PSYCHOMETRIKA
2013
DOI: 10.1007/S11336-013-9383-8

COMMENT ON THE ASYMPTOTICS OF A DISTRIBUTION-FREE GOODNESS OF FIT TEST STATISTIC

MICHAEL W. BROWNE

OHIO STATE UNIVERSITY

ALEXANDER SHAPIRO

GEORGIA INSTITUTE OF TECHNOLOGY

In a recent article Jennrich and Satorra (Psychometrika 78: 545–552, 2013) showed that a proof by Browne (British Journal of Mathematical and Statistical Psychology 37: 62–83, 1984) of the asymptotic distribution of a goodness of fit test statistic is incomplete because it fails to prove that the orthogonal component function employed is continuous. Jennrich and Satorra (Psychometrika 78: 545–552, 2013) showed how Browne's proof can be completed satisfactorily but this required the development of an extensive and mathematically sophisticated framework for continuous orthogonal component functions. This short note provides a simple proof of the asymptotic distribution of Browne's (British Journal of Mathematical and Statistical Psychology 37: 62–83, 1984) test statistic by using an equivalent form of the statistic that does not involve orthogonal component functions and consequently avoids all complicating issues associated with them.

Key words: covariance structures, statistical test, asymptotics.

A recent paper by Jennrich and Satorra (2013) showed that the proof of asymptotic chi-squaredness of the goodness of fit test statistic in Proposition 4 of Browne (1984) is incomplete. This is because the orthogonal complement function employed in the proof is not shown to be continuous. The completion of the proof proposed in Jennrich and Satorra (2013, p. 550) requires the provision and use of an extensive mathematical framework for continuous orthogonal component functions. We shall provide a simple proof that avoids incorporating an orthogonal component function in the test statistic. Browne (1984) provided two equivalent expressions for this statistic. The first (Browne, 1984, expression (2.20a)) involves an orthogonal component function and was employed by Jennrich and Satorra (2013). The second (Browne, 1984, expression (2.20b)) involves no orthogonal component function and will be employed here. This avoids all complicating issues associated with orthogonal component functions.

We use the following notation and terminology throughout the paper. For a $p \times q$ matrix $\boldsymbol{X}$, of full column rank $q < p$, it is said that a $p \times (p-q)$ matrix $\boldsymbol{X}_c$ is an orthogonal complement of $\boldsymbol{X}$ if $\boldsymbol{X}'\boldsymbol{X}_c = \boldsymbol{0}$ and $\boldsymbol{X}_c$ has full column rank $p-q$. This orthogonal complement is not defined uniquely. For a $p \times p$ symmetric matrix $\boldsymbol{S}$ we denote by $\mathrm{vecs}(\boldsymbol{S})$ the corresponding $p^* \times 1$ column vector formed from the $p^* = p(p+1)/2$ non-duplicated elements of $\boldsymbol{S}$. We use the following matrix identity which holds for any $p \times q$ matrix $\boldsymbol{X}$ of full column rank q, and positive definite $p \times p$ matrix $\boldsymbol{A}$ (e.g. Khatri, 1966, Lemma 1; Rao, 1973, p. 77):

$$\boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{X}\big(\boldsymbol{X}'\boldsymbol{A}^{-1}\boldsymbol{X}\big)^{-1}\boldsymbol{X}'\boldsymbol{A}^{-1} = \boldsymbol{X}_c\big(\boldsymbol{X}_c'\boldsymbol{A}\boldsymbol{X}_c\big)^{-1}\boldsymbol{X}_c'. \tag{1}$$

This identity has frequently been used for deriving asymptotics in the analysis of covariance structures, e.g., Browne (1984, p. 69), Shapiro (1986, p. 145).

Requests for reprints should be sent to Michael W. Browne, Department of Psychology, The Ohio State University, 1835 Neil Avenue, Columbus, OH 43210-1222, USA. E-mail: browne.4@osu.edu

© 2013 The Psychometric Society

Published online: 05 December 2013

Let us recall the basic setting in Browne (1984). Suppose that $\boldsymbol{S}$ is the usual unbiased estimator of a $p \times p$ population covariance matrix $\boldsymbol{\Sigma}_0$ obtained from N independent observations on a $p \times 1$ vector variate which has a distribution with *finite fourth-order moments*. Let $\boldsymbol{s} = \text{vecs}(\boldsymbol{S})$ and $\boldsymbol{\sigma}_0 = \text{vecs}(\boldsymbol{\Sigma}_0)$. It follows by the Central Limit Theorem that $N^{1/2}(\boldsymbol{s} - \boldsymbol{\sigma}_0)$ converges in distribution to a multivariate normal with a null mean vector and $p^* \times p^*$ covariance matrix $\boldsymbol{U}$. The components of matrix $\boldsymbol{U}$ can be estimated by using second and fourth order sample moments of the considered observations (formulas (3.1)–(3.3) in Browne, 1984). So let $\widehat{\boldsymbol{U}}$ be a consistent sample estimate of the population covariance matrix $\boldsymbol{U}$. That is, we assume that $\widehat{\boldsymbol{U}}$ converges in probability to its population value $\boldsymbol{U}$, i.e., $\widehat{\boldsymbol{U}} = \boldsymbol{U} + o_p(1)$. Because of the assumption of finite fourth-order moments, this is ensured by the Law of Large Numbers if the standard sample estimate $\widehat{\boldsymbol{U}}$ is used.

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 Θ. For the sake of simplicity we assume that the model holds for the true (population) covariance matrix $\boldsymbol{\Sigma}_0$, i.e., there exists $\boldsymbol{\theta}_0 \in \Theta$ such that $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$. (A more general case of population drift can be handled in a similar way.) 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 and the matrix $\boldsymbol{\Delta}_0$* has full column rank q.

Consider the following distribution free statistic, introduced in Browne (1984, expression (2.20a)):

$$T = N(\boldsymbol{s} - \hat{\boldsymbol{\sigma}})' \hat{\boldsymbol{\Delta}}_c \big(\hat{\boldsymbol{\Delta}}_c' \widehat{\boldsymbol{U}} \hat{\boldsymbol{\Delta}}_c\big)^{-1} \hat{\boldsymbol{\Delta}}_c' (\boldsymbol{s} - \hat{\boldsymbol{\sigma}}), \tag{2}$$

where $\hat{\boldsymbol{\Delta}}_c$ is an orthogonal complement of matrix $\hat{\boldsymbol{\Delta}}$. By using matrix identity (1) with $\boldsymbol{X} = \hat{\boldsymbol{\Delta}}$ and $\boldsymbol{A} = \widehat{\boldsymbol{U}}$, we can write the statistic T in the following equivalent form (Browne, 1984, expression (2.20b)):

$$T = N(\boldsymbol{s} - \hat{\boldsymbol{\sigma}})' \big[\widehat{\boldsymbol{U}}^{-1} - \widehat{\boldsymbol{U}}^{-1} \hat{\boldsymbol{\Delta}} \big(\hat{\boldsymbol{\Delta}}' \widehat{\boldsymbol{U}}^{-1} \hat{\boldsymbol{\Delta}}\big)^{-1} \hat{\boldsymbol{\Delta}}' \widehat{\boldsymbol{U}}^{-1}\big] (\boldsymbol{s} - \hat{\boldsymbol{\sigma}}). \tag{3}$$

It follows that the T statistic, written in the form (2), does not depend on a particular choice of the orthogonal complement matrix $\hat{\boldsymbol{\Delta}}_c$.

Now one can proceed basically in the same way as in Browne (1984). We make the following assumption.

(A) The estimator $\hat{\boldsymbol{\theta}}$ is $O_p(N^{-1/2})$-consistent. That is, $N^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is bounded in probability, denoted $N^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_p(1)$.

The above condition means that the estimator $\hat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}_0$ at a rate of $O_p(N^{-1/2})$. In particular this holds if $N^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges in distribution.

Proposition 1. *Suppose that assumption* (A) *holds. Then*

$$T = N(\boldsymbol{s} - \boldsymbol{\sigma}_0)' \big[\boldsymbol{U}^{-1} - \boldsymbol{U}^{-1} \boldsymbol{\Delta}_0 \big(\boldsymbol{\Delta}_0' \boldsymbol{U}^{-1} \boldsymbol{\Delta}_0\big)^{-1} \boldsymbol{\Delta}_0' \boldsymbol{U}^{-1}\big] (\boldsymbol{s} - \boldsymbol{\sigma}_0) + o_p(1). \tag{4}$$

Proof: Since $\boldsymbol{\sigma}(\cdot)$ is continuously differentiable we can write

$$\boldsymbol{\sigma}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\sigma}(\boldsymbol{\theta}_0) = \boldsymbol{\Delta}_0(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o\big(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|\big). \tag{5}$$

Together with assumption (A) this implies that

$$\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0 = \boldsymbol{\Delta}_0\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) + o_p\big(N^{-1/2}\big). \tag{6}$$

Denote by $\boldsymbol{\Delta}_c$ an orthogonal complement of matrix $\boldsymbol{\Delta}_0$. Recall that $\widehat{\boldsymbol{U}} = \boldsymbol{U} + o_p(1)$. Also since $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$ and $\boldsymbol{\Delta}(\cdot)$ is continuous, it follows that $\hat{\boldsymbol{\Delta}} = \boldsymbol{\Delta}_0 + o_p(1)$. Together with (6) this implies

$$\begin{aligned}
&\big[\widehat{\boldsymbol{U}}^{-1} - \widehat{\boldsymbol{U}}^{-1}\hat{\boldsymbol{\Delta}}\big(\hat{\boldsymbol{\Delta}}'\widehat{\boldsymbol{U}}^{-1}\hat{\boldsymbol{\Delta}}\big)^{-1}\hat{\boldsymbol{\Delta}}'\widehat{\boldsymbol{U}}^{-1}\big](\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0) \\
&\quad = \big[\boldsymbol{U}^{-1} - \boldsymbol{U}^{-1}\boldsymbol{\Delta}_0\big(\boldsymbol{\Delta}_0'\boldsymbol{U}^{-1}\boldsymbol{\Delta}_0\big)^{-1}\boldsymbol{\Delta}_0'\boldsymbol{U}^{-1}\big]\boldsymbol{\Delta}_0\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) + o_p\big(N^{-1/2}\big).
\end{aligned} \tag{7}$$

Hence by the identity (1) and since $\boldsymbol{\Delta}_c'\boldsymbol{\Delta}_0 = \mathbf{0}$, we can write

$$\begin{aligned}
&\big[\widehat{\boldsymbol{U}}^{-1} - \widehat{\boldsymbol{U}}^{-1}\hat{\boldsymbol{\Delta}}\big(\hat{\boldsymbol{\Delta}}'\widehat{\boldsymbol{U}}^{-1}\hat{\boldsymbol{\Delta}}\big)^{-1}\hat{\boldsymbol{\Delta}}'\widehat{\boldsymbol{U}}^{-1}\big](\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0) \\
&\quad = \boldsymbol{\Delta}_c(\boldsymbol{\Delta}_c\boldsymbol{U}\boldsymbol{\Delta}_c)^{-1}\boldsymbol{\Delta}_c'\boldsymbol{\Delta}_0\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) + o_p\big(N^{-1/2}\big) = o_p\big(N^{-1/2}\big).
\end{aligned} \tag{8}$$

Consequently

$$\begin{aligned}
T &= N(\boldsymbol{s} - \hat{\boldsymbol{\sigma}})'\big[\widehat{\boldsymbol{U}}^{-1} - \widehat{\boldsymbol{U}}^{-1}\hat{\boldsymbol{\Delta}}\big(\hat{\boldsymbol{\Delta}}'\widehat{\boldsymbol{U}}^{-1}\hat{\boldsymbol{\Delta}}\big)^{-1}\hat{\boldsymbol{\Delta}}'\widehat{\boldsymbol{U}}^{-1}\big](\boldsymbol{s} - \hat{\boldsymbol{\sigma}}) \\
&= N(\boldsymbol{s} - \boldsymbol{\sigma}_0 + \boldsymbol{\sigma}_0 - \hat{\boldsymbol{\sigma}})'\big[\widehat{\boldsymbol{U}}^{-1} - \widehat{\boldsymbol{U}}^{-1}\hat{\boldsymbol{\Delta}}\big(\hat{\boldsymbol{\Delta}}'\widehat{\boldsymbol{U}}^{-1}\hat{\boldsymbol{\Delta}}\big)^{-1}\hat{\boldsymbol{\Delta}}'\widehat{\boldsymbol{U}}^{-1}\big](\boldsymbol{s} - \boldsymbol{\sigma}_0 + \boldsymbol{\sigma}_0 - \hat{\boldsymbol{\sigma}}) \\
&= N(\boldsymbol{s} - \boldsymbol{\sigma}_0)'\big[\boldsymbol{U}^{-1} - \boldsymbol{U}^{-1}\boldsymbol{\Delta}_0\big(\boldsymbol{\Delta}_0'\boldsymbol{U}^{-1}\boldsymbol{\Delta}_0\big)^{-1}\boldsymbol{\Delta}_0'\boldsymbol{U}^{-1}\big](\boldsymbol{s} - \boldsymbol{\sigma}_0) + o_p(1).
\end{aligned} \tag{9}$$

This completes the proof. □

Using identity (1) we can also write (4) as

$$T = N(\boldsymbol{s} - \boldsymbol{\sigma}_0)'\boldsymbol{\Delta}_c(\boldsymbol{\Delta}_c\boldsymbol{U}\boldsymbol{\Delta}_c)^{-1}\boldsymbol{\Delta}_c'(\boldsymbol{s} - \boldsymbol{\sigma}_0) + o_p(1), \tag{10}$$

where $\boldsymbol{\Delta}_c$ can be *any* orthogonal complement of matrix $\boldsymbol{\Delta}_0$.

It may be noted that, up to the asymptotically negligible term $o_p(1)$, the right hand side of (4) is the optimal value of the standard Generalized Least Squares problem

$$\min_{\boldsymbol{\zeta} \in \mathbb{R}^{p^*}} (\boldsymbol{Z} - \boldsymbol{\Delta}_0\boldsymbol{\zeta})'\boldsymbol{U}^{-1}(\boldsymbol{Z} - \boldsymbol{\Delta}_0\boldsymbol{\zeta}), \tag{11}$$

where $\boldsymbol{Z} = N^{1/2}(\boldsymbol{s} - \boldsymbol{\sigma}_0)$. Since $\boldsymbol{Z}$ asymptotically has a normal distribution with a null mean vector and covariance matrix $\boldsymbol{U}$, it follows that the optimal value of (11) asymptotically has a chi-square distribution with $p^* - q$ degrees of freedom (e.g., Seber, 1977, p. 64). Consequently because of (4) and by the Slutsky theorem we obtain the following result.

Proposition 2. *Suppose that assumption* (A) *holds. Then the test statistic T converges in distribution to a chi-square distribution with $p^* - q$ degrees of freedom.*

The question of existence of a continuous selection of an orthogonal complement matrix was raised in Jennrich and Satorra (2013, p. 548): "A basic question is: Does there exist a continuous orthogonal complement function that can be evaluated? We suspect not." The following observation may be made. Let $\bar{\boldsymbol{X}}$ be a $p \times q$ matrix of full column rank $q < p$, and $\bar{\boldsymbol{X}}_c$ be an orthogonal complement of $\bar{\boldsymbol{X}}$. Consider the following matrix valued function:

$$\boldsymbol{F}(\boldsymbol{X}) = \big[\boldsymbol{I}_p - \boldsymbol{X}\big(\boldsymbol{X}'\boldsymbol{X}\big)^{-1}\boldsymbol{X}'\big]\bar{\boldsymbol{X}}_c. \tag{12}$$

Since there is a neighborhood of $\bar{\boldsymbol{X}}$ such that every matrix in that neighborhood has full column rank q, and hence $\boldsymbol{X}'\boldsymbol{X}$ is invertible, this function is well defined in a neighborhood of $\bar{\boldsymbol{X}}$. Clearly $\boldsymbol{F}(\bar{\boldsymbol{X}}) = \bar{\boldsymbol{X}}_c$, $\boldsymbol{X}'\boldsymbol{F}(\boldsymbol{X}) = \mathbf{0}$ and $\boldsymbol{F}(\boldsymbol{X})$ is continuous for all $\boldsymbol{X}$ in a neighborhood of $\bar{\boldsymbol{X}}$. That is, formula (12) gives a *closed form* expression for a continuous selection for the orthogonal complement matrix. It may be noted that the matrix $\boldsymbol{P} = \boldsymbol{I}_p - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is just the orthogonal projection matrix onto the null space of matrix $\boldsymbol{X}$.

References

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical & Statistical Psychology*, *37*, 62–83.

Jennrich, R., & Satorra, A. (2013). Continuous orthogonal complement functions and distribution-free goodness of fit tests in moment structure analysis. *Psychometrika*, *78*, 545–552.

Khatri, C.G. (1966). A note on a MANOVA model applied to a problem in growth curves. *Annals of the Institute of Statistical Mathematics*, *18*, 75–86.

Rao, C.R. (1973). *Linear statistical inference and its applications*. New York: Wiley.

Seber, G.A.F. (1977). *Linear regression analysis*. New York: Wiley.

Shapiro, A. (1986). Asymptotic theory of overparameterized structural models. *Journal of the American Statistical Association*, *81*, 142–149.

Manuscript Received: 30 APR 2013