Two Simple Approximations to the Distributions of Quadratic Forms*

Ke-Hai Yuan
University of Notre Dame
and
Peter M. Bentler
University of California, Los Angeles

September 3, 2007

*This research was supported by NSF grant DMS04-37167, and grants DA01070 and DA00017 from the National Institute on Drug Abuse. Correspondence concerning this article should be addressed to Ke-Hai Yuan (kyuan@nd.edu).

Abstract

Many test statistics are asymptotically equivalent to quadratic forms of normal variables, which are further equivalent to $T = \sum_{i=1}^{d} \lambda_i z_i^2$ with z_i being independent and following $N(0,1)$. Two approximations to the distribution of T have been implemented in popular software and are widely used in evaluating various models. It is important to know how accurate these approximations are when compared to each other and to the exact distribution of T. The paper systematically studies the quality of the two approximations and examines the effect of λ_i's and the degrees of freedom d by analysis and Monte Carlo. The results imply that one approximation can be as good as the exact distribution when d is large. When the coefficient of variation of the λ_i's is small, another approximation is also adequate for practical model inference. The results are applied to a study of alcoholism and psychological symptoms.

Key words: Adjusted chi-square, rescaled statistic, coefficient of variation, Kolmogorov-Smirnov statistic, covariance structure analysis.

1. Introduction

In many statistical problems, the statistics for testing null hypotheses are asymptotically equivalent to quadratic forms of normal variables, which may not follow a chi-square distribution. Examples include the general likelihood ratio (LR) statistic when the distribution is misspecified (Foutz & Srivastava, 1977; Vuong, 1989); the Pearson chi-square statistic for contingency tables when the true covariance matrix of the estimated cells cannot be consistently estimated (Rao & Scott, 1984); test statistics in covariance structure analysis when the discrepancy function is specified using the normality assumption but the true underlying population distribution of the sample is unknown (Shapiro, 1983); test statistics for dimension reduction in inverse regression when the underlying distribution of the predictors is unknown (Li, 1991, 1992; Bura & Cook, 2001; Cook & Ni, 2005); the likelihood ratio statistic in testing the number of components in a normal mixture model when the null hypothesis holds (Lo, Mendell & Rubin, 2001). The quadratic forms are also the building blocks for the commonly used F-statistics in ANOVA and regression. The distribution of a quadratic form of normal variables can be characterized by a linear combination of independent chi-square variates, each with one degree of freedom. Because the exact distribution of a linear combination of independent chi-square variates is difficult to obtain in general, various approximations to its distribution have been proposed (Solomon & Stephens, 1977). Two relatively simple ones are widely used in practice, one is to rescale the involved statistic, the other is to adjust the chi-square distribution. The purpose of this paper is to study these two distribution approximations using analysis and Monte Carlo. In section 2 we will review the two approximations and their use in practice. In section 3 we will study the effect of the coefficients on the approximations. Section 4 presents Monte Carlo results. We will illustrate the application of the two approximations using a covariance structure model in section 5. Conclusions will be provided in section 6.

2. Two Approximations to the Distribution of Quadratic Forms

Let $\mathbf{x} \sim N_p(\mathbf{0}, \mathbf{\Gamma})$ and $T = \mathbf{x}'\mathbf{W}\mathbf{x}$ be a quadratic form in $\mathbf{x}$. The matrix $\mathbf{\Gamma}$ is typically of full rank while $\mathbf{W}$ is nonnegative definite. Let the rank of $\mathbf{W}$ be d and the nonzero

eigenvalues of $\mathbf{W\Gamma}$ be λ_1, λ_2, ..., λ_d. There exists

$$T = \mathbf{x}'\mathbf{W}\mathbf{x} = \sum_{i=1}^{d} \lambda_i z_i^2, \tag{1}$$

where $z_i \sim N(0,1)$ and are independent. The first approximation to the distribution of T is to rescale T by referring $T_R = c^{-1}T$ to χ_d^2, where $c = \sum_{i=1}^{d} \lambda_i/d$. We will use the notation

$$T_R \dot{\sim} \chi_d^2 \text{ or } T \dot{\sim} c\chi_d^2 \tag{2}$$

to imply approximating the distribution of T_R by χ_d^2 or that of T by $c\chi_d^2$. It is obvious that $E(T_R) = d$, so that the rescaling is actually a mean correction. A more sophisticated correction is

$$T \dot{\sim} a\chi_b^2, \tag{3}$$

where a and b are determined by matching the first two moments of T with those of $a\chi_b^2$. Straightforward calculation leads to

$$a = \frac{\sum_{i=1}^{d} \lambda_i^2}{\sum_{i=1}^{d} \lambda_i} \text{ and } b = \frac{(\sum_{i=1}^{d} \lambda_i)^2}{\sum_{i=1}^{d} \lambda_i^2}.$$

These approximations were originally proposed by Welch (1938) and further studied by Satterthwaite (1941) and Box (1954). When both $\mathbf{\Gamma}$ and $\mathbf{W}$ can be consistently estimated, c, a and b will be estimated as

$$\hat{c} = \mathrm{tr}(\hat{\mathbf{W}}\hat{\mathbf{\Gamma}})/d, \quad \hat{a} = \mathrm{tr}[(\hat{\mathbf{W}}\hat{\mathbf{\Gamma}})^2]/\mathrm{tr}(\hat{\mathbf{W}}\hat{\mathbf{\Gamma}}), \quad \hat{b} = [\mathrm{tr}(\hat{\mathbf{W}}\hat{\mathbf{\Gamma}})]^2/\mathrm{tr}[(\hat{\mathbf{W}}\hat{\mathbf{\Gamma}})^2].$$

In many cases, a generalized Wald statistic (Boos, 1992) is also available for testing the same null hypothesis. Such a statistic asymptotically follows a chi-square distribution. However, referring the generalized Wald statistic to χ_d^2 may perform worse than referring T_R to χ_d^2 or referring T to $a\chi_b^2$. In dealing with the effect of survey design on analyzing multiway contingency tables, Rao and Scott (1984) noted that the approximations in (2) and (3) are practically adequate and may perform better than a Wald statistic. In the context of covariance structure analysis, Satorra and Bentler (1988, 1994) proposed using the two approximations when T is the normal distribution based LR statistic. Monte Carlo results in Hu, Bentler and Kano (1992) showed that the approximation in (2) performed very well and much better than a generalized Wald statistic (see Browne, 1984). The rescaled

statistic T_R in (2) has been in standard software (EQS, LISREL, MPLUS) for many years and used in numerous publications by researchers in psychology, education, sociology, medicine, business, etc. The adjusted distribution for T in (3) has also been in popular software (e.g., MPLUS) and widely used in analyzing categorical data.

Although these two approximations have been used for inference on a variety of models, their relative merits are not well-understood. In the context of covariance structure analysis, Fouladi (1997) reported that (3) performs better than (2). In testing the dimensionality of the space of the effective predictors using inverse regression, Bura and Cook (2003) also found that (3) performs better than (2). However, Bentler and Xie (2000) found that (2) performs much better than (3). These conclusions are based on examples and simulated type I errors, not the whole distribution approximation. Satorra and Bentler (1994) reported a few percentiles of T and T_R using a small simulation, they did not contrast the two approximations. As we shall see, the performance of the two approximations depends on the values of the coefficients λ_i's in (1). None of the above studies have controlled these coefficients. Actually, in any of these contexts, it is rather difficult to control the λ_i's when $\mathbf{\Gamma}$ and $\mathbf{W}$ are derived from models. Even when all the λ_i's can be specified, their effect on (2) and (3) will be confounded with sampling errors due to finite sample sizes.

Because researchers commonly use these approximations for inference, it is important to know how these approximations perform when compared to a statistic that exactly follows a chi-square distribution. There is also a need to compare the relative goodness of the two approximations. When studying them through a LR or Pearson chi-square statistic, we will not be able to separate the approximation of the distribution of the statistic with that of a quadratic form from those in (2) and (3). So we will work directly on quadratic forms with known λ_i's in the following sections.

3. Effect of the Coefficients λ_i's on the Approximating Distributions

In this section we study the effect of the λ_i's on the approximations in (2) and (3) by analysis, and relate the a and b to the coefficient of variations of the λ_i's. We will also introduce the Kolmogorov-Smirnov distance and a related measure of mean distance between two distributions, which will be used for studying the empirical performance of the two approximations in the next section.

Consider when $\lambda_1 = \lambda_2 = \ldots = \lambda_d = \lambda$, then $c = \lambda$, $a = \lambda$ and $b = d$, and the approximations in (2) and (3) are perfect. When all the λ_i's in (1) change proportionally, i.e., λ_i becoming $\tau\lambda_i$, then T changes to τT; c changes to τc; a changes to τa and b remains the same. In such a case, the qualities of the approximations in (2) and (3) do not change. So it is the relative sizes of the λ_i's that affect the two approximations.

When $\sum_{i=1}^d \lambda_i$ is a constant while the λ_i's change, the distribution of $\sum_{i=1}^d \lambda_i z_i^2$ will change. But the scaling factor c remains the same. So the quality of the approximation in (2) is affected as variations occur among the λ_i's. It is obvious that the relative sizes of the λ_i's also affect the approximation in (3). To see how a and b change when the λ_i's change, we rewrite b as

$$b = (\sum_{i=1}^{d} \tau_i^2)^{-1},$$

where $\tau_i = \lambda_i / \sum_{i=1}^d \lambda_i$. Because $\sum_{i=1}^d \tau_i = 1$, $\sum_{i=1}^d \tau_i^2$ reaches its minimum when $\tau_1 = \tau_2 = \ldots = \tau_d = 1/d$. This implies that b reaches its maximum value at d when all the λ_i's are equal; b decreases as the λ_i's depart from each other. Because $ab = \sum_{i=1}^d \lambda_i$, when holding $\sum_{i=1}^d \lambda_i$ constant, a will increase when the λ_i's depart from each other. Of course, when $\sum_{i=1}^d \lambda_i$ decreases, it is very likely that both a and b decrease.

We may use the coefficient of variation of the λ_i's,

$$\mathrm{CV}(\lambda) = \frac{\mathrm{SD}(\lambda)}{\bar{\lambda}} = \frac{\{\sum_{i=1}^d (\lambda_i - \bar{\lambda})^2/d\}^{1/2}}{\bar{\lambda}},$$

to measure the relative variations among the λ_i's, where $\bar{\lambda} = \sum_{i=1}^d \lambda_i/d$. When $\mathrm{CV}(\lambda) = 0$, both the approximations in (2) and (3) are perfect. They become poorer as $\mathrm{CV}(\lambda)$ increases. Actually, both a and b are closely related to $\mathrm{CV}(\lambda)$. It follows from

$$\mathrm{CV}^2(\lambda) = \frac{\sum_{i=1}^d \lambda_i^2/d - \bar{\lambda}^2}{\bar{\lambda}^2} = d\frac{\sum_{i=1}^d \lambda_i^2}{(\sum_{i=1}^d \lambda_i)^2} - 1$$

and $ab = \sum_{i=1}^d \lambda_i$ that

$$b = \frac{d}{\mathrm{CV}^2(\lambda) + 1} \quad \text{and} \quad a = \bar{\lambda}[\mathrm{CV}^2(\lambda) + 1].$$

So the approximations in (2) and (3) are equivalent only when $\mathrm{CV}(\lambda) = 0$. The distribution approximation in (2) can be regarded as approximating (3) by treating $\mathrm{CV}(\lambda) = 0$ even when it is not. So we would expect that the difference between (2) and (3) becomes obvious when $\mathrm{CV}(\lambda)$ increases.

The approximation will also depend on the degrees of freedom. As d increases, according to the central limit theorem, the distribution of T may even be approximately described by a normal distribution, and so may $a\chi^2_b$, T_R and χ^2_d. Thus, we may expect that the approximations in (2) and (3) will improve as d increases, which will be examined by Monte Carlo in the next section.

We will use the well-known Kolmogorov-Smirnov (KS) statistic to evaluate the two approximations. The KS statistic measures the distance between the empirical distribution function (EDF) $\hat{F}(t)$ and the proposed target distribution function $G(t)$; $F(t)$ will be reserved for the true cumulative distribution function (CDF) of T. Suppose we have N independent observations on T. Let the ordered statistics be $t_{(1)} \leq t_{(2)} \leq \ldots t_{(N)}$, the KS is calculated by

$$\text{KS} = \max_{1 \leq i \leq N} \text{KS}_i \text{ with } \text{KS}_i = \max\{|\frac{i-1}{N} - G(t_{(i)})|, |\frac{i}{N} - G(t_{(i)})|\}.$$

Because KS is decided by one point on the real line, it does not tell us the whole picture of the approximation. The other measure we use is the average or the mean of the KS_i,

$$\text{MKS} = \sum_{i=1}^{N} \text{KS}_i/N,$$

which is a reasonable measure of the overall discrepancy between $\hat{F}(t)$ and $G(t)$. The maximum value of the KS is 1.0, which implies that $\hat{F}(t)$ and $G(t)$ do not have any overlap. To see the maximum value of MKS, we may assume that $\hat{F}(t)$ is above $G(t)$ or $G(t_{(1)}) = 1.0$, then $\text{KS}_i = 1 - (i-1)/N$ and

$$\begin{aligned} \text{MKS} &= \frac{1}{N}\sum_{i=1}^{N}[1 - \frac{(i-1)}{N}] \\ &= 1 - \frac{1}{N^2}[\frac{N(N+1)}{2} - N] \approx \frac{1}{2}. \end{aligned}$$

The KS and MKS will be used to measure the distance between the EDF of T_R and the CDF of χ^2_d as well as that between the EDF of T and the CDF of $a\chi^2_b$ in the next section.

4. Empirical Results

We will study the effect of both $\text{CV}(\lambda)$ and the degrees of freedom d on the two approximations. We start with $d = 2$ and the vector of the λ_i's as $\boldsymbol{\lambda}_k = (1, k)'$. The $\text{CV}(\lambda)$, KS

and MKS for $k = 2$ to 10, over $N = 2000$ replications, are reported in Table 1(a). The KS and MKS under χ^2_d are obtained by referring a simulated chi-square variate to the chi-square distribution χ^2_d, they represent what KS and MKS are like under a perfect situation. Because KS does not depend on $F(t)$ (Serfling, 1980, p. 62), we can also regard that KS and MKS under χ^2_d correspond to the discrepancy between the EDF of T and the CDF of T. The KS and MKS under $c\chi^2_d$ correspond to the approximation in (2), those under $a\chi^2_b$ correspond to the approximation in (3). Because people in practice commonly refer the LR statistic to a nominal chi-square distribution without checking the distribution of the sample, we also include the KS and MKS under $L\chi^2_d$ in Table 1(a) when referring a linear combination of independent chi-square variates to χ^2_d.

When $k = 2$, $\mathrm{CV}(\lambda) = .333$, the KS under $c\chi^2_d$ is the smallest, the MKS under $c\chi^2_d$ is also comparable to that for the ideal case. As k or $\mathrm{CV}(\lambda)$ increases, the KS and MKS under χ^2_d fluctuate; the KS and MKS under $c\chi^2_d$ also fluctuate, but they tend to increase; those under $a\chi^2_b$ also tend to increase, but the speed is a lot smaller; those under $L\chi^2_d$ are always the greatest. The biggest number in each column is marked in boldface, the KS corresponding to $L\chi^2_d$ is about 15 times of the perfect case; that corresponding to $c\chi^2_d$ is 4 times of the perfect case; that corresponding to $a\chi^2_b$ is about twice of the perfect case. Comparisons of the largest MKS's are similar to those of the KS's. Each number in the last row of Table 1(a) is the average of the previous rows, according to which the approximation in (3) is a lot better than that in (2). Actually, only when $k = 2$ does the approximation in (2) enjoy smaller KS and MKS than those for (3); we put this condition in boldface in the first column of the table.

Insert Table 1 about here

To see the effect of the degrees of freedom on the approximations in (2) and (3), we next study the conditions of $d = 6$ with $\boldsymbol{\lambda}_k = (\mathbf{1}'_3, k\mathbf{1}'_3)'$ and $d = 10$ with $\boldsymbol{\lambda}_k = (\mathbf{1}'_5, k\mathbf{1}'_5)'$, where $\mathbf{1}_j$ represents a vector of j 1's. The KS and MKS are reported in Table 1(b) and 1(c), respectively. Although the degrees of freedom increased, the $\mathrm{CV}(\lambda)$ for a given k is the same due to the same two distinct λ_i's. The patterns of KS and MKS under $L\chi^2_d$ and $c\chi^2_d$ in Table 1(b) and (c) are about the same as in (a), they tend to increase as $\mathrm{CV}(\lambda)$ increases.

However, the KS and MKS under $a\chi_b^2$ may just fluctuate. Actually, the greatest KS or MKS under $a\chi_b^2$ in either Table1(b) or (c) is smaller than that under χ_d^2. Comparing the averaged KS or MKS at the bottom of Table 1(a), (b) and (c), we notice that those corresponding to χ_d^2 tend to be stable as d changes, since the distribution of KS does not depend on F; the KS and MKS corresponding to $c\chi_d^2$ also appear not affected when d changes; the KS and MKS corresponding to $L\chi_d^2$ obviously increase when d increases; those corresponding to $a\chi_b^2$ tend to decrease as d increases. At $d = 10$, with three decimals, the averaged KS for $a\chi_b^2$ is identical to that for χ_d^2, and so is the averaged MKS.

Mean and covariance structure analysis typically involves many variables, the degrees of freedom can be much larger than those studied in Table 1; there can be many predictors in regression and the degrees of freedom can also be very large in testing the number of principal Hessian directions when using inverse regression. It is most likely that, as the dimension increases, the corresponding CV(λ) also change. To further compare the two approximations in (2) and (3) under these conditions, we choose (a) $d = 10$ with ten conditions on λ_i's: $\boldsymbol{\lambda}_1 = (1, 1.1, 1.2, \ldots, 1.9)'$, $\boldsymbol{\lambda}_2 = (1, 1.2, 1.4, \ldots, 2.8)'$, ..., $\boldsymbol{\lambda}_{10} = (1, 2, 3, \ldots, 10)'$; (b) $d = 30$ with ten conditions on λ_i's: $\boldsymbol{\lambda}_1 = (1, 1.1, 1.2, \ldots, 3.9)'$, $\boldsymbol{\lambda}_2 = (1, 1.2, 1.4, \ldots, 6.8)'$, ..., $\boldsymbol{\lambda}_{10} = (1, 2, 3, \ldots, 30)'$; and (c) $d = 50$ with ten conditions on λ_i's: $\boldsymbol{\lambda}_1 = (1, 1.1, 1.2, \ldots, 5.9)'$, $\boldsymbol{\lambda}_2 = (1, 1.2, 1.4, \ldots, 10.8)'$, ..., $\boldsymbol{\lambda}_{10} = (1, 2, 3, \ldots, 50)'$. The KS and MKS using $N = 2000$ as well as the associated CV(λ) are reported in Table 2(a), (b) and (c), respectively. Except when $d = 10$ and $k = 5$ where KS and MKS under $c\chi_d^2$ are smaller than those under both χ_d^2 and $a\chi_b^2$, all the other KS and MKS corresponding to the approximation in (3) are smaller than those corresponding to the approximation in (2). The KS and MKS under $a\chi_b^2$ in Table 2(a) are almost as small as those under χ_d^2; the KS and MKS on average under $a\chi_b^2$ in Table 2(b) are even smaller than those under χ_d^2, due to sampling errors. The average KS and MKS under $a\chi_b^2$ are identical to those under χ_d^2 in Table 2(c). As d and CV(λ) increase, the KS and MKS corresponding to $L\chi_d^2$ reach their maximum; then it is meaningless to approximate the linear combination of chi-square variate by the nominal chi-square distribution.

Insert Table 2 about here

In the practice of principal components and factor analysis, when ordering the eigenvalues of a sample covariance matrix from large to small, it often happens that the first few drop

dramatically in size, the remaining ones slowly decrease. The phenomenon that most smaller eigenvalues sit on a line is called the scree test in factor analysis (see Gorsuch, 1983, pp. 165–169). We also include the following conditions to mimic such a phenomenon: $d = 10$ and $\boldsymbol{\lambda}_{10} = (1, 1.1, 1.2, \ldots, 1.7, 1.8, 10)'$, 9 eigenvalues are evenly spaced except the largest one; $d = 20$ and $\boldsymbol{\lambda}_{20} = (1, 1.1, 1.2, \ldots, 2.6, 2.7, 10, 20)'$, 18 eigenvalues are evenly spaced except the largest two; ..., $d = 100$ and $\boldsymbol{\lambda}_{100} = (1, 1.1, 1.2, \ldots, 9.7, 9.8, 9.9, 10, 20, 30, \ldots, 100)'$, 90 eigenvalues are evenly spaced except the largest ten. Table 3 contains the CV(λ) as well as the KS and MKS for these conditions. The CV(λ) increases as d increases, the KS and MKS under χ^2_d remain stable as they should be; those under $L\chi^2_d$ reach their maximum values after $d = 50$ or 40; the KS and MKS under $c\chi^2_d$ tend to increase due to the increase in CV(λ); but the KS and MKS under $a\chi^2_b$ tend to decrease due to the increasing of d although CV(λ) increases.

Insert Table 3 about here

We may conclude from Tables 1 to 3 that, when controlling CV(λ), the approximation in (2) is almost not affected by the degrees of freedom while the approximation in (3) improves as the degrees of freedom increase. For a given d, when CV(λ) increases, the approximation in (2) tends to become worse; the approximation in (3) also tends to become worse when d is small. At a large d, the approximation in (3) is almost not affected by the change of CV(λ).

5. Illustration of Applications with a Covariance Structure Model

Neumann (1994) contains a study on the relationship of alcoholism and psychological symptoms, based on a data set with 10 variables and 335 participants. The 10 variables are Family History for Psychopathology, Family History for Alcoholism; the Age of First Problem with Alcohol, the Age of First Detoxification from Alcohol; Alcohol Severity Score, Alcohol Use Inventory; SCL-90 Psychological Inventory, the Sum of the Minnesota Multiphasic Personality Inventory Scores; the Lowest Level of Psychosocial Functioning during the past year, the Highest Level of Psychosocial Functioning during the past year. With two indicators for each latent construct, these 10 variables are respectively measuring: Family History, Age of Onset, Alcohol Symptoms, Psychopathology Symptoms, and Global Functioning. Let $\mathbf{x}$ represent the vector of the first two variables and $\mathbf{y}$ represent the vector

of the next eight variables. Using LISREL notation (Jöreskog & Sörbom, 1996, pp. 1–3), Neumann's theoretical model for these variables is

$$\mathbf{x} = \boldsymbol{\mu}_x + \boldsymbol{\Lambda}_x \xi + \boldsymbol{\delta}, \quad \mathbf{y} = \boldsymbol{\mu}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{4a}$$

and

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\xi + \boldsymbol{\zeta}, \tag{4b}$$

where $\boldsymbol{\mu}_x = E(\mathbf{x})$, $\boldsymbol{\mu}_y = E(\mathbf{y})$, $E(\xi) = 0$, $E(\boldsymbol{\eta}) = \mathbf{0}$, $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $E(\boldsymbol{\zeta}) = \mathbf{0}$,

$$\boldsymbol{\Lambda}_x = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}, \quad \boldsymbol{\Lambda}_y = \begin{pmatrix} 1 & \lambda_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \lambda_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \lambda_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_6 \end{pmatrix}',$$

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \beta_{21} & 0 & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 & 0 \\ 0 & \beta_{42} & \beta_{43} & 0 \end{pmatrix}, \quad \boldsymbol{\Gamma} = \begin{pmatrix} \gamma_{11} \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

and $\boldsymbol{\varepsilon}$, $\boldsymbol{\delta}$, and $\boldsymbol{\zeta}$ are vectors of errors whose elements are all uncorrelated.

Fitting the sample covariance matrix by the model in (4), using the normal distribution based maximum likelihood, leads to a LR statistic $T_{ML} = 48.961$ and a corresponding p-value of 0.012 when referred to χ^2_{29}. Because the normalized Mardia's measure of multivariate kurtosis for the data set is 14.763, which is highly significant, we may not trust the LR statistic. The estimates for the constants a, b and c in (2) and (3) are $\hat{a} = 1.372$, $\hat{b} = 21.858$ and $\hat{c} = 1.034$; the p-values when referring $T_R = 47.345$ to χ^2_{29} and T_{ML} to $\hat{a}\chi^2_{\hat{b}}$ are respectively 0.017 and 0.031. So the approximation in (3) gives a little better support for the theoretical model than that in (2). Although there is not much difference among the p-values, $\hat{b}$ is much smaller than the nominal degrees of freedom $d = 29$. The estimates of the 29 nonzero eigenvalues of the quadratic form corresponding to the LR statistic are 3.024, 2.198, 1.847, 1.652, 1.555, 1.472, 1.355, 1.266, 1.203, 1.156, 1.077, 1.059, 1.034, 0.961, 0.868, 0.815, 0.799, 0.770, 0.731, 0.671, 0.664, 0.610, 0.547, 0.536, 0.512, 0.452, 0.421, 0.401, 0.333, with $\mathrm{CV}(\hat{\lambda}) = 0.572$. So both the d and $\mathrm{CV}(\hat{\lambda})$ are close to the condition at the bottom of Table 2(b), where the approximation in (3) performs more accurately than that in (2).

Notice that the average of the sample eigenvalues $\hat{c}$ is close to 1.0. We may wonder whether all the population λ_i's are equal to 1.0 or the discrepancy among the $\hat{\lambda}_i$'s is due to

sampling error. Ten samples each with size 335 are generated from $N_{10}(\mathbf{0}, \mathbf{I})$. Each of which is further transformed to have the same sample covariance matrix $\mathbf{S}$ as in Neumann's data, using

$$\mathbf{y}_j = \mathbf{S}^{1/2}\mathbf{S}_I^{-1/2}\mathbf{x}_j, \quad j = 1, 2, \ldots, 335,$$

where $\mathbf{x}_j$ is a simulated observation, $\mathbf{S}_I$ is the sample covariance matrix of a simulated sample, and $\mathbf{y}_j$ is the transformed observation. Because the transformed $\mathbf{y}_j$ can be regarded as following a multivariate normal distribution, all the corresponding λ_i's are equal to 1.0 and a nonzero $\mathrm{CV}(\hat{\lambda}) = 0$ is due to sampling error. The ten $\mathrm{CV}(\hat{\lambda})$'s of the simulated samples are 0.443, 0.441, 0.418, 0.412, 0.440, 0.417, 0.434, 0.459, 0.411, 0.443. None of them is greater than $\mathrm{CV}(\hat{\lambda}) = 0.572$ of Neumann's data and model. So we cannot regard the population λ_i's in Neumann's data and model as being equal.

The limited simulation results on $\mathrm{CV}(\hat{\lambda})$ indicate that $\hat{\lambda}_i$'s are more dispersed than λ_i's. Analytical results (see Muirhead, 1982, p. 388) imply that larger $\hat{\lambda}_i$'s tend to over-estimate their population values and smaller $\hat{\lambda}_i$'s tend to under-estimate their population values. Thus, $\mathrm{CV}(\hat{\lambda})$ tends to be greater than $\mathrm{CV}(\lambda)$. Although the λ_i's for the real data and model are not equal, the corresponding $\mathrm{CV}(\lambda)$ should be a lot smaller than $\mathrm{CV}(\hat{\lambda}) = 0.572$. If $\mathrm{CV}(\lambda)$ is less than 0.5, by referring to Table 2(b), the approximation in (3) should give an equally reliable evaluation of the model as when the true asymptotic distribution of T_{ML} is known and used to evaluate the model. Thus, although the p-values among the different approximations are not much different, the analysis does give us more confidence on the p-value obtained by using the approximation in (3).

Because $\hat{c} \approx 1$ for Neumann's data and model, we used simulated normal samples to roughly check whether $\mathrm{CV}(\lambda) = 0$. When $\hat{c}$ is greater than 1 with practical data, one may simulate samples from a multivariate t-distribution with $m = 2(2\hat{c} - 1)/(\hat{c} - 1)$ degrees of freedom for the same purpose.

6. Conclusion

In this paper, we quantified the conditions that may affect the two widely used approximations. The quality of the two approximations was studied by varying and controlling the conditions. Because the true CDF, $F(t)$, of a quadratic form is hard to evaluate, we used the

EDF $\hat{F}(t)$ to estimate it. In addition to using Monte Carlo, one may use a numerical method to approximate $F(t)$, which can be defined through an integral with an infinite upper limit. The procedure involves replacing the infinity limit by a finite number and followed by a numerical integration (see Farebrother, 1990). Errors will occur when replacing the infinite limit by a finite limit and when using a numerical method to calculate the area under a continuous curve. The amount of error depends on the chosen upper limit and the number of rectangles or trapezoids used in the numerical integration, it also depends on the value of x. The amount of computation in the numerical method can be huge although the error can be made arbitrarily small. Comparing to using numerical method to evaluate $F(t)$, the EDF $\hat{F}(t)$ approximates $F(t)$ by Monte Carlo. The error in $\hat{F}(t)$ can be characterized by its standard deviation. Using

$$\text{Var}[\hat{F}(x)] = F(x)[1 - F(x)]/N \leq 1/(4N)$$

and $N = 2000$, the error in $\hat{F}(t)$ is in the magnitude of $1/(8000)^{1/2} = 0.011$. The error can be made smaller if we choose a larger N. But $N = 2000$ is enough for our purpose, that is, we can clearly tell the pros and cons of each of the two approximations under varied conditions.

The approximation in (2) may perform equally well as that in (3) when both CV(λ) and the degrees of freedom are small. The approximation in (3) generally performs better, especially when d is large. When CV(λ) is not large, say less than 0.5 and d is greater than 10, the approximation in (3) can be as good as knowing the exact distribution of T.

In practice, the λ_i's are never known, only $\hat{\lambda}_i$'s are available. Because CV($\hat{\lambda}$) tends to be greater than CV(λ). When CV($\hat{\lambda}$) is small, one may use either (2) or (3) for inference. When the degrees of freedom are large, (3) should be the choice.

In Monte Carlo studies with LR or other statistics, it may happen that the λ_i's are equal (see Yuan & Bentler, 1998). Then the approximation in (2) uses the correct assumption about the λ_i's and thus it will perform better than that in (3). Conflicting results on controlling type I errors by the two approximations (e.g., Fouladi, 1997; Bentler & Xie, 2000; Bura & Cook, 2003) would most likely be resolved if the population λ_i's were known.

In addition to the KS and MKS, we have also obtained the quantile and quantile (QQ) plots for each of the conditions in Tables 1 to 3. These plots indicate that the major discrepancy between the proposed distribution in (2) or (3) and the corresponding empirical

distribution occurs almost always on the right tail. So a larger KS or MKS also implies a poorer tail approximation. Because there are a total of 67 conditions, we elect not to include these QQ plots in the paper.

References

Bentler, P. M., & Xie, J. (2000). Corrections to test statistics in principal Hessian directions. *Statistics & Probability Letters, 47*, 381–89.

Boos, D. D. (1992). On generalized score tests. *American Statistician, 46*, 327–333.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in one-way classification. *Annals of Mathematical Statistics, 25*, 290–302.

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

Bura, E., & Cook, R. D. (2001). Extending sliced inverse regression: The weighted chi-squared test. *Journal of the American Statistical Association, 96*, 996–1003.

Bura, E., & Cook, R. D. (2003). Assessing corrections to the weighted chi-squared test for dimension. *Communication in Statistics: Simulation and Computation, 32*, 127–146.

Cook, R. D., & Ni, L. (2005). Sufficient dimension reduction via inverse regression: a minimum discrepancy approach. *Journal of the American Statistical Association, 100*, 410–428.

Farebrother, R. W. (1990). Algorithm AS 256: The distribution of a quadratic form in normal variables. *Applied Statistics, 39*, 294–309

Fouladi, R. T. (1997). Type I error control of some covariance structure analysis techniques under conditions of multivariate nonnormality. *Computing Science and Statistics, 29*, 526–532.

Foutz, R. V., & Srivastava, R. C. (1977). The performance of the likelihood ratio test when the model is incorrect. *Annals of Statistics, 5*, 1183–1194.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112*, 351–362.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8 user's reference guide*. Chicago: Scientific Software International.

Li, K. C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association, 86*, 316–342.

Li, K. C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma. *Journal of the American Statistical Association, 87*, 1025–1039.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*, 767–778.

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory.* New York: Wiley.

Neumann, C. S. (1994). *Structural equation modeling of symptoms of alcoholism and psychopathology.* Dissertation. Lawrence KS: University of Kansas.

Rao, J. N. K., & Scott, A. J. (1984). On chi-squared tests for multi-way contingency tables with cell proportions estimated from survey data. *Annals of Statistics, 12*, 46–60.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *1988 Proceedings of Business and Economics Sections* (pp. 308–313). Alexandria, VA: American Statistical Association.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Newbury Park, CA: Sage.

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika*, 6, 309–316.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics.* New York: Wiley.

Shapiro, A. (1983). Asymptotic distribution theory in the analysis of covariance structures (a unified approach). *South African Statistical Journal, 17*, 33–81.

Solomon, H., & Stephens, M. A. (1977). Distribution of a sum of weighted chi-square variables. *Journal of the American Statistical Association, 72*, 881–885.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and nonnested hypotheses. *Econometrica, 57*, 307–333.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika, 29*, 350–361.

Yuan, K.-H., & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology, 51*, 289–309.

Table 1(a). KS and MKS with $\boldsymbol{\lambda}_k = (1, k)'$ $(d = 2)$

		KS				MKS			
k	CV(λ)	χ^2_d	$L\chi^2_d$	$c\chi^2_d$	$a\chi^2_b$	χ^2_d	$L\chi^2_d$	$c\chi^2_d$	$a\chi^2_b$
2	0.333	0.018	0.150	0.017	0.029	0.006	0.098	0.008	0.011
3	0.500	0.012	0.235	0.037	0.033	0.004	0.149	0.016	0.014
4	0.600	0.011	0.278	0.054	0.050	0.004	0.183	0.030	0.020
5	0.667	0.028	0.330	0.074	0.046	0.011	0.207	0.038	0.018
6	0.714	0.018	0.376	0.064	0.045	0.007	0.233	0.038	0.020
7	0.750	0.018	0.408	0.077	0.060	0.010	0.256	0.039	0.022
8	0.778	0.012	0.430	0.084	0.056	0.003	0.267	0.047	0.021
9	0.800	**0.030**	0.421	**0.120**	0.062	**0.014**	0.269	**0.063**	**0.026**
10	0.818	0.015	**0.462**	0.109	**0.066**	0.004	**0.289**	0.053	0.021
Ave	0.662	0.018	0.343	0.071	0.050	0.007	0.217	0.037	0.019

Table 1(b). KS and MKS with $\boldsymbol{\lambda}_k = (\mathbf{1}'_3, k\mathbf{1}'_3)'$ $(d = 6)$

		KS				MKS			
k	CV(λ)	χ^2_d	$L\chi^2_d$	$c\chi^2_d$	$a\chi^2_b$	χ^2_d	$L\chi^2_d$	$c\chi^2_d$	$a\chi^2_b$
2	0.333	0.020	0.269	0.020	0.024	0.008	0.178	0.010	0.008
3	0.500	0.014	0.409	0.040	0.018	0.007	0.270	0.021	0.007
4	0.600	0.021	0.510	0.057	0.025	0.007	0.324	0.030	0.010
5	0.667	0.018	0.572	0.079	0.029	0.006	0.361	0.041	0.013
6	0.714	0.016	0.623	0.086	**0.033**	0.006	0.382	0.046	0.014
7	0.750	**0.042**	0.666	0.098	0.025	**0.018**	0.400	**0.051**	0.011
8	0.778	0.015	0.709	**0.103**	0.022	0.004	0.421	0.048	0.009
9	0.800	0.031	0.754	0.071	0.026	0.014	0.433	0.040	**0.016**
10	0.818	0.018	**0.760**	0.101	0.020	0.006	**0.441**	0.049	0.006
Ave	0.662	0.022	0.586	0.073	0.025	0.009	0.357	0.037	0.010

Table 1(c). KS and MKS with $\boldsymbol{\lambda}_k = (\mathbf{1}'_5, k\mathbf{1}'_5)'$ $(d = 10)$

		KS				MKS			
k	CV(λ)	χ^2_d	$L\chi^2_d$	$c\chi^2_d$	$a\chi^2_b$	χ^2_d	$L\chi^2_d$	$c\chi^2_d$	$a\chi^2_b$
2	0.333	0.022	0.332	0.025	0.012	0.010	0.220	0.012	0.004
3	0.500	0.015	0.523	0.047	0.016	0.004	0.336	0.019	0.005
4	0.600	0.016	0.630	0.071	0.025	0.007	0.386	0.032	0.012
5	0.667	0.029	0.698	0.083	0.031	0.015	0.422	0.043	0.015
6	0.714	**0.034**	0.786	0.054	**0.033**	**0.018**	0.453	0.032	**0.016**
7	0.750	0.016	0.815	0.069	0.014	0.005	0.461	0.037	0.005
8	0.778	0.010	0.838	0.085	0.022	0.004	0.467	0.043	0.007
9	0.800	0.019	0.859	**0.094**	0.012	0.005	0.476	**0.047**	0.005
10	0.818	0.019	**0.882**	0.082	0.015	0.006	**0.483**	0.044	0.006
Ave	0.662	0.020	0.707	0.068	0.020	0.008	0.412	0.034	0.008

Table 2(a). KS and MKS with $\boldsymbol{\lambda}_k = \mathbf{1}_{10} + k(0, 0.1, 0.2, \ldots, 0.9)'$ $(d = 10)$

		KS				MKS			
k	CV(λ)	χ^2_d	$L\chi^2_d$	$c\chi^2_d$	$a\chi^2_b$	χ^2_d	$L\chi^2_d$	$c\chi^2_d$	$a\chi^2_b$
1	0.198	0.022	0.304	0.022	0.019	0.010	0.208	0.010	0.007
2	0.302	0.015	0.505	0.020	0.015	0.004	0.331	0.007	0.003
3	0.367	0.016	0.632	0.041	0.029	0.007	0.388	0.016	0.011
4	0.410	0.029	0.701	**0.053**	**0.035**	0.015	0.424	**0.025**	0.015
5	0.442	**0.034**	0.799	0.032	0.032	**0.018**	0.460	0.015	**0.018**
6	0.466	0.016	0.824	0.029	0.012	0.005	0.466	0.016	0.005
7	0.484	0.010	0.853	0.036	0.018	0.004	0.475	0.018	0.006
8	0.500	0.019	0.875	0.040	0.013	0.005	0.482	0.018	0.004
9	0.512	0.019	0.903	0.032	0.023	0.006	0.488	0.018	0.007
10	0.522	0.022	**0.924**	0.036	0.024	0.010	**0.491**	0.017	0.006
Ave	0.420	0.020	0.732	0.034	0.022	0.008	0.421	0.016	0.008

Table 2(b). KS and MKS with $\boldsymbol{\lambda}_k = \mathbf{1}_{30} + k(0, 0.1, 0.2, \ldots, 2.9)'$ $(d = 30)$

		KS				MKS			
k	CV(λ)	χ^2_d	$L\chi^2_d$	$c\chi^2_d$	$a\chi^2_b$	χ^2_d	$L\chi^2_d$	$c\chi^2_d$	$a\chi^2_b$
1	0.353	0.020	0.899	0.027	0.017	0.005	0.488	0.012	0.005
2	0.444	0.022	0.987	0.047	0.021	0.010	0.500	0.021	0.007
3	0.485	0.013	0.996	0.039	0.011	0.004	0.500	0.020	0.004
4	0.509	0.027	0.999	0.032	0.016	0.009	0.500	0.018	0.006
5	0.525	0.027	1.000	0.036	0.022	**0.016**	0.500	0.019	0.011
6	0.535	0.020	1.000	0.042	0.022	0.007	0.500	0.021	0.008
7	0.543	0.019	1.000	0.043	0.019	0.007	0.500	0.023	0.007
8	0.550	0.021	1.000	0.039	0.013	0.005	0.500	0.020	0.004
9	0.554	**0.028**	1.000	0.034	0.017	0.008	0.500	0.017	0.007
10	0.558	0.015	**1.000**	**0.050**	**0.026**	0.006	**0.500**	**0.024**	**0.013**
Ave	0.506	0.021	0.988	0.039	0.018	0.008	0.499	0.019	0.007

Table 2(c). KS and MKS with $\boldsymbol{\lambda}_k = \mathbf{1}_{50} + k(0, 0.1, 0.2, \ldots, 4.9)'$ $(d = 50)$

		KS				MKS			
k	CV(λ)	χ^2_d	$L\chi^2_d$	$c\chi^2_d$	$a\chi^2_b$	χ^2_d	$L\chi^2_d$	$c\chi^2_d$	$a\chi^2_b$
1	0.418	0.015	0.997	0.031	0.010	0.005	0.500	0.015	0.003
2	0.489	0.025	1.000	0.026	0.020	0.009	0.500	0.015	0.005
3	0.518	0.029	1.000	0.043	0.026	0.014	0.500	0.021	0.014
4	0.534	0.018	1.000	0.033	0.017	0.006	0.500	0.017	0.007
5	0.545	0.017	1.000	0.032	0.018	0.006	0.500	0.016	0.006
6	0.551	0.013	1.000	0.040	0.022	0.004	0.500	0.022	0.006
7	0.557	0.010	1.000	0.054	0.016	0.004	0.500	0.024	0.007
8	0.560	0.028	1.000	**0.057**	**0.037**	0.012	0.500	**0.030**	0.013
9	0.563	**0.040**	1.000	0.054	0.028	**0.018**	0.500	0.028	**0.014**
10	0.566	0.015	**1.000**	0.049	0.016	0.005	**0.500**	0.027	0.006
Ave	0.530	0.021	1.000	0.042	0.021	0.008	0.500	0.022	0.008

Table 3. KS and MKS with $\boldsymbol{\lambda}_d = (1, 1.1, 1.2, \ldots, 1 + .1[d - d/10 - 1], 10, 20, \ldots, d)'$

		KS				MKS			
d	CV(λ)	χ^2_d	$L\chi^2_d$	$c\chi^2_d$	$a\chi^2_b$	χ^2_d	$L\chi^2_d$	$c\chi^2_d$	$a\chi^2_b$
10	1.147	0.022	0.519	0.116	**0.062**	0.010	0.328	0.067	**0.030**
20	1.352	0.022	0.838	0.154	0.054	0.008	0.472	0.083	0.029
30	1.462	0.022	0.960	0.157	0.042	0.010	0.498	0.088	0.024
40	1.531	0.017	0.993	0.181	0.051	0.007	0.500	0.098	0.023
50	1.579	**0.029**	1.000	0.172	0.039	**0.014**	0.500	0.096	0.018
60	1.613	0.013	1.000	0.171	0.033	0.005	0.500	0.093	0.016
70	1.640	0.018	1.000	0.185	0.034	0.008	0.500	0.100	0.018
80	1.660	0.012	1.000	0.190	0.041	0.004	0.500	0.103	0.024
90	1.677	0.028	1.000	**0.191**	0.029	0.008	0.500	0.103	0.014
100	1.691	0.019	**1.000**	0.191	0.029	0.005	**0.500**	**0.107**	0.013
Ave	1.535	0.020	0.931	0.171	0.041	0.008	0.480	0.094	0.021

$\boldsymbol{\lambda}_{10} = (1, 1.1, 1.2, \ldots, 1.7, 1.8, 10)'$, $\boldsymbol{\lambda}_{20} = (1, 1.1, 1.2, \ldots, 2.6, 2.7, 10, 20)'$, ...,
$\boldsymbol{\lambda}_{100} = (1, 1.1, 1.2, \ldots, 9.7, 9.8, 9.9, 10, 20, 30, \ldots, 100)'$