EFFICIENT BOOTSTRAP TESTS FOR THE GOODNESS OF FIT IN COVARIANCE STRUCTURE ANALYSIS

Masanori Ichikawa* and Sadanori Konishi**

Asymptotic chi-square tests, such as the normal theory likelihood ratio test, are often used to evaluate the goodness-of-fit of a covariance structure analysis model. Another approach is to use the bootstrap test, which is known to have the desired asymptotic level if model restrictions are taken into account in designing a resampling algorithm. The bootstrap test is, however, computationally very tedious and the problem of nonconvergence and improper solutions often arise in bootstrap resampling. In this paper, we propose a bootstrap test which is based on an approximation, by a quadratic form, to the minimum value of a discrepancy function calculated from each bootstrap sample. Hence, the proposed bootstrap test is efficient in the sense of the amount of computing needed and is free from the problem of nonconvergence and improper solutions with resampling. A Monte Carlo experiment is conducted to compare the performance of the proposed method with that of asymptotic chi-square tests for each combination of three distributions and four sample sizes.

1. Introduction

A class of covariance structure analysis (CSA) models includes the factor analysis, path analysis, and related structural models that play an essential role in the fields of research such as psychology, education, economics, sociology, etc. (see, for example, Bollen, 1989). An important process in CSA is to assess the goodness-of-fit of a model. For this purpose, asymptotic chi-square tests are often used in practice. When the distribution of observations are multivariate normal, the parameters are estimated by the maximum likelihood method, and the goodness-of-fit of a model is evaluated by using the likelihood ratio test. Under a wide class of nonnormal distributions, the asymptotically distribution free method by Browne (1984), a class of generalized least-squares methods, and associated test statistics can be used.

Another approach to testing goodness-of-fit of a CSA model is to use the bootstrap method introduced by Efron (1979). It is known that if the structural hypothesis concerning a population covariance matrix is taken into account in designing the resampling algorithm, then the bootstrap test has the desired asymptotic level under the conditions that are usually met in practice (Beran & Srivastava 1985). It should be noted, however, that the bootstrap tests are computationally tedious, and there exists the problem of nonconvergence and

Key Words and Phrases: Bootstrap, covariance structure analysis, goodness of fit.

* Tokyo University of Foreign Studies, 3 11-1 Asahi-cho, Fuchu-shi, Tokyo 183-8534, Japan. E-mail: Ichikawa.M@tufs.ac.jp. Request for reprints should be sent to Masanori Ichikawa.

** Graduate School of Mathematics, Kyushu University.
improper solutions with resampling which may require caution for bootstrap inference (Ichikawa & Konishi, 1995).

In this paper, we propose a bootstrap test for the goodness-of-fit of a CSA model: we show that the proposed bootstrap test is efficient in the sense of computational burden and is free from the problem of nonconvergence and improper solutions. We also compare by a Monte Carlo experiment the performance of the proposed bootstrap test with the asymptotic chi-square tests under three distributions and four sample sizes.

2. Testing Goodness-of-fit in Covariance Structure Analysis

Under the CSA model the covariance matrix $\Sigma$ of the observed vector variable $\mathbf{x} = (x_1, \ldots, x_p)'$ can be written as $\Sigma = \Sigma(\theta)$, where $\theta (\in \Theta)$ is a $q \times 1$ vector of parameters. Let $\mathbf{X} = \{x_1, \ldots, x_N\}$ be a random sample of size $N = n + 1$ from the probability distribution $G$ whose covariance matrix is $\Sigma(\theta_0)$. Let $\mathbf{S}$ be the sample covariance matrix

$$\mathbf{S} = \frac{1}{n} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})',$$

where $\bar{x} = N^{-1} \sum_{i=1}^{N} x_i$. In CSA the estimate is usually obtained by minimizing a discrepancy function $F(\mathbf{S}, \Sigma)$, which is a scalar valued function of two $p \times p$ positive definite matrices with the properties (Browne, 1982):

(i) $F(\mathbf{S}, \Sigma) \geq 0$ for all $\mathbf{S}$ and $\Sigma$,
(ii) $F(\mathbf{S}, \Sigma) = 0$ if and only if $\mathbf{S} = \Sigma$,
(iii) $F(\mathbf{S}, \Sigma)$ is a twice continuously differentiable function of $\mathbf{S}$ and $\Sigma$.

An estimate $\hat{\theta}$ of $\theta_0$ may be obtained by minimizing the discrepancy function so that

$$F(\mathbf{S}, \Sigma(\hat{\theta})) = \min \{F(\mathbf{S}, \Sigma(\theta))|\theta \in \Theta\}.$$

The problem of interest here is to test the null hypothesis $H_0 : \Sigma = \Sigma(\theta)$ against the general alternative $H_1 : \Sigma > 0$. Let $\Sigma = \Sigma(\theta)$. Under an appropriate choice of $F$, the asymptotic null distribution of

$$T = n F(\mathbf{S}, \Sigma),$$

the minimum value of $F$ multiplied by $n$, is central chi-square with $d$ degrees of freedom, where $d = p(p + 1)/2 - q$. Hence, the asymptotic chi-square test, $\varphi_A$, of size $\alpha$ can be constructed as

$$\varphi_A = \begin{cases} 
1 & \text{if } T > \chi^2_d(1 - \alpha), \\
0 & \text{otherwise},
\end{cases}$$

where $\chi^2_d(1 - \alpha)$ is the $100(1 - \alpha)$ percentile point of the chi-square distribution with $d$ degrees of freedom.
2.1 Bootstrap test

Another approach to testing goodness-of-fit in CSA is to use the bootstrap developed by Efron (1979). The basic idea behind the bootstrap is to estimate the unknown probability distribution \( G \) by empirical distribution \( \hat{G} \). Let

\[
y_i = \sum^{1/2} S^{-1/2} x_i, \quad (i = 1, \ldots, N).
\]

In our testing problem, \( \hat{G} \) is the distribution that has probability \( 1/N \) on each point \( y_i \). The motivation behind (1) is to take account of the covariance structure implied by the null hypothesis. If \( \hat{G} \) is constructed in this way then the bootstrap test has the desired asymptotic level (Beran & Srivastava, 1985), that is, the probability of type I error converges to the nominal level \( \alpha \) as \( N \) tends to infinity. Let \( Y^* = \{y_1^*, \ldots, y_N^*\} \) be a random sample of size \( N \) from \( \hat{G} \), called the bootstrap sample. In other words, the bootstrap sample \( Y^* \) is a random sample of size \( N \) drawn with replacement from the population of \( N \) objects \( y_1, \ldots, y_N \). Let

\[
S^* = \frac{1}{n} \sum_{i=1}^{N} (\hat{y}^*_i - \bar{y}^*)(\hat{y}^*_i - \bar{y}^*)',
\]

where \( \bar{y}^* = N^{-1} \sum_{i=1}^{N} y_i^* \). Further, let \( \hat{\theta}^* \) be the bootstrap version of the estimate so that

\[
F(S^*, \Sigma(\hat{\theta}^*)) = \min\{F(S^*, \Sigma(\theta))|\theta \in \Theta\}.
\]

In the bootstrap test the sampling distribution of \( T = nF(S, \Sigma) \) based on the random sample from \( G \) is approximated by the bootstrap distribution of \( T^* = nF(S^*, \Sigma^*) \) from \( \hat{G} \), where \( \Sigma^* = \Sigma(\hat{\theta}^*) \). Hence, the bootstrap test, \( \varphi_B \), of size \( \alpha \) can be constructed as

\[
\varphi_B = \begin{cases} 
1 & \text{if } T > T^*(1 - \alpha), \\
0 & \text{otherwise,}
\end{cases}
\]

where \( T^*(1 - \alpha) \) is the 100(1 - \( \alpha \)) percentile point of \( T^* \). In practice, \( T^*(1 - \alpha) \) is approximated by the 100(1 - \( \alpha \)) percentile point of \( B \) realizations \( T_1^*, \ldots, T_B^* \) that are calculated by a Monte Carlo resampling algorithm.

It should be noted, however, that the bootstrap test is computationally tedious; we have to fit the CSA model to each bootstrap sample by minimizing the discrepancy function using an iterative numerical procedure. The number \( B \) of the bootstrap sample required for the bootstrap test is usually 1,000 to 2,000, about ten times as large as that needed for the standard error estimation. Moreover, as noted by Ichikawa and Konishi (1995), there exists the problem of nonconvergence and improper solutions with resampling: the iterative process for finding the minimum value of \( F \) may not terminate within a set number of iterations or the value of \( \theta \) corresponding to the minimum value of \( F \) may be outside of the admissible parameter space \( \Theta \).
2.2 Proposed method

To overcome the difficulties associated with the ordinary bootstrap test, we propose an efficient bootstrap test. Let \( s = \text{vech} \mathbf{S} \) and \( \sigma(\theta_0) = \text{vech} \mathbf{\Sigma}(\theta_0) \) be vectors consisting of the nonduplicated \( p(p+1)/2 \) elements of \( \mathbf{S} \) and \( \mathbf{\Sigma}(\theta_0) \), respectively. Further, let

\[
\Delta(\theta) = \frac{\partial \text{vech} \mathbf{\Sigma}(\theta)}{\partial \theta^r},
\]

\[
\mathbf{W}(\Sigma_1, \Sigma_2) = \frac{\partial^2 F(\Sigma_1, \Sigma_2)}{\partial \text{vech} \Sigma_1 \partial (\text{vech} \Sigma_2)^r}.
\]

Shapiro (1985) has shown that if the model is identified and certain regularity conditions are met, the minimum value of a discrepancy function can be approximated by a quadratic form as follows

\[
T = T + o_p(1),
\]

where

\[
T = n \{s - \sigma(\theta_0)\}'\{\mathbf{W}_0 - \mathbf{W}_0 \Delta_0 (\Delta_0' \mathbf{W}_0 \Delta_0)^{-1} \Delta_0' \mathbf{W}_0\}\{s - \sigma(\theta_0)\},
\]

(2)

with \( \Delta_0 = \Delta(\theta_0) \) and \( \mathbf{W}_0 = \mathbf{W}(\Sigma(\theta_0), \Sigma(\theta_0)) \), respectively (see Browne, 1974).

Let

\[
\tilde{T} = n \{s^* - \sigma(\theta)\}'\{\tilde{\mathbf{W}} - \tilde{\mathbf{W}} \tilde{\Delta} (\tilde{\Delta}' \tilde{\mathbf{W}} \tilde{\Delta})^{-1} \tilde{\Delta}' \tilde{\mathbf{W}}\}\{s^* - \sigma(\tilde{\theta})\},
\]

(3)

which is obtained, as in the case of the ordinary bootstrap test described above, by replacing \( s, \sigma(\theta_0), \Delta_0, \) and \( \mathbf{W}_0 \) in (2) by their bootstrap version \( s^* = \text{vech} \mathbf{S}^* \), \( \sigma(\theta) = \text{vech} \mathbf{\Sigma}(\theta), \tilde{\Delta} = \Delta(\tilde{\theta}) \), and \( \tilde{\mathbf{W}} = \mathbf{W}(\tilde{\Sigma}, \tilde{\Sigma}) \), respectively. Our approach is to approximate the sampling distribution of \( T \) from \( \mathcal{G} \) by the bootstrap distribution of \( \tilde{T}^* \) from \( \tilde{\mathcal{G}} \). Hence, the proposed bootstrap test, \( \varphi_{\tilde{B}} \), can be constructed as

\[
\varphi_{\tilde{B}} = \begin{cases} 
1 & \text{if } \tilde{T} > \tilde{T}^*(1 - \alpha), \\
0 & \text{otherwise},
\end{cases}
\]

where \( \tilde{T}^*(1 - \alpha) \) is the \( 100(1 - \alpha) \) percentile point of \( \tilde{T}^* \). As in the case of the ordinary bootstrap test described earlier, \( \tilde{T}^*(1 - \alpha) \) is approximated by the \( 100(1 - \alpha) \) percentile point of \( B \) realizations \( \tilde{T}^*_1, \ldots, \tilde{T}^*_B \) that are calculated by a bootstrap resampling algorithm.

Since the matrix of quadratic form in (3) is common to all bootstrap samples and only \( s^* \) depends on each bootstrap sample, we can calculate \( \tilde{T}^* \) without fitting a CSA model to \( \mathbf{S}^* \). As a consequence, the proposed bootstrap test \( \varphi_{\tilde{B}} \) is efficient and is free from the problem of nonconvergence and improper solutions in the bootstrap samples. We did not consider the problem of calculating the power of the proposed bootstrap test. The asymptotic theory or the bootstrap method may be used for this problem. Further work remains to be done from theoretical and practical viewpoints, however.
3. A Monte Carlo Experiment

3.1 Methods

We conducted a Monte Carlo experiment to demonstrate the performance of the proposed test. In our experiment a confirmatory factor analysis model (Jöreskog, 1969) was used as a class of CSA models. Under the model, the vector \( \mathbf{x} \) of the observed variables is expressed as \( \mathbf{x} = \mu + \Lambda \mathbf{f} + \mathbf{e} \), where \( \Lambda = (\lambda_{ik}) \) is a \( p \times m \) matrix of factor loadings, and \( \mathbf{f} = (f_1, \ldots, f_m)' \) and \( \mathbf{e} = (e_1, \ldots, e_p)' \) are unobservable random vectors. The elements of \( \mathbf{f} \) and \( \mathbf{e} \) are called the common factors and the unique factors, respectively. It is assumed that \( \mathbb{E} (\mathbf{f}) = 0 \), \( \mathbb{E} (\mathbf{e}) = 0 \), and \( \mathbb{E} (\mathbf{f} \mathbf{e}' ) = 0 \). It is further assumed that the unique factors are mutually uncorrelated and hence the covariance matrix \( \Psi \) is diagonal with the \( i \)-th diagonal element \( \psi_i (> 0) \). The common factors are, on the contrary, allowed to be correlated with covariance matrix \( \Phi = (\phi_{kl}) \). Then the covariance matrix \( \Sigma \) of \( \mathbf{x} \) can be written as \( \Sigma (\theta) = \Lambda \Phi \Lambda' + \Psi \), where \( \theta = (\psi', \phi', \lambda')' \) and \( \psi, \phi \) and \( \lambda \) are the vectors formed from the free parameters in \( \Psi, \Phi \), and \( \Lambda \), respectively.

In practice, the confirmatory factor analysis model used had 15 observed variables with three common factors; each observed variable was influenced by one, and only one, common factor. The values of the parameters were chosen as the same as those used in Hu, Bentler and Kano (1992). That is, \( \Lambda \) was taken as

\[
\Lambda = \begin{pmatrix}
\lambda_1 & 0 & 0 \\
0 & \lambda_2 & 0 \\
0 & 0 & \lambda_3
\end{pmatrix},
\]

where \( \lambda_1 = \lambda_2 = \lambda_3 = (0.70, 0.70, 0.75, 0.80, 0.80)' \). All the common factors were assumed to have unit variances and the factor correlations were taken to be 0.3, 0.4, and 0.5. The values of the unique variances were taken so that the resulting diagonal elements of \( \Sigma (\theta) \) were equal to unity. Hence, the number \( q \) of free parameters in the model was 33.

The true distribution was chosen as a contaminated normal distribution

\[
(1 - \varepsilon)\mathcal{N}_p (\mathbf{0}, \mathbf{V}) + \varepsilon \mathcal{N}_p (\mathbf{0}, \tau^2 \mathbf{V}),
\]

a class of elliptical distributions, under which the covariance matrix can be written as \( \Sigma (\theta) = \{1 + \varepsilon (\tau^2 - 1)\} \mathbf{V} \), see, for example, Muirhead (1982). Accordingly, \( F \) was chosen as

\[
F (\Sigma, \Sigma (\theta)) = \text{tr} \mathbf{S} \Sigma (\theta)^{-1} - \log |\mathbf{S} \Sigma (\theta)^{-1}| - p,
\]

the discrepancy function of the normal theory maximum likelihood method. If \( \varepsilon = 0.0 \) then the distribution of the observed variables is multivariate normal and \( T = n F (\mathbf{S}, \Sigma) \) has an asymptotic chi-square distribution. If \( \varepsilon \neq 0.0 \) then, from the elliptical distribution theory by Shapiro and Browne (1987), \( T = n F (\mathbf{S}, \Sigma)/\hat{\eta} \)
has an asymptotic chi-square distribution where \( \hat{\eta} = b_{2,p}/\{p(p + 2)\} \) is the estimated kurtosis parameter and \( b_{2,p} \) is the multivariate measure of kurtosis due to Mardia (1970).

Our Monte Carlo experiment was carried out as follows:

1. A random sample \( X \) was generated from the probability distribution \( G \).
2. The confirmatory factor analysis model was fit to \( S \) and the asymptotic chi-square test statistic \( T (= n F(S, \Sigma) \text{ if } \varepsilon = 0.0 \text{ and } = n F(S, \Sigma)/\hat{\eta} \text{ if } \varepsilon \neq 0) \) was calculated.
3. The value of \( T \) was compared with \( \chi^2_{33}(0.95) \), the asymptotic chi-square test \( \varphi_A \).
4. Bootstrap sample was drawn \( B = 1,000 \) times and \( T^* \) was calculated from each bootstrap sample. Then \( T^*(0.95) \) was obtained.
5. The value of \( T \) was compared with \( T^*(0.95) \), the proposed bootstrap test \( \varphi_B \).

The steps from 1 to 5 were repeated \( R = 5,000 \) times. The value of the variance inflation parameter \( \tau^2 \) was chosen as 3.0. The experiments were carried out for all combinations of \( \varepsilon = \{0.0, 0.1, 0.3\} \) and \( n = \{150, 250, 500, 1000, 2000\} \).

3.2 Results

Table 1 shows the frequency of rejecting the null hypothesis by the asymptotic chi-square test \( \varphi_A \) and the proposed bootstrap test \( \varphi_B \) under various conditions. As a whole, the asymptotic chi-square tests rejected the null hypothesis too often; on the contrary, the bootstrap test accepted the null hypothesis too often. It can be seen that the asymptotic chi-square test based on the elliptical distribution theory broke down when the sample size was small. This may be due to the poor performance of the kurtosis parameter estimation.

4. Concluding Remarks

In this paper, we proposed an efficient bootstrap test for the goodness-of-fit in covariance structure analysis. As demonstrated by the numerical experiment, the proposed method generally performed over the asymptotic chi-square tests, in particular, when sample size was small.

An advantage of the proposed method is that it is free from the problem of nonconvergence and improper solutions from bootstrap samples. It is known that the occurrence of this problem is by no means an exception in CSA. For example, Ichikawa and Konishi (1997) reported in their simulation study the average frequency of nonconvergence and improper solutions before obtaining 1,000 proper estimates from bootstrap samples. In the case of an unrestricted factor analysis
model with ten variables and two factors, the average frequency was 4.2 when the observed variables were normally distributed and $N = 50$; it was 75.9 when the common factors and the unique factors were not mutually independent but just uncorrelated. In practice, the frequency could be higher if a wrong model is fit and/or the degree of freedom is small. However, there is no clear solution to the problem of nonconvergence and improper solutions in usual bootstrap tests to date.

Finally, we note that the proposed bootstrap test can be applied to arbitrary minimum discrepancy function methods. In that case, the matrix $W_0$ in (2) and hence $W$ in (3) must be suitably adopted.

Acknowledgements

The authors would like to thank the referees and the Editor for their careful reading of the manuscript and helpful comments.

References


*Received May 2001. Final version accepted October 2001.*