Improved tests for the equality of normal coefficients of variation

K. Krishnamoorthy & Meesook Lee

Computational Statistics

ISSN 0943-4062

Comput Stat
DOI 10.1007/s00180-013-0445-2

ONLINE FIRST

Computational Statistics 28 · 4 2013

OBITUARY

A memorial for the Late Professor Wolfgang Polasek
Y. Mori · H. Herwartz · J. Symanzik 1383

ORIGINAL PAPERS

Linear latent variable models: the lava-package
K.K. Holst · E. Budtz-Jørgensen 1385

Simplification of joint confidence regions for the parameters of the Pareto distribution
J. Zhang 1453

A single-index model procedure for interpolation intervals in time series
A.M. Alonso · A.E. Sipols · S. Quintas 1463

Maximum likelihood estimation of multinomial probit factor analysis models for multivariate t-distribution
J. Jiang · X. Liu · K. Yu 1485

An algorithm to construct Monte Carlo confidence intervals for an arbitrary function of probability distribution parameters
H. Tyralis · D. Koutsoyiannis · S. Kozanis 1501

Parameter estimation for the discretely observed fractional Ornstein–Uhlenbeck process and the Yuima R package
A. Brouste · S.M. Iacus 1529

Nonlinear nonparametric mixed-effects models for unsupervised classification
L. Azzimonti · F. Ieva · A. Maria Paganoni 1549

(Contents continued on outside back cover)

Springer

Springer

Comput Stat
DOI 10.1007/s00180-013-0445-2

ORIGINAL PAPER

Improved tests for the equality of normal coefficients of variation

K. Krishnamoorthy · Meesook Lee

Received: 27 September 2012 / Accepted: 30 July 2013

Abstract The problem of testing the equality of coefficients of variation of independent normal populations is considered. For comparing two coefficients, we consider the signed-likelihood ratio test (SLRT) and propose a modified version of the SLRT, and a generalized test. Monte Carlo studies on the type I error rates of the tests indicate that the modified SLRT and the generalized test work satisfactorily even for very small samples, and they are comparable in terms of power. Generalized confidence intervals for the ratio of (or difference between) two coefficients of variation are also developed. A modified LRT for testing the equality of several coefficients of variation is also proposed and compared with an asymptotic test and a simulation-based small sample test. The proposed modified LRTs seem to be very satisfactory even for samples of size three. The methods are illustrated using two examples.

Keywords Fixed point iteration · Generalized p value · Modified LRT · Parametric bootstrap · Power signed-LRT

1 Introduction

The coefficient of variation (CV) is defined as the ratio of the standard deviation to the mean. This is a commonly used measure of variation, because it is not affected

Electronic supplementary material The online version of this article (doi:10.1007/s00180-013-0445-2) contains supplementary material, which is available to authorized users.

K. Krishnamoorthy (✉)
Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70504, USA
e-mail: krishna@louisiana.edu

M. Lee
Department of Mathematics, South Louisiana Community College, Lafayette, LA 70504, USA

by the units of measurement. In practical situations where the CV is an appropriate measure of variability, the variable is usually positive. For a normal population, the ratio of the mean to the standard deviation has to be on the order of three or more, for the probability of a negative value is negligible. This means that the CV must be at most .33 in practical situations where the CV is a suitable measure of variability (Johnson and Welch 1940). The problem of interval estimating or testing the CV for the normal case has been well addressed in the literature. Johnson and Welch (1940) have proposed an exact method of finding confidence intervals (CIs) for the normal CV. This method is computationally involved, and so many authors have provided closed-form approximate CIs. Among the closed-form approximate CIs, the modified McKay CI proposed by Vangel (1996) seems to be the best as long as the CV is .4 or less. It should be noted that the exact method involves finding the noncentrality parameter of a noncentral t distribution, which is not an issue nowadays.

The problem of comparing two or more coefficients of variation (CsV) of different normal populations arises in many practical situations. In some chemical experiments the CV is used to judge the precision of measurements, and two measurement methods are compared on the basis of their respective CsV. Miller and Karson (1977) noted that the CV can be used as a measure of relative risk in financial analysis, and a test for equality of the CsV of two stocks is useful to determine whether they involve similar risk. CsV are also used to compare two measurement methods. Hamer et al. (1995) used the CV to assess homogeneity of bone test samples produced by a new method. Some asymptotic tests for testing the equality of two CsV are proposed in the literature (e.g., see Bhoj and Ahsanullah 1993 and Fung and Tsang 1998). These asymptotic tests are not accurate even for moderately large samples, and they may be used only for very large samples. Feltz and Miller (1996) have proposed an asymptotic test that appears to be satisfactory, and performs better than many other approximate tests proposed in the literature.

In this article, we first consider a signed-likelihood ratio test (SLRT), and propose a modified version of the SLRT as suggested by DiCiccio et al. (2001) in a general setup. The modified version of the SLRT (M-SLRT) turns out to be far superior than the SLRT in terms of controlling type I error rates. Furthermore, the M-SLRT works satisfactorily even for samples of size three. We also propose a similar modification of the likelihood ratio test (LRT) for testing the equality of several CsV. The modified LRT also works well even for very small samples. For comparing two coefficients of variation, we also propose a test based on the GV approach. The GV test is based on the generalized pivotal quantity (GPQ) for individual coefficients of variations, and is quite comparable with the M-SLRT for the two-sample problem. The GV test and the M-SLRT are compared with the F approximate test proposed by Forkman (2009).

The rest of the article is organized as follows. In the following section, we describe the methods for obtaining the maximum likelihood estimates (MLEs) over unrestricted parameter space and over a restricted parameter space. For the two-sample problem, closed-form MLEs are available in the literature. For more than two samples, we propose a simple fixed point iteration to compute the MLEs. In Sect. 3, we describe the SLRT, M-SLRT, and the GV test for the two-sample problem. The type I error rates and powers of all the tests are evaluated using Monte Carlo simulation, and compared. In Sect. 4, we describe the LRT for the equality of several coefficients of variation, and propose a modification to the LRT. The LRT, the modified LRT and the asymptotic test by Feltz and Miller (1996) are evaluated by Monte Carlo simulation. Two examples involving real-life data are illustrated in Sect. 5. Some concluding remarks are provided in Sect. 6.

2 The MLEs

Let $(\bar{X}_i, S_i^2)$ denote the mean and variance (unbiased estimate) of a random sample of size n_i from a $N(\mu_i, \tau_i^2\mu_i^2)$ distribution, $i = 1, \ldots, k$. Let $(\bar{x}_i, s_i^2)$ be an observed value of $(\bar{X}_i, S_i^2)$, $i = 1, \ldots, k$, and let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$ and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_k)$. The log-likelihood function under the unrestricted parameter space $\{\mu_i > 0, \tau_i > 0, \ i = 1, \ldots, k\}$, after omitting the constant terms, is given by

$$l(\boldsymbol{\mu}, \boldsymbol{\tau}) = -\sum_{i=1}^{k} n_i \ln(\tau_i\mu_i) - \sum_{i=1}^{k} \frac{n_i(V_i^2 + (\bar{X}_i - \mu_i)^2)}{2\tau_i^2\mu_i^2}, \tag{1}$$

where $V_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2, i = 1, 2$. The MLEs over the unrestricted parameter space are (see Bhoj and Ahsanullah 1993) given by

$$\widehat{\mu}_i = \bar{X}_i \ \text{ and } \ \widehat{\tau}_i = \frac{V_i}{\bar{X}_i}, \quad i = 1, \ldots, k.$$

The log-likelihood function under $H_0 : \tau_1 = \cdots = \tau_k = \tau$, after omitting the constant terms, can be expressed as

$$l(\boldsymbol{\mu}, \tau) = -\sum_{i=1}^{k} n_i \ln(\tau\mu_i) - \sum_{i=1}^{k} \frac{n_i(V_i^2 + (\bar{X}_i - \mu_i)^2)}{2\tau^2\mu_i^2}, \tag{2}$$

where τ denotes the unknown common coefficient of variation under H_0. For the case of $k = 2$, Gerig and Sen (1980) derived the constrained MLEs given by

$$\widehat{\mu}_1^* = n_1\bar{X}_1\widehat{\mu}_2^*/(n\widehat{\mu}_2^* - n_2\bar{X}_2),$$
$$\widehat{\tau}^{*2} = (V_2^2 + \bar{X}_2^2 - \bar{X}_2\widehat{\mu}_2^*)/\widehat{\mu}_2^{*2},$$

and

$$\widehat{\mu}_2^* = (-b + \sqrt{b^2 - 4ac})/(2a),$$

where $a = n\widehat{\tau}_1^2 + n_2, b = [2n_2\widehat{\tau}_1^2 + (2n_2 - n_1)]\bar{X}_2, c = [n_2^2(\widehat{\tau}_1^2 + 1) - n_1^2(\widehat{\tau}_2^2 + 1)]\bar{X}_2^2/n$, and $n = n_1 + n_2$. For the case $n_1 = n_2 = n$, the MLEs can be simplified further; see Lohrding (1969).

To find the constrained MLEs for the case $k \geq 3$, note that the partial derivatives of (2) with respect to $\mu_1, \ldots, \mu_k$ and τ yield the following equations:

$$\tau^2\mu_i^2 + \bar{X}_i\mu_i - (V_i^2 + \bar{X}_i^2) = 0, \quad i = 1, \ldots, k, \tag{3}$$

and

$$\tau^2 = \sum_{i=1}^{k} w_i (V_i^2 + (\bar{X}_i - \mu_i)^2)/\mu_i^2, \tag{4}$$

where $w_i = n_i / \sum_{j=1}^{k} n_j$, $i = 1, \ldots, k$. Analytic solutions for the above equations are not available for $k \geq 3$. However, the MLEs can be obtained using a simple fixed point iteration as follows. Solving (3) for μ_i, we get

$$\mu_i(\tau^2) = \frac{-\bar{X}_i + \sqrt{\bar{X}_i^2 + 4\tau^2(V_i^2 + \bar{X}_i^2)}}{2\tau^2}, \quad i = 1, \ldots, k.$$

Substituting the above $\mu_i(\tau^2)$ for μ_i in (4), we get the following fixed point iteration:

$$\tau_{j+1}^2 = \sum_{i=1}^{k} w_i (V_i^2 + (\bar{X}_i - \mu_i(\tau_j^2))^2)/\mu_i^2(\tau_j^2), \quad j = 0, 1, \ldots \tag{5}$$

We can use

$$\tau_0^2 = \sum_{i=1}^{k} w_i V_i^2 / \bar{X}_i^2$$

as an initial value for the above fixed point iteration. Even though we are unable to verify any sufficient condition under which the above iteration process converges, we observed in our simulation study that the fixed point iteration scheme converges in a few iterations (in most cases fewer than five iterations) with the initial value defined above, and the stopping rule $|\tau_{j+1}^2 - \tau_j^2| < 10^{-7}$. Once the iterative process is stopped at the mth step, the corresponding $\mu_i(\tau_m^2)$ is the MLE of $\mu_i, i = 1, \ldots, k$. Let us denote the constrained MLEs obtained using the above iterative scheme by $\widehat{\mu}_1^*, \ldots, \widehat{\mu}_k^*$ and $\widehat{\tau}^*$.

3 Tests and confidence intervals for comparing two coefficients of variation

Let $(\bar{X}_i, S_i^2)$ denote the mean and variance of a random sample of size n_i from a $N(\mu_i, \tau_i^2\mu_i^2)$ distribution, $i = 1, 2$. Let $(\bar{x}_i, s_i^2)$ be an observed value of $(\bar{X}_i, S_i^2)$, $i = 1, 2$. We shall describe some tests for $\tau_1 - \tau_2$ and for the ratio τ_1/τ_2.

3.1 The signed-likelihood ratio test

Note that the log-likelihood function (1) at $(\mu_1, \tau_1, \mu_2, \tau_2) = (\widehat{\mu}_1, \widehat{\tau}_1, \widehat{\mu}_2, \widehat{\tau}_2)$ simplifies to

$$l(\widehat{\mu}_1, \widehat{\tau}_1, \ldots, \widehat{\mu}_k, \widehat{\tau}_k) = -\sum_{i=1}^{k} n_i \ln(V_i) - \sum_{i=1}^{k} \frac{n_i}{2}. \tag{6}$$

It follows from (2) and (4) that

$$l(\widehat{\mu}_1^*, \ldots, \widehat{\mu}_k^*, \widehat{\tau}^*) = -\sum_{i=1}^{k} n_i \ln(\widehat{\tau}^* \widehat{\mu}_i^*) - \sum_{i=1}^{k} \frac{n_i}{2}. \tag{7}$$

The SLRT statistic for testing $\tau_1 = \tau_2$ is given by

$$\begin{aligned} R &= \operatorname{sign}(\widehat{\tau}_1 - \widehat{\tau}_2) \left\{2[\ln(\widehat{\mu}_1, \widehat{\tau}_1, \widehat{\mu}_2, \widehat{\tau}_2) - \ln(\widehat{\mu}_1^*, \widehat{\mu}_2^*, \widehat{\tau}^*)]\right\}^{1/2}, \\ &= \operatorname{sign}(\widehat{\tau}_1 - \widehat{\tau}_2) \left\{2\left[n_1 \ln\left(\frac{\widehat{\tau}^* \widehat{\mu}_1^*}{V_1}\right) + n_2 \ln\left(\frac{\widehat{\tau}^* \widehat{\mu}_2^*}{V_2}\right)\right]\right\}^{1/2}, \end{aligned} \tag{8}$$

where $\operatorname{sign}(x) = 1$ if $x > 0$, -1 otherwise. The SLRT statistic R has an asymptotic standard normal distribution. For testing

$$H_0 : \tau_1 = \tau_2 \text{ versus } H_a : \tau_1 \neq \tau_2, \tag{9}$$

the null hypothesis is rejected at the level of α if $|R| > z_{1-\alpha/2}$, where z_p is the $100p$ percentile of the standard normal distribution. For testing

$$H_0 : \tau_1 \leq \tau_2 \text{ versus } H_a : \tau_1 > \tau_2, \tag{10}$$

the null hypothesis is rejected if $R > z_{1-\alpha}$.

The SLRT is in general known to be liberal for small to moderate samples (see Verrill and Johnson 2007a). Barndorff-Nielsen (1991) proposed a modification to the SLRT in a general setup so that the modified SLRT statistic follows a standard normal distribution up to an order of $O(n^{-3/2})$. However, development of the modified SLRT using Barndroff-Nielsen's approach is quite involved, and so we choose an equivalent simpler approach proposed in DiCiccio et al. (2001). To describe their approach, let $m(R)$ and $v(R)$ denote respectively the mean and variance of R evaluated at the constrained MLEs $(\widehat{\mu}_1^*, \widehat{\mu}_2^*, \tau^*)$. Now define the modified statistic

$$R_M = \frac{R - m(R)}{\sqrt{v(R)}}. \tag{11}$$

Then, as noted in DiCiccio et al. (2001), R_M follows a standard normal distribution up to an order of $O(n^{-3/2})$ while R itself follows a standard normal distribution with $O(n^{-1/2})$. Another third-order accurate method consists of estimating the p value of an observed value of R at $(\widehat{\mu}_1^*, \widehat{\mu}_2^*, \widehat{\tau}^*)$, or equivalently, percentiles of R at $(\mu_1, \mu_2, \tau) = (\widehat{\mu}_1^*, \widehat{\mu}_2^*, \widehat{\tau}^*)$. The H_0 in (10) is rejected if an observed value of R is greater than the estimated $100(1-\alpha)$ percentile of R at $(\widehat{\mu}_1^*, \widehat{\mu}_2^*, \widehat{\tau}^*)$. As noted in DiCiccio et al. (2001), analytic expressions are often not available for the mean $m(R)$ and variance $v(R)$, or for the percentiles of R. However, these quantities can be easily approximated by a parametric bootstrap (PB) method, that is, based on simulated samples from $N(\widehat{\mu}_1^*, \widehat{\tau}^* \widehat{\mu}_1^*)$ and $N(\widehat{\mu}_2^*, \widehat{\tau}^* \widehat{\mu}_2^*)$. A PB approach can also be used to calculate the percentiles of R. We shall refer to the tests based on these modified approaches as M-SLRT. Calculation details of the M-SLRT are given in the following Algorithm 1.

Algorithm 1 For given data sets, compute $\bar{x}_i$ and $v_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$.

Using the above quantities, compute the MLEs $(\widehat{\mu}_1, \widehat{\tau}_1, \widehat{\mu}_2, \widehat{\tau}_2)$ and the constrained MLEs $\widehat{\mu}_1^*$, $\widehat{\mu}_2^*$ and $\widehat{\tau}^*$ given in Sect. 2. Compute the SLRT statistic R in (8). Set $s_i^* = \widehat{\tau}^*\widehat{\mu}_i^*, i = 1, 2$.

For i = 1 to N_2

Generate $\bar{x}_i^* \sim N(\widehat{\mu}_i^*, s_i^{*2}/n_i)$ and $v_i^{*2} \sim s_i^{*2}\chi^2_{n_i-1}/n_i, i = 1, 2$.
Compute the unconstrained MLEs $(\widehat{\mu}_{B1}, \widehat{\mu}_{B2}, \widehat{\tau}_{B1}, \widehat{\tau}_{B2})$ and the constrained MLEs $(\widehat{\mu}_{B1}^*, \widehat{\mu}_{B2}^*, \widehat{\tau}_B^*)$ based on $(\bar{x}_1^*, v_1^*, \bar{x}_2^*, v_2^*)$.
Compute the SLRT statistic in (8) using the constrained and the unconstrained MLEs in the above step, and denote it by R^*.

(end do loop)

Compute the mean m^* and variance v^* of the R^*'s generated above, and compute the modified SLRT $R_M = \frac{R-m^*}{\sqrt{v^*}}$. The modified SLRT rejects the H_0 in (9) if $|R_M| > z_{1-\alpha/2}$, and rejects H_0 in (10) if $R_M > z_{1-\alpha}$.

3.2 Generalized variable approach

A generalized variable (GV) test is based on the so called generalized pivotal quantity (GPQ) for τ_1/τ_2. The GPQ for τ_i can be obtained from the ones for μ_i and σ_i by substitution. The GPQs for a normal mean and variances are now well-known and they are described in many articles. For example, see Krishnamoorthy and Mathew (2003), Weerahandi (Example 1.2, 2004), Tian (2005), and Tian and Wu (2007).

The GPQ for μ_i is given by

$$G_{\mu_i} = \bar{x}_i - \frac{Z_i}{\sqrt{\chi^2_{n_i-1}/(n_i-1)}}\frac{s_i}{\sqrt{n_i}}, \quad i = 1, 2,$$

and the one for σ_i is given by

$$G_{\sigma_i} = \sqrt{\frac{(n_i-1)s_i^2}{\chi^2_{n_i-1}}}, \quad i = 1, 2,$$

where all the random variables $Z_1, Z_2, \chi^2_{n_1-1}$ and $\chi^2_{n_2-1}$ are mutually independent with $Z_i \sim N(0, 1), i = 1, 2$.

Using these GPQs for μ_i and σ_i, a GPQ for τ_i is obtained by substitution, and is given by

$$G_{\tau_i} = \left(\frac{\bar{x}_i\sqrt{W_i}}{s_i} - \frac{Z_i}{\sqrt{n_i}}\right)^{-1}. \tag{12}$$

where $W_i = \chi^2_{n_i-1}/(n_i - 1)$. Note that this GPQ for τ_i could be negative even though τ_i is assumed to positive. This is not a problem in finding a CI (or a test) for the ratio, because a CI for τ_1/τ_2 can be obtained from the one for τ_1^2/τ_2^2.

A GPQ for τ_1^2/τ_2^2 is given by

$$Q_{\tau_1^2/\tau_2^2} = G^2_{\tau_1}/G^2_{\tau_2} = \left(\frac{\bar{x}_2\sqrt{W_2}}{s_2} - \frac{Z_2}{\sqrt{n_2}}\right)^2 \Big/ \left(\frac{\bar{x}_1\sqrt{W_1}}{s_1} - \frac{Z_1}{\sqrt{n_1}}\right)^2. \tag{13}$$

The 100α percentile, and the $100(1-\alpha)$ percentile of $Q_{\tau_1^2/\tau_2^2}$ form a $1-2\alpha$ CI for τ_1^2/τ_2^2, and the square root of this CI is a $1-2\alpha$ CI for τ_1/τ_2.

To find a CI for the difference $\tau_1 - \tau_2$, we propose the GPQ

$$Q_{\tau_1-\tau_2} = \sqrt{G^2_{\tau_1}} - \sqrt{G^2_{\tau_2}}. \tag{14}$$

For a given $(\bar{x}_1, s_1, \bar{x}_2, s_2)$, the distribution of $Q_{\tau_1^2/\tau_2^2}$ (or that of $Q_{\tau_1-\tau_2}$) does not depend on any unknown parameters, and so they can be estimated using Monte Carlo simulation as shown in Algorithm 2.

The generalized test for (10) is described as follows. Let $Q_{\tau_1^2/\tau_2^2;\alpha}$ denote the α quantile of $Q_{\tau_1^2/\tau_2^2}$. Note that $Q_{\tau_1^2/\tau_2^2;\alpha}$ is the $1-\alpha$ lower confidence limit for τ_1^2/τ_2^2. The null hypothesis in (10) is rejected if this lower confidence limit is greater than one, or equivalently,

$$P_1 = P\left(Q_{\tau_1^2/\tau_2^2} < 1\right) \le \alpha.$$

The above probability is referred to as the generalized p value. The generalized p value for testing (9) is given by

$$2\min\{P_1, 1 - P_1\}.$$

These generalized p values and the generalized CLs can be estimated using Monte carlo simulation as shown in the following Algorithm 2.

Algorithm 2 For given samples from $N(\mu_1, \tau_1^2\mu_1^2)$ and $N(\mu_2, \tau_2^2\mu_2^2)$, compute $(\bar{x}_1, s_1, \bar{x}_2, s_2)$.

For i = 1 to N_2

Generate $Z_j \sim N(0, 1)$ and $W_j \sim \chi^2_{n_j-1}/(n_j - 1)$, $j = 1, 2$.
Calculate $Q_i = Q_{\tau_1^2/\tau_2^2} = G^2_{\tau_1}/G^2_{\tau_2}$ using (13).
Set $I_i = 1$ if $Q_i < 1$; else set $I_i = 0$.

(end do loop)

The proportion $P_1 = \frac{1}{N_2}\sum_{i=1}^{N_2} I_i$ is a Monte Carlo estimate of the generalized p value for testing (10). Furthermore, $2\min\{P_1, 1 - P_1\}$ is a Monte carlo estimate of the generalized p value for testing hypotheses in (9). The 100α percentile of these Q_i's generated above is a $100\alpha\%$ lower confidence limit for τ_1^2/τ_2^2.

3.3 Forkman's test

Forkman (2009) proposed a test based on a statistic that has an approximate F distribution. To outline the test, let $\tilde{\tau}_i = S_i/\bar{X}_i$, where S_i^2 is the usual unbiased estimate of σ_i^2, $i = 1, 2$. The test statistic proposed by Forkman is given by

$$F = \frac{\tilde{\tau}_1^2/[1 + \tilde{\tau}_1^2(n_1 - 1)/n_1]}{\tilde{\tau}_2^2/[1 + \tilde{\tau}_2^2(n_2 - 1)/n_2]}. \tag{15}$$

Forkman has shown that the above statistic has an approximate F distribution with degrees of freedom $n_1 - 1$ and $n_2 - 1$. This test rejects H_0 in (10) when an observed value of F in (15) is greater than $F_{n_1-1,n_2-1;1-\alpha}$.

3.4 Simulation studies

To judge the accuracy of the proposed tests in the preceding sections and to compare them, we estimated their type I error rates and powers for testing one-sided hypotheses. Earlier simulation studies (e.g., Verrill and Johnson 2007b) and our own studies indicated that the SLRT is satisfactory only for large approximately equal sample sizes. Specifically, we observed that the SLRT works satisfactorily only when both sample sizes are at least 40 and they are not heavily unbalanced. So the SLRT is not included in our comparison study for small to moderate samples.

The type I error rates and powers of the M-SLRT and the generalized test for testing hypotheses in (10) are evaluated using the following Algorithm 3.

Algorithm 3 For given parameters (τ_1, τ_2) and sample sizes n_1 and n_2, set $\sigma_1 = \sigma_2 = 1$, $\mu_1 = \sigma_1/\tau_1$ and $\mu_2 = \sigma_2/\tau_2$.

For i = 1 to N_1

Generate $\bar{x}_j \sim N(\mu_j, \sigma_j^2/n_j)$ and $v_j^2 \sim \chi^2_{n_j-1}\sigma_j^2/n_j$, $j = 1, 2$.
Use Algorithm 1 to compute the modified SLRT R_M, and Algorithm 2 to find the generalized p value P_1 for testing (10).
Set $I_i = 1$ if $R_M > z_{1-\alpha}$; else set $I_i = 0$. Set $h_i = 1$ if $P_1 \leq \alpha$; else set $h_i = 0$.

(end do loop)

Compute $\widehat{\alpha}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} I_i$ and $\widehat{\alpha}_2 = \frac{1}{N_1}\sum_{i=1}^{N_1} h_i$. If $\tau_1 = \tau_2$, then $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ are Monte Carlo estimates of the type 1 error rates of the M-SLRT and the generalized test, respectively. If $\tau_1 > \tau_2$, then these estimates are the powers for testing (10). Type I error rates of a left-tailed test or of a two-tailed test can be estimated similarly for Fortran programs to estimate type I error rates and powers, see online supplementary files.

We used Algorithm 3 with $N_1 = 10{,}000$ and Algorithm 2 with $N_2 = 10{,}000$ to estimate the type I error rates of the generalized test, and used the same values for N_1 and N_2 to estimate the type I error rates of the M-SLRT. The type I error rates of the Forkman test are evaluated using simulation of 10,000 runs. The estimated type I error rates for testing one-sided hypotheses at the nominal level .05 are given in Table 1. Estimated type I error rates clearly indicate that all three tests control type I error rates very satisfactorily even for samples of sizes three.

Table 1 Type I error rates of the tests at level .05

$H_0: \tau_1 \leq \tau_2$ versus $H_a: \tau_1 > \tau_2$

$\tau_1 = \tau_2$	(n_1, n_2)														
	(3, 3)			(3, 5)			(5, 9)			(10, 10)			(10, 15)		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
.05	.050	.049	.050	.052	.051	.050	.052	.054	.049	.049	.052	.049	.050	.051	.050
.10	.050	.049	.049	.049	.053	.050	.051	.051	.049	.049	.052	.051	.049	.053	.049
.20	.048	.055	.050	.049	.049	.051	.047	.051	.050	.049	.052	.052	.049	.049	.050
.30	.050	.049	.050	.050	.051	.052	.046	.050	.052	.050	.054	.052	.050	.052	.051
.40	.050	.049	.050	.051	.052	.051	.049	.052	.051	.051	.049	.052	.049	.050	.051
.50	.050	.052	.050	.054	.048	.053	.053	.052	.053	.054	.053	.052	.051	.050	.049

$H_0: \tau_1 \geq \tau_2$ versus $H_a: \tau_1 < \tau_2$

$\tau_1 = \tau_2$	(n_1, n_2)														
	(3, 3)			(3, 5)			(5, 9)			(10, 10)			(10, 15)		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
.05	.050	.050	.049	.049	.051	.049	.052	.050	.050	.047	.047	.050	.048	.048	.052
.10	.048	.050	.050	.050	.047	.051	.051	.049	.050	.052	.048	.050	.049	.048	.051
.20	.050	.054	.051	.053	.050	.049	.046	.047	.049	.052	.051	.050	.052	.050	.050
.30	.049	.050	.050	.051	.051	.049	.053	.047	.048	.052	.051	.050	.051	.049	.049
.40	.047	.049	.050	.049	.048	.048	.049	.053	.047	.051	.049	.050	.052	.052	.048
.50	.048	.054	.048	.054	.050	.046	.047	.050	.045	.050	.053	.048	.054	.051	.046

1. Modified SLRT, 2. GV test, 3. Forkman's test

To judge the power properties of the tests, we estimated powers of the M-SLRT, the generalized test and Forkman's test, and presented them in Table 2. Examination of powers in Table 2 clearly indicates that all three tests have desirable power properties: For a fixed (n_1, n_2) the powers are increasing with increasing $\tau_1 - \tau_2$. For a fixed (τ_1, τ_2), the powers are increasing with increasing sample sizes. Finally, we note that all three tests are quite similar in terms of powers. As the Forkman test is much simpler to use than the other two tests, it is preferable to the other tests for testing the ratio of two coefficients of variation. The GV test is recommended for testing the difference between two coefficients of variation because it is relatively easier to implement than the M-SLRT. Finally, we note that the GV approach is useful for finding confidence intervals for the difference or the ratio of coefficients of variation, whereas the Forkman test is not useful to find a confidence interval, and the MLRT is not easy to invert to find a confidence interval. A web-based program based on Verrill and Johnson's (2007b) simulation approach can be used to find a confidence interval for a CV; visit "http://www1.fpl.fs.fed.us/covratio.html".

We also evaluated type I error rates at other nominal levels .01 and .1. The type I error rates of all three test are very close to the nominal levels, and so these values are not reported here.

Table 2 Powers of the tests at level .05

$H_0 : \tau_1 \le \tau_2$ versus $H_a : \tau_1 > \tau_2$

$\tau_2 = .05$	τ_1														
(n_1, n_2)	.05			.1			.15			.2			.25		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
(5, 5)	.049	.049	.050	.331	.334	.329	.625	.625	.624	.795	.810	.793	.889	.882	.886
(5, 7)	.049	.051	.050	.421	.416	.419	.731	.745	.732	.875	.865	.875	.936	.940	.936
(8, 8)	.049	.048	.050	.525	.528	.528	.857	.860	.858	.959	.961	.959	.986	.987	.987
(10, 10)	.050	.050	.050	.628	.630	.628	.929	.930	.927	.987	.987	.986	.997	.997	.997
(10, 15)	.051	.048	.050	.724	.726	.727	.964	.960	.962	.995	.994	.994	.999	.999	.999

1. Modified SLRT, 2. GV test, 3. Forkman's test

4 Tests for the equality of several coefficients of variation

Let $\bar{X}_i$ and S_i^2 denote the mean and variance of a random sample of size n_i from a $N(\mu_i, \sigma_i^2)$ distribution, $i = 1, \ldots, k$. We shall now develop tests for

$$H_0 : \tau_1 = \cdots = \tau_k \ \text{ versus } \ H_a : \tau_i \neq \tau_j \text{ for some } i \neq j, \tag{16}$$

and compare them via Monte Carlo simulation.

4.1 The LRT and a modified LRT

Using (6) and (7), the LRT statistic can be written as

$$\begin{aligned} \Lambda &= 2[\ln(\widehat{\mu}_1, \widehat{\tau}_1, \ldots, \widehat{\mu}_k, \widehat{\tau}_k) - \ln(\widehat{\mu}_1^*, \ldots, \widehat{\mu}_k^*, \widehat{\tau}^*)] \\ &= 2\left[\sum_{i=1}^{k} n_i \ln\left(\frac{\widehat{\tau}^* \widehat{\mu}_i^*}{V_i}\right)\right], \end{aligned} \tag{17}$$

and it has an asymptotic χ^2_{k-1} distribution. The LRT rejects the H_0 in (16) when $\Lambda > \chi^2_{k-1;1-\alpha}$.

As in the two-sample case in Sect. 3, we can improve the LRT by using its estimated mean and variance. Let $m(\Lambda)$ and $v(\Lambda)$ denote the mean and variance of Λ, respectively. Then

$$\Lambda_M = \sqrt{2(k-1)}\left(\frac{\Lambda - m(\Lambda)}{\sqrt{v(\Lambda)}}\right) + (k-1) \sim \chi^2_{k-1}. \tag{18}$$

Note that Λ_M is obtained so that the mean and variance of Λ_M are the same as those of χ^2_{k-1}. This modified LRT rejects H_0 when $\Lambda_M > \chi^2_{k-1;1-\alpha}$. Expressions for $m(\Lambda)$ and $v(\Lambda)$ are difficult to obtain, and as in the two-sample problem, we shall estimate them using simulated samples from $N(\widehat{\mu}_1^*, \widehat{\tau}^{*2}\widehat{\mu}_1^{*2}), \ldots, N(\widehat{\mu}_k^*, \widehat{\tau}^{*2}\widehat{\mu}_k^{*2})$, and algorithm similar to Algorithm 1. We refer to the test based on (17) as the LRT, and the one based on (18) as the modified LRT (MLRT).

Remark 1 Verrill and Johnson (2007a,b) proposed a simulation based small sample approach for testing equality of several coefficients of variation. Their approach is the alternative approach based on the percentile of R at $(\widehat{\mu}_1^*, \ldots, \widehat{\mu}_k^*, \widehat{\tau}^*)$. As mentioned earlier in Sect. 3, DiCiccio et al. (2001) noted that this approach is also third order accurate. Therefore, the MLRT based on (18) and the one proposed in Verrill and Johnson should be similar. Indeed, a reviewer's limited simulation study indicted that the power and size properties of these two tests are very similar.

4.2 An asymptotic test

Feltz and Miller (1996) proposed an asymptotic test which is outlined as follows. Let $m_i = n_i - 1$, and define $\widehat{\tau}_c = M^{-1}\left(\sum_{j=1}^k m_j S_j/\bar{x}_j\right)$, where $M = \sum_{j=1}^m m_j$. The test statistic for testing (16) is given by

$$A = \widehat{\tau}_c^{\,-2}\left(.5 + \widehat{\tau}_c^2\right)^{-1}\left[\sum_{i=1}^k m_i(S_i/\bar{X}_i)^2 - M\widehat{\tau}_c^2\right], \tag{19}$$

which has an asymptotic χ^2_{k-1} distribution. This test rejects H_0 in (16) if $A > \chi^2_{k-1;1-\alpha}$.

4.3 Comparison of the MLRT and the asymptotic test

To understand the size and power properties of the MLRT and the asymptotic tests, we estimate the type I error rates and powers of these tests in various settings using Monte Carlo simulation. We evaluated the type I error rates of the LRT and the asymptotic tests for some moderate and large sample sizes, and reported them in Table 3. We see in Table 3 that the asymptotic test is clearly better than the LRT in controlling type I error rates. The LRT is not satisfactory even for samples of sizes 40. Overall, we see that the asymptotic test is certainly preferable to the LRT for simplicity and for controlling type I error rates.

In Table 4, we present the type I error rates of the MLRT and the asymptotic test for some small samples. As in the two-sample case, the type I error rates of the MLRT were estimated using $N_1 = 10{,}000$ and $N_2 = 10{,}000$, and those of the asymptotic test were estimated using simulation consisting of 100,000 runs. The type I error rates clearly indicate that the MLRT is very satisfactory for all sample size and parameter configurations reported in Table 4. In particular, we see for the case of $(n_1, \ldots, n_7) = (2, \ldots, 2)$ that the type I error rates of the MLRT are very close to the nominal level .05. The asymptotic test also performs satisfactorily in terms of controlling type I error rates except for very small sample sizes and/or the coefficients of variation are large. Overall, we see that the MLRT outperforms the asymptotic test in maintaining the type I error rates very close to the nominal level.

Table 3 Type I error rates of the LRT and the asymptotic test for the equality of several coefficients of variation

$H_0 : \lambda_1 = \cdots = \lambda_k$ versus $H_a : \lambda_i \neq \lambda_j$ for some $i \neq j$; $\alpha = .05$

(n_1, n_2, n_3)	Methods	$k = 3$ $\tau_1 = \tau_2 = \tau_3$							
		.05	.10	.15	.20	.30	.35	.4	.5
(15, 15, 15)	LRT	.065	.067	.067	.067	.065	.067	.068	.067
	Asym	.052	.054	.054	.052	.051	.052	.052	.048
(20, 20, 20)	LRT	.062	.063	.063	.061	.063	.061	.062	.063
	Asym	.053	.054	.053	.052	.052	.050	.051	.049
(15, 15, 60)	LRT	.067	.066	.067	.067	.065	.065	.064	.066
	Asym	.050	.050	.051	.052	.050	.051	.051	.052
(30, 30, 30)	LRT	.058	.058	.059	.058	.057	.058	.057	.057
	Asym	.052	.053	.052	.052	.050	.050	.050	.048
(40, 40, 40)	LRT	.055	.055	.055	.055	.056	.056	.057	.055
	Asym	.051	.050	.051	.051	.051	.051	.050	.049
$(n_1, \ldots, n_5)$	Methods	$k = 5$ $\tau_1 = \cdots = \tau_5$							
		.05	.10	.15	.20	.30	.35	.4	.5
(15, 15, 15, 15, 15)	LRT	.072	.072	.071	.070	.072	.071	.071	.071
	Asym	.054	.055	.053	.053	.054	.053	.054	.054
(20, 20, 20, 20, 20)	LRT	.066	.065	.066	.065	.065	.066	.066	.066
	Asym	.054	.053	.054	.052	.053	.053	.053	.054
(15, 15, 60, 10, 5)	LRT	.095	.094	.094	.094	.091	.093	.091	.091
	Asym	.048	.048	.048	.049	.050	.055	.055	.065
(10, 10, 30, 30, 10)	LRT	.078	.077	.078	.079	.077	.079	.078	.078
	Asym	.051	.050	.051	.051	.053	.054	.054	.059
(40, 40, 40, 40, 40)	LRT	.057	.058	.055	.057	.058	.057	.057	.057
	Asym	.052	.052	.049	.052	.052	.051	.051	.051

Finally, we compare the MLRT and the asymptotic test in terms of power for various sample size and parameter configurations. The estimated powers of the tests are given in Table 5. We first observe from the first part ($k = 3$) of the table that the asymptotic test has a peculiar power property. Specifically, at $(\tau_1, \tau_2, \tau_3) = (.1, .2, .2)$ and $(n_1, n_2, n_3) = (5, 5, 5)$ the power of the asymptotic test is .189, at $(n_1, n_2, n_3) = (5, 10, 5)$ the power is .153, and at $(n_1, n_2, n_3) = (5, 10, 10)$ the power is .138. The power is decreasing with increasing sample sizes! However, the power at the sample sizes (10, 10, 10) is .448 which is comparable to the power of the MLRT. We also observe similar peculiar power behavior of the asymptotic test when $(\tau_1, \tau_2, \tau_3) = (.1, .4, .4)$ and $(.1, .5, .4)$. This asymptotic test has this unusual power property when larger sample sizes are associated with the larger coefficients of variation. On the other

Table 4 Type I error rates of the modified LRT for testing equality of several coefficients of variation

$H_0 : \lambda_1 = \cdots = \lambda_k$ versus $H_a : \lambda_i \neq \lambda_j$ for some $i \neq j$; $\alpha = .05$

(n_1, n_2, n_3)	Methods	$k = 3$ $\tau_1 = \tau_2 = \tau_3$								
		.05	.10	.15	.20	.25	.30	.35	.4	.5
(3, 3, 3)	MLRT	.049	.052	.050	.053	.049	.048	.051	.050	.053
	Asym	.060	.058	.057	.056	.054	.053	.049	.047	.040
(4, 4, 12)	MLRT	.048	.052	.054	.052	.050	.050	.050	.049	.049
	Asym	.048	.047	.047	.047	.046	.048	.048	.050	.057
(5, 3, 21)	MLRT	.051	.053	.051	.047	.050	.052	.051	.049	.046
	Asym	.040	.040	.041	.040	.042	.044	.048	.051	.061
(3, 6, 4)	MLRT	.051	.051	.052	.046	.053	.045	.051	.052	.051
	Asym	.054	.054	.053	.053	.053	.051	.049	.051	.047
(8, 12, 7)	MLRT	.051	.048	.052	.048	.052	.052	.050	.048	.048
	Asym	.054	.055	.054	.054	.053	.052	.051	.052	.050
(10, 10, 10)	MLRT	.050	.050	.053	.051	.049	.051	.049	.048	.051
	Asym	.055	.055	.055	.054	.054	.054	.052	.050	.048

hand, the power of the MLRT increases with increasing sample sizes; see the powers for the case of $k = 3$ in Table 5.

Overall, we observe that the MLRT controls the type I error rates even for very small sample sizes, and posses some desirable power properties. As noted earlier in Remark 1, the MLRT and the simulation-based approach by Verrill and Johnson (2007b) are similar in terms of size and power properties, and so we recommend either of the test procedures for practical applications.

5 Examples

Example 1 For the purpose of promoting the quality and the standards of medical laboratory technology, Hong Kong Medical Technology Association conducted a quality assurance program for medical laboratories in Hong Kong in 1989 . In the specialty of hematology and serology, one normal and one abnormal hematology and serology blood samples were sent to participants for measurements of Hb, RBC, MCV, Hct, WBC and Platelet in each survey. The summary statistics of data on normal blood samples collected from the third surveys of 1995 and 1996 are given in Fung and Tsang (1998), and they are reproduced here in Table 6.

We calculated p values for testing the equality of CsV of the 1995 and 1996 surveys and presented them in Table 7. We see from this table that the p values of all three tests are in close agreement. Furthermore, the conclusions based on all the tests are the same for each pair of variables at the level of significance .05. These results are in agreement with our earlier comparison of the tests. We also presented 95 % generalized CIs for the ratios of the CsV and for the differences of the CsV in Table 7. We once again

Table 4 continued

$H_0 : \lambda_1 = \cdots = \lambda_k$ versus $H_a : \lambda_i \neq \lambda_j$ for some $i \neq j$; $\alpha = .05$

$(n_1, \ldots, n_4)$	Methods	$k = 4$, $\tau_1 = \cdots = \tau_4$								
		.05	.10	.15	.20	.25	.30	.35	.4	.5
(3, 3, 3, 3)	MLRT	.051	.053	.051	.050	.049	.050	.050	.051	.048
	Asym	.065	.067	.064	.064	.062	.061	.059	.057	.050
(5, 2, 5, 2)	MLRT	.051	.048	.054	.049	.048	.049	.048	.050	.050
	Asym	.050	.051	.051	.052	.052	.053	.055	.058	.067
(4, 4, 12, 3)	MLRT	.048	.051	.048	.048	.050	.049	.049	.049	.049
	Asym	.048	.049	.049	.049	.051	.052	.053	.058	.068
(5, 3, 2, 4)	MLRT	.047	.052	.049	.049	.047	.053	.050	.055	.050
	Asym	.054	.055	.055	.056	.055	.055	.056	.057	.061
(3, 6, 4, 2)	MLRT	.051	.051	.047	.048	.048	.051	.053	.050	.053
	Asym	.051	.052	.053	.053	.052	.054	.054	.057	.063
(8, 12, 7, 20)	MLRT	.052	.049	.051	.050	.051	.047	.054	.053	.051
	Asym	.053	.053	.054	.053	.053	.053	.053	.054	.055

$(n_1, \ldots, n_7)$	Methods	$k = 7$, $\tau_1 = \tau_2 = \cdots = \tau_3 = \tau_7$								
		.05	.10	.15	.20	.25	.30	.35	.4	.5
(2, 2, 2, 2, 2, 2, 2)	MLRT	.049	.047	.045	.050	.052	.045	.052	.046	.053
	Asym	.088	.087	.085	.086	.086	.087	.086	.089	.091
(5, 2, 5, 2, 3, 3, 2)	MLRT	.052	.050	.048	.047	.051	.050	.051	.050	.051
	Asym	.058	.059	.060	.062	.065	.067	.072	.080	.101
(4, 4, 12, 3, 2, 4, 6)	MLRT	.050	.050	.048	.050	.050	.050	.049	.049	.051
	Asym	.052	.053	.052	.054	.056	.060	.065	.072	.094
(5, 3, 2, 4, 2, 2, 10)	MLRT	.052	.052	.046	.049	.048	.053	.052	.050	.054
	Asym	.047	.048	.050	.051	.055	.059	.068	.076	.103
(3, 6, 4, 2, 20, 10, 7)	MLRT	.047	.049	.048	.047	.048	.046	.049	.052	.050
	Asym	.046	.047	.050	.049	.051	.056	.061	.068	.091
(2, 3, 2, 3, 2, 3, 10)	MLRT	.051	.052	.051	.048	.049	.050	.051	.051	.052
	Asym	.047	.049	.049	.052	.056	.061	.070	.080	.111

observe that the comparison results for the CsV based on the p values and the CIs are in agreement for all the pairs.

An R program that was used to compute the CIs for the ratio and the difference of CsV is posted at http://www.ucs.louisiana.edu/~kxk4695.

Example 2 This example involves comparison of the intra-assay variability of three types of assays of murine gamma interferon: the inhibition of viral cytopathic effect assay, inhibition of WEHI-279 assay, and ELISA assay. Each of these assays had five replications ($n_1 = n_2 = n_3 = 5$) at an interferon concentration of 7.5 U/ml. The data

Table 5 Powers of the MLRT and the Feltz-Miller's asymptotic test (in parentheses) for testing equality of several coefficients of variation

$H_0 : \lambda_1 = \cdots = \lambda_k$ versus $H_a : \lambda_i \neq \lambda_j$ for some $i \neq j$; $\alpha = .05$

(τ_1, τ_2, τ_3)	(n_1, n_2, n_3)				
	(5, 5, 5)	(5, 10, 5)	(5, 10, 10)	(10, 10, 10)	(10, 20, 20)
(.1, .1, .1)	.050 (.058)	.046 (.055)	.051 (.054)	.053 (.056)	.049 (.050)
(.1, .2, .2)	.181 (.189)	.210 (.153)	.248 (.138)	.454 (.448)	.561 (.434)
(.1, .4, .4)	.614 (.504)	.707 (.391)	.771 (.326)	.973 (.973)	.999 (.981)
(.1, .5, .4)	.716 (.572)	.822 (.442)	.863 (.402)	.994 (.989)	.999 (.994)
(.1, .05, .4)	.922 (.932)	.965 (.982)	.999 (.999)	.999 (.999)	1 (.999)
$(\tau_1, \tau_2, \tau_3, \tau_4)$	(n_1, n_2, n_3, n_4)				
	(5, 5, 5, 5)	(5, 10, 5, 5)	(5, 5, 10, 10)	(10, 10, 10, 10)	(10, 30, 50, 100)
(.1, .1, .1, .1)	.048 (.061)	.050 (.057)	.047 (.053)	.052 (.057)	.047 (.048)
(.1, .2, .2, .2)	.162 (.171)	.180 (.144)	.21 5 (.124)	.419 (.386)	.570 (.316)
(.1, .2, .3, .3)	.355 (.358)	.364 (.384)	.542 (.380)	.816 (.790)	.990 (.955)
(.1, .4, .4, .4)	.548 (.384)	.601 (.309)	.711 (.246)	.982 (.946)	.999 (.876)
(.1, .3, .4, .5)	.614 (.500)	.624 (.511)	.799 (.489)	.987 (.965)	1 (.997)
$(\tau_1, \tau_2, \tau_3, \tau_4)$	(n_1, n_2, n_3, n_4)				
	(15, 15, 15, 15)	(20, 20, 20, 20)	(30, 30, 30, 30)	(40, 20, 40, 20)	(40, 10, 40, 40)
(.33, .33, .33, .33)	.047 (.051)	.051 (.057)	.050 (.057)	.053 (.052)	.051 (.049)
(.33, .3, .25, .2)	.290 (.298)	.395 (.399)	.596 (.591)	.481 (.475)	.741 (.739)
(.2, .33, .33, .33)	.343 (.313)	.476 (.434)	.698 (.662)	.664 (.653)	.845 (.817)
(.33, .33, .20, .40)	.475 (.450)	.640 (.602)	.848 (.829)	.870 (.888)	.952 (.938)
(.2, .1, .1, .1)	.760 (.787)	.879 (.892)	.968 (.961)	.973 (.973)	.995 (.995)

Table 6 Summary statistics of normal blood samples

	1995							1996					
	HB	RBC	MCV	Hct	WBC	Plt		HB	RBC	MCV	Hct	WBC	Plt
n_1	65	65	63	64	65	64	n_2	73	73	72	72	73	71
$\bar{x}_1$	13.39	4.409	84.13	0.3717	8.466	247.1	$\bar{x}_2$	12.97	4.275	85.68	0.3658	7.818	227.31
v_1	0.2192	0.0826	3.390	0.0181	0.4655	17.91	v_2	0.2929	0.0863	2.946	0.0137	0.4798	21.61
$\widehat{\tau}_1$	0.0165	0.0189	0.0406	0.0490	0.0554	0.0730	$\widehat{\tau}_2$	0.0227	0.0203	0.0346	0.0378	0.0618	0.0957

for the three assays are reported in Table 8. Feltz and Miller (1993) used the above data and computed the asymptotic statistic as $A = 5.5304$ with $\widehat{\tau}_c = .2973$, and the critical value $\chi^2_{2;.95} = 5.991$. So the asymptotic test does not reject the hypothesis of equal coefficients of variation. The statistics that are required for finding the LRT and the MLRT are computed as follows. The constrained MLEs are $\widehat{\mu}_1^* = 6.3374$,

Table 7 P values for testing (9) and 95 % generalized CIs (L_r, U_r) for the ratio of CsV and 95 % CIs (L_d, U_d) for the difference of CsV

	p values of the tests							95 % CIs for the ratio τ_1/τ_2					
	HB	RBC	MCV	Hct	WBC	Plt		HB	RBC	MCV	Hct	WBC	Plt
GV	.0097	.5505	.1929	.0321	.3815	.0313	L_r	.571	.733	.921	1.024	.706	.597
Forkman	.0084	.5368	.2036	.0353	.3567	.0287	U_r	.924	1.182	1.501	1.663	1.142	.976
M-SLRT	.0098	.5481	.1933	.0332	.3761	.0312	L_d	−.011	−.006	−.003	.001	−.021	−.044
							U_d	−.002	.003	.016	.023	.008	−.002

Table 8 Assays of murine gamma interferon and test results for (16)

	ELISA	WEHI	Viral inhibition	Test results		
				Methods	Statistics	p value
	1	2	3	Asymptotic	5.5304	.063
Mean	6.8	8.5	6.0	MLRT	6.617	.037
CV	0.090	0.462	0.340			

$\widehat{\mu}_2^* = 9.1133$, $\widehat{\mu}_3^* = 6.0343$, and $\widehat{\tau}^* = .2929$. The LRT statistic is 9.145. To find the MLRT statistic, the PB estimate of $m(\Lambda) = 2.7748$ and the PB estimate of $\sqrt{v(\Lambda)} = 2.7596$. Using (18), we get

$$\text{MLRT statistic} = 2 \times \left(\frac{9.145 - 2.7748}{2.7596} \right) + 2 = 6.617.$$

Note that the MLRT statistic is greater than $\chi^2_{2;.95}$ (p value .037), and so the MLRT rejects the null hypothesis of equal coefficient of variation.

To apply the Verrill and Johnson (2007b) simulation-based approach, the 95th percentile of the LRT statistic at the constrained MLEs is estimated using simulation as 8.263. The LRT statistic (17) is 9.145. Since the LRT statistic is greater than the 95th percentile, the null hypothesis of equal coefficients of variation is rejected at the 5 % level. The p value is estimated by the probability that the LRT statistic is greater than 9.145 at the constrained MLEs. Using simulation of 10,000 runs, the p value is estimated as .037, which is the same as the one based on the MLRT.

An R program that was used to compute the MLRT statistic and the p value is posted at http://www.ucs.louisiana.edu/~kxk4695.

6 Concluding remarks

For comparing two CsV, we have developed generalized CIs for the ratio and the difference of two CsV. Alternatively, we can develop a CI for the difference between (or the ratio of) two CsV by inverting the SLRT or M-SLRT for $H_0 : \tau_1 - \tau_2 = d$ versus

$H_a : \tau_1 - \tau_2 \neq d$. In particular, the set of values of d for which the null hypotheses are not rejected form a CI for d. The CIs based on the SLRT should be liberal having coverage probability less than the nominal confidence level. The CIs based on the M-SLRT should be more accurate, however, finding them is numerically involved. On the basis of comparison of tests, we expect that the generalized CIs should be similar to those based on the M-SLRT. Note that the generalized CIs are conceptually simple, and are easier to obtain than those based on the M-SLRT.

For comparing several coefficients of variation, we used the idea of DiCiccio et al. (2001) to modify the usual LRT statistic. As observed in our simulation studies in Sect. 4.3, the chi-square approximation to the MLRT statistic seems to be very accurate even for very small samples, and the MLRT performs like an exact test in controlling type I error rates. Furthermore, the MLRT (or the small sample simulation-based approach by Verrill and Johnson 2007a,b) has some desirable power properties, and performs better than other approximate tests. As noted earlier, we have provided R codes to compute generalized CIs and to test the equality of several coefficients of variation at http://www.ucs.louisiana.edu/~kxk4695. We hope that the results of this article are useful to statisticians and practitioners of statistics in other areas of sciences.

Acknowledgments The authors are grateful to two reviewers for providing useful references and comments which improved the earlier version of the paper substantially.

References

Barndorff-Nielsen OE (1991) Modified signed log likelihood ratio. Biometrika 78:557–563

Bhoj D, Ahsanullah M (1993) Testing equality of coefficients of variation of two populations. Biometric J 35:355–359

DiCiccio TJ, Martin MA, Stern SE (2001) Simple and accurate one-sided inference from signed roots of likelihood ratios. Can J Stat 29:67–76

Feltz GJ, Miller GE (1996) An asymptotic test for the equality of coefficients of variation from k normal populations. Stat Med 15:647–658

Forkman J (2009) Estimator and tests for common coefficients of variation in normal distributions. Commun Stat Theory Methods 38:233–251

Fung WK, Tsang TS (1998) A simulation study comparing tests for the equality of coefficients of variation. Stat Med 17:2003–2014

Gerig TM, Sen AR (1980) MLE in two normal samples with equal but unknown population coefficients of variation. J Am Stat Assoc 75:704–708

Hamer AJ, Strachan JR, Black MM, Ibbotson C, Elson RA (1995) A new method of comparative bone strength measurement. J Med Eng Technol 19:1–5

Johnson NL, Welch BL (1940) Application of the nonecentral t-distribution. Biometrika 31:362–389

Krishnamoorthy K, Mathew T (2003) Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals. J Stat Plan Inference 115:103–121

Lohrding RK (1969) A test of equality of two normal population means assuming homogeneous coefficient of variation. Ann Math Stat 40:1374–1385

Miller EG, Karson MJ (1977) Testing the equality of two coefficients of variation. American Statistical Association: Proceedings of the Business and Economics Section, Part I, pp 278–283

Tian L (2005) Inferences on the common coefficient of variation. Stat Med 24:2213–2220

Tian L, Wu J (2007) Inferences on the common mean of several log-normal populations: the generalized variable approach. Biometr J 49:944–951

Vangel MG (1996) Confidence intervals for a normal coefficient of variation. Am Stat 50:21–26

Verrill, SP, Johnson RA (2007a) Confidence bounds and hypothesis tests for normal distribution coefficients of variation. United States Department of Agriculture, Forest Service, Research paper FPL-RP-638

Verrill S, Johnson RA (2007b) Confidence bounds and hypothesis tests for normal distribution coefficients of variation. Commun Stat Theory Methods 36:2187–2206

Weerahandi S (2004) Generalized inference in repeated measures: exact methods in MANOVA and mixed models. Wiley, Hoboken, NJ