On the benefit of robust Bayesian confirmatory factor analysis

Radislav Vaisman[a] Mitchell Scovell[b], Nikolai Kinaev[c], and Javier Fernandez[d]

[a]School of Mathematics and Physics, The University of Queensland, Brisbane 4072, Australia [b]Environment, CSIRO, Brisbane QLD 4102, Australia [c]Energy, CSIRO, Brisbane QLD 4069, Australia [d]School of Agriculture and Food Sustainability, The University of Queensland, Brisbane 4072, Australia

ARTICLE HISTORY
Compiled November 12, 2024

ABSTRACT
Data normality is a typical assumption when performing confirmatory factor analysis. However, always using confirmatory factor analysis under the normally assumption will often lead to poor model fit in real-world applications. In this work, we construct an efficient sampler for a robust Bayesian confirmation factor analysis by utilizing the Student's t distribution and provide a novel way to calculate the corresponding marginal likelihood. We demonstrate that the proposed strategy is both computationally efficient and performs favorably when compared to standard approaches that rely on the normality assumption. Finally, we provide a research toolkit for robust Bayesian confirmatory factor analysis and model comparison.

KEYWORDS
Robust Bayesian confirmatory factor analysis; data normality; Markov chain monte carlo; Bayes factors; importance sampling

1. Introduction

Confirmatory factor analysis (CFA) is a commonly used technique to test a theoretically informed explanation for the association between observed variables. Joreskog (1969) introduced a general approach for performing CFA inference using maximum likelihood. Numerous scientific domains, including social science, psychology, medicine, health science, engineering, and many more, utilize CFA (Depaoli, 2021; Lawal et al., 2020; Lesia, Aigbavboa, & Thwala, 2024; Shao, Elahi Shirvan, & Alamer, 2022; Sureshchandar, 2023).The main value of CFA is that it enables the empirical testing of theories in a way that aligns with the hypothetico-deductive approach of scientific inquiry (Haig, 2005, 2018). Confirmatory factor analysis, in other words, tests whether observed data conflict with a theory.

Despite the importance of CFA, its application can be limited due to its dependence on the assumption of data normality. This is assumption is problematic as real-world data do not always come from normal distributions. Consequently, if CFA is used blindly, always using the assumption of normality, the analysis may be compromised and incorrect results and conclusions may be drawn. This work addresses the normality problem by presenting a robust Bayesian CFA analysis method that can be used for

CONTACT Radislav Vaisman. Email: r.vaisman@uq.edu.au

analysis, with the assumption of normally distributed data. In order to accomplish this, we renounce the normalcy assumption and instead, introduce an approach using the more robust Student's t distribution (Tong & Zhang, 2020). For example, extending the use of confirmatory factor analysis to account for non-normality can significantly enhance their applicability in most areas of agriculture research. In life and biological sciences, commonly observed genomics or phenomics data often deviates from the normal distribution. Examples include measurements of plant phenotypic traits such as leaf nitrogen concentration or time to flowering stages (Stroup, 2015). By allowing a Student's t distribution error assumption into CFA, researchers can better handle and understand these non-normally distributed data, providing a more robust and adequate analysis. This approach allows for more reliable validation of theoretical concepts established for complex variables for this example, say water or nutrient use efficiencies from observable data, ultimately improving the understanding of the underpinning biological processes.

Understanding the connection between data variables, both observable and latent, is the primary goal of the CFA. For instance, one may wish to know how well the observed response to each mental health survey item represents a symptom of anxiety condition (here, the anxiety level is assumed to be unobserved). In other words, CFA quantifies the indirect link between latent and observable variables, where the observable data is measured, for example, through survey questions. A crucial component that needs to be considered is a *factor loading matrix*, which is a set of numerical quantities that measure the degree to which observable variables and latent variables correspond. As soon as factor loadings are available (through a statistical inference procedure), the researcher can determine which factor loadings exhibit connections that are statistically significant. For instance, we may establish that factor loadings higher than a given threshold, for example 0.8, are appropriate for evaluating a particular latent factor. Moreover, if the entire survey, or, a majority of survey questions have factor loadings that are higher than this threshold, we can conclude that the observed data are consistent with the theorised latent factors. For an extensive CFA overview, we refer to Depaoli (2021); Roos & Bauldry (2022) and Harrington (2009).

Under the CFA setting, one can perform inference using maximum likelihood (ML), generalized least squares (GLS), or Bayesian approach (Browne, 1973, 1984; Depaoli, 2021). Both the GLS and the ML methods, in general, operate under the presumption that the observed data is derived from multivariate normal distributions. In other words, the data is assumed to be normally distributed. Indeed, these estimators exhibit good qualities under the normality assumption. For instance, it is possible to prove that the ML estimator is asymptotically identical to the "best" GLS estimator (Lee, 2007). Because the normality assumption is restrictive, asymptotically distribution-free methods can be used when the observed data is still assumed to be independent and identically distributed, but, the multivariate normality assumption is discarded (Browne, 1984).

In this work, we focus on the Bayesian inference task, which offers a number of benefits over traditional estimation techniques, particularly when compared to GLS and ML methods (Asparouhov, Muthén, & Morin, 2015). The benefit of the Bayesian approach applied to latent variables models, including the CFA, are highlighted by Lee (2007). First, the Bayesian setup makes it possible to include important prior information, offer the posterior distribution's full characterization of the relevant parameters, and produce superior outcomes with smaller data sizes. The application of the Bayesian approach began to gain traction in social sciences. For instance, Jackman (2009) states that "Bayesian inference is straightforward and direct" when the

posterior probability statement in Bayesian analysis is used. To be more precise, the Bayesian paradigm takes into account the characterization of existing knowledge following the data observation. Lastly, compared to the classical (frequentist) inference techniques, the posterior distribution offers a much more natural way to examine essential inference summaries such as confidence intervals through means, standard deviations, and other helpful statistics.

Assuming that the research data is not necessarily normally distributed (a common scenario in survey data (Cain, Zhang, & Yuan, 2017)), and may or may not contain outliers, the main focus of this study is the introduction of a robust Bayesian estimation procedure of the corresponding CFA model including model comparison via Bayes factors (Gelman & Rubin, 1992). No prior checks for normality or the presence of outliers are required when using the suggested model. To address the above challenge, we suggest to replace the normal (Gaussian) error term in the CFA model with the error term that follows the Student's t distribution. We show that the proposed Bayesian inference procedure can handle the error term distribution substitution rather directly, and that the new CFA model will be less influenced by outliers (see for example (Tong & Zhang, 2020)). The derivation of an efficient sampling scheme from the relevant posterior distribution presents the main challenge. In this work, we provide an effective Gibbs (Gelman, Carlin, Stern, & Rubin, 2003) sampler approach that can handle the CFA model with the Student's t distribution error term. This process allows to obtain accurate estimators of key parameters, such as the factor loading coefficients and the accompanied confidence intervals. The study makes the following specific contributions.

(1) First, we provide an efficient Gibbs sampler construction method for the Student's t distribution error term setup. This allows to perform robust Bayesian inference under the CFA setting.
(2) Our second contribution is the development of an effective way to calculate the corresponding marginal likelihood. Numerical evaluation shows that the proposed importance sampling estimator (Rubinstein, Ridder, & Vaisman, 2014) has a small relative error (Rubino & Tuffin, 2009) and thus, a relatively modest number of samples is sufficient for providing reliable estimates of the marginal likelihood. Thus, our estimator opens the way for fast and reliable model comparison using Bayes factors.
(3) Finally, we offer a research `R` package that computes marginal likelihood estimators and implements both the normal CFA (nCFA) and the Student's t CFA (tCFA) models. Since the package was written in `C++`, the software is computationally efficient and can handle data sets of a reasonable size. We believe that practitioners can benefit from this package.

To conclude, we propose a robust approach for Bayesian CFA analysis. Our experimental research indicates that the tCFA model can offer a considerable advantage over the traditional nCFA model; for this reason, we believe that tCFA should be taken into account in practical situations.

The structure of this document is as follows. We explicitly define the CFA problem and provide a brief description of the appropriate conjugate Bayesian analysis for the nCFA model in Section 2. In Section 3, we demonstrate how the Markov Chain Monte Carlo (MCMC) sampler for the nCFA model, can be extended to handle the Student's t distribution setting, hence eliminating the need for the data normality assumption. Section 4 shows how one can construct a simple importance sampling estimator for marginal likelihood calculation. An experimental investigation that shows how well the

suggested methods work with both synthetic and real-world data examples is presented in Section 5. Finally, we conclude with Section 6, which provides a summary of our findings and potential avenues for further research.

2. The nCFA — Bayesian CFA under the data normality assumption

We consider the CFA model

$$\mathbf{y}_i = \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i, \tag{1}$$

where $i = 1, \ldots, n$ is the number of observations, $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ is the whole observable data, where $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,p})^\top$. Similarly, $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_n)$ is the latent data, where $\boldsymbol{\omega}_i = (\omega_{i,1}, \ldots, \omega_{i,q})^\top$, and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n)$ is the measurement error, where $\boldsymbol{\epsilon}_i = (\epsilon_{i,1}, \ldots, \epsilon_{i,p})^\top$. In this manuscript, we denote the factor loading matrix by $\boldsymbol{\Lambda} = (\lambda_{k,j})_{1\leq k\leq p,\, 1\leq j\leq q}$. We further assume that for all $1 \leq i \leq n$, $\boldsymbol{\omega}_i$ and $\boldsymbol{\epsilon}_i$ follow a multivariate normal distribution. Specifically, we set $\boldsymbol{\omega}_i \sim \mathsf{MVN}(0, \boldsymbol{\Phi})$ and $\boldsymbol{\epsilon}_i \sim \mathsf{MVN}(0, \boldsymbol{\Psi}_\epsilon)$, where $\boldsymbol{\Phi}$ is a $q \times q$ general covariance matrix, and $\boldsymbol{\Psi}_\epsilon = \mathrm{diag}(\psi_1, \ldots, \psi_p)$ is a diagonal covariance matrix. The probability distributions utilized in this manuscript are described in Section 1 of the Online Supplementary Materials record. Since $\boldsymbol{\epsilon}_i$ has the multivariate Gaussian distribution, we proclaim this model to pursue the normality assumption.

Following (Depaoli, 2021; Lee, 2007; Lee & Song, 2012), consider the collection of conjugate prior distributions for the set of parameters of interest, namely, for $(\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\omega})$. Specifically, let

$$\begin{aligned} \lambda_{k,j} &\sim \mathsf{N}(0, \sigma^2_{k,j}) \text{ for } k = 1, \ldots, p \text{ and } j = 1, \ldots, q, \\ \boldsymbol{\Phi} &\sim \mathsf{IW}(A, a), \\ \psi_k &\sim \mathsf{IG}(\alpha_k, \beta_k) \text{ for } k = 1, \ldots, p, \end{aligned} \tag{2}$$

and, to ensure identifiability, we fix one of the factor loadings (in each column of the factor loading matrix), to a constant value of 1. The sampling from the posterior distribution

$$p(\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\omega}|\mathbf{y}) = \prod_{i=1}^{n} p(\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\omega}_i|\mathbf{y}_i), \tag{3}$$

can be performed efficiently via Gibbs sampler (Gilks, Richardson, & Spiegelhalter, 1996). Specifically, from (2) and (3), it can be shown (see for example (Lee, 2007), or, consider a more general case in Section 3 of this manuscript), that:

$$\lambda_{k,j}|\mathbf{y}, \boldsymbol{\Lambda}_{-\lambda_{k,j}}, \boldsymbol{\Phi}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\omega} \sim \mathsf{N}\left(\sigma^2\left(\sum_{i=1}^{n} \frac{\omega_{i,j}(y_{i,k} - \sum_{l=1}^{q} \lambda_{kl}\omega_{il}\mathbb{1}_{\{l\neq j\}})}{\psi_k}\right), \sigma^2\right), \tag{4}$$

were $\boldsymbol{\Lambda}_{-\lambda_{k,j}}$ is the set of all elements of $\boldsymbol{\Lambda}$ except of $\lambda_{k,j}$ and

$$\sigma^2 = \left(\frac{1}{\sigma^2_{k,j}} + \sum_{i=1}^{n} \frac{\omega^2_{i,j}}{\psi_k}\right)^{-1} \text{ for } k = 1, \ldots, p \text{ and } j = 1, \ldots, q,$$

$$\boldsymbol{\Phi}|\mathbf{y}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\omega} \sim \mathsf{IW}\left(A + \boldsymbol{\omega}^\top\boldsymbol{\omega}, a + n\right),$$

$$\psi_k|\mathbf{y}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\omega} \sim \mathsf{IG}\left(0.5n + \alpha_k, \beta_k + \frac{1}{2}\sum_{i=1}^{n}\left(y_{i,k} - \boldsymbol{\Lambda}_k\boldsymbol{\omega}_i\right)^2\right) \text{ for } k = 1, \ldots, p,$$

were $\boldsymbol{\Lambda}_k$ is the kth row of $\boldsymbol{\Lambda}$,

$$\boldsymbol{\omega}_i|\mathbf{y}, \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}_\epsilon \sim \mathsf{MVN}\left(\Sigma'(\boldsymbol{\Lambda}^\top\boldsymbol{\Psi}_\epsilon^{-1}\mathbf{y}_i), \Sigma'\right),$$

were $\Sigma' = \left(\boldsymbol{\Phi}^{-1} + \boldsymbol{\Lambda}^\top\boldsymbol{\Psi}_\epsilon^{-1}\boldsymbol{\Lambda}\right)^{-1}$ for $1 \leq i \leq n$.

That is, we conclude that the set of conditional distributions (4) can be readily used for efficient sampling from the posterior distribution in (3).

3. The tCFA — robust Bayesian CFA

Consider the CFA model from (1), but, in this section, we assume that measurement errors follow the Student's t distribution. It is well known that the usage of t distribution allows for greater robustness; see for example (Lange, Little, & Taylor, 1989). Formally, for the robust model (tCFA), instead of $\epsilon_{i,k} \sim \mathsf{N}(0, \psi_k)$ for $1 \leq i \leq n$ and $1 \leq k \leq p$, define $\epsilon_{i,k} \sim \mathsf{t}(\nu_k, 0, \psi_k)$. In this case, the set of parameters of interest is extended to be $(\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\omega}, \boldsymbol{\nu})$, where $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)^\top$ is the vector of degrees of freedom that correspond to the Student's t distribution. Note that $\mathbf{y}_i|\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\omega}_i, \boldsymbol{\nu}$ is not Gaussian any more. Instead, we are dealing with Student's t distribution with location $\boldsymbol{\Lambda}\boldsymbol{\omega}_i$, scale of $\boldsymbol{\Psi}_\epsilon$ and $\boldsymbol{\nu}$ degree of freedom. Specifically, it holds that:

$$y_{i,k}|\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\omega}_i, \boldsymbol{\nu} \sim \mathsf{t}\left(\boldsymbol{\Lambda}_k\boldsymbol{\omega}_i, \psi_k, \nu_k\right), \quad \forall\, 1 \leq i \leq n,\ 1 \leq k \leq p. \tag{5}$$

Geweke (1993), proposed to introduce latent variables with the view to handle Student's t distribution measurement errors. In this paper, we mimic this strategy to perform inference under the robust CFA setting. Specifically, instead of handling the Student's t distribution directly, Geweke (1993) uses the data augmentation approach and exploits the fact that for a latent variable $z \sim \mathsf{IG}(\nu/2, \nu/2)$, and for $X|z \sim \mathsf{N}(\mu, z\sigma^2)$, it holds that $X \sim \mathsf{t}(\mu, \sigma^2, \nu)$. That is, one can rewrite (5) via

$$y_{i,k}|\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\omega}_i, \boldsymbol{\nu}, z_{i,k} \sim \mathsf{N}\left(\boldsymbol{\Lambda}_k\boldsymbol{\omega}_i, z_{i,k}\psi_k\right), \quad \forall\, 1 \leq i \leq n,\ 1 \leq k \leq p, \tag{6}$$

where $z_k \sim \mathsf{IG}(\nu_k/2, \nu_k/2)$.

Finally, to fully specify the Bayesian model, we will need to define the set of priors. The benefit of Geweke's strategy is fully revealed here, since (6) allows to extend the set of conjugate priors from (2). That is, we just need to introduce a prior over the vector $\boldsymbol{\nu}$. In this case, we choose to implement the uniform prior, namely

$$\nu_k \sim \mathsf{U}(2, \nu_{max}) \text{ for } k = 1, \ldots, p, \tag{7}$$

where ν_{max} is a constant such as $\nu_{max} = 100$. From, (2), (7), and (6), we next derive the corresponding conditional distributions. The latter consequently leads to the construction of an efficient Gibbs sampler. The conditional distributions for $(\boldsymbol{\Phi}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\omega}, \boldsymbol{\nu}, \boldsymbol{Z})$ (where $\boldsymbol{Z}$ is an $n \times p$ matrix of latent variables), are provided in (8). Since the derivation

of (8) is quite standard, we provide the corresponding technical details in Section 2 of the Online Supplementary Materials record.

$$\lambda_{k,j}|\mathbf{y}, \mathbf{\Lambda}_{-\lambda_{k,j}}\mathbf{\Phi}, \mathbf{\Psi}_\epsilon, \boldsymbol{\omega}, \boldsymbol{\nu}, \boldsymbol{Z} \sim \mathsf{N}\left(\sigma^2\left(\sum_{i=1}^{n}\frac{\omega_{i,j}(y_{i,k}-\sum_{l=1}^{q}\lambda_{kl}\omega_{il}\mathbb{1}_{\{l\neq j\}})}{\psi_k z_{i,k}}\right), \sigma^2\right), \quad (8)$$

$$\text{where } \sigma^2 = \left(\frac{1}{\sigma^2_{k,j}} + \sum_{i=1}^{n}\frac{\omega^2_{i,j}}{\psi_k z_{i,k}}\right)^{-1} \text{ for } k = 1, \ldots, p \text{ and } j = 1, \ldots, q,$$

$$\mathbf{\Phi}|\mathbf{y}, \mathbf{\Lambda}, \mathbf{\Psi}_\epsilon, \boldsymbol{\omega}, \boldsymbol{\nu}, \boldsymbol{Z} \sim \mathsf{IW}\left(A + \boldsymbol{\omega}^\top\boldsymbol{\omega}, a + n\right),$$

$$\psi_k|\mathbf{y}, \mathbf{\Lambda}, \mathbf{\Phi}, \boldsymbol{\omega}, \boldsymbol{\nu}, \boldsymbol{Z} \sim \mathsf{IG}\left(0.5n + \alpha_k, \beta_k + \frac{1}{2}\sum_{i=1}^{n}(z_{i,k})^{-1}(y_{i,k} - \mathbf{\Lambda}_k\boldsymbol{\omega}_i)^2\right),$$

$$\text{for } k = 1, \ldots, p,$$

$$\boldsymbol{\omega}_i|\mathbf{y}, \mathbf{\Lambda}, \mathbf{\Phi}, \mathbf{\Psi}_\epsilon, \boldsymbol{\nu}, \boldsymbol{Z} \sim \mathsf{MVN}\left(\Sigma'(\mathbf{\Lambda}^\top(\mathbf{\Psi}_\epsilon\boldsymbol{Z}_i)^{-1}\mathbf{y}_i), \Sigma'\right),$$

$$\text{where } \Sigma' = \left(\mathbf{\Phi}^{-1} + \mathbf{\Lambda}^\top(\mathbf{\Psi}_\epsilon\boldsymbol{Z}_i)^{-1}\mathbf{\Lambda}\right)^{-1} \text{ for } 1 \leq i \leq n,$$

$$z_{i,k}|\mathbf{y}, \mathbf{\Lambda}, \mathbf{\Phi}, \mathbf{\Psi}_\epsilon, \boldsymbol{\omega}, \boldsymbol{\nu} \sim \mathsf{IG}\left(0.5 + 0.5\nu_k, 0.5\nu_k + \frac{\psi_k^{-1}}{2}(y_{i,k} - \mathbf{\Lambda}_k\boldsymbol{\omega}_i)^2\right).$$

In order to complete the specification of the Gibbs sampler, we show how to handle a non-conjugate conditional distribution $\boldsymbol{\nu}|\mathbf{y}, \mathbf{\Lambda}, \mathbf{\Phi}, \mathbf{\Psi}_\epsilon, \boldsymbol{\omega}, \boldsymbol{Z}$.

3.1. Conditional distribution of the degree of freedom $\boldsymbol{\nu}$

It holds that:

$$p(\boldsymbol{\nu}|\mathbf{y}, \mathbf{\Lambda}, \mathbf{\Phi}, \mathbf{\Psi}_\epsilon, \boldsymbol{\omega}, \boldsymbol{Z}) \propto p(\boldsymbol{Z}|\boldsymbol{\nu})p(\boldsymbol{\nu}) \propto \prod_{i=1}^{n}\prod_{k=1}^{p} p(z_{i,k}|\nu_k)$$
$$\propto \prod_{i=1}^{n}\prod_{k=1}^{p}\frac{(0.5\nu_k)^{0.5\nu_k}}{\Gamma(0.5\nu_k)}(z_{i,k})^{-(0.5\nu_k+1)}\mathrm{e}^{-\frac{0.5\nu_k}{z_{i,k}}},$$

and thus, for $k = 1, \ldots, p$, we arrive at:

$$p(\nu_k|\mathbf{y}, \mathbf{\Lambda}, \mathbf{\Phi}, \mathbf{\Psi}_\epsilon, \boldsymbol{\omega}) \propto \prod_{i=1}^{n}\frac{(0.5\nu_k)^{0.5\nu_k}}{\Gamma(0.5\nu_k)}(z_{i,k})^{-(0.5\nu_k+1)}\mathrm{e}^{-\frac{0.5\nu_k}{z_{i,k}}} \quad (9)$$
$$= \frac{(0.5\nu_k)^{0.5n\nu_k}}{(\Gamma(0.5\nu_k))^n}\left(\prod_{i=1}^{n}z_{i,k}\right)^{-(0.5\nu_k+1)}\exp\left\{-0.5\nu_k\sum_{i=1}^{n}(z_{i,k})^{-1}\right\}.$$

The distribution in (9) is not standard, but, it can be sampled via Metropolis Hastings (Gelman et al., 2003; Gilks et al., 1996) method as follows. First, note that

$$\log p(\nu_k|\mathbf{y}, \mathbf{\Lambda}, \mathbf{\Phi}, \mathbf{\Psi}_\epsilon, \boldsymbol{\omega}, \boldsymbol{Z}) = 0.5n\nu_k\log(0.5\nu_k) - n\log\Gamma(0.5\nu_k)$$

$$- (0.5\nu_k + 1)\sum_{i=1}^{n} \log z_{i,k} - 0.5\nu_k \sum_{i=1}^{n} (z_{i,k})^{-1} + \text{constant},$$

an in addition, it holds that

$$\frac{d \log(0.5\nu_k)}{d\nu_k} = \frac{1}{0.5\nu_k}\frac{d0.5\nu_k}{d\nu_k} = \frac{1}{\nu_k}, \text{ and } \frac{d}{d\nu_k}\log(\Gamma(0.5\nu_k)) = \frac{d0.5}{d0.5\nu_k}\log(\Gamma(0.5\nu_k)),$$

$$\begin{aligned}\frac{d \log p(\nu_k|\boldsymbol{Z})}{d\nu_k} &= \frac{n}{2}\log(\nu_k/2) + \frac{n}{2} - n\frac{d\log\Gamma(\nu_k/2)}{d\nu_k} - \frac{1}{2}\sum_{i=1}^{n}\log z_{i,k} - 0.5\sum_{i=1}^{n}(z_{i,k})^{-1} \\ &= \frac{n}{2}\log(\nu_k/2) + \frac{n}{2} - \frac{n}{2}\frac{d\log\Gamma(\nu_k/2)}{d(\nu_k/2)} - \frac{1}{2}\sum_{i=1}^{n}\log z_{i,k} - 0.5\sum_{i=1}^{n}(z_{i,k})^{-1},\end{aligned}$$

and

$$\begin{aligned}\frac{d^2 \log p(\nu_k|\boldsymbol{Z})}{d^2\nu_k} &= \frac{n}{2\nu_k} - \frac{n}{2}\left(\frac{d}{d\nu_k}\frac{d}{d(\nu_k/2)}\log\Gamma(\nu_k/2)\right) \\ &= \frac{n}{2\nu_k} - \frac{n}{2}\left(\frac{d1/2}{d\nu_k/2}\frac{d}{d(\nu_k/2)}\log\Gamma(\nu_k/2)\right) = \frac{n}{2\nu_k} - \frac{n}{4}\frac{d^2}{d^2(\nu_k/2)}\log\Gamma(\nu_k/2).\end{aligned}$$

Since $\frac{d}{d(\nu_k/2)}\log\Gamma(\nu_k/2)$ and $\frac{d^2}{d^2(\nu_k/2)}\log\Gamma(\nu_k/2)$ is the *digamma* and the *trigamma* functions, respectively; these functions can be calculated using standard statistical packages.

Finally, in order to sample efficiently from the target in (9), we will approximate the Metropolis proposal distribution via a normal distribution with the mean located on the target's mode and the standard deviation set to be the negative Hessian of the target's logarithm evaluated at the mode. Specifically, the mode can be found using the Newton–Raphson method (Ypma, 1995) by solving numerically

$$\frac{n}{2}\log(\nu_k/2) + \frac{n}{2} - \frac{n}{2}\frac{d\log\Gamma(\nu_k/2)}{d(\nu_k/2)} - \frac{1}{2}\sum_{i=1}^{n}\log z_{i,k} - 0.5\sum_{i=1}^{n}(z_{i,k})^{-1} = 0.$$

Provided that the mode is located in ν_k^*, the corresponding hessian is computed at ν_k^*, via $H|_{\nu_k^*} = \frac{n}{2\nu_k^*} - \frac{n}{4}\frac{d^2}{d^2(\nu_k/2)}\log\Gamma(\nu_k^*/2)$. Consequently, the Metropolis-Hastings proposal is $\nu_k' \sim \mathsf{N}(\nu_k^*, H|_{\nu_k^*})$, and the sampled ν_k' is accepted with probability $\alpha = \min\left\{1, \frac{f(\nu_k')}{f(\nu_k)}\right\}$, where $f(\cdot)$ is the target proposal in (9).

4. Marginal likelihood estimation using importance sampling

In this section we propose a method which is based on importance sampling (IS) (Rubinstein et al., 2014) for marginal likelihood (and Bayes factor) estimation. Our numerical evaluation indicates that this simple idea results in good estimates with small relative errors (RE) (Rubino & Tuffin, 2009). The basic concept is as follows.

The marginal likelihood, namely, the integral $\int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$, can be written as:

$$\int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} = \int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\frac{g(\boldsymbol{\theta})}{g(\boldsymbol{\theta})}\mathrm{d}\boldsymbol{\theta} = \int \frac{p(\mathbf{y}|\boldsymbol{\theta})\,p(\boldsymbol{\theta})}{g(\boldsymbol{\theta})}g(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} = \mathbb{E}_{g(\boldsymbol{\theta})}\left[\frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{g(\boldsymbol{\theta})}\right].$$

The central difficulty here is to specify a sampling distribution $g(\boldsymbol{\theta})$. As soon as a good sampling distribution is available, one can obtain an estimator of the marginal likelihood via

$$\widehat{p(\mathbf{y})}_{\mathrm{IS}} = \frac{1}{N}\sum_{i=1}^{N}\frac{p(\mathbf{y}|\boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i)}{g(\boldsymbol{\theta}_i)},$$

where $\boldsymbol{\theta}_i$ for $i = 1,\ldots,N$ are independent and identically distributed samples from $g(\boldsymbol{\theta})$. While there exists an extensive IS theory, our contribution in this paper is that we show how to obtain a simple sampling distribution $g(\boldsymbol{\theta})$, which also provides good results in the sense of low variance and thus low RE. The RE, which is defined via: $\mathrm{RE} = \mathrm{Var}\left(\widehat{p(\mathbf{y})}_{\mathrm{IS}}\right)\Big/\mathbb{E}\left[\widehat{p(\mathbf{y})}_{\mathrm{IS}}\right]$ is used to examine the efficiency of CMC estimators. Specifically, the RE is proportional to the minimum sample size N needed to guarantee a prespecified estimator precision. For a detailed example, please see (Herr, Vaisman, Scovell, & Kinaev, 2024, Section D). Note that $\boldsymbol{\theta} = \{\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\omega}\}$ and $\boldsymbol{\theta} = \{\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\omega}, \boldsymbol{\nu}\}$ for the normal and for the Student's t distribution cases, respectively. The latent variable representation of the Student's t distribution allows us to handle both cases simultaneously. Our proposition is to treat each component of $\boldsymbol{\theta}$ independently, that is, we set

$$g(\boldsymbol{\theta}) = g(\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}_\epsilon, \boldsymbol{\nu}) = g_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda})g_{\boldsymbol{\Phi}}(\boldsymbol{\Phi})g_{\boldsymbol{\Psi}_\epsilon}(\boldsymbol{\Psi}_\epsilon)g_{\boldsymbol{\omega}}(\boldsymbol{\omega})g_{\boldsymbol{\nu}}(\boldsymbol{\nu}).$$

The distribution $g(\boldsymbol{\theta})$ will be derived from the posterior samples. Suppose now that $\boldsymbol{\theta}^{(i)}$ for $i = 1,\ldots,M$, are MCMC samples that (approximately) follow the desired posterior distribution. The derivation of $g(\boldsymbol{\theta})$ is based on the method of moments (Casella & Berger, 2002) as follows.

(1) For the factor loading matrix $\boldsymbol{\Lambda}$, define $g_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}) = \prod_{k=1}^{p}\prod_{j=1}^{q} g_{\lambda_{k,j}}(\lambda_{k,j})$, for $k = 1,\ldots,p$ and $j = 1,\ldots,q$. Set $g_{\lambda_{k,j}}(\lambda_{k,j})$ to be a Normal distribution density function, namely, let $\lambda_{k,j} \sim \mathsf{N}\left(\widehat{\mu}_{\lambda_{k,j}}, \widehat{\sigma^2}_{\lambda_{k,j}}\right)$, where

$$\widehat{\mu}_{\lambda_{k,j}} = \frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\theta}^{(i)}(\lambda_{k,j}) \text{ and } \widehat{\sigma^2}_{\lambda_{k,j}} = \frac{1}{M-1}\sum_{m=1}^{M}\left(\boldsymbol{\theta}^{(i)}(\lambda_{k,j}) - \widehat{\mu}_{\lambda_{k,j}}\right)^2.$$

(2) Similarly, for the latent variable vector $\boldsymbol{\omega}$, define $g_{\boldsymbol{\omega}}(\boldsymbol{\omega}) = \prod_{i=1}^{n}\prod_{j=1}^{q} g_{\omega_{i,j}}(\omega_{i,j})$, for $i = 1,\ldots,n$ and $j = 1,\ldots,q$. Set $g_{\omega_{i,j}}(\omega_{i,j})$ to be a Normal distribution density function, that is, let $\omega_{i,j} \sim \mathsf{N}\left(\widehat{\mu}_{\omega_{i,j}}, \widehat{\sigma^2}_{\omega_{i,j}}\right)$, where

$$\widehat{\mu}_{\omega_{i,j}} = \frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\theta}^{(i)}(\omega_{i,j}), \text{ and } \widehat{\sigma^2}_{\omega_{i,j}} = \frac{1}{M-1}\sum_{m=1}^{M}\left(\boldsymbol{\theta}^{(i)}(\omega_{i,j}) - \widehat{\mu}_{\omega_{i,j}}\right)^2.$$

(3) For the covariance matrix $\boldsymbol{\Phi}$, define $g_{\boldsymbol{\Phi}}(\boldsymbol{\Phi})$ to follow the Inverse Wishart distribu-

tion. Let $\boldsymbol{\theta}^{(i)}(\boldsymbol{\Phi})$ be the ith posterior sample that corresponds to the covariance matrix $\boldsymbol{\Phi}$. Then, set $\boldsymbol{\Phi} \sim \mathsf{IW}(A', a')$, and estimate the parameters of $g_{\boldsymbol{\Phi}}(\boldsymbol{\Phi})$, namely, A' and a', as follows. First, note that:

$$\mathbb{E}[\boldsymbol{\Phi}] = \frac{A'}{a' - q - 1} \quad \text{for } a' > q + 1, \tag{10}$$

and that:

$$\mathrm{Var}(\boldsymbol{\Phi}_{i',j'}) = \frac{(a' - q + 1)A'^2_{i',j'} + (a' - q - 1)A_{i',i'}A_{j',j'}}{(a' - q)(a' - q - 1)^2(a' - q - 3)} \text{ for } 1 \leq i', j' \leq q.$$

That is, for any $1 \leq i' \leq q$,

$$\mathrm{Var}(\boldsymbol{\Phi}_{i',i'}) = \frac{(a' - q + 1)A'^2_{i',i'} + (a' - q - 1)A_{i',i'}A_{i',i'}}{(a' - q)(a' - q - 1)^2(a' - q - 3)} = \frac{2A'^2_{i',i'}}{(a' - q - 1)^2(a' - q - 3)},$$

and since $\mathbb{E}[\boldsymbol{\Phi}_{i',j'}]^2 = \frac{A'^2_{i,i}}{(a'-q-1)^2}$, it holds that:

$$\frac{\mathbb{E}[\boldsymbol{\Phi}_{i',i'}]^2}{\mathrm{Var}(\boldsymbol{\Phi}_{i',i'})} = \frac{\frac{A'^2_{i,i}}{(a'-q-1)^2}}{\frac{2A'^2_{i',i'}}{(a'-q-1)^2(a'-q-3)}} = \frac{1}{2}(a' - q - 3). \tag{11}$$

Finally, using the posterior samples, calculate

$$\widehat{\mu}_{\boldsymbol{\Phi}} = \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{\theta}^{(i)}(\boldsymbol{\Phi}), \text{ and } \widehat{\sigma^2}_{\boldsymbol{\Phi}_{i',i'}} = \frac{1}{M-1}\sum_{m=1}^{M}\left(\boldsymbol{\theta}^{(i)}(\boldsymbol{\Phi}_{i',i'}) - \widehat{\mu}_{\boldsymbol{\Phi}_{i',i'}}\right)^2,$$

for some $1 \leq i' \leq q$ and thus, the a' and the A' parameters can be estimated (using (11) and (10)), via:

$$a' = 2\frac{\widehat{\mu}_{\boldsymbol{\Phi}_{i',i'}}}{\widehat{\sigma^2}_{\boldsymbol{\Phi}_{i',i'}}} + q + 3, \text{ and } A' = (a' - q - 1)\widehat{\mu}_{\boldsymbol{\Phi}}.$$

(4) For the covariance matrix $\boldsymbol{\Psi}_\epsilon$, define $g_{\boldsymbol{\Psi}_\epsilon}(\boldsymbol{\Psi}_\epsilon) = \prod_{k=1}^{p} g_{\psi_k}(\psi_k)$, where for $k = 1, \ldots, p$, set $g_{\psi_k}(\psi_k)$ to be a Gamma distribution density function, namely, let $\psi_k \sim \mathsf{Gamma}\left(\widehat{\alpha}_{\psi_k}, \widehat{\beta}_{\psi_k}\right)$. Since it holds that for $X \sim \mathsf{Gamma}(\alpha, \beta)$, $\mathbb{E}[X] = \alpha\beta$ and $\mathrm{Var}(X) = \alpha\beta^2$, we arrive at:

$$\frac{\mathrm{Var}(X)}{\mathbb{E}[X]} = \beta \text{ and } \frac{\mathbb{E}[X]}{\beta} = \alpha. \tag{12}$$

By computing

$$\widehat{\mu}_{\psi_k} = \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{\theta}^{(i)}(\psi_k), \text{ and } \widehat{\sigma^2}_{\psi_k} = \frac{1}{M-1}\sum_{m=1}^{M}\left(\boldsymbol{\theta}^{(i)}(\psi_k) - \widehat{\mu}_{\psi_k}\right)^2,$$

the $\widehat{\alpha}_{\psi_k}$ and the $\widehat{\beta}_{\psi_k}$ parameters are estimated using (12) via

$$\widehat{\beta}_{\psi_k} = \frac{\widehat{\sigma^2}_{\psi_k}}{\widehat{\mu}_{\psi_k}} \text{ and } \widehat{\alpha}_{\psi_k} = \frac{\widehat{\alpha}_{\psi_k}}{\widehat{\beta}_{\psi_k}}.$$

(5) Similarly to the calculation presented for $\boldsymbol{\Psi}_\epsilon$, for the vector of degrees of freedom, we define $g_{\boldsymbol{\nu}}(\boldsymbol{\nu}) = \prod_{k=1}^{p} g_{\nu_k}(\nu_k)$, where for $k = 1, \ldots, p$, set $g_{\nu_k}(\nu_k)$ to be the Gamma distribution density; specifically, let $\nu_k \sim \mathsf{Gamma}\left(\widehat{\alpha}_{\nu_k}, \widehat{\beta}_{\nu_k}\right)$. By computing

$$\widehat{\mu}_{\nu_k} = \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{\theta}^{(i)}(\nu_k), \text{ and } \widehat{\sigma^2}_{\nu_k} = \frac{1}{M-1}\sum_{m=1}^{M}\left(\boldsymbol{\theta}^{(i)}(\nu_k) - \widehat{\mu}_{\nu_k}\right)^2,$$

the $\widehat{\alpha}_{\nu_k}$ and the $\widehat{\beta}_{\nu_k}$ parameters are estimated using (12) via $\widehat{\beta}_{\nu_k} = \frac{\widehat{\sigma^2}_{\nu_k}}{\widehat{\mu}_{\nu_k}}$ and $\widehat{\alpha}_{\nu_k} = \frac{\widehat{\alpha}_{\nu_k}}{\widehat{\beta}_{\nu_k}}$.

5. Results

In this section, we examine the performance of both the normal (nCFA) and the Student's t (tCFA) models. The nCFA and the tCFA methods are applied to synthetic and real data sets. Specifically, the following case studies are considered.

(1) To test the nCFA and the tCFA models under a controlled setting, we consider synthetic CFA instances that were generated from the CFA model in (1). In particular, we took $p = 8$ and $q = 3$. The covariance matrix for latent variables and the factor loading matrix, were set to be:

$$\boldsymbol{\Phi} = \begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1.5 & -0.5 \\ 0 & -0.5 & 2 \end{pmatrix} \text{ and } \boldsymbol{\Lambda}^\top = \begin{pmatrix} 1 & 0.5 & 0.8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0.5 & 0.8 \end{pmatrix},$$

respectively. To reliably benchmark the nCFA and the tCFA models, we simulated 200 datasets for each $n \in \{100, 150, 200, 250\}$. For each n, 100 datasets were generated from the Gaussian model using the $\boldsymbol{\Psi}_\epsilon = \text{diag}(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8)$ covariance matrix. The remaining 100 datasets were generated from the g-and-k distribution, which is more flexible as compared to the normal distribution and thus can model outliers (Drovandi & Pettitt, 2011; Prangle, 2020). In particular, the g-and-k distribution is defined via a quantile function:

$$\begin{aligned} &Q(q_{\text{gk}}|A_{\text{gk}}, B_{\text{gk}}, g_{\text{gk}}, k_{\text{gk}}, c_{\text{gk}}) \\ &\quad = A_{\text{gk}} + B_{\text{gk}}\left[1 + c_{\text{gk}}\frac{1 - \exp\{-g_{\text{gk}} z(q_{\text{gk}})\}}{1 + \exp\{-g_{\text{gk}} z(q_{\text{gk}})\}}\right](1 + z(q_{\text{gk}})^2)^k z(q_{\text{gk}}), \end{aligned} \tag{13}$$

where $z(q_{\text{gk}})$ is the q_{gk}-th quantile of the standard normal distribution function. For this section, we take $A_{\text{gk}} = -1$, $B_{\text{gk}} = 1.5$, $g_{\text{gk}} = 2$, $k_{\text{gk}} = 0.4$, and $c_{\text{gk}} = 0.8$.

The g-and-k distribution creates a considerable divergence from the normal distribution assumption, which is the driving force for the proposed controlled setup. The details are provided in Table 1, which summarizes multivariate skewness, kurtosis, and the corresponding p-values (Mardia, 1970). The average Mardia's skewness and kurtosis are within the reported range (between the 75th and the 95th percentiles) of the 254 multivariate distributions collected from authors of articles published in Psychological Science and the American Education Research Journal (Cain et al., 2017).

Table 1. The average multivariate skewness, kurtosis, and p-values including corresponding standard deviations (in parentheses). The p-values correspond to Multivariate Normality test, indicating a departure from the multivariate normal distribution of the g-and-k measurement error model.

	Gaussian measurement error		g-and-k measurement error	
n	skewness (p-value)	kurtosis (p-value)	skewness (p-value)	kurtosis (p-value)
100	6.82 (0.82)	76.8 (1.79)	46.6 (9.11)	121 (12.5)
	0.62 (0.26)	0.30 (0.27)	6.43×10^{-37} (6.39×10^{-36})	1.46×10^{-8} (1.45×10^{-7})
150	4.63 (0.59)	77.7 (1.69)	47.2 (9.16)	129 (14.1)
	0.57 (0.28)	0.34 (0.29)	8.01×10^{-100} (7.96×10^{-99})	1.22×10^{-33} (1.21×10^{-32})
200	3.47 (0.46)	78.3 (1.48)	48.3 (8.11)	135 (13.8)
	0.58 (0.29)	0.38 (0.29)	6.70×10^{-164} (1.00×10^{-300})	4.14×10^{-64} (4.12×10^{-63})
250	2.78 (0.39)	78.6 (1.33)	48.2 (8.17)	137 (14.1)
	0.57 (0.29)	0.41 (0.29)	3.18×10^{-208} (1.00×10^{-300})	1.09×10^{-82} (1.08×10^{-81})

For each data size n, and for each measurement error distribution (normal and g-and-k), we applied the nCFA and the tCFA algorithms. Our numerical results indicate that on average, tCFA outperforms the nCFA for both normal and g-and-k distribution measurement errors. However, the benefit of tCFA increases with n for the g-and-k distribution measurement error model. This is not very surprising, since the Student's t distribution is more robust and thus, it is expected to handle outliers in a better fashion.

(2) In our second case study, we consider the CFA analysis of the well-known *Holzinger & Swineford 1939* dataset (Rosseel, 2012). This dataset is used as a working example in both the `lavaan` and the `blavaan` structural equation modeling packages (Merkle & Rosseel, 2018; Rosseel, 2012). The dataset consists of mental ability scores of children from different schools. Following Rosseel (2012), we also consider 9 out of the original 26 scores, and use the CFA model which consists of 3 latent factors (visual, textual, and speed), that are measured by x_1, x_2 and x_3 for the visual factor, by x_4, x_5 and x_6 for the textual factor, and by x_7, x_8 and x_9 for the speed factor. Our results indicate the tCFA method outperforms the nCFA model.

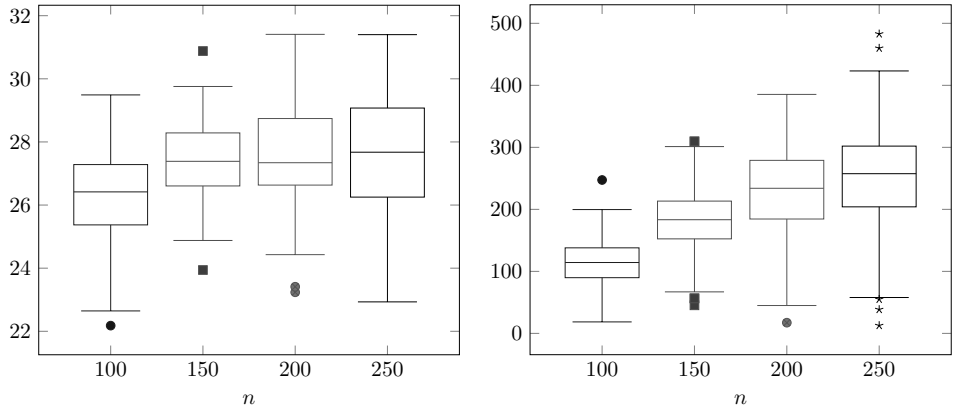
The experimental setup. Both the nCFA and the tCFA models, were implemented in `C++` version 11. The code was compiled and wrapped to a research `R` package (`R` version 4.2.2). Section 5 of the Online Supplementary Materials record contains examples on how to use the package. All tests, including the compilation of the `R` package were performed on Intel Core i7-6920HQ CPU 2.90GHz processor with 32GB of RAM running 64 bit Debian 12 "bookworm". The software and the research data, are freely at `https://osf.io/ehqsf/?view_only=21a5075217474427bb44054bbaf755f5`. By conducting several preliminary experiments (not reported here), we obtained a set of parameters (burnin and sample size), that work well in the sense of MCMC convergence and IS estimator RE, for the case studies under consideration. In particular, for the

synthetic case study, both the burnin and the sample size parameters of the MCMC sampler were fixed to be 10,000. For the second case study, we used the 10,000 burnin and 50,000 sample size. Chain convergence was tested using the Gelman-Rubin $\widehat{R}$ diagnostics (Gelman & Rubin, 1992). For both case studies, the IS marginal likelihood estimator was computed using 1000 samples.

Finally, for all data instances, the same set of priors was used. While we understand the importance of proper prior selection and both prior and posterior predictive checks (Kruschke, 2021), the purpose of this study is to compare the nCFA and the tCFA methods, thus, the identical prior set was applied. Specifically, we used the following prior distributions: $\lambda_{k,j} \sim \mathsf{N}(0, 10^2)$ for $k = 1, \ldots, 8$ and $j = 1, \ldots, 3$, $\mathbf{\Phi} \sim \mathsf{IW}(\mathbf{I}_3, 4)$, $\psi_k \sim \mathsf{IG}(1.0, 1.0)$ for $k = 1, \ldots, 8$, and $\nu_k \sim \mathsf{U}(2, 100)$ for $k = 1, \ldots, 8$. Here, $\mathbf{I}_3$ is the 3×3 identity matrix. To ensure identifiability of the models, we fix $\lambda_{k,j} = 1$ for $(k, j) \in \{(1, 1), (4, 2), (6, 3)\}$. Sections 5.1 and 5.2 present detailed results for the synthetic and the *Holzinger & Swineford 1939* experiments, respectively.

5.1. Synthetic data case study

As noted above, for all MCMC runs in this section, we selected the burnin and the sample size to be 10,000. The marginal likelihood estimator was obtained based on 1000 samples. Our numerical evaluation indicates that the RE of the IS marginal likelihood estimator is very small. In particular, we observed that the average RE for all cases is about 6.97×10^{-4}, which indicates a strong stability of the proposed estimator.



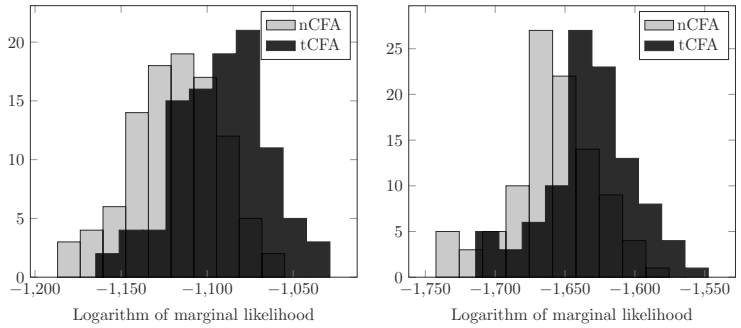
(a) Summary of logarithm of Bayes factors as a function of the instance size n for the Gaussian measurement error synthetic data. For $n = 100, 150, 200, 250$, the estimated log BFs are $26.351 \pm 1.405, 27.397 \pm 1.237, 27.565 \pm 1.554, 27.637 \pm 1.832$, respectively.

(b) Summary of logarithm of Bayes factors as a function of the instance size n for the g- and k measurement error synthetic data. For $n = 100, 150, 200, 250$, the estimated log BFs are $115.031 \pm 37.863, 181.718 \pm 52.456, 229.060 \pm 70.551, 251.037 \pm 88.471$, respectively.

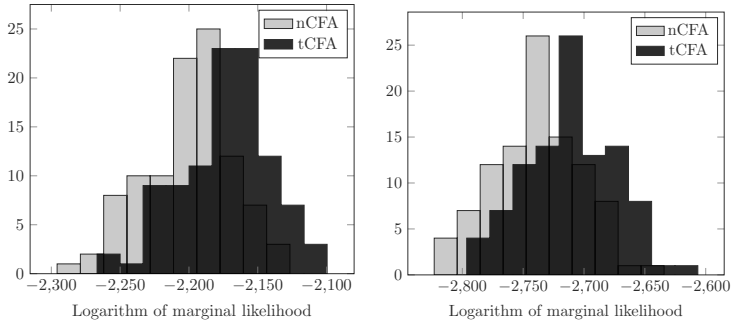
Figure 1. Summary of logarithm of Bayes factors as a function of the instance size n for the Gaussian and the g-and k measurement error synthetic data.

Figure 1 depicts the logarithm of the corresponding Bayes factor for the Gaussian and the g-and-k measurement error data. It is apparent that the tCFA model is preferable as compared to the nCFA model. However, for the g-and-k measurement error data, the Bayes factor grows with n. Figure 2 and Figure 3 show the histograms of the logarithm of marginal likelihood (for both the nCFA and the tCFA methods), for

the Gaussian and the g-and-k measurement error data, respectively. Again, one can observe that tCFA introduces a superior performance as compared to nCFA.

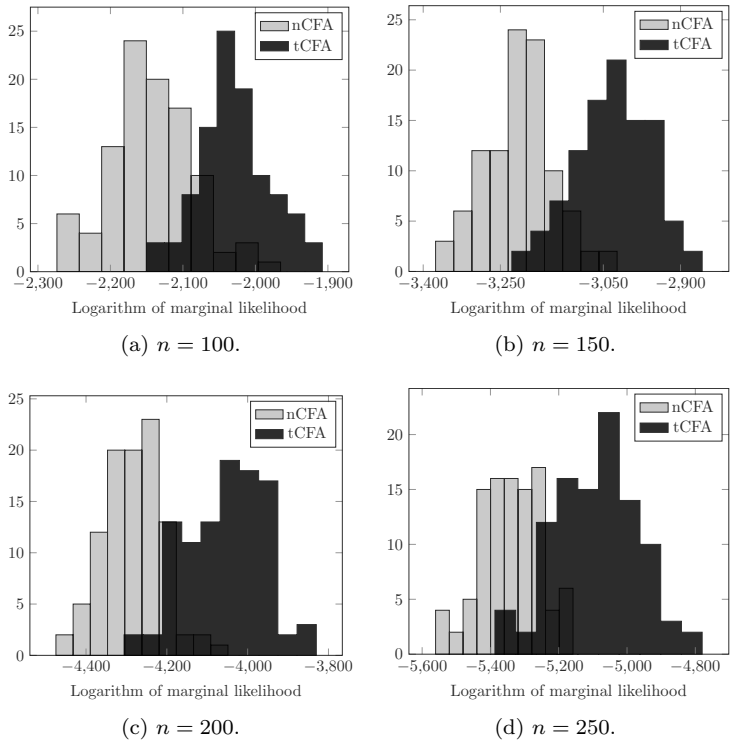


(a) $n = 100$. (b) $n = 150$.



(c) $n = 200$. (d) $n = 250$.

Figure 2. Histograms of marginal likelihoods for the nCFA and the tCFA models when applied to the Gaussian measurement error synthetic data.



(a) $n = 100$. (b) $n = 150$.

(c) $n = 200$. (d) $n = 250$.

Figure 3. Histograms of marginal likelihoods for the nCFA and the tCFA models when applied to the g-and k measurement error synthetic data.

Remark 1 (Bayes factors)**.** The tCFA model is favorable according to the Bayes factor. This is especially true for data examples that have g-and-k measurement errors. It is important to note that a better model will lead to gains in the corresponding squared errors of point estimators. To see this, we consider a typical synthetic instance of size $n = 250$ that was generated with g-and-k measurement errors. Table 2 and Table 3 in Section 4 of the Online Supplementary Materials record, show the convergence diagnostic (including bias and squared errors) of nCFA and tCFA for this specific dataset. The estimated logarithm Bayes factor is 292.4 in favor of the tCFA model. Tables 2 and 3 are very instructive in the sense that the average squared error and the average bias are 23.42 and 2.79 for the nCFA model and 1.17 and 0.801 for the tCFA model. Moreover, we generated 100 instances of size $n = 250$ with g-and-k measurement errors. The corresponding summary of the squared error and bias of the nCFA and the tCFA models are presented in Table 4 in the Online Supplementary Materials record. This summary indicates that on average, the tCFA model outperforms the nCFA model, with the exception of $\mathbf{\Phi}$'s bias. The reason for the tCFA's performance gains is the flexibility of the Student's t distribution. Specifically, note that Table 3 shows that the estimated mean degrees of freedom are between 2.1 and 3.1. These degrees of freedom allow thicker tails and therefore provide a better model fit for the g-and-k distributed measurement errors.

5.2. The Holzinger Swineford 1939 school data

As noted in the experimental setup section, for this example, we set the 10,000 burnin and the 50,000 sample sizes. Three independent MCMC chains were created for both the nCFA and the tCFA models. Similar to the synthetic case study, the IS marginal likelihood estimator is very stable. Based on 1000 samples and multiple independent runs, the average estimator is -3920.2 for the nCFA and -3888.9 for the tCFA models. The corresponding average RE is stable and is equal to 3.81×10^{-4} and to 2.50×10^{-4} for the nCFA and the tCFA models, respectively. That is, the logarithm of Bayes factor is $\log \mathrm{BF} \approx -3888.9 - (-3920.2) \approx 31.3$, that is, the tCFA model is strongly preferable. This is not very surprising as the data deviates from the normality assumption. Specifically, the Mardia's skewness and kurtosis are 6.81 and 102.9, respectively. The corresponding p-values are 2.19×10^{-14} and 0.016, respectively, and thus we can conclude that the null hypothesis (data normality) is rejected at the 0.05 significance level. Complete posterior summaries for the nCFA and the tCFA models are summarized in Tables 2 and 3, respectively.

Table 2. MCMC summary for nCFA model for the Holzinger Swineford 1939 school data. The execution time for each chain is about 48.491 seconds on average. The marginal likelihood estimation took about 8.987 seconds on average.

Parameter	mean	sd	2.5%	50%	97.5%	$\widehat{R}$
ψ_1	0.612	0.113	0.391	0.612	0.835	1.00
ψ_2	1.131	0.105	0.940	1.127	1.351	1.00
ψ_3	0.834	0.097	0.650	0.831	1.030	1.00
ψ_4	0.388	0.049	0.297	0.386	0.489	1.00
ψ_5	0.455	0.059	0.346	0.452	0.577	1.00
ψ_6	0.366	0.044	0.284	0.364	0.458	1.00
ψ_7	0.834	0.089	0.672	0.830	1.022	1.00
ψ_8	0.521	0.091	0.346	0.520	0.703	1.00
ψ_9	0.550	0.092	0.367	0.551	0.728	1.00
$\lambda_{2,1}$	0.601	0.117	0.389	0.595	0.846	1.00
$\lambda_{3,1}$	0.791	0.128	0.565	0.782	1.067	1.00
$\lambda_{5,2}$	1.128	0.068	1.002	1.126	1.268	1.00
$\lambda_{6,2}$	0.938	0.059	0.828	0.936	1.058	1.00
$\lambda_{8,3}$	1.230	0.168	0.942	1.216	1.601	1.00
$\lambda_{9,3}$	1.195	0.223	0.831	1.168	1.705	1.00
$\mathbf{\Phi}_{1,1}$	0.741	0.141	0.487	0.733	1.037	1.00
$\mathbf{\Phi}_{1,2}$	0.380	0.078	0.235	0.377	0.542	1.00
$\mathbf{\Phi}_{1,3}$	0.243	0.053	0.147	0.240	0.354	1.00
$\mathbf{\Phi}_{2,2}$	0.961	0.112	0.757	0.956	1.197	1.00
$\mathbf{\Phi}_{2,3}$	0.164	0.047	0.078	0.162	0.264	1.00
$\mathbf{\Phi}_{3,3}$	0.350	0.086	0.200	0.344	0.534	1.00

Table 3. MCMC summary for the tCFA model for the Holzinger Swineford 1939 school data. The execution time for each chain is about 105.617 seconds on average. The marginal likelihood estimation took about 10.37 seconds on average.

Parameter	mean	sd	2.5%	50%	97.5%	$\widehat{R}$
ψ_1	0.577	0.111	0.361	0.575	0.801	1.00
ψ_2	1.038	0.122	0.793	1.039	1.280	1.00
ψ_3	0.804	0.097	0.621	0.802	1.003	1.00
ψ_4	0.368	0.050	0.275	0.367	0.471	1.00
ψ_5	0.434	0.059	0.324	0.432	0.556	1.00
ψ_6	0.328	0.052	0.223	0.329	0.429	1.01
ψ_7	0.801	0.091	0.634	0.796	0.990	1.00
ψ_8	0.469	0.092	0.293	0.467	0.655	1.00
ψ_9	0.519	0.091	0.336	0.520	0.695	1.00
$\lambda_{2,1}$	0.601	0.116	0.392	0.594	0.849	1.00
$\lambda_{3,1}$	0.803	0.128	0.579	0.793	1.080	1.00
$\lambda_{5,2}$	1.129	0.068	1.002	1.127	1.270	1.00
$\lambda_{6,2}$	0.926	0.059	0.815	0.924	1.047	1.00
$\lambda_{8,3}$	1.197	0.163	0.915	1.184	1.558	1.00
$\lambda_{9,3}$	1.172	0.225	0.816	1.142	1.691	1.00
$\mathbf{\Phi}_{1,1}$	0.734	0.139	0.480	0.728	1.024	1.00
$\mathbf{\Phi}_{1,2}$	0.380	0.078	0.235	0.377	0.542	1.00
$\mathbf{\Phi}_{1,3}$	0.243	0.053	0.147	0.240	0.354	1.00
$\mathbf{\Phi}_{2,2}$	0.964	0.112	0.759	0.958	1.200	1.00
$\mathbf{\Phi}_{2,3}$	0.168	0.048	0.081	0.165	0.269	1.00
$\mathbf{\Phi}_{3,3}$	0.357	0.088	0.201	0.352	0.544	1.00
ν_1	49.68	27.57	7.21	48.19	97.26	1.01
ν_2	43.30	27.57	7.14	37.73	96.41	1.01
ν_3	63.32	23.16	19.28	65.31	98.36	1.00
ν_4	56.18	26.23	10.95	56.82	97.86	1.00
ν_5	59.07	24.90	13.88	60.40	98.12	1.01
ν_6	38.60	28.46	4.97	30.75	95.78	1.01
ν_7	62.05	23.82	17.20	63.96	98.31	1.01
ν_8	37.59	27.46	5.73	29.42	95.49	1.00
ν_9	49.50	27.47	7.87	47.52	97.31	1.01

Finally, we examined the mixing properties of the nCFA and the tCFA models. For the Holzinger Swineford 1939 school data. The Gelman-Rubin diagnostics $\widehat{R}$ in Tables 2 and 3, indicates that the chains reached convergence as desired. Finally, some additional graphical convergence diagnostics summaries are presented in Section 3 of the Online Supplementary Materials record.

6. Discussion

In this study, we proposed a Bayesian approach for performing robust confirmatory factor analysis. Specifically, we demonstrated how to efficiently carry out inference tasks under the Student's t distribution measurement error assumption, as well as how to estimate the corresponding marginal likelihood and perform model selection. Our numerical analysis shows that when dealing with real-world (non-normal) data, the Student's t distribution approach can result in a significant benefit. Additionally, we provided a research `R` package for confirmatory factor analysis that is effective in implementing the measurement error models for both the Gaussian and Student's t distributions. One of the significant contribution of this work is that this package was implemented in `C++` and thus it can be readily used in applied research for real-life data sizes. In our package, we also implemented the marginal likelihood estimator importance sampling procedure, and therefore, our research software provides a full work flow machinery for performing robust Bayesian confirmatory factor analysis.

Our method has many applications in applied settings when researchers are concerned about the influence of outliers when performing CFA. The approach outlined has clear benefits for researchers who, based on theory, assume that latent variables of interest infer the presence of more extreme values than what would be expected from a normal distribution. Such latent variables are not uncommon (Cain et al., 2017; Micceri, 1989) in subfields of social science like psychopathology, political science and social psychology. For instance, depression (Tomitaka, 2020) and political attitudes (Alizadeh & Cioffi-Revilla, 2014) can exhibit more extreme values than what would be expected from a normal distribution, particularly in certain subpopulations. The results show that under such conditions (i.e., knowing that extreme values are less rare events) our approach outperforms a typical approach to CFA, which assumes multivariate normality.

Despite the advantages of the robust Bayesian approach, this study is subject to a few limitations. The primary flaw is that factor and ordinal variables cannot yet be handled by our software. This issue can be resolved by standard Bayesian latent variable approaches. However, in the current study, these were not implemented. The proposed `R` package is not a commercial software and only introduce a single-threaded implementation of the proposed methods. That is, we believe that a parallel utilization will allow to handle even larger real-world instances. Finally, it is important to note that the sample sizes for both the MCMC chains and the IS estimator, are of crucial importance. Our package does not implement convergence measures and relies on external `R` packages such as `coda`.

We believe that this manuscript opens the way for significant research, and, that the following directions are of interest.

(1) It is imperative, as previously noted, to expand the suggested approach to incorporate factor and ordinary variable management. Furthermore, it is of crucial importance to extend the concepts presented in this manuscript to general structural equation models.
(2) We presented a number of computational experiments with certain prior distributions. Therefore, it is important to conduct an extensive research that take into account the sensitivity of inference under additional prior distribution settings. In addition, it will be of interest to extend the computational experimental investigation by considering non-Gaussian synthetic instances obtained via copula-based techniques (Foldnes & Grønneberg, 2015; Foldnes & Olsson, 2016; Grønneberg, Foldnes, & Marcoulides, 2022).
(3) Finally, a parallel implementation that can handle larger real-life instances is of great interest. For example, since the marginal likelihood computation is an importance sampling estimator, its computational performance can be significantly accelerated by using multiple central processing units or a graphics processing unit.

Acknowledgement(s): We are thoroughly grateful to the editor and to anonymous reviewers for their valuable and constructive remarks and suggestions. This work was supported by the Analytics for the Australian Grains Industry grant and by the Commonwealth Scientific and Industrial Research Organisation (hydrogen energy systems future Science Platform).

Disclosure Statement: No potential conflict of interest was reported by the author(s).

Disclosure Statement: The data are available at the open science repository.

Disclosure Statement: The associated files for this manuscript are available at `https://osf.io/ehqsf/?view_only=21a5075217474427bb44054bbaf755f5`.

References

Alizadeh, M., & Cioffi-Revilla, C. (2014). Distributions of opinion and extremist radicalization: Insights from agent-based modeling. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (Vol. 8851, pp. 348–358). Cham: Springer International Publishing.

Asparouhov, T., Muthén, B., & Morin, A. J. S. (2015). Bayesian Structural Equation Modeling With Cross-Loadings and Residual Covariances: Comments on Stromeyer et al. *Journal of management*, *41*(6), 1561–1577.

Browne, M. W. (1973). Generalized least squares estimators in the analysis of covariance structures. *Research bulletin (Educational Testing Service.)*, *1973*(1), i–36.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British journal of mathematical & statistical psychology*, *37*(1), 62–83.

Cain, M. K., Zhang, Z., & Yuan, K.-H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior research methods*, *49*(5), 1716–1735.

Casella, G., & Berger, R. (2002). *Statistical inference* (2nd ed.). Pacific Grove, CA: Duxbury/Thomson Learning.

Depaoli, S. (2021). *Bayesian Structural Equation Modeling*. New York, NY: Guilford Publications.

Drovandi, C. C., & Pettitt, A. N. (2011). Likelihood-free Bayesian estimation of multivariate quantile distributions. *Computational statistics & data analysis*, *55*(9), 2541–2556.

Foldnes, N., & Grønneberg, S. (2015). How General is the Vale–Maurelli Simulation Approach? *Psychometrika*, *80*(4), 1066–1083.

Foldnes, N., & Olsson, U. H. (2016). A simple simulation technique for nonnormal data with prespecified skewness, kurtosis, and covariance matrix. *Multivariate Behavioral Research*.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian Data Analysis* (3rd ed.). Oxfordshire, UK: Taylor & Francis.

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*, 457–472.

Geweke, J. (1993). Bayesian treatment of the independent student-t linear model. *Journal of applied econometrics (Chichester, England)*, *8*(S1), S19–S40.

Gilks, W. R. W. R., Richardson, S. S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. London, UK: Chapman & Hall.

Grønneberg, S., Foldnes, N., & Marcoulides, K. M. (2022). covsim: An R Package for Simulating Non-Normal Data for Structural Equation Models Using Copulas. *Journal of statistical*

software, *102*(3), 1–45.
Haig, B. D. (2005). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research*, *40*(3), 303–329.
Haig, B. D. (2018). Exploratory factor analysis, theory generation, and scientific method. In *Method matters in psychology: Essays in applied philosophy of science* (pp. 65–88). Cham: Springer International Publishing.
Harrington, D. (2009). *Confirmatory factor analysis*. Oxford, UK: Oxford University Press.
Herr, D. Z., Vaisman, R., Scovell, M., & Kinaev, N. (2024). On alternative monte carlo methods for parameter estimation in gamma process models with intractable likelihood. *IEEE transactions on reliability*, 1–15.
Jackman, S. (2009). *Bayesian Analysis for the Social Sciences* (1st ed., Vol. 846). Newark, NJ: Wiley.
Joreskog, K. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*(2), 183–202.
Kruschke, J. K. (2021). Bayesian Analysis Reporting Guidelines. *Nature human behaviour*, *5*(10), 1282–1291.
Lange, K., Little, R. J. A., & Taylor, J. (1989). Robust statistical modeling using the t-distribution. *Journal of the American Statistical Association*, *84*(408), 881–896.
Lawal, O. A., Awosoga, O., Santana, M. J., James, M. T., Southern, D. A., Wilton, S. B., ... Sajobi, T. (2020). Psychometric evaluation of a Canadian version of the Seattle Angina Questionnaire (SAQ-CAN). *Health and quality of life outcomes*, *18*(1), 377–377.
Lee, S.-Y. (2007). *Structural equation modeling: a Bayesian approach*. Chichester, UK and Hoboken, NJ: Wiley.
Lee, S.-Y., & Song, X.-Y. (2012). *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences* (1st ed.). Newark, NJ: Wiley.
Lesia, M. P., Aigbavboa, C. O., & Thwala, W. D. (2024). Factors influencing residential location choice in South Africa: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). *Journal of housing and the built environment*, *39*(1), 133–160.
Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*(3), 519–530.
Merkle, E. C., & Rosseel, Y. (2018). blavaan: bayesian structural equation models via parameter expansion. *Journal of statistical software*, *85*(4), 1–30.
Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological bulletin*, *105*(1), 156–166.
Prangle, D. (2020). gk: An R Package for the g-and-k and Generalised g-and-h Distributions. *The R journal*, *12*(1), 7.
Roos, J. M., & Bauldry, S. (2022). *Confirmatory factor analysis*. Los Angeles, CA: SAGE Publications, Inc.
Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of statistical software*, *48*(2), 1–36.
Rubino, G., & Tuffin, B. (2009). *Rare event simulation using Monte Carlo methods*. Oxford, UK: Wiley.
Rubinstein, R. Y., Ridder, A., & Vaisman, R. (2014). *Fast sequential Monte Carlo methods for counting and optimization* (1st ed.). Hoboken, NJ: John Wiley & Sons, Inc.
Shao, K., Elahi Shirvan, M., & Alamer, A. (2022). How Accurate Is Your Correlation? Different Methods Derive Different Results and Different Interpretations. *Frontiers in psychology*, *13*, 901412–901412.
Stroup, W. W. (2015). Rethinking the analysis of non-normal data in plant and soil science. *Agronomy Journal*, *107*(2), 811–827.
Sureshchandar, G. (2023). Quality 4.0 – a measurement model using the confirmatory factor analysis (CFA) approach. *The International journal of quality & reliability management*, *40*(1), 280–303.
Tomitaka, S. (2020). Patterns of item score and total score distributions on depression rating scales in the general population: evidence and mechanisms. *Heliyon*, *6*(12), e05862–e05862.

Tong, X., & Zhang, Z. (2020). Robust Bayesian Approaches in Growth Curve Modeling: Using Student's t Distributions versus a Semiparametric Method. *Structural equation modeling*, *27*(4), 544–560.

Ypma, T. J. (1995). Historical Development of the Newton-Raphson Method. *SIAM review*, *37*(4), 531–551.