

MODIFIED DISTRIBUTION-FREE GOODNESS-OF-FIT TEST STATISTIC

SO YEON CHUN

GEORGETOWN UNIVERSITY

MICHAEL BROWNE

OHIO STATE UNIVERSITY

ALEXANDER SHAPIRO

GEORGIA INSTITUTE OF TECHNOLOGY

Covariance structure analysis and its structural equation modeling extensions have become one of the most widely used methodologies in social sciences such as psychology, education, and economics. An important issue in such analysis is to assess the goodness of fit of a model under analysis. One of the most popular test statistics used in covariance structure analysis is the asymptotically distribution-free (ADF) test statistic introduced by Browne (Br J Math Stat Psychol 37:62–83, 1984). The ADF statistic can be used to test models without any specific distribution assumption (e.g., multivariate normal distribution) of the observed data. Despite its advantage, it has been shown in various empirical studies that unless sample sizes are extremely large, this ADF statistic could perform very poorly in practice. In this paper, we provide a theoretical explanation for this phenomenon and further propose a modified test statistic that improves the performance in samples of realistic size. The proposed statistic deals with the possible ill-conditioning of the involved large-scale covariance matrices.

Key words: covariance structures, distribution-free test statistic, asymptotics, Chi-square distribution, ill-conditioned problem.

1. Introduction

Structural equation modeling or covariance structure analysis has become a popular tool of quantitative research in the social sciences (e.g., Bollen, 1989; Lee, 1990; Kano, 2002; Tomarken & Waller, 2005; Xu & Mackenzie, 2012; Wu & Browne, 2015). In the analysis of covariance structures, it is hypothesized that the $p \times p$ population covariance matrix Σ can be represented as a matrix function $\Sigma(\theta)$ of q unknown parameters. Assuming that the population has multivariate normal distribution, the maximum likelihood (ML) estimate $\hat{\theta}$ is obtained by minimizing

$$F_{\text{ML}}(\theta) = \text{tr}(S\Sigma^{-1}(\theta)) - \log |S\Sigma^{-1}(\theta)| - p, \quad (1.1)$$

where S is the sample covariance matrix based on a sample of size n . The associated ML test statistic for evaluating the hypothetical model is $T_{\text{ML}} = nF_{\text{ML}}(\hat{\theta})$ (Browne, 1982).

It is known that under the assumption of multivariate normality and the null hypothesis (and some mild regularity conditions), T_{ML} is asymptotically distributed as Chi-square with $p^* - q$ degrees of freedom, where $p^* = p(p + 1)/2$. The ML method has been implemented in virtually all structural equation modeling software packages and widely used in practice. Although it could be guaranteed that T_{ML} has an asymptotic Chi-square distribution even without the assumption

Correspondence should be made to So Yeon Chun, McDonough School of Business, Georgetown University, Washington, DC 20057, USA. Email: soyeon.chun@georgetown.edu

of multivariate normality of the data in some situations (Amemiya & Anderson, 1990; Browne & Shapiro, 1987), it has been demonstrated that the test statistic T_{ML} can break down completely for nonnormally distributed data in many empirical studies (e.g., Hu et al., 1992).

Since the assumption of multivariate normality of data sets is often unrealistic, a test statistic that does not require any specific distribution assumptions of the observed data, distribution-free ADF test statistic, was proposed by Browne (1984). Unfortunately, it has shown in various empirical studies that unless sample sizes are extremely large (e.g., more than 4000 samples depending on the model and data analyzed), this ADF can perform very poorly (e.g., Hoogland & Boomsma, 1998; Boomsma & Hoogland, 2001).

The goal of this paper is twofold. First, we aim to give a theoretical explanation for some of the observed behaviors of this statistic. The second goal is to suggest a modification, which could lead to improved behavior in samples of realistic size. Different from existing modifications of Browne's test statistic suggested in the literature (e.g., Yuan & Bentler, 1998, 1999), our approach is mainly concerned with the potential ill-conditioning of the large-scale covariance matrices involved in test statistic calculations. We provide an asymptotic analysis of the proposed test statistic and further explain why and how the suggested adjustments could improve the behavior of ADF statistic. In addition, we present an illustrative numerical example using Monte Carlo method to study and compare finite sample behavior of ADF statistic and the proposed statistic.

2. Preliminary Discussion

Let $\mathbf{X} = (X_1, \dots, X_p)'$ be a $p \times 1$ random vector with mean $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}]$ and covariance matrix

$$\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] = \mathbb{E}[\mathbf{X}\mathbf{X}'] - \boldsymbol{\mu}\boldsymbol{\mu}'.$$

For a $p \times 1$ vector \mathbf{a} we have that $\text{Var}(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}$, and hence $\mathbf{a}'(\mathbf{X} - \boldsymbol{\mu})$ is identically zero iff $\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} = 0$. It follows that the matrix $\boldsymbol{\Sigma}$ is singular iff the elements of random vector $\mathbf{X} - \boldsymbol{\mu}$ are linearly dependent. Moreover, $\boldsymbol{\Sigma}$ has rank $r < p$, iff there exist $p - r$ linearly independent vectors $\mathbf{a}_1, \dots, \mathbf{a}_{p-r}$ such that $\mathbf{a}_i'\boldsymbol{\Sigma}\mathbf{a}_i = 0, i = 1, \dots, p - r$. This leads to the following result.

Proposition 1. *Rank of the covariance matrix $\boldsymbol{\Sigma}$ is equal to the maximal number of linearly independent elements of $\mathbf{X} - \boldsymbol{\mu}$.*

Let $\mathbf{X}^1, \dots, \mathbf{X}^n$ be an iid sample of vector \mathbf{X} , and

$$\mathbf{S} = \frac{1}{n-1} \sum_{v=1}^n (\mathbf{X}^v - \bar{\mathbf{X}})(\mathbf{X}^v - \bar{\mathbf{X}})'$$

be the corresponding sample covariance matrix. We have that if $\mathbf{a}'(\mathbf{X} - \boldsymbol{\mu}) \equiv 0$, then $\mathbf{a}'(\mathbf{X}^v - \bar{\mathbf{X}}) \equiv 0, v = 1, \dots, n$, and hence, $\text{rank}(\mathbf{S}) \leq \text{rank}(\boldsymbol{\Sigma})$. In fact if $n > p$ and random vector \mathbf{X} has continuous distribution, then $\text{rank}(\mathbf{S}) = \text{rank}(\boldsymbol{\Sigma})$ w.p.1. Denote by $\boldsymbol{\sigma} = \text{vecs}(\boldsymbol{\Sigma})$ and $\mathbf{s} = \text{vecs}(\mathbf{S})$ the corresponding $p^* \times 1$ vectors formed from the above (including diagonal elements, i.e., nonduplicated elements of the respective matrices). Recall that \mathbf{s} is an unbiased estimator of $\boldsymbol{\sigma}$, i.e., the expected value $\mathbb{E}[\mathbf{s}] = \boldsymbol{\sigma}$.

Consider the $p^* \times 1$ vector \mathbf{Y} with elements $(X_i - \mu_i)(X_j - \mu_j), i \leq j$, and let $\boldsymbol{\Gamma}$ be the covariance matrix of random vector \mathbf{Y} . That is, $\mathbb{E}[\mathbf{Y}] = \boldsymbol{\sigma}$ and

$$\boldsymbol{\Gamma} = \mathbb{E}[(\mathbf{Y} - \boldsymbol{\sigma})(\mathbf{Y} - \boldsymbol{\sigma})'] = \mathbb{E}[\mathbf{Y}\mathbf{Y}'] - \boldsymbol{\sigma}\boldsymbol{\sigma}', \quad (2.1)$$

with the typical element

$$\begin{aligned} [\mathbf{\Gamma}]_{ij,kl} &= \mathbb{E} \left\{ [(X_i - \mu_i)(X_j - \mu_j) - \sigma_{ij}][(X_k - \mu_k)(X_l - \mu_l) - \sigma_{kl}] \right\} \\ &= \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)(X_l - \mu_l)] - \sigma_{ij}\sigma_{kl}. \end{aligned} \quad (2.2)$$

Assuming existence (finiteness) of the fourth-order moments, we have by the central limit theorem that $\sqrt{n}(s - \sigma)$ converges in distribution to normal $N(\mathbf{0}, \mathbf{\Gamma})$. Now, let us discuss properties of the $p^* \times p^*$ covariance matrix $\mathbf{\Gamma}$.

Proposition 2. *The following properties hold:*

- (i) Rank of $\mathbf{\Gamma}$ is equal to the maximal number of linearly independent elements of $\mathbf{Y} - \sigma$.
- (ii) Let r be the rank of the covariance matrix $\mathbf{\Sigma}$. Then

$$\text{rank}(\mathbf{\Gamma}) \leq r(r + 1)/2. \quad (2.3)$$

Proof. Property (i) follows immediately from Proposition 1.

Now suppose that $\text{rank}(\mathbf{\Sigma}) = r$, it follows that there are r linearly independent elements of vector $\mathbf{X} - \mu$ such that the other elements of $\mathbf{X} - \mu$ are linear functions of these r elements. By permuting elements of $\mathbf{X} - \mu$ if necessary, we can assume that the first r elements of $\mathbf{X} - \mu$ are linearly independent. Then for $i > r$ and $j > r$ we have that $X_i - \mu_i = \sum_{k=1}^r a_k(X_k - \mu_k)$ and $X_j - \mu_j = \sum_{l=1}^r b_l(X_l - \mu_l)$ for some coefficients a_k and b_l . It follows that

$$(X_i - \mu_i)(X_j - \mu_j) = \sum_{k,l=1}^r a_k b_l (X_k - \mu_k)(X_l - \mu_l).$$

That is, the component Y_{ij} of \mathbf{Y} is a linear combination of elements Y_{kl} , $k, l = 1, \dots, r$. Since there are $r(r + 1)/2$ elements Y_{kl} , $k \leq l = 1, \dots, r$, this completes the proof. \square

It follows that the covariance matrix $\mathbf{\Gamma}$ is singular iff random variables $(X_i - \mu_i)(X_j - \mu_j) - \sigma_{ij}$, $i \leq j$, are linearly dependent. In particular, this implies that if $\mathbf{\Gamma}$ is singular, then the distribution of \mathbf{X} is degenerate (cf., Jennrich & Satorra, 2013). It also follows that if the matrix $\mathbf{\Sigma}$ is singular, then the matrix $\mathbf{\Gamma}$ is singular. However, the converse of this is not true in general. That is, it can happen that $\mathbf{\Sigma}$ is nonsingular, while the corresponding matrix $\mathbf{\Gamma}$ is singular. For example, let $p = 1$ and consider the following random variable X , which takes two values, $X = 1$ with probability $1/2$ and $X = -1$ with probability $1/2$. This random variable has zero mean and positive variance, while $Y = X^2$ is constantly 1. As an another example, for $p = 2$, let $X_1 \sim N(0, 1)$, and $X_2 = X_1$ for $|X_1| \leq 1$ and $X_2 = -X_1$ for $|X_1| > 1$. The random variables X_1 and X_2 have zero mean and are linearly independent, while $X_1^2 = X_2^2$. This also shows that the inequality (2.3) can be strict.

Now let \mathbf{Y}^v , $v = 1, \dots, n$, be $p^* \times 1$ vector composed from elements $(X_i^v - \bar{X}_i)(X_j^v - \bar{X}_j)$, $i \leq j$, and $\bar{\mathbf{Y}} = n^{-1} \sum_{v=1}^n \mathbf{Y}^v$. A sample estimate of $\mathbf{\Gamma}$ is

$$\tilde{\mathbf{\Gamma}} = \frac{1}{n-1} \sum_{v=1}^n (\mathbf{Y}^v - \bar{\mathbf{Y}})(\mathbf{Y}^v - \bar{\mathbf{Y}})', \quad (2.4)$$

with typical elements

$$[\tilde{\Gamma}]_{ij,kl} = \frac{1}{n-1} \sum_{v=1}^n [(X_i^v - \bar{X}_i)(X_j^v - \bar{X}_j) - \bar{Y}_{ij}] [(X_k^v - \bar{X}_k)(X_l^v - \bar{X}_l) - \bar{Y}_{kl}]. \quad (2.5)$$

Note that $\mathbf{s} = (n-1)^{-1} \sum_{v=1}^n \mathbf{Y}^v$, so the estimate

$$\check{\Gamma} = \frac{1}{n-1} \sum_{v=1}^n (\mathbf{Y}^v - \mathbf{s})(\mathbf{Y}^v - \mathbf{s})' \quad (2.6)$$

is slightly different from the estimate $\tilde{\Gamma}$. Note also that $\mathbb{E}[(X_i^v - \mu_i)(X_j^v - \mu_j)] = \sigma_{ij}$, while $\mathbb{E}[(X_i^v - \bar{X}_i)(X_j^v - \bar{X}_j)] \neq \sigma_{ij}$, i.e., $\mathbb{E}[\mathbf{Y}^v] \neq \boldsymbol{\sigma}$, and that both estimators $\tilde{\Gamma}$ and $\check{\Gamma}$ are biased. By the law of large numbers, both estimators are consistent, i.e., converge to Γ w.p.1 as $n \rightarrow \infty$.

3. Asymptotically Distribution-Free Test Statistic

Let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ be a covariance structural model, and $\boldsymbol{\Delta}(\boldsymbol{\theta}) = \partial \boldsymbol{\sigma}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ be the corresponding $p^* \times q$ Jacobian matrix. Here, $\boldsymbol{\theta}$ is $q \times 1$ parameter vector varying in a specified parameter space $\Theta \subset \mathbb{R}^q$. For the sake of simplicity we assume first that the model is correctly specified, i.e., there exists $\boldsymbol{\theta}_0 \in \Theta$ such that $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$. In order to evaluate power of the considered tests, we also consider misspecified models later assuming Pitman population drift, with resulting noncentral Chi-square asymptotic distribution. Let $\hat{\boldsymbol{\theta}}$ be a consistent estimate of $\boldsymbol{\theta}_0$, and consider $\hat{\boldsymbol{\sigma}} = \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})$ and the estimate $\hat{\boldsymbol{\Delta}} = \boldsymbol{\Delta}(\hat{\boldsymbol{\theta}})$ of the population Jacobian matrix $\boldsymbol{\Delta}_0 = \boldsymbol{\Delta}(\boldsymbol{\theta}_0)$. We assume throughout the paper that the mapping $\boldsymbol{\theta} \mapsto \boldsymbol{\sigma}(\boldsymbol{\theta})$ is continuously differentiable. For an $m \times k$ matrix \mathbf{A} , we denote by \mathbf{A}_c its orthogonal complement, i.e., \mathbf{A}_c is an $m \times (m - \nu)$ matrix of full column rank $m - \nu$ such that $\mathbf{A}_c' \mathbf{A} = \mathbf{0}$ with $\nu = \text{rank}(\mathbf{A})$. All respective orthogonal complements can be obtained from a particular one by the transformation $\mathbf{A}_c \mapsto \mathbf{A}_c \mathbf{M}$, where \mathbf{M} is an arbitrary $(m - \nu) \times (m - \nu)$ nonsingular matrix.

Consider the following distribution-free test statistic, introduced in Browne (1984),

$$T_B = n(\mathbf{s} - \hat{\boldsymbol{\sigma}})' \hat{\boldsymbol{\Delta}}_c (\hat{\boldsymbol{\Delta}}_c' \hat{\boldsymbol{\Delta}}_c)^{-1} \hat{\boldsymbol{\Delta}}_c' (\mathbf{s} - \hat{\boldsymbol{\sigma}}), \quad (3.1)$$

where $\hat{\Gamma}$ is a consistent estimator of Γ , and $\hat{\boldsymbol{\Delta}}_c$ is an orthogonal complement of matrix $\hat{\boldsymbol{\Delta}}$. The right-hand side of (3.1) does not depend on a particular choice of the orthogonal complement $\hat{\boldsymbol{\Delta}}_c$. By using a matrix identity, the statistic T_B can be written in the following equivalent form [Browne, 1984; expression (2.20b)],

$$T_B = n(\mathbf{s} - \hat{\boldsymbol{\sigma}})' \left[\hat{\Gamma}^{-1} - \hat{\Gamma}^{-1} \hat{\boldsymbol{\Delta}} (\hat{\boldsymbol{\Delta}}' \hat{\Gamma}^{-1} \hat{\boldsymbol{\Delta}})^{-1} \hat{\boldsymbol{\Delta}}' \hat{\Gamma}^{-1} \right] (\mathbf{s} - \hat{\boldsymbol{\sigma}}), \quad (3.2)$$

provided that matrix $\hat{\Gamma}$ is nonsingular. Let us consider the following assumptions.

- (A1) The population vector \mathbf{X} has finite fourth-order moments.
- (A2) The population Jacobian matrix $\boldsymbol{\Delta}_0$ has full column rank q .
- (A3) The estimator $\hat{\boldsymbol{\theta}}$ is $O_p(n^{-1/2})$ consistent estimator of $\boldsymbol{\theta}_0$, i.e., $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is bounded in probability.

(A4) The covariance matrix $\mathbf{\Gamma}$ is nonsingular.

Under the above regularity conditions (A1)–(A4), the test statistic T_B converges in distribution to a Chi-square distribution with $p^* - q$ degrees of freedom (cf., Browne & Shapiro, 2015, see also Theorem 1 below). The required regularity conditions are rather mild, and no assumptions about population distribution are made (apart from existence of fourth-order moments). The asymptotic distribution of $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ does not need to be normal, in particular the true value $\boldsymbol{\theta}_0$ of the parameter vector can be a boundary point of the parameter space Θ . Even the assumption (A2) can be relaxed by considering overparameterized models and assuming instead that $\boldsymbol{\theta}_0$ is a regular point of the mapping $\boldsymbol{\theta} \mapsto \boldsymbol{\sigma}(\boldsymbol{\theta})$. In that case, the number of degrees of freedom should be corrected to $p^* - \text{rank}(\mathbf{\Delta}_0)$ (cf., Shapiro, 1986). While it was previously argued that the requirement for the matrix $\mathbf{\Gamma}$ to be nonsingular is essential (e.g., Jennrich & Satorra, 2013), this assumption can be relaxed to the following.

(A5) The $(p^* - q) \times (p^* - q)$ matrix $\mathbf{\Delta}'_c \mathbf{\Gamma} \mathbf{\Delta}_c$ is nonsingular (nonsingularity of this matrix does not depend on a particular choice of the orthogonal complement $\mathbf{\Delta}_c$).

It is important to note that, even if $\mathbf{\Gamma}$ is nonsingular, it could still be ill-conditioned. Conditioning of a symmetric positive definite matrix is often measured by its condition number, defined as $\text{cond}(\mathbf{\Gamma}) = \lambda_{\max}(\mathbf{\Gamma})/\lambda_{\min}(\mathbf{\Gamma})$, where $\lambda_{\max}(\mathbf{\Gamma})$ and $\lambda_{\min}(\mathbf{\Gamma})$ are the respective largest and smallest eigenvalues of $\mathbf{\Gamma}$. The largest and smallest eigenvalues have the following variational representation

$$\lambda_{\max}(\mathbf{\Gamma}) = \max_{\mathbf{a}'\mathbf{a}=1} \mathbf{a}'\mathbf{\Gamma}\mathbf{a}, \quad \lambda_{\min}(\mathbf{\Gamma}) = \min_{\mathbf{a}'\mathbf{a}=1} \mathbf{a}'\mathbf{\Gamma}\mathbf{a}. \quad (3.3)$$

It follows that by adding a row and column to a symmetric positive definite matrix (and hence increasing its dimension by one), its largest eigenvalue becomes larger and smallest eigenvalue becomes smaller (Sturmiian separation theorem, e.g., Bellman, 1960, p. 117). Therefore, matrices of large dimensions tend to be ill-conditioned. It also follows from the variational representation (3.3) that $\lambda_{\max}(\cdot)$ is a convex function, and hence by Jensen's inequality $\mathbb{E}[\lambda_{\max}(\hat{\mathbf{\Gamma}})] \geq \lambda_{\max}(\mathbb{E}[\hat{\mathbf{\Gamma}}])$. It follows that $\lambda_{\max}(\hat{\mathbf{\Gamma}})$ tends to be bigger than $\lambda_{\max}(\mathbf{\Gamma})$, and similarly $\lambda_{\min}(\hat{\mathbf{\Gamma}})$ tends to be smaller than $\lambda_{\min}(\mathbf{\Gamma})$. Consequently, $\text{cond}(\hat{\mathbf{\Gamma}})$ tends to be even larger than $\text{cond}(\mathbf{\Gamma})$.

If $\text{cond}(\mathbf{\Gamma})$ is large, then even very small changes in elements of $\mathbf{\Gamma}$ can result in large changes of its inverse. In that case, the right-hand side of (3.2) could become very sensitive to small perturbations of the estimate $\hat{\mathbf{\Gamma}}$. As a result, one would need a very large sample for a reasonable convergence of the distribution of T_B to the Chi-square distribution. This was observed empirically (e.g., Huang & Bentler, 2015) and is considered to be a serious drawback of the distribution-free test statistic T_B . We can cite, for example: "One of major limitations associated with this approach [ADF] in addressing nonnormality has been its excessively demanding sample-size requirement. It is now well known that unless sample sizes are extremely large the ADF estimator performs very poorly and can yield severely distorted estimated values and standard errors" (Byrne, 2012, p. 315).

It is numerically demonstrated in Yuan and Bentler (1998) that the test statistic T_B is "sensitive to model degrees of freedom $p^* - q$, rather than to the model complexity, defined as number of parameters." Indeed, we observe that in the representation (3.1), one needs to invert matrix $\hat{\mathbf{\Delta}}'_c \hat{\mathbf{\Gamma}} \hat{\mathbf{\Delta}}_c$ of order $(p^* - q) \times (p^* - q)$, rather than the $p^* \times p^*$ matrix $\hat{\mathbf{\Gamma}}$. Considering the corresponding population counterparts, observe that the eigenvalues, and hence the condition number, of matrix $\mathbf{\Delta}'_c \mathbf{\Gamma} \mathbf{\Delta}_c$ depend on the choice of the orthogonal complement $\mathbf{\Delta}_c$. Suppose that matrix $\mathbf{\Delta}'_c \mathbf{\Gamma} \mathbf{\Delta}_c$ is nonsingular (assumption (A5)) and consider matrix

$$\mathbf{\Xi} = \mathbf{\Delta}_c (\mathbf{\Delta}'_c \mathbf{\Gamma} \mathbf{\Delta}_c)^{-1} \mathbf{\Delta}'_c, \quad (3.4)$$

which is independent of a particular choice of Δ_c . Suppose further that the orthogonal complement matrix Δ_c has orthonormal column vectors, i.e., $\Delta_c' \Delta_c = I_{p^*-q}$. Existence of such orthonormal complement matrix is ensured by applying the transformation $\Delta_c \mapsto \Delta_c M$ if necessary. If matrix Δ_c is orthonormal, then the nonzero eigenvalues of matrix Ξ coincide with the eigenvalues of matrix $(\Delta_c' \Gamma \Delta_c)^{-1}$, which are inverse of the eigenvalues of matrix $\Delta_c' \Gamma \Delta_c$, and do not depend on a particular choice of the orthonormal complement matrix Δ_c .

Unless stated otherwise, we assume that the matrix Δ_c is orthonormal. By the above discussion, it makes sense to measure sensitivity of the statistic T_B in terms of the condition number of $\Delta_c' \Gamma \Delta_c$ rather than the condition number of Γ .

If we replace sample estimates in (3.2) with their true counterparts, then the corresponding statistic T_{true} is calculated as

$$\begin{aligned} T_{\text{true}} &= n(s - \hat{\sigma})' \Delta_c (\Delta_c' \Gamma \Delta_c)^{-1} \Delta_c' (s - \hat{\sigma}) \\ &= n(s - \hat{\sigma})' \left[\Gamma^{-1} - \Gamma^{-1} \Delta (\Delta' \Gamma^{-1} \Delta)^{-1} \Delta' \Gamma^{-1} \right] (s - \hat{\sigma}). \end{aligned} \quad (3.5)$$

We will show that T_{true} could have a distribution well approximated by the corresponding Chi-square distribution even when the matrices Γ and $\Delta_c' \Gamma \Delta_c$ are ill-conditioned. Note that T_{true} cannot be computed in practice since true values of the required parameters are not known.

It is pointed out in Yuan and Bentler (1998) that since for a random symmetric positive definite matrix A , it holds that $\mathbb{E}[A^{-1}] \succeq (\mathbb{E}A)^{-1}$ (in the sense that matrix $\mathbb{E}[A^{-1}] - (\mathbb{E}A)^{-1}$ is positive semidefinite), “in the typical situation with large models and small to moderate sample sizes, T_B rejects correct models far too frequently.” By using a second-order approximation of the difference $(\hat{\Delta}_c' \hat{\Gamma} \hat{\Delta}_c)^{-1} - (\hat{\Delta}_c' \Gamma \hat{\Delta}_c)^{-1}$, we can write the following approximation of the difference $T_B - T_{\text{true}}$

$$T_B - T_{\text{true}} \approx Z_n' \left[-\Xi (\hat{\Gamma} - \Gamma) \Xi + \Xi (\hat{\Gamma} - \Gamma) \Xi (\hat{\Gamma} - \Gamma) \Xi \right] Z_n, \quad (3.6)$$

where $Z_n = n^{1/2}(s - \hat{\sigma})$ and matrix Ξ is defined in (3.4). When Ξ is “large,” which happens when matrix $\Delta_c' \Gamma \Delta_c$ is ill-conditioned, even small differences between the respective elements of $\hat{\Gamma}$ and Γ are amplified by the matrix Ξ . This gives another explanation for sensitivity of the distribution of T_B to the number of degrees of freedom $p^* - q$, which is the dimension of matrix $\Delta_c' \Gamma \Delta_c$. Since matrix Ξ is positive semidefinite, the second-order term in the expansion (3.6) of $T_B - T_{\text{true}}$ is always nonnegative. If $\hat{\Gamma}$ is an unbiased estimator of Γ , then $\mathbb{E}[\Xi (\hat{\Gamma} - \Gamma) \Xi] = \mathbf{0}$. Even if $\hat{\Gamma}$ is a slightly biased, for “large” Ξ , the second-order term typically dominates. This indicates that T_B tends to be stochastically bigger than T_{true} , and hence T_B rejects correct models too often.

So far, we have provided theoretical explanations on the reasons why ADF statistic could perform poorly and discussed some empirical evidence documented in the literature. In the next section, we propose a modified test statistic, which could improve the behavior of ADF statistic.

4. Modified Test Statistic

As discussed earlier, under assumptions (A1)–(A3) and (A5), the test statistic T_B has asymptotically Chi-square distribution with $p^* - q$ degrees of freedom. Let us now consider the case where matrix $\Delta_c' \Gamma \Delta_c$ is singular. Let $\xi_1 \geq \dots \geq \xi_{p^*-q}$ be the eigenvalues of matrix $\Delta_c' \Gamma \Delta_c$. Since it is assumed that matrix Δ_c is orthonormal, the eigenvalues $\xi_1, \dots, \xi_{p^*-q}$ coincide with the first (largest) $p^* - q$ eigenvalues of matrix Ξ defined in (3.4). By making transformation $\Delta_c \mapsto \Delta_c M$ with an appropriate orthogonal matrix M , we can make matrix $\Delta_c' \Gamma \Delta_c$ diagonal with diagonal elements $\xi_1, \dots, \xi_{p^*-q}$. We assume that matrix $\Delta_c' \Gamma \Delta_c$ is not null so that $\xi_1 > 0$.

Let

$$r = \text{rank}(\mathbf{\Delta}'_c \mathbf{\Gamma} \mathbf{\Delta}_c) + q. \quad (4.1)$$

It follows that $r - q > 0$, and since $\mathbf{\Delta}'_c \mathbf{\Gamma} \mathbf{\Delta}_c$ is singular, we have that $r - q < p^* - q$, and $\xi_{r-q} > 0$ while $\xi_{r-q+1} = \dots = \xi_{p^*-q} = 0$. Consider partitioning $\mathbf{\Delta}_c = [\mathbf{\Delta}_{1c}, \mathbf{\Delta}_{2c}]$, where $\mathbf{\Delta}_{1c}$ formed from first $r - q$ columns of $\mathbf{\Delta}_c$, and $\mathbf{\Delta}_{2c}$ formed from the columns corresponding to zero eigenvalues $\xi_{r-q+1}, \dots, \xi_{p^*-q}$. It makes sense then to reduce dimension of matrix $\mathbf{\Delta}_c$ by removing columns corresponding to zero eigenvalues $\xi_{r-q+1}, \dots, \xi_{p^*-q}$, and hence to replace $\mathbf{\Delta}_c$ with its $p^* \times (r - q)$ submatrix $\mathbf{\Delta}_{1c}$. Similarly, even if the population matrix $\mathbf{\Delta}'_c \mathbf{\Gamma} \mathbf{\Delta}_c$ is nonsingular but is ill-conditioned, it would be worthwhile to remove unstable components associated with its small eigenvalues. This is similar to a common practice in principal component analysis (PCA).

We can approach this in the following way. Let \mathbf{Y} be a $p^* \times r$ matrix such that the $r \times r$ matrix $\mathbf{Y}' \mathbf{\Gamma} \mathbf{Y}$ is nonsingular. Let us make the following replacements in the right-hand side of (3.2): $\hat{\mathbf{\Gamma}}$ by $\mathbf{Y}' \hat{\mathbf{\Gamma}} \mathbf{Y}$, $\hat{\mathbf{\Delta}}$ by $\mathbf{Y}' \hat{\mathbf{\Delta}}$ and $s - \hat{\sigma}$ by $\mathbf{Y}'(s - \hat{\sigma})$. That is, consider the following modified statistic

$$T_{\mathbf{Y}} = n(s - \hat{\sigma})' \mathbf{Y} [(\mathbf{Y}' \hat{\mathbf{\Gamma}} \mathbf{Y})^{-1} - (\mathbf{Y}' \hat{\mathbf{\Gamma}} \mathbf{Y})^{-1} \mathbf{Y}' \hat{\mathbf{\Delta}} (\hat{\mathbf{\Delta}}' \mathbf{Y} (\mathbf{Y}' \hat{\mathbf{\Gamma}} \mathbf{Y})^{-1} \mathbf{Y}' \hat{\mathbf{\Delta}})^{-1} \hat{\mathbf{\Delta}}' \mathbf{Y} (\mathbf{Y}' \hat{\mathbf{\Gamma}} \mathbf{Y})^{-1}] \mathbf{Y}'(s - \hat{\sigma}). \quad (4.2)$$

Alternatively, we can write this statistic as

$$T_{\mathbf{Y}} = n(s - \hat{\sigma})' \mathbf{Y} \hat{\mathbf{\Delta}}_{\mathbf{Y}} (\hat{\mathbf{\Delta}}'_{\mathbf{Y}} \mathbf{Y}' \hat{\mathbf{\Gamma}} \mathbf{Y} \hat{\mathbf{\Delta}}_{\mathbf{Y}})^{-1} \hat{\mathbf{\Delta}}'_{\mathbf{Y}} \mathbf{Y}'(s - \hat{\sigma}), \quad (4.3)$$

where $\hat{\mathbf{\Delta}}_{\mathbf{Y}}$ is an orthogonal complement of matrix $\mathbf{Y}' \hat{\mathbf{\Delta}}$. That is, $\hat{\mathbf{\Delta}}_{\mathbf{Y}}$ is an $r \times (r - q)$ matrix of full column rank such that $\hat{\mathbf{\Delta}}'_{\mathbf{Y}} \mathbf{Y}' \hat{\mathbf{\Delta}}_{\mathbf{Y}} = \mathbf{0}$. Note that the right-hand side of (4.3) does not depend on a particular choice of the orthogonal complement $\hat{\mathbf{\Delta}}_{\mathbf{Y}}$. The $p^* \times (r - q)$ matrix $\tilde{\mathbf{\Delta}}_c = \mathbf{Y} \hat{\mathbf{\Delta}}_{\mathbf{Y}}$ is orthogonal of matrix $\hat{\mathbf{\Delta}}$ and can be viewed as a sample counterpart of the matrix $\mathbf{\Delta}_{1c}$.

Using $\tilde{\mathbf{\Delta}}_c$ we can write

$$T_{\mathbf{Y}} = n(s - \hat{\sigma})' \tilde{\mathbf{\Delta}}_c (\tilde{\mathbf{\Delta}}_c' \hat{\mathbf{\Gamma}} \tilde{\mathbf{\Delta}}_c)^{-1} \tilde{\mathbf{\Delta}}_c' (s - \hat{\sigma}). \quad (4.4)$$

Note that because of the equivalence of (4.2) and (4.3), the expression in the right-hand side of (4.3) does not depend on a particular choice of the orthogonal complement $\hat{\mathbf{\Delta}}_{\mathbf{Y}}$.

Theorem 1. *Suppose that the assumptions (A1)–(A3) hold and the matrix $\mathbf{Y}' \mathbf{\Gamma} \mathbf{Y}$ is nonsingular. Then the test statistic $T_{\mathbf{Y}}$ converges in distribution to a Chi-square distribution with $r - q$ degrees of freedom.*

Proof. Because of the assumption (A1), we have by the central limit theorem that $n^{1/2}(s - \sigma_0)$ converges in distribution to normal $N(\mathbf{0}, \mathbf{\Gamma})$. Here $\mathbf{\Sigma}_0 = \mathbf{\Sigma}(\theta_0)$ is the population covariance matrix and $\sigma_0 = \text{vecs}(\mathbf{\Sigma}_0)$. Since $\sigma(\cdot)$ is differentiable, we can write

$$\sigma(\theta) - \sigma(\theta_0) = \mathbf{\Delta}(\theta - \theta_0) + o(\|\theta - \theta_0\|),$$

where $\mathbf{\Delta} = \mathbf{\Delta}(\theta_0)$ (for simplicity of notation, we drop here the subscript in $\mathbf{\Delta}_0$). Moreover, because of (A3), we have that $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$, and hence

$$\hat{\sigma} - \sigma_0 = \sigma(\hat{\theta}) - \sigma(\theta_0) = \mathbf{\Delta}(\hat{\theta} - \theta_0) + o_p(n^{-1/2}).$$

Let $\Delta_{\#}$ be the population counterpart of matrix $\hat{\Delta}_{\#} = \hat{\Upsilon} \hat{\Delta}_{\Upsilon}$. Since $\Delta'_{\#} \Delta = \mathbf{0}$, it follows that

$$\Delta'_{\#}(s - \hat{\sigma}) = \Delta'_{\#}(s - \sigma_0) + \Delta'_{\#}(\sigma_0 - \hat{\sigma}) = \Delta'_{\#}(s - \sigma_0) + o_p(n^{-1/2}).$$

Now since $n^{1/2}(s - \sigma_0)$ converges in distribution, it follows that $s - \sigma_0 = O_p(n^{-1/2})$, and hence we obtain that

$$\hat{\Delta}'_{\#}(s - \sigma_0) = \Delta'_{\#}(s - \sigma_0) + o_p(n^{-1/2}).$$

Since $n^{1/2} \Delta'_{\#}(s - \sigma_0)$ converges in distribution to normal $N(\mathbf{0}, \Delta'_{\#} \Gamma \Delta_{\#})$, it follows by Slutsky's theorem that T_{Υ} converges in distribution to $\mathbf{Z}' \mathbf{U}^{-1} \mathbf{Z}$, where $\mathbf{U} = \Delta'_{\#} \Gamma \Delta_{\#}$ and $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{U})$. It follows that T_{Υ} converges in distribution to Chi-square with $r - q$ degrees of freedom. \square

It is important to note that the true rank of Γ (or rather of $\Delta'_c \Gamma \Delta_c$) is not known in practice. Therefore, we could take Υ to be composed of eigenvectors of $\hat{\Gamma}$ corresponding to its largest eigenvalues. Choice of the number of the largest eigenvalues could be based on heuristic, similar to the heuristic of PCA. It is also possible to approach this by choosing largest eigenvalues of $\hat{\Delta}'_c \hat{\Gamma} \hat{\Delta}_c$.

Next, in order to evaluate power of the modified statistic T_{Υ} , we can use the following approach. Assume that the population value $\sigma_{0,n}$ depends on the sample size n and converges to $\sigma^* = \sigma(\theta^*)$, $\sigma^* \in \Theta$, in such way that $n^{1/2}(\sigma_{0,n} - \sigma^*)$ tends to a vector μ . This is the so-called population drift assumption and referred to as a sequence of local alternatives [see, e.g., McManus, (1991) for a historical overview of this assumption]. Under this additional assumption of local alternatives, the statistic T_{Υ} converges in distribution to a noncentral Chi-square distribution with $r - q$ degrees of freedom and the noncentrality parameter

$$\delta = \mu' \Delta_{1c} (\Delta'_{1c} \Gamma \Delta_{1c})^{-1} \Delta'_{1c} \mu. \quad (4.5)$$

Consequently, power of test statistic T_{Υ} is measured in terms of the noncentrality parameter δ and the number of degrees of freedom $df = r - q$. This leads to a certain loss of power compared with the test statistic T_B . In particular, if $\Delta'_{1c} \mu = \mathbf{0}$ for some μ , then $\delta = 0$. This is the price one would be willing to pay for making the ADF approach work better.

5. Illustrative Numerical Example

In our Monte Carlo experiments, we consider the following factor analysis model $\Sigma = \Lambda \Lambda' + \Psi$, where $\Lambda = [\lambda_{k\ell}]$ is $p \times m$ matrix of factor loadings and $\Psi = \text{diag}(\psi_{11}, \dots, \psi_{pp})$ is $p \times p$ diagonal matrix of residual variances. In order to make this model identifiable for $m > 1$, for example, we can set the upper triangular part of Λ to zero, i.e., set $\lambda_{k\ell} = 0$ for $\ell > k$. Here, the parameter vector θ consists of $mp - m(m - 1)/2$ elements of Λ (with $\lambda_{k\ell}$, $\ell > k$, removed) and p diagonal elements of Ψ . So the dimension of θ is $q = mp - m(m - 1)/2 + p$. With this structure, we can further write this model as

$$\sigma_{ij} = \begin{cases} \sum_{\ell=1}^m \lambda_{i\ell} \lambda_{j\ell} + \psi_{ii}, & i = j, \\ \sum_{\ell=1}^m \lambda_{i\ell} \lambda_{j\ell}, & i \neq j. \end{cases}$$

TABLE 1.
Model 1: generated parameters.

Λ^*	0.65	0	0	Ψ^*	1.24
	0.99	0	0		0.21
	0.64	0	0		1.57
	0	0.39	0		0.65
	0	0.69	0		0.34
	0	0.75	0		1.60
	0	0	0.05		1.76
	0	0	0.30		2.84
	0	0	0.73		2.42

TABLE 2.
Model 2: generated parameters.

Λ^*	0.46	0	Ψ^*	0.91
	0.61	0.39		2.96
	0.08	0.22		0.47
	0.88	0.11		2.57
	0.64	0.37		0.80
	0.82	0.61		2.46
	0.12	0.78		0.99
	1.00	0.35		1.94
	0.65	0.48		0.65

Therefore, the elements $\delta_{ij,k\ell}$ of the Jacobian matrix $\mathbf{\Delta}(\boldsymbol{\theta})$ can be computed by using partial derivatives $\delta_{ij,k\ell} = \frac{\partial \sigma_{ij}}{\partial \lambda_{k\ell}}$, $1 \leq i \leq j \leq p$, $k = 1, \dots, p$, $\ell = 1, \dots, m$, $\ell \leq k$, with

$$\frac{\partial \sigma_{ij}}{\partial \lambda_{k\ell}} = \begin{cases} 2\lambda_{i\ell}, & i = j = k, \\ \lambda_{j\ell}, & i \neq j, k = i, \\ \lambda_{i\ell}, & i \neq j, k = j, \\ 0 & \text{otherwise,} \end{cases}$$

$$\frac{\partial \sigma_{ij}}{\partial \psi_{kk}} = \begin{cases} 1, & i = j = k, \\ 0, & \text{otherwise.} \end{cases}$$

With the factor analysis model described above, we first demonstrate that if $\boldsymbol{\Gamma}$ is ill-conditioned, then the corresponding Chi-square distribution could be a poor approximation of the distribution of statistic $T_{\mathbb{B}}$ defined in (3.2). Then, we illustrate that the correction proposed in Sect. 4 can make significant improvements.

Detailed simulation procedures and simulation parameters are as follows. In our experiment, we set $p = 9$, $m = 3$, $q = 18$ for Model 1 (confirmatory factor analysis model), $p = 9$, $m = 2$, $q = 26$ for Model 2 (exploratory factor analysis model) and construct the population covariance matrix $\boldsymbol{\Sigma}^*$ with specific values of elements of $\boldsymbol{\Lambda}^*$ and diagonal elements of $\boldsymbol{\Psi}^*$ generated as shown in Tables 1 and 2. Note that these models are identifiable since the upper triangular part of $\boldsymbol{\Lambda}^*$ is zero. For both models, the factor correlations are fixed to zero.

TABLE 3.
Model 1: comparison of T_{true} , T_{B} , and T_{γ} , KS distances and rejection rates.

Test statistic	df	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	K	AK
T_{true}	27	1.400	5.744	10.967	0.02093	0.00928
T_{B}	27	63.467	74.833	79.867	0.75416	0.43867
T_{γ}	22	5.067	7.367	9.333	0.52544	0.32432
	21	2.600	4.700	6.067	0.39737	0.24774
	20	1.733	4.100	5.300	0.25292	0.13966
	19	2.233	4.667	7.100	0.21319	0.12182
	18	3.900	7.233	10.267	0.21942	0.12394
	17	4.833	8.900	11.033	0.15297	0.07961
	16	7.7333	12.967	15.700	0.08007	0.04805

TABLE 4.
Model 2: comparison of T_{true} , T_{B} , and T_{γ} , KS distances and rejection rates.

Test statistic	df	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	K	AK
T_{true}	19	2.033	7.033	13.067	0.04222	0.02325
T_{B}	19	31.000	43.367	50.467	0.47444	0.29511
T_{γ}	17	6.300	11.967	16.133	0.25193	0.14143
	16	3.900	9.267	13.900	0.17090	0.08553
	15	4.400	10.233	15.567	0.12166	0.05659
	14	5.267	12.367	17.300	0.08694	0.05218
	13	3.100	9.533	14.667	0.12908	0.06354

Next, we generate the ill-conditioned sample estimate $\hat{\Gamma}$ with the data from an elliptical distribution ($M = 3000$ simulation samples with the sample size $n = 500$ for each model) as follows. Let $\mathbf{X} \sim N(\mathbf{0}, \Sigma)$ be a random vector having (multivariate) normal distribution and W be a random variable independent of \mathbf{X} . Then the random vector $\mathbf{Y} = W\mathbf{X}$ has an elliptical distribution with zero mean vector, covariance matrix $\alpha\Sigma$, where $\alpha = \mathbb{E}[W^2]$, and the kurtosis parameter $\kappa = \frac{\mathbb{E}[W^4]}{(\mathbb{E}[W^2])^2} - 1$ (see Chun & Shapiro, 2009 for more details). We involve a random variable W taking two values, 2 with probability 0.2 and 0.5 with probability 0.8. The kurtosis parameter of the elliptical distributions is then $\kappa = 2.25$. Note that in this case, $\mathbb{E}[W^2] = 1$, so that the covariance matrices of \mathbf{X} and \mathbf{Y} are equal to each other. For the Monte Carlo samples, condition numbers of $\hat{\Gamma}$ range from 198.83 to 1589.7 (Model 1), and 241.097 to 2721.653 (Model 2), with a mean of 529.89 and 728.492, respectively (while the condition number of Γ is 113.39 and 259.115 for respective Models 1 and 2). This is in accordance with an observation in Sect. 3 that the condition number of sample covariance matrix $\hat{\Gamma}$ tends to be larger than the condition number of Γ .

For each simulation sample, we obtain consistent estimates $\hat{\sigma}$ (through the generalized least squares method) and calculate three different test statistics T_{B} , T_{true} , T_{γ} as follows. Using the sample estimates $\hat{\sigma}$, $\hat{\Gamma}$ and $\hat{\Delta}$, the statistic T_{B} can be calculated as defined in Eq. (3.2). The statistic T_{true} is computed as in (3.5) by replacing sample estimates in (3.2) with their true counterparts. We will show that T_{true} could have a distribution well approximated by the corresponding Chi-square distribution, while this Chi-square distribution could give a poor approximation of the corresponding statistic T_{B} . We make corrections by taking $\mathbf{Y} = [\hat{e}_1, \dots, \hat{e}_r]$, where $\hat{e}_1, \dots, \hat{e}_r$

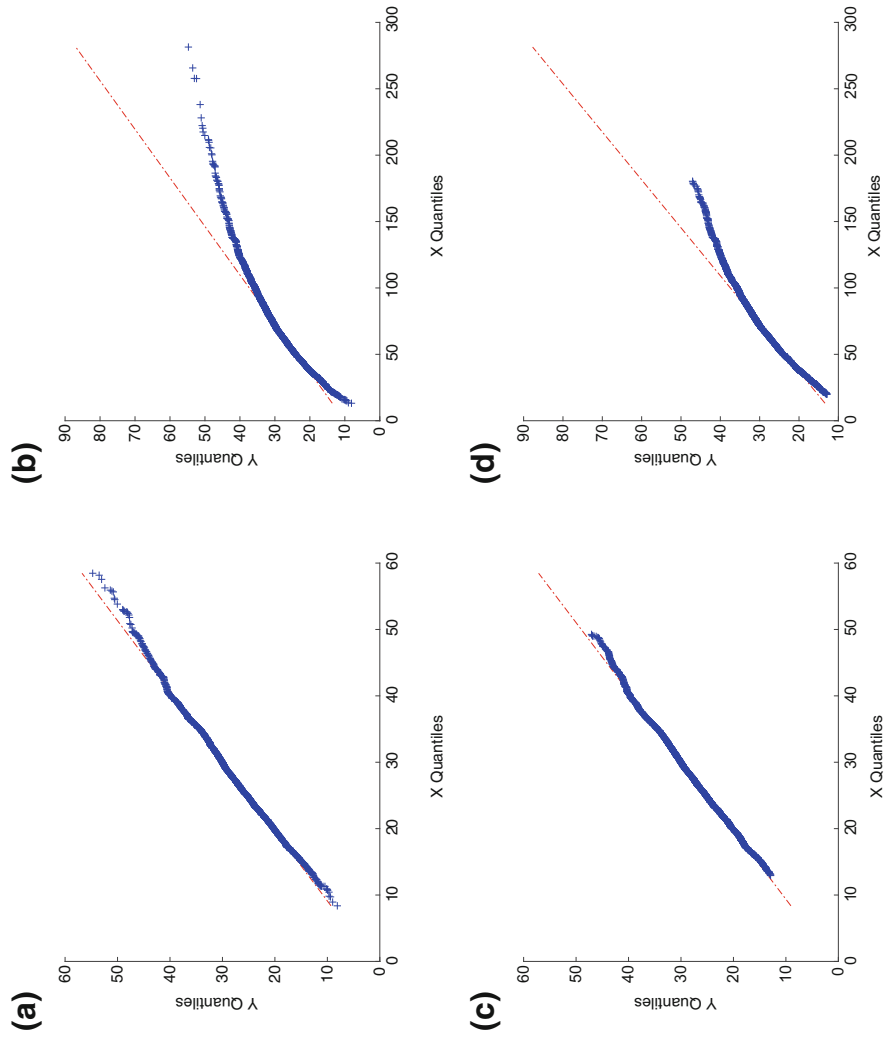


FIGURE 1. Model 1: Q-Q plots of T_{true} and T_B against Chi-square distribution with $df = 27$. **a** T_{true} . **b** T_B . **c** T_{true} (only for 1st–99th percentiles). **d** T_B (only for 1st–99th percentiles).

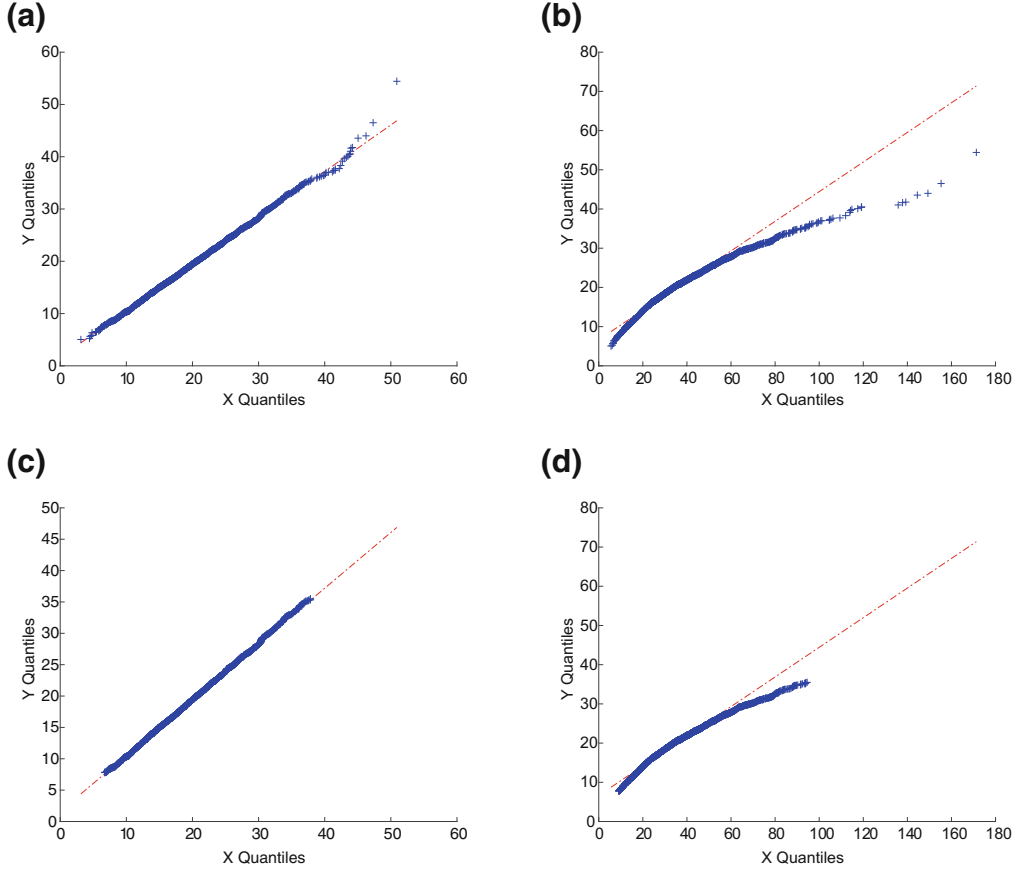


FIGURE 2.

Model 2: Q–Q plots of T_{true} and T_B against Chi-square distribution with $df = 19$. **a** T_{true} . **b** T_B . **c** T_{true} (only for 1st–99th percentiles). **d** T_B (only for 1st–99th percentiles).

are eigenvectors of $\hat{\Gamma}$ corresponding to its largest eigenvalues, and we can calculate T_γ as defined in Eq. (4.2). In our experiments, we try to remove the smallest eigenvalues which are smaller than 0.05–1.2% of the largest eigenvalue for each simulation sample. With this correction, we will show that the corresponding Chi-square asymptotics provided in Theorem 1 give a reasonably good fit for the statistic T_γ . As we discussed earlier, it is also possible to make corrections based on the largest eigenvalues of $\hat{\Delta}'_c \hat{\Gamma} \hat{\Delta}_c$.

We use several discrepancy measures to compare the fit of each statistic. One is the Kolmogorov–Smirnov (KS) distance defined as

$$K = \sup_{t \in \mathbb{R}} |\hat{F}_M(t) - F(t)|, \quad (5.1)$$

where $\hat{F}_M(t) = \frac{\#\{T_i \leq t\}}{M}$ is the empirical cumulative distribution function (cdf) based on the $M = 3000$ Monte Carlo samples of each test statistic T_B , T_{true} , T_γ , and $F(t)$ is the theoretical cdf of the respective approximations of the test statistic. As the KS distance only depends on the extreme cases, we also consider the *average* Kolmogorov–Smirnov distance (AK) (cf., Yuan et al., 2007) as

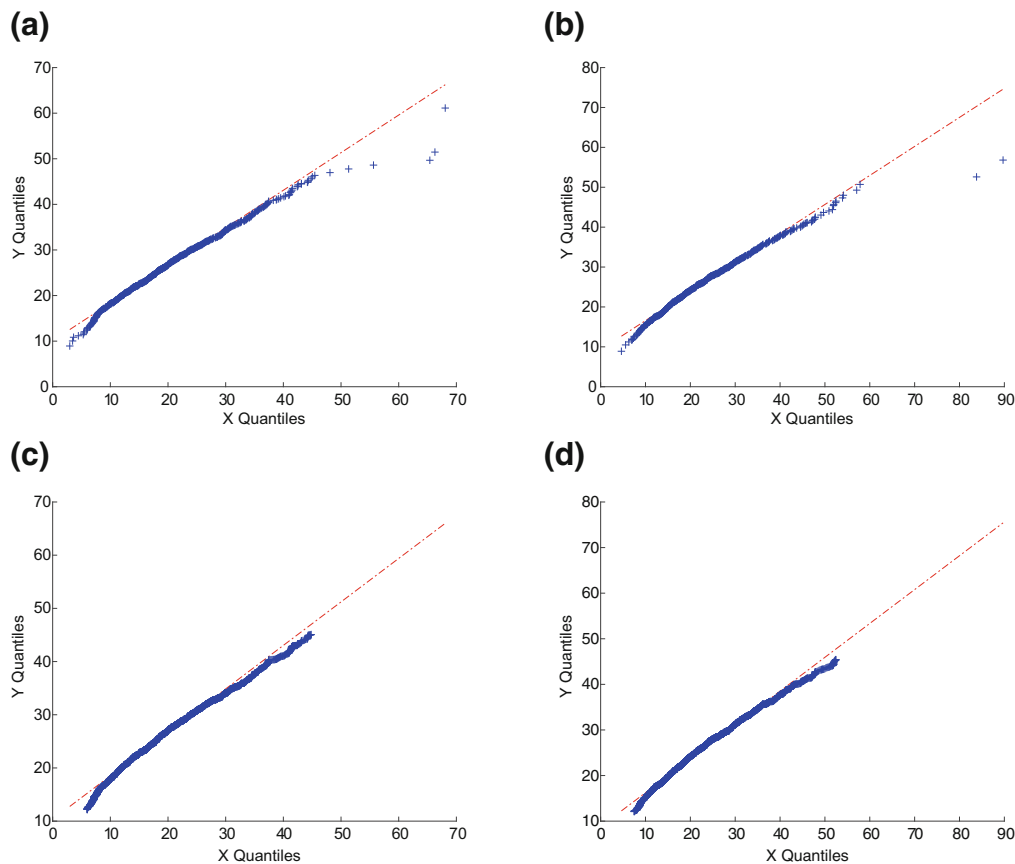


FIGURE 3.

Model 1: Q–Q plots of T_γ . **a** $df = 17$. **b** $df = 19$. **c** $df = 17$ (only for 1st–99th percentiles). **d** $df = 19$ (only for 1st–99th percentiles).

$$AK = \frac{1}{M} \sum_{i=1}^M K_i, \quad (5.2)$$

where

$$K_i = \max \left\{ \left| \frac{i-1}{M} - F(T_{(i)}) \right|, \left| \frac{i}{M} - F(T_{(i)}) \right| \right\},$$

with $T_{(1)} \leq \dots \leq T_{(M)}$ being the respective order statistics.

In addition to the KS distances, quantile comparisons through the empirical rejection rates are done to investigate the quality of each approximation with respect to the validity of confidence intervals. That is, we count the number of rejections for $\alpha = 1, 5, 10\%$ for each statistic. If the statistic is well approximated by the corresponding theoretical Chi-square distribution, the empirical rejection rates should be close to 1, 5, 10%, respectively. Finally, quantile–quantile (Q–Q) plots for each statistic against its theoretical Chi-square distributions are provided.

Tables 3 and 4 present KS distances and empirical rejection rates for each α level for three different statistics. It can be seen that distribution of T_{true} is well approximated by the corresponding Chi-square distribution and has better empirical rejection rates (rejection rates are close to the

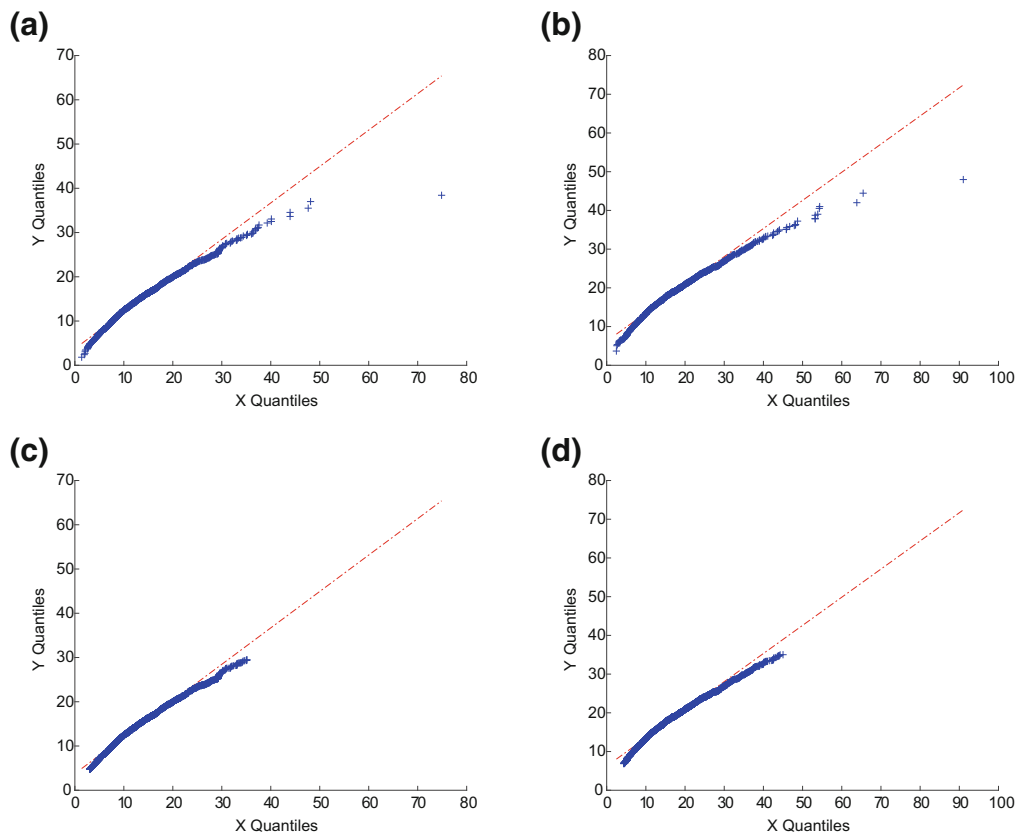


FIGURE 4.

Model 2: Q–Q plot examples of T_γ . **a** $df = 13$. **b** $df = 14$. **c** $df = 13$ (only for 1st–99th percentiles). **d** $df = 14$ (only for 1st–99th percentiles).

α levels set). As expected, the approximation of T_{true} by the ADF statistic T_B is extremely poor. It rejects the true values too often (e.g., empirical rejection rate is about 75% where $\alpha = 5\%$ for Model 1 in Table 3), and KS distance numbers are very large. On the other hand, when we make corrections by removing some of the smallest eigenvalues, we observe the significantly improved fit. Empirical rejection rates are much closer to the α levels, and KS distance numbers are much smaller. For example, when $df = 20$ for Model 1 (Table 3), empirical rejection rates of T_γ are 1.7, 4.1, and 5.3% where the rates of T_{true} are 1.4, 5.7, and 11%, and of T_B are 63, 75, and 80%. That is, the proposed test statistic T_γ has the considerably improved empirical rejection rate compared to T_B . For Model 2, the performance of T_γ and even the performance of T_{true} are slightly worse than for Model 1, but the proposed test statistic T_γ still has the significantly improved empirical rejection rates compared to T_B . For example, when $df = 16$ (Table 4), empirical rejection rates of T_γ are 3.9, 9.3, and 13.9% where the rates of T_{true} are 2, 7, and 13%, and the rates of T_B are 31, 43, and 50%. That is, the proposed test statistic T_γ has the considerably improved empirical rejection rates compared to T_B for both models. We also observe that the KS distance numbers are much smaller than those of T_B . Note that for T_γ , since we remove a certain percentage of the eigenvalues, we ended up with several different degrees of freedom depending on how many eigenvalues are removed.

The following Q–Q plots confirm the similar results. We first generate plots with all the data points, and then with only the 1st–99th percentiles to see the relevant deviations. Figures 1 and

TABLE 5.
Model 1: comparison of T_{true} , T_{B} , and T_{γ} , KS distances and rejection rates.

Sample size	Test statistic	df	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	K	AK
$n = 250$	T_{true}	27	2.000	7.800	13.400	0.03570	0.01530
	T_{B}	27	92.300	95.500	96.800	0.93330	0.49300
	T_{γ}	22	9.100	10.900	11.700	0.73800	0.41300
		21	3.800	4.300	4.400	0.69600	0.37890
		20	0.900	1.500	1.500	0.75450	0.38290
		19	0.400	1.000	1.200	0.47650	0.26400
		18	1.100	2.500	2.900	0.38020	0.19910
		17	2.400	4.800	6.400	0.26540	0.14440
$n = 500$	T_{true}	27	0.900	5.900	11.200	0.01873	0.00690
		27	62.900	73.000	79.300	0.74631	0.43490
	T_{γ}	22	5.300	7.500	9.600	0.47119	0.30118
		21	2.400	3.600	5.000	0.47981	0.29478
		20	1.400	2.100	3.400	0.24742	0.14378
		19	2.800	6.300	7.700	0.24105	0.12268
		18	4.100	7.900	10.800	0.22315	0.12082
		17	6.400	11.000	14.200	0.09829	0.04887
$n = 1000$	T_{true}	27	1.100	5.600	10.600	0.01600	0.00520
		27	38.100	55.100	62.100	0.59450	0.36360
	T_{γ}	22	3.900	7.300	10.500	0.36210	0.20220
		21	3.300	5.900	8.900	0.35480	0.20530
		20	2.300	6.000	8.500	0.26060	0.13980
		19	4.000	7.800	10.200	0.21790	0.12150
		18	5.900	11.300	13.800	0.16460	0.07650
		17	8.000	13.400	18.100	0.11790	0.05830
	16	10.200	17.500	20.300	0.07500	0.04160	

2 show the comparison between the Q–Q plots for T_{true} and T_{B} for Models 1 and 2, respectively. While we observe a reasonably straight line for T_{true} , a clearly curved pattern is shown for T_{B} . Again, as shown in Figs. 3 (Model 1) and 4 (Model 2), T_{γ} significantly improves the fit even though there still exist a few outliers. In summary, T_{γ} performs significantly better than the T_{B} but not as good as the T_{true} . However, it is important to note that, in real applications, T_{true} is *not* available.

Next, we extend our simulation study and compare test statistics with different sample sizes, and we also study the model with larger p and provide additional results. Regarding different samples sizes, we generate the ill-conditioned sample estimate $\hat{\Gamma}$ with the data from an elliptical distribution with $M = 1000$ simulation samples with different sample sizes $n = 250, 500, 1000$ for Model 1 (other simulation parameters remain the same). Table 5 provides empirical rejection rates (for $\alpha = 1, 5, 10\%$) and KS distances for three different statistics with different sample sizes. As expected, the performance of the ADF statistic T_{B} is very sensitive to the sample size and improves as the sample size increases. However, it is still a lot worse than the performances of T_{true} and T_{γ} even when $n = 1000$. Both T_{true} and T_{γ} display slightly better fits with larger sample sizes, and the proposed statistic T_{γ} notably improves the performance compared to T_{B} for all the sample sizes considered.

TABLE 6.
Model 3: generated parameters.

λ^*	0.40	0	0	Ψ^*	1.96
	0.80	0	0		1.16
	0.67	0	0		0.35
	0.94	0	0		0.49
	0.76	0	0		2.66
	0.70	0	0		2.28
	0.72	0	0		1.50
	0	0.86	0		2.73
	0	0.71	0		1.30
	0	0.79	0		1.50
	0	0.41	0		0.30
	0	0.83	0		0.22
	0	0.58	0		1.39
	0	0.85	0		2.25
	0	0	0.74		1.28
	0	0	0.58		0.25
	0	0	0.85		0.19
	0	0	0.81		1.87
	0	0	0.98		1.47
	0	0	0.75		0.84
	0	0	0.66		0.67

TABLE 7.
Model 3: comparison of T_{true} , T_{B} , and T_{Y} , KS distances and rejection rates.

Test statistic	df	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	K	AK
T_{true}	189	1.800	6.700	13.700	0.08221	0.03820
T_{B}	189	100.000	100.000	100.000	1.00000	0.50000
T_{Y}	85	2.900	4.800	5.100	0.21327	0.11575
	84	3.700	5.100	5.800	0.26352	0.13213
	83	5.800	7.900	8.900	0.24327	0.11680
	82	6.400	7.700	8.400	0.18880	0.07926
	81	6.300	8.000	8.500	0.21726	0.10431
	80	6.400	7.700	8.600	0.16112	0.07347
	79	3.500	4.100	4.800	0.30633	0.13112
	78	4.300	4.900	5.100	0.27748	0.10235
	77	3.100	3.100	3.500	0.23082	0.07028

Lastly, we compare statistics for a larger model with $p = 21$, $m = 3$, $q = 42$ (Model 3) with constructed population covariance matrix Σ^* with specific values of elements of Λ^* and diagonal elements of Ψ^* generated as shown in Table 6. The factor correlations are fixed to zero. We generate the ill-conditioned matrix $\hat{\Gamma}$ with the data from an elliptical distribution (with $\kappa = 2.25$, $M = 1000$, and $n = 1000$). For the Monte Carlo samples, condition numbers of $\hat{\Gamma}$ range from 22,441 to 66,500, with a mean of 40,580 (while the condition number of Γ is 1607.3). Note that the condition number of sample covariance matrix $\hat{\Gamma}$ is significantly larger than the condition number of Γ .

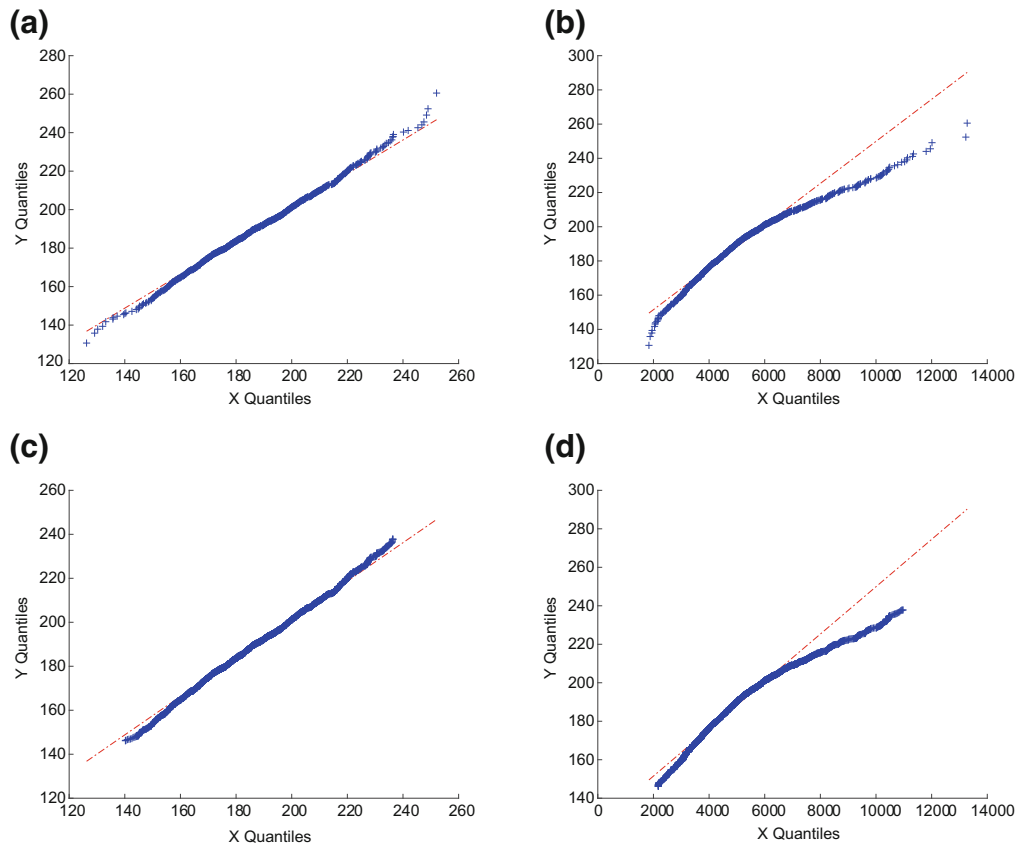


FIGURE 5.

Model 3: Q-Q plots of T_{true} and T_B against Chi-square distribution with $df = 189$. **a** T_{true} . **b** T_B . **c** T_{true} (only for 1st–99th percentiles). **d** T_B (only for 1st–99th percentiles).

Table 7 presents KS distances and empirical rejection rates for each α level for three different statistics. Similar to the results for Models 1 and 2, T_{true} is well approximated by the corresponding Chi-square distribution and has relatively good empirical rejection rates (rejection rates are close to the α levels set) while p (and the corresponding degrees of freedom and condition numbers) is significantly larger than in Models 1 and 2. On the other hand, performance of the ADF statistic T_B is extremely poor and a lot worse than the cases of Models 1 and 2. In particular, empirical rejection rates are 100% for all three levels of α and KS distances are extremely large. That is, T_B breaks down completely. However, the proposed test statistic T_γ again demonstrates a considerably improved fit. T_γ has the remarkably improved empirical rejection rate and a lot smaller KS distances compared to T_B . Q-Q plots shown in Figs. 5 and 6 lead to similar conclusions. Therefore, we could confirm that T_γ works reasonably well and performs significantly better than T_B even for a larger model.

6. Conclusions

Browne's ADF test statistic has become very popular in structural equation modeling in applied research. However, it was found empirically that, unless the sample size is very large,

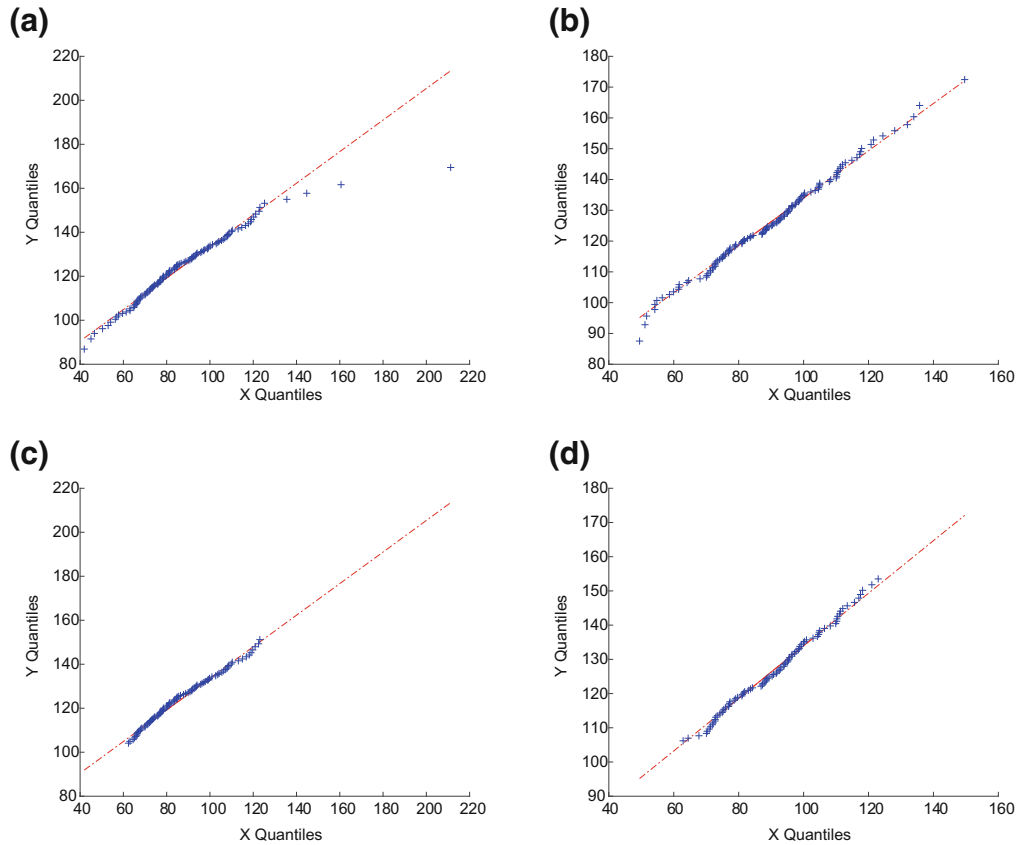


FIGURE 6.

Model 3: Q-Q plot examples of T_{γ} . **a** $df = 82$. **b** $df = 83$. **c** $df = 82$ (only for 1st–99th percentiles). **d** $df = 83$ (only for 1st–99th percentiles).

it could perform poorly in practice. In this paper, we first provide a theoretical explanation of this phenomenon by demonstrating the tendency of ill-conditioning of large covariance matrices. When the $p^* \times p^*$ covariance matrix of sample covariances is ill-conditioned, small variations of the involved estimates produce large variations of the computed ADF statistic. As a result, the corresponding rejection rate is significantly larger than it should be. This theoretical observation is also confirmed by our numerical experiments.

Next, we suggest a modified test statistic that could improve the performance of ADF test statistic and provide an asymptotic analysis. We make approximations while removing small eigenvalues of the corresponding $p^* \times p^*$ covariance matrix, which is in a similar spirit of the principal component analysis. By performing Monte Carlo numerical experiments, we compare the behavior of the proposed test statistic against ADF statistic. We demonstrate that the modified statistic considerably improves the performance of ADF statistic in terms of the empirical rejection rate and KS distances.

Acknowledgments

Funding was provided for the third author by National Science Foundation (Grant No. CMMI1232623).

References

- Amemiya, Y., & Anderson, T. W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *Annals of Statistics*, *18*, 1453–1463.
- Bellman, R. E. (1960). *Introduction to matrix analysis*. New York: McGraw-Hill Book Company.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Boomsma, A., & Hoogland, J. J. (2001). The robustness of lisrel modeling revisited. In *Structural equation modeling: Present and future: A Festschrift in Honor of Karl Jöreskog* (pp. 139–168). Chicago: Scientific Software International.
- Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge: Cambridge University Press.
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62–83.
- Browne, M. W., & Shapiro, A. (1987). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Psychology*, *41*, 193–208.
- Browne, M. W., & Shapiro, A. (2015). Comments on the asymptotics of a distribution-free goodness of fit test statistic. *Psychometrika*, *80*, 196–199.
- Byrne, B. M. (2012). Choosing structural equation modeling computer software: Snapshots of lisrel, eqs, amos, and mplus. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling, chap. 19* (pp. 307–324). New York: Guilford Press.
- Chun, S. Y., & Shapiro, A. (2009). Normal versus noncentral chi-square asymptotics of misspecified models. *Multivariate Behavioral Research*, *44*(6), 803–827.
- Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods and Research*, *26*(3), 329–367.
- Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*, 351–362.
- Huang, Y., & Bentler, P. M. (2015). Behavior of asymptotically distribution free test statistics in covariance versus correlation structure analysis. *Structural Equation Modeling*, *22*, 489–503.
- Jennrich, R., & Satorra, A. (2013). The nonsingularity of γ in covariance structure analysis of nonnormal data. *Psychometrika*, *79*, 51–59.
- Kano, Y. (2002). Variable selection for structural models. *Journal of Statistical Planning and Inference*, *108*(1–2), 173–187.
- Lee, S.-Y. (1990). Multilevel analysis of structural equation models. *Biometrika*, *77*(4), 763–772.
- McManus, D. A. (1991). Who invented local power analysis? *Econometric Theory*, *7*, 265–268.
- Shapiro, A. (1986). Asymptotic theory of overparametrized structural models. *Journal of the American Statistical Association*, *81*, 142–149.
- Tomarken, A., & Waller, N. G. (2005). Structural equation modeling as a data-analytic framework for clinical science: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology*, *1*, 31–65.
- Wu, Hao, & Browne, M. W. (2015). Quantifying adventitious error in a covariance structure as a random effect. *Psychometrika*, *80*(3), 571–600.
- Xu, J., & Mackenzie, G. (2012). Modelling covariance structure in bivariate marginal models for longitudinal data. *Biometrika*, *99*(3), 649–662.
- Yuan, K. H., & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modelling. *British Journal of Mathematical and Statistical Psychology*, *51*, 289–309.
- Yuan, K. H., & Bentler, P. M. (1999). F tests for mean and covariance structure analysis. *Journal of Educational and Behavioral Statistics*, *24*(3), 225–243.
- Yuan, K. H., Hayashi, K., & Bentler, P. M. (2007). Normal theory likelihood ratio statistic for mean and covariance structure analysis under alternative hypotheses. *Journal of Multivariate Analysis*, *98*(6), 1262–1282.

Manuscript Received: 7 DEC 2015

Final Version Received: 12 APR 2017

Published Online Date: 8 JUN 2017