Thailand Statistician
January 2022; 20(1): 207-226
http://statassoc.or.th
Contributed paper

Simultaneous Confidence Intervals for All Differences of Coefficients of Variation of Normal Distributions with an Application to PM2.5 Dispersion

Warisa Thangjai [a], Sa-Aat Niwitpong [b] and Suparat Niwitpong*[b]
[a] Department of Statistics, Ramkhamhaeng University, Bangkok, Thailand.
[b] Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand.
*Corresponding author; e-mail: suparat.n@sci.kmutnb.ac.th

Received: 12 October 2020
Revised: 19 January 2021
Accepted: 24 January 2021

Abstract

This paper considers the problem of constructing simultaneous confidence intervals (SCIs) for all pairwise differences of coefficients of variation from several normal distributions. The proposed approaches are based on the generalized confidence interval (GCI) approach, the method of variance estimates recovery (MOVER) approach, and the computational approach. The performances of these approaches, using the biased estimator of the coefficient of variation, are compared with the performances of those approaches using the shrinkage estimator. A simulation study based comparison of these SCIs in terms of the coverage probability, average length, and standard error. The simulation results indicated that the GCI approach and the MOVER approach can provide the SCIs with satisfying coverage probabilities regardless of sample sizes. Furthermore, the performances of the biased estimator are better than the performances of the shrinkage estimator. Finally, an application to PM2.5 dispersion in the Northern Thailand is given to illustrate the proposed simultaneous confidence intervals.

Keywords: Simultaneous confidence intervals, coefficient of variation, simulation, coverage probability, average length.

1. Introduction

For a random sample of sample size n from a normal distribution with mean μ and variance σ^2, let $\bar{X}=\sum_{j=1}^{n} X_j / n$ and $S^2=\sum_{j=1}^{n}(X_j-\bar{X})^2/(n-1)$ be the sample mean and the sample variance, respectively. The normal distribution is the most important distribution in statistics. It fits many phenomena such as height, weight, blood pressure, and intelligence quotient (IQ) scores. Furthermore, this distribution is widely used in many areas. For instance, in hydrology, the distribution of rainfall is thought to be the normal distribution according to the central limit theorem, see Waylen et al. (1996) and Abdullah and Al-Mazroui (1998). In real-life situations, the amount of rainfall is used to assess the amount of water available to meet the various demands of agriculture, industry, and other human activities. In climatology, the coefficient of variation is used to analyze the rainfall data (Ananthakrishnan and Soman 1989). Estimation of the coefficient of variation received interest in several researchers such as Ahmed (1994), Ahmed (1995), Ahmed et al. (2012), and Hayter (2015). Furthermore, the multiple comparisons of several treatments are corresponding to the simultaneous confidence intervals. The simultaneous confidence intervals for coefficients of variation of normal distributions are useful and important problem in real life. For instance, the simultaneous confidence intervals for coefficients of variation are used to compare rainfall variability in different regions. In demography, the differences of coefficients of variation of numbers of birth on various days can estimate using the simultaneous confidence intervals. Moreover, in clinical trials, the simultaneous confidence intervals of coefficients of variation are used to study the patient-level cost in different groups.

It is well known that if X is a sample taken from a normal distribution with mean μ and variance σ^2, then $Y = \exp(X)$ is a log-normal distribution with mean μ_Y and variance σ_Y^2. The log-normal distribution is used to describe the natural phenomena. Application of log-normal distribution can be found in biology, medicine, hydrology, economics, and environment. For example, the log-normal distribution is used to estimate annual maximum values of daily rainfall. Moreover, the log-normal distribution is used to analyze air pollution level. Air pollution is a risk factor for a number of pollution-related diseases such as heart disease and lung cancer. Air pollution is a mixture of gaseous pollutants and particulate matter (PM). PM2.5 is particle that has a diameter of less than 2.5 micrometers, while PM10 has a diameter of less than 10 micrometers. The World Health Organization sets 25 micrograms as World Health Organization's safe level, while Pollution Control Department sets 50 micrograms as Thailand's safe level. PM2.5 has been a common problem in North of Thailand. It mainly occurs from January to April, but peaks in March. Air quality monitoring stations in North of Thailand found levels higher than 50 micrograms. According to an air pollution report of Pollution Control Department, the PM2.5 level in North of Thailand reached peaked at more than 350 micrograms, a level considered very unhealthy. In statistic, the air quality can be described by the coefficient of variation (Zhao et al. 2019). Real data set from the PM2.5 level in regions of northern Thailand; see Section 4.

In many fields, testing the equality of coefficients of variation in several populations is of interest. As a result, the problem of testing the equality of coefficients of variation of several normal populations has been widely studied. For example, Miller and Karson (1977) proposed testing equality of coefficients of variation in two normal distributions. Doornbos and Dijkstra (1983) proposed a multi sample test for the equality of coefficients of variation in k normal populations. Gupta and Ma (1996) proposed bisection method to obtain the estimator of parameters for testing the equality of the coefficients of variation in k normal populations. Fung and Tsang (1998) presented parametric and nonparametric tests for the equality of coefficients of variation in k normal populations. Ahmed (2002) developed simultaneous estimation of coefficients of variation of normal distributions. Recently, Hayter and Kim (2015) studied small-sample tests for the equality of coefficients of variation of two normal distributions.

Simultaneous confidence intervals (SCIs) provide the confidence regions for multivariate parameter, comprising individual intervals for the separate components of the parameter. The SCIs is useful for multiple comparisons of treatments such as treatment means, treatment variances, and treatment coefficients of variation. For example, Thangjai et al. (2018) constructed the SCIs for all differences of means of two-parameter exponential distributions. In addition, Thangjai and Niwitpong

(2020) proposed the SCIs for all differences of coefficients of variation of two-parameter exponential distributions. In this paper, we proposed three approaches to construct the SCIs for all differences of coefficients of variation from k normal populations. Firstly, the concepts of generalized pivotal quantity (GPQ) and generalized confidence interval (GCI) were introduced by Weerahandi (1993). Several researchers have used the concept of the GCI approach for constructing the confidence interval for coefficient of variation of normal distribution, i.e., see Tian (2005) and Thangjai et al. (2018). Secondly, the concept of MOVER approach, was introduced by Zou and Donner (2008) and Zou et al. (2009), was proposed to construct the confidence interval for the sum of two parameters. Moreover, the confidence interval for the difference parameters of two populations based on MOVER approach was proposed by Donner and Zou (2010). Recently, the MOVER approach has been successfully used to construct confidence interval, such as see Zou et al. (2009), Donner and Zou (2010), Suwan and Niwitpong (2013), Niwitpong (2015), Wongkhao et al. (2015), Niwitpong and Wongkhao (2016), Sangnawakij et al. (2015), and Sangnawakij and Niwitpong (2017). Finally, the computational approach was provided by Pal et al. (2007). The computational approach based on simulation and numerical computation uses the maximum likelihood estimates (MLEs). The computational approach was used to test equality of several populations; see Gokpinar et al. (2013), Jafari and Abdollahnezhad (2015), and Gokpinar and Gokpinar (2015). The daily PM2.5 levels at different areas are important aspects of air quality. The coefficients of variation of the PM2.5 data in different areas are different values. The coefficients of variation in the different areas are used to compare the PM2.5 level variability. To our knowledge there is no previous research paper on SCIs for all differences of coefficients of variation from k normal populations. Therefore, this paper will propose the GCI approach, the MOVER approach, and the computational approach for constructing the SCIs for all differences of coefficients of variation of PM2.5 levels in different areas of northern Thailand.

2. Simultaneous Confidence Intervals

Assume initially a single sample $X = (X_1, X_2, \ldots, X_n)$ from the normal distribution with mean μ and variance σ^2. The coefficient of variation is $\theta = \sigma / \mu$.

Let n and k be sample size and sample case, respectively. For k sample case, let X_{ij} be random sample from normal distribution based on the i^{th} sample. The coefficient of variation based on the i^{th} sample is defined by

$$\theta_i = \frac{\sigma_i}{\mu_i}. \tag{1}$$

Let $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij} / n_i$ and $S_i^2 = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 / (n_i - 1)$ be sample mean and sample variance for normal data for the i^{th} sample and let $\bar{x}_i$ and s_i^2 be observed sample mean and observed sample variance, respectively. The maximum likelihood estimator (MLE) of θ_i is a biased estimator. It is defined by

$$\hat{\theta}_i = \frac{\hat{\sigma}_i}{\hat{\mu}_i} = \frac{S_i}{\bar{X}_i}. \tag{2}$$

Ahmed (2002) proposed the shrinkage estimators for θ_i. In this paper, the Stein-type shrinkage estimator is used for θ_i because the Stein-type shrinkage estimator is the best estimator. The shrinkage estimator is defined as

$$\hat{\theta}_i^A = \hat{\theta}_i - (k-3)\Lambda^{-1}\left(\hat{\theta}_i - \frac{\sum_{i=1}^{k} n_i\hat{\theta}_i}{n}\right), \tag{3}$$

where $\Lambda = n \cdot \left(\hat{\theta}_i - \frac{\sum_{i=1}^{k} n_i\hat{\theta}_i}{n}\right)^2 \cdot \left(\frac{n_i}{n\hat{\tau}_i^2}\right)$, $\hat{\tau}_i^2 = \frac{1}{2}\left(\frac{\sum_{i=1}^{k} n_i\hat{\theta}_i}{n}\right)^2 + \left(\frac{\sum_{i=1}^{k} n_i\hat{\theta}_i}{n}\right)^4$, and $n = n_1 + n_2 + \ldots + n_k$.

According to Feltz and Miller (1996) and Tian (2005), the asymptotic variance of $\hat{\theta}_i$ is defined by

$$Var(\hat{\theta}_i) = \frac{\theta_i^2(0.5+\theta_i^2)}{n_i - 1}. \tag{4}$$

The interest of this paper is constructing SCIs for all differences of coefficients of variation $\theta_i - \theta_l$, where $i, l = 1, 2, \ldots, k$ and $i \neq l$.

2.1. Generalized confidence interval approach

Definition 1 Let $X = (X_1, X_2, \ldots, X_n)$ be the random sample from a distribution $F_X(x;\theta,\delta)$, where θ is a scalar parameter of interest and δ is a vector of nuisance parameters. Let $x = (x_1, x_2, \ldots, x_n)$ be the observed value of $X = (X_1, X_2, \ldots, X_n)$. The random quantity $R(X;x,\theta,\delta)$ is called the generalized pivotal quantity if it has the following two properties; see Weerahandi (1993):

(i) The distribution of $R(X;x,\theta,\delta)$, $X = x$, is free of all unknown parameters.

(ii) The observed value of $R(X;x,\theta,\delta)$, $X = x$, is the quantity.

Let $R(\alpha)$ be the $(\alpha)^{\text{th}}$ quantile of $R(X;x,\theta,\delta)$. Therefore, $[R(\alpha/2), R(1-\alpha/2)]$ becomes the $100(1-\alpha)\%$ two-sided generalized confidence interval for parameter of interest. It is well known that $\bar{X}_i$ and S_i^2 are independent and

$$\bar{X}_i \sim N\left(\mu_i, \frac{\sigma_i^2}{n_i}\right) \text{ and } \frac{(n_i-1)S_i^2}{\sigma_i^2} \sim \chi^2_{n_i-1}, \tag{5}$$

where $\chi^2_{n_i-1}$ denotes the chi-squared distribution with $n_i - 1$ degrees of freedom. The generalized pivotal quantity of μ_i based on the i^{th} sample is defined by

$$R_{\mu_i} = \bar{x}_i - \frac{Z_i}{\sqrt{U_i}}\sqrt{\frac{(n_i-1)s_i^2}{n_i}}, \tag{6}$$

where Z_i denotes the standard normal distribution and U_i denotes the chi-squared distribution with $n_i - 1$ degrees of freedom. Furthermore, the generalized pivotal quantity of σ_i^2 based on the i^{th} sample is defined by

$$R_{\sigma_i^2} = \frac{(n_i-1)s_i^2}{V_i}, \tag{7}$$

where V_i denotes the chi-squared distribution with $n_i - 1$ degrees of freedom. Therefore, the generalized pivotal quantity of θ_i based on the i^{th} sample is defined by

$$R_{\theta_i} = \frac{\sqrt{R_{\sigma_i^2}}}{R_{\mu_i}}, \tag{8}$$

where R_{μ_i} and $R_{\sigma_i^2}$ are defined in (6) and (7), respectively. For $i,l = 1,2,\ldots,k$ and $i \neq l$, the generalized pivotal quantity of $\theta_{ii} = \theta_i - \theta_l$ is defined by

$$R_{\theta_{il}} = R_{\theta_i} - R_{\theta_l} = \frac{\sqrt{R_{\sigma_i^2}}}{R_{\mu_i}} - \frac{\sqrt{R_{\sigma_l^2}}}{R_{\mu_l}}. \tag{9}$$

Therefore, the $100(1-\alpha)\%$ two-sided simultaneous confidence intervals for all differences of coefficients of variation θ_{il} based on GCI approach are defined by

$$SCI_{il(GCI)} = \left[R_{\theta_{il}}(\alpha/2), R_{\theta_{il}}(1-\alpha/2)\right], \tag{10}$$

where $R_{\theta_{il}}(\alpha/2)$ and $R_{\theta_{il}}(1-\alpha/2)$ denote the $(\alpha/2)^{\text{th}}$ and the $(1-\alpha/2)$ -th quantiles of $R_{\theta_{il}}$, respectively. The values of $R_{\theta_{il}}(\alpha/2)$ and $R_{\theta_{il}}(1-\alpha/2)$ are easily obtained by Monte Carlo method.

The values of $R_{\theta_{il}}(\alpha/2)$ and $R_{\theta_{il}}(1-\alpha/2)$ can be estimated using Monte Carlo procedure as follows:

Algorithm 1

Step 1 Calculate the values of $\bar{x}_i$, $\bar{x}_l$, s_i^2, and s_l^2, the observed values of $\bar{X}_i$, $\bar{X}_l$, S_i^2, and S_l^2, respectively, where $i,l = 1,2,\ldots,k$ and $i \neq l$.

Step 2 Generate Z_i and Z_l from the standard normal distributions and generate U_i and U_l from the chi-squared distributions with $n_i - 1$ and $n_l - 1$ degrees of freedom. Compute R_{μ_i} and R_{μ_l}, where $i,l = 1,2,\ldots,k$ and $i \neq l$.

Step 3 Generate V_i and V_l from the chi-squared distributions with $n_i - 1$ and $n_l - 1$ degrees of freedom. Compute $R_{\sigma_i^2}$ and $R_{\sigma_l^2}$.

Step 4 Compute R_{θ_i} , R_{θ_l} , and $R_{\theta_{il}}$.

Step 5 Repeat Step 1 to Step 4 a large number of times (say, $m = 1000$ times). From these m values, obtained the $1000 R_{\theta_{il}}$ and its $(\alpha/2)^{\text{th}}$ and $(1-\alpha/2)^{\text{th}}$ quantiles as $R_{\theta_{il}}(\alpha/2)$ and $R_{\theta_{il}}(1-\alpha/2)$.

Theorem 1 *Let X_{ij} be the random sample from the normal distribution with mean μ_i and variance σ_i^2, where $i = 1,2,\ldots,k$ and $j = 1,2,\ldots,n_i$. The coefficients of variation are $\theta_i = \sigma_i/\mu_i$ and $\theta_l = \sigma_l/\mu_l$, where $i,l = 1,2,\ldots,k$ and $i \neq l$. Let $\hat{\theta}_i = S_i/\bar{X}_i$ and $\hat{\theta}_l = S_l/\bar{X}_l$ be the estimators of θ_i and θ_l, respectively. Also, let θ_{il} be the difference between θ_i and θ_l. Let n_i be the sample size based on the i^{th} sample. Assume that the ratio $n_i/N \to r_i \in (0,1)$ as $N \to \infty$, where $N = n_1 + n_2 + \ldots + n_k$. Then the joint coverage probability is*

$$P\left(R_{\theta_{il}}(\alpha/2) \leq \theta_{il} \leq R_{\theta_{il}}(1-\alpha/2), \forall i \neq l\right) \to 1-\alpha. \tag{11}$$

Proof: Let $\hat{\theta}_{il}$ be the difference between $\hat{\theta}_i$ and $\hat{\theta}_l$, where $i,l = 1,2,\ldots,k$ and $i \neq l$. The mean and variance of $R_{\theta_{il}}$ are

$$E(R_{\theta_{il}}) = \theta_i - \theta_l = \theta_{il} \text{ and } Var(R_{\theta_{il}}) = Var(\hat{\theta}_i) + Var(\hat{\theta}_l) = Var(\hat{\theta}_{il}).$$

According to Zhang (2014) and Kharrati-Kopaei and Eftekhar (2017), then

$$P\left(\theta_{il(\alpha/2)} \le \theta_{il} \le \theta_{il(1-\alpha/2)}\right) = P\left(\theta_{il} \in (E(R_{\theta_{il}}) \pm t_{1-\alpha/2}\sqrt{Var(R_{\theta_{il}})}), \forall i \neq l\right)$$

$$= P\left(\theta_{il} \in (\hat{\theta}_{il} \pm t_{1-\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_{il})}), \forall i \neq l\right),$$

where $t_{1-\alpha/2}$ denotes the $(1-\alpha/2)^{\text{th}}$ quantile of the Student's t-distribution with $n_i - 1$ degrees of freedom of

$$Q_n = \max_{i \neq l}\left|\frac{R_{\theta_{il}} - E(R_{\theta_{il}})}{\sqrt{\widehat{Var}(R_{\theta_{il}})}}\right| = \max_{i \neq l}\left|\frac{\hat{\theta}_{il} - \theta_{il}}{\sqrt{\widehat{Var}(\hat{\theta}_{il})}}\right|.$$

Therefore,

$$P\left(\theta_{il} \in (\hat{\theta}_{il} \pm t_{1-\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_{il})}), \forall i \neq l\right) \to 1-\alpha,$$

implies that

$$P\left(R_{\theta_{il}}(\alpha/2) \le \theta_{il} \le R_{\theta_{il}}(1-\alpha/2), \forall i \neq l\right) \to 1-\alpha.$$

Hence, Theorem 1 is proved.

From Equation (3), using the generalized pivotal quantities of μ_i and σ_i^2 in (5) and (6), the $100(1-\alpha)\%$ two-sided simultaneous confidence intervals for all differences of coefficients of variation θ_{il} based on GCI approach using the shrinkage estimator are similar to (10) defined by

$$SCI_{il(A.GCI)} = [R^A_{\theta_{il}}(\alpha/2), R^A_{\theta_{il}}(1-\alpha/2)], \tag{12}$$

where $R^A_{\theta_{il}}(\alpha/2)$ and $R^A_{\theta_{il}}(1-\alpha/2)$ denote the $(\alpha/2)^{\text{th}}$ and the $(1-\alpha/2)^{\text{th}}$ quantiles of $R^A_{\theta_{il}}$, respectively.

2.2. Method of variance estimates recovery approach

According to Donner and Zou (2010), the $100(1-\alpha)\%$ two-sided confidence interval for coefficient of variation of normal distribution is given by

$$l_i = \left[\bar{x}_i - \sqrt{\max\{0, \bar{x}_i^2 + a_i c_i (a_i - 2)\}}\,\right] s_i / c_i \tag{13}$$

and

$$u_i = \left[\bar{x}_i + \sqrt{\max\{0, \bar{x}_i^2 + b_i c_i (b_i - 2)\}}\,\right] s_i / c_i, \tag{14}$$

where $a_i = \sqrt{(n_i - 1)/\chi^2_{1-\alpha/2, n_i - 1}}$, $b_i = \sqrt{(n_i - 1)/\chi^2_{\alpha/2, n_i - 1}}$, $c_i = \bar{x}_i^2 - z^2_{\alpha/2} s_i^2 / n_i$, and $i = 1,2,\ldots,k$.

For $i = 1,2,$ Donner and Zou (2010) introduced the MOVER approach to construct a $100(1-\alpha)\%$ two-sided confidence interval $[L_{12}, U_{12}]$ of $\theta_1 - \theta_2$, where θ_1 and θ_2 denote the parameters of interest and L_{12} and U_{12} denote the lower limit and the upper limit of the confidence interval. The $[l_i, u_i]$ contains the parameter values for θ_i, where $i = 1,2.$ The lower limit L_{12} is defined by

$$L_{12} = \hat{\theta}_1 - \hat{\theta}_2 - \sqrt{(\hat{\theta}_1 - l_1)^2 + (u_2 - \hat{\theta}_2)^2}. \tag{15}$$

The upper limit U_{12} is defined by

$$U_{12} = \hat{\theta}_1 - \hat{\theta}_2 + \sqrt{(u_1 - \hat{\theta}_1)^2 + (\hat{\theta}_2 - l_2)^2}. \tag{16}$$

For $i,l = 1,2,\ldots,k$ and $i \neq l$, the lower limit L_{il} and the upper limit U_{il} are defined by

$$L_{il} = \hat{\theta}_i - \hat{\theta}_l - \sqrt{(\hat{\theta}_i - l_i)^2 + (u_l - \hat{\theta}_l)^2} \tag{17}$$

and

$$U_{il} = \hat{\theta}_i - \hat{\theta}_l + \sqrt{(u_i - \hat{\theta}_i)^2 + (\hat{\theta}_l - l_l)^2}, \tag{18}$$

where $\hat{\theta}_p = s_p / \bar{x}_p$, $l_p = \left[\bar{x}_p - \sqrt{\max\{0, \bar{x}_p^2 + a_p c_p (a_p - 2)\}}\right] s_p / c_p$,

$u_p = \left[\bar{x}_p + \sqrt{\max\{0, \bar{x}_p^2 + b_p c_p (b_p - 2)\}}\right] s_p / c_p$, $a_p = \sqrt{(n_p - 1) / \chi^2_{1-\alpha/2, n_p - 1}}$, $b_p = \sqrt{(n_p - 1) / \chi^2_{\alpha/2, n_p - 1}}$,

and $c_p = \bar{x}_p^2 - z_{\alpha/2}^2 s_p^2 / n_p$.

Therefore, the $100(1-\alpha)\%$ two-sided simultaneous confidence intervals for all differences of coefficients of variation θ_{il} based on MOVER approach are defined by

$$SCI_{il(MOVER)} = \left[\hat{\theta}_i - \hat{\theta}_l - \sqrt{(\hat{\theta}_i - l_i)^2 + (u_l - \hat{\theta}_l)^2}, \hat{\theta}_i - \hat{\theta}_l + \sqrt{(u_i - \hat{\theta}_i)^2 + (\hat{\theta}_l - l_l)^2}\right]. \tag{19}$$

Theorem 2 *Suppose that $X_{ij} \sim N(\mu_i, \sigma_i^2)$, where $i = 1,2,\ldots,k$ and $j = 1,2,\ldots,n_i$. Let θ_i and $\hat{\theta}_i - \hat{\theta}_l - \sqrt{(\hat{\theta}_i - l_i)^2 + (u_l - \hat{\theta}_l)^2}, \hat{\theta}_i - \hat{\theta}_l + \sqrt{(u_i - \hat{\theta}_i)^2 + (\hat{\theta}_l - l_l)^2}$ be the coefficients of variation based on the i^{th} sample and l^{th} sample, respectively. And let $\hat{\theta}_i$ and $\hat{\theta}_l$ be the estimators of θ_i and θ_l, respectively. Let θ_{il} be a difference between θ_i and θ_l, where $i,l = 1,2,\ldots,k$ and $i \neq l$. Let L_{il} and U_{il} be the lower limit and the upper limit of the confidence interval for θ_{il}. Then the joint coverage probability is*

$$P\left(L_{il} \leq \theta_{il} \leq U_{il}, \forall i \neq l\right) \to 1 - \alpha. \tag{20}$$

Proof: For $i,l = 1,2,\ldots,k$ and $i \neq l$, the lower limit L_{il} and the upper limit U_{il} of the confidence interval for $\theta_{il} = \theta_i - \theta_l$ are obtained by

$$L_{il} = \hat{\theta}_i - \hat{\theta}_l - \sqrt{(\hat{\theta}_i - l_i)^2 + (u_l - \hat{\theta}_l)^2} = \hat{\theta}_{il} - \sqrt{(\hat{\theta}_i - l_i)^2 + (u_l - \hat{\theta}_l)^2}$$

and

$$U_{il} = \hat{\theta}_i - \hat{\theta}_l + \sqrt{(u_i - \hat{\theta}_i)^2 + (\hat{\theta}_l - l_l)^2} = \hat{\theta}_{il} + \sqrt{(u_i - \hat{\theta}_i)^2 + (\hat{\theta}_l - l_l)^2}.$$

The variance estimates for $\hat{\theta}_i$ at $\theta_i = l_i$ and $\hat{\theta}_l$ at $\theta_l = l_l$ are

$$\widehat{Var}(\hat{\theta}_i) = \frac{(\hat{\theta}_i - l_i)^2}{z_{\alpha/2}^2} \text{ and } \widehat{Var}(\hat{\theta}_l) = \frac{(\hat{\theta}_l - l_l)^2}{z_{\alpha/2}^2},$$

where $z_{\alpha/2}$ denotes the $(\alpha/2)^{\text{th}}$ quantile of the standard normal distribution. The variance estimates for $\hat{\theta}_i$ at $\theta_i = u_i$ and $\hat{\theta}_l$ at $\theta_l = u_l$ are

$$\widehat{Var}(\hat{\theta}_i) = \frac{(u_i - \hat{\theta}_i)^2}{z_{\alpha/2}^2} \text{ and } \widehat{Var}(\hat{\theta}_l) = \frac{(u_l - \hat{\theta}_l)^2}{z_{\alpha/2}^2}.$$

Hence, the lower limit and the upper limit are

$$L_{il} = \hat{\theta}_{il} - z_{\alpha/2}\sqrt{\frac{(\hat{\theta}_i - l_i)^2}{z_{\alpha/2}^2} + \frac{(u_l - \hat{\theta}_l)^2}{z_{\alpha/2}^2}} = \hat{\theta}_{il} - z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_i) + \widehat{Var}(\hat{\theta}_l)}$$

and

$$U_{il} = \hat{\theta}_{il} + z_{\alpha/2}\sqrt{\frac{(u_i - \hat{\theta}_i)^2}{z_{\alpha/2}^2} + \frac{(\hat{\theta}_l - l_l)^2}{z_{\alpha/2}^2}} = \hat{\theta}_{il} + z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_i) + \widehat{Var}(\hat{\theta}_l)}.$$

Following Zhang (2014) and Kharrati-Kopaei and Eftekhar (2017), then

$$\begin{aligned} P(L_{il} \le \theta_{il} \le U_{il}) &= P\left(\hat{\theta}_{il} - z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_i) + \widehat{Var}(\hat{\theta}_l)} \le \theta_{il} \le \hat{\theta}_{il} + z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_i) + \widehat{Var}(\hat{\theta}_l)}\right) \\ &= P\left(\theta_{il} \in (\hat{\theta}_{il} \pm z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_i) + \widehat{Var}(\hat{\theta}_l)}), \forall i \ne l\right), \end{aligned}$$

where $z_{\alpha/2}$ denotes the $(\alpha/2)^{\text{th}}$ quantile of the standard normal distribution of

$$Q_n' = \max_{i \ne l}\left|\frac{\hat{\theta}_{il} - \theta_{il}}{\sqrt{\widehat{Var}(\hat{\theta}_i) + \widehat{Var}(\hat{\theta}_l)}}\right|.$$

Then

$$P\left(\theta_{il} \in (\hat{\theta}_{il} \pm z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_i) + \widehat{Var}(\hat{\theta}_l)}), \forall i \ne l\right) \to 1 - \alpha,$$

implies that

$$P\left(L_{il} \le \theta_{il} \le U_{il}, \forall i \ne l\right) \to 1 - \alpha.$$

Hence, Theorem 2 is proved.

In addition, the $100(1-\alpha)\%$ two-sided simultaneous confidence intervals for all differences of coefficients of variation θ_{il} based on MOVER approach using the shrinkage estimator are defined by

$$SCI_{il(A.MOVER)} = \left[\hat{\theta}_i^A - \hat{\theta}_l^A - \sqrt{(\hat{\theta}_i^A - l_i)^2 + (u_l - \hat{\theta}_l^A)^2}, \hat{\theta}_i^A - \hat{\theta}_l^A + \sqrt{(u_i - \hat{\theta}_i^A)^2 + (\hat{\theta}_l^A - l_l)^2}\right]. \quad (21)$$

2.3. Computational approach

Again, let $\theta_i = \sigma_i / \mu_i$ be the coefficient of variation of normal distribution, where $i = 1, 2, \ldots, k$. According to Doornbos and Dijkstra (1983) and Gokpinar and Gokpinar (2015), the restricted maximum likelihood estimator (RML) of θ_i is given by

$$\hat{\theta}_{i(RML)} = \sqrt{\frac{(1 + 2(S_i / \bar{X}_i)^2)^2 - 1}{4(1 + (S_i / \bar{X}_i)^2)}}. \quad (22)$$

The RML of μ_i is given by

$$\hat{\mu}_{i(RML)} = \frac{2(1 + (S_i / \bar{X}_i)^2)\bar{X}_i}{1 + \sqrt{1 + 4(1 + (S_i / \bar{X}_i)^2)\hat{\theta}_{i(RML)}^2}}. \quad (23)$$

Let $X_{ij(RML)}$ be an artificial sample from the normal distribution with mean $\hat{\mu}_{i(RML)}$ and variance $\hat{\sigma}_{i(RML)}^2 = \hat{\mu}_{i(RML)}^2 \hat{\theta}_{i(RML)}^2$, where $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_i$. Let $\bar{X}_{i(RML)}$ and $S_{i(RML)}^2$ be sample mean and sample variance for normal data for the i^{th} artificial sample and let $\bar{x}_{i(RML)}$ and $s^2_{i(RML)}$ be observed values of $\bar{X}_{i(RML)}$ and $S^2_{i(RML)}$, respectively. The difference of coefficient of variation estimators based on the artificial sample is defined by

$$\hat{\theta}_{il(RML)} = \hat{\theta}_{i(RML)} - \hat{\theta}_{l(RML)} = \frac{S_{i(RML)}}{\bar{X}_{i(RML)}} - \frac{S_{l(RML)}}{\bar{X}_{l(RML)}}, \tag{24}$$

where $i \neq l$. Therefore, the $100(1-\alpha)\%$ two-sided simultaneous confidence intervals for all differences of coefficients of variation θ_{il} based on computational approach are defined by

$$SCI_{il(CA)} = \left[\hat{\theta}_{il(RML),(\alpha/2)}, \hat{\theta}_{il(RML),(1-\alpha/2)}\right], \tag{25}$$

where $\hat{\theta}_{il(RML),(\alpha/2)}$ and $\hat{\theta}_{il(RML),(1-\alpha/2)}$ denote the $(\alpha/2)^{\text{th}}$ and the $(1-\alpha/2)^{\text{th}}$ quantiles of $\hat{\theta}_{il(RML)}$, respectively. The computational approach is presented in Algorithm 2 and bellow:

Algorithm 2

Step 1. Obtain the MLE of the parameters as $\hat{\mu}_i = \bar{X}_i$ and $\hat{\sigma}_i = S_i$. Then $\hat{\theta}_i - \hat{\theta}_l = \frac{S_i}{\bar{X}_i} - \frac{S_l}{\bar{X}_l}$, where $i = 1, 2, \ldots, k$ and $i \neq l$.

Step 2. Calculate the value of $\hat{\theta}_{i(RML)}$ as given by (22) and calculate the value of $\hat{\mu}_{i(RML)}$ as given by (23).

Step 3. Generate artificial sample $X_{ij(RML)}$ from $N(\hat{\mu}_{i(RML)}, \hat{\mu}^2_{i(RML)}\hat{\theta}^2_{i(RML)})$ a large number of times (say, $m = 1000$ times), where $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_i$. For each of these replicated samples, recalculate the MLE of $\theta_{il(RML)}$, where $i, l = 1, 2, \ldots, k$ and $i \neq l$. Let these recalculated MLE values of $\theta_{il(RML)}$ be $\hat{\theta}_{il(RML),1}, \hat{\theta}_{il(RML),2}, \ldots, \hat{\theta}_{il(RML),m}$.

Step 4. Let $\hat{\theta}_{il(RML),(1)} \leq \hat{\theta}_{il(RML),(2)} \leq \ldots \leq \hat{\theta}_{il(RML),(m)}$ be the ordered values of $\hat{\theta}_{il(RML),g}$, where $g = 1, 2, \ldots, m$, $i, l = 1, 2, \ldots, k$, and $i \neq l$.

Step 5. Find the lower bound is defined by $\hat{\theta}_{il(RML),((\alpha/2)m)}$ and find the upper bound is defined by $\hat{\theta}_{il(RML),((1-\alpha/2)m)}$.

Theorem 3 *Suppose that* $X_{ij} \sim N(\mu_i, \sigma_i^2)$ *and* $X_{ij(RML)} \sim N(\hat{\mu}_{i(RML)}, \hat{\sigma}^2_{i(RML)})$, *where* $i = 1, 2, \ldots, k$ *and* $j = 1, 2, \ldots, n_i$. *Let* $\hat{\theta}_{il(RML),(\alpha/2)}$ *and* $\hat{\theta}_{il(RML),(1-\alpha/2)}$ *be the lower limit and upper limit of the confidence interval for* $\hat{\theta}_{il(RML)}$. *Then*

$$P\left(\hat{\theta}_{il(RML),(\alpha/2)} \leq \hat{\theta}_{il(RML)} \leq \hat{\theta}_{il(RML),(1-\alpha/2)}, \forall i \neq l\right) \to 1-\alpha. \tag{26}$$

Proof: Let $\hat{\theta}_{il(RML)} = \hat{\theta}_{i(RML)} - \hat{\theta}_{l(RML)}$, where $i = 1, 2, \ldots, k$ and $i \neq l$. The mean and variance of $\hat{\theta}_{il} = \hat{\theta}_i - \hat{\theta}_l$ are

$$E(\hat{\theta}_{il}) = \hat{\theta}_{il(RML)} \text{ and } Var(\hat{\theta}_{il}) = Var(\hat{\theta}_i) + Var(\hat{\theta}_l).$$

Using Pal et al. (2007), Zhang (2014), and Kharrati-Kopaei and Eftekhar (2017), then

$$P(\hat{\theta}_{il(RML),(\alpha/2)} \le \hat{\theta}_{il(RML)} \le \hat{\theta}_{il(RML),(1-\alpha/2)}) = P\left(\hat{\theta}_{il(RML)} \in (\hat{\theta}_{il} \pm t_{1-\alpha/2}\sqrt{V\hat{a}r(\hat{\theta}_{il})}), \forall i \ne l\right),$$

where $t_{1-\alpha/2}$ denotes the $(1-\alpha/2)^{\text{th}}$ quantile of the Student's t-distribution with $n_i - 1$ degrees of freedom of

$$Q'' = \max_{i \ne l}\left|\frac{\hat{\theta}_{il} - \hat{\theta}_{il(RML)}}{\sqrt{\widehat{Var}(\hat{\theta}_{il})}}\right|.$$

Therefore,

$$P\left(\hat{\theta}_{il(RML)} \in (\hat{\theta}_{il} \pm t_{1-\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_{il})}), \forall i \ne l\right) \to 1-\alpha,$$

implies that

$$P\left(\hat{\theta}_{il(RML),(\alpha/2)} \le \hat{\theta}_{il(RML)} \le \hat{\theta}_{il(RML),(1-\alpha/2)}, \forall i \ne l\right) \to 1-\alpha.$$

Hence, Theorem 3 is proved.

By performing similar steps with the idea that the difference of coefficients of variation estimator based on the artificial sample using the shrinkage estimator is defined by

$$\hat{\theta}^{A}_{il(RML)} = \hat{\theta}^{A}_{i(RML)} - \hat{\theta}^{A}_{l(RML)}. \tag{27}$$

It is easy to see that the $100(1-\alpha)\%$ two-sided simultaneous confidence intervals for all differences of coefficients of variation θ_{il} based on computational approach using the shrinkage estimator are defined by

$$SCI_{il(A.CA)} = \left[\hat{\theta}^{A}_{il(RML),(\alpha/2)}, \hat{\theta}^{A}_{il(RML),(1-\alpha/2)}\right],, \tag{28}$$

where $\hat{\theta}^{A}_{il(RML),(\alpha/2)}$ and $\hat{\theta}^{A}_{il(RML),(1-\alpha/2)}$ denote the $(\alpha/2)^{\text{th}}$ and the $(1-\alpha/2)^{\text{th}}$ quantiles of $\hat{\theta}^{A}_{il(RML)}$, respectively.

3. Simulation Studies

In this section, simulation studies are carried out to evaluate the performance of the simultaneous confidence intervals based on GCI approach (SCI_{GCI} and $SCI_{A.GCI}$), MOVER approach (SCI_{MOVER} and $SCI_{A.MOVER}$), and computational approach (SCI_{CA} and $SCI_{A.CA}$) for differences of coefficients of variation of several normal distributions. The performance of these three approaches was evaluated through the coverage probabilities, average lengths, and standard errors of the simultaneous confidence intervals.

In simulations, four configuration factors are considered to evaluate the performance of the three simultaneous confidence interval approaches: (1) sample cases: k = 5; (2) population means: $\mu_1 = \mu_2 = \ldots = \mu_5 = 1$; (3) population coefficients of variation: $\theta_1, \theta_2, \ldots, \theta_5$; (4) sample sizes: $n_1, n_2, \ldots, n_5$. The specific combinations are given in the following table. The nominal confidence level was chosen to be 0.95.

The following algorithm was used to estimate the coverage probabilities of six simultaneous confidence intervals.

Algorithm 3

Step 1. Generate X_{ij}, a random sample of sample size n_i from normal population with parameters μ_i and σ_i^2, where $i = 1, 2, \ldots k$ and $j = 1, 2, \ldots, n_i$. Calculate $\bar{x}_i$ and s_i (the observed values of $\bar{X}_i$ and S_i).

Step 2. Construct two-sided simultaneous confidence intervals based on GCI approach ($SCI_{il(GCI)}$ and $SCI_{il(A.GCI)}$) and record whether or not all the values of θ_{il} fall in their corresponding simultaneous confidence intervals.

Step 3. Construct two-sided simultaneous confidence intervals based on MOVER approach ($SCI_{il(MOVER)}$ and $SCI_{il(A.MOVER)}$) and record whether or not all the values of θ_{il} fall in their corresponding simultaneous confidence intervals.

Step 4. Construct two-sided simultaneous confidence intervals based on computational approach ($SCI_{il(CA)}$ and $SCI_{il(A.CA)}$) and record whether or not all the values of θ_{il} fall in their corresponding simultaneous confidence intervals.

Step 5. Repeat Step 1 to Step 4 a large number of times, $M = 5000$. Then, the fraction of times that all θ_{il} are in their corresponding SCIs provides an estimate of the coverage probability.

The simulation results from $k = 5$ are presented in the Table 1. From Table 1, the results show that the coverage probabilities of SCI_{GCI} and $SCI_{A.GCI}$ are close to nominal confidence level 0.95. However, the average lengths and standard errors of $SCI_{A.GCI}$ are smaller than those of SCI_{GCI} for all cases. The coverage probabilities of SCI_{MOVER} performs satisfactorily for all cases. For $n_i \le 50$, it is seen that SCI_{GCI}, $SCI_{A.GCI}$, and SCI_{MOVER} have similar coverage probabilities but the average lengths of $SCI_{A.GCI}$ are smaller than the average lengths of SCI_{GCI} and SCI_{MOVER}. For $n_i > 50$, $SCI_{A.GCI}$ performs as well as SCI_{GCI} and SCI_{MOVER}. In addition, $SCI_{A.CA}$ is better than the other simultaneous confidence intervals in terms of average lengths when sample sizes are large ($n_i > 50$) and the coefficients of variation are same value. Moreover, Figures 1-4 present the line graphs of these simulation results.

4. Example

PM2.5 level data is given by Pollution Control Department. The PM2.5 level data from 1 March 2020 to 15 April 2020 are presented in Table 2. Area 1, area 2, area 3, area 4, and area 5 are Mae Moh district of Lampang, Mueang Chiang Rai district of Chiang Rai, Mueang Phayao district of Phayao, Mae Chaem district of Chiang Mai, and Mueang Chiang Mai district of Chiang Mai, respectively. The Shapiro-Wilk normality test indicated that the log-PM2.5 level data of five areas follow normal distributions with p-values 0.8978, 0.9448, 0.9345, 0.8291, and 0.9943 for area 1, area 2, area 3, area 4, and area 5, respectively. Table 3 presents the sample statistics of five areas. Using three approaches, the 95% simultaneous confidence intervals for the differences of coefficients of variation are presented in Table 4. The computational approach is the shortest lengths. Hence, the result confirms the simulation study in the previous section for five sample cases.

Table 1 The coverage probabilities (CP), average lengths (AL) and standard errors (s.e.) of 95% of two-sided simultaneous confidence intervals for all differences of coefficients of variation of normal distributions: 5 sample cases

$n(5)$	$\theta(5)$	SCI_{GCI}		$SCI_{A.GCI}$		SCI_{MOVER}		$SCI_{A.MOVER}$	
		CP	**AL (s.e.)**	**CP**	**AL (s.e.)**	**CP**	**AL (s.e.)**	**CP**	**AL (s.e.)**
10(5)	0.1(5)	0.9493	0.1694 (0.0074)	0.9485	0.1520 (0.0066)	0.9545	0.1751 (0.0074)	0.9880	0.1705 (0.0126)
	0.1(2),0.3,0.5(2)	0.9594	0.7603 (0.1268)	0.9469	0.6419 (0.1023)	0.9549	0.7150 (0.1125)	0.9235	0.8740 (0.1418)
20(5)	0.1(5)	0.9491	0.1019 (0.0031)	0.9485	0.0968 (0.0030)	0.9520	0.1038 (0.0030)	0.9886	0.1077 (0.0065)
	0.1(2),0.3,0.5(2)	0.9518	0.4126 (0.0575)	0.9466	0.3844 (0.0529)	0.9510	0.4081 (0.0562)	0.8975	0.7151 (0.1173)
10(2),20,30(2)	0.1(5)	0.9511	0.1226 (0.0099)	0.9494	0.1129 (0.0084)	0.9531	0.1248 (0.0101)	0.9859	0.1254 (0.0116)
	0.1(2),0.3,0.5(2)	0.9494	0.3521 (0.0343)	0.9459	0.3322 (0.0334)	0.9500	0.3539 (0.0335)	0.8910	0.7038 (0.1133)
30(5)	0.1(5)	0.9486	0.0794 (0.0020)	0.9477	0.0767 (0.0019)	0.9503	0.0804 (0.0019)	0.9883	0.0874 (0.0054)
	0.1(2),0.3,0.5(2)	0.9541	0.3166 (0.0423)	0.9508	0.3026 (0.0401)	0.9541	0.3151 (0.0418)	0.8795	0.6650 (0.1118)
50(5)	0.1(5)	0.9492	0.0593 (0.0012)	0.9494	0.0581 (0.0012)	0.9512	0.0597 (0.0011)	0.9891	0.0666 (0.0040)
	0.1(2),0.3,0.5(2)	0.9505	0.2326 (0.0299)	0.9476	0.2266 (0.0290)	0.9508	0.2322 (0.0297)	0.8405	0.6214 (0.1095)
30(2),50,100(2)	0.1(5)	0.9486	0.0611 (0.0038)	0.9481	0.0595 (0.0036)	0.9504	0.0615 (0.0038)	0.9864	0.0668 (0.0056)
	0.1(2),0.3,0.5(2)	0.9497	0.1805 (0.0160)	0.9492	0.1772 (0.0159)	0.9512	0.1809 (0.0159)	0.8000	0.6199 (0.1129)
100(5)	0.1(5)	0.9495	0.0408 (0.0006)	0.9494	0.0404 (0.0006)	0.9501	0.0409 (0.0005)	0.9889	0.0452 (0.0021)
	0.1(2),0.3,0.5(2)	0.9509	0.1596 (0.0200)	0.9499	0.1575 (0.0197)	0.9513	0.1593 (0.0199)	0.7808	0.5825 (0.1071)
200(5)	0.1(5)	0.9504	0.0284 (0.0004)	0.9493	0.0282 (0.0004)	0.9512	0.0284 (0.0003)	0.9890	0.0319 (0.0015)
	0.1(2),0.3,0.5(2)	0.9499	0.1108 (0.0137)	0.9487	0.1101 (0.0136)	0.9509	0.1107 (0.0137)	0.7115	0.5720 (0.1119)
500(5)	0.1(5)	0.9494	0.0178 (0.0002)	0.9493	0.0178 (0.0002)	0.9502	0.0178 (0.0001)	0.9886	0.0202 (0.0009)
	0.1(2),0.3,0.5(2)	0.9503	0.0694 (0.0085)	0.9497	0.0693 (0.0085)	0.9506	0.0694 (0.0085)	0.6244	0.5734 (0.1156)
100(2),200,500(2)	0.1(5)	0.9505	0.0301 (0.0022)	0.9502	0.0299 (0.0021)	0.9518	0.0302 (0.0022)	0.9853	0.0337 (0.0032)
	0.1(2),0.3,0.5(2)	0.9510	0.0818 (0.0062)	0.9512	0.0815 (0.0062)	0.9517	0.0818 (0.0061)	0.6265	0.6179 (0.1297)
500(3),1000(2)	0.1(5)	0.9483	0.0158 (0.0005)	0.9479	0.0158 (0.0005)	0.9495	0.0158 (0.0005)	0.9875	0.0178 (0.0011)
	0.1(2),0.3,0.5(2)	0.9491	0.0535 (0.0052)	0.9484	0.0535 (0.0052)	0.9494	0.0535 (0.0052)	0.5382	0.5654 (0.1108)
1000(5)	0.1(5)	0.9500	0.0126 (0.0001)	0.9500	0.0126 (0.0001)	0.9506	0.0126 (0.0000)	0.9887	0.0141 (0.0006)
	0.1(2),0.3,0.5(2)	0.9480	0.0489 (0.0060)	0.9477	0.0489 (0.0060)	0.9487	0.0489 (0.0060)	0.5831	0.5455 (0.1066)
200(2),500,1000(2)	0.1(5)	0.9471	0.0207 (0.0015)	0.9474	0.0206 (0.0015)	0.9479	0.0207 (0.0015)	0.9844	0.0231 (0.0022)
	0.1(2),0.3,0.5(2)	0.9497	0.0560 (0.0044)	0.9499	0.0559 (0.0044)	0.9506	0.0560 (0.0043)	0.5316	0.6036 (0.1256)

Table 1 (Continued)

$n(5)$	$\theta(5)$	SCI_{CA} CP	SCI_{CA} AL (s.e.)	$SCI_{A.CA}$ CP	$SCI_{A.CA}$ AL (s.e.)
10(5)	0.1(5)	0.9486	0.1290 (0.0055)	0.9012	0.0870 (0.0163)
	0.1(2),0.3,0.5(2)	0.9218	0.5720 (0.0921)	0.2914	0.3400 (0.0807)
20(5)	0.1(5)	0.9489	0.0897 (0.0027)	0.9341	0.0673 (0.0122)
	0.1(2),0.3,0.5(2)	0.9353	0.3663 (0.0507)	0.2123	0.2653 (0.0624)
10(2),20,30(2)	0.1(5)	0.9392	0.1006 (0.0065)	0.8884	0.0719 (0.0136)
	0.1(2),0.3,0.5(2)	0.9405	0.3160 (0.0338)	0.2239	0.2279 (0.0493)
30(5)	0.1(5)	0.9478	0.0730 (0.0018)	0.9433	0.0591 (0.0113)
	0.1(2),0.3,0.5(2)	0.9429	0.2936 (0.0390)	0.1796	0.2268 (0.0534)
50(5)	0.1(5)	0.9491	0.0564 (0.0011)	0.9507	0.0471 (0.0089)
	0.1(2),0.3,0.5(2)	0.9431	0.2227 (0.0286)	0.1386	0.1696 (0.0383)
30(2),50,100(2)	0.1(5)	0.9444	0.0575 (0.0033)	0.9351	0.0450 (0.0084)
	0.1(2),0.3,0.5(2)	0.9475	0.1746 (0.0160)	0.1200	0.1342 (0.0274)
100(5)	0.1(5)	0.9491	0.0398 (0.0006)	0.9582	0.0329 (0.0060)
	0.1(2),0.3,0.5(2)	0.9479	0.15621 (0.0196)	0.0941	0.1214 (0.0271)
200(5)	0.1(5)	0.9499	0.0281 (0.0004)	0.9626	0.0243 (0.0046)
	0.1(2),0.3,0.5(2)	0.9477	0.1096 (0.0136)	0.0655	0.0901 (0.0206)
500(5)	0.1(5)	0.9490	0.0177 (0.0002)	0.9632	0.0153 (0.0029)
	0.1(2),0.3,0.5(2)	0.9492	0.0691 (0.0085)	0.0340	0.0556 (0.0126)
100(2),200,500(2)	0.1(5)	0.9489	0.0296 (0.0021)	0.9536	0.0239 (0.0047)
	0.1(2),0.3,0.5(2)	0.9506	0.0811 (0.0062)	0.0597	0.0642 (0.0127)
500(3),1000(2)	0.1(5)	0.9472	0.0158 (0.0005)	0.9634	0.0129 (0.0023)
	0.1(2),0.3,0.5(2)	0.9495	0.0534 (0.0052)	0.0316	0.0412 (0.0084)
1000(5)	0.1(5)	0.9494	0.0125 (0.0001)	0.9643	0.0103 (0.0018)
	0.1(2),0.3,0.5(2)	0.9479	0.0488 (0.0060)	0.0215	0.0376 (0.0081)
200(2),500,1000(2)	0.1(5)	0.9465	0.0205 (0.0015)	0.9555	0.0166 (0.0032)
	0.1(2),0.3,0.5(2)	0.9496	0.0558 (0.0044)	0.0393	0.0452 (0.0090)

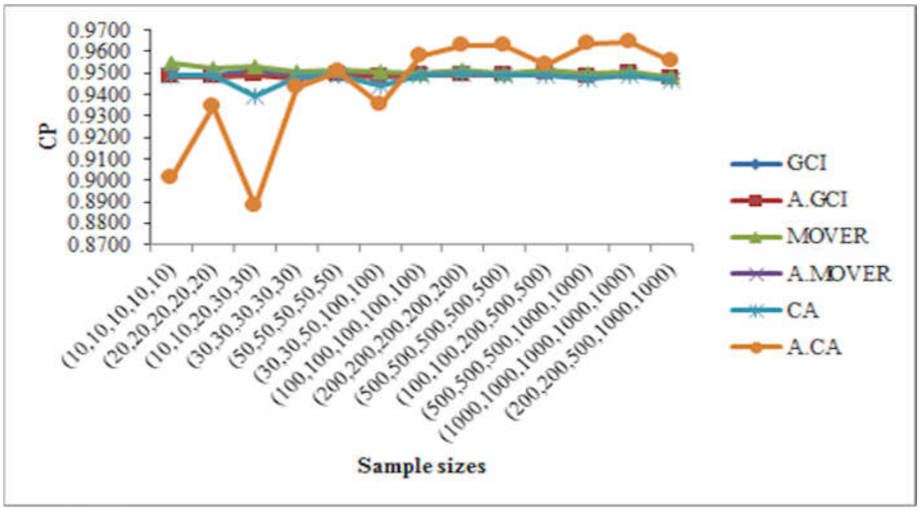

Figure 1 The CP of 95% of two-sided simultaneous confidence intervals for all differences of coefficients of variation of normal distributions: 5 sample cases and $(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5)=(0.1,0.1,0.1,0.1,0.1)$

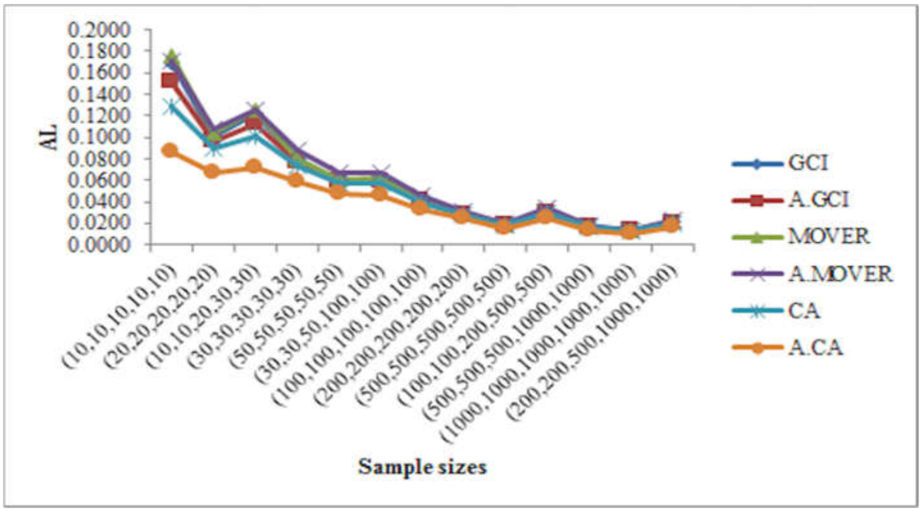

Figure 2 The AL of 95% of two-sided simultaneous confidence intervals for all differences of coefficients of variation of normal distributions: 5 sample cases and $(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5)=(0.1,0.1,0.1,0.1,0.1)$

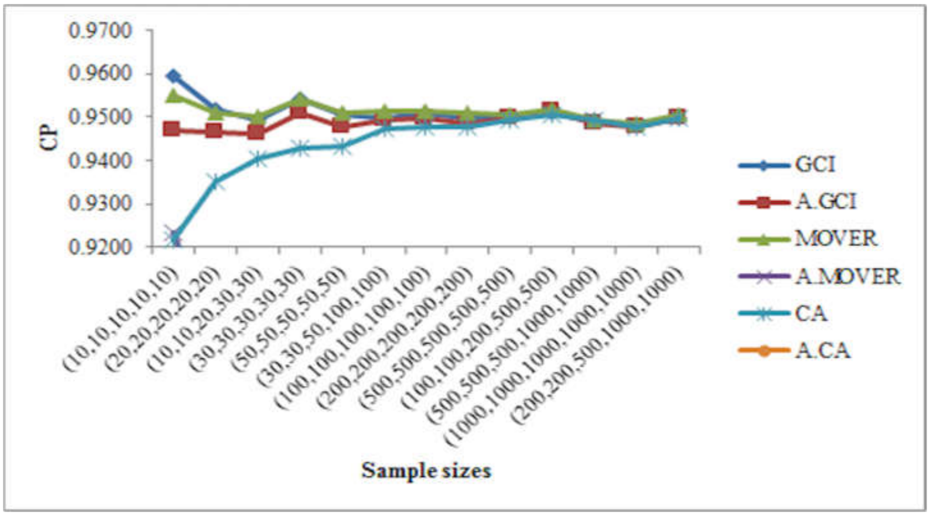

Figure 3 The CP of 95% of two-sided simultaneous confidence intervals for all differences of coefficients of variation of normal distributions: 5 sample cases and $(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5) = (0.1,0.1,0.3,0.5,0.5)$

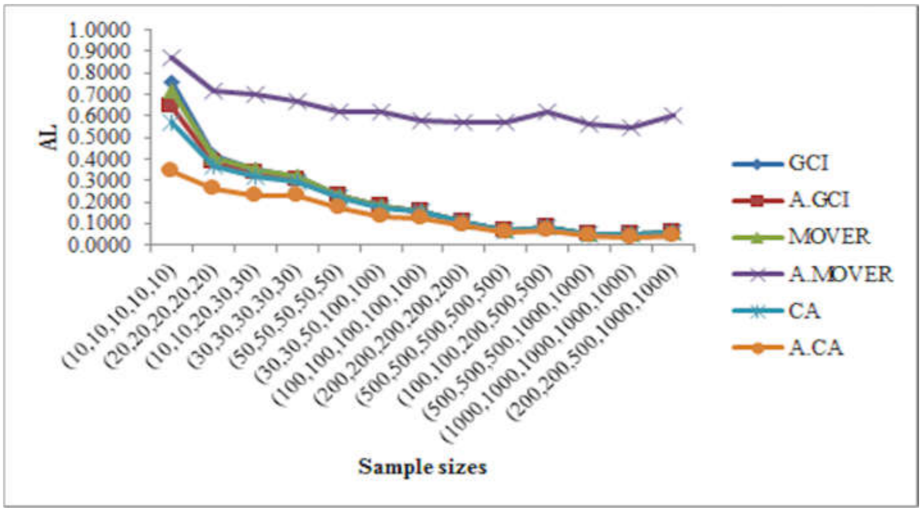

Figure 4 The AL of 95% of two-sided simultaneous confidence intervals for all differences of coefficients of variation of normal distributions: 5 sample cases and $(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5) = (0.1,0.1,0.3,0.5,0.5)$

Table 2 PM2.5 level data of five areas ($\mu g / m^3$)

Area 1		Area 2		Area 3		Area 4		Area 5	
46	48	65	67	63	63	67	58	48	40
42	35	53	52	60	46	50	41	34	31
31	26	43	40	37	32	32	30	22	20
30	36	62	74	41	51	33	43	31	43
42	49	73	115	69	131	50	52	45	59
52	57	105	85	112	95	36	65	41	36
80	81	134	212	182	246	77	104	75	97
71	42	251	169	185	56	126	81	90	57
44	46	66	94	52	61	45	44	69	53
56	53	129	150	102	78	60	58	48	52
45	56	91	84	64	92	54	51	64	189
59	69	101	100	118	98	52	43	145	73
67	59	109	111	91	79	56	58	141	110
53	88	132	175	95	141	65	107	80	68
95	104	182	149	137	135	96	76	63	62
92	103	119	112	117	115	74	59	66	83
84	76	124	113	108	92	75	80	117	58
65	68	181	141	92	89	81	125	84	
73	63	138	83	94	75	93	66		
62	52	100	104	85	60	80	79		
37	43	65		48	69	55			
66	48			82	57				
32	49			32	48				

Source: Pollution Control Department (http://aqmthai.com/aqi.php)

Table 3 Sample statistics

Sample statistics	Areas				
	Area 1	Area 2	Area 3	Area 4	Area 5
n_i	46	41	46	41	35
$\overline{y}_i$	58.1522	111.0488	88.5870	65.2927	68.4000
$s^2_{Y_i}$	383.0208	2164.2480	1829.5810	550.3622	1345.6590
$\overline{x}_i$	4.0079	4.6246	4.3819	4.1182	4.0999
$s^2_{X_i}$	0.1143	0.1806	0.2068	0.1246	0.2584
$\hat{\theta}_i$	0.0843	0.0919	0.1038	0.0857	0.1240

Table 4 The 95% simultaneous confidence intervals for all pairwise differences of coefficients of variation of log-PM2.5 level data

Comparison	CI_{GCI}		$CI_{A.GCI}$		CI_{MOVER}		$CI_{A.MOVER}$	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Area 2 / Area 1	-0.0187	0.0406	-0.0194	0.0356	-0.0200	0.0372	-0.0201	0.0349
Area 3 / Area 1	-0.0097	0.0510	-0.0086	0.0525	-0.0089	0.0503	-0.0126	0.0466
Area 4 / Area 1	-0.0261	0.0314	-0.0264	0.0291	-0.0255	0.0295	-0.0234	0.0294
Area 5 / Area 1	0.0080	0.0854	0.0059	0.0787	0.0070	0.0813	0.0028	0.0759
Area 3 / Area 2	-0.0214	0.0414	-0.0182	0.0425	-0.0196	0.0438	-0.0229	0.0414
Area 4 / Area 2	-0.0386	0.0238	-0.0354	0.0229	-0.0363	0.0231	-0.0325	0.0242
Area 5 / Area 2	-0.0065	0.0763	-0.0030	0.0737	-0.0032	0.0745	-0.0070	0.0706
Area 4 / Area 3	-0.0496	0.0120	-0.0479	0.0105	-0.0494	0.0120	-0.0434	0.0161
Area 5 / Area 3	-0.0179	0.0645	-0.0176	0.0611	-0.0161	0.0632	-0.0190	0.0654
Area 5 / Area 4	0.0043	0.0819	0.0046	0.0799	0.0042	0.0803	-0.0011	0.0720

Table 4 (Continued)

Comparison	CI_{CA}		$CI_{A.CA}$	
	Lower	Upper	Lower	Upper
Area 2 / Area 1	-0.0196	0.0340	-0.0133	0.0232
Area 3 / Area 1	-0.0075	0.0465	-0.0023	0.0227
Area 4 / Area 1	-0.0227	0.0256	-0.0060	0.0206
Area 5 / Area 1	0.0037	0.0727	0.0037	0.0487
Area 3 / Area 2	-0.0175	0.0408	-0.0087	0.0189
Area 4 / Area 2	-0.0339	0.0217	-0.0121	0.0173
Area 5 / Area 2	-0.0044	0.0652	-0.0019	0.0451
Area 4 / Area 3	-0.0472	0.0109	-0.0074	0.0017
Area 5 / Area 3	-0.0165	0.0563	-0.0035	0.0344
Area 5 / Area 4	0.0032	0.0717	0.0005	0.0382

From above examples, the numerical results from the data confirm the simulation results in the previous section in term of the average length. In simulation, the computational confidence interval has the shortest average lengths because the coverage probabilities provide less than the nominal confidence level 0.95. The coverage probability and length of each approach in two examples are computed by using only 1 sample, whereas the coverage probability and average length in the simulation are computed by repeating 5000 random samples. Moreover, the coverage probability and average length are considered to compare the confidence intervals. First, the coverage probability is considered that whether or not the 95% confidence interval have the coverage probability in a range of between $\left[c - z_{\alpha/2}\sqrt{\frac{c(1-c)}{M}}, c + z_{\alpha/2}\sqrt{\frac{c(1-c)}{M}}\right]$ = [0.9440, 0.9560], where c is the nominal confidence level and M is a number of simulation runs. Second, the average length is considered when the coverage probability is in the range of [0.9440, 0.9560]. Therefore, the computational approach is not recommended to construct the simultaneous confidence intervals for all differences of coefficients of variation when the sample sizes are small. The GCI and MOVER approaches are recommended to construct the simultaneous confidence intervals.

Furthermore, computation of confidence interval for all pairwise differences of coefficients of variation is useful to help assess the variability between groups of observations.

5. Discussion

In this paper, the generalized confidence interval (GCI) approach, the method of variance estimates recovery (MOVER) approach, and the computational approach were constructed for simultaneous confidence intervals for all differences of coefficients of variation. These three approaches are applied using two estimators: the biased estimator and the shrinkage estimator. The coverage probabilities, average lengths and standard errors of the proposed simultaneous confidence intervals were evaluated via Monte Carlo simulations. The simulation results indicated that the coverage probabilities of simultaneous confidence interval based on the GCI approach are close to nominal confidence level 0.95. The GCI approach can be considered as an alternative to estimate the simultaneous confidence intervals for differences of coefficients of variation. The MOVER approach performs satisfactorily: its coverage probability is close to the nominal confidence level 0.95. It was similar to the results of Donner and Zou (2010) and Niwitpong (2015). Moreover, the MOVER approach can be used to construct the simultaneous confidence intervals for all pairwise differences of coefficients of variation from several normal distributions for all sample sizes. Additionally, the computational approach can be used for estimating the simultaneous confidence intervals when sample sizes are large. For Stein-type shrinkage estimator, the coverage probabilities of $SCI_{A.CA}$ underestimate the nominal confidence level 0.95 when sample sizes are small and coefficients of variation are same value, whereas the coverage probabilities of $SCI_{A.CA}$ are close to nominal confidence level 0.95 when sample sizes are large and coefficients of variation are same value. Furthermore, the coverage probabilities of $SCI_{A.CA}$ underestimate the nominal confidence level 0.95 when coefficients of variation are different values for all cases.

Finally, the GCI approach is based on the concept of generalized pivotal quantities, whereas the computational approach is based on the maximum likelihood estimates. Both the GCI approach and the computational approach are based on simulated data. The MOVER approach uses exact formula. Therefore, the MOVER approach may be more useful than other approaches.

6. Conclusions

Based on simulation results of Ahmed (2002), we chose the Stein-type shrinkage estimator in Ahmed (2002) which was one of the best estimator. The performance of three approaches of the coefficient of variation is compared with the performance of three approaches using the shrinkage estimator of Ahmed (2002). From the simulation results, coverage probabilities of the coefficients of variation estimator are better than the coverage probabilities of the shrinkage estimator of Ahmed (2002) for some cases. In general, the average lengths of confidence intervals based on the shrinkage estimator are slightly shorter than the average lengths of proposed approaches, especially, $SCI_{A.CA}$ is recommended when the sample sizes are large $(n_i > 50)$ and coefficients of variation are same value.

Acknowledgements

This research was funded by King Mongkut's University of Technology North Bangkok. Grant No. KMUTNB-63-KNOW-032.

References

Abdel-Karim AH. Construction of simultaneous confidence intervals for ratios of means of lognormal distributions. Commun Stat - Simul Comput. 2015; 44: 271-283.

Abdullah MA, Al-Mazroui MA. Climatological study of the southwestern region of Saudi Arabia. I. Rainfall analysis. Climate Res. 1998; 9: 213-223.

Ahmed SE. Improved estimation of the coefficient of variation. J Appl Stat. 1994; 21: 565-573.

Ahmed SE. A pooling methodology for coefficient of variation. Sankhya. 1995; 57: 57-75.

Ahmed SE. Simultaneous estimation of coefficient of variation. J Stat Plan Infer. 2002; 104: 31-51.

Ahmed AS, Abd Ellah AH, Abou-Elheggag NA, Modhesh AA. Estimation of the coefficient of variation for non-normal model using progressive first-failure-consoring data. J Appl Stat. 2012; 39: 2741-2758.

Ananthakrishnan R, Soman K. Statistical distribution of daily rainfall and its association with the coefficient of variation of rainfall series. Int J Climatol. 1989; 9: 485-500.

Briggs AH, Nixon R, Dixon S, Thompson SG. Parametric Modelling of cost data: Some simulation evidence. Health Econ. 2005; 14: 421-428.

Donner A, Zou GY. Closed-form confidence intervals for function of the normal standard deviation. Stat Meth Med Res. 2010; 21: 347-359.

Doornbos R, Dijkstra JB. A multi sample test for the equality of coefficients of variation in normal populations. Commun Stat - Simul Comput. 1983; 12: 147-158.

Feltz CJ, Miller GE. An asymptotic test for the equality of coefficients of variation from k populations. Stat Med. 1996; 15: 647-658.

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

Gokpinar F, Gokpinar E. A computational approach for testing equality of coefficients of variation in k normal populations. Hacett J Math Stat. 2015; 44: 1197-1213.

Gokpinar EY, Polat E, Gokpinar F, Gunay S. A new computational approach for testing equality of inverse Gaussian means under heterogeneity. Hacett J Math Stat. 2013; 42: 581-590.

Gupta RC, Ma S. Testing the equality of the coefficients of variation in k normal populations. Commun Stat - Theory Methods. 1996; 25: 115-132.

Hayter AJ. Confidence bounds on the coefficient of variation of a normal distribution with applications to win-probabilities. J Stat Comput Simul. 2015; 85: 3778-3791.

Hayter AJ, Kim J. Small-sample tests for the equality of two normal cumulative probabilities, coefficients of variation, and Sharpe ratios. J Stat Theory Pract. 2015; 9: 23-36.

Jafari AA, Abdollahnezhad K. Inferences on the means of two log-normal distributions; A computational approach test. Commun Stat - Simul Comput. 2015; 44: 1659-1672.

Kharrati-Kopaei M, Eftekhar S. Simultaneous confidence intervals for comparing several inverse Gaussian means under heteroscedasticity. J Stat Comput Simul. 2017; 87: 777-790.

Miller EC, Karson MJ. Testing equality of two coefficients of variation. In: Proceedings of the Business and Economics Section; 1977. pp. 278-283.

Niwitpong S. Confidence intervals for the difference between coefficients of variation of normal distribution with bounded parameters. Far East J Math Sci. 2015; 98: 649-663.

Niwitpong S, Wongkhao A. Confidence intervals for the difference between inverse of normal means. Adv Appl Stat. 2016; 48(5): 337-347.

Pal N, Lim WK, Ling CH. A computational approach to statistical inferences. J Appl Prob Stat. 2007; 2: 13-35.

Sangnawakij P, Niwitpong S. Confidence intervals for coefficients of variation in two-parameter exponential distribution. Commun Stat - Simul Comput, 2017; 46: 6618-6630.

Sangnawakij P, Niwitpong S, Niwitpong S. Confidence intervals for the ratio of coefficients of variation of the gamma distributions. Lect Notes Comput Sci. 2015; 9376: 193-203.

Suwan S, Niwitpong S. Estimated variance ratio confidence interval of nonnormal distributions. Far East J Math Sci. 2013; 4: 339-350.

Thangjai W, Niwitpong S, Niwitpong S. Simultaneous confidence intervals for all differences of means of two-parameter exponential distributions. Stud Comput Intell. 2018; 760: 298-308.

Thangjai W, Niwitpong S. Simultaneous confidence intervals for all differences of coefficients of variation of two-parameter exponential distributions. Thail Stat. 2020; 18(2): 135-149.

Thangjai W, Niwitpong S, Niwitpong S. Adjusted generalized confidence intervals for the common coefficient of variation of several normal populations. Commun Stat - Simul Comput. 2020; 49(1): 194-206.

Tian L. Inferences on the common coefficient of variation. Stat Med. 2005; 24: 2213-2220.

Tsou TS. A robust score test for testing several coefficients of variation with unknown underlying distributions. Commun Stat - Theory Methods. 2009; 38: 1350-1360.

Waylen PR, Qusesada ME, Caviedes CN. Temporal and Spatial variability of annual precipitation in Costa Rica and the Southern Oscillation. Int J Climatol. 1996; 16: 173-193.

Weerahandi S. Generalized confidence intervals. J Am Stat Assoc. 1993; 88: 899-905.

Wongkhao A, Niwitpong S, Niwitpong S. Confidence intervals for the ratio of two independent coefficients of variation of normal distribution. Far East J Math Sci. 2015; 98: 741-757.

Zhang G. Simultaneous confidence intervals for several inverse Gaussian populations. Stat Prob Lett. 2014; 92: 125-131.

Zhao D, Chen H, Yu E, Luo T. PM2.5/PM10 ratios in eight economic regions and their relationship with meteorology in China. Adv Meteorol. 2019; 2019: 1-15.

Zou GY, Donner A. Construction of confidence limits about effect measures: A general approach. Stat Med. 2008; 27: 1693-1702.

Zou GY, Taleban J, Hao CY. Confidence interval estimation for lognormal data with application to health economics. Comput Stat Data Anal. 2009; 53: 3755-3764.