Research Article Environmetrics

Received: 15 August 2013, Revised: 14 November 2013, Accepted: 2 January 2014, Published online in Wiley Online Library: 20 February 2014

(wileyonlinelibrary.com) DOI: 10.1002/env.2261

Small sample inference for gamma parameters: one-sample and two-sample problems

K. Krishnamoorthy[a]* and Luis León-Novelo[a]

Signed-likelihood ratio tests (SLRTs) are derived for testing the shape and scale parameters and the mean of a gamma distribution. The properties of the proposed tests are evaluated by Monte Carlo simulation and compared with the other tests available in the literature. SLRTs are also developed for two-sample problems of comparing two shape parameters, two scale parameters and two means, and their merits are evaluated by Monte Carlo simulation. Evaluation studies indicate that the SLRTs are accurate even for small samples and are comparable with or better than other tests. Furthermore, simple parametric bootstrap (PB) methods based on the maximum likelihood estimates are proposed for interval estimation. The PB confidence intervals are satisfactory even for small samples. The methods are illustrated and compared using two examples. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: constrained MLEs; hypothesis test; modified LRT; t-percentile bootstrap; power; third-order accurate; type I error rates

1. INTRODUCTION

Gamma distributions are widely used for the analysis of meteorological data, pollution data, and lifetime data. In meteorology, the gamma model has been used extensively to fit rainfall data on fairly large space and timescales, ranging from individual storms up to monthly and yearly distributions. Specifically, gamma distributions are used to model the amounts of daily rain fall in a region (Das, 1955; Stephenson *et al.*, 1999) and to fit hydrological data sets (Aksoy, 2000). Ashkar and Ouarda (1998) used a two-parameter gamma distribution to fit annual maximum flood series data in order to construct confidence intervals (CIs) for a quantile. Schickedanz and Krause (1970) used rainfall data for different seasons to compare the scale parameters of two gamma distributions. Husak *et al.* (2006) noted that among probability distributions that could be successfully utilized to parameterize rainfall distributions, the gamma distribution is one of the most widely understood, making it a good choice for modeling rainfall distributions. In exposure/pollution data analysis, gamma models are used as alternatives to lognormal models. Maxim *et al.* (2006) have observed that the gamma distribution is a possible distribution for concentrations of carbon/coke fibers in plants that produce green or calcined petroleum coke. Gibbons (1994), Bhaumik and Gibbons (2006), Krishnamoorthy, Mathew and Mukherjee (2008) and Bhaumik, Kapur and Gibbons (2009) noted that gamma distributions are potentially useful for applications in many fields, including environmental monitoring, groundwater monitoring, industrial hygiene, genetic research, and industrial quality control. In some of the aforementioned applications, sample sizes are typically small. For example, Bhaumik and Gibbons (2006) have pointed out that assessing environmental impact on the basis of a small number of samples obtained from an area of concern is a common problem in environmental monitoring. These authors noted that the distribution of an analyte of concern in environmental monitoring problem is typically non-normal and illustrated the relevance of the gamma distribution to environmental data. In industrial exposure assessment, the sample sizes are often small because of the cost of sampling and burden on workers. If the sample sizes are moderate to large, nonparametric methods can be used to avoid distributional misspecification problem; however, such nonparametric methods are usually less powerful than their parametric counterparts. It should also be noted that for very small sample sizes, such as four, nonparametric methods may not yield any meaningful results. So the purpose of this article is to provide accurate small sample inference procedures for one-sample and two-sample problems involving gamma distributions.

The probability density function of a gamma distribution with the shape parameter a and the scale parameter b is given by

$$f(x|a,b) = \frac{1}{\Gamma(a)b^a} e^{-x/b} x^{a-1}, \; x > 0, a > 0, b > 0 \tag{1}$$

* Correspondence to: K. Krishnamoorthy, Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70508, U.S.A. E-mail: krishna@louisiana.edu

a Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70508, U.S.A.

Krishnamoorthy *et al.* (2008) have shown that approximate solutions for some problems involving gamma distributions can be readily obtained using cube root transformation (Wilson and Hilferty, 1931) and normal-based methods. Specifically, tolerance intervals, CIs for a survival probability, prediction intervals, and CIs for the stress-strength reliability involving gamma distributions can be easily obtained using cube root transformation. However, inferential procedures for the gamma parameters are not simple to obtain because the parameters are not of the more convenient location-scale form.

Let $X_1, \ldots, X_n$ be a sample from a gamma(a, b) distribution. The arithmetic mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and the geometric mean $\widetilde{G} = \left(\prod_{i=1}^{n} X_i\right)^{1/n}$ are jointly sufficient statistics for a and b. Bain and Engelhardt (1975) have derived an approximate test for a based on an approximation to the distribution of statistic $\bar{X}/\widetilde{G}$. An exact uniformly most powerful unbiased test for b based on the conditional distribution of $\bar{X}$ given $\widetilde{G}$ exists, but practical implementation requires percentage points, which is extremely complicated; see Engelhardt and Bain (1977). The mean of the gamma distribution, $\mu = ab$, is a function of both parameters. No exact method for finding confidence limits for μ when both a and b are unknown is available. Several authors have provided approximate solutions for estimating the parameters and the means in one-sample and two-sample problems; see Grice and Bain (1980), Shiue *et al.* (1988), Shiue and Bain (1990), Keating *et al.* (1990), Tripathi *et al.* (1993), Wong (1992, 1993), Bhaumik and Gibbons (2006), and Bhaumik *et al.* (2009) and the references therein.

Even though several approximate methods have been proposed in the literature, only a few papers were considered likelihood-based methods. Obviously, the standard likelihood method based on the Fisher information matrix can be easily obtained, but such methods are satisfactory only for large samples. A few modifications to the likelihood ratio test (LRT) statistic in a general setup are proposed in the literature so that the modified LRT statistic has the standard normal distribution with an error of $O(n^{-3/2})$. Fraser *et al.* (1997) have applied a version of modified LRT by Fraser and Reid (1995) to find a test for the mean of a gamma distribution. Their numerical study indicated that the modified LRT is accurate even for samples of size three. There are a few higher order versions of the signed-likelihood ratio test (SLRT) statistic, which differ in terms of ease of implementation. For example, see Barndorff-Nielsen (1991), Skovgaard (1996), and DiCiccio *et al.* (2001). In particular, DiCiccio *et al.* (2001) have proposed a simulation based methods to improve upon the accuracy of the normal approximation of the SLRT statistic. In the present paper, we shall apply these simulation based methods for various one-sample and two-sample problems involving gamma models.

The rest of the article is organized as follows. In the following section, we provide some preliminary results on finding the maximum likelihood estimates (MLEs), the Fisher information matrix, and an improved version of the SLRT by DiCiccio *et al.* (2001). In Section 3, we describe the signed-likelihood ratio tests (SLRTs) for one-sample problems of testing the shape and the scale parameters, and the mean, and some other available tests. We also evaluate the tests in terms of type I error rates and powers. In Section 4, we address the two-sample problems and describe the SLRTs. We also provide simple parametric bootstrap (PB) methods for finding CIs in one-sample and two-sample problems in Section 5. All the methods are evaluated by Monte Carlo (MC) simulation. In Section 6, we made some recommendations as to the choice of the methods for applications. Two practical examples are used to illustrate various tests in Section 7. Some concluding remarks are given in Section 8.

2. PRELIMINARIES

Let $X_1, \ldots, X_n$ be a sample from a gamma(a, b) distribution. Let $\bar{X}$ and $\widetilde{G}$ denote, respectively, the arithmetic mean and geometric mean of the sample. The log-likelihood function is expressed as

$$l\left(a, b|\bar{X}, \widetilde{G}\right) = -n\ln\Gamma(a) - na\ln b - n\bar{X}/b + (a-1)n\ln\widetilde{G} \tag{2}$$

The MLE $\widehat{a}$ is the solution of the equation

$$\ln(a) - \psi(a) = \ln\left(\bar{X}/\widetilde{G}\right) \tag{3}$$

where ψ is the digamma function. Letting $s = \ln(\bar{X}/\widetilde{G})$, an approximation to $\widehat{a}$ is given by

$$\widehat{a} \simeq \frac{3 - s + \sqrt{(s-3)^2 + 24s}}{12s} \tag{4}$$

Using the aforementioned approximate with MLE as the initial value a_0, the MLE can be evaluated by the Newton–Raphson iterative scheme

$$a_1 = a_0 - \frac{\ln a_0 - \psi(a_0) - s}{1/a_0 - \psi'(a_0)}$$

where $\psi'(x) = \frac{\partial\psi(x)}{\partial x}$ is the trigamma function. The MLE of b is $\widehat{b} = \bar{X}/\widehat{a}$. Note that the MLE $\widehat{a}$ is implicitly a function of $\bar{X}/\widetilde{G}$, and so it is invariant under a scale transformation of the samples, and $\widehat{b} = \bar{X}/\widehat{a}$ is scale equivariant.

Approximate variance estimates of the MLEs are usually obtained from the estimated inverse Fisher information matrix given by

$$\begin{pmatrix} \widehat{\text{var}}(\widehat{a}) & \widehat{\text{cov}}(\widehat{a},\widehat{b}) \\ \widehat{\text{cov}}(\widehat{a},\widehat{b}) & \widehat{\text{var}}(\widehat{b}) \end{pmatrix} = \frac{1}{n(\widehat{a}\psi'(\widehat{a}) - 1)} \begin{pmatrix} \widehat{a} & -\widehat{b} \\ -\widehat{b} & \widehat{b}^2\psi'(\widehat{a}) \end{pmatrix} \tag{5}$$

The usual Wald tests for the parameters a and b are based on the asymptotic variance estimates in (5), which are valid only for large samples. In order to describe the improved versions of the LRT by DiCiccio *et al.* (2001), let $(\xi, ,\theta')'$ be a vector parameter in a statistical model, where ξ is a real-valued function of the parameters of interest. Furthermore, let $l(\xi, \boldsymbol{\theta})$ denote the log-likelihood function for $(\xi, \boldsymbol{\theta}')'$ based on observed data. We also denote the MLE of $(\xi, \boldsymbol{\theta}')'$ by $(\widehat{\xi}, \widehat{\boldsymbol{\theta}}')'$. Furthermore, for a fixed ξ, the MLE of the nuisance parameter $\boldsymbol{\theta}$ will be denoted by $\widehat{\boldsymbol{\theta}}_\xi$. For inference concerning ξ, the SLRT statistic, denoted by $R(\xi)$, is given by

$$R(\xi) = \text{sign}\left(\widehat{\xi} - \xi\right)\left[2\left\{l\left(\widehat{\xi}, \widehat{\boldsymbol{\theta}}\right) - l\left(\xi, \widehat{\boldsymbol{\theta}}_\xi\right)\right\}\right]^{1/2}$$

where sign(x) is defined as $+1$ if $x > 0$ and -1 if $x < 0$. In general, it is known that $R(\xi)$ follows a standard normal distribution up to an error of $O(n^{-1/2})$.

Two third-order accurate methods are proposed in DiCiccio *et al.* (2001). For a fixed value of ξ, let $m(\xi, \widehat{\boldsymbol{\theta}}_\xi)$ denote the mean of $R(\xi)$ and $SD(\xi, \widehat{\boldsymbol{\theta}}_\xi)$ denote the standard deviation of $R(\xi)$, both evaluated at $(\xi, \widehat{\boldsymbol{\theta}}_\xi)$. That is, $m(\xi, \widehat{\boldsymbol{\theta}}_\xi) = E_{\xi, \widehat{\boldsymbol{\theta}}_\xi}(R(\xi))$ and $SD(\xi, \widehat{\boldsymbol{\theta}}_\xi) = \sqrt{E_{\xi, \widehat{\boldsymbol{\theta}}_\xi}(R(\xi) - m(\xi, \widehat{\boldsymbol{\theta}}_\xi))^2}$. The standardized SLRT (SSLRT) statistic is $\frac{R(\xi) - m(\xi, \hat{\boldsymbol{\theta}}_\xi)}{SD(\xi, \widehat{\boldsymbol{\theta}}_\xi)}$. DiCiccio *et al.* (2001) have shown that SSLRT statistic has a standard normal distribution up to an error of $O(n^{-3/2})$. Another third-order accurate method consists of computing the percentiles $R(\xi)$ or the p-value of an observed statistic $R_0(\xi)$ with respect to the distribution of $\widehat{\boldsymbol{\theta}}_\xi$ while ξ is fixed. As pointed out in DiCiccio *et al.* (2001), analytic expressions are difficult to obtain for the mean, SD, or the percentiles. However, these quantities can be easily approximated by MC simulation of the SLRT statistic $R(\xi)$ when $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_\xi$, for a fixed ξ. Despite the asymptotic equivalence of these two methods, DiCiccio *et al*. noted that the SSLRT fails to take skewness properly into account, while the method on the basis of the p-value is able to account for skewness to some extent. Furthermore, based on applications of these two methods to some examples, DiCiccio *et al*. found that the method based on the p-value is often more accurate than the SSLRT. Even though our investigation indicated that these two methods produced similar results for various problems involving gamma distributions, in order to save space, we shall illustrate only the method based on the MC estimate of the p-value in the sequel.

3. ONE-SAMPLE TESTS

In the following sections, we shall describe some tests and CIs for the parameters a and b, and the mean $\mu = ab$ based on the sample mean $\bar{X}$ and the geometric mean $\widetilde{G}$.

3.1. Tests for the shape parameter

3.1.1. *The SLRT for a*

Consider testing

$$H_0 : a \leqslant a_0 \text{ vs. } H_a : a > a_0 \tag{6}$$

where a_0 is a specified value. It is easy to verify that, for a fixed a, the MLE of b is given by $\widehat{b}_a = \bar{X}/a$. The SLRT statistic is expressed as

$$\begin{aligned} R(a_0) &= \text{sign}(\widehat{a} - a_0)\left\{2\left[\ln l(\widehat{a}, \widehat{b}) - \ln l(a_0, \widehat{b}_{a_0})\right]\right\}^{1/2} \\ &= \text{sign}(\widehat{a} - a_0)\sqrt{2n}\left[\ln\frac{\Gamma(a_0)}{\Gamma(\widehat{a})} + (\widehat{a} - a_0)[\ln(\widetilde{G}/\bar{X}) - 1] + (\widehat{a}\ln\widehat{a} - a_0\ln a_0)\right]^{1/2} \end{aligned} \tag{7}$$

where $\widehat{a}$ and $\widehat{b}$ are the MLEs.

As the testing problem is invariant under scale transformation $X \to cX$, where c is a positive constant, the distribution of $R(a_0)$ depends only on a_0. Therefore, the null distribution can be evaluated empirically by MC simulation. In particular, for an observed value of SLRT statistic $R_0(a_0)$, the null hypothesis in (6) is rejected if the p-value $P(R(a_0) > R_0(a_0)) < \alpha$, where $0 < \alpha < 0.5$ is a specified level of significance. Note that this p-value can be estimated using simulated samples from a gamma$(a_0, 1)$ distribution. So this test is exact, except for the simulation error.

3.1.2. *The Bhaumik-Kapur-Gibbons (BKG) test*

For testing hypotheses in (6), Bhaumik, Kapur and Gibbons (2009) proposed the following test statistic $T_1 = 2na_0s$, where $s = \ln(\bar{X}/\widetilde{G})$. The null distribution is approximated by a constant times the chi-square distribution, $c\chi^2_\nu$, where c and ν are determined by the method of moments. Specifically, c and ν are determined by the $2na_0E(s) = c\nu$ and $(2na_0)^2\text{var}(s) = 2c^2\nu$. The mean and variance of s are estimated using simulated samples from the gamma$(a_0, 1)$ distribution. The null hypothesis in (6) is rejected if $T_1 < c\chi^2_{\nu;\alpha}$. We shall refer to the test as the BKG test.

3.2. Tests for the scale parameter

Consider testing

$$H_0 : b \leqslant b_0 \text{ vs. } H_a : b > b_0 \tag{8}$$

where b_0 is a specified value. In the following, we shall describe the SLRT and the BKG test for the scale parameter.

3.2.1. The SLRT for b

For a fixed b, the MLE $\widehat{a}_b$ of a is the solution of the equation

$$\psi(a) - \ln \frac{\widetilde{G}}{b} = 0 \tag{9}$$

Let $d = \ln\left(\widetilde{G}/b\right)$. Noting that $\ln(x-1) \leqslant \psi(x) \leqslant \ln(x)$, we see that $\ln(a-1) - d \leqslant \psi(a) - d \leqslant \ln(a) - d$. So the root of the equation (9) lies in the interval $(e^d, e^d + 1) = \left(\widetilde{G}/b, \widetilde{G}/b + 1\right)$, and a root bracketing method such as the usual bisection method can be used to find the root. An enhanced version of the bisection method called Illinois method (Thisted, 1988) converges in a fewer steps and is about three times faster than the usual bisection method. The SLRT statistic for testing $b = b_0$ is expressed as

$$\begin{aligned} R(b_0) &= \text{sign}\left(\widehat{b} - b_0\right)\left\{2\left[\ln l(\widehat{a}, \widehat{b}|\bar{X}, \widetilde{G}) - \ln l\left(\widehat{a}_{b_0}, b_0|\bar{X}, \widetilde{G}\right)\right]\right\}^{1/2} \\ &= \text{sign}\left(\widehat{b} - b_0\right)\sqrt{2n}\left\{\ln \frac{\Gamma(\widehat{a}_{b_0})}{\Gamma(\widehat{a})} + \widehat{a}\left(\ln \frac{\widetilde{G}}{\widehat{b}} - 1\right) - \widehat{a}_{b_0} \ln \frac{\widetilde{G}}{b_0} + \frac{\bar{X}}{b_0}\right\}^{1/2} \end{aligned} \tag{10}$$

As the SLRT statistic $R(b_0)$ is invariant under scale transformation, its distribution does not depend on the scale parameter, but it may depend on the shape parameter. However, on the basis of our extensive simulation study, we observed that the distribution of the SLRT statistic $R(b_0)$ does not depend on the shape parameter, and it depends only the sample size. To show some evidence, we estimated quantiles of $R(b_0)$ based on 1,000,000 simulated samples each of size 4 from a gamma(1,1), gamma(11,1), and gamma(100,1) distributions and plotted them in Figure 1. The first plot represents the quantiles for gamma(1, 1) and gamma(11, 1) distributions, and the second plot represents quantiles for gamma(1, 1) and gamma(100, 1) distributions. These two plots clearly indicate that the distributions of SLRT statistic do not depend on the shape parameter. We have striking simulation evidence to indicate that the distribution of the SLRT statistic does not depend on any parameters. As the standard invariance argument is not applicable to gamma distribution, we are unable to prove theoretically that the null distribution is free of a.

In view of our findings in the preceding paragraph, the p-value of the SLRT for testing (8) is given by $P(R^*(1) > R_0(b_0)|a = 1, b = 1)$, where $R_0(b_0)$ is an observed value of $R(b_0)$ and $R^*(1)$ is the test statistic in (10) based on a random sample of size n from a gamma(1,1) distribution. Note that, for an observed SLRT statistic $R_0(b_0)$, this p-value can be estimated by MC simulation.

3.2.2. The BKG test for the scale parameter

Bhaumik *et al.* (2009) proposed an approximate test as follows. These authors first noted that $Z = 2n\bar{X}/b_0$ follows a $\chi^2_{2n\widehat{a}}$ distribution approximately. The test statistic is defined as $T_2 = 2\widehat{a}ns/c + Z$, where $s = \ln(\bar{X}/\widetilde{G})$, $2\widehat{a}ns/c \sim \chi^2_\nu$ approximately, c and ν are determined by $2n\widehat{a}E(s) = c\nu$ and $(2n\widehat{a})^2\text{var}(s) = 2c^2\nu$. The $E(s)$ and var(s) are estimated by MC method based on samples generated from gamma$(\widehat{a}, b_0)$. For an observed value T_2^0 of T_2, the BKG test rejects the null hypothesis in (8) whenever $P\left(\chi^2_{\nu+2n\widehat{a}} > T_2^0\right) < \alpha$.

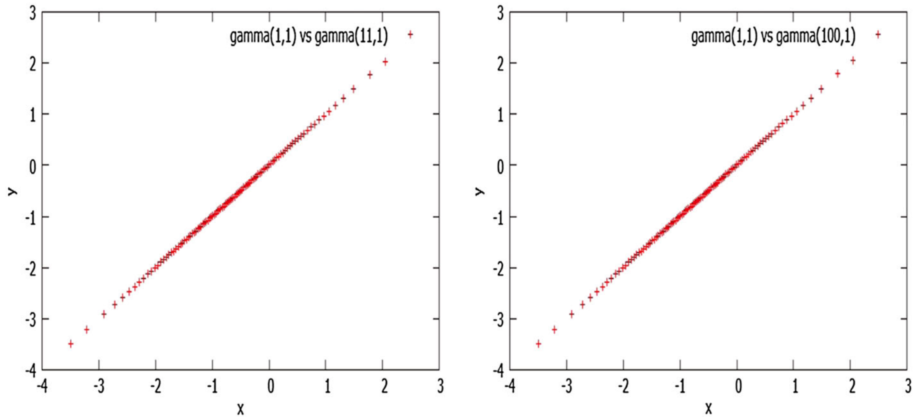

Figure 1. Quantiles of the signed-likelihood ratio test (SLRT) statistic $R(b_0)$ based on samples each of size 4 from gamma$(1, 1)$, gamma(1,11), and gamma(100,1) distributions

Instead of the chi-square approximation, Bhaumik *et al*. also have proposed a PB approach to find the percentiles of T_2. Our simulation studies (not reported here) indicated that the test based on the chi-square approximation and the one based on the PB approach are similar with respect to type I error rates, and so we shall not consider the PB approach here.

3.3. Tests for the mean

The SLRT statistic is given by

$$R(\mu_0) = \text{sign}(\widehat{\mu} - \mu_0)\left\{2\left[l\left(\widehat{a}, \widehat{b}|\bar{X}, \widetilde{G}\right) - l\left(\widehat{a}_{\mu_0}, \mu_0|\bar{X}, \widetilde{G}\right)\right]\right\}^{1/2} \tag{11}$$

where $l(a, b|\bar{X}, \widetilde{G})$ is the log-likelihood function in (2),

$$l\left(a, \mu|\bar{X}, \widetilde{G}\right) = -n\ln\Gamma(a) - na\ln(a/\mu) - na\bar{X}/\mu + (a-1)n\ln\widetilde{G}$$

and $\widehat{a}_{\mu_0}$ is the MLE of a at $\mu = \mu_0$. This constrained MLE $\widehat{a}_{\mu_0}$ is obtained as the root of the equation

$$\ln a - \psi(a) = \ln\frac{\mu_0}{\widetilde{G}} + \frac{\bar{X}}{\mu_0} - 1 \tag{12}$$

As the aforementioned equation is similar to (3), a Newton–Raphson iterative scheme is readily obtained.

3.3.1. *The modified LRT for the mean*

The modified LRT by Fraser *et al.* (1997) is given by

$$MLRT(\mu_0) = R(\mu_0) - \frac{1}{R(\mu_0)}\ln\left(\frac{R(\mu_0)}{Q}\right) \tag{13}$$

where $R(\mu_0)$ is defined in (11), and $Q = \sqrt{n\widehat{a}}(\widehat{\mu}/\mu_0 - 1)(\psi'(\widehat{a}) - 1/\widehat{a})^{\frac{1}{2}}/(\psi'(\widehat{a}_{\mu_0}) - 1/\widehat{a}_{\mu_0})^{\frac{1}{2}}$, and $\psi'(x)$ is the trigamma function. This MLRT has third-order accuracy in the sense that the standard normal approximation to the distribution of MLRT(μ_0) is accurate up to $O(n^{-3/2})$. For testing

$$H_0 : \mu \leqslant \mu_0 \ \text{ vs. } \ H_a : \mu > \mu_0 \tag{14}$$

the MLRT rejects the null hypothesis if $MLRT(\mu_0) > z_{1-\alpha}$. For a two-sided alternative hypothesis, the MLRT rejects the null hypothesis if $|MLRT(\mu_0)| > z_{1-\alpha/2}$.

3.3.2. *The SLRT for the mean*

For a given level of significance, and an observed value $R_0(\mu_0)$, the SLRT rejects the null hypothesis in (14) if

$$P_{\mu_0, \widehat{a}_{\mu_0}}(R(\mu_0) \geqslant R_0(\mu_0)) < \alpha \tag{15}$$

The aforementioned probability can be estimated by MC simulation based on samples generated from gamma($\widehat{a}_{\mu_0}, \mu_0/\widehat{a}_{\mu_0}$) distribution.

3.3.3. *The BKG test for the mean*

Bhaumik *et al.* (2009) have proposed a few approximate tests for the mean $\mu = ab$. These authors recommended two comparable tests, and we shall consider the one in Equation (7) of their paper. Let μ_0 be a specified value of μ under H_0, $\gamma_0 = (n\mu_0)^{1/3}$, and let $\widehat{\gamma} = (n\bar{X})^{1/3}$. Bhaumik *et al*. (2011) showed that

$$T_3(\mu_0) = \frac{9\gamma_0(n-1)(\widehat{\gamma} - \gamma_0)^2}{2n\mu_0\ln(\bar{X}/\widetilde{G})} \sim F_{1,n-1}, \text{ approximately} \tag{16}$$

For testing

$$H_0 : \mu = \mu_0 \text{ vs. } H_a : \mu \neq \mu_0 \tag{17}$$

the null hypothesis is rejected if $T_3 > F_{1,n-1;1-\alpha}$. It should be noted that the aforementioned test is not applicable for testing one-sided hypotheses. The BKG test for the mean is easy to apply as the test statistic involves only the arithmetic mean and geometric mean of the sample. Furthermore, the two roots of the equation $T_3(\mu_0) = F_{1,n-1;1-\alpha}$ form a $100(1-\alpha)\%$ CI for μ.

3.4. Type I error rates and power studies

To judge the error rates of the one-sample tests in the preceding sections, and to compare them with other available tests, we estimated the type I error rates and power of the tests using MC simulation. To estimate the type I error rates of the tests, we choose $b = 1$ without loss of generality. The estimated type I error rates for the tests on the shape parameter are reported in Table 1 for various values a and sample sizes ranging from 3 to 15. All MC estimates are based on 100,000 runs. The estimated type I error rates of the SLRT are very close to the nominal level 0.05. The BKG test could be anti-conservative when both sample size and a are very small; otherwise, it also performs satisfactorily. The powers of the tests are practically the same in situations where the BKG test controls the type I error rates satisfactorily; see Figure 2. Both tests involve simulation, but the SLRT is exact and performs better than the BKG test in some cases, and so the SLRT is recommended for applications.

MC estimates of the type I error rates of the SLRT and those of the BKG test on the scale parameter b are reported in Table 2 for various values of a. We observe that the SLRT controls the type I error rates very satisfactorily regardless of values of a and n. The BKG test is not satisfactory, and it could be too conservative for small a and liberal for large a. In general, the BKG test for the scale parameter is not satisfactory in controlling type I error rates. We also plotted the powers of the SLRT for testing $H_0 : b \leqslant 1$ vs. $H_a : b > 1$ when sample

Table 1. MC estimates of type I error rates of the SLRT and (BKG) tests for the shape parameter

a	$n = 3$	$n = 5$	$n = 8$	$n = 10$	$n = 15$
0.25	0.050(0.070)	0.050(.059)	0.050(0.056)	0.052(0.055)	0.051(0.052)
0.50	0.050(0.062)	0.051(.058)	0.051(0.056)	0.050(0.053)	0.051(0.053)
0.75	0.051(0.057)	0.050(.055)	0.050(0.052)	0.049(0.052)	0.049(0.050)
1	0.050(0.054)	0.050(.052)	0.049(0.052)	0.049(0.051)	0.050(0.051)
1.5	0.050(0.051)	0.049(.051)	0.050(0.052)	0.051(0.051)	0.049(0.051)
2	0.051(0.050)	0.051(.052)	0.049(0.050)	0.051(0.051)	0.051(0.051)
3	0.051(0.050)	0.051(.051)	0.050(0.049)	0.049(0.050)	0.051(0.050)
4	0.050(0.053)	0.051(.051)	0.051(0.051)	0.052(0.051)	0.049(0.050)

MC, Monte Carlo; SLRT, signed-likelihood ratio test.

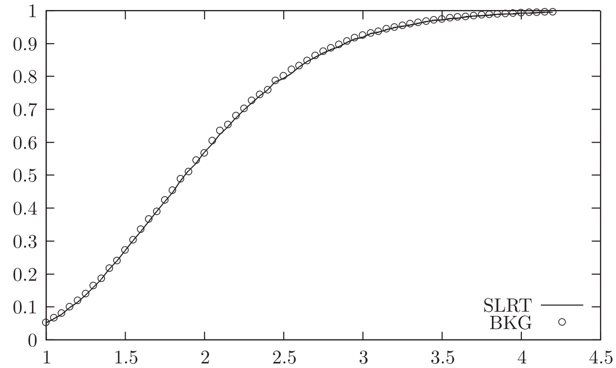

Figure 2. Powers of the signed-likelihood ratio test (SLRT) and BKG's test for testing $H_0 : a \leqslant 1$ vs. $H_a : a > 1$ at the level 0.05 and $n = 10$

Table 2. MC estimates of type I error rates of the SLRT and (BKG) tests for the scale parameter

a	$n = 5$	$n = 10$	$n = 15$	$n = 20$	$n = 30$
0.50	0.050(0.019)	0.049(0.021)	0.052(0.035)	0.051(0.024)	0.052(0.030)
0.75	0.050(0.036)	0.050(0.033)	0.051(0.045)	0.049(0.056)	0.049(0.051)
1	0.050(0.055)	0.051(0.068)	0.049(0.068)	0.049(0.070)	0.051(0.060)
1.5	0.049(0.060)	0.050(0.069)	0.050(0.106)	0.050(0.091)	0.051(0.113)
2	0.052(0.082)	0.050(0.106)	0.050(0.114)	0.051(0.103)	0.048(0.136)
3	0.050(0.104)	0.051(0.126)	0.051(0.160)	0.050(0.169)	0.049(0.168)
4	0.050(0.106)	0.048(0.159)	0.051(0.182)	0.049(0.193)	0.049(0.226)
10	0.049(0.191)	0.049(0.228)	0.051(0.240)	0.050(0.260)	0.050(0.253)

MC, Monte Carlo; SLRT, signed-likelihood ratio test.

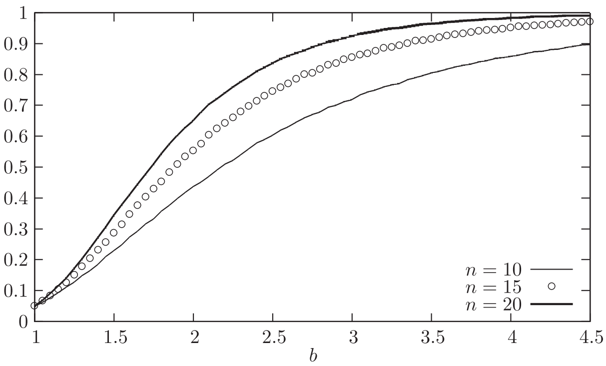

Figure 3. Power of the signed-likelihood ratio test (SLRT) for testing $H_0 : b \leqslant 1$ vs. $H_a : b > 1$ for different sample sizes

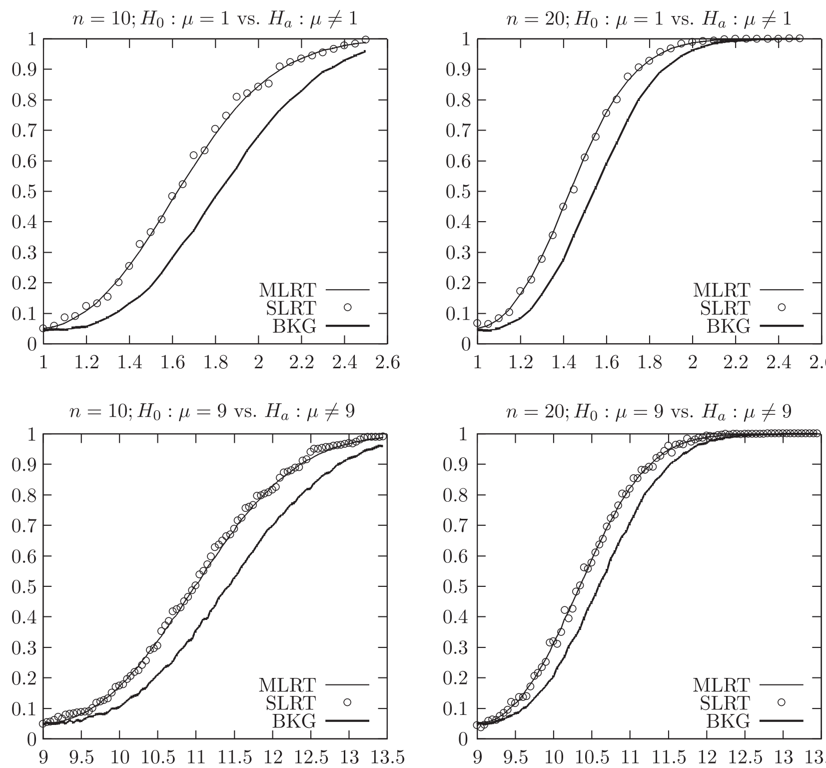

Figure 4. Power of the tests for testing $H_0 : \mu = \mu_0$ vs. $H_a : \mu \neq \mu_0$ for different sample sizes

sizes $n = 10, 15$, and 20 in Figure 3. The powers were plotted as a function b in the interval (1, 4.5). This plot clearly indicates that the power of the SLRT is increasing with increasing b and also increasing with increasing sample sizes. Thus, the SLRT has all natural properties of an efficient test. Power study for the BKG test was not carried out because this test does not control the type I error rates satisfactorily.

MC estimates of powers of the SLRT, MLRT, and the test by BKG on the mean are plotted in Figure 4 for sample sizes $n = 10, 15$, and 20. The powers were estimated as a function of μ in the interval (1, 2.5) for testing $H_0 : \mu = 1$ vs. $H_a : \mu \neq 1$. The four plots in Figure 4 clearly indicate that all three tests are controlling type I error rates very close to the nominal level 0.05. Regarding powers, the powers of the MLRT and the SLRT are practically the same, and they are greater than those of the BKG test. The powers were also estimated for testing some other values of μ_0, and they are not reported here because the power comparisons were very similar to the one for testing $\mu = 1$ and $\mu = 9$. Among these three tests for the mean, the MLRT is recommended because it does not involve simulation.

4. TWO-SAMPLE TESTS

Let $\bar{X}_i$ and $\widetilde{G}_i$ denote the mean and geometric mean of a sample of size n_i from a gamma(a_i, b_i) distribution, $i = 1, 2$. In the following, we shall describe SLRTs based on $(\bar{X}_1, \widetilde{G}_1, \bar{X}_2, \widetilde{G}_2)$ for two-sample problems.

4.1. Test for $a_1 - a_2$

Consider testing the hypotheses

$$H_0 : a_1 = a_2 \text{ vs. } H_a : a_1 \neq a_2 \tag{18}$$

4.1.1. *SLRT for* $a_1 - a_2$

The joint log-likelihood function based on samples from gamma(a_1, b_1) and gamma(a_2, b_2) distributions can be written as

$$l(a_1, b_1, a_2, b_2) = \ln\left(a_1, b_1 | \bar{X}_1, \widetilde{G}_1\right) + \ln\left(a_2, b_2 | \bar{X}_2, \widetilde{G}_2\right) \tag{19}$$

where $l(a, b|\bar{X}, \widetilde{G})$ is as defined in (2). The log-likelihood function under the constraint $a_1 = a_2 = a$ is expressed as

$$\begin{aligned} l(a, b_1, b_2) = & -(n_1 + n_2) \ln \Gamma(a) - a(n_1 \ln b_1 + n_2 \ln b_2) \\ & + (a-1)\left(n_1 \ln \widetilde{G}_1 + n_2 \ln \widetilde{G}_2\right) - \frac{n_1 \bar{X}_1}{b_1} - \frac{n_2 \bar{X}_2}{b_2} \end{aligned}$$

Equating $\partial l(a, b_1, b_2)/\partial b_i$ to 0, and solving for b_i, we find $b_1 = \bar{X}_1/a$ and $b_2 = \bar{X}_2/a$. The partial differential equation $\partial l\left(a, \frac{\bar{X}_1}{a}, \frac{\bar{X}_2}{a}\right) / \partial a = 0$ simplifies to

$$\ln(a) - \psi(a) = w_1 \ln \frac{\bar{X}_1}{\widetilde{G}_1} + w_2 \ln \frac{\bar{X}_2}{\widetilde{G}_2} \tag{20}$$

where $w_1 = n_1/(n_1 + n_2)$ and $w_2 = 1 - w_1$. Letting $s = w_1 \ln \frac{\widetilde{G}_1}{\bar{X}_1} + w_2 \ln \frac{\widetilde{G}_2}{\bar{X}_2}$, the procedure for the one-sample case can be used to find the MLE of the unknown common a. Denoting the constrained MLE satisfying (20) by $\widehat{a}_c$, we obtain the constrained MLE of b_i as $\widehat{b}_{ic} = \bar{X}_i/\widehat{a}_c$. The SLRT statistic for testing $a_1 = a_2$ is expressed as

$$R(d_a) = \text{sign}(\widehat{a}_1 - \widehat{a}_2) \left\{ 2\left[l(\widehat{a}_1, \widehat{b}_1, \widehat{a}_2, \widehat{b}_2) - l\left(\widehat{a}_c, \frac{\bar{X}_1}{\widehat{a}_c}, \frac{\bar{X}_2}{\widehat{a}_c}\right)\right]\right\}^{1/2} \tag{21}$$

Note that we used the notation $R(d_a)$ to denote the LRT statistic for testing the difference $a_1 - a_2$. For this testing problem also, we observed that the null distribution of the SLRT statistic $R(d_a)$ does not depend on any unknown parameters, and it depends only on sample sizes. To see this, we plotted estimated quantiles of $R(d_a)$ for samples of size $n_1 = 4$ from a gamma(1, 1) distribution and samples of size $n_2 = 5$ from a gamma(1,5) distributions along with quantiles of $R(d_a)$ for samples from gamma(5, 1) and gamma(5,20) in plot (a) of Figure 5. Plot (b) represents quantiles of $R(d_a)$ based on samples from (gamma(1,1), gamma(1,5)) and those based on (gamma(20,1), gamma(20,50)) distributions. These two plots clearly indicate that the null distribution $R(d_a)$ does not depend on any unknown parameters, and so the p-value $P(R^*(d_a) > R_0(d_a))$, where $R_0(d_a)$ is an observed value of $R(d_a)$ in (21) and $R^*(d_a)$ is the same based on independent samples of sizes n_1 and n_2 from a gamma(1,1) distribution. For a given $R_0(d_a)$, this p-value can be estimated by MC simulation.

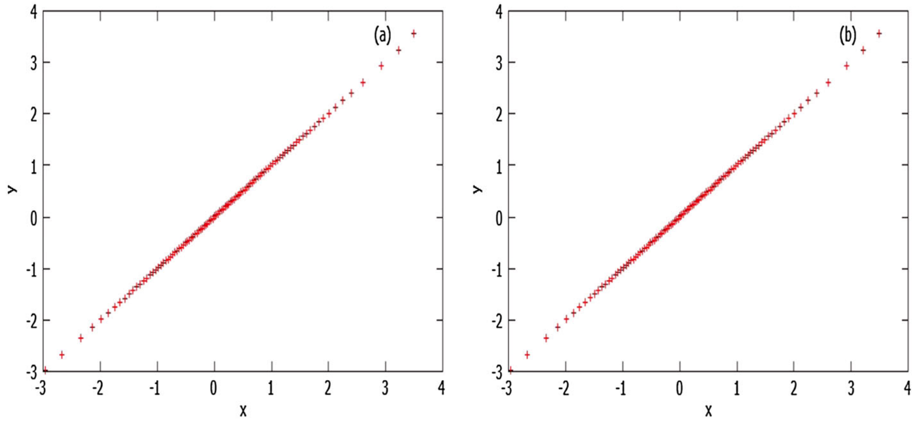

Figure 5. Quantiles of the signed-likelihood ratio test (SLRT) statistic $R(d_a)$ based on samples from different gamma distributions: (a) (gamma(1,1),gamma(1,5)) vs. (gamma(5,1),gamma(5,20)); (b) (gamma(1,1),gamma(1,5)) vs. (gamma(20,1),gamma(20,50)); $n_1 = 4, n_2 = 5$

4.1.2. The Shiue–Bain–Engelhardt (SBE) Test

The approximate test by Shiue *et al.* (1988) is described as follows. Let $s_i = \ln(\bar{X}_i/\widetilde{G}_i)$, and $\nu_i = (n_1 - 1)\left(1 + 1/(1 + 4.3\widehat{a}_i)^2\right)$, $i = 1, 2$. Then, SBE test rejects the null hypothesis of equal shape parameter at the level of significance α, if

$$\frac{n_1 s_1/(n_1-1)}{n_2 s_2/(n_2-1)} > F_{\nu_1,\nu_2;1-\alpha} \tag{22}$$

where $F_{m,n;p}$ denotes the $100p$ percentile of the F distribution with degrees of freedoms (dfs) m and n.

4.2. Tests for $b_1 - b_2$

Consider testing

$$H_0 : b_1 = b_2 \text{ vs. } H_a : b_1 \neq b_2 \tag{23}$$

The log-likelihood function when $b_1 = b_2 = b$ can be written as

$$l(a_1, a_2, b) = l\left(a_1, b|\bar{X}_1, \widetilde{G}_1\right) + l\left(a_2, b|\bar{X}_2, \widetilde{G}_2\right) \tag{24}$$

where $l(a, b|\bar{X}, \widetilde{G})$ is given in (2). It is easy to see that the constrained MLEs are

$$\widehat{b}_c = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1\widehat{a}_{1c} + n_2\widehat{a}_{2c}}$$

where $\widehat{a}_{1c}$ and $\widehat{a}_{2c}$ are solutions of

$$\begin{aligned} \ln(n_1a_1 + n_2a_2) - \psi(a_1) &= \ln\frac{\left(n_1\bar{X}_1 + n_2\bar{X}_2\right)}{\widetilde{G}_1} \\ \ln(n_1a_1 + n_2a_2) - \psi(a_2) &= \ln\frac{\left(n_1\bar{X}_1 + n_2\bar{X}_2\right)}{\widetilde{G}_2} \end{aligned} \tag{25}$$

The MLEs $\widehat{a}_{1c}$ and $\widehat{a}_{2c}$ can be obtained iteratively; see Appendix A. The SLRT statistic can be expressed as

$$R(d_b) = \text{sign}(\widehat{b}_1 - \widehat{b}_2)\left\{2\left[l(\widehat{a}_1, \widehat{b}_1, \widehat{a}_2, \widehat{b}_2) - l\left(\widehat{a}_{1c}, \widehat{a}_{2c}, \widehat{b}_c\right)\right]\right\}^{1/2} \tag{26}$$

where $l(a_1, a_2, b)$ is given in (24), and $l(a_1, b_1, a_2, b_2)$ is given in (19). For an observed value $R_0(d_b)$ of $R(d_b)$ and a given nominal level α, the null hypothesis in (23) is rejected if the p-value $P(|R(d_b)| > |R_0(d_b)|) < \alpha$. The p-value can be estimated as described in the succeeding text.

As the testing problem is scale invariant, the null distribution of the SLRT statistic $R(d_b)$ may depend on (a_1, a_2). However, as in the case of two-sample test for the shape parameters, we find strong simulation evidence to indicate that the null distribution of $R(d_b)$ does not depend on any unknown parameters, and it depends only on sample sizes. In other words, for given sample sizes, the percentiles of the $R(d_b)$ are not affected by the values of the shape parameters a_1 and a_2. For example, we estimated quantiles of $R(d_b)$ based on independent samples of sizes ($n_1 = 4, n_2 = 7$) from gamma(1, 1) distributions along with quantiles of $R(d_b)$ based on independent samples from

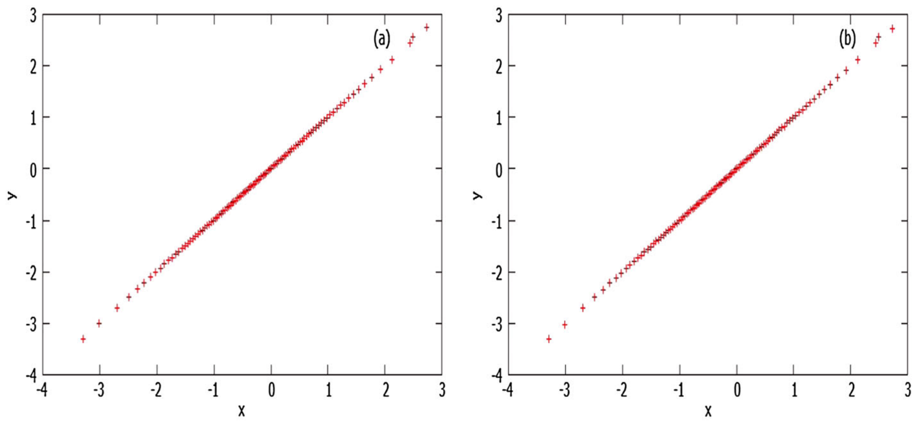

Figure 6. Quantiles of the signed-likelihood ratio test (SLRT) statistic $R(d_b)$ based on samples from different gamma distributions: (a) (gamma(1, 1),gamma(1, 1)) vs. (gamma(12,5),gamma(4, 5)); (b) (gamma(1, 1),gamma(1, 1)) vs. (gamma(5, 12),gamma(30, 12)); $n_1 = 4, n_2 = 7$

gamma(12, 5) and gamma(4,5) distributions and plotted them in Figure 6 plot (a). Plot (b) represents quantiles of $R(d_b)$ based on samples from (gamma(1,1), gamma(1,1)) and those based on (gamma(5,12), gamma(30,12)) distributions. These two plots clearly indicate that the quantiles based on different distributions are the same, and so the null distribution $R(d_b)$ does not depend on any unknown parameters. As a result, the p-value of the SLRT for $b_1 = b_2$ can be estimated as in the case of two-sample test for the shape parameters.

4.3. Tests for the difference between two means

Let $\mu_i = a_i b_i, i = 1, 2$, and consider testing

$$H_0 : \mu_1 = \mu_2 \text{ vs. } H_a : \mu_1 \neq \mu_2 \tag{27}$$

4.3.1. *The SLRT for $\mu_1 - \mu_2$*

Denoting the unknown common mean under H_0 by μ, the log-likelihood function under H_0 can be expressed as

$$\sum_{i=1}^{2} l(a_i, \mu) = \sum_{i=1}^{2} \left(-n_i \ln \Gamma(a_i) - n_i a_i \ln \frac{\mu}{a_i} - n_i \bar{X}_i \frac{a_i}{\mu} + n_i (a_i - 1) \ln \widetilde{G}_i \right) \tag{28}$$

The SLRT statistic is defined as

$$R(d_\mu) = \text{sign}(\widehat{\mu}_1 - \widehat{\mu}_2) \left\{ 2 \left[\sum_{i=1}^{2} l(\widehat{a}_i, \widehat{b}_i | \bar{X}_i, \widetilde{G}_i) - \sum_{i=1}^{2} l(\widehat{a}_{ic}, \widehat{\mu}_c) \right] \right\}^{1/2} \tag{29}$$

where $l(a_i, b_i | \bar{X}_i, \widetilde{G}_i)$ is defined in (2) and $l(a_i, \mu)$ is defined in (28). Details for calculation of the constrained MLEs are given in Appendix B. For an observed value $R_0(d_\mu)$ of $R(d_\mu)$, the SLRT rejects H_0 in (27) when

$$P\left(|R(d_\mu)| \geqslant |R_0(d_\mu)| \right) < \alpha \tag{30}$$

An MC estimate of the aforementioned p-value can be obtained based on samples generated from gamma$(\widehat{a}_{1c}, \widehat{\mu}_c/\widehat{a}_{1c})$ and gamma$(\widehat{a}_{2c}, \widehat{\mu}_c/\widehat{a}_{2c})$ distributions.

4.3.2. *A computational approach test*

Chang *et al.* (2011) have proposed a test for equality of several gamma means, referred to as the computational approach test (CAT), which is based on the test statistic $\sum_{i=1}^{k} (\ln \widehat{\mu}_i - \overline{\ln \widehat{\mu}})^2$, where $\widehat{\mu}_i$ is the MLE of $\mu_i, i = 1, \ldots, k$. For the two-sample case, the test statistic simplifies to $(\ln \widehat{\mu}_1 - \ln \widehat{\mu}_2)^2$. The percentiles (under $\mu_1 = \mu_2$) of the test statistic is estimated based on simulated samples from gamma$(\widehat{a}_{1c}, \widehat{\mu}_c/\widehat{a}_{1c})$

Table 3. MC estimates of type I error rates of the SLRT and the (SBE) test $H_0 : a_1 \leqslant a_2$ vs. $H_a : a_1 > a_2$

$\alpha = .01$	(n_1, n_2)				
a	(5,4)	(4,9)	(9,6)	(20,4)	(15,15)
0.5	0.009(0.009)	0.009(0.008)	0.009(0.008)	0.010(0.008)	0.011(0.009)
1	0.010(0.010)	0.010(0.009)	0.010(0.008)	0.010(0.009)	0.010(0.010)
1.5	0.010(0.010)	0.010(0.010)	0.010(0.010)	0.010(0.009)	0.010(0.010)
2	0.010(0.010)	0.011(0.010)	0.010(0.009)	0.010(0.010)	0.010(0.010)
2.5	0.010(0.010)	0.010(0.010)	0.010(0.010)	0.010(0.010)	0.010(0.009)
3	0.010(0.009)	0.010(0.010)	0.010(0.010)	0.010(0.010)	0.010(0.010)
$\alpha = 0.05$					
a	(5,4)	(4,9)	(9,6)	(20,4)	(15,15)
0.5	0.049(0.043)	0.049(0.049)	0.049(0.057)	0.053(0.047)	0.051(0.047)
0.1	0.049(0.047)	0.049(0.051)	0.050(0.055)	0.050(0.051)	0.050(0.051)
1.5	0.049(0.049)	0.050(0.051)	0.051(0.052)	0.052(0.050)	0.050(0.050)
2	0.050(0.050)	0.049(0.051)	0.051(0.054)	0.051(0.050)	0.048(0.050)
2.5	0.049(0.049)	0.050(0.049)	0.051(0.053)	0.050(0.049)	0.047(0.050)
3	0.050(0.050)	0.050(0.050)	0.050(0.051)	0.049(0.051)	0.051(0.050)

Mc, Monte Carlo; SBE, Shiue–Bain–Engelhardt; SLRT, signed-likelihood ratio test.

and gamma($\widehat{a}_{2c}, \widehat{\mu}_c/\widehat{a}_{2c}$). The test rejects the null hypothesis in (27) at the level α, if an observed value of the test statistic is larger than $100(1-\alpha)$ percentile of the $(\ln \widehat{\mu}_1 - \ln \widehat{\mu}_2)^2$.

4.4. Type I error rates and power studies for two-sample problems

We evaluated the type I error rates and powers of the tests for the two-sample problems addressed in the preceding sections using MC simulation. The MC estimates of type I error rates of the SLRT and the SBE test for $H_0 : a_1 \leqslant a_2$ vs. $H_a : a_1 > a_2$ are given in Table 3. Examination of the estimated type I error rates indicates that the SBE test and the SLRT for the shape parameters are very satisfactory and they are quite comparable. We also plotted power surfaces of these two tests in Figure 7 for testing two-sided hypotheses for values of a_1 and a_2 in the interval (.5, 4.5) while (b_1, b_2) is fixed at $(1, 2)$. The power surfaces coincide with one another, indicating that these two tests have very similar power properties. Between the SLRT and the SBE test, the latter is preferable because it is simple to apply.

The estimated type I error rates of the SLRT for $H_0 : b_1 \leqslant b_2$ vs. $H_a : b_1 > b_2$ are given in Table 4. Furthermore, the type I error rates were evaluated when $b_1 = b_2 = 1$ and for some values of (a_1, a_2) as given in Table 4. We see in Table 4 that the estimated type I error rates

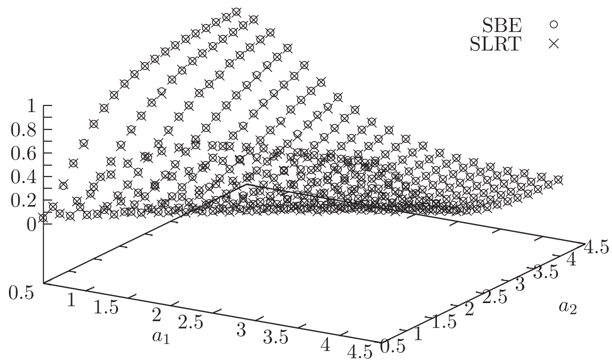

Figure 7. Power surfaces of the signed-likelihood ratio test (SLRT) and Shiue–Bain–Engelhardt's (SBE's) test for $H_0 : a_1 = a_2$ vs. $H_a : a_1 \neq a_2$ at the level 0.05; $b_1 = 1, b_2 = 2; n_1 = n_2 = 10$

Table 4. MC estimates of type I error rates of the SLRT for testing $H_0 : b_1 \leqslant b_2$ vs. $H_a : b_1 > b_2$

	$b_1 = b_2 = 1$											
	$\alpha = 0.01$ (n_1, n_2)				$\alpha = 0.05$ (n_1, n_2)				$\alpha = 0.10$ (n_1, n_2)			
(a_1, a_2)	(4,4)	(5,7)	(4,10)	(5,15)	(4,4)	(5,7)	(4,10)	(5,15)	(4,4)	(5,7)	(4,10)	(5,15)
(0.5,0.5)	0.010	0.010	0.011	0.011	0.050	0.048	0.053	0.051	0.101	0.098	0.100	0.105
(0.5,3)	0.011	0.010	0.010	0.009	0.051	0.051	0.050	0.050	0.101	0.096	0.103	0.103
(1.5,4)	0.011	0.010	0.010	0.010	0.049	0.052	0.051	0.051	0.100	0.097	0.101	0.104
(2,6)	0.009	0.011	0.011	0.010	0.051	0.048	0.051	0.048	0.099	0.101	0.102	0.102
(2,10)	0.010	0.009	0.010	0.010	0.049	0.051	0.047	0.050	0.100	0.103	0.102	0.100
(8,0.5)	0.010	0.009	0.011	0.011	0.050	0.050	0.050	0.047	0.099	0.100	0.098	0.100

Mc, Monte Carlo; SLRT, signed-likelihood ratio test.

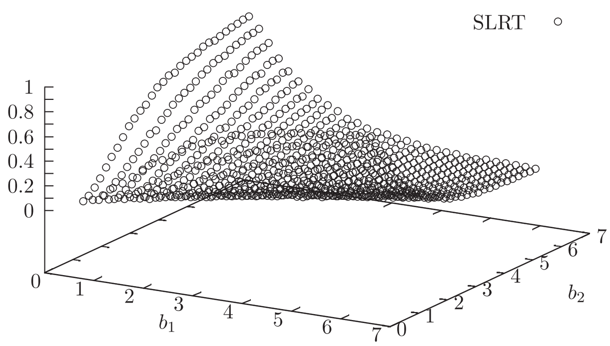

Figure 8. Power surface of the signed-likelihood ratio test (SLRT) for $H_0 : b_1 = b_2$ vs. $H_a : b_1 \neq b_2$ at the level 0.05; $a_1 = 2, a_2 = 4; n_1 = n_2 = 10$

Table 5. MC estimates of type I error rates of the SLRT and (CAT) for testing the equality of two gamma means

$\alpha = 0.05$ (a_1, a_2, μ)	(5,5) SLRT(CAT)	(5,10) SLRT(CAT)	(10,10) SLRT(CAT)	(15,10) SLRT(CAT)	(15,15) SLRT(CAT)
(0.5,0.5,1)	0.049(0.062)	0.048(0.050)	0.051(0.046)	0.050(0.052)	0.050(0.047)
(1.5,1,2)	0.049(0.034)	0.052(0.047)	0.050(0.059)	0.050(0.051)	0.050(0.040)
(2,12,3)	0.045(0.020)	0.051(0.040)	0.053(0.039)	0.051(0.059)	0.051(0.057)
(3,10,4)	0.045(0.020)	0.050(0.033)	0.051(0.045)	0.047(0.054)	0.047(0.053)
(4,0.5,5)	0.053(0.039)	0.051(0.049)	0.050(0.051)	0.051(0.053)	0.051(0.058)
(10,20,6)	0.040(0.011)	0.048(0.040)	0.050(0.057)	0.050(0.040)	0.050(0.048)
.					
$\alpha = 0.10$ (a_1, a_2, μ)					
(0.5,0.5,1)	0.111(0.121)	0.108(0.109)	0.105(0.098)	0.099(0.108)	0.103(0.116)
(1.5,1,2)	0.120(0.084)	0.094(0.127)	0.100(0.126)	0.096(0.103)	0.101(0.107)
(2,12,3)	0.103(0.092)	0.102(0.101)	0.097(0.116)	0.100(0.113)	0.096(0.108)
(3,10,4)	0.113(0.062)	0.104(0.109)	0.099(0.106)	0.102(0.095)	0.101(0.102)
(4,0.5,5)	0.103(0.105)	0.103(0.103)	0.100(0.096)	0.103(0.104)	0.098(0.111)
(10,20,6)	0.113(0.072)	0.094(0.120)	0.096(0.097)	0.097(0.123)	0.102(0.118)
.					

MC, Monte Carlo; CAT, computational approach test; SLRT, signed-likelihood ratio test.

Table 6. Powers of the SLRT and (CAT) for testing the equality of two gamma means at $\alpha = 0.05$ $n_1 = n_2 = 15$

	μ_1							
μ_2	1	1.5	2	2.5	3	3.5	4	4.5
1	0.05(0.05)	0.29(0.29)	0.68(0.65)	0.91(0.88)	0.97(0.97)	0.98(0.99)	0.99(0.99)	0.99 (0.99)
1.5	0.29(0.29)	0.05(0.06)	0.14(0.16)	0.39(0.43)	0.67(0.69)	0.83(0.82)	0.91(0.93)	0.96 (0.96)
2	0.69(0.68)	0.19(0.15)	0.05(0.06)	0.13(0.13)	0.30(0.33)	0.47(0.48)	0.65(0.68)	0.80 (0.78)
2.5	0.89(0.88)	0.41(0.43)	0.12(0.11)	0.05(0.06)	0.09(0.10)	0.24(0.25)	0.35(0.38)	0.50 (0.53)
3	0.95(0.96)	0.66(0.65)	0.29(0.29)	0.10(0.10)	0.05(0.05)	0.09(0.07)	0.15(0.18)	0.28 (0.28)
3.5	0.97(0.96)	0.82(0.82)	0.49(0.48)	0.23(0.23)	0.10(0.09)	0.05(0.05)	0.09(0.08)	0.14 (0.16)
4	0.97(0.97)	0.91(0.91)	0.66(0.67)	0.37(0.36)	0.17(0.15)	0.07(0.08)	0.05(0.03)	0.07 (0.06)
4.5	0.98(0.98)	0.96(0.95)	0.79(0.79)	0.53(0.55)	0.28(0.29)	0.14(0.13)	0.07(0.06)	0.05 (0.05)

CAT, computational approach test; SLRT, signed-likelihood ratio test.

are very close to the corresponding nominal levels for all the cases considered. Thus, our simulation studies indicate that the SLRT is very satisfactory for testing the difference between two shape parameters. The power surface of the SLRT is plotted in Figure 8 for $n_1 = n_2 = 10$, which clearly indicates that the type I error rates are very close to the nominal level 0.05, and the power is an increasing function of $|b_1 - b_2|$.

Type I error rates of the SLRT for testing the equality of two means are estimated as follows. We first generated 2500 samples each of size n_i from a gamma$(a_i, \mu/a_i)$ distribution, $i = 1, 2$. For each sample, we calculated the SLRT statistic (29) and estimated the mean and SD of the SLRT statistic based on 5000 samples of size n_i generated from gamma$(\widehat{a}_{ic}, \widehat{\mu}_c/\widehat{a}_{ic})$ distribution, $i = 1, 2$. The SLRT was calculated for each of 2500 samples, and the proportion of the 2500 statistics that lead to the rejection of the null hypothesis is an MC estimate of the type I error rate at the parameter values of (a_1, a_2, μ). The estimates of type I error rates of CAT can be obtained similarly. The MC estimates of type I error rates of the SLRT and CAT are given in Table 5 for some selected values of (a_1, a_2, μ). We see in Table 5 that the CAT is slightly liberal or conservative for some cases. For smaller values of sample sizes, the SLRT is better than the CAT with respect to type I error rates. The CAT maybe satisfactory for sample of sizes 15 or larger. Powers of these two tests were estimated for sample size $n_1 = n_2 = 15$ and reported in Table 6. The powers of the SLRT and CAT are quite comparable, except that the type I error rates of the CAT is slightly different from the nominal level for some cases.

5. CONFIDENCE INTERVALS

We shall now propose parametric t-percentile bootstrap CIs for the gamma parameters, means, CIs for comparing two shape parameters and for comparing two gamma means. As mentioned in Shao and Tu (1995, p. 16), the PB is more efficient than the nonparametric counterpart when the parametric model is approximately correct. Hall (1988) examined the theoretical properties of seven bootstrap methods in the parametric and nonparametric contexts, and concluded that the t-percentile and the accelerated bias correction are the most promising methods. Empirical comparisons of four bootstrap techniques in Shao and Tu (1995, p. 106) for an example indicated that the CIs based on the t-percentile approach are better than the accelerated bias correction bootstrap. The t-percentile method may not be the best for all parametric inference. We choose t-percentile method for finding CIs for gamma parameters because approximate expressions for the required variance estimates can be readily obtained from the Fisher information matrix (5). Furthermore, we show via simulation studies that the CIs based on the t-percentile perform satisfactorily in terms coverage probabilities, and as shown in the sequel, quite comparable with other likelihood approaches given in the preceding sections for some problems.

5.1. Parametric bootstrap confidence intervals

Let $\bar{X}$ and $\widetilde{G}$ denote the mean and geometric mean, respectively, based on a sample of size n from a gamma(a, b) distribution. Let $\widehat{a}$ and $\widehat{b}$ denote the MLEs based on $(\bar{X}, \widetilde{G})$. Similarly, let $\bar{X}^*$ and $\widetilde{G}^*$ denote the mean and geometric mean, respectively, based on a bootstrap sample of size n generated from the gamma$(\widehat{a}, \widehat{b})$ distribution. Let $(\widehat{a}^*, \widehat{b}^*)$ denote the MLEs based on $(\bar{X}^*, \widetilde{G}^*)$.

5.1.1. *CIs for the shape parameter*

Let $Q_{a;\alpha}$ denote the 100α percentile of

$$Q_a = \frac{\widehat{a}^* - \widehat{a}}{\widehat{\sigma}_{\widehat{a}^*}} = \frac{\widehat{a}^* - \widehat{a}}{\{\widehat{a}^*/[n(\widehat{a}^*\psi'(\widehat{a}^*) - 1)]\}^{1/2}} \tag{31}$$

where the variance estimate $\widehat{\sigma}^2_{\widehat{a}^*}$ in the aforementioned expression is obtained from the inverse Fisher information matrix (5). The $100(1-2\alpha)$ percent PB CI for the shape parameter a is given by

$$\left(\widehat{a} - Q_{a;\alpha}\widehat{\sigma}_{\widehat{a}},\ \ \widehat{a} - Q_{a;1-\alpha}\widehat{\sigma}_{\widehat{a}}\right) \tag{32}$$

where $\widehat{\sigma}^2_{\widehat{a}} = \widehat{a}/[n(\widehat{a}\psi'(\widehat{a}) - 1)]$. The following algorithm can be used to estimate the percentiles $Q_{a;\alpha}$ and $Q_{a;1-\alpha}$.

Algorithm 1

1. For a given sample of size n, calculate the MLEs $\widehat{a}$ and $\widehat{b}$.
2. Generate a bootstrap sample of size n from gamma$(\widehat{a}, \widehat{b})$ distribution and calculate the MLEs $\widehat{a}^*$ and $\widehat{b}^*$ based on the bootstrap sample.
3. Set $Q = \frac{\widehat{a}^* - \widehat{a}}{\left\{\widehat{a}^*/\left[n\left(\widehat{a}^*\psi'(\widehat{a}^*) - 1\right)\right]\right\}^{1/2}}$
4. Repeat steps 2 and 3 for a large number of times, say, 10000.
5. The 100α lower percentile and the 100α upper percentile of Q's are estimates of $Q_{a;\alpha}$ and $Q_{a;1-\alpha}$, respectively.

5.1.2. *CIs for the scale parameter*

To find the PB CI for the scale parameter b, we note that

$$Q_b = \frac{\widehat{b}^* - \widehat{b}}{\widehat{\sigma}_{\widehat{b}^*}} = \frac{\widehat{b}^* - \widehat{b}}{\left\{\widehat{b}^{*2}\psi'(\widehat{a}^*)/\left[n\left(\widehat{a}^*\psi'(\widehat{a}^*) - 1\right)\right]\right\}^{1/2}} \tag{33}$$

where the variance estimate of $\widehat{b}^*$ is obtained from the Fisher information matrix. Letting $\widehat{\sigma}^2_{\widehat{b}} = \widehat{b}^2\psi'(\widehat{a})/\left[n\left(\widehat{a}\psi'(\widehat{a}) - 1\right)\right]$, the PB CI for b is given by

$$\left(\widehat{b} - Q_{b;1-\alpha}\widehat{\sigma}_{\widehat{b}},\ \ \widehat{b} - Q_{b;\alpha}\widehat{\sigma}_{\widehat{b}}\right) \tag{34}$$

where $Q_{b;\alpha}$ is the 100α percentile of Q_b. The aforementioned PB CI can be estimated using an algorithm similar to Algorithm 1.

5.1.3. *Confidence interval for the mean*

Recall that the mean of a gamma distribution is given by $\mu = ab$, and so the MLE of μ is $\widehat{\mu} = \widehat{a}\widehat{b} = \bar{X}$. The variance estimate

$$\widehat{\sigma}^2_{\bar{X}} = \frac{\widehat{a}\widehat{b}^2}{n} = \frac{\bar{X}^2}{n\widehat{a}}$$

The PB pivotal is given by

$$Q_\mu = \frac{\widehat{\mu}^* - \widehat{\mu}}{\widehat{\sigma}_{\bar{X}^*}} = \frac{(\bar{X}^* - \bar{X})}{\bar{X}^*/\sqrt{n\widehat{a}^*}} \tag{35}$$

where the MLEs $\widehat{a}^*$ and $\bar{X}^*$ are based on the bootstrap sample from gamma$(\widehat{a},\widehat{b})$ distribution. The $100(1-2\alpha)$ percent PB CI for μ is given by

$$\left(\bar{X} - Q_{\mu;1-\alpha}\frac{\bar{X}}{\sqrt{n\widehat{a}}},\ \bar{X} - Q_{\mu;\alpha}\frac{\bar{X}}{\sqrt{n\widehat{a}}}\right) \tag{36}$$

where $Q_{\mu;\alpha}$ is the 100α percentile of Q_μ defined in (35).

5.2. Confidence intervals: two-sample case

5.2.1. *Confidence intervals for $a_1 - a_2$*

Let $(\widehat{a}_i,\widehat{b}_i)$ denote the MLE of (a_i,b_i) based on a sample of size n_i from a gamma(a_i,b_i) distribution, $i=1,2$. The PB pivotal quantities for various two-sample problems can be readily obtained from those for the one-sample problems. For ease of reference, we shall outline the PB methods for finding CIs for the difference between shape parameters, scale parameters, and for the difference between the means.

The PB pivotal to estimate the difference $a_1 - a_2$ is given by

$$Q_{a_1-a_2} = \frac{(\widehat{a}_1^* - \widehat{a}_2^*) - (\widehat{a}_1 - \widehat{a}_2)}{\sqrt{\widehat{\sigma}^2_{\widehat{a}_1^*} + \widehat{\sigma}^2_{\widehat{a}_2^*}}} \tag{37}$$

where $\widehat{\sigma}^2_{\widehat{a}_i^*} = \widehat{a}_i^*/\left[n_i\left(\widehat{a}_i^*\psi'(\widehat{a}_i^*) - 1\right)\right]$, $i=1,2$. The $100(1-2\alpha)$ percent CI for $a_1 - a_2$ is given by

$$\left(\widehat{a}_1 - \widehat{a}_2 - Q_{a_1-a_2;1-\alpha}\left[\widehat{\sigma}^2_{\widehat{a}_1} + \widehat{\sigma}^2_{\widehat{a}_2}\right]^{1/2},\ \widehat{a}_1 - \widehat{a}_2 - Q_{a_1-a_2;\alpha}\left[\widehat{\sigma}^2_{\widehat{a}_1} + \widehat{\sigma}^2_{\widehat{a}_2}\right]^{1/2}\right) \tag{38}$$

where $Q_{a_1-a_2;\alpha}$ denotes the 100α percentile of $Q_{a_1-a_2}$. This percentile, for a given $(\widehat{a}_1,\widehat{b}_1,\widehat{a}_2,\widehat{b}_2)$, can be obtained using MC simulation as described in Algorithm 1 for the one-sample case.

5.2.2. *Confidence intervals for the difference between two means*

To find a PB CI for the difference between two means $\mu_1 = a_1b_1$ and $\mu_2 = a_2b_2$, the PB pivotal can be expressed as

$$Q_{\mu_1-\mu_2} = \frac{(\bar{X}_1^* - \bar{X}_2^*) - (\bar{X}_1 - \bar{X}_2)}{\left\{\frac{\bar{X}_1^{*2}}{n_1\widehat{a}_1^*} + \frac{\bar{X}_2^{*2}}{n_2\widehat{a}_2^*}\right\}^{1/2}} \tag{39}$$

where $\bar{X}_i^*$ is the mean of a bootstrap sample generated from gamma$(\widehat{a}_i,\widehat{b}_i)$ distribution, $i=1,2$. The $100(1-2\alpha)$ percent CI for $\mu_1 - \mu_2$ is given by

$$\left(\bar{X}_1 - \bar{X}_2 - Q_{\mu_1-\mu_2;1-\alpha}\left\{\frac{\bar{X}_1^2}{n_1\widehat{a}_1} + \frac{\bar{X}_2^2}{n_2\widehat{a}_2}\right\}^{1/2},\ \bar{X}_1 - \bar{X}_2 - Q_{\mu_1-\mu_2;\alpha}\left\{\frac{\bar{X}_1^2}{n_1\widehat{a}_1} + \frac{\bar{X}_2^2}{n_2\widehat{a}_2}\right\}^{1/2}\right) \tag{40}$$

where $Q_{\mu_1-\mu_2;\alpha}$ denotes the 100α percentile of $Q_{\mu_1-\mu_2}$. This percentile, for a given $(\widehat{a}_1,\widehat{b}_1,\widehat{a}_2,\widehat{b}_2)$, can be obtained using MC simulation as described in Algorithm 1 for the one-sample case.

Remark 1. The PB approach for estimating the difference between two scale parameters is not satisfactory in terms of coverage probabilities. In some cases, the coverage probabilities could be as low as 0.8 when the nominal level is 0.95. So, the PB t-percentile approach is not recommended for finding CIs for the difference between two scale parameters.

5.3. Coverage properties of the PB confidence intervals

To judge the performance of the PB CIs described in the preceding sections, we estimated the coverage probabilities as follows. For an assumed set of values of (n,a,b), we generated 2500 samples, each of size n, from the gamma(a,b) distribution. Based on each generated sample, we calculated the 95% PB CI for the parameter of interest based on 5000 bootstrap samples. The percentage of these 5000 PB CIs that include the parameter of interest is a MC estimate of the coverage probability. As noted earlier, the MLE of the shape parameter is scale invariant, and the MLE of the scale parameter is scale equivariant, and so the coverage probabilities do not depend on the values of b. Thus, for coverage studies, we assume $b=1$ without loss of generality.

Table 7. Coverage probabilities of $1-\alpha$ PB CIs for a gamma mean

	$1-\alpha=0.90$			$1-\alpha=0.95$			$1-\alpha=0.99$		
	n			n			n		
a	5	10	15	5	10	15	5	10	15
0.5	0.902	0.897	0.902	0.948	0.949	0.950	0.987	0.990	0.992
1	0.895	0.902	0.895	0.946	0.952	0.947	0.990	0.990	0.986
1.5	0.901	0.894	0.901	0.950	0.950	0.954	0.985	0.990	0.988
2	0.897	0.907	0.897	0.951	0.950	0.952	0.986	0.992	0.990
5	0.901	0.910	0.901	0.949	0.950	0.952	0.992	0.992	0.991
10	0.907	0.890	0.901	0.946	0.944	0.952	0.986	0.992	0.992
20	0.905	0.899	0.899	0.953	0.951	0.955	0.993	0.989	0.987
30	0.900	0.892	0.909	0.947	0.948	0.945	0.990	0.993	0.991

CI, confidence interval; PB, parametric bootstrap.

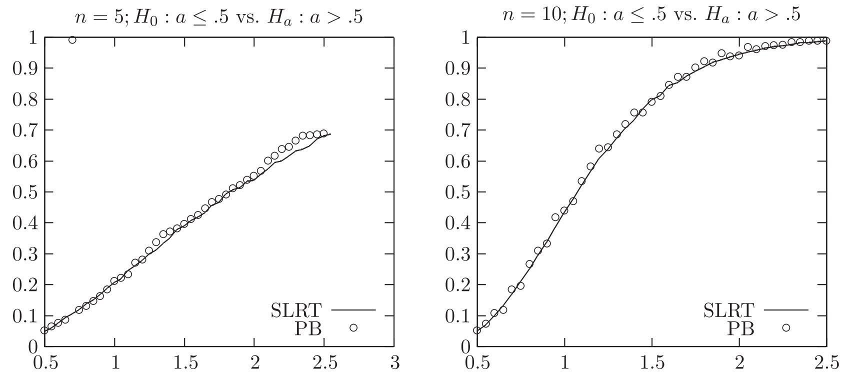

Figure 9. Powers of the signed-likelihood ratio test (SLRT) and the parametric bootstrap (PB) test as a function of a; $\alpha = 0.05$

The estimated coverage probabilities for the mean are given in Table 7 for sample sizes $n = 5, 10$, and 15, and for the confidence coefficients 0.90, 0.95, and 0.99. We see in Table 7 that the coverage probabilities are very close to the nominal levels even for sample of size 5. Thus, the PB CIs are very satisfactory in terms of coverage probabilities for estimating the mean of a gamma distribution. We also estimated the coverage probabilities of PB CIs for the scale parameters and the shape parameters. In general, the PB CIs for the shape parameters are as good as the CIs for the mean, and they are satisfactory for samples of size as small as 5 and the shape parameter $a \geqslant 0.5$. We also noticed similar performance of the PB CIs for the scale parameter except that they are slightly liberal for small parameter values and sample sizes. These coverage estimates are not reported here in order to save space.

In order to judge the power of the test (PB test) based on the PB CIs for the shape parameter a, we estimated the powers and plotted them along with those of the SLRT in Figure 9. The power plots indicate that the PB test is as good as the SLRT even for sample size 5. So the PB CIs for a are expected to be as good as the ones obtained by inverting the SLRT for a. We also made similar comparison for powers of the PB test and the MLRT for the mean in Figure 10. The power plots indicate that for sample size 10, the powers of the MLRT for testing $H_0 : \mu = 1$ vs. $H_a : \mu \neq 1$ are larger than those of the PB test. The difference between powers decreasing with increasing sample size and/or the null value of the mean is not small. In particular, we see that the powers of the MLRT and PB test for $H_0 : \mu = 3$ vs. $H_a : \mu \neq 3$ are not appreciably different when sample size is 10. This comparison holds for interval estimation in the sense that the PB CIs are comparable with those based on the MLRT for sample sizes around 15 or more. If the mean is known to be large, then PB CIs for the mean can be used even for small sample sizes. We also compared the powers of the PB test and the SLRT for the scale parameter (not reported here) and noted that these two tests have similar power properties for the cases where the PB test controls the type I error rates close to the nominal level.

MC estimates of coverage probabilities of CIs for the difference between two means are given in Table 8. Our preliminary simulation studies indicated that the coverage probabilities are not much affected by the values of the scale parameters, and so we chose $b_1 = b_2 = 1$ for coverage studies. The estimated coverage probabilities in Table 8 indicate that the PB CIs for the difference between two means are satisfactory except for small samples and small values of shape parameters. Even in these cases, the coverage probabilities are not much lower than the nominal level. Specifically, we observe from Table 8 that the PB CIs for the difference between two means are satisfactory for moderate sample sizes, and they could be slightly liberal when both sample sizes and the shape parameters are small.

The coverage results of the PB CIs for the difference between two shape parameters are very similar to those of the PB CIs for the difference between two means, and so they are not reported here. In general, we observed that the coverage probabilities are slightly smaller than the nominal levels when the shape parameters and sample sizes are small, and they are close to the nominal level for sample sizes 20 or larger, and parameters are not too small.

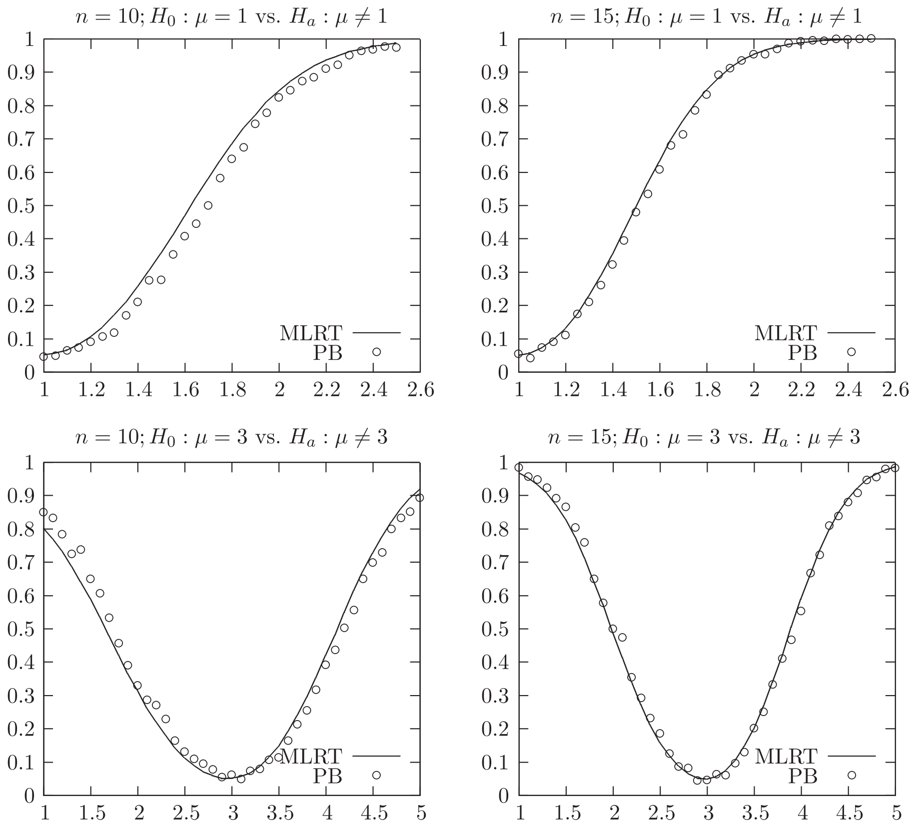

Figure 10. Powers of the modified likelihood ratio test (MLRT) and the parametric bootstrap (PB) test as a function of μ; $\alpha = 0.05$

Table 8. Coverage probabilities of 95% PB CIs for the difference between two gamma means

	n				n			
(a_1, a_2)	(5,5)	(10,10)	(15,15)	(30,30)	(5,10)	(15,10)	(20,10)	(15,30)
(0.5,1.5)	0.921	0.931	0.923	0.932	0.915	0.929	0.925	0.935
(1,3)	0.923	0.932	0.941	0.940	0.931	0.934	0.928	0.930
(1.5,5)	0.928	0.940	0.937	0.940	0.935	0.943	0.937	0.942
(2,3)	0.944	0.940	0.952	0.944	0.940	0.940	0.944	0.937
(5,15)	0.950	0.943	0.949	0.950	0.950	0.947	0.952	0.944
(10,3)	0.935	0.938	0.946	0.950	0.937	0.947	0.954	0.946
(9,2)	0.943	0.940	0.946	0.949	0.944	0.944	0.946	0.952
(10,9)	0.953	0.951	0.944	0.944	0.946	0.943	0.945	0.945

CI, confidence interval; PB, parametric bootstrap.

Overall, we see that the PB CIs for the one-sample case are quite satisfactory even for small samples, and they do not require additional computation of the constrained MLEs. However, the tests based on the PB CIs are slightly less powerful than the corresponding LRTs described in Section 3 for very small sample sizes. For moderate sample sizes, the PB method can be used for both hypothesis testing and for finding confidence intervals.

6. RECOMMENDATIONS

We have considered several methods for various problems involving hypothesis tests and CIs for gamma parameters. In order to help readers to select the best methods considered in this article, we recommend the following methods on the basis of simplicity and accuracy. The R codes for the recommended methods are posted at http://www.ucs.louisiana.edu/~kxk4695, and they are also available at this journal's web page.

Recommended Methods

One-Sample Problems		Two-Sample Problems	
Test for shape parameter	SLRT	Test for the equality of shape parameters	SBE test
Test for scale parameter	SLRT	Test for the equality of shape parameters	SLRT
Test for the mean	MLRT	Test for the equality of means	SLRT
CI for the shape and scale parameters	PB	CI for $a_1 - a_2$	PB
		CI for $\mu_1 - \mu_2$	PB
CI for the mean	MLRT or PB		

Table 9. Vinyl chloride concentrations in monitoring wells

5.1	2.4	.4	.5	2.5	.1	6.8	1.2	.5	.6	5.3	2.3	1.8
1.2	1.3	1.1	.9	3.2	1.	.9	.4	.6	8	.4	2.7	.2
2	.2	.5	.8	2	2.9	.1	4					

7. EXAMPLES

Example 1. To illustrate the methods for one-sample problems described earlier, we consider the data given in Table 1 of Bhaumik and Gibbons (2006). The data, reproduced in Table 9, represent vinyl chloride concentrations (in μg/L) collected from clean upgradient monitoring wells. A quantile–quantile plot by Bhaumik and Gibbons showed an excellent fit of these data to a gamma distribution.

The MLEs are $\widehat{a} = 1.063$ and $\widehat{b} = 1.769$. Note that the sample size is $n = 34$. To find a 90% CI for the mean concentration, we need to estimate the lower 5th and the upper 5th percentiles of Q_μ in (35). On the basis 100,000 bootstrap samples (each of size 34), we estimated the lower 5th percentile of Q_μ as -2.138 and the upper 5th percentile as 1.402. Noting that the mean $\bar{X} = 1.879$, the 95% PB CI (36) for the mean is calculated as

$$\left(1.879 - 1.402\frac{1.879}{\sqrt{34 \times 1.063}}, 1.879 + 2.138\frac{1.879}{\sqrt{34 \times 1.063}}\right) = (\mathbf{1.44, 2.55})$$

The CI based on the MLRT in Section 3.3 can be obtained using the aforementioned PB CI as the starting values as follows. Note that

$$MLRT(1.44) = 1.67 \simeq z_{.95} \text{ and } MLRT(2.55) = -1.63 \simeq z_{.025},$$

where $MLRT(\mu_0)$ is defined in (13). In view of the above, the MLR CI should be close to the PB CI (1.44, 2.55). Indeed, we find by trial-error that $MLRT(1.445) = 1.651$ and $MLRT(2.56) = -1.645$, and so the MLR CI is (**1.45, 2.56**), which is practically the same as the PB CI. The 95% CI (L, U) by Bhaumik *et al.* (2009) is determined by the two roots of the equation $T_3(\mu_0) = F_{1,33;.95} = 2.864$, where $T_3(\mu_0)$ is given in (16). Using the confidence limits in the preceding paragraph as starting values and trial-error, we find $T_3(1.404) = 2.864$ and $T_3(2.595) = 2.863$. Thus, the 95% CI by Bhaumik *et al.* (2009) is given by (**1.40, 2.60**). Even though these three CIs are not appreciably different, the BKG CI is wider than the other two CIs. Note that the CIs are in agreement with our earlier conclusion on power comparison of the BKG test and the MLRT for the mean. Specifically, the BKG test is less powerful than the MLRT; as a result, it yielded a wider CI. We also see from the aforementioned CIs that the 95% upper confidence limit for the mean concentration is approximately 2.56 μg/L.

Suppose it is desired to test $H_0 : a \leqslant 0.5$ vs. $H_a : a > 0.5$, where a is the true shape parameter of the distribution of vinyl chloride concentrations. For this test, we calculated the SLRT statistic (7) as 3.198, and the p-value on the basis of 100,000 simulation runs was obtained as 0.002. To find a 95% CI for the shape parameter of the concentration distribution, we generated 100,000 bootstrap samples (each of size 34) and estimated the lower 2.5th percentile of Q_a in (31) as -2.019 and the upper 2.5th percentile as 1.821. The standard deviation $\widehat{\sigma}_{\widehat{a}}$ is 0.2282. Using (32), we obtain

$$(1.063 - 1.821 \times 0.2282, 1.063 + 2.019 \times 0.2282) = (0.647, 1.523).$$

A 90% PB CI for the scale parameter of the concentration distribution is similarly computed using (34) and is given by

$$(1.769 - 1.125 \times 0.4804, 1.769 + 2.721 \times 0.4804) = (1.23,\ 3.08).$$

Table 10. Single-cloud data for 1968 and 1970

1 Seeded rain	1 Seeded rain	2 Control rain	2 Control rain
129.6	7.7	26.1	28.6
31.4	1656.0	26.3	830.1
2745.6	978.0	87.0	345.5
489.1	198.6	95.0	1202.6
430.0	703.4	1.0[a]	36.6
302.8	1697.8	372.4	4.9
119.0	334.1	17.3	4.9
4.1	118.3	24.4	41.1
92.4	255.0	11.5	29.0
17.5	115.3	321.2	163.0
200.7	242.5	68.5	244.3
274.7	32.7	81.2	147.8
274.7	40.6	47.3	21.7

[a]The actual entry 0 is replaced to fit a gamma model.

To illustrate the test procedures for the scale parameter, let us consider testing $H_0 : b \leqslant 1.3$ vs. $H_a : b > 1.3$. The SLRT statistic in (10) is calculated as 1.204. The p-value $P(R^*(1.3) > 1.204)$, where $R^*(1.3)$ is the SLRT statistic (10) based on samples from gamma$(1, 1)$ distribution, is estimated by MC simulation as 0.080. To apply the BKG test, we calculated $c = 1.116$ and $\nu = 33.97$ using MC simulation with 100,000 runs. Using these values, the p-value of the BKG test is calculated as 0.039. Thus, the BKG test rejects H_0 at the level 0.05, whereas the SLRT does not reject H_0. Note that the result based on the SLRT is in agreement with the 90% PB CI for the scale parameter given earlier.

The R programs that were used to obtain the results for the aforementioned example are posted at www.ucs.louisiana.edu/~kxk4695 and also available at *Environmetrics* Web site.

Example 2. Experimental Meteorology Laboratory conducted randomized pyrotechnic seeding experiments on single clouds in south Florida during 1968 and 1970. Overall, 26 seeded and 26 control clouds were compared in the experiment to judge the effect of seeding. The data (in acre-feet per cloud) are taken from Simpson (1972), and they are given in Table 10. As noted in Simpson (1972), several articles have used gamma models to analyze and compare the data. Using Minitab, we found that the seeded rain data fit a gamma distribution very well (p-value >.250), whereas the data on control rain barely fit a gamma model (p-value > 0.057). We shall use the data to illustrate some two-sample methods described in earlier sections.

The calculated statistics for seeded rain are as follows: $\bar{X}_1 = 441.98$, $\widehat{a}_1 = 0.6396$, $\widehat{b}_1 = 691.05$. For control rain, $\bar{X}_2 = 164.59$, $\widehat{a}_2 = 0.5608$, $\widehat{b}_2 = 293.51$. The mean difference $\bar{X}_1 - \bar{X}_2 = 277.4$.

Let μ_1 denote the mean amount of seeded rain, and let μ_2 denote the same for the control rain. To test the effect of seeding, one may want to test $H_0 : \mu_1 \leqslant \mu_2$ vs. $H_a : \mu_1 > \mu_2$. To obtain the SLRT statistic in (29), we calculated $\widehat{a}_{1c} = 0.6058$, $\widehat{a}_{2c} = 0.4940$, and $\widehat{\mu}_c = 317.4$. Using these statistics in (29), we found the SLRT statistic as 2.604. The p-value $P(R(d_\mu) \geqslant 2.604)$ was estimated based on 10,000 samples generated from gamma$(0.6058, 317.4/0.6058)$ and gamma $(0.4940, 317.4/0.4940)$ as 0.006. This p-value indicates that there is a seeding effect on rainfall. To find a 95% PB CI for $\mu_1 - \mu_2$, the standard error of $\bar{X}_1 - \bar{X}_2$ given in (39) and is calculated as 116.64. Furthermore, the (0.025, 0.975) percentiles of $Q_{\mu_1-\mu_2}$ are estimated based on 100,000 bootstrap samples as $(-2.590, 1.585)$. Substituting these quantities in (40), we obtain the 95% PB CI for $\mu_1 - \mu_2$ as

$$(277.4 - 1.585 \times 116.64,\ 277.4 + 2.590 \times 116.64) = (92.5, 579.5).$$

The aforementioned CI indicates that on the average seeding effect on rainfall exceeded by 92.5 to 579.5 acre-feet.

To illustrate the test for the difference between two shape parameters, let us consider testing $H_0 : a_1 = a_2$ vs. $H_a : a_1 \neq a_2$, where a_1 is the shape parameter for seeded rain and a_2 is the shape parameter for the control rain. Using (21), we calculated the SLRT statistic as 0.3993. The p-value $P(R^*(d_a) > 0.3993)$, where $R^*(d_a)$ is the SLRT statistic (21), was estimated using MC simulation with 100,000 runs as **0.697**. To apply the SBE test (22), the values of $\nu_1 = 26.778$ and $\nu_2 = 27.148$ and the value of the F statistic is 0.8599. The p-value of the SBE test is **0.699**. Note that all p-values are in close agreement, and they all indicate that the shape parameters of the seeded and controlled rain are not significantly different.

To test $H_0 : b_1 \leqslant b_2$ vs. $H_a : b_1 > b_2$, the LRT statistic in (26) is calculated as 1.709. To estimate the p-value, we generated 100,000 pairs of independent samples of sizes $n_1 = 26$ and $n_2 = 26$ from the gamma$(1, 1)$ distribution and calculated the SLRT statistic $R(d_b)$ in (26) for each pair of samples. The percentage of 100,000 statistics exceeding 1.709 was found to be **0.049**. This p-value indicates that the difference between the scale parameters is barely significant at the level of 5%.

The R programs that were used to obtain the results for the aforementioned two-sample problems are posted at www.ucs.louisiana.edu/~kxk4695, and also available as online materials at *Environmetrics* Web page.

8. CONCLUDING REMARKS

We have explored the improved likelihood methods for developing tests for the gamma parameters and means in one-sample and two-sample problems. We have compared the improved likelihood methods with other approximate approaches proposed in the literature and made some recommendations for practical applications. For all the problems considered, the improved likelihood ratio tests are comparable with or better than the existing methods. Furthermore, the tests based on the simulated p-values of the SLRT statistics are third-order accurate and are very satisfactory for small samples. On the basis of our extensive simulation studies, we have conjectured that the null distributions of the SLRT statistics for the scale parameter, difference between two scale parameters, and for the difference between two shape parameters do not depend on any unknown parameters. Our plots on percentiles comparison (Figures 1, 5, and 6) and our simulation studies on type I error rates support this conjecture. As the SLRT statistics for various problems are scale invariant and the family of gamma distributions is scale invariant, the null distributions of the SLRT statistics do not depend on the scale parameters. However, the family of gamma distributions is not location-scale invariant, so standard invariant arguments are not applicable to show that the null distributions are parameter-free. At present, proofs for our conjectures are not clear.

Acknowledgements

The authors are grateful to two reviewers for providing useful comments and suggestions, which enhanced the first version of the paper.

REFERENCES

Aksoy H. 2000. Use of gamma distribution in hydrological analysis. *Turkey Journal of Engineering and Environmental Sciences* **24**:419–428.

Ashkar F, Ouarda TBMJ. 1998. Approximate confidence intervals for quantiles of gamma and generalized gamma distributions. *Journal of Hydrologic Engineering* **3**:43–51.

Bain LJ, Engelhardt M. 1975. A two-moment chi-square approximation for the statistic $\log(\bar{x}/\tilde{x})$. *Journal of the American Statistical Association* **70**: 948–950.

Barndorff-Nielsen OE. 1991. Modified signed log-likelihood ratio. *Biometrika* **79**:557–563.

Bhaumik DK, Gibbons RD. 2006. One-sided approximate prediction intervals for at least p of m observations from a gamma population at each of r locations. *Technometrics* **48**:112–119.

Bhaumik DK, Kapur K, Gibbons RD. 2009. Testing parameters of a gamma distribution for small samples. *Technometrics* **51**:326–334.

Chang C-H, Lin J-J, Pal N. 2011. Testing the equality of several gamma means: a parametric bootstrap method with applications. *Computational Statistics* **26**:55–76.

Das SC. 1955. Fitting truncated type III curves to rainfall data. *Australian Journal of Physics* **8**:298–304.

DiCiccio TJ, Martin MA, Stern SE. 2001. Simple and accurate one-sided inference from signed roots of likelihood ratios. *The Canadian Journal of Statistics* **29**:67–76.

Engelhardt M, Bain LJ. 1977. Uniformly most powerful unbiased tests on the scale parameter of a gamma distribution with a nuisance shape parameter. *Technometrics* **19**:77–81.

Fraser DAS, Reid N. 1995. Ancillaries and third order significance. *Utilitas Mathematica* **47**:33–53.

Fraser DAS, Reid N, Wong A. 1997. Simple and accurate inference for the mean of the gamma model. *The Canadian Journal of Statistics* **25**:91–99.

Gibbons RD. 1994. *Statistical Methods for Groundwater Monitoring*. Wiley: New York.

Grice JV, Bain LJ. 1980. Inferences concerning the mean of the gamma distribution. *Journal of the American Statistical Association* **75**:929–933.

Hall P. 1988. Theoretical comparison of bootstrap confidence intervals. *The Annals of Statistics* **16**:927–953.

Husak GJ, Michaelsen J, Funk C. 2006. Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. *International Journal of Climatology* **27**:935–944.

Keating JP, Glaser RE, Ketchum NS. 1990. Testing hypotheses about the shape parameter of a gamma distribution. *Technometrics* **32**:67–82.

Krishnamoorthy K, Mathew T, Mukherjee S. 2008. Normal based methods for a gamma distribution: prediction and tolerance interval and stress-strength reliability. *Technometrics* **50**:69–78.

Maxim LD, Galvin JB, Niebo R, Segrave AM, Kampa OA, Utell MJ. 2006. Occupational exposure to carbon/coke fibers in plants that produce green or calcined petroleum coke and potential health effects: fiber concentrations. *Inhalation Toxicology* **18**:17–32.

Schickedanz PT, Krause GF. 1970. A test for the scale parameters of two gamma distributions using the generalized likelihood ratio. *Journal of Applied Meteorology* **9**:13–16.

Shao J, Tu D. 1995. *The Jackknife and Bootstrap*. Springer: New York.

Shiue W-K, Bain LJ, Engelhardt M. 1988. Test of equal gamma-distribution means with unknown and unequal shape parameters. *Technomterics* **30**:169–174.

Shiue W-K, Bain LJ. 1990. Simple approximate inference procedures for the mean of the gamma model. *Journal of Statistical Computation and Simulation* **34**:67–73.

Simpson J. 1972. Use of the gamma distribution in single-cloud rainfall analysis. *Monthly Weather Review* **100**:309–312.

Skovgaard IM. 1996. An explicit large deviation approximation to one parameter tests. *Bernoulli* **2**:145–165.

Stephenson DB, Kumar KR, Doblas-Reyes FJ, Royer JF, Chauvin E, Pezzulli S. 1999. Extreme daily rainfall events and their impact on ensemble forecasts of the Indian monsoon. *Monthly Weather Review* **127**:1954–1966.

Thisted RA. 1988. *Elements of Statistical Computing*. Chapman & Hall: New York.

Tripathi RC, Gupta RC, Pair RK. 1993. Statistical tests involving several independent gamma distributions. *Annals of the Institute of Mathematical Statistics* **45**:777–786.

Wilson EB, Hilferty MM. 1931. The distribution of chi-squares. *Proceedings of the National Academy of Sciences* **17**:684–688.

Wong ACM. 1992. Inferences on the shape parameter of a gamma distribution: a conditional approach. *Technometrics* **34**:348–351.

Wong ACM. 1993. A note on inference for the mean parameter of the gamma distribution. *Statistics & Probability Letters* **17**:61–66.

APPENDIX A

To find the roots of the equations in (25), we consider the following equivalent equations:

$$\begin{aligned} f_1(a_1,a_2) &= n_1\left[\ln(\bar{\bar{a}}) - \psi(a_1) - \ln\frac{\bar{\bar{X}}}{\widetilde{G}_1}\right] = 0 \\ f_2(a_1,a_2) &= n_2\left[\ln(\bar{\bar{a}}) - \psi(a_2) - \ln\frac{\bar{\bar{X}}}{\widetilde{G}_2}\right] = 0 \end{aligned} \tag{A.1}$$

where $\bar{\bar{a}} = (n_1a_1 + n_2a_2)/(n_1+n_2)$ and $\bar{\bar{X}} = (n_1\bar{X}_1 + n_2\bar{X}_2)/(n_1+n_2)$. The partial derivative $f_{ii}(a_1,a_2) = \partial f_i/\partial a_i = n_i^2/(n_1a_1 + n_2a_2) - n_i\psi'(a_i)$, $i = 1,2$. Furthermore, $f_{12} = \partial f_1/\partial a_2 = f_{21} = \partial f_2/\partial a_1 = n_1n_2/(a_1n_1 + a_2n_2)$. In terms of the partial derivatives, we obtain the following Newton–Raphson iterative scheme:

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} a_{10} \\ a_{20} \end{pmatrix} - \begin{pmatrix} f_{11}(a_{10},a_{20}) & f_{12}(a_{10},a_{20}) \\ f_{21}(a_{10},a_{20}) & f_{22}(a_{10},a_{20}) \end{pmatrix}^{-1} \begin{pmatrix} f_1(a_{10},a_{20}) \\ f_2(a_{10},a_{20}) \end{pmatrix}$$

The roots can be obtained using the aforementioned scheme with the starting value

$$a_{i0} \simeq \frac{3 - \tilde{s}_i + \sqrt{(\tilde{s}_i - 3)^2 + 24\tilde{s}_i}}{12\tilde{s}_i}$$

where $\tilde{s}_i = \ln\left[\left(n_1\bar{X}_1 + n_2\bar{X}_2\right)/\tilde{X}_i\right]$, $i = 1,2$.

APPENDIX B

Differentiating (28) with respect to a_i, and setting the derivative to zero, we obtain

$$g_i(a_1,a_2) = n_i\left[\ln a_i - \psi(a_i) - \ln\frac{\mu}{\widetilde{G}_i} - \frac{\bar{X}_i}{\mu} + 1\right] = 0,\ i = 1,2 \tag{B.1}$$

The equation $\partial\sum_{i=1}^k l(a_i,\mu)/\partial\mu = 0$ yields

$$\mu = \frac{\sum_{i=1}^k n_ia_i\bar{X}_i}{\sum_{i=1}^k n_ia_i} \tag{B.2}$$

Let $g_{ij} = \partial g_i(a_1,a_2)/\partial a_j$. After substituting the aforementioned expression for μ in (B.1), and taking partial derivatives, we obtain

$$g_{ii}(a_1,a_2) = \frac{n_i}{a_i} - n_i\psi'(a_i) + \frac{n_i^2(\bar{X}_i - \mu)^2}{\mu^2(n_1a_1 + n_2a_2)},\ i = 1,2$$

and

$$g_{ij}(a_1,a_2) = \frac{n_in_j(\bar{X}_i - \mu)(\bar{X}_j - \mu)}{\mu^2(n_1a_1 + n_2a_2)},\ i \neq j$$

where μ is as given in (B.2). In terms of the partial derivatives, we obtain the following Newton–Raphson iterative scheme:

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} a_{10} \\ a_{20} \end{pmatrix} - \begin{pmatrix} g_{11}(a_{10},a_{20}) & g_{12}(a_{10},a_{20}) \\ g_{21}(a_{10},a_{20}) & g_{22}(a_{10},a_{20}) \end{pmatrix}^{-1} \begin{pmatrix} g_1(a_{10},a_{20}) \\ g_2(a_{10},a_{20}) \end{pmatrix}$$

The starting values for the aforementioned iterative process are obtained as follows. Let

$$\mu^* = \frac{\sum_{i=1}^k n_i\bar{X}_i}{\sum_{i=1}^k n_i} \text{ and } s_i(\mu^*) = \ln\frac{\mu^*}{\widetilde{G}_i} + \frac{\bar{X}_i}{\mu^*} - 1,\ i = 1,2$$

Noting that the equations in (B.1) are similar to (3), we find an initial approximation for a_i as

$$a_{i0} = \frac{3 - s_i(\mu^*) + \sqrt{(s_i(\mu^*) - 3)^2 + 24s_i(\mu^*)}}{12s_i(\mu^*)},\ i = 1,2$$

The iterative process with the aforementioned initial values converge in a few steps, in most cases, fewer than five. Finally, note that if $\widehat{a}_{1c}$ and $\widehat{a}_{2c}$ are the roots of the aforementioned iterative scheme, then $\widehat{\mu}_c = (n_1\widehat{a}_{1c}\bar{X}_1 + n_2\widehat{a}_{2c}\bar{X}_2)/(n_1\widehat{a}_{1c} + n_2\widehat{a}_{2c})$.

