Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 8(12): 29713–29733.
DOI:10.3934/math.20231521
Received: 09 August 2023
Revised: 21 October 2023
Accepted: 30 October 2023
Published: 02 November 2023

Research article

Confidence intervals for the difference between coefficients of variation of zero-inflated gamma distributions

Hongping Guo[1,*], **Yuhang Qian**[1], **Yiran Zhu**[1], **Xinming Dai**[2] **and Xiao Wang**[3,*]

[1] School of Mathematics and Statistics, Hubei Normal University, Huangshi 435002, China
[2] Department of Applied Statistics, Social Science and Humanities, New York University, New York 10003, United States
[3] School of Mathematics and Statistics, Qingdao University, Qingdao 266071, China

* **Correspondence:** Email: guohongping@hbnu.edu.cn, xwang1020@qdu.edu.cn.

Abstract: The problem of constructing confidence intervals (CIs) for the difference between coefficients of variation of two zero-inflated gamma distributions was considered. As gamma distribution does not have closed form maximum likelihood estimators, the parameters of gamma distribution have to be estimated numerically. To this end, we proposed here four different generalized confidence intervals (GCIs) based on fiducial inference, Box-Cox transformation, parametric bootstrap and the method of variance of estimates recovery (MOVER). Performances of the four GCIs were evaluated and compared via extensive simulation. The simulation results showed that all four methods returned satisfactory results according to coverage probabilities, even for the setting of small sample sizes.

Keywords: confidence interval; zero-inflated gamma distribution; coefficient of variation; rainfall data
Mathematics Subject Classification: 62F25, 62P12

1. Introduction

Many disciplines, including meteorology, biomedicine and economics, frequently encounter zero-inflated nonnegative continuous data [1], in which zero-inflated refers to the part of the data consisting of zero values. For example, daily precipitation data often exhibits a mix of zero values representing dry days and a right-skewed positive part indicating rainy or snowy days. As one of the most common right-skewed models, gamma distribution has been widely used to analyze precipitation data [2, 3]. In such cases, zero values are neglected since gamma distribution deals with positive values only. As a consequence, the analytical results produced are usually biased [4].

In order to overcome the aforementioned problem, researchers started to analyze the original data that contains zero values by using zero-inflated right-skewed models. In such models, the number of zero values follows a binomial distribution, and the non-zero values follow right-skewed distributions (e.g., lognormal, gamma, Pareto and others). Hasan and Krishnamoorthy [5] proposed confidence intervals (CIs) for the mean and a percentile based on zero-inflated lognormal data. Wang and Li [6] proposed generalized CIs for the mean of zero-inflated Pareto distribution based on fiducial inference. Bugallo and Esteban et al. [7] predicted wild fires in Spain based on a zero-inflated negative binomial mixed model, and they used the parametric bootstrap method to estimate mean squared errors and constructed prediction intervals. More specifically, zero-inflated gamma distribution is one of the most commonly used models to analyze real precipitation data. In this regard, Vännman [8] derived the distribution of the estimated mean and the standardized statistic for the zero-inflated gamma distribution, and subsequently obtained a CI for the mean based on large sample approximation. Muralidharan and Kale [9] provided maximum likelihood estimates for the three parameters in the zero-inflated gamma distribution and derived the asymptotic distribution. They further obtained a 95% CI for the mean. Wang et al. [4] proposed CIs for the means of gamma distributions containing zero values based on fiducial inference, parametric bootstrap and the method of variance of estimates recovery (MOVER). Kaewprasert, Niwitpong S-A and Niwitpong S [10] proposed CIs for the common mean of several zero-inflated gamma distributions based on fiducial inference and the highest posterior density methods.

In practical applications, it is frequently necessary to compare the mean difference or ratio of means between two or more zero-inflated semicontinuous distributions, given the approximation of this distribution to many scientific datasets. For instance, Zou, Taleban and Huo [11] employed the MOVER method to construct CIs for the mean difference of two zero-inflated lognormal distributions, illustrating the methodology using economically healthy data. Ren, Liu and Pu [12] proposed an exact fiducial inference approach to construct simultaneous CIs for the mean difference of multiple zero-inflated gamma distributions. They applied this method to daily precipitation data across different periods or regions. Kaewprasert, Niwitpong S-A and Niwitpong S [13] constructed credible and highest posterior density intervals for the mean and the difference between the means of delta-gamma distributions. More recently, Kaewprasert, Niwitpong S-A and Niwitpong S [14] introduced eight Bayesian methods for constructing CIs for the ratio of means in zero-inflated gamma distributions. They applied these methods to analyze rainfall datasets from northern Thailand. However, when examining the fluctuations of zero-inflated semicontinuous data in various regions or periods, neither the mean difference nor the ratio of means can effectively capture these dynamics as they do not account for the influence of variance.

To potentially address this problem, the coefficient of variation (CV) is a classical and valuable statistical metric for comparing variabilities can be considered. CV is a relative measure of variability that quantifies the magnitude of a standard deviation in relation to the corresponding mean. It is defined as the ratio of the standard deviation to the mean of the distribution. Notably, the CV is a standardized and dimensionless measure, enabling comparisons of the magnitude of changes across different population data even if they possess disparate units of measurement, as highlighted by Albatineh, Kibria and Zogheib [15]. Previous researches, such as the study conducted by Ananthakrishnan and Soman [16], has explored the relationship between CVs of rainfall series and the statistical distribution of daily rainfall. Puggard, Niwitpong S-A and Niwitpong S [17] proposed CIs based on the generalized

CI, MOVER, large-sample, Bayesian credible interval and highest posterior density interval to estimate the common CV for several Birnbaum-Saunders distributions. Buntao and Niwitpong S-A [18] investigated the difference between the CV of two delta-lognormal distributions. To our knowledge, although there are many published works focused on constructing CIs for CVs of various situations, there is no work that mainly focuses on constructing CIs for the difference between two CVs of zero-inflated gamma distributions. The fact motivated us to develop appropriate methods in such cases. In addition, despite the CIs for the difference between two CVs of zero-inflated lognormal distribution analyze rainfall data, in special cases, such as the data analyzed in this article, zero-inflated gamma distribution is a more suitable model to be used.

In this study, our main focus is on the construction of CIs for the difference between two zero-inflated gamma CVs. We consider four fundamental methods for estimating gamma parameters and three methods for estimating the number of zero values in the dataset. Thereafter, 12 combinations of each method are constructed by MOVER. The four basic methods for estimating gamma parameters are the fiducial inference method proposed by Krishnamoorthy and Wang [19] (denoted as Fiducial-1), another fiducial inference method introduced by Ren, Liu and Pu [12] (denoted as Fiducial-2), the Box-Cox transformation method (denoted as BC) and the parametric bootstrap method (denoted as PB). Fiducial inference, initially proposed by Fisher [20], is widely employed for constructing CIs, requiring only sample information. The BC method, as suggested by Gao and Tian [21], was originally developed to construct CIs for the difference and ratio of two gamma means. In this work, we extend the concepts to construct CIs for the difference between two zero-inflated gamma CVs. The PB method, which has been demonstrated to perform favorably in constructing one-sided and two-sided CIs for the mean of zero-inflated normal distributions with small sample sizes, was compared against the percentile-t bootstrap interval, the maximum likelihood method for biasing and the likelihood ratio test method [22]. Binomial distribution is the model that has been proven to be appropriate to fit the number of zero values in the dataset. To this end, we introduce three different methods to estimate the binomial parameter. These methods are respectively the fiducial inference method proposed by Krishnamoorthy and Wang [19] (denoted as Fiducial-1), another fiducial inference method introduced by Ren, Liu and Pu [12] (denoted as Fiducial-2) and an exact estimation of the CI proposed by Wilson [23] (denoted as Wilson).

The MOVER method, first introduced by Zou and Donner [24], was employed to construct a universal CI for the difference of two parameters. Tang [25] applied the MOVER method to construct a general framework for CIs of difference or ratio effect parameters under stratified sampling. Li, Tang and Wong [26] studied the performance of various unconditional MOVER methods, which combined two independent CIs of a single Poisson rate into their ratio CIs. Donner and Zou [27] utilized the MOVER method to construct CIs as a function of the mean and standard deviation of a normal distribution, with simulations demonstrating that the CIs constructed using MOVER performed comparably to exact CIs.

The rest of the article is organized as follows. Section two presents the definition of the difference between two zero-inflated gamma CVs. Section three investigates the construction of CIs for the difference between two zero-inflated gamma CVs. Section four presents simulation experiments designed to compare the coverage probabilities and average lengths of the proposed methods. Section five applies the developed methodologies to analyze real monthly rainfall data from Beijing and Zhengzhou. Finally, in section six, the conclusions drawn from this study are summarized.

2. The coefficients of variation (CV) difference of zero-inflated gamma distributions

Suppose that $X_i = (X_{i1}, X_{i2}, \dots, X_{in_i}), i = 1, 2$ are the ith group of samples from the zero-inflated gamma distribution as follows:

$$X_i \sim g_i(x; \delta_i, a_i, b_i) = \delta_i I(x = 0) + (1 - \delta_i) I(x > 0) f(x; a_i, b_i), \tag{2.1}$$

where δ_i is the proportion of zero values, $I(\cdot)$ is the indicator function and $f(x; a_i, b_i)$ is the probability density function (pdf) of the gamma distribution with shape parameter a_i and scale parameter b_i given by

$$f(x; a_i, b_i) = \frac{\frac{1}{b_i^{a_i}} x^{a_i - 1} e^{-\frac{1}{b_i} x}}{\Gamma(a_i)}. \tag{2.2}$$

Let $N_{i(0)} = \sum_{j=1}^{n_i} I(X_{ij} = 0)$. It can be easily seen that $N_{i(0)} \sim bin(n_i, \delta_i)$ and $N_{i(1)} = n_i - N_{i(0)}$, and δ_i is the binomial parameter. Moreover, $n_{i(1)}$ and $n_{i(0)}$ denote the numbers of positive and zero observations, respectively.

The mean and variance of X_i from zero-inflated gamma distribution are $E(X_i) = (1 - \delta_i) a_i b_i$ and $Var(X_i) = (1 - \delta_i) a_i b_i^2$, respectively. Therefore, the coefficient of variation (CV) of X_i is denoted as

$$\eta_i = CV(X_i) = \frac{\sqrt{Var(X_i)}}{E(X_i)} = \sqrt{\frac{1}{(1 - \delta_i) a_i}}. \tag{2.3}$$

From (2.3), we can see that the CV of a zero-inflated gamma distributed random variable depends on the shape parameter a_i and the binomial parameter δ_i, but does not depend on the scale parameter b_i. In this paper, we aim to construct CIs for the CV difference of two independent zero-inflated gamma distributions as follows:

$$\gamma = \eta_1 - \eta_2 = \sqrt{\frac{1}{(1 - \delta_1) a_1}} - \sqrt{\frac{1}{(1 - \delta_2) a_2}}. \tag{2.4}$$

3. CIs for the CV difference of zero-inflated gamma distributions

3.1. Fiducial-1 method

Fiducial inference is widely used for constructing CIs because it only requires the sample information. Due to the fact that the cubic root of the sample from the gamma distribution approximately distributed as the normal distribution, Krishnamoorthy and Wang [19] proposed approximate fiducial quantities for gamma distribution based on the normal distribution. Suppose that the first part of $X_i (i = 1, 2)$ are nonzero samples: $X_{ij} > 0$ for $j = 1, 2, \dots, n_{i(1)}$; $X_{ij} = 0$ for $j = n_{i(1)} + 1, n_{i(1)} + 2, \dots, n_i$. Let $Y_{ij} = X_{ij}^{\frac{1}{3}}, X_{ij} > 0$ and $X_{ij} \sim gamma(a_i, b_i)$, then Y_{ij}'s are approximately normally distributed as $N(\mu_i, \sigma_i^2)$, where

$$\mu_i = (b_i a_i)^{\frac{1}{3}} (1 - \frac{1}{9 a_i}) \tag{3.1}$$

and

$$\sigma_i^2 = \frac{b_i^{\frac{1}{3}}}{9 a_i^{\frac{1}{3}}}. \tag{3.2}$$

Denote the sample mean and variance of Y_i are

$$\overline{Y}_i = \frac{1}{n_{i(1)}} \sum_{j=1}^{n_{i(1)}} Y_{ij},\ S_i^2 = \frac{1}{n_{i(1)} - 1} \sum_{j=1}^{n_{i(1)}} (Y_{ij} - \overline{Y}_i)^2. \tag{3.3}$$

Assume

$$\overline{Y}_i \xrightarrow{d} \mu_i + Z_i \frac{\sigma_i}{\sqrt{n_{i(1)}}},\ S_i^2 \xrightarrow{d} \sigma_i^2 \frac{Q_i}{n_{i(1)} - 1}, \tag{3.4}$$

where $\xrightarrow{d}$ is distributed as $Z_i \sim N(0,1)$ and $Q_i \sim \chi^2_{n_{i(1)}-1}$. Replace $(\overline{Y}_i, S_i^2)$ with their observed values $(\bar{y}_i, s_i^2)$, then the fiducial quantities for μ_i and σ_i^2 are

$$G_{\mu_i} = \bar{y}_i + \frac{Z_i \sqrt{n_{i(1)} - 1}}{\sqrt{Q_i}} \frac{s_i}{\sqrt{n_{i(1)}}} \tag{3.5}$$

and

$$G_{\sigma_i^2} = \frac{(n_{i(1)} - 1)s_i^2}{Q_i}. \tag{3.6}$$

Take (3.5) and (3.6) into (3.1) and (3.2), then we can obtain that the fiducial quantities for a_i, b_i are

$$G_{a_i} = \frac{1}{9}\{(1 + \frac{G^2_{\mu_i}}{2G_{\sigma_i^2}}) + [(1 + \frac{G^2_{\mu_i}}{2G_{\sigma_i^2}})^2 - 1]^{\frac{1}{2}}\} \tag{3.7}$$

and

$$G_{b_i} = 27 G_{a_i}^{\frac{1}{2}} (G^2_{\sigma_i})^{\frac{2}{3}}. \tag{3.8}$$

Then, we introduce the estimation for binomial parameter δ_i proposed by Li, Zou and Tian [28]; that is

$$G_{\delta_i} = \frac{n_{i(0)} + \frac{Z^2_{W_i}}{2}}{n_i + Z^2_{W_i}} - \frac{Z_{W_i}}{n_i + Z^2_{W_i}} [n_{i(0)}(1 - \frac{n_{i(0)}}{n_i}) + \frac{Z^2_{W_i}}{4}]^{\frac{1}{2}}, \tag{3.9}$$

where $Z_{W_i} = \frac{N_{i(0)} - n_i\delta_i}{\sqrt{n_i\delta_i(1-\delta_i)}} \sim N(0,1)$, $\sim$ stands for convergence distribution. The distribution of G_{δ_i} does not depend on any unknown parameters. When $N_{i(0)} = n_{i(0)}$, the observed value of G_{δ_i} is equal to δ_i.

Therefore, the fiducial quantity for the CV difference of two independent zero-inflated gamma distribution is

$$G_\gamma = G_{\eta_1} - G_{\eta_2} = \sqrt{\frac{1}{(1 - G_{\delta_1})G_{a_1}}} - \sqrt{\frac{1}{(1 - G_{\delta_2})G_{a_2}}}. \tag{3.10}$$

The CI of $100(1-\alpha)\%$ for γ is denoted as $[G_{\gamma;(\frac{\alpha}{2})}, G_{\gamma;(1-\frac{\alpha}{2})}]$, where $G_{\gamma;(\frac{\alpha}{2})}$ and $G_{\gamma;(1-\frac{\alpha}{2})}$ are the $100(\frac{\alpha}{2})$, $100(1-\frac{\alpha}{2})$ quartiles for G_γ's distribution, respectively.

The CV difference of two independent zero-inflated gamma distributions can be calculated in the following algorithm:

Algorithm 1:

Step1: Determine $n_{i(0)}$ and $n_{i(1)}$, where $n_{i(0)}$ is a realization of $N_{i(0)} \sim bin(n_i, \delta_i), n_{i(1)} = n_i - n_{i(0)}, i = 1, 2$.

Step2: For nonzero sample of size $n_{i(1)}$ from $gamma(a_i, b_i)$, calculate the transformed sample mean $\bar{y}_i$ and variance s_i^2.

Step3: Generate random variables $Z_i \sim N(0,1)$ and $Q_i \sim \chi^2_{n_{i(1)}-1}$, and then calculate $G_{\mu_i}$and $G_{\sigma_i^2}$ using (3.5) and (3.6).
Step4: Calculate G_{a_i} using (3.7).
Step5: Generate random variable $Z_{W_i} \sim N(0,1)$, and calculate G_{δ_i} using (3.9).
Step6: Calculate G_γ using (3.10).
Step7: Repeat Steps 2–6 for 10,000 times.
Step8: Calculate the $100(1-\alpha)\%$ CI for γ, denoted as $[l_{f1}, u_{f1}]$.

3.2. Fiducial-2 method

Different from traditional fiducial inference, Ren, Liu and Pu [12] proposed an exact method to fiducial inference for gamma parameters. To be specific, for the nonzero part of the zero-inflated gamma distribution, there exists a relationship between the arithmetic mean ($\overline{X}_i$) and the geometric mean $\tilde{X}_i$ of X_i as follows:

$$\frac{\overline{X}_i}{\tilde{X}_i} = \frac{(\frac{1}{n_{i(1)}})\sum_{j=1}^{n_{i(1)}} G^{-1}(\mathrm{U}_{ij}; a_i, 1)}{[\prod_{j=1}^{n_{i(1)}} G^{-1}(\mathrm{U}_{ij}; a_i, 1)]^{\frac{1}{n_{i(1)}}}}, \tag{3.11}$$

where $G(x; a_i, b_i)$ is the cumulative distribution function of $gamma(a_i, b_i)$, $\mathrm{U}_{ij} \sim \mathrm{U}(0,1)$, for $i = 1, 2$ and $j = 1, 2, \ldots, n_{i(1)}$. The fiducial quantity for the shape parameter a_i (i.e., G_{a_i}) can be solved by the sample observations using (3.11).

For the binomial distribution part, the transformed parameter converges to the standard normal distribution; that is

$$2\sqrt{n_i + 2 + \frac{1}{n_i}}(arcsin\sqrt{\tilde{\delta}_i} - arcsin\sqrt{\delta_i}) \xrightarrow{d} N(0,1), \tag{3.12}$$

where $\tilde{\delta}_i = \frac{n_{i(0)}+0.5}{n_i+1}$, $i = 1, 2$. Then, the fiducial quantity for the parameter δ_i is

$$G_{\delta_i} = sin^2(arcsin\sqrt{\tilde{\delta}_i^{obs}} - \frac{Z_i}{2\sqrt{n_i + 2 + \frac{1}{n_i}}}), \tag{3.13}$$

where $Z_i \sim N(0,1)$ and $\tilde{\delta}_i^{obs}$ is the observed value of $\tilde{\delta}_i$.

Once G_{a_i} and G_{δ_i} are obtained, the fiducial quantity G_γ can be calculated using (3.10), and the CI of $100(1-\alpha)\%$ for γ can be further obtained.

The algorithm for estimating the *CV* difference of two independent zero-inflated gamma distributions in Fiducial-2 method is as follows:

Algorithm 2:
Step1: Determine $n_{i(0)}$ and $n_{i(1)}$, where $n_{i(0)}$ is a realization of $N_{i(0)} \sim bin(n_i, \delta_i)$, $n_{i(1)} = n_i - n_{i(0)}$, $i = 1, 2$.
Step2: For nonzero sample of size $n_{i(1)}$ from $gamma(a_i, b_i)$, calculate the arithmetic mean $\overline{X}_i$ and the geometric mean $\tilde{X}_i$.
Step3: Generate random variables $\mathrm{U}_{ij} \sim \mathrm{U}(0,1)$ and calculate G_{a_i} using (3.11).
Step4: Generate random variable $Z_i \sim N(0,1)$ and calculate G_{δ_i} using (3.13).
Step5: Calculate G_γ using (3.10).
Step6: Repeat Steps 2–6 for 10,000 times.
Step7: Calculate the $100(1-\alpha)\%$ CI for γ, denoted as $[l_{f2}, u_{f2}]$.

3.3. Box-Cox transformation (BC) method

Box-Cox transformation (BC) is a statistical technique used to stabilize variance and/or normalize the distribution of a variable. For the samples distributed as gamma distribution, Gao and Tian [21] proposed to convert it into a normal distribution through BC transformation as follows:

$$Y_{ij} = X_{ij}^{(\lambda_i)} = \begin{cases} \frac{X_{ij}^{\lambda_i}-1}{\lambda_i}, \lambda_i > 0, \\ log(X_{ij}), \lambda_i = 0, \end{cases} \tag{3.14}$$

where $X_{ij} \sim gamma(a_i, b_i)$ and $Y_{ij} \sim N(\mu_i, \sigma_i^2)$, for $i = 1, 2$ and $j = 1, \ldots, n_{i(1)}$. The estimation of parameter λ_i can be derived through the maximum likelihood estimation (MLE) method, denoted as $\hat{\lambda}_i$. Under two different values of $\hat{\lambda}_i$, the parameters in the gamma distribution exhibit different relationships with those in the normal distribution.

If $\hat{\lambda}_i \neq 0$, the relationships can be expressed as

$$\frac{\Gamma(G_{a_i} + \hat{\lambda}_i) {G_{b_i}}^{\hat{\lambda}_i}}{\Gamma(G_{a_i})} = \hat{\lambda}_i G_{\mu_i} + 1 \tag{3.15}$$

and

$$\frac{\Gamma(G_{a_i} + 2\hat{\lambda}_i) {G_{b_i}}^{2\hat{\lambda}_i}}{\Gamma(G_{a_i})} = (G_{\mu_i}^2 + G_{\sigma_i^2})\hat{\lambda}_i^2 + 2\hat{\lambda}_i G_{\mu_i} + 1. \tag{3.16}$$

If $\hat{\lambda}_i = 0$, the relationships can be expressed as

$$G_{\mu_i} = \psi(G_{a_i}) + log(G_{b_i}) \tag{3.17}$$

and

$$G_{\sigma_i^2} = \psi^{(1)}(G_{a_i}), \tag{3.18}$$

where $\Gamma(\cdot)$ is the gamma function $\psi(\cdot)$ is the digamma function, and $\psi^{(1)}(\cdot)$ is the trigamma function.

Due to the fact that the fiducial quantities G_{μ_i} and $G_{\sigma_i^2}$ can be calculated using (3.5) and (3.6), the fiducial quantity G_{a_i} can be obtained according to the different expressions in the above scenarios.

For the binomial distribution part, we choose to use the fiducial quantity G_{δ_i} in (3.9), and G_γ can be similarly obtained using (3.10). Furthermore, the CI of $100(1-\alpha)\%$ for γ can be obtained.

The algorithm for estimating the CV difference of two independent zero-inflated gamma distributions in the BC method is as follows:

Algorithm 3:

Step1: Determine $n_{i(0)}$ and $n_{i(1)}$, where $n_{i(0)}$ is a realization of $N_{i(0)} \sim bin(n_i, \delta_i), n_{i(1)} = n_i - n_{i(0)}, i = 1, 2$.

Step2: For nonzero sample of size $n_{i(1)}$ from $gamma(a_i, b_i)$, perform BC transformation in (3.14), and calculate $\hat{\lambda}_i$ by the MLE.

Step3: Generate random variables $Z_i \sim N(0, 1)$ and $Q_i \sim \chi^2_{n_{i(1)}-1}$, and then calculate G_{μ_i} and $G_{\sigma_i^2}$ using (3.5) and (3.6).

Step4: Calculate G_{a_i} using (3.15) and (3.16), or (3.17) and (3.18).

Step5: Generate random variable $Z_{W_i} \sim N(0, 1)$, and calculate G_{δ_i} using (3.9).

Step6: Calculate G_γ using (3.10).

Step7: Repeat Steps 2–6 for 10,000 times.

Step8: Calculate the $100(1-\alpha)\%$ CI for γ, denoted as $[l_{bc}, u_{bc}]$.

3.4. Parametric bootstrap (PB) method

Parametric bootstrap (PB) method is a resampling technique used to estimate the sampling distribution of a statistic by generating new samples from a parametric model. In our study, based on the sample $X_{ij}, i = 1, 2, j = 1, 2, \ldots, n_i$ from the zero-inflated gamma distribution, we can calculate the MLE of δ_i (denoted as $\hat{\delta}_i$). Then, PB samples of size n_i are generated based on $Bin(n_i, \hat{\delta}_i)$, and the MLE of $\hat{\delta}_i$ (denoted as $\hat{\delta}_i^*$) can be calculated. For $gamma(a_i, b_i)$, we use the positive part of samples with size $n_{i(1)}$ to compute the $MLEs$ of a_i, b_i (denoted as $\hat{a}_i, \hat{b}_i$). Finally, PB samples are generated by $gamma(\hat{a}_i, \hat{b}_i)$ and the MLE of $\hat{a}_i$ (denoted as $\hat{a}_i^*$) can be obtained.

The CV difference between two zero-inflated gamma distributions can be calculated by the following equation:

$$G_{PB} = \sqrt{\frac{1}{(1 - \hat{\delta}_1^*)\hat{a}_1^*}} - \sqrt{\frac{1}{(1 - \hat{\delta}_2^*)\hat{a}_2^*}}. \tag{3.19}$$

The CI of $100(1 - \alpha)\%$ for PB is denoted as $[G_{PB;(\frac{\alpha}{2})}, G_{PB;(1-\frac{\alpha}{2})}]$, where $G_{PB;(\frac{\alpha}{2})}$ and $G_{PB;(1-\frac{\alpha}{2})}$ are the $100(\frac{\alpha}{2})$, $100(1 - \frac{\alpha}{2})$ quartiles for G_{PB}'s distribution, respectively.

The PB method for constructing CI of the CV difference of two independent zero-inflated gamma distributions is implemented as follows:

Algorithm 4:

Step1: Determine $n_{i(0)}$ and $n_{i(1)}$, where $n_{i(0)}$ is a realization of $N_{i(0)} \sim bin(n_i, \delta_i), n_{i(1)} = n_i - n_{i(0)}, i = 1, 2$.
Step2: Compute the MLE $\hat{\delta}_i$.
Step3: Generate bootstrap samples of size n_i from $bin(n_i, \hat{\delta}_i)$, and calculate the MLE $\hat{\delta}_i^*$.
Step4: For nonzero samples of size $n_{i(1)}$ from $gamma(a_i, b_i)$, calculate the MLE $\hat{a}_i$.
Step5: Generate bootstrap samples of size $n_{i(1)}$ from $gamma(\hat{a}_i, \hat{b}_i)$, and calculate the MLE $\hat{a}_i^*$.
Step6: Calculate G_{PB} using (3.19).
Step7: Repeat Steps 2–6 for 10,000 times.
Step8: Calculate the $100(1 - \alpha)\%$ CI for PB, denoted as $[l_{pb}, u_{pb}]$.

3.5. MOVER method

The method of variance estimates recovery (MOVER) was proposed for constructing CIs for function of parameters, such as the difference or ratio between two parameters. Based on MOVER, the lower and upper confidence limits for $\theta_1 + \theta_2$ are

$$L = \hat{\theta}_1 + \hat{\theta}_2 - \sqrt{(\hat{\theta}_1 - l_1)^2 + (\hat{\theta}_2 - l_2)^2} \tag{3.20}$$

and

$$U = \hat{\theta}_1 + \hat{\theta}_2 - \sqrt{(u_1 - \hat{\theta}_1)^2 + (u_2 - \hat{\theta}_2)^2}. \tag{3.21}$$

Similarly, the lower and upper confidence limits for $\theta_1 - \theta_2$ are

$$L' = \hat{\theta}_1 - \hat{\theta}_2 - \sqrt{(\hat{\theta}_1 - l_1)^2 + (u_2 - \hat{\theta}_2)^2} \tag{3.22}$$

and

$$U' = \hat{\theta}_1 - \hat{\theta}_2 + \sqrt{(u_1 - \hat{\theta}_1)^2 + (\hat{\theta}_2 - l_2)^2}. \tag{3.23}$$

$\hat{\theta}_1$ and $\hat{\theta}_2$ are independent estimators, and (l_1, u_1) and (l_2, u_2) are $100(1-\alpha)\%$ confidence limits for parameters θ_1 and θ_2.

For the CV of zero-inflated gamma distribution, we can perform a log-transformation, i.e., $log(\eta_i) = log(\sqrt{\frac{1}{(1-\delta_i)a_i}}) = -\frac{1}{2}[log(1-\delta_i) + log(a_i)]$. Denote $\theta_{i1} \triangleq log(1-\delta_i)$ and $\theta_{i2} \triangleq log(a_i)$, and we can obtain the CI for $log(\eta_i)$ according to (3.20), (3.21) and $\eta_i = exp[-\frac{1}{2}(\theta_{i1} + \theta_{i2})]$.

For the CV difference of two independent zero-inflated gamma distribution, we can first obtain the MLE of two CVs, i.e., $\hat{\eta}_i = \sqrt{\frac{1}{(1-\hat{\delta}_i)\hat{a}_i}}$, $i = 1, 2$, then the CI for the CV difference can be calculated using (3.22) and (3.23).

In the preceding sections, we have presented four methods (Fiducial-1, Fiducial-2, BC, PB) for constructing CIs of the CVs difference between zero-inflated gamma distributions. These methods involve four distinct approaches for constructing the pivot quantity for the shape parameter a_i and two different approaches (Fiducial-1, Fiducial-2) for constructing the pivot quantity for the binomial parameter δ_i. Additionally, we introduce another exact estimation of the CI for δ_i proposed by Wilson [23], given as follows:

$$[\delta_{il}, \delta_{iu}] = [\frac{\hat{\delta}_i + \frac{Z^2_{\frac{\alpha}{2}}}{2n_i} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\delta}_i(1-\hat{\delta}_i)}{n_i} + \frac{Z^2_{\frac{\alpha}{2}}}{4n_i^2}}}{1 + \frac{Z^2_{\frac{\alpha}{2}}}{n_i}}, \frac{\hat{\delta}_i + \frac{Z^2_{\frac{\alpha}{2}}}{2n_i} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\delta}_i(1-\hat{\delta}_i)}{n_i} + \frac{Z^2_{\frac{\alpha}{2}}}{4n_i^2}}}{1 + \frac{Z^2_{\frac{\alpha}{2}}}{n_i}}], \tag{3.24}$$

where $\hat{\delta}_i = \frac{n_{i(0)}}{n_i}$ and $Z_{\frac{\alpha}{2}}$ represents the upper $\frac{\alpha}{2}$ quartiles of the standard normal distribution. Consequently, there are a total of four interval estimates for θ_{i2} and three interval estimates for θ_{i1}. In this way, we can obtain twelve MOVER combination methods for constructing CIs of the CV difference.

4. Simulation experiments

To judge the accuracies of the proposed methods, we conduct Monte Carlo simulation studies to compare each method according to the following measurements.

Coverage probability (CP): The ratio of true parameters falling within the CIs.

Average length (AL): The average length of all estimated CIs.

Left error rate (ER_L): The percentage of a true parameter of interest falls to the left of the lower confidence limits (L_{cl}).

Right error rate (ER_R): The percentage of a true parameter of interest falls to the right of the upper confidence limits (U_{cl}).

The formulas are shown in below:

$$CP = \frac{C(L_{cl} \le \gamma \le U_{cl})}{t}, \tag{4.1}$$

$$AL = \frac{\sum_{i=1}^{t}(U_{cl_i} - L_{cl_i})}{t}, \tag{4.2}$$

$$ER_L = \frac{C(\gamma \le L_{cl})}{t}, \tag{4.3}$$

$$ER_R = \frac{C(U_{cl} \leq \gamma)}{t}, \tag{4.4}$$

where $C(L_{cl} \leq \gamma \leq U_{cl})$ refers to the number of experiments where the difference γ falls within the CIs and t is the number of experiments, which is chosen to be $1,000$. To evaluate the performances of CIs, we tend to favor those CIs with coverage probabilities equal to or greater than the nominal level, and at the same time, with relatively short interval lengths. Left and right tail error rates of the CIs are also considered. In general, we tend to favor those CIs with balanced tail error rates.

In the simulation studies, we set sample sizes (n_1, n_2) to be 10, 20, 50 and 132, proportion of zero values (δ_1, δ_2) to be from 0.1 to 0.6, shape parameters (a_1, a_2) to be two and five, nominal level α to be 0.05 and the run size t to be 1000. As the problem we consider here is scale invariant [19], the scale parameter is set to be one for all scenarios. In addition, as sample sizes for the practical problems in section five are 132 and 132, we illustrate our proposed CIs work properly for these situations in the simulation studies.

Five estimation methods proposed by Fiducial-1 (F1), Fiducial-2 (F2), BC transform, PB and Wilson (Wil) are introduced in this article. Different combinations of these methods are obtained by MOVER. For example, in Tables 1–3, F1F2 means that a is estimated by F1 and δ is estimated by F2. The values in bold in the tables correspond to the method whose coverage possibility is closest to 0.95 for each set of parameters. The results show that the MOVER comprised by F2 and F2 produced the most satisfactory results according to coverage probabilities. To this end, MOVER with F2F2 will be further investigated in the following.

Tables 4–6 illustrate F1, F2, BC transformation, PB and the MOVER method comprised by F2 and F2. From the results, we can conclude that:

1) F1, F2, BC and MOVER CIs produce satisfactory results according to coverage probabilities for small to large sample sizes.

2) Among these four methods, the average interval lengths for F1 are shortest for small to moderate sample sizes. As the sample sizes get large, the average interval lengths become similar for all four methods.

3) All these four methods produce well balanced CIs.

4) Coverage probabilities for PB CIs are conservative for small to moderate sample sizes. As the sample sizes get larger, the coverage probabilities get close to nominal level.

5) The computational costs for MOVER are comparable to those of the F2 approach, whereas F1 demonstrates the most efficient time. Time comparison for all these five methods, including F1, F2, BC transformation, PB and the MOVER method comprised by F2 and F2 (1,000 times running measured in seconds), are illustrated in Figure 1.

All the simulation results and time comparisons are conducted using R, and the codes can be shared upon request.

Table 1. CP of the MOVER methods for small proportion of zeros.

$\delta_1 : \delta_2 = 0.1 : 0.1$ $n_1 : n_2$	$a_1 : a_2$	F1F1	F1F2	F1Wil	F2F1	F2F2	F2Wil	BCF1	BCF2	BCWil	PBF1	PBF2	PBWil
10:10	2:2	0.963	0.960	0.961	0.961	**0.954**	0.956	0.963	0.956	0.958	0.958	0.956	0.957
	2:5	0.971	0.959	0.962	0.966	0.958	0.960	0.964	**0.956**	**0.956**	0.938	0.939	0.943
	5:5	0.974	0.965	0.970	0.973	**0.963**	0.965	0.975	0.964	0.967	0.974	0.966	0.964
10:20	2:2	0.968	0.954	0.953	0.962	**0.950**	0.953	0.959	0.955	0.955	0.934	0.933	0.935
	2:5	0.978	0.968	0.971	0.973	0.968	0.966	0.971	0.966	**0.965**	0.896	0.902	0.914
	5:5	0.967	**0.950**	0.952	0.962	0.953	0.952	0.962	0.954	0.953	0.945	0.938	0.935
10:50	2:2	0.980	0.972	0.969	0.974	0.967	**0.962**	0.975	0.967	0.967	0.881	0.871	0.883
	2:5	0.978	0.964	0.965	0.970	**0.953**	0.954	0.967	0.961	0.960	0.861	0.877	0.890
	5:5	0.976	0.965	0.960	0.973	0.961	**0.958**	0.972	0.965	0.964	0.891	0.887	0.890
10:132	2:2	0.979	0.956	0.960	0.976	**0.949**	**0.949**	0.966	0.954	0.958	0.864	0.865	0.872
	2:5	0.966	0.946	0.955	0.965	0.941	0.948	0.960	**0.949**	0.952	0.830	0.849	0.863
	5:5	0.964	0.946	0.945	0.966	0.944	0.939	0.966	**0.950**	0.942	0.869	0.874	0.880
50:50	2:2	0.981	**0.950**	0.952	0.980	0.946	0.951	0.978	0.953	0.948	0.978	0.948	0.945
	2:5	0.991	0.964	0.966	0.989	**0.956**	0.957	0.991	0.960	0.960	0.978	0.959	0.960
	5:5	0.983	0.962	0.963	0.982	0.958	0.963	0.982	0.957	**0.956**	0.981	0.961	0.961
50:132	2:2	0.987	0.967	0.964	0.987	0.956	0.959	0.985	**0.954**	**0.954**	0.981	0.957	0.957
	2:5	0.980	0.955	0.958	0.974	0.946	0.947	0.974	**0.950**	0.949	0.956	0.932	0.930
	5:5	0.979	0.946	0.944	0.977	0.941	0.944	0.977	0.943	0.945	0.975	0.947	**0.948**
132:132	2:2	0.996	0.960	0.962	0.996	0.953	**0.952**	0.995	0.958	0.958	0.995	0.955	0.953
	2:5	0.994	0.968	0.969	0.995	0.963	0.966	0.994	0.966	0.967	0.993	0.964	**0.962**
	5:5	0.985	0.955	0.952	0.985	**0.950**	0.949	0.984	0.949	0.947	0.987	0.951	0.948

Table 2. CP of the MOVER methods for moderate proportion of zeros.

$\delta_1 : \delta_2 = 0.3 : 0.3$ $n_1 : n_2$	$a_1 : a_2$	F1F1	F1F2	F1Wil	F2F1	F2F2	F2Wil	BCF1	BCF2	BCWil	PBF1	PBF2	PBWil
10:10	2:2	0.980	0.966	0.961	0.977	**0.958**	0.962	0.986	0.964	0.967	0.973	0.963	0.962
	2:5	0.978	0.958	0.958	0.976	**0.951**	0.953	0.980	0.955	0.961	0.947	0.941	0.941
	5:5	0.981	0.969	0.968	0.980	0.969	**0.965**	0.987	0.973	0.971	0.979	0.973	0.972
10:20	2:2	0.995	0.969	0.970	0.993	0.968	0.969	0.992	**0.965**	0.970	0.968	0.938	0.935
	2:5	0.973	0.942	**0.946**	0.970	0.941	0.941	0.969	0.944	0.941	0.884	0.869	0.871
	5:5	0.986	0.959	0.961	0.985	0.959	0.959	0.989	0.964	0.964	0.973	0.942	**0.946**
10:50	2:2	0.985	**0.950**	0.947	0.985	0.939	0.944	0.980	0.958	0.954	0.930	0.882	0.887
	2:5	0.976	**0.950**	0.953	0.977	**0.950**	0.948	0.965	0.951	0.953	0.856	0.830	0.836
	5:5	0.988	0.959	**0.958**	0.988	**0.958**	0.959	0.991	0.969	0.970	0.936	0.887	0.887
10:132	2:2	0.981	**0.950**	0.952	0.980	0.945	0.942	0.980	0.949	0.954	0.907	0.838	0.841
	2:5	0.978	0.958	0.956	0.977	0.953	**0.952**	0.977	0.963	0.964	0.864	0.833	0.838
	5:5	0.991	**0.958**	0.963	0.989	0.960	0.961	0.993	0.972	0.974	0.913	0.866	0.865
50:50	2:2	0.999	**0.949**	0.948	0.999	0.943	0.939	0.999	0.943	0.941	0.999	0.943	0.942
	2:5	0.997	0.955	0.953	0.996	0.945	0.944	0.994	**0.951**	0.957	0.993	0.954	0.953
	5:5	1.000	0.953	0.955	1.000	0.952	0.954	1.000	0.953	0.951	1.000	**0.950**	0.952
50:132	2:2	1.000	0.956	0.958	1.000	0.953	0.947	1.000	**0.951**	0.948	1.000	0.939	0.941
	2:5	0.999	0.967	0.969	0.999	0.958	0.961	0.999	0.960	0.958	0.996	0.940	**0.947**
	5:5	0.999	0.953	0.952	0.999	0.947	**0.949**	0.999	**0.951**	**0.949**	0.999	0.944	0.944
132:132	2:2	1.000	0.952	0.948	1.000	0.944	0.944	1.000	**0.950**	0.949	1.000	0.944	0.945
	2:5	1.000	0.952	**0.950**	1.000	0.943	0.948	1.000	0.946	0.944	1.000	0.944	0.944
	5:5	1.000	0.964	0.961	1.000	0.959	0.958	1.000	0.962	0.963	1.000	**0.957**	0.961

Table 3. CP of the MOVER methods for large proportion of zeros.

$\delta_1 : \delta_2 = 0.6 : 0.6$ $n_1 : n_2$	 $a_1 : a_2$	F1F1	F1F2	F1Wil	F2F1	F2F2	F2Wil	BCF1	BCF2	BCWil	PBF1	PBF2	PBWil
10:10	2:2	0.983	0.958	0.958	0.984	0.959	0.956	0.991	0.973	0.968	0.971	**0.946**	0.945
	2:5	0.985	0.967	**0.966**	0.986	0.968	0.968	0.991	0.969	0.974	0.917	0.900	0.894
	5:5	0.983	0.966	0.966	0.985	0.966	0.966	0.989	0.976	0.974	0.957	**0.944**	0.938
10:20	2:2	0.994	0.963	0.962	0.993	0.960	0.963	0.998	0.974	0.974	**0.952**	0.900	0.899
	2:5	0.982	0.952	0.958	0.979	**0.949**	**0.949**	0.986	0.959	0.959	0.895	0.843	0.839
	5:5	0.992	0.973	0.972	0.992	0.973	0.970	0.997	0.981	0.982	**0.962**	0.917	0.913
10:50	2:2	0.992	0.955	0.954	0.992	**0.952**	0.953	0.995	0.964	0.963	0.937	0.823	0.813
	2:5	0.986	0.955	**0.953**	0.983	0.957	**0.953**	0.990	0.964	0.962	0.862	0.770	0.765
	5:5	0.996	0.969	0.968	0.996	0.970	0.970	0.999	0.979	0.978	**0.946**	0.842	0.834
10:132	2:2	0.995	0.963	0.962	0.996	**0.957**	0.961	0.995	0.971	0.972	0.912	0.756	0.749
	2:5	0.986	0.965	0.965	0.986	**0.957**	0.961	0.989	0.963	0.966	0.861	0.751	0.743
	5:5	0.990	0.959	0.964	0.989	**0.956**	0.957	0.999	0.978	0.971	0.924	0.788	0.774
50:50	2:2	1.000	0.966	0.965	1.000	**0.958**	0.961	1.000	0.961	0.958	1.000	0.974	0.970
	2:5	1.000	0.964	0.965	1.000	0.965	0.959	1.000	0.964	0.957	1.000	0.959	**0.954**
	5:5	1.000	0.952	0.953	1.000	0.954	**0.951**	1.000	0.955	0.954	1.000	0.962	0.965
50:132	2:2	1.000	0.955	0.955	1.000	0.949	0.955	1.000	0.952	0.953	1.000	**0.950**	0.944
	2:5	1.000	0.962	0.964	1.000	0.958	0.959	1.000	0.958	0.959	1.000	0.944	**0.947**
	5:5	1.000	0.959	0.957	1.000	0.957	0.954	1.000	0.957	**0.951**	1.000	0.941	0.942
132:132	2:2	1.000	0.966	0.964	1.000	0.963	0.962	1.000	0.965	**0.961**	1.000	0.965	0.964
	2:5	1.000	0.941	0.940	1.000	0.941	0.940	1.000	0.940	0.942	1.000	**0.943**	**0.943**
	5:5	1.000	0.955	0.956	1.000	**0.954**	**0.954**	1.000	0.958	0.956	1.000	0.956	0.959

Table 4. CP, AL and tail error rates for small proportion of zeros.

$\delta_1 : \delta_2 = 0.1 : 0.1$		Fiducial-1		Fiducial-2		BC		PB		MOVER	
		CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
$n_1 : n_2$	$a_1 : a_2$	ER_L	ER_R	ER_L	ER_R	ER_L	ER_R	ER_L	ER_R	ER_L	ER_R
10:10	2:2	0.952	1.256	0.946	1.237	0.947	1.287	**0.949**	0.902	0.954	1.314
		0.023	0.025	0.027	0.027	0.025	0.028	0.023	0.028	0.023	0.023
	2:5	**0.949**	1.084	0.948	1.068	0.952	1.092	0.937	0.769	0.958	1.127
		0.024	0.027	0.025	0.027	0.018	0.030	0.003	0.060	0.022	0.020
	5:5	0.955	0.867	0.952	0.854	**0.952**	0.856	0.956	0.590	0.963	0.909
		0.023	0.022	0.023	0.025	0.028	0.020	0.021	0.023	0.019	0.018
10:20	2:2	0.951	1.041	0.938	1.025	0.951	1.055	0.931	0.796	**0.950**	1.071
		0.027	0.022	0.033	0.029	0.022	0.027	0.004	0.065	0.026	0.024
	2:5	0.967	0.949	0.964	0.932	**0.958**	0.959	0.900	0.711	0.968	0.961
		0.014	0.019	0.019	0.017	0.016	0.026	0.003	0.097	0.017	0.015
	5:5	0.944	0.710	0.942	0.701	**0.951**	0.703	0.933	0.519	0.953	0.731
		0.035	0.021	0.036	0.022	0.028	0.021	0.005	0.062	0.029	0.018
10:50	2:2	0.970	0.913	**0.960**	0.903	0.963	0.928	0.870	0.702	0.967	0.914
		0.014	0.016	0.022	0.018	0.013	0.024	0.001	0.129	0.018	0.015
	2:5	0.963	0.882	**0.953**	0.872	0.960	0.899	0.878	0.665	**0.953**	0.872
		0.022	0.015	0.030	0.017	0.016	0.024	0.000	0.122	0.031	0.016
	5:5	0.957	0.620	**0.952**	0.613	0.960	0.616	0.882	0.457	0.961	0.624
		0.022	0.021	0.029	0.019	0.018	0.022	0.001	0.117	0.022	0.017
10:132	2:2	0.959	0.878	0.944	0.869	0.954	0.892	0.864	0.666	**0.949**	0.866
		0.023	0.018	0.038	0.018	0.016	0.030	0.001	0.135	0.029	0.022
	2:5	0.954	0.857	0.941	0.848	**0.947**	0.866	0.848	0.647	0.941	0.839
		0.021	0.025	0.031	0.028	0.018	0.035	0.000	0.152	0.031	0.028
	5:5	**0.946**	0.603	0.945	0.596	0.943	0.594	0.878	0.436	0.944	0.593
		0.031	0.023	0.035	0.020	0.028	0.029	0.000	0.122	0.033	0.023
50:50	2:2	**0.953**	0.448	0.942	0.441	0.942	0.448	0.945	0.416	0.946	0.449
		0.023	0.024	0.029	0.029	0.027	0.031	0.027	0.028	0.027	0.027
	2:5	0.957	0.378	**0.953**	0.373	0.962	0.377	0.961	0.351	0.956	0.379
		0.021	0.022	0.028	0.019	0.020	0.018	0.013	0.026	0.027	0.017
	5:5	**0.958**	0.296	**0.958**	0.293	0.959	0.292	0.963	0.274	**0.958**	0.299
		0.025	0.017	0.024	0.018	0.023	0.018	0.022	0.015	0.024	0.018
50:132	2:2	0.964	0.368	0.953	0.363	**0.950**	0.367	0.955	0.347	0.956	0.366
		0.019	0.017	0.027	0.020	0.021	0.029	0.009	0.036	0.026	0.018
	2:5	0.957	0.336	0.945	0.331	**0.947**	0.335	0.932	0.316	0.946	0.333
		0.018	0.025	0.028	0.027	0.016	0.037	0.007	0.061	0.027	0.027
	5:5	**0.945**	0.241	0.940	0.238	0.941	0.239	**0.945**	0.226	0.941	0.242
		0.032	0.023	0.033	0.027	0.029	0.030	0.014	0.041	0.033	0.026
132:132	2:2	0.960	0.265	0.954	0.262	**0.953**	0.267	0.956	0.256	**0.953**	0.264
		0.022	0.018	0.027	0.019	0.023	0.024	0.023	0.021	0.025	0.022
	2:5	0.969	0.225	0.963	0.222	0.961	0.226	**0.960**	0.217	0.963	0.223
		0.013	0.018	0.016	0.021	0.020	0.019	0.014	0.026	0.019	0.018
	5:5	0.948	0.175	0.947	0.173	0.943	0.174	0.947	0.169	**0.950**	0.175
		0.027	0.025	0.027	0.026	0.028	0.029	0.028	0.025	0.025	0.025

Table 5. CP, AL and tail error rates for moderate proportion of zeros.

$\delta_1 : \delta_2 = 0.3 : 0.3$		Fiducial-1		Fiducial-2		BC		PB		MOVER	
		CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
$n_1 : n_2$	$a_1 : a_2$	ER_L	ER_R	ER_L	ER_R	ER_L	ER_R	ER_L	ER_R	ER_L	ER_R
10:10	2:2	**0.952**	1.960	0.941	1.991	0.953	2.245	0.959	1.242	0.958	2.139
		0.023	0.025	0.029	0.030	0.022	0.024	0.023	0.018	0.022	0.020
	2:5	0.947	1.718	0.943	1.740	0.953	1.960	0.933	1.049	**0.951**	1.876
		0.029	0.024	0.032	0.025	0.019	0.028	0.005	0.062	0.028	0.021
	5:5	**0.957**	1.389	**0.957**	1.377	0.968	1.475	0.964	0.802	0.969	1.487
		0.026	0.017	0.026	0.017	0.017	0.016	0.021	0.015	0.020	0.011
10:20	2:2	0.965	1.562	**0.961**	1.575	0.964	1.704	0.921	1.073	0.968	1.665
		0.015	0.020	0.020	0.019	0.017	0.019	0.000	0.079	0.016	0.016
	2:5	0.939	1.396	0.936	1.391	**0.941**	1.515	0.857	0.941	**0.941**	1.446
		0.026	0.035	0.028	0.036	0.021	0.037	0.000	0.143	0.026	0.033
	5:5	**0.949**	1.094	0.955	1.081	0.962	1.150	0.935	0.705	0.959	1.144
		0.031	0.020	0.027	0.018	0.021	0.016	0.001	0.064	0.026	0.015
10:50	2:2	0.943	1.412	0.932	1.420	**0.952**	1.667	0.873	0.974	0.939	1.450
		0.029	0.028	0.039	0.029	0.022	0.026	0.000	0.127	0.034	0.027
	2:5	0.951	1.292	0.946	1.292	0.943	1.485	0.821	0.877	**0.950**	1.311
		0.020	0.029	0.026	0.028	0.013	0.044	0.000	0.179	0.024	0.026
	5:5	0.960	0.937	**0.956**	0.928	0.970	1.009	0.875	0.605	0.958	0.955
		0.024	0.016	0.028	0.016	0.018	0.011	0.000	0.125	0.027	0.015
10:132	2:2	**0.949**	1.295	0.942	1.309	0.955	1.479	0.833	0.874	0.945	1.321
		0.024	0.027	0.029	0.029	0.013	0.032	0.000	0.167	0.027	0.028
	2:5	0.956	1.282	**0.951**	1.287	0.960	1.485	0.826	0.860	0.953	1.286
		0.021	0.023	0.025	0.024	0.018	0.022	0.000	0.174	0.024	0.023
	5:5	**0.958**	0.915	0.961	0.917	0.971	0.974	0.850	0.570	0.960	0.918
		0.023	0.019	0.021	0.018	0.015	0.013	0.000	0.150	0.021	0.019
50:50	2:2	**0.949**	0.618	0.937	0.609	0.941	0.613	0.939	0.565	0.943	0.623
		0.022	0.029	0.028	0.035	0.030	0.029	0.029	0.032	0.025	0.032
	2:5	0.953	0.518	0.944	0.512	**0.951**	0.513	0.954	0.475	0.945	0.524
		0.018	0.029	0.024	0.032	0.018	0.031	0.008	0.038	0.026	0.029
	5:5	**0.951**	0.407	0.944	0.404	0.948	0.401	0.955	0.370	0.952	0.414
		0.027	0.022	0.030	0.026	0.025	0.027	0.024	0.021	0.028	0.020
50:132	2:2	**0.953**	0.501	0.946	0.494	0.944	0.500	0.935	0.467	**0.953**	0.501
		0.022	0.025	0.023	0.031	0.016	0.040	0.011	0.054	0.021	0.026
	2:5	0.963	0.462	0.953	0.456	**0.952**	0.460	0.941	0.430	0.958	0.460
		0.016	0.021	0.022	0.025	0.018	0.030	0.008	0.051	0.021	0.021
	5:5	0.946	0.330	0.944	0.327	0.955	0.326	0.942	0.306	**0.947**	0.332
		0.029	0.025	0.031	0.025	0.023	0.022	0.014	0.044	0.029	0.024
132:132	2:2	0.958	0.361	0.935	0.356	0.943	0.362	0.942	0.346	**0.944**	0.360
		0.017	0.025	0.027	0.038	0.023	0.034	0.024	0.034	0.024	0.032
	2:5	**0.949**	0.306	0.941	0.303	0.942	0.305	0.941	0.294	0.943	0.306
		0.026	0.025	0.036	0.023	0.030	0.028	0.028	0.031	0.037	0.020
	5:5	0.960	0.237	0.956	0.235	**0.954**	0.235	0.957	0.228	0.959	0.238
		0.021	0.019	0.023	0.021	0.021	0.025	0.021	0.022	0.021	0.020

Table 6. CP, AL and tail error rates for middle proportion of zeros.

$\delta_1 : \delta_2 = 0.3 : 0.3$		Fiducial-1		Fiducial-2		BC		PB		MOVER	
		CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
$n_1 : n_2$	$a_1 : a_2$	ER_L	ER_R	ER_L	ER_R	ER_L	ER_R	ER_L	ER_R	ER_L	ER_R
10:10	2:2	**0.943**	4.786	0.938	6.838	0.959	9.159	0.926	2.031	0.959	7.390
		0.028	0.029	0.032	0.030	0.020	0.020	0.033	0.041	0.021	0.020
	2:5	**0.947**	4.464	**0.947**	5.922	0.960	7.950	0.881	1.673	0.968	6.409
		0.015	0.037	0.018	0.035	0.010	0.030	0.001	0.118	0.011	0.021
	5:5	0.941	3.872	**0.944**	4.898	0.958	6.084	0.915	1.280	0.966	5.303
		0.028	0.031	0.029	0.028	0.020	0.023	0.042	0.043	0.018	0.017
10:20	2:2	**0.944**	3.740	0.942	4.683	0.963	6.458	0.879	1.856	0.960	5.028
		0.022	0.034	0.025	0.033	0.022	0.015	0.002	0.119	0.017	0.023
	2:5	0.946	3.448	0.939	4.459	0.955	5.778	0.817	1.617	**0.949**	4.746
		0.020	0.034	0.030	0.032	0.021	0.024	0.000	0.183	0.022	0.029
	5:5	**0.961**	2.871	0.962	3.332	0.971	4.271	0.890	1.174	0.973	3.598
		0.015	0.024	0.014	0.024	0.014	0.015	0.004	0.106	0.009	0.018
10:50	2:2	**0.948**	3.303	0.945	4.358	0.962	6.224	0.790	1.611	**0.952**	4.512
		0.026	0.026	0.026	0.028	0.019	0.019	0.000	0.210	0.024	0.024
	2:5	**0.950**	3.092	0.951	4.117	0.956	5.525	0.735	1.438	0.957	4.307
		0.025	0.025	0.022	0.027	0.025	0.019	0.000	0.265	0.023	0.020
	5:5	**0.962**	2.487	0.963	2.977	0.974	3.949	0.795	1.025	0.970	3.105
		0.018	0.020	0.019	0.018	0.015	0.011	0.000	0.205	0.016	0.014
10:132	2:2	0.958	2.999	0.960	3.906	0.967	5.639	0.728	1.407	**0.957**	3.938
		0.013	0.029	0.014	0.026	0.012	0.020	0.000	0.272	0.015	0.028
	2:5	0.963	3.013	**0.956**	4.095	0.962	5.412	0.719	1.405	0.957	4.140
		0.013	0.024	0.015	0.029	0.017	0.021	0.000	0.281	0.015	0.028
	5:5	**0.954**	2.440	0.957	3.049	0.974	3.945	0.740	0.921	0.956	3.138
		0.015	0.031	0.011	0.031	0.012	0.013	0.000	0.260	0.015	0.029
50:50	2:2	0.962	1.263	**0.953**	1.246	0.956	1.249	0.969	1.085	0.958	1.295
		0.018	0.020	0.020	0.027	0.022	0.022	0.014	0.017	0.020	0.022
	2:5	0.964	1.049	0.960	1.037	0.958	1.036	**0.952**	0.901	0.965	1.077
		0.013	0.023	0.016	0.024	0.013	0.029	0.006	0.042	0.017	0.018
	5:5	0.946	0.820	0.948	0.812	**0.949**	0.800	0.963	0.695	0.954	0.847
		0.027	0.027	0.027	0.025	0.028	0.023	0.020	0.017	0.023	0.023
50:132	2:2	0.955	1.000	0.948	0.992	0.953	0.993	0.941	0.889	**0.949**	1.012
		0.021	0.024	0.028	0.024	0.017	0.030	0.006	0.053	0.027	0.024
	2:5	0.959	0.917	0.965	0.906	**0.955**	0.903	0.942	0.808	0.958	0.919
		0.019	0.022	0.017	0.018	0.014	0.031	0.005	0.053	0.022	0.020
	5:5	0.954	0.656	**0.949**	0.651	0.947	0.646	0.938	0.575	0.957	0.666
		0.022	0.024	0.023	0.028	0.019	0.034	0.004	0.058	0.017	0.026
132:132	2:2	0.966	0.696	0.960	0.689	**0.958**	0.697	0.963	0.657	0.963	0.702
		0.020	0.014	0.023	0.017	0.023	0.019	0.023	0.014	0.021	0.016
	2:5	0.941	0.583	0.933	0.577	0.935	0.581	**0.944**	0.549	0.941	0.587
		0.023	0.036	0.032	0.035	0.026	0.039	0.018	0.038	0.029	0.030
	5:5	0.956	0.456	0.948	0.451	**0.950**	0.452	0.955	0.428	0.954	0.461
		0.019	0.025	0.020	0.032	0.022	0.028	0.018	0.027	0.021	0.025

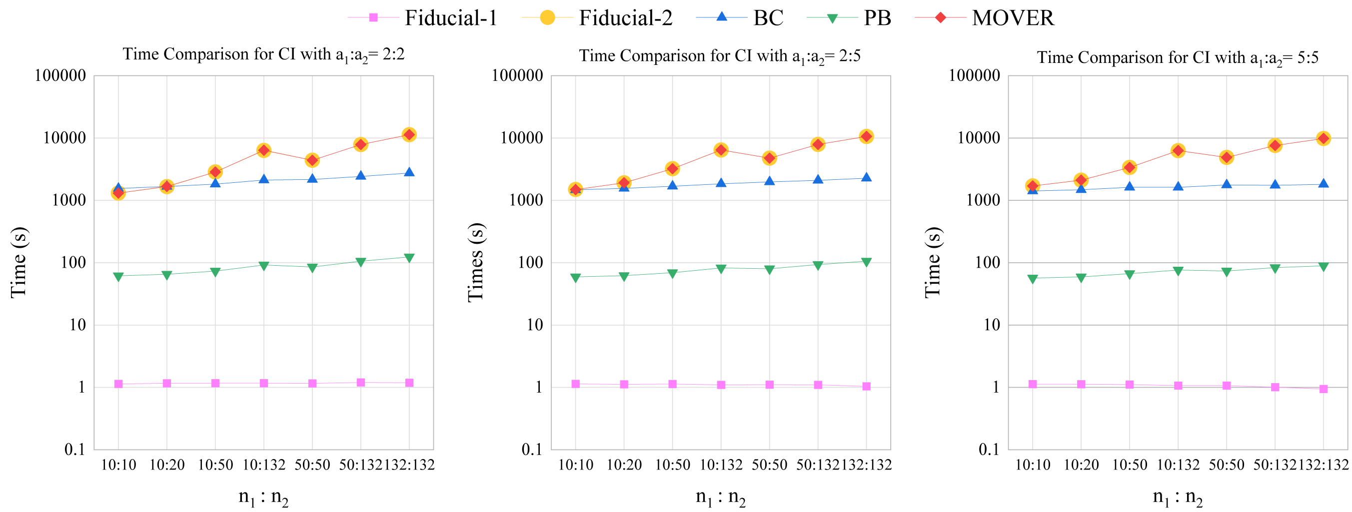

Figure 1. Time Comparison for CI.

5. Empirical analysis

To illustrate the proposed CIs, we study the monthly precipitation data from the National Meteorological Science Data Center for Beijing and Zhengzhou for 11 years, from 2009 to 2019. As the density curves shown in Figure 2, the two precipitation data for Beijing and Zhengzhou contain a large number of zeros, and the non-zero parts follow gamma distirbution well. The Akaike information criterias (AICs) of several distributions to fit the non-zero part of the data sets are calculated and listed in Table 7. AICs illustrated in Table 7 show that the data sets fit gamma distribution well, as the AICs for gamma distribution is smallest among other distributions.

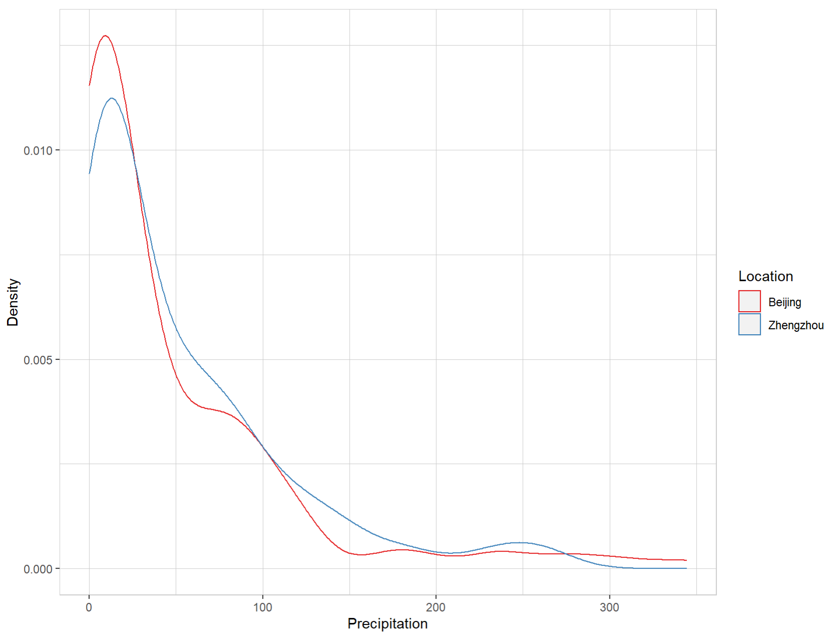

Figure 2. Density curves for rainfall data in Beijing and Zhengzhou.

Table 7. AICs of different distributions to fit rainfall data in Beijing and Zhengzhou.

	Gamma	Normal	Lognormal	Cauchy	Exponential	Wibull	t
Beijing	1116.668	1278.887	1139.087	1232.678	1131.436	1116.893	1231.427
Zhengzhou	1250.871	1391.725	1282.288	1378.279	1259.561	1252.355	1369.012

In addition, we use the goft package in R, which is usually used for data fitting and parameter estimation. In the goft package, the gamma_test function is used to check if the data follows gamma distribution; p-values greater than 0.05 means failure to reject the null, which is the data following gamma distribution. The gamma_fit function estimates the parameters of the gamma distribution. These two functions are used to check if the data follows gamma distribution and to estimate the corresponding parameters in this study. The p-values of the non-zero part of the monthly precipitation data of Beijing and Zhengzhou are 0.9682 and 0.4429, respectively, which means strong evidence to support the data follows gamma distribution. As a consequence, the parameters of the two are estimated by the gamma_fit function as (0.64, 85.55), (0.72, 74.75), respectively. These results combined with AICs in Table 7 indicate that gamma distribution is the most appropriate model among all the considered models to analyze the data.

For the two data sets, the maximum likelihood estimates for each indicator were δ_1=0.144, a_1=0.64, $\delta_2 = 0.045, a_2 = 0.72$, and the difference in CV between the two data sets was 0.1511631. CIs for the four basic methods, F1, F2, BC transformation, PB and the best combination of MOVER with F2F2 are constructed. As shown in Table 8, the intervals are $(-0.0823, 0.3278)$, $(-0.0737, 0.3409)$, $(-0.0797, 0.3592)$, $(-0.0624, 0.3377)$ and $(-0.0661, 0.3427)$, and the corresponding interval lengths are 0.4101, 0.4146, 0.4389, 0.4001 and 0.4088. All the CIs indicate a greater precipitation in Beijing than that in Zhengzhou. We can see that the interval length for MOVER with F2F2 is shorter than the interval length of F1, F2 and BC transformation, and is just slightly longer than the interval length of PB. Since MOVER with F2F2 outperforms PB according to coverage probabilities and tail errors, one can safely use MOVER with F2F2 to analyze such zero-inflated data in practical purposes.

Table 8. CI and interval length obtained by above five methods.

Methods	Fiducial-1	Fiducial-2	BC	PB	MOVER(F2F2)
CI	(-0.0823, 0.3278)	(-0.0737, 0.3409)	(-0.0797, 0.3592)	(-0.0624, 0.3377)	(-0.0661, 0.3427)
Interval length	0.4101	0.4146	0.4389	0.4001	0.4088

6. Conclusions

In this study, four inferential methods, F1, F2, BC transformation, PB and MOVER, were used to construct CIs for the difference between two zero-inflated gamma CVs. The performances of all these CIs were evaluated and compared by the Monte Carlo simulation.

Simulation studies have shown that the F1 CIs and the MOVER CIs combining F2 and F2 return more satisfactory results than other CIs according to the criterions we used in this article. More specifically, F1 CIs are well balanced and the coverage probabilities of such CIs are close to nominal level, even for small sample sizes. MOVER CIs return more satisfactory results than F1 according to coverage probabilities and interval length. In addition, the F1 method is more time efficient than MOVER.

For practical application purposes, one can safely use F1 for some specific situations (small sample sizes, large proportion of zeros, time efficiency). If the sample size is large and the computational cost is not a concern, MOVER consisting of F2F2 is the best choice as it returns more satisfactory results for all criterion used in this study.

One potential future work is to focus on finding possible closed form estimations of gamma parameters. In such a way, we should be able to construct exact CIs. Another possible future work is to extend the current work to analyze censored data, such as the work proposed in [29].

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was supported by the Natural Science Foundation of Hubei Province (Grant No. 2022CFB942), the Teaching and Research Project of Hubei Normal University (Grant No. 2022013), the Open Subject for Teacher Education Research of Hubei Normal University (Grant No. 2022TEZ06), the National Natural Science Foundation of China (Grant No. 12101346) and the Natural Science Foundation of Shandong Province (Grant No. ZR2021QA044).

Conflict of interest

The authors declare no conflict of interest.

References

1. L. Liu, Y. Shih, R. Strawderman, D. Zhang, B. Johnson, H. Chai, Statistical analysis of zero-inflated nonnegative continuous data: a review, *Statist. Sci.*, **34** (2019), 253–279. http://dx.doi.org/10.1214/18-STS681

2. D. Stephenson, K. Kumar, F. Doblas-Reyes, J. Royer, F. Chauvin, S. Pezzulli, Extreme daily rainfall events and their impact on ensemble forecasts of the Indian monsoon, *Mon. Weather Rev.*, **127** (1999), 1954–1966. http://dx.doi.org/10.1175/1520-0493(1999)127¡1954:EDREAT¿2.0.CO;2

3. X. Wang, C. Zou, L. Yi, J. Wang, X. Li, Fiducial inference for gamma distributions: two-sample problems, *Commun. Stat.-Simul. C.*, **50** (2021), 811–821. http://dx.doi.org/10.1080/03610918.2019.1568471

4. X. Wang, M. Li, W. Sun, Z. Gao, X. Li, Confidence intervals for zero-inflated gamma distribution, *Commun. Stat.-Simul. C.*, in press. http://dx.doi.org/10.1080/03610918.2022.2104315

5. M. Hasan, K. Krishnamoorthy, Confidence intervals for the mean and a percentile based on zero-inflated lognormal data, *J. Stat. Comput. Simul.*, **88** (2018), 1499–1514. http://dx.doi.org/10.1080/00949655.2018.1439033

6. X. Wang, X. Li, Generalized confidence intervals for zero-inflated Pareto distribution. *Mathematics*, **9** (2021), 3272. http://dx.doi.org/10.3390/math9243272
7. M. Bugallo, M. Esteban, M. Marey-Pérez, D. Morales, Wildfire prediction using zero-inflated negative binomial mixed models: application to Spain, *J. Environ. Manage.*, **328** (2023), 116788. http://dx.doi.org/10.1016/j.jenvman.2022.116788
8. V. Kerstin, On the distribution of the estimated mean from nonstandard mixtures of distributions, *Commun. Stat.-Theor. M.*, **24** (1995), 1569–1584. http://dx.doi.org/10.1080/03610929508831572
9. K. Muralidharan, B. Kale, Modified gamma distribution with singularity at zero, *Commun. Stat.-Simul. C.*, **31** (2002), 143–158. http://dx.doi.org/10.1081/SAC-9687286
10. T. Kaewprasert, S. A. Niwitpong, S. Niwitpong, Confidence interval estimation for the common mean of several zero-inflated gamma distributions, *Symmetry*, **15** (2023), 67. http://dx.doi.org/10.3390/sym15010067
11. G. Zou, J. Taleban, C. Huo, Confidence interval estimation for lognormal data with application to health economics. *Comput. Stat. Data Anal.*, **53** (2009), 3755–3764. http://dx.doi.org/10.1016/j.csda.2009.03.016
12. P. Ren, G. Liu, X. Pu, Simultaneous confidence intervals for mean differences of multiple zero-inflated gamma distributions with applications to precipitation, *Commun. Stat.-Simul. C.*, **52** (2023), 4705–4716. http://dx.doi.org/10.1080/03610918.2021.1966466
13. T. Kaewprasert, S. A. Niwitpong, S. Niwitpong, Bayesian estimation for the mean of delta-gamma distributions with application to rainfall data in Thailand, *PeerJ*, **10** (2022), 13465. http://dx.doi.org/10.7717/peerj.13465
14. T. Kaewprasert, S. A. Niwitpong, S. Niwitpong, Bayesian confidence intervals for the ratio of the means of zero-inflated gamma distributions with application to rainfall data. *Commun. Stat.-Simul. C.*, in press. http://dx.doi.org/10.1080/03610918.2023.2198176
15. A. Albatineh, B. Kibria, B. Zogheib, Asymptotic sampling distribution of inverse coefficient of variation and its applications: revisited, *International Journal of Advanced Statistics and Probability*, **2** (2014), 15–20. http://dx.doi.org/10.14419/ijasp.v2i1.1475
16. R. Ananthakrishnan, M. Soman, Statistical distribution of daily rainfall and its association with the coefficient of variation of rainfall series, *Int. J. Climatol.*, **9** (1989), 485–500. http://dx.doi.org/10.1002/joc.3370090504
17. W. Puggard, S. A. Niwitpong, S. Niwitpong, Confidence intervals for common coefficient of variation of several Birnbaum-Saunders distributions, *Symmetry*, **14** (2022), 2101. http://dx.doi.org/10.3390/sym14102101
18. N. Niwitpong, S. A. Niwitpong. Confidence intervals for the ratio of coefficients of variation of delta-lognormal distribution, *Appl. Math. Sci.*, **7** (2013), 3811–3818. http://dx.doi.org/10.12988/ams.2013.35248
19. K. Krishnamoorthy, X. Wang, Fiducial confidence limits and prediction limits for a gamma distribution: censored and uncensored cases, *Environmetrics*, **27** (2016), 479–493. http://dx.doi.org/10.1002/env.2408

20. R. Fisher, The fiducial argument in statistical inference, *Annals of Eugenics*, **6** (1935), 391–398. http://dx.doi.org/10.1111/j.1469-1809.1935.tb02120.x
21. Y. Gao, L. Tian, Confidence interval estimation for the difference and ratio of the means of two gamma distributions, *Commun. Stat.-Simul. C.*, in press. http://dx.doi.org/10.1080/03610918.2022.2116646
22. X. Zhou, W. Tu, Confidence intervals for the mean of diagnostic test charge data containing zeros, *Biometrics*, **56** (2000), 1118–1125. http://dx.doi.org/10.1111/j.0006-341X.2000.01118.x
23. E. Wilson, Probable inference, the law of succession, and statistical inference, *J. Am. Stat. Assoc.*, **22** (1927), 209–212. http://dx.doi.org/10.2307/2276774
24. G. Zou, A. Donner, Construction of confidence limits about effect measures: a general approach, *Stat. Med.*, **27** (2008), 1693–1702. http://dx.doi.org/10.1002/sim.3095
25. Y. Tang, MOVER confidence intervals for a difference or ratio effect parameter under stratified sampling, *Stat. Med.*, **41** (2022), 194–207. http://dx.doi.org/10.1002/sim.9230
26. H. Li, M. Tang, W. Wong, Confidence intervals for ratio of two Poisson rates using the method of variance estimates recovery, *Comput. Stat.*, **29** (2014), 869–889. http://dx.doi.org/10.1007/s00180-013-0467-9
27. A. Donner, G. Zou, Closed-form confidence intervals for functions of the normal mean and standard deviation, *Stat. Methods Med. Res.*, **21** (2012), 347–359. http://dx.doi.org/10.1177/0962280210383082
28. X. Li, X. Zhou, L. Tian, Interval estimation for the mean of lognormal data with excess zeros, *Stat. Probabil. Lett.*, **83** (2013), 2447–2453. http://dx.doi.org/10.1016/j.spl.2013.07.004
29. X. Wang, X. Li, L. Zhang, Z. Liu, M. Li, Fiducial inference on gamma distributions: two-sample problems with multiple detection limits, *Environ. Ecol. Stat.*, **29** (2022), 453–475. http://dx.doi.org/10.1007/s10651-022-00528-5