



## OPEN ACCESS

## EDITED BY

Han-Ying Liang,  
Tongji University, China

## REVIEWED BY

Steffen Zitzmann,  
University of Tübingen, Germany  
Irina Moustaki,  
London School of Economics and Political  
Science, United Kingdom

## \*CORRESPONDENCE

Karl Schweizer  
✉ k.schweizer@psych.uni-frankfurt.de

## SPECIALTY SECTION

This article was submitted to  
Statistics and Probability,  
a section of the journal  
Frontiers in Applied Mathematics and Statistics

RECEIVED 11 November 2022

ACCEPTED 13 March 2023

PUBLISHED 28 March 2023

## CITATION

Schweizer K, DiStefano C and French B (2023) A  
maximum likelihood approach for asymmetric  
non-normal data using a transformational  
measurement model.  
*Front. Appl. Math. Stat.* 9:1095769.  
doi: 10.3389/fams.2023.1095769

## COPYRIGHT

© 2023 Schweizer, DiStefano and French. This  
is an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# A maximum likelihood approach for asymmetric non-normal data using a transformational measurement model

Karl Schweizer <sup>1\*</sup>, Christine DiStefano<sup>2</sup> and Brian French<sup>3</sup>

<sup>1</sup>Institute of Psychology, Goethe University Frankfurt, Frankfurt, Germany, <sup>2</sup>Department of Educational Studies, University of South Carolina, Columbia, MO, United States, <sup>3</sup>College of Education, Washington State University, Pullman, WA, United States

A transformational measurement model for structural equation modeling (SEM) of asymmetric non-normal data is proposed. This measurement model aligns with the expectation-maximization (EM) algorithm of the maximum likelihood estimation (MLE) method, creating adaptability to data that deviate from normality. Distinctive properties of the connection of the measurement model and EM algorithm are maintenance of the normality assumption, which is at the core of EM algorithm, and applicability to asymmetric non-normality of observed data mediated by distortion coefficients. An evaluation using a mixture of normal and severely asymmetric non-normal data analyzed by MLE for asymmetric non-normal data (MLE for ASN) demonstrated efficiency of the model. Comparisons with robust DWLS and WLS yielded better fit results under MLE for ASN estimation.

## KEYWORDS

maximum likelihood estimation, confirmatory factor analysis, EM algorithm, DWLS, WLS

## Introduction

A popular method for parameter estimation in structural equation modeling (SEM) is the maximum likelihood estimation (MLE) method which was introduced in the late 1960's [1]. A major component of this estimation technique is the maximum likelihood function derived from the density function of a multivariate normal distribution. Recently, the long-standing popularity of this estimation method has been hampered, as the estimator has come under scrutiny because the validity of MLE outcomes is impaired when investigating non-normal data [2–4]. These outcomes suggest the restriction of MLE to investigations of data which closely follows a normal distribution.

Since much of the data investigated in applied situation are non-normally distributed, replacement and modification of MLE are options to explore. There are multiple ways to conduct alternative MLE including modified fitting functions [5, 6] and *post-hoc* corrections which take data characteristics into consideration (e.g., Satorra-Bentler method [7]). Furthermore, estimation methods based on a least-square approach are available [8, 9]. Recent studies compare two popular approaches: robust MLE with diagonally weighted least square estimation [10, 11]. These methods perform equally well in the presence of data commonly encountered by researchers such as: number of ordinal categories between four and ten, sample sizes between 200 and 1,000, and symmetric as well as asymmetric distributions. But, aside from the overall similarity in many situations, when MLE was applied to non-normal (asymmetrically distributed) data, parameter estimates by MLE yielded larger bias.

This essay reports on another possible modification of MLE for the purpose of improving its performance when analyzing data showing various degrees of non-normality. It takes the individual deviation of each observed variable into consideration and is not restricted to estimates of model fit. This modification is accomplished by a transformational measurement model that exerts influence on parameter estimation by use of the expectation-maximization (EM) algorithm [12, 13]. We characterize it as transformational because it is expected to shift the range of applicability from normally distributed data to asymmetric non-normal data.

In a way, the modification brought about by the transformational measurement model narrows the gap between MLE in factor analysis (including SEM) and outside of factor analysis where normality is rarely observed [14]. While in the context of factor analysis, MLE expects normally distributed data; however, in other applications MLE is recommended for dealing with non-normality [15]. An elaboration in line with this recommendation, for example, is an adaptation of MLE to the Rayleigh distribution [16].

For establishing the transformational measurement model as source of influence on EM algorithm, components of the model of the covariance matrix of MLE have to be related to corresponding components of the EM algorithm. This means re-parametrization of EM algorithm [17, 18]. In the course of re-parametrization it is necessary to assure that the performance of the algorithm is not impaired and that at the same time deviation from normality is taken into consideration.

Since non-normality as descriptive term applies to a variety of different distributions that may require different treatments, we restrict our research to a subset of non-normal distributions. The focus is on distributions that can be perceived as derived from a normal distribution by distortion (e.g., due to a scale restriction) in such a way that some properties are retained: the continuous scale, the existence of a major (i.e., global) peak as well as a mostly monotone increase up to this peak and a mostly monotone decrease following the peak. A major difference between original normal and non-normal data distributions is lack of symmetry. We characterize such distributions as *asymmetric non-normal*, abbreviated as *asn* (see Tables 1–3). The version of MLE described in this paper is optimized for investigating *asn* data.

The specification for asymmetric non-normal data is a characteristic that distinguishes the proposed version of MLE from other versions that are considered for non-normal data in general. Another characteristic is that parameters and fit statistics are estimated in the framework of a model reflecting the deviation of data from normality. This means that there is no *post-hoc* correction as found in other MLE versions. We refer to the proposed version of MLE as MLE for ASN.

The research reported in this paper was guided by two aims: (1) conceptualization of a version of MLE for asymmetric non-normal data within the framework of regular MLE for SEM by combining it with a transformational measurement model, and (2) demonstrating the efficiency of this version in investigating the structure of asymmetric non-normal data. MLE for ASN was also applied to asymmetric non-normal data with one-dimensional and two-dimensional underlying structures and

compared with weighted least squares and diagonally weighted least squares methods.

The theoretical argument shows the following structure: at first, the model of the covariance matrix of MLE [19] is related to corresponding components of EM algorithm [13]. Subsequently, changes are introduced to fit the model to asymmetric non-normal data. Finally, the transformational measurement model is extracted.

## The outset

We begin this investigation by considering the MLE fitting function,  $F$ , that is proposed for fitting the parameters of the theoretical model to data *via* the maximum likelihood criterion [1, 9, 15, 17]. This outset is selected since it provides information on what must be taken into consideration in order to demonstrate acceptable model-data fit. Let  $\mathbf{S}$  ( $\mathbf{S} \in \mathbb{R}^{p \times p}$ ) be the  $p \times p$  empirical covariance matrix,  $\mathbf{\Sigma}$  ( $\mathbf{\Sigma} \in \mathbb{R}^{p \times p}$ ) the  $p \times p$  model-implied covariance matrix that is specified by assigning value(s) to parameter(s),  $\vartheta$ , and  $F(\cdot)$  the MLE fitting function for model evaluation. MLE by definition refers to an iterative search for  $\vartheta$  which provides the best fit of  $\mathbf{\Sigma}$  to  $\mathbf{S}$  on the basis of

$$F[\mathbf{\Sigma}(\vartheta), \mathbf{S}] \quad (1)$$

Whereas  $\mathbf{S}$  is given and  $\mathbf{\Sigma}$  reflects the assumptions of the tested model,  $\vartheta$  is estimated while minimizing  $F$  in a succession of cycles of EM algorithm [12].

The model-implied covariance matrix,  $\mathbf{\Sigma}$  ( $\mathbf{\Sigma} \in \mathbb{R}^{p \times p}$ ), can show different degrees of complexity. For the purpose of the present study, we concentrate on the basic version with one latent variable for centered data associated with the congeneric measurement model [20] typically used with confirmatory factor analysis and structural equation modeling. This model includes two components that are expected to account for common and unique variation of data, respectively:

$$\mathbf{\Sigma} = \lambda\phi\lambda' + \Theta, \quad (2)$$

with  $p \times 1$  vector  $\lambda$  ( $\lambda \in \mathbb{R}^{p \times 1}$ ) of factor loadings on the latent variable of the corresponding measurement model,  $\xi$  [ $\xi \sim N(0, \sigma^*)$ ], variance parameter  $\varphi$  ( $\varphi \in \mathbb{R}^+$ ), and  $p \times p$  diagonal matrix of residual variances  $\Theta$  ( $\Theta \in \mathbb{R}^{p \times p}$ ). Since  $\lambda$  and  $\phi$  cannot be estimated at the same time, scaling is necessary [21]. This is frequently done by setting  $\phi = 1$ .

$\mathbf{\Sigma}$  can also be viewed as a matrix composed of variances and covariances. Since variances are the more complex entries, which include two components, it is sufficient to concentrate the discussion on these components. The variance of  $i$ th manifest variable,  $\sigma_{ii}$  ( $\sigma_{ii} \in \mathbb{R}^+$ ) ( $i = 1, \dots, p$ ), is defined as

$$\sigma_{ii} = \lambda_i\phi\lambda_i + \theta_i \quad (3)$$

where  $\lambda_i$  ( $\lambda_i \in \mathbb{R}$ ) is the factor loading of the  $i$ th manifest variable on latent variable  $\xi$ ,  $\phi$  ( $\phi \in \mathbb{R}^+$ ) the variance parameter and  $\theta_i$  ( $\theta_i \in \mathbb{R}^+$ ) the residual variance. In this model,  $\phi$  measures the dispersion of the latent variable that is common to all manifest variables and,  $\theta_i$ , the dispersion that is unique for the  $i$ th manifest variable.

In MLE the estimation of parameters is conducted by the EM algorithm. This method is a tool for the iterative estimation of parameters according to the maximum likelihood principle [22]. It is employed in situations such as missing information under the normal distribution assumption and mixture distributions [23]. The EM algorithm is considered important throughout statistics. Of importance for the present research is its application to factor analysis [13]. Employed as part of MLE, EM performs parameter estimation under the assumption of the normal distribution. Therefore, in the present study the consequences of deviations from normality reflected by the transformational measurement model for parameter estimation by EM algorithm need to be analyzed in some detail.

As the EM algorithm for factor analysis plays a major role in MLE for ASN, we proceed by recognizing the relationship between the parameters of Equation 3 and the parameters of EM algorithm. Furthermore, distributional assumptions need to be taken into consideration. EM-theory for factor analysis [13] assumes latent variables (i.e., factors) established by unobservable factor scores, which are symbolized as  $Z$  ( $Z \in \mathfrak{R}^{n \times p}$ ). These factor scores are assumed to follow a normal distribution. This means that latent variables can be considered as normal through factor scores. EM theory also considers observed scores that are symbolized by  $Y$  ( $Y \in \mathfrak{R}^{n \times q}$ ). In MLE they are also assumed to follow a normal distribution [1, 19].

Adaptation of the model described by Equation 3 to the notation of EM algorithm requires linking parameters  $\lambda$ ,  $\theta$  and  $\phi$  of  $\lambda$ ,  $\phi$  and  $\Theta$  to parameters of  $\beta$ ,  $\tau^2$  and  $R$  [13] in corresponding order. Matrix  $\beta$  is defined as the “regression coefficient matrix” that “is commonly called factor-loading matrix” (p. 70), and matrix  $\tau^2$  as matrix “of residual variances ... commonly called the uniquenesses” (p. 70). Furthermore, matrix  $R$  is defined as correlation matrix of the unobservable factor scores. We assume dimensionalities and data types for the EM matrices corresponding to the matrices of Equation 2. The symbol equivalent to  $\phi$  is  $R^*$  since in a one-factor model  $R$  reduces to a  $1 \times 1$  matrix, i.e., a scalar that is signified by the added star. It can be set free for estimation (see [13], Case 3). After a few EM cycles, this free parameter is likely to no more demonstrate the property of a (Pearson) correlation but may serve as estimate of common latent variation. Finally,  $\sigma_{ii}$  (Equation 3) is re-written using parameters of  $\beta$ ,  $\tau^2$  and  $R$  as

$$\sigma_{ii} = \beta_i R^* \beta_i + \tau_{ii}^2. \quad (4)$$

## The preparation of EM algorithm for data non-normality

In this section, the model for normally distributed data is slightly modified to make it suitable for investigating asymmetric non-normal data while the basic definitions and assumptions of EM algorithm except of the ones regarding  $Y$  are retained.

The investigation of asymmetric non-normal data requires the establishment of a relationship between normally distributed data and asymmetric non-normal data. Recall at this point that asymmetric non-normal data are assumed to originate from normally distributed data by distortion (see Introductory section). We consider such a relationship at the level of variances (and

covariances) since variances (and covariances) serve as input to SEM. We assume that the effect of distortion is reflected by distortion coefficient  $\alpha_i^2$  ( $\alpha_i \in \mathfrak{R}^+$ ) ( $i = 1, \dots, p$ ) linking the variance of the manifest variable representing asymmetric non-normal data [ $\sigma_{ii(\text{asn})}$ ] to the variance of the (corresponding) manifest variable representing normally distributed data [ $\sigma_{ii(n)}$ ]:

$$\sigma_{ii(\text{asn})} = \alpha_i^2 \sigma_{ii(n)}. \quad (5)$$

In the following several transformations are described that are necessary for performing parameter estimation that is in line with Equation 5 by available SEM software including EM algorithm. In the first step,  $\sigma_{ii(n)}$  is replaced by its constituents (see Equation 3) so that its common and residual components can be treated separately. In the next step, the distortion coefficient is associated with each one of the two components so that

$$\sigma_{ii(\text{asn})} = \alpha_i^2 (\lambda_i \phi \lambda_i)_{ii(n)} + \alpha_i^2 \theta_i. \quad (6)$$

In the following steps,  $\alpha_i^2$  and  $\theta_i$  are merged to give  $\theta_{i(\text{asn})}$  on one hand and on the other hand scalar  $\alpha_i^2$  ( $\alpha_i \in \mathfrak{R}^+$ ) is subdivided and used for multiplying the common component included in parentheses from the left-hand side and the right-hand side. Since  $\phi$  represents the dispersion of normally distributed latent variable  $\xi$ , this is signified by adding a subscript; it is written as  $\phi_{(n)}$ :

$$\sigma_{ii(\text{asn})} = \alpha_i (\lambda_i \phi_{(n)} \lambda_i) \alpha_i + \theta_{i(\text{asn})}. \quad (7)$$

Finally, we introduce an assumption that enables equal treatments of all manifest variables. It is assumed that the latent variable equally contributes to all manifest variables. This is a useful assumption for simulation studies and can also apply to empirical data under appropriate conditions. It allows setting  $\lambda_i = 1$  ( $i = 1, \dots, p$ ) so that the product of variance parameter and factor loadings,  $\lambda_i \phi \lambda_i$ , can be replaced by  $\phi_{\text{common}(n)}$  ( $= 1 \phi_{(n)} 1$ ) that is the same for all manifest variables ( $i = 1, \dots, p$ ):

$$\sigma_{ii(\text{asn})} = \alpha_i \phi_{\text{common}(n)} \alpha_i + \theta_{i(\text{asn})}. \quad (8)$$

This expression includes a parameter that reflects normality of the latent variable,  $\phi_{\text{common}(n)}$ , and distinguishes parts that relate to asymmetric non-normality,  $\alpha_i$ , or are influenced by it,  $\theta_{i(\text{asn})}$ . It implies re-scaling:  $\phi$  is set free while the replacement of the factor loading is fixed.

Following the final transformation we note the compatibility with EM algorithm. According to Case 3 of EM algorithm [13],  $\beta_i R^* \beta_i$  (see Equation 4) can be fixed while  $R^*$  be free. Therefore, we set  $\beta_i$  equal to  $\alpha_i$  (Equation 8) and  $R^*$  equal to  $\phi_{\text{common}(n)}$  (Equation 8). Furthermore, since the description of EM algorithm does not include a distributional assumption regarding  $\tau^2$ ,  $\tau^2$  is set equal to  $\theta_{i(\text{asn})}$ . Writing Equation 8 accordingly gives

$$\sigma_{ii(\text{asn})} = \beta_{(\text{EM, fixed})i} R_{(\text{EM, estimated})}^* \beta_{(\text{EM, fixed})i} + \tau_{(\text{EM, estimated})i}^2 \quad (9)$$

that is in line with Equation 4. Subscripts added; parentheses indicate a role in EM algorithm and whether the element is estimated or fixed.

Parameter estimation by EM algorithm occurs in cycles. Within each cycle  $R^*$  under the condition of Case 3 is estimated by

means of a regression coefficient,  $\delta$ , applied to the empirical covariance matrix. Here,  $\delta$  is determined by  $\beta$  as well as the previous estimate of  $R^*$  [13]. Furthermore, residual matrix,  $\Delta$ , with  $\beta$  and the previous estimate of  $R^*$  as components also exerts influence on revised  $R^*$  [13]. This means that the fixed  $\beta$  (i.e.,  $\alpha^2$ ) parameter upholds the relationship between what is expected to follow the normal distribution and what appears as distributional distortion.

## Quantification of deviation from normality

This section presents a description of quantifying distortion coefficients,  $\alpha_i^2$  ( $i = 1, \dots, p$ ) (see Equation 5). Distortion coefficients are assumed to reflect the effect of distorting normally distributed data regarding their variances. Since  $\sigma_{ii(n)}$  and  $\sigma_{ii(asn)}$  of Equation 5 are unknown,  $\alpha_i^2$  is not obtainable by rearranging the ingredients of Equation 5. Here, what is possible is using the possible proportionality of  $\alpha_i^2$  ( $i = 1, \dots, p$ ) and  $\sigma_{ii(asn)}$  ( $i = 1, \dots, p$ ). For this purpose all  $\alpha$ s and  $\sigma_{(asn)}$ s are arranged as  $p \times 1$  vectors. They show proportionality if  $\sigma_{ii(n)}$  ( $i = 1, \dots, p$ ) can be assumed to be constant:

$$\begin{bmatrix} \alpha_1^2 \\ \alpha_2^2 \\ \alpha_3^2 \\ \vdots \\ \alpha_p^2 \end{bmatrix} \propto \begin{bmatrix} \sigma_{11(asn)} \\ \sigma_{22(asn)} \\ \sigma_{33(asn)} \\ \vdots \\ \sigma_{pp(asn)} \end{bmatrix}. \quad (10)$$

The equality assumption of factor loadings leading to Equation 8 justifies Equation 10.

Although  $\sigma_{(asn)}$ s are unknown, they demonstrate a property that can be employed for arriving at values that approximate  $\alpha$ s. This property is that each  $\sigma_{ii(asn)}$  ( $i = 1, \dots, p$ ) is expected to reflect the observed variance of corresponding asymmetric non-normal data,  $s_{ii}$  ( $i = 1, \dots, p$ ) of  $\mathbf{S}$ . This expectation implies that  $\sigma_{ii(asn)}$  shows a size similar to the size of  $s_{ii}$ :

$$\sigma_{ii(asn)} \approx s_{ii}. \quad (11)$$

Together Equations 10 and 11 suggest a vector of variances obtained from distorted data for adjusting the statistical model to the effect of distortion. Furthermore, Equations 8 and 9 suggest the replacement of  $\alpha^2$ s by their square roots as well as  $s$ s by their square roots: let  $\hat{\alpha}$  ( $\hat{\alpha} \in \mathbb{R}^{p \times 1}$ ) be the  $p \times 1$  vector of observation-based distortion coefficients corresponding to the standard deviations observed in distorted data. Then

$$\hat{\alpha} = \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \vdots \\ \hat{\alpha}_p \end{bmatrix} = \begin{bmatrix} s_1^{1/2} \\ s_2^{1/2} \\ s_3^{1/2} \\ \vdots \\ s_p^{1/2} \end{bmatrix}. \quad (12)$$

The hat (^) on top of  $\alpha$  and  $\alpha$ s is added to signify that observations serve as source.

## The transformational measurement model for asymmetric non-normal data

Finally, the transformational measurement model is achieved by determining the measurement model associated with Equation 8 (i.e., finding the measurement model giving rise to Equation 8) with deviation coefficients according to Equation 12:

$$\mathbf{x} = \hat{\alpha}\xi + \delta \quad (13)$$

including  $p \times 1$  vector  $\mathbf{x}$  ( $\mathbf{x} \in \mathbb{R}^{p \times 1}$ ) of centered manifest variables,  $p \times 1$  vector  $\hat{\alpha}$  ( $\hat{\alpha} \in \mathbb{R}^{p \times 1}$ ) of distortion-reflecting fixations replacing fixed factor loadings (see Equation 12), latent variable  $\xi$  and  $p \times 1$  vector  $\delta$  ( $\delta \in \mathbb{R}^{p \times 1}$ ) of residual variables. This model is proposed for investigating the hypothesis that the systematic variation of data is due to one underlying dimension (i.e., structural validity) if data show asymmetric non-normality.

## Checking the compatibility with fitting function F

This section addresses the question whether fitting function F (Equation 1) is appropriate for asymmetric non-normal data. There is the original version of F derived from the density function of normally distributed data [1, 15]. Furthermore, there is an extended version that is implemented in current SEM software. It is an extended version that shows a special characteristic which broadens the field of application. It is defined as

$$F = \log |\Sigma| + \text{tr}(\mathbf{S}\Sigma^{-1}) - \log |\mathbf{S}| - p \quad (14)$$

where  $\Sigma$  ( $\Sigma \in \mathbb{R}^{p \times p}$ ) and  $\mathbf{S}$  ( $\mathbf{S} \in \mathbb{R}^{p \times p}$ ) represent the  $p \times p$  model-implied and empirical covariance matrices in corresponding order and positive integer  $p$  the number of manifest variables [9, 17].

The special characteristic of this version is that in the case of a correct model (including correct parameter estimates),  $\Sigma_c$ , and data (an empirical covariance matrix) free of random influences,  $\mathbf{S}_f$ :  $\log |\Sigma_c| = \log |\mathbf{S}_f|$ , so that their difference is zero. Furthermore, in this case:  $\mathbf{S}_f \times \Sigma_c^{-1} = \mathbf{I}$  and  $\text{tr}(\mathbf{I}) = p$ . Moreover, the difference of the trace of  $\mathbf{S}_f \times \Sigma_c^{-1}$  and  $p$  is also zero so that,

$$F[\Sigma_c, \mathbf{S}_f] = \log |\Sigma_c| + \text{tr}(\mathbf{S}_f \Sigma_c^{-1}) - \log |\mathbf{S}_f| - p = 0. \quad (15)$$

In sum, in the case a correct model and data with no error influence, F can be expected to arrive at a value of zero, and it is not necessary to make assumptions regarding the distribution of data (but, the matrices involved in the equation must allow for the computation of a determinant and inverse).

The presence of random influences may not invalidate the previous reasoning. In this case the model has to correctly account for the systematic variation of data so that only unsystematic variation due to random influences remains that is reflected by the outcome of F. If there is reason for assuming that deviations from expectations due to random influences follow normal distributions, the outcome can be considered as a  $\chi^2$  statistic, as is in investigating normally distributed data.

Although the fitting function, F, was developed for normally distributed data, its extended version appears to be applicable to a wider range of data types if the described characteristics

hold. This means that there is the possibility that  $\Sigma$  accounts for systematic variation of  $S$  computed from asymmetric non-normal data because of the adjustment by  $\hat{\alpha}$ . When there are also remainders due to random influences, the outcome of computing  $F$  may follow a  $\chi^2$  distribution.

It should be noted that the F-based  $\chi^2$  is currently only of importance as a component for the calculation of some fit indices. Furthermore, model fit may also be evaluated by fit indices that do not include  $\chi^2$  as a primary ingredient (e.g., SRMR).

## An evaluation using simulated data

The major aim of the evaluation was to demonstrate that the described measurement model in combination with MLE will improve performance in investigating the latent structure of data in the presence of asymmetric non-normality. This premise gave rise to two hypotheses: first, we hypothesized that when analyzing data illustrating asymmetric non-normality MLE for ASN would lead to better overall model fit compared to regular MLE. This hypothesis is based on the expectation that only MLE for ASN, but not regular MLE, could account for all systematic variation in distorted data. Second, we hypothesized that MLE for ASN would not provide an advantage over regular MLE in accounting for systematic variation unrelated to asymmetric non-normality. This hypothesis aimed at systematic variation that was neither reached by an incorrect model nor related to distributional deviation. MLE for ASN was not expected to account for this kind of systematic variation. Another aim was the comparison of MLE for ASN with two established estimation methods suggested to account to non-normality: robust DWLS and WLS.

To evaluate our claims, we conducted a simulation study. Structured random data with and without skewness manipulation were investigated by a measurement model corresponding to the model for data generation, that is a model assuming equal contributions of a common source to all items. For this purpose we generated normally distributed random data showing an underlying structure and manipulated their distribution for obtaining asymmetric non-normal data. In most investigations the structure of data was unidimensional. In additional investigations with incorrect (and corresponding correct) models the underlying structure was two-dimensional.

## Method

Data generation started with three sets of  $500 \times 20$  matrices of continuous and normally distributed data  $[N(0,1)]$  using a generation method described by [24]. Either one or two unrelated underlying dimensions characterized the matrices. The expected factor loading of the first set was 0.40, in the second set 0.35 and in the third set 0.30. We refer to these as data with *expected first-level factor loadings*, *expected second-level factor loadings* and *expected third-level factor loadings* in corresponding order. The expected residuals were set to the difference between 1 and the common variation of items (the squares of the corresponding expected factor loadings). Data matrices with two underlying dimensions were

realized as the combination of two subsets of columns. In this case, columns 1–10 of a matrix were associated with the first dimension and columns 11–20 with the second one.

When analyzing real data outside of the current research project, the typical observation was that the items' deviation from a symmetric distribution varied. We attempted to simulate this characteristic of real data by restricting deviation from symmetry to 10 of the 20 columns and generated different degrees of deviation. This transformation from symmetric to asymmetric was conducted according to the following formula:  $x_{asn} = (x \times k)^{1/2}/k$ , where  $x > 0$ . That is, that the data of a column were multiplied by integer  $k$  in the first step. Then the square root was computed and the result divided by  $k$ . The integer,  $k$ , was set to 4 in transforming two columns of a matrix, to 16 in another two, to 64 in further two, and also to 256 and 1,024 in two plus two more columns such that the amount of non-normality varied across columns.

The measurement model for the statistical investigation included one latent variable with fixed factor loadings (see Equation 13) whereas the variance parameter of the corresponding model-implied covariance matrix was set free for estimation, as also were the parameters included in the main diagonal of  $\Theta$  (see Equations 8, 9). The factor loadings were fixed according to Equation 12.

Parameter estimation was conducted by means of the MLE option in LISREL [25] (regular MLE and MLE for ASN). In addition to MLE, DWLS and WLS were employed for parameter estimation. The LISREL version of DWLS additionally performed robust estimation according to a Satorra-Bentler *post-hoc* correction method.

## Results

In the following the results of investigating model fit are reported for the different data conditions (data with no ASN and data with ASN) using regular MLE and MLE for ASN (first section), for correct and incorrect models (second section), and for different estimation methods (MLE for ASN, DWLS, and WLS) (third section).

### Effects of MLE versions in normal and non-normal data (hypothesis 1)

Fit results for normally distributed and asymmetric non-normal data investigated by regular MLE and MLE for ASN are included in Table 1.

The three sections of this table comprise fit results observed when investigating data with an expected factor loading of 0.40 (expected first-level factor loadings), of 0.35 (expected second-level factor loadings) and of 0.30 (expected third-level factor loadings).

The first row of the upper section includes the mean  $\chi^2$ , RMSEA, SRMR, NNFI, CFI, and AIC for the one-factor model with regular MLE applied to normally distributed data and the second row corresponding standard deviations (SD). Using commonly accepted guidelines (e.g.,  $RMSEA \leq 0.6$ ,  $SRMR \leq 0.8$ ,  $NNFI \geq 0.95$ , and  $CFI \geq 0.95$ ), all model fit indices indicated good

TABLE 1 Mean fit results and SDs (in italics) observed in investigating data showing (no) asymmetric non-normality (ASN) by one-factor models using regular MLE and MLE for ASN ( $N = 500$ ).

Characteristics of data (and estimation)		$\chi^2$	RMSEA	SRMR	NNFI	CFI	AIC
<b>Expected first-level factor loadings (0.40)</b>							
No ASN (MLE)	Mean	258.0	0.021	0.057	0.97	0.97	300.0
	SD	78.3	<i>0.018</i>	<i>0.016</i>	<i>0.03</i>	<i>0.03</i>	78.3
ASN (MLE)	Mean	300.3	0.034	0.069	0.95	0.95	342.3
	SD	33.0	<i>0.005</i>	<i>0.005</i>	<i>0.01</i>	<i>0.01</i>	33.0
ASN (MLE for ASN)	Mean	229.9	0.019	0.054	0.98	0.98	271.9
	SD	25.7	<i>0.008</i>	<i>0.005</i>	<i>0.01</i>	<i>0.01</i>	25.7
<b>Expected second-level factor loadings (0.35)</b>							
No ASN (MLE)	Mean	235.5	0.018	0.052	0.96	0.96	277.5
	SD	56.6	<i>0.015</i>	<i>0.011</i>	<i>0.04</i>	<i>0.04</i>	56.6
ASN (MLE)	Mean	262.7	0.027	0.059	0.94	0.94	304.7
	SD	27.9	<i>0.005</i>	<i>0.004</i>	<i>0.02</i>	<i>0.02</i>	27.9
ASN (MLE for ASN)	Mean	216.0	0.015	0.049	0.97	0.97	258.0
	SD	24.0	<i>0.008</i>	<i>0.004</i>	<i>0.02</i>	<i>0.02</i>	24.0
<b>Expected third-level factor loadings (0.30)</b>							
No ASN (MLE)	Mean	217.6	0.014	0.048	0.96	0.95	259.6
	SD	39.8	<i>0.012</i>	<i>0.007</i>	<i>0.06</i>	<i>0.05</i>	39.8
ASN (MLE)	Mean	234.7	0.021	0.052	0.93	0.93	276.7
	SD	24.2	<i>0.006</i>	<i>0.003</i>	<i>0.03</i>	<i>0.03</i>	24.2
ASN (MLE for ASN)	Mean	205.6	0.012	0.046	0.97	0.97	247.6
	SD	22.2	<i>0.008</i>	<i>0.003</i>	<i>0.03</i>	<i>0.03</i>	22.2

model fit [26]. The third row provides the mean results observed in investigating asymmetric non-normal data also using regular MLE and the fourth row corresponding SDs. All statistics showed a numeric impairment in comparison to values presented in the first row; however, only NNFI and CFI values were down to the cutoff for good model fit. The mean results obtained by MLE for ASN applied to asymmetric non-normal data are included in the fifth row and the corresponding SDs in the sixth rows. All fit indices obtained by MLE for ASN indicated good model fit (i.e., were lower respectively larger than the corresponding cutoffs).

The fit results reported in the second and third sections (expected second-level and third-level factor loadings) of Table 1 showed similar patterns as in the first one although numerically slightly worse degrees of model fit were signified. In these sections, asymmetric non-normal data led to the switch from good to acceptable model fit according to NNFI and CFI when there was no MLE for ASN. In contrast, MLE for ASN yielded good model fit.

In sum, investigations by regular MLE revealed the expected decrease in model fit in the presence of asymmetric non-normality while adjustment to asymmetric non-normality resulted in the return to the original degree of model fit (confirmation of hypothesis 1).

## Effects of MLE versions in unrelated systematic variation (hypothesis 2)

Fit results for correct and incorrect models used in investigating normally distributed and asymmetric non-normal data with regular MLE and MLE for ASN are reported in Table 2.

The first to fourth rows of this table include fit results for the correctly specified model and the remaining rows for the misspecified (i.e., incorrectly specified) model. The fit results for the correct model, but with no adaptation to asymmetric non-normality (first row), signified good model fit according to RMSEA and SRMR whereas NNFI and CFI illustrated model misfit. The report for the correct model with adaptation to asymmetric non-normality (third row) revealed improvement regarding NNFI and CFI to the level of acceptable model fit. All fit results for the incorrect models (see fifth to eighth rows) were numerically worse than the fit results for the correct models (see first to fourth rows). The impairment in model fit was very obvious in NNFI and CFI results. These values were far below of what was considered as acceptable. Furthermore, in misspecified models, there was virtually no fit improvement due to adaptation to asymmetric non-normality, as was suggested by the second hypothesis.

TABLE 2 Mean fit results and SDs (in italics) observed in investigating data showing (no) asymmetric non-normality (ASN) by one-factor and two-factor models using regular MLE and MLE for ASN ( $N = 500$ ).

Characteristics of model		$\chi^2$	RMSEA	SRMR	NNFI	CFI	AIC
Two factors, MLE	Mean	266.2	0.029	0.055	0.86	0.86	312.2
	SD	26.6	0.005	0.003	0.04	0.04	26.6
Two factors, MLE for ASN	Mean	236.8	0.022	0.051	0.91	0.91	282.8
	SD	25.2	0.007	0.003	0.04	0.04	25.2
One factor, MLE	Mean	534.3	0.06	0.07	0.59	0.59	576.3
	SD	71	0.006	0.005	0.05	0.05	71
One factor, MLE for ASN	Mean	524.8	0.059	0.069	0.62	0.62	566.8
	SD	72.5	0.006	0.005	0.05	0.05	72.5

TABLE 3 Mean fit results and SDs (in italics) observed for MLE for ASN, DWLS and WLS in investigating data showing asymmetric non-normality (ASN) ( $N = 500$ ).

Estimation method		NT $\chi^2$	SB $\chi^2$	RMSEA	SRMR	NNFI	CFI	AIC
Expected first-level factor loadings (0.40)								
MLE for ASN	Mean	229.4	–	0.019	0.054	0.98	0.98	271.4
	SD	26.1	–	0.008	0.004	0.01	0.01	26.1
DWLS	Mean	300.7	288.5	0.032	0.066	0.95	0.95	330.5
	SD	30.2	28.8	0.005	0.005	0.01	0.01	28.8
WLS	Mean	428.7 <sup>a</sup>	–	0.050	0.126	0.38	0.38	470.7
	SD	44.5	–	0.004	0.015	0.08	0.08	44.5
Expected second-level factor loadings (0.35)								
MLE for ASN	Mean	216.0		0.015	0.049	0.97	0.97	258.0
	SD	24.0		0.008	0.004	0.02	0.02	24.0
DWLS	Mean	262.6	255.7	0.026	0.057	0.94	0.94	297.7
	SD	26.4	26.0	0.005	0.004	0.02	0.02	26.0
WLS	Mean	400.0 <sup>a</sup>		0.047	0.101	0.40	0.40	442.0
	SD	45.1		0.005	0.012	0.09	0.09	45.1
Expected third-level factor loadings (0.30)								
MLE for ASN	Mean	205.4		0.011	0.046	0.97	0.97	247.4
	SD	22.9		0.008	0.003	0.04	0.03	22.9
DWLS	Mean	235.5	231.8	0.020	0.051	0.93	0.93	273.5
	SD	23.1	23.0	0.006	0.003	0.03	0.03	20.0
WLS	Mean	373.7 <sup>a</sup>		0.044	0.086	0.40	0.40	415.7
	SD	39.3		0.005	0.009	0.09	0.09	39.3

<sup>a</sup>LISREL reports minimum fit function  $\chi^2$  for WLS.

In sum, MLE for ASN did improve model fit when the model was correct. In contrast, in incorrect models there was virtually no improvement (results in line with hypothesis 2).

## Comparison of estimation methods

Finally, when analyzing asymmetric non-normal data, model fit was compared under MLE for ASN and alternative estimation

methods, DWLS and WLS. In order to establish comparability, the alternative estimation methods were applied in combination with the one-factor model showing factor loadings fixed to the value of one, i.e., factor loadings reflecting the assumption of equal contributions of the latent source to all manifest variables.

Table 3 provides the mean fit results.

Results presented in Table 3 follow the same structure as Table 1. There are three parts according to the expected factor loadings considered in data generation: expected factor loading

of 0.40 (expected first-level factor loadings), of 0.35 (expected second-level factor loadings) and of 0.30 (expected third-level factor loadings). In each part the first and second rows repeat fit results of Table 1 for MLE for ASN, followed by DWLS and WLS results in corresponding order.

In each part, the fit results observed in investigating the data by MLE for ASN signified the numerically best model fit, DWLS the second best fit and the one obtained by WLS, ranked third. Furthermore, we investigated the difference between the statistics using the reported SDs. In data characterized by expected first-level factor loadings the difference between MLE for ASN and DWLS was larger than two SDs for  $\chi^2$ , NNFI, CFI and AIC, and it was exactly two SDs for SRMR in data characterized by expected second-order factor loadings. WLS differed from MLE for ASN and from DWLS by more than two SDs in all comparison. Moreover, comparisons by CFI difference test [27] revealed substantial differences between all estimation methods for all data types.

## Discussion

The maximum likelihood estimation method for factor analysis is closely linked to the concept of a multivariate normal distribution. The density function of a multivariate normal distribution provided the outset for creating the maximum likelihood fitting function [1], and unobservable data as the core component of parameter estimation by EM algorithm are assumed to follow a normal distribution [13]. Furthermore, normality of the remainders of fitting  $\Sigma$  to  $S$  is necessary to justify the assumption that the output of the maximum likelihood fitting function follows a  $\chi^2$  distribution [19]. Therefore, it is no surprise that analyzing non-normal data by regular MLE may not yield expected fit results.

Although the primary output of MLE, the  $\chi^2$  statistic, is currently rarely used in evaluating model fit, many alternative fit indices that are preferred instead [26] include the  $\chi^2$  statistic as ingredient in their calculations. The fit indices with cutoffs used in evaluating model fit show different degrees of sensitivity for deviation from a normal distribution in the reported study. NNFI and CFI showed considerable sensitivity for deviation whereas RMSEA and SRMR indicated good model fit under all conditions. The replacement of MLE by the other estimation methods revealed that SRMR was also to some degree sensitive to deviation from normality. Although RMSEA showed numeric impairment due to deviation from normality, there was no substantial difference according to the corresponding cutoff [27]. The mean RMSEA always signified good model fit, even in incorrect models.

The modification of MLE by using the transformational measurement model for the investigation of asymmetric non-normal data led to the expected improvement in model fit over regular MLE. When analyzing data in a variety of data types, the CFI difference [27] signified fit improvement. A substantial improvement was also observed by the RMSEA difference [27] in data with expected first-level factor loadings. Both MLE versions discriminated virtually equally well between correct and incorrect models. But there was no improvement due to MLE for ASN if systematic variation that was not otherwise reached was unrelated to distributional deviation.

To arrive at general statements regarding the performances of the estimation methods, we base the comparisons among these methods on the means reported in Tables 1–3 for one-factor models regarding the cutoffs for good model-data fit (RMSEA  $\leq$  0.6, SRMR  $\leq$  0.8, NNFI  $\geq$  0.95, and CFI  $\geq$  0.95). The percentages of reaching these cutoffs were 100 for MLE ASN, 66.6 for regular MLE, 66.6 for DWLS and 25 for WLS. The similar performances for regular MLE and DWLS are overall in line with results reported for these methods observed in investigating data showing less strong deviation from normality [10, 11]. The better performance of MLE for ASN is presumably restricted to data showing considerable deviation from normality. It is likely to do less well in overall normally distributed data.

An important but restricting assumption of MLE for ASN is the assumption of equal contributions of the latent source to all responses that justifies setting the factor loadings to equal sizes. The assumption of one latent variable as the source of responding is characteristic of customary confirmatory factor analysis [20] and the version of EM algorithm for confirmatory factor analysis [13]. But, in applied research, equal contributions of the latent source are rarely assumed so that factor loadings are mostly free for estimation. In investigating simulated data with expected factor loadings of the same size for all manifest variables, fixed factor loadings do not mean a disadvantage in comparison to free factor loadings [28]. Under appropriate conditions it may also apply for empirical data. Nevertheless, a version allowing for free factor loadings is desirable that, however, would require a more basic change of EM algorithm.

One limitation of the present study is the restriction to data distributions derived from a normal distribution by distortion. We concentrated on such asymmetric non-normal data because of our experiences with empirical research. The reason was that we observed that problems in reaching good model fit were frequently associated with data showing asymmetric non-normality. Another limitation is that distributions related to the normal distribution like beta, binomial and gamma distributions [29] are not taken into consideration. A further limitation is that we concentrated on model fit and omitted investigating bias in estimates. Future studies may include different simulation conditions to examine the viability of MLE for ASN under additional distribution and model conditions while including a variety of outcomes (e.g., parameter bias).

In summary, we present an alternative version of the maximum likelihood estimation (MLE) method for confirmatory factor analysis which includes a transformational measurement model. This transformational measurement model enables the investigation of the latent structure of asymmetric non-normal data. Furthermore, a simulation study is reported that demonstrates the efficiency of the method. An advantage in investigating asymmetric non-normal data over otherwise efficient robust DWLS and WLS is made apparent. We look forward to future research studies addressing these strengths and limitations of MLE for ASN.

## Conclusion

The transformational measurement model opens up maximum likelihood estimation to data that deviate from a normal



distribution that is otherwise a precondition for maximum likelihood estimation in factor analysis.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

KS conceptualized the study, performed the computations, and contributed to the writing. CD and BF shared ideas and contributed substantially to the writing. All authors contributed to the article and approved the submitted version.

## References

- Jöreskog KG. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*. (1969) 34:183–202. doi: 10.1007/BF02289343
- Jobst LJ, Auerswald M, Moshagen M. The effect of latent and error non-normality on measures of fit in structural equation modeling. *Educ Psychol Meas*. (2021) 82:911–37. doi: 10.1177/00131644211046201
- Lai K. Estimating standardized SEM parameters given non-normal data and incorrect model: methods and comparisons. *Struct Equation Model*. (2018) 25:1–21. doi: 10.1080/10705511.2017.1392248
- West SG, Finch JE, Curran PJ. Structural equation models with non-normal variables: problems and remedies. In: Hoyle R, editor. *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage (1995). p. 56–75.
- Yu C, Muthén B. Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. In: *Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans)* (2002).
- Deng L, Yang M, Marcoulides KM. Structural equation modeling with many variables: a systematic review of issues and developments. *Front Psychol*. (2018) 9:1–14. doi: 10.3389/fpsyg.2018.00580
- Satorra A, and Bentler PM. Corrections to test statistics and standard errors in covariance structure analysis. In: von Eye A, Clogg CC, editors. *Latent Variables Analysis: Applications for Developmental Research*. Thousand Oaks: Sage (1994). p. 399–419.
- Arruda EH, Bentler PM, A. Regularized GLS for structural equation modeling. *Struct Equation Model*. (2017) 24:657–65. doi: 10.1080/10705511.2017.1318392
- Browne MW. Asymptotically distribution-free methods for the analysis of covariance structures. *Br J Math Stat Psychol*. (1984) 37:62–83. doi: 10.1111/j.2044-8317.1984.tb00789.x
- Li CH. Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behav Res Methods*. (2016) 48:936–49. doi: 10.3758/s13428-015-0619-7
- Li CH. Statistical estimation of structural equation models with a mixture of continuous and categorical observed variables. *Behav Res Methods*. (2021) 53:2191–213. doi: 10.3758/s13428-021-01547-z
- Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via the EM algorithm. *J Roy Stat Soc B*. (1977) 39:1–38. doi: 10.1111/j.2517-6161.1977.tb01600.x
- Rubin DB, Thayer DT. EM algorithms for ML factor analysis. *Psychometrika*. (1982) 47:69–76. doi: 10.1007/BF02293851
- Micceri T. The unicorn, the normal curve, and other improbable creatures. *Psychol Bull*. (1989) 105:156–66. doi: 10.1037/0033-2909.105.1.156
- Myung JJ. Tutorial on maximum likelihood estimation. *J Math Psychol*. (2003) 47:90–100. doi: 10.1016/S0022-2496(02)00028-7
- Lalitha S, Mishra A. Modified maximum likelihood estimation for Rayleigh distribution. *Commun Stat A Theor*. (2011) 25:389–401. doi: 10.1080/03610929608831702
- Gelman A, Hill J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York, NY: Cambridge University Press (2007).
- Hecht M, Gische C, Vogel D, Zitzmann S. Integrating out nuisance parameters for computationally more efficient Bayesian estimation—an illustration and tutorial. *Struct Equation Model*. (2020) 27:483–93. doi: 10.1080/10705511.2019.1647432
- Jöreskog KG. A general method for analysis of covariance structure. *Biometrika*. (1970) 57:239–57. doi: 10.2307/2334833
- Graham JM. Congeneric and (essentially) tau-equivalent estimates of score reliability. *Educ Psychol Meas*. (2006) 66:930–44. doi: 10.1177/0013164406288165
- Schweizer K, Troche S, DiStefano C. Scaling the variance of a latent variable while assuring constancy of the model. *Front Psychol*. (2019) 10:887. doi: 10.3389/fpsyg.2019.00887
- McLachlan GJ. *Computation: Expectation-Maximization Algorithm. International Encyclopedia of the Social and Behavioral Sciences (Second Edition)*. Amsterdam: Elsevier. (2016).
- Gumedze FN, Dunne TT. Parameter estimation and inference in the linear mixed model. *Linear Algebra Appl*. (2011) 435:1920–44. doi: 10.1016/j.laa.2011.04.015
- Jöreskog KG, Sörbom D. *Interactive LISREL: User's Guide*. Lincolnwood, IL: Scientific Software International Inc. (2001).
- Jöreskog KG, Sörbom D. *LISREL 8.80*. Lincolnwood, IL: Scientific Software International Inc (2006).
- DiStefano C. Examining fit with structural equation models. In: Schweizer K, DiStefano C, editors. *Principles and Methods of Test Construction*. Göttingen: Hogrefe Publishing (2016). p. 166–193.
- Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct Equation Model*. (2002) 9:233–55. doi: 10.1207/S15328007SEM0902\_5
- Schweizer K, Ren X, Wang T, Zeller F. Does the constraint of factor loadings impair model fit and accuracy in parameter estimation? *Int J Stat Prob*. (2015) 4:40–50. doi: 10.5539/ijsp.v4n4p40
- Leemis LM, McQueston JT. Univariate distribution relationships. *Am Stat*. (2008) 62:45–53. doi: 10.1198/000313008X270448

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.