsymbol{\beta}_i'\mathbf{w})\right]}{\sum_{c'=1}^{m_i} \exp\left[\sum_{v=1}^{c'} (\mathbf{a}_i'\mathbf{z} + b_{iv} + \boldsymbol{\beta}_i'\mathbf{w})\right]}, \tag{4}$$

where b_{iv} represents threshold parameters for item i and where a logit link function is assumed. A second model is the graded response model (GRM; Samejima, 1969), with

$$P_i(y_i = c|\mathbf{z}, \mathbf{w}) = P_i^*(c|\mathbf{z}, \mathbf{w}) - P_i^*(c+1|\mathbf{z}, \mathbf{w}), \tag{5}$$

where

$$P_i^*(c|\mathbf{z}, \mathbf{w}) = \frac{1}{1 + \exp(-\mathbf{a}_i'\mathbf{z} - b_{iv} - \boldsymbol{\beta}_i'\mathbf{w})}, \tag{6}$$

with $P_i^*(1|\mathbf{z}, \mathbf{w}) = 1$ and $P_i^*(m_i + 1|\mathbf{z}, \mathbf{w}) = 0$.

2. *Continuous responses*. Here we define $P_i(y_i|\mathbf{z}, \mathbf{w})$ as a conditional density function. Following Huber et al. (2004), we assume a normal distribution with an identity link and obtain

$$P_i(y_i|\mathbf{z}, \mathbf{w}) = \exp\left[\frac{y_i(b_i + \boldsymbol{\beta}_i'\mathbf{w} + \mathbf{a}_i'\mathbf{z}) - (b_i + \boldsymbol{\beta}_i'\mathbf{w} + \mathbf{a}_i'\mathbf{z})^2/2}{\phi_i} - \frac{y_i^2}{2\phi_i} - \frac{\log(2\pi\phi_i)}{2}\right]. \tag{7}$$

If the continuous data are not normally distributed (e.g, positively skewed response times), it is common to apply a log-transformation (as in De Boeck & Scalise, 2019; van der Linden, 2006, 2007; Wang et al., 2018) before applying Equation (7) in the field of educational and psychological measurement.

3. *Count responses*. We consider Poisson and negative-binomial distributions with a log link function for count data. In the former case, we have

$$P_i(y_i = c|\mathbf{z}, \mathbf{w}) = \frac{\lambda_i^c}{c!}\exp(-\lambda_i), \tag{8}$$

where $\lambda_i = \exp(b_i + \boldsymbol{\beta}_i'\mathbf{w} + \mathbf{a}_i'\mathbf{z})$. With a negative-binomial distribution, we have the conditional probability mass function (Niku et al., 2017)

$$P_i(y_i = c|\mathbf{z}, \mathbf{w}) = \frac{\Gamma\left(c + \frac{1}{\phi_i}\right)}{c!\Gamma\left(\frac{1}{\phi_i}\right)}\left(\frac{\exp(b_i + \boldsymbol{\beta}_i'\mathbf{w} + \mathbf{a}_i'\mathbf{z})}{\frac{1}{\phi_i} + \exp(b_i + \boldsymbol{\beta}_i'\mathbf{w} + \mathbf{a}_i'\mathbf{z})}\right)^c \left(\frac{1}{1 + \phi_i\exp(b_i + \boldsymbol{\beta}_i'\mathbf{w} + \mathbf{a}_i'\mathbf{z})}\right)^{\frac{1}{\phi_i}}, \tag{9}$$

where $\Gamma(\cdot)$ denotes the gamma function $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$.

2.3 | Laplace approximations for generalized linear latent variable models

As mentioned above, Equation (3) does not have an explicit solution, requiring approximation methods for parameter estimation. In this paper we utilize Laplace approximations to approximate the integrals in the likelihood function. We define $h_f(\mathbf{z}) = -\log P(\mathbf{y}_f|\mathbf{z})\psi(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\hat{h} = h_f(\hat{\mathbf{z}}_f)$, and $\mathbf{H}_f = \partial^2\hat{h}/\partial\mathbf{z}\partial\mathbf{z}'$, where $\hat{\mathbf{z}}_f$ represents the posterior modes of the latent scores of individual $f \in 1, \ldots, N$. The second-order Laplace approximation of the marginal log-likelihood for an individual f can then be written as (Shun, 1997)

$$\tilde{l}_f^{\text{Lap2}}(\boldsymbol{\theta}|\mathbf{y}) = \frac{P}{2}\log(2\pi) - \frac{1}{2}|\mathbf{H}_f| - \hat{h} + \log(1 + \epsilon_f), \tag{10}$$

with

$$\epsilon_f = -\frac{1}{2}\left[\frac{1}{4}\sum_{jklm}^{P}\frac{\partial^4 \hat{h}}{\partial z_j \partial z_k \partial z_l \partial z_m} b_{jl} b_{km} - \frac{1}{4}\sum_{jklrst}^{P}\frac{\partial^3 \hat{h}}{\partial z_j \partial z_k \partial z_l}\frac{\partial^3 \hat{h}}{\partial z_r \partial z_s \partial z_t} b_{jr} b_{kl} b_{st} \right.$$
$$\left. - \frac{1}{6}\sum_{jklrst}^{P}\frac{\partial^3 \hat{h}}{\partial z_j \partial z_k \partial z_l}\frac{\partial^3 \hat{h}}{\partial z_r \partial z_s \partial z_t}\frac{1}{6} b_{jr} b_{ks} b_{lt}\right], \tag{11}$$

where b_{jk} represents the entry of row j and column k in $\mathbf{H}_f^{-1}$. By setting $\epsilon_f = 0$ in Equation (10), the second-order Laplace approximation reduces to the first-order Laplace approximation. To efficiently compute ϵ_f, it is necessary to consider the particular model structure used and identify unique and zero entries of ϵ_f. Readers are directed to Andersson et al. (2023b) for details of the filtering procedure used to compute ϵ_f. We utilize the same estimation approach that Jin and Andersson (2020) and Andersson et al. (2023b) proposed for categorical observed variables and extend it to support continuous and count data measurement models, where the derivatives in Equation (10) and its gradient are derived analytically. Each entry $\theta \in \boldsymbol{\theta}$ of the gradient is given by

$$\nabla_f^{\theta} = \frac{\partial l_f^{\text{Lap2}}(\boldsymbol{\theta}|\mathbf{y})}{\partial \theta} + \frac{\partial \hat{\mathbf{z}}_f}{\partial \theta}\frac{\partial l_f^{\text{Lap2}}(\boldsymbol{\theta}|\mathbf{y})}{\partial \mathbf{z}}\Bigg|^{\mathbf{z}=\hat{\mathbf{z}}_f}, \tag{12}$$

where the second term is needed to account for the dependence between $\boldsymbol{\theta}$ and $\hat{\mathbf{z}}_f$. The needed derivatives (up to the fifth order) are presented in Appendix 2. A quasi-Newton method using the Broyden–Fletcher–Goldfarb–Shanno algorithm is utilized to maximize the approximated marginal log-likelihood function. The estimation approach has been implemented in the R package lamle (Andersson et al., 2023a).

The approximation error of the Laplace approximations is $O(I^{-1})$ for the first-order Laplace approximation and $O(I^{-2})$ for the second-order Laplace approximation (Bianconcini, 2014; Huber et al., 2004). Hence, the Laplace-approximated likelihood functions approach the true likelihood as the number of observed outcome variables increases, but at different rates. As a consequence of this, we expect that the Laplace-approximated estimators have the same statistical properties as the true maximum likelihood estimator (MLE) with finite sample sizes when the number of observed outcome variables is large, with the second-order Laplace approximation attaining the properties of the true MLE at a faster rate than the first-order Laplace approximation. The simulation study presented in Section 4 examines the finite-sample properties of the estimators based on first- and second-order Laplace approximations.

2.4 | Inference and model evaluation

To estimate standard errors and confidence intervals we propose to use the observed information matrix (Andersson et al., 2023b). The observed information matrix can be obtained analytically by computing the Jacobian matrix of the observed gradient in Equation (12). However, this requires accounting for the dependence between the parameter estimates and the mode estimates, which makes the computation excessively laborious. Instead, we propose approximating this matrix by computing numerical derivatives of the gradient with respect to the unknown parameters. We accomplish this by defining a vector-valued function which, for a given input vector of model parameters, computes the mode of the latent variable vector for each individual and then computes the gradient in Equation (12). The derivatives of this function are then approximated with a finite-difference approach to obtain the approximated observed information matrix.

To evaluate the model fit, we adopt a method from the factor analysis and item response theory literature that uses differences between the sample and model-implied correlation matrix of the observed outcome variables to define the standardized root mean squared residual (SRMSR) statistic. Let r_{jk} be the sample Pearson correlation and let $\hat{\rho}_{jk}$ be the model-implied Pearson correlation, for the observed outcome variables j and k. The SRMSR is then defined as (Maydeu-Olivares & Joe, 2014)

$$\mathrm{SRMSR} = \sum_{j<k} \frac{(r_{jk} - \hat{\rho}_{jk})^2}{I(I-1)/2}.$$

In the implementation here, we computed the model-implied correlations $\hat{\rho}_{jk}$ by simulating a data matrix from the estimated model with a very large sample size and then computing the correlation matrix from the resulting data matrix.

3 | MOTIVATING EXAMPLE

3.1 | Sample and data

In this section we provide an example based on the computer-based assessment of mathematics (CBAM) in PISA 2012, which aims to assess 15-year-old students' mathematical literacy and reflects the importance of using digital tools to solve mathematics tasks (OECD, 2013). Students can, for example, rotate representations of three-dimensional objects and draw points and lines to facilitate their thinking processes. The full CBAM instrument consists of 41 items from 15 units, and the items are organized into four clusters. Each student was given two clusters with 40 minutes' total testing time (OECD, 2013). PISA released three units out of 15 and the data are available on the website of the Organisation for Economic Co-operation and Development. We chose unit CM015 (*CD Production*) to illustrate the practical use of the proposed method and to guide our simulation design. CM015 presents an interactive graph and a price calculator and asks participants to enter the number of copies to discover its relationship with the cost of copying CDs using duplication and replication methods. Three items were included in CM015, with one multiple-choice and two constructed-response items. As an example, we used the Australian data set because it had the largest sample size ($N = 1824$) participating in this unit.

3.2 | Measures

Three types of indicators were extracted from each task: task scores (P1–P3), response time (T1–T3), and the number of actions (A1–A3). We pre-processed the data by log-transforming and centring the response time to deal with its positively skewed distribution (van der Linden, 2006) and by excluding outliers in terms of response times and the number of actions that were beyond the range from $Q_1 - 3 \times \mathrm{IQR}$ to $Q_3 + 3 \times \mathrm{IQR}$, where Q_1, Q_3, and IQR represent the first quantile, the third quantile, and the interquartile range, respectively. These procedures are introduced to reduce the potential influence of the high skewness and extreme values on the model estimation. After excluding these outliers and missing values, 1029 respondents remained and were used for the following analysis. A summary of the three indicators is presented in Figure 1.

It is worth noting that previous studies have extracted similar indicators from process and performance data. For example, two studies used task scores, log-transformed response time, and log-transformed action counts as observed indicators for cognitive proficiency, speed, and exploration behaviour (Costa & Chen, 2023; De Boeck & Scalise, 2019). In their studies, all the indicators were regarded as continuous variables. In another study, finer-grained process indicators, the longest duration and the number of non-targeted operations, were defined to measure latent variables representing planning and non-targeted exploration, respectively (Zhang et al., 2023). These process indicators were subsequently recoded as ordinal data in their paper. Different from these studies (Costa & Chen, 2023;

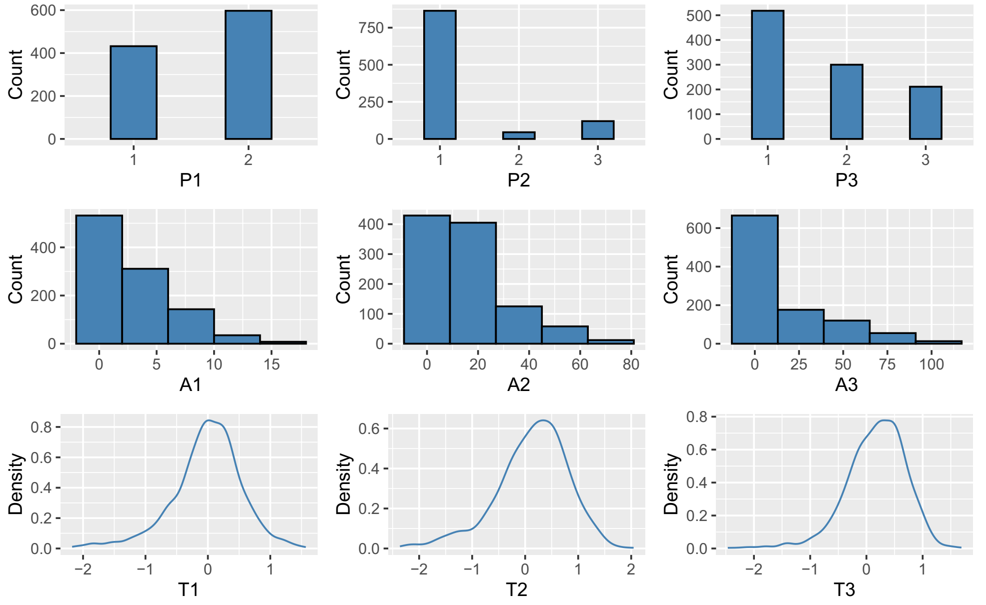


FIGURE 1 Summary of observed indicators. P1–P3, A1–A3, and T1–T3 represent scored responses, the number of actions, and transformed response times of items 1–3, respectively.

De Boeck & Scalise, 2019; Zhang et al., 2023), this study treats the measures as they are; that is, task scores are considered as binary or ordinal data, log-transformed response times as continuous data, and the number of actions as count data.

3.3 | Analysis

A two-step procedure was used in the motivating example. The first step included three unidimensional measurement models. Specifically, P1–P3 were used to measure mathematical literacy, indicating 'an individuals capacity to formulate, employ, and interpret mathematics in a variety of contexts' (OECD, 2013, p. 17); T1–T3 were used to measure speed, indicating how fast or slowly respondents completed the tasks; and A1–A3 were used to measure interactivity, indicating the tendency to interact with the computer. Second-order Laplace approximations were used for model estimation. All models converged and the parameter estimates are used as a reference for the following simulation studies.

In the next step, we combined the three indicators in a single model using GLLVMs with the residual correlations of indicators from the same stimulus or item considered (ModRes, see Figure 6) or not (ModInd, see Figure 2). Equality constraints were added to residual factor loadings from the same item for simplification. In total, we estimated 2 (model structure: ModInd or ModRes) × 2 (count type: Poisson or negative-binomial) × 2 (algorithms: Lap1 or Lap2) = 8 models. By comparison, previous studies using similar indicators primarily adopted confirmatory factor analysis (Costa & Chen, 2023; De Boeck & Scalise, 2019) or item factor analysis (Zhang et al., 2023). However, converting continuous variables and count variables into a few categories can cause loss of detailed information embedded in the data (De Leon & Chough, 2013).

3.4 | Results

In this subsection we present the results of multidimensional GLLVMs in terms of model fit and timing information (Table 1). Lap2 achieved convergence in all models, whereas Lap1 failed to converge

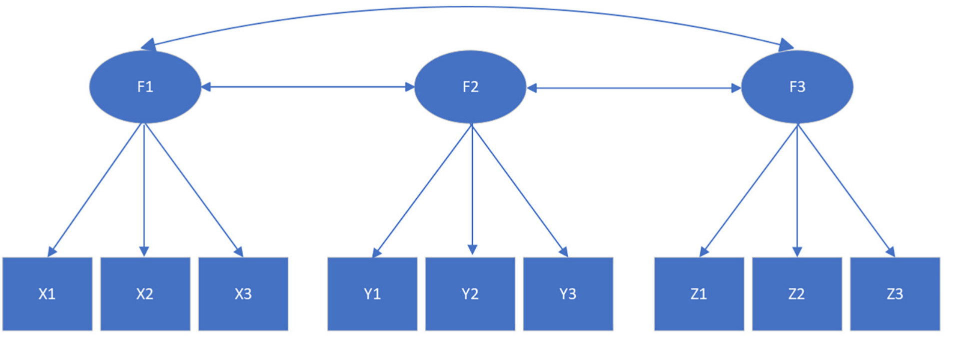


FIGURE 2 Model illustration of three-dimensional GLLVMs. X, Y, and Z denote three different types of indicators. F1–F3 denote latent variables.

TABLE 1 Model fit and timing (seconds) of generalized linear latent variable models using the empirical data

		ModInd			ModRes		
Count	Method	BIC	SRMSR	Timing	BIC	SRMSR	Timing
Poisson	Lap1	32,925	0.127	15.7	–	–	–
Poisson	Lap2	32,897	0.126	16.8	28,325	0.088	71.5
Negbin	Lap1	28,372	0.086	11.4	–	–	–
Negbin	Lap2	28,353	0.086	13.1	27,923	0.070	73.9

Note: Lap1 failed to compute the observed information matrix for ModRes, regardless of the count data model used, and indicated non-convergence. Lap1, first-order Laplace; Lap2, second-order Laplace; BIC, Bayesian information criterion; SRMSR, standardized root mean squared residuals; Negbin, negative-binomial.

with ModRes regardless of the type of count data model. Table 1 indicates that the estimation time of ModInd was around 15 s, whereas ModRes required longer time (approximately 70 s). A negative-binomial distribution fitted better than the Poisson distribution when the residual factors were not considered. After accounting for the residual correlations among the observed variables (ModRes), both the BIC and SRMSR values decreased, which indicated a substantially better fit to the data. This conclusion was in line with existing studies (Costa & Chen, 2023; De Boeck & Scalise, 2019; Zhang et al., 2023), which reported significant within-task relationships after considering the covariance of the latent variables. This underscores the importance of integrating task-level correlations into the joint model of performance and process data, something which several previous studies have not considered (Man et al., 2022; Man & Harring, 2023; Qiao et al., 2022).

Given that ModRes with a negative-binomial distribution using Lap2 had the lowest BIC, we concluded that this model best represented the observed data and present the item parameter estimates in Table 2. The factor loadings were all positive, with varied magnitudes found for the performance model and the action model. The residual factor loadings indicated a consistent magnitude of within-task relationships across the CM015 unit. In terms of the relationships of the latent variables, our results mirrored those of the study of Costa and Chen (2023). This similarity is expected since we utilized the same tasks, albeit in different countries. Specifically, the latent variable correlations of mathematical literacy with interactivity (.876) and speed (.658) are moderately to highly positive. This suggests that individuals who interact more actively with the problem environment (e.g., the computer interface), or who invest more time in the tasks, generally exhibit higher mathematical literacy. This makes sense because a correct solution requires appropriate operations and sufficient time to represent, plan, and implement the strategy.

In empirical studies investigating the relationship between response accuracy and response times, a common phenomenon is the speed–accuracy trade-off (Schnipke & Scrams, 2002). This refers to

TABLE 2 Parameter estimates (standard error) of the final model

	Indicator	Slope	Intercept 1	Intercept 2	Scale
Performance	P1	1.361 (0.125)	0.441 (0.088)	–	–
	P2	2.695 (0.312)	−3.346 (0.317)	−3.987 (0.355)	–
	P3	3.093 (0.288)	0.120 (0.139)	−3.042 (0.264)	–
Action	A1	0.515 (0.087)	1.101 (0.029)	–	0.080 (0.020)
	A2	0.421 (0.025)	2.679 (0.025)	–	0.209 (0.017)
	A3	1.046 (0.039)	2.315 (0.044)	–	0.272 (0.038)
Time	T1	0.288 (0.019)	−0.018 (0.017)	–	0.108 (0.012)
	T2	0.481 (0.023)	0.132 (0.021)	–	0.079 (0.015)
	T3	0.388 (0.018)	0.170 (0.017)	–	0.009 (0.013)
Residual	Task 1	0.355 (0.018)	–	–	–
	Task 2	0.382 (0.018)	–	–	–
	Task 3	0.354 (0.023)	–	–	–

Note: P2 and P3 have three categories and thus have two intercept parameters. Correlations between latent variables: Cor(mathematical literacy, interactivity) = .876 (SE = 0.023), Cor(mathematical literacy, speed) = .658 (SE = 0.035), and Cor(interactivity, speed) = .584 (SE = 0.032).

the phenomenon that when an individual works at a quick pace, her/his accuracy tends to decline, whereas working at a slower pace tends to increase response accuracy. Van der Linden (2007) argued that the speed–accuracy trade-off is a within-person trade-off; however, the relationship between ability and speed at a cross-person level can be positive, negative, or non-significant depending on the item content and assessment settings (Schnipke & Scrams, 2002). In the current empirical studies, we found a positive relationship between mathematical literacy and speed, similar to the results of Costa and Chen (2023), but different from the negative relationship of De Boeck and Scalise (2019) that employed collaborative problem-solving tasks. Different task domains can be an explanation for the varied results.

In addition, respondents with a higher level of interactivity typically spent more time on the tasks, evidenced by a correlation of .584 between interactivity and speed. This relationship can be interpreted in two ways: performing more operations inherently requires more time; and respondents who were more engaged in the tasks tended to make more effort to solve the tasks. Note that although interactivity and speed are moderately correlated, they represent different latent traits. Speed describes the pace of work, while interactivity refers to the extent to which respondents tend to conduct operations on the computer and indicates exploratory behaviour (Costa & Chen, 2023).

4 | SIMULATIONS

We conducted two simulation studies to assess the performance of Laplace approximations in the context of mixed data using newly developed code written in C++ and R 4.1.2 (Andersson et al., 2023a; R Core Team, 2021). In Simulation 1, we considered three-dimensional GLLVMs with three types of indicators: ordinal, continuous, and count data. In Simulation 2, we also considered residual correlations between indicators from the same task.

We evaluated the performance of the proposed methods in terms of convergence rate, estimation time, and the recovery of model parameters. Convergence was determined by satisfying three criteria: the algorithm stopped before 500 iterations, the approximated observed information matrix was positive definite, and all parameter estimates had absolute bias lower than 5. Regarding parameter recovery, we computed the bias and the mean squared error (MSE) to assess the accuracy and precision of parameter estimates via

$$\text{bias}_\theta = \sum_{r=1}^{R} (\hat{\theta}^r - \theta)/R, \tag{13}$$

and

$$\text{MSE}_\theta = \sum_{r=1}^{R} (\hat{\theta}^r - \theta)^2/R, \tag{14}$$

where θ and $\hat{\theta}^r$ denote the true value and the parameter estimate in replication $r \in 1, \ldots, R$, respectively.

4.1 | First simulation study

4.1.1 | Design

In Simulation 1, we considered three correlated latent variables with ordinal, continuous, and count data as indicators, respectively. We illustrate the model in Figure 2. Four experimental factors were manipulated: the distribution of the count data model (Type: Poisson or negative-binomial distributions); the number of observed outcome variables per dimension (J: 3 or 6); the covariance between the latent variables (ρ: small and large), and sample size (N: 250 and 1000). This design resulted in $2 \times 2 \times 2 \times 2 = 16$ conditions. We used 1000 replications for each condition. To determine the ranges of the simulated parameters, we used the result from the motivating example and the PISA 2018 item parameter pool. Specifically, the item pool provides the estimates of item parameters in terms of task scores, and we used the 10% and 90% quantiles as the range of the item parameters for the ordinal data model. For the models corresponding to response time and number of actions, we generated the item parameters based on the above motivating example. The item parameters and the covariance of the latent variables used in the simulation are presented in Appendix 1: Tables A1–A4. Latent variables were randomly simulated from a multivariate normal distribution with a zero mean vector and unit variances. The observed data were then generated based on Equations 4–9. With the data sets generated, we estimated unidimensional models for each type of outcome variable, and if convergence was achieved, then the parameter estimates were used as the starting values to estimate the three-dimensional model with the first- and second-order Laplace approximation methods. If convergence failed for any of the unidimensional models, we used non-informative default starting values instead.

4.1.2 | Results

The convergence rates and estimation times of each algorithm are presented in Table 3. Averaged across all conditions, Lap1 and Lap2 reached high convergence rates, exceeding 98%. Lap2 achieved a convergence rate higher than or equal to Lap1 under all conditions in Simulation 1. Increasing the number of observed outcome variables, the sample size, or the covariance of the latent variables improved the convergence rates. Among the converged replications, both methods completed the estimation procedure within an average of 4 and 13 seconds for small and large sample sizes, respectively. As expected, Lap1 required less time than Lap2 in all conditions, but the difference was minor. For the experimental factors, Table 3 suggests that increasing the number of observed outcome variables or the sample size increased estimation time.

Next, we summarize the recovery of the parameters. Overall, the estimators showed small bias, indicating that the methods recovered the true parameters accurately. Both Lap1 and Lap2 produced small and similar biases in the parameter estimates for both the continuous data and the Poisson-distributed data, with an average absolute bias smaller than 0.002. To illustrate the differences between the Lap1

TABLE 3 Convergence rate and timing (seconds) of Lap1 and Lap2 in Simulation 1

Sample size	Type	J	ρ	Convergence rate		Timing	
				Lap1	Lap2	Lap1	Lap2
250	Pois	3	Small	93.3%	97.5%	2.3	2.7
250	Pois	3	Large	99.3%	100%	2.3	2.8
250	Pois	6	Small	100%	100%	3.4	3.9
250	Pois	6	Large	100%	100%	3.3	3.7
250	Negbin	3	Small	94.8%	95.4%	2.3	2.9
250	Negbin	3	Large	98.7%	99.5%	2.4	2.8
250	Negbin	6	Small	100%	100%	4.0	4.6
250	Negbin	6	Large	100%	100%	3.8	4.4
1000	Pois	3	Small	93.6%	100%	7.7	9.4
1000	Pois	3	Large	99.7%	100%	7.8	9.6
1000	Pois	6	Small	100%	100%	14.8	17.1
1000	Pois	6	Large	100%	100%	14.2	16.1
1000	Negbin	3	Small	97.9%	99.9%	8.0	9.7
1000	Negbin	3	Large	99.1%	100%	7.5	10.0
1000	Negbin	6	Small	100%	100%	14.5	16.9
1000	Negbin	6	Large	100%	100%	13.7	15.9
250	Overall			98.3%	99.0%	3.0	3.5
1000	Overall			98.8%	100%	11.0	13.1

Note: Lap1, first-order Laplace; Lap2, second-order Laplace; Pois, Poisson; Negbin, negative-binomial; J, the number of observed outcome variables per dimension; ρ, covariance of latent variables.

and Lap2 results, we present only the results related to relatively large biases, namely cases with average absolute bias greater than 0.01. These cases involved the ordinal data model, the count data model with a negative-binomial distribution, and the covariances of the latent variables. Because sample size plays an important role in the results, we separately present the parameter recovery for small and large sample sizes.

We summarize the average absolute bias for the conditions with a small sample size in Table 4 and plot the biases in Figures 3a–5a. Overall, Lap2 outperformed Lap1 in estimating the slope parameters in the negative-binomial model, and especially in estimating the covariance of the latent variables (average absolute bias: Lap1 = 0.0219 and Lap2 = 0.0039), under all conditions. However, Lap1 produced smaller absolute biases than Lap2 in the estimation of the slope and threshold parameters in the graded response model when the number of observed outcome variables was three and the covariance was small, as shown in Table 4. This result was somewhat unexpected, and we further investigated the fifth condition shown in Table 3 by comparing the results to AGHQ with 13 quadrature points, which has a fifth-order accuracy (Jin & Andersson, 2020) and can essentially be viewed as the true MLE. Following a similar approach of Andersson et al. (2023b) and Zhang and Chen (2022), we then compared Lap1 and Lap2 to the AGHQ estimates which indicated that Lap2 had more similar estimates to AGHQ. Regarding the effects of the simulating factors, Table 4 suggests that both the number of observed outcome variables and the magnitude of the covariance showed a positive effect on the estimation accuracy, especially for Lap2.

For sample size 1000, we summarize the average absolute bias under each simulating factor in Table 5. The results show that Lap2 produced smaller absolute biases in estimating all the parameters than Lap1 did, with the largest differences in the slope parameters for the ordinal data (average absolute bias: Lap1 = 0.0367 and Lap2 = 0.0086) and the covariances of the latent variables (average absolute bias: Lap1 = 0.0172 and Lap2 = 0.0016). With respect to the influence of the simulating factors, Table 5

TABLE 4 Average absolute bias of selected parameter estimates under the simulating factors in Simulation 1 (sample size = 250)

			Ordinal model		Negbin model		
Factor	**Level**	**Method**	**Slope**	**Thresholds**	**Slope**	**Scale**	**Covariance**
J	3	Lap1	0.0417	0.0073	0.0141	0.0078	0.0341
	6	Lap1	0.0266	0.0102	0.0050	0.0049	0.0097
	3	Lap2	0.0558	0.0268	0.0033	0.0080	0.0056
	6	Lap2	0.0210	0.0106	0.0024	0.0061	0.0022
ρ	Small	Lap1	0.0406	0.0103	0.0094	0.0077	0.0205
	Large	Lap1	0.0226	0.0081	0.0067	0.0040	0.0233
	Small	Lap2	0.0464	0.0206	0.0028	0.0076	0.0043
	Large	Lap2	0.0188	0.0114	0.0027	0.0059	0.0036
Overall		Lap1	0.0316	0.0092	0.0080	0.0059	0.0219
		Lap2	0.0326	0.0160	0.0027	0.0067	0.0039

Note: Lap1, first-order Laplace; Lap2, second-order Laplace; Negbin, negative-binomial; J, the number of observed outcome variables per dimension; ρ, covariance of latent variables.

suggests that increasing the number of observed outcome variables or the covariance between the latent variables improves the accuracy of the parameter estimates, especially for Lap1, and that when estimating the covariance parameters, Lap2 was less influenced by the simulating factors and recovered the covariances satisfactorily under all conditions, whereas Lap1 showed a relatively large bias, especially when the number of observed outcome variables was three. We also plot the bias of the parameter estimates for the ordinal data, the count data, and the covariances in Figures 3b, 4b, and 5b, respectively. The figures suggest that Lap1 can produce relatively large biases for certain parameters, whereas the biases of the model parameters were generally small using Lap2.

In terms of the estimation precision, Lap2 had a larger MSE than Lap1 when the sample size, the number of observed outcome variables, and the true covariance were small. In other cases, Lap1 and Lap2 shared a similar level of estimation precision. The detailed results regarding the MSE can be found in the Appendix S1.[1]

In summary, both algorithms achieved high convergence rates within a short estimation time, but Lap2 had slightly higher convergence rates when the number of observed outcome variables was three. When the sample size was large, Lap2 estimated the model parameters accurately and outperformed Lap1, especially in terms of the estimation of the graded response model and the covariances of the latent variables. However, a small sample size and few observed outcome variables can result in a larger bias and MSE of the slope and threshold parameters with ordinal data, particularly for Lap2.

4.2 | Second simulation study

4.2.1 | Design

In Simulation 2 we considered residual correlations of indicators from the same stimulus to capture the task-specific effect for a single stimulus (Figure 6). We specified that the residual latent variables impose the same effect on the indicators; namely, we set equal residual factor loading across indicators from the same stimulus (e.g., equal residual factor loadings for X1, Y1, and Z1). Accordingly, we added

[1]https://osf.io/nec8m/?view_only=fc93a3e633ea47eba597357722fe8c83.

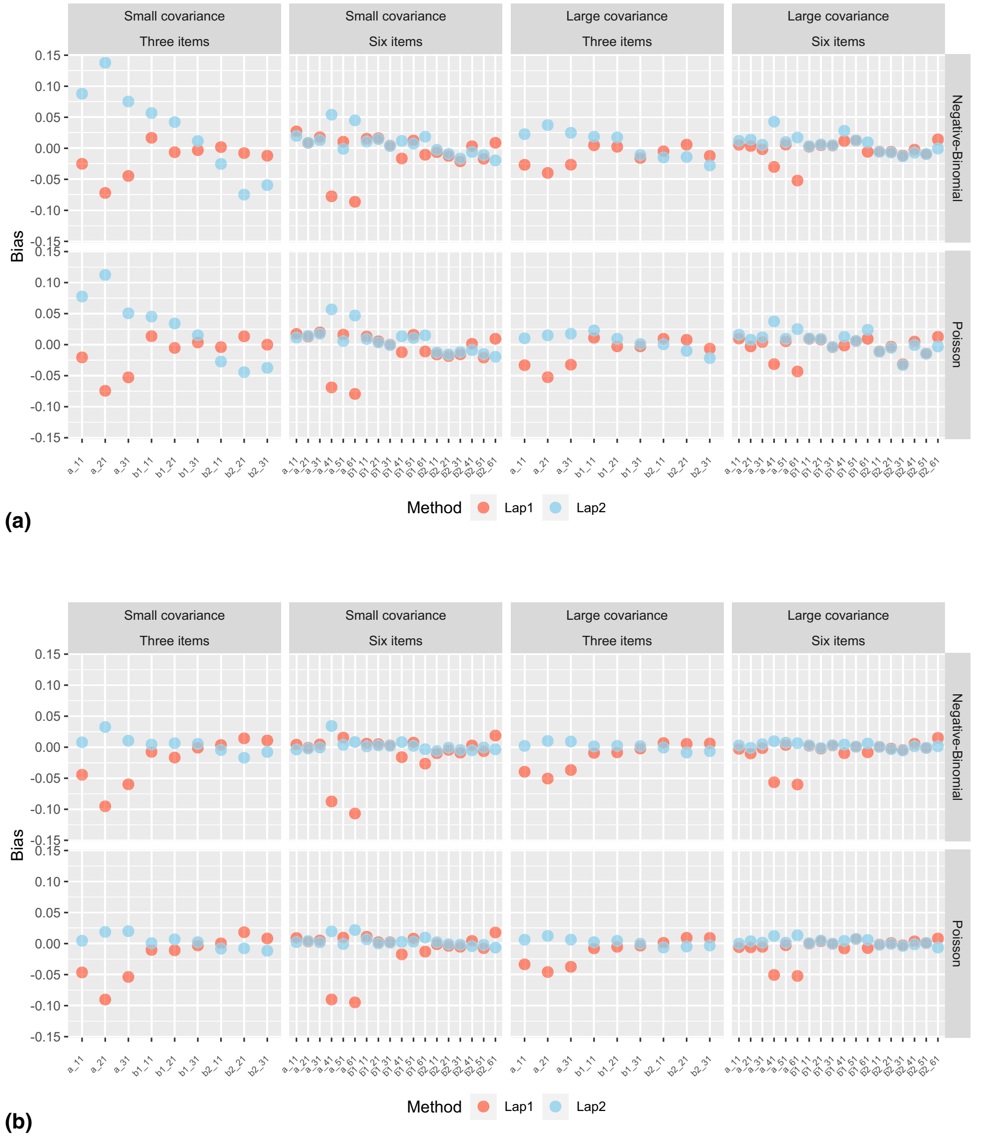


FIGURE 3 Bias of the slope (a) and threshold ($b1$ and $b2$) parameters in the ordinal data model in Simulation 1. a_ip represents the slope parameter of latent variable F_p on item i. 'Three items' and 'six items' indicate three or six observed outcome variables per dimension. (a) Bias of the estimates of the ordinal data model when the sample size was 250. (b) Bias of the estimates of the ordinal data model when the sample size was 1000.

the magnitude of the residual factor loadings (small or large) to the simulation design. Specifically, large residual factor loadings were generated from $U(0.4, 0.8)$, the same distribution as the slope parameter for the continuous data and count data, whereas the small residual factor loading was set to half of the large one. The true values of the residual factor loadings are presented in Table A4. In sum, Simulation 2 resulted in 2 (Poisson or negative-binomial distribution) $\times$ 2 (3 or 6 observed outcome variables per dimension) $\times$ 2 (small or large covariance) $\times$ 2 (small or large residual factor loading) $\times$ 2 (small or large sample size) = 32 conditions, and we generated 1000 data sets under each condition. The number of latent variables in Simulation 2 was either six or nine. Both Lap1 and Lap2 were applied to analyse the data sets.

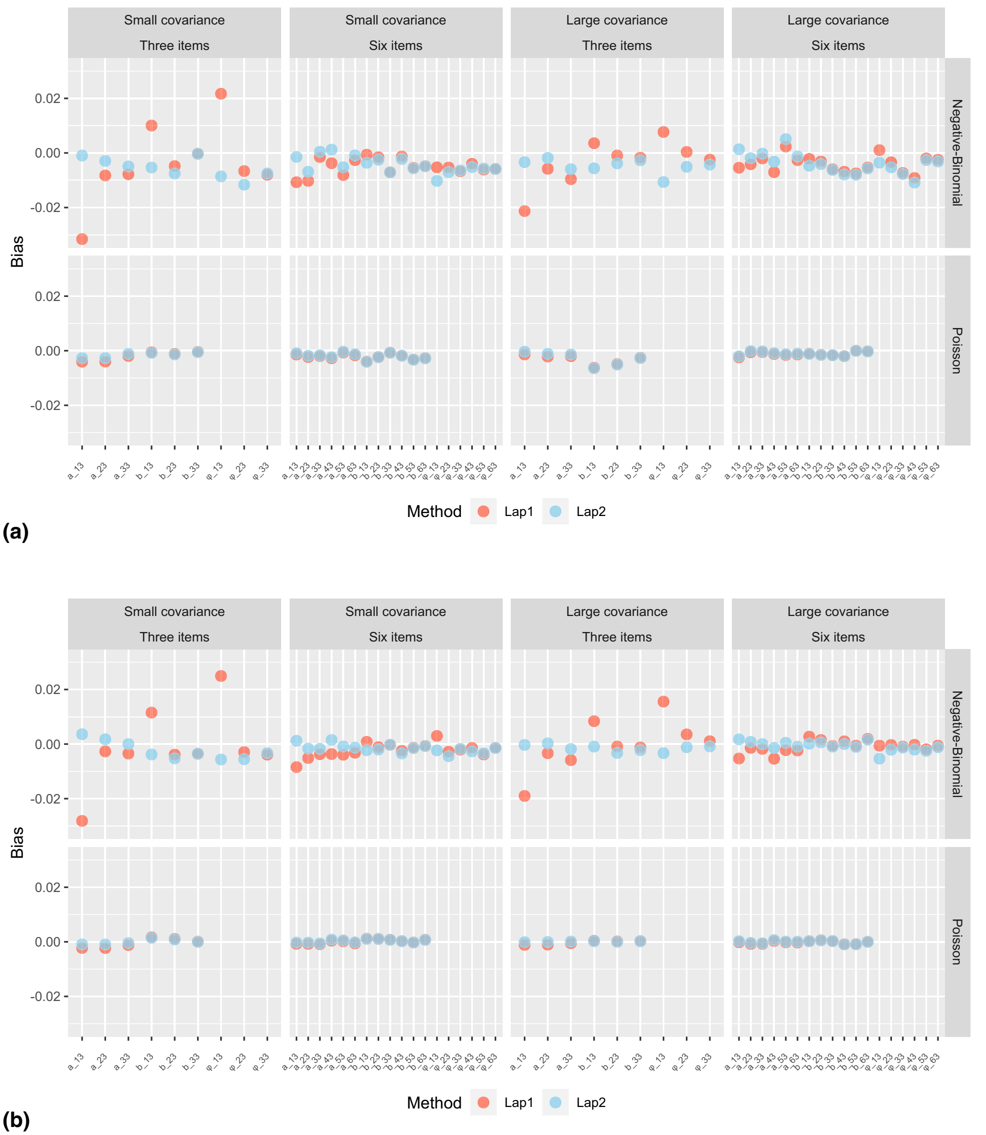
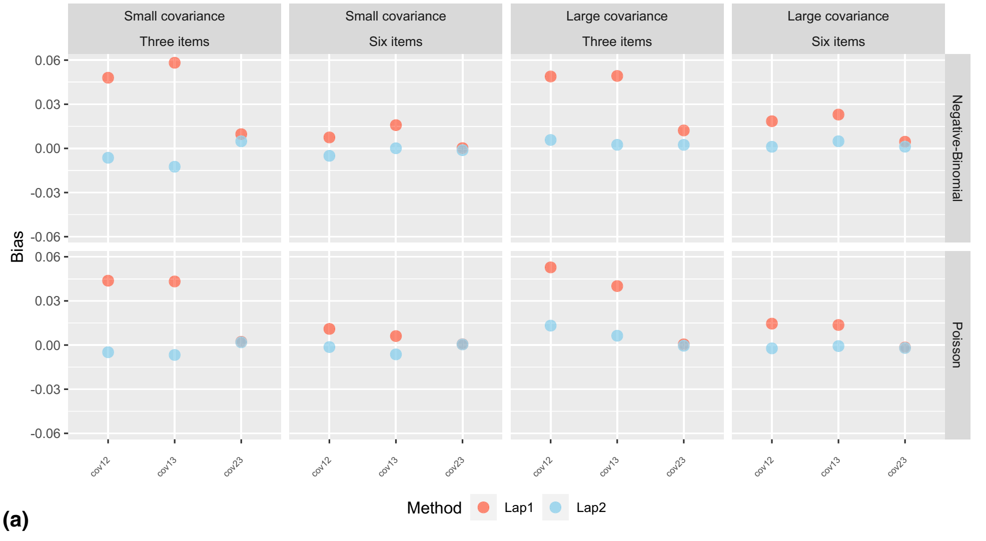


FIGURE 4 Bias of the slope (a), threshold (b), and scale (φ) parameters in the count data model in Simulation 1. a_ip represents the slope parameter of latent variable F_p on item i. 'Three items' and 'six items' indicate three or six observed outcome variables per dimension. (a) Bias of the estimates of the count data model when the sample size was 250. (b) Bias of the estimates of the count data model when the sample size was 1000.

4.2.2 | Results

We now summarize the results from Simulation 2. In line with Simulation 1, Lap2 reached a higher average convergence rate than Lap1 did in both small and large sample size conditions, as shown in Tables 6 and 7. Both Lap1 and Lap2 achieved over 99% convergence rates when the number of observed outcome variables was six, while decreasing the number of observed outcome variables can lead to a decrease in convergence rates, especially for Lap1. When there were three observed outcome variables per dimension, increasing the residual factor loadings was associated with higher convergence rates for large sample sizes, but this effect was less pronounced when the sample size was 250. In terms of estimation time, Lap2 required more time than Lap1, especially as the number



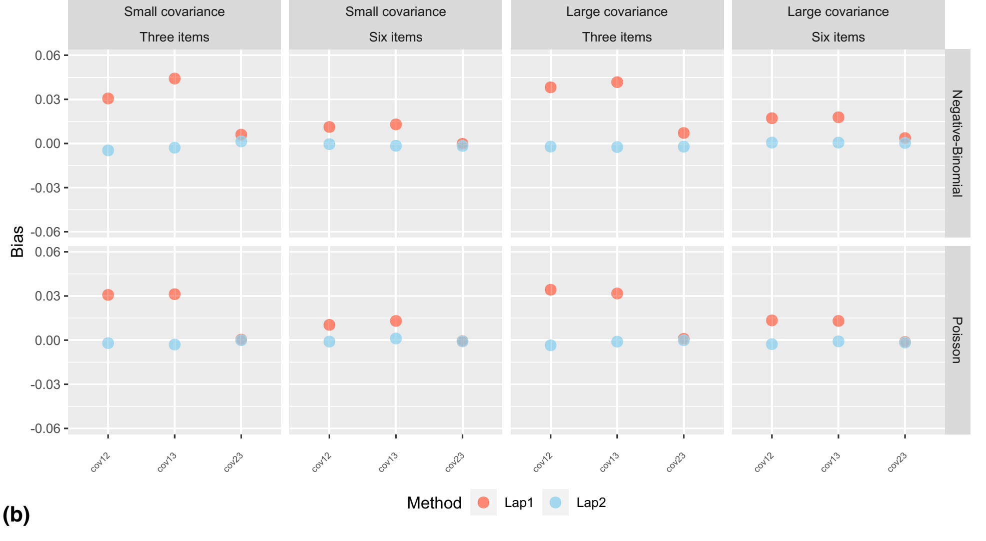


FIGURE 5 Bias of the covariance parameters in Simulation 1. cov_{pq} represents the covariance estimate of latent variable F_p and F_q. 'Three items' and 'six items' indicate three or six observed outcome variables per dimension. (a) Bias of the estimates of the covariance when the sample size was 250. (b) Bias of the estimates of the covariance when the sample size was 1000.

of observed outcome variables or sample size increased. For both algorithms, an increase in the number of observed outcome variables, the sample size, or the magnitude of residual factor loadings resulted in a longer estimation time.

Next, we summarize parameter recovery in Simulation 2. First, regardless of the algorithm used, the sample size, the number of observed outcome variables, and the covariance of the latent variables positively influenced the estimation accuracy and precision. Second, larger biases (i.e., absolute bias larger than 0.01) were observed in parameter estimates for ordinal data, count data with a negative-binomial distribution, and covariance. Third, Lap2 generally outperformed Lap1 in estimating the model parameters when the sample size was 1000. However, Lap1 outperformed Lap2 in estimating the ordinal data model parameters when both the sample size and the number of observed outcome variables were small.

TABLE 5 Average absolute bias of selected parameter estimates under the simulating factors in Simulation 1 (sample size = 1000)

Factor	Level	Method	Ordinal model		Negbin model		Covariance
			Slope	Thresholds	Slope	Scale	
J	3	Lap1	0.0528	0.0074	0.0104	0.0087	0.0247
	6	Lap1	0.0286	0.0064	0.0039	0.0016	0.0096
	3	Lap2	0.0117	0.0052	0.0013	0.0033	0.0021
	6	Lap2	0.0070	0.0032	0.0011	0.0025	0.0011
ρ	Small	Lap1	0.0456	0.0087	0.0069	0.0051	0.0160
	Large	Lap1	0.0278	0.0048	0.0052	0.0028	0.0184
	Small	Lap2	0.0110	0.0046	0.0015	0.0034	0.0017
	Large	Lap2	0.0062	0.0031	0.0009	0.0022	0.0015
Overall		Lap1	0.0367	0.0067	0.0061	0.0039	0.0172
		Lap2	0.0086	0.0038	0.0012	0.0028	0.0016

Note: Lap1, first-order Laplace; Lap2, second-order Laplace; Negbin, negative-binomial; J, the number of observed outcome variables; ρ, covariance of latent variables.

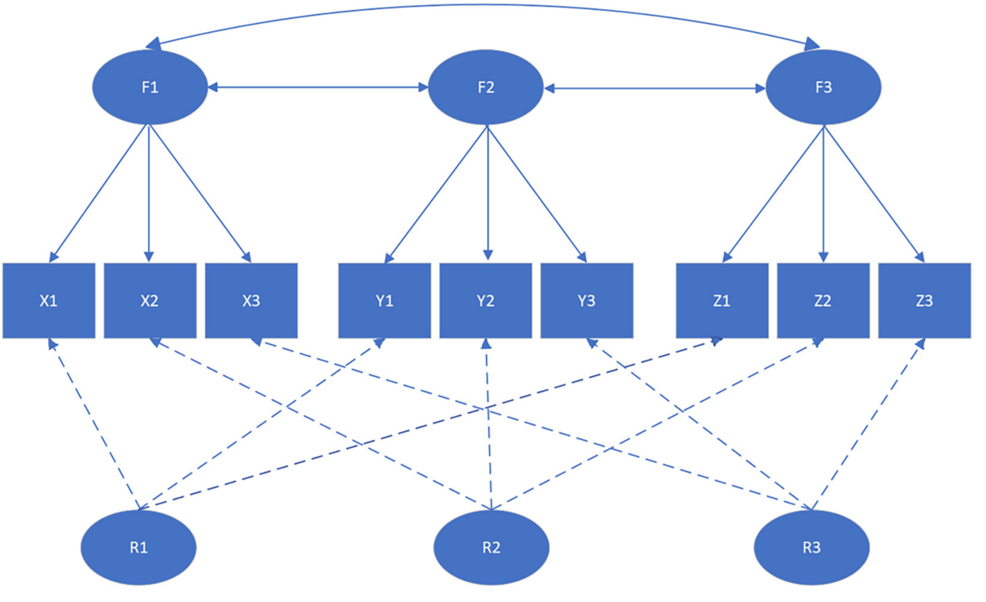


FIGURE 6 Model illustration including residual latent variables, where X, Y, and Z denote three different types of indicators, F1–F3 are latent variables measured by the X-, Y-, and Z-variables, respectively, and R1–R3 are residual latent variables.

Simulation 2 introduced residual factors to the model and we therefore focus on the impact of these. We plot the biases of the residual factor loading estimates with negative-binomial count data in Figure 7 and summarize the influences of the magnitude of the residual factor loadings in Table 8. The figure indicates that Lap2 produced less biased estimates of residual factor loadings, especially when the sample size was 250. We also manipulated the magnitude of the residual factor loadings in Simulation 2. As Figure 7 reveals, a higher magnitude of residual factor loadings was related to less biased estimates of residual factor loadings. However, Table 8 suggests that higher residual factor loadings were associated with greater biases in other parameter estimates.

TABLE 6 Convergence rates and timing (seconds) of Lap1 and Lap2 in Simulation 2 (sample size = 250)

					Convergence rate		Timing	
Sample size	Type	J	ρ	Residuals	Lap1	Lap2	Lap1	Lap2
250	Pois	3	Small	Small	92.1%	95.2%	4.1	15.1
250	Pois	3	Small	Large	93.2%	95.6%	4.3	16.0
250	Pois	3	Large	Small	95.2%	100%	3.9	15.1
250	Pois	3	Large	Large	94.0%	99.1%	4.2	15.3
250	Negbin	3	Small	Small	91.5%	91.8%	4.1	16.7
250	Negbin	3	Small	Large	88.6%	91.9%	4.3	16.9
250	Negbin	3	Large	Small	91.5%	98.0%	4.2	17.0
250	Negbin	3	Large	Large	94.6%	98.9%	4.2	17.1
250	Pois	6	Small	Small	100%	100%	11.5	92.1
250	Pois	6	Small	Large	99.5%	99.7%	12.3	98.7
250	Pois	6	Large	Small	100%	100%	10.3	83.7
250	Pois	6	Large	Large	99.0%	99.3%	12.6	101.8
250	Negbin	6	Small	Small	100%	100%	12.8	108.4
250	Negbin	6	Small	Large	99.8%	100%	12.8	109.7
250	Negbin	6	Large	Small	100%	100%	12.6	106.4
250	Negbin	6	Large	Large	100%	100%	12.9	110.9
Overall					96.2%	98.1%	8.2	58.8

Note: Lap1, first-order Laplace; Lap2, second-order Laplace; Pois, Poisson; Negbin, negative-binomial; J, the number of observed outcome variables; ρ, covariance of latent variables.

TABLE 7 Convergence rates and timing (seconds) of Lap1 and Lap2 in Simulation 2 (sample size = 1000)

					Convergence rate		Timing	
Sample size	Type	J	ρ	Residuals	Lap1	Lap2	Lap1	Lap2
1000	Pois	3	Small	Small	87.6%	97.1%	13.7	56.5
1000	Pois	3	Small	Large	93.3%	97.6%	15.1	63.5
1000	Pois	3	Large	Small	84.7%	95.5%	13.2	52.7
1000	Pois	3	Large	Large	91.1%	99.8%	15.3	59.6
1000	Negbin	3	Small	Small	82.9%	97.7%	14.0	61.8
1000	Negbin	3	Small	Large	92.0%	98.2%	15.2	69.1
1000	Negbin	3	Large	Small	81.7%	95.2%	14.4	60.8
1000	Negbin	3	Large	Large	90.5%	99.8%	15.5	66.0
1000	Pois	6	Small	Small	100%	100%	40.1	351.4
1000	Pois	6	Small	Large	99.6%	99.7%	47.4	410.8
1000	Pois	6	Large	Small	100%	100%	38.0	333.9
1000	Pois	6	Large	Large	99.2%	99.7%	48.7	423.0
1000	Negbin	6	Small	Small	100%	100%	44.8	413.7
1000	Negbin	6	Small	Large	100%	100%	44.7	416.8
1000	Negbin	6	Large	Small	100%	100%	43.9	405.9
1000	Negbin	6	Large	Large	100%	100%	45.0	415.5
Overall					93.9%	98.8%	29.32	228.81

Note: Lap1, first-order Laplace; Lap2, second-order Laplace; Pois, Poisson; Negbin, negative-binomial; J, the number of observed outcome variables; ρ, covariance of latent variables.

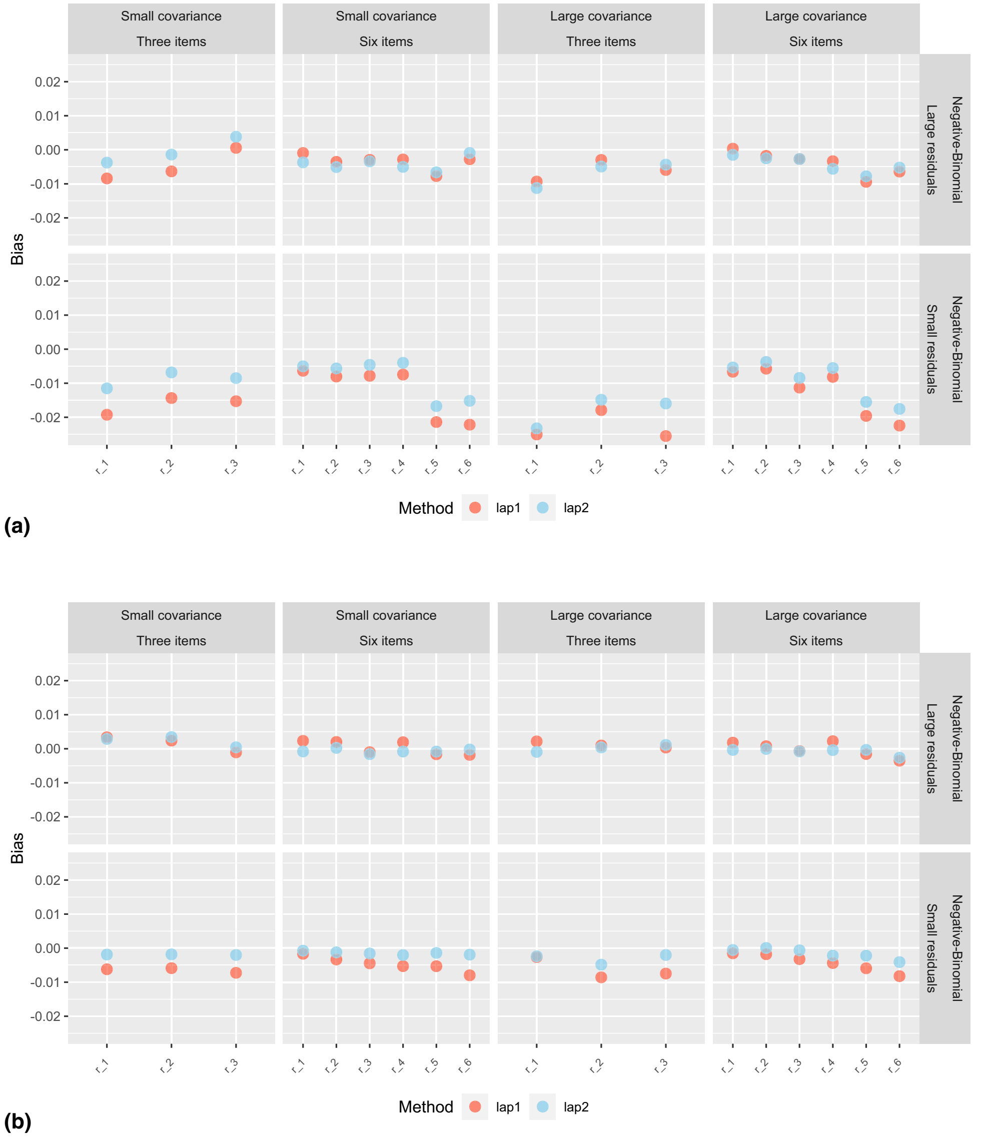


FIGURE 7 Bias of the estimates of the residual factor loadings in Simulation 2. 'Three items' and 'six items' indicate three or six observed outcome variables per dimension. (a) Bias of residual factor loading estimates when the sample size was 250. (b) Bias of residual factor loading estimates when the sample size was 1000.

5 | DISCUSSION

The advent of complex measurement tools has facilitated research by providing more detailed information on the response process. As a result, the data often consist of different types. In this paper we implemented first- and second-order Laplace approximations to jointly model a mixture of ordinal, continuous, and count data within the framework of GLLVMs. An empirical study demonstrated the usage of the proposed methods in practice, and two simulation studies were conducted to examine the performance of both algorithms in the scenario of computer-based assessment with process indicators and performance data. The results indicated that Lap2 had a higher convergence rate and better parameter recovery compared to Lap1, especially when the sample size was large. However, Lap2 took longer to estimate, especially with complex models that incorporated residual factors.

TABLE 8 Absolute bias of parameter estimates under different levels of the residual factor loadings in Simulation 2

			Ordinal model		Negbin model		
Sample size	*Res*	Method	Slope	Thresholds	Slope	Scale	Covariance
250	Small	Lap1	0.0369	0.0083	0.0101	0.0067	0.0263
	Large	Lap1	0.0384	0.0094	0.0141	0.0080	0.0275
	Small	Lap2	0.0377	0.0191	0.0025	0.0104	0.0050
	Large	Lap2	0.0477	0.0212	0.0045	0.0112	0.0057
1000	Small	Lap1	0.0402	0.0075	0.0098	0.0066	0.0213
	Large	Lap1	0.0440	0.0074	0.0116	0.0100	0.0214
	Small	Lap2	0.0110	0.0044	0.0017	0.0030	0.0017
	Large	Lap2	0.0155	0.0060	0.0017	0.0032	0.0033
Overall		Lap1	0.0399	0.0081	0.0114	0.0078	0.0241
		Lap2	0.0280	0.0127	0.0026	0.0070	0.0079

Note: Lap1, first-order Laplace; Lap2, second-order Laplace; Negbin, negative-binomial; *Res*, the residual factor loadings.

The experimental factors impacted the results in the following ways. First, higher sample sizes improved the estimation accuracy, especially for Lap2. Second, a larger number of observed outcome variables had a positive influence on convergence and parameter recovery. As the number of observed outcome variables increased, both Lap1 and Lap2 approximated the marginal log-likelihood better and the error of the estimators decreased (Huber et al., 2004). Moreover, higher-order Laplace approximations have a faster rate of approximation error decrease (Andersson & Xin, 2021), which means that fewer indicators are needed to achieve a high accuracy. Third, the magnitude of the covariance between latent variables had a positive effect on the estimation. Fourth, the magnitude of the residual factor loadings had some influence on the convergence and estimation time but a mixed effect on parameter recovery. Larger residual factor loadings implied stronger task-specific effects, which should be considered in the model specification. In the empirical study, we found that considering the residual correlation improved the model fit and that the residual factor loadings were substantial. In this study, we imposed equality restrictions on residual factor loadings from the same stimulus. If there are prior hypotheses about the residual factor loadings, it is flexible to take them into account and specify such a hypothesized model as long as it is identifiable.

5.1 | Contributions and limitations

The study makes significant contributions in several ways. First, we derived the second-order Laplace approximation likelihood form of Poisson and negative-binomial distributions for count data, which extended existing research that only included first-order Laplace approximations for count data (Niku et al., 2017). Employing Lap2 can improve the estimation accuracy in terms of the slope and scale parameters compared to using Lap1 for data with a negative-binomial distribution. Second, the current study provided a fast yet accurate solution for a combination of count data, continuous data, and ordinal data within the framework of GLLVMs. Compared to a Bayesian or quadrature approach, Laplace approximations greatly increase the computational efficiency in high-dimensional GLLVMs (Huber et al., 2004). Our research considered different types of observed variables and potential residual correlations between the indicators in a single model. This extended the study of Niku et al. (2017) by considering different types of indicators at the same time, and the research related to joint modelling of responses and response times within a hierarchical framework (van der Linden, 2007) by incorporating count data. Third, compared to Andersson et al. (2023b) who only considered categorical data, we compared the first- and second-order Laplace approximation in the case of GLLVMs with a mixture of ordinal, count, and continuous indicators, which advanced our knowledge of the performance of both algorithms in the mixed-data situation, correlations of latent variables, and magnitudes of residual factor loadings.

On the other hand, some limitations of the paper should be noted. First, the Laplace approximations require computing higher-order derivatives for the distributions of the observed outcome variables. This means that substantial derivations are necessary to support additional distributions. In the current study, only certain types of continuous, count, and ordinal data were considered. However, it is feasible to consider other types/distributions of indicators, which is a potential direction for future studies. Second, we only compared the first- and second-order Laplace approximations and did not compare them against quadrature-based methods. This was due to the computational expense of quadrature-based methods. Other approaches, such as Bayesian methods, were not considered in this study because of the focus on likelihood-based estimation. Bayesian approaches employing MCMC methods have been applied to estimate joint models of continuous, count, and binary data (Man et al., 2022; Man & Harring, 2023) but these implementations have so far not considered residual correlations between the indicators. It is worth noting that Bayesian methods can be time-intensive in estimating model parameters (Ulitzsch et al., 2020a). In addition, practitioners may face challenges in deciding on parameter priors. Third, the GLLVMs used in this study did not consider potential latent subgroups. In empirical studies of response times and responses, mixture modelling that distinguishes respondents with aberrant behaviour from those with solution behaviour have been considered (Wang et al., 2018). A future direction could be to extend GLLVMs by incorporating finite-mixture models to account for latent subgroups.

5.2 | Practical suggestions

In this study, we used process and performance data from computer-based assessments to demonstrate the application of the proposed method. The method can also be applied to game-based assessments to model the number of trials, response times, and performance scores (Landers et al., 2022), or psychophysiological multimodal data including item responses, response times, and visual fixation counts using eye-tracking equipment (Man et al., 2022; Man & Harring, 2023). Additionally, the method has the potential to be applied to broader areas beyond psychological and educational assessments. For example, a combination of different data types often occurs in ecological data such as species counts and biomass in biology (Niku et al., 2017) and patient data relevant to symptoms such as presence/absence, frequency, and scale scores in health (Daniels & Normand, 2006).

For practitioners dealing with a mixture of different types of data, we offer some suggestions. First, a proper treatment of data pre-processing is necessary for achieving accurate estimates. For example, when dealing with highly positively skewed data, such as response times, it is recommended to apply a logarithm transformation to account for the skewness of the response time distributions (van der Linden, 2006). As with other statistical analyses, the presence of outliers can result in inaccurate parameter estimates, necessitating the identification and proper corrections of outliers (Chambers et al., 2004). Second, when there are more than two latent variables, Laplace approximations have a great advantage over numerical quadrature or Bayesian approaches in terms of computational efficiency. Within Laplace approximations, Lap1 is faster than Lap2 and the efficiency advantage increases with the dimension of latent variables and the complexity of model structures. For example, the difference between the average time for estimating three-dimensional models (Simulation 1) was 2 seconds, while the value increased to 350 seconds for nine-dimensional models with residual correlations (Simulation 2) when the sample size was 1000. Third, when both the sample size and the number of variables were small (e.g., 250 participants and three variables in our simulations), Lap1 and Lap2 showed their advantages in estimating different model parameters. In general, when practitioners are primarily interested in investigating the relationships between latent variables or task-specific relationships (Costa & Chen, 2023), we recommend using Lap2, as it produced more accurate and precise estimates of the covariance and residual factor loadings in our simulations and was closer to the gold standard estimator. Finally, starting values have a large impact on the estimation of GLLVMs. This is because the observed likelihood can be multimodal when GLLVMs have a complex mean and latent variable structure (Niku et al., 2019). If researchers or practitioners have prior knowledge of the estimates based on existing literature or studies, it is possible to make use of that information. If no prior knowledge is available, it is possible to make use of the data provided to determine the starting values (Niku et al., 2019). In our simulation studies, we first fitted unidimensional measurement models and obtained the estimates as starting values if the unidimensional models converged. This greatly reduced the estimation time and increased the convergence rate.

CONFLICT OF INTEREST STATEMENT

All authors declare that they have no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Database – CBA PISA 2012 at https://www.oecd.org/pisa/pisaproducts/database-cbapisa2012.htm.

ORCID

Maoxin Zhang https://orcid.org/0000-0002-6250-3610
Björn Andersson https://orcid.org/0000-0002-9007-2440
Shaobo Jin https://orcid.org/0000-0001-6538-3477

REFERENCES

Andersson, B., Jin, S., & Zhang, M. (2023a). Lamle: Maximum Likelihood Estimation of Latent Variable Models Maximum Likelihood Estimation of Latent Variable Models. R package version 0.3.1. https://CRAN.R-project.org/package=lamle

Andersson, B., Jin, S., & Zhang, M. (2023b). Fast estimation of multiple group generalized linear latent variable models for categorical observed variables. *Computational Statistics & Data Analysis*, *182*, 107710.

Andersson, B., & Xin, T. (2021). Estimation of latent regression item response theory models using a second-order Laplace approximation. *Journal of Educational and Behavioral Statistics*, *46*(2), 244–265.

Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach*. John Wiley & Sons.

Bianconcini, S. (2014). Asymptotic properties of adaptive maximum likelihood estimators in latent variable models. *Bernoulli*, *20*, 1507–1531.

Chambers, R., Hentges, A., & Zhao, X. (2004). Robust automatic methods for outlier and error detection. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *167*(2), 323–339.

Costa, D. R., & Chen, C.-W. (2023). Exploring the relationship between process data and contextual variables among Scandinavian students on PISA 2012 mathematics tasks. *Large-Scale Assessments in Education*, *11*(1), 1–28.

Daniels, M. J., & Normand, S.-L. T. (2006). Longitudinal profiling of health care units based on continuous and discrete patient outcomes. *Biostatistics*, *7*(1), 1–15.

De Boeck, P., & Scalise, K. (2019). Collaborative problem solving: Processing actions, time, and performance. *Frontiers in Psychology*, *10*, 1280.

De Leon, A. R., & Chough, K. C. (2013). *Analysis of mixed data: Methods & applications*. CRC Press.

Huber, P., Ronchetti, E., & Victoria-Feser, M.-P. (2004). Estimation of generalized linear latent variable models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *66*(4), 893–908.

Jin, S., & Andersson, B. (2020). A note on the accuracy of adaptive Gauss-Hermite quadrature. *Biometrika*, *107*, 737–744.

Joe, H. (2008). Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics & Data Analysis*, *52*, 5066–5074.

Landers, R. N., Armstrong, M. B., Collmus, A. B., Mujcic, S., & Blaik, J. (2022). Theory-driven game-based assessment of general cognitive ability: Design theory, measurement, prediction of performance, and test fairness. *Journal of Applied Psychology*, *107*(10), 1655–1677.

Man, K., & Harring, J. R. (2023). Detecting preknowledge cheating via innovative measures: A mixture hierarchical model for jointly modeling item responses, response times, and visual fixation counts. *Educational and Psychological Measurement*, *83*(5), 1059–1080.

Man, K., Harring, J. R., & Zhan, P. (2022). Bridging models of biometric and psychometric assessment: A three-way joint modeling approach of item responses, response times, and gaze fixation counts. *Applied Psychological Measurement*, *46*(5), 361–381.

Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research*, *49*(4), 305–328.

Meng, X.-B., Tao, J., & Chang, H.-H. (2015). A conditional joint modeling approach for locally dependent item responses and response times. *Journal of Educational Measurement*, *52*(1), 1–27.

Moustaki, I. (1996). A latent trait and a latent class model for mixed observed variables. *British Journal of Mathematical and Statistical Psychology*, *49*(2), 313–334.

Moustaki, I., & Knott, M. (2000). Generalized latent trait models. *Psychometrika*, *65*(3), 391–411.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*, 159–176.

Muthén, L. K., & Muthén, B. (2017). *Mplus user's guide: Statistical analysis with latent variables, user's guide*. Muthén & Muthén.

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A*, *135*(3), 370–384.

Niku, J., Brooks, W., Herliansyah, R., Hui, F. K., Taskinen, S., & Warton, D. I. (2019). Efficient estimation of generalized linear latent variable models. *PLoS ONE*, *14*(5), e0216129.

Niku, J., Warton, D. I., Hui, F. K., & Taskinen, S. (2017). Generalized linear latent variable models for multivariate count and biomass data in ecology. *Journal of Agricultural, Biological and Environmental Statistics*, *22*(4), 498–522.

Olsson, U., Drasgow, F., & Dorans, N. J. (1982). The polyserial correlation coefficient. *Psychometrika*, *47*, 337–347.

OECD. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. OECD Publishing.

Qiao, X., Jiao, H., & He, Q. (2022). Multiple-group joint modeling of item responses, response times, and action counts with the Conway-Maxwell-Poisson distribution. *Journal of Educational Measurement*, *60*(2), 255–281.

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software manual]. https://www.R-project.org/

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, *2*(1), 1–21.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, *69*(2), 167–190.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores*. Psychometrika, Monograph Supplement No. 17.

Schnipke, D. L., & Scrams, D. J. (2002). Exploring issues of examinee behavior: Insights gained from response-time analyses. In C. N. Mills, M. Potenza, J. J. Fremer, & W. Ward (Eds.), *Computer-based testing: Building the foundation for future assessments* (1st ed., pp. 237–266). Lawrence Erlbaum Associates.

Shun, Z. (1997). Another look at the salamander mating data: A modified Laplace approximation approach. *Journal of the American Statistical Association*, *92*(437), 341–349.

Steinfeld, N. (2016). "I agree to the terms and conditions": (How) do users read privacy policies online? An eye-tracking experiment. *Computers in Human Behavior*, *55*, 992–1000.

Ulitzsch, E., von Davier, M., & Pohl, S. (2020a). A hierarchical latent response model for inferences about examinee engagement in terms of guessing and item-level non-response. *British Journal of Mathematical and Statistical Psychology*, *73*, 83–112.

Ulitzsch, E., von Davier, M., & Pohl, S. (2020b). Using response times for joint modeling of response and omission behavior. *Multivariate Behavioral Research*, *55*(3), 425–453.

van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, *31*(2), 181–204.

van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, *72*(3), 287–308.

Wang, C., Xu, G., & Shang, Z. (2018). A two-stage approach to differentiating normal and aberrant behavior in computer based testing. *Psychometrika*, *83*, 223–254.

Zhang, M., Andersson, B., & Greiff, S. (2023). Investigating planning and non-targeted exploration in PIAAC 2012: Validating their measures based on process data and investigating their relationships with problem-solving competency. *Journal of Intelligence*, *11*(8), 156.

Zhang, S., & Chen, Y. (2022). Computation for latent variable model estimation: A unified stochastic proximal framework. *Psychometrika*, *87*(4), 1473–1502.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhang, M., Andersson, B., & Jin, S. (2024). Fast estimation of generalized linear latent variable models for performance and process data with ordinal, continuous, and count observed variables. *British Journal of Mathematical and Statistical Psychology*, *00*, 1–31. https://doi.org/10.1111/bmsp.12337

APPENDIX 1

TRUE VALUES OF THE PARAMETERS IN THE SIMULATION STUDIES

TABLE A1 Distributions of true parameters

Data type	Parameter	Distribution
Ordinal data	a slope parameter	$U(0.74, 1.69)$
	$b1$ threshold parameter	$U(0.2, 1.25)$
	$b2$ threshold parameter	$U(-1.25, -0.2)$
Continuous data	a slope parameter	$U(0.4, 0.8)$
	b intercept parameter	$U(-0.2, 0.2)$
	ϕ scale parameter	$U(0.1, 0.3)$
Count data	a slope parameter	$U(0.4, 0.8)$
	b intercept parameter	$U(1, 3)$
	ϕ scale parameter (negative binomial)	$U(0.5, 1)$
Covariance parameters	ρ covariance: small	$U(0.2, 0.4)$
	ρ covariance: large	$U(0.6, 0.8)$

TABLE A2 Covariance matrix for the latent variables when covariance is small

	F1	F2	F3
F1	1	0.363	0.377
F2	0.363	1	0.204
F3	0.377	0.204	1

TABLE A3 Covariance matrix for the latent variables when covariance is large

	F1	F2	F3
F1	1	0.743	0.632
F2	0.743	1	0.630
F3	0.632	0.630	1

TABLE A4 True residual factor loadings

Stimulus	Small	Large
1	0.633	0.316
2	0.715	0.357
3	0.627	0.313
4	0.649	0.325
5	0.426	0.213
6	0.505	0.253

APPENDIX 2

DERIVATIVES

The derivatives required in Equation (10) are presented as follows. Let $h_i = -\log P_{iy_i}$ and define $\mathbf{1}()$ as an indicator function.

Ordinal data

We give the results for the GPCM and refer to Jin and Andersson (2020) for results with the GRM. Let P_{ic} represent $P_i(y_i = c|\mathbf{z}, \mathbf{w})$. The derivatives of h_i with respect to $\mathbf{z}$ are

$$\frac{\partial h_i}{\partial z_j} = -a_{ij}\left(y_i - \sum_{c=1}^{m_i} cP_{ic}\right),$$

$$\frac{\partial^2 h_i}{\partial z_j \partial z_k} = a_{ij}\sum_{c=1}^{m_i} c\frac{\partial P_{ic}}{\partial z_k},$$

$$\frac{\partial^3 h_i}{\partial z_j \partial z_k \partial z_l} = a_{ij}\sum_{c=1}^{m_i} c\frac{\partial^2 P_{ic}}{\partial z_k \partial z_l},$$

$$\frac{\partial^4 h_i}{\partial z_j \partial z_k \partial z_l \partial z_m} = a_{ij}\sum_{c=1}^{m_i} c\frac{\partial^3 P_{ic}}{\partial z_k \partial z_l \partial z_m},$$

and

$$\frac{\partial^5 h_i}{\partial z_j \partial z_k \partial z_l \partial z_m \partial z_n} = a_{ij}\sum_{c=1}^{m_i} c\frac{\partial^4 P_{ic}}{\partial z_k \partial z_l \partial z_m \partial z_n}.$$

The derivatives of h_i with respect to $u \in \{a_i, b_{i2}, \ldots b_{im_i}\}$ are

$$\frac{\partial h_i}{\partial u} = -\frac{\frac{\partial P_{iy_f}}{\partial u}}{P_{iy_f}},$$

$$\frac{\partial^2 h_i}{\partial z_j \partial u} = -\mathbf{1}(u = a_{ij})\left(y_i - \sum_{c=1}^{m_i} c\frac{\partial P_{ic}}{\partial u}\right) + a_{ij}\sum_{c=1}^{m_i} c\frac{\partial P_{ic}}{\partial u},$$

$$\frac{\partial^3 h_i}{\partial z_j \partial z_k \partial u} = \mathbf{1}(u = a_{ij})\sum_{c=1}^{m_i} c\frac{\partial P_{ic}}{\partial z_k} + a_{ij}\sum_{c=1}^{m_i} c\frac{\partial^2 P_{ic}}{\partial z_k \partial u},$$

$$\frac{\partial^4 h_i}{\partial z_j \partial z_k \partial z_l \partial u} = \mathbf{1}(u = a_{ij})\sum_{c=1}^{m_i} c\frac{\partial^2 P_{ic}}{\partial z_k \partial z_l} + a_{ij}\sum_{c=1}^{m_i} c\frac{\partial^3 P_{ic}}{\partial z_k \partial z_l \partial u},$$

and

$$\frac{\partial^5 h_i}{\partial z_j \partial z_k \partial z_l \partial z_m \partial u} = \mathbf{1}(u = a_{ij})\sum_{c=1}^{m_i} c\frac{\partial^3 P_{ic}}{\partial z_k \partial z_l \partial z_m} + a_{ij}\sum_{c=1}^{m_i} c\frac{\partial^4 P_{ic}}{\partial z_k \partial z_l \partial z_m \partial u}.$$

The derivatives of P_{ic} with respect to $\mathbf{z}$ in the above equations are

$$\frac{\partial P_{ic}}{\partial z_k} = P_{ic} a_{ik} \left[c - \sum_{c'=1}^{m_i} c' P_{ic'} \right],$$

$$\frac{\partial^2 P_{ic}}{\partial z_k \partial z_l} = \frac{\partial P_{ic}}{\partial z_l} a_{ik} \left[c - \sum_{c'=1}^{m_i} c P_{ic} \right] - P_{ic} a_{ik} \sum_{c=1}^{m_i} c' \frac{\partial P_{ic'}}{\partial z_l},$$

$$\begin{aligned}\frac{\partial^3 P_{ic}}{\partial z_k \partial z_l \partial z_m} = {} & \frac{\partial^2 P_{ic}}{\partial z_l \partial z_m} a_{ik} \left[c - \sum_{c'=1}^{m_i} c' P_{ic'} \right] - \frac{\partial P_{ic}}{\partial z_l} a_{ik} \sum_{c'=1}^{m_i} c' \frac{\partial P_{ic'}}{\partial z_m} \\ & - \frac{\partial P_{ic}}{\partial z_m} a_{ik} \sum_{c'=1}^{m_i} c' \frac{\partial P_{ic'}}{\partial z_l} - P_{ic} a_{ik} \sum_{c'=1}^{m_i} c' \frac{\partial^2 P_{ic'}}{\partial z_l \partial z_m},\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^4 P_{ic}}{\partial z_k \partial z_l \partial z_m \partial z_n} = {} & \frac{\partial^3 P_{ic}}{\partial z_l \partial z_m \partial z_n} a_{ik} \left[c - \sum_{c'=1}^{m_i} c' P_{ic'} \right] - \frac{\partial^2 P_{ic}}{\partial z_l \partial z_m} a_{ik} \sum_{c'=1}^{m_i} c' \frac{\partial P_{ic'}}{\partial z_n} \\ & - \frac{\partial^2 P_{ic}}{\partial z_l \partial z_n} a_{ik} \sum_{c'=1}^{m_i} c' \frac{\partial P_{ic'}}{\partial z_m} - \frac{\partial P_{ic}}{\partial z_l} a_{ik} \sum_{c'=1}^{m_i} c' \frac{\partial^2 P_{ic'}}{\partial z_m \partial z_n} \\ & - \frac{\partial^2 P_{ic}}{\partial z_m \partial z_n} a_{ik} \sum_{c'=1}^{m_i} c' \frac{\partial P_{ic'}}{\partial z_l} - \frac{\partial P_{ic}}{\partial z_m} a_{ik} \sum_{c'=1}^{m_i} c' \frac{\partial^2 P_{ic'}}{\partial z_l \partial z_n} \\ & - \frac{\partial P_{ic}}{\partial z_n} a_{ik} \sum_{c'=1}^{m_i} c' \frac{\partial^2 P_{ic'}}{\partial z_l \partial z_m} - P_{ic} a_{ik} \sum_{c'=1}^{m_i} c' \frac{\partial^3 P_{ic'}}{\partial z_l \partial z_m \partial z_n}.\end{aligned}$$

The derivatives of P_{ic} with respect to $a_{ij} \in \mathbf{a}_i$ and $b_{iv} \in \mathbf{b}_i$ are

$$\frac{\partial P_{ic}}{\partial a_{ij}} = P_{ic} c z_j - P_{ic} z_j \sum_{c'=1}^{m_i} c' P_{ic'},$$

and

$$\frac{\partial P_{ic}}{\partial b_{iv}} = \mathbf{1}(c \geq v) P_{ic} - P_{ic} \sum_{c'=v}^{m_i} P_{ic'}.$$

Then we have

$$\frac{\partial^2 P_{ic}}{\partial z_k \partial u} = \frac{\partial P_{ic}}{\partial u} a_{ik} \left[c - \sum_{c'=1}^{m_i} c' P_{ic'} \right] + \mathbf{1}(u = a_{ik}) P_{ic} \left[c - \sum_{c'=1}^{m_i} c' P_{ic'} \right] - P_{ic} a_{ik} \sum_{c'=1}^{m_i} c' \frac{\partial P_{ic'}}{\partial u},$$

$$\begin{aligned}\frac{\partial^3 P_{ic}}{\partial z_k \partial z_l \partial u} = {} & \mathbf{1}(u = a_{ik}) \left[\frac{\partial P_{ic}}{\partial z_l} \left(c - \sum_{c'=1}^{m_i} c' P_{ic} \right) - P_{ic} \sum_{c'=1}^{m_i} c' \frac{\partial P_{ic'}}{\partial z_l} \right] \\ & + a_{ik} \left[\frac{\partial^2 P_{ic}}{\partial z_l \partial u} \left(c - \sum_{c'=1}^{m_i} c' P_{ic'} \right) - \frac{\partial P_{ic}}{\partial z_l} \sum_{c'=1}^{m_i} c' \frac{\partial P_{ic'}}{\partial u} \right. \\ & \left. - \frac{\partial P_{ic}}{\partial u} \sum_{c'=1}^{m_i} c' \frac{\partial P_{ic'}}{\partial z_l} - P_{ic} \sum_{c'=1}^{m_i} c' \frac{\partial^2 P_{ic'}}{\partial z_l \partial u} \right],\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^4 P_{ic}}{\partial z_k \partial z_l \partial z_m \partial u} = \; & \mathbf{1}(u = a_{ik}) \left[\frac{\partial^2 P_{ic}}{\partial z_l \partial z_m} \left(c - \sum_{c'=1}^{m_i} c' P_{ic'} \right) - \frac{\partial P_{ic}}{\partial z_l} \sum_{c'=1}^{m_i} c' \frac{\partial P_{ic'}}{\partial z_m} \right. \\ & \left. - \frac{P_{ic}}{\partial z_m} \sum_{c'=1}^{m_i} c' \frac{\partial P_{ic'}}{\partial z_l} - P_{ic} \sum_{c'=1}^{m_i} c' \frac{\partial^2 P_{ic'}}{\partial z_l \partial z_m} \right] \\ & + a_{ik} \left[\frac{\partial^3 P_{ic}}{\partial z_l \partial z_m \partial u} \left(c - \sum_{c'=1}^{m_i} c' P_{ic'} \right) - \frac{\partial^2 P_{ic}}{\partial z_l \partial z_m} \sum_{c'=1}^{m_i} c' \frac{\partial P_{ic'}}{\partial u} \right. \\ & - \frac{\partial^2 P_{ic}}{\partial z_l \partial u} \sum_{c'=1}^{m_i} c' \frac{\partial P_{ic'}}{\partial z_m} - \frac{\partial P_{ic}}{\partial z_l} \sum_{c'=1}^{m_i} c' \frac{\partial^2 P_{ic'}}{\partial z_m \partial u} \\ & - \frac{\partial^2 P_{ic}}{\partial z_m \partial u} \sum_{c'=1}^{m_i} c' \frac{\partial P_{ic'}}{\partial z_l} - \frac{\partial P_{ic}}{\partial z_m} \sum_{c'=1}^{m_i} c' \frac{\partial^2 P_{ic'}}{\partial z_l \partial u} \\ & \left. - \frac{\partial P_{ic}}{\partial u} \sum_{c'=1}^{m_i} c' \frac{\partial^2 P_{ic'}}{\partial z_l \partial z_m} - P_{ic} \sum_{c'=1}^{m_i} c' \frac{\partial^3 P_{ic'}}{\partial z_l \partial z_m \partial u} \right].\end{aligned}$$

The derivatives with respect to $\boldsymbol{\beta}_{id}$ are equal to the product of w_d and the derivatives with respect to b_i.

Continuous data

The derivatives of h_i with respect to $\mathbf{z}$ for continuous data (Huber et al., 2004) are given as follows. Note that only the first and second derivatives exist in this case.

$$\frac{\partial h_i}{\partial z_j} = \frac{a_{ij}}{\phi_i}(b_i + \boldsymbol{\beta}_i'\mathbf{w} + \mathbf{a}_i'\mathbf{z} - y_i)$$

$$\frac{\partial^2 h_i}{\partial z_j \partial z_k} = \frac{a_{ij} a_{ik}}{\phi_i}$$

The derivatives of h_i with respect to a_{ij}, b_i, and ϕ_i with continuous data are

$$\frac{\partial h_i}{\partial a_{ij}} = \frac{z_j}{\phi_i}(b_i + \boldsymbol{\beta}_i'\mathbf{w} + \mathbf{a}_i'\mathbf{z} - y_i),$$

$$\frac{\partial h_i}{\partial b_i} = \frac{(b_i + \boldsymbol{\beta}_i'\mathbf{w} + \mathbf{a}_i'\mathbf{z} - y_i)}{\phi_i},$$

$$\frac{\partial h_i}{\partial \phi_i} = \frac{2y_i(b_i + \boldsymbol{\beta}_i'\mathbf{w} + \mathbf{a}_i'\mathbf{z}) - (b_i + \boldsymbol{\beta}_i'\mathbf{w} + \mathbf{a}_i'\mathbf{z})^2 - t_i^2 + \phi_i}{2\phi_i^2},$$

$$\frac{\partial^2 h_i}{\partial z_j \partial a_{ix}} = \frac{a_{ij} z_x}{\phi_i} + \mathbf{1}(x = j)\frac{b_i + \boldsymbol{\beta}_i'\mathbf{w} + \mathbf{a}_i'\mathbf{z} - y_i}{\phi_i},$$

$$\frac{\partial^2 h_i}{\partial z_j \partial b_i} = \frac{a_{ij}}{\phi_i},$$

$$\frac{\partial^2 h_i}{\partial z_j \partial \phi_i} = -\frac{a_{ij}(b_i + \boldsymbol{\beta}_i'\mathbf{w} + \mathbf{a}_i'\mathbf{z} - y_i)}{\phi_i^2},$$

$$\frac{\partial^3 h_i}{\partial z_j \partial z_k \partial a_{ix}} = \frac{a_{ik}}{\phi_i},$$

and

$$\frac{\partial^3 h_i}{\partial z_j \partial z_k \partial \phi_i} = -\frac{a_{ij}a_{ik}}{\phi_i^2}.$$

The derivatives with respect to β_{id} are equal to the product of w_d and the derivatives with respect to b_i.

Count data: Poisson distribution

The first- to fifth-order derivatives of h_i with respect to $\mathbf{z}$ are

$$\frac{\partial h_i}{\partial z_j} = (\lambda_i - y_i)a_{ij},$$

$$\frac{\partial^2 h_i}{\partial z_j \partial z_k} = a_{ij}\frac{\partial \lambda_i}{\partial z_k} = \lambda_i a_{ij} a_{ik},$$

$$\frac{\partial^3 h_i}{\partial z_j \partial z_k \partial z_l} = a_{ij}a_{ik}\frac{\partial \lambda_i}{\partial z_l} = \lambda_i a_{ij} a_{ik} a_{il},$$

$$\frac{\partial^4 h_i}{\partial z_j \partial z_k \partial z_l \partial z_m} = \lambda_i a_{ij} a_{ik} a_{il} a_{im},$$

and

$$\frac{\partial^5 h_i}{\partial z_j \partial z_k \partial z_l \partial z_m \partial z_n} = \lambda_i a_{ij} a_{ik} a_{il} a_{im} a_{in}.$$

The derivatives of h_i with respect to $u \in \{\mathbf{a}_i, b_i\}$ are

$$\frac{\partial h_i}{\partial u} = (1 - \frac{y_i}{\lambda_i})\frac{\partial \lambda_i}{\partial u} = (\lambda_i - y_i) z_j^{\mathbf{1}(u=a_{ij})},$$

$$\frac{\partial^2 h_i}{\partial z_j \partial a_{ix}} = a_{ij}\lambda_i z_x + (\lambda_i - y_i)\mathbf{1}(x=j)$$

$$\frac{\partial^2 h_i}{\partial z_j \partial b_i} = a_{ij}\lambda_i,$$

$$\frac{\partial^3 h_i}{\partial z_j \partial z_k \partial a_{ix}} = a_{ij} a_{ik} \lambda_i z_x + \lambda_i a_{ik} \mathbf{1}(x=j) + \lambda_i a_{ij} \mathbf{1}(x=k),$$

$$\frac{\partial^3 h_i}{\partial z_j \partial z_k \partial b_i} = a_{ij} a_{ik} \lambda_i,$$

$$\frac{\partial^4 h_i}{\partial z_j \partial z_k \partial z_l \partial a_{ix}} = a_{ij} a_{ik} a_{il} \lambda_i z_x + \lambda_i a_{ik} a_{il} \mathbf{1}(x=j) + \lambda_i a_{ij} a_{il} \mathbf{1}(x=k) + \lambda_i a_{ij} a_{ik} \mathbf{1}(x=l),$$

$$\frac{\partial^4 h_i}{\partial z_j \partial z_k \partial z_l \partial b_i} = a_{ij} a_{ik} a_{il} \lambda_i,$$

$$\begin{aligned}\frac{\partial^5 h_i}{\partial z_j \partial z_k \partial z_l \partial z_m \partial a_{ix}} = {} & a_{ij} a_{ik} a_{il} a_{im} \lambda_i z_x + \lambda_i a_{ik} a_{il} a_{im} \mathbf{1}(x=j) + \lambda_i a_{ij} a_{il} a_{im} \mathbf{1}(x=k) + \\ & \lambda_i a_{ij} a_{ik} a_{im} \mathbf{1}(x=l) + \lambda_i a_{ij} a_{ik} a_{il} \mathbf{1}(x=m),\end{aligned}$$

and

$$\frac{\partial^5 h_i}{\partial z_j \partial z_k \partial z_l \partial z_m \partial b_i} = a_{ij} a_{ik} a_{il} a_{im} \lambda_i.$$

The derivatives with respect to β_{id} are equal to the product of w_d and the derivatives with respect to b_i.

Count data: Negative-binomial distribution

With a negative-binomial distribution, we have that

$$\begin{aligned}h_i = {} & -\log\Gamma\left(y_i + \frac{1}{\phi_i}\right) + \log(y_i!) + \log\Gamma\left(\frac{1}{\phi_i}\right) - y_i(b_i + \boldsymbol{\beta}_i' \mathbf{w} + \mathbf{a}_i' \mathbf{z}) \\ & + y_i \log\left[\frac{1}{\phi_i} + \exp(b_i + \boldsymbol{\beta}_i' \mathbf{w} + \mathbf{a}_i' \mathbf{z})\right] + \frac{1}{\phi_i} \log\left[1 + \phi_i \exp(b_i + \boldsymbol{\beta}_i' \mathbf{w} + \mathbf{a}_i' \mathbf{z})\right].\end{aligned}$$

Let $\eta_i = b_i + \boldsymbol{\beta}_i' \mathbf{w} + \mathbf{a}_i' \mathbf{z}$. The derivatives with respect to $\mathbf{z}$ are

$$\frac{\partial h_i}{\partial z_j} = -y_i a_{ij} + (\phi_i y_i + 1) \frac{\exp(\eta_i)}{1 + \phi_i \exp(\eta_i)} a_{ij},$$

$$\frac{\partial^2 h_i}{\partial z_j \partial z_k} = (\phi_i y_i + 1) \frac{\exp(\eta_i)}{[1 + \phi_i \exp(\eta_i)]^2} a_{ij} a_{ik},$$

$$\frac{\partial^3 h_i}{\partial z_j \partial z_k \partial z_l} = -(\phi_i y_i + 1) \frac{\exp(\eta_i)[\phi_i \exp(\eta_i) - 1]}{\left[1 + \phi_i \exp(\eta_i)\right]^3} a_{ij} a_{ik} a_{il},$$

$$\frac{\partial^4 h_i}{\partial z_j \partial z_k \partial z_l \partial z_m} = (\phi_i y_i + 1) \frac{\exp(\eta_i)(\phi_i^2 \exp(2\eta_i) - 4\phi_i \exp(\eta_i) + 1)}{\left[1 + \phi_i \exp(\eta_i)\right]^4} a_{ij} a_{ik} a_{il} a_{im},$$

and

$$\frac{\partial^5 h_i}{\partial z_j \partial z_k \partial z_l \partial z_m \partial z_n} = -(\phi_i y_i + 1)\frac{\exp(\eta_i)\left[\phi_i^3 \exp(3\eta_i) - 11\phi_i^2 \exp(2\eta_i) + 11\phi_i \exp(\eta_i) - 1\right]}{\left[1 + \phi_i \exp(\eta_i)\right]^5} \times a_{ij} a_{ik} a_{il} a_{im} a_{in}.$$

The derivatives with respect to b_i are

$$\frac{\partial h_i}{\partial b_i} = -y_i + \left(y_i + \frac{1}{\phi_i}\right)\frac{\phi_i \exp(\eta_i)}{1 + \phi_i \exp(\eta_i)},$$

$$\frac{\partial^2 h_i}{\partial z_j \partial b_i} = (\phi_i y_i + 1)\frac{\exp(\eta_i)}{1 + \phi_i \exp(\eta_i)} a_{ij} - (\phi_i y_i + 1)\frac{\phi_i \exp(2\eta_i)}{[1 + \phi_i \exp(\eta_i)]^2} a_{ij},$$

$$\frac{\partial^3 h_i}{\partial z_j \partial z_k \partial b_i} = (\phi_i y_i + 1)\frac{\exp(\eta_i)}{[1 + \phi_i \exp(\eta_i)]^2} a_{ij} a_{ik} - 2(\phi_i y_i + 1)\frac{\phi_i \exp(2\eta_i)}{[1 + \phi_i \exp(\eta_i)]^3} a_{ij} a_{ik},$$

$$\begin{aligned}\frac{\partial^4 h_i}{\partial z_j \partial z_k \partial z_l \partial b_i} = &-(\phi_i y_i + 1)\frac{\exp(\eta_i)[\phi_i \exp(\eta_i) - 1]}{[1 + \phi_i \exp(\eta_i)]^3} a_{ij} a_{ik} a_{il} - (\phi_i y_i + 1)\frac{\phi_i \exp(2\eta_i)}{[1 + \phi_i \exp(\eta_i)]^3} a_{ij} a_{ik} a_{il} \\ &+ 3(\phi_i y_i + 1)\frac{\phi_i \exp(2\eta_i)[\phi_i \exp(\eta_i) - 1]}{[1 + \phi_i \exp(\eta_i)]^4} a_{ij} a_{ik} a_{il},\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^5 h_i}{\partial z_j \partial z_k \partial z_l \partial z_m \partial b_i} = &(\phi_i y_i + 1)\frac{\exp(\eta_i)[\phi_i^2 \exp(2\eta_i) - 4\phi_i \exp(\eta_i) + 1]}{[1 + \phi_i \exp(\eta_i)]^4} a_{ij} a_{ik} a_{il} a_{im} \\ &+ (\phi_i y_i + 1)\frac{\exp(\eta_i)[2\phi_i^2 \exp(2\eta_i) - 4\phi_i \exp(\eta_i)]}{[1 + \phi_i \exp(\eta_i)]^4} a_{ij} a_{ik} a_{il} a_{im} \\ &- 4(\phi_i y_i + 1)\frac{\phi_i \exp(2\eta_i)[\phi_i^2 \exp(2\eta_i) - 4\phi_i \exp(\eta_i) + 1]}{[1 + \phi_i \exp(\eta_i)]^5} a_{ij} a_{ik} a_{il} a_{im}.\end{aligned}$$

The derivatives with respect to a_{ic} are

$$\frac{\partial h_i}{\partial a_{ic}} = -y_i z_c + \left(y_i + \frac{1}{\phi_i}\right)\frac{\phi_i \exp(\eta_i) z_c}{1 + \phi_i \exp(\eta_i)},$$

$$\begin{aligned}\frac{\partial^2 h_i}{\partial z_j \partial a_{ic}} = &-y_i \mathbf{1}(c = j) + (\phi_i y_i + 1)\frac{\exp(\eta_i)}{1 + \phi_i \exp(\eta_i)} a_{ij} z_c \\ &- (\phi_i y_i + 1)\frac{\phi_i \exp(2\eta_i)}{[1 + \phi_i \exp(\eta_i)]^2} a_{ij} z_c + (\phi_i y_i + 1)\frac{\exp(\eta_i)}{1 + \phi_i \exp(\eta_i)} \frac{a_{ij}}{a_{ic}} \mathbf{1}(c = j),\end{aligned}$$

$$\begin{aligned}\frac{\partial^3 h_i}{\partial z_j \partial z_k \partial a_{ic}} = &(\phi_i y_i + 1)\frac{\exp(\eta_i)}{[1 + \phi_i \exp(\eta_i)]^2} a_{ij} a_{ik} z_c - 2(\phi_i y_i + 1)\frac{\phi_i \exp(2\eta_i)}{[1 + \phi_i \exp(\eta_i)]^3} a_{ij} a_{ik} z_c \\ &+ (\phi_i y_i + 1)\frac{\exp(\eta_i)}{[1 + \phi_i \exp(\eta_i)]^2} \frac{a_{ij} a_{ik}}{a_{ic}} \mathbf{1}(c \in \{j, k\}),\end{aligned}$$

$$\frac{\partial^4 h_i}{\partial z_j \partial z_k \partial z_l \partial a_{ic}} = -(\phi_i y_i + 1)\frac{\exp(\eta_i)[\phi_i \exp(\eta_i) - 1]}{[1+\phi_i \exp(\eta_i)]^3} a_{ij} a_{ik} a_{il} z_c$$
$$-(\phi_i y_i + 1)\frac{\phi_i \exp(2\eta_i)}{[1+\phi_i \exp(\eta_i)]^3} a_{ij} a_{ik} a_{il} z_c$$
$$+3(\phi_i y_i + 1)\frac{\phi_i \exp(2\eta_i)[\phi_i \exp(\eta_i) - 1]}{[1+\phi_i \exp(\eta_i)]^4} a_{ij} a_{ik} a_{il} z_c$$
$$-(\phi_i y_i + 1)\frac{\exp(\eta_i)[\phi_i \exp(\eta_i) - 1]}{[1+\phi_i \exp(\eta_i)]^3} \frac{a_{ij} a_{ik} a_{il}}{a_{ic}} \mathbf{1}(c \in \{j,k,l\}),$$

and

$$\frac{\partial^5 h_i}{\partial z_j \partial z_k \partial z_l \partial z_m \partial a_{ic}} = (\phi_i y_i + 1)\frac{\exp(\eta_i)[\phi_i^2 \exp(2\eta_i) - 4\phi_i \exp(\eta_i) + 1]}{[1+\phi_i \exp(\eta_i)]^4} a_{ij} a_{ik} a_{il} a_{im} z_c$$
$$+(\phi_i y_i + 1)\frac{\exp(\eta_i)[2\phi_i^2 \exp(2\eta_i) - 4\phi_i \exp(\eta_i)]}{[1+\phi_i \exp(\eta_i)]^4} a_{ij} a_{ik} a_{il} a_{im} z_c$$
$$-4(\phi_i y_i + 1)\frac{\phi_i \exp(2\eta_i)[\phi_i^2 \exp(2\eta_i) - 4\phi_i \exp(\eta_i) + 1]}{[1+\phi_i \exp(\eta_i)]^5} a_{ij} a_{ik} a_{il} a_{im} z_c$$
$$+(\phi_i y_i + 1)\frac{\exp(\eta_i)[\phi_i^2 \exp(2\eta_i) - 4\phi_i \exp(\eta_i) + 1]}{[1+\phi_i \exp(\eta_i)]^4} \frac{a_{ij} a_{ik} a_{il} a_{im}}{a_{ic}} \mathbf{1}(c \in \{j,k,l,m\}).$$

The derivatives with respect to ϕ_i are, with ψ denoting the digamma function,

$$\frac{\partial h_i}{\partial \phi_i} = -\frac{\log[1+\phi_i \exp(\eta_i)]}{\phi_i^2} + \left(y_i + \frac{1}{\phi_i}\right)\frac{\exp(\eta_i)}{1+\phi_i \exp(\eta_i)} - \frac{y_i}{\phi_i} + \frac{\psi\left(y_i + \frac{1}{\phi_i}\right)}{\phi_i^2} - \frac{\psi\left(\frac{1}{\phi_i}\right)}{\phi_i^2},$$

$$\frac{\partial^2 h_i}{\partial z_j \partial \phi_i} = y_i \frac{\exp(\eta_i)}{1+\phi_i \exp(\eta_i)} a_{ij} - (\phi_i y_i + 1)\frac{\exp(2\eta_i)}{[1+\phi_i \exp(\eta_i)]^2} a_{ij},$$

$$\frac{\partial^3 h_i}{\partial z_j \partial z_k \partial \phi_i} = y_i \frac{\exp(\eta_i)}{[1+\phi_i \exp(\eta_i)]^2} a_{ij} a_{ik} - 2(\phi_i y_i + 1)\frac{\exp(2\eta_i)}{[1+\phi_i \exp(\eta_i)]^3} a_{ij} a_{ik},$$

$$\frac{\partial^4 h_i}{\partial z_j \partial z_k \partial z_l \partial \phi_i} = -y_i \frac{\exp(\eta_i)[\phi_i \exp(\eta_i) - 1]}{[1+\phi_i \exp(\eta_i)]^3} a_{ij} a_{ik} a_{il} - (\phi_i y_i + 1)\frac{\exp(2\eta_i)}{[1+\phi_i \exp(\eta_i)]^3} a_{ij} a_{ik} a_{il}$$
$$+3(\phi_i y_i + 1)\frac{\exp(2\eta_i)[\phi_i \exp(\eta_i) - 1]}{[1+\phi_i \exp(\eta_i)]^4} a_{ij} a_{ik} a_{il},$$

and

$$\frac{\partial^5 h_i}{\partial z_j \partial z_k \partial z_l \partial z_m \partial \phi_i} = y_i \frac{\exp(\eta_i)[\phi_i^2 \exp(2\eta_i) - 4\phi_i \exp(\eta_i) + 1]}{[1+\phi_i \exp(\eta_i)]^4} a_{ij} a_{ik} a_{il} a_{im}$$
$$+(\phi_i y_i + 1)\frac{\exp(\eta_i)[2\phi_i \exp(2\eta_i) - 4\exp(\eta_i)]}{[1+\phi_i \exp(\eta_i)]^4} a_{ij} a_{ik} a_{il} a_{im}$$
$$-4(\phi_i y_i + 1)\frac{\exp(2\eta_i)[\phi_i^2 \exp(2\eta_i) - 4\phi_i \exp(\eta_i) + 1]}{[1+\phi_i \exp(\eta_i)]^5} a_{ij} a_{ik} a_{il} a_{im}.$$

The derivatives with respect to β_{id} are equal to the product of w_d and the derivatives with respect to b_i.