Austrian Journal of Statistics
2023, Volume 52, 63–81.
http://www.ajs.or.at/
doi:10.17713/ajs.v52i5.1534

Alternative Confidence Interval for Variability Parameters in the Normal Distribution with Applications to Stock Exchange Index Data Set

Patarawan Sangnawakij
Thammasat University

Wanwarat Anlamlert
Thammasat University

Abstract

This work focusses on estimation the standard deviation (SD) and coefficient of variation (CV) in the normal distribution. These two measures are useful in applications and widely used to report the spread or variability of continuous data. We develop the confidence intervals for these parameters using the two pivotal quantity methods with the unbiased estimator of SD. The first confidence interval uses the pivot function based on a chi-square distribution and the second one is based on a generalized pivotal quantity. The performance of our approaches is conducted via simulations. We show that the confidence intervals for SD and CV based on the new pivots have coverage probabilities greater than existing confidence intervals. Furthermore, they have acceptable short expected lengths. We also provide two real-data sets on the SET50 index of Thailand to demonstrate the proposed methods.

Keywords: coefficient of variation, corrected value, pivot method, SET50 index, standard deviation.

1. Introduction

An important characteristic of any dataset, especially continuous outcome, is the variation in the data. This represents how much the values differ or spread out from the central or mean. Measures of variation widely used in practices are variance, standard deviation (SD), and coefficient of variation (CV). The two former measures are closely related. As in computation, the SD is presented by the root of average distance between square of each point value and its mean. In other word, the SD is the root of variance, and gives the same scale as the data or mean. However, the variance is expressed in squared unit which may be difficult to intuitively understand. If the data sets are concentrated in the same mean and expressed in the same unit, the SD would be used to compare the distribution of those data sets. For the CV, it is defined as the ratio of the SD to the mean. This is used to report the dispersion of a variable as the previous measure (Pearson 1896). In contrast to the SD or variance, the CV is a unit-free measure and usually computed as a percentage. Therefore, it can be used to compare several variables expressed in different units. This is often used in economics to make an assessment of income inequalities, and in social sciences to assess demographic heterogeneity and social aggregates such as race, gender, and age. If the SD or CV estimator provides a single statistic for estimating the unknown parameters of a population, in general it is called the point estimator.

In statistical inference, interval estimation is also an important tool used to estimate the unknown parameter. It is applied in survival areas, including scientific and applied researches. This method provides two statistics, usually noted as lower and upper limits, which guarantee that the unknown parameter contains within a given probability or confidence level (Casella and Berger 2002). This is called the confidence interval. According to the normal distribution, the traditional confidence interval for the true SD is constructed based on the pivotal quantity which follows a chi-square distribution with $n-1$ degrees of freedom (Bonett 2006), where n is the sample size. This pivot is a function of $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$, which is the unbiased estimator for the population variance, where X_i is the i-th observed value and $\bar{X}$ is the sample mean (Choi and Son 1990; Turner, Robinson, and Mavris 2013). We will show this pivot by formula again in the next section. However, it is easy to verify that the sample SD, or $S = \sqrt{S^2}$, is the biased estimator of its population SD. If S is used, particularly when sample sizes are small, we must caution that it may have an effect to the performance of estimator in estimating the SD. This issue may not only occur in point estimation, but can be found in interval estimation. Furthermore, this problem can be lead to a low performance in estimating the CV, as the traditional estimator, namely $\hat{\tau}$, is depended on S.

Normal distribution is a probability model applied in widely research areas (Clarren, Chudley, Wong, Friesen, and Brant 2010; Javari 2017; Protopapas and Parrila 2018; Sreejith and Pathy 2012; Zhaohong, Mengqiang, Yan, and Xin 2009), and involves in basis probability theories of statistical inference. The construction of confidence intervals for parameter and function of parameters in this distribution is therefore the subject of this work. From the literature review, a few studies have addressed the construction of confidence intervals for the population SD. This is because the conventional estimator given in (5) is usually used. However, in this paper we will derive the confidence interval using the new pivotal quantity to be an alternative method. For the CV, although there have been studies introduced the confidence intervals, they were generally constructed based on the normal approximation or large-sample methods (Donner and Zou 2010; Mahmoudvand and Hassani 2009; Panichkitkosolkul 2009; Vangel 1996). Furthermore, confidence intervals proposed in the previous works were related to the traditional estimator $\hat{\tau}$, and the performance of those confidence intervals did not investigate in many situations via simulations, such as high-level variances or CVs. Therefore, in the current work we will introduce the new method to build the confidence interval for the population CV using the proposed pivot and the method of variance estimates recovery. The performance of estimators will be also studied in a comprehensive situation.

The rest of this paper is organized as follows. Section 2 describes the unbiased estimator for the SD in the normal distribution. Then it is used to calculate the CV. In Section 3 and Section 4, the methods used to construct the new confidence intervals for the SD and CV parameters are offered, respectively. In Section 5, we conduct the performance of confidence intervals using simulation in several situations. Then, data application is presented in Section 6. Since stock index values around the world are volatility during coronavirus outbreak 2021 to early 2022 (Demir, Kizys, Rouatbi, and Zaremba 2021; Vera 2022; Zhang, Wang, Haq, and Nosheen 2022), also in Thailand, it is interesting to measure the dispersion of index values using a suitable approach. The two data examples on the total values of SET50 index of Thailand will be applied here. The conclusions with a brief discussion are given in Section 7.

2. Point estimation for variability parameter

2.1. Notations

Suppose that $X = (X_1, X_2, \cdots, X_n)$ is a random variable of size n from a normal distribution with two unknown parameters, mean $E(X) = \mu$ and standard deviation $SD(X) = \sigma$. This is denoted as $X \sim N(\mu, \sigma)$. The probability density function of X is given by

$$f_X(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \tag{1}$$

where $-\infty < x < \infty$, $-\infty < \mu < \infty$, and $\sigma > 0$ or the variance of X is $\sigma^2 > 0$ (Baron 2014). The population CV of X is denoted as $\tau = \sigma/\mu$. Since the parameters given in model (1) are unknown, it must be estimated using a function of random sample of size n. Consider the unbiased estimators for μ and σ^2, they are given by

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \tag{2}$$

respectively. Hence, σ and τ can be estimated by $S = \sqrt{S^2}$ and $\hat{\tau} = S/\bar{X}$, respectively. Note that the traditional point estimators given above are simple and widely used in basic researches. Another popular method is maximum likelihood (ML) estimation. This is based on finding the parameter value that maximizes the likelihood of observations given the parameter. The ML estimator has good properties, especially the invariance qualification (Rossi 2010). Under (1), the ML estimator for σ^2 is given by $S^2_{ml} = \sum_{i=1}^{n}(X_i - \bar{X})^2/n$. This is the asymptotically unbiased estimator of σ^2. Panichkitkosolkul (2009) used S^2_{ml} and introduced the estimator for τ as $\hat{\tau}_{wp} = S_{ml}/\bar{X}$, where $S_{ml} = \sqrt{S^2_{ml}}$.

To investigate the performance of S^2 in terms of unbiasedness, the function of random sample $A = (n-1)S^2/\sigma^2$ is applied. A has a chi-square distribution with $n-1$ degrees of freedom (or $A \sim \chi^2_{n-1}$). This follows that the variance and expected value of S^2 are $Var(S^2) = 2\sigma^4/(n-1)$ and $E(S^2) = \sigma^2$. From the later property, it can be concluded that S^2 given in (2) is the unbiased estimator for σ^2. However, if the capability of S is conducted, we have the fact that its expected value equals to

$$E(S) = \left(\frac{\sqrt{2}\Gamma(n/2)}{\sqrt{n-1}\Gamma((n-1)/2)}\right)\sigma, \tag{3}$$

and the bias of S is

$$Bias(S) = \sigma\left(\frac{\sqrt{2}\Gamma(n/2)}{\sqrt{n-1}\Gamma((n-1)/2)} - 1\right),$$

where $\Gamma(\cdot)$ is a gamma function. Since $E(S) \neq \sigma$, S is the *biased estimator* of σ, which can be allowed under- or over-estimation. Evidently, the lack of a good property of point estimator can be occurred in estimating σ if S is applied. To find a function of random sample which has the expected value equal to σ, (3) is adjusted. It can be rewritten as

$$E\left(\frac{\sqrt{n-1}\Gamma((n-1)/2)}{\sqrt{2}\Gamma(n/2)}S\right) = \sigma.$$

Thus, the unbiased estimator for σ is accomplished as

$$S_{ub} = \left(\frac{\sqrt{n-1}\Gamma((n-1)/2)}{\sqrt{2}\Gamma(n/2)}\right)S = g(n)S. \tag{4}$$

Equation (4) can be computed if $n > 1$. Furthermore, if n is large, $g(n)$ converges to one, and S_{ub} and S are identical. According to (3), the traditional statistic $\hat{\tau} = S/\bar{X}$ can be also provided a large bias in estimation. In this paper, S_{ub} is then suggested to construct the estimated CV. It is given by $\hat{\tau}_{ub} = S_{ub}/\bar{X}$. We wish this estimator would be better performance than the existing traditional estimator in some cases.

2.2. Simulation study for point estimator

This section provides a simulation to display the performance of S and S_{ub}. The data were generated from an $N(\mu = 0, \sigma)$, where σ was given by 1, 3, 9, and 15, reflecting variation of the data from small to large. The sample size (n) was given as 10, 20, 30, 50, and 100. For each scenario, we performed 10,000 simulation runs using R (R Core Team 2022). The bias and variance of S_{ub} obtained from simulations were approximated by

$$B(S_{ub}) = \frac{\sum_{i=1}^{10{,}000}(S_{ub})_i}{10{,}000} - \sigma \quad \text{and} \quad Var(S_{ub}) = \frac{\sum_{i=1}^{10{,}000}\left((S_{ub})_i - E(\hat{\sigma})\right)^2}{10{,}000},$$

respectively. Those equations were also applied to estimate the bias and variance of S. The performance of two estimators is shown in Figure 1. Evidently, the biases of S_{ub} were very closer to zero than those of S in all situations of the study. Biases of S_{ub} did not depend on the true parameter σ and n. Meanwhile, S provided small biases if σ was small ($\sigma \leq 3$). Furthermore, it can be seen that these two estimators had similar variances. Thus, S_{ub} is strongly recommended to estimate the SD in the normal distribution.

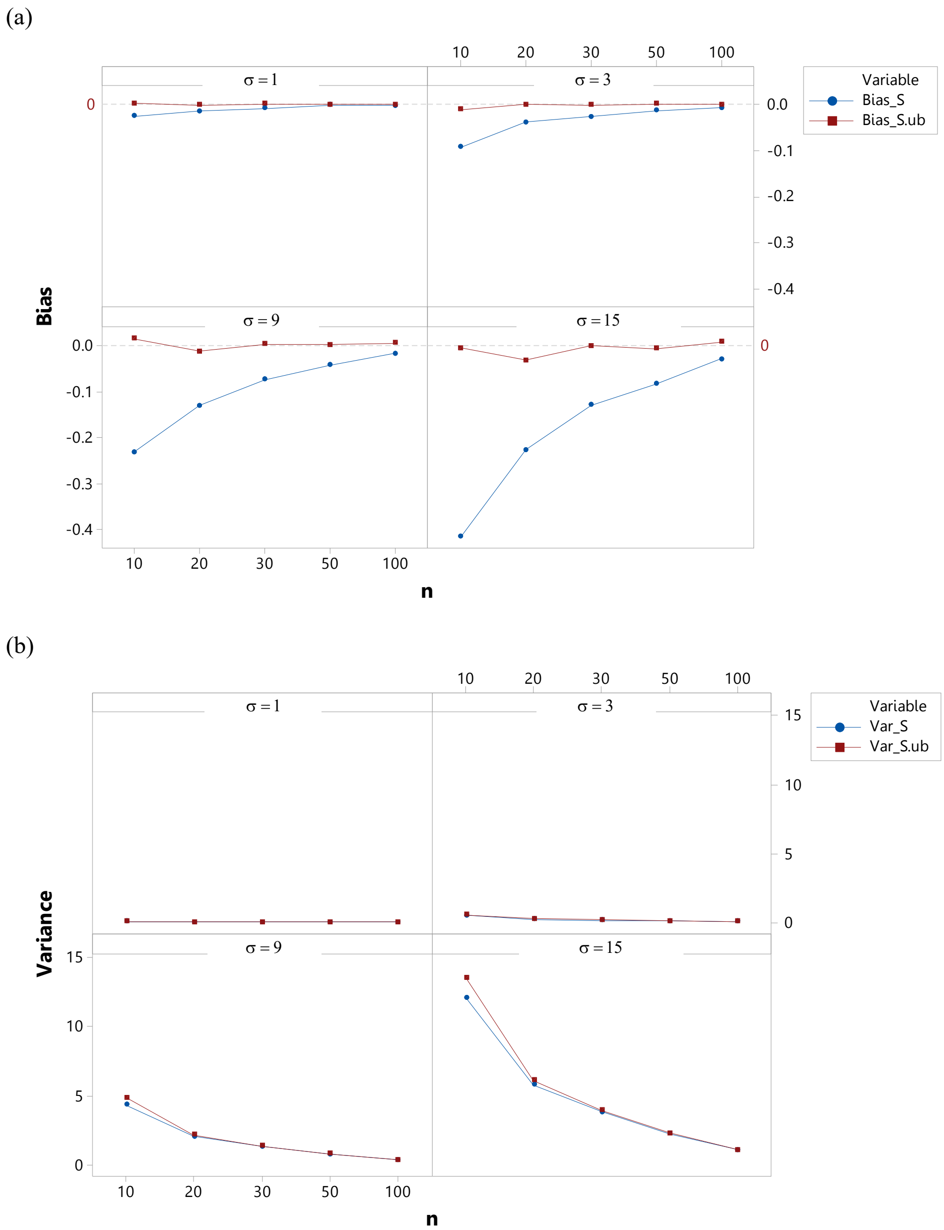


Figure 1: Bias and variance of the two estimators (S and S_{ub}) using simulations

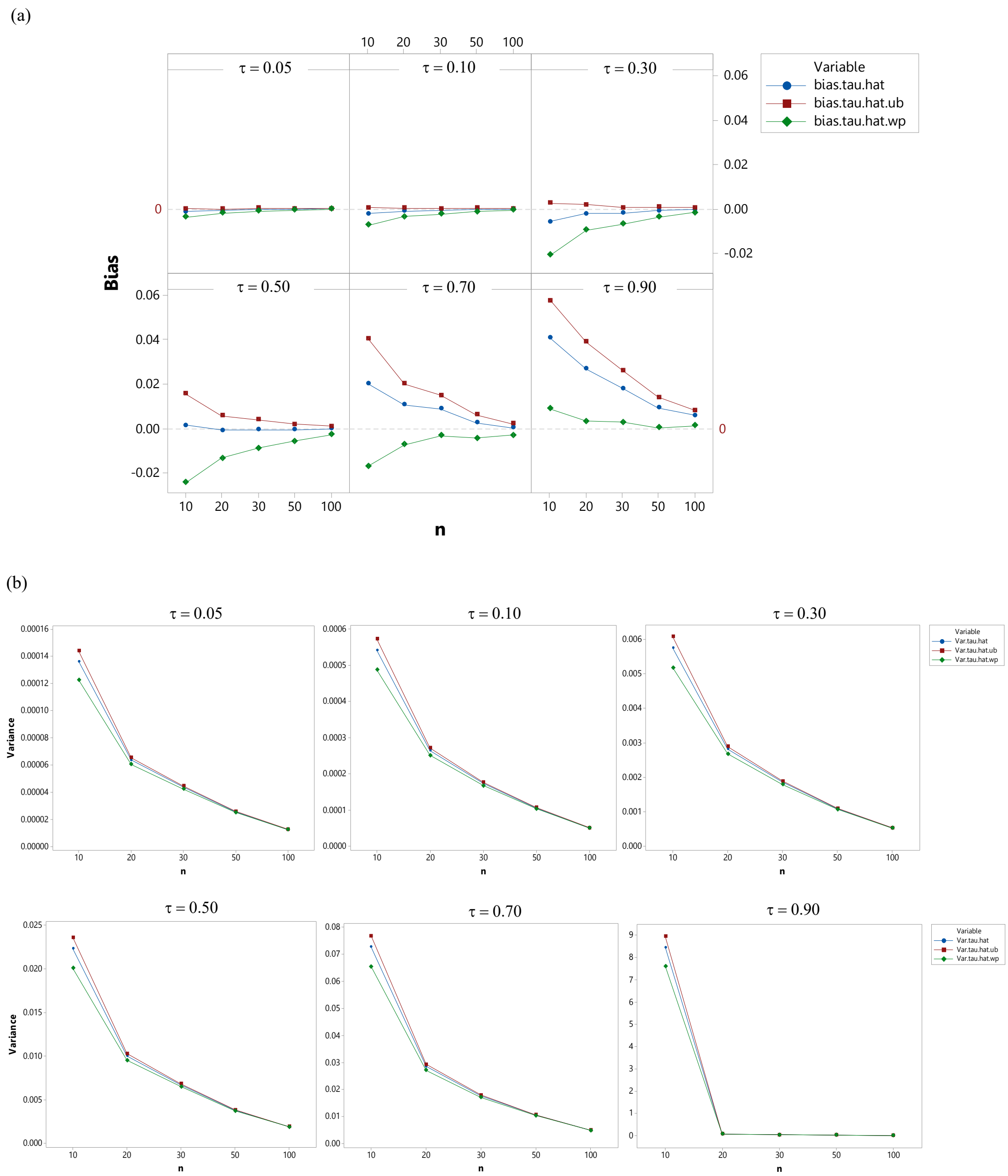


Figure 2: Bias (a) and variance (b) of the three estimated coefficients of variation ($\hat{\tau}$, $\hat{\tau}_{wp}$, and $\hat{\tau}_{ub}$) using simulations

To evaluate the performance of estimator for CV, we generated the data from an $N(\mu, \sigma = 5)$. The true CV (τ) was varied and given as 0.05, 0.1, 0.3, 0.5, 0.7, and 0.9. Then, μ was computed by $\mu = \sigma/\tau$. We used the sample sizes and number of simulation runs as noted in the previous setting. $\hat{\tau}$, $\hat{\tau}_{wp}$, and $\hat{\tau}_{ub}$, were investigated in terms of bias. The results are given in Figure 2. It can be seen that biases of all estimators converged to zero when n was increased. $\hat{\tau}_{ub}$ is better than the others and worked well to estimate τ if $\tau \leq 0.3$ or small deviation situations, as its bias was closest to zero. $\hat{\tau}$ performed better than the compared estimators if $0.3 < \tau \leq 0.5$. In high variation situations or $\tau > 0.5$, $\hat{\tau}_{wp}$ has the smallest bias values. These are very interesting results and it remains the user to apply the suitable method.

From the results, S_{ub} provides a good performance in estimation in general cases. We then use this estimator for constructing the confidence intervals for σ and τ. The details are given in the next sections.

3. Confidence interval for standard deviation

Confidence interval estimation for σ in the normal distribution has not been often introduced. A crucial reason is that in general a $(1-\alpha)100\%$ confidence interval for σ, which is given by

$$CI_{trad} = \left(\sqrt{\frac{(n-1)}{\chi^2_{1-\alpha/2,n-1}}}S, \sqrt{\frac{(n-1)}{\chi^2_{\alpha/2,n-1}}}S \right), \tag{5}$$

is very popular and widely used in basic statistic courses and researches, where $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-\alpha/2,n-1}$ are the $(\alpha/2)$th and $(1-\alpha/2)$th quantiles of a χ^2_{n-1} distribution (Banik, Albatineh, Shawiesh, and Kibria 2014). This traditional interval estimator is constructed based on the probability statement $P(\chi^2_{\alpha/2,n-1} \leq A \leq \chi^2_{1-\alpha/2,n-1}) = 1-\alpha$, where $A = (n-1)S^2/\sigma^2$. However, as we show in the simulation that S_{ub} has small bias and variance. In the following subsections, the new confidence intervals for σ are introduced using S_{ub} with the two approaches: adjusted pivot method and generalized pivotal method.

3.1. Adjusted pivot method

Let $X_i \sim N(\mu, \sigma)$, for $i = 1, 2, \cdots, n$, and X_i's are independent. At the beginning, we assume that the pivotal quantity $B = (n-1)S^2_{ub}/\sigma^2$ has a χ^2_{n-1} distribution. This mimics the existing pivot A. Hence, it is easy to obtain a $(1-\alpha)100\%$ confidence interval for σ using B, which is given as

$$CI_{adj} = \left(\sqrt{\frac{(n-1)}{\chi^2_{1-\alpha/2,n-1}}}S_{ub}, \sqrt{\frac{(n-1)}{\chi^2_{\alpha/2,n-1}}}S_{ub} \right). \tag{6}$$

However, the pivot B used in the above procedure is simply obtained by substituting S_{ub} into S of (5). A question arises whether B follows a χ^2_{n-1} distribution or exactly provides a good property of confidence interval for σ. An interesting idea introduced in this current work is proposed. We adjust the constant in B to obtain the new pivotal quantity C, where C still has a χ^2_{n-1} distribution, and is related to

$$P\left(\chi^2_{\alpha/2,n-1} \leq C \leq \chi^2_{1-\alpha/2,n-1}\right) = 1-\alpha. \tag{7}$$

The confidence interval for σ obtained from (7) must be satisfied the two properties.

1) The confidence interval has an acceptable coverage probability or close to a given probability level at $1-\alpha$
2) The confidence interval has a short length interval.

Here, we suggest to use $n - aa$ in the numerator of B, instead of $n-1$, where aa denotes *the corrected value* which is a number that greater than or equal to 1. The correction used here is done to adjust the random function C, to have an exact chi-square distribution with $n-1$ degrees of freedom. Thus, the proposed pivotal quantity is given by

$$C = \frac{(n-aa)S_{ub}^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Note that if $aa = 1$, the two pivots B and C are identical. An important reason that the degree of freedom should not be changed, but it is still equal to $n-1$, is that in the normal distribution it only losses one degree for estimating the mean inside the variance estimator. This is a crucial rule in statistical theory. Using (7), the proposed $(1-\alpha)100\%$ confidence interval for σ using the pivot C is of the form

$$CI_{adj_{aa}} = \left(\sqrt{\frac{(n-aa)}{\chi_{1-\alpha/2,n-1}^2}} S_{ub}, \sqrt{\frac{(n-aa)}{\chi_{\alpha/2,n-1}^2}} S_{ub} \right). \tag{8}$$

Note that although this confidence interval has an obvious closed-form solution, the Monte-Carlo simulation method is needed to verify the optimum value aa under the two conditions. The details to do so will be given in Section 5.

3.2. Generalized pivot method

The generalized confidence interval (GCI) is a method for constructing the confidence interval for parameter. This method uses the generalized pivotal quantity, which is a function of random variables that the distribution does not depend on unknown parameter. We follow the method to find the generalized pivotal quantity from Weerahandi (1993) for obtaining the confidence interval for σ. The general details are given as follows. Let $X = (X_1, X_2, \cdots, X_n)$ be a random sample of size n with the probability density function $f(x; \mu, \sigma^2)$, and $x = (x_1, x_2, \cdots, x_n)$ be the observed values of X. Following Weerahandi, the generalized pivotal quantity $G(X; x, \mu, \sigma^2)$ requires the two conditions.

1) If x is fixed, $G(X; x, \mu, \sigma^2)$ has a distribution free from unknown parameters.
2) The observed value of $G(X; x, \mu, \sigma^2)$, denoted as $g(x; x, \mu, \sigma^2)$, independents from nuisance parameter.

Then the $(1-\alpha)100\%$ confidence interval for the parameter of interest is given by the percentiles $(\alpha/2)100$th and $(1-\alpha/2)100$th of $G(X; x, \mu, \sigma^2)$.

Now, we are interested in estimating σ of the normal distribution. Let $X = (X_1, X_2, \cdots, X_n)$ be a normal random sample with the probability density function given in (1). Hence, the log-likelihood function of σ^2 on the elimination of nuisance parameter μ (as used in the profile likelihood method) is

$$\log L(\sigma^2) = -\frac{1}{2} \sum_{i=1}^{n} (x_i - \bar{x})^2 - \frac{n}{2} \log(2\pi\sigma^2),$$

and the expected Fisher's information is given by $I(\sigma^2) = n/(2\sigma^4)$. According to the statistical inference, the score statistic is built as

$$U = \frac{\partial \log L(\sigma^2)}{\partial \sigma^2} \sim N(0, I(\sigma^2)).$$

This follows that

$$W = \frac{U}{\sqrt{I(\sigma^2)}} = \frac{1}{\sqrt{2n}} \left(\frac{(n-1)S^2}{\sigma^2} - n \right) \sim N(0,1), \tag{9}$$

approximately. Moreover, W is the pivotal quantity, as its distribution does not depend on parameter. This function is used to construct the new generalized pivotal quantity. From (9), it can be re-written into the form:

$$R_{\sigma^2} = \left[\left(W + \frac{n}{\sqrt{2n}} \right) \frac{\sqrt{2n}}{(n-1)S^2} \right]^{-1}.$$

This function satisfies the two conditions of Weerahandi (1993), i.e., 1) R_{σ^2} is free from the parameter and 2) when the observed values is given, $r_{\sigma^2} = \sigma$ or r_{σ^2} is not depend on nuisance parameter. R_{σ^2} is therefore the generalized pivotal quantity for σ^2. The $(1-\alpha)100\%$ confidence interval for σ^2 is then given as $(R_{\sigma^2}(\alpha/2), R_{\sigma^2}(1-\alpha/2))$. Taking the square root of those limits, the proposed confidence interval for σ is obtained by

$$CI_{gci.S} = \left(\sqrt{R_{\sigma^2}(\alpha/2)}, \sqrt{R_{\sigma^2}(1-\alpha/2)} \right). \tag{10}$$

In computation, the iterative method is required, as there is no closed-form solution for $CI_{gci.S}$. We apply the following process.

Step 1: compute $\bar{X}$ and S from the data

Step 2: sample W from a standard normal distribution and compute R_{σ^2}

Step 3: repeat Step 2) for B times, say $B = 1,000$, and rank B values of R_{σ^2} from small to large

Step 4: compute the $(1-\alpha)100\%$ percentile confidence interval for σ.

In addition, S_{ub} is used in the estimation to compare with (10). The GCI based on this unbiased estimator is denoted as $CI_{gci.S_{ub}}$, and applies the above algorithm in calculation.

4. Confidence interval for coefficient of variation

An important statistical tool to measure the spread of data in terms of relative variability is the CV. Dimensionless unit is a benefit of this estimator. So, CV is used in comparing the variability of several data sets with different units or means. There have been researches done confidence intervals for the population CV or τ. Lists of papers related to this work are given in the following.

Panichkitkosolkul (2009) introduced the confidence interval for τ in the normal distribution. It was developed from Vangel (1996). A $(1-\alpha)100\%$ confidence interval is given by

$$CI_{wp} = \left(\hat{\tau}_{ml} \left[\left(\frac{w_1}{n} - 1 \right) \hat{\tau}^2_{ml} + \frac{w_1}{n-1} \right]^{-\frac{1}{2}}, \hat{\tau}_{ml} \left[\left(\frac{w_2}{n} - 1 \right) \hat{\tau}^2_{ml} + \frac{w_2}{n-1} \right]^{-\frac{1}{2}} \right), \tag{11}$$

where $\hat{\tau}_{wp} = S_{ml}/\bar{X}$, $w_1 = \chi^2_{1-\alpha/2,n-1} + w$, $w_2 = \chi^2_{\alpha/2,n-1} + w$, and w is denoted as a suitable constant, which is suggested as $w = 2$. The performance of CI_{wp} was conducted using simulations and compared with the confidence interval proposed by Vangel, where the latter used $\hat{\tau}$ instead of $\hat{\tau}_{wp}$. It was found that CI_{wp} had the coverage probability greater than or close to the nominal level when $\tau = 0.1$, 0.2, and 0.3, as well as Vangel's confidence interval. However, the expected length of CI_{wp} was slightly shorter.

Method of variance estimates recovery (MOVER) is an approach used to construct the confidence interval for a function of parameters given in probability models. It was developed by Zou, Donner, and their colleges (Donner and Zou 2010; Zou and Donner 2013; Zou, Huo, and Taleban 2009), and also applied in many works (Li, Tang, and Wong 2014; Sangnawakij and Niwitpong 2020; Yang, Tian, and Liu 2021). Suppose that $\hat{\theta}_1$ and $\hat{\theta}_2$ are two independently estimators for θ_1 and θ_2, respectively. The MOVER interval can be used to estimate $\theta_1+\theta_2$, θ_1/θ_2, or $\theta_1-\theta_2$, under the normal approximation method. It gives an explicit closed-form solution, simple to use. The MOVER algorithm as originally needs the separate confidence intervals for each of two parameters. Then, we combine them into a single confidence interval for the parameter of interest using recovering variances. Donner and Zou (2010) introduced the confidence interval for τ in the normal distribution. The single confidence interval for σ used in their work is given in (5) and for μ is $\bar{X} \pm z_{\alpha/2}S/\sqrt{n}$. Therefore, a $(1-\alpha)100\%$ confidence interval for τ using the MOVER is given by

$$CI_{dz} = \left(\left(\bar{X} - \sqrt{\bar{X}^2 + d_1 d_3 (d_1 - 2)}\right)\frac{S}{d_3}, \left(\bar{X} + \sqrt{\bar{X}^2 + d_2 d_3 (d_2 - 2)}\right)\frac{S}{d_3}\right), \tag{12}$$

where $d_1 = \sqrt{(n-1)/\chi^2_{1-\alpha/2,n-1}}$, $d_2 = \sqrt{(n-1)/\chi^2_{\alpha/2,n-1}}$, and $d_3 = \bar{X}^2 - z^2_{\alpha/2}S^2/n$. In simulation, CI_{dz} provided the coverage probability close to the target level, and was better than Vangel's confidence interval. However, it seemed the expected length of CI_{dz} was greater.

According to the MOVER method, we have an idea to construct the new confidence interval for τ. Since S and $\bar{X}$ are independent by Basu's theorem (Basu 1955), S_{ub} which is the function of S is independent from $\bar{X}$ as well. This is under the condition of the MOVER approach. We substitute S_{ub} into S and rewrite the formula into a general form. The new proposed confidence interval for τ is then given by

$$CI_{mov} = (L_{mov}, U_{mov}), \tag{13}$$

where

$$L_{mov} = \frac{S_{ub}\bar{X} - \sqrt{(S_{ub}\bar{X})^2 - l_1 u_2 (2S_{ub} - l_1)(2\bar{X} - u_2)}}{u_2(2\bar{X} - u_2)}$$

and

$$U_{mov} = \frac{S_{ub}\bar{X} + \sqrt{(S_{ub}\bar{X})^2 - u_1 l_2 (2S_{ub} - u_1)(2\bar{X} - l_2)}}{l_2(2\bar{X} - l_1)}.$$

l_1 and u_1 are the lower and upper limits for σ, here we refer to $CI_{adj_{aa}}$ given in (8), depending on the corrected value aa. Moreover, l_2 and u_2 are the lower and upper confidence limits for μ, given as $\bar{X} \pm z_{\alpha/2}S_{ub}/\sqrt{n}$. Writing the confidence interval in the general form as noted in (13) is simple and convenient to use when we have a new confidence interval for the single parameter. Another confidence interval is constructed based on the MOVER, but we use the GCI for σ presented in subsection 3.2. A $(1-\alpha)100\%$ confidence interval for τ is given by

$$CI_{mov-gci.S} = (L_{mov-gci.S}, U_{mov-gci.S}), \tag{14}$$

where $L_{mov-gci.S}$ and $U_{mov-gc.Si}$ are similar to L_{mov} and U_{mov} given in (13), but we replace $\left(\sqrt{R_{\sigma^2}(\alpha/2)}, \sqrt{R_{\sigma^2}(1-\alpha/2)}\right)$ in (l_1, u_1). Again, S_{ub} is applied and the confidence interval is denoted as $CI_{mov-gci.S_{ub}}$.

The performance of the confidence intervals presented in Sections 3 and 4 will be investigated via simulations. We provide the scope of study and simulation results in the next section.

5. Simulation study for confidence interval

The performance of the confidence intervals for σ was conducted in this section. Here, the proposed intervals, namely CI_{adj}, $CI_{adj_{aa}}$, $CI_{gci.S}$, and $CI_{gci.S_{ub}}$, were compared with the traditional estimator CI_{trad}. Simulation settings as given in subsection 2.2 were applied. Each scenario was repeated 10,000 times. On average, we computed the coverage probability and expected length of the confidence interval from

$$CP = \frac{n(L(X) \le \sigma \le U(X))}{10,000} \quad \text{and} \quad EL = \frac{\sum_{i=1}^{10,000}(U(X)_i - L(X)_i)}{10,000},$$

respectively, where $n(L(X) \leq \sigma \leq U(X))$ is the number that σ lies within the lower and upper limits of a confidence interval. In decision, we prefer a confidence interval that has the coverage probability greater than or close to the nominal probability level at 0.95 and has the short interval length.

Table 1: The performance of 95% confidence intervals for σ in the normal distribution using the corrected values aa = 2, 3, and 4

		Coverage probability			Expected length		
σ	n	$CP_{adj.aa=2}$	$CP_{adj.aa=3}$	$CP_{adj.aa=4}$	$EL_{adj.aa=2}$	$EL_{adj.aa=3}$	$EL_{adj.aa=4}$
1	10	0.9541	0.9470	0.9214	1.0738	1.0014	0.9308
	20	0.9498	0.9488	0.9382	0.6797	0.6617	0.6430
	30	0.9527	0.9446	0.9432	0.5377	0.5277	0.5185
	50	0.9492	0.9436	0.9430	0.4076	0.4265	0.4289
	100	0.9530	0.9491	0.9449	0.2813	0.2806	0.2792
3	10	0.9530	0.9469	0.9159	3.2054	3.0122	2.7832
	20	0.9519	0.9441	0.9362	2.0434	1.9827	1.9287
	30	0.9520	0.9499	0.9436	1.6134	1.5891	1.5547
	50	0.9542	0.9494	0.9424	1.2204	1.1890	1.1698
	100	0.9521	0.9492	0.9494	0.8458	0.8412	0.8376
9	10	0.9500	0.9426	0.9183	9.6706	9.0431	8.3559
	20	0.9493	0.9461	0.9388	6.1236	5.9358	5.7904
	30	0.9485	0.9477	0.9403	4.8466	4.7516	4.6695
	50	0.9500	0.9487	0.9405	3.6610	3.5581	3.4310
	100	0.9508	0.9457	0.9476	2.5409	2.5291	2.5148
15	10	0.9552	0.9444	0.9182	16.0848	15.0336	13.9368
	20	0.9495	0.9469	0.9387	10.1992	9.9226	9.6302
	30	0.9485	0.9529	0.9446	8.0748	7.9365	7.7886
	50	0.9511	0.9526	0.9465	6.0963	6.3454	5.9533
	100	0.9491	0.9472	0.9485	4.2350	4.2096	4.1902

The corrected values aa = 2, 3, and 4 were firstly considered, in order to use in $CI_{adj_{aa}}$. As can be seen in Table 1, coverage probabilities of $CI_{adj_{aa=2}}$ were satisfied the target level at 0.95 in general cases. The average lengths of $CI_{adj_{aa=4}}$ and $CI_{adj_{aa=3}}$ were slightly smaller than those of $CI_{adj_{aa=2}}$, respectively, but in almost cases these two estimators had coverage probabilities much lower than 0.95. We also observed that the coverage probability tended to more decrease if $aa > 2$. In conclusion, $aa = 2$ is the *optimal value* for our method. The efficacy of $CI_{adj_{aa=2}}$ was then compared with CI_{trad}, CI_{adj}, $CI_{gci.S}$, and $CI_{gci.S_{ub}}$. The results are shown in Figure 3. It can be concluded that CI_{adj} had the coverage probability lower than 0.95 in all cases in the study. Meanwhile, CI_{trad}, $CI_{adj_{aa=2}}$, $CI_{gci.S}$, and $CI_{gci.S_{ub}}$ had coverage probabilities greater than or close to 0.95. The proposed interval estimators had coverage probabilities greater than the original method CI_{trad}. Additionally, $CI_{adj_{aa=2}}$ provided the shortest expected lengths in all cases. $CI_{gci.S}$ and $CI_{gci.S_{ub}}$ performed well in terms of coverage probability, but for large variation they had acceptable expected length when $n > 20$. These results strongly confirm that $CI_{adj_{aa=2}}$, which uses the unbiased estimator and suitable corrected value, is a superior method to estimate the true SD.

Next, the performance of proposed confidence intervals for τ, $CI_{mov_{aa=2}}$, $CI_{mov-gci.S}$, and $CI_{mov-gci.S_{ub}}$, were compared with the existing intervals, CI_{wp} and CI_{dz}, and given in Figure 4. $CI_{mov_{aa=2}}$ and CI_{dz} provided the coverage probability close to 0.95 in all cases in simulations. $CI_{mov-gci.S}$ and $CI_{mov-gci.S_{ub}}$ had the coverage probabilities greater than 0.95, while their expected lengths were slightly larger than $CI_{mov_{aa=2}}$ and CI_{dz}. It can be seen that for small sample sizes with high distribution or $\tau \geq 50\%$, CI_{wp} has a limitation as it was undefined. This is not mentioned and studied in the original paper (Panichkitkosolkul 2009). We therefore conclude that our confidence intervals could be used to estimate the population CV in all situations.

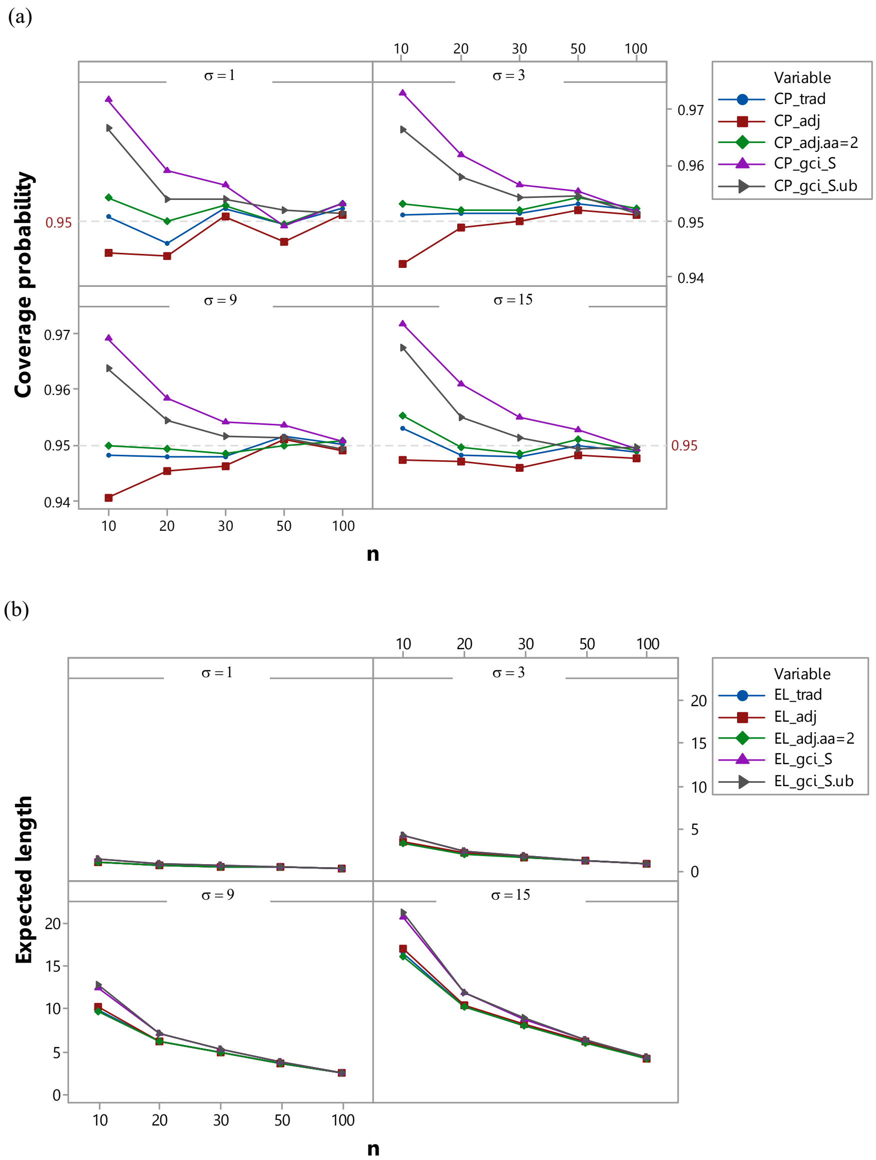
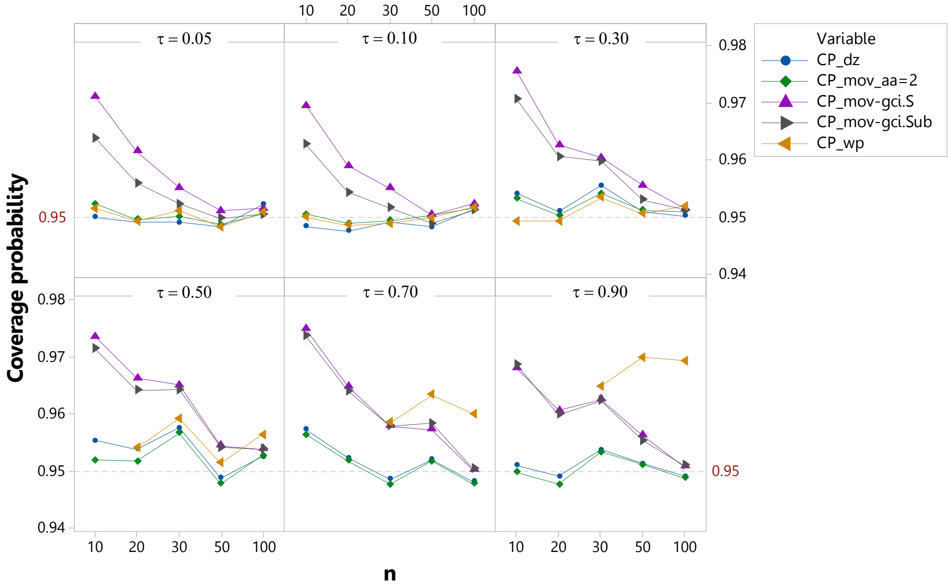


Figure 3: Coverage probability (a) and expected length (b) of the 95% confidence intervals for σ using simulations

(a)



(b)

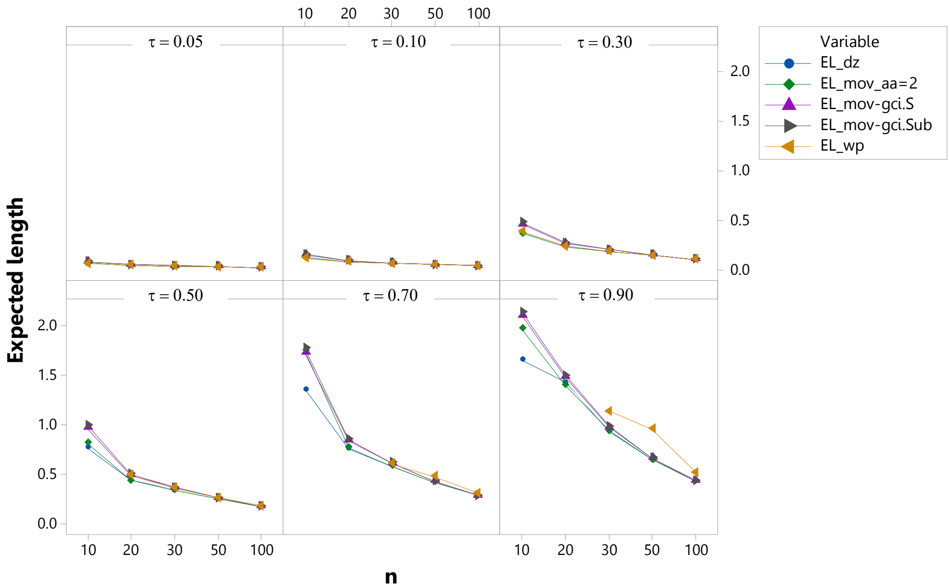
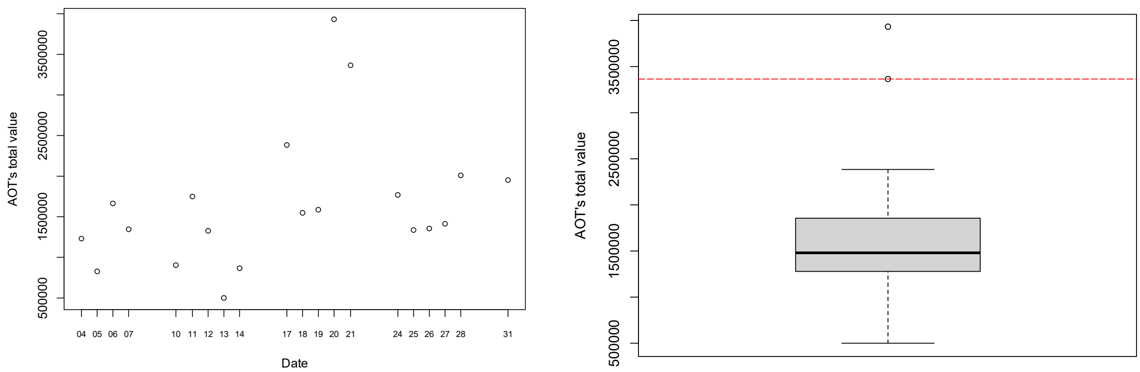


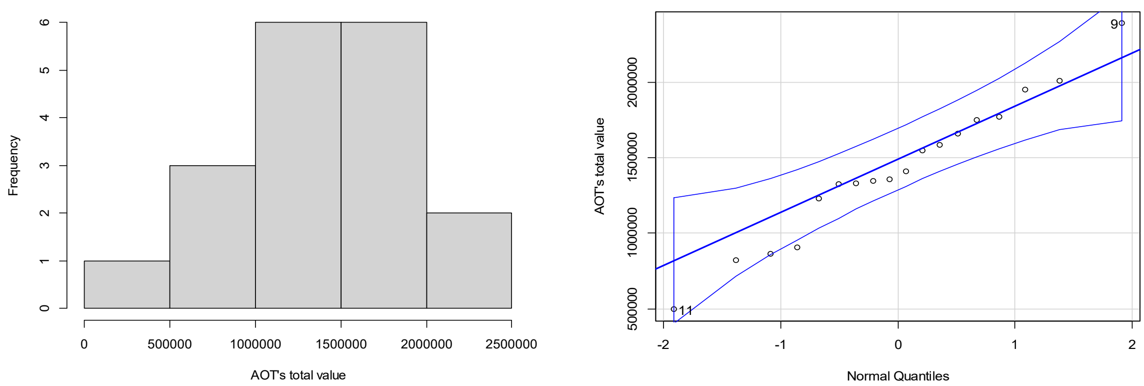
Figure 4: Coverage probability (a) and expected length (b) of the 95% confidence intervals for τ using simulations

6. Data application

This paper uses the real data on a major stock market index in Thailand. SET50 Index is the stock prices of the top 50 listed companies on Stock Exchange of Thailand (SET). It is a capitalization-weighted price index that compares the current market value of all listed common stocks with their market values on the base data (Fixed Income and Other Product Department 2013). Due to the Covid-19 Omicron situation in Thailand and the global financial volatility, Thai economic growth slowed in early 2022. This included transportation and logistics sectors (Banchongduang 2020). We select the SET50 total value (Million Baht: MB) from Airports of Thailand Public Company Limited (AOT) in January 2022, a total of 20 observations, in this application. The data obtained from the SET's website (https://www.set.or.th/th/market/index/set/overview) are shown in Figure 5 (a). We also note here that AOT is top 10 of SET50 on NVDR. From the box-plot, it indicates that the indexes in two days of the month are outliers, where they are extremely greater than the most other data in group. In this study, the outliers were excluded from the analysis. The data used in computation are given by histogram, see Figure 5 (b). Furthermore, the probability plot shows that the AOT total values follow the normal distribution (Shapiro-Wilk normality test statistic = 0.98 and p-value = 0.97). In summary, the AOT company had total value in January 2022 on average of 1,430,617 MB (median = 1,383,411 MB).



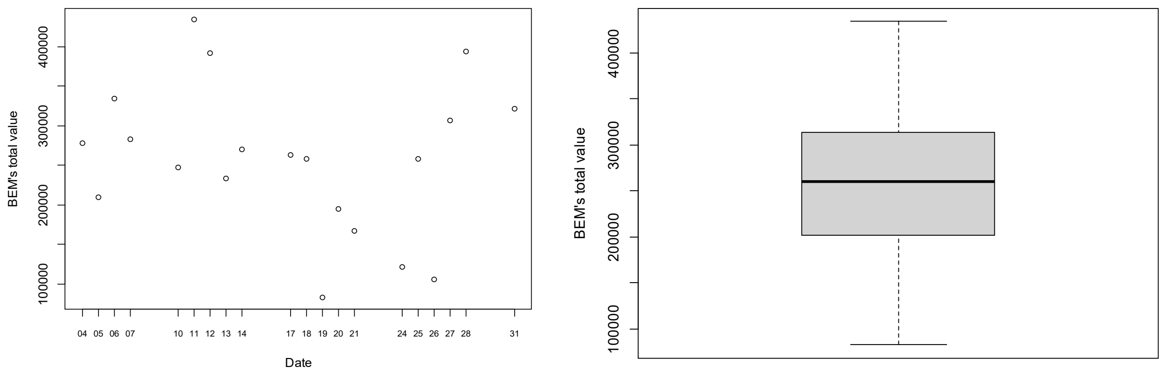
(a) Total value of AOT Index in January 2022 (20 observations) and boxplot



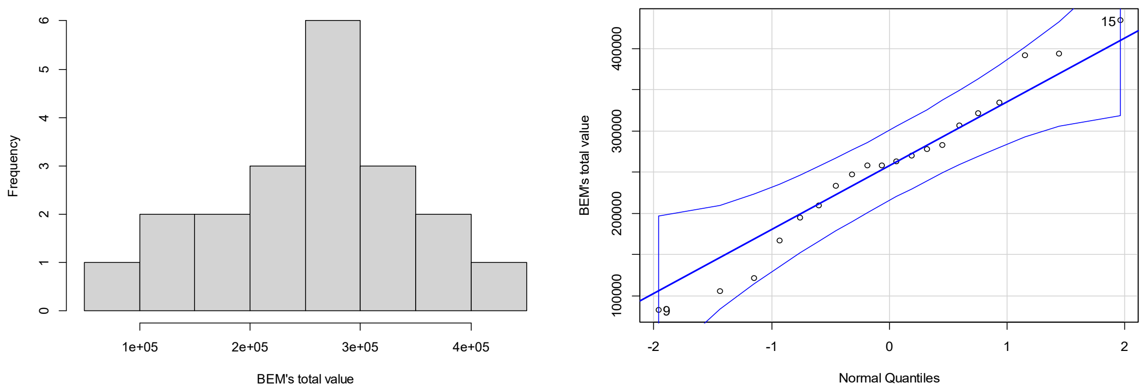
(b) Histogram and QQ-plot of total value of AOT Index (18 observations), excluding outliers

Figure 5: Plots of the SET50 total value in January 2022 (in million Thai Baht) of AOT company

Another example used in this study was the total values of Bangkok Expressway and Metro Public Company Limited (BEM). This company is a public transportation corporation in Thailand. It operates the expressway and metro systems in Bangkok city. The BEM index is also included in SET50 quotation and one of the indexes that is investment grade. In this section, we use the total value of BEM index in January 2022. The original data obtained from The SET's website are shown in Figure 6 (a). The data from 20 observations are checked for normality and shown by the QQ plot in Figure 6 (b). Clearly, the data has a normal distribution (Shapiro-Wilk normality statistic = 0.9734 with a p-value = 0.8242), where there has no outlier. Furthermore, it was found that the mean of total value was 257,374 MB (median = 260,394 MB).



(a) Total value of BEM Index in January 2022 (20 observations) and boxplot



(b) Histogram and QQ-plot of total value of BEM Index (20 observations)

Figure 6: Plots of the SET50 total value in January 2022 (in million Thai Baht) of BEM company

The estimated SDs and CVs of the two datasets obtained from various methods considered in this paper were computed and given in Table 2. The 95% confidence intervals for SD of the total value for the AOT company were large, depending on the unit. Those for CV were between 23% and 56%. From the latter measurement, it can be indicated that the stock index value of AOT company had fairly high volatility in this period, because the CV estimate was greater than 30% (Couto, Peternelli, and Barbosa 2013). For the BEM company, the 95% confidence intervals for SD of total value were smaller than the SD of AOT's total value. However, in fact, comparison of the dispersion expressed in the different means as in these two datasets must be measured by the CV as remarked in Section 1. Approximately, the BEM's total values had the 95% confidence intervals for CV between 25% and 62%. Since there are many methods that provide the estimated values for SD and CV, the question arises whether which one should be used to report the findings. We refer to the simulation results given in Section 5. The confidence intervals from the adjusted method using $aa = 2$ outperform the comparison methods. Hence, the confidence intervals for SD of total value were approximately 346,590 to 692,430 MB for AOT company and 70,630 to 135,650 MB for BEM. The confidence intervals for CV of total value for AOT and BEM were 23.53% to 50.26%, and 26.52% to 55.08%, respectively. We note that in the period of study the total value of both companies had high distribution and the BEM total value provided the volatility index slightly greater than AOT. The lengths of interval given in this example also match the simulation results. A relevant `R` code useful in applications is presented in the Supplemental Material.

Table 2: Estimated standard deviation (SD), coefficient of variation (CV), and the 95% confidence interval of volatility total value from the AOT and BEM data examples

Method	SD ($\times$1,000 MB)	Interval length	Method	CV (%)	Interval length
AOT:					
S	469.15	-	$\hat{\tau}$	32.79	-
S_{ub}	476.10	-	$\hat{\tau}_{ub}$	33.27	-
CI_{trad}	(352.05, 703.33)	351.28	$\hat{\tau}_{wp}$	31.86	-
CI_{adj}	(357.26, 713.74)	356.48	CI_{dz}	(23.84, 50.88)	27.04
$CI_{adj_{aa=2}}$	(346.59, 692.43)	345.84	$CI_{mov_{aa=2}}$	(23.53, 50.26)	26.73
$CI_{gci.S}$	(354.44, 780.13)	425.68	$CI_{mov-gci.S}$	(23.92, 55.59)	31.67
$CI_{gci.S_{ub}}$	(359.69, 791.68)	431.99	$CI_{mov-gci.S_{ub}}$	(24.27, 56.42)	32.15
			CI_{wp}	(23.39, 50.55)	27.16
BEM:					
S	94.18	-	$\hat{\tau}$	36.59	-
S_{ub}	95.42	-	$\hat{\tau}_{ub}$	37.07	-
CI_{trad}	(71.62, 137.55)	65.93	$\hat{\tau}_{wp}$	35.66	-
CI_{adj}	(72.57, 139.37)	66.80	CI_{dz}	(26.82, 55.66)	28.84
$CI_{adj_{aa=2}}$	(70.63, 135.65)	65.02	$CI_{mov_{aa=2}}$	(26.52, 55.08)	28.56
$CI_{gci.S}$	(72.28, 154.62)	82.34	$CI_{mov-gci.S}$	(27.02, 62.16)	35.14
$CI_{gci.S_{ub}}$	(73.24, 156.67)	83.43	$CI_{mov-gci.S_{ub}}$	(27.38, 62.98)	35.60
			CI_{wp}	(26.41, 55.63)	29.22

7. Concluding remarks and discussion

The adjusted confidence interval based on the new pivotal quantity for the SD is introduced in this work. We show in the simulation section that our proposed method ($CI_{adj_{aa=2}}$) provides a good performance in terms of coverage probability and expected length. In advantage, $CI_{adj_{aa=2}}$ has the closed-form solution and provides lower expected length than the classical confidence interval for the SD, where the latter is usually used in elementary statistic learning and applications. Reasons that $CI_{adj_{aa=2}}$ is satisfied are highlighted in the following.

i) It is constructed using S_{ub}. This is the unbiased estimator for σ, one of the most important properties of a point estimator.

ii) The pivot $(n-2)S_{ub}^2/\sigma^2$ is suitable and follows the chi-square distribution with $n-1$ degrees of freedom. As can be seen in Figure 7, the empirical cumulative distribution function (CDF) plots from this pivot are very close to the theoretical distribution.

Furthermore, we propose the confidence interval for the CV based the MOVER approach or $CI_{mov_{aa=2}}$. This also uses S_{ub}. $CI_{mov_{aa=2}}$ has a good performance in terms of coverage probability and provides the acceptable short length of interval. Moreover, it can be used in wider situations of normally distributed data, small, medium, or high dispersion, and sample sizes. We therefore recommend $CI_{adj_{aa=2}}$ and $CI_{mov_{aa=2}}$ in estimating the true SD and CV, respectively, for all cases in the normal distribution. The proposed generalized confidence intervals (CI_{gci} and $CI_{mov-gci}$) required computational algorithms can be used as alternative methods to estimate the SD and CV, respectively, if sample size is greater than 30. Since the proposed generalized pivot is useful and can be applied in interval estimation in, it would be used to construct the confidence interval for two or more populations in the future work.

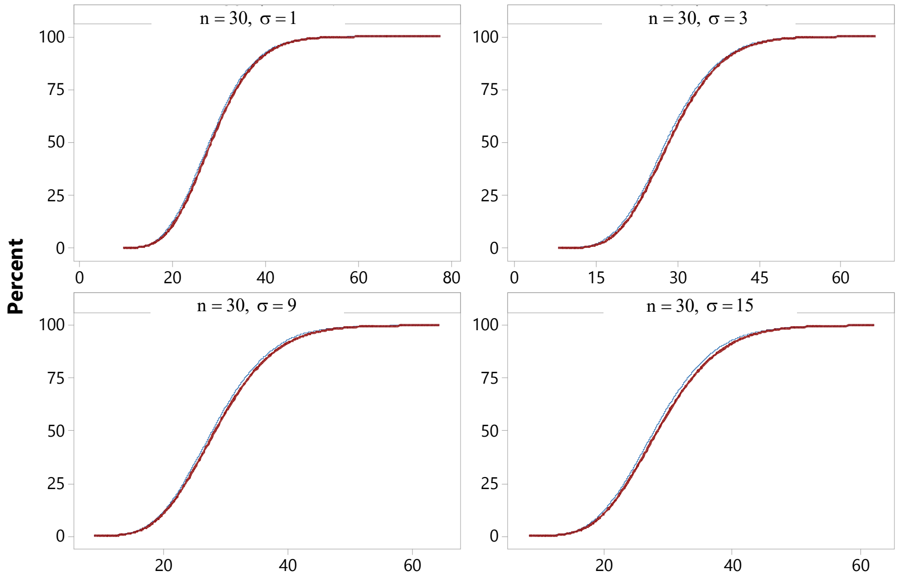


Figure 7: CDF (blue line) of chi-square distribution with df $= n-1$ and empirical CDF plots for $(n-2)S_{ub}^2/\sigma^2$ from simulations with $n = 30$ and various values of σ

Supplemental Materials

The R-code used in data application is available with this paper. It is given in the separated file.

Acknowledgments

The authors would like to thank the Editor and reviewers for all suggestions which lead to considerable improvements of the paper. The first author gratefully acknowledges research funder. This study was supported by Thammasat University Research Fund, Contact No, TUFT 9/2565.

References

Banchongduang S (2020). *BoT in Inflation Warning*. Bangkok Post. URL https://www.bangkokpost.com/business/2271611.

Banik S, Albatineh AN, Shawiesh M, Kibria BM (2014). “Estimating the Population Standard Deviation with Confidence Interval: A Simulation Study under Skewed and Symmetric Conditions.” *International Journal of Statistics in Medical Research*, **3**(4), 356–367. doi:https://doi.org/10.6000/1929-6029.2014.03.04.4.

Baron M (2014). *Probability and Statistics for Computer Scientists*. CRC Press.

Basu D (1955). “On Statistics Independent of a Complete Sufficient Statistic.” *Sankhya A*, **15**(4), 377–380. URL http://www.jstor.org/stable/25048259.

Bonett D (2006). “Approximate Confidence Interval for Standard Deviation of Nonnormal Distributions.” *Computational Statistics and Data Analysis*, **50**(3), 775–782. doi:https://doi.org/10.1016/j.csda.2004.10.003.

Casella G, Berger R (2002). *Statistical Inference*. Duxbury Press.

Choi YA, Son J (1990). "Comparisons on Approximating Methods." *Yongnam Statistical Letters*, **1**(5-7), 59–65.

Clarren SK, Chudley AE, Wong L, Friesen J, Brant R (2010). "Normal Distribution of Palpebral Fissure Lengths in Canadian School Age Children." *Canadian Journal of Clinical Pharmacology*, **17**(1), 67–78. URL https://www.jptcp.com/index.php/jptcp/article/view/536.

Couto MF, Peternelli LA, Barbosa MH (2013). "Classification of the Coefficients of Variation for Sugarcane Crops." *Ciência Rural*, **43**(6), 957–961. doi:https://doi.org/10.1590/S0103-84782013000600003.

Demir E, Kizys R, Rouatbi W, Zaremba A (2021). "COVID-19 Vaccinations and the Volatility of Energy Companies in International Markets." *Journal of Risk and Financial Management*, **14**(12), 1–21. doi:https://doi.org/10.3390/jrfm14120611.

Donner A, Zou G (2010). "Interval Estimation for a Difference between Intraclass kappa Statistics." *Statistical Methods in Medical Research*, **21**(1), 347–359. doi:https://doi.org/10.1111/j.0006-341x.2002.00209.x.

Fixed Income and Other Product Department (2013). *SET Index Methodology*. The Stock Exchange of Thailand. URL https://www.set.or.th/en/products/index/files/2013-01-SET50-100-IndexRule-EN.pdf.

Javari M (2017). "Spatial Monitoring and Variability of Daily Rainfall in Iran." *International Journal of Applied Environmental Sciences*, **12**(5), 801–824. URL https://www.ripublication.com/ijaes17/ijaesv12n5_07.pdf.

Li HQ, Tang ML, Wong WK (2014). "Confidence Intervals for Ratio of Two Poisson Rates Using the Method of Variance Estimates Recovery." *Computational Statistics*, **25**, 869–889. doi:https://doi.org/10.1007/s00180-013-0467-9.

Mahmoudvand R, Hassani H (2009). "Two New Confidence Intervals for the Coefficient of Variation in a Normal Distribution." *Journal of Applied Statistics*, **36**(4), 429–442. doi:https://doi.org/10.1080/02664760802474249.

Panichkitkosolkul W (2009). "Improved Confidence Intervals for a Coefficient of Variation of a Normal Distribution." *Thailand Statistician*, **7**(2), 193–199. URL https://ph02.tci-thaijo.org/index.php/thaistat/article/view/34315.

Pearson K (1896). "Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia." *Philosophical Transactions of the Royal Society A*, **187**, 253–318. doi:https://doi.org/10.1098/rsta.1896.0007.

Protopapas A, Parrila R (2018). "Is Dyslexia a Brain Disorder?" *Brain Sciences*, **8**(4), 1–18. doi:https://doi.org/10.3390/brainsci8040061.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rossi G (2010). "Maximum Likelihood Estimation of the Cox-Ingersoll-Ross Model Using Particle Filter." *Computational Economics*, **36**, 1–16. doi:https://doi.org/10.1007/s10614-010-9208-0.

Sangnawakij P, Niwitpong S (2020). "Interval Estimation for the Common Coefficient of Variation of Gamma Distributions." *Thailand Statistician*, **18**(3), 346–353. URL https://ph02.tci-thaijo.org/index.php/thaistat/article/view/241285.

Sreejith S, Pathy G (2012). "Study on Impact of Foreign Exchange Value of Indian Rupee on Relationship between International Crude Oil Prices and Petrol Prices in Indian Market." *ZENITH International Journal of Business Economics and Management Research*, **2**(4), 227–231. URL http://www.zenithresearch.org.in/images/stories/pdf/2012/NOV/ZIJBEMR/21_ZIJBMER_NOV2012_VOL2_ISSUE11.pdf.

Turner AJ, Robinson S, Mavris D (2013). "Heuristics for the Regression of Stochastic Simulations." *Journal of Simulation*, **7**(4), 229–239. doi:https://doi.org/10.1057/jos.2013.1.

Vangel MG (1996). "Confidence Intervals for a Normal Coefficient of Variation." *The American Statistician*, **15**(1), 21–26. doi:https://doi.org/10.1080/00031305.1996.10473537.

Vera JE (2022). "The Persistence of Financial Volatility after COVID-19." *Finance Research Letters*, **44**, 1–9. doi:https://doi.org/10.1016/j.frl.2021.102056.

Weerahandi S (1993). "Generalized Confidence Intervals." *Journal of the American Statistical Association*, **88**(423), 899–905. doi:https://doi.org/10.1080/01621459.1993.10476355.

Yang Z, Tian GL, Liu X (2021). "Simultaneous Confidence Interval Construction for Many-to-One Comparisons of Proportion Differences Based on Correlated Paired Data." *Journal of Applied Statistics*, **48**(8), 1442–1456. doi:https://doi.org/10.1080/02664763.2020.1795815.

Zhang N, Wang A, Haq N, Nosheen S (2022). "The Impact of COVID-19 Shocks on the Volatility of Stock Markets in Technologically Advanced Countries." *Economic Research-Ekonomska Istrazivanja*, **35**(1), 2191–2216. doi:https://doi.org/10.1080/1331677X.2021.1936112.

Zhaohong W, Mengqiang X, Yan Y, Xin L (2009). "The Forecasting Algorithm for Cloud Model of Similar Normal Distributed Data." *Computer Applications and Software*, **9**, 78–79. URL http://en.cnki.com.cn/Article_en/CJFDTotal-JYRJ200909023.htm.

Zou GY, Donner A (2013). "Construction of Confidence Limits about Effect Measures: A General Approach." *Statistics in Medicine*, **27**(10), 1693–1702. doi:https://doi.org/10.1002/sim.3095.

Zou GY, Huo CY, Taleban J (2009). "Simple Confidence Intervals for Lognormal Means and Their Differences with Environmental Applications." *Environmetrics*, **20**(2), 172–180. doi:https://doi.org/10.1002/env.919.

Affiliation:

Patarawan Sangnawakij
Department of Mathematics and Statistics
Faculty of Sciences and Technology, Thammasat University
Pathum Thani 12120, Thailand
E-mail: patarawan.s@gmail.com
URL: https://sci.tu.ac.th/en/

Wanwarat Anlamlert
Department of Mathematics and Statistics
Faculty of Sciences and Technology, Thammasat University
Pathum Thani 12120, Thailand
E-mail: wanwarat@mathstat.sci.tu.ac.th
URL: https://sci.tu.ac.th/en/

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 52 *Submitted:* 2022-04-25
2023 *Accepted:* 2022-10-06