

Normal Versus Noncentral Chi-square Asymptotics of Misspecified Models

So Yeon Chun and Alexander Shapiro
*School of Industrial and Systems Engineering
Georgia Institute of Technology*

The noncentral chi-square approximation of the distribution of the likelihood ratio (LR) test statistic is a critical part of the methodology in structural equation modeling. Recently, it was argued by some authors that in certain situations normal distributions may give a better approximation of the distribution of the LR test statistic. The main goal of this article is to evaluate the validity of employing these distributions in practice. Monte Carlo simulation results indicate that the noncentral chi-square distribution describes behavior of the LR test statistic well under small, moderate, and even severe misspecifications regardless of the sample size (as long as it is sufficiently large), whereas the normal distribution, with a bias correction, gives a slightly better approximation for extremely severe misspecifications. However, neither the noncentral chi-square distribution nor the theoretical normal distributions give a reasonable approximation of the LR test statistics under extremely severe misspecifications. Of course, extremely misspecified models are not of much practical interest. We also use the Thurstone data (Thurstone & Thurstone, 1941) from a classic study of mental ability for our illustration.

It is well recognized that no model can represent real data exactly (e.g., see Browne & Cudeck, 1993). Therefore, even reasonably good models are often rejected for larger sample sizes by standard test statistics. This motivated investigations of the statistical properties of test statistics under alternative hypotheses. A classical result states that under a sequence of local alternatives, that is, the so-called population drift, and certain regularity conditions, likelihood ratio (LR)

Correspondence concerning this article should be addressed to So Yeon Chun, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205. E-mail: gth586a@mail.gatech.edu

test statistics asymptotically have a noncentral chi-square distribution. Thus, the noncentral chi-square distribution is widely used for model evaluation and power analysis of testing in structural equation modeling (SEM). In practice this means that rather than assuming an exact fit of the data to a considered model, one can estimate the population discrepancy with the model by employing an estimate of the corresponding noncentrality parameter. Usage of noncentral chi-square asymptotics has a long history in the statistics literature (e.g., see McManus, 1991, for a historical overview). In the analysis of covariance (moment) structures it goes back to Shapiro (1983) and Steiger, Shapiro, & Browne (1985).

One of the criticisms of this approach is that the assumption of the population drift, where the population covariance matrix is assumed to depend on the sample size, is unrealistic. Recently this issue was discussed in a number of publications with a suggestion that the normal distribution could sometimes be a better alternative for approximating the true distribution of the LR test statistics (e.g., Golden, 2003; Olsson, Foss, & Breivik, 2004; Yuan, 2008; Yuan, Hayashi, & Bentler, 2007).

In this article, we empirically compare the noncentral chi-square distribution with the normal distribution in describing the behavior of the LR test statistics T_{ML} under a variety of sample sizes and model misspecifications. Our simulation results may be of some practical assistance to researchers facing model evaluation so that they can derive reasonable inferences.

This article is organized as follows: Theoretical background regarding noncentral chi-square and normal approximations is given in the next section. Then the results of Monte Carlo experiments aimed at evaluation of the appropriateness of using the noncentral chi-square and normal distributions for LR test statistics are given. In particular, the Kolmogorov-Smirnov distance and quantile-quantile (QQ) plots are provided as measures of the distributions' fit. We also use the Thurstone data (Thurstone & Thurstone, 1941) from a classic study of mental ability for our illustration. The Discussion section gives some remarks and suggestions for future directions of research.

THEORETICAL BACKGROUND

Let us start with a critical look at the noncentral chi-square distribution. Let Y_1, \dots, Y_k be a sequence of independent random variables having normal distributions with standard deviation 1 and respective means μ_1, \dots, μ_k , i.e., $Y_i \sim N(\mu_i, 1)$, $i = 1, \dots, k$. Then the random variable $V = Y_1^2 + \dots + Y_k^2$ has noncentral chi-square distribution with k degrees of freedom and noncentrality parameter $\delta = \mu_1^2 + \dots + \mu_k^2$, denoted $V \sim \chi_k^2(\delta)$. Note that the distribution of V depends only on the sum $\mu_1^2 + \dots + \mu_k^2$ and not on the individual means μ_i . Therefore we can assume that $\mu_1 = \mu$ and $\mu_2 = \dots = \mu_k = 0$. In that case

$\delta = \mu^2$ and

$$V = (Z_1 + \mu)^2 + Z_2^2 + \dots + Z_k^2 = \underbrace{Z_1^2 + Z_2^2 + \dots + Z_k^2}_W + 2\mu Z_1 + \mu^2, \quad (1)$$

where $Z_i \sim N(0, 1)$ are independent standard normal random variables.

The right-hand side of Equation (1) can be considered the sum of two components, namely, the sum $W = Z_1^2 + \dots + Z_k^2$, which has a (central) chi-square distribution with k degrees of freedom, and the term $2\mu Z_1 + \mu^2$, which has normal distribution $N(\mu^2, 4\mu^2)$. Moreover, variables Z_1^2 and Z_1 are uncorrelated, and hence these two terms are uncorrelated. Recall that the expected value of W is k and its variance is $2k$. For large values of the noncentrality parameter δ , the term $2\mu Z_1 + \mu^2$ becomes dominant and hence the corresponding noncentral chi-square distribution could be well approximated by the normal distribution $N(k + \delta, 2k + 4\delta)$. It also could be noted that the random variable W is given by the sum of k independent identically distributed random variables, and hence by the Central Limit Theorem its distribution approaches normal with increase of the number of degrees of freedom k . In other words, a noncentral chi-square distribution can be well approximated by the respective normal distribution if the number of degrees of freedom k is large even if the noncentrality parameter δ is small or even zero. That is, a noncentral chi-square distribution can be approximately normal if either the noncentrality parameter is large or the number of degrees of freedom is large or both.

Consider a covariance structure model $\Sigma = \Sigma(\theta)$ relating parameter vector $\theta \in \mathbb{R}^q$ to $p \times p$ population covariance matrix. Let X_1, \dots, X_n be a random sample from the considered population, and $S = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'$ be the corresponding sample covariance matrix. Recall that S is an unbiased estimate of the population covariance matrix Σ_0 . The popular test statistic for testing the model is $T_{ML} = n\hat{F}_{ML}$, where

$$\hat{F}_{ML} = \min_{\theta} F_{ML}(S, \Sigma(\theta)) \quad (2)$$

and

$$F_{ML}(S, \Sigma) = \log |\Sigma| + \text{tr}(S \Sigma^{-1}) - \log |S| - p. \quad (3)$$

We say that the *normality assumption* holds if the population, from which the random sample is drawn, has *normal* distribution, that is, $X_i \sim N(\mu, \Sigma_0)$, $i = 1, \dots, n$. In that case $\frac{n-1}{n}S$ becomes the Maximum Likelihood estimator¹ of the population covariance matrix and T_{ML} becomes the corresponding likelihood

¹Of course, for large n the factor $\frac{n-1}{n}$ is close to one, and for asymptotic results this correction does not matter.

ratio test statistic. This is why T_{ML} is referred to as the ML test statistic. Of course, this test statistic can be computed whether or not the population distribution is normal. We discuss this point later.

The classical result, going back to Wilks (1938), is that if the model is correct, that is, $\Sigma_0 = \Sigma(\theta_0)$ for some value θ_0 of the parameter vector, then under the normality assumption and mild regularity conditions the asymptotic distribution of the test statistic T_{ML} is central chi-square with $df = p(p+1)/2 - q$ degrees of freedom. Let us briefly outline arguments behind this theoretical result. Consider the function

$$f(\mathbf{Z}) = \min_{\theta} F_{ML}(\mathbf{Z}, \Sigma(\theta)) \tag{4}$$

of a $p \times p$ positive definite symmetric matrix variable \mathbf{Z} . Note that here \mathbf{Z} is a general (matrix valued) variable whereas \mathbf{S} denotes the sample covariance matrix, so that for $\mathbf{Z} = \mathbf{S}$ we have that $\hat{F}_{ML} = f(\mathbf{S})$.

In the subsequent analysis we use notation $\mathbf{s}, \boldsymbol{\sigma}, \mathbf{z}$, for the $p^2 \times 1$ dimensional vectors² obtained by stacking columns of the respective matrices $\mathbf{S}, \Sigma, \mathbf{Z}$, that is, $\mathbf{s} = \text{vec}(\mathbf{S})$, and so on. Observe that the ML discrepancy function F_{ML} has the following properties: For any positive definite symmetric matrices \mathbf{Z} and Σ , it holds that $F_{ML}(\mathbf{Z}, \Sigma) \geq 0$ and $F_{ML}(\mathbf{Z}, \Sigma) = 0$ if and only if $\mathbf{Z} = \Sigma$. This implies that $f(\mathbf{z}) \geq 0$ for any \mathbf{z} , and $f(\mathbf{z}) = 0$ for $\mathbf{z} = \boldsymbol{\sigma}_0$. That is, if the model is correct, then the function $f(\mathbf{z})$ attains its minimum (equal zero) at $\mathbf{z} = \boldsymbol{\sigma}_0$, and hence vector $\partial f(\boldsymbol{\sigma}_0)/\partial \mathbf{z}$, of partial derivatives at $\mathbf{z} = \boldsymbol{\sigma}_0$, is zero. By using the *second order* Taylor expansion of $f(\mathbf{z})$ at the point $\mathbf{z} = \boldsymbol{\sigma}_0$, we can approximate

$$f(\mathbf{s}) \approx f(\boldsymbol{\sigma}_0) + (\mathbf{s} - \boldsymbol{\sigma}_0)' [\partial f(\boldsymbol{\sigma}_0)/\partial \mathbf{z}] + (\mathbf{s} - \boldsymbol{\sigma}_0)' \mathbf{Q} (\mathbf{s} - \boldsymbol{\sigma}_0), \tag{5}$$

where $\mathbf{Q} = \frac{1}{2} \partial^2 f(\boldsymbol{\sigma}_0)/\partial \mathbf{z} \partial \mathbf{z}'$ is half the Hessian matrix of second order partial derivatives of $f(\mathbf{z})$ at $\mathbf{z} = \boldsymbol{\sigma}_0$. Because $T_{ML} = nf(\mathbf{s})$ and by the aforementioned the first two terms $f(\boldsymbol{\sigma}_0)$ and $(\mathbf{s} - \boldsymbol{\sigma}_0)' [\partial f(\boldsymbol{\sigma}_0)/\partial \mathbf{z}]$ in the earlier expansion vanish, it follows that

$$T_{ML} \approx [n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}_0)]' \mathbf{Q} [n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}_0)]. \tag{6}$$

Now by the Central Limit Theorem (CLT) we have that $n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}_0)$ converges in distribution to a (multivariate) normal³ with zero mean vector and a covariance

²Note that because matrices $\mathbf{S}, \Sigma, \mathbf{Z}$ are symmetric, the corresponding $p^2 \times 1$ dimensional vectors have no more than $p(p+1)/2$ nonduplicated elements. We use here the respective $p^2 \times 1$, rather than $p(p+1)/2 \times 1$, dimensional vectors for the sake of an algebraic convenience. Note also the corresponding gradient vectors $\partial f(\boldsymbol{\sigma})/\partial \mathbf{z}$ have the same structure of duplicated components.

³For this to hold we only need to verify that the population distribution has finite fourth order moments.

matrix Γ , given by

$$\Gamma = \mathbb{E}\{\text{vec}[(X_i - \mu)(X_i - \mu)']\text{vec}'[(X_i - \mu)(X_i - \mu)']\} - \sigma_0\sigma_0'. \quad (7)$$

This implies that T_{ML} converges in distribution to the distribution of the quadratic form $Y'QY$, where Y is a random vector having normal $N(\mathbf{0}, \Gamma)$ distribution. If the population has normal distribution, then the matrix Γ has a specific structure, which is a function of the covariance matrix Σ_0 alone, that is, does not involve calculation of fourth order moments of the population distribution. We denote this matrix by Γ_N in order to emphasize that it is computed under the assumption of normality. The point is that under the normality assumption and standard regularity conditions, we have here that $Q\Gamma_N Q = Q$ and matrix Q has rank $p(p + 1)/2 - q$. Then invoking some algebraic manipulations it is possible to show that the distribution of the quadratic form $Y'QY$ is (central) chi-square with $df = p(p + 1)/2 - q$ degrees of freedom (cf. Shapiro, 1983, Theorem 5.5).

It is worthwhile to point out the following: In this derivation the only place where the assumption about normality of the population distribution was used is verification of the equation $Q\Gamma_N Q = Q$, which is based on a particular structure of the covariance matrix Γ_N . In some cases this equation can be verified, and hence asymptotic chi-squaredness of the distribution of T_{ML} can be established, even without the normality assumption. This is a basis of the so-called asymptotic robustness theory of the ML discrepancy test statistic (cf. Browne & Shapiro, 1988).

Suppose now that the model is *misspecified*, that is, the population covariance matrix Σ_0 is different from $\Sigma(\theta)$ for any value of the parameter vector θ . We still have that $T_{ML} = nf(s)$ with function $f(\cdot)$ defined in Equation (4) and, by the CLT, $n^{1/2}(s - \sigma_0)$ converges in distribution to (multivariate) normal $N(\mathbf{0}, \Gamma)$. However, now the term

$$F_{ML}^* = \min_{\theta} F(\Sigma_0, \Sigma(\theta)), \quad (8)$$

representing the discrepancy between the *population* value Σ_0 of the covariance matrix and the model, is strictly positive. Consequently, the first term $f(\sigma_0) = F_{ML}^*$ in the second order Taylor expansion, given in the right-hand side of Equation (5), does not vanish and is strictly positive. It follows that for large n the statistic T_{ML} can be approximated by nF_{ML}^* and will grow to infinity as $n \rightarrow \infty$. A more precise statement is that $n^{-1}T_{ML} = \hat{F}_{ML}$ converges with probability one (w.p.1) to F_{ML}^* . Also by employing the *first order* Taylor expansion at the point $\mathbf{z} = \sigma_0$, that is, by using first two terms in the right-hand side of Equation (5), we can write

$$n^{1/2}[f(s) - f(\sigma_0)] \approx [n^{1/2}(s - \sigma_0)]'[\partial f(\sigma_0)/\partial \mathbf{z}]. \quad (9)$$

It is possible to show that

$$\frac{\partial f(\boldsymbol{\sigma}_0)}{\partial \mathbf{z}} = \left. \frac{\partial F_{ML}(\mathbf{z}, \boldsymbol{\sigma}^*)}{\partial \mathbf{z}} \right|_{\mathbf{z}=\boldsymbol{\sigma}_0}, \tag{10}$$

where $\boldsymbol{\sigma}^* = \boldsymbol{\sigma}(\boldsymbol{\theta}^*)$ and $\boldsymbol{\theta}^*$ is the minimizer of the function $F_{ML}(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$, provided that this minimizer is unique (Equation (10) follows by the so-called Danskin Theorem). Recalling that $f(s) = \hat{F}_{ML}$ and $f(\boldsymbol{\sigma}_0) = F_{ML}^*$ and that $n^{1/2}(s - \boldsymbol{\sigma}_0)$ converges in distribution to $Y \sim N(\mathbf{0}, \boldsymbol{\Gamma})$, we obtain that $n^{1/2}(\hat{F}_{ML} - F_{ML}^*)$ converges in distribution to $\boldsymbol{\gamma}'Y \sim N(0, \boldsymbol{\gamma}'\boldsymbol{\Gamma}\boldsymbol{\gamma})$, where

$$\boldsymbol{\gamma} = \left. \frac{\partial F_{ML}(\mathbf{z}, \boldsymbol{\sigma}^*)}{\partial \mathbf{z}} \right|_{\mathbf{z}=\boldsymbol{\sigma}_0} = \text{vec} [(\boldsymbol{\Sigma}^*)^{-1} - \boldsymbol{\Sigma}_0^{-1}]. \tag{11}$$

This implies the following result (see Shapiro, 1983, Sec. 5, for technical details):

- Let $\boldsymbol{\theta}^*$ be the unique minimizer of $F_{ML}(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$. Then $n^{1/2}(\hat{F}_{ML} - F_{ML}^*)$ converges in distribution to normal $N(0, \boldsymbol{\gamma}'\boldsymbol{\Gamma}\boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ is given in Equation (11) and $\boldsymbol{\sigma}^* = \boldsymbol{\sigma}(\boldsymbol{\theta}^*)$. In other words we can approximate the distribution of $T_{ML} = n\hat{F}_{ML}$ by the normal distribution with mean nF_{ML}^* and variance $n\boldsymbol{\gamma}'\boldsymbol{\Gamma}\boldsymbol{\gamma}$.

The (asymptotic) covariance matrix $\boldsymbol{\Gamma}$ depends on the population distribution. In particular, if the population distribution is normal, then (cf. Shapiro, 2009)

$$\begin{aligned} \boldsymbol{\gamma}'\boldsymbol{\Gamma}_N\boldsymbol{\gamma} &= 2 \text{tr} \left[(\boldsymbol{\Sigma}^{*-1} - \boldsymbol{\Sigma}_0^{-1}) \boldsymbol{\Sigma}_0 (\boldsymbol{\Sigma}^{*-1} - \boldsymbol{\Sigma}_0^{-1}) \boldsymbol{\Sigma}_0 \right] \\ &= 2 \text{tr} \left[(\boldsymbol{\Sigma}^{*-1} \boldsymbol{\Sigma}_0 - \mathbf{I}_p)^2 \right]. \end{aligned} \tag{12}$$

If the population distribution is normal, and hence T_{ML} becomes the likelihood ratio test statistic, then the aforementioned result can also be derived from Vuong (1989). Note, however, that the aforementioned asymptotic normality of T_{ML} holds even without the normality assumption, although in that case the right-hand side of Equation (12) may be not a correct formula for the asymptotic variance $\boldsymbol{\gamma}'\boldsymbol{\Gamma}\boldsymbol{\gamma}$. We discuss this issue further later.

Theoretically this is a correct result. However, in any real application the question is, “How good is this normal approximation for a *finite* sample?” Let us point to the obvious deficiencies of the normal approximation. Any normal distribution is symmetric around its mean. On the other hand, as it was mentioned earlier, the test statistic T_{ML} is always nonnegative and its distribution is typically skewed, especially when $\boldsymbol{\Sigma}_0$ is not “too far” from the model and hence the (population) discrepancy F_{ML}^* is close to zero. In the extreme case when the

model is correct, we have that $F_{ML}^* = 0$ and $\boldsymbol{\gamma} = \mathbf{0}$, and hence the normal approximation, of $n^{1/2}\hat{F}_{ML}$, degenerates into the identically zero distribution. This should be not surprising because in that case T_{ML} converges (in distribution) to a finite limit and hence $n^{1/2}\hat{F}_{ML} = n^{-1/2}T_{ML}$ tends (in probability) to zero. Of course, our primary interest is in situations when the fit is not “too bad,” and this is exactly where the normal approximation may not work well. Another deficiency of the earlier construction of normal approximation is that it is based on the *first order* Taylor expansion and does not take into account the third (quadratic) term in the right-hand side of Equation (5). It is possible to make a bias correction based on this quadratic term (cf. Shapiro, 1983, and see later), but the skewness problem may still persist.

In order to resolve these problems we can use the following idea: Instead of a second order Taylor expansion at the population point (covariance matrix) $\boldsymbol{\Sigma}_0$, let us consider the respective expansion at the point $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}(\boldsymbol{\theta}^*)$ satisfying the model. (Recall that $\boldsymbol{\Sigma}^*$ is the closest to $\boldsymbol{\Sigma}_0$, in terms of the F_{ML} discrepancy function, covariance matrix satisfying the considered model.) That is,

$$f(\mathbf{s}) \approx f(\boldsymbol{\sigma}^*) + (\mathbf{s} - \boldsymbol{\sigma}^*)'[\partial f(\boldsymbol{\sigma}^*)/\partial \mathbf{z}] + (\mathbf{s} - \boldsymbol{\sigma}^*)' \mathbf{Q}^*(\mathbf{s} - \boldsymbol{\sigma}^*), \quad (13)$$

where $\mathbf{Q}^* = \frac{1}{2}\partial^2 f(\boldsymbol{\sigma}^*)/\partial \mathbf{z}\partial \mathbf{z}'$. The aforementioned approximation Equation (13) could be reasonable if $\boldsymbol{\sigma}^*$ is close to $\boldsymbol{\sigma}_0$, that is, if the discrepancy between $\boldsymbol{\Sigma}_0$ and the model is not too bad. Again we have that first two terms in the right-hand side of Equation (13) vanish and hence

$$T_{ML} = nf(\mathbf{s}) \approx [n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}^*)]' \mathbf{Q}^*[n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}^*)]. \quad (14)$$

Because \mathbf{S} is an unbiased estimate of $\boldsymbol{\Sigma}_0$, i.e., $\mathbb{E}[\mathbf{S}] = \boldsymbol{\Sigma}_0$, we have that $\mathbb{E}[\mathbf{s} - \boldsymbol{\sigma}^*] = \boldsymbol{\sigma}_0 - \boldsymbol{\sigma}^*$. Therefore we can approximate the distribution of T_{ML} by the distribution of the quadratic form $\mathbf{Y}'\mathbf{Q}^*\mathbf{Y}$, where $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ with $\boldsymbol{\mu} = n^{1/2}(\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}^*)$. This suggests approximating the distribution of T_{ML} by a noncentral chi-square distribution with $df = p(p+1)/2 - q$ degrees of freedom and noncentrality parameter $\delta = n(\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}^*)'\mathbf{Q}^*(\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}^*)$. Again by Equation (13) we have that

$$(\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}^*)'\mathbf{Q}^*(\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}^*) \approx f(\boldsymbol{\sigma}_0) = F_{ML}^*, \quad (15)$$

and hence we can use $\delta = nF_{ML}^*$ as the noncentrality parameter as well. Because $F_{ML}^* > 0$ we have here that the noncentrality parameter δ tends to infinity as $n \rightarrow \infty$. In order to reconcile this problem we may assume that the population value $\boldsymbol{\sigma}_{0,n}$ depends on the sample size n in such a way that $n^{1/2}(\boldsymbol{\sigma}_{0,n} - \boldsymbol{\sigma}^*)$ converges to a fixed limit. This assumption implies that $\boldsymbol{\sigma}_{0,n}$ converges to $\boldsymbol{\sigma}^*$ at a rate of $O(n^{-1/2})$ and referred to as a *sequence of local alternatives* or the *population drift*. Note also that the approximation Equation (15) makes sense

only if σ_0 is close to σ^* , that is, if the misspecification is not “too serious,” and can be poor otherwise (e.g., Sugawara & MacCallum, 1993).

The concept of the population drift is just a mathematical fabrication allowing to make an exact mathematical statement. It could be pointed out, however, that the assumption about existence of an abstract population from which we can sample indefinitely, and hence to arrive at a limiting distribution as the sample size tends to infinity, is also a mathematical abstraction. In practice the sample is always finite, and the real question is how good a considered approximation is for a given sample. This, of course, depends on a particular application. One could be also tempted to use the second order Taylor approximation of the discrepancy function at the *population* point Σ_0 . However, for misspecified models the corresponding quadratic form does not have a (noncentral) chi-square distribution, even under the normality assumption (cf. Shapiro, 1983, Theorem 5.4(c)). Consequently asymptotics based on such approximation could be difficult to use in practice.

The noncentrality parameter $\delta = nF_{ML}^*$ can be large for two somewhat different reasons, namely, it can happen that F_{ML}^* is large, that is, the fit is bad, or that the sample size n is large amplifying a reasonably small discrepancy F_{ML}^* , and of course it could be both. If the noncentrality parameter is large because of the large sample size, whereas F_{ML}^* is reasonably small, then the noncentral chi-square approximation can be still reasonable. As it was discussed at the beginning of this section, for large δ the distribution $\chi_k^2(\delta)$ by itself can be approximately normal.

Let us finally mention that by taking into account the last (quadratic) term in the right-hand side of Equation (5), we can make the following correction for the normal distribution approximation. The expected value of this quadratic term can be approximated by $n^{-1}\text{tr}(\mathbf{\Gamma}\mathbf{Q})$. In order to apply bias correction based on that term one would need to estimate matrices $\mathbf{\Gamma}$ and \mathbf{Q} , which may be not easy and will involve an error in any such estimation. Alternatively the term $\text{tr}(\mathbf{\Gamma}_N\mathbf{Q})$ can be approximated by the number of degrees of freedom $df = p(p+1)/2 - q$. The variance of this quadratic term can be approximated by $n^{-2}(2df + 4\delta)$. Therefore, assuming that the population distribution is normal, we can use the corrected normal distribution approximation of the distribution of T_{ML} with mean $nF_{ML}^* + df = \delta + df$ and variance

$$2n \text{tr} \left[\left(\Sigma^{*-1} \Sigma_0 - I_p \right)^2 \right] + 2df + 4\delta. \quad (16)$$

Similar analysis can be performed for the Generalized Least Squares (GLS) discrepancy function

$$F_{GLS}(\mathbf{S}, \Sigma) = \frac{1}{2} \text{tr} \{ [(\mathbf{S} - \Sigma)\mathbf{S}^{-1}]^2 \}. \quad (17)$$

In that respect it is worthwhile to point out the following: The second order Taylor expansion of the GLS discrepancy function, at a point satisfying the model, coincides with the corresponding second order Taylor expansion of the ML discrepancy function. Therefore, if the model is correct, then the test statistics T_{ML} and T_{GLS} are asymptotically equivalent (cf. Browne, 1974). In that case the numerical values of T_{ML} and T_{GLS} , for a given sample covariance matrix S , should be close to each other. On the other hand for misspecified models, as the population covariance matrix moves away from the model, the test statistics T_{ML} and T_{GLS} diverge and the corresponding estimates of the noncentrality parameter based on these statistics could be quite different from each other. As far as the asymptotic normality is concerned the following result, similar to the ML case, holds like this:

- Let θ^* be the unique minimizer of $F_{GLS}(\Sigma_0, \Sigma(\theta))$ and $\gamma = \left. \frac{\partial F_{GLS}(\mathbf{z}, \Sigma^*)}{\partial \mathbf{z}} \right|_{\mathbf{z}=\sigma_0}$, where $\sigma^* = \sigma(\theta^*)$. Then $n^{1/2}(\hat{F}_{GLS} - F_{GLS}^*)$ converges in distribution to normal $N(0, \gamma' \Gamma \gamma)$.

In particular, if the population distribution is normal, then the asymptotic variance associated with the GLS test statistic is given by the following formula (cf. Shapiro, 2009):

$$\gamma' \Gamma_N \gamma = 2 \operatorname{tr} \left[\left(\Sigma_0^{-1} \Sigma^* \Sigma_0^{-1} \Sigma^* - \Sigma_0^{-1} \Sigma^* \right)^2 \right]. \quad (18)$$

Note that here Σ^* corresponds to the minimizer θ^* of the GLS discrepancy function and vector γ is given by derivatives of the GLS discrepancy function, and formula Equation (18) for the asymptotic variance is different from the corresponding formula Equation (12) for the ML discrepancy function.

Nonnormal Distributions

The asymptotic normality of \hat{F}_{ML} , that is, convergence in distribution of $n^{1/2}(\hat{F}_{ML} - F_{ML}^*)$ to $N(0, \gamma' \Gamma \gamma)$, holds without the assumption that the population has normal distribution as well. The asymptotic variance $\gamma' \Gamma \gamma$ can be estimated directly from the data by using formulas Equations (7) and (11). That is, components of the matrix Γ and vector γ can be estimated by replacing the respective fourth and second order moments with their sample estimates. Note, however, that estimation of matrix Γ involves estimation of $p(p+1)(p+2)(p+3)/4$ distinct fourth order moments, which can result in a significant estimation error. Therefore it could be desirable to consider specific situations where estimation of fourth order moments can be avoided. One such case, other than normal, is the case of elliptical distributions.

Suppose now that the population distribution is elliptical. The elliptical class of distributions incorporates a single additional kurtosis parameter, κ , and is convenient for investigating the sensitivity of normal theory methods to the kurtosis of the population distribution. Note that kurtosis parameter $\kappa = \frac{1}{3}\gamma$, where γ is the (marginal) kurtosis of the multivariate distribution (e.g., Muirhead & Waternaux, 1980). The basic asymptotic result that we need here is that the corresponding matrix $\mathbf{\Gamma}$ has the following structure (e.g., Muirhead & Waternaux, 1980):

$$\mathbf{\Gamma} = (1 + \kappa)\mathbf{\Gamma}_N + \kappa\boldsymbol{\sigma}_0\boldsymbol{\sigma}'_0. \quad (19)$$

Here, as it was defined before, $\mathbf{\Gamma}_N$ is the asymptotic covariance matrix of $n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}_0)$ obtained under the assumption that the population has normal distribution. Consequently

$$\boldsymbol{\gamma}'\mathbf{\Gamma}\boldsymbol{\gamma} = (1 + \kappa)\boldsymbol{\gamma}'\mathbf{\Gamma}_N\boldsymbol{\gamma} + \kappa(\boldsymbol{\gamma}'\boldsymbol{\sigma}_0)^2, \quad (20)$$

where $\boldsymbol{\gamma}'\mathbf{\Gamma}_N\boldsymbol{\gamma}$ is given by the right-hand side of Equation (12) and represents the asymptotic variance of $n^{1/2}(\hat{F}_{ML} - F_{ML}^*)$ under the normality assumption. Also by Equation (11) we have

$$\kappa(\boldsymbol{\gamma}'\boldsymbol{\sigma}_0)^2 = \kappa \left[\text{tr} \left(\boldsymbol{\Sigma}^{*-1} \boldsymbol{\Sigma}_0 - \mathbf{I}_p \right) \right]^2. \quad (21)$$

Let us also note that assuming that the model is invariant under a constant scaling factor, we have here that under a sequence of local alternatives the test statistic $(1 + \kappa)^{-1}T_{ML}$ asymptotically has a noncentral chi-square distribution with $df = p(p + 1)/2 - q$ degrees of freedom and noncentrality parameter $(1 + \kappa)^{-1}\delta$, where $\delta = nF_{ML}^*$ (cf. Shapiro & Browne, 1987). Therefore, similar to Equation (16), we can use the corrected normal distribution approximation of the distribution of T_{ML} with mean $nF_{ML}^* + (1 + \kappa)df$ and variance

$$(1 + \kappa)\boldsymbol{\gamma}'\mathbf{\Gamma}_N\boldsymbol{\gamma} + \kappa \left[\text{tr} \left(\boldsymbol{\Sigma}^{*-1} \boldsymbol{\Sigma}_0 - \mathbf{I}_p \right) \right]^2 + (1 + \kappa)^2(2df + 4\delta). \quad (22)$$

NUMERICAL ILLUSTRATIONS

In this section we discuss Monte Carlo experiments aimed at an empirical evaluation of the suitability of the noncentral chi-square and normal distributions for the LR test statistic. We consider factor analysis models $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$ under varying conditions of model misspecification and sample size. Our study also includes different numbers of variables and factors. Furthermore, we use both

normal and nonnormal (elliptically distributed) data to investigate the robustness of test statistics to nonnormality of the population distribution. Because of the space limitations we present here only a sample of our experimental results. For an extensive list of the respective tables and figures the interested reader is referred to a technical report that can be found on the Web site: <http://ideas.repec.org/p/pramprapa/17310.html>

Normally Distributed Data

Our experiments included six sample sizes— $n = 50, 100, 200, 400, 800, 1,000$ —with various degrees of model misspecification ranging from small to severe.

The population covariance matrices employed in Monte Carlo simulations were constructed as follows. First, a $p \times p$ covariance matrix $\Sigma^* = \Lambda^* \Lambda^{*'} + \Psi^*$, satisfying the Factor Analysis model, was constructed with specific values of elements of matrix Λ^* and diagonal elements of matrix Ψ^* , as shown in Table 1 for Model 1. Model 1 has 7 variables and one factor. Model 2 has 12 variables and three factors. Next, misspecified covariance matrices were generated of the form $\Sigma_0 = \Sigma^* + tE$, where E is a $p \times p$ symmetric matrix and $t > 0$ is a scaling factor controlling the level of misspecification. The matrix E was chosen in such a way that the corresponding matrix Σ_0 is positive definite and $\Sigma^* = \Sigma(\theta^*)$, where θ^* is the minimizer of the right-hand side of Equation (8). That is, for $S = \Sigma_0$ the estimated covariance matrix obtained by applying the ML procedure is the specified matrix Σ^* , and hence $F_{ML}^* = F_{ML}(\Sigma_0, \Sigma^*)$.

In order to construct matrix E , producing a largest possible range of the discrepancy values, we used procedures developed in Cudeck and Browne (1992) and Chun and Shapiro (2008). Given the population covariance matrix Σ_0 , we randomly generated $M = 50,000$ sample covariance matrices, corresponding to the specified population covariance matrix Σ_0 and the sample size n , from the Wishart distribution $W_p\left(\frac{1}{n-1}\Sigma_0, n-1\right)$. We used the Matlab function “wishrnd” to generate random matrices having Wishart distribution. For each covariance matrix, sample values $T_i, i = 1, \dots, M$, for the LR test statistics were calculated

TABLE 1
Generated Parameters for Model 1

Λ^*	0.6916	Ψ^*	0.8727
	1.2404		0.6480
	0.7971		1.0672
	0.9011		1.0614
	0.5761		3.0594
	1.5620		1.8551
	0.8117		1.3567

by ML estimation. Estimation of factor loading matrix Λ was done by Matlab function "factoran."

For Model 1, the maximum discrepancy F_{ML}^* (corresponding to the largest value of the scaling parameter t) was computed to be 1.360. By using different values of the scaling parameter t we generated population covariance matrices, of the form $\Sigma_0 = \Sigma^* + tE$, with discrepancy values in the ranges of 0.025 to 1.360. Similarly, population covariance matrices for Model 2 were generated with discrepancy values from 0.01 to 0.5. Discrepancy misspecification and corresponding population values of RMSEA are shown in Table 2. The RMSEA stands for *Root Mean Square Error of Approximation*, and its (population) value is defined as

$$\text{RMSEA} = \sqrt{\frac{F_{ML}^*}{df}}$$

(cf. Browne & Cudeck, 1992; Steiger & Lind, 1980). In the present case $df = 14$ for Model 1 and $df = 33$ for Model 2.

We compare the noncentral chi-square distribution with the normal distribution for describing the behavior of the ML test statistic $T_{ML} = n\hat{F}_{ML}$. In the text and tables to follow, the *noncentral chi-square distribution* is referred to as *ncx*. For the comparison we specify normal distributions with four different mean and variance values, namely, mean $\delta = nF_{ML}^*$ and variance $n\gamma'\Gamma\gamma$, with $\gamma'\Gamma\gamma$ given in Equation (12) (referred to as *nm*); corrected mean $nF_{ML}^* + df$ and variance given in Equation (16) (referred to as *nm2*); mean and variance estimated directly from the simulated values T_1, \dots, T_M by computing their average and sample variance (referred to as *nm3*); and mean $nF_{ML}^* + df$ and variance $2df + 4\delta$ (referred to as *nm4*). That is, *nm* corresponds to the direct normal approximation, *nm2* corresponds to the normal approximation with the bias correction, *nm3* corresponds to the normal approximation with mean and variance estimated directly from the sample, and *nm4* corresponds to the normal approximation of the respective noncentral chi-square distribution. We refer to *nm*, *nm2*, and *nm4* as *theoretical* normal approximations because their parameters (mean and variance) can be estimated from the data. On the other hand, sample mean and variance used in *nm3* can be computed only in a simulation study.

TABLE 2
Degree of Discrepancy Misspecification for Model 1

F_{ML}^*	0.025	0.090	0.185	0.318	0.474	0.655	0.863	1.097	1.360
RMSEA	0.042	0.080	0.116	0.151	0.184	0.216	0.248	0.280	0.312

We used several discrepancy measures to compare the fit of each distribution. One is the Kolmogorov-Smirnov (K) distance defined as

$$K = \sup_{t \in \mathbb{R}} |\hat{F}_M(t) - F(t)|, \quad (23)$$

where $\hat{F}_M(t) = \frac{\#\{T_i \leq t\}}{M}$ is the empirical cumulative distribution function (cdf) based on Monte Carlo sample T_1, \dots, T_M of M computed values of the test statistic, and $F(t)$ is the theoretical cdf of the respective approximations ncx , nm , $nm2$, $nm3$, and $nm4$ of the test statistic. We also consider the average Kolmogorov-Smirnov distance (AK), defined as

$$AK = \frac{1}{M} \sum_{i=1}^M K_i, \quad (24)$$

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. The computed values of the KS distances are denoted as $ncxK$, nmK , $nm2K$, $nm3K$, and $nm4K$, respectively, and the computed values of the AK distances are denoted as $ncxAK$, $nmAK$, $nm2AK$, $nm3AK$, and $nm4AK$, respectively. These measures were used in Yuan et al. (2007).

Table 3 contains Kolmogorov-Smirnov distances (K) for Model 1 with sample sizes $n = 400$ and $n = 1,000$ and nine degrees of misspecification F_{ML}^* ranging from 0.025 to 1.360. Corresponding $\delta = nF_{ML}^*$ values are from 9.8 to 544 for $n = 400$ and from 24.50 to 1,360 for $n = 1,000$. From this table we can compare the performance of each distribution for different degrees of discrepancy F_{ML}^* for Model 1. We can see that, for small to severe misspecification F_{ML}^* (with respective RMSEA values ranging from 0.042 to 0.116), $ncxK$ is smaller than nmK and $nm2K$, but the status of those measures is reverse for extremely severe misspecifications (with RMSEA values greater than 0.151).

This shows that for small, moderate, and even severe misspecifications, the noncentral distribution gives a better approximation. On the other hand, for extremely severe misspecifications the normal distribution with bias correction ($nm2$) gives a slightly better approximation. However, models with extremely severe misspecifications are rejected anyway, say by the RMSEA criterion, and are not of much practical interest. Moreover, these results indicate that neither noncentral chi-square or theoretical normal is a reasonable approximation for severely misspecified models. For all values of F_{ML}^* , we observe that $ncxK \leq nm4K$, and these values are getting close to each other as F_{ML}^* increases,

TABLE 3
Kolmogorov-Smirnov distance(K) for Model 1, $df = 14$

n	F_{ML}^*	δ	RMSEA	$ncxK$	nmK	$nm2K$	$nm3K$	$nm4K$
400	0.025	9.80	0.042	0.009	0.686	0.063	0.045	0.042
	0.090	36.12	0.080	0.022	0.421	0.051	0.031	0.037
	0.190	75.88	0.116	0.034	0.313	0.041	0.024	0.042
	0.318	127.32	0.151	0.044	0.252	0.037	0.019	0.048
	0.474	189.56	0.184	0.053	0.211	0.038	0.016	0.057
	0.655	262.16	0.216	0.065	0.174	0.047	0.013	0.068
	0.863	345.16	0.248	0.090	0.117	0.080	0.008	0.094
	1.097	438.88	0.280	0.174	0.084	0.170	0.004	0.181
	1.360	544.00	0.312	0.360	0.266	0.327	0.006	0.365
	1,000	0.025	24.50	0.042	0.012	0.487	0.065	0.038
0.090		90.30	0.080	0.025	0.283	0.049	0.024	0.033
0.190		189.70	0.116	0.035	0.217	0.036	0.018	0.039
0.318		318.30	0.151	0.044	0.188	0.030	0.015	0.044
0.474		473.90	0.184	0.051	0.166	0.028	0.012	0.052
0.655		655.40	0.216	0.061	0.143	0.031	0.010	0.061
0.863		862.90	0.248	0.082	0.101	0.059	0.007	0.083
1.097		1,097.20	0.280	0.196	0.145	0.179	0.004	0.200
1.360		1,360.00	0.312	0.490	0.442	0.435	0.005	0.493

^a Kolmogorov-Smirnov distance (K) for different sample sizes n with discrepancy F_{ML}^* and noncentral parameter $\delta = nF_{ML}^*$.

^b ncx stands for $\chi_{df}^2(\delta)$, nm stands for $N(\delta, 2n \text{tr}[(\Sigma^{*-1}\Sigma_0 - I_p)^2])$, $nm2$ stands for $N(\delta + df, 2n \text{tr}[(\Sigma^{*-1}\Sigma_0 - I_p)^2])$, $nm3$ stands for Normal with sample mean and variance, and $nm4$ stands for $N(\delta + df, 2df + 4\delta)$.

implying that for large δ the noncentral chi-square distribution by itself can be approximated by a normal distribution, as it was discussed at the beginning of the section Theoretical Background. Note that the noncentrality parameter $\delta = nF_{ML}^*$ gets larger because the discrepancy F_{ML}^* gets bigger with fixed n here. It also could be noted that for large discrepancies the normal distribution with sample mean and variance (column $nm3K$) gives a good approximation. This, however, is of little practical interest because sample mean and variance could be computed only in simulation experiments.

Average Kolmogorov-Smirnov distance(AK) for Model 1 with sample sizes $n = 400$ and $n = 1,000$ and F_{ML}^* values ranging from 0.025 to 1.36 are calculated as well. The patterns of changes in AK are very similar to those of K in Table 3 except that the respective values are smaller here. This is the result of the different calculation in Equations (23) and (24). Thus, we could get a similar conclusion, namely, the noncentral chi-square and the normal distributions are becoming similar in describing T_{ML} as F_{ML}^* increases, but the noncentral chi-square is better than the normal distribution (nm) or normal

with bias correction (*nm2*) for small, moderate, and severe misspecifications. Again, normal distribution with bias correction is a little better description for the distribution of T_{ML} under extremely severe misspecifications. Note that neither *ncx* nor *nm2* is a reasonable approximation under extremely severe misspecifications.

The results in Table 3 do not tell us much about the effect of the sample size for a fixed discrepancy F_{ML}^* . Table 4 is designed to show the effect of sample size on *AK* for each distribution for Model 1. We present three values of F_{ML}^* for the comparison. The value of the noncentrality parameter $\delta = nF_{ML}^*$ varies from 4.52 to 1,097.20. As we can see, *ncxAK* is smaller than *nmAK*, *nm2AK*, and *nm4K* for all sample sizes n except $n = 50$ when $F_{ML}^* = 0.090$, confirming our analysis. For $F_{ML}^* = 0.474$, normal approximation with bias correction (*nm2*) is slightly better than the noncentral chi-square for the sample size $n \geq 400$. The normal (*nm*) provides a better description on the behavior of T_{ML} when discrepancy is extremely large, that is $F_{ML}^* = 1.097$, but none of the distributions gives a reasonable description for T_{ML} under extremely severe misspecifications. Our simulation results also show that sample size effect was not as important as the degree of misspecification of the model.

TABLE 4
Average Kolmogorov-Smirnov Distance(AK) for Model 1 With Different Sample Sizes,
 $df = 14$

F_{ML}^*	RMSEA	n	δ	<i>ncxAK</i>	<i>nmAK</i>	<i>nm2AK</i>	<i>nm3AK</i>	<i>nm4AK</i>
0.090	0.080	50	4.52	0.036	0.483	0.021	0.026	0.020
		100	9.03	0.011	0.439	0.026	0.025	0.019
		200	18.06	0.010	0.369	0.030	0.022	0.020
		400	36.12	0.022	0.421	0.051	0.031	0.037
		800	72.24	0.014	0.213	0.030	0.013	0.018
		1,000	90.30	0.025	0.283	0.049	0.024	0.033
0.474	0.184	50	23.70	0.019	0.326	0.028	0.018	0.028
		100	47.39	0.028	0.234	0.030	0.015	0.036
		200	94.78	0.028	0.170	0.024	0.011	0.032
		400	189.56	0.053	0.211	0.038	0.016	0.057
		800	379.12	0.029	0.092	0.017	0.007	0.030
		1,000	473.90	0.051	0.166	0.028	0.012	0.052
1.097	0.280	50	54.86	0.097	0.148	0.091	0.008	0.104
		100	109.72	0.110	0.064	0.099	0.005	0.115
		200	219.44	0.111	0.031	0.099	0.003	0.115
		400	438.88	0.174	0.084	0.170	0.004	0.181
		800	877.76	0.128	0.074	0.111	0.002	0.130
		1,000	1,097.20	0.196	0.145	0.179	0.004	0.200

^a Average Kolmogorov-Smirnov distance (*AK*) for different sample sizes n with discrepancy F_{ML}^* and noncentral parameter $\delta = nF_{ML}^*$.

Validity of confidence intervals for fit indices and methods of power estimation, which rely upon the test statistic T_{ML} , depend on the quality of employed theoretical approximations. In that respect, we generated 50,000 sample test statistics for Model 1 and calculated the empirical quantile (denoted $Q - T_{ML}$) and percentage of samples from the simulation that covered theoretical distribution quantile (denoted $P - T_{ML}$) under four underlying distribution assumptions with two noncentrality parameter values, $\delta = 36.12$ (Table 5) and $\delta = 189.56$ for $n = 400$. Here $Q - ncx$ are the quantiles from $\chi^2_{df}(\delta)$ and $P - ncx$ is the percentage of samples that is less than computed quantile $Q - ncx$. Other measures are defined for the four normal distributions in a similar way. Values in parentheses are the differences between empirical values and respective theoretical values from each distribution. For $\delta = 36.12$, measures from $\chi^2_{df}(\delta)$ are very similar

TABLE 5
Quantile Comparison for Model 1 ($n = 400, \delta = 36.12$)

$Q - T_{ML}$	22.0196	28.7702	32.7764	68.9375	75.4350	87.9215
$P - T_{ML}$	1%	5%	10%	90%	95%	99%
$Q - ncx$	23.8941	30.2580	34.0014	67.4783	73.2377	84.7275
$Q_{diff} - ncx$	(-1.8745)	(-1.4878)	(-1.2250)	(1.4592)	(2.1973)	(3.1940)
$P - ncx$	1.65%	6.63%	11.96%	88.39%	93.60%	98.45%
$P_{diff} - ncx$	(-0.65)	(-1.63)	(-1.96)	(1.61)	(1.40)	(0.55)
$Q - nm$	12.1095	19.1462	22.8974	49.3624	53.1137	60.1503
$Q_{diff} - nm$	(9.9101)	(9.6240)	(9.8790)	(19.5751)	(22.3213)	(27.7712)
$P - nm$	0.01%	0.44%	1.24%	50.95%	61.24%	77.19%
$P_{diff} - nm$	(0.99)	(4.56)	(8.76)	(39.05)	(33.76)	(21.81)
$Q - nm2$	11.2630	22.6489	28.7187	71.5412	77.6109	88.9969
$Q_{diff} - nm2$	(10.7566)	(6.1213)	(4.0577)	(-2.6037)	(-2.1759)	(-1.0754)
$P - nm2$	0.00%	1.16%	4.95%	92.39%	96.10%	99.17%
$P_{diff} - nm2$	(1.00)	(3.84)	(5.05)	(-2.39)	(-1.10)	(-0.17)
$Q - nm3$	17.0173	26.7196	31.8918	68.3821	73.5543	83.2566
$Q_{diff} - nm3$	(5.0023)	(2.0506)	(0.8846)	(0.5554)	(1.8807)	(4.6649)
$P - nm3$	0.18%	3.29%	8.77%	89.41%	93.80%	98.10%
$P_{diff} - nm3$	(0.82)	(1.71)	(1.23)	(0.59)	(1.20)	(0.90)
$Q - nm4$	19.5741	28.5253	33.2972	66.9627	71.7345	80.6857
$Q_{diff} - nm4$	(2.4455)	(0.2449)	(-0.5208)	(1.9748)	(3.7005)	(7.2358)
$P - nm4$	0.51%	4.78%	10.85%	87.80%	92.54%	97.32%
$P_{diff} - nm4$	(0.49)	(0.22)	(-0.85)	(2.20)	(2.46)	(1.68)

^aEmpirical quantile ($Q - T_{ML}$) and percentage of samples from the simulation that covered theoretical distribution quantile for $\delta = 36.12$ with $n = 400$.

^b $Q - distribution$ are the quantiles from $\chi^2_{df}(\delta)$ and $P - ncx$ is the percentage of samples that is less than computed quantile $Q - distribution$.

^cValues in parentheses (Q_{diff}, P_{diff}) are the differences between empirical values and respective theoretical values from each distribution.

to empirical values. On the other hand, theoretical values from the three normal distributions (nm , $nm2$, $nm4$) are quite different from empirical ones. Moreover, we can observe that normal quantile values show skewness problem, which was pointed out in the section Theoretical Background. For the large value of $\delta = 189.56$, measures from $nm2$ are more similar than those from ncx , but none of them are close to empirical ones. Also, skewness problem still exists.

Figure 1 and Figure 2 provide the QQ plots for T_{ML} against ncx and $nm2$ for $n = 400$ with $\delta = 36.1229$ from Model 1. When $\delta = 36.1229$, $\chi_{df}^2(\delta)$ describes the behavior of T_{ML} pretty well (Figure 1), whereas normal distribution with bias correction ($nm2$) works poorly (Figure 2). These plots confirm the skewness problem again. When $\delta = 189.555$, two plots (T_{ML} against ncx and $nm2$) show very similar patterns because $\chi_{df}^2(\delta)$ and normal distribution gets similar in terms of performance of describing T_{ML} . We could not see a difference between them from the plots.

We get similar results for Model 2. These results can be found in the technical report on the Web site: <http://ideas.repec.org/p/prapa/mprapa/17310.html> That is, $ncxK$ is smaller than nmK and $nm2K$ for small, moderate, and severe misspecification. Similarly, $ncxAK$ is smaller than $nmAK$ and $nm2AK$ for most cases.

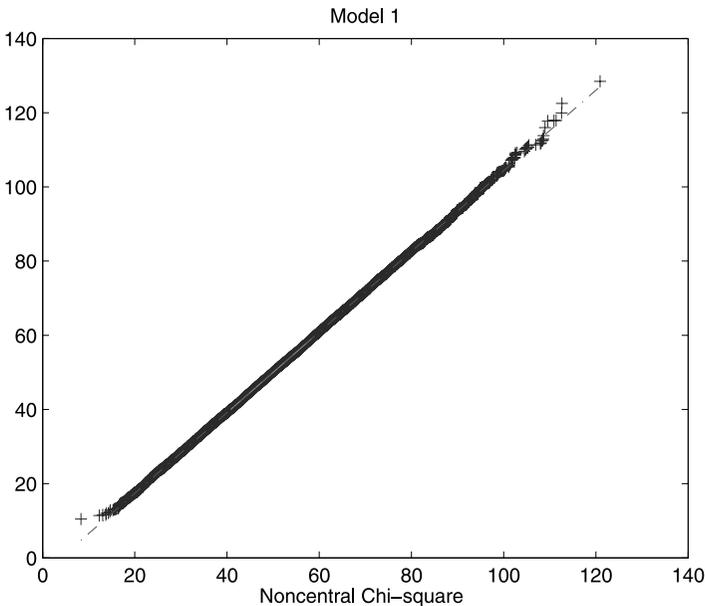


FIGURE 1 QQ plot of T_{ML} against ncx with $\delta = 36.12$ for Model 1.

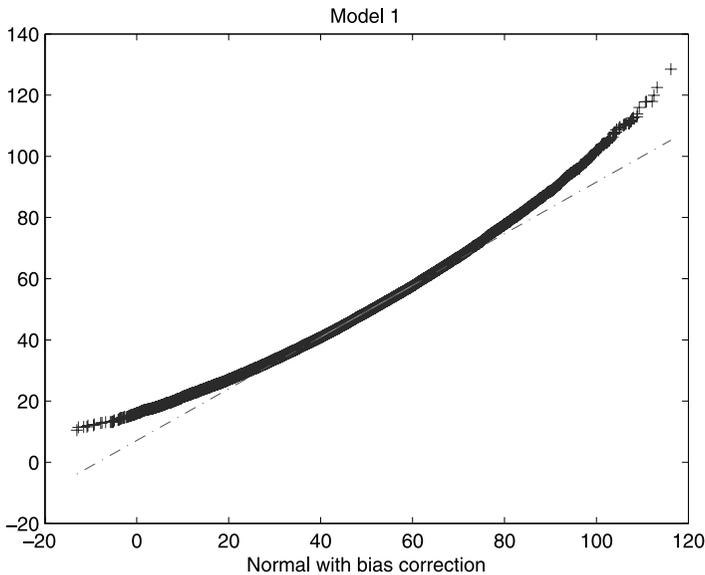


FIGURE 2 QQ plot of T_{ML} against $nm2$ with $\delta = 36.12$ for Model 1.

That is, $\chi_{df}^2(\delta)$ is a better approximation for T_{ML} under small to severe misspecification. Normal with bias correction ($nm2$) is slightly better for extremely severe misspecification, but none of these distributions gives a reasonable approximation in that case.

QQ plots for T_{ML} against $\chi_{df}^2(\delta)$ and normal distributions for $n = 400$ with $\delta = 39.95$ and $\delta = 80.05$ from Model 2 are produced. We could see that $\chi_{df}^2(\delta)$ describes the behavior of T_{ML} pretty well, whereas normal distribution with bias correction ($nm2$) shows poor performance. Skewness problem of normal approximation is very clear.

Nonnormally Distributed Data

We also use nonnormally (elliptically) distributed data to empirically illustrate the robustness of LR test statistics as we explained in the section Nonnormal Distributions. In order to generate data with an elliptical distribution we proceed 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 the Appendix).

We consider the same structure as in Model 1 discussed in the section Normally Distributed Data but with elliptically distributed data. That is, we directly calculate sample covariance matrices from the generated elliptically distributed data instead of using Wishart distribution. See Table 1 for generated parameters and Table 2 for discrepancy misspecification values. We generated two sets of elliptical distributions with different kurtosis parameter κ . Model 3 involves elliptically distributed data with random variable W taking two values, 1.2 with probability 0.45 and 0.8 with probability 0.55. Model 4 involves W taking two values, 2 with probability 0.2 and 0.5 with probability 0.8. The kurtosis parameter of these elliptical distributions is $\kappa = 0.1584$ (Model 3) and $\kappa = 2.25$ (Model 4). Note that in both cases $\mathbb{E}[W^2] = 1$ so that the covariance matrices of X and Y are equal to each other.

Kolmogorov-Smirnov distance (K) and Average Kolmogorov-Smirnov distance (AK) for Model 3 with sample sizes $n = 400$ and $n = 1,000$ and nine degrees of misspecification, $F_{ML}^* = 0.025, \dots, 1.360$, are calculated. For more detail, see tables and figures on the Web site: <http://ideas.repec.org/p/pram/prapa/17310.html> We observe that for small to severe misspecification F_{ML}^* (with RMSEA values ranging from 0.042 to 0.116), $ncxK$ is smaller than $nm2K$ and $ncxAK$ is smaller than $nm2AK$, but the status of those measures reverse for extremely severe misspecifications (with RMSEA values greater than 0.151). This implies that for small, moderate, and severe misspecifications, $\chi_{df}^2(\delta)$ is a better approximation. On the other hand, for extremely severe misspecifications the normal distribution with bias correction ($nm2$) gives a slightly better approximation, but none of distributions gives reasonable description for T_{ML} under extremely misspecified model. These results are consistent with the corresponding results of the section Normally Distributed Data.

Quantile comparisons are done to investigate the quality of each theoretical approximation with respect to the validity of confidence intervals or fit indices and methods of power estimation. We calculated the empirical quantile (denoted $Q - T_{ML}$ and $(1 + \kappa)^{-1}T_{ML}$) and percentage of samples from the simulation that covered theoretical distribution quantile (denoted $P - T_{ML}$) with $M = 50,000$ sample test statistics of Model 3 under four underlying distribution assumptions with two noncentrality parameter values, $\delta = 36.12$ and $\delta = 189.56$ for $n = 400$. Here $Q - ncx$ are the quantiles from $\chi_{df}^2((1 + \kappa)^{-1}\delta)$ and $P - ncx$ is the percentage of samples that is less than computed quantile $Q - ncx$. Other measures are defined for the normal distributions in a similar way. Values in parentheses are the differences between empirical values and respective theoretical values from each distribution. For both $\delta = 36.12$ and $\delta = 189.56$, measures from $\chi_{df}^2(\delta)$ are very similar to empirical values. On the other hand, theoretical values from the normal distributions (nm , $nm2$) are very different from empirical ones. Again, we can observe that normal quantile values show skewness problem, which was pointed out in the section Theoretical Background.

QQ plots for T_{ML} against $\chi_{df}^2(\delta)$ and normal distributions for $n = 400$ with $\delta = 36.1229$ and $\delta = 189.555$ from Model 3 are produced. For both $\delta = 36.1229$ and $\delta = 189.555$, $\chi_{df}^2(\delta)$ describes the behavior of T_{ML} pretty well, whereas normal distribution with bias correction (*nm2*) works poorly. We could confirm strong skewness problem of normal approximation.

Similar results are obtained for Model 4. Tables and figures for Model 4 can be found on the Web site: <http://ideas.repec.org/p/pramprapa/17310.html> It is interesting to see that $\chi_{df}^2(\delta)$ describes the behavior of T_{ML} better than normal distribution under small, moderate, severe, and even extremely severe misspecification for Model 4. QQ plots confirm same conclusion, especially clear skewness of normal approximation.

Empirical Data

We consider the Thurstone data (Thurstone & Thurstone, 1941). The data matrix is generated by 60 test scores from a classic study of mental ability. We use a nine variable Thurstone problem, which is discussed in detail by McDonald (1999). The nine variables are “Sentences,” “Vocabulary,” “Sentence completion,” “First Letters,” “Four letter words,” “Suffixes,” “Letter Series,”

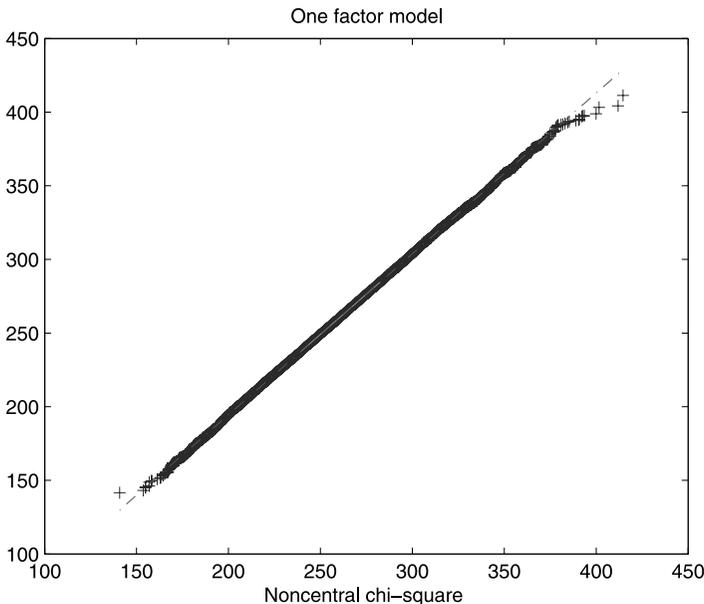


FIGURE 3 QQ plot of T_{ML} against $nc\chi$ for Model Thurstone – 1.

“Pedigrees,” and “Letter Grouping,” which measure verbal ability, word fluency, and reasoning ability.

We apply one factor model (denoted *Thurstone – 1*) and three factor model (denoted *Thurstone – 3*) to these data with 213 observations. We estimate parameters in factor analysis and calculate RMSEA values for each model. Tables with detailed information can be found on the Web site: <http://ideas.repec.org/p/pramprapa/17310.html> Note that one factor model indicates an extremely poor fit (with RMSEA value 0.2036) whereas three factor model shows a good fit (with RMSEA value 0.0408). In order to evaluate statistical properties of the corresponding LR test statistics we employ the parametric bootstrap approach (see Efron & Tibshirami, 1993, Sec. 6.5). That is, in the Monte Carlo sampling the (unknown) population covariance matrix is replaced by the sample covariance matrix. Consequently, we randomly generate 50,000 sample covariance matrices from the respective Wishart distribution and calculate the LR test statistics T_{ML} . QQ plots for T_{ML} against noncentral chi-square and normal distribution for *Thurstone – 1* and *Thurstone – 3* models are provided (Figures 3–4 and Figures 5–6). Noncentral chi-square distribution describes the distribution of test statistics pretty well for both models whereas normal distribution with bias correction shows a poor performance, especially for the three factor model. For both models the skewness problem of normal approximation is present and is especially bad for the three factor model (Figure 6).

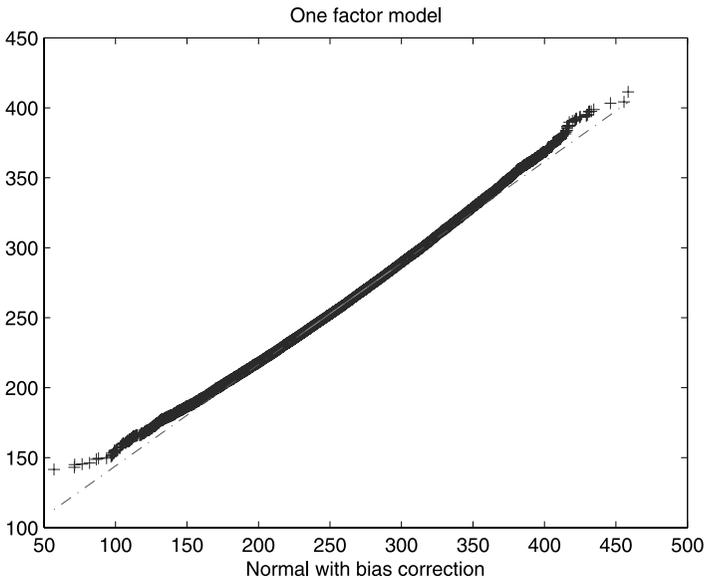


FIGURE 4 QQ plot of T_{ML} against $nm2$ for Model Thurstone – 1.

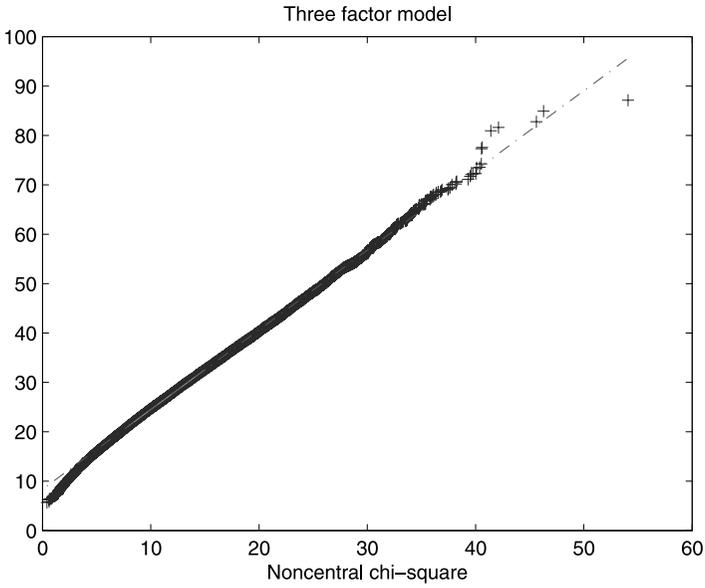


FIGURE 5 QQ plot of T_{ML} against n_{cx} for Model Thurstone - 3.

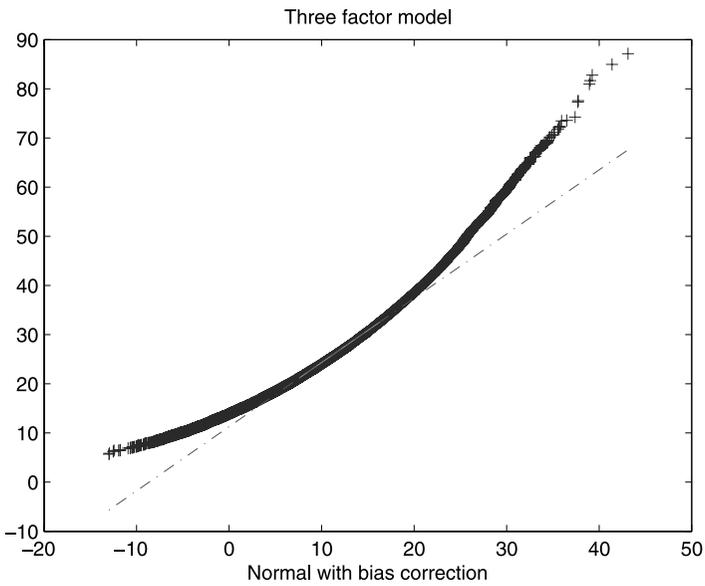


FIGURE 6 QQ plot of T_{ML} against $nm2$ for Model Thurstone - 3.

DISCUSSION

The noncentral chi-square distribution is widely used to describe the behavior of LR test statistics T_{ML} in SEM for the computation of fit indices and evaluation of statistical power. Recently, it was suggested by several authors that T_{ML} could be better described by the normal than the noncentral chi-square distribution. In this article, we discuss the underlying theory of both approximations, normal and noncentral chi-square, and present some numerical experiments aimed at empirical comparison of the performance of two distributions in describing the distribution of the test statistic T_{ML} .

Monte Carlo experiments are conducted for several factor analysis models. Furthermore, we use both normal and nonnormal data to investigate the robustness of test statistics to nonnormality. For each model, we considered different sample sizes ranging from 50 to 1,000 and varying conditions of model misspecification ranging from small to extremely severe. Several discrepancy measures based on the Kolmogorov-Smirnov distance were used to compare the noncentral chi-square distribution with normal distributions. Respective quantiles are compared in order to investigate the behavior of tails in each distribution as well. Empirical results indicate that the distribution of T_{ML} is described well by the noncentral chi-square distribution under small, moderate, and even severe misspecifications irrespective of the sample size. For the extremely misspecified model, the normal distribution with a bias correction is slightly better than the noncentral chi-square distribution.

It could be noted that normal distribution with *estimated* sample mean and variance gives a better approximation for *larger* discrepancy values (see columns *nm3K* and *nm3AK* in the tables). This, however, is of a little practical significance because the corresponding mean and variance could be computed only in simulation experiments and will be unavailable for a given data set.

In summary, the noncentral chi-square approximation of the ML test statistic is valid under reasonable misspecifications and models. The normal distribution with a bias correction may perform slightly better under extreme misspecifications. However, neither the noncentral chi-square distribution nor the theoretical normal distributions give reasonable approximations of LR test statistics under extremely severe misspecifications. Of course, extremely misspecified models are unacceptable anyway for a reasonable statistical inference. These findings may differ with variations in model complexity, model parameterization, and underlying data structure.

ACKNOWLEDGMENTS

Research of So Yeon Chun was partly supported by the National Science Foundation award DMS-0510324 and of Alexander Shapiro by the NSF awards DMS-0510324 and DMI-0619977.

REFERENCES

- Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, 8, 1–24.
- Browne, M. W., & Cudeck, R. (1992). Alternative methods of assessing model fit. *Sociological Methods and Research*, 21, 230–258.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*. (pp. 136–162). Beverly Hills, CA: Sage.
- Browne, M. W., & Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Psychology*, 41, 193–208.
- Chun, S. Y., & Shapiro, A. (2008). *Construction of covariance matrices with a specified discrepancy function minimizer, with application to factor analysis*. Available from <http://www.optimization-online.org>. To appear in *SIAM Journal on Matrix Analysis and Applications*.
- Cudeck, R., & Browne, M. W. (1992). Constructing a covariance matrix that yields a specified minimizer and a specified minimum discrepancy function value. *Psychometrika*, 57, 357–369.
- Efron, B., & Tibshirami, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Golden, R. M. (2003). Discrepancy risk model selection test theory for comparing possibly misspecified or nonnested models. *Psychometrika*, 68, 229–249.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- McManus, D. A. (1991). Who invented local power analysis? *Econometric Theory*, 7, 265–268.
- Muirhead, R. J., & Waternaux, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. *Biometrika*, 67, 31–43.
- Olsson, U. H., Foss, T., & Breivik, E. (2004). Two equivalent discrepancy functions for maximum likelihood estimation: Do their test statistics follow a non-central chi-square distribution under model misspecification?. *Sociological Methods Research*, 32, 453–500.
- Shapiro, A. (1983). Asymptotic distribution theory in the analysis of covariance structures (a unified approach). *South African Statistical Journal*, 17, 33–81.
- Shapiro, A. (2009). Asymptotic normality of test statistics under alternative hypotheses. *Journal of Multivariate Analysis*, 100, 936–945.
- Shapiro, A., & Browne, M. W. (1987). Analysis of covariance structures under elliptical distributions. *Journal of the American Statistical Association*, 82, 1092–1097.
- Steiger, J. H., & Lind, J. (1980). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.
- Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, 50, 253–264.
- Sugawara, H. M., & MacCallum, R. C. (1993). Effect of estimation method on incremental fit indexes for variance structure models. *Applied Psychological Measurement*, 17, 365–377.
- Thurstone, L. L., & Thurstone, T. G. (1941). *Factorial studies of intelligence*. Chicago, IL: The University of Chicago Press.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307–333.
- Wilks, S. S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 9, 60–62.
- Yuan, K. H. (2008). Noncentral chi-square versus normal distributions in describing the likelihood ratio statistic: The univariate case and its multivariate implication. *Multivariate Behavioral Research*, 43, 109–136.
- 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, 1262–1282.

APPENDIX

Let $\mathbf{X} \sim N(\mathbf{0}, \Sigma)$ be a random vector having normal distribution and W be a random variable independent of \mathbf{X} . Then $\mathbf{Y} = W\mathbf{X}$ has elliptical distribution with $\mathbb{E}[\mathbf{Y}] = \mathbf{0}$ and characteristic function

$$\phi(\mathbf{t}) = \mathbb{E}[\exp(iW\mathbf{t}'\mathbf{X})] = \mathbb{E}[\mathbb{E}\{\exp(iW\mathbf{t}'\mathbf{X})|W\}] = \mathbb{E}[\exp\{-\frac{1}{2}W^2\mathbf{t}'\Sigma\mathbf{t}\}].$$

That is, $\phi(\mathbf{t}) = \psi(\mathbf{t}'\Sigma\mathbf{t})$, where $\psi(z) = \mathbb{E}[\exp\{-\frac{1}{2}W^2z\}]$. Then it follows that the covariance matrix of \mathbf{Y} is $\alpha\Sigma$, where $\alpha = -2\psi'(0) = \mathbb{E}[W^2]$. It also follows that the kurtosis parameter is

$$\kappa = \frac{\psi''(0) - \psi'(0)^2}{\psi'(0)^2} = \frac{\mathbb{E}[W^4]}{(\mathbb{E}[W^2])^2} - 1$$

(cf. Muirhead & Waternaux, 1980). For example, if W can take two values, a with probability p and b with probability $1 - p$, then $\alpha = a^2p + b^2(1 - p)$ and

$$1 + \kappa = \frac{a^4p + b^4(1 - p)}{(a^2p + b^2(1 - p))^2}.$$