*Appl. Math. J. Chinese Univ.*
2024, 39(1): 24-33

# Identifying the best common factor model via exploratory eactor analysis

HE Bao-hua[1,2] TANG Rui[3] TAGN Qi-yi[4]

**Abstract**. Currently, there is no solid criterion for judging the quality of the estimators in factor analysis. This paper presents a new evaluation method for exploratory factor analysis that pinpoints an appropriate number of factors along with the best method for factor extraction. The proposed technique consists of two steps: testing the normality of the residuals from the fitted model via the Shapiro-Wilk test and using an empirical quantified index to judge the quality of the factor model. Examples are presented to demonstrate how the method is implemented and to verify its effectiveness.

## §1 Introduction

Exploratory factor analysis (EFA) is a collection of procedures for estimating the common factor, usually determined by the method of parameter estimation and the number of factors to be retained. For decades, various criteria have been proposed for determining the number of factors, such as the eigenvalues of the correlation matrix [3,4], $k_G$ of the reduced correlation matrix[8], the measure of sampling adequacy (MSA)[10], the parallel analysis procedure [18], and some other new sophisticated methods (see [11] and the references therein). Regarding the technique for extracting the common factor, some commonly used methods are principal component method (PCM), principal factor method (PFM, with SMC), maximum likelihood (ML), unweighted least squares (ULS), weighted (generalized) least squares (GLS), and alpha-factor analysis (AF). Unfortunately, the criteria for judging the quality of a common factor have not been well defined. That is, these criteria typically focus on the data characteristic rather than on the statistical test used to assess the model. Furthermore, as [9] explains, there does not seem to exist a single solution that makes the best "sense".

Received: 2017-04-22. Revised: 2019-04-15.
MR Subject Classification: 62H25.
Keywords: factor analysis, Shapiro-Wilk, normality, residuals.
Digital Object Identifier(DOI): https://doi.org/10.1007/s11766-024-3544-7.
Supported by the National Basic Research Program of China(2010CB126200) and the National Natural Science Foundation of China(30370914).

To evaluate the model fit, [2] introduced the chi-squared statistic

$$\chi_k^2 = \Big(n - \frac{2p+4k+11}{6}\Big) \ln \frac{|\hat{\Sigma}|}{|\mathbf{S}|}, \tag{1}$$

where $n$ is the number of samples, $p$ is the number of variables, $k \leq p$ is the number of common factors, $\hat{\Sigma}$ is the fitted covariance matrix, and $\mathbf{S}$ is the sample covariance matrix. However, this statistic has two drawbacks: it can only be applied via the ML method and it tends to overestimate the number of factors [16]. As an example in [6], Table 1 shows 4 factors can barely be extracted by using this test as the p-value = 0.0158 is scarcely bigger than 1% ( but less than 5%). Using our proposed criteria, we will show that the desired number of factors should be 3 and that the ML approach is inappropriate.

Table 1. Number of factors that should be extracted per Bartlett's (1954) goodness-of-fit statistic, as shown in [10].

| k | $\chi_k^2$ | df | P-value |
|---|---|---|---|
| 2 | 219.27 | 13 | $< 0.0001$ |
| 3 | 25.60 | 7 | 0.0006 |
| 4 | 8.29 | 2 | 0.0158 |

One alternative method is the root mean square residual (RMS) index [1].

$$RMS = \sqrt{\frac{\sum_{i=2}^{p}\sum_{j=1}^{i-1}\epsilon_{ij}^2}{p(p-1)}}, \tag{2}$$

where the numerator is the sum of the squared residuals in the upper triangle, excluding diagonal elements. In contrast to Bartlett's (1954) statistic, the RMS statistic can be used to test the "consistent" number of factors [19].

Thus, residual analysis is of great importance and is used in many procedures designed to detect various types of disagreement between data and an assumed model[5]. Practice dictates that in EFA, if the residuals are normally distributed, then the factor model is acceptable.

In this paper, we use the Shapiro-Wilk test [15,17] to examine the normality of the residuals. We aim to examine whether there is any systematic bias in the residuals of the fitted model, rather than the general fit index of the magnitude of the residuals between the simulated and the original correlation matrices. Based on the Shapiro-Wilk test and the RMS statistic, we propose an index Q for judging the quality of the factor model. Examples are provided to show the effectiveness and practicality of the criteria as applied to factor analysis.

## §2 Goodness of Fit

Without loss of generality, $\Sigma = LL^{'} + \Psi$, where $L$ is the matrix of factor loadings and $\Psi$ is the diagonal matrix of errors, or specific variance. If the reproduced matrix $LL^{'}$ is equal to $\Sigma$, the residual matrix should be zero; however, the reality is that $||\Sigma - LL^{'}|| > 0$. If $\hat{\Sigma} = \hat{L}\hat{L}^{'} + \hat{\Psi}$

is a good approximation of $\Sigma$, then $RMS$ is close to 0, and the residuals, $\Sigma - \hat{\Sigma}$, should be approximately normal.

The Shapiro-Wilk test, as a normality test of residuals has been proved a very useful routine tool for fitting the goodness of $\Psi$. Let $\epsilon$ denote the $p^2$-vector form of the residual matrix $\Psi$. Then the Shapiro-Wilk test statistic is given by

$$W = \frac{\left(\sum_{\ell=1}^{p^2} a_\ell \epsilon_{(\ell)}\right)^2}{\sum_{\ell=1}^{p^2} (\epsilon_\ell - \bar{\epsilon})^2}, \tag{3}$$

where $\epsilon_{(\ell)}$ is the $\ell^{th}$ order statistic and $\bar{\epsilon}$ is the sample mean of the residuals. The constants $a_\ell$ are determined by

$$(a_1, a_2, \ldots, a_n) = \frac{m^T V^{-1}}{\sqrt{m^T V^{-1} V^{-1} m}}, \tag{4}$$

where $m = (m_1, m_2, \ldots, m_n)^T$ is the vector of expected values of standard normal order statistics and $V$ is the corresponding covariance matrix.

In order to use the Shapiro-Wilk test for normality, first compute $\hat{\Sigma} - \Sigma$. Let $r_{ij}$ and $\sigma_{ij}$ denote the elements of the matrices $\hat{\Sigma}$ and $\Sigma$, respectively. Then we apply a similar transformation as [7] by converting the elements of the covariance matrices via

$$r'_{ij} = \frac{1}{2} \ln\left(\frac{1 + r_{ij}}{1 - r_{ij}}\right) \quad \text{and} \quad \sigma'_{ij} = \frac{1}{2} \ln\left(\frac{1 + \sigma_{ij}}{1 - \sigma_{ij}}\right) \tag{5}$$

which then yield "more normal" residuals per

$$\epsilon'_{ij} = r'_{ij} - \sigma'_{ij}, \quad i < j \leq p. \tag{6}$$

After testing for normality, the next step is to evaluate the accuracy of the model. Based on the Shapiro-Wilk statistic, $W$, we propose the combination index, $Q$, for judging the quality of the factor model. Define the index as

$$Q = W(1 - k \cdot RMS)^2. \tag{7}$$

Note that $Q$ is composed of two parts: the $W$ statistic measures the normality of the residuals and the term $1 - k \cdot RMS$ measures the accuracy of the model. The closer to normality the data are, the smaller the $W$ statistic will be. That is, the test rejects the hypothesis of normality for large $W$. Thus, if the residuals satisfy the normality assumption and since the root mean square is typically much less than $1/k$, the $Q$-index is usually less than one. Furthermore, the greater the value of $Q$, the better the model fits the data. This follows because as the fit of the model improves, the RMS term converges to zero, which implies that $Q$ approaches $W$.

## §3 Monte Carlo Simulation

In the simulation study, we generate 1000 data sets with a three typical distributions:
1) $x \sim N_6(0_6, \Sigma)$. It's Multivariate normal distribution $N_6(0_6, \Sigma)$.
2) $x \sim Mt_6(0_6, \Sigma)$. It's Multivariate t-distribution $Mt_6(0_6, \Sigma)$
3) $x \sim GAMMA(0_6, \Sigma)$ It's Centerized Gamma distribution, i.e. Gamma($\Sigma$)-E(Gamma($\Sigma$))
where $0_q$ is a q-dimensional zero vector and $\Sigma$ is the sample correlation covariance matrix. Since

$\Sigma$ is symmetric, let $i \leq j$ and

$$a = \begin{cases} 1 \quad if\, i = j \\ 0.75 - 0.1|i-j| \quad if\, i \neq j \end{cases} (i, j = 1, .., 6)$$

Not surprisingly, the distributions in Normals(e.g.Gaussian and t-distribution, see Table 2 and Table 3) do not need Fishers z transformation below:

Table 2. the results of the residual analysis(Gaussian distribution).

| | PCM | | | MLM | | | PFM | | |
|---|---|---|---|---|---|---|---|---|---|
| m | W | P | Q | W | P | Q | W | P | Q |
| 2 | 0.9484 | 0.4993 | 0.6856 | 0.8890 | 0.0648 | 0.8397 | 0.8627 | 0.0264 | 0.7650 |
| 3 | 0.9396 | 0.3779 | 0.5910 | **0.9241** | **0.2224** | **0.9986** | 0.8627 | 0.0264 | 0.7184 |
| 4 | 0.9386 | 0.3646 | 0.5376 | 0.8466 | 0.0155 | 0.6787 | 0.8627 | 0.0264 | 0.6732 |

Table 3. The results of the residual analysis(t-distribution).

| | PCM | | | MLM | | | PFM | | |
|---|---|---|---|---|---|---|---|---|---|
| m | W | P | Q | W | P | Q | W | P | Q |
| 2 | 0.9755 | 0.9299 | 0.7521 | 0.6674 | 0.0001 | 0.9568 | 0.9776 | 0.9501 | 0.9276 |
| 3 | 0.9318 | 0.2877 | 0.8247 | **0.8781** | **0.4445** | **0.9991** | 0.9776 | 0.9501 | 0.9030 |
| 4 | 0.8641 | 0.0277 | 0.8018 | 0.6285 | 0.0001 | 0.9799 | 0.8633 | 0.7493 | 0.8894 |

However, Non-normal(e.g.Gamma distribution) do need Fishers z transformation: From Table 4, we can see none of methods meet the normality (in the sense of significance level 5%):

Table 4. the results of the residual analysis(before Fishers z transfor-mation).

| | PCM | | | MLM | | | PFM | | |
|---|---|---|---|---|---|---|---|---|---|
| m | W | P | Q | W | P | Q | W | P | Q |
| 2 | 0.4512 | <0.0001 | 0.3145 | 0.6417 | <0.0001 | 0.6514 | 0.6172 | 0.0021 | 0.4789 |
| 3 | **0.9349** | **0.0352** | **0.8460** | 0.4526 | 0.0001 | 0.3992 | 0.6658 | <0.0001 | 0.5752 |
| 4 | 0.8202 | 0.0132 | 0.5523 | 0.7243 | <0.0001 | 0.6879 | 0.9345 | <0.0001 | 0.6486 |

Table 4 is the results of the residual analysis before Fishers z transfor-mation. We can see none of methods meet the normality, as all p-values are less than 0.05, so we do the Fishers z transformation and then the Shapiro-Wilks test. The W and Q values are below.

Finally, from all the tables above we can confirm that k=3 is an optimal choice.However, by comparing the other methods, we find the methods of PCM perform well for distributions in Normals, while MLM for Non- normals .

Table 5. The results of the residual analysis(After Fishers z transformation).

| | PCM | | | MLM | | | PFM | | |
|---|---|---|---|---|---|---|---|---|---|
| m | W | P | Q | W | P | Q | W | P | Q |
| 2 | 0.9147 | <0.0001 | 0.6149 | 0.9846 | 0.0033 | 0.4267 | 0.6139 | <0.0001 | 0.5124 |
| 3 | **0.9528** | **0.1277** | **0.8247** | 0.4623 | <0.0001 | 0.4076 | 0.7095 | <0.0001 | 0.6111 |
| 4 | 0.7432 | 0.0001 | 0.4561 | 0.5412 | <0.0001 | 0.4165 | 0.9714 | 0.0001 | 0.6937 |

# §4 Examples

## 4.1 Simulated Population Correlation Matrix

To illustrate the method, we begin with an artificial example presented by [13]. The example uses a modified version of simulation methods developed by [20]. The correlations among 12 measured variables were constructed as arising from effects of three major common factors and 50 minor factors simulated to represent a type of model error. The data generating parameters are shown in Table 6.

Table 6. Data generating parameters for simulated population correlation matrix as provided by [18].

| | Major Domain Factors | | | | |
|---|---|---|---|---|---|
| Variable | 1 | 2 | 3 | Uniqueness | Minor Variance |
| 1 | .95 | 0 | 0 | .000 | .098 |
| 2 | .95 | 0 | 0 | .000 | .098 |
| 3 | .95 | 0 | 0 | .000 | .098 |
| 4 | .95 | 0 | 0 | .000 | .098 |
| 5 | .95 | 0 | 0 | .000 | .098 |
| 6 | 0 | .70 | 0 | .413 | .098 |
| 7 | 0 | .70 | 0 | .413 | .098 |
| 8 | 0 | .70 | 0 | .413 | .098 |
| 9 | 0 | .70 | 0 | .413 | .098 |
| 10 | 0 | 0 | .50 | .653 | .098 |
| 11 | 0 | 0 | .50 | .653 | .098 |
| 12 | 0 | 0 | .50 | .653 | .098 |

The three major domain factors are clearly of unequal strength. Under the simulation method, variance in each measured variable not accounted for by these three major factors was attributed partly to a unique factor for each measured variable with the remainder attributable to effects of the 50 minor factors. Using the parameters in Table 6, [13] generated the population correlation matrix in Table 7.

We determine the Shapiro-Wilk test statistic for four different factor estimation methods, namely PCM, PFM (with SMC), ULS, and ML. The test statistics with p-values and the

Table 7. Simulated Population Correlation Matrix per [13] based on parameters in Table 6.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | **1.00** | | | | | | | | | | | |
| 2 | **.94** | **1.00** | | | | | | | | | | |
| 3 | **.87** | **.88** | **1.00** | | | | | | | | | |
| 4 | **.89** | **.90** | **.95** | **1.00** | | | | | | | | |
| 5 | **.96** | **.94** | **.89** | **.86** | **1.00** | | | | | | | |
| 6 | -.01 | -.01 | .06 | .08 | -.04 | **1.00** | | | | | | |
| 7 | -.06 | -.06 | .06 | .03 | -.06 | **.53** | **1.00** | | | | | |
| 8 | .00 | -.06 | -.04 | .02 | -.06 | **.49** | **.52** | **1.00** | | | | |
| 9 | -.01 | .05 | -.02 | -.06 | .04 | **.45** | **.45** | **.42** | **1.00** | | | |
| 10 | .06 | .07 | -.02 | .02 | .05 | .02 | -.06 | -.04 | .02 | **1.00** | | |
| 11 | .04 | .05 | -.05 | -.06 | .07 | -.05 | -.05 | -.04 | .05 | **.29** | **1.00** | |
| 12 | .01 | -.06 | -.02 | -.07 | .04 | -.06 | .00 | .02 | .00 | **.21** | **.27** | **1.00** |

corresponding Q index values are shown in Table 8.

Table 8. Shapiro-Wilk statistic for 4 factor estimation methods, with corresponding p-values and Q indices. Only PFM and ULS satisfy the normality assumption. Note that they also have the largest emprircal quantified index values.

| | W | P-value | Q |
|---|---|---|---|
| PCM | 0.8307 | $< 0.0001$ | 0.5135 |
| PFM | 0.9844 | 0.5766 | 0.8224 |
| ULS | 0.9790 | 0.3249 | 0.8311 |
| ML | 0.5987 | $< 0.0001$ | 0.4193 |

It is apparent that the normality assumption is only satisfied by PFM and ULS. Furthermore, as depicted in Table 9, PFM and ULS estimate the true major domain factors the best, whereas PCM somewhat overestimates factors 2 and 3 and the maximum likelihood method performs poorly when extracting the third factor. This observation agrees with the result given by [13].

Hence, this example shows that the best extraction methods were the ones whose residuals were normally distributed, as shown via Shapiro-Wilk hypothesis tests. This corresponds to the methods with the largest empirical quantified index values.

## 4.2 Semantic Differential Rating of Words

In a study conducted by [6], semantic differential ratings were obtained based on the following 8 scales: (1) friendly/unfriendly, (2) good/bad, (3) nice/awful, (4) brave/not brave, (5) big/little, (6) strong/weak, (7) moving/still, and (8) fast/slow. They measured the connotative meaning of 487 words as provided by fifth-grade students and studied the mean semantic differential ratings. For one list of 292 words, the intercorrelations among the ratings are provided

Table 9. Factor loadings obtained for the 4 extraction methods obtained via oblique rotation. PFM and ULS retrieve the major factor loading fairly well, whereas PCM slightly overestimates the loadings for factors 2 and 3 and ML poorly estimates the loadings for factor 3.

| | PCM | | | PFM | | | ULS | | | ML | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| $X_1$ | **0.97** | -0.02 | 0.05 | **0.96** | -0.02 | 0.08 | **0.96** | -0.02 | 0.07 | **0.96** | -0.02 | 0.15 |
| $X_2$ | **0.97** | -0.02 | 0.03 | **0.96** | -0.02 | 0.06 | **0.96** | -0.03 | 0.05 | **0.95** | -0.02 | 0.09 |
| $X_3$ | **0.95** | 0.03 | -0.06 | **0.95** | 0.03 | -0.10 | **0.94** | 0.03 | -0.09 | **0.96** | 0.03 | -0.10 |
| $X_4$ | **0.96** | 0.03 | -0.08 | **0.95** | 0.03 | -0.13 | **0.95** | 0.03 | -0.11 | **0.97** | 0.01 | **-0.25** |
| $X_5$ | **0.97** | -0.04 | 0.08 | **0.96** | -0.04 | 0.13 | **0.96** | -0.04 | 0.11 | **0.96** | -0.02 | **0.28** |
| $X_6$ | 0.02 | **0.80** | -0.05 | 0.02 | **0.69** | -0.06 | 0.02 | **0.72** | -0.05 | 0.03 | **0.71** | -0.18 |
| $X_7$ | -0.02 | **0.81** | -0.06 | -0.01 | **0.71** | -0.08 | -0.01 | **0.74** | -0.07 | -0.01 | **0.74** | -0.12 |
| $X_8$ | -0.03 | **0.78** | -0.02 | -0.03 | **0.69** | -0.03 | -0.03 | **0.69** | -0.03 | -0.02 | **0.67** | -0.11 |
| $X_9$ | 0.00 | **0.73** | 0.10 | 0.00 | **0.65** | 0.11 | 0.00 | **0.62** | 0.09 | -0.01 | **0.67** | 0.23 |
| $X_{10}$ | 0.04 | -0.01 | **0.69** | 0.03 | -0.01 | **0.44** | 0.03 | -0.01 | **0.45** | 0.04 | -0.02 | 0.06 |
| $X_{11}$ | 0.01 | -0.02 | **0.76** | 0.01 | -0.02 | **0.52** | 0.01 | -0.02 | **0.63** | 0.00 | -0.02 | **0.25** |
| $X_{12}$ | -0.03 | 0.00 | **0.67** | -0.03 | 0.00 | **0.45** | -0.02 | 0.00 | **0.42** | -0.02 | 0.00 | 0.21 |

in Table 10.

Table 10. Correlation matrix of mean semantic differential ratings for a list of 292 words based on 8 scales.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1.00 | | | | | | | |
| 2 | .95 | 1.00 | | | | | | |
| 3 | .96 | .98 | 1.00 | | | | | |
| 4 | .68 | .70 | .68 | 1.00 | | | | |
| 5 | .33 | .35 | .31 | .52 | 1.00 | | | |
| 6 | .60 | .63 | .61 | .79 | .61 | 1.00 | | |
| 7 | .21 | .19 | .19 | .43 | .31 | .42 | 1.00 | |
| 8 | .30 | .31 | .31 | .57 | .29 | .57 | .68 | 1.00 |

[19] points out that the reduced correlation matrix ($k_G = 4$) is dominated by one root and that, while Kaiser's[10] measure of sampling adequacy suggest 2 factors is sufficient, despite small residuals, the 2-factor solution is statistically a poor fit. Furthermore, they note that the 3-factor solution appears to result in a sing factor that tries to explain more than 100% of the common variance. Thus, they declare that the data set is inappropriate for factor analysis.

To test their assertion, we investigated the normality of the residuals and the $Q$ values for the same methods used in the previous example.

Per Table 11, only one scenario met the normality assumption at the 5% level of significance. It should also be observed that the 3-factor solution obtained via ULS also has the largest value for $Q$. Therefore, ULS with three factors yields the best solution.

Applying ULS and extracting 3 factors, we get the factor loading matrix and specific variance

Table 11. Results of Shapiro-Wilk test on the fitted model of the semantic attributes. The asterisk denotes the normality assumption is satisfied ($p > 0.05$). The 3-factor solution via ULS appears to be the best model since it has the largest $Q$-index value.

| | PCM | | PFM | | ULS | | ML | |
|---|---|---|---|---|---|---|---|---|
| | W | Q | W | Q | W | Q | W | Q |
| 2 | 0.9023 | 0.6504 | 0.9061 | 0.7273 | 0.8821 | 0.7103 | 0.6170 | 0.4769 |
| 3 | 0.8802 | 0.6574 | 0.8510 | 0.7428 | 0.9822* | 0.9279 | 0.8084 | 0.7528 |
| 4 | 0.8529 | 0.6656 | 0.8510 | 0.7084 | 0.6857 | 0.6642 | 0.8552 | 0.8350 |

as shown in Table 12.

Table 12. Loading matrix of the semantic attributes data using oblique rotation. Most of the variability is explained by the 3 factors. Factors 1, 2 and 3 appear to represent a person's level of kindness, courage, and movement, respectively.

| | Factor Loadings | | | |
|---|---|---|---|---|
| Variable | 1 | 2 | 3 | Specific Variance |
| friendly/unfriendly | **0.9492** | 0.0234 | -0.0042 | 0.0985 |
| good/bad | **0.9585** | 0.0616 | -0.0320 | 0.0765 |
| nice/awful | **1.0150** | -0.0369 | 0.0124 | -0.0317 |
| brave/not brave | 0.3117 | **0.5446** | 0.1512 | 0.5834 |
| big/little | -0.1649 | **0.8993** | -0.1719 | 0.1344 |
| strong/weak | 0.1085 | **0.8105** | 0.0483 | 0.3290 |
| moving/still | -0.1121 | 0.1605 | **0.6160** | 0.5822 |
| fast/slow | 0.0067 | -0.0797 | **1.0464** | -0.1013 |
| Eigenvalues | 2.9985 | 1.8001 | 1.5304 | |
| Cumulative % of explained variability | 37.48 | 59.98 | 79.11 | |

According to the eigenvalues, the three factors explain around 79% of the variability in the model. Furthermore, careful inspection of the factor loadings enables us to describe that factor 1 is a measure of an individual's *propensity for kindness*. Factor 2, which consists primarily of the attributes brave/not brave, big/little, and strong/weak, might represent a person's *courage* or *strength of character*. Similarly, the third factor appears to represent *movement* or one's *pace of life*.

## §5 Conclusion and Discussion

Unlike current subjective or unquantifiable criteria used to identify common factors, the criteria we propose can be used to judge whether the data are suitable for EFA, which specifically determines an appropriate method of factor extraction, and the number of the common factors. In addition, the method proposed in this paper does not easily leads to systematic bias. Since the $Q$ index is related not only to the magnitude of the residuals (i.e. RMS), but also to the structure of the model (i.e. $W$), it naturally emphasizes the best factor extraction methods to use for a given data set, and even ranks them (i.e. larger $Q$ is better).

Residual analysis can detect the systematic bias in a model, for any parameter bias in the estimation of a common factor model will be reflected in the abnormality of the residuals. Therefore, checking the normality of the residuals is particularly suitable for exploratory factor analysis. In order to apply the method to CFA (confirmatory factor analysis), further study is needed. It is also important to recognize the limitation that the use of the Q-index is most appropriate when $RMS << 1/k$. While the examples presented in this paper and other datasets that we have applied our method to have satisfied this requirement, the root mean square is not necessarily bounded above by $1/k$. Further research is needed to determine under what scenarios or how frequently this requirement would not be met.

Measures of model fit are also one of the most important aspects of structural equation modeling (SEM). There are already dozens of model fit indices in the process of structural equation modeling [14,21]. However, how to determine the adequacy of an SEM remains difficult and unsolved due to the conflicting opinions on which model best fits the observed data. Use of our proposed Q-index in conjunction with checking for normality distributed residuals would help alleviate this problem.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

## References

[1] H Akaike. *Factor analysis and AIC*, Psychometrika, 1987, 52(3): 317-332.

[2] M S Bartlett. *A note on the multiplying factors for various $x^2$ approximations*, Journal of the Royal Statistical Society Series B, 1954, 16(2): 296-298.

[3] R B Cattell. *The scree test for the number of factors*, Multivariate behavioral research, 1966, 1(2): 245-276.

[4] R B Cattell, S Vogelmann. *A comprehensive trial of the scree and KG criteria for determining the number of factors*, Multivariate Behavioral Research, 1977, 12(3): 289-325.

[5] R D Cook, S Weisberg. *Residuals and Influence in Regression*, Chapman and Hall, 1982, 223.

[6] F J D Vesta, R T Walls. *Factor analysis of the semantic attributes of 487 words and some relationships to the conceptual behavior of fifth-grade children*, Journal of Educational Psychology, 1970, 61(6p2): 1-15.

[7] R A Fisher. *Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population*, Biometrika,1915, 10(4): 507-521.

[8] L Guttman. *Some necessary conditions for common-factor analysis*, Psychometrika, 1954, 19(2): 149-161.

[9] R A Johnson, D W Wichern. *Applied Multivariate Statistical Analysis*, Prentice Hall, 2007.

[10] H F Kaiser. *The application of electronic computers to factor analysis*, Educational and Psychological Measurement,1960, 20(1): 141-151.

[11] G Kapetanios. *A testing procedure for determining the number of factors in approximate factor models with large datasets*, Journal of Business and Economic Statistics, 2010, 28(3): 397-409.

[12] R C MacCallum, M W Browne, L Cai. *Factor analysis models as approximations conference presentation*, 2004.

[13] R C MacCallum, M W Browne, L Cai. *Factor analysis models as approximations: some history and some implications*, Routledge Academic, 2007, 153-176.

[14] S A Mulaik. *Linear Causal Modeling with Structural Equations*, Chapman and Hall/CRC, 2009.

[15] J P Royston. *An extension of shapiro and wilk?s w test for normality to large samples*, Applied Statistics, 1982, 31(2): 115-124.

[16] J R Schott. *Testing the equality of the smallest latent roots of a correlation matrix*, Biometrika, 1988, 75(4): 794-796.

[17] S S Shapiro, M B Wilk. *An analysis of variance test for normality*, Biometrika, 1965, 52(3/4): 591-611.

[18] B Thompson, L G Daniel. *Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines*, Educational and Psychological Measurement, 1996, 56(2): 197-208.

[19] N Timm. *Applied multivariate analysis*, Springer texts in statistics, Springer, 2002.

[20] L Tucker, R Koopman, R Linn. *Evaluation of factor analytic research procedures by means of simulated correlation matrices*, Psychometrika, 1969, 34(4): 421-459.

[21] K H Yuan. *Fit indices versus test statistics*, Multivariate Behavioral Research, 2005, 40(1): 115-148.

[1]School of Big Data, Fuzhou University of International Studies and Trade, Fuzhou 350202, China.
[2]School of Mathematics, Sun Yat-Sen University, Guangzhou 510275, China.
[3]Department of Computer Science, University of Bath, UK.
[4]Institute of Insect Science, Zhejiang University, Hangzhou 310058, China.
Email: mrhephd@163.com