



Confidence Intervals for the Coefficient of Quartile Variation of a Zero-inflated Lognormal Distribution

Noppadon Yosboonruang¹, Sa-Aat Niwitpong^{2*}

¹ Department of Statistics, School of Science, University of Phayao, Phayao 56000, Thailand

² Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand

Abstract

There are many types of skewed distribution, one of which is the lognormal distribution that is positively skewed and may contain true zero values. The coefficient of quartile variation is a statistical tool used to measure the dispersion of skewed and kurtosis data. The purpose of this study is to establish confidence and credible intervals for the coefficient of quartile variation of a zero-inflated lognormal distribution. The proposed approaches are based on the concepts of the fiducial generalized confidence interval, and the Bayesian method. Coverage probabilities and expected lengths were used to evaluate the performance of the proposed approaches via Monte Carlo simulation. The results of the simulation studies show that the fiducial generalized confidence interval and the Bayesian based on uniform and normal inverse Chi-squared priors were appropriate in terms of the coverage probability and expected length, while the Bayesian approach based on Jeffreys' rule prior can be used as alternatives. In addition, real data based on the red cod density from a trawl survey in New Zealand is used to illustrate the performances of the proposed approaches.

Keywords:

Bayesian; Coverage Probability; Lognormal; Fiducial Generalized Confidence Interval; Prior Distribution.

Article History:

Received:	19	March	2021
Revised:	02	July	2021
Accepted:	19	July	2021
Published:	01	August	2021

1- Introduction

Fisheries research, which encompasses many of the natural science fields, frequently provides non-negative data following a skewed distribution [1-3]. Especially, trawl survey data has been used in several studies to estimate the mean fish abundance [1, 4, 5]. The coefficient of variation (CV) is well known as a measure of the dispersion of data that have different units or radically different means. Many researchers have established confidence intervals for the CV of both normal and non-normal distributions [6-8]. However, when trawl survey data are positively skewed with kurtosis, then the coefficient of quartile variation (CQV) is a good alternative statistical tool for measuring the dispersion [9].

The CQV is computed based on the first and third quartiles (Q_1 and Q_3 , respectively). Bonett [9] proposed an approach to establish confidence intervals for the CQVs of several skewed distributions, the performances of which showed its appropriateness for highly skew distributions. After that, Altunkaynak and Gamgam [10] compared the bootstrap confidence interval with Bonett's confidence interval for the CQVs of the same distributions mentioned in Bonett [9]. The results show that the bootstrap confidence intervals performed better than Bonett's approach for small sample sizes. Among the skewed distributions, the lognormal distribution is one of the most widely used for statistical inference and applications. Krishnamoorthy and Mathew [11] used the concepts of the generalized p -value and generalized confidence

* CONTACT: sa-aat.n@sci.kmutnb.ac.th

DOI: <http://dx.doi.org/10.28991/esj-2021-01289>

© 2021 by the authors. Licensee ESJ, Italy. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<https://creativecommons.org/licenses/by/4.0/>).

interval (GCI) to develop confidence intervals for lognormal means and compared them with the large-sample approach and concluded that the proposed approaches were better than the latter approach. Hannig et al. [12] established the fiducial GCI (FGCI) for the ratio of the means of lognormal distributions that performs well for small sample sizes. Lin and Wang [13] established confidence intervals using the quadratic approach for the means of lognormal distributions; their simulation results reveal that their proposed approach performed well in all cases. Nam and Kwon [14] presented Wald-type, Fieller-type, log, and the method of variance estimate recovery (MOVER) approaches for confidence interval estimation of the ratio of lognormal CVs, all of which performed well for medium sample sizes. Later, Hasan and Krishnamoorthy [15] proposed MOVER and fiducial approaches to establish the confidence intervals for the ratio of CVs of lognormal distributions; their approaches were effective for small sample size cases. In addition, Thangjai and Niwitpong [16] established confidence intervals for the signal-to-noise ratio of lognormal distributions using the GCI, large-sample, and MOVER approaches and recommended the GCI approach under all of the circumstances tested. Recently, Thangjai et al. [17] suggested the Bayesian approach for confidence interval construction for both a single lognormal CV and the difference between the CVs of two lognormal distributions.

There have been many studies involving lognormal distributions that are zero-inflated, such as diagnostic test charging [18], the concentration of air contaminants (airborne chlorine) at an industrial site in the United States [19], fish spotter data in fisheries research [20], red cod densities from trawl surveys in New Zealand [4, 5], and rainfall measurements [21, 22]. Furthermore, many researchers have constructed confidence intervals for the parameters of zero-inflated lognormal distributions. Zhou and Tu [23] suggested percentile- t bootstrap interval and likelihood approaches to establish the confidence intervals for the mean of diagnostic test charging data that included zero values. Tian [24] recommended constructing confidence intervals using the concept of generalized variables for the mean. Tian and Wu [25] established confidence intervals for the mean based on adjusted signed log-likelihood ratio statistics, the results of which showed that the proposed approach was suitable in all cases examined. Fletcher [4] proposed three approaches: Aitchison's estimator, a modification of Cox's approach, and a profile likelihood interval to establish confidence intervals for the mean; the profile likelihood interval was the best in terms of the coverage probability. Wu and Hsieh [5] compared the GCI approach to construct confidence intervals for the mean with Aitchison's, modified Land's, profile likelihood, maximum likelihood, and bootstrap approaches and found that GCI was appropriate in all cases. Li et al. [26] presented generalized pivotal quantity and fiducial approaches for interval estimation of the mean; their results demonstrate that the fiducial approach was suitable under all circumstances examined. Recently, Hasan and Krishnamoorthy [27] established confidence intervals for the mean based on the fiducial and MOVER approaches and reported that the former performed the best for small sample sizes. Moreover, Maneerat et al. [22, 28, 29] recommended Bayesian approaches for estimating the confidence intervals for the mean and its functions. Likewise, Yosboonruang et al. [21, 30] also recommended that Bayesian approaches are reasonable for establishing confidence intervals for the CV and the difference between the CVs of zero-inflated lognormal data.

So far, there have not been any published reports focusing on the CQV for a highly skewed zero-inflated lognormal distribution. Herein, we propose fiducial and Bayesian approaches to construct confidence intervals for the CQV of a zero-inflated lognormal distribution. We also report the results of their efficacies via simulation and empirical studies followed by a discussion and conclusions on the research.

2- Methods

$X = (X_1, X_2, \dots, X_m, \dots, X_n)$ is a non-negative random sample from a zero-inflated lognormal distribution. Suppose that the non-zero observations follow a lognormal distribution denoted by $Y_i = \ln(X_i) \sim N(\mu, \sigma^2)$ for $i = 1, 2, \dots, m$, and the zero observations follow a binomial distribution with parameter $\delta' = (n - m)/n$, then the probability of non-zero observations is $\delta = m/n$. The probability density function for a zero-inflated lognormal distribution can be expressed as;

$$f(x; \mu, \sigma^2, \delta) = \begin{cases} \delta' & ; x = 0 \\ \delta \frac{1}{\sqrt{2\pi}\sigma} \left(\frac{1}{x}\right) \exp\left\{-\frac{[\ln(x) - \mu]^2}{2\sigma^2}\right\} & ; x > 0. \end{cases}$$

As proposed by Aitchison [31], the mean and variance of X can be written as:

$$E(X) = \delta \exp(\mu + \sigma^2/2)$$

And,

$$Var(X) = \delta \exp(2\mu + \sigma^2) [\exp(\sigma^2) - \delta],$$

Respectively. In this study, we focus on the CQV based on the lower and upper quartiles (Q_1 and Q_3), which is defined as:

$$\zeta = \frac{Q_3 - Q_1}{Q_3 + Q_1},$$

where Q_1 and Q_3 are the 25th and 75th percentiles, respectively.

According to Hasan and Krishnamoorthy [27], the quartiles can be defined as;

$$Q_r = \exp \left[\mu + \Phi^{-1} \left(\frac{p - (1 - \delta)}{\delta} \right) \sigma \right], \quad (1)$$

where $p = r/4 > 1 - \delta$ and Φ is the standard normal distribution function. Therefore, we can respectively represent Q_1 and Q_3 in Equation (1) as:

$$Q_1 = \exp \left[\mu + \Phi^{-1} \left(\frac{0.25 - (1 - \delta)}{\delta} \right) \sigma \right], \quad \delta > 0.75 \quad (2)$$

And;

$$Q_3 = \exp \left[\mu + \Phi^{-1} \left(\frac{0.75 - (1 - \delta)}{\delta} \right) \sigma \right], \quad \delta > 0.25. \quad (3)$$

In the following sub-sections, we present the methods used to establish the confidence interval for the CQV of a zero-inflated lognormal distribution.

2-1- The FGCI Method

The idea of fiducial inference was first suggested by Fisher [32], after which Hannig et al. [12] illustrated methods to construct the fiducial generalized pivotal quantity (FGPQ). Subsequently, Hannig [33] used Fisher's fiducial concept to prove and express a generalized fiducial recipe that is a generalization of the FGPQ. Recall that X is a zero-inflated lognormal distribution. From Equation 1, the parameters of interest are δ , μ , and σ , and thus their FGPQs are required. According to Li et al. [26] and Hasan and Krishnamoorthy [27], we can express the respective FGPQs for δ , μ , and σ as;

$$R_\delta \sim \text{Beta}(n - m + 0.5, m + 0.5), \quad (4)$$

$$R_\mu = \hat{\mu} - Z \sqrt{\frac{(m-1)\hat{\sigma}^2}{mU}}, \quad (5)$$

And;

$$R_\sigma = \sqrt{\frac{(m-1)\hat{\sigma}^2}{U}}, \quad (6)$$

where $Z \sim N(0,1)$ and $U \sim \chi_{m-1}^2$. By substituting Equations 4 to 6 into Equation 1, we obtain

$$R_{Q_r} = \exp \left[R_\mu + \Phi^{-1} \left(\frac{p - (1 - R_\delta)}{R_\delta} \right) R_\sigma \right], \quad p > 1 - R_\delta. \quad (7)$$

Subsequently, following Equations 2, 3, and 7, the FGPQ for the CQV can be expressed as;

$$R_\zeta = \frac{R_{Q_3} - R_{Q_1}}{R_{Q_3} + R_{Q_1}}.$$

This allows us to establish the $100(1 - \alpha)\%$ FGCI for CQV as;

$$CI_\zeta^{fgci} = [R_\zeta(\alpha/2), R_\zeta(1 - \alpha/2)],$$

where $R_\zeta(\alpha)$ is the 100α -th percentile of R_ζ . The steps to compute the FGCI for the CQV are shown in Algorithm 1.

Algorithm 1

1. Generate x_i , $i = 1, 2, K$, m, K , n from a zero-inflated lognormal distribution;

2. Compute $\hat{\delta}$ and $\hat{\sigma}^2$;
3. Generate $Z \sim N(0,1)$, $U \sim \chi_{m-1}^2$, and $Beta(n-m+0.5, m+0.5)$;
4. Compute R_δ , R_μ , and R_σ ;
5. Compute R_{Q_δ} , R_{Q_μ} , and R_{Q_σ} ;
6. Repeat 3-5 2,000 times;
7. Compute the 95% FGCI for the CQV;
8. Repeat 1-7 10,000 times.

2-2- The Bayesian Approach

Since X is a mixed distribution comprising lognormal and binomial distributions, the joint likelihood function can be written as;

$$L(\delta', \mu, \sigma^2 | x) \propto (\delta')^{n-m} \delta^m \prod_{i=1}^m \frac{1}{\sqrt{\sigma^2}} \exp\left\{-\frac{[\ln(x_i) - \mu]^2}{2\sigma^2}\right\}. \quad (8)$$

This leads to obtaining the Fisher information matrix by using the second derivative of the log-likelihood function as follows:

$$I(\delta', \mu, \sigma^2) = -E\left\{\frac{\partial^2 \ln[L(\delta', \mu, \sigma^2 | x)]}{\partial^2(\delta', \mu, \sigma^2)}\right\} = \text{diag}\left[\frac{n}{\delta'\delta}, \frac{n\delta}{\sigma^2}, \frac{n\delta}{2(\sigma^2)^2}\right].$$

Three different priors to construct credible intervals are present as follows.

Jeffreys' Rule Prior

According to the Fisher information matrix, Jeffreys' rule priors for unknown parameters from binomial and lognormal distributions are derived from the square root of the determinant of the Fisher information matrix. Therefore, the Jeffreys' rule prior for δ' from a binomial distribution by following Harvey and van der Merwe's method [34] is $p(\delta') \propto (\delta')^{-\frac{1}{2}} \delta^{\frac{1}{2}}$. For parameter σ^2 from a lognormal distribution, the Jeffreys' rule prior becomes $p(\sigma^2) = 1/\sigma^3$, and so the Jeffreys' rule prior density function for a zero-inflated lognormal distribution is;

$$p(\delta', \mu, \sigma^2) \propto \frac{1}{\sigma^3} (\delta')^{-\frac{1}{2}} \delta^{\frac{1}{2}}. \quad (9)$$

By combining the likelihood function in Equation 8 and the prior density function in Equation (9), we can express the joint posterior density function as;

$$p(\delta', \mu, \sigma^2 | x) \propto \frac{1}{\text{Beta}\left(n-m+\frac{1}{2}, m+\frac{3}{2}\right)} (\delta')^{n-m-\frac{1}{2}} \delta^{m+\frac{1}{2}} \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{m}}} \exp\left[-\frac{1}{2\sigma^2}(\mu - \hat{\mu})^2\right] \frac{\left[\frac{m\hat{\sigma}^2}{2}\right]^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)} (\sigma^2)^{-\frac{m}{2}} \exp\left[-\frac{m\hat{\sigma}^2}{2\sigma^2}\right], \quad (10)$$

where $\hat{\mu} = \sum_{i=1}^m \ln(x_i)/m$ and $\hat{\sigma}^2 = \sum_{i=1}^m [\ln(x_i) - \hat{\mu}]^2 / (m-1)$.

Moreover, we can obtain the respective posterior distribution of δ' , μ , and σ^2 by using the integral of the function in Equation 10 with respect to the others as follows;

$$p(\delta' | x) = \iint p(\delta', \mu, \sigma^2 | x) d\mu d\sigma^2 \propto \frac{1}{\text{Beta}\left(n-m+\frac{1}{2}, m+\frac{3}{2}\right)} (\delta')^{\left(n-m+\frac{1}{2}\right)-1} \delta^{\left(m+\frac{3}{2}\right)-1}, \quad (11)$$

$$p(\mu | \sigma^2, x) = \int p(\delta', \mu, \sigma^2 | x) d\delta' \propto \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{m}}} \exp\left[-\frac{1}{2\sigma^2}(\mu - \hat{\mu})^2\right], \quad (12)$$

and

$$p(\sigma^2|x) = \iint p(\delta', \mu, \sigma^2|x) d\delta' d\mu \propto \frac{\left[\frac{m\hat{\sigma}^2}{2}\right]^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)} (\sigma^2)^{-1-\frac{m}{2}} \exp\left[-\frac{m\hat{\sigma}^2}{2\sigma^2}\right]. \quad (13)$$

Thus, we can obtain $\delta'|x \sim \text{Beta}(n-m+1/2, m+3/2)$, $\mu|\sigma^2, x \sim N(\hat{\mu}, \sigma^2/m)$, and $\sigma^2|x \sim \text{Inv-Gamma}(m/2, m\hat{\sigma}^2/2)$, respectively.

The Uniform Prior

Since the uniform prior density function for binomial and lognormal distributions is constant [35, 36], then it must be so for a zero-inflated lognormal distribution (i.e., $p(\delta', \mu, \sigma^2) \propto 1$). The joint posterior density function is composed in a similar fashion to Equation 10 as follows:

$$p(\delta', \mu, \sigma^2|x) \propto \frac{1}{\text{Beta}(n-m+1, m+1)} (\delta')^{n-m} \delta^m \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{m}}} \exp\left[-\frac{1}{2\frac{\sigma^2}{m}} (\mu - \hat{\mu})^2\right] \frac{\left[\frac{(m-2)\hat{\sigma}^2}{2}\right]^{\frac{m-2}{2}}}{\Gamma\left(\frac{m-2}{2}\right)} (\sigma^2)^{-1-\frac{m-2}{2}} \exp\left[-\frac{(m-2)\hat{\sigma}^2}{2\sigma^2}\right],$$

where $\hat{\mu} = \sum_{i=1}^m \ln(x_i)/m$ and $\hat{\sigma}^2 = \sum_{i=1}^m [\ln(x_i) - \hat{\mu}]^2 / (m-1)$. Likewise, the posterior distributions of δ , μ , and σ^2 are $\delta'|x \sim \text{Beta}(n-m+1, m+1)$, $\mu|\sigma^2, x \sim N(\hat{\mu}, \sigma^2/m)$, and $\sigma^2|x \sim \text{Inv-Gamma}[(m-2)/2, (m-2)\hat{\sigma}^2/2]$, respectively.

The Normal Inverse Chi-squared Prior

The normal inverse Chi-squared prior was first proposed by Maneerat et al. [22] to establish the highest posterior density interval for the delta-lognormal mean. By following this concept, the joint posterior density function can be expressed as

$$p(\delta', \mu, \sigma^2|x) \propto \frac{1}{\text{Beta}(n-m+1, m+1)} (\delta')^{n-m} \delta^m \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^m [(x_i - \hat{\mu}) + (\hat{\mu} - \mu)]^2\right\} (\sigma^2)^{-\left(\frac{m-1}{2}\right)-1}$$

Accordingly, by using the integral in the same way as for Equations 11 to 13, the respective posterior distributions of δ' , μ , and σ^2 are beta, Student's t , and inverse Chi-squared distributions defined as $\delta'|x \sim \text{Beta}(n-m+1, m+1)$, $\mu|\sigma^2, x \sim t_{m-1}\left\{\hat{\mu}, \sum_{i=1}^m [\ln(x_i) - \hat{\mu}]^2 / [m(m-1)]\right\}$, and $\sigma^2|x \sim \text{Inv-}\chi^2\left\{m-1, \sum_{i=1}^m [\ln(x_i) - \hat{\mu}]^2 / (m-1)\right\}$, respectively.

Algorithm 2

1. Generate x_i , $i = 1, 2, \dots, K$, m, K, n from a zero-inflated lognormal distribution;
2. Compute $\hat{\delta}$ and $\hat{\sigma}^2$;
3. Generate the posterior distribution for $\delta'|x$, $\mu|\sigma^2, x$, and $\sigma^2|x$;
4. Compute Q_1 , Q_3 , and ζ ;
5. Repeat 3-4 2,000 times;
6. Compute the 95% credible interval for the CQV;
7. Repeat 1-6 10,000 times.

The research methodology of this study is shown in Figure 1.

3- Results

3-1-Simulation Studies

The coverage probabilities and expected lengths of the confidence intervals under different scenarios were investigated via Monte Carlo simulation using the R statistical program. A confidence interval with a coverage

probability that is equal to or greater than the nominal confidence level of 0.95 together with the minimum expected length shows the best performing method for a particular set of criteria. The proposed FGCI and Bayesian approaches based on Jeffreys' rule, uniform, and normal inverse Chi-squared priors were tested. For this simulation, we defined the sample sizes (n) as 15, 30, 50, or 100; μ as $-\sigma^2/2$ (following [27]); and σ^2 as 1, 2, 3, 5, or 10. Following the conditional of δ' in Equation 2, δ was varied as 0.80, 0.85, 0.90, or 0.95.

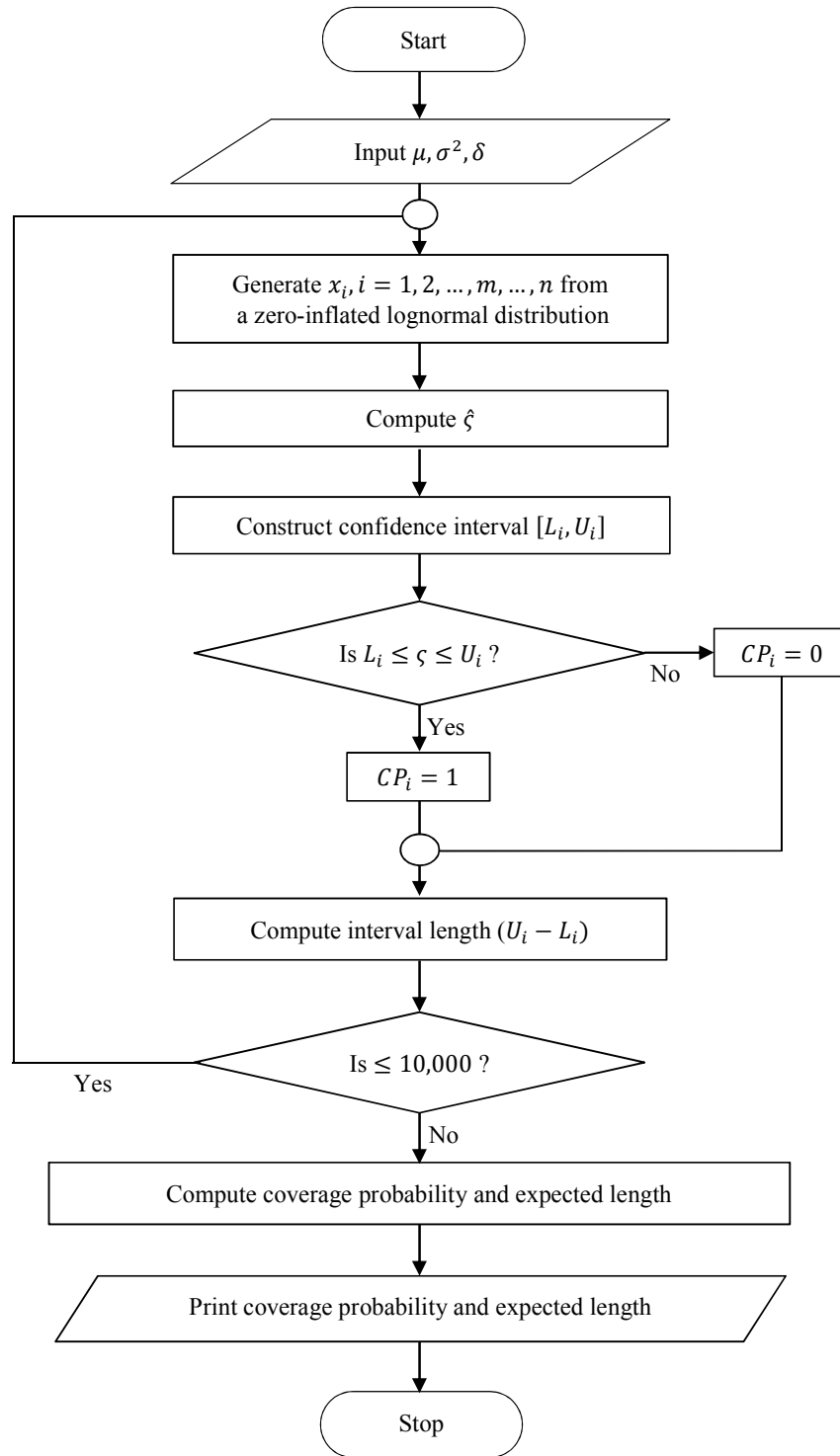


Figure 1. Flowchart of the research methodology.

The FGCI and Bayesian approaches were replicated 2,000 times for 10,000 simulation runs. The results in Table 1 report that the coverage probabilities of FGCI were greater than or close to the nominal confidence level of 0.95 for all cases when $\delta \geq 0.85$. Likewise, the performances of the Bayesian methods based on the uniform and normal inverse

Chi-squared priors were similar to that of FGCI in terms of coverage probability for $\delta = 0.85$ and 0.90 , while for $\delta = 0.95$, the coverage probabilities of the Bayesian-uniform prior method were greater than the others and close to the target for all cases of $n \geq 30$, which is the same as the results for the Bayesian-Jeffreys' rule prior method when $\delta = 0.90$. Moreover, the coverage probabilities were greater than 0.95 for the Bayesian-normal inverse Chi-squared prior method when $\delta = 0.95$ and $n = 50$ or 100 , and for Bayesian-Jeffreys' rule prior method when $\delta = 0.85$ or 0.95 and $n = 50$ or 100 . For $\delta = 0.80$, the coverage probabilities were greater than 0.95 when $\sigma^2 = 10$ for the Bayesian-Jeffreys' rule prior method and $\sigma^2 = 5$ or 10 for the Bayesian-uniform prior and Bayesian-normal inverse Chi-squared prior methods for all sample sizes, as was FGCI for $n = 100$ only.

In terms of the expected length, the Bayesian-normal inverse Chi-squared prior method had the shortest expected lengths for almost all cases for $\delta = 0.80$ or 0.85 except for $\delta = 0.85$, $n = 100$ and $\sigma^2 \leq 3$, for which the Bayesian-Jeffreys' rule prior method had the shortest expected lengths. For $\delta = 0.90$, the expected lengths of the Bayesian-normal inverse Chi-squared prior method were shorter than the others when $\sigma^2 \geq 2$ together with $n = 15$ or 30 . Likewise, the Bayesian-Jeffreys' rule prior method had the shortest expected lengths for $n = 50$ or 100 and $\sigma^2 \leq 5$ whereas the Bayesian-normal inverse Chi-squared prior method had the shortest expected lengths for the other cases. For $\delta = 0.95$, the expected lengths of FGCI were shorter than the others when $n = 15$ or 30 for almost all cases while the Bayesian-normal inverse Chi-squared prior method had the shortest expected lengths for $n = 50$ and $\sigma^2 \geq 3$. Moreover, the expected lengths of the Bayesian-Jeffreys' rule prior method were shorter than the others for $n = 100$ and $\sigma^2 = 2, 3$, or 5 , while the Bayesian-normal inverse Chi-squared prior method had the shortest expected lengths for $n = 100$ and $\sigma^2 = 1$ or 10 . However, the expected lengths of the FGCI and Bayesian methods were slightly different. Summaries of the coverage probabilities and expected lengths of the proposed methods are shown in Figures 2 and 3, respectively. However, the expected lengths of the FGCI and Bayesian methods were not different. Summaries of the coverage probabilities and expected lengths of the proposed methods are shown in Figures 2 and 3, respectively.

Table 1. The coverage probabilities and expected lengths for the 95% confidence intervals and credible intervals for the CQV of a zero-inflated lognormal distribution.

δ	n	σ^2	Coverage probabilities (Expected lengths)				
			FGCI	B-J	B-U	B-NIC	
0.80	15	1	0.9009	0.8305	0.9093	0.8992	
			(0.3895)	(0.3705)	(0.3968)	(0.3876)	
		2	0.9038	0.8578	0.9284	0.9211	
			(0.3227)	(0.3081)	(0.3168)	(0.3108)	
		3	0.8980	0.8816	0.9442	0.9354	
			(0.2616)	(0.2461)	(0.2473)	(0.2429)	
		5	0.8969	0.9197	0.9645	0.9601	
			(0.1762)	(0.1600)	(0.1569)	(0.1539)	
		10	0.8981	0.9719	0.9931	0.9911	
			(0.0801)	(0.0681)	(0.0653)	(0.0637)	
		30	1	0.9171	0.8635	0.9137	0.9118
				(0.3208)	(0.3093)	(0.3220)	(0.3189)
	2		0.9201	0.8899	0.9330	0.9286	
			(0.2547)	(0.2476)	(0.2508)	(0.2486)	
	3		0.9228	0.9084	0.9486	0.9441	
			(0.1969)	(0.1892)	(0.1884)	(0.1869)	
	5		0.9186	0.9383	0.9677	0.9648	
			(0.1226)	(0.1140)	(0.1110)	(0.1101)	
	10		0.9171	0.9736	0.9887	0.9876	
			(0.0466)	(0.0405)	(0.0386)	(0.0382)	
	50		1	0.9357	0.8963	0.9275	0.9263
				(0.2804)	(0.2713)	(0.2792)	(0.2778)
		2	0.9360	0.9093	0.9385	0.9363	
			(0.2151)	(0.2105)	(0.2124)	(0.2114)	
3		0.9325	0.9220	0.9463	0.9450		
		(0.1605)	(0.1559)	(0.1552)	(0.1544)		
5		0.9349	0.9445	0.9671	0.9663		
		(0.0940)	(0.0886)	(0.0866)	(0.0861)		
10		0.9325	0.9758	0.9849	0.9849		
		(0.0317)	(0.0281)	(0.0268)	(0.0267)		

100	1	0.9507 (0.2320)	0.9223 (0.2245)	0.9414 (0.2293)	0.9407 (0.2288)
	2	0.9461 (0.1708)	0.9252 (0.1678)	0.9435 (0.1692)	0.9427 (0.1688)
	3	0.9513 (0.1231)	0.9394 (0.1208)	0.9547 (0.1205)	0.9549 (0.1203)
	5	0.9507 (0.0672)	0.9533 (0.0646)	0.9661 (0.0636)	0.9665 (0.0635)
	10	0.9517 (0.0197)	0.9751 (0.0179)	0.9833 (0.0173)	0.9823 (0.0172)
	15	1	0.9564 (0.3808)	0.9155 (0.3616)	0.9554 (0.3906)
2		0.9581 (0.3212)	0.9252 (0.3071)	0.9614 (0.3183)	0.9557 (0.3121)
3		0.9580 (0.2621)	0.9333 (0.2476)	0.9673 (0.2504)	0.9626 (0.2457)
5		0.9585 (0.1786)	0.9536 (0.1631)	0.9771 (0.1604)	0.9734 (0.1574)
10		0.9582 (0.0818)	0.9786 (0.0700)	0.9876 (0.0672)	0.9851 (0.0655)
30		1	0.9635 (0.3035)	0.9345 (0.2910)	0.9621 (0.3071)
	2	0.9614 (0.2518)	0.9404 (0.2449)	0.9641 (0.2516)	0.9615 (0.2493)
	3	0.9593 (0.1997)	0.9445 (0.1933)	0.9655 (0.1949)	0.9639 (0.1932)
	5	0.9615 (0.1284)	0.9592 (0.1212)	0.9756 (0.1193)	0.9739 (0.1182)
	10	0.9632 (0.0513)	0.9786 (0.0455)	0.9857 (0.0436)	0.9845 (0.0431)
	0.85	1	0.9634 (0.2528)	0.9404 (0.2430)	0.9601 (0.2536)
2		0.9655 (0.2083)	0.9488 (0.2034)	0.9671 (0.2084)	0.9653 (0.2073)
3		0.9627 (0.1631)	0.9517 (0.1594)	0.9679 (0.1613)	0.9658 (0.1604)
5		0.9657 (0.1008)	0.9653 (0.0971)	0.9754 (0.0963)	0.9747 (0.0957)
10		0.9665 (0.0368)	0.9803 (0.0336)	0.9864 (0.0325)	0.9870 (0.0323)
100		1	0.9657 (0.1913)	0.9511 (0.1849)	0.9639 (0.1904)
	2	0.9636 (0.1561)	0.9536 (0.1530)	0.9637 (0.1558)	0.9644 (0.1555)
	3	0.9637 (0.1202)	0.9565 (0.1182)	0.9674 (0.1195)	0.9669 (0.1192)
	5	0.9710 (0.0719)	0.9689 (0.0703)	0.9769 (0.0703)	0.9762 (0.0701)
	10	0.9670 (0.0241)	0.9749 (0.0228)	0.9794 (0.0224)	0.9787 (0.0223)

	15	1	0.9654 (0.3655)	0.9370 (0.3455)	0.9599 (0.3794)	0.9550 (0.3699)
		2	0.9678 (0.3165)	0.9362 (0.3029)	0.9548 (0.3187)	0.9500 (0.3118)
		3	0.9670 (0.2639)	0.9419 (0.2506)	0.9607 (0.2566)	0.9551 (0.2516)
		5	0.9665 (0.1838)	0.9508 (0.1697)	0.9656 (0.1682)	0.9613 (0.1650)
		10	0.9688 (0.0866)	0.9663 (0.0750)	0.9694 (0.0721)	0.9654 (0.0704)
0.90	30	1	0.9736 (0.2812)	0.9551 (0.2692)	0.9716 (0.2868)	0.9693 (0.2835)
		2	0.9683 (0.2431)	0.9517 (0.2360)	0.9606 (0.2455)	0.9597 (0.2429)
		3	0.9683 (0.1983)	0.9496 (0.1924)	0.9591 (0.1966)	0.9553 (0.1945)
		5	0.9709 (0.1315)	0.9574 (0.1255)	0.9645 (0.1250)	0.9599 (0.1238)
		10	0.9671 (0.0548)	0.9673 (0.0495)	0.9692 (0.0478)	0.9667 (0.0473)
0.90	50	1	0.9645 (0.2206)	0.9520 (0.2132)	0.9638 (0.2231)	0.9634 (0.2215)
		2	0.9682 (0.1932)	0.9562 (0.1888)	0.9669 (0.1949)	0.9648 (0.1935)
		3	0.9690 (0.1575)	0.9610 (0.1540)	0.9692 (0.1572)	0.9675 (0.1562)
		5	0.9671 (0.1036)	0.9609 (0.1006)	0.9651 (0.1010)	0.9635 (0.1003)
		10	0.9680 (0.0408)	0.9665 (0.0382)	0.9677 (0.0373)	0.9650 (0.0371)
	100	1	0.9676 (0.1553)	0.9584 (0.1521)	0.9668 (0.1557)	0.9663 (0.1552)
		2	0.9650 (0.1374)	0.9554 (0.1351)	0.9640 (0.1376)	0.9639 (0.1370)
		3	0.9628 (0.1126)	0.9576 (0.1108)	0.9654 (0.1122)	0.9657 (0.1117)
		5	0.9638 (0.0733)	0.9605 (0.0718)	0.9647 (0.0721)	0.9655 (0.0719)
		10	0.9644 (0.0277)	0.9677 (0.0267)	0.9720 (0.0265)	0.9718 (0.0264)
0.95	15	1	0.9669 (0.3399)	0.9431 (0.3208)	0.9652 (0.3589)	0.9591 (0.3488)
		2	0.9672 (0.3071)	0.9376 (0.2937)	0.9501 (0.3159)	0.9449 (0.3081)
		3	0.9596 (0.2615)	0.9285 (0.2493)	0.9433 (0.2608)	0.9352 (0.2550)
		5	0.9619 (0.1869)	0.9355 (0.1747)	0.9483 (0.1760)	0.9427 (0.1724)
		10	0.9674 (0.0905)	0.9528 (0.0795)	0.9546 (0.0770)	0.9488 (0.0753)

30	1	0.9646 (0.2456)	0.9526 (0.2368)	0.9647 (0.2529)	0.9620 (0.2493)
	2	0.9581 (0.2252)	0.9395 (0.2188)	0.9526 (0.2296)	0.9495 (0.2265)
	3	0.9586 (0.1910)	0.9401 (0.1857)	0.9514 (0.1922)	0.9456 (0.1897)
	5	0.9612 (0.1347)	0.9452 (0.1296)	0.9513 (0.1314)	0.9468 (0.1298)
	10	0.9602 (0.0603)	0.9530 (0.0557)	0.9536 (0.0547)	0.9495 (0.0540)
50	1	0.9561 (0.1894)	0.9451 (0.1845)	0.9560 (0.1921)	0.9545 (0.1905)
	2	0.9618 (0.1761)	0.9515 (0.1725)	0.9590 (0.1778)	0.9597 (0.1762)
	3	0.9599 (0.1500)	0.9485 (0.1469)	0.9574 (0.1503)	0.9558 (0.1489)
	5	0.9564 (0.1042)	0.9467 (0.1014)	0.9536 (0.1025)	0.9519 (0.1016)
	10	0.9603 (0.0447)	0.9567 (0.0424)	0.9562 (0.0420)	0.9544 (0.0417)
100	1	0.9563 (0.1352)	0.9496 (0.1328)	0.9551 (0.1355)	0.9534 (0.1349)
	2	0.9588 (0.1259)	0.9528 (0.1241)	0.9567 (0.1259)	0.9562 (0.1254)
	3	0.9586 (0.1069)	0.9522 (0.1054)	0.9568 (0.1066)	0.9565 (0.1061)
	5	0.9589 (0.0737)	0.9517 (0.0724)	0.9554 (0.0728)	0.9535 (0.0725)
	10	0.9576 (0.0306)	0.9531 (0.0296)	0.9535 (0.0295)	0.9522 (0.0247)

Note: B-J, B-U, and B-NIC represented the Bayesian based on Jeffrey's rule, uniform, and normal inverse Chi-squared priors, respectively.

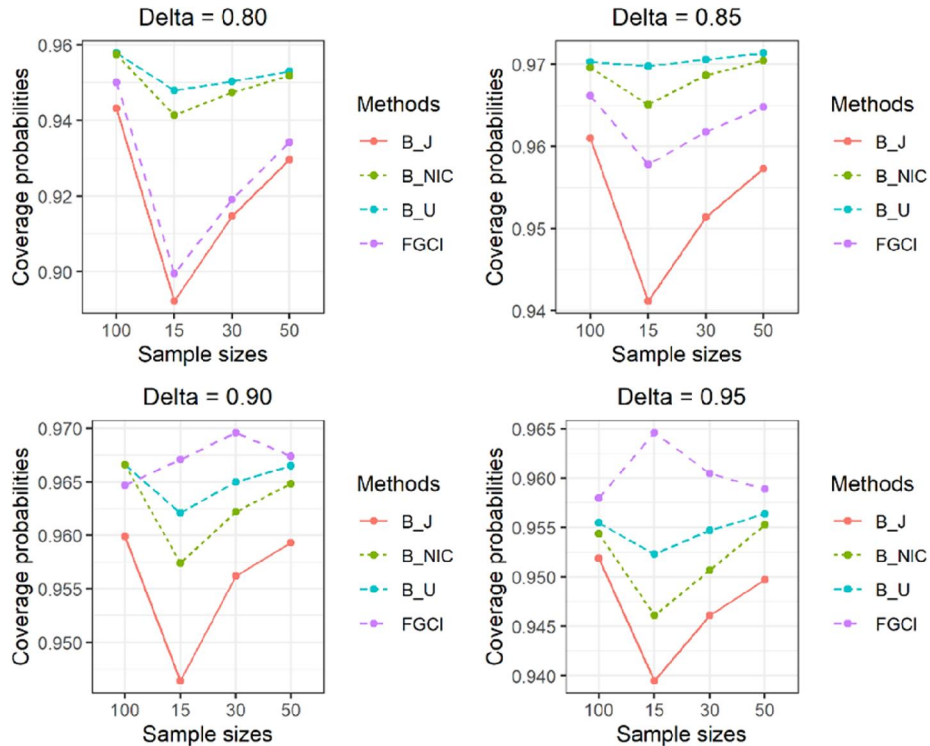


Figure 2. The coverage probabilities of the 95% confidence intervals and the credible intervals for the CQV by the probabilities of non-zero values.

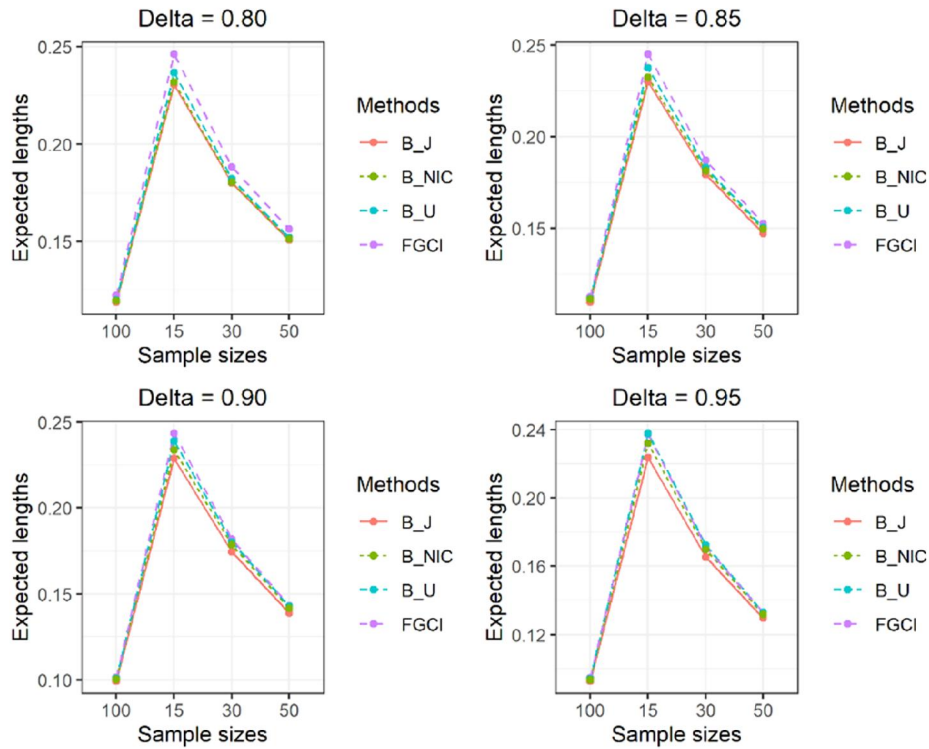


Figure 3. The expected lengths of the 95% confidence intervals and the credible intervals for the CQV by the probabilities of non-zero values.

3-2- An Example Using Real Data

To illustrate the performance of the proposed approaches, we used the data on red cod density (kg/km²) from a trawl survey implemented by the National Institute of Water & Atmospheric Research in New Zealand. This dataset contained data from 67 trawls of which 13 records had no red cod. The remaining 54 trawls were as follows [4]:

10.8, 13.2, 18.2, 19.6, 34.2, 37.0, 41.5, 42.3, 46.1, 46.3, 52.7, 53.8, 55.5, 59.2, 64.5, 66.0, 70.2, 70.6, 74.7, 76.8, 77.6, 78.8, 85.0, 88.1, 89.9, 90.8, 95.4, 100.9, 114.1, 123.2, 131.8, 132.7, 135.1, 141.4, 147.4, 183.0, 223.0, 235.3, 246.5, 253.5, 267.1, 276.4, 293.7, 298.6, 465.2, 584.2, 639.2, 639.3, 663.3, 915.7, 1004.2, 1402.2, 1563.2, 2948.8.

The data are positively skewed and the log-transformation creates a normal distribution, as shown in Figures 4 and 5, respectively. The minimum Akaike information criterion (AIC) was used to test the distribution of the positive data, the results of which in Table 2 indicate that it is a lognormal distribution. Accordingly, the red cod densities from 67 trawls conformed to a zero-inflated lognormal distribution. The summary statistics for this data are $n = 67$, $m = 54$, $\hat{\delta} = 0.81$, $\hat{\mu} = 4.8636$, $\hat{\sigma}^2 = 1.4854$, and $\hat{\xi} = 0.8348$. The 95% confidence intervals and credible intervals for the CQV of the red cod densities are reported in Table 3. Similar to the findings of the simulation study, the Bayesian-uniform prior method performed the best in term of the interval length.

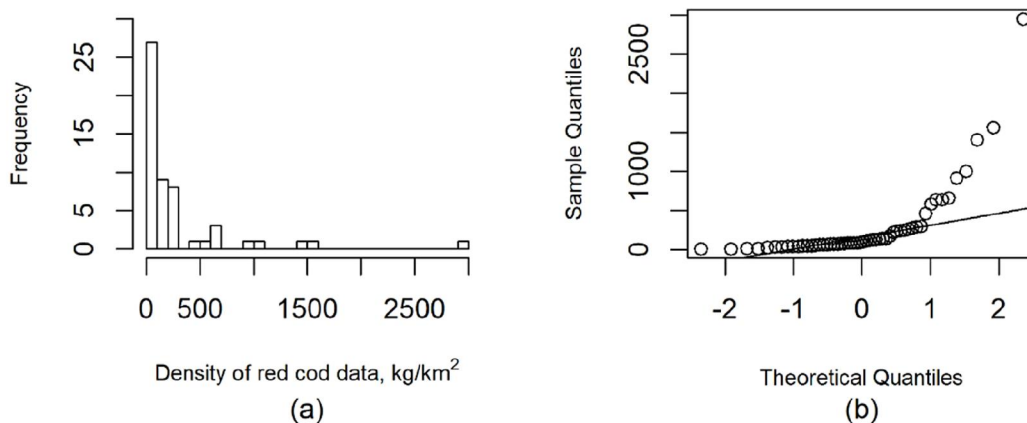


Figure 4. Distribution of the red cod data (a) a histogram (b) a normal Q-Q plot.

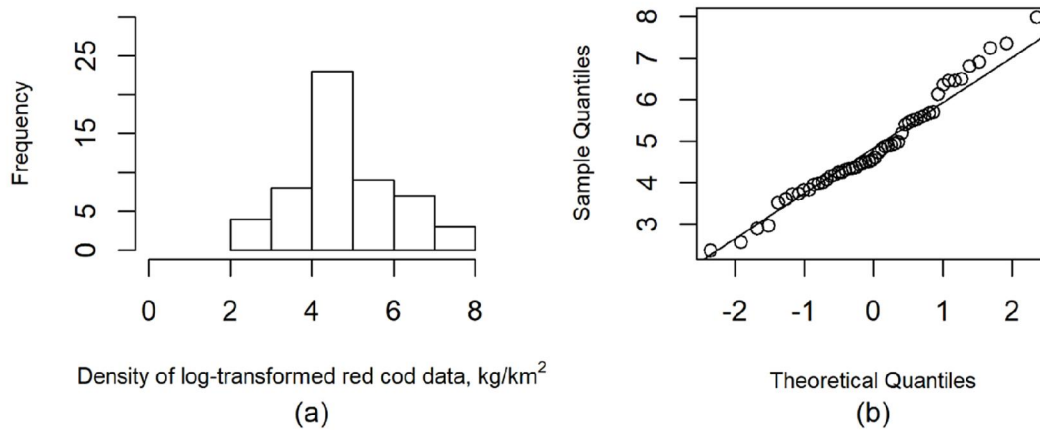


Figure 5. Distribution of log-transformed red cod data (a) a histogram (b) a normal Q-Q plot.

Table 2. AIC values for fitting the distributions.

Distributions	Normal	Lognormal	Cauchy	Exponential	Gamma	Weibull
AIC	826.5334	702.8736	734.7968	722.5387	720.9743	716.9117

Table 3. The 95% confidence intervals and credible intervals for the CQV of the red cod data from the trawls survey in New Zealand.

Approaches	FGCI	B-J	B-U	B-NIC
CIs	(0.7148,0.9563)	(0.7193,0.9544)	(0.7270,0.9555)	(0.7248,0.9580)
Lengths	0.2415	0.2351	0.2285	0.2332

4- Discussion

As mentioned earlier, the CQV can be used to measure the dispersion of highly skewed distributions with kurtosis, such as a zero-inflated lognormal distribution. Interval estimation methods for the CQV of a zero-inflated lognormal distribution by using FGCI and the Bayesian approach are proposed herein. In the present study, it was found that when the probability of non-zero values was 0.80 for large variances, the coverage probabilities of the Bayesian priors were close to or greater than the target. This is due to the posterior distributions of the mean and variance being higher than the variance estimator together with δ' being nearly equal to p in Equation 2, all of which affect obtaining the CQV estimator to generate the quantile functions. FGCI performed well for large sample cases only because the FGPO values for the parameters depend on the sample size of non-zero values. In cases where the probability of non-zero values was greater than 0.80, the coverage probabilities of FGCI were stable and close to the target for all cases. This was similar to the performances of the Bayesian method based on uniform and normal inverse Chi-squared priors where the probability of non-zero values was equal to 0.85 or 0.90. In addition, the performance of the Bayesian-Jeffreys' rule prior method was good for almost all cases with a large sample size or a large variance together with a small sample size. The results of the empirical study using red cod density data from a trawl survey in New Zealand were the same as those of the simulation study in that the Bayesian-uniform prior method is suitable due to the lower and upper bounds covering the CQV of that dataset together with the shortest widths.

5- Conclusion

Herein, we proposed approaches based on FGCI and the Bayesian method to establish confidence intervals for the CQV of a zero-inflated lognormal distribution. The Bayesian approach comprised three priors: Jeffreys' rule, uniform, and normal inverse Chi-squared. Since the CQV is appropriate for extremely skewed data, the red cod data from a trawl survey was used to evaluate the performance of the proposed approaches. The simulation studies show that the confidence intervals constructed with the FGCI and the Bayesian-uniform prior methods performed well for cases with a high proportion of non-zero values since the coverage probabilities were greater than or closed to the target. The performance of the Bayesian-normal inverse Chi-squared prior method was similar to the FGCI and Bayesian-uniform prior methods except for cases with a high proportion of non-zero values together with a small sample size. Therefore, the FGCI and Bayesian methods based on uniform and normal inverse Chi-squared priors are suitable for constructing confidence intervals for the CQV of a zero-inflated lognormal distribution. Last, the Bayesian-Jeffreys' rule prior method is suitable for large sample sizes and a few cases with a large variance.

6- Declarations

6-1-Author Contributions

Conceptualization, S.N.; methodology, N.Y.; software, N.Y.; validation, N.Y. and S.N.; formal analysis, N.Y. and S.N.; investigation, S.N.; resources, N.Y.; data curation, N.Y.; writing—original draft preparation, N.Y.; writing—review and editing, S.N.; visualization, N.Y.; supervision, S.N.; project administration, N.Y. and S.N.; funding acquisition, S.N. All authors have read and agreed to the published version of the manuscript.

6-2-Data Availability Statement

The dataset of red cod densities, taken from a fisheries trawl survey in New Zealand, were compiled in a research article by Fletcher (2008) as <https://doi.org/10.1007/s10651-007-0046-8>.

6-3-Funding

The second author is grateful to King Mongkut's University of Technology North Bangkok. Grant No. KMUTNB-FF-65-22.

6-4-Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

- [1] Pennington, Michael. "Efficient Estimators of Abundance, for Fish and Plankton Surveys." *Biometrics* 39, no. 1 (March 1983): 281. doi:10.2307/2530830.
- [2] Syrjala, S. "Critique on the Use of the Delta Distribution for the Analysis of Trawl Survey Data." *ICES Journal of Marine Science* 57, no. 4 (August 2000): 831–842. doi:10.1006/jmsc.2000.0571.
- [3] Terceiro, M. "The statistical properties of recreational catch rate data for some fish stocks off the northeast U.S. coast." *Fishery Bulletin* 101, no. 3 (Apr 2003): 653-672.
- [4] Fletcher, D. "Confidence intervals for the mean of the delta-lognormal distribution." *Environmental and Ecological Statistics* 15, no. 2 (Jun 2008): 175-189. doi:10.1007/s10651-007-0046-8.
- [5] Wu, W.-H., and H.-N. Hsieh. "Generalized confidence interval estimation for the mean of delta-lognormal distribution: an application to New Zealand trawl survey data." *Journal of Applied Statistics* 41, no. 7 (Jan 2014): 1471-1485. doi:10.1080/02664763.2014.881780.
- [6] Verrill, S., and R. A. Johnson. "Confidence bounds and hypothesis tests for normal distribution coefficients of variation." *Communications in Statistics - Theory and Methods* 36, no. 12 (Aug 2007): 2187-2206. doi:10.1080/03610920701215126.
- [7] Mahmoudvand, R., and H. Hassani. "Two new confidence intervals for the coefficient of variation in a normal distribution." *Journal of Applied Statistics* 36 no. 4 (Apr 2009): 429-442. doi:10.1080/02664760802474249.
- [8] Hayter, A. J. "Confidence bounds on the coefficient of variation of a normal distribution with applications to win-probabilities." *Journal of Statistical Computation and Simulation* 85, no. 18 (Apr 2015): 3778-3791. doi:10.1080/00949655.2015.1035654.
- [9] Bonett, Douglas G. "Confidence Interval for a Coefficient of Quartile Variation." *Computational Statistics & Data Analysis* 50, no. 11 (July 2006): 2953–2957. doi:10.1016/j.csda.2005.05.007.
- [10] Altunkaynak, B., and H. Gamgam. "Bootstrap confidence intervals for the coefficient of quartile variation." *Communications in Statistics - Simulation and Computation* 48, no. 7 (2019): 2138-2146. doi:10.1080/03610918.2018.1435800.
- [11] Krishnamoorthy, K., and T. Mathew. "Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals." *Journal of Statistical Planning and Inference* 115, no. 1 (Jul 2003): 103-121. doi:10.1016/S0378-3758(02)00153-2.
- [12] Hannig, J., H. Iyer, and P. Patterson. "Fiducial generalized confidence intervals." *Journal of the American Statistical Association* 101 no. 473 (Mar 2006): 254-269. doi:10.1198/016214505000000736.
- [13] Lin, S. H., and R. S. Wang. "Modified method on the means for several log-normal distributions." *Journal of Applied Statistics* 40, no. 1 (Jan 2013): 194-208. doi:10.1080/02664763.2012.740622.
- [14] Nam, J., and D. Kwon. "Inference on the ratio of two coefficients of variation of two lognormal distributions." *Communications in Statistics - Theory and Methods* 46, no. 17 (May 2016):8575-8587. doi:10.1080/03610926.2016.1185118.

- [15] Hasan, Md Sazib, and K. Krishnamoorthy. "Improved Confidence Intervals for the Ratio of Coefficients of Variation of Two Lognormal Distributions." *Journal of Statistical Theory and Applications* 16, no. 3 (September 2017): 345. doi:10.2991/jsta.2017.16.3.6.
- [16] Thangjai, W., and S.-A. Niwitpong. "Confidence intervals for the signal-to-noise ratio and difference of signal-to-noise ratios of log-normal distributions." *Stats* 2, no. 1 (Feb 2019): 164-173. doi:10.3390/stats2010012.
- [17] Thangjai, Warisa, Sa-Aat Niwitpong, and Suparat Niwitpong. "Bayesian Confidence Intervals for Coefficients of Variation of PM10 Dispersion." *Emerging Science Journal* 5, no. 2 (April 1, 2021): 139–154. doi:10.28991/esj-2021-01264.
- [18] Callahan, C. M., J. G. Kesterson, and W. M. Tierney. "Association of symptoms of depression with diagnostic test charges among older adults." *Annals of Internal Medicine* 126, no. 6 (Mar 1997): 426-432. doi:10.7326/0003-4819-126-6-199703150-00002.
- [19] Owen, W. J., and T. A. DeRouen. "Estimation of the mean for lognormal data containing zeroes and left-censored values, with applications to the measurement of worker exposure to air contaminants." *Biometrics* 36 no. 4 (Dec 1980): 707-719. doi:10.2307/2556125.
- [20] Lo, N. C., L. D. Jacobson, and J. L. Squire. "Indices of relative abundance from fish spotter data based on delta-lognormal models." *Canadian Journal of Fisheries and Aquatic Sciences* 49 (Dec 1992): 2515-2526. doi:10.1139/f92-278.
- [21] Yosboonruang, N., S.-A. Niwitpong, and S. Niwitpong. "Measuring the dispersion of rainfall using Bayesian confidence intervals for coefficient of variation of delta-lognormal distribution: a study from Thailand." *PeerJ* 7 (Jul 2019): e7344. doi:10.7717/peerj.7344.
- [22] Maneerat, P., S.-A. Niwitpong, and S. Niwitpong. "Bayesian confidence intervals for a single mean and the difference between two means of delta-lognormal distributions." *Communications in Statistics - Simulation and Computation* (May 2019). doi:10.1080/03610918.2019.1616095.
- [23] Zhou, X. H., and W. Tu. "Confidence intervals for the mean of diagnostic test charge data containing zeros." *Biometrics* 56 no. 4 (Dec 2000): 1118-1125. doi:10.1111/j.0006-341x.2000.01118.x.
- [24] Tian, L. "Inferences on the common coefficient of variation." *Statistics in Medicine* 24, no. 14 (Apr 2005): 2213-2210. doi:10.1002/sim.2088.
- [25] Tian, L., and J. Wu. "Confidence intervals for the mean of lognormal data with excess zeros." *Biometrical Journal* 48, no. 1 (Feb 2006): 149-156. doi:10.1002/bimj.200510155.
- [26] Li, X., X. Zhou, and L. Tian. "Interval estimation for the mean of lognormal data with excess zeros." *Statistics & Probability Letters* 83, no. 11 (Nov 2013): 2447-2453. doi:10.1016/j.spl.2013.07.004.
- [27] Hasan, M. S., and K. Krishnamoorthy. "Confidence intervals for the mean and a percentile based on zero-inflated lognormal data." *Journal of Statistical Computation and Simulation* 88, no. 8 (Feb 2018): 1499-1514. doi:10.1080/00949655.2018.1439033.
- [28] Maneerat, P., S.-A. Niwitpong, and S. Niwitpong. "A Bayesian approach to construct confidence intervals for comparing the rainfall dispersion in Thailand." *PeerJ* 8 (Aug 2020): e8502. doi:10.7717/peerj.8502.
- [29] Maneerat, P., S.-A. Niwitpong, and S. Niwitpong. "Bayesian confidence intervals for the difference between variances of delta-lognormal distributions." *Biometrical Journal* 62, no. 7 (Jun 2020): 1769-1790. doi:10.1002/bimj.201900079.
- [30] Yosboonruang, N., S.-A. Niwitpong, and S. Niwitpong. "The Bayesian confidence intervals for measuring the difference between dispersions of rainfall in Thailand." *PeerJ* 8 (Aug 2020): e9662. doi:10.7717/peerj.9662.
- [31] Aitchison, John. "On the Distribution of a Positive Random Variable Having a Discrete Probability Mass at the Origin." *Journal of the American Statistical Association* 50, no. 271 (September 1955): 901–908. doi:10.1080/01621459.1955.10501976.
- [32] Fisher, R. A. "Inverse Probability." *Mathematical Proceedings of the Cambridge Philosophical Society* 26, no. 4 (October 1930): 528–535. doi:10.1017/s0305004100016297.
- [33] Hannig, J. "On generalized fiducial inference." *Statistica Sinica* 19, no. 2 (Apr 2009): 491-544.
- [34] Harvey, J., and A. J. van der Merwe. "Bayesian confidence intervals for means and variances of lognormal and bivariate lognormal distributions." *Journal of Statistical Planning and Inference* 142, no. 6 (Jun 2012): 1294-1309. doi:10.1016/j.jspi.2011.12.006.
- [35] Bolstad, William M., and James M. Curran. "Introduction to Bayesian Statistics, Third Edition" (August 25, 2016). doi:10.1002/9781118593165.
- [36] Kalkur, T. A., and A. Rao. "Bayes estimator for coefficient of variation and inverse coefficient of variation for the normal distribution." *International Journal of Statistics and Systems* 12, no. 6 (2017): 721-732.