Springer Proceedings in Mathematics & Statistics

Marie Wiberg · Dylan Molenaar
Jorge González · Ulf Böckenholt
Jee-Seon Kim *Editors*

Quantitative Psychology

The 85th Annual Meeting of the Psychometric Society, Virtual

Springer

Springer Proceedings in Mathematics & Statistics

Volume 353

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including operation research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

More information about this series at http://www.springer.com/series/10533

Marie Wiberg • Dylan Molenaar • Jorge González
Ulf Böckenholt • Jee-Seon Kim
Editors

Quantitative Psychology

The 85th Annual Meeting of the Psychometric Society, Virtual

Springer

Editors
Marie Wiberg
Department of Statistics, USBE
Umeå University
Umeå, Västerbottens Län, Sweden

Jorge González
Facultad de Matemáticas
Pontificia Universidad Católica de Chile
Santiago, Chile

Jee-Seon Kim
Department of Educational Psychology
University of Wisconsin-Madison
Madison, WI, USA

Dylan Molenaar
Department of Psychology
University of Amsterdam
Amsterdam, The Netherlands

Ulf Böckenholt
Kellogg School of Management
Northwestern University
Evanston, IL, USA

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-030-74771-8 ISBN 978-3-030-74772-5 (eBook)
https://doi.org/10.1007/978-3-030-74772-5

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume represents presentations given at the 85th annual meeting of the Psychometric Society, that due to the pandemic of covid-19 was held virtually. This is the first IMPS meeting held only over internet and it was given during July 14-17, 2020. There were 230 abstracts submitted (154 oral presentations, 89 posters, and 3 symposia). The virtual meeting attracted 378 participants, 54 of whom also participated in the virtual short course pre-conference workshop. There were three keynote presentations, three invited presentations, eight spotlight speaker presentations, and one dissertation award presentation.

Since the 77th meeting in Lincoln, Nebraska, Springer publishes the proceedings volume from the annual meeting of the Psychometric Society to allow presenters at the annual meeting to spread their ideas quickly to the wider research community, while still undergoing a thorough review process. This is especially important now as meeting in person was difficult in 2020. The previous eight volumes of the meetings were received successfully, and we expect these proceedings to be successful as well.

The authors were asked to use their presentation at the meeting as the basis of their chapters, possibly extended with new ideas or additional information. The result is a selection of 42 stateof- the-art chapters addressing a diverse set of psychometric topics, including but not limited to item response theory, factor analysis, test equating, cognitive diagnostic models, response time, IRT as well as psychometric applications within different fields.

Umeå, Västerbottens Län, Sweden — Marie Wiberg

Amsterdam, The Netherlands — Dylan Molenaar

Santiago, Chile — Jorge González

Evanston, IL, USA — Ulf Böckenholt

Madison, WI, USA — Jee-Seon Kim

Contents

A Rotation Criterion That Encourages a Hierarchical Factor Structure

Chen Tian and Yang Liu

1 Introduction

In Yung, Thissen, and McLeod's terminology (1999), a hierarchical factor model may have several layers of factors: Each manifest variable loads on exactly one of the factors in each layer. Hierarchical factor structures are common in educational and psychological testing. For example, the big five personality traits can be divided into many aspects, and each aspect can further be divided into facets (e.g., Allen & DeYoung, 2017). These relationships between personality traits, aspects, facets, and manifest variables can be represented and analyzed using the higher-order factor model, a special case of hierarchical factor model with proportionality constraints. Another example is the testlet effect: Both the construct and the testlet factors contribute to the observed responses in a compensatory fashion. In addition to explaining the correlated errors, testlets may also explicitly represent higher-order facets within the hierarchy of interested constructs (Cooke et al., 2007).

Despite the wide-spread usage in theorizing constructs, it remains challenging to directly obtain hierarchical structures in Exploratory Factor Analysis (EFA) since there lacks a suitable rotation criterion. Rotation to a partially specified target may be used, but it requires fully specifying the positions of zero loadings in the target matrix (Browne, 1972, 2001). For circumstances where we have limited prior knowledge on the exact pattern of factor-item dependencies, Jennrich and Bentler (2011) discussed rotation criterions that encourage a bifactor structure which is the simplest hierarchical model with one general factor and one layer of specific factors. A rotation criterion function measures the discrepancy from an exact bifactor structure, which requires each item to load on at most one specific

C. Tian (✉) · Y. Liu
University of Maryland, College Park, MD, USA
e-mail: ctian1@terpmail.umd.edu

M. Wiberg et al. (eds.), *Quantitative Psychology*, Springer Proceedings in Mathematics & Statistics 353, https://doi.org/10.1007/978-3-030-74772-5_1

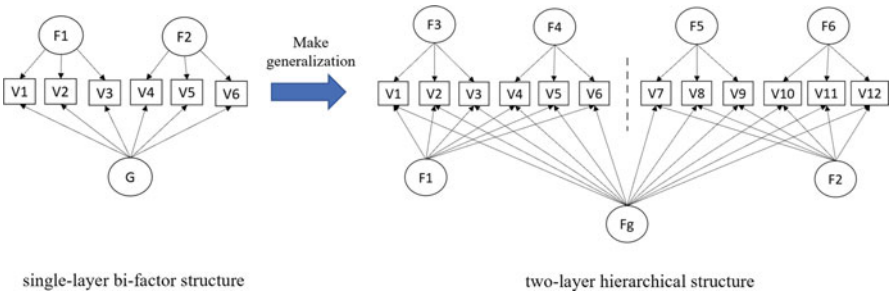

Fig. 1 Generalizing the rotation cretirion from a bi-factor structure to a hierarchical structure

factor. One example from Jennrich and Bentler (2011) is to apply the quartimin rotation criterion to the specific factors: $\sum_{i=1}^{p}\sum_{r=2}^{k}\sum_{s=r+1}^{k}\lambda_{ir}^2\lambda_{is}^2$, where λ_{ir} is the loading of item i on factor r, p is the number of items, and k is the number of factors.

Inspired by Jennrich and Bentler's exploratory bifactor analysis using the quartimin rotation criterion, the goal of this study is to propose a generalized rotation criterion for a two-layer hierarchical structures in EFA. Fig. 1 displays an example of such a structure: The corresponding factor loading matrix is expressed as eq. (1), in which asterisks denote non-zero loading entries, and the columns from left to right represent F_g and F1-F6. In the sequel, we say that a higher level factor is the parent of an adjacent lower level factor if all the items loading on the lower level factor also load on the higher level factor, and that two lower level factors having the same parent are sibling factors. In Fig. 1, F1 is the parent of F3 and F4, and F2 is the parent of F5 and F6.

$$\Lambda = \left(\begin{array}{cc|c|cccc} * & * & 0 & * & 0 & 0 & 0 \\ * & * & 0 & * & 0 & 0 & 0 \\ * & * & 0 & * & 0 & 0 & 0 \\ * & * & 0 & 0 & * & 0 & 0 \\ * & * & 0 & 0 & * & 0 & 0 \\ * & * & 0 & 0 & * & 0 & 0 \\ * & 0 & * & 0 & 0 & * & 0 \\ * & 0 & * & 0 & 0 & * & 0 \\ * & 0 & * & 0 & 0 & * & 0 \\ * & 0 & * & 0 & 0 & 0 & * \\ * & 0 & * & 0 & 0 & 0 & * \\ * & 0 & * & 0 & 0 & 0 & * \end{array}\right) \tag{1}$$

2 Methods

2.1 Proposed Rotation Criterion Function

The proposed rotation criterion function should first be able to encourage a simple structure within each layer, and this can be achieved by summing up the quartimin criterion applied to different layers. In our case with F1 and F2 in the first level, there is only one pair to constrain, so the corresponding term is $\sum_{i=1}^{p} \lambda_{i1}^2 \lambda_{i2}^2$. Minimizing this non-negative term encourages either λ_{i1}^2 or λ_{i2}^2 in the F1-F2 pair to be close to 0. With four factors in the second level, F3-F6, there are $\binom{4}{2} = 6$ pairs of factors, and the corresponding quartimin term is $\sum_{i=1}^{p} \lambda_{i3}^2 \lambda_{i4}^2 + \lambda_{i3}^2 \lambda_{i5}^2 + \lambda_{i3}^2 \lambda_{i6}^2 + \lambda_{i4}^2 \lambda_{i5}^2 + \lambda_{i4}^2 \lambda_{i6}^2 + \lambda_{i5}^2 \lambda_{i6}^2$. Summing up those terms from two layers allows us to simplify the within-layer structure simultaneously for the two layers.

The rotation criterion should also be able to constrain the between-layer relationship to avoid items being loaded on the same second-level factor but different first-level factors. Ideally, this constraint on the parent-child relationship can be achieved using indicator functions. In the hierarchical structure shown in eq. (1), we consider the product of two sums for each child factor. If F3 is a child of F1, as shown in Fig. 1 and Eq. (1), for all the items loading on F1, the sum of their squared loadings on F3, $\sum_{i=1}^{p} I\left(\lambda_{i1}^2 \neq 0\right) \cdot \lambda_{i3}^2$, should be non-zero; for all the items loading on F2, the sum of their squared loadings on F3, $\sum_{i=1}^{p} I\left(\lambda_{i2}^2 \neq 0\right) \cdot \lambda_{i3}^2$, should be zero. Taking the product of these two sums encourages F3 to be the child of either F1 or F2: In other words, it penalizes the case when F3 is the child of both F1 and F2. Similarly, we can encourage that F1 is the only parent of F4, and F2 is the only parent of F5 and F6.

The proposed rotation criterion function designed for a two-layer binary-split hierarchical structure can be written in the following equation:

$$
\begin{aligned}
P(\Lambda) = \sum_{i=1}^{p} & \left(\lambda_{i1}^2 \lambda_{i2}^2 + \lambda_{i3}^2 \lambda_{i4}^2 + \lambda_{i3}^2 \lambda_{i5}^2 + \lambda_{i3}^2 \lambda_{i6}^2 + \lambda_{i4}^2 \lambda_{i5}^2 + \lambda_{i4}^2 \lambda_{i6}^2 + \lambda_{i5}^2 \lambda_{i6}^2\right) \\
& + \left(\sum_{i=1}^{p} I\left(\lambda_{i1}^2 = 0\right) \cdot \lambda_{i3}^2\right) \times \left(\sum_{i=1}^{p} I\left(\lambda_{i2}^2 = 0\right) \cdot \lambda_{i3}^2\right) \\
& + \left(\sum_{i=1}^{p} I\left(\lambda_{i1}^2 = 0\right) \cdot \lambda_{i4}^2\right) \times \left(\sum_{i=1}^{p} I\left(\lambda_{i2}^2 = 0\right) \cdot \lambda_{i4}^2\right) \\
& + \left(\sum_{i=1}^{p} I\left(\lambda_{i1}^2 = 0\right) \cdot \lambda_{i5}^2\right) \times \left(\sum_{i=1}^{p} I\left(\lambda_{i2}^2 = 0\right) \cdot \lambda_{i5}^2\right) \\
& + \left(\sum_{i=1}^{p} I\left(\lambda_{i1}^2 = 0\right) \cdot \lambda_{i6}^2\right) \times \left(\sum_{i=1}^{p} I\left(\lambda_{i2}^2 = 0\right) \cdot \lambda_{i6}^2\right)
\end{aligned} \tag{2}
$$

The first seven terms constrain the within-layer relationship between factors in the same level, and the others consider the parent-child relationship between two layers of factors. As indicated by the notation, the proposed rotation criterion assumes the position of parent and children factors in the loading matrix: The second and third columns should be the parent factors F1 and F2, and the fourth to seventh columns should be the child factors F3 to F6. This rotation criterion can be generalized to cases where one parent has more than two children by adding more terms that control the relationship between a low-level factor and all other high-level factors. For example, if we have three first-level factors, the between-layer term for F3 can be generalized to $\left(\sum_{i=1}^{p} I\left(\lambda_{i1}^2 = 0\right) \cdot \lambda_{i3}^2\right) \times \left(\sum_{i=1}^{p} I\left(\lambda_{i2}^2 = 0\right) \cdot \lambda_{i3}^2\right) + \left(\sum_{i=1}^{p} I\left(\lambda_{i1}^2 = 0\right) \cdot \lambda_{i3}^2\right) \times \left(\sum_{i=1}^{p} I\left(\lambda_{i3}^2 = 0\right) \cdot \lambda_{i3}^2\right) + \left(\sum_{i=1}^{p} I\left(\lambda_{i2}^2 = 0\right) \cdot \lambda_{i3}^2\right) \times \left(\sum_{i=1}^{p} I\left(\lambda_{i3}^2 = 0\right) \cdot \lambda_{i3}^2\right)$, which ensures that F3 has only one parent. It can also be generalized beyond two layers of factors by adding another collection of terms constraining the children-grandchildren relationship.

2.2 Computational Techniques

The non-continuous indicator function was approximated by the smooth exponential function such that the criterion function is differentiable as required by the optimization algorithm. The exponential function used in this simulation is $y = e^{-\alpha x}$, with $\alpha = 1{,}000{,}000$.

The Riemannian trust-region algorithm (Liu, 2020) was used to perform the orthogonal rotation, a second-order optimization algorithm for numerical search on the orthogonal group, i.e., the space of rotation matrices. It converges much faster than the gradient projection algorithm (Jennrich, 2001) with fewer iterations.

As the criterion function is sensitive to starting values and may converge to local minima, we used multiple random starts and chose the best as the final solution. It is a common practice for rotation criterions that are sensitive to starting values (e.g., Kiers, 1994; Rozeboom, 1992). In this study, we used 30 random starts and pick the solution with the minimum resulted function value from all 30 solutions.

2.3 Simulation Design

To understand the tolerance of the criterion function to the non-perfect and diverse EFA practices, three design factors were manipulated in the simulation study: loading matrices having (a) different numbers of rows/items; (b) magnitude of small errors replacing zeros; and (c) equal/unequal numbers of manifest variables among sibling factors. This study considers hierarchical structures with binary split: The numbers of items are 16, 32, and 64. To generate a scenario that is more realistic than the condition with exact hierarchical patterns, slight departures from

the exact structure were generated by substituting exact zero loadings with random variates independently sampled from Uniform [−0.05, 0.05] or Uniform [−0.1, 0.1]. The non-zero loadings were simulated from Uniform [0.3, 1]. For the third factor involving the balancing condition of items loading on sibling factors, we considered three levels: balanced loading matrices with 50:50 items loading on sibling factors (*balanced*), unbalanced loading matrices with 35:65 items loading on sibling factors (*unbalanced 1*), and unbalanced loading matrices with 20:80 items loading on sibling factors (*unbalanced 2*). Note that F3 and F4 are siblings because their parent is F1, and F3 and F5 are not siblings because they do not have the same parent. For the unbalanced conditions, we may have non-integer numbers of items per factor (e.g. 16×0.2=3.4), and those numbers were rounded to the closest integers.

An initial loading matrix with an unrecognizable structure, which mimics the result of EFA, was the matrix to be rotated using the proposed rotation criterion. We need to make sure the initially unrecognizable matrix is finally recognizable. Therefore, the initial matrix was created by randomly rotating the true loading matrix. Thirty replications were done for each of the 3×3×3 conditions. The convergence tolerance was set to 10^{-5}. To evaluate the results, we calculate the scaled Frobenius-norm error between the true matrix and the rotated loading matrix averaged across 30 replications:

$$\frac{1}{\sqrt{pk}}\left\|\hat{\Lambda}-\Lambda\right\|_F=\sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{k}\left(\hat{\lambda_{ij}}-\lambda_{ij}\right)^2}{pk}}, \tag{3}$$

where k is the total number of factors or columns in the complete loading matrix Λ and p is the number of items.

3 Results

For the 30 replications of each of the 3×3×3 conditions, all the final solutions converged, and the mean of minimized criterion function values over 30 replications are summarized in Table 1. The results show that when the true loading matrix has some errors, the function value will be greater than 0 because the true value itself is greater than 0. Holding the error range and the extent of balance constant, the more items we have, the larger the minimized function value; holding the number of items and the error range constant, the more balanced the true loading matrix, the larger the minimized function value.

The scaled Frobenius-norm error was summarized in Table 2 and depicted in Fig. 2. The values in Table 2 reflect the estimated error per entry and can be seen as the "averaged distance" between true and estimated loading matrices, considering the size of the matrices. Table 2 shows that holding the error range and the number of items constant, the more unbalanced the true matrix is, the larger the distances

Table 1 The mean of minimized criterion function values over 30 replications for all conditions

		Balanced (50%)	Unbalanced 1 (35%)	Unbalanced 2 (20%)
16 items	0 errors	0	0	0
	Unif (−0.05, 0.05)	0.007	0.006	0.005
	Unif (−0.1, 0.1)	0.023	0.018	0.008
32 items	0 errors	0	0	0.001
	Unif (−0.05, 0.05)	0.019	0.019	0.016
	Unif (−0.1, 0.1)	0.071	0.069	0.033
64 items	0 errors	0	0	0
	Unif (−0.05, 0.05)	0.043	0.043	0.043
	Unif (−0.1, 0.1)	0.174	0.163	0.11

Table 2 The scaled Frobenius-norm error averaged across 30 replications for all conditions

		Balanced (50%)	Unbalanced 1 (35%)	Unbalanced 2 (20%)
16 items	0 errors	0	0	0.042
	Unif (−0.05, 0.05)	0.057	0.081	0.258
	Unif (−0.1, 0.1)	0.180	0.241	0.327
32 items	0 errors	0	0	0.017
	Unif (−0.05, 0.05)	0.035	0.044	0.172
	Unif (−0.1, 0.1)	0.114	0.223	0.335
64 items	0 errors	0	0	0
	Unif (−0.05, 0.05)	0.034	0.035	0.047
	Unif (−0.1, 0.1)	0.079	0.123	0.328

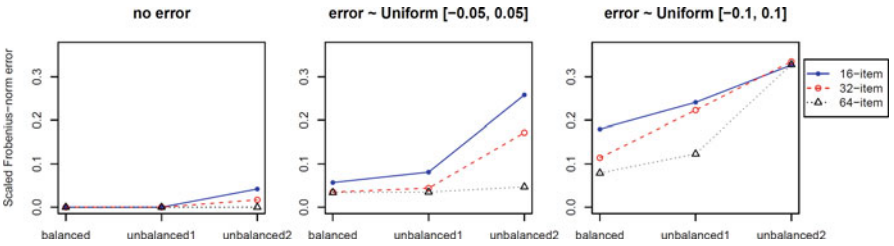

Fig. 2 The scaled Frobenius-norm error averaged across 30 replications for all conditions

between true and estimated loading matrix. Holding the balance extent and the number of items constant, the larger the errors, the larger the distances. Holding the balance extent and the error range constant, the more items we have, which means more pieces of information, the smaller the distances are.

4 Discussions

This proposed function is a generalization of Jennrich and Bentler's exploratory bifactor analysis to a two-layer hierarchical structure, which facilitates recovering and testing of a more complicated hierarchical structure in EFA with limited prior knowledge about item-factor dependency. Starting from an initial matrix with an unrecognizable structure, we can find a rotation matrix such that the rotated matrix is as close to a matrix with the hierarchical structure as possible. Our simulation results suggest that the proposed criterion function is generally robust to realistic scenarios when slight to moderate departures from a perfect hierarchical structure are present and when the matrix is moderately unbalanced. We also observe that parameter recovery is worsened by the magnitude of imperfect factor structure and the unbalancedness while improved as the number of items increases (Fig. 2). When the error ranges from -0.1 to 0.1, given that the meaningful loading value ranges from 0.3 to 1, we may have a too-small signal-to-noise ratio to recover the true hierarchical loading matrix. The negative effect of errors was intensified when we also have a severely unbalanced true matrix.

There are some limitations of the current study. First, the proposed rotation criterion requires us to know exactly how many factors we have at each layer and locks the positions of factors of a specific layer. In other words, the criterion function implicitly assumes that the general factor lies in the first column, the two first-level factors lie in the second and third columns, and the four second-level factors lie in the fourth to seventh columns. The problem is that, if we reorder the columns of a loading matrix with a perfect hierarchical structure in some ways, say, let the general factor lie in the fourth column and a second-level factor lie in the first column, then the function value is not zero, even though the re-ordered matrix has a perfect hierarchical structure. An ideal criterion function should always give a zero for matrices with the perfect hierarchical structure regardless of the ordering of columns, which is not satisfied by the current criterion function. In our numerical experiments, we observed local minima for the simpler bifactor structure as well, but not as severe as the two-layer structure which has more constrains to the column-wise relationships. Second, the proposed criterion function has many local minima. Some initial matrices may be easily rotated to a point which does not give a zero function value, then be further distorted when forcing the function value to be zero. Using random starts is only a heuristic solution, and more investigation is needed.

References

Allen, T. A., & DeYoung, C. G. (2017). Personality neuroscience and the five factor model. In *Oxford handbook of the five factor model* (pp. 319–352). Oxford University Press.

Browne, M. (1972). Orthogonal rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology, 25*(1), 115–120.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*(1), 111–150.

Cooke, D. J., Michie, C., & Skeem, J. (2007). Understanding the structure of the psychopathy checklist–revised: An exploration of methodological confusion. *The British Journal of Psychiatry, 190*(S49), s39–s50.

Jennrich, R. I. (2001). A simple general procedure for orthogonal rotation. *Psychometrika, 66*(2), 289–306.

Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika, 76*(4), 537–549.

Kiers, H. A. L. (1994). SIMPLIMAX: Oblique rotation to an optimal target with simple structure. *Psychometrika, 59*, 567–579.

Liu, Y. (2020). Riemannian Newton and trust-region algorithms for analytic rotation in exploratory factor analysis. *British Journal of Mathematical and Statistical Psychology*.

Rozeboom, W. W. (1992). The glory of suboptimal factor rotation: Why local minima in analytic optimization of simple structure are more blessing than curse. *Multivariate Behavioral Research, 27*, 585–599.

Yung, Y. F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika, 64*(2), 113–128.

Comparison Between Different Estimation Methods of Factor Models for Longitudinal Ordinal Data

Silvia Bianconcini and Silvia Cagnone

1 Introduction

In recent years, common statistical applications have dealt with multivariate longitudinal data with the purpose of measuring changes in constructs over time, such as attitudes, opinions, performances and abilities. In this context, both the multifaceted nature of the data and the longitudinal evolution of the underlying constructs have to be studied jointly, and Generalised Linear and Latent Variable Models (GLLVMs) (Dunson, 2003; Cagnone et al., 2009) represent a useful framework. GLLVMs assume that the entire set of the responses given by an individual to a certain number of items at different occasions, called the response pattern, can be expressed as a function of one or more latent variables and random effects through a monotone differentiable link function.

A potential barrier to the application of these latent variable models is the computational challenge presented by typically large datasets. Panel studies usually have several thousands of respondents which, when combined with multiple waves of measurement and a large choice set, renders unfeasible existing estimation approaches (likelihood-based and Bayesian one). Even when cross-sectional models are used, if the observed variables are of different nature, continuous and discrete, the estimation of these models is cumbersome. It can be carried out using a full information maximum likelihood method via either the EM algorithm or direct maximisation, but, in both cases, the integrals involved in the likelihood computation have no analytical solutions and need to be approximated. This problem is more evident in presence of longitudinal data since the number of latent variables and random effects increases proportionally to the number of observed items. In

S. Bianconcini (✉) · S. Cagnone
Department of Statistical Sciences, University of Bologna, Bologna, Italy
e-mail: silvia.bianconcini@unibo.it; silvia.cagnone@unibo.it

M. Wiberg et al. (eds.), *Quantitative Psychology*, Springer Proceedings in Mathematics & Statistics 353, https://doi.org/10.1007/978-3-030-74772-5_2

presence of multidimensional longitudinal data, classical quadrature techniques, such as the adaptive Gauss Hermite, that represents the gold standard in the GLLVM framework, is already unfeasible when four items are observed at three different time points. Alternatively, the widely applied Laplace approximation is known to provide inaccurate estimates in presence of discrete data.

Alternative methods that can be used to produce estimators with desired statistical properties and that, in addition, simplify the estimation process, are greatly needed. The most popular method that offers reduction in estimation complexity is the composite likelihood approach, introduced by Lindsay in 1988 (Lindsay, 1988) and further discussed, among the others, by Varin and Vidoni (2005). The composite likelihood estimator is obtained by maximising the univariate and/or bivariate likelihood products that contain the greatest quantity of model parameter information. The immediate effect of the composite likelihood estimation is the reduction of the number of integrations required in the likelihood computation.

Another approach that has been recently proposed in the literature is the dimension-wise quadrature, developed by Bianconcini et al. (2017). It consists in reducing the dimension of the multidimensional integrals by truncating the Taylor series expansion of the integrand. This makes the computation feasible also when the number of latent variables is large. The proposed approach provides a higher order approximation than the Laplace one but does not require any derivative computation, hence it is very simple to implement. Furthermore, the corresponding estimators are asymptotically as accurate as the adaptive Gauss Hermite estimators.

This paper investigates the use of pairwise likelihood methods and dimension-wise quadratures for estimating latent variable models for multivariate longitudinal ordinal data. A simulation study is carried out to compare the performance of these estimators under different common empirical scenarios, and their behaviour is also highlighted through a real data example.

2 Generalized Linear Latent Variable Models for Longitudinal Ordinal Data

Let $y_{t1}, y_{t2}, \ldots, y_{tp}$ be a vector of p ordinal observed variables at time (t, $t = 1, \ldots, T$) each of them with c_j, $j = 1, \ldots, p$, categories, and $z_1, z_2, \ldots, z_T$ a latent variable that accounts for the associations among the p items at each time point. Let $u_1, u_2, \ldots, u_p$ be p random effects that account for the associations of the same item at different time points. We consider the Generalized Linear Latent Variable Models (GLLVM) approach for longitudinal data that has been discussed by Dunson (2003) for mixed observed variables, and by Cagnone et al. (2009) in the specific case of ordinal data. According to the GLLVM approach, we define the joint density of the observed variables as

$$f(\mathbf{y}) = \int_{R^q} g(\mathbf{y} \mid \mathbf{z}, \mathbf{u}) h(\mathbf{z}, \mathbf{u}) d\mathbf{z} d\mathbf{u}$$

where $g(\mathbf{y} \mid \mathbf{z}, \mathbf{u})$ is referred to as measurement part of the model and $h(\mathbf{z}, \mathbf{u})$ as structural part of the model. The dimension of the integral depends on the number of observed variables and the number of time points, that is $q = p + T$.

The measurement part of the model is defined as a generalized linear model with random component given by

$$g(\mathbf{y}|\mathbf{z}, \mathbf{u}) = \prod_{t=1}^{T} \prod_{j=1}^{p} g(y_{tj}|z_t, u_j) = \tag{1}$$

$$= \prod_{t=1}^{T} \prod_{j=1}^{p} \prod_{r=1}^{c_j} (\gamma_{tj(r)}(z_t, u_j) - \gamma_{tj(r-1)}(z_t, u_j))^{y_{tj(r)}}, r = 1, \ldots, c-1$$

where the first equality comes from the conditional independence assumption between items and over time. Each conditional marginal density $g(y_{tj}|z_t, u_j)$ follows a multinomial distribution of parameter $\gamma_{tj(r)}(z_t, u_j)$ that is the cumulative probability that an individual responds to item j at time t up to category r. $y_{tj(r)}$ is a dummy variable that assumes value 1 up to category r.

The systematic component defines the linear predictor

$$\eta_{tj(r)} = \tau_{tj(r)} - \lambda_{tj} z_t - u_j$$

where $\tau_{tj(r)}$'s are item, time and category-dependent thresholds and λ_{tj}'s are item and time-dependent factor loadings. The link between the systematic component and the conditional means of the random component distributions is $\eta_{tj(r)} = \nu_{tj(r)}(\gamma_{tj(r)}(z_t, u_j))$ where $\nu_{tj(r)}$ is the link function which can be any monotonic, differentiable function. We consider the logit link function.

As for the structural part of the model, we consider the specification given by Cagnone et al. (2009), that is we assume that the latent variables follow an autoregressive no stationary process of first order as follows

$$z_t = \phi z_{t-1} + \delta_t, \tag{2}$$

where ϕ is the autoregressive coefficient, $\delta_t \sim N(0, 1)$ and $z_1 \sim N(0, \sigma_1^2)$.

Moreover, the joint density $h(\mathbf{z}, \mathbf{u}) \sim N(\mathbf{0}, \mathbf{\Psi})$ where

$$\mathbf{\Psi} = \begin{bmatrix} \mathbf{\Gamma} & 0 \\ 0 & \mathbf{\Omega} \end{bmatrix}.$$

$\mathbf{\Omega} = diag_{j=1,\ldots,p}\{\sigma_{uj}^2\}$ and the inverse of $\mathbf{\Gamma}$ has a well known special pattern whose expression can be found in Cagnone et al. (2009).

3 Model Estimation

Model estimation is usually performed by using a full maximum likelihood method. Given a sample of size n, the log-likelihood is given by:

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(\mathbf{y}_i, \boldsymbol{\theta}) = \sum_{i=1}^{n} \log \int_{R^q} g(\mathbf{y}_i \mid \mathbf{z}_i, \mathbf{u}_i) h(\mathbf{z}_i, \mathbf{u}_i) d\mathbf{z}_i d\mathbf{u}_i \tag{3}$$

where $\boldsymbol{\theta}$ is the vector of parameters to be estimated. A problem related to the maximization of the log-likelihood is that, in general, the multidimensional integral in (3) is not solvable analytically.

Among the remedies proposed in the literature, numerical quadrature-based methods represent a widespread solution to this problem and, among them, the adaptive Gauss Hermite quadrature is considered the gold standard approach (Rabe-Hesketh et al., 2005; Schilling & Bock, 2005). Alternatively, the Laplace approximation avoids the integral computation and represents the easiest method to implement (Liu & Pierce, 1994; Huber et al., 2004). The adaptive Gauss Hermite quadrature provides more accurate estimates than the Laplace approximation, but it is computationally unfeasible as the number of latent variables and random effects increases. To overcome these limitations, recent solutions proposed for factor models for longitudinal ordinal data are the pairwise likelihood approach (Vasdekis et al., 2012) and the dimension-wise quadrature method (Bianconcini et al., 2017).

3.1 Pairwise Likelihood Approach

The pairwise likelihood estimator is obtained by maximizing bivariate likelihood products that contain the greatest quantity of model parameter information (Lindsay, 1988; Cox & Reid, 2004). The immediate effect of the pairwise likelihood estimation is the reduction of the number of integrations in the expression of the likelihood (3). Indeed, the contribution of any given individual to the pairwise log-likelihood is given by

$$pl(\theta; \mathbf{y}) = \sum_{i=1}^{n} pl(\theta; \mathbf{y}_i) \tag{4}$$

with $pl(\theta; \mathbf{y}_i) = \sum_{t,t',j,j'} \log f(y_{tji}, y_{t'j'i}; \theta)$. In particular, for the factor model for longitudinal ordinal data described in Sect. 2, the bivariate density for a pair of responses $(y_{tji}, y_{t'j'i})$ is given by

$$f(y_{tji}, y_{t'j'i}; \theta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(y_{tji}, y_{t'j'i}, z_{ti}, z_{t'i}, u_{ji}, u_{j'i}) dz_{ti} dz_{t'i} du_{ji} du_{j'i}$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(y_{tji}|z_{ti}, u_{ji}) f(y_{t'j'i}|z_{t'i}, u_{j'i}) \times \tag{5}$$

$$\times h(z_{ti}, z_{t'i}, u_{ji}, u_{j'i}) dz_{ti} dz_{t'i} du_{ji} du_{j'i}.$$

The dimensional of the integrals involved in the expression of the likelihood components is four and if $j = j'$ or $t = t'$ it reduces to three. Thus, the three/four-dimensional integral can be easily approximated using a the Gauss Hermite quadrature method.

The resulting estimators are generally consistent and asymptotically normally distributed (Varin & Vidoni, 2005).

3.2 Dimension-Wise Quadrature Method

Let $\mathbf{b} = (\mathbf{z}, \mathbf{u})$ denote the vector of latent variables and random effects and consider the following representation of the marginal density function

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

where $\mathbf{b}_{mo}$ is the maximum of the integrand $g(\mathbf{y} \mid \mathbf{z}, \mathbf{u}) h(\mathbf{z}, \mathbf{u})$ and $\boldsymbol{\Sigma}_{mo} = \mathbf{C}_{mo}\mathbf{C}'_{mo}$ is minus the inverse of the corresponding Hessian matrix evaluated in the mode $\mathbf{b}_{mo}$. $\phi(\cdot)$ the normal density function.

The dimension-wise method is applied to the expected value $E_{\phi}[m(\mathbf{b}^*)]$. It is based on the Taylor expansion of $m(\mathbf{b}^*)$ around 0 up to the s term as follows

$$\hat{m}(\mathbf{b}^*) = \sum_{w=1}^{s} t_w \tag{7}$$

where each component t_w considers all the derivatives of $m(\mathbf{b}^*)$ taken with respect to w latent factors, that is

$$t_w = \sum_{j_1, j_2, \ldots, j_w} \sum_{k_1 < k_2 < \ldots < k_w} \frac{1}{j_1! j_2! \ldots j_w!} \frac{\partial^{j_1 + j_2, \ldots, + j_w} m}{\partial b_{k_1}^{*j_1} \partial b_{k_2}^{*j_2} \ldots \partial b_{k_w}^{*j_w}} (\mathbf{0}) b_{k_1}^{*j_1} b_{k_2}^{*j_2} \ldots b_{k_w}^{*j_w} \tag{8}$$

The approximated function $\hat{m}(\mathbf{b}^*)$ admits the following equivalent representation (Bianconcini et al., 2017)

$$\hat{m}(\mathbf{b}^*) = \sum_{l=0}^{s}(-1)^l \binom{q-s+l-1}{l} m_{s-l}(\mathbf{b}^*) \tag{9}$$

where $m_{s-l}(\mathbf{b}^*) = m(0, \cdots, b^*_{k_1}, 0 \cdots, 0, b^*_{k_{s-l}}, 0, \cdots, 0)$. Thus, m_{s-l} is a function of just $s-l$ variables being all the remaining fixed to 0. Replacing Eq. (9) in Eq. (6), we obtain the approximate density function

$$f_a(\mathbf{y};\theta) = |\mathbf{C}_{mo}| \left[\sum_{l=0}^{s-1}(-1)^l \binom{q-s+l-1}{l} \int_{R^{s-l}} \sum_{k_1<\ldots<k_{s-l}} m_{s-l}(\mathbf{b}^*) \times \right. \tag{10}$$
$$\left. \phi(b^*_{k_1}) \cdots \phi(b^*_{k_{s-i}}) db^*_{k_1} .. db^*_{k_{s-l}} \right].$$

The dimension of the integrals in expression (10) depends on the choice of s. If $s = 1$, we obtain a linear combination of unidimensional integrals, if $s = 2$, we obtain a linear combination of uni- and bi-dimensional integrals and so on. For small values of s, the integrals can be easily approximated using the Gauss Hermite quadrature method. In the extreme cases of $s = 0$ and $s = q$ the solution is equivalent to the classical Laplace and to the adaptive Gauss-Hermite quadrature ones, respectively.

The dimension-wise quadrature estimators share the same accuracy as the adaptive Gauss-Hermite method (Bianconcini et al., 2017), but avoiding the main computational limitations of the latter.

4 Simulation Study

The performance of the two approximation methods are compared through a simulation study.

We consider a first scenario where the number of factors is fixed to four, that is we assume to have observed two items at two different occasions ($q = 4$, $p = 2$, $T = 2$). This simple scenario allows us to also consider the Adaptive Gauss Hermite (AGH) quadrature that is feasible with number of quadrature points nq equal to 8 (AGH8) and 15 (AGH15). The AGH with 15 quadrature points can be considered the benchmark providing an almost exact representation of the function (Bianconcini et al., 2017). In this scenario, the AGH corresponds to the dimension-wise method (DWM) with $s = 4$. We also estimate the classical Laplace approximation, that corresponds to DWM with $s = 0$. Finally, we consider the dimension-wise quadrature with $s = 1, 2$, and 3. We set sample sizes equal to $n = 200$ and 1000, and, as suggested by Vasdekis et al. (2012), we consider 100 replications for each condition of the study.

Table 1 reports the bias and Root Mean Square Error (rmse) of the parameter estimates in the case of $n = 200$. The first loading is fixed to 1 for identification reasons and measurement invariance of thresholds and loadings over time is assumed.

As known, the Laplace approximation produces strongly biased estimates but with generally lower rmse than AGH15. The accuracy of the DWM estimates improves as s increases, and with similar results to AGH15 when $s = 3$.

The pairwise method also produces similar estimates to those of AGH15. Thus, for finite samples ($n = 200$), the pairwise method and the dimension-wise with $s = 3$ perform similarly. The main differences in the performance of the pairwise and dimension-wise methods are evident in larger samples, that is for $n = 1000$.

Figures 1 and 2 show the density estimates for λ_2 and ϕ, respectively.

The Laplace estimator is very inaccurate whereas the DWM estimator based on $s = 3$ is the closest to the AGH15 one. The pairwise behaves in between DWM with $s = 2$ and DWM with $s = 3$. It is interesting to notice that even if DWM with $s = 1$ is less accurate than the same method with higher values of s and pairwise, it is far superior to the Laplace estimator. The parameter estimators not shown here have a similar performance.

Table 1 Results for the 4-factor model, $n = 200$

	Pairwise		*Laplace* $s = 0$		*DWM* $s = 1$		*DWM* $s = 2$		*DWM* $s = 3$	
True	Bias	rmse	Bias	rmse	Bias	rmse	Bias	rmse	Bias	rmse
$\lambda_1 = 1$										
$\lambda_2 = 1.61$	0.00	0.70	−0.67	0.67	−0.35	0.83	−0.15	0.87	0.06	0.82
$\phi = 0.8$	−0.02	0.19	0.11	0.12	0.01	0.15	0.00	0.17	−0.05	0.18
$\sigma_1^2 = 2$	0.51	1.29	0.69	1.39	1.10	2.84	0.92	2.26	0.59	1.50
$\sigma_{u_1}^2 = 1$	−0.12	0.71	−0.96	1.11	−0.76	1.06	−0.46	0.99	−0.13	0.75
$\sigma_{u_2}^2 = 1.5$	0.19	0.95	−0.44	0.92	−0.52	1.22	0.20	1.28	0.50	1.62
	AGH8 $s = 4$		*AGH15* $s = 4$							
True	Bias	rmse	Bias	rmse						
$\lambda_1 = 1$										
$\lambda_2 = 1.61$	0.03	0.76	0.05	0.78						
$\phi = 0.8$	−0.02	0.17	−0.02	0.17						
$\sigma_1^2 = 2$	0.59	1.58	0.57	1.58						
$\sigma_{u_1}^2 = 1$	−0.10	0.84	−0.14	0.81						
$\sigma_{u_2}^2 = 1.5$	0.34	1.25	0.48	1.59						

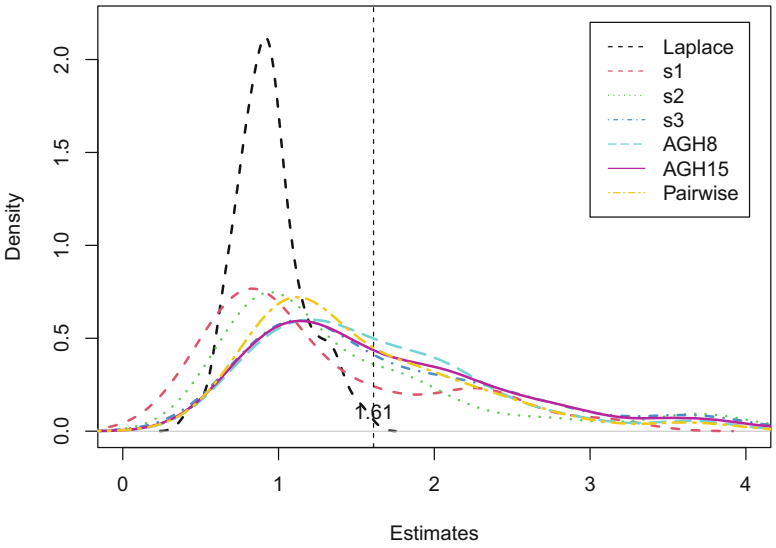

Fig. 1 Densities estimation of λ_2 (true value: 1.61) for the 4-factor model, $n = 1000$

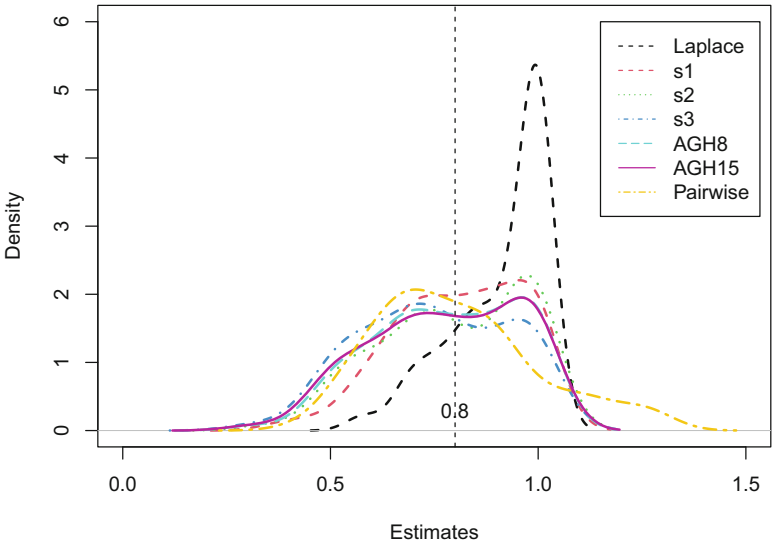

Fig. 2 Densities estimation of ϕ (true value: 0.8) for the 4-factor model, $n = 1000$

In Fig. 3 the values of the log-likelihood for an increasing number of quadrature point nq, that is 3, 5, 8, 11, and 15 are reported for the analyzed methods. We can observe that in all the cases pairwise performs worse than AGH, DWM with $s = 2$ and $s = 3$, whereas the log-likelihood values based on DWM with $s = 3$ and AGH

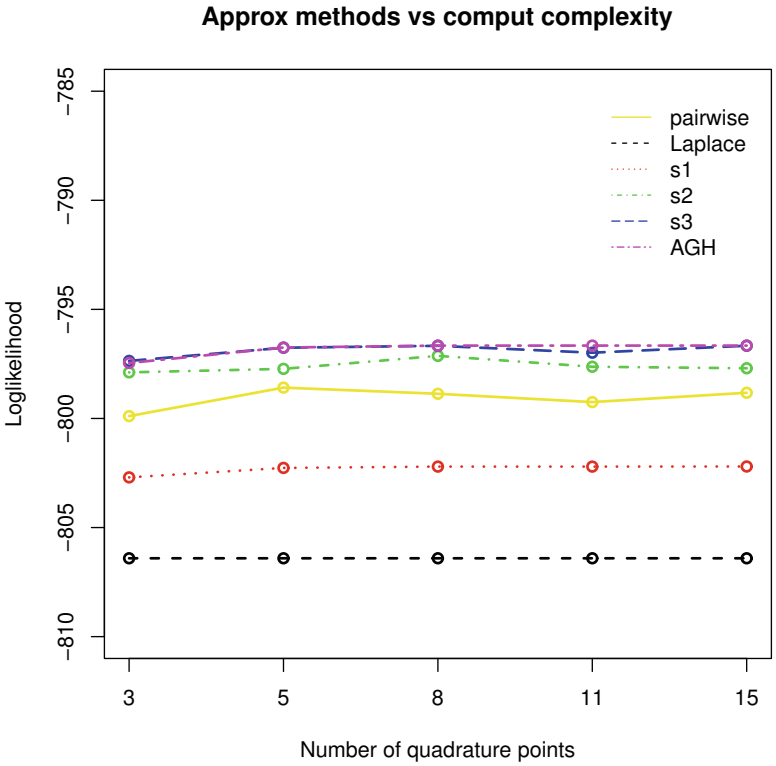

Fig. 3 Log-likelihood for increasing number of quadrature points, 4-factor model, $n = 1000$

are almost identical, independently on the number of quadrature points used in the integral approximations.

In the second simulation scenario we consider a factor model with seven factors, that is we assume to observe three items at four different occasions ($q = 7$, $p = 3$, $T = 4$). In this setting the AGH is not feasible. As before, we consider sample sizes equal to 200 and 1000, and 100 replications for each condition of the study.

The results are reported in Tables 2 and 3 for $n = 200$ and $n = 1000$, respectively. It is interesting to notice that, in this scenario, for both small and large sample sizes, even DWM with $s = 2$ outperforms the pairwise method in terms of bias as well as rmse. The pairwise performs more similarly to DWM with $s = 1$.

5 Real Data Analysis

Lastly, the two different approximation techniques considered through this research are compared using a real dataset, taken from the British Household Panel Survey

Table 2 Results for the 7-factor model, $n = 200$

	Pairwise		*Laplace* $s = 0$		*DWM* $s = 1$		*DWM* $s = 2$		*DWM* $s = 3$	
True	Bias	rmse	Bias	rmse	Bias	rmse	Bias	rmse	Bias	rmse
$\lambda_1 = 1$										
$\lambda_2 = 1.61$	0.23	0.39	−0.32	0.35	−0.15	0.34	−0.03	0.25	0.03	0.29
$\lambda_3 = 0.66$	0.05	0.20	−0.06	0.15	−0.10	0.17	−0.01	0.15	0.00	0.16
$\phi = 0.4$	−0.04	0.11	0.08	0.13	0.02	0.11	0.00	0.01	−0.01	0.10
$\sigma_1^2 = 2$	0.01	0.59	−0.29	0.44	−0.20	0.41	−0.03	0.11	0.00	0.33
$\sigma_{u_1}^2 = 1$	0.17	0.33	−0.32	0.50	−0.21	0.47	0.00	0.22	0.08	0.52
$\sigma_{u_2}^2 = 1.5$	0.38	0.61	−0.18	0.51	−0.07	0.49	0.03	0.26	0.05	0.52
$\sigma_{u_3}^2 = 2$	0.19	0.57	0.32	0.86	0.45	1.09	0.16	0.60	0.14	0.75

Table 3 Results for the 7-factor model, $n = 1000$

	Pairwise		*Laplace* $s = 0$		*DWM* $s = 1$		*DWM* $s = 2$	
True	Bias	rmse	Bias	rmse	Bias	rmse	Bias	rmse
$\lambda_1 = 1$								
$\lambda_2 = 1.61$	0.28	0.31	−0.33	0.33	−0.19	0.22	−0.05	0.12
$\lambda_3 = 0.66$	0.04	0.10	−0.06	0.09	−0.10	0.12	−0.01	0.07
$\phi = 0.4$	−0.05	0.07	0.09	0.10	0.03	0.06	0.01	0.11
$\sigma_1^2 = 2$	−0.08	0.25	−0.29	0.33	−0.20	0.25	−0.03	0.05
$\sigma_{u_1}^2 = 1$	0.28	0.31	−0.37	0.40	−0.26	0.32	−0.06	0.06
$\sigma_{u_2}^2 = 1.5$	0.42	0.47	−0.21	0.30	−0.11	0.24	−0.01	0.23
$\sigma_{u_3}^2 = 2$	0.14	0.28	0.17	0.37	0.26	0.46	0.05	0.31

(BHPS). The data are composed of an annual nationally representative sample of approximately 10,000 individuals (5000 households) aged 16 years and over. The main objective of the BHPS is to study social and economic changes in Britain at individual and household levels. To analyze the data, five waves were selected (1992, 1994, 1996, 1998, 2000), as well as three ordinal items that indicate social and political attitudes. After eliminating the missing values, we are left with a sample size of 3784 individuals. The manifest items that are to be analyzed are taken from the section on values and opinions and indicate respondents views on responsibility of the private sector, Government and trade unions for labor conditions. The items are:

1. Private enterprise is the best way to solve the UKs economic problems [Enterp].
2. It is the Governments responsibility to provide a job for everyone who wants one [Govern].

3. Strong trade unions are needed to protect the working conditions and wages of employees [TrUnion].

Permitted responses to these questions were agree strongly (AS), agree (A), not agree/disagree (Neither A nor D), disagree (D), disagree strongly (DS). Due to the fact that a small proportion of respondents fall into the first and last item categories, the first two and last two categories have been collapsed; thus leaving us with three categories for each item. Finally, the response categories of item Enterp have been reversed.

A one factor model is fitted with the time-dependent factors (latent variables) as an AR(1) autoregressive model. Previously, Vasdekis et al. (2012) analyzed the same data by allowing all model item parameters that are associated with the measurement model (thresholds and slopes) to vary across time points. The analysis proceeded with the fitting of the models, exclusively with: equal thresholds, equal loadings, and, finally, equal thresholds and loadings. The model with equal loadings, but not thresholds, across time was the preferred, and we quote here the results for this model. The model is estimated using the pairwise likelihood method and by applying the diwension-wise quadrature with different levels of approximation (s ranging from 0 to 3). The choice of s in the dimension-wise quadrature has been done by increasing its value until the mean of the relative absolute differences in parameter estimates ($Av(\Delta)$) was sufficiently small (order 10^{-3}). The stability in the estimated parameters is achieved with $s = 3$. Hence, Table 4 gives the estimated model parameters using the pairwise likelihood method and the dimension-wise quadrature with s set equal to three.

As observed in the simulation study, the two methods provide similar results. Loadings are all found to be positive and close to one-another. The large estimated variances for the random effects are indicative of the presence of a large amount of heterogeneity in the responses within each item over time. Heterogeneity that clearly could not be entirely accounted for by the autoregressive model was fitted upon the time-dependent latent variables. The latter are characterised by a strong correlation between subsequent time-dependent factors, which is judged by the large estimated autoregressive parameter ϕ equal to 0.83, with an estimated standard error equal to 0.002, for the pairwise method, and equal to 0.896 (with standard error equal to 0.005) for the dimension-wise quadrature.

Table 4 Estimated factor loadings with standard errors in brackets for the non-stationary model with time-specific latent variables and estimated variances for the item-specific random effects based on pairwise likelihood and dimension-wise quadrature ($s = 3$) methods

	Pairwise		*DWM*	
Item	λ_j	σ_{u_j}	λ_j	σ_{u_j}
Enterp	1.00	3.65 (0.21)	1.00	3.30 (0.15)
Govern	1.30 (0.07)	6.67 (0.37)	1.29 (0.04)	6.64 (0.33)
TrUnion	1.27 (0.07)	5.80 (0.40)	1.05 (0.03)	6.36 (0.71)

6 Discussion

The estimation of models for longitudinal data with random components, such as latent variable models, involve multidimensional integrations. A way to avoid the high dimensional integrations is to base the estimation and inference on lower data dimensions such as bivariate or trivariate. It is found that composite likelihood estimation, using the bivariate marginal likelihoods, and dimension-wise quadratures for a latent variable model with longitudinal ordinal responses, produce estimates with desirable properties. These results are endorsed by the simulation studies. In presence of multidimensional longitudinal data, these approximate likelihood estimation methods also allow the fit of more realistic models to data sets with many items and a few factors on each time dimension.

Both the simulation results and the real data application have shown the similar performance of the pairwise approach and the dimension-wise quadrature with $s = 3$. However, the latter is more advantageous than the pairwise likelihood method, since its implementation is straightforward and does not depend on the specified model. On the other hand, the pairwise approach becomes unfeasible in a longitudinal setting where the latent variable model has more than one factor at each occasion, and the matrix of the factor loadings does not have a simple structure.

Further work needs to be done to corroborate the findings of this study. The properties of the dimension-wise based and pairwise estimators should be investigated and compared theoretically. Furthermore, more efficient versions of the pairwise likelihood estimation based on weighted estimators should be also considered. The real application from the British Household Panel study has shown the potential of the proposed methodologies to the analysis of real data sets. More complex models for the structural part of the model and the inclusion of covariates in the measurement model are useful additions.

References

Bianconcini, S., Cagnone, S., & Rizopoulos, D. (2017). Approximate likelihood inference in generalized linear latent variable models based on the dimension-wise quadrature. *Electronic Journal of Statistics, 11*, 4404–4423.

Cagnone, S., Moustaki, I., & Vasdekis, V. (2009). Latent variable models for multivariate longitudinal ordinal responses. *British Journal of Mathematical and Statistical Psychology, 62*, 401–415.

Cox, D. R., & Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika, 91*, 729–737.

Dunson, D. (2003). Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association, 98*, 555–563.

Liu, Q., & Pierce, D.A. (1994). A note on Gauss-Hermite quadrature. *Biometrika, 81*, 624–629.

Lindsay, B. (1988). Composite likelihood methods. In N. U. Prabhu (Ed.), *Statistical inference from stochastic processes* (pp. 221–239). Providence: American Mathematical Society.

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