PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 121, Number 3, July 1994

COMPARISONS OF LOGNORMAL POPULATION MEANS

HANFENG CHEN

(Communicated by Wei-Yin Loh)

ABSTRACT. Comparisons of two lognormal population means are investigated. For large samples, the conventional test for significance of the population means is the ordinary Student t-test with normal critical value. By Chen and Loh's result (Ann. Statist. **20** (1992)), however, the transformed t-test based on log-data is asymptotically more powerful than the ordinary t-test. In this paper, a new power transformation (hence a new transformed t-test) is proposed. The new transformed t-test is proven to be asymptotically more powerful than the one based on log-data. Both small sample and large sample properties of the proposed estimate for the power-transformation parameter are studied. A simulation study shows that the advantages of the new test over the t-test based on log-data are overwhelming and evident for sizes of the two samples as small as 20 and 30, or even 10 and 15. The simulation results also show that the new test has greater asymptotic power than Rao's efficient score test.

1. INTRODUCTION

The lognormal model has been widely used to fit skewed positive data, such as the sizes of organisms and the numbers of species in biology, the rainfalls in meteorology, the sizes of incomes in economics, and so on (see Crow and Shimizu [6, Chapters 9–14]). By a direct definition, a positive random variable X is said to be lognormally distributed if $\log X$ is normally distributed with mean μ and variance σ^2. The lognormal distribution is then denoted by $\Lambda(\mu, \sigma^2)$. This is the so-called two-parameter definition of the lognormal distribution. The distribution of X is thus fully specified by the two parameters μ and σ^2, and this seems to be the simplest natural specification, as Aitchison and Brown [2] commented. By this definition, the mean ν and variance τ^2 of X are given by

$$\nu = \exp(\mu + \sigma^2/2) \quad \text{and} \quad \tau^2 = \exp(2\mu + \sigma^2)\{\exp(\sigma^2) - 1\}. \tag{1}$$

The density $f(x)$ of X takes the form

$$f(x) = (2\pi)^{-1/2}(x\sigma)^{-1} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}, \qquad x > 0. \tag{2}$$

Received by the editors October 9, 1992.

1991 *Mathematics Subject Classification*. Primary 62F03; Secondary 62F05.

Key words and phrases. Asymptotic relative efficiency, Box-Cox transformation, two-sample t-test.

Research partially supported by a Bowling Green State University FRC Basic Grant for 1992 and the Ohio Supercomputer Center.

In the paper, we consider the two-sample problem with a law of error following the lognormal distribution. Let $X_1, \dots, X_m$ be a sample of size m from $\Lambda(\mu_1, \sigma_1^2)$, and $Y_1, \dots, Y_n$ a sample of size n from $\Lambda(\mu_2, \sigma_2^2)$. Let ν_1 and τ_1^2 be the mean and variance of the X-population, and similarly, ν_2 and τ_2^2 the mean and variance of the Y-population. Relations between μ_i, σ_i^2 and ν_i, τ_i^2 are given by (1). By the two-sample problem, it is assumed that $\tau_1 = \tau_2$. We desire to test

$$H_0: \nu_1 = \nu_2 \quad \text{versus} \quad H_1: \nu_1 \neq \nu_2. \tag{3}$$

First, it may be noted that the difference between the two-sample and the one-sample problem in which only a single lognormal population involved is essential. With the aid of the best-developed normal theory, research results on the one-sample problem have been rich and mature (see Land [9] and the references therein). Nevertheless, when two (or more than two) lognormal populations are involved, the normal theory is of no help for the problem, for the difference of lognormal means, which is the quantity of interest, is no longer a linear function of the corresponding normal parameters so that standard procedures fail to apply. As usual, we thus turn to an asymptotic analysis. Throughout the paper, therefore, it is assumed that $m \to \infty$ and $n \to \infty$ with $m/N \to k \in (0, 1)$, where $N = m + n$ is the grant sample size.

The conventional test for H_0 is the Student t-test

$$T = (mn/N)^{1/2}(\overline{Y} - \overline{X})/S,$$

where $\overline{Y}$ and $\overline{X}$ are the sample means, and S the pooled sample standard deviation. Under null hypothesis H_0, the limit distribution of T is standard normal. Recently, Chen and Loh [5] argued that the Box-Cox t-test $T(\widehat{0})$ should be used instead of the Student t-test. Here $T(\lambda)$ is the transformed t-test defined by

$$T(\lambda) = (mn/N)^{1/2}\{\overline{Y}(\lambda) - \overline{X}(\lambda)\}/S(\lambda), \tag{4}$$

where $\overline{Y}(\lambda)$, $\overline{X}(\lambda)$, and $S(\lambda)$ are the sample means and pooled standard deviation of the transformed data through the Box-Cox (Box and Cox [3]) power transformation $h(x, \lambda) = (x^\lambda - 1)/\lambda$ if $x \neq 0$, and $\log(x)$ otherwise (so $\overline{Y}(\lambda) = n^{-1}\sum h(Y_i, \lambda)$, for example), and $\widehat{0}$ is the Box-Cox estimate for λ (also called the maximum likelihood estimate) chosen to minimize the function

$$L(\lambda) = S^2(\lambda)/\exp[2\lambda N^{-1}\{m\overline{X}(0) + n\overline{Y}(0)\}].$$

(The notation $\widehat{0}$ is employed here to indicate the fact that $\widehat{0} \approx 0$ for the present problem.) Chen and Loh [5] proved that the Box-Cox transformed t-test is asymptotically more efficient than the Student t-test; for the lognormal model in particular, testing power gained through the transformation $h(x, \widehat{0})$ is remarkable (Table 3 there). However, it can be noted that the estimating function $L(\lambda)$ (hence the estimate for λ and the induced test) does not use knowledge of the lognormal model assumption. In fact, the test is designed essentially as a nonparametric procedure. Therefore the Box-Cox transformed t-test may still be very far from the best.

In the present paper, we propose a new estimate for λ, denoted by $\hat{\lambda}$, and hence a new transformed t-test $T(\hat{\lambda})$ for H_0. The basic idea is to choose

$\hat{\lambda}$ such that the asymptotic efficacy of $T(\hat{\lambda})$ is maximized among all possible power transformations. The new test is then expected to be more powerful than the Box-Cox t-test or the ordinary Student t-test. To motivate this, consider Pitman-type alternatives as follows:

$$H_1^N : \nu_2 = \nu_1 + cN^{-1/2}. \tag{5}$$

Let λ be fixed temporarily. Under the alternatives, it can be seen that the asymptotic efficacy of $T(\lambda)$ is proportional to

$$Q(\lambda) = \begin{cases} \lambda^2 \exp\{\sigma^2(1-2\lambda)\}/\{\exp(\lambda^2\sigma^2) - 1\} & \text{if } \lambda \neq 0, \\ \sigma^{-2}\exp(\sigma^2/2) & \text{if } \lambda = 0. \end{cases} \tag{6}$$

Here σ^2 is the common variance of $\log(X)$ and $\log(Y)$ populations under the null hypothesis H_0, and it can be replaced by the estimator $\hat{\sigma}^2 = S^2(0)$, the pooled sample variance of transformed data by logarithm. The function $Q(\lambda)$ after substitution is denoted by $\widehat{Q}(\lambda)$; i.e.,

$$\widehat{Q}(\lambda) = \begin{cases} \lambda^2 \exp\{\hat{\sigma}^2(1-2\lambda)\}/\{\exp(\lambda^2\hat{\sigma}^2) - 1\} & \text{if } \lambda \neq 0, \\ \hat{\sigma}^{-2}\exp(\hat{\sigma}^2/2) & \text{if } \lambda = 0. \end{cases}$$

Definition 1. The estimate $\hat{\lambda}$ is the maximizer of $\widehat{Q}(\lambda)$.

With the estimate $\hat{\lambda}$, the test for H_0 is to reject H_0 if $|T(\hat{\lambda})| \geq z_\alpha$, where z_α is the upper (50α)th percentile of standard normal distribution.

Section 2 discusses the small sample and large sample properties of $\hat{\lambda}$. Lower and upper bounds on $\hat{\lambda}$ are provided. It is proven that under the null hypothesis, $\hat{\lambda}$ converges almost surely to a limit λ_0 and $N^{1/2}(\hat{\lambda} - \lambda_0)$ has normal limit distribution. Section 3 considers the asymptotic null and alternative distributions of the new test. It is then shown that the new transformed t-test is more powerful than the Box-Cox transformed t-test, and hence than the ordinary t-test, as expected. Section 4 reports a simulation study with discussion. Since the lognormal model is parametric, it would be desirable to compare $T(\hat{\lambda})$ with certain popular parametric test. In §4, Rao's efficient score test as a parametric testing procedure is included in the simulation study.

All proofs for the results in the paper are put in Appendix A.

2. Properties of the estimate $\hat{\lambda}$

It can be noted that finding $\hat{\lambda}$ needs numerical computation. The following lemma ensures $\hat{\lambda}$'s existence and uniqueness.

Lemma 1. *For $\hat{\sigma} > 0$, $\widehat{Q}(\lambda)$ is log concave downward on $(-\infty, \infty)$.*

Figure 1 displays a nice concave downward graph of $\hat{q}(\lambda) = \log \widehat{Q}(\lambda)$ with $\hat{\sigma} = 1$. Noting that $\hat{q}(-\infty) = \hat{q}(\infty) = -\infty$ and using the lemma yield the following corollary.

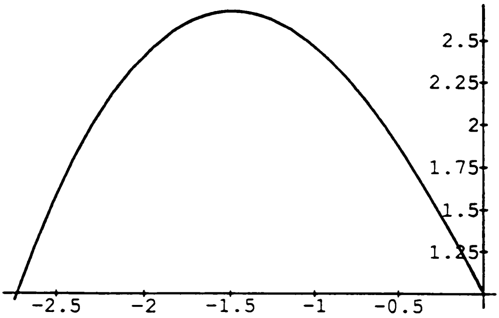

FIGURE 1. The graph of $\log \widehat{Q}(\lambda)$ with $\hat{\sigma} = 1$. $\hat{\lambda}$ is found to be -1.49.

Corollary 1. *For all observations with $\hat{\sigma} > 0$, $\widehat{Q}(\lambda)$ has only one local maximum which also serves as the globe maximum over $(-\infty, \infty)$.*

The corollary above has practical as well as theoretic value in finding $\hat{\lambda}$. For instance, one can first try an initial interval of λ. If the maximizer of $\widehat{Q}(\lambda)$ on the interval happens to be an interior point of the interval, terminate and the maximizer is exactly the $\hat{\lambda}$; otherwise the maximizer can be an indicator with which the direction of next try is advised. A further study gives the idea how to choose an interval smartly. The derivative of $\hat{q}(\lambda)$ is given by

$$\hat{q}'(\lambda) = 2\begin{cases} \lambda^{-1} - \hat{\sigma}^2 - \lambda\hat{\sigma}^2 \exp(\lambda^2\hat{\sigma}^2)/\{\exp(\lambda^2\hat{\sigma}^2) - 1\} & \text{if } \lambda \neq 0, \\ -\hat{\sigma}^2 & \text{if } \lambda = 0. \end{cases}$$

Setting $\hat{q}(\lambda) = 0$ and letting $x = \lambda\hat{\sigma}$, we have the estimating equation

$$\exp(x^2) = (\hat{\sigma}x - 1)/(x^2 + \hat{\sigma}x - 1), \qquad x \neq 0. \tag{7}$$

Equation (7) has a unique root from Corollary 1. To see where the root could be, consider the function of the rhs of (7). It has a horizontal asymptote at $y = 0$, and two vertical asymptotes at $x = x_1$ and x_2, respectively, where

$$x_1 = -(\hat{\sigma}/2)\{1 + (1 + 4/\hat{\sigma}^2)^{1/2}\} < 0, \qquad x_2 = -(\hat{\sigma}/2)\{1 - (1 + 4/\hat{\sigma}^2)^{1/2}\} > 0.$$

First note that the unique root of (7) must be greater than x_1 since the rhs of (7) is negative for $x < x_1$. Next, it is easy to see that the root cannot be positive since the function $t(x) = (x^2 + \hat{\sigma}x - 1)\exp(x^2) - (\hat{\sigma}x - 1)$ is positive for all $x > 0$ by observing that $t(0) = 0$ and $t'(x) = (2x^3 + 2\hat{\sigma}x^2)\exp(x^2) + \hat{\sigma}\{1 - \exp(x^2)\} > 0$ for $x > 0$. Therefore the unique root of (7) must be between x_1 and 0. This has established the following.

Theorem 1. *For $\hat{\sigma}^2 > 0$, $-\{1 + (1 + 4/\hat{\sigma}^2)^{-1/2}\}/2 < \hat{\lambda} < 0$.*

The lower bound is sharp enough in a sense of numerical computation. For example, when $\hat{\sigma}^2$ is even as small as 0.1, the lower bound is -3.7 so that one only needs to search interval $(-3.7, 0)$ for $\hat{\lambda}$. Table 1 shows us that the actual values are usually between -2 and -1.

TABLE 1. Numerical results of λ_0 for some values of σ

σ	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
$-\lambda_0$	1.84	1.71	1.59	1.49	1.41	1.34	1.28	1.24	1.20

Theorem 2. *Under null hypothesis* H_0, *we have*

(a) $\hat{\lambda} \to \lambda_0$ *almost surely, where* λ_0 *is the maximizer of* $Q(\lambda)$, *and*

(b) *in distribution,* $N^{1/2}(\hat{\lambda} - \lambda_0) \to N(0, \gamma^2)$, *where*

$$\gamma^2 = 2\lambda_0^4\sigma^4 \left\{ \frac{(\lambda_0 + 1)\exp(2\lambda_0^2\sigma^2) + (\lambda_0^3\sigma^2 - \lambda_0 - 2)\exp(\lambda_0^2\sigma^2) + 1}{(\lambda_0^2\sigma^2 + 1)\exp(2\lambda_0^2\sigma^2) - (2\lambda_0^4\sigma^4 + \lambda_0^2\sigma^2 + 2)\exp(\lambda_0^2\sigma^2) + 1} \right\}^2 .$$

The numerical solution of $\hat{\lambda}$ may be obtained in reality by specifying an interval appropriately, say $[-B, 0]$, where B is a positive constant. Theorem 1 suggests B does not need to be large. From our experience of Monte Carlo trials and Table 1, $B = 4$ is large enough. On the other hand, the theoretic framework can go with this reality. To be general, we assume B can be any positive constant or ∞. When $B = \infty$, it simply gives the unrestriction procedure. Let $\hat{\lambda}^*$ be the output of numerical computation. Since $\widehat{Q}(\lambda)$ is logconcave downward, we have $\hat{\lambda}^* = \max\{\hat{\lambda}, -B\}$. Let $\lambda_0^* = \max\{\lambda_0, -B\}$.

Corollary 2. *Under null hypothesis* H_0, *we have*

(a) $\hat{\lambda}^* \to \lambda_0^*$ *almost surely, and*

(b) $N^{1/2}(\hat{\lambda}^* - \lambda_0^*)$ *is bounded in probability*

From now on, we use $\hat{\lambda}^*$ instead of $\hat{\lambda}$ since $\hat{\lambda}^*$ is more general.

3. Null distribution and testing power

In this section, we study the asymptotic null and alternative distributions of $T(\hat{\lambda}^*)$. The following useful lemma was obtained by Doksum and Wong [7] and Carroll [4].

Lemma 2. *Under null hypothesis* H_0, $T(\hat{\lambda}^*) - T(\lambda_0^*) = o_P(1)$.

By this lemma, we immediately have

Corollary 3. *Under null hypothesis* H_0, $T(\hat{\lambda}^*)$ *has limit distribution* $N(0, 1)$.

Theorem 3. *Under alternative* H_1^N, $T(\hat{\lambda}^*)$ *has normal limit distribution* $N(\xi_1, 1)$, *where* $\xi_1^2 = c^2[k(1-k)]^2 Q(\lambda_0^*)$.

We know that under H_1^N, the Box-Cox transformed t-test $T(\widehat{0})$ has limit $N(\xi_2, 1)$ (Chen and Loh [5]), where $\xi_2^2 = c^2[k(1-k)]^2 Q(0)$. Therefore, we have that the Pitman asymptotic relative efficiency of $T(\hat{\lambda}^*)$ against $T(\widehat{0})$ is

$$e(\hat{\lambda}^*, \widehat{0}) = \frac{Q(\lambda_0^*)}{Q(0)} = \frac{\sigma^2\lambda_0^{*2}\exp(\sigma^2/2 - 2\lambda_0^*\sigma^2)}{\exp(\sigma^2\lambda_0^{*2}) - 1}.$$

By Lemma 1 and Theorem 1, we have

Theorem 4. *The test* $T(\hat{\lambda}^*)$ *is asymptotically more powerful than* $T(\widehat{0})$, *i.e.,* $e(\hat{\lambda}^*, \widehat{0}) > 1$.

The inequality in the theorem is actually very conservative and the values of the asymptotic relative efficiency may be much larger than 1. For example, when $\sigma = 0.4$, $e(\hat{\lambda}^*, \widehat{0}) = 7.131$; when $\sigma = 1.0$, $e(\hat{\lambda}^*, \widehat{0}) = 101.9$.

4. Monte Carlo study and discussion

A simulation study was carried out to assess the small sample performance of the transformed t-test $T(\hat{\lambda}^*)$, comparing to the test $T(\widehat{0})$. In addition, since the lognormal is a parametric model, we include Rao's efficient score test R (Rao [11, p. 418]) for comparison. The test R is described in Appendix B.

The simulation results reported here are for the 135 combinations of:

1. Ten values of σ—$0.4\,(0.2)\,2.0$.
2. Three sets of sample sizes (m, n)—$(10, 15)$, $(20, 30)$, and $(50, 70)$.
3. Five different location shifts including zero shift (null hypothesis)—$0, 0.10, 0.15, 0.20$, and 0.25.

In all combinations, the nominal testing power α was 5% and B was 4 in defining $\hat{\lambda}^*$. As a matter of fact, $\hat{\lambda}^* = -B$ never happened in the simulation, though. For each combination, 40,000 Monte Carlo trials were performed so that two times the estimated maximum standard error of the simulation is 0.005. All data sets were generated through use of the RNLNL routine in Version 1.1 of the October 1987 IMSL library and the simulation was done on the Cray supercomputer at Columbus, Ohio. Table 2 contains the results for the case $m = 10$, $n = 15$, Table 3 for $m = 20$, $n = 30$, and Table 4 for $m = 50$, $n = 70$. Some findings may be summarized as follows.

Table 2. Monte Carlo simulated rejection rate estimates for tests $T(\hat{\lambda}^*)$, $T(\widehat{0})$, and Rao's efficient score test R with sample sizes $m = 10$ and $n = 15$. The nominal level for all tests was 0.05, and 40,000 Monte Carlo trials were performed, giving a maximum simulation standard error of 0.0025. The same simulated data were used to obtain the level and power of the tests. Average values of $\hat{\lambda}$'s in simulation are given in parentheses.

shift	test	σ = 0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
	$T(\widehat{0})$	.066	.067	.066	.064	.066	.065	.066	.066	.068
.00	R	.059	.063	.065	.064	.066	.062	.061	.059	.058
	$T(\hat{\lambda}^*)$	.055	.050	.043	.038	.032	.029	.027	.025	.024
	$(-E\hat{\lambda}^*)$	(1.84)	(1.71)	(1.59)	(1.49)	(1.41)	(1.35)	(1.29)	(1.25)	(1.21)
	$T(\widehat{0})$	.122	.099	.089	.086	.083	.084	.085	.090	.091
.10	R	.139	.115	.108	.104	.096	.097	.095	.096	.097
	$T(\hat{\lambda}^*)$	.130	.101	.085	.080	.074	.076	.081	.087	.088
	$(-E\hat{\lambda}^*)$	(1.85)	(1.74)	(1.63)	(1.53)	(1.45)	(1.39)	(1.34)	(1.29)	(1.25)
	$T(\widehat{0})$	.193	.134	.117	.106	.102	.103	.105	.110	.109
.15	R	.227	.169	.152	.138	.133	.130	.132	.129	.131
	$T(\hat{\lambda}^*)$	.216	.148	.126	.114	.109	.112	.116	.121	.121
	$(-E\hat{\lambda}^*)$	(1.86)	(1.74)	(1.64)	(1.54)	(1.47)	(1.40)	(1.35)	(1.30)	(1.26)
	$T(\widehat{0})$	.283	.184	.144	.133	.132	.125	.126	.128	.130
.20	R	.343	.240	.210	.189	.182	.174	.171	.170	.169
	$T(\hat{\lambda}^*)$	.322	.212	.172	.154	.151	.146	.149	.155	.154
	$(-E\hat{\lambda}^*)$	(1.86)	(1.75)	(1.65)	(1.55)	(1.48)	(1.41)	(1.37)	(1.31)	(1.27)
	$T(\widehat{0})$	.390	.245	.185	.168	.156	.151	.146	.148	.150
.25	R	.475	.333	.280	.252	.232	.222	.218	.213	.208
	$T(\hat{\lambda}^*)$	.448	.294	.227	.206	.193	.190	.185	.188	.186
	$(-E\hat{\lambda}^*)$	(1.87)	(1.76)	(1.65)	(1.56)	(1.49)	(1.42)	(1.36)	(1.32)	(1.28)

TABLE 3. Monte Carlo simulated rejection rate estimates for tests $T(\hat{\lambda}^*)$, $T(\widehat{0})$, and Rao's efficient score test R with sample sizes $m = 20$ and $n = 30$. The nominal level for all tests was 0.05, and 40,000 Monte Carlo trials were preformed, giving a maximum simulation standard error of 0.0025. The same simulated data were used to obtain the level and power of the tests. Average values of $\hat{\lambda}$'s in simulation are given in parentheses.

		σ								
shift	test	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
	$T(\hat{0})$	.055	.056	.058	.057	.059	.056	.058	.057	.059
.00	R	.061	.066	.066	.063	.064	.063	.060	.060	.059
	$T(\hat{\lambda}^*)$	.050	.047	.041	.035	.032	.030	.027	.025	.024
	$(-E\hat{\lambda}^*)$	(1.84)	(1.71)	(1.59)	(1.49)	(1.41)	(1.35)	(1.29)	(1.25)	(1.21)
	$T(\hat{0})$	.169	.117	.100	.096	.095	.093	.099	.100	.108
.10	R	.213	.163	.150	.143	.139	.141	.140	.143	.149
	$T(\hat{\lambda}^*)$	.201	.145	.129	.126	.134	.144	.159	.171	.195
	$(-E\hat{\lambda}^*)$	(1.85)	(1.73)	(1.62)	(1.53)	(1.45)	(1.38)	(1.33)	(1.28)	(1.25)
	$T(\hat{0})$	.300	.188	.151	.139	.131	.129	.129	.132	.134
.15	R	.386	.282	.241	.225	.220	.213	.215	.215	.216
	$T(\hat{\lambda}^*)$	.368	.258	.225	.214	.218	.232	.247	.260	.271
	$(-E\hat{\lambda}^*)$	(1.86)	(1.74)	(1.63)	(1.54)	(1.46)	(1.40)	(1.34)	(1.29)	(1.26)
	$T(\hat{0})$	.473	.282	.218	.191	.179	.169	.172	.179	.179
.20	R	.576	.423	.358	.325	.312	.303	.299	.294	.295
	$T(\hat{\lambda}^*)$	.569	.392	.335	.316	.318	.319	.332	.342	.341
	$(-E\hat{\lambda}^*)$	(1.86)	(1.75)	(1.64)	(1.55)	(1.47)	(1.40)	(1.35)	(1.30)	(1.26)
	$T(\hat{0})$	.634	.392	.292	.252	.226	.216	.217	.218	.217
.25	R	.755	.570	.484	.444	.412	.398	.386	.375	.367
	$T(\hat{\lambda}^*)$	.741	.541	.457	.425	.408	.408	.405	.403	.398
	$(-E\hat{\lambda}^*)$	(1.87)	(1.76)	(1.65)	(1.56)	(1.48)	(1.41)	(1.36)	(1.31)	(1.27)

Testing power. The estimation method $\hat{\lambda}^*$ is proposed for the transformed t-test by a power-transformation to gain testing power as much as possible. The simulating results indeed support the faith. When sample sizes are as small as $m = 20$ and $m = 30$, even as $m = 10$ and 15, the testing power of $T(\hat{\lambda}^*)$ is overwhelmingly greater than that of $T(\widehat{0})$. When the sample sizes are large like $(50, 70)$, $T(\hat{\lambda}^*)$ has greater power than Rao's efficient score test R within the parameter combinations considered. Comparing the results for the sample sizes $(10, 15)$, $(20, 30)$, and $(50, 70)$, one can see that the power of $T(\hat{\lambda}^*)$ approaches 1 faster than that of R, which leads us to conclude that $T(\hat{\lambda})$ has bigger asymptotic power than R.

Significance level. The small sample significance level of $T(\hat{\lambda}^*)$ is lower than the nominal one, especially for $\sigma > 1$, while those of $T(\widehat{0})$ and R are higher. This, together with the comment above on testing power, appears to conclude that $T(\hat{\lambda}^*)$ has better small sample performance than $T(\widehat{0})$ does. But the accuracy for $T(\hat{\lambda}^*)$'s distribution to be approximated by $N(0, 1)$ strongly depends on σ's values.

Convergence of $\hat{\lambda}$. Comparing Table 1 with the simulating results, we see that $\hat{\lambda}$ approaches λ_0 very fast. When $m = 10$ and $n = 15$, margin of the estimation error is already within 0.01 on average under null hypothesis. And the effect of location-shift on the estimation seems to be slight.

TABLE 4. Monte Carlo simulated rejection rate estimates for tests $T(\hat{\lambda}^*)$, $T(\widehat{0})$, and Rao's efficient score test R with sample sizes $m = 50$ and $n = 70$. The nominal level for all tests was 0.05, and 40,000 Monte Carlo trials were performed, giving a maximum simulation standard error of 0.0025. The same simulated data were used to obtain the level and power of the tests. Average values of $\hat{\lambda}$'s in simulation are given in parentheses.

		σ								
shift	test	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
	$T(\widehat{0})$	.054	.054	.054	.055	.052	.054	.052	.053	.053
.00	R	.063	.067	.066	.066	.064	.059	.061	.057	.057
	$T(\hat{\lambda}^*)$	.050	.047	.041	.040	.034	.030	.028	.026	.024
	$(-E\hat{\lambda}^*)$	(1.84)	(1.71)	(1.59)	(1.49)	(1.41)	(1.34)	(1.29)	(1.24)	(1.21)
	$T(\widehat{0})$	.313	.191	.153	.141	.139	.138	.143	.155	.167
.10	R	.413	.307	.266	.253	.256	.255	.265	.280	.288
	$T(\hat{\lambda}^*)$	.399	.291	.274	.291	.324	.374	.420	.463	.492
	$(-E\hat{\lambda}^*)$	(1.85)	(1.73)	(1.62)	(1.52)	(1.44)	(1.38)	(1.32)	(1.28)	(1.24)
	$T(\widehat{0})$	.589	.361	.273	.239	.225	.224	.229	.239	.252
.15	R	.713	.541	.472	.440	.432	.431	.437	.440	.449
	$T(\hat{\lambda}^*)$	.713	.550	.509	.513	.544	.575	.598	.617	.609
	$(-E\hat{\lambda}^*)$	(1.86)	(1.74)	(1.63)	(1.54)	(1.46)	(1.39)	(1.34)	(1.29)	(1.25)
	$T(\widehat{0})$	.817	.555	.421	.359	.335	.323	.321	.333	.342
.20	R	.912	.763	.674	.633	.616	.604	.600	.596	.599
	$T(\hat{\lambda}^*)$	.912	.775	.715	.706	.707	.710	.703	.690	.666
	$(-E\hat{\lambda}^*)$	(1.86)	(1.75)	(1.64)	(1.54)	(1.47)	(1.40)	(1.34)	(1.30)	(1.26)
	$T(\widehat{0})$	.944	.727	.572	.488	.448	.432	.420	.419	.428
.25	R	.982	.906	.834	.792	.763	.750	.738	.726	.719
	$T(\hat{\lambda}^*)$	.983	.910	.857	.825	.803	.784	.759	.728	.700
	$(-E\hat{\lambda}^*)$	(1.87)	(1.75)	(1.65)	(1.55)	(1.47)	(1.41)	(1.35)	(1.30)	(1.26)

APPENDIX A: PROOFS

Proof of Lemma 1. It suffices to verify that the second derivative $\hat{q}''(\lambda)$ of $\hat{q}(\lambda) = \log \widehat{Q}(\lambda)$ is negative at all λ. For convenience, let $g_1(x) = \hat{q}''(x/\hat{\sigma})$. It can be seen that

$$g_1(x) = \begin{cases} -2\hat{\sigma}^2[x^{-2} + \{1 - (1+2x^2)\exp(-x^2)\}/\{1-\exp(-x^2)\}^2] & \text{if } x \neq 0, \\ -\hat{\sigma}^2 \quad \text{if } x = 0. \end{cases}$$

Then $g_1(x)$ is negative on $(-\infty, \infty)$ if and only if $g_2(x) = \{1 - \exp(-x)\}^2 + x\{1 - \exp(-x) - 2x\exp(-x)\}$ is positive on $(0, \infty)$. Again $g_2(x) > 0$ for all $x > 0$ if and only if $g_3(x) = (1+x)\exp(x) + \exp(-x) - 2x^2 - x - 2$ is positive for all $x > 0$. To prove the latter statement, study the first three derivatives of $g_3(x)$ given by the following:

$$g_3'(x) = (2+x)\exp(x) - \exp(-x) - 4x - 1,$$
$$g_3''(x) = (3+x)\exp(x) + \exp(-x) - 4,$$
$$g_3'''(x) = (4+x)\exp(x) - \exp(-x).$$

It is clear that $g_3'''(x) > 0$ for $x > 0$. Noting $g''(0) = g'(0) = g(0) = 0$, we see that $g_3'''(x) > 0$ for $x > 0$ implies $g_3''(x) > 0$ for $x > 0$ which in turn implies $g_3'(x) > 0$ for $x > 0$ and hence $g(x) > 0$ for all $x > 0$, completing the proof. □

Proof of Theorem 2. Part (a) is immediate from Lemma 1 and Corollary 1, and Corollary II.2 of Andersen and Gill [1]. To prove (b), applying the Taylor expansion to function $\widehat{Q}$ obtains: $0 = \widehat{Q}'(\hat{\lambda}) = \hat{Q}'(\lambda_0) + \hat{Q}''(\xi)(\hat{\lambda} - \lambda_0)$ where ξ is between $\hat{\lambda}$ and λ_0. Thus

$$N^{1/2}(\hat{\lambda} - \lambda_0) = -N^{1/2}\widehat{Q}'(\lambda_0)/\widehat{Q}''(\xi).$$

Now consider $\widehat{Q}'(\lambda_0)$ as a function of $\hat{\sigma}^2$, denoted by $g(\hat{\sigma}^2)$. Noting that $g(\sigma^2) = Q'(\lambda_0) = 0$ and in distribution $N^{1/2}(\hat{\sigma}^2 - \sigma^2) \to N(0, 2\sigma^4)$, using the delta-method yields that in distribution $N^{1/2}\widehat{Q}'(\lambda_0) \to N(0, 2\sigma^4 g'^2(\sigma^2))$. It is clear that with probability one $\widehat{Q}''(\xi) \to Q''(\lambda_0)$. Finally, it follows that $N^{1/2}(\hat{\lambda} - \lambda_0) \to N(0, \gamma^2)$ with $\gamma^2 = 2\sigma^4 g'^2(\sigma_2)/Q''^2(\lambda_0)$. The proof of part (b) is then completed by calculating g' and Q''. □

Proof of Corollary 2. If $\lambda_0 \geq -B$, then $\lambda_0^* = \lambda_0$ so that $|\hat{\lambda}^* - \lambda_0^*| \leq |\hat{\lambda} - \lambda_0|$. If $\lambda_0 < -B$, then $\lambda_0^* = -B$ so that $|\hat{\lambda}^* - \lambda_0^*| = |\max\{\hat{\lambda}, -B\} + B| \leq |\hat{\lambda} - \lambda_0|$, since $|\max\{\hat{\lambda}, -B\} + B| = 0$ if $\hat{\lambda} < -B$. By Theorem 2, the proof is completed. □

Proof of Theorem 3. The basic idea is to use Lemma 2 and LeCam's third lemma (LeCam and Young [10]). Let

$$l_N = \sum_{i=1}^{n} \log\{f(Y_i - cN^{-1/2})/f(Y_i)\},$$

where f is given by (2). By the LeCam's third lemma, it suffices to prove that under the null hypothesis, $\{T(\hat{\lambda}^*), l_N\}$ is asymptotically jointly bivariate normal $(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \sigma_{12})$ with $\theta_2 = -\sigma_2^2/2$, $\theta_1 + \sigma_{12} = \xi_1$, and $\sigma_1^2 = 1$. Under the null hypothesis, by Lemma 2, $T(\hat{\lambda}^*) - T(\lambda_0^*) = o_P(1)$ so that, instead of $\{T(\hat{\lambda}^*), l_N\}$, we only need to deal with $\{T(\lambda_0^*), l_N\}$, more simply with (Z_N, l_N), where

$$Z_N = (mn/N)^{1/2}\{\overline{Y}(\lambda_0^*) - \overline{X}(\lambda_0^*)\}/\delta,$$

and δ^2 is the variance of $h(X_1, \lambda_0^*)$ under the null hypothesis. The remaining of the proof is the same as the treatment in Chapter VI.2 of Hájek and Šidák [8]. □

Appendix B: Rao's efficient test R

Under the null-hypothesis, both $\log X$ and $\log Y$ are identically and normally distributed, say $N(\mu, \sigma^2)$. Let $\hat{\sigma}_0^2$ and $\hat{\mu}_0$ be the maximum likelilhood estimators for μ and σ^2 under the null-hypothesis. Let $\hat{\nu}_0 = \exp(\hat{\mu}_0 + \hat{\sigma}_0^2/2)$ and $\hat{\tau}_0^2 = \exp(2\hat{\mu}_0 + \hat{\sigma}_0^2)\{\exp(\hat{\sigma}_0^2) - 1\}$. Then it can be seen that Rao's efficient scores ϕ_i are as follows: with $\pi = 2\hat{\sigma}_0^2(\hat{\tau}_0 + \hat{\nu}_0)N^{1/2}$,

$$\phi_1 = \pi^{-1}\left\{m\hat{\tau}_0 - \hat{\tau}_0\sum_{i=1}^{m}(\log X_i)^2 + \hat{\sigma}_0(2\hat{\tau}_0 + \hat{\nu}_0)\sum_{i=1}^{m}\log X_i\right\},$$

$$\phi_2 = \pi^{-1}\left\{ n\hat{\tau}_0 - \hat{\tau}_0 \sum_{j=1}^{n} (\log Y_j)^2 + \hat{\sigma}_0 (2\hat{\tau}_0 + \hat{\nu}_0) \sum_{j=1}^{n} \log Y_j \right\},$$

$$\phi_3 = \pi^{-1}\hat{\nu}_0 \left\{ -N + \sum_{i=1}^{m} (\log X_i)^2 + \sum_{j=1}^{n} (\log Y_j)^2 - \hat{\sigma}_0 \left(\sum_{i=1}^{m} \log X_i + \sum_{j=1}^{n} \log Y_j \right) \right\}.$$

If $\mathcal{I} = (\mathcal{I}_{rs})$ is the 3×3 information matrix, then $\mathcal{I}_{11} = (m/N)a$, $\mathcal{I}_{22} = (1 - m/N)a$, $\mathcal{I}_{33} = (m/N)d$, $\mathcal{I}_{12} = 0$, $\mathcal{I}_{13} = (m/N)b$, $\mathcal{I}_{23} = (1 - m/N)b$, where

$$a = \{\hat{\tau}_0^2 + \hat{\sigma}_0^2 (2\hat{\tau}_0 + \hat{\nu}_0)^2\} / \{4\hat{\sigma}_0^4 (\hat{\tau}_0 + \hat{\nu}_0)^2 \hat{\nu}_0^2\},$$
$$b = (2\hat{\tau}_0 + \hat{\nu}_0)\hat{\sigma}_0 (1 - \hat{\sigma}_0) / \{4\hat{\sigma}_0^4 (\hat{\tau}_0 + \hat{\nu}_0)^2 \hat{\nu}_0\},$$

and

$$d = (\hat{\sigma}_0^2 + 2) / \{4\hat{\sigma}_0^4 (\hat{\tau}_0 + \hat{\nu}_0)^2\}.$$

Then Rao's efficient score test is given by

$$R = \phi' \mathcal{I}^{-1} \phi$$

where $\phi' = (\phi_1, \phi_2, \phi_3)$. The asymptotic null-distribution of R is χ_1^2.

REFERENCES

1. P. K. Andersen and R. D. Gill, *Cox's regression model for counting processes: a large sample study*, Ann. Statist. **10** (1982), 1100–1120.
2. J. Aitchison and J. A. C. Brown, *The lognormal distribution*, Cambridge Univ. Press, London and New York, 1957.
3. G. E. P. Box and D. R. Cox, *An analysis of transformations*, J. Roy. Statist. Soc. Ser. B **26** (1964), 211–252.
4. R. J. Carroll, *Tests for regression parameters in power transformations model*, Scand. J. Statist. **9** (1983), 217–222.
5. H. Chen and W-Y Loh, *Bounds on AREs of tests following Box-Cox transformations*, Ann. Statist. **20** (1992), 1485–1500.
6. E. L. Crow and K. Shimizu, *Lognormal distributions*, Marcel Dekker, New York, 1988.
7. K. A. Doksum and C-W Wong, *Statistical tests based on transformed data*, J. Amer. Statist. Assoc. **48** (1983), 411–417.
8. J. Hájek and K. Šidák, *Theory of rank tests*, Academia, Prague, 1967.
9. C. E. Land, *Hypothesis tests and interval estimates*, Lognormal Distributions (E. L. Crow and K. Shimizu, eds.), Marcel Dekker, New York, 1988.
10. L. M. LeCam and G. Yang, *Asymptotics in statistics: some basic concepts*, Springer-Verlag, New York, 1990.
11. C. R. Rao, *Linear statistical inference and its applications*, 2nd ed., Wiley, New York, 1973.

DEPARTMENT OF MATHEMATICS AND STATISTICS, BOWLING GREEN STATE UNIVERSITY, BOWLING GREEN, OHIO 43403

E-mail address: `hchen@andy.bgsu.edu`