



## Power Comparison of Empirical Likelihood Ratio Tests : Small Sample Properties through Monte Carlo Studies

Tanizaki, Hisashi

---

**(Citation)**

Kobe University Economic Review, 50:13-25

**(Issue Date)**

2004

**(Resource Type)**

departmental bulletin paper

**(Version)**

Version of Record

**(JaLCDOI)**

<https://doi.org/10.24546/00399708>

**(URL)**

<https://hdl.handle.net/20.500.14094/00399708>



# POWER COMPARISON OF EMPIRICAL LIKELIHOOD RATIO TESTS: SMALL SAMPLE PROPERTIES THROUGH MONTE CARLO STUDIES\*

By HISASHI TANIZAKI

There are various kinds of nonparametric tests. In this paper, we consider testing population mean, using the empirical likelihood ratio test. The empirical likelihood ratio test is useful in a large sample, but it has size distortion in a small sample. For size correction, various corrections have been considered. Here, we utilize the Bartlett correction and the bootstrap method. The purpose of this paper is to compare the  $t$  test and the empirical likelihood ratio tests with respect to the sample power as well as the empirical size through Monte Carlo experiments.

## 1. Introduction

In the case of testing population mean, we test the null hypothesis using the  $t$  test, assuming that the population is normal. However, conventionally it is not known whether a population is normal. In a small sample, when a population is not normal, we cannot obtain proper results if we apply the  $t$  test. In a large sample, when both mean and variance are finite, the distribution of sample mean is approximated as a normal distribution from the central limit theorem, where a population distribution is not assumed.

In this paper, we consider testing population mean without any assumption on the underlying distribution. There are numerous papers on the distribution-free test. Here, we introduce the empirical-likelihood-based test and compare it with the  $t$  test. In the past research, Thomas and Grunkemeier (1975) made an attempt to construct confidence intervals using the empirical likelihood ratio. Owen (1988) proposed various test statistics using the empirical likelihood in univariate cases, which are based on Thomas and Grunkemeier (1975). Moreover, Owen (1990) obtained the confidence intervals in multivariate cases and Owen (1991) considered the testing procedure in linear regression models, using the empirical likelihood. As for the other studies, see Baggerly (1998), Chen and Qin (1993), DiCiccio, Hall and Romano (1989), Hall (1990), Kitamura (1997, 2001), Lazar and Mykland (1998), Qin (1993) and Qin and Lawless (1994). Owen (2001) is one of the best references in this field.

In a lot of literature, e.g., Chen (1996), Chen and Hall (1993), DiCiccio, Hall and Romano (1991), DiCiccio and Romano (1989) and Hall (1990), the empirical likelihood is Bartlett correctable. Also, see Jing and Wood (1996), where the empirical likelihood is not Bartlett correctable in a special case. In this paper, we show that in a small sample the Bartlett-corrected empirical likelihood ratio test is improved to some extent but it still has size

---

\* This research is partially supported by Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research (C)(2), 14530033, 2002-2005, and Grants-in-Aid for the 21st Century COE Program.

distortion. As shown in Namba (2004), the bootstrap empirical likelihood ratio test is preferred. In this paper, the  $t$  test and the empirical-likelihood-based test are compared with respect to the sample powers and the empirical size through Monte Carlo studies. The empirical-likelihood-based test is a large sample test and accordingly it has size distortion. To improve the size distortion, we also consider using the Bartlett correction and the bootstrap method.

## 2. Empirical Likelihood ratio Test

In this section, we introduce the testing procedure on population mean, using the empirical likelihood ratio test. For example, see Owen (2001) for the empirical likelihood ratio test.

Let  $X_1, X_2, \dots, X_n$  be mutually independently distributed random variables and  $x_1, x_2, \dots, x_n$  be the observed values of  $X_1, X_2, \dots, X_n$ . For now,  $x_1, x_2, \dots, x_n$  are assumed to be the distinct values in  $\{X_1, X_2, \dots, X_n\}$ . We consider the discrete approximation of the distribution function of  $X$ , which is approximated as  $\text{Prob}(X_i=x_i)=p_i$ , where  $0 < p_i < 1$  and  $\sum_{i=1}^n p_i=1$  have to be satisfied. The joint distribution of  $X_1, X_2, \dots, X_n$ , i.e., the likelihood function (especially, called the nonparametric likelihood), is given by:

$$L(p) = \prod_{i=1}^n \text{Prob}(X_i=x_i) = \prod_{i=1}^n p_i,$$

where  $p=(p_1, p_2, \dots, p_n)$ . Let  $\hat{p}_i$  be the estimate of  $p_i$ , which is given by  $\hat{p}_i = 1/n$  because  $x_1, x_2, \dots, x_n$  are regarded as the realizations randomly generated from the same distribution. Therefore, the empirical likelihood is represented as:

$$L(\hat{p}) = \prod_{i=1}^n \hat{p}_i = \prod_{i=1}^n \frac{1}{n},$$

where  $\hat{p}=(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_n)$ . The empirical likelihood ratio  $R(p)$  is defined as a ratio of the nonparametric likelihood and the empirical likelihood, which is given by:

$$R(p) = \frac{L(p)}{L(\hat{p})} = \prod_{i=1}^n np_i. \quad (1)$$

In this paper, we consider testing the population mean. The expectation of  $X$  is given by:

$$E(X) = \sum_{i=1}^n x_i \text{Prob}(X_i=x_i) = \sum_{i=1}^n x_i p_i = \mu,$$

which is rewritten as:

$$\sum_{i=1}^n (x_i - \mu) p_i = 0. \quad (2)$$

Then, given  $(x_1, x_2, \dots, x_n)$  and  $\mu$ , we want to obtain the  $p_1, p_2, \dots, p_n$  which maximize (1)

with the two constraints, i.e.,  $\sum_{i=1}^n p_i = 1$  and (2). Substituting the solutions of  $p_1, p_2, \dots, p_n$  into (1), the maximum value of the empirical likelihood ratio is obtained as a function of  $\mu$ . That is, we solve the Lagrangian:

$$G = \sum_{i=1}^n \log(np_i) - n\lambda' \left( \sum_{i=1}^n (x_i - \mu)p_i \right) + \gamma \left( \sum_{i=1}^n p_i - 1 \right),$$

with respect to  $p_1, p_2, \dots, p_n, \lambda$  and  $\gamma$ , where  $\lambda$  and  $\gamma$  are denoted by the Lagrangian multipliers. In the second term of the Lagrangian,  $n$  is multiplied for simplicity of calculation. Since the logarithm of the empirical likelihood ratio,  $\log R(p) = \sum_{i=1}^n \log(n p_i)$ , is concave for all  $0 < p_i < 1, i=1, 2, \dots, n$ , we have a global maximum of  $G$ .

However, we have the problem that all the solutions of  $p_1, p_2, \dots, p_n, \lambda$  and  $\gamma$  are not obtained in a closed form. Solving the Lagrangian,  $p_i$  is given by:

$$p_i = \frac{1}{n} \frac{1}{1 + (x_i - \mu)'\lambda}, \quad (3)$$

which depends on  $\lambda$ . In order to obtain the solution of  $\lambda$ , multiplying  $(x_i - \mu)$  in both sides of (3) and summing up with respect to  $i$ , we have the following expression:

$$\sum_{i=1}^n \frac{x_i - \mu}{1 + (x_i - \mu)'\lambda} = 0. \quad (4)$$

Since (4) is an implicit function of  $\lambda$ , we may solve (4) with respect to  $\lambda$  by the iterative procedure such as the Newton-Raphson optimization method or a simple grid search. Thus, given  $\mu$ , we can obtain the  $\lambda$  which satisfies (4). As a computational remark, the condition  $0 < p_i < 1$  is required. This condition is equivalent to:

$$1 + (x_i - \mu)'\lambda > \frac{1}{n}, \quad (5)$$

which comes from (3). Moreover, (5) is rewritten as:

$$-\frac{1 - 1/n}{x_{\max} - \mu} < \lambda < -\frac{1 - 1/n}{x_{\min} - \mu}, \quad (6)$$

where  $x_{\min}$  and  $x_{\max}$  represent minimum and maximum values of  $x_1, x_2, \dots, x_n$ , respectively. Therefore, we have to perform the above iterative procedure with the condition (6). Substituting the solution of  $\lambda$  into (3), we can obtain  $(p_1, p_2, \dots, p_n)$ , given  $(x_1, x_2, \dots, x_n)$  and  $\mu$ . In other words, substituting (3) into (1), the logarithm of the empirical likelihood ratio,  $\log R(p) = \sum_{i=1}^n \log(n p_i)$ , with the constraints  $\sum_{i=1}^n p_i = 1$  and (2) is rewritten as follows:

$$-\sum_{i=1}^n \log(1 + (x_i - \mu)'\lambda) = \log \tilde{R}(x; \mu), \quad (7)$$

where  $x = (x_1, x_2, \dots, x_n)$  and  $\lambda$  has to satisfy (4).  $\tilde{R}(x; \mu)$  corresponds to the maximum value of  $R(p)$  given  $x_1, x_2, \dots, x_n, \mu$ .

We have assumed that  $x_1, x_2, \dots, x_n$  are the distinct values in  $\{X_1, X_2, \dots, X_n\}$ . Therefore,  $p_i$  is interpreted as the probability where  $X$  takes  $x_i$ . Even when some of  $x_1, x_2, \dots, x_n$  take the same value, the above discussion still holds without any modification. However, the interpretation of  $p_i$  should be changed. Suppose that  $x_i$  takes the same value as  $x_j$  for  $i \neq j$ , i.e.,  $x_i=x_j=x_*$ , but it is different from the others.  $\text{Prob}(X=x_*) = \text{Prob}(X=x_i) + \text{Prob}(X=x_j) = p_i + p_j = p_*$  represents the probability where  $X$  takes  $x_i=x_j=x_*$ , where  $p_i$  should be equal to  $p_j$ . In this case, we need to interpret  $p_i$  as a specific weight, rather than the probability.

Now, we consider the null hypothesis  $H_0: \mu = \mu_0$ . Replace the actual data  $x_i$  in (7) by the corresponding random variable  $X_i$ . Then, as  $n$  goes to infinity, under the null hypothesis  $H_0$  we have the following asymptotic property:

$$-2 \log \tilde{R}(X; \mu_0) = 2 \sum_{i=1}^n \log(1 + (X_i - \mu_0)' \lambda) \sim \chi^2(k), \quad (8)$$

which is shown in Owen (2001), where  $X = (X_1, X_2, \dots, X_n)$ . Let  $\chi^2_\alpha(k)$  be  $100 \times \alpha$  percent point from the  $\chi^2$  distribution with  $k$  degrees of freedom. Remember that  $\mu$  is a  $k \times 1$  vector. Thus, we reject the null hypothesis  $H_0: \mu = \mu_0$ , when  $-2 \log \tilde{R}(X; \mu_0) = 2 \sum_{i=1}^n \log(1 + (X_i - \mu_0)' \lambda) > \chi^2_\alpha(k)$ .

Note as follows. Under the null hypothesis  $H_0: \mu = \mu_0$ , we sometimes have the case where the  $\lambda$  which satisfies (4), (5) and  $\mu = \mu_0$  does not exist. The reason why we have this case is because the null hypothesis  $H_0: \mu = \mu_0$  does not hold. Therefore, it is plausible to consider that the empirical likelihood ratio  $-2 \log \tilde{R}(X; \mu_0)$  takes an extremely large value.

**Bartlett-Corrected Empirical Likelihood Ratio Test:** As mentioned above, the empirical likelihood ratio test statistic  $-2 \log \tilde{R}(X; \mu_0)$  is asymptotically distributed as a  $\chi^2(k)$  random variable. However, in a small sample, it is not necessarily chi-squared and the approximation might be quite poor. Therefore, we consider applying size correction to (8). In the past, various size correction methods have been proposed. In this paper, we adopt the Bartlett correction, which is discussed in Chen (1996), Chen and Hall (1993), DiCiccio, Hall and Romano (1991), DiCiccio and Romano (1989), Hall (1990), Jing and Wood (1996) and Owen (2001). By the Bartlett correction, asymptotically we have the following:

$$-2 \log \tilde{R}(X; \mu_0) = 2 \sum_{i=1}^n \log(1 + (X_i - \mu_0)' \lambda) - (1 + \frac{a}{n}) \chi^2(k),$$

or equivalently,

$$-(1 + \frac{a}{n})^{-1} 2 \log \tilde{R}(X; \mu_0) = (1 + \frac{a}{n})^{-1} 2 \sum_{i=1}^n \log(1 + (X_i - \mu_0)' \lambda) \sim \chi^2(k),$$

where  $\lambda$  has to satisfy (4), (5) and  $\mu = \mu_0$ , and  $a$  is denoted by  $a = \frac{1}{2} (m_4/m_2^2) - \frac{1}{3} (m_3^2/m_2^3)$ .  $m_j$  denotes the  $j$ th moment of  $X$  about the mean  $E(X)$ , i.e.,  $m_j = E((X - E(X))^j)$ . Since  $m_j$  is unknown, replacing  $m_j$  by its estimate  $\hat{m}_j = (1/n) \sum_{i=1}^n (x_i - \bar{x})^j$ , we perform the testing procedure as follows:

$$-(1 + \frac{\hat{a}}{n})^{-1} 2 \log \tilde{R}(X; \mu_0) = (1 + \frac{\hat{a}}{n})^{-1} 2 \sum_{i=1}^n \log(1 + (X_i - \mu_0)' \lambda) - \chi^2(k), \quad (9)$$

where  $\lambda$  has to satisfy (4), (5) and  $\mu = \mu_0$ , and  $\hat{a}$  indicates the estimate of  $a$ , which is shown as:  $\hat{a} = \frac{1}{2} (\hat{m}_4 / \hat{m}_2^2) - \frac{1}{3} (\hat{m}_3^2 / \hat{m}_2^3)$ . Thus, we reject the null hypothesis  $H_0: \mu = \mu_0$ , when  $-(1 + \hat{a}/n)^{-1} 2 \log \tilde{R}(x; \mu_0) = (1 + \hat{a}/n)^{-1} 2 \sum_{i=1}^n \log(1 + (x_i - \mu_0)' \lambda) > \chi_{\alpha}^2(k)$ .

**Bootstrap Empirical Likelihood Ratio Test:** As another size correction method, we consider deriving an empirical distribution of  $-2 \log \tilde{R}(X; \mu)$  by the bootstrap method.

Let  $x_1^*, x_2^*, \dots, x_n^*$  be the bootstrap sample from  $x_1, x_2, \dots, x_n$ . That is, one of the original data  $x_1, x_2, \dots, x_n$  is randomly chosen for  $x_i^*$ . We repeat this procedure for all  $i=1, 2, \dots, n$  with replacement and we have the bootstrap sample  $x_1^*, x_2^*, \dots, x_n^*$ . We solve (1) subject to the constraints  $0 < p_i < 1$ ,  $\sum_{i=1}^n p_i = 1$  and  $\sum_{i=1}^n (x_i^* - \bar{x}) p_i = 0$ , where  $\bar{x} = (1/n) \sum_{i=1}^n x_i$ . As a result, in (2) – (7),  $x_i$  and  $\mu$  are replaced by  $x_i^*$  and  $\bar{x}$ , respectively. That is, instead of (7), we obtain the following test statistic:

$$-2 \log \tilde{R}(x^*; \bar{x}), \quad (10)$$

where  $x^* = (x_1^*, x_2^*, \dots, x_n^*)$ . Whenever the bootstrap sample  $\{x_1^*, x_2^*, \dots, x_n^*\}$  is resampled from  $\{x_1, x_2, \dots, x_n\}$  with replacement, we have a different value of (10). Resampling the bootstrap sample  $M$  times, we have  $M$  empirical log-likelihood ratios, where  $M = 10^4$  is taken in the Monte Carlo studies in the next section. This procedure is equivalent to deriving the empirical distribution of  $-2 \log \tilde{R}(X; \bar{x})$ , where the sample mean  $\bar{x}$  is given. Sorting the  $M$  values by size, we can obtain the percent points from the  $M$  empirical log-likelihood ratios. The percent points based on the  $M$  values are compared with the empirical log-likelihood ratio based on the original data. Thus, we can test the null hypothesis  $H_0: \mu = \mu_0$ . See, for example, Hall (1987), Namba (2004) and Owen (2001) for the bootstrap empirical likelihood ratio test.

Sometimes we have the case where  $x_i^* > \bar{x}$  or  $x_i^* < \bar{x}$  for all  $i=1, 2, \dots, n$ . In this case, we consider that (10) lies on the side of the distribution and accordingly takes an extremely large value.

### 3. Monte Carlo Experiments

In Section 2, we have introduced the empirical likelihood ratio tests based on (8), (9) and (10), which are applied to testing the population mean. In this section, the  $t$  test, the conventional empirical likelihood ratio test (8), the Bartlett-corrected empirical likelihood ratio test (9) and the bootstrap empirical likelihood ratio test (10) are compared with respect to empirical size and sample power through Monte Carlo studies.

The setup of the Monte Carlo experiments is as follows:

(i) For the underlying distribution of univariate random variable  $X_i$ , we take the following

seven distributions.

**Standard Normal Distribution (N):**

$$f(x) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} x^2\right), \quad -\infty < x < \infty.$$

**Uniform distribution (U):**

$$f(x) = \frac{1}{2}, \quad -3 < x < 3.$$

**$\chi^2(1)$  Distribution (X):**

$$x = \frac{y-1}{2}, \text{ where } f(y) = \frac{1}{2\pi} y^{-1/2} \exp\left(-\frac{1}{2} y\right), \quad 0 < y < \infty.$$

**$t(3)$  Distribution (T):**

$$x = \frac{y}{3}, \text{ where } f(y) = \frac{2}{\pi} (1+y^2/3)^{-2}, \quad -\infty < y < \infty.$$

**Double Exponential Distribution (D):**

$$x = \frac{y}{2}, \text{ where } f(y) = \frac{1}{2} \exp(-|y|), \quad -\infty < y < \infty.$$

**Logistic Distribution (L):**

$$x = \frac{y}{\pi}, \text{ where } f(y) = \frac{\exp(y)}{(1+\exp(y))^2}, \quad -\infty < y < \infty.$$

**Gumbel Distribution (G):**

$$x = \frac{y}{\pi}, \text{ where } f(y) = \exp(-(y-\alpha)-e^{-(y-\alpha)}) \text{ and } \alpha \approx -0.5772, \quad -\infty < y < \infty.$$

The mean and variance of  $X_i$  are normalized to be zero and one for  $i=1,2,\dots,n$ .

- (ii) The sample size is given by  $n = 20, 50, 100$ .
- (iii) The null hypothesis  $H_0: \mu=\mu_0$  and the alternative one  $H_1: \mu=\mu_1$  are taken for  $\mu_0=0$  and  $\mu_1=.0, .1, .2, .4$ .
- (iv) The significance level  $\alpha$  is given by  $\alpha=.10, .05$ .
- (v) For each significance level, perform  $10^4$  simulation runs and obtain the number of rejections of the null hypothesis  $H_0: \mu=\mu_0$ . Dividing the number of the rejections by  $10^4$ , the sample power is computed.

Under the setup above, we perform Monte Carlo experiments. The obtained results are expected as follows:

Since the normal distribution is assumed in (N), the  $t$  test indicates the same empirical size as the significance level and it gives us the most powerful test.

In the case of  $\mu_1=0$ , the power corresponds to the size, i.e., the significance level  $\alpha$ . However, the empirical likelihood ratio test shown in (8) indicates a large sample test. Therefore, in a small sample, it might be expected that the empirical likelihood ratio test (8) has size distortion. The Bartlett-corrected empirical likelihood ratio test (9) and the bootstrap empirical likelihood ratio test (10) should be improved with respect to the empirical size in a small sample.

The conventional empirical likelihood ratio test, the Bartlett-corrected empirical likelihood ratio test and the bootstrap empirical likelihood ratio test should be equivalent to the  $t$  test as  $n$  is large, because the three test statistics have the same distribution, i.e., chi-square distribution, in a large sample.

The results are obtained in Table 1, where the three kinds of empirical likelihood ratio tests are shown with the  $t$  test. (8) represents the conventional empirical likelihood ratio test which simply utilizes the chi-square percent points with one degree of freedom. (9) indicates the Bartlett-corrected empirical likelihood ratio test, where the chi-square percent point with one degree of freedom is regarded as the percent point for the test statistic. (10) shows the bootstrap empirical likelihood ratio test, where the distribution of the test statistic is constructed by the bootstrap sample. Let  $\hat{\delta}$  be a value in the table. The standard error of  $\hat{\delta}$  is given by  $\hat{\delta}(1-\hat{\delta})/10^4$ , and accordingly it is at most  $.5(1-.5)/10^4 = .005$ . Note that  $10^4$  in the square root indicates the number of simulation runs.

The results are summarized as follows.

**Empirical Size:** First, consider the case of  $\mu_1=0$  in Table 1, which represents the empirical size. Let  $\delta$  and  $\hat{\delta}$  be the true size and the empirical size, respectively. Under the null hypothesis  $H_0: \delta = \alpha$ , asymptotically we have  $(\hat{\delta} - \alpha)/\sqrt{\alpha(1-\alpha)/10^4} \sim N(0,1)$ , which comes from the central limit theorem. In the case of  $\mu_1=0$  in Table 1, \*\* and \* show that the null hypothesis  $H_0: \delta = \alpha$  is rejected at significance levels 1 % and 5 % by the both-sided test. That is, in the case of  $\mu_1=0$ , the values without \*\* and \* are preferred, because they show correct sizes. As for (N), the  $t$  test indicates the correct sizes for all  $\alpha = .10, .05$  and  $n = 20, 50, 100$ . These results are plausible, because the  $t$  test gives us the uniformly most powerful test under normality assumption. For (U), similarly the empirical size  $\hat{\delta}$  in the  $t$  test is very close to the significance level  $\alpha$ . For (T), (D) and (L), the  $t$  test is not an appropriate test especially for small  $n$ , because the size is statistically different from the significance level. That is, the  $t$  test gives us the correct sizes and the powers for (N), but it yields the over-estimated sizes or the under-estimated sizes for (U), (T), (D) and (L).

When  $n$  is small, for any distribution the conventional empirical likelihood ratio test (8) has size distortion and it over-estimates empirical sizes. For example, comparing the empirical likelihood ratio test (8) and the  $t$  test in the case of (N),  $\mu_1=0$  and  $n=20$ , the empirical likelihood ratio test (8) has larger empirical sizes than the  $t$  test for both  $\alpha = .10$  and  $.05$  (we

have 0.1340 and 0.0808 for the empirical likelihood ratio test (8), and 0.0971 and 0.0503 for the  $t$  test). As  $n$  increases (for example,  $n=100$ ), the empirical likelihood ratio test (8) as well as the  $t$  test indicate the correct sizes.

In a small sample, we perform Bartlett correction to improve the size distortion in the empirical likelihood ratio test. See (8) in Table 1 for the Bartlett-corrected empirical likelihood ratio test. We can observe that the size distortion decreases but we still have the size distortion to some extent. For example, in the case of (N),  $n=20$ ,  $\mu_1=0$  and  $\alpha=0.10$ , (8) is .1340 and (9) is .1216, which indicates that (9) is closer to  $\alpha=0.10$  than (8). (8) is .0808 for  $\alpha=0.05$  and (9) is .0735 for  $\alpha=0.05$ . Therefore, (9) is better than (8), but it is not practical.

The bootstrap empirical likelihood ratio test (10) is also implemented to improve the size distortion. Both (8) and (9) are size-distorted in most of the cases, but (10) shows a correct empirical size in a lot of cases. Thus, the size distortion is improved when (10) is applied. As a result, we can conclude that (10) is better than (8) and (9) for the empirical size criterion and that the empirical size in (10) is very close to the significance level.

When  $n$  is large, the empirical likelihood ratio tests (8) - (10) and  $t$  test show the size which is equal to the significance level. That is, in the case of  $\mu_1=0$  and  $n=100$ , there are a lot of values without the superscript \*. Thus, all the tests perform better in a large sample.

**Sample Power:** Next, we consider the case of  $\mu_1=0$  in Table 1, which indicates the sample power. In the case of  $\mu_1=0.1, 0.2, 0.4$  and the empirical likelihood ratio tests (8) - (10), , , and indicate comparison with the  $t$  test. Let  $\delta_t$  be the value in  $t$  Test of Table 1. We put the superscript when  $(\hat{\delta}-\delta_t)/\delta_t(1-\delta_t)/10^4$  is greater than 1.9600, and the superscript when it is greater than 2.5758. We put the superscript if  $(\hat{\delta}-\delta_t)/\delta_t(1-\delta_t)/10^4$  is less than -1.9600, and the superscript if it is less than -2.5758. Note that in a large sample we have the following:  $(\hat{\delta}-\delta_t)/\delta_t(1-\delta_t)/10^4 \sim N(0,1)$  under the null hypothesis  $H_0: \delta=\delta_t$  and the alternative one  $H_1: \delta \neq \delta_t$ . Therefore, the values with the superscript indicate a more powerful test than the  $t$  test. In addition, the number of the superscript shows the degree of the sample power. Contrarily, the values with the superscript represent a less powerful test than the  $t$  test. Taking an example of the case (N),  $n=20$  and  $\mu_1=0.1$ , the sample power in (8) is given by  $\hat{\delta}=0.1687$  while that in  $t$  Test is  $\delta_t=0.1325$ . We want to test  $H_0: \delta=\delta_t$  against  $H_1: \delta \neq \delta_t$ . In this case, the test statistic is given by  $(0.1687-0.1325)/\sqrt{0.1325(1-0.1325)/10^4}=10.677$ , which is greater than 2.5758. Therefore,  $H_0: \delta=\delta_t$  is rejected at significance level  $\alpha=0.01$ .

In the case of (N), the  $t$  test is the most powerful test. For small  $n$ , both (8) and (9) have size distortion, and they over-estimate the sample powers. Accordingly, it is not meaningful to compare (8), (9) and the  $t$  test with respect to the sample powers. The empirical likelihood ratio test (10) is very close to the  $t$  test in all the cases of (N). Thus, (10) performs better than (8) and (9).

For (U), the empirical sizes in (8) and (9) are over-estimated, while those in (10) and the  $t$  test are sometimes under-estimated but correctly estimated in many cases. In addition, the sample powers in (8) and (9) are larger than those in the  $t$  test. (10) is slightly better than the  $t$

TABLE 1. Empirical Size and Sample Power

n	$\mu_1 \setminus \alpha$	Empirical Likelihood Ratio Test						<i>t</i> Test	
		(8)		(9)		(10)			
		.10	.05	.10	.05	.10	.05	.10	.05
<b>Normal Distribution(N)</b>									
20	.0	0.1340**	0.0808**	0.1216**	0.0735**	0.1021	0.0512	0.0971	0.0503
	.1	0.1687	0.1055	0.1576	0.0976	0.1321	0.0700	0.1325	0.0673
	.2	0.2648	0.1802	0.2480	0.1666	0.2142	0.1226	0.2180	0.1337
	.4	0.5805	0.4624	0.5624	0.4399	0.5041	0.3381	0.5306	0.3925
50	.0	0.1147**	0.0617**	0.1095**	0.0583**	0.1007	0.0510	0.0988	0.0473
	.1	0.1963	0.1197	0.1891	0.1157	0.1754	0.0978	0.1791	0.1029
	.2	0.4190	0.3015	0.4081	0.2929	0.3841	0.2551	0.3968	0.2770
	.4	0.8783	0.8035	0.8745	0.7959	0.8558	0.7428	0.8706	0.7872
100	.0	0.1070*	0.0571**	0.1043	0.0558**	0.0976	0.0497	0.0992	0.0503
	.1	0.2707	0.1788	0.2674	0.1750	0.2553	0.1594	0.2597	0.1670
	.2	0.6379	0.5167	0.6325	0.5113	0.6159	0.4792	0.6290	0.5054
	.4	0.9905	0.9778	0.9903	0.9766	0.9892	0.9685	0.9903	0.9765
<b>Uniform Distribution(U)</b>									
20	.0	0.1149**	0.0651**	0.1057	0.0600**	0.0932*	0.0493	0.0923*	0.0469
	.1	0.1464	0.0885	0.1379	0.0815	0.1232	0.0662	0.1237	0.0677
	.2	0.2456	0.1615	0.2338	0.1502	0.2151	0.1196	0.2103	0.1284
	.4	0.5805	0.4553	0.5656	0.4386	0.5360	0.3721	0.5184	0.3767
50	.0	0.1056	0.0550*	0.1028	0.0524	0.0974	0.0469	0.0962	0.0470
	.1	0.1853	0.1101	0.1802	0.1065	0.1728	0.0972	0.1713	0.0966
	.2	0.4147	0.2952	0.4093	0.2893	0.3927	0.2690	0.3926	0.2708
	.4	0.8933	0.8216	0.8910	0.8173	0.8840	0.7917	0.8781	0.7929
100	.0	0.0993	0.0527	0.0977	0.0518	0.0933*	0.0482	0.0943	0.0483
	.1	0.2744	0.1764	0.2717	0.1736	0.2662	0.1612	0.2657	0.1644
	.2	0.6517	0.5283	0.6490	0.5252	0.6398	0.5054	0.6391	0.5117
	.4	0.9935	0.9828	0.9934	0.9822	0.9930	0.9806	0.9924	0.9792
<b><math>\chi^2(1)</math> Distribution(X)</b>									
20	.0	0.1777**	0.1211**	0.1653**	0.1096**	0.1094**	0.0691**	0.1540**	0.1076**
	.1	0.1923	0.1190	0.1737	0.1063	0.0888	0.0390	0.1147	0.0628
	.2	0.3165	0.2160	0.2916	0.1966	0.1526	0.0552	0.1506	0.0643
	.4	0.8184	0.7444	0.8012	0.7208	0.6336	0.3961	0.5771	0.3641
50	.0	0.1451**	0.0861**	0.1363**	0.0822**	0.0994	0.0594**	0.1273**	0.0829**
	.1	0.2195	0.1361	0.2078	0.1268	0.1359	0.0393	0.1415	0.0623
	.2	0.5215	0.4044	0.5051	0.3865	0.4102	0.1925	0.3672	0.2140
	.4	0.9896	0.9785	0.9884	0.9756	0.9621	0.8211	0.9659	0.9108
100	.0	0.1258**	0.0744**	0.1214**	0.0713**	0.0949	0.0474	0.1151**	0.0679**
	.1	0.2978	0.2058	0.2892	0.1971	0.2388	0.0845	0.2198	0.1144
	.2	0.7585	0.6619	0.7498	0.6496	0.7154	0.4799	0.6664	0.4993
	.4	0.9998	0.9998	0.9998	0.9998	0.9953	0.9122	0.9998	0.9992

TABLE 1. Empirical Size and Sample Power &lt;Continued&gt;

n	$\mu_1 \setminus \alpha$	Empirical Likelihood Ratio Test						<i>t</i> Test	
		(8)		(9)		(10)			
		.10	.05	.10	.05	.10	.05	.10	.05
<b>t (3) Distribution(T)</b>									
20	.0	0.1590**	0.1015**	0.1443**	0.0894**	0.0998	0.0474	0.0952	0.0445*
	.1	0.2140	0.1441	0.1926	0.1274	0.1420	0.0774	0.1419	0.0784
	.2	0.3434	0.2584	0.3225	0.2367	0.2577	0.1623	0.2792	0.1803
	.4	0.6732	0.5849	0.6551	0.5600	0.5770	0.4384	0.6355	0.5178
50	.0	0.1261**	0.0736**	0.1163**	0.0668**	0.0916**	0.0446*	0.0922**	0.0435**
	.1	0.2399	0.1619	0.2279	0.1507	0.1946	0.1129	0.2020	0.1180
	.2	0.4917	0.3878	0.4773	0.3686	0.4316	0.3040	0.4679	0.3458
	.4	0.8838	0.8214	0.8737	0.8117	0.8317	0.7354	0.8947	0.8332
100	.0	0.1262**	0.0720**	0.1187**	0.0678**	0.0967	0.0477	0.0976	0.0475
	.1	0.3181	0.2241	0.3074	0.2135	0.2746	0.1726	0.2950	0.1924
	.2	0.6796	0.5779	0.6688	0.5674	0.6364	0.5031	0.6796	0.5764
	.4	0.9715	0.9562	0.9685	0.9553	0.9512	0.8880	0.9814	0.9676
<b>Double Exponential Distribution (D)</b>									
20	.0	0.1503**	0.0937**	0.1367**	0.0831**	0.1001	0.0455*	0.0924*	0.0408**
	.1	0.1919	0.1304	0.1769	0.1193	0.1333	0.0720	0.1336	0.0706
	.2	0.2981	0.2169	0.2789	0.2022	0.2238	0.1295	0.2383	0.1469
	.4	0.5971	0.5005	0.5767	0.4762	0.5014	0.3483	0.5575	0.4325
50	.0	0.1241**	0.0737**	0.1159**	0.0695**	0.1003	0.0501	0.0958	0.0461
	.1	0.2027	0.1301	0.1919	0.1224	0.1718	0.0944	0.1751	0.1018
	.2	0.4322	0.3170	0.4172	0.3048	0.3813	0.2466	0.4116	0.2905
	.4	0.8653	0.7854	0.8568	0.7757	0.8226	0.6775	0.8685	0.7911
100	.0	0.1164**	0.0666**	0.1130**	0.0635**	0.1012	0.0502	0.0994	0.0481
	.1	0.2821	0.1892	0.2761	0.1819	0.2530	0.1519	0.2705	0.1717
	.2	0.6431	0.5239	0.6364	0.5159	0.6107	0.4649	0.6400	0.5185
	.4	0.9875	0.9701	0.9866	0.9692	0.9794	0.9321	0.9884	0.9728
<b>Logistic Distribution (L)</b>									
20	.0	0.1348**	0.0805**	0.1235**	0.0720**	0.0963	0.0465	0.0930*	0.0437**
	.1	0.1755	0.1158	0.1642	0.1060	0.1340	0.0714	0.1292	0.0680
	.2	0.2735	0.1975	0.2590	0.1823	0.2196	0.1265	0.2273	0.1387
	.4	0.5886	0.4787	0.5686	0.4541	0.5023	0.3487	0.5424	0.4074
50	.0	0.1174**	0.0688**	0.1123**	0.0644**	0.1000	0.0508	0.0956	0.0471
	.1	0.1952	0.1207	0.1868	0.1158	0.1690	0.0940	0.1728	0.0985
	.2	0.4239	0.3101	0.4141	0.3001	0.3847	0.2543	0.4042	0.2829
	.4	0.8740	0.7971	0.8675	0.7871	0.8432	0.7131	0.8709	0.7904
100	.0	0.1101**	0.0598**	0.1075*	0.0568**	0.0974	0.0484	0.0988	0.0475
	.1	0.2764	0.1820	0.2723	0.1781	0.2554	0.1541	0.2660	0.1711
	.2	0.6434	0.5239	0.6385	0.5156	0.6172	0.4747	0.6393	0.5136
	.4	0.9887	0.9741	0.9882	0.9732	0.9847	0.9563	0.9900	0.9758

TABLE 1. Empirical Size and Sample Power &lt;Continued&gt;

n	$\mu_1 \setminus \alpha$	Empirical Likelihood Ratio Test						t Test	
		(8)		(9)		(10)		.10	.05
		.10	.05	.10	.05	.10	.05	.10	.05
<b>Gumbel Distribution(G)</b>									
20	.0	0.1378**	0.0842**	0.1266**	0.0773**	0.1017	0.0535	0.1047	0.0562**
	.1	0.1619	0.0935	0.1484	0.0853	0.1066	0.0410	0.1081	0.0464
	.2	0.2641	0.1747	0.2453	0.1607	0.1789	0.0777	0.1791	0.0950
	.4	0.6466	0.5268	0.6253	0.5008	0.5335	0.3094	0.5338	0.3663
50	.0	0.1171**	0.0672**	0.1115**	0.0640**	0.0955	0.0490	0.1004	0.0520
	.1	0.1913	0.1146	0.1825	0.1078	0.1512	0.0648	0.1480	0.0765
	.2	0.4449	0.3273	0.4342	0.3166	0.3922	0.2211	0.3811	0.2461
	.4	0.9338	0.8844	0.9297	0.8783	0.9195	0.8029	0.9108	0.8341
100	.0	0.1107**	0.0588**	0.1078**	0.0565**	0.0943	0.0442**	0.0990	0.0510
	.1	0.2791	0.1847	0.2742	0.1776	0.2507	0.1297	0.2450	0.1446
	.2	0.6810	0.5699	0.6760	0.5638	0.6600	0.5040	0.6434	0.5104
	.4	0.9970	0.9943	0.9966	0.9939	0.9968	0.9816	0.9965	0.9909

test but (9) is not significantly different from the *t* test. It might be concluded that (10) is slightly more powerful than the *t* test in the case where the underlying distribution is given by (U).

In the case of (X), (T), (D), (L) and (G), both (7) and (8) have size distortion, where all the empirical sizes in (8) and (9) are over-estimated. Therefore, it is useless for (X), (T), (D), (L) and (G) to compare (8) and (9) with the *t* test. (10) has no size distortion in a lot of cases, while the empirical sizes in the *t* test are sometimes under-estimated. In most cases, (10) is less than the *t* test with respect to the sample powers. Therefore, the *t* test is more powerful than (10), especially in the case of (T), (D) and (L).

In the case of  $n=100$ , the sample powers take similar values for all  $\mu_1=0.1, 0.2, 0.4$ . Therefore, it is shown from Table 1 that all the tests give us the same sample power in a large sample.

#### 4. Summary

In this paper, we have compared three empirical likelihood ratio tests with the *t* test, where we test the population mean. In the case of testing the population mean, conventionally we assume a normal distribution. Under the normality assumption, the appropriate test statistic has a *t* distribution. Thus, we need the normality assumption for the *t* test. However, the underlying distribution is not known in general. Therefore, the normality assumption is not necessarily plausible. In this paper, we consider testing the population mean without any assumption on the underlying distribution.

In the case of a large sample, the sample mean is approximated to be normal when mean and

variance exist. This result comes from the central limit theorem. Therefore, using the normal distribution, we can test the hypothesis on the population mean. The asymptotic distribution of the  $t$  distribution is given by the standard normal distribution. Thus, when the sample size is large we can apply the  $t$  test to any underlying distribution.

The empirical likelihood ratio test statistic is asymptotically distributed as a chi-square random variable, which is also a large sample test, where the chi-square percent points are compared with the empirical likelihood ratio test statistic. The conventional empirical likelihood ratio test with chi-square percent points yields the size distortion. Therefore, we have considered the empirical likelihood ratio tests based on the Bartlett correction and the bootstrap method.

In this paper, the normal distribution, the uniform distribution, the  $t(3)$  distribution, the double exponential distribution, or the logistic distribution is assumed for the underlying distribution. Using the three empirical likelihood ratio tests shown in (8) - (10) and the  $t$  test, we have obtained the empirical sizes and the sample powers for each distribution through Monte Carlo experiments. When the sample size  $n$  is large, we have obtained the result that all the tests show the similar empirical sizes and sample powers. However, when  $n$  is small, the conventional empirical likelihood ratio test (8) indicates a large size distortion, and accordingly it is not practically useful. Therefore, we have performed the size correction by Bartlett correction, but the empirical likelihood ratio test with Bartlett correction, shown in (9), still has size distortion although it improves to some extent. Next, using the bootstrap method, we have obtained the empirical distribution of the the empirical likelihood ratio test statistic and derived the critical region. Based on the critical region, we have performed the empirical likelihood ratio test (10). The obtained empirical sizes and sample powers are plausible, using (10).

## REFERENCES

Baggerly, K.A. (1998), "Empirical Likelihood as a Goodness-of-Fit Measure," *Biometrika*, 85, pp. 535-547.

Chen, S.X. (1996), "Empirical Likelihood Confidence Intervals for Nonparametric Density Estimation," *Biometrika*, 83, pp. 329-341.

Chen, S.X. and Hall, P. (1993), "Smoothed Empirical Likelihood Confidence Intervals for Quantiles," *The Annals of Statistics*, 21, pp. 1166-1181.

Chen, J. and Qin, J. (1993), "Empirical Likelihood Estimation for Finite Populations and the Effective Usage of Auxiliary Information," *Biometrika*, 80, pp. 107-116.

DiCiccio, T.J., Hall, P. and Romano, J.P. (1989), "Comparison of Parametric and Empirical Likelihood Functions," *Biometrika*, 76, pp. 465-476.

DiCiccio, T.J., Hall, P. and Romano, J.P. (1991), "Empirical Likelihood is Bartlett-Correctable," *The Annals of Statistics*, 19, pp. 1053-1061.

DiCiccio, T.J. and Romano, J.P. (1989), "On Adjustments Based on the Signed of the Empirical Likelihood Ratio Statistic," *Biometrika*, 76, pp. 447-456.

Hall, P. (1987), "On the Bootstrap and Likelihood-Based Confidence Regions," *Biometrika*, 74, pp. 481-493.

Hall, P. (1990), "Pseudo-Likelihood Theory for Empirical Likelihood," *The Annals of Statistics*, 18, pp. 121-140.

Jing, B.Y. and Wood, A.T.A. (1996), "Exponential Empirical Likelihood is Not Bartlett Correctable," *The Annals of Statistics*, 24, pp. 365-369.

Kitamura, Y. (1997), "Empirical Likelihood Methods with Weakly Dependent Processes," *The Annals of Statistics*, 25, pp. 2084-2102.

Kitamura, Y. (2001), "Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions," *Econometrica*, 69, pp. 1661-1672.

Lazar, N. and Mykland, P.A. (1998), "An Evaluation of the Power and Conditionality Properties of Empirical Likelihood," *Biometrika*, 85, pp. 523-534.

Namba, A. (2004), "Simulation Studies on Bootstrap Empirical Likelihood Tests," *Communications in Statistics, Simulation and Computation*, 33, pp. 99-108.

Owen, A.B. (1988), "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75, pp. 237- 249.

Owen, A.B. (1990), "Empirical Likelihood Ratio Confidence Regions," *The Annals of Statistics*, 18, pp. 90-120.

Owen, A.B. (1991), "Empirical Likelihood for Linear Models," *The Annals of Statistics*, 19, pp. 1725-1747.

Owen, A.B. (2001), *Empirical Likelihood* (Monographs on Statistics and Applied Probability 92), Chapman \& Hall/CRC.

Qin, J. (1993), "Empirical Likelihood in Biased Sample Problems," *The Annals of Statistics*, 21, pp. 1182-1196.

Qin, J. and Lawless, J. (1994), "Empirical Likelihood and General Estimating Equations," 22, pp. 300-325.

Thomas, D.R. and Grunkemeier, G.L. (1975), "Confidence Interval Estimation of Survival Probabilities for Censored Data," *Journal of the American Statistical Association*, 70, pp. 865-871.