Proceedings of the 90th Annual International Meeting of the Psychometric Society, Minneapolis, United States (2025).

Bayesian item factor analysis when binary indicator variables have skewed marginal distributions

R. Noah Padgett [*†] and Sonja D. Winter [‡]

†Department of Epidemiology, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA
‡College of Education & Human Development, University of Missouri, Columbia, MO, USA
*Corresponding author. Email: npadgett@hsph.harvard.edu

Abstract

Bayesian estimation approaches for item factor analyses (IFA) can provide a valuable tool to overcome the difficulty in parameter estimation in frequentist approaches when the marginal response distribution is highly skewed. Highly skewed marginal response distributions are problematic in Bayesian IFA models with binary indicator variables depending on how the latent response distribution is parameterized. We provide an alternative implementation of Bayesian IFA under a DELTA, or a unit total variance, parameterization of the latent response distribution. We use data on psychopathology to examine the (in)consistency in factor loading and threshold estimates under relatively small sample sizes. Using a small-scale simulation using the full psychopathology dataset as a population (N=9,282), we repeatedly sampled (n=200) sampled from the full population that contains several highly skewed indicators (e.g., <3% of the population are positive for a psychology). The study highlights how the proposed DELTA parameterization for Bayesian IFA may overcome the limitations of the traditional fixed-residual variance THETA parameterization to yield interpretable and valid estimates for factor loading and threshold parameters in IFA when marginal distributions are highly skewed.

Keywords: thresholds, factor loadings, parameterization, Bayesian, item factor analysis

1. Introduction

Item factor models are often used in psychopathology research to account for diagnostic co-occurrence or comorbidity among supposedly distinct forms of psychopathology that are outlined in diagnostic manuals. The majority of people who meet diagnostic criteria for one mental disorder also meet criteria for a second or third diagnosis (Kessler et al., 2005; McGrath et al., 2020), suggesting that there may be a common underlying cause underlying subsets of psychological disorders (e.g., internalizing and externalizing or an overall p-factor). The utility of these psychopathology item factor models in informing theory (Watts et al., 2020), evaluating intervention efficacy (e.g., Constantinou et al., 2019), and supporting prevention efforts (e.g., Lynch et al., 2021) depends on reliable and accurate estimation of item parameters. However, psychological disorders differ significantly in their prevalence, ranging from relatively common (e.g., about 12% meet diagnostic criteria for specific phobia; Kessler et al., 2012) to extremely rare (e.g., <1% of people meet diagnostic criteria for agoraphobia; Roest et al., 2019). It is known items with such low prevalence negatively impacts the estimation of factor models (Bainter, 2017). Improving parameter recovery in the presence of a few highly skewed indicators would enhance the study of psychopathology and other fields in which item endorsement varies widely. Due to the rarity of some pathologies, we would expect the threshold (item location) parameters to be quite extreme in some samples due to random chance,

and using the currently implemented methods of estimating thresholds can lead to difficulties in interpreting estimates due to the uncertainty in the resulting location parameters. The current paper contributes to developing a robust approach to estimating these models under such ill-conditioned data structures. Next, we review some highlights from the methodological literature on estimation of item factor models.

1.1 Background on methodological research

Estimation methods for item factor analysis generally fail to provide accurate parameter recovery and interval estimate coverage when the marginal distribution of indicators is skewed (Forero et al., 2009; Moshagen & Musch, 2014; Wirth & Edwards, 2007). There is promise in using a Bayesian approach to help improve the parameter recovery and interval estimate coverage rates over maximum likelihood (Bainter, 2017). Specifically, the Bayesian approach uses priors to help regularize uncertain threshold estimates. However, current Bayesian approaches still lack adequate parameter recovery when the response distribution is highly skewed for all indicators (Bainter, 2017; Padgett et al., 2024). The response distribution is, however, not always severely skewed for all indicators, but only a subset. Thus, an open question when using Bayesian methods to estimate item factor models is how frequently a "rare" item can occur in a factor while still obtaining adequate coverage? Motivated by the challenges of item factor analysis faced in psychopathology research, our goal is to help provide an answer to this question.

In our experience, the regularizing effect of priors, such as the Dirichlet distribution, placed on thresholds can be quite extreme (Padgett et al., 2024). We hypothesize that one reason for this extreme regularizing effect of the Dirichlet prior could be how the category response probabilities are mapped onto the latent response distribution. How to parameterize the latent response distribution is an important consideration when using IFAs. The parameterization of the latent response distribution determines how the item location (or item thresholds) can be interpreted in the context of the population being studied, and once estimated, the thresholds determine the location and scale of the latent response distribution (Wu & Estabrook, 2016). Depending on the number of response categories, the thresholds can uniquely identify the relative mean and variance of a latent response distribution across contexts (e.g., between groups or across time). Failing to estimate the thresholds accurately can lead to misrepresentation of the latent structure, which, when moving from relatively simple single-group models to multiple groups, can lead to even worse estimation performance (e.g., interval estimates inconsistently containing the estimand of interest). This may be especially true in the context of Bayesian IFA where the threshold prior can have a regularizing effect on estimates even when using relatively diffuse priors (Padgett et al., 2024). The mapping approach in available Bayesian SEM software, such as blavaan, relies on the THETA parameterization (Merkle et al., 2023). The THETA parameterization uses a fixed residual variance for the latent response distribution (generally unit variances), leading to an unconstrained total latent response variance. The lack of conditional constraints on the latent response distribution leads to Bayesian IFA models with relatively straightforward prior specification for factor loadings and latent factor covariance matrices when implemented in software such as JAGS or Stan for general posterior approximation. However, under the THETA parameterization, severe regularizing effects can arise due to the latent response distribution possibly having a large total variance relative to the prior variance. The mismatch in variances leads to the threshold prior to be inappropriately scaled relative to the latent response distribution.

In this paper, we investigate whether respecifying the Bayesian IFA model by using the DELTA parameterization, where the total variance of the latent response distribution has a fixed unit variance, can remedy this mismatch between prior and latent response distribution and result in improved coverage. In addition to introducing a pseudo-DELTA parameterization, we aim to understand the effects of prior specifications, even differences among relatively diffuse specifications of the

normal and Dirichlet priors. Insight from this study will help provide guidance to researchers and methodologists on how to specify Bayesian IFAs in a principled manner.

The remainder of this paper is organized as follows. Next, we discuss the Bayesian IFA in more detail, including the latent response parameterization and prior specification. Then, using data for the National Comorbidity Survey Replication (NCS-R)(Kessler & Merikangas, 2004), we illustrate the impact of latent response parameterization on resulting parameter estimates using a simulation study. We end this paper with comments on parameterization of these types of factor models and areas of for future development of these methods.

2. Bayesian Factor Analysis with Binary Indicators

This study focuses on using Bayesian latent variable models with binary indicators. To begin, consider the measurement model for items $i = 1, \ldots, I$ with response categories that possibly vary among items. In the measurement model for cases, the latent variable η for $n = 1, \ldots, N$ cases can be denoted by

$$\mathbf{y}_n^* = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_n + \boldsymbol{\varepsilon}_n \tag{1}$$

$$y_{ni} = 0 \Leftrightarrow y_{ni}^* < \tau_i \tag{2}$$

$$y_{ni} = 1 \Leftrightarrow y_{ni}^* > \tau_i \tag{3}$$

In a Bayesian perspective, the model can be described distributionally by the latent factor variables $\boldsymbol{\eta}_n \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Phi})$, $\boldsymbol{\varepsilon}_n \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Theta})$, $\boldsymbol{\Lambda}$ is a matrix of factor loadings with possibly many zero entries to define the different factors, and $\boldsymbol{\Theta}$ is the latent response residual variance-covariance matrix. The latent response distribution can be described as $\boldsymbol{y}_n^* \sim \text{MVN}(\boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_n, \boldsymbol{\Theta})$. If the above model were implemented directly, this would be considered a *conditional* specification of the latent response distribution. However, in this work, we followed the approach utilized in blavaan (Merkle et al., 2021) by implementing the *marginal* specification of the latent variable model resulting in $\boldsymbol{\eta}$ not being directly sampled. Marginalizing the latent response distribution with respect to the latent factors results in the simplified distributional expression of $\boldsymbol{y}_n^* \sim \text{MVN}(\boldsymbol{\nu}, \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta})$. In IFA with binary indicators, the latent response distribution intercepts ($\boldsymbol{\nu}$) cannot be jointly identified with the threshold parameters (Wu & Estabrook, 2016). Due to the lack of identification, the vector latent response intercepts are fixed to $\mathbf{0}$. The other major identification concern is the scale, or variance, of the latent response distribution. Next, we will discuss the latent response parameterization.

2.1 Latent Response Parameterization

The general model described above can be parameterized in various ways to achieve identification (in a nontechnical, general sense). The two main approaches in an SEM tradition are using either a fixed total variance of the latent response distribution (DELTA parameterization) or a fixed residual variance for the latent response distribution (THETA parameterization). Kamata and Bauer (2008) provided an in-depth discussion of parameterizations for binary factor analysis, which we briefly summarize here.

Binary observed variables lead to several complexities in how the model in (1) can be identified due to the limitations of the amount of information that is available to inform model parameters. The information from a binary observed variable needs to inform the latent response distribution and a threshold model. These two pieces interact to obtain the overall model-implied endorsement probability. However, there is not enough information to uniquely identify all parameters (e.g., the threshold–τ, latent response intercept–ν, latent response residual variance–θ, factor loading–λ, latent factor variance–ϕ). The lack of information means that different sets of model parameters can lead to the same model-implied endorsement probabilities (a non-unique solution). We resolve this

indeterminancy by selecting which parameters are fixed for a given analysis: selecting between the THETA and DELTA parameterization, for instance.

The choice of parameterization for the latent response distribution is especially crucial for Bayesian item factor analysis because prior distributions need to be selected to maintain these constraints. The simplest parameterization to construct prior for is the fixed residual variance (THETA) parameterization because we aren't constrained in how we select priors for the factor loadings and factor (co)variances. However, the THETA parameterization results an unconstrained total variance for the latent response distribution. Prior simulation studies of Bayesian item factor analysis has pointed to this might be the cause of issues with item location parameter recovery when the observed endorsement rate is extreme (Bainter, 2017; Padgett et al., 2024). Padgett and colleagues suggested the poor parameter recovery may be due to the uncertainty associated with the latent response total variance. In this paper, we provide a way to resolve this issue.

While technically trickier, the fixed total variance (DELTA) parameterization is possible for Bayesian item factor analysis. Our proposal is to employ a "pseudo-DELTA" parameterization where an intermediate step is added to the Bayesian model between sampling from the priors and employing the likelihood function. The intermediate step is to standardized the latent response distribution. The steps are

1. Start with the fixed residual variance (THETA) parameterization.
2. Sample priors for factor loadings and factor (co)variance(s).
3. Build total model implied latent response covariance matrix: $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$
4. Standardized $\boldsymbol{\Sigma}$ to a correlation matrix
5. Use likelihood with standardized $\boldsymbol{\Sigma}$.

Using the intermediate step of standardizing the latent response distribution is expected to mimic the results obtained when using the DELTA parameterization from a frequentist estimation approach. This additional step fits well into existing Bayesian estimation algorithms, such as the one used in blavaan (Merkle & Rosseel, 2016) using Stan, by simply being an additional step in the transform parameters block. A major advantage of this mid-step standardization is that the likelihood is unchanged. The likelihood is the same regardless of using the correlation matrix or covariance matrix for the latent response distribution because these variables are indeterminant with respect to scale. But, by using the mid-step standardization, we achieve a constrained scale for the latent response distribution and thus the thresholds can be sampled without the latent response distribution variance changing every iteration. Additional details on the implemention of the methods are shown Appendix A.

The following simulation based on psychopathology data aims to provide some evidence that the "pseudo-DELTA" can be used in Bayesian item factor analysis to obtain estimates for thresholds and factor loadings that align with those from frequentist methods.

3. Methods

For this simulation study, we repeatedly (m=50) analyzed subsamples from a dataset on psychopathology classification of patients to evaluate how parameterization of the latent response distribution impacts the parameter estimates. The full dataset consists of a sample of 9,282 patients, and for this simulation study, we took random samples of size 200. For each sample, we estimated the item factor model using the THETA parameterization and the DELTA parameterization using frequentist and Bayesian estimation methods to evaluate the impact of parameterization and estimation method on parameter recovery.

3.1 Data

For this illustration, we used data from the National Comorbidity Survey Replication (NCS-R) study (Kessler & Merikangas, 2004), which is a large national probability sample of adults aged 18 to 99 (N = 9,282, M_{age} = 45, 55% female) collected between 2001 and 2003 in the United States. Recruitment and consent procedures were approved by the Human Subjects Committees of Harvard Medical School and the University of Michigan. We selected this data set because it has been used in a large number of well-cited studies on the latent structure of psychopathology (e.g., Ruscio et al., 2008). In addition, these data are publicly available (https://www.icpsr.umich.edu/web/ICPSR/studies/20240).

We included the following diagnoses in the item factor model: Alcohol Dependence (ALD), Drug Dependence (DRD), Conduct Disorder (CD), Major Depressive Disorder (MDD), Dysthemic Disorder (or Persistent Depressive Disorder; DYS), Generalized Anxiety Disorder (GAD), Agoraphobia (AGO), Social Anxiety Disorder (SO), Specific Phobia (SP), and Panic Disorder (PDS). The NCS-R assessed psychiatric diagnoses using trained, non-clinician interviewers who administered a modified version of the World Health Organization Composite International Diagnostic Interview (for details, see Kessler & Üstün, 2004). We used binary indicators of lifetime diagnoses of psychological disorders that were generated without hierarchical exclusion rules (with the exception of major depressive disorder, which is only available with hierarchical exclusion rules), typical in psychiatric epidemiology (Kessler & Merikangas, 2004). Diagnoses were based on the Diagnostic and Statistical Manual of Mental Disorders, version IV (American Psychiatric Association [APA], 1994).

3.2 Model Specification

The model specification was informed by Krueger 1999, whose paper has been cited over 2400 times. Krueger 1999 specified a three correlated factors model with externalizing (tendencies toward poor behavioral and emotional control), distress (tendencies toward anxious misery), and fear (tendencies towards phobic responses to stressors) dimensions. In our model (see Figure 1), the externalizing factor was composed of alcohol dependence, drug dependence, and conduct disorder; the distress factor was composed of major depression, dysthymia, and generalized anxiety disorder; and the fear factor was composed of agoraphobia, panic disorder, social anxiety disorder, and specific phobia. Although Krueger 1999 included antisocial personality disorder diagnoses, we used conduct disorder diagnoses because NCS-R only makes the latter available.

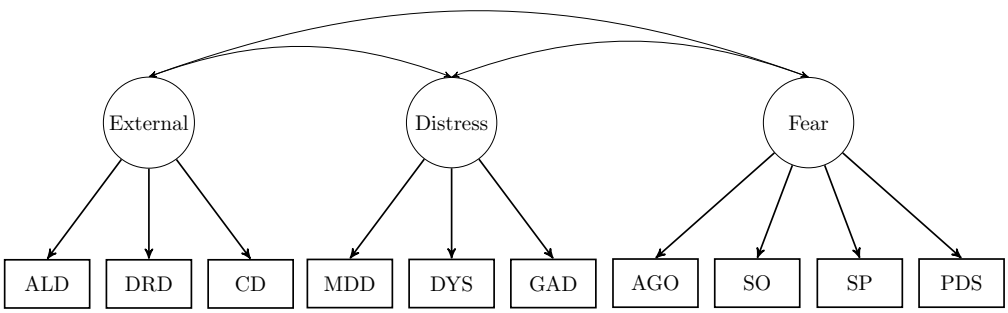


Figure 1. Path diagram of the three-factor latent structure of psychopathology. *Note.* External, externalizing factor (tendencies toward poor behavioral and emotional control); Distress, distress factor (tendencies toward anxious misery); Fear, fear factor (tendencies towards phobic responses to stressors); ALD, Alcohol Dependence; DRD, Drug Dependence; CD, Conduct Disorder; MDD, Major Depressive Disorder; DYS, Dysthemic Disorder (or Persistent Depressive Disorder; GAD, Generalized Anxiety Disorder; AGO, Agoraphobia; SO, Social Anxiety Disorder; SP, Specific Phobia; and PDS, Panic Disorder.

3.3 Analysis strategy

First, "population values" for all parameters of the item factor model were obtained by fitting the model to the full dataset using lavaan (Rosseel, 2012) with diagonally weighted least squares (DWLS). Next, 50 random samples of size (N = 200) were drawn from the population dataset. With each sampled dataset, the three-factor model was estimated using (1) DWLS using the THETA parameterization, (2) DWLS using the DELTA parameterization, (3) Bayesian estimation under the THETA parameterization, and (4) Bayesian estimation under the pseudo-DELTA parameterization. Once estimated, point estimates for the standardized thresholds and factor loadings were extracted. For Bayesian models, the posterior median was used as the point estimate. The impact of latent response distribution parameterization is illustrated by examining the distribution of point estimates across parameterizations. All code and data to replicate analyses are available in our online repository (https://doi.org/10.17605/osf.io/398Q4).

3.4 Priors

When the models were estimated using Bayesian methods, the priors were selected based on blavaan (Merkle & Rosseel, 2016), Mplus (Asparouhov & Muthen, 2021), and Padgett et al. (2024). All priors were the same regardless of the latent response distribution parameterization. First, a fixed loading reference-variable approach was used to identify the scale of the latent factors. The prior for factor loadings thus depended on the variable. The variables ALD, MDD, and AGO had a fixed loading of 1.0 for identification. All other factor loadings had the relatively diffuse normal distribution prior of $\lambda \sim \text{Normal}(0.0, 10.0)$. The prior for the latent variable covariance matrix is decomposed into a Lewandowski-Kurowicka-Joe (LKJ) prior (Lewandowski et al., 2009) for the factor correlations and a factor standard deviations component with prior of $\phi_{sd} \sim \text{Gamma}(1.0, 0.5)$. We used the Dirichlet induced prior for the threshold prior using a relatively diffuse specification, $\tau \sim$ Induced-Dirichlet$(0.5, 0.5)$, resulting in a prior equivalent sample size of approximately 1 case. The latent response distribution residual variances were fixed to 1.0, and the latent response distribution intercepts were fixed to 0.0.

Initially, all models were specified using a fixed loading on each factor to identify the model and latent factor variance. Upon review of the results, the DWLS estimates for the standardized factor loadings were not infrequently greater than 1.0. To aid in comparability with the results obtained using Bayesian methods, we re-fit the DWLS estimated models using the fixed factor variance method and bounds for nonnegative variances to avoid standardized loadings greater than 1.0. We extracted one replication where this issue occurred for reproducing the issue using lavaan in our companion repository on the OSF (see the 'Delta parameterization bounds' folder and accompanying data and R script).

4. Results

In the full NCS-R dataset, the proportion of patients with each diagnosis are (from most common to to least common): Major Depressive Disorder 17%, Specific Phobia 12.9%, Social Anxiety Disorder 12.3%, Generalized Anxiety Disorder 8.1%, Panic Disorder 4.9%, Alcohol Dependence 4.8%, Conduct Disorder 4.4%, Dysthemic Disorder (or Persistent Depressive Disorder) 4.2%, Drug Dependence 2.7%, and Agoraphobia 2.5%. The resulting estimates for the standardized factor loadings estimates under each parameterization are summarized in Table 1, and the thresholds are summarized in Table 2.

The rows of Table 1 are arranged by magnitude of the population value to illustrate how differences in factor loadings may influence the estimated value. Within the DELTA parameterization, two major points stick out to us. First, is that estimation with DWLS can lead to standardized factor loadings equal to exactly 1.0. Bayesian approach, though some estimates approach quite close to 1.0. Secondly, for moderate standardized factor loadings (those below 0.8), there does not appear to be a

Table 1. Variation in standardized factor loadings in simulation under alternative latent response parameterizations and estimation approaches.

		DWLS				Bayesian			
Variable	Population value	Avg.	MSE	Q[0.10,0.90]	%≥ .99	Avg.	MSE	Q[0.10,0.90]	%≥ .99
DELTA parameterization									
DRD	0.90	0.89	0.02	[0.74,1.00]	30%	0.93	0.01	[0.77, 1.00]	30%
ALD	0.88	0.83	0.02	[0.65,0.97]	10%	0.73	0.04	[0.50, 0.91]	0%
DYS	0.86	0.88	0.02	[0.67,1.00]	38%	0.90	0.02	[0.72, 0.99]	18%
AGO	0.85	0.83	0.05	[0.70,1.00]	12%	0.62	0.09	[0.36,0.85]	0%
GAD	0.76	0.75	0.02	[0.60,0.91]	4%	0.80	0.02	[0.62,0.96]	0%
SO	0.75	0.70	0.06	[0.52,0.87]	0%	0.75	0.03	[0.57, 0.93]	8%
PDS	0.72	0.71	0.04	[0.49,0.89]	0%	0.69	0.04	[0.45, 0.97]	0%
SP	0.70	0.69	0.04	[0.57,0.85]	0%	0.73	0.02	[0.53, 0.92]	2%
MDDH	0.69	0.67	0.02	[0.49,0.82]	4%	0.68	0.01	[0.52, 0.81]	0%
CD	0.65	0.66	0.04	[0.39,0.93]	4%	0.68	0.05	[0.35, 0.93]	0%
THETA parameterization									
DRD	0.90	0.88	0.01	[0.74,0.95]	0%	0.81	0.02	[0.66,0.89]	0%
ALD	0.88	0.83	0.02	[0.65,0.95]	0%	0.63	0.10	[0.38,0.81]	0%
DYS	0.86	0.87	0.01	[0.68,0.95]	0%	0.81	0.01	[0.68,0.90]	0%
AGO	0.85	0.82	0.05	[0.70,0.95]	0%	0.51	0.15	[0.26,0.77]	0%
GAD	0.76	0.75	0.01	[0.61,0.91]	0%	0.78	0.01	[0.67,0.86]	0%
SO	0.75	0.70	0.06	[0.52,0.87]	0%	0.74	0.01	[0.62,0.88]	0%
PDS	0.72	0.71	0.04	[0.49,0.90]	0%	0.66	0.03	[0.43,0.84]	0%
SP	0.70	0.69	0.04	[0.57,0.85]	0%	0.73	0.01	[0.56,0.83]	0%
MDDH	0.69	0.67	0.02	[0.49,0.84]	0%	0.63	0.02	[0.48,0.79]	0%
CD	0.65	0.67	0.04	[0.40,0.95]	0%	0.65	0.03	[0.40,0.84]	0%

Note. Population values obtained from estimating the factor model in lavaan using DWLS using the full sample size (N=9,282); Avg., average point estimate across replications; MSE, mean square error of the point estimate across replications; Q[0.10,0.90], 10th and 90th-quantile of the distribution of point estimates; %≥ .99, percent of estimates greater than or equal to 0.99; ALD, Alcohol Dependence; DRD, Drug Dependence; CD, Conduct Disorder; MDD, Major Depressive Disorder; DYS, Dysthemic Disorder (or Persistent Depressive Disorder; GAD, Generalized Anxiety Disorder; AGO, Agoraphobia; SO, Social Anxiety Disorder; SP, Specific Phobia; and PDS, Panic Disorder.

meaningful difference in the loading recovery even in relatively small sample sizes of 200. However, this is not the case for thresholds. The threshold parameters were not recovered on average when using the THETA parameterization with Bayesian estimation, which is a known issue when the scale is highly skewed (Bainter, 2017).

When the models were estimated using the THETA parameterization, we see that the higher standardized factor loadings were more likely to be underestimated. For example, the loading for ALD in the population is 0.88, but estimates as 0.83 on average using DWLS and on 0.63 on average using a Bayesian approach. When the population loading was lower than 0.80, both estimation approaches were closer to the population value. This is a potentially concerning result as under estimation of standardized factor loadings could lead to mistakenly throwing out variables as indicators of factors due to poor loadings. On the other side though, for the estimates under the DELTA parameters, we found that that the standard loading point estimates could be nearly or exactly 1.0 across replications a fair number of times. For example, the DYS standardized factor loading

Table 2. Variation in threshold estimates in simulation under alternative latent response parameterizations and estimation approaches.

		DWLS			Bayesian		
Variable	Population value	Mean	MSE	Q[0.10,0.90]	Mean	MSE	Q[0.10,0.90]
DELTA parameterization							
AGO	1.96	1.98	0.03	[1.81,2.17]	1.98	0.03	[1.79,2.19]
DRD	1.93	2.01	0.04	[1.81,2.33]	1.98	0.04	[1.77,2.25]
DYS	1.73	1.75	0.03	[1.51,1.96]	1.72	0.03	[1.50,1.92]
CD	1.71	1.78	0.05	[1.55,2.05]	1.75	0.04	[1.53,2.02]
ALD	1.67	1.71	0.04	[1.51,1.96]	1.69	0.03	[1.51,1.93]
PDS	1.65	1.66	0.03	[1.48,1.96]	1.63	0.03	[1.45,1.92]
GAD	1.40	1.40	0.01	[1.28,1.55]	1.39	0.01	[1.27,1.54]
SO	1.16	1.17	0.01	[1.06,1.28]	1.16	0.01	[1.02,1.26]
SP	1.13	1.14	0.01	[1.01,1.28]	1.12	0.01	[0.99,1.27]
MDDH	0.95	0.96	0.01	[0.86,1.06]	0.95	0.01	[0.85,1.06]
THETA parameterization							
AGO	1.96	1.98	0.03	[1.81,2.17]	1.62	0.25	[1.16,2.08]
DRD	1.93	2.01	0.04	[1.81,2.33]	1.20	0.64	[0.86,1.65]
DYS	1.73	1.75	0.03	[1.51,1.96]	0.97	0.67	[0.71,1.39]
CD	1.71	1.78	0.05	[1.55,2.05]	1.29	0.26	[0.99,1.71]
ALD	1.67	1.71	0.04	[1.51,1.96]	1.28	0.25	[0.90,1.70]
PDS	1.65	1.66	0.03	[1.48,1.96]	1.18	0.33	[0.83,1.58]
GAD	1.40	1.40	0.01	[1.28,1.55]	0.82	0.37	[0.62,1.07]
SO	1.16	1.17	0.01	[1.06,1.28]	0.73	0.20	[0.54,0.91]
SP	1.13	1.14	0.01	[1.01,1.28]	0.75	0.17	[0.57,0.92]
MDDH	0.95	0.96	0.01	[0.86,1.06]	0.68	0.10	[0.49,0.85]

Note. Population values obtained from estimating the factor model in lavaan using DWLS using the full sample size (N=9,282); Mean, average point estimate across replications; MSE, mean square error of the point estimate across replications; Q[0.10,0.90], 10th and 90th-quantile of the distribution of point estimates; ALD, Alcohol Dependence; DRD, Drug Dependence; CD, Conduct Disorder; MDD, Major Depressive Disorder; DYS, Dysthemic Disorder (or Persistent Depressive Disorder; GAD, Generalized Anxiety Disorder; AGO, Agoraphobia; SO, Social Anxiety Disorder; SP, Specific Phobia; and PDS, Panic Disorder.

under the DWLS estimation approach resulted 38% of replication with an estimate 1.0 which was fixed during estimation to avoid a negative residual variance. While using a Bayesian approach never lead to estimates of exactly 1.0, the posterior median was above 0.99 in 18% of replications for the DYS item. The impact of the latent response distribution parameterization on threshold estimation was even more severe, as summarized in Table 2. For these standardized loadings, we found that the DELTA parameterization recovered the population values well on average using DWLS or the Bayesian pseudo-delta method. Using the THETA parameterization, the estimated thresholds being recovered depended on using DWLS and could not be recovered well using Bayesian methods.

To dig even deeper in the concerns about estimating thresholds and factor loadings in item factor analysis, we focus next on the Dysthemic Disorder variable due the extreme variability we observed across replications and the finding that this variable has a high loading. The resulting distributions of point estimates comparing across parameterizations and estimation methods are shown in Figure 2. We see a potentially troubling issue with the THETA parameterization that only

one of the loading point estimates got above 0.90, but that point has other issue as well; namely, that the standardized threshold is essentially 0.0. This occurred because the latent response distribution variance exploded to a very large number resulting the standardized estimate going to practically zero. This convergence to zero as the loading increased for the threshold parameters is a concerning finding for the THETA parameterization using Bayesian methods. This might point to a limitation of the traditional THETA parameterization in Bayesian item factor analysis. However, of note is how the "pseudo-DELTA" parameterization aligned quite well with the DWLS DELTA parameterization across the full range of factor loadings, providing evidence that the intermediary step of standardizing the latent response distribution can overcome the limitations of the THETA parameterization.

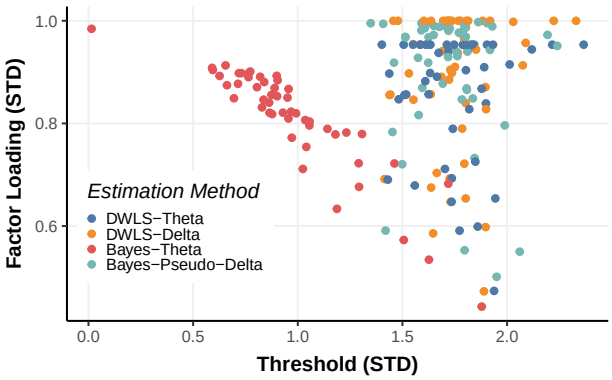
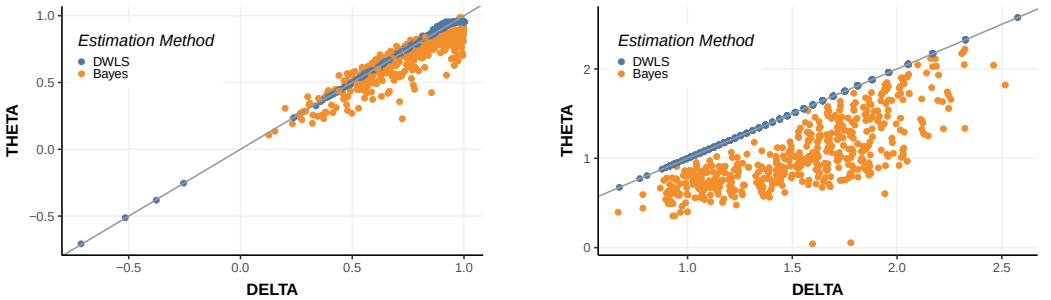


Figure 2. Dysthemic Disorder (or Persistent Depressive Disorder) diagnosis thresholds are severely impacted by parameterization. *Note.* Each point represents the DWLS point estimate or posterior median with a sample size of N=200, and all points were horizontally jittered slightly (±0.10) to aid to seeing the distinct points due to overlap between DWLS point estimates. All reported estimates are based on the standardized solution.

The differences between the parameterizations for factor loadings are shown in Figure 3a and for thresholds in Figure 3b. The loadings are relatively consistent between parameterizations, though under Bayesian estimation, the DELTA parameters were more likely to be greater (estimates below the diagonal line. This is very clearly seen in the threshold parameter estimates where the nearly all THETA parameterization posterior medians were lower than when using the proposed pseudo-DELTA parameterization for Bayesian estimation.

5. Conclusion

We aimed to investigate whether a pseudo-DELTA parameterization of the item factor analysis model resulted in improved parameter recovery in the presence of a few highly skewed indicators compared to existing frequentist and Bayesian approaches. The results from our small simulation study based on a large psychopathology sample indicate that the DELTA parameterization may have advantages over alternative parameterizations. Specifically, when binary factor indicators exhibit skewed marginal distributions, the DELTA parameterization provides more reliable threshold parameter estimates than the commonly implemented THETA parameterization for Bayesian item factor analysis. In addition, the Bayesian DELTA parameterization less frequently resulted in point estimates on the boundary of the parameter space of standardized loadings (i.e., estimates near 1.0). Future research will need to further test whether the proposed simple modification—transforming the latent response covariance matrix into a correlation matrix—yields well-calibrated credible intervals without requiring complex or highly informative priors. The results here are promising, though. For applied researchers working with sparse response patterns (e.g., rare disorders, uncommon behaviors), adopting the



(a) Standardized Factor Loadings

(b) Standardized Thresholds

Figure 3. Comparison of point estimates between parameterization for factor loadings and thresholds for all items. *Note.* Each point represents the DWLS point estimate or posterior median with a sample size of N=200. All reported estimates are based on the standardized solution.

"pseudo-DELTA" parameterization may substantially improve the interpretability and trustworthiness of threshold and factor loading point estimates in Bayesian item factor analysis.

Although the results from the current study are promising, future studies should conduct a more extensive simulation study comparing existing approaches to the Bayesian pseudo-DELTA parameterization proposed here. Such a study should examine how characteristics of the data (e.g., sample size, number of response options, and severity and pervasiveness of skewness in item responses) affect the performance of the pseudo-DELTA parameterization. In addition, the current study looked at one relatively diffuse prior specification for the threshold parameters. Future research should examine alternative prior specifications that are more informative or that are not aligned with the data to provide further guidance to applied researchers interested in using this method.

Acknowledgement

Funding Statement This research was not supported by grants.

Competing Interests None.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders: DSM-IV* (4th ed.). Autor.

Asparouhov, T., & Muthen, B. (2021). Bayesian analysis of latent variable models using mplus [Retrieved on [2023-12-18] from https://www.statmodel.com/download/BayesAdvantages18.pdf].

Bainter, S. A. (2017). Bayesian estimation for item factor analysis models with sparse categorical indicators. *Multivariate behavioral research*, *52*(5), 593–615. https://doi.org/10.1080/00273171.2017.1342203

Constantinou, M. P., Goodyer, I. M., Eisler, I., Butler, S., Kraam, A., Scott, S., Pilling, S., Simes, E., Ellison, R., Allison, E., et al. (2019). Changes in general and specific psychopathology factors over a psychosocial intervention. *Journal of the American Academy of Child & Adolescent Psychiatry*, *58*(8), 776–786.

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A monte carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 625–641. https://doi.org/10.1080/10705510903203573

Kamata, A., & Bauer, D. J. (2008). A note on the relation between factor analytic and item response theory models. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(1), 136–153. https://doi.org/10.1080/10705510701758406

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month dsm-iv disorders in the national comorbidity survey replication. *Archives of general psychiatry*, *62*(6), 617–627.

Kessler, R. C., & Merikangas, K. R. (2004). The national comorbidity survey replication (ncs-r): Background and aims. *International journal of methods in psychiatric research*, *13*(2), 60–68.

Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M., & Wittchen, H.-U. (2012). Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the united states. *International journal of methods in psychiatric research*, *21*(3), 169–184.

Kessler, R. C., & Üstün, T. B. (2004). The world mental health (wmh) survey initiative version of the world health organization (who) composite international diagnostic interview (cidi). *International journal of methods in psychiatric research*, *13*(2), 93–121.

Krueger, R. F. (1999). The structure of common mental disorders. *Archives of general psychiatry*, *56*(10), 921–926.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008

Lynch, S. J., Sunderland, M., Newton, N. C., & Chapman, C. (2021). A systematic review of transdiagnostic risk and protective factors for general and specific psychopathology in young people. *Clinical psychology review*, *87*, 102036.

McGrath, J. J., Lim, C., Plana-Ripoll, O., Holtz, Y., Agerbo, E., Momen, N., Mortensen, P., Pedersen, C., Abdulmalik, J., Aguilar-Gaxiola, S., et al. (2020). Comorbidity within mental disorders: A comprehensive analysis based on 145 990 survey respondents from 27 countries. *Epidemiology and Psychiatric Sciences*, *29*, e153.

Merkle, E., Ariyo, O., Winter, S. D., & Garnier-Villarreal, M. (2023). Opaque prior distributions in bayesian latent variable models. *Methodology*, *1*(3), 228–255. https://doi.org/10.5964/meth.11167

Merkle, E., Fitzsimmons, E., Uanhoro, J., & Goodrich, B. (2021). Efficient bayesian structural equation modeling in stan. *Journal of Statistical Software*, *100*, 1–22. https://doi.org/10.18637/jss.v100.i06

Merkle, E., & Rosseel, Y. (2016). Blavaan: Bayesian latent variable analysis. *R Package Version 0. 1-3. Available online at: https://cran. r-project. org/package= blavaan*.

Moshagen, M., & Musch, J. (2014). Sample size requirements of the robust weighted least squares estimator. *Methodology*, *10*(2), 60–70. https://doi.org/10.1027/1614-2241/a000068

Padgett, R. N., Morgan, G. B., & Lomas, T. (2024). Regularizing threshold priors with sparse response patterns in bayesian factor analysis with categorical indicators [Advanced online publication]. *Psychological Methods*. https://doi.org/10.1037/met0000682

Roest, A. M., de Vries, Y. A., Lim, C. C., Wittchen, H.-U., Stein, D. J., Adamowski, T., Al-Hamzawi, A., Bromet, E. J., Viana, M. C., de Girolamo, G., et al. (2019). A comparison of dsm-5 and dsm-iv agoraphobia in the world mental health surveys. *Depression and anxiety*, *36*(6), 499–510.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. http://www.jstatsoft.org/v48/i02/

Ruscio, A. M., Brown, T. A., Chiu, W. T., Sareen, J., Stein, M. B., & Kessler, R. C. (2008). Social fears and social phobia in the usa: Results from the national comorbidity survey replication. *Psychological medicine*, *38*(1), 15–28.

Watts, A. L., Lane, S. P., Bonifay, W., Steinley, D., & Meyer, F. A. (2020). Building theories on top of, and not independent of, statistical models: The case of the p-factor. *Psychological inquiry*, *31*(4), 310–320.

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*(1), 58–79. https://doi.org/10.1037/1082-989X.12.1.58

Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, *81*(4), 1014–1045. https://doi.org/10.1007/s11336-016-9506-0

Appendix 1. Stan implementation of pseudo-Delta method

The Bayesian item factor analysis model was implemented in Stan using the data-augmentation approach described in section 2. The general structure of the program was previously used in Padgett et al. (2024), and showed good coverage and parameter recovery under ideal conditions without sparse response categories. The model has two major components, a transform parameters block and model block where the priors and data likelihood are sampled. Another important section is the functions block where we define the transformation from covariance to correlation.

```
function {
  matrix cov2cor(matrix u) {
    matrix[rows(u), cols(u)] w;
    for (k in 1:rows(u)){
      for(j in 1:cols(u)){
        w[k,j] = u[k,j]/( (u[k,k]^0.5)*(u[j,j]^0.5) );
      }
    }
    return w;
  }
}
```

```
... //other Stan code for data and other functions
// D = number of items
// N = number of cases/persons
// m = number of latent factors
transformed parameters {
  vector[D] ystar[N];
  matrix[D,m] Lambda; // matrix for factor loadings
  matrix[m,m] Psi; // full factor covariance matrix
  matrix[m, m] L_Psi_full; // Cholesky factor of factor covariances
  matrix[D,D] Sigma_theta; //covariances of latent response variables
  matrix[D,D] Sigma_delta; //correlations of latent response variables
  real ystar_jacobian;
  Lambda = fill_loading_matrix(lambda_free, pos_matrix);
  L_Psi_full = diag_pre_multiply(L_Psi_sd, L_Psi);
  Psi = multiply_lower_tri_self_transpose(L_Psi_full);
  Sigma_theta = diag_matrix(rep_vector(1, D)) + quad_form_sym(Psi,
      transpose(Lambda));
  //*** HERE IS THE NEW STEP ***
  //Convert from covariance (theta parameterization) to correlation (
      delta parameterization):
  Sigma_delta = cov2cor(Sigma_theta);
  //This has no effect on the likelihood, but has a significant impact
       on sampling efficiency.
  //***
  // Implement the data-augmentation + jacobian adjustment
  ystar_jacobian = 0;
  for(n in 1:N){
    for(d in 1:D){
      real lb;   real ub;
      if(y[n, d]==1){
        lb = -10;
        ub = tau[d,1];
      } else if(y[n, d] == nlevs[d]){
        lb = tau[d,nlevs[d]-1];
        ub = 10;
      } else {
        lb = tau[d,y[n,d]-1];
        ub = tau[d,y[n,d]];
      }
      ystar[n,d] = lb + (ub - lb)*u[n,d];
      ystar_jacobian += log(ub-lb);
    }
}}
model {
  for(n in 1:N){
    multi_normal_lpdf(ystar[n,1:D] | rep_vector(0,D), Sigma_delta);
  }
  target += ystar_jacobian;
  ...//priors
}
```