International Journal of Mathematics, Statistics and Operations Research
Vol. 6; Number 1; 2026, pp. 173-199
https://doi.org/10.47509/IJMSOR.2026.v06i01.10

ARF INDIA
Academic Open Access Publishing
www.arfjournals.com

Size and Power Analysis of Classical and Modified Coefficient of Variation Tests: Insights from Monte Carlo Simulations

B.M. Golam Kibria[1] and Shipra Banik[2]
[1]Department of Mathematics and Statistics, Florida International University, University Park, Miami FL 33199, USA. E-mail: kibriag@fiu.edu
[2]Department of Physical Sciences, Independent University, Bangladesh, Dhaka 1212, Bangladesh.
Email: banik@iub.edu.bd

Abstract

This study presents a comprehensive evaluation of the size and power properties of several coefficient of variation (CV) tests under varying distributional assumptions and sample sizes. Using extensive Monte Carlo simulations, we assess the performance of both classical and modified CV tests across normal, gamma, chi-square, and log-normal distributions. The findings reveal that while many tests perform adequately under normality with large sample sizes, only a few maintain robustness under non-normal conditions and small samples. Notably, the median-modified versions of the Sharma and Krishna, Curto and Pinto, and ADJ tests consistently demonstrate strong power and effective size control across diverse settings. The parametric bootstrap test also performs well, particularly for skewed distributions such as the log-normal. In contrast, the classical *t*-test remains overly conservative, and the standard versions of the Miller, Vangel, and Panichkitkosolkul tests exhibit weak power. These results highlight the importance of using modified CV tests tailored to distributional characteristics and sample size, thereby improving the reliability of statistical inference in applied research.

Keywords: Coefficient of variation; Gamma distribution; Hypothesis testing; Log-normal distribution; Nominal size; Simulation study; Statistical power; Skewness

1. Introduction

The coefficient of variation (CV) is a widely used statistical measure that quantifies relative variability by expressing the standard deviation as a proportion of the mean. Unlike absolute measures such as variance or standard deviation, the CV is unit less, allowing comparisons across datasets with different scales or units. Mathematically, it is defined as:

$$CV = (\text{Standard Deviation} / \text{Mean}) \times 100$$

Due to its scale-invariant nature, the CV is especially useful in fields like finance, biology, healthcare, engineering and others. In finance, for example, it is used to assess the risk-return trade-off of investments; in biology, to compare variability across species; and in manufacturing, to evaluate process stability and quality control. Despite its versatility, the CV's statistical properties especially under skewed distributions are complex and not yet fully understood. Many real-world datasets, such as income levels, survival times, mechanical failure rates and others,

Article History
Received: 09 May 2025; Revised: 30 May 2026; Accepted: 10 June 2026; Published: 30 June 2026

To cite this paper
B.M. Golam Kibria & Shipra Banik (2026). Size and Power Analysis of Classifical and Modified Coefficient of Variation Tests: Insights from Onte Carlo Simulations. *International Journal of Mathematics, Statistics and Operations Research.* 6(1), 173-199.

exhibit skewness. In such cases, skewness can distort the CV's behavior, potentially undermining the validity of inferences based on it. This makes it essential to evaluate the statistical size and power of CV-based hypothesis tests under skewed conditions. It is well known size and power are key components of hypothesis testing. The size of a test refers to its type I error rate, while power reflects its ability to detect true effects. For the CV, assessing these properties is crucial to determine the reliability of tests when applied to skewed distributions. Inadequate size control or low power may lead to incorrect conclusions, missed effects and flawed decision-making. Therefore, understanding how these tests perform under varying conditions is critical to their practical application. It is known that simulation studies provide a powerful framework for this evaluation. By systematically varying factors such as sample size, skewness and distribution type, simulations allow researchers to examine test behavior under controlled yet realistic conditions. This is particularly valuable when theoretical results do not fully capture the complexities of real-world data. Simulation results can highlight strengths and limitations of different testing methods such as inflated type I error rates or reduced power thereby guiding more and robust methodological choices.

In recent years, several approaches for constructing confidence intervals for the CV have been proposed in the statistical literature (Amiri and Zwanzig (2010), Curto and Pinto (2009), Banik and Kibria (2011) and many others. Notably, Banik and Kibria (2011) conducted a simulation study evaluating the performance of various confidence interval techniques for the population coefficient of variation. However, while the CV is extensively applied across numerous fields, formal hypothesis testing procedures for the CV have received relatively limited attention among statisticians. Although a few test statistics have been introduced, there remains a scarcity of research assessing their performance in practical settings. Furthermore, many of the existing methods are developed under the assumption of normality, which often does not hold in real-world data scenarios especially when dealing with small samples or inherently skewed distributions. Indeed, skewed distributions are common in areas such as survival analysis, reliability engineering and economics (Baklizi and Kibria (2009), Shi and Kibria (2007), Banik and Kibria (2009), Almonte and Kibria (2009) and others highlighting the need for a more robust understanding of CV-based tests in skewed distribution contexts. The primary objective of this paper is to investigate the size and power properties of CV-based tests under skewed

distributions. Through an extensive simulation study, we examine the performance of these tests across a range of sample sizes and distributional shapes, including the gamma, log-normal and chi-square distributions and other distributions. While some previous work has addressed this topic, notably Miller (1991), Sharma and Krishna (1994), Vangel (1996), Panichkitkosolkul (2009), Banik et al. (2012), Gulhar et al. (2012), Abu-Shawiesh et al. (2019), comprehensive evaluations of CV test performance remain limited. To address this gap, this paper aims to provide empirically grounded recommendations for researchers and practitioners, enhancing the reliability of statistical inference when using the CV in skewed data contexts. The organization of this paper is as follows: Section 2 reviews existing methods and proposed some modified methods for testing the CV. Section 3 presents the simulation study and compares the empirical size and power of selected tests. Section 4 concludes with key findings and practical recommendations.

2. Testing procedures of a CV parameter

To describe the test statistics, suppose X_i, i = 1, 2, …, n represents a random sample of size n drawn from a normal population with mean μ and standard deviation σ. Below is a concise summary of the various methods recommended by different researchers for testing the population CV:

2.1 Classical t- test

Testing procedure for the CV parameter is first proposed by Hendricks and Robey (1936) and later is reviewed by Koopman et al (1964) and Iglewicz (1967). The null and alternative hypotheses are defined respectively below:

$$H_0: CV = CV_0 \tag{1}$$

and

$$H_1: CV > CV_0 \tag{2}$$

To test (1) vs. (2), the test statistic is defined as:

$$t_{cal} = \frac{\widehat{CV} - CV_0}{S_{CV}}$$

where $\widehat{CV}$ is the sample CV, CV_0 is the specified value of CV under H_0, $S_{cv} = \frac{\widehat{CV}}{\sqrt{2n}}$ is the standard error of the $\widehat{CV}$. Under H_0, t_{cal} has an approximate *t*-distribution with $n-1$ degrees of freedom (df). Then at α level of significance, we will reject H_0, when $t_{cal} > t_{(n-1),\alpha}$, where $t_{(n-1),\alpha}$ is the upper (1-α) quantile of the t-distribution with (n-1) df.

2.2 McKay's test

For testing (1) vs. (2), McKay (1932) proposed the following test statistics,

$$\chi^2_{Mck} = \frac{n(CV_0{}^2 + \widehat{CV}^4)}{\widehat{CV}^2(\widehat{CV}^2 + 1)},$$

where CV and $\widehat{CV}$ are defined as above. We will reject H_0, if $\chi^2_{Mck} > \chi^2_{(n-1),\alpha}$, where $\chi^2_{(n-1),\alpha}$ is the upper (1-α) quantile of a central chi-square distribution with (n-1) df.

2.3 Miller test

Following Miller (1991), we define the following test statistic for testing (1) vs. (2),

$$Z_{Mil} = \frac{\widehat{CV} - CV_0}{\sqrt{\frac{CV_0{}^4 + 0.5CV_0{}^2}{n-1}}},$$

where $\widehat{CV}$ and CV are defined as above. We will reject H_0 at α level of significance, when $Z_{Mil} > Z_\alpha$, where Z_α is the upper (1-α) quantile of a standard normal distribution.

2.4 Sharma and Krishna (SK) test

The testing procedure for testing (1) vs. (2) is defined as

$$Z_{SK} = \sqrt{n}\left(\frac{1}{\widehat{CV}} - \frac{1}{CV_0}\right),$$

We will reject H_0 at α level of significance, when $Z_{Mil} > Z_\alpha$, where Z_α is the upper (1-α) quantile of a standard normal distribution. For details, see Sharma and Krishna (1994).

2.5 Curto and Pinto (CP) test

The testing procedure for testing (1) vs. (2) is defined as

$$Z_{CP} = \frac{\widehat{CV} - CV_0}{\sqrt{SE(\widehat{CV})}}$$

where $SE(\widehat{CV}) \cong \sqrt{V_{GMM}/n}$, $V_{GMM} = \frac{\partial f(\theta)}{\partial \theta} \Sigma \frac{\partial f(\theta)}{\partial \theta'}$, $\frac{\partial f(\theta)}{\partial \theta'} = \begin{bmatrix} \frac{-\sigma}{\mu^2} \\ \frac{1}{2\sigma\mu} \end{bmatrix}$,

$\widehat{\Sigma} = \widehat{\Omega}_0 + \sum_{j=1}^{m} \omega(j,m)(\widehat{\Omega}_j + \widehat{\Omega}'_j)$, $\widehat{\Omega}_j = \frac{1}{n}\sum_{t=j+1}^{n} \varphi(Y_t, \hat{\theta})\varphi(Y_{t-jj}, \hat{\theta})'$,

$\varphi(X_t, \hat{\theta}) = \begin{bmatrix} X_t - \hat{\mu} \\ (X_t - \hat{\mu})^2 - \hat{\sigma}^2 \end{bmatrix}$, $\omega(j,m) = 1 - \frac{j}{m+1}$, m is the truncated lag that must satisfy the condition $m/n \to \infty$ as n increases without bound to ensure consistency. We will reject H_0 at α level of significance, when $Z_{CP} > Z_\alpha$, where Z_α is the upper (1-α) quantile of a standard normal distribution. For details, see Curto and Pintu (2009).

2.6 Gulhar, Kibria, Albatineh and Ahmed's (GKAA) test

The testing procedure for testing (1) vs (2) is defined as

$$\chi^2_{GKAA} = (n-1)\left(\frac{\widehat{CV}}{CV_0}\right)^2,$$

where $\widehat{CV}$ and CV are defined as above. We will reject H_0, if $\chi^2_{GKAA} > \chi^2_{(n-1),\alpha}$, where $\chi^2_{(n-1),\alpha}$ is the upper (1-α) quantile of a central chi-square distribution with (n-1) df. For details, see Gulhar et al. (2012).

2.7 Abu-Shawiesh, Akyuz and Kibria Large (AAKL) sample test

The testing procedure for testing (1) vs (2) is defined as

$$Z_{AAKL} = 2log\left(\frac{CV_0}{\widehat{CV}}\right)/\sqrt{A},$$

where $A = \frac{G_2 + 2n/(n-1)}{n}$, $G_2 = \frac{n-1}{(n-2)(n-3)}[(n-1)g_2 + 6]$, $g_2 = \frac{m_4}{m_2^2} - 3$, $m_4 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^4$ and $m_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. We will reject H_0, when $Z_{AAKL} > Z_\alpha$, reject H_0. For details, see Abu-Shawiesh et al. (2019).

2.8 Abu-Shawiesh, Akyuz and Kibria Augmented Large (AAKL_AUG) sample test

The testing procedure for testing (1) vs (2) is defined as

$$Z_{AAKL\_AUG} = \frac{log\left(\frac{CV_0^2}{\widehat{CV}^2}\right) - C}{\sqrt{B}},$$

where $B = \widehat{var}(\log(S^2))$, $C = \frac{\hat{K}_{e,5}+{}^{2n}/_{(n-1)}}{2n}$, $\hat{K}_{e,5} = (\frac{n+1}{n-1})G_2(1+\frac{5G_2}{n})$ and $G_2$is defined above.

We will reject H_0, when $Z_{AAK\_AUG} > Z_\alpha$. For details, see Abu-Shawiesh et al (2019).

2.9 Abu-Shawiesh, Akyuz & Kibria Augmented DF (ADJ) test

The testing procedure for testing (1) vs (2) is defined as

$$\chi^2_{ADJ} = \left(\frac{\hat{r}\widehat{CV}}{CV_0}\right)^2,$$

where $\hat{r} = \frac{2n}{\gamma+[\frac{2n}{n-1}]}$, $\hat{\gamma} = \frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{S^4} - \left[\frac{3(n-)^2}{(n-2)(n-3)}\right]$ and S^2 is the sample variance.

We will reject H_0, if $\chi^2_{ADJ} > \chi^2_{(n-1),\alpha}$, where $\chi^2_{(n-1),\alpha}$ is the upper (1-α) quantile of a central chi-square distribution with (n-1) df. For details, see Abu-Shawiesh et al (2019).

2.10 Modified McKay's (MMckay) test

It is proposed by Vangel (1996), modified the Mckay test in the following way. The testing procedure for testing (1) vs (2) is defined as

$$\chi^2_{MMckay} = \frac{n(CV_0{}^2+\widehat{CV}^4)-2\widehat{CV}^4}{\widehat{CV}^2(\widehat{CV}^2+1)},$$

where $\widehat{CV}$ and CV are defined as above. We will reject H_0, if $\chi^2_{MMckay} > \chi^2_{(n-1),\alpha}$, where $\chi^2_{(n-1),\alpha}$ is the upper (1-α) quantile of a central chi-square distribution with (n-1) df. For details, see Vangel (1996).

2.11 Panichkitkosolkuls (PAN) test

The testing procedure for testing (1) vs (2) is defined as

$$\chi^2_{PAN} = \frac{n}{\hat{k}+1}\left(\frac{CV_0{}^2}{\hat{k}^2} + \hat{k}^2 - \frac{2\hat{k}^2}{n}\right)$$

where $\tilde{k} = \frac{\sqrt{\sum_{i=1}^{n}(X_i-\bar{x})^2}}{\sqrt{n\bar{x}}}$ and $\widehat{CV}$ is defined as above. We will reject H_0, if $\chi^2_{PAN} > \chi^2_{(n-1),\alpha}$, where $\chi^2_{(n-1),\alpha}$ is the upper (1-α) quantile of a central chi-square distribution with (n-1) df. For details, see Panichkitkosolkul*s* (2009).

2.12 Parametric Bootstrap (PB) test

A bootstrap test is a statistical method used to estimate the sampling distribution of a test statistic by resampling with replacement from the observed data. It is particularly useful when the theoretical distribution of the statistic is complex or unknown. The parametric bootstrap technique is described below:

Suppose we have an original sample $X_1, X_2,\ldots, X_n$ with a sample mean and a sample standard deviation and find the estimated CV of the original sample. Draw B bootstrap samples from the original dataset. Each bootstrap sample is a random resampling of the original dataset with replacement. For each bootstrap sample, compute the CV (denoted as CV^*) based on the resampled data. The collection of B bootstrap CV values $\{CV^{*(1)}, CV^{*(2)},\ldots,CV^{*(B)}\}$ forms an empirical distribution of the CV under H_0 that the true CV is equal to a specified value. The test statistics for testing the CV parameter using bootstrap samples is essentially based on comparing the observed CV with the distribution of CV values obtained from the bootstrap samples is defined as:

$$T^*_{n(i)} = \frac{CV_0 - \widehat{CV}}{S_{\widehat{CV}}}$$

where $\widehat{CV}$ is the sample CV, $T^*_{(n),\alpha/2}$ and $T^*_{(n),1-\alpha/2}$ are the $\alpha/2^{th}$ and $(1-\alpha/2)^{th}$ sample quantiles of $T_i^* = \frac{CV_i^* - \overline{\overline{CV}}}{\hat{\sigma}_{CV}}$, $\hat{\sigma}_{CV} = \sqrt{\frac{1}{B}\sum_{i=1}^{n}(CV_i^* - \overline{\overline{CV}})^2}$ and $\overline{\overline{CV}} = \frac{1}{n}\sum_{i=1}^{n} CV_i^*$ is a bootstrap CV.

We will reject H_0 at α level of significance when $T^*_{n(i)} > T^*_{(n),\alpha}$. The size and power of the test is estimated as the proportion of times H_0 is rejected when H_0 and H_1 are true.

2.13 Nonparametric-Bootstrap (NPB) test

Consider the hypotheses: (1) and (2). Calculate the sample CV from the observed data. Create bootstrap samples by randomly resampling with replacement from the original data. Typically, generate B bootstrap samples, where each sample has the same size n as the original sample. For each bootstrap sample, calculate the CV, following the same formula as for the original sample. The bootstrap distribution of the CV will allow us to estimate how likely it is to observe a CV as extreme (or more extreme) than the observed one, assuming H_0 is true. To calculate the p-value, we compare the observed CV to the bootstrap distribution. The p-value is the proportion of bootstrap samples whose CV is greater than or equal to the observed CV, defined as:

p-value = (Number of bootstrap samples with CV ≥ CV0) / B

If the p-value is less than the chosen significance level (e.g., α=0.05), reject H_0. Otherwise, fail to reject H_0. The size and power of the test is estimated as the proportion of times H_0 is rejected when hypotheses are true.

Now, we will propose some new test statistics for testing the population CV in the following section.

2.14 Proposed median version tests

The median is widely recognized as the most appropriate measure of central tendency for skewed distributions, as it is less influenced by extreme values or outliers. Unlike the mean, which can be significantly affected by unusually large or small observations, the median represents the middle value of an ordered dataset. This provides a more accurate reflection of the distribution's central point without being distorted by skewness. In a skewed distribution, where one tail is longer or more dispersed than the other, the median serves as a more reliable indicator of central location. Shi and Kibria (2007) emphasized that for skewed data distributions, it is often more appropriate to define the sample standard deviation relative to the median rather than the mean. The primary reason is that the median serves as a robust measure of central tendency and is less affected by extreme observations or outliers, which commonly occur in skewed distributions. In contrast, the mean can be substantially influenced by such extremes, potentially leading to misleading inferences. Consequently, statistical measures that rely on the mean may not adequately capture the underlying variability of skewed data. Motivated by this insight, and to improve the reliability of inference under skewness, we propose several test statistics in this section for evaluating the coefficient of variation (CV) parameter based on median-centered formulations.

2.14.1 Median based classical test (t_me)

The test statistic for the CV can be defined as:

$$t_{Me} = \frac{\widetilde{CV} - CV_0}{S_{\widetilde{CV}}},$$

where $\widetilde{CV} = \frac{\tilde{S}}{\tilde{X}}$ is the sample CV, $\tilde{S} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \tilde{X})^2}$ is the modified sample standard deviation, $\tilde{X}$ is the sample median, CV_0 is the specified CV under H_0 and $S_{\widetilde{CV}} = \frac{\widetilde{CV}}{\sqrt{2n}}$ is the standard error of the CV. We will reject (1) when $t_{Me} > t_{(n-1),\alpha}$, where $t_{(n-1),\alpha}$ is the upper (α) percentile from a student's t distribution with (n-1) df.

2.14.2 Median version of McKay's test (McKay_Me)

To test hypothesis (1) vs. (2), the modified version of McKay test is as follows,

$$\chi^2_{Mk\_Me} = \frac{n(\widetilde{CV}^2 + \widetilde{CV}^4)}{\widetilde{CV}^2(\widetilde{CV}^2 + 1)},$$

where CV and $\widetilde{CV}$ are defined as above. We refer to this test statistic as Mk_Me, as it is based on a modified sample standard deviation. We will reject H_0, when $\chi^2_{Mk\_Me} > \chi^2_{(n-1),\alpha}$.

2.14.3 Median version of the Miller test (Miller_Me)

The testing procedure for testing (1) vs (2) is defined as

$$Z_{Mi\_Me} = \frac{\widetilde{CV} - CV_0}{\sqrt{\frac{{CV_0}^4 + 0.5{CV_0}^2}{n-1}}}$$

where $\widetilde{CV}$ and CV are defined as above. We will reject H_0, when $Z_{Mi\_Me} > Z_\alpha$.

2.14.4 Median version of the Sharma and Krishna (SK_Me) test

The testing procedure for testing (1) vs (2) is defined as

$$Z_{SK\_Me} = \sqrt{n}\left(\frac{1}{\widetilde{CV}} - \frac{1}{CV_0}\right)$$

where $\widetilde{CV}$ and CV are defined as above. We will reject H_0, when $Z_{SK\_Me} > Z_\alpha$.

2.14.5 Median version of the Curto and Pinto test (CP_Me)

For testing (1) vs (2), the test statistics is defined as

$$Z_{CP\_Me} = \frac{\widetilde{CV} - CV_0}{\sqrt{SE(\widetilde{CV})}}$$

where $\left(\widetilde{CV}\right) = \frac{\widetilde{CV} - CV_0}{\sqrt{\frac{1}{n}(\widetilde{CV}^4 + 0.5\widetilde{CV}^2)}}$, $\widetilde{CV}$ and CV_0 are defined as above. We will reject H_0, when $Z_{CP\_Me} > Z_\alpha$.

2.14.6 Median version of the Gulhar, Kibria, Albatineh and Ahmed's (GKAA_Me) test

The testing procedure for testing (1) vs (2) is defined as

$$\chi^2_{GKAA\_Me} = (n-1)\left(\frac{\widetilde{CV}}{CV_0}\right)^2$$

where $\widetilde{CV}$ and CV are defined as above. We will reject H_0, when $\chi^2_{GKAA\_Me} > \chi^2_{(n-1),\alpha}$.

2.14.7 Median version Abu-Shawiesh, Akyuz and Kibria Large (AAKL_Me) sample test

The testing procedure for testing (1) vs (2) is defined as

$$Z_{AAKL\_Me} = 2log\left(\frac{CV_0}{\widetilde{CV}}\right)/\sqrt{A}$$

where $A = \frac{G_2+2n/(n-1)}{n}$, $G_2 = \frac{n-1}{(n-2)(n-3)}[(n-1)g_2+6]$, $g_2 = \frac{m_4}{m_2^2} - 3$, $m_4 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^4$ and $m_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. We will reject H_0, when $Z_{AAKL\_Me} > Z_\alpha$.

2.14.8 Median version of ADJ (ADJ_Me) test

The testing procedure for testing (1) vs (2) is defined as

$$\chi^2_{ADJ\_Me} = \left(\frac{\hat{r}\widehat{CV}}{CV_0}\right)^2$$

We will reject H_0 if $\chi^2_{ADJ\_Me} > \chi^2_{(n-1),\alpha}$, where $\chi^2_{(n-1),\alpha}$ is the upper (1-α) quantile of a central chi-square distribution with (n-1) df.

2.14.9 Median version of Abu-Shawiesh, Akyuz and Kibria Augmented Large (AAKL_AUG_Me) sample test

The testing procedure for testing (1) vs (2) is defined as

$$Z_{AAKL\_AUG\_Me} = \frac{log\left(\frac{{CV_0}^2}{\widetilde{CV}^2}\right) - C}{\sqrt{B}}$$

where $B = \widehat{var}(\log(S^2))$, $C = \frac{\hat{K}_{e,5}+2n/(n-1)}{2n}$, $\hat{K}_{e,5} = (\frac{n+1}{n-1})G_2(1+\frac{5G_2}{n})$ and G_2 is defined above.

We will reject H_0, when $Z_{AAKL\_AUG\_Me} > Z_\alpha$.

2.14.10 Modified McKay's test (MMckay_Me)

The median version of the Vangel (1996) testing procedure for testing (1) vs (2) is defined as

$$\chi^2_{MMK\_Me} = \frac{n\left({CV_0}^2 + \widetilde{CV}^4\right) - 2\widetilde{CV}^4}{\widetilde{CV}^2(\widetilde{CV}^2+1)}$$

where $\widetilde{CV}$ and CV are defined as above. We will reject H_9, when $\chi^2_{MMK\_Me} > \chi^2_{(n-1),\alpha}$.

2.14.11 Median version of the Panichkitkosolkuls (Pan_Me) test

The testing procedure for testing (1) vs (2) is defined as

$$\chi^2_{PAN\_Me} = \frac{n}{\tilde{k}+1}\left(\frac{{CV_0}^2}{\tilde{k}^2} + \tilde{k}^2 - \frac{2\tilde{k}^2}{n}\right)$$

where $\tilde{k} = \frac{\sqrt{\sum_{i=1}^{n}(X_i - Me)^2}}{\sqrt{nMe}}$ and $\widetilde{CV}$ is defined as above. We will reject H_0, if $\chi^2_{PAN,\_Me} > \chi^2_{(n-1),\alpha}$

3. Simulation design

To compare the performance of our various considered test statistics, a simulation study has been conducted in this section. It consists of two subsections: simulation design and simulations results discussion

3.1 Simulation Design

The simulation technique involves generating data from different distributions, applying the CV test to the generated samples and estimating the test's size and power across various sample sizes and distributions. The structure of the simulation is as follows:

- Sample sizes: n=10, 20, 30, 50, 100, 200 and 300.
- For each n, generate 10,000 random samples from each distribution: N(3,2), Gamma(2,2), $\chi^2_{(1)}$ and Lognormal(2,1).
- For each sample, compute CV. Apply the CV test, which typically involves comparing the sample's CV to a critical value under H_0.
- Bootstrap sampling: For each generated sample, perform 5,000 bootstrap resamples to estimate the sampling distribution of the CV.
- Hypothesis testing: For each sample, conduct a hypothesis test (e.g., testing whether the CV equals a specified value under H_0). The critical region for rejecting H_0 is determined using the bootstrap distribution.
- Estimation of size and power: Estimate the size of the test as the proportion of times the H_0 is rejected when it is true. Estimate the power of the test as the proportion of times H_0 is rejected when it is false (using distributions such as Gamma(2,2), $\chi^2_{(1)}$ and Lognormal(2,1)).
- A conventional significance level of α=0.05 is adopted throughout the study.

The simulation results are presented in Tables 3.1 to 3.8.

3.2 Discussion of simulation results

In this section, we examined and analyzed the results based on our simulation results. Table 3.1 presents (see the Figure 3.1 for visualization) type 1 error rates for all the tests we considered. The vital findings are outlined below:

Table 3.1: Size of the CV tests when data generated from the N(3,2) distribution

Test Statistic	n=10	n=20	n=30	n=50	n=100	n=200	n=300
t_{cal}	0.009	0.023	0.034	0.033	0.044	0.051	0.060
χ^2_{Mck}	0.201	0.133	0.122	0.096	0.089	0.074	0.074
Z_{Mil}	0.016	0.030	0.037	0.034	0.040	0.043	0.047
Z_{SK}	0.021	0.025	0.025	0.021	0.021	0.031	0.007
Z_{CP}	0.033	0.030	0.029	0.025	0.031	0.034	0.032
χ^2_{GKAA}	0.075	0.076	0.075	0.066	0.068	0.065	0.074
Z_{AAKL}	0.111	0.086	0.086	0.078	0.082	0.075	0.078
χ^2_{ADJ}	0.007	0.009	0.010	0.022	0.037	0.043	0.012
Z_{AAKL_AUG}	0.005	0.007	0.009	0.0114	0.0106	0.0144	0.0301
χ^2_{MMckay}	0.197	0.129	0.121	0.095	0.087	0.072	0.074
χ^2_{PAN}	0.174	0.079	0.051	0.057	0.059	0.052	0.054
t_{Me}	0.020	0.042	0.044	0.050	0.058	0.055	0.071
$\chi^2_{Mk\_Me}$	0.187	0.128	0.121	0.093	0.091	0.077	0.079
Z_{Mi_Me}	0.030	0.052	0.049	0.050	0.054	0.055	0.059
Z_{SK_Me}	0.036	0.031	0.028	0.020	0.029	0.042	0.043
Z_{CP_Me}	0.074	0.064	0.057	0.050	0.048	0.050	0.048
$\chi^2_{GKAA\_Me}$	0.101	0.097	0.095	0.090	0.086	0.084	0.087
Z_{AAKL_Me}	0.104	0.079	0.085	0.077	0.085	0.079	0.084
$\chi^2_{ADJ\_Me}$	0.014	0.020	0.021	0.035	0.041	0.052	0.049
$Z_{AAKL\_AUG\_Me}$	0.015	0.019	0.0206	0.0196	0.0196	0.023	0.039
$\chi^2_{MMK\_Me}$	0.185	0.125	0.118	0.092	0.090	0.076	0.079
$\chi^2_{PAN\_Me}$	0.174	0.079	0.051	0.057	0.059	0.052	0.053
$T^*_{n(i)}$	0.006	0.012	0.019	0.017	0.035	0.041	0.043
NPB	0.247	0.154	0.136	0.103	0.093	0.074	0.069

As n increases, the size of most test statistics converges toward the nominal level of 0.05, except χ^2_{Mck} Z_{AAKL}, $\chi^2_{MMK\_Me}$, χ^2_{PAN}, $\chi^2_{Mk\_Me}$, $\chi^2_{GKAA\_Me}$, $\chi^2_{PAN\_Me}$, $\chi^2_{PAN\_Me}$ and the NPB test. However, these tests can be used for very large samples. In contrast, tests like t_{cal}, χ^2_{ADJ} and $T^*_{n(i)}$ demonstrate very low size values, particularly for small n, indicating a conservative tendency. The t_{cal}, in particular, shows highly conservative behavior, with exceptionally low size values (e.g., 0.009 for n=10), while the t_{Me} test performs slightly better, yielding values closer to 0.05, though it remains somewhat conservative. χ^2_{Mck} and $\chi^2_{Mk\_Me}$ initially display high size values (e.g., 0.201 for n=10), which decrease as n increases but still remain somewhat inflated, suggesting a liberal tendency (elevated type I error) for small n. The $T^*_{n(i)}$is highly conservative for small n (e.g., 0.006 for n=10) but gradually approaches 0.05 as n increases. Conversely, the NPB starts with an extremely large size (0.247 for n=10), indicating a strong liberal bias, but stabilizes as n grows.

Consistently stable tests are observed according to our study results: $\boldsymbol{Z_{Mil}}$, $\boldsymbol{Z_{SK}}$, $\boldsymbol{Z_{CP}}$ and $\boldsymbol{\chi^2_{GKAA}}$ maintain relatively stable sizes across different n, showing more balanced behavior. These might be better choices for small n due to their controlled type I error rates. Comparison between standard and modified (Me) versions are: In many cases, the "Me" versions of tests (e.g., $\boldsymbol{\chi^2_{Mk\_Me}}$, $\boldsymbol{\chi^2_{MMK\_Me}}$, $\boldsymbol{\chi^2_{GKAA\_Me}}$) slightly improve the performance, but differences are not always substantial. Some "Me" versions are more conservative than their standard counterparts, particularly for small n.

Best and worst performing tests according to our simulation plan for N(3,2) DGP are:

(i) Best performing (closer to 0.05 for all n): Z_{Mil}, Z_{SK}, Z_{CP} and χ^2_{GKAA}.
(ii) Worst performing (highly conservative or liberal): χ^2_{Mck} (too liberal for small n), NPB (highly liberal for small n) and t_{cal} and χ^2_{ADJ} (too conservative).

Thus it may be concluded that for N(3,2) DGP, for small sample sizes (n≤30), Z_{Mil}, Z_{SK}, Z_{CP} and χ^2_{GKAA} seem to be the best choices. For larger sample sizes (n≥50), most tests perform adequately but t_{cal} remains slightly conservative. NPB and χ^2_{Mck} should be used cautiously for small n due to inflated type I error rates.

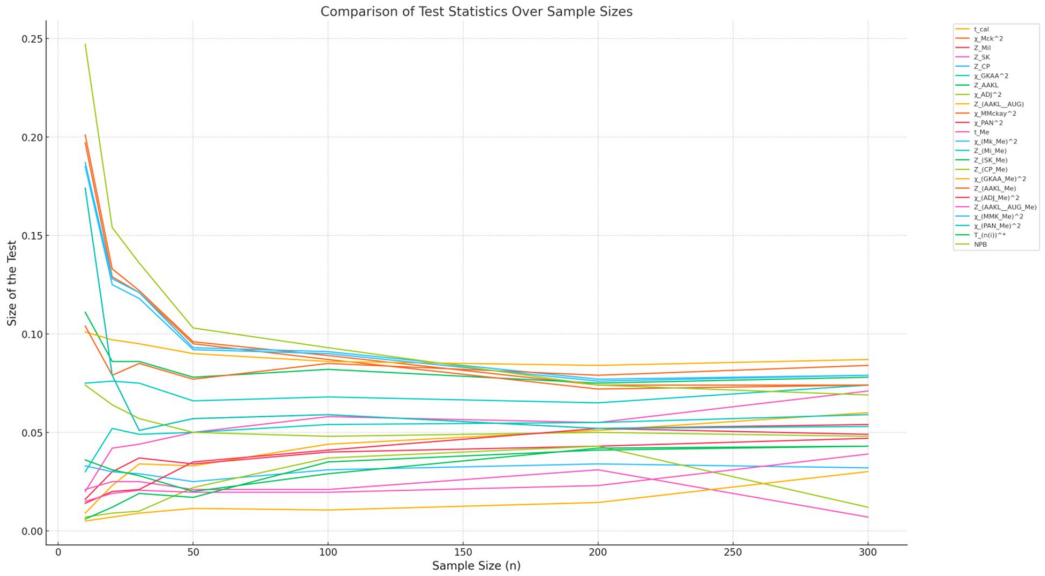

Fig 3.1: Size of the CV tests for N(3,2) DGP

Table 3.2: Power of the CV tests when data generated from the N(3,2) distribution

Test Statistic	n=10	n=20	n=30	n=50	n=100	n=200	n=300
t_{cal}	0.793	0.971	0.995	0.982	1.000	1.000	1.000
χ^2_{Mck}	0.853	0.971	0.994	0.979	1.000	1.000	1.000
Z_{Mil}	0.462	0.871	0.974	0.945	0.999	1.000	1.000
Z_{SK}	0.889	0.992	1.000	0.991	1.000	1.000	1.000
Z_{CP}	0.849	0.963	0.992	0.971	0.999	1.000	1.000
χ^2_{GKAA}	0.520	0.910	0.986	0.965	1.000	1.000	1.000
Z_{AAKL}	0.781	0.968	0.991	0.980	1.000	1.000	1.000
χ^2_{ADJ}	0.165	0.485	0.810	0.746	0.991	1.000	1.000
Z_{AAKL_AUG}	0.785	0.963	0.996	0.982	1.000	1.000	1.000
χ^2_{MMckay}	0.836	0.968	0.994	0.979	1.000	1.000	1.000
χ^2_{PAN}	0.797	0.925	0.974	0.887	0.987	1.000	1.000
t_{Me}	0.767	0.955	0.992	0.973	1.000	1.000	1.000
$\chi^2_{Mk\_Me}$	0.821	0.955	0.989	0.968	0.999	1.000	1.000
Z_{Mi_Me}	0.438	0.841	0.961	0.923	0.999	1.000	1.000
Z_{SK_Me}	0.867	0.983	0.998	0.983	1.000	1.000	1.000
Z_{CP_Me}	0.817	0.953	0.989	0.962	0.999	1.000	1.000
$\chi^2_{GKAA\_Me}$	0.500	0.883	0.975	0.947	0.999	1.000	1.000
Z_{AAKL_Me}	0.750	0.947	0.986	0.969	1.000	1.000	1.000
$\chi^2_{ADJ\_Me}$	0.147	0.464	0.780	0.722	0.984	1.000	1.000
$Z_{AAKL\_AUG\_Me}$	0.755	0.950	0.989	0.974	1.000	1.000	1.000
$\chi^2_{MMK\_Me}$	0.811	0.953	0.988	0.966	0.999	1.000	1.000
$\chi^2_{PAN\_Me}$	0.797	0.925	0.974	0.887	0.999	1.000	1.000
$T^*_{n(i)}$	0.821	0.965	0.993	0.977	1.000	1.000	1.000
NPB	0.898	0.981	0.997	0.987	1.000	1.000	1.000

Table 3.2 presents the power of various CV tests for different n. It is observed from Table 3.2 is that as n increases, the power of all tests improves, often reaching 1.000 at n ≥100. We will compare power among various tests only those have attained the nominal level 0.05. For n=10, χ^2_{ADJ}, $\chi^2_{ADJ\_Me}$, Z_{Mil}, Z_{Mi_Me}, χ^2_{GKAA}, $\chi^2_{GKAA\_Me}$ have low power. t_{cal}, Z_{SK}, t_{Me}, Z_{CP} and tests generally show high power across all n. Z_{Mil}, tests exhibit lower power, especially at small n, making them less reliable for detecting variation in small samples. χ^2_{ADJ} starts with only 0.165 power at n=10 but improves significantly for larger n. Most "Me" versions have slightly lower power than their original counterparts but follow similar trends. Some tests, such as Z_{SK_Me}, remain competitive with their standard versions. NPB outperforms $T^*_{n(i)}$, especially at small and moderate n. Both bootstrap methods reach power 1.000 at n≥100, making them robust choices for larger datasets.

According to our simulation findings, it is found that for small sample sizes, Z_{SK} and NPB perform best. For moderate sample sizes, Z_{SK}, NPB and t_{cal} remain strong choices. For large samples, all tests perform well, but bootstrap methods and Z_{SK} remain among the most reliable.

The following Figure 3.2 is the visualization of the power of various CV tests across different n. One can see how power improves with larger n and which tests perform better overall.

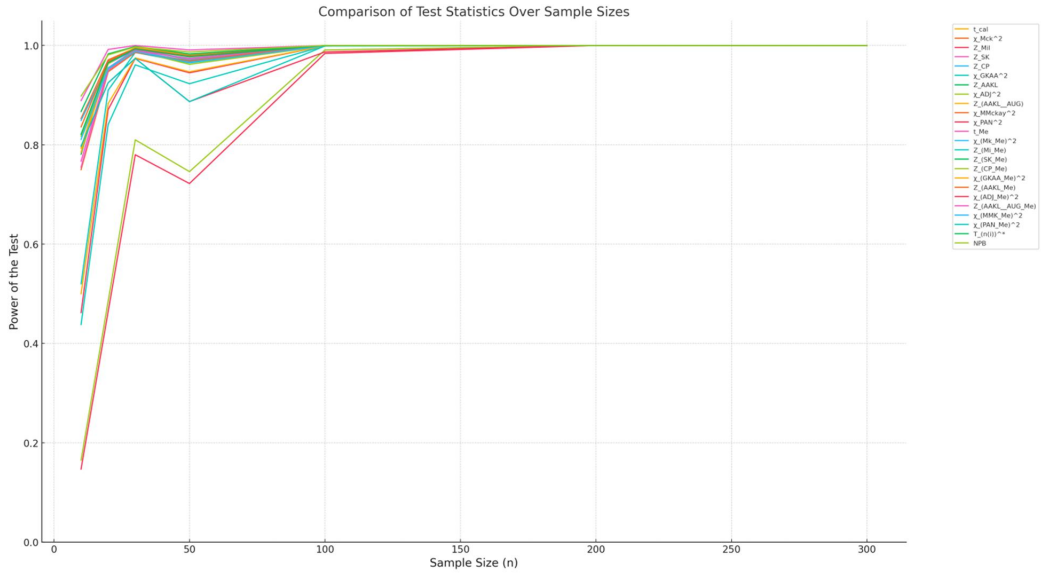

Fig 3.2: Power of the CV tests for N(3,2) DGP

Table 3.3: Size of the CV test when data generated from the G(2,2) distribution with skewness 1.414

Test Statistic	n=10	n=20	n=30	n=50	n=100	n=200	n=300
t_{cal}	0.006	0.022	0.028	0.046	0.053	0.067	0.061
χ^2_{Mck}	0.177	0.121	0.098	0.080	0.061	0.045	0,051
Z_{Mil}	0.000	0.000	0.001	0.006	0.012	0.015	0.024
Z_{SK}	0.089	0.106	0.107	0.128	0.130	0.149	0.151
Z_{CP}	0.042	0.045	0.047	0.048	0.052	0.054	0.053
χ^2_{GKAA}	0.048	0.059	0.062	0.077	0.077	0.083	0.073
Z_{AAKL}	0.082	0.093	0.095	0.097	0.098	0.100	0.083
χ^2_{ADJ}	0.023	0.028	0.030	0.032	0.035	0.040	0.044
Z_{AAKL_AUG}	0.076	0.087	0.089	0.085	0.088	0.087	0.084
χ^2_{MMckay}	0.117	0.114	0.092	0.074	0.057	0.043	0.047
χ^2_{PAN}	0.131	0.058	0.053	0.052	0.053	0.051	0.053
t_{Me}	0.008	0.025	0.030	0.043	0.058	0.060	0.054
$\chi^2_{Mk\_Me}$	0.180	0.118	0.100	0.089	0.057	0.048	0.054
Z_{Mi_Me}	0.005	0.008	0.010	0.017	0.023	0.034	0.043
Z_{SK_Me}	0.079	0.059	0.033	0.017	0.004	0.001	0.000
Z_{CP_Me}	0.045	0.048	0.053	0.052	0.048	0.049	0.049
$\chi^2_{GKAA\_Me}$	0.039	0.046	0.045	0.047	0.051	0.056	0.052
Z_{AAKL_Me}	0.052	0.064	0.077	0.082	0.095	0.096	0.083
$\chi^2_{ADJ\_Me}$	0.027	0.034	0.037	0.039	0.040	0.043	0.047
$Z_{AAKL\_AUG\_Me}$	0.043	0.046	0.067	0.078	0.084	0.089	0.085
$\chi^2_{MMK\_Me}$	0.056	0.057	0.063	0.071	0.087	0.097	0.086
$\chi^2_{PAN\_Me}$	0.127	0.054	0.052	0.051	0.054	0.052	0.054
$T^*_{n(i)}$	0.010	0.021	0.026	0.029	0.030	0.031	0.032
NPB	0.154	0.096	0.087	0.075	0.072	0.069	0.062

Table 3.3 the empirical size of various CV tests when applied to data generated from a G(2,2) distribution. We observed that some tests show decreasing size as n increases (χ^2_{Mck}, $\chi^2_{Mk\_Me}$, Z_{SK_Me}, NPB), indicating they become more conservative with larger samples. Other tests Z_{SK}, Z_{AAKL}, $\chi^2_{MMK\_Me}$ show an increasing size trend, meaning they become more liberal at larger n. A few tests: Z_{CP}, χ^2_{ADJ}, parametric bootstrap) remain fairly stable across all n.

The following tests are found with size closest to 0.05 (good control of type I error)

(i) Z_{CP}, χ^2_{GKAA}, Z_{AAKL}, χ^2_{ADJ} and $T^*_{n(i)}$ maintain size close to 0.05 across all n, making them reliable choices for controlling type I error.
(ii) t_{cal} approaches 0.05 for larger n, improving its validity.
(iii) χ^2_{PAN} test stabilizes near 0.05 for larger n.

The following tests with overly conservative size (type I error too low); $\boldsymbol{Z_{Mil}}$, and $\boldsymbol{Z_{Mi\_Me}}$ are extremely conservative, with sizes close to **zero** at small n and only slightly
increasing with larger n. These tests are too restrictive and may fail to reject H_0 even when appropriate. $\boldsymbol{t_{cal}}$ is also highly conservative for small n but improves with increasing n. The following tests with inflated size (type I error too high); $\boldsymbol{\chi^2_{Mck}}$ and $\boldsymbol{\chi^2_{Mk\_Me}}$ have excessively large size at small n (e.g., 0.177 at n=10), meaning they are too liberal and reject H_0 too often. NPB has inflated size at small n (0.154 for n=10), which decreases as n increases, suggesting instability. $\boldsymbol{\chi^2_{PAN}}$ starts at 0.117 for n=10 and gradually decreases, showing a lack of stability.

Best tests for controlling size (close to 0.05) are observed: Z_{CP}, χ^2_{GKAA}, Z_{AAKL}, χ^2_{ADJ} and $T^*_{n(i)}$. These tests are reliable across all n values**.** Some tests that are too conservative (reject too rarely, low power risk)**:** Z_{Mil}, Z_{Mi_Me} and t_{cal} (for small n). These tests may be too strict especially for small n. Some tests that are observed too liberal (reject too often, high false positives) are**:** χ^2_{Mck}, $\chi^2_{Mk\_Me}$ NPB and χ^2_{PAN} (for small n*)*. These tests may be unreliable for small samples due to excessive type I errors.

The plot (Figure 3.3) visualizes the size of different CV tests when data is generated from a Gamma(2,2) distribution with skewness 1.414. This made us clear some certain tests may not be well-calibrated for slightly positively skewed distributions, particularly at small n.

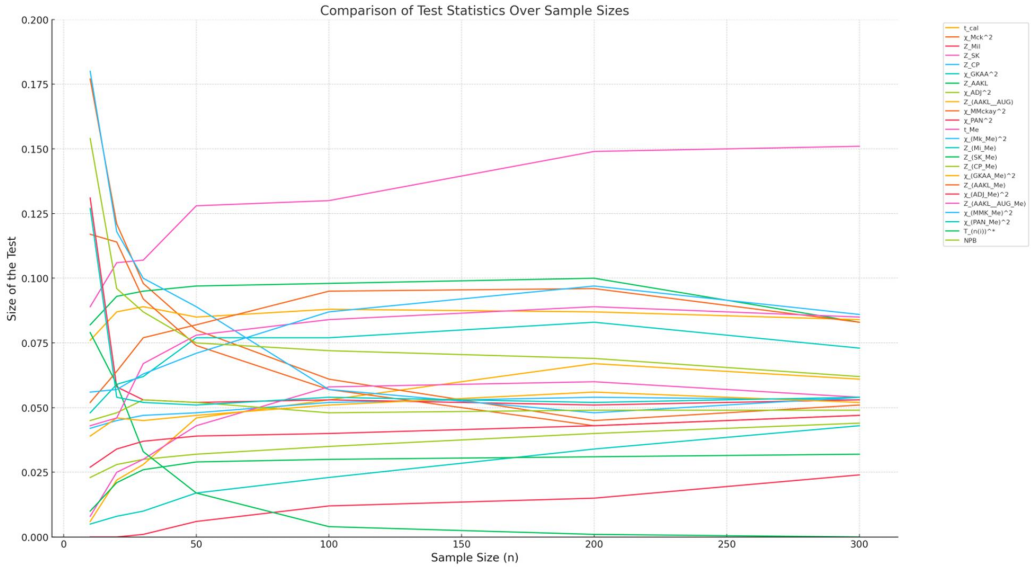

Fig 3.3: Size of the CV tests for G(2,2) DGP

Table 3.4: Power of the CV test when data generated from the G(2,2) distribution with skewness 1.414

Test Statistic	n=10	n=20	n=30	n=50	n=100	n=200	n=300
t_{cal}	0.301	0.400	0.498	0.599	0.603	0.702	0.710
χ^2_{Mck}	0.127	0.281	0.363	0.445	0.494	0.623	0.625
Z_{Mil}	0.116	0.218	0.222	0.419	0.523	0.734	0.745
Z_{SK}	0.352	0.370	0.370	0.379	0.392	0.397	0.401
Z_{CP}	0.193	0.244	0.325	0.504	0.681	0.767	0.801
χ^2_{GKAA}	0.255	0.371	0.384	0.487	0.588	0.799	0.823
Z_{AAKL}	0.353	0.473	0.582	0.587	0.695	0.700	0.731
χ^2_{ADJ}	0.112	0.136	0.208	0.310	0.414	0.470	0.483
Z_{AAKL_AUG}	0.360	0.479	0.590	0.632	0.699	0.745	0.762
χ^2_{MMckay}	0.120	0.190	0.358	0.442	0.533	0.622	0.634
χ^2_{PAN}	0.295	0.435	0.521	0.608	0.701	0.769	0.782
t_{Me}	0.574	0.665	0.730	0.828	0.941	0.993	0.996
$\chi^2_{Mk\_Me}$	0.523	0.554	0.591	0.679	0.827	0.958	0.963
Z_{Mi_Me}	0.280	0.409	0.488	0.619	0.810	0.961	0.968
Z_{SK_Me}	0.873	0.903	0.921	0.953	0.987	0.999	1.000
Z_{CP_Me}	0.655	0.693	0.732	0.814	0.916	0.983	0.984
$\chi^2_{GKAA\_Me}$	0.406	0.564	0.652	0.781	0.924	0.991	0.994
Z_{AAKL_Me}	0.456	0.602	0.685	0.803	0.931	0.992	0.995
$\chi^2_{ADJ\_Me}$	0.112	0.177	0.237	0.314	0.472	0.695	0.714
$Z_{AAKL\_AUG\_Me}$	0.506	0.613	0.690	0.825	0.917	0.991	0.994
$\chi^2_{MMK\_Me}$	0.507	0.533	0.575	0.665	0.819	0.954	0.962
$\chi^2_{PAN\_Me}$	0.745	0.851	0.923	0.984	0.999	1.000	1.000
$T^*_{n(i)}$	0.146	0.233	0.422	0.412	0.692	0.778	0.812
NPB	0.251	0.308	0.378	0.557	0.726	0.796	0.813

The results describe the power of various CV tests when data is generated from a Gamma(2,2) distribution with skewness 1.414 are tabulated in the Table 3.4. Our observations are: For most tests, power increases as n grows, which is expected. Some tests (e.g., Z_{SK_Me}, $\chi^2_{PAN\_Me}$) achieve near-perfect power (>0.99) at large n, making them highly effective for detecting differences in CV. Z_{SK_Me}, $\chi^2_{PAN\_Me}$, Z_{CP_Me}, Z_{AAKL_Me}, and $\chi^2_{GKAA\_Me}$ show strong performance across all n. t_{cal}, χ^2_{Mck}, χ^2_{PAN} and χ^2_{GKAA} start with low power but improve significantly at larger n. Z_{Mil}, χ^2_{ADJ} and NPB struggle with low power at smaller n, indicating weaker ability to detect true effects. $T^*_{n(i)}$ shows moderate power growth. NPB struggles at small n but improves at larger n. If detecting differences in CV is crucial, using Z_{SK_Me}, $\chi^2_{PAN\_Me}$, or

Z_{CP_Me} is beneficial, especially at small to moderate n. Some classical methods Z_{Mil} and (e.g., χ^2_{ADJ}) may not be reliable for detecting CV differences, particularly at small n.

Here is the plot (see Figure: 3.4) showing the power of different CV tests as a function of sample size when data is generated from a G(2,2) distribution.

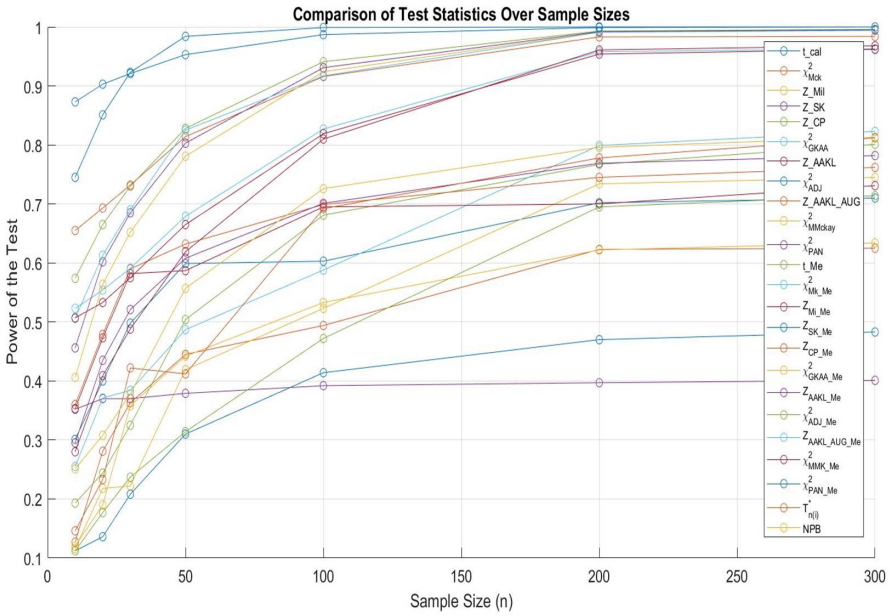

Fig 3.4: Power of the CV tests for G(2,2) DGP

Table 3.5: Size of the CV test when data generated from the chi-square distribution with 1 df and skewness 2

Test Statistic	n=10	n=20	n=30	n=50	n=100	n=200	n=300
t_{cal}	0.003	0.024	0.047	0.062	0.089	0.128	0.112
χ^2_{Mck}	0.156	0.096	0.072	0.050	0.034	0.017	0.020
Z_{Mil}	0.004	0.019	0.020	0.023	0.026	0.025	0.029
Z_{SK}	0.128	0.172	0.213	0.234	0.270	0.296	0.245
Z_{CP}	0.000	0.022	0.035	0.028	0.025	0.027	0.032
χ^2_{GKAA}	0.032	0.055	0.061	0.065	0.064	0.062	0.061
Z_{AAKL}	0.018	0.019	0.020	0.026	0.029	0.038	0.042
χ^2_{ADJ}	0.009	0.016	0.025	0.032	0.041	0.045	0.047
Z_{AAKL_AUG}	0.020	0.022	0.025	0.029	0.031	0.040	0.043
χ^2_{MMckay}	0.130	0.079	0.056	0.049	0.046	0.045	0.048
χ^2_{PAN}	0.015	0.020	0.043	0.061	0.064	0.062	0.053
t_{Me}	0.018	0.026	0.031	0.035	0.043	0.054	0.055
$\chi^2_{Mk\_Me}$	0.029	0.037	0.044	0.052	0.058	0.061	0.063
Z_{Mi_Me}	0.009	0.020	0.034	0.043	0.045	0.046	0.052
Z_{SK_Me}	0.103	0.094	0.088	0.073	0.075	0.093	0.083
Z_{CP_Me}	0.004	0.029	0.033	0.039	0.043	0.048	0.052
$\chi^2_{GKAA\_Me}$	0.025	0.031	0.035	0.038	0.045	0.051	0.055
Z_{AAKL_Me}	0.038	0.032	0.039	0.047	0.042	0.047	0.049
$\chi^2_{ADJ\_Me}$	0.085	0.084	0.070	0.041	0.052	0.051	0.053
$Z_{AAKL\_AUG\_Me}$	0.040	0.037	0.043	0.051	0.047	0.049	0.053
$\chi^2_{MMK\_Me}$	0.270	0.163	0.114	0.073	0.055	0.053	0.054
$\chi^2_{PAN\_Me}$	0.019	0.025	0.052	0.055	0.062	0.057	0.053
$T^*_{n(i)}$	0.013	0.022	0.027	0.028	0.039	0.043	0.046
NPB	0.019	0.025	0.032	0.034	0.037	0.039	0.041

Table 3.5 presents the size of various CV tests when data are drawn from a chi-square distribution with 1 df, which is highly skewed (skewness = 2)**.** The goal is to evaluate the performance of different methods across increasing sample sizes (n = 10, 20, 30, 50, 100, 200, 300**).** We observed that t_{cal} values increase as n grows, indicating that the test statistic becomes more sensitive to detecting deviations in larger n. χ^2_{Mck} statistic decrease as n increases, suggesting that this method stabilizes with larger n. Z_{Mil} statistics remain relatively stable across different Z_{SK}statistics show an increasing trend, suggesting that the test statistics grow with n. Z_{CP} statistics fluctuate slightly but remain relatively small. χ^2_{GKAA}, Z_{AAKL} and χ^2_{ADJ} statistics exhibit a gradual increase in test statistic values with n. χ^2_{MMckay}statistics show a decreasing trend, meaning the test statistic stabilizes with larger n. χ^2_{PAN} statistic initially increases with n,

then stabilizes. Bootstrap statistics $T^*_{n(i)}$ and NPB) show increasing trends, meaning they become more sensitive as n increases.

Overall, it is found that

(i) Some statistics (χ^2_{Mck} and χ^2_{MMckay}) stabilize with larger n.

(ii) Others (Z_{SK}, χ^2_{GKAA}, Z_{AAKL}, $T^*_{n(i)}$) increase with n, suggesting they may be more conservative in smaller n.

(iii) t_{cal} and related statistics tend to become more sensitive with increasing n.

(iv) $T^*_{n(i)}$ provides alternative estimations, generally increasing but stabilizing.

Table 3.6: Power of the CV test when data generated from the chi-square distribution with 1 df and skewness 2

Test Statistic	n=10	n=20	n=30	n=50	n=100	n=200	n=300
t_{cal}	0.159	0.277	0.495	0.514	0.623	0.738	0.743
χ^2_{Mck}	0.106	0.257	0.339	0.424	0.512	0.609	0.620
Z_{Mil}	0.150	0.180	0.283	0.314	0.412	0.513	0.527
Z_{SK}	0.709	0.717	0.742	0.752	0.757	0.766	0.782
Z_{CP}	0.202	0.391	0.481	0.567	0.606	0.711	0.716
χ^2_{GKAA}	0.158	0.209	0.345	0.574	0.626	0.733	0.737
Z_{AAKL}	0.126	0.261	0.378	0.403	0.506	0.734	0.739
χ^2_{ADJ}	0.119	0.216	0.338	0.442	0.516	0.776	0.783
Z_{AAKL_AUG}	0.134	0.267	0.387	0.415	0.574	0.798	0.803
χ^2_{MMckay}	0.091	0.249	0.331	0.420	0.556	0.708	0.714
χ^2_{PAN}	0.101	0.251	0.342	0.448	0.511	0.608	0.614
t_{Me}	0.528	0.560	0.579	0.608	0.626	0.643	0.652
$\chi^2_{Mk\_Me}$	0.246	0.358	0.405	0.556	0.624	0.810	0.816
Z_{Mi_Me}	0.169	0.230	0.301	0.477	0.540	0.623	0.629
Z_{SK_Me}	0.947	0.946	0.949	0.948	0.949	0.954	0.961
Z_{CP_Me}	0.323	0.447	0.501	0.663	0.799	0.863	0.869
$\chi^2_{GKAA\_Me}$	0.525	0.569	0.585	0.613	0.661	0.751	0.803
Z_{AAKL_Me}	0.532	0.549	0.567	0.590	0.691	0.736	0.743
$\chi^2_{ADJ\_Me}$	0.513	0.721	0.830	0.933	0.994	1.000	1.000
$Z_{AAKL\_AUG\_Me}$	0.536	0.553	0.574	0.596	0.695	0.743	0.752
$\chi^2_{MMK\_Me}$	0.231	0.343	0.497	0.559	0.621	0.789	0.793
$\chi^2_{PAN\_Me}$	0.166	0.200	0.306	0.432	0.576	0.843	0.847
$T^*_{n(i)}$	0.180	0.280	0.377	0.571	0.638	0.722	0.725
NPB	0.273	0.347	0.440	0.524	0.675	0.759	0.762

Table 3.6 presents the power of various CV tests when data are generated from a highly skewed chi-square distribution with 1 df. We found that the

t_{cal} has moderate power, starts at 0.159 (n=10) and increases steadily to 0.743 (n=300). Performance is reasonable but not the strongest option. χ^2_{Mck} test has lower power than t_{cal}, especially for small n (0.106 for n=10). This test improves with sample n but still lags behind many other statistics. Z_{Mil} test has weak power compared to other methods, especially for small n. By n=300, power is only 0.527, indicating inefficiency in detecting deviations. Z_{SK} test has consistently high power, starting at 0.709 (n=10) and reaching 0.782 (n=300). one of the strongest performers, showing that it can effectively detect variations in CV. The Z_{CP} test has decent power, starts at 0.247 (n=10) and grows to 0.716 (n=300). It performs well but not the best. χ^2_{GKAA}, Z_{AAKL}, and χ^2_{ADJ} Methods has moderate to high power, with χ^2_{ADJ}being the strongest: χ^2_{ADJ} (0.119 to 0.776) shows a steady increase. Z_{AAKL} (0.126 - 0.734) is slightly weaker but still effective. χ^2_{MMckay} test has weak power for small samples, improving over time (0.091 - 0.714). Not the best for detecting differences in small n. χ^2_{PAN} test similar trend as χ^2_{MMckay} with weak early power but improvements for larger samples. Bootstrap Methods ($T^*_{n(i)}$ and NPB) have reliable performance, especially for moderate and large n. NPB is particularly strong (0.273 - 0.762). $T^*_{n(i)}$also performs well but is slightly weaker.

It is observed that modified versions (Me) generally outperform their classical counterparts.

(i) t_{cal} (0.743) vs. t_{Me} (0.652) – t_{cal} is better.
(ii) χ^2_{Mck} (0.620) vs. $\chi^2_{Mk\_Me}$ (0.816) – The modified version is much stronger.
(iii) Z_{CP} (0.716) vs. Z_{CP_Me} (0.869) – Large power improvement.
(iv) Z_{SK} (0.782) vs. Z_{SK_Me} (0.961) – One of the best performing tests.

Best overall tests:

(i) Z_{SK_Me} (0.947 - 0.961) – Highest power across all sample sizes.
(ii) $\chi^2_{ADJ\_Me}$ (0.513 - 1.000) – Reaches maximum power for large n.
(iii) Z_{CP_Me} (0.323 - 0.869) – Strong growth and efficiency.

Worst overall tests:

(i) Z_{Mil} (0.150 - 0.527) – Consistently weak power.
(ii) χ^2_{MMckay} (0.091 - 0.714) – Slowest improvement over increasing n.

Best tests for small samples (n≤30):

(i) Z_{SK}, $\chi^2_{ADJ\_Me}$, . t_{Me} – High power even for small n.
(ii) Avoid Z_{Mil} and χ^2_{MMckay}– Low power for small samples.

Best tests for large samples (n≥100):

(i) Z_{SK_Me}, $\chi^2_{ADJ\_Me}$, Z_{CP_Me} – Top performers.
(ii) Bootstrap methods ($T^*_{n(i)}$and NPB) also perform well.

Tests with poor power performance:

(i) Z_{Mil} and χ^2_{Mck} – Poor power in detecting deviations
(ii) χ^2_{MMckay} and χ^2_{PAN}– Weak in small n.

It is observed that

(i) $Z_{SK_Me}$and $\chi^2_{ADJ_Me}$are the strongest tests overall.
(ii) Bootstrap methods offer stable power growth and should be considered for general use.
(iii) χ^2_{MMckay} and Z_{Mil} perform poorly, particularly for small n.

Table 3.7: Size of the CV test when data generated from the lognormal (2,1) distribution and skewness 6.19

Test Statistic	n=10	n=20	n=30	n=50	n=100	n=200	n=300
t_{cal}	0.000	0.012	0.023	0.027	0.034	0.039	0.044
χ^2_{Mck}	0.103	0.072	0.065	0.061	0.044	0.032	0.035
Z_{Mil}	0.001	0.011	0.017	0.019	0.021	0.022	0.025
Z_{SK}	0.098	0.111	0.148	0.176	0.187	0.196	0.198
Z_{CP}	0.008	0.034	0.042	0.048	0.056	0.053	0.056
χ^2_{GKAA}	0.034	0.052	0.059	0.053	0.045	0.048	0.049
Z_{AAKL}	0.020	0.025	0.029	0.030	0.039	0.045	0.047
χ^2_{ADJ}	0.010	0.021	0.026	0.033	0.046	0.049	0.050
Z_{AAKL_AUG}	0.024	0.028	0.032	0.035	0.043	0.048	0.049
χ^2_{MMckay}	0.112	0.064	0.058	0.054	0.051	0.046	0.047
χ^2_{PAN}	0.019	0.027	0.046	0.053	0.051	0.059	0.053
t_{Me}	0.013	0.022	0.029	0.030	0.038	0.045	0.047
$\chi^2_{Mk\_Me}$	0.030	0.041	0.046	0.047	0.052	0.056	0.053
Z_{Mi_Me}	0.005	0.014	0.025	0.033	0.035	0.042	0.045
Z_{SK_Me}	0.095	0.087	0.076	0.065	0.067	0.070	0.065
Z_{CP_Me}	0.014	0.039	0.043	0.054	0.057	0.060	0.056
$\chi^2_{GKAA\_Me}$	0.031	0.037	0.039	0.043	0.054	0.056	0.054
Z_{AAKL_Me}	0.043	0.045	0.048	0.053	0.052	0.051	0.049
$\chi^2_{ADJ\_Me}$	0.067	0.073	0.067	0.054	0.058	0.057	0.054
$Z_{AAKL\_AUG\_Me}$	0.048	0.047	0.051	0.052	0.054	0.055	0.053
$\chi^2_{MMK\_Me}$	0.173	0.127	0.117	0.102	0.087	0.075	0.071
$\chi^2_{PAN\_Me}$	0.023	0.027	0.043	0.048	0.054	0.052	0.051
$T^*_{n(i)}$	0.017	0.025	0.032	0.038	0.043	0.047	0.049
NPB	0.021	0.028	0.037	0.039	0.043	0.035	0.038

Table 3.7 presents the empirical size (type I error rate) of various CV tests when applied to data generated from a lognormal(2,1) distribution, which has a high skewness of 6.19.

General conclusion can be drawn from this table results

(i) t_{cal} is too conservative and should not be used for small sample sizes with skewed data.
(ii) Z_{SK} test is too liberal and inflates type I errors.
(iii) Bootstrap-based methods provide stable type I error control across sample sizes.
(iv) Adjusted versions **of** tests (e.g., $\chi^2_{Mk\_Me}$, χ^2_{MMckay}) improve performance for larger n.
(v) Z_{CP}, χ^2_{GKAA}, and Z_{AAKL}tests appear to be reasonable choice**s** as they control the type I error effectively.

Table 3.8: Power of the CV test when data generated from the lognormal (2,1) distribution and skewness 6.18

Test Statistic	n=10	n=20	n=30	n=50	n=100	n=200	n==300
t_{cal}	0.123	0.212	0.364	0.423	0.501	0.623	0.634
χ^2_{Mck}	0.096	0.245	0.297	0.345	0.421	0.549	0.554
Z_{Mil}	0.130	0.151	0.191	0.231	0.279	0.354	0.367
Z_{SK}	0.512	0.603	0.642	0.652	0.687	0.698	0.721
Z_{CP}	0.174	0.214	0.343	0.437	0.460	0.511	0.518
χ^2_{GKAA}	0.139	0.199	0.215	0.351	0.423	0.531	0.337
Z_{AAKL}	0.101	0.187	0.245	0.311	0.432	0.545	0.553
χ^2_{ADJ}	0.134	0.196	0.256	0.323	0.425	0.514	0.527
Z_{AAKL_AUG}	0.121	0.192	0.252	0.323	0.445	0.567	0.576
χ^2_{MMckay}	0.123	0.149	0.242	0.319	0.423	0.554	0.563
χ^2_{PAN}	0.095	0.143	0.221	0.323	0.405	0.532	0.545
t_{Me}	0.243	0.398	0.406	0.512	0.543	0.587	0.593
$\chi^2_{Mk\_Me}$	0.165	0.265	0.332	0.443	0.587	0.608	0.612
Z_{Mi_Me}	0.161	0.237	0.365	0.587	0.638	0.762	0.774
Z_{SK_Me}	0.754	0.789	0.797	0.806	0.849	0.887	0.892
Z_{CP_Me}	0.476	0.490	0.565	0.643	0.693	0.765	0.778
$\chi^2_{GKAA\_Me}$	0.428	0.469	0.587	0.621	0.687	0.765	0.775
Z_{AAKL_Me}	0.467	0.493	0.583	0.607	0.685	0.698	0.712
$\chi^2_{ADJ\_Me}$	0.476	0.565	0.676	0.734	0.865	0.879	0.890
$Z_{AAKL\_AUG\_Me}$	0.478	0.504	0.598	0.623	0.695	0.767	0.780
$\chi^2_{MMK\_Me}$	0.367	0.487	0.557	0.676	0.754	0.889	0.894
$\chi^2_{PAN\_Me}$	0.245	0.276	0.386	0.486	0.575	0.954	0.967
$T^*_{n(i)}$	0.204	0.256	0.387	0.476	0.586	0.643	0.654
NPB	0.187	0.265	0.376	0.465	0.598	0.698	0.734

The above table evaluates the power of different coefficient of variation (CV) tests for different sample sizes **(**n = 10, 20, 30, 50, 100, 200 and 300**)** when data is generated from a highly skewed lognormal(2,1) distribution.

From the Table 3.8, it is observed that

(i) Sharma and Krishna and its "Me" version have the highest power, making them the best choices for detecting changes in the CV under skewed data.
(ii) Bootstrap methods perform well but are not the strongest choices.
(iii) Adjusted versions ("Me") of many tests significantly improve power, suggesting that modifications enhance performance.
(iv) t_{cal} is not the best choice, as other tests consistently outperform it in power.
(v) Z_{Mil} and χ^2_{MMckay} tests perform poorly in terms of power, making them less reliable.

4. Conclusion

This study provides a comprehensive evaluation of the size and power characteristics of various CV tests under different distributional assumptions and sample sizes. Using extensive simulation experiments, we assessed the performance of several CV tests across normal, gamma, chi-square, and log-normal distributions. The results reveal key performance patterns that inform the appropriate use of these tests. Among the evaluated methods, the proposed Sharma and Krishna_Me, Curto and Pinto_Me, and ADJ_Me tests exhibit robust performance across a wide range of conditions, showing strong power while effectively maintaining the nominal significance level 0.05. These tests are particularly suitable for data from non-normal distributions, such as gamma and chi-square, where classical tests often fail to provide reliable results. The parametric bootstrap test also performs well, especially in skewed distributions like log-normal, by stabilizing type I error rates. Conversely, the classical *t*-test remains overly conservative, particularly under non-normality and tests such as the Miller and Vangel tests exhibit weak power, making them less suitable for detecting differences in CV in many practical situations.

Overall, the findings emphasize the importance of selecting CV tests based on both the distributional properties of the data and the sample size. The modified tests evaluated in this study offer improved reliability over traditional approaches, enhancing the accuracy and robustness of statistical inference in applied research contexts.

Despite its comprehensive simulation framework, this study has several limitations that should be acknowledged. First, the evaluation is restricted to a selected set of distributions—normal, gamma, chi-square, and log-normal—which, while commonly encountered in practice, do not cover all possible distributional forms, such as heavy-tailed or multimodal distributions. Second,

the simulation design focuses on specific parameter settings and sample size ranges; thus, the findings may not generalize to extremely small samples or to scenarios involving high-dimensional data. Third, the study considers only hypothesis tests for a single coefficient of variation and does not address extensions to multiple-sample or dependent data settings. Future research could address these limitations by exploring a broader class of distributions, alternative dependence structures, and more complex testing frameworks.

References

Abu-Shawiesh, M.O.A, Akyüz, H.E. and Kibria, B.M.G (2019), Performance of some confidence intervals for estimating the population coefficient of variation under both symmetric and skewed distributions, *Statistics, Optimization and Information Computing*, 7, 277-290.

Almonte, C. and Kibria, B.M.G. (2009). On some classical, bootstrap and transformation confidence intervals for estimating the mean of an asymmetrical population. *Model Assisted Statistics and Applications*, 4(2), 91-104.

Amiri, S. and Zwanzig, S. (2010). An improvement of the nonparametric bootstrap test for the comparison of the coefficient of variations. *Communication in Statistics-Simulation and Computation*, 39(9), 1726-1734.

Baklizi, A. and Kibria, B.M.G. (2009). One and two sample confidence intervals for estimating the mean of skewed populations: An empirical comparative study. *Journal of Applied Statistics* 36, 601-609.

Banik, S. and Kibria, B.M.G. (2009). On some discrete distributions and their applications with real life data. *Journal of Modern Applied Statistical Methods*, 8(2), 423-447.

Banik, S. and Kibria, B.M.G. (2011), Estimating the population coefficient of variation by confidence intervals. *Communications in Statistics-Simulation and Computation*, 40(8), 1236-1261.

Banik, S., Kibria, B.M.G. and Sharma, D. (2012), Testing the population coefficient of variation, *Journal of Modern Applied Statistical Methods*, vol.11, no. 2, 325-335.

Curto, J.D. and Pinto, J.C. (2009), The coefficient of variation asymptotic distribution in the case of non-iid random variables. *Journal of Applied Statistics*, 36, 21-32.

Gulhar, M., Kibria, B.M.G., Albatineh, A.N. and Ahmed N.U. (2012), A comparison of some confidence intervals for estimating the population coefficient of variation: A simulation study, *SORT*, 36, 46–68.

Hendricks, W.A. and Robey, W.K. (1936), The sampling distribution of the coefficient of variation. *Annals of Mathematical Statistics*, 1, 129-132.

Iglewicz, B. (1967), Some properties of the coefficient of variation. Unpublished PhD thesis, Virginia Polytechnic Institute, USA.

Koopmans, L.H., Owen, D.B., and Rosenblatt, J.I. (1964), Confidence intervals for the coefficient of variation for the normal and lognormal distributions. *Biometrika*, 51, 25-32.

McKay, A.T.(1932), Distribution of the coefficient of variation and the extended t distribution, *J. R. Stat. Soc*. 95, pp.695-698.

Miller, G.E.(1991), Asymptotic test statistics for coefficients of variation, *Comm. Statist. Theory and Methods*, 20, 3351–3363.

Panichkitkosolkul, W. (2009), Improved confidence intervals for a coefficient of variation of a normal distribution, *Thailand Statist. 7,* 193–199.

Sharma, K.K. and Krishna, H. (1994), Asymptotic sampling distribution of inverse coefficient-of-variation and its applications, *IEEE Trans. Reliab*. 43, 630–633.

Shi, W. and Kibria, B.M.G. (2007), On some confidence intervals for estimating the mean of a skewed population. *Int J Math Educ Sci Technol*, 38, 412–421.

Vangel, M.G.(1996), Confidence intervals for a normal coefficient of variation, *Amer. Statist*. vol.50, 21–26.