

**ADJUSTED EMPIRICAL LIKELIHOOD METHOD AND PARAMETRIC
HIGHER ORDER ASYMPTOTIC METHOD WITH APPLICATIONS TO
FINANCE**

HANGJING WANG

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN MATHEMATICS AND STATISTICS
YORK UNIVERSITY
TORONTO, ONTARIO
FEBRUARY 2019

©Hangjing Wang 2019

Abstract

In recent years, applying higher order likelihood-based method to obtain inference for a scalar parameter of interest is becoming more popular in statistics because of the extreme accuracy that it can achieve. In this dissertation, we applied higher order likelihood-based method to obtain inference for the correlation coefficient of a bivariate normal distribution with known variances, and the mean parameter of a normal distribution with known coefficient of variation. Simulation results show that the higher order method has remarkable accuracy even when sample size is small.

The empirical likelihood (EL) method extends the traditional parametric likelihood-based inference method to a nonparametric setting. The EL method has several nice properties, however, it is subject to the convex hull problem, especially when sample size is small. In order to overcome this difficulty, Chen et al. (2008) proposed the adjusted empirical likelihood (AEL) method which adjusts the EL function by adding one “artificial”

point created from the observed sample. In this dissertation, we extended the AEL inference to the situation with nuisance parameters. In particular, we applied the AEL method to obtain inference for the correlation coefficient. Simulation results show that the AEL method is more robust than its competitors.

For the application to finance, we apply both the higher order parametric method and the AEL method to obtain inference for the Sharpe ratio. The Sharpe ratio is the prominent risk-adjusted performance measure used by practitioners. Simulation results show that the higher order parametric method performs well for data from normal distribution, but it is very sensitive to model specifications. On the other hand, the AEL method has the most robust performance under a variety of model specifications.

Keywords: adjusted empirical likelihood, ancillary direction, coverage probability, curved exponential family, modified signed log-likelihood ratio statistic, nonparametric, nuisance parameter, standardized maximum likelihood estimate departure.

Acknowledgements

First and foremost I would like to express my greatest appreciation to my supervisors, Professor Augustine Wong and Professor Yuejiao Fu. Without their valuable guidance, constant encouragement and extensive knowledge, this dissertation would not be possible.

I would also like to express my sincere appreciation to Professor Xin Gao, Professor Wei Liu as members of my supervisory committee.

Last but not least, I deeply thank my family, especially my parents for always believing in me, for their continuous love and their supports. A special thank goes to my beloved husband for his kindness and wisdom. This dissertation is dedicated to my lovely daughter who brings me happiness everyday.

Table of Contents

Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Tables	viii
List of Figures	x
1 Higher Order Parametric Inference	1
1.1 Review of likelihood-based inference	1
1.2 Higher order likelihood-based asymptotic inference	5
1.3 Inference on population mean for $N(\mu, c^2\mu^2)$	12
1.3.1 Background of the problem	12

1.3.2	The higher order statistic for the inference of $N(\mu, c^2\mu^2)$	14
1.3.3	Numerical studies	18
1.3.4	Discussion	30
1.4	Inference for bivariate normal correlation coefficient ρ	30
1.4.1	Background of the correlation coefficient ρ	30
1.4.2	The higher order statistic for bivariate normal correlation coefficient	32
1.4.3	Simulation study	40
1.4.4	Discussion	49
2	Adjusted Empirical Likelihood Method	50
2.1	Literature review	51
2.2	Adjusted empirical likelihood inference for the mean	61
2.3	Adjusted empirical likelihood with nuisance parameters	65
2.4	The AEL-based inference for the correlation coefficient	78
3	Application on the Sharpe Ratio	84
3.1	Higher order parametric inference for the Sharpe ratio sr	85
3.2	The AEL-based inference for the Sharpe ratio	91

3.2.1	Quantile-quantile plots	93
3.2.2	Simulation study	95
3.2.3	Real data analysis	100
3.3	Unbounded confidence region problem	104
4	Future Works	109
5	Bibliography	113

List of Tables

1.1	Comparing the methods proposed in Bhat and Rao (2007) and the proposed method using $n = 2$ and various combinations of c and μ	20
1.2	Comparing the methods proposed in Bhat and Rao (2007) and the proposed method for the case $n = 100$ and various combination of c and μ	21
1.3	Relative efficiency of different estimators with respect to \bar{X} for $n = 2$. . .	26
1.4	Relative efficiency of different estimators with respect to \bar{X} for $n = 3$. . .	27
1.5	Relative efficiency of different estimators with respect to \bar{X} for $n = 15$. .	28
1.6	Relative efficiency of different estimators with respect to \bar{X} for $n = 100$.	29
1.7	The lower tail error	45
1.8	The upper tail error	47
2.1	Coverage probabilities of the population mean	63

2.2	Coverage probabilities of the correlation coefficient	82
3.1	Simulated results of 90% and 95% confidence intervals for the Sharpe ratio	90
3.2	Coverage probabilities of the Sharpe ratio	98
3.3	Confidence Intervals of the Sharpe ratio for Apple Inc. return data	103
3.4	Coverage probability and median/mean length.	108

List of Figures

1.1	Average bias for different methods with sample size $n = 3$	42
1.2	Average bias for different methods with sample size $n = 5$	42
1.3	Average bias for different methods with sample size $n = 7$	43
1.4	Average bias for different methods with sample size $n = 10$	43
1.5	Average bias for different methods with sample size $n = 15$	44
1.6	Average bias for different methods with sample size $n = 20$	44
3.1	Quantile-quantile plot of data generated from $N(1, 0.25)$	94
3.2	Quantile-quantile plot of data generated from χ_4^2	94
3.3	Quantile-quantile plot of Apple Inc. return data	102

1 Higher Order Parametric Inference

1.1 Review of likelihood-based inference

As a formal concept, likelihood had appeared in jurisprudence, commerce and scholasticism long before it was given a rigorous mathematical foundation. The use of likelihood was popularized in statistics by Sir R.A. Fisher in his 1922 paper “On the mathematical foundations of theoretical statistics”. In that paper, Fisher proposed inferences based on likelihood functions that he termed as “maximum likelihood”, which fixed the terminology statisticians use today.

Let $X_1, X_2, \dots, X_n \in \mathbb{R}^d$ be the independent and identically distributed random vectors with density function $f(x; \theta)$, where $\theta \in \Theta \subset \mathbb{R}^p$ is a p -dimensional parameter with entries $(\theta_1, \theta_2, \dots, \theta_p)$. Their observed values will be denoted by x_1, x_2, \dots, x_n . Let $L(\theta)$ be the

likelihood function of the sample, then

$$L(\theta) \propto \prod_{i=1}^n f(x_i; \theta).$$

Denote the log-likelihood function by $\ell(\theta) = \log L(\theta)$. The maximum likelihood estimator (MLE) of θ is defined as $\hat{\theta} = \arg \sup_{\theta} \ell(\theta)$. Let $I(\theta) = E \left[-\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^r} \right]$ be the Fisher expected information matrix. Define the score function as $U(\theta) = \partial \ell(\theta) / \partial \theta$. At the true value θ_0 , it can be shown easily that $U(\theta_0)$ has mean 0 and variance is the Fisher expected information matrix evaluated at θ_0 . Then under the following regularity conditions (Shao, 2009),

1. $\theta \in \Theta$ and Θ is an open set in \mathbb{R}^p .
2. for every x in the range of x_1 , $f(x; \theta)$ is twice continuously differentiable in θ ,
3. $\frac{\partial}{\partial \theta^r} \int \frac{\partial f(x; \theta)}{\partial \theta} dx = \int \frac{\partial^2 f(x; \theta)}{\partial \theta \partial \theta^r} dx$,
4. $E \left[\frac{\partial}{\partial \theta} \log f(x; \theta) \frac{\partial}{\partial \theta^r} \log f(x; \theta) \right]$ is positive definite,
5. $E \left[\sup_{\theta \in \Theta} \left\| \frac{\partial^2 f(x; \theta)}{\partial \theta \partial \theta^r} \right\| \right] < \infty$,

we have

$$U(\theta_0)^r I(\theta_0)^{-1} U(\theta_0) \xrightarrow{d} \chi_p^2, \quad (1.1)$$

where θ_0 is the true value of θ and χ_p^2 denotes a χ^2 distribution with p degrees of freedom. Statistical tests based on (1.1) is known as the score test, also known as the Rao's test (Rao, 1947).

Moreover, it can be shown that the mean of $\hat{\theta}$ is asymptotically θ_0 and the corresponding variance is asymptotically the inverse of the Fisher expected information evaluated at θ_0 . Then under the same regularity conditions, Wald (1943) shows that

$$(\hat{\theta} - \theta_0)^\tau I(\theta_0)(\hat{\theta} - \theta_0) \xrightarrow{d} \chi_p^2, \quad (1.2)$$

and Wilks (1938) shows that

$$W(\theta_0) = 2[\ell(\hat{\theta}) - \ell(\theta_0)] \xrightarrow{d} \chi_p^2. \quad (1.3)$$

In practice, the parameter of interest ψ can usually be expressed as a function of θ . Suppose ψ is a q -dimensional parameter with true value ψ_0 . If ψ is a linear function of θ such that $\psi = c^\tau \theta$ for some constant matrix $c \in \mathbb{R}^{p \times q}$. Then Wald's result becomes

$$(\hat{\psi} - \psi_0)^\tau [c^\tau I(\theta_0)^{-1} c]^{-1} (\hat{\psi} - \psi_0) \xrightarrow{d} \chi_q^2$$

where $\hat{\psi} = c^\tau \hat{\theta}$. Similarly, the Wilks' result becomes

$$W(\psi_0) = 2[\ell(\hat{\theta}) - \ell(\hat{\theta}_\psi)] \xrightarrow{d} \chi_q^2$$

where $\hat{\theta}$ denote the global MLE as before and $\hat{\theta}_\psi$ is the constrained MLE with respect to the likelihood function $\ell(\theta)$ under the constraint $\psi(\theta) = c^\tau \theta = \psi_0$.

If the parameter of interest $\psi = \psi(\theta)$ is an arbitrary function of θ , the inference of ψ can still be based on Wald's or Wilks' result. By applying the delta method to Wald's result, we have

$$(\hat{\psi} - \psi_0)^\tau \left[\psi_\theta(\hat{\theta}) I(\theta_0)^{-1} \psi_\theta(\hat{\theta})^\tau \right]^{-1} (\hat{\psi} - \psi_0) \xrightarrow{d} \chi_q^2$$

where $\psi_\theta(\hat{\theta}) = (\partial\psi(\theta)/\partial\theta)|_{\hat{\theta}}$. Similarly, the Wilks' theorem for the likelihood ratio test becomes

$$W(\psi_0) = 2[\ell(\hat{\theta}) - \ell(\hat{\theta}_\psi)] \xrightarrow{d} \chi_q^2. \quad (1.4)$$

where $\hat{\theta}$ denote the global MLE as before and $\hat{\theta}_\psi$ is the constrained MLE with respect to the likelihood function $\ell(\theta)$ under the constraint $\psi(\theta) = \psi_0$. In general, $\hat{\theta}_\psi$ can be obtained using the method of Lagrange multiplier.

In this dissertation, the focus is when ψ is a scalar parameter of interest, that is $q = 1$. The details of Wald's and Wilks' results are discussed in Section 1.2. Note that the inference based on (1.1), (1.2) and (1.3) have asymptotic accuracy $O(n^{-\frac{1}{2}})$. The rest of Chapter 1 is devoted to improve the accuracy of inference to a higher order of accuracy. In Chapter 2, we relax the parametric setup to accommodate nonparametric families. In Chapter 3, we

apply the higher order parametric method in Chapter 1 and the AEL method in Chapter 2 to obtain inference on Sharpe ratio. We end this dissertation with a future work in Chapter 4.

1.2 Higher order likelihood-based asymptotic inference

Let $\psi = \psi(\theta)$ be a scalar parameter of interest and $\lambda = \lambda(\theta)$ is a vector of nuisance parameters. Again denote $\hat{\theta}$ to be the overall MLE of θ and let $\hat{\theta}_\psi$ be the constrained MLE for a given $\psi(\theta) = \psi$.

Two widely used methods for obtaining asymptotic confidence interval for ψ are based on the MLE of θ and the signed log-likelihood ratio statistic. It is well-known that $\hat{\theta}$ is asymptotically normally distributed with mean θ and variance $var(\hat{\theta})$. Since $\hat{\psi} = \psi(\hat{\theta})$, by applying the delta method, we have

$$var(\hat{\psi}) \approx \psi_\theta(\hat{\theta})var(\hat{\theta})\psi_\theta(\hat{\theta})^\tau.$$

Since the Fisher's expected information can be difficult to calculate, we can approximate it by using the observed information evaluated at $\hat{\theta}$ and hence, and we have

$$\widehat{var}(\hat{\theta}) \approx \left[-\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^\tau} \right]^{-1} \Bigg|_{\theta=\hat{\theta}}.$$

Therefore, $(\hat{\psi} - \psi) / \sqrt{\widehat{var}(\hat{\psi})}$ is asymptotically distributed as the standard normal distribution.

Thus, the p -value function of ψ can be approximated by

$$p_{MLE}(\psi) = \Phi \left(\frac{\hat{\psi} - \psi}{\sqrt{\widehat{var}(\hat{\psi})}} \right)$$

where $\Phi()$ is the cumulative distribution function of the standard normal distribution. Or

equivalently, a $100(1 - \alpha)\%$ confidence interval for ψ based on the MLE is

$$\left(\hat{\psi} - z_{\alpha/2} \sqrt{\widehat{var}(\hat{\psi})}, \hat{\psi} + z_{\alpha/2} \sqrt{\widehat{var}(\hat{\psi})} \right)$$

where $z_{\alpha/2}$ is the $100(1 - \alpha/2)^{th}$ percentile of the standard normal distribution.

Alternatively, under the regularity conditions (Shao, 2009), the signed log-likelihood ratio statistic

$$r(\psi) = \text{sgn}(\hat{\psi} - \psi) \{2[\ell(\hat{\theta}) - \ell(\hat{\theta}_\psi)]\}^{1/2} \quad (1.5)$$

is also asymptotically distributed according to a standard normal distribution. Hence, the p -value function of ψ is

$$p_{LR}(\psi) = \Phi(r(\psi))$$

and the $100(1 - \alpha)\%$ confidence interval for ψ based on the signed log-likelihood ratio statistic is

$$\{\psi : |r(\psi)| < z_{\alpha/2}\} .$$

Note that both methods have rates of convergence $O(n^{-1/2})$ only. Although in practice, the MLE-based method is often preferred because of the simplicity in calculations, Doganaksoy and Schmee (1993) illustrated that the signed log-likelihood statistic method has better coverage property than the MLE-based method in the cases that they examined.

In recent years, various adjustments have been proposed to improve the accuracy of the signed log-likelihood ratio statistic. In particular, Barndorff-Nielsen (1986, 1991) derived the modified signed log-likelihood ratio statistic for models with known ancillary statistic based on the Laplace method. Fraser et al. (1999) derived the version of the modified signed log-likelihood ratio statistic that can be applied to any model with log-likelihood function $\ell(\theta)$ and it takes the following form

$$r^*(\psi) = r(\psi) - \frac{1}{r(\psi)} \log \frac{r(\psi)}{Q(\psi)} \quad (1.6)$$

where $r(\psi)$ is the signed log-likelihood ratio statistic as defined in (1.5) and $Q(\psi)$ is the standardized maximum likelihood estimate departure calculated in the canonical parameter scale of an exponential family model. Assume the model cannot be expressed as a natural exponential family model, or when it is expressed as a natural exponential model, the dimension of the canonical parameter is not the same as the dimension of the original parameter. Fraser et al. (1999) derived a systematic approach for calculating $Q(\psi)$ based on

a locally defined canonical parameter. Let $z(\theta, x)$ be a pivotal quantity. Then the ancillary direction V is defined as

$$V = \left\{ \frac{\partial z(\theta, x)}{\partial x^\tau} \right\}^{-1} \left\{ \frac{\partial z(\theta, x)}{\partial \theta} \right\} \Big|_{(x, \hat{\theta})} \quad (1.7)$$

and the locally defined canonical parameter $\varphi(\theta)$ is

$$\varphi(\theta)^\tau = \frac{\partial \ell(\theta)}{\partial x} V. \quad (1.8)$$

Hence, the standardized maximum likelihood estimate departure calculated in the locally defined canonical parameter scale is

$$Q(\psi) = \text{sgn}(\hat{\psi} - \psi) \frac{|\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)|}{\sqrt{\widehat{\text{var}}(\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi))}} \quad (1.9)$$

where

$$\widehat{\text{var}}(\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)) = \frac{|J_{(\lambda\tau)}(\hat{\theta}_\psi)|}{|J_{\varphi\varphi^\tau}(\hat{\theta})|} \quad (1.10)$$

with

$$\chi(\theta) = \frac{\varphi^\psi(\hat{\theta}_\psi)}{\|\varphi^\psi(\hat{\theta}_\psi)\|} \varphi(\theta) \quad (1.11)$$

being a rotated coordinate of $\varphi(\theta)$ that agrees with $\psi(\theta)$ at $\hat{\theta}_\psi$. Let $\varphi_\theta(\theta)$ and $\varphi_\lambda(\theta)$ be the derivatives of $\varphi(\theta)$ with respect to θ and λ , respectively. Then $\varphi^\psi(\hat{\theta})$ is the row of $\varphi_\theta^{-1}(\theta)$ that

corresponds to ψ , $\|\varphi^\psi(\hat{\theta})\|$ is the Euclidean distance of the vector $\varphi_\theta^{-1}(\theta)$, and

$$|j_{\varphi\varphi^\tau}(\hat{\theta})| = |j_{\theta\theta^\tau}(\hat{\theta})| |\varphi_\theta(\hat{\theta})|^{-2} \quad \text{and} \quad |j_{(\lambda\lambda^\tau)}(\hat{\theta}_\psi)| = |j_{\lambda\lambda^\tau}(\hat{\theta}_\psi)| |\varphi_\lambda(\hat{\theta}_\psi)^\tau \varphi_\lambda(\hat{\theta}_\psi)|^{-1}$$

where

$$j_{\theta\theta^\tau}(\hat{\theta}) = - \left. \frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^\tau} \right|_{\hat{\theta}} \quad (1.12)$$

and

$$j_{\lambda\lambda^\tau}(\hat{\theta}_\psi) = - \left. \frac{\partial^2 \ell(\theta)}{\partial \lambda \partial \lambda^\tau} \right|_{\hat{\theta}_\psi} .$$

If the canonical parameter is explicitly available and its dimension is the same as the dimension of θ , then we can directly go to equation (1.9) to obtain $Q(\psi)$. Once we have $Q(\psi)$, the modified signed log-likelihood ratio statistic $r^*(\psi)$ can be obtained from equation (1.6). It is shown in Fraser et al. (1999) that $r^*(\psi)$ is asymptotically distributed as the standard normal distribution with rate of convergence $O(n^{-3/2})$. Thus the p -value function of ψ based on the modified signed log-likelihood ratio statistic is

$$p_{r^*}(\psi) = \Phi(r^*(\psi)) \quad (1.13)$$

and the corresponding $100(1 - \alpha)\%$ confidence interval for ψ based on the signed log-likelihood ratio statistic is

$$\{\psi : |r^*(\psi)| < z_{\alpha/2}\} .$$

We can summarize the method by Fraser and Reid (1999) in the following algorithm:

Algorithm 1.2.1.

Given

The likelihood function $\ell(\theta)$.

Aim

Obtain p -value function of ψ .

Step 1

Based on the likelihood function $\ell(\theta)$, we can obtain the MLE $\hat{\theta}$, $\ell(\hat{\theta})$, $\hat{\psi} = \psi(\hat{\theta})$ as well as $j_{\theta\theta^r}(\hat{\theta})$ defined in (1.12).

Step 2

By maximizing $\ell(\theta)$ subject to the constraint $\psi(\theta) = \psi$, we can obtain the constrained MLE $\hat{\theta}_\psi$.

Step 3

Compute the signed log-likelihood ratio statistic $r(\psi)$ defined in (1.5).

Step 4

If the canonical parameter $\varphi(\theta)$ is explicitly available, and also the dimension of the canonical parameter is the same as the dimension of the original parameter, then go to Step 7. Otherwise, obtain a pivotal quantity $z(\theta, x)$.

Step 5

Obtain the ancillary direction V as in (1.7).

Step 6

Calculate the local canonical parameter $\varphi(\theta)$ given by (1.8).

Step 7

Compute the MLE departure $(\hat{\psi} - \psi)$ in the $\varphi(\theta)$ scale via (1.11) and (1.10).

Step 8

Use (1.9) to calculate $Q(\psi)$, which is the standardized maximum likelihood estimate departure in the canonical parameter scale.

Step 9

Obtain the modified signed log-likelihood ratio statistic $r^(\psi)$ as in (1.6).*

Step 10

Obtain the p -value function of ψ based on the modified signed log-likelihood ratio statistic $r^(\psi)$ as in (1.13).*

1.3 Inference on population mean for $N(\mu, c^2\mu^2)$

1.3.1 Background of the problem

Normal distribution is one of the most widely known and commonly used distributions in statistics. Even in the introductory statistics courses, we discussed inference about the mean of a normal distribution. Usually we assume that the population mean and the population standard deviation are unrelated parameters. However, in many physical and biological applications the population standard deviation is often found to be proportional to the mean. That is, the mean and standard deviation are related. The ratio of the standard deviation to the mean is defined as the coefficient of variation (CV) in statistics. The focus of this section is to make inference on the normal mean when CV is known.

In practice, this problem arises more frequently than we might anticipate. For example, in environmental studies, inference about the mean of the pollutant is of special interest. And in those studies, the standard deviation of a pollutant is often assumed to be directly related to the mean of the pollutant (Niwitpong, 2012). In agricultural studies, it is customary to

conduct multi-location trials. From the results of a few locations, the CV can be calculated and subsequently used as a known value for studying the mean of the experiment conducted in a new location (Bhat and Rao, 2007). Brazauskas and Ghorai (2007) also gives examples of this problem emerging from biological and medical experiments. From the theoretical point of view, estimating a normal mean with known CV is also an interesting problem because it has a scalar parameter but a two-dimensional minimal sufficient statistic. In other words, we have a curved exponential family model, and standard inferential methods cannot be directly applied (see Efron (1975)).

In literature, many authors have studied point estimation of a normal mean with known CV. For example, a consistent estimator was obtained by Searls (1964) based on truncation of extreme observations. Khan (1968) derived the best unbiased estimator with minimum variance. Gleser and Healy (1976) obtained the uniformly minimum risk estimator when the loss function is the squared error. Sen (1979) proposed a simple and consistent estimator but the proposed estimator is biased. Guo and Pal (2003) worked out an estimator based on the scaled quadratic loss function. Chaturvedi and Tomer (2003) extended the method in Singh (1998) and proposed a three-stage procedure and an accelerated sequential procedure to estimate the normal mean. By various ways of combining the minimal sufficient statistic,

Anis (2008) proposed three simple but biased estimators. And most recently, Srisodaphol and Tongmol (2012) suggested that the estimator based on jackknife technique is preferred as it has the smallest mean square error.

Despite the large literature devoted to point estimation, very few literature is available for interval estimation and hypothesis test for the normal mean with known CV. Hinkley (1977) derived two locally most powerful test for right alternatives based on an ancillary statistic. Bhat and Rao (2007) examined the likelihood ratio test and the Wald test. Niwitpong (2012) proposed two confidence intervals for the normal mean based on the work of Searls (1964).

In this section, we extended the approach of Bhat and Rao (2007) and applied the modified signed log-likelihood ratio test for the normal mean with known CV. The proposed method is known to have third-order accuracy. Moreover, a new estimator is obtained from the modified signed log-likelihood ratio statistic.

1.3.2 The higher order statistic for the inference of $N(\mu, c^2\mu^2)$

Let x_1, \dots, x_n be a random sample from a normal distribution with mean μ and variance σ^2 .

We follow the set up in Srisodaphol and Tongmol (2012) that the coefficient of variation $c = \frac{\sigma}{\mu}$ is known. Without loss of generality, we assume $\mu > 0$, and hence c is positive. The

log-likelihood function is

$$\ell(\mu) = -n \log \mu + \frac{nt_1}{c^2\mu} - \frac{nt_2}{2c^2\mu^2}, \quad (1.14)$$

where $(t_1, t_2) = (\sum_{i=1}^n x_i/n, \sum_{i=1}^n x_i^2/n)$ is the minimal sufficient statistic. This is a curved exponential model as defined in Efron (1975) with a two-dimensional minimal sufficient statistic but only a one-dimensional parameter. Classical statistical methods cannot be directly applied to obtain the p -value function of μ . We apply Algorithm 1.2.1 to obtain inference for μ .

Step 1 – Step 3 of Algorithm 1.2.1 follows naturally from the likelihood function. Since we assumed μ is positive, and c is positive and known, the overall MLE of μ is

$$\hat{\mu} = \frac{-t_1 + \sqrt{t_1^2 + 4c^2t_2}}{2c^2}$$

and the observed information evaluated at MLE is

$$j_{\mu\mu}(\hat{\mu}) = -\frac{n(c^2\hat{\mu}^2 + 2t_1\hat{\mu} - 3t_2)}{c^2\hat{\mu}^4}.$$

The signed log-likelihood ratio statistic is

$$r = r(\mu) = \text{sgn}(\hat{\mu} - \mu)\{2[\ell(\hat{\mu}) - \ell(\mu)]\}^{1/2}.$$

Following the rationale of Step 4 of the algorithm, since the model belongs to a curved exponential family, the dimension of the canonical parameter is larger than the dimension

of the original parameter. Hence, we need to obtain the locally defined canonical parameter $\varphi(\mu)$ which depends on the pivotal quantity $z(\mu, x)$. In this case, the pivotal quantity for the i^{th} observation is

$$z_i = z(\mu, x_i) = \frac{x_i - \mu}{c\mu}$$

and we have

$$\frac{\partial z_i}{\partial x_i} = \frac{1}{c\mu}, \quad \frac{\partial z_i}{\partial \mu} = -\frac{x_i}{c\mu^2}.$$

The i^{th} component of the ancillary direction in Step 5 is

$$V_i = -\left(\frac{\partial z_i}{\partial x_i}\right)^{-1} \left(\frac{\partial z_i}{\partial \mu}\right) \Big|_{(x_i, \hat{\mu})} = \frac{x_i}{\hat{\mu}}.$$

Moreover

$$\frac{\partial \ell(\mu)}{\partial x_i} = \frac{1}{c^2\mu} - \frac{x_i}{c^2\mu^2}$$

and the locally defined canonical parameter $\varphi(\mu)$ of Step 6 is

$$\varphi(\mu) = \sum_{i=1}^n \frac{\partial \ell(\mu)}{\partial x_i} V_i = \frac{n}{c^2\hat{\mu}} \left(\frac{t_1}{\mu} - \frac{t_2}{\mu^2} \right)$$

with

$$\varphi_\mu(\mu) = \frac{\partial \varphi(\mu)}{\partial \mu} = -\frac{n}{c^2\hat{\mu}\mu} \left(\frac{t_1}{\mu} - \frac{2t_2}{\mu^2} \right).$$

Since there is no nuisance parameter involved in this problem, Step 7 of Algorithm 1.2.1 follows from simplifying (7) and (6). We have

$$\chi(\mu) = \varphi(\mu)$$

and

$$\widehat{\text{var}}(\chi(\hat{\mu}) - \chi(\mu)) = |j_{\mu\mu}(\hat{\mu})|^{-1} |\varphi_{\mu}(\hat{\mu})|^2 .$$

By Step 8, the maximum likelihood departure in $\varphi(\mu)$ scale is

$$Q = \left[t_1 \left(\frac{1}{\hat{\mu}} - \frac{1}{\mu} \right) - t_2 \left(\frac{1}{\hat{\mu}^2} - \frac{1}{\mu^2} \right) \right] \frac{\sqrt{n}\hat{\mu}}{c} \frac{1}{\sqrt{t_2 + c^2\hat{\mu}^2}} . \quad (1.15)$$

Thus $r^*(\mu)$ is calculated from Step 9 of the algorithm and the p -value function of μ , $p(\mu)$, can be obtained from Step 10 of the algorithm.

In addition, we proposed a new estimator of μ which is a by-product of the modified signed log-likelihood ratio method. We denote our new estimator as $\tilde{\mu}$ which satisfies

$$\Phi(r^*(\tilde{\mu})) = 0.5$$

or equivalently

$$r^*(\tilde{\mu}) = 0 .$$

Although the explicit form of $\tilde{\mu}$ is not available, it can be obtained easily by simple numerical methods.

1.3.3 Numerical studies

Our first simulation study is to compare the accuracy of the confidence intervals obtained from the Wald method (Wald) and the likelihood ratio method (LR) as discussed in Bhat and Rao (2007), and also those obtained by the proposed method (r^*). We consider the extreme case of $n = 2$. For each combinations of $c = 1, 10, 20$ and $\mu = 2, 5, 10$, ten thousand replications are performed. For each generated sample, the 95% confidence interval for μ is calculated. The performance of a method is judged using the following criteria:

- the coverage probability (CP)

Proportion of the true μ falls within the 95% confidence interval

- the lower tail error rate (LE)

Proportion of the true μ falls below the lower limit of the 95% confidence interval

- the upper tail error rate (UE)

Proportion of the true μ falls above the upper limit of the 95% confidence interval

- the average absolute bias (AB)

$$AB = \frac{|LE - 0.025| + |UE - 0.025|}{2} .$$

The desired values are 0.95, 0.025, 0.025 and 0, respectively. These values reflect the desired properties of the accuracy and symmetry of the interval estimates of μ . Results are recorded in Table 1.1. The Wald method gives unsatisfactory coverage probability. LR gives decent coverage probability. Both the Wald method and the likelihood ratio method gives asymmetric intervals. However, the proposed modified signed log-likelihood ratio method gives excellent results in all four criteria even for this extreme sample size case. Table 1.2 recorded a large sample situation ($n = 100$) with $c = 5$ and $\mu = 10$. For such a large sample, the Wald method gives decent coverage probability but also gives asymmetric intervals. Both LR and r^* give similar coverage probability with r^* having a smaller average bias.

c	μ	Wald				LR				r^*			
		CP	LE	UE	AB	CP	LE	UE	AB	CP	LE	UE	AB
1	2	0.7784	0.2216	0	0.1108	0.9308	0.0553	0.1390	0.0207	0.9500	0.0231	0.0269	0.0019
	5	0.7680	0.2320	0	0.1160	0.9295	0.0591	0.0114	0.0239	0.9493	0.0279	0.0228	0.0026
	10	0.7728	0.2272	0	0.1136	0.9331	0.0545	0.0124	0.0211	0.9465	0.0265	0.0270	0.0018
10	2	0.7808	0.2192	0	0.1096	0.9357	0.0543	0.0100	0.0222	0.9521	0.0233	0.0246	0.0011
	5	0.7770	0.2230	0	0.1115	0.9361	0.0529	0.0110	0.0210	0.9537	0.0237	0.0226	0.0019
	10	0.7676	0.2324	0	0.1162	0.9274	0.0611	0.0115	0.0248	0.9494	0.0266	0.0240	0.0013
20	2	0.7773	0.2227	0	0.1114	0.9319	0.0539	0.0142	0.0199	0.9502	0.0237	0.0261	0.0012
	5	0.7770	0.2230	0	0.1115	0.9347	0.0537	0.0116	0.0211	0.9509	0.0235	0.0256	0.0011
	10	0.7785	0.2215	0	0.1108	0.9338	0.0529	0.0133	0.0198	0.9499	0.0245	0.0256	0.0006

Table 1.1: Comparing the methods proposed in Bhat and Rao (2007) and the proposed method using $n = 2$ and various combinations of c and μ .

c	μ	Wald				LR				r^*			
		CP	LE	UE	AB	CP	LE	UE	AB	CP	LE	UE	AB
1	2	0.9455	0.0409	0.0136	0.0136	0.9491	0.0281	0.0228	0.0026	0.9497	0.0250	0.0253	0.0001
	5	0.9419	0.0464	0.0117	0.0173	0.9501	0.0308	0.0191	0.0059	0.9522	0.0268	0.0210	0.0029
	10	0.9489	0.0383	0.0128	0.0128	0.9516	0.0258	0.0226	0.0016	0.9503	0.0241	0.0256	0.0008
10	2	0.9431	0.0425	0.0144	0.0141	0.9481	0.0279	0.0240	0.0020	0.9490	0.0248	0.0262	0.0007
	5	0.9425	0.0447	0.0128	0.0159	0.9478	0.0292	0.0230	0.0031	0.9473	0.0275	0.0252	0.0013
	10	0.9459	0.0424	0.0117	0.0153	0.9532	0.0250	0.0218	0.0016	0.9545	0.0217	0.0238	0.0023
20	2	0.9423	0.0439	0.0138	0.0151	0.9480	0.0289	0.0231	0.0029	0.9477	0.0265	0.0258	0.0011
	5	0.9466	0.0415	0.0119	0.0148	0.9504	0.0273	0.0223	0.0025	0.9516	0.0238	0.0246	0.0008
	10	0.9462	0.0424	0.0114	0.0155	0.9528	0.0255	0.0217	0.0019	0.9521	0.0229	0.0250	0.0011

Table 1.2: Comparing the methods proposed in Bhat and Rao (2007) and the proposed method for the case $n = 100$ and various combination of c and μ .

Anis (2008) compares the relative efficiency of ten point estimators of μ (denoted as T_1, T_2, \dots, T_{10}) with the “standard” estimator \bar{X} and concluded that T_6 , which is the maximum likelihood estimator, performs best. Moreover, T_3 , which is easy to compute, is comparable to T_8, T_9 and T_{10} .

To be specific, let $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ and $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Then $E(S) = \alpha_n \sigma$ and $Var(S) = (1 - \alpha_n^2) \sigma^2$, where $\alpha_n = \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \sqrt{\frac{2}{n-1}}$. Set $S^{*2} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ and $C_n = \frac{1}{c\alpha_n \sqrt{n-1}}$. Further recall that for any estimators $\hat{\theta}_1, \hat{\theta}_2$ of θ , the relative efficiency of $\hat{\theta}_1$ over $\hat{\theta}_2$ is

$$RE(\hat{\theta}_1, \hat{\theta}_2) = \frac{MSE(\hat{\theta}_2)}{MSE(\hat{\theta}_1)},$$

where MSE is the mean square error of the estimator.

The following is the list of estimators under consideration:

- T_1 : The estimator T_1 is defined as

$$T_1 = \xi \bar{X} + (1 - \xi)S,$$

where $\xi = 1 - \frac{c^2}{c^2 - n(c^2 - 2c\alpha_n + 1)}$. Note that T_1 is the estimator that has the smallest

mean square error (MSE) among all estimators of the form $T(\beta) = \beta \bar{X} + (1 - \beta)S$

with $0 \leq \beta \leq 1$. The relative efficiency of T_1 over \bar{X} is

$$\text{RE}(T_1, \bar{X}) = 1 + \frac{c^2}{n(c^2 - 2c\alpha_n + 1)}$$

- T_2 : The estimator T_2 can be written as

$$T_2 = r(\bar{X} + S),$$

with $r = \frac{n(1 + c\alpha_n)}{c^2 + nc^2 + n + 2nc\alpha_n}$. T_2 is the estimator that has the smallest MSE among all estimators of the form $T(\beta) = \beta\bar{X} + S$. The relative efficiency of T_2 over \bar{X} is

$$\text{RE}(T_2, \bar{X}) = \frac{c^2 + nc^2 + n + 2nc\alpha_n}{n(1 + n(1 - \alpha_n^2))}$$

- T_3 : As a generalization of T_1 and T_2 , T_3 can be expressed as

$$T_3 = W_1\bar{X} + W_2S,$$

where $W_1 = \frac{n(1 - \alpha_n^2)}{c^2 + n(1 - \alpha_n^2)}$ and $W_2 = \frac{c\alpha_n}{c^2 + n(1 - \alpha_n^2)}$. T_3 is the estimator that has minimizes the MSE of $T(\beta_1, \beta_2) = \beta_1\bar{X} + \beta_2S$. The relative efficiency of T_3 over \bar{X} is

$$\text{RE}(T_3, \bar{X}) = \frac{c^2 + n(1 - \alpha_n^2)}{n(1 - \alpha_n^2)}$$

- T_4 : The fourth estimator discussed in Anis (2008) was proposed by Khan (1968) as

$$T_4 = C_n \sqrt{n}S^*.$$

- T_5 : The fifth estimator is similar to T_1 and takes the form

$$T_5 = AT_4 + (1 - A)\bar{X},$$

where $A = \frac{c^2\alpha_n^2}{c^2\alpha_n^2 + n(1 - \alpha_n^2)}$. Khan (1968) proved that T_4 has the smallest variance among all unbiased estimators that are linear in \bar{X} and S^* .

- T_6 : The sixth estimator to be compared is the maximum likelihood estimator

$$T_6 = \hat{\mu} = \frac{\sqrt{4c^2S^{*2} + (1 + 4c^2)\bar{X}^2} - \bar{X}}{2c^2}.$$

- T_7 : Searls (1964) proposed the estimator $T_7 = \frac{n\bar{X}}{n + c^2}$.

- T_8 : The eighth estimator under comparison is again linear in \bar{X} and S^* . It takes the form

$$T_8 = \beta\bar{X} + (1 - \beta)\frac{S^*}{c},$$

where $\beta = \frac{2n - 1 - 2\alpha_n\sqrt{n(n-1)}}{c^2 + 2n - 1 - 2\alpha_n\sqrt{n(n-1)}}$. This estimator was proposed by Sen (1979).

- T_9 : Let $\delta_n = \frac{(n-1)c^2\xi_n^2}{n} - 1$ and $\xi_n = \frac{1}{c\alpha_n}\sqrt{\frac{n}{n-1}}$. Gleser and Healy introduced

$$T_9 = \frac{n\delta_n\bar{X} + c^2\xi_n S^*}{c^2 + (c^2 + n)\delta_n}.$$

- T_{10} : The last we are interested in is the minimum discrimination information estimator proposed by Soofi and Gokhale (1991)

$$T_{10} = \frac{\sqrt{\bar{X}^2 + 4(c^2 + 1)S^{*2}} + \bar{X}}{2(1 + c^2)}.$$

We mimic the simulation study discussed in Anis (2008) to compare our proposed estimator $\tilde{\mu}$ to the ten estimators discussed in Anis (2008). As in Anis (2008), we chose $\mu = 100$, for the combinations of $c = 0.05, 0.10, 0.15, 0.20, 0.25, 0.50, 0.75, 1.00, 1.50, 2.00, 2.50, 3.00$ and $n = 2, 3, 15, 100$, ten thousand replications were performed. For each generated sample, we calculated the relative efficiency of the estimator with the “standard” estimator \bar{X} . Results are reported in Table 1.3–1.6.

From Table 1.3–1.6, we observe that our proposed estimator $\tilde{\mu}$ has the best performance when CV takes small values because it has the largest relative efficiency. When CV is relatively large, T_6 performs best and our proposed estimator $\tilde{\mu}$ ranks second. The interval estimate based on the maximum likelihood estimator (T_6) does not give satisfactory coverage probability properties. On the other hand, the interval estimate based on the modified signed log-likelihood ratio statistic ($\tilde{\mu}$) has the best coverage probability properties. Thus, the proposed method is the recommended method.

CV	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	$\tilde{\mu}$
0.05	1.002	0.621	1.004	0.002	1.001	1.005	1.002	1.004	1.004	1.005	1.013
0.10	1.005	0.683	1.014	0.009	1.010	1.015	1.004	1.013	1.014	1.015	1.020
0.15	1.012	0.738	1.031	0.019	1.022	1.034	1.009	1.030	1.031	1.034	1.033
0.20	1.031	0.816	1.063	0.036	1.041	1.069	1.022	1.062	1.063	1.069	1.053
0.25	1.048	0.843	1.081	0.052	1.049	1.090	1.032	1.080	1.081	1.089	1.090
0.50	1.264	1.248	1.335	0.216	1.218	1.422	1.115	1.326	1.335	1.389	1.310
0.75	1.760	1.752	1.765	0.494	1.487	2.027	1.276	1.748	1.765	1.902	1.948
1.00	2.296	2.436	2.442	0.893	1.915	3.060	1.523	2.409	2.442	2.687	2.433
1.50	2.287	3.960	4.125	1.974	2.949	5.669	2.177	4.071	4.125	4.557	4.443
2.00	2.124	5.899	6.534	3.544	4.556	9.501	2.966	6.391	6.534	7.024	7.288
2.50	1.911	8.390	9.626	5.415	6.370	14.40	4.181	9.467	9.626	10.21	10.95
3.00	1.876	11.16	13.42	7.912	8.951	20.75	5.438	13.11	13.42	13.97	15.43

Table 1.3: Relative efficiency of different estimators with respect to \bar{X} for $n = 2$

CV	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	$\tilde{\mu}$
0.05	1.002	0.670	1.006	0.003	1.004	1.006	1.001	1.006	1.006	1.006	1.032
0.10	1.003	0.729	1.015	0.012	1.013	1.017	1.002	1.014	1.015	1.017	1.041
0.15	1.008	0.800	1.036	0.028	1.031	1.041	1.006	1.035	1.036	1.040	1.058
0.20	1.012	0.848	1.058	0.047	1.051	1.064	1.007	1.055	1.058	1.064	1.083
0.25	1.031	0.917	1.090	0.076	1.072	1.100	1.019	1.088	1.090	1.100	1.115
0.50	1.244	1.364	1.401	0.308	1.306	1.462	1.092	1.395	1.401	1.445	1.398
0.75	1.765	1.833	1.833	0.694	1.661	2.041	1.186	1.813	1.833	1.928	1.893
1.00	2.438	2.358	2.519	1.237	21.99	2.940	1.318	2.481	2.519	2.678	2.628
1.50	2.321	3.959	4.443	2.789	3.775	5.597	1.694	4.342	4.443	4.665	4.782
2.00	1.897	6.094	7.248	4.833	5.683	9.201	2.367	7.126	7.248	7.490	7.885
2.50	1.732	8.799	11.33	7.854	8.939	14.31	3.165	11.12	11.33	11.57	11.87
3.00	1.630	11.09	14.93	10.98	11.92	19.80	3.996	14.62	14.93	15.08	16.74

Table 1.4: Relative efficiency of different estimators with respect to \bar{X} for $n = 3$

CV	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	$\tilde{\mu}$
0.05	1.000	0.731	1.006	0.005	1.006	1.006	1.000	1.006	1.006	1.006	1.017
0.10	1.000	0.803	1.020	0.019	1.020	1.020	1.000	1.019	1.020	1.020	1.029
0.15	1.003	0.862	1.042	0.041	1.039	1.043	1.002	1.042	1.042	1.043	1.051
0.20	1.003	0.949	1.080	0.073	1.078	1.082	1.001	1.079	1.080	1.082	1.082
0.25	1.006	1.022	1.119	0.115	1.116	1.122	1.003	1.117	1.119	1.122	1.123
0.50	1.069	1.471	1.473	0.454	1.453	1.485	1.020	1.472	1.473	1.484	1.468
0.75	1.442	2.024	2.093	1.078	2.047	2.119	1.043	2.088	2.093	2.112	2.055
1.00	2.876	2.628	2.889	1.852	2.830	2.983	1.072	2.872	2.889	2.917	2.887
1.50	1.521	4.243	5.421	4.218	5.284	5.647	1.152	5.380	5.421	5.442	5.285
2.00	1.252	6.147	8.895	7.559	8.608	9.375	1.269	8.836	8.895	8.910	8.788
2.50	1.179	8.285	13.12	11.69	12.64	13.91	1.433	13.07	13.13	13.15	13.21
3.00	1.151	10.68	17.84	16.24	17.19	18.88	1.574	17.74	17.84	17.82	18.63

Table 1.5: Relative efficiency of different estimators with respect to \bar{X} for $n = 15$

CV	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	$\tilde{\mu}$
0.05	1.000	0.721	1.004	0.005	1.003	1.004	1.000	1.004	1.004	1.004	1.008
0.10	1.000	0.813	1.023	0.020	1.023	1.023	1.000	1.023	1.023	1.023	1.023
0.15	1.000	0.855	1.036	0.043	1.036	1.036	1.000	1.035	1.036	1.036	1.048
0.20	1.001	0.950	1.077	0.078	1.976	1.077	1.001	1.077	1.077	1.077	1.083
0.25	1.001	1.047	1.127	1.127	1.127	1.128	1.000	1.127	1.127	1.128	1.128
0.50	1.008	1.518	1.518	0.493	1.516	1.521	1.002	1.517	1.518	1.520	1.498
0.75	1.084	2.050	2.142	1.148	2.138	2.147	1.005	2.140	2.142	2.145	2.116
1.00	2.956	2.652	2.959	1.949	2.938	2.967	1.011	2.957	2.959	2.962	2.979
1.50	1.090	4.143	5.459	4.463	5.430	5.495	1.021	5.457	5.459	5.466	5.449
2.00	1.039	5.944	8.785	7.788	8.763	8.860	1.041	8.767	8.785	8.778	8.909
2.50	1.030	8.047	13.16	12.18	13.11	13.27	1.054	13.14	13.16	13.16	13.36
3.00	1.025	10.58	18.78	17.80	18.73	18.98	1.086	18.74	18.78	18.76	18.79

Table 1.6: Relative efficiency of different estimators with respect to \bar{X} for $n = 100$

1.3.4 Discussion

In this section we applied the modified signed log-likelihood ratio method to obtain inference for the mean parameter of a normal distribution when the coefficient of variation is known. A by-product of the method is the availability of an efficient point estimator of the mean. Theoretically, this method has rate of convergence $O(n^{-3/2})$ and simulation results show the extreme numerical accuracy of the method even when the sample size is small.

1.4 Inference for bivariate normal correlation coefficient ρ

1.4.1 Background of the correlation coefficient ρ

Interests in the bivariate correlation coefficient can be traced back to 1885 when Sir Francis Galton defined the theoretical concept of the bivariate correlation coefficient. Ten years later, Karl Pearson developed the sample correlation coefficient (sometimes referred to as Pearson's r , see Pearson, 1920), which is a measure of linear correlation (or dependence) between two variables. A recent study by Fosdick and Raftery (2012) reviewed and compared several estimators of the bivariate normal correlation coefficient ρ with the assumption that the means are zero and the variances are known. This problem is of special

interest because, as shown in Fosdick and Raftery (2012), when the means and variances are known, the sample correlation coefficient is not a good estimator of ρ , and the MLE requires solving a cubic polynomial equation and may have multiple roots. Inference for the correlation coefficient in bivariate normal with known variances, especially when the sample size is small, is also an important problem in applied statistics. For example, the United Nations is interested in projecting the total fertility rate (TFR) in all countries. The model was proposed in Alkema et al. (2012) which works well for the projection of an individual country but has not taken into consideration the correlation between different countries. In order to take the correlations between the normalized forecast errors in different countries into account, pairs of these errors are treated as samples from a bivariate normal distribution with means zero and variances equal to one. A necessary and major requirement of the projection is to test whether the correlations between the countries are nonzero. Note that the United Nations TFR data set contains typically five to ten samples for each country. Standard inference methods may not give accurate results because these methods generally require large sample sizes.

A large amount of literature has been devoted to obtain better point estimators for ρ (see Gajjar and Subrahmanian, 1978; Spruill and Gaswirth, 1982; Rodgers and Nicewander,

1988; Barnard et al., 2000; Liechty et al., 2004; Berger and Sun, 2008; Ghosh et al., 2010; and references therein). Fosdick and Raftery (2012) compared the performance of various point estimators. However, much fewer literature on inference for ρ exists, especially when the sample size is small. Fosdick and Raftery (2012) proposed several Bayesian methods, and in particular, using the uniform, arc-sine and Jeffreys priors, to obtain inference for ρ . They concluded that the three priors give similar inference results.

In this section, we applied the modified sign log-likelihood ratio method to obtain the inference for ρ . Simulation results show that the proposed method is extremely accurate even when sample size is extremely small.

1.4.2 The higher order statistic for bivariate normal correlation coefficient

Let $\omega_i = (x_i, y_i)$, $i = 1, 2, \dots, n$ be a random sample drawn from the bivariate normal distribution with a known mean (μ_x, μ_y) and variance-covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \quad (1.16)$$

where σ_x^2 and σ_y^2 are known. We take $\mu_x = \mu_y = 0$ and $\sigma_x^2 = \sigma_y^2 = 1$. Hence, the variance-covariance matrix is

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Denote

$$SS_x = \sum_{i=1}^n x_i^2, \quad SS_y = \sum_{i=1}^n y_i^2, \quad \text{and} \quad SS_{xy} = \sum_{i=1}^n x_i y_i.$$

Inference for ρ is an interesting theoretical problem because the model has only a scalar parameter of interest but has a two-dimensional minimal sufficient statistic $(SS_x + SS_y, SS_{xy})$.

In other words, it belongs to the curved exponential family as defined in Efron (1975) and standard inferential methods cannot be directly applied to obtain inference for ρ . Therefore, we follow the method sketch in Algorithm 1.2.1 for the inference of ρ .

For this problem, the log-likelihood function for ρ can be written as

$$\ell(\rho) = -n \cdot \frac{\log(1 - \rho^2)}{2} - \frac{SS_x + SS_y - 2\rho SS_{xy}}{2(1 - \rho^2)} \quad (1.17)$$

and the score function is

$$\ell_\rho(\rho) = \frac{d\ell(\rho)}{d\rho} = -\frac{n\rho^3 - SS_{xy}\rho^2 + (SS_x + SS_y - n)\rho - SS_{xy}}{(1 - \rho^2)^2}. \quad (1.18)$$

Step 1 of Algorithm 1.2.1 is very tricky here. To obtain the MLE of ρ , denoted as $\hat{\rho}$, we have to find the roots of the score equation $\ell_\rho(\rho) = 0$. In other words, we have to solve for

the roots of a cubic polynomial equation. As a result, we will have at most three real roots.

Here we propose a systematic way to find $\hat{\rho}$ by a detailed analysis of the cubic polynomial equation. The proposed method would significantly reduce the computational cost in simulations.

Proposition 1.4.1. *If $SS_{xy} > 0$, $\hat{\rho}$ is the only root of $\ell_{\rho}(\rho) = 0$ in $(0, 1)$. If $SS_{xy} < 0$, $\hat{\rho}$ is the only root of $\ell_{\rho}(\rho) = 0$ in $(-1, 0)$.*

Proof. From (7), we have

$$\ell_{\rho}(\rho) = -\frac{n\rho^3 - SS_{xy}\rho^2 + (SS_x + SS_y - n)\rho - SS_{xy}}{(1 - \rho^2)^2}$$

Let

$$h(\rho) = n\rho^3 - SS_{xy}\rho^2 + (SS_x + SS_y - n)\rho - SS_{xy}.$$

Then $\hat{\rho}$ is a root of $h(\rho) = 0$. The cubic equation $h(\rho) = 0$ has either three real roots or one real root and two complex conjugate roots. Note that

$$h(-1) = -(SS_x + SS_y + 2SS_{xy}) < 0,$$

$$h(1) = SS_x + SS_y - 2SS_{xy} > 0.$$

By the Intermediate-Value Theorem, $h(\rho)$ must have at least one real root in $(-1, 1)$. Thus, if $h(\rho) = 0$ has one real root and two complex roots, $\hat{\rho}$ must be taken as the only real root

of $h(\rho) = 0$. Note that $h(0) = -SS_{xy}$. We have $\hat{\rho} \in (0, 1)$ when $SS_{xy} > 0$ and $\hat{\rho} \in (-1, 0)$ when $SS_{xy} < 0$, which concurs with the claim in the Proposition.

Next we prove the Proposition is true when $h(\rho) = 0$ has three real roots. Here we assume that the roots of $h(\rho) = 0$ are all distinct. Multiple roots can be easily seen to yield the same results. Let the three distinct roots of $h(\rho) = 0$ be $\rho_1 < \rho_2 < \rho_3$, we have

$$\rho_1 + \rho_2 + \rho_3 = \frac{SS_{xy}}{n}, \quad (1.19)$$

$$\rho_1\rho_2\rho_3 = \frac{SS_{xy}}{n}. \quad (1.20)$$

Note that ρ_2 is a local minimum of $l(\rho)$ and therefore, $\hat{\rho}$ must be chosen between ρ_1 and ρ_3 .

When $SS_{xy} > 0$, $h(0) = -SS_{xy} < 0$, together with (9), we have either $0 < \rho_1 < \rho_2 < \rho_3$ or $\rho_1 < \rho_2 < 0 < \rho_3$. Suppose $0 < \rho_1 < \rho_2 < \rho_3$, then (8) and (9) together give us

$$\begin{aligned} \frac{SS_{xy}}{n} &= \rho_1 + \rho_2 + \rho_3 \\ &> 3\sqrt[3]{\rho_1\rho_2\rho_3} \\ &= 3\sqrt[3]{\frac{SS_{xy}}{n}} \end{aligned}$$

which implies $SS_{xy} > 3\sqrt[3]{n}$. Hence

$$\rho_3 > \sqrt[3]{\rho_1\rho_2\rho_3} = \sqrt[3]{\frac{SS_{xy}}{n}} \geq \sqrt[3]{3} > 1$$

which suggests ρ_3 fails to be a suitable estimator for $\hat{\rho}$ in this case. Since $h(1) > 0$, we have $\rho_2 > 1$ as well. Therefore, $\hat{\rho} = \rho_1$ is the only root of $h(\rho) = 0$ in $(0, 1)$.

Now suppose $\rho_1 < \rho_2 < 0 < \rho_3 < 1$. If $\rho_1 < -1$, it fails to be a proper estimator of $\rho \in (-1, 1)$. If $\rho_1 \in (-1, 0)$, from (1.17) we get

$$\begin{aligned}
\ell(\rho_1) &= -n \frac{\log(1 - \rho_1^2)}{2} + \frac{SS_x + SS_y - 2\rho_1 SS_{xy}}{2(1 - \rho_1^2)} \\
&\leq -n \frac{\log(1 - \rho_1^2)}{2} + \frac{SS_x + SS_y + 2\rho_1 SS_{xy}}{2(1 - \rho_1^2)} \\
&= -n \frac{\log(1 - (-\rho_1)^2)}{2} + \frac{SS_x + SS_y - 2(-\rho_1)SS_{xy}}{2(1 - (-\rho_1)^2)} \\
&= \ell(-\rho_1) \\
&\leq \ell(\rho_3)
\end{aligned}$$

It implies that ρ_3 is the maximum likelihood estimator, which means in this case, $\hat{\rho} = \rho_3$ is again the only root in $(0, 1)$. □

This is a systematic method to obtain $\hat{\rho}$ which could dramatically simplify the computation. The signed log-likelihood ratio statistic for ρ in Step 3 of the algorithm is then

$$r(\rho) = \text{sgn}(\hat{\rho} - \rho)[2(\ell(\hat{\rho}) - \ell(\rho))]^{\frac{1}{2}}$$

and, thus, the $p_{LR}(\rho)$ and confidence interval for ρ can be obtained from the signed log-likelihood ratio statistic.

In order to apply the higher order asymptotic method as discussed in Section 1.2, we need to first obtain the pivotal quantity as in Step 4 of Algorithm 1.2.1. Since we have a bivariate normal model, the pivotal quantity takes the form

$$z(\rho, \omega) = \Sigma^{-1/2}\omega.$$

Note that $\Sigma^{-1/2}$ is not unique. We will try two ways, the singular value decomposition and the Cholesky decomposition, to obtain $\Sigma^{-1/2}$. First, we use the singular value decomposition to obtain $\Sigma^{-1/2}$, which takes the form

$$\Sigma^{-\frac{1}{2}} = \frac{1}{2} \begin{pmatrix} \frac{1}{\sqrt{1+\rho}} + \frac{1}{\sqrt{1-\rho}} & \frac{1}{\sqrt{1+\rho}} - \frac{1}{\sqrt{1-\rho}} \\ \frac{1}{\sqrt{1+\rho}} - \frac{1}{\sqrt{1-\rho}} & \frac{1}{\sqrt{1+\rho}} + \frac{1}{\sqrt{1-\rho}} \end{pmatrix}.$$

Hence, the pivotal quantity is

$$z(\rho, \omega) = \frac{1}{2} \left(\frac{x+y}{\sqrt{1+\rho}} + \frac{x-y}{\sqrt{1-\rho}}, \frac{x+y}{\sqrt{1+\rho}} - \frac{x-y}{\sqrt{1-\rho}} \right)^{\tau}.$$

Since

$$\frac{\partial z(\rho, \omega)}{\partial \omega} = \Sigma^{-\frac{1}{2}}$$

and

$$\frac{\partial z(\rho, \omega)}{\partial \rho} = -\frac{1}{4} \left(\frac{x+y}{(1+\rho)^{\frac{3}{2}}} - \frac{x-y}{(1-\rho)^{\frac{3}{2}}}, \frac{x+y}{(1+\rho)^{\frac{3}{2}}} + \frac{x-y}{(1-\rho)^{\frac{3}{2}}} \right)^{\tau},$$

the ancillary direction V in Step 5 of the algorithm can be calculated as

$$V = -\left(\frac{\partial z(\rho, \omega)}{\partial \omega}\right)^{-1} \left(\frac{\partial z(\rho, \omega)}{\partial \rho}\right) \Big|_{(x, y, \hat{\rho})} = \frac{1}{2(1 - \hat{\rho}^2)}(y - \hat{\rho}x, x - \hat{\rho}y)^\tau.$$

Thus, the locally defined canonical parameter $\varphi(\rho)$ in Step 6 is

$$\begin{aligned} \varphi(\rho) &= \frac{\partial \ell(\rho)}{\partial \omega} \cdot V \\ &= \frac{(SS_x + SS_y)(\rho + \hat{\rho}) - 2SS_{xy}(1 + \rho\hat{\rho})}{2(1 - \hat{\rho}^2)(1 - \rho^2)}. \end{aligned}$$

For this problem, there is no nuisance parameter. Combining Step 7 and Step 8, the standardized maximum likelihood estimate departure calculated in the locally defined canonical parameter scale, $Q(\rho)$, can be simplified to

$$\begin{aligned} Q(\rho) &= \text{sgn}(\hat{\rho} - \rho) |\varphi(\hat{\rho}) - \varphi(\rho)| |j_{\rho\rho}(\hat{\rho})|^{\frac{1}{2}} |\varphi_\rho(\hat{\rho})|^{-1} \\ &= \frac{\hat{\rho} - \rho}{1 - \rho^2} \left| \frac{2SS_{xy}(\rho + 2\hat{\rho} + \rho\hat{\rho}^2) - (SS_x + SS_y)(1 + 2\rho\hat{\rho} + \hat{\rho}^2)}{2SS_{xy}(3\hat{\rho} + \hat{\rho}^3) - (SS_x + SS_y)(1 + 3\hat{\rho}^2)} \right|. \quad (1.21) \\ &\quad \left| 3n\hat{\rho}^2 - 2SS_{xy} \cdot \hat{\rho} + SS_x + SS_y - n \right|^{\frac{1}{2}} \end{aligned}$$

Finally, we can calculate the r^* statistic in Step 9 of Algorithm 1.2.1 and $p_{r^*}(\rho)$ can be obtained with rate of convergence $O(n^{-3/2})$ under the standard normal reference distribution.

The second way we use to obtain $\Sigma^{-1/2}$ is the Cholesky decomposition. The Cholesky decomposition of Σ is

$$\Sigma^{\frac{1}{2}} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1 - \rho^2} \end{pmatrix}.$$

Therefore,

$$\Sigma^{-\frac{1}{2}} = \begin{pmatrix} 1 & 0 \\ -\frac{\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{pmatrix}.$$

The pivotal quantity in Step 4 of Algorithm 1.2.1 is

$$z(\rho, \omega) = \Sigma^{-\frac{1}{2}}(x, y)^\tau = \left(x, \frac{y - \rho x}{\sqrt{1-\rho^2}} \right)^\tau.$$

Hence, we have

$$\frac{\partial z(\rho, \omega)}{\partial \omega} = \Sigma^{-\frac{1}{2}}$$

and

$$\frac{\partial z(\rho, \omega)}{\partial \rho} = \left(0, -\frac{x - \rho y}{(1-\rho^2)^{\frac{3}{2}}} \right)^\tau.$$

Thus the ancillary direction in Algorithm Step 5 is

$$V = -\left(\frac{\partial z(\rho, \omega)}{\partial \omega} \right)^{-1} \left(\frac{\partial z(\rho, \omega)}{\partial \rho} \right) \Big|_{(x, y, \hat{\rho})} = \left(0, \frac{x - \hat{\rho} y}{1 - \hat{\rho}^2} \right)^\tau.$$

Consequently, the local canonical parameter of Step 6 is

$$\varphi(\rho) = \frac{\partial \ell(\rho)}{\partial \omega} \cdot V = \frac{\rho S S_x + \hat{\rho} S S_y - (1 + \rho \hat{\rho}) S S_{xy}}{(1 - \hat{\rho}^2)(1 - \rho^2)}$$

and the standardized maximum likelihood estimate departure in the locally defined canonical

parameter scale $Q(\rho)$ of Algorithm Step 8 can be again calculated using (1.21).

Now we have two ways to decompose Σ : The singular value decomposition and the Cholesky decomposition. These result in two slightly different ways to construct r^* for the inference of ρ . Results from simulation studies show that the two ways of obtaining $\Sigma^{-1/2}$ give almost the same results.

1.4.3 Simulation study

In this section, we first perform simulations to compare the accuracy of the confidence intervals obtained from the signed log-likelihood ratio statistic method (LR), the Bayesian method discussed in Fosdick and Raftery (2012) using the uniform prior (Uniform)

$$\pi_{uniform}(\rho) \propto 1$$

the Jeffreys prior (Jeffreys)

$$\pi_{Jeff}(\rho) \propto \frac{\sqrt{1 + \rho^2}}{1 - \rho^2}$$

and the arc-sine prior (arcsine)

$$\pi_{arcsine}(\rho) = \frac{1}{\pi} \frac{1}{\sqrt{1 - \rho^2}}$$

and the proposed modified signed log-likelihood ratio method (r_S^*) using singular value decomposition and the modified signed log-likelihood ratio method (r_C^*) using Cholesky

decomposition. For each combination of $n = 2, 5, 7, 10, 15, 20$, and $\rho = -0.9$ to 0.9 with a step size of 0.2 , $N = 10,000$ replications are performed. For each generated sample, the 95% confidence interval for ρ is calculated. Results for the lower tail error and upper tail error are recorded in Table 1.7. It is clear that the signed log-likelihood ratio statistic method and the Bayesian method using either Uniform prior or arc-sine prior do not give satisfactory results. The Bayesian method using the Jeffreys prior does not perform well when the sample size is small and ρ is near 0. The proposed method consistently performed better than any of the other methods discussed in this paper even for extremely small sample sizes. Moreover, using either the singular value decomposition or the Cholesky decomposition does not significantly alter the accuracy of the proposed method. Figures 1.1 - 1.6 plotted the average bias of the six methods discussed. The results show that the proposed method, using either the singular value decomposition or the Cholesky decomposition, is consistently around the desired value 0, whereas the other methods are far away from 0.

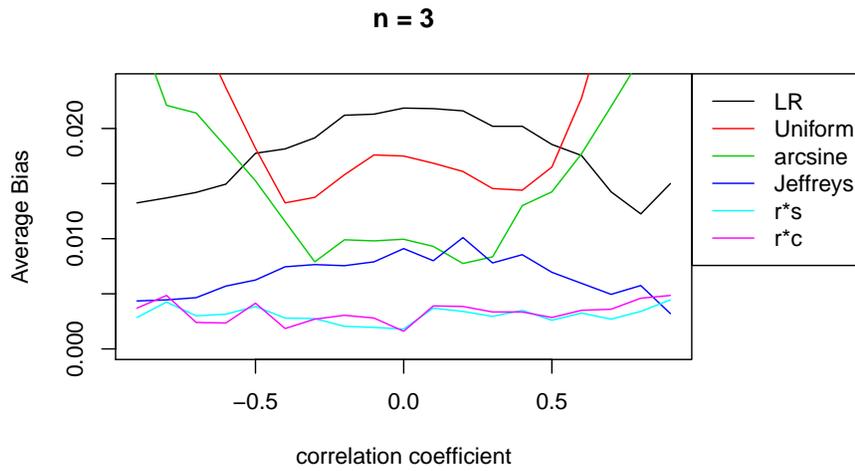


Figure 1.1: Average bias for different methods with sample size $n = 3$

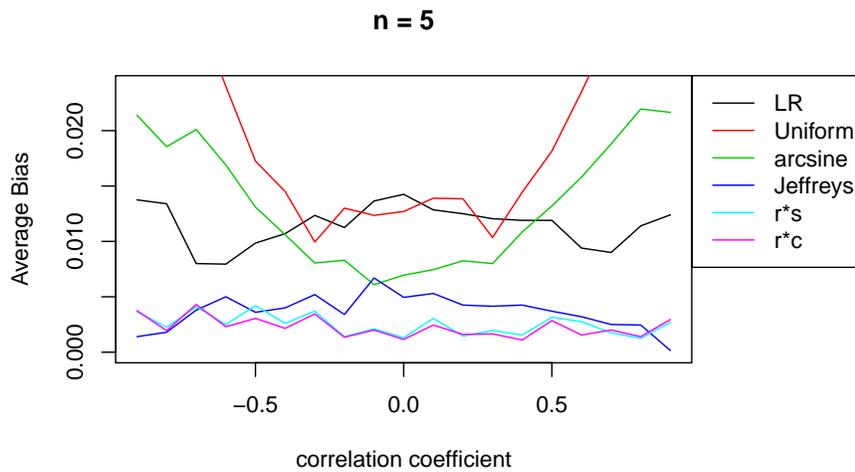


Figure 1.2: Average bias for different methods with sample size $n = 5$

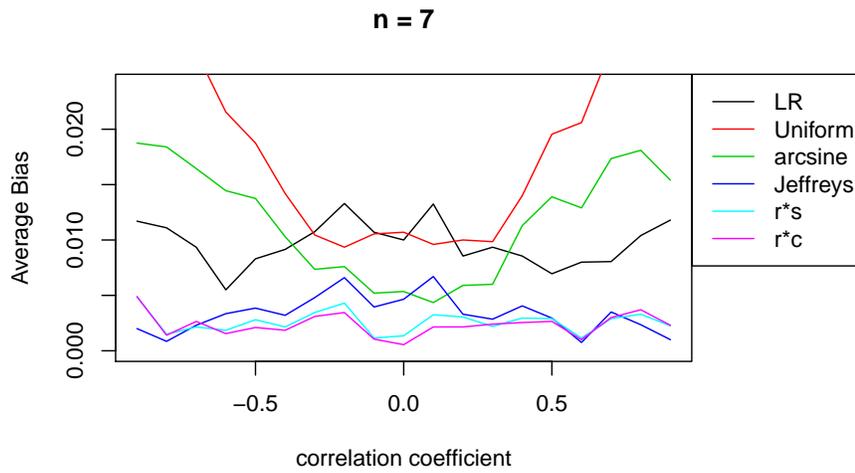


Figure 1.3: Average bias for different methods with sample size $n = 7$

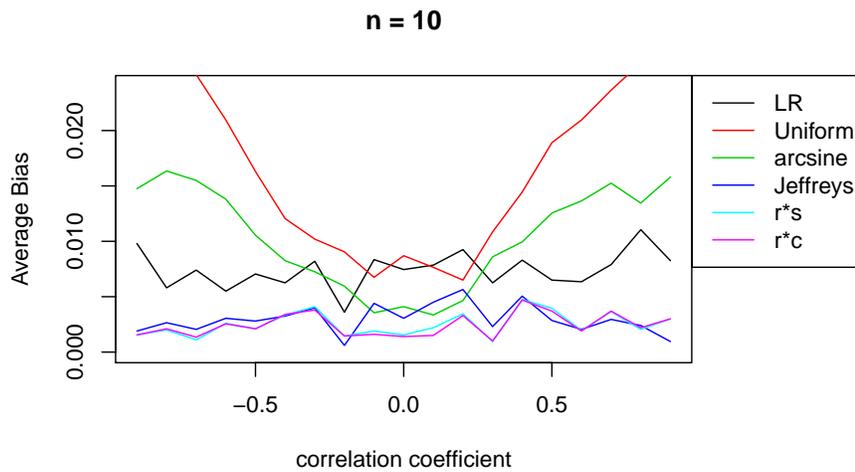


Figure 1.4: Average bias for different methods with sample size $n = 10$

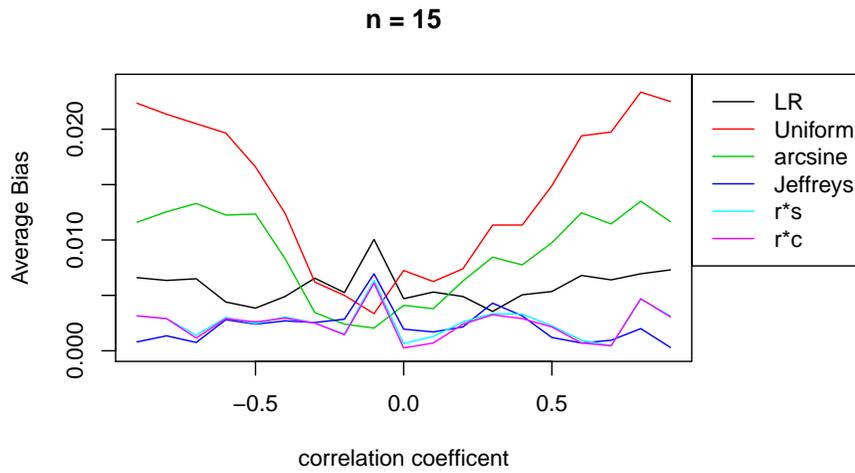


Figure 1.5: Average bias for different methods with sample size $n = 15$

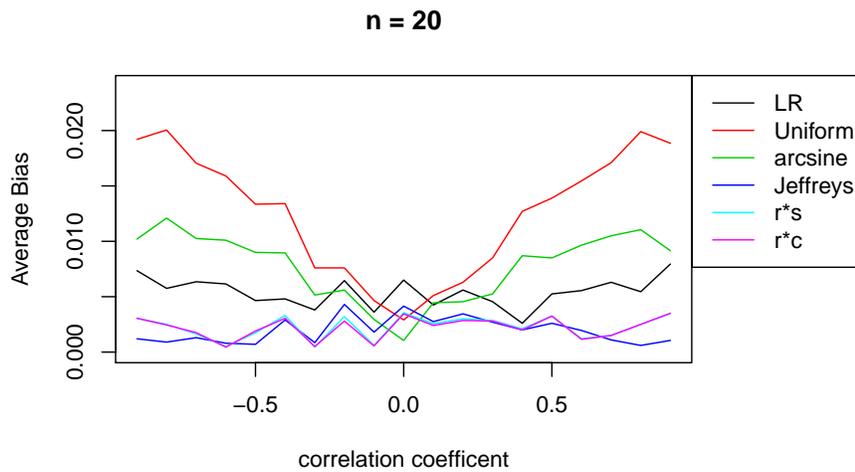


Figure 1.6: Average bias for different methods with sample size $n = 20$

n	Method	$\rho = -0.9$	-0.7	-0.5	-0.3	-0.1	0	0.1	0.3	0.5	0.7	0.9
3	LR	0.0458	0.0474	0.0474	0.0451	0.0486	0.0457	0.0462	0.0418	0.0358	0.0306	0.0177
	Uniform	0.1079	0.0594	0.0369	0.0207	0.0095	0.0077	0.0051	0.0013	0.0005	0.0001	0.0000
	arcsine	0.0607	0.0464	0.0360	0.0257	0.0173	0.0159	0.0130	0.0089	0.0048	0.0031	0.0034
	Jeffreys	0.0316	0.0338	0.0363	0.0359	0.0320	0.0351	0.0308	0.0305	0.0299	0.0245	0.0275
	r_S^*	0.0256	0.0244	0.0248	0.0250	0.0272	0.0241	0.0306	0.0282	0.0287	0.0303	0.0293
	r_C^*	0.0256	0.0250	0.0268	0.0266	0.0290	0.0244	0.0297	0.0274	0.0278	0.0318	0.0298
5	LR	0.0440	0.0409	0.0395	0.0398	0.0388	0.0383	0.0392	0.0339	0.0301	0.0231	0.0166
	Uniform	0.0893	0.0634	0.0376	0.0258	0.0161	0.0128	0.0071	0.0046	0.0031	0.0013	0.0009
	arcsine	0.0504	0.0486	0.0349	0.0283	0.0219	0.0185	0.0146	0.0110	0.0091	0.0068	0.0077
	Jeffreys	0.0262	0.0323	0.0318	0.0319	0.0331	0.0307	0.0289	0.0270	0.0264	0.0246	0.0250
	r_S^*	0.0295	0.0261	0.0242	0.0270	0.0261	0.0256	0.0299	0.0286	0.0290	0.0281	0.0277
	r_C^*	0.0296	0.0263	0.0246	0.0269	0.0266	0.0251	0.0287	0.0282	0.0284	0.0285	0.0282
7	LR	0.0418	0.0395	0.0380	0.0394	0.0393	0.0345	0.0361	0.0304	0.0269	0.0232	0.0168
	Uniform	0.0739	0.0583	0.0423	0.0294	0.0177	0.0145	0.0129	0.0061	0.0048	0.0038	0.0029
	arcsine	0.0472	0.0440	0.0385	0.0303	0.0216	0.0201	0.0179	0.0145	0.0102	0.0103	0.0107
	Jeffreys	0.0265	0.0294	0.0322	0.0318	0.0277	0.0301	0.0320	0.0274	0.0250	0.0252	0.0257
	r_S^*	0.0307	0.0264	0.0246	0.0267	0.0270	0.0262	0.0282	0.0262	0.0300	0.0296	0.0256
	r_C^*	0.0307	0.0266	0.0249	0.0264	0.0270	0.0255	0.0271	0.0264	0.0294	0.0296	0.0258

Table 1.7: The lower tail error

n	Method	$\rho = -0.9$	-0.7	-0.5	-0.3	-0.1	0	0.1	0.3	0.5	0.7	0.9
10	LR	0.0355	0.0352	0.0376	0.0368	0.0339	0.0306	0.0336	0.0284	0.0277	0.0219	0.0184
	Uniform	0.0614	0.0547	0.0402	0.0316	0.0220	0.0172	0.0140	0.0079	0.0072	0.0075	0.0048
	arcsine	0.0396	0.0410	0.0354	0.0310	0.0245	0.0222	0.0196	0.0124	0.0126	0.0132	0.0117
	Jeffreys	0.0226	0.0272	0.0288	0.0306	0.0306	0.0296	0.0284	0.0240	0.0250	0.0271	0.0240
	r_S^*	0.0274	0.0242	0.0272	0.0291	0.0258	0.0244	0.0275	0.0266	0.0328	0.0291	0.0289
	r_C^*	0.0273	0.0240	0.0275	0.0286	0.0253	0.0239	0.0264	0.0267	0.0323	0.0291	0.0289
15	LR	0.0327	0.0335	0.0326	0.0373	0.0377	0.0298	0.0277	0.0283	0.0231	0.0195	0.0190
	Uniform	0.0523	0.0487	0.0425	0.0269	0.0253	0.0176	0.0163	0.0105	0.0088	0.0078	0.0082
	arcsine	0.0374	0.0394	0.0381	0.0261	0.0278	0.0209	0.0199	0.0155	0.0147	0.0143	0.0149
	Jeffreys	0.0248	0.0263	0.0294	0.0248	0.0306	0.0276	0.0279	0.0225	0.0256	0.0236	0.0250
	r_S^*	0.0282	0.0261	0.0248	0.0296	0.0321	0.0255	0.0257	0.0299	0.0282	0.0256	0.0271
	r_C^*	0.0282	0.0258	0.0246	0.0294	0.0318	0.0254	0.0254	0.0294	0.0281	0.0257	0.0269
20	LR	0.0325	0.0332	0.0307	0.0315	0.0299	0.0307	0.0316	0.0253	0.0229	0.0201	0.0189
	Uniform	0.0470	0.0453	0.0363	0.0290	0.0227	0.0227	0.0152	0.0140	0.0111	0.0104	0.0101
	arcsine	0.0358	0.0360	0.0317	0.0279	0.0235	0.0247	0.0186	0.0192	0.0169	0.0154	0.0162
	Jeffreys	0.0236	0.0261	0.0253	0.0257	0.0265	0.0297	0.0244	0.0277	0.0275	0.0241	0.0259
	r_S^*	0.0286	0.0255	0.0238	0.0257	0.0255	0.0278	0.0289	0.0274	0.0285	0.0277	0.0272
	r_C^*	0.0286	0.0255	0.0236	0.0259	0.0258	0.0280	0.0285	0.0274	0.0287	0.0276	0.0272

Table 1.7: The lower tail error (continued)

n	Method	$\rho = -0.9$	-0.7	-0.5	-0.3	-0.1	0	0.1	0.3	0.5	0.7	0.9
3	LR	0.0193	0.0310	0.0381	0.0432	0.0440	0.0480	0.0474	0.0486	0.0513	0.0479	0.0477
	Uniform	0.0000	0.0004	0.0005	0.0018	0.0053	0.0073	0.0112	0.0196	0.0335	0.0638	0.1033
	arcsine	0.0018	0.0036	0.0055	0.0099	0.0131	0.0142	0.0184	0.0256	0.0333	0.0471	0.0593
	Jeffreys	0.0229	0.0245	0.0262	0.0294	0.0338	0.0331	0.0352	0.0351	0.0340	0.0344	0.0289
	r_S^*	0.0301	0.0304	0.0325	0.0305	0.0267	0.0277	0.0268	0.0277	0.0265	0.0249	0.0296
	r_C^*	0.0318	0.0298	0.0315	0.0288	0.0266	0.0276	0.0281	0.0293	0.0279	0.0254	0.0299
5	LR	0.0165	0.0251	0.0302	0.0349	0.0385	0.0402	0.0365	0.0402	0.0437	0.0411	0.0414
	Uniform	0.0008	0.0014	0.0031	0.0059	0.0092	0.0118	0.0151	0.0247	0.0394	0.0597	0.0868
	arcsine	0.0076	0.0084	0.0087	0.0122	0.0159	0.0176	0.0205	0.0270	0.0355	0.0444	0.0510
	Jeffreys	0.0266	0.0247	0.0254	0.0285	0.0303	0.0292	0.0317	0.0313	0.0310	0.0296	0.0253
	r_S^*	0.0278	0.0321	0.0326	0.0304	0.0281	0.0270	0.0238	0.0247	0.0273	0.0254	0.0277
	r_C^*	0.0279	0.0323	0.0307	0.0300	0.0274	0.0272	0.0238	0.0249	0.0273	0.0255	0.0277
7	LR	0.0184	0.0208	0.0286	0.0321	0.0321	0.0355	0.0404	0.0383	0.0370	0.0393	0.0404
	Uniform	0.0029	0.0043	0.0048	0.0085	0.0112	0.0141	0.0179	0.0258	0.0439	0.0584	0.0686
	arcsine	0.0097	0.0111	0.0110	0.0156	0.0180	0.0192	0.0234	0.0265	0.0380	0.0450	0.0415
	Jeffreys	0.0275	0.0252	0.0255	0.0278	0.0302	0.0292	0.0314	0.0283	0.0309	0.0318	0.0237
	r_S^*	0.0290	0.0279	0.0302	0.0302	0.0253	0.0265	0.0283	0.0282	0.0242	0.0262	0.0289
	r_C^*	0.0291	0.0287	0.0291	0.0298	0.0251	0.0256	0.0272	0.0284	0.0241	0.0264	0.0288

Table 1.8: The upper tail error

n	Method	$\rho = -0.9$	-0.7	-0.5	-0.3	-0.1	0	0.1	0.3	0.5	0.7	0.9
10	LR	0.0159	0.0204	0.0235	0.0296	0.0328	0.0343	0.0321	0.0341	0.0353	0.0377	0.0349
	Uniform	0.0048	0.0046	0.0076	0.0112	0.0145	0.0154	0.0207	0.0296	0.0450	0.0548	0.0673
	arcsine	0.0101	0.0100	0.0143	0.0165	0.0184	0.0196	0.0237	0.0296	0.0377	0.0437	0.0433
	Jeffreys	0.0236	0.0231	0.0268	0.0274	0.0282	0.0265	0.0306	0.0286	0.0307	0.0288	0.0259
	r_S^*	0.0257	0.0264	0.0270	0.0291	0.0280	0.0275	0.0269	0.0253	0.0251	0.0282	0.0271
	r_C^*	0.0258	0.0267	0.0267	0.0290	0.0279	0.0267	0.0266	0.0253	0.0251	0.0283	0.0271
15	LR	0.0195	0.0205	0.0251	0.0242	0.0324	0.0296	0.0329	0.0288	0.0338	0.0323	0.0336
	Uniform	0.0076	0.0077	0.0093	0.0145	0.0186	0.0179	0.0212	0.0332	0.0386	0.0473	0.0532
	arcsine	0.0142	0.0128	0.0134	0.0192	0.0237	0.0209	0.0225	0.0324	0.0342	0.0372	0.0382
	Jeffreys	0.0236	0.0248	0.0246	0.0299	0.0333	0.0263	0.0255	0.0311	0.0268	0.0255	0.0244
	r_S^*	0.0281	0.0267	0.0297	0.0246	0.0306	0.0258	0.0269	0.0232	0.0263	0.0253	0.0292
	r_C^*	0.0281	0.0265	0.0298	0.0244	0.0304	0.0249	0.0260	0.0229	0.0262	0.0252	0.0292
20	LR	0.0178	0.0205	0.0214	0.0239	0.0273	0.0323	0.0269	0.0338	0.0334	0.0327	0.0348
	Uniform	0.0086	0.0112	0.0096	0.0138	0.0180	0.0215	0.0254	0.0310	0.0389	0.0446	0.0478
	arcsine	0.0154	0.0155	0.0137	0.0176	0.0206	0.0232	0.0275	0.0297	0.0339	0.0364	0.0345
	Jeffreys	0.0240	0.0235	0.0239	0.0260	0.0271	0.0286	0.0299	0.0277	0.0277	0.0263	0.0238
	r_S^*	0.0275	0.0278	0.0273	0.0252	0.0257	0.0293	0.0238	0.0283	0.0280	0.0247	0.0298
	r_C^*	0.0275	0.0280	0.0274	0.0251	0.0247	0.0289	0.0237	0.0282	0.0278	0.0246	0.0298

Table 1.8: The upper tail error (continued)

1.4.4 Discussion

In this section, the modified signed log-likelihood ratio method has been applied to obtain inference for the correlation coefficient of the bivariate normal distribution with known variances. Although the derived formulas for the method seem complicated, they can be very easily implemented in *R* or any other statistical software. Simulation results from Table 1.7 show that the proposed method outperformed the commonly used Bayesian methods with uniform, arc-sine and Jeffreys priors respectively, especially when the sample size is small.

2 Adjusted Empirical Likelihood Method

In this chapter, we discuss the empirical likelihood (EL) method, which essentially extends the traditional parametric likelihood-based inference method to a non-parametric setting. The EL method has several nice properties: it does not impose prior constraints on region shape, does not require construction of a pivotal statistic, and admits a Bartlett correction which allows low coverage error (Hall & La Scala, 1990). However, the EL method is subject to the convex hull problem, especially when sample size is small. In order to overcome this difficulty, Chen et al. (2008) proposed the adjusted empirical likelihood (AEL) method which adjusts the EL function by adding one “artificial” point created from the observed sample. In this dissertation, we extended the AEL inference to the situation with nuisance parameters.

2.1 Literature review

Empirical likelihood-type method was first used by Thomas & Grunkemeier (1975) to study the survival probabilities estimated by the Kaplan-Meier curve. Art Owen (1988, 1990) formalized empirical likelihood as a unified inference method under more general settings. Let $X_1, X_2, \dots, X_n \in \mathbb{R}^d$ be the independent and identically distributed random vectors following distribution F with mean μ and nonsingular covariance matrix. Their observed values will be denoted by x_1, x_2, \dots, x_n . The empirical likelihood function for the population distribution F is given by

$$L(F) = \prod_{i=1}^n F(\{x_i\}) = \prod_{i=1}^n p_i, \quad (2.1)$$

where $F(\{x_i\})$ is the probability of getting the value x_i in a sample from F and $p_i = Pr(X = x_i)$. Suppose we want to construct a confidence region for the mean μ . The profile EL function of μ is defined to be

$$L_{EL}(\mu) = \sup \left\{ \prod_{i=1}^n p_i : p_i \geq 0, i = 1, \dots, n; \sum_{i=1}^n p_i = 1; \sum_{i=1}^n p_i x_i = \mu \right\}.$$

Hence the corresponding profile empirical log-likelihood function is

$$l_{EL}(\theta) = \log L_{EL}(\mu) = \sup \left\{ \sum_{i=1}^n \log p_i : p_i \geq 0, i = 1, \dots, n; \sum_{i=1}^n p_i = 1; \sum_{i=1}^n p_i x_i = \mu \right\}.$$

Let μ_0 be the true value of μ and $\tilde{\mu} = \arg \sup_{\mu} l_{EL}(\mu)$ be the maximum EL estimator of μ .

Note that $\tilde{\mu}$ equals the sample mean \bar{x}_n . Similar to the parametric Wilks' theorem (Wilks, 1938), Owen (1990) showed that under mild conditions the corresponding EL ratio statistic

$$W_0(\mu_0) = 2[l_{EL}(\tilde{\mu}) - l_{EL}(\mu_0)]$$

converges to χ_d^2 in distribution as the sample size n approaches infinity. This result can be used to construct an approximate $100(1 - \alpha)\%$ confidence region of μ ,

$$I_{EL} = \{\mu : W_0(\mu) \leq \chi_d^2(1 - \alpha)\},$$

where $\chi_d^2(1 - \alpha)$ is the $100(1 - \alpha)\%$ quantile of the χ_d^2 distribution, and α is a pre-specified significance level. .

The EL-based confidence region has several nice properties: it does not impose prior constraints on region shape, is transformation invariant and Bartlett correctable (Hall & La Scala, 1990).

Qin & Lawless (1994) applied the EL to inference for parameters that are generated from estimating equations. For a general p -dimensional parameter $\theta(F)$ associated to random vectors X_1, X_2, \dots, X_n from some unknown d -variate distribution F , Qin & Lawless (1994) linked generalized estimating equations with the defined EL functions. Suppose θ is

associated with F via a vector $g(x, \theta)$ of $r \geq p$ functionally independent unbiased estimating functions. Then for each $j = 1, 2, \dots, r$, we have an estimating equation $E_F\{g_j(x, \theta)\} = 0$, which can be written in the vector form as

$$E_F\{g(x, \theta)\} = 0. \quad (2.2)$$

The profile EL function is defined as

$$L_{EL}(\theta) = \sup \left\{ \prod_{i=1}^n p_i : p_i \geq 0, i = 1, \dots, n; \sum_{i=1}^n p_i = 1; \sum_{i=1}^n p_i g(x_i, \theta) = 0 \right\}. \quad (2.3)$$

and the profile log-EL function is given by

$$l_{EL}(\theta) = \sup \left\{ \sum_{i=1}^n \log p_i : p_i \geq 0, i = 1, \dots, n; \sum_{i=1}^n p_i = 1; \sum_{i=1}^n p_i g(x_i, \theta) = 0 \right\}. \quad (2.4)$$

For example, suppose we have information relating the first and second moments of the unknown distribution F with mean θ , and $E(X^2) = m(\theta)$, where $m(\cdot)$ is a known function. The goal is to estimate θ . The information about F can be expressed by estimating function

$$g(X, \theta) = (X - \theta, X^2 - m(\theta))^T.$$

In this example, we have $r = 2 > p = 1$.

The constrained maximization problem in (2.4) can be solved by applying the method of Lagrangian multipliers. Let λ and $t = (t_1, \dots, t_r)^T$ be Lagrangian multipliers and define

$$H = \sum_i \log p_i + \lambda(1 - \sum_i p_i) - nt^T \sum_i p_i g(x_i, \theta). \quad (2.5)$$

Then maximizing (2.4) is equivalent to maximizing H unconditionally. Setting the first partial derivative of (2.5) with respect to p_i equal to 0, we have

$$\frac{\partial H}{\partial p_i} = \frac{1}{p_i} - \lambda - nt^\tau g(x_i, \theta) = 0,$$

$$\sum_{i=1}^n p_i \frac{\partial H}{\partial p_i} = n - \lambda = 0 \quad \Rightarrow \quad \lambda = n$$

and

$$\hat{p}_i = \frac{1}{n[1 + t^\tau g(x_i, \theta)]},$$

where the vector of Lagrangian multipliers t can be expressed as a function of θ by solving the following equations

$$\sum_{i=1}^n \hat{p}_i g(x_i, \theta) = 0. \quad (2.6)$$

Now the profile EL function (2.3) can be expressed as

$$L_{EL}(\theta) = \prod_{i=1}^n \left\{ \left(\frac{1}{n} \right) \frac{1}{1 + t^\tau g(x_i, \theta)} \right\}, \quad (2.7)$$

and the profile log-EL function becomes

$$l_{EL}(\theta) = - \sum_{i=1}^n \log [1 + t^\tau g(x_i, \theta)] - n \log n. \quad (2.8)$$

Note that (2.6) can be rewritten as

$$\sum_{i=1}^n \frac{g(x_i, \theta)}{1 + t^\tau g(x_i, \theta)} = 0. \quad (2.9)$$

A necessary and sufficient condition for the existence of a solution $\tilde{t} = \tilde{t}(\theta)$ in (2.9) is that 0 must be an inner point of the convex hull expanded by $\{g(x_i, \theta), i = 1, 2, \dots, n\}$. Let $\tilde{\theta}$ be the maximum EL estimator of θ . Suppose the true value of the parameter is θ_0 . Assume the following regularity conditions (Qin & Lawless, 1994):

1. $E[g(x, \theta_0)g(x, \theta_0)^T]$ is positive definite,
2. $\partial g(x, \theta)/\partial \theta$ is continuous in a neighborhood of θ_0 ,
3. $\|\partial g(x, \theta)/\partial \theta\|$ and $\|g(x, \theta)\|^3$ are bounded by some integrable function $G(x)$ in this neighborhood,
4. The rank of $E[\partial g(x, \theta_0)/\partial \theta]$ is p ,
5. $\partial^2 g(x, \theta)/\partial \theta \partial \theta^T$ is continuous in θ in some neighborhood of θ_0 ,
6. $\|\partial^2 g(x, \theta)/\partial \theta \partial \theta^T\|$ can be bounded by some integrable function $H(x)$ in this neighborhood.

Qin & Lawless (1994) further proved that under the above regularity conditions, the EL ratio statistic

$$W_0(\theta_0) = 2[l_{EL}(\tilde{\theta}) - l_{EL}(\theta_0)]$$

converges to χ_p^2 in distribution as the sample size n approaches infinity. One of the results worth noticing is Corollary 5 of Qin & Lawless (1994) where they proved the convergence of the EL ratio statistic to the limiting chi-square distribution in the presence of nuisance parameters.

As pointed out by Chen et al. (2008) and Owen (2001), the true parameter θ_0 is the unique solution of the estimating equations in (2.2). Hence under mild moment conditions, the convex hull of $\{g(x_i, \theta), i = 1, 2, \dots, n\}$ contains 0 as its inner point with probability 1 as $n \rightarrow \infty$. However, if θ is not close to θ_0 or when the sample size n is small, the convex hull is not guaranteed to contain 0. Thus, there is a nonzero probability that the solution to (2.9) does not exist. It results computational issues when solving the constrained optimization problem in the definition of the EL function. This is known as the convex hull problem in the EL literature. In order to overcome this difficulty, Chen et al. (2008) proposed the adjusted empirical likelihood (AEL) method which adjusts the EL function by adding one “artificial” point created from the observed sample. In the following, we will review the general idea of the AEL method.

For convenience, denote

$$g_i = g_i(\theta) = g(x_i, \theta)$$

and

$$\bar{g}_n = \bar{g}_n(\theta) = \frac{1}{n} \sum_{i=1}^n g_i.$$

Let $a_n = o_p(n)$ be a given positive constant. We define a new point g_{n+1} by

$$g_{n+1} = g_{n+1}(\theta) = -\frac{a_n}{n} \sum_{i=1}^n g_i = -a_n \bar{g}_n.$$

Comparing with (2.3), Chen et al. (2008) proposed the profile adjusted empirical log-likelihood ratio function as

$$l_{AEL}(\theta) = \sup \left\{ \sum_{i=1}^{n+1} \log[(n+1)p_i] : p_i \geq 0, i = 1, \dots, n+1; \sum_{i=1}^{n+1} p_i = 1; \sum_{i=1}^{n+1} p_i g_i = 0 \right\}.$$

By (2.8), we have

$$l_{AEL}(\theta) = - \sum_{i=1}^{n+1} \log \left[1 + t^\tau g(x_i, \theta) \right], \quad (2.10)$$

where t is obtained by solving

$$\sum_{i=1}^{n+1} \frac{g(x_i, \theta)}{1 + t^\tau g(x_i, \theta)} = 0. \quad (2.11)$$

The introduction of the additional point does not affect the asymptotic properties of the EL ratio statistic but will guarantee a solution \tilde{t} of (2.9). Under mild regularity conditions, the AEL ratio statistic $W(\theta_0) = 2[l_{AEL}(\tilde{\theta}) - l_{AEL}(\theta_0)]$ converges to χ_p^2 in distribution as the sample size n approaches infinity. Chen et al. (2008) showed that the AEL method improves the coverage probabilities of the original EL method.

Besides the AEL, a number of approaches have been proposed to solve the convex hull problem. Emerson & Owen (2009) proposed the balanced empirical likelihood (BEL). They observed that when the parameter under discussion is the mean μ , the extra sample point in AEL can be expressed as

$$x_{n+1} = \mu - a_n(\bar{x} - \mu).$$

Let $u = \frac{\bar{x} - \mu}{\|\bar{x} - \mu\|}$ and $c_u = (u^T s^{-1} u)^{-\frac{1}{2}}$, where s is the sample covariance matrix. Using the distributional information estimated from the sample, they modified the AEL extra point to

$$x_{n+1} = \mu - a_n c_u u$$

and added a second point

$$x_{n+2} = 2\bar{x} - \mu + a_n c_u u$$

into the sample. Note that x_{n+1} and x_{n+2} are symmetric about \bar{x} as implied by the name BEL. They further proved that the EL ratio statistic $W(\theta_0) = 2[l(\tilde{\theta}) - l(\theta_0)]$ based on the new sample set converges to the corresponding chi-square distribution as the sample size approaches infinity.

Restricting to the inference of the mean, Tsao (2013) and Tsao & Wu (2013) proposed the extended empirical likelihood (EEL) method. Their idea was to transform the ordinary

EL function so that the convex hull is extended to contain the origin as an inner point. To be specific, they modified the original EL function $l(\mu)$ into $l^*(\mu) = l(h_n^{-1}(\mu))$ via a function $h_n(\mu) = \bar{x} + \gamma_n(\mu - \bar{x})$, where γ_n is a constant such that $\gamma_n \geq 0$ and $\gamma_n \rightarrow 1$ as $n \rightarrow \infty$. A suitable choice of γ_n will guarantee a solution for (2.9). Under mild regularity conditions, the EEL ratio statistic $W(\mu_0) = 2[l^*(\tilde{\mu}) - l^*(\mu_0)]$ converges to χ_p^2 in distribution as the sample size n approaches infinity.

Bartolucci (2007) and Lahiri & Mukhopadhyay (2012) applied a penalty to the ordinary EL function. They proposed the penalized empirical likelihood (PEL) method, which is suitable for the inference on the population mean when the dimension of the observations p can grow faster than the sample size n . Denote the j -th component of the observation x_i by x_{ij} . Lahiri & Mukhopadhyay (2012) proposed the PEL function

$$l_{PEL}(\mu) = \sup \left\{ \sum_{i=1}^n \log p_i - \lambda \sum_{j=1}^p \delta_j \left[\sum_{i=1}^n p_i (x_{ij} - \mu_j) \right]^2 : \sum_{i=1}^n p_i = 1 \right\},$$

where $\{\delta_j, 1 \leq j \leq p\}$ are component specific weights and $\lambda > 0$ is the overall penalty factor.

In Lahiri & Mukhopadhyay (2012), the asymptotic distributions of the PEL ratio statistic were derived under different component-wise dependence structures of the observations, namely, the non-Ergodic, the long-range dependence and the short-range dependence structure. Note that the limiting distribution of the PEL ratio statistic is different from the

usual chi-squared limit of the ordinary EL ratio statistics.

The above work of AEL, EEL and PEL all focused on solving the problem when the convex hull of the data does not contain the true parameter. Note that the corresponding EL statistics all converge to the limiting distributions at the rate of $O(n^{-1})$. One of the advantages of the EL inference is that the EL ratio statistic adopts Bartlett corrections. Let $W_0(\theta)$ be the ordinary EL ratio statistic which converges to the chi-square distribution with rate $O(n^{-1})$. Although $W_0(\theta)$ converges to a limiting chi-square distribution, $E(W_0(\theta))$ does not match the expectation of the chi-square distribution at higher orders. Thus, a correction of the form $W_0(\theta)/b$ will correct the expectation of $W_0(\theta)$ to match that of the χ^2 distribution at higher orders, given that b approaches 1 as the sample size increases to infinity. The correction will generally raise the convergence rate of $W_0(\theta)$ to $O(n^{-2})$ in distribution. Related work can be found in Diccio et al. (1991) where the Bartlett correction for smooth functions of means were established. Chen & Cui (2006) gave a summary of the Bartlett correction on the EL ratio statistic when there are nuisance parameters in presence. Chen & Liu (2010) proved that if the adjustment a_n is chosen properly, the convergence rate of the AEL ratio statistic will reach the same order as the Bartlett correction.

2.2 Adjusted empirical likelihood inference for the mean

As an illustration of the empirical likelihood methods, we apply the EL and AEL to the inference of the population mean. Recall that in order to guarantee the existence of a solution for t in (2.9), Chen et al. (2008) proposed to replace the ordinary log-EL function by the log-AEL function $l_{AEL}(\theta)$ given in (2.10) so that t is given by solving (2.11). As long as $a_n = o_p(n)$, the behavior of the AEL function $l_{AEL}(\theta)$ is analogous to the ordinary EL function $l_{EL}(\theta)$, whereas (2.11) always has a solution. Hence the AEL method has resolved the convex hull constraint problem. The AEL preserves all the first-order asymptotic properties of the EL. The following theorem was proved in Theorem 1 of Chen et al. (2008).

Theorem 1 of Chen et al. (2008). *Let θ^T be a $p \times 1$ vector. For the null hypothesis $H_0: \theta = \theta_0$, the adjusted profile EL ratio test statistic is $W(\theta_0) = 2[l_{AEL}(\tilde{\theta}) - l_{AEL}(\theta_0)]$, where $\tilde{\theta}$ maximizes $l_{AEL}(\theta)$. Under H_0 , $W(\theta_0) \xrightarrow{d} \chi_p^2$ as $n \rightarrow \infty$.*

Let μ be the population mean. The estimating function is $X - \mu$. When $a_n = o_p(n^{2/3})$, Chen et al. (2008) showed that the first order asymptotic properties of EL are retained by AEL, and they also found letting $a_n = \max(1, 0.5 \log n)$ is useful in many applications.

We conduct simulation studies to compare the performance of the following methods

with different adjustment levels: (1) the EL method (EL); (2) the AEL method (AEL) with the adjustment level $a_n = \max(1, 0.5 \log n)$ which is suggested by Chen et al. (2008); (3) the AEL method (AEL₁) with $a_n = 0.25 \log n$; (4) the AEL method (AEL₂) with $a_n = \sqrt{n}$. The coverage probabilities are based on 5000 simulations for nominal levels 0.8, 0.9, 0.95, 0.99. The data are drawn from the $N(0, 1)$, t_5 and χ_1^2 distributions, respectively. Simulations are performed under sample size $n = 10, 20, 30, 50$. When the convex hull problem occurs to the EL method for a set of generated data, we set the corresponding $W(\mu_0) = -\infty$ according to the convention. The simulated coverage probabilities for the various cases are shown in Table 2.1.

From Table 2.1, we can see that the simulated coverage probability from all methods excluding the AEL₂ is getting close to the nominal value when the sample size reaches $n = 50$. The AEL method always performs better than other methods regardless of the distribution or the sample size.

Data	Method	$n=10$				$n=20$			
		0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
$N(0, 1)$	EL	0.7392	0.8394	0.8972	0.9580	0.7710	0.8768	0.9326	0.9792
	AEL	0.7996	0.8914	0.9454	0.9990	0.8108	0.9066	0.9528	0.9902
	AEL ₁	0.7672	0.8594	0.9164	0.9696	0.7918	0.8926	0.9426	0.9842
	AEL ₂	1.0000	1.0000	1.0000	1.0000	0.9176	0.9954	1.0000	1.0000
t_5	EL	0.7308	0.8268	0.8806	0.9472	0.7638	0.8632	0.9192	0.9720
	AEL	0.7890	0.8802	0.9372	0.9986	0.8004	0.8948	0.9422	0.9854
	AEL ₁	0.7538	0.8466	0.9018	0.9606	0.7820	0.8790	0.9292	0.9766
	AEL ₂	1.0000	1.0000	1.0000	1.0000	0.9092	0.9958	1.0000	1.0000
χ_1^2	EL	0.7018	0.7978	0.8578	0.9266	0.7370	0.8422	0.8972	0.9568
	AEL	0.7410	0.8320	0.8860	0.9678	0.7720	0.8696	0.9230	0.9668
	AEL ₁	0.7066	0.800	0.8582	0.9296	0.7540	0.8560	0.9076	0.9594
	AEL ₂	1.0000	1.0000	1.0000	1.0000	0.8804	0.9672	1.0000	1.0000

Table 2.1: Coverage probabilities of the population mean

Data	Method	$n=30$				$n=50$			
		0.80	0.90	0.95	0.99	0.80	0.90	0.95	0.99
$N(0, 1)$	EL	0.7830	0.8822	0.9382	0.9870	0.7926	0.8940	0.9436	0.9878
	AEL	0.8102	0.9068	0.9524	0.9920	0.8082	0.9102	0.9594	0.9916
	AEL ₁	0.7962	0.8916	0.9464	0.9888	0.8014	0.9008	0.9482	0.9898
	AEL ₂	0.8852	0.9714	0.9986	1.0000	0.8678	0.9498	0.9856	1.0000
t_5	EL	0.7706	0.8766	0.9326	0.9844	0.7778	0.8790	0.9352	0.9886
	AEL	0.8008	0.8970	0.9501	0.9894	0.7978	0.8972	0.9502	0.9912
	AEL ₁	0.7868	0.8864	0.9410	0.9856	0.7878	0.8878	0.9424	0.9900
	AEL ₂	0.8762	0.9688	0.9976	1.0000	0.8666	0.9486	0.9862	1.0000
χ_1^2	EL	0.7670	0.8596	0.9194	0.9696	0.7756	0.8796	0.9344	0.9810
	AEL	0.7910	0.8846	0.9352	0.9768	0.7924	0.8920	0.9446	0.9860
	AEL ₁	0.7786	0.8714	0.9274	0.9732	0.7828	0.8858	0.9398	0.9830
	AEL ₂	0.8676	0.9478	0.9838	1.0000	0.8538	0.9410	0.9770	0.9994

Table 2.1: Coverage probabilities of the population mean (continued)

2.3 Adjusted empirical likelihood with nuisance parameters

In this section, we discuss some mathematical details of the AEL method. Chen et al. (2008) discussed the AEL-based inference for arbitrary p -dimensional parameters without nuisance parameters. Building upon Chen et al. (2008) and Qin & Lawless (1994), the focus of this chapter is to investigate the asymptotic properties of the AEL when nuisance parameters present. To be specific, suppose the p -dimensional parameter $\theta = (\theta_1, \theta_2)$ consists a q -dimensional parameter of interest θ_1 as well as a $(p - q)$ -dimensional nuisance parameter θ_2 . The goal is to test the null hypothesis $H_0 : \theta_1 = \theta_1^0$ for some given θ_1^0 , then the results of Chen et al. (2008) have to be reconstructed before applied to the inference with nuisance parameters. Consequently, when there are nuisance parameters in presence, the corresponding asymptotic property is needed for the likelihood ratio test, which will be given in Theorem 2.3.1 of this section. First, we develop a lemma about positive definite matrices. If a matrix M is positive semidefinite, we denote it by $M \geq 0$; if M is positive definite, we write $M > 0$. For any matrices G and H , let $G \geq H$ denote that $G - H$ is positive semidefinite, and let $G > H$ denote that $G - H$ is positive definite.

Lemma 2.3.1. *Let M be a $p \times p$ symmetric positive definite block matrix of the form*

$$M = \begin{pmatrix} A & B \\ B^\tau & C \end{pmatrix},$$

where A is a $q \times q$ matrix, B is a $q \times (p - q)$ matrix, and C is a $(p - q) \times (p - q)$ matrix.

Then C is positive definite and

$$\begin{pmatrix} A & B \\ B^\tau & C \end{pmatrix}^{-1} \geq \begin{pmatrix} 0 & 0 \\ 0 & C^{-1} \end{pmatrix}.$$

Proof. Since M is a symmetric positive matrix, we have

$$C > 0 \quad \text{and} \quad A - BC^{-1}B^\tau > 0.$$

See Theorem 16.1 in Gallier (2011). Noting that M has the following factorization

$$\begin{pmatrix} A & B \\ B^\tau & C \end{pmatrix} = \begin{pmatrix} I & BC^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A - BC^{-1}B^\tau & 0 \\ 0 & C \end{pmatrix} \begin{pmatrix} I & 0 \\ (BC^{-1})^\tau & I \end{pmatrix},$$

we have

$$\begin{pmatrix} A & B \\ B^\tau & C \end{pmatrix}^{-1} = \begin{pmatrix} I & 0 \\ (BC^{-1})^\tau & I \end{pmatrix}^{-1} \begin{pmatrix} (A - BC^{-1}B^\tau)^{-1} & 0 \\ 0 & C^{-1} \end{pmatrix} \begin{pmatrix} I & BC^{-1} \\ 0 & I \end{pmatrix}^{-1}.$$

Further note that

$$\begin{aligned}
& \begin{pmatrix} 0 & 0 \\ 0 & C^{-1} \end{pmatrix} \\
&= \begin{pmatrix} I & 0 \\ (BC^{-1})^\tau & I \end{pmatrix}^{-1} \begin{pmatrix} I & 0 \\ (BC^{-1})^\tau & I \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & C^{-1} \end{pmatrix} \begin{pmatrix} I & BC^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} I & BC^{-1} \\ 0 & I \end{pmatrix}^{-1} \\
&= \begin{pmatrix} I & 0 \\ (BC^{-1})^\tau & I \end{pmatrix}^{-1} \begin{pmatrix} 0 & 0 \\ 0 & C^{-1} \end{pmatrix} \begin{pmatrix} I & BC^{-1} \\ 0 & I \end{pmatrix}^{-1}.
\end{aligned}$$

Above two factorizations lead to

$$\begin{aligned}
& \begin{pmatrix} A & B \\ B^\tau & C \end{pmatrix}^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & C^{-1} \end{pmatrix} \\
&= \begin{pmatrix} I & 0 \\ (BC^{-1})^\tau & I \end{pmatrix}^{-1} \left[\begin{pmatrix} (A - BC^{-1}B^\tau)^{-1} & 0 \\ 0 & C^{-1} \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & C^{-1} \end{pmatrix} \right] \begin{pmatrix} I & BC^{-1} \\ 0 & I \end{pmatrix}^{-1} \\
&= \begin{pmatrix} I & 0 \\ (BC^{-1})^\tau & I \end{pmatrix}^{-1} \begin{pmatrix} (A - BC^{-1}B^\tau)^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} I & BC^{-1} \\ 0 & I \end{pmatrix}^{-1}.
\end{aligned}$$

Since $A - BC^{-1}B^\tau > 0$, we have

$$(A - BC^{-1}B^\tau)^{-1} > 0,$$

which leads to

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix}^{-1} \geq \begin{pmatrix} 0 & 0 \\ 0 & C^{-1} \end{pmatrix}.$$

□

In order to prove the main theorem, we also need the following two results about idempotent matrices. The proof of these two results can be found in Rao (1973, p186-187).

Result 2.3.1. *Let Y be a random vector with independent entries that identically follows standard normal distribution. A necessary and sufficient condition that $Y'AY$ has a χ^2 distribution is that A is idempotent, that is, $A^2 = A$, in which case the degrees of freedom of χ^2 is*

$$v = \text{rank}(A) = \text{trace}(A).$$

Result 2.3.2. *If A , B , $A - B$ are matrices of non-negative quadric forms and A and B are idempotent, then $A - B$ is also idempotent.*

Based on Lemma 2.3.1 and the above two results, we have the following theorem which gives the asymptotic properties of the AEL ratio test statistic. The theorem is a nonparametric analogue of the theorem in Wilks (1938) on the asymptotic distribution of the likelihood ratio. The difference is that Wilks' theorem is based on parametric likelihood

and ours is based on the nonparametric adjusted empirical likelihood. Moreover, it takes into consideration of nuisance parameters.

Theorem 2.3.1. *Let $\theta^\tau = (\theta_1, \theta_2)^\tau$, where θ_1 and θ_2 are $q \times 1$ and $(p - q) \times 1$ vectors, respectively. For $H_0 : \theta_1 = \theta_1^0$, the profile AEL ratio test statistic is*

$$W(\theta_1^0) = 2[l_{AEL}(\tilde{\theta}_1, \tilde{\theta}_2) - l_{AEL}(\theta_1^0, \tilde{\theta}_2^0)],$$

where $\tilde{\theta}^\tau = (\tilde{\theta}_1, \tilde{\theta}_2)^\tau$ maximizes $l_{AEL}(\theta) = l_{AEL}(\theta_1, \theta_2)$, and $\tilde{\theta}_2^0$ maximizes $l_{AEL}(\theta_1^0, \theta_2)$ with respect to θ_2 . Under H_0 , we have

$$W(\theta_1^0) \xrightarrow{d} \chi_q^2$$

as $n \rightarrow \infty$.

Proof. For simplicity, denote $l(\theta) = -l_{AEL}(\theta)$. Then $\tilde{\theta}^\tau = (\tilde{\theta}_1, \tilde{\theta}_2)^\tau$ minimizes $l(\theta) = l(\theta_1, \theta_2)$, and $\tilde{\theta}_2^0$ minimizes $l(\theta_1^0, \theta_2)$ with respect to θ_2 . Under this new notation, the test statistic becomes

$$W(\theta_1^0) = 2[l(\theta_1^0, \tilde{\theta}_2^0) - l(\tilde{\theta}_1, \tilde{\theta}_2)].$$

First, the following notations are needed in this proof. Let

$$Q_{1n}(\theta, t) = \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{g_i(\theta)}{1 + t^\tau g_i(\theta)},$$

$$Q_{2n}(\theta, t) = \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{1}{1+t^\tau g_i(\theta)} \left(\frac{\partial g_i(\theta)}{\partial \theta} \right)^\tau t.$$

Let $\tilde{\theta}$ and $\tilde{t} = t(\tilde{\theta})$ be the solution of

$$Q_{1n}(\tilde{\theta}, \tilde{t}) = 0, \quad Q_{2n}(\tilde{\theta}, \tilde{t}) = 0.$$

The existence of $\tilde{\theta}$ and $\tilde{t} = t(\tilde{\theta})$ in a neighborhood of the true parameter θ_0 is proved in Chen et. al (2008) and Qin & Lawless (1994). Note that

$$\begin{aligned} \frac{\partial Q_{1n}(\theta, 0)}{\partial \theta} &= \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{\partial g_i(\theta)}{\partial \theta}, & \frac{\partial Q_{1n}(\theta, 0)}{\partial t^\tau} &= -\frac{1}{n+1} \sum_{i=1}^{n+1} g_i(\theta) g_i(\theta)^\tau, \\ \frac{\partial Q_{2n}(\theta, 0)}{\partial \theta} &= 0, & \frac{\partial Q_{2n}(\theta, 0)}{\partial t^\tau} &= \frac{1}{n+1} \sum_{i=1}^{n+1} \left(\frac{\partial g_i(\theta)}{\partial \theta} \right)^\tau. \end{aligned}$$

Taylor expansion of $Q_{1n}(\tilde{\theta}, \tilde{t})$ and $Q_{2n}(\tilde{\theta}, \tilde{t})$ at $(\theta_0, 0)$ gives

$$\begin{aligned} 0 &= Q_{1n}(\tilde{\theta}, \tilde{t}) \\ &= Q_{1n}(\theta_0, 0) + \frac{\partial Q_{1n}(\theta_0, 0)}{\partial \theta} (\tilde{\theta} - \theta_0) + \frac{\partial Q_{1n}(\theta_0, 0)}{\partial t^\tau} (\tilde{t} - 0) + o_p(\delta_n) \end{aligned}$$

$$\begin{aligned} 0 &= Q_{2n}(\tilde{\theta}, \tilde{t}) \\ &= Q_{2n}(\theta_0, 0) + \frac{\partial Q_{2n}(\theta_0, 0)}{\partial \theta} (\tilde{\theta} - \theta_0) + \frac{\partial Q_{2n}(\theta_0, 0)}{\partial t^\tau} (\tilde{t} - 0) + o_p(\delta_n), \end{aligned}$$

where $\delta_n = \|\tilde{\theta} - \theta_0\| + \|\tilde{t}\|$. Observing that $Q_{2n}(\theta_0, 0) = 0$, we have

$$S_n \begin{pmatrix} \tilde{t} \\ \tilde{\theta} - \theta_0 \end{pmatrix} = \begin{pmatrix} -Q_{1n}(\theta_0, 0) + o_p(\delta_n) \\ o_p(\delta_n) \end{pmatrix}, \quad (2.12)$$

where

$$S_n = \left(\begin{array}{cc} \frac{\partial Q_{1n}}{\partial t^\tau} & \frac{\partial Q_{1n}}{\partial \theta} \\ \frac{\partial Q_{2n}}{\partial t^\tau} & 0 \end{array} \right) \Bigg|_{(\theta_0, 0)}.$$

Now we solve (2.12) for an expression of \tilde{t} . By the law of large numbers, as $n \rightarrow \infty$

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial g_i(\theta)}{\partial \theta} \rightarrow E\left(\frac{\partial g(\theta)}{\partial \theta}\right).$$

Therefore,

$$\frac{\partial g_{n+1}(\theta)}{\partial \theta} = -\frac{a_n}{n} \sum_{i=1}^n \frac{\partial g_i(\theta)}{\partial \theta} = o_p(n).$$

Hence applying the law of large numbers again

$$\begin{aligned} \frac{\partial Q_{1n}(\theta, 0)}{\partial \theta} &= \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{\partial g_i(\theta)}{\partial \theta} \\ &= \frac{1}{n+1} \sum_{i=1}^n \frac{\partial g_i(\theta)}{\partial \theta} + \frac{1}{n+1} \frac{\partial g_{n+1}(\theta)}{\partial \theta} \\ &= E\left(\frac{\partial g(\theta)}{\partial \theta}\right) + o_p(1). \end{aligned}$$

Similarly, we can obtain

$$\frac{\partial Q_{2n}}{\partial t^\tau} = E \frac{\partial g(\theta)^\tau}{\partial \theta} + o_p(1) \quad \text{and} \quad -\frac{\partial Q_{1n}}{\partial t^\tau} = E g(\theta) g(\theta)^\tau + o_p(1).$$

Thus as $n \rightarrow \infty$

$$S_n \rightarrow \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & 0 \end{pmatrix} = \begin{pmatrix} -Egg^\tau & E\frac{\partial g}{\partial \theta} \\ E\frac{\partial g^\tau}{\partial \theta} & 0 \end{pmatrix} \Big|_{\theta=\theta_0}.$$

We can see that

$$S_n^{-1} \rightarrow \begin{pmatrix} S_{11}^{-1} + S_{11}^{-1}S_{12}S_{22.1}^{-1}S_{21}S_{11}^{-1} & -S_{11}^{-1}S_{12}S_{22.1}^{-1} \\ -S_{22.1}^{-1}S_{21}S_{11}^{-1} & S_{22.1}^{-1} \end{pmatrix},$$

where $S_{22.1}^{-1} = \left[\left(E\frac{\partial g}{\partial \theta} \right)^\tau (Egg^\tau)^{-1} \left(E\frac{\partial g}{\partial \theta} \right) \right]^{-1}$. Consequently, (2.12) can be solved as

$$\begin{pmatrix} \tilde{t} \\ \tilde{\theta} - \theta_0 \end{pmatrix} = S_n^{-1} \begin{pmatrix} -Q_{1n}(\theta_0, 0) + o_p(\delta_n) \\ o_p(\delta_n) \end{pmatrix},$$

which means

$$\tilde{\theta} - \theta_0 = S_{22.1}^{-1}S_{21}S_{11}^{-1}Q_{1n}(\theta_0, 0) + o_p(\delta_n) \tag{2.13}$$

$$\tilde{t} = -(S_{11}^{-1} + S_{11}^{-1}S_{12}S_{22.1}^{-1}S_{21}S_{11}^{-1})Q_{1n}(\theta_0, 0) + o_p(\delta_n).$$

Note that by Central Limit Theorem

$$\begin{aligned} Q_{1n}(\theta_0, 0) &= \frac{1}{n+1} \sum_{i=1}^{n+1} g_i(\theta_0) \\ &= \frac{n^{\frac{1}{2}}}{n+1} \cdot n^{-\frac{1}{2}} \sum_{i=1}^n g_i(\theta_0) - n^{-\frac{1}{2}} \cdot \frac{a_n}{n+1} \cdot n^{-\frac{1}{2}} \sum_{i=1}^n g_i(\theta_0) \\ &= n^{-\frac{1}{2}} \cdot n^{-\frac{1}{2}} \sum_{i=1}^n g_i(\theta_0) + o_p(n^{-\frac{1}{2}}), \end{aligned}$$

which implies

$$\sqrt{n}Q_{1n}(\theta_0, 0) \longrightarrow N(0, E g g^\tau) \quad \text{and} \quad Q_{1n} = O_p(n^{-\frac{1}{2}}). \quad (2.14)$$

From (2.13), we know that

$$\delta_n = \|\tilde{\theta} - \theta_0\| + \|\tilde{t}\| = O_p(n^{-\frac{1}{2}}).$$

Therefore, we have obtained the desired result

$$\tilde{t} = -(S_{11}^{-1} + S_{11}^{-1}S_{12}S_{22.1}^{-1}S_{21}S_{11}^{-1})Q_{1n}(\theta_0, 0) + o_p(n^{-\frac{1}{2}}) \quad (2.15)$$

and

$$\tilde{\theta} - \theta_0 = S_{22.1}^{-1}S_{21}S_{11}^{-1}Q_{1n}(\theta_0, 0) + o_p(n^{-\frac{1}{2}}).$$

In particular, we can see that

$$\tilde{t} = O_p(n^{-\frac{1}{2}}) \quad \text{and} \quad \tilde{\theta} - \theta_0 = O_p(n^{-\frac{1}{2}}).$$

Now we are ready to compute $l(\tilde{\theta}) = l(\tilde{\theta}_1, \tilde{\theta}_2)$. Taylor expansion yields

$$\begin{aligned} l(\tilde{\theta}_1, \tilde{\theta}_2) &= \sum_{i=1}^{n+1} \log[1 + \tilde{t}^\tau g_i(\tilde{\theta})] \\ &= \sum_{i=1}^{n+1} \left(\tilde{t}^\tau g_i(\tilde{\theta}) - \frac{1}{2} (\tilde{t}^\tau g_i(\tilde{\theta}))^2 \right) + o_p(1) \\ &= \tilde{t}^\tau \sum_{i=1}^{n+1} g_i(\tilde{\theta}) - \frac{1}{2} \tilde{t}^\tau \left(\sum_{i=1}^{n+1} g_i(\tilde{\theta}) g_i(\tilde{\theta})^\tau \right) \tilde{t} + o_p(1). \end{aligned} \quad (2.16)$$

Note that expanding $g_i(\tilde{\theta})$ at θ_0 , we get

$$g_i(\tilde{\theta}) = g_i(\theta_0) + \frac{\partial g_i(\theta_0)}{\partial \theta}(\tilde{\theta} - \theta_0) + O_p(n^{-1}),$$

for $i = 1, 2, \dots, n$. Hence

$$\begin{aligned} \sum_{i=1}^n g_i(\tilde{\theta}) &= \sum_{i=1}^n g_i(\theta_0) + \sum_{i=1}^n \frac{\partial g_i(\theta_0)}{\partial \theta}(\tilde{\theta} - \theta_0) + O_p(1) \\ &= nQ_{1n}(\theta_0, 0) + nS_{12}S_{22.1}^{-1}S_{21}S_{11}^{-1}Q_{1n}(\theta_0, 0) + o_p(n^{\frac{1}{2}}) \end{aligned}$$

and

$$g_{n+1}(\tilde{\theta}) = -\frac{a_n}{n} \sum_{i=1}^n g_i(\tilde{\theta}) = o_p(n^{\frac{1}{2}}).$$

Consequently, we can obtain the first term of (2.16) as

$$\tilde{\tau}^\tau \sum_{i=1}^{n+1} g_i(\tilde{\theta}) = -n Q_{1n}(\theta_0, 0)^\tau (S_{11}^{-1} + S_{11}^{-1}S_{12}S_{22.1}^{-1}S_{21}S_{11}^{-1})Q_{1n}(\theta_0, 0) + o_p(1).$$

Now we calculate the second term of (2.16). For $i = 1, 2, \dots, n$,

$$g_i(\tilde{\theta})g_i(\tilde{\theta})^\tau = g_i(\theta_0)g_i(\theta_0)^\tau + O_p(n^{-\frac{1}{2}}).$$

Thus

$$\sum_{i=1}^n g_i(\tilde{\theta})g_i(\tilde{\theta})^\tau = \sum_{i=1}^n g_i(\theta_0)g_i(\theta_0)^\tau + O_p(n^{\frac{1}{2}}) = -nS_{11} + O_p(n^{\frac{1}{2}}).$$

Note that

$$g_{n+1}(\tilde{\theta})g_{n+1}(\tilde{\theta})^\tau = o_p(n^{\frac{1}{2}})o_p(n^{\frac{1}{2}}) = o_p(n).$$

We have

$$\tilde{t}^\tau \left(\sum_{i=1}^{n+1} g_i(\tilde{\theta}) g_i(\tilde{\theta})^\tau \right) \tilde{t} = -n Q_{1n}(\theta_0, 0)^\tau (S_{11}^{-1} + S_{11}^{-1} S_{12} S_{22.1}^{-1} S_{21} S_{11}^{-1}) Q_{1n}(\theta_0, 0) + o_p(1).$$

Finally, we have

$$l(\tilde{\theta}_1, \tilde{\theta}_2) = -\frac{n}{2} Q_{1n}(\theta_0, 0)^\tau (S_{11}^{-1} + S_{11}^{-1} S_{12} S_{22.1}^{-1} S_{21} S_{11}^{-1}) Q_{1n}(\theta_0, 0) + o_p(1). \quad (2.17)$$

Similarly, we can apply the above process to $l(\theta_1^0, \tilde{\theta}_2^0)$. The procedures are sketched as

follows. Let $\tilde{\theta}_2^0$ and $\tilde{t}_0 = t(\theta_1^0, \tilde{\theta}_2^0)$ satisfy

$$Q_{1n}(\theta_1^0, \tilde{\theta}_2^0, \tilde{t}_0) = 0 \quad \text{and} \quad Q_{2n}(\theta_1^0, \tilde{\theta}_2^0, \tilde{t}_0) = 0.$$

Expanding Q_{1n} and Q_{2n} at $(\theta_1^0, \theta_2^0, 0)$ will produce the linear equations

$$H_n \begin{pmatrix} \tilde{t}_0 \\ \tilde{\theta}_2^0 - \theta_2^0 \end{pmatrix} = \begin{pmatrix} -Q_{1n}(\theta_0, 0) + o_p(\delta'_n) \\ o_p(\delta'_n) \end{pmatrix}, \quad (2.18)$$

where $\theta_0 = (\theta_1^0, \theta_2^0)$ is the true value of θ , $\delta'_n = \|\tilde{\theta}_2^0 - \theta_2^0\| + \|\tilde{t}_0\|$ and as $n \rightarrow \infty$

$$H_n \longrightarrow \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & 0 \end{pmatrix} = \begin{pmatrix} -E g g^\tau & E \frac{\partial g}{\partial \theta_2} \\ E \frac{\partial g}{\partial \theta_2}^\tau & 0 \end{pmatrix} \Bigg|_{\theta=\theta_0}.$$

Note that $H_{11} = S_{11}$. Solving (2.18) gives us

$$\tilde{t}_0 = -(H_{11}^{-1} + H_{11}^{-1} H_{12} H_{22.1}^{-1} H_{21} H_{11}^{-1}) Q_{1n}(\theta_0, 0) + o_p(n^{-\frac{1}{2}}) \quad (2.19)$$

and

$$\tilde{\theta}_2^0 - \theta_2^0 = H_{22,1}^{-1} H_{21} H_{11}^{-1} Q_{1n}(\theta_0, 0) + o_p(n^{-\frac{1}{2}}).$$

By Taylor expansion, the above estimations yield

$$l(\theta_1^0, \tilde{\theta}_2^0) = -\frac{1}{2} n Q_{1n}(\theta_0, 0)^\tau (H_{11}^{-1} + H_{11}^{-1} H_{12} H_{22,1}^{-1} H_{21} H_{11}^{-1}) Q_{1n}(\theta_0, 0) + o_p(1). \quad (2.20)$$

Using (2.20) and (2.17), we can write

$$\begin{aligned} W(\theta_1^0) &= 2l(\theta_1^0, \tilde{\theta}_2^0) - 2l(\tilde{\theta}_1, \tilde{\theta}_2) \\ &= [(Egg^\tau)^{-\frac{1}{2}} \sqrt{n} Q_{1n}(\theta_0, 0)]^\tau (A - B) [(Egg^\tau)^{-\frac{1}{2}} \sqrt{n} Q_{1n}(\theta_0, 0)] + o_p(1), \end{aligned}$$

where

$$\begin{aligned} A &= (Egg^\tau)^{-\frac{1}{2}} \left(E \frac{\partial g}{\partial \theta} \right) \left[\left(E \frac{\partial g}{\partial \theta} \right)^\tau (Egg^\tau)^{-1} \left(E \frac{\partial g}{\partial \theta} \right) \right]^{-1} \left(E \frac{\partial g}{\partial \theta} \right)^\tau (Egg^\tau)^{-\frac{1}{2}} \\ B &= (Egg^\tau)^{-\frac{1}{2}} \left(E \frac{\partial g}{\partial \theta_2} \right) \left[\left(E \frac{\partial g}{\partial \theta_2} \right)^\tau (Egg^\tau)^{-1} \left(E \frac{\partial g}{\partial \theta_2} \right) \right]^{-1} \left(E \frac{\partial g}{\partial \theta_2} \right)^\tau (Egg^\tau)^{-\frac{1}{2}} \end{aligned}$$

and all the evaluations related to g are performed at the true value θ_0 . By assumption, $E \frac{\partial g}{\partial \theta}$

has rank k and Egg^τ is positive definite. Therefore, both A and B are non-negative definite and

idempotent. By Lemma 2.3.1

$$\begin{aligned}
& \left(E \frac{\partial g}{\partial \theta}\right) \left[\left(E \frac{\partial g}{\partial \theta}\right)^\tau (E g g^\tau)^{-1} \left(E \frac{\partial g}{\partial \theta}\right) \right]^{-1} \left(E \frac{\partial g}{\partial \theta}\right)^\tau \\
& \geq \left(E \frac{\partial g}{\partial \theta_1}, E \frac{\partial g}{\partial \theta_2}\right) \begin{pmatrix} 0 & 0 \\ 0 & \left[\left(E \frac{\partial g}{\partial \theta_2}\right)^\tau (E g g^\tau)^{-1} \left(E \frac{\partial g}{\partial \theta_2}\right) \right]^{-1} \end{pmatrix} \begin{pmatrix} E \frac{\partial g}{\partial \theta_1}^\tau \\ E \frac{\partial g}{\partial \theta_2}^\tau \end{pmatrix} \\
& = \left(E \frac{\partial g}{\partial \theta_2}\right) \left[\left(E \frac{\partial g}{\partial \theta_2}\right)^\tau (E g g^\tau)^{-1} \left(E \frac{\partial g}{\partial \theta_2}\right) \right]^{-1} \left(E \frac{\partial g}{\partial \theta_2}\right)^\tau,
\end{aligned}$$

which means that $A - B$ is non-negative definite. Thus by Result 2.3.2, $A - B$ is also idempotent.

From (2.14), we can see that $(E g g^\tau)^{-\frac{1}{2}} \sqrt{n} Q_{ln}(\theta_0, 0)$ follows the multivariate standard normal distribution asymptotically. Note that $tr(A) = p$ and $tr(B) = p - q$. We have

$$tr(A - B) = p - (p - q) = q.$$

The requirement of Lemma 2.3.1 is satisfied, which implies

$$W(\theta_1^0) \xrightarrow{d} \chi_q^2.$$

□

It is worth noticing that Theorem 2.3.1 holds true as long as $a_n = o_p(n)$. In application, a_n with higher orders is usually not recommended, since the AEL ratios are decreasing functions of the adjustment level a_n (Chen and Huang, 2013). We will discuss the adjustment level a_n later in Section 3.3.

2.4 The AEL-based inference for the correlation coefficient

In this section, we apply the AEL method with nuisance parameters on the inference of the correlation coefficient ρ . Let $U_i = (X_i, Y_i)$, $i = 1, 2, \dots, n$ be the independent and identically distributed random vectors with mean (μ_x, μ_y) and variance-covariance matrix given in (1.16) which is

$$\Sigma = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}, \quad (2.21)$$

where $\sigma_{xy} = \rho\sigma_x\sigma_y$.

The estimating functions are as following:

$$X - \mu_x, \quad Y - \mu_y, \quad (X - \mu_x)^2 - \sigma_x^2, \quad (Y - \mu_y)^2 - \sigma_y^2, \quad \text{and} \quad (X - \mu_x)(Y - \mu_y) - \rho\sigma_x\sigma_y. \quad (2.22)$$

Consider ρ as the parameter of interest and $\mu_x, \mu_y, \sigma_x, \sigma_y$ as the nuisance parameters. By Theorem 2.3.1, the AEL test statistic $W(\rho)$ created from estimating equations based on (2.22) will follow χ_1^2 distribution under the null hypothesis $H_0 : \rho = \rho_0$ v.s. $H_a : \rho \neq \rho_0$.

The simulation study compares our proposed method with various existing methods (Banik & Kibria, 2017) in coverage probability. Let r be the sample correlation coefficient and $z = \frac{1}{2} \log\left(\frac{1+r}{1-r}\right)$.

- Classical: The details of the classical test for Pearson's correlation coefficient can be

found in (Lock et al. 2013). Testing for null hypothesis $H_0 : \rho = 0$ v.s. $H_a : \rho \neq 0$, the classical test statistic is

$$t_C = \sqrt{\frac{(n-2)r^2}{1-r^2}},$$

which follows the t -distribution with degrees of freedom $(n-2)$ under the null hypothesis.

- Fisher: For testing $H_0 : \rho = \rho_0$ v.s. $H_a : \rho \neq \rho_0$, the test statistic is computed by Fisher's transformation z (Fisher, 1915), and is given as follows:

$$t_F = \sqrt{n-3}(z - z_0),$$

where $z = \frac{1}{2} \log\left(\frac{1+r}{1-r}\right)$ and $z_0 = \frac{1}{2} \log\left(\frac{1+\rho_0}{1-\rho_0}\right)$. The distribution of t_F under H_0 is the standard normal.

- GL: Gorsuch and Lehmann test statistics are modifications of the classical and Fisher tests. Gorsuch and Lehmann (2010) proposed the following two statistics:

– GL1: Test statistic

$$t_{GL1} = \sqrt{n-1}\left(\frac{r}{1-r^2}\right)$$

follows t_{n-1} distribution under $H_0 : \rho = 0$.

– GL2: Test statistic

$$t_{GL2} = \sqrt{n-3}(z - z_0)$$

follows t_{n-1} distribution under $H_0 : \rho = \rho_0$.

- BK: This test was proposed by Banik and Kibria (2017). Let $Y_1 = \beta_0 + \beta_1 Y_2 + e$ be the regression model of Y_1 on Y_2 , where $e|Y_2 \sim N(0, \sigma^2)$. Denote the mean square error of the model by MSE . Then the test statistic for $H_0 : \rho = 0$ v.s. $H_a : \rho \neq 0$ is given by

$$t_{BK} = \frac{r}{\sqrt{\frac{MSE}{\sum_{i=1}^n (y_{1i} - \bar{y}_1)^2}}}$$

The distribution of t_{BK} is t_{n-2} under the null hypothesis.

The simulation compares the above methods with our proposed method. We generate data for $\rho = 0, 0.5, 0.9$ respectively under bivariate normal distribution with $\mu_x = 1, \mu_y = 1, \sigma_x = 1, \sigma_y = 1$ and bivariate chi-square distribution with degrees of freedom 4 and 6. The simulation is conducted with sample size $n = 20, 50$ for nominal values $n\alpha = 0.9, 0.95$. Coverage probabilities are obtained from 3000 repetitions for each case. We set up the adjustment $a_n = 0.5 \log n$ for the AEL method. The results are summarized in Table 2.2.

From Table 2.2, we can see that the AEL method has the most robust performance for different underlying population distributions. Note that for the bivariate normal data

with sample size 50 or larger, when the value of the true correlation coefficient is non-zero, the AEL method has comparable performance with the Fisher and GL2 methods. While when the data comes from bivariate chi-square distribution, for testing the true correlation coefficient is a non-zero value, the Fisher and GL2 methods perform poorly, and the AEL method significantly outperforms all other methods.

Bivariate normal		$\rho = 0$		$\rho = 0.5$		$\rho = 0.9$	
<i>nv</i>	Method	<i>n</i> = 20	<i>n</i> = 50	<i>n</i> = 20	<i>n</i> = 50	<i>n</i> = 20	<i>n</i> = 50
0.9	AEL	0.8683	0.8897	0.8633	0.8843	0.8690	0.8837
	Classical	0.8950	0.9057	NA	NA	NA	NA
	Fisher	0.8957	0.9057	0.9007	0.8943	0.9013	0.8963
	GL1	0.8653	0.8903	NA	NA	NA	NA
	GL2	0.9113	0.9103	0.9180	0.9013	0.9200	0.9030
	BK	0.8770	0.8973	NA	NA	NA	NA
0.95	AEL	0.9247	0.9407	0.9237	0.9380	0.9293	0.9397
	Classical	0.9473	0.9517	NA	NA	NA	NA
	Fisher	0.9467	0.9513	0.9520	0.9523	0.9540	0.9520
	GL1	0.9163	0.9270	NA	NA	NA	NA
	GL2	0.9623	0.9587	0.9640	0.9563	0.9640	0.9577
	BK	0.9370	0.9473	NA	NA	NA	NA

Table 2.2: Coverage probabilities of the correlation coefficient

Bivariate chi-square		$\rho = 0$		$\rho = 0.5$		$\rho = 0.9$	
<i>nv</i>	Method	<i>n</i> = 20	<i>n</i> = 50	<i>n</i> = 20	<i>n</i> = 50	<i>n</i> = 20	<i>n</i> = 50
0.9	AEL	0.8417	0.8783	0.8040	0.8413	0.8130	0.8427
	Classical	0.8950	0.9003	NA	NA	NA	NA
	Fisher	0.8950	0.9007	0.7693	0.7443	0.7743	0.7460
	GL1	0.8580	0.8897	NA	NA	NA	NA
	GL2	0.9143	0.9083	0.7880	0.7530	0.7957	0.7540
	BK	0.8723	0.8960	NA	NA	NA	NA
0.95	AEL	0.9073	0.9237	0.8800	0.9027	0.8877	0.9050
	Classical	0.9463	0.9507	NA	NA	NA	NA
	Fisher	0.9460	0.9503	0.8403	0.8210	0.8477	0.8227
	GL1	0.9190	0.9410	NA	NA	NA	NA
	GL2	0.9600	0.9553	0.8693	0.8317	0.8727	0.8357
	BK	0.9343	0.9477	NA	NA	NA	NA

Table 2.2: Coverage probabilities of the correlation coefficient (continued)

3 Application on the Sharpe Ratio

We have discussed two methods for statistical inference in previous chapters: the higher order parametric method in Chapter 1 and the AEL method in Chapter 2. In this chapter, we apply the inferential methods onto the Sharpe ratio sr .

In financial economics, Sharpe ratio (Sharpe, 1966) provides a measure of a fund's excess returns relative to its volatility. Let μ be an expected return of an asset, and σ be the corresponding standard deviation. Then the Sharpe ratio is defined as

$$sr = \frac{\mu - R_f}{\sigma} \quad (3.1)$$

where R_f is a known risk-free rate of return. In this case, we have parameter $\theta = (\mu, \sigma^2)$, and the parameter of interest is $sr = (\mu - R_f)/\sigma$.

Let x_1, \dots, x_n be a random sample from a distribution with mean μ and variance σ^2 . For simplicity, let assume that x_1, \dots, x_n are asset returns already adjusted for the risk-free return rate R_f . Then by (3.1), we have $sr = \mu/\sigma$. Note that the larger the Sharpe ratio

is, the more return the investor is getting per unit of risk. It is the standard convention in economics and finance research to report the Sharpe ratio. Therefore, the Sharpe ratio is very well studied as a measure of the mutual fund performance in the financial economic areas such as the portfolio analysis, the pricing of capital asset under conditions of risk and the general behavior of stock market prices. The popularity of the Sharpe ratio in financial economics is not only from its simplicity; the study of the Sharpe ratio will also directly result in deeper understandings in portfolio selections.

3.1 Higher order parametric inference for the Sharpe ratio sr

In this section, we review some work on the application of the higher order methods to the inference of the Sharpe ratio. The inference for the Sharpe ratio sr with higher order accuracy has been established in Liu, Rekkas and Wong (2012). The methodology there is essentially the same as the algorithm we summarized in Algorithm 1.2.1. The difference between their calculation and our process is that while computing $\chi(\theta)$, Liu, Rekkas and Wong (2012) adopted a calibration introduced by Fraser and Reid (1995). The details are as follows.

The standardized maximum likelihood estimate in Liu, Rekkas and Wong (2012) is

calculated as

$$\chi(\theta) = \psi_{\theta^r}(\hat{\theta}_\psi) \varphi_{\theta^r}^{-1}(\hat{\theta}_\psi) \varphi(\theta), \quad (3.2)$$

and the variance of the departure can be approximately estimated by

$$\widehat{\text{var}}(\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)) = \frac{\psi_{\theta^r}^r(\hat{\theta}_\psi) \tilde{j}_{\theta\theta^r}^{-1}(\hat{\theta}_\psi) \psi_{\theta^r}(\hat{\theta}_\psi) |\tilde{j}_{\theta\theta^r}(\hat{\theta}_\psi)| |\varphi_{\theta^r}(\hat{\theta}_\psi)|^{-2}}{|j_{\theta\theta^r}(\hat{\theta})| |\varphi_{\theta^r}(\hat{\theta})|^{-2}}, \quad (3.3)$$

where

$$\tilde{j}_{\theta\theta^r}(\hat{\theta}_\psi) = - \left. \frac{\partial^2 \tilde{\ell}(\theta)}{\partial \theta \partial \theta^r} \right|_{\hat{\theta}_\psi}.$$

In the above expression, $\tilde{\ell}(\theta)$, which is called the tilted log-likelihood, is defined as

$$\tilde{\ell}(\theta) = l(\theta) + \hat{\alpha}(\varphi(\theta) - \varphi),$$

where $\hat{\alpha}$ is the estimate of the Lagrange multiplier α , which is used to find the constraint MLE $\hat{\theta}_\psi$ under the constraint $\psi(\theta) = \psi$.

Except for the expressions given above, the steps to obtain for the inference of sr is the same as in Algorithm 1.2.1. Suppose the population distribution is normal with mean μ and variance σ . The likelihood function is

$$l(\theta) = l(\mu, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

The MLE of θ is

$$\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2) = (\bar{x}, \sum_{i=1}^n (x_i - \bar{x})^2 / n).$$

Therefore, we have

$$\hat{sr} = \frac{\hat{\mu}}{\hat{\sigma}},$$

and it is not hard to obtain

$$j_{\theta\theta^r}(\hat{\theta}) = \begin{pmatrix} \frac{n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{n}{2\hat{\sigma}^4} \end{pmatrix}.$$

Consequently,

$$|j_{\theta\theta^r}(\hat{\theta})| = \frac{n^2}{2\hat{\sigma}^6}.$$

The constraint MLE $\hat{\theta}_\psi = (\hat{\mu}_{sr}, \hat{\sigma}_{sr}^2)$ under the constraint $sr(\mu, \sigma^2) = sr$ can be obtained as

$$\hat{\sigma}_{sr} = \frac{-sr \cdot \bar{x} + \sqrt{(sr \cdot \bar{x})^2 + 4\left(\sum_{i=1}^n x_i/n\right)}}{2},$$

and

$$\hat{\mu}_{sr} = sr \cdot \hat{\sigma}_{sr},$$

with the estimate of the Lagrange multiplier α given by

$$\hat{\alpha} = n\left(sr - \frac{\bar{x}}{\hat{\mu}_{sr}}\right).$$

Therefore, the tilted log-likelihood $\tilde{\ell}(\theta)$ is

$$\tilde{\ell}(\theta) = \ell(\theta) + \hat{\alpha}\left(\frac{\mu}{\sigma} - sr\right).$$

The second partial derivatives of $\tilde{\ell}(\theta)$ are listed as below

$$\tilde{\ell}_{\mu\mu}(\theta) = \ell_{\mu\mu}(\theta),$$

$$\tilde{\ell}_{\mu\sigma^2}(\theta) = \ell_{\mu\sigma^2}(\theta) - \frac{\hat{\alpha}}{2\sigma^3},$$

$$\tilde{\ell}_{\sigma^2\sigma^2}(\theta) = \ell_{\sigma^2\sigma^2}(\theta) + \frac{3\hat{\alpha}\mu}{4\sigma^5},$$

where $\ell_{\mu\mu}(\theta)$, $\ell_{\mu\sigma^2}(\theta)$ and $\ell_{\sigma^2\sigma^2}(\theta)$ are the corresponding second partial derivatives of the original likelihood function $\ell(\theta)$, which are given by

$$\ell_{\mu\mu}(\theta) = -\frac{n}{\sigma^2},$$

$$\ell_{\mu\sigma^2}(\theta) = -\frac{1}{\sigma^4} \sum_{i=1}^n (x_i - \mu),$$

$$\ell_{\sigma^2\sigma^2}(\theta) = \frac{n}{2\sigma^2} - \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2.$$

Now it follows that

$$\tilde{j}_{\theta\theta^\tau}(\hat{\theta}_\psi) = \begin{pmatrix} -\tilde{\ell}_{\mu\mu}(\theta) & -\tilde{\ell}_{\mu\sigma^2}(\theta) \\ -\tilde{\ell}_{\mu\sigma^2}(\theta) & -\tilde{\ell}_{\sigma^2\sigma^2}(\theta) \end{pmatrix}.$$

The canonical parameters of the normal distribution is given by

$$\varphi(\theta)^\tau = \left(\frac{\mu}{\sigma^2}, \frac{1}{\sigma^2} \right).$$

Therefore,

$$\varphi_{\theta}(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & -\frac{\mu}{\sigma^4} \\ 0 & -\frac{1}{\sigma^4} \end{pmatrix},$$

and

$$\varphi_{\theta}^{-1}(\theta) = \begin{pmatrix} \sigma^2 & -\mu\sigma^2 \\ 0 & -\sigma^4 \end{pmatrix}.$$

With our parameter of interest

$$\psi(\theta) = sr(\mu, \sigma^2) = \mu/\sigma,$$

we have

$$\psi_{\theta}(\theta)^{\tau} = \left(\frac{1}{\sigma}, -\frac{\mu}{2\sigma^3} \right).$$

Now all the ingredients in (3.2) and (3.3) are directly obtainable and we are ready to find $Q(sr)$ and $r(sr)$ for $r^*(sr)$ so that our inference follows subsequently.

Recall that our proposed method for the inference of the Sharpe ratio sr is based on the assumption that the sample is from the normal distribution. Here we are interested in the simulation study about how much our method depends on the normality assumption. In order to achieve our goal, data are generated from $N(1, 0.25)$, χ_4^2 and χ_6^2 distributions with sample size $n = 20, 50$. For each distribution under each sample size n , we generate 5000 repetitions. The 0.90 and 0.95 confidence intervals are created for each sample based on

the proposed method. Then the proportion of the true sr that fall within, below or above the 5000 confidence intervals, which are defined as CP, LE and UE respectively in Section 1.3.3, are recorded correspondingly in Table 3.1. If the proposed method works well for the sample, then the proportion of the confidence intervals that contain the true sr should be close to the theoretical significance level $1 - \alpha$.

$1 - \alpha$	Distribution	$n = 20$			$n = 50$		
		CP	LE	UE	CP	LE	UE
0.9	$N(1, 0.25)$	0.901	0.045	0.054	0.904	0.047	0.049
	χ_4^2	0.967	0.009	0.024	0.954	0.017	0.029
	χ_6^2	0.951	0.017	0.032	0.948	0.020	0.032
0.95	$N(1, 0.25)$	0.948	0.022	0.030	0.950	0.023	0.027
	χ_4^2	0.988	0.003	0.009	0.982	0.007	0.011
	χ_6^2	0.981	0.006	0.013	0.978	0.009	0.013

Table 3.1: Simulated results of 90% and 95% confidence intervals for the Sharpe ratio

From Table 3.1, we can see that the proposed method works well for data from the normal with the proportion of confidence intervals containing the true sr quite close to

the theoretical value. However, if the data is generated from distributions with strong deformation from normality, for example the skewed families like the χ^2 -distributions used in our simulation, then the proportion can be way off the theoretical confidence.

The above simulation work shows that the parametric model is very sensitive to the correctness of the model specification. The method may not work well if the parametric family is incorrectly specified. This motivates us to consider methods that adopt more general background distributions. In Section 3.2, we apply the AEL method to the inference of the Sharpe ratio, which results in inferences that are more robust to the background distributions.

3.2 The AEL-based inference for the Sharpe ratio

In this section, we conduct simulation studies on the finite sample performance of the AEL method in presence of nuisance parameters. In particular, the investigations majorly focus on the inference of the Sharpe ratio sr , whose setup was briefly discussed in the context of normal data in Section 3.1. In order to define an AEL function for the parameter sr , a nuisance parameter, either the mean μ or the variance σ^2 , should be included in the

estimating equations. The estimating functions can be either

$$X - \mu \quad \text{and} \quad (X - \mu)^2 - \left(\frac{\mu}{sr}\right)^2 \quad (3.4)$$

or

$$X - \sigma \cdot sr \quad \text{and} \quad (X - \sigma \cdot sr)^2 - \sigma^2. \quad (3.5)$$

Suppose we want to perform a test with the null hypothesis

$$H_0: sr = sr_0.$$

The AEL ratio can be used as the test statistic. In the context of Theorem 2.3.1, we can set $\theta_1 = sr$ to be the parameter of interest, and let $\theta_2 = \mu$ or $\theta_2 = \sigma$ be the nuisance parameters. Thus by Theorem 2.3.1, the AEL ratio statistic $W(sr)$ will converge asymptotically to the χ_1^2 distribution under H_0 . Therefore, χ_1^2 can be used as a reference distribution for the test. Simulation shows that tests for sr based on estimating equation (3.4) and (3.5) yield essentially the same result. Therefore, throughout this chapter, we set the AEL ratio test statistic with the null hypothesis $H_0: sr = sr_0$ as

$$W(sr_0) = 2 [l(sr_0, \tilde{\sigma}_0) - l(\tilde{sr}, \tilde{\sigma})], \quad (3.6)$$

where $W(sr_0)$ follows χ_1^2 distribution when the sample size n approaches infinity.

3.2.1 Quantile-quantile plots

Figures 3.1 & 3.2 below are the Quantile-quantile plots of sample quantiles for the AEL ratio statistic $W(sr_0)$ against the theoretical χ_1^2 quantiles. The plots are each based on 5000 repetitions at sample size $n = 500$. We set $a_n = 0.5 \log n$ as suggested by Chen et al. (2008). In Figure 3.1, the data are generated from the normal distribution with mean 1 and variance 0.25, and the data used in Figure 3.2 are generated from the χ_4^2 distribution.

Figure 3.1 suggests that if data are generated from normal distributions, the calibration of the AEL method has higher accuracy. This implies that we can expect a more accurate coverage probabilities for normal data. However, if the data are generated from χ_4^2 , Figure 3.2 shows some deviance between the sample $W(sr_0)$ quantiles and the asymptotic χ_1^2 quantiles. This suggests that for data generated from skewed distributions, the convergence rate of the AEL ratio statistic to the corresponding asymptotic χ^2 -distribution can be slow.

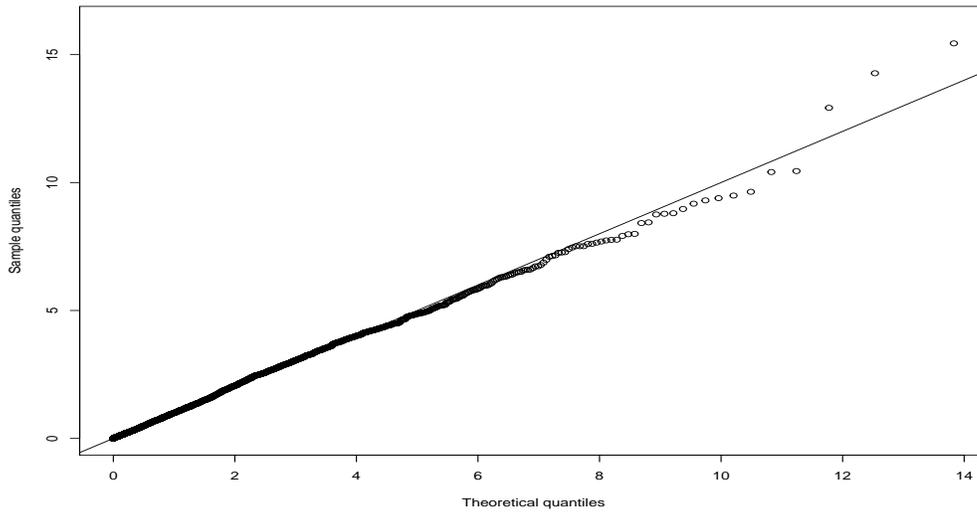


Figure 3.1: Quantile-quantile plot of data generated from $N(1, 0.25)$.

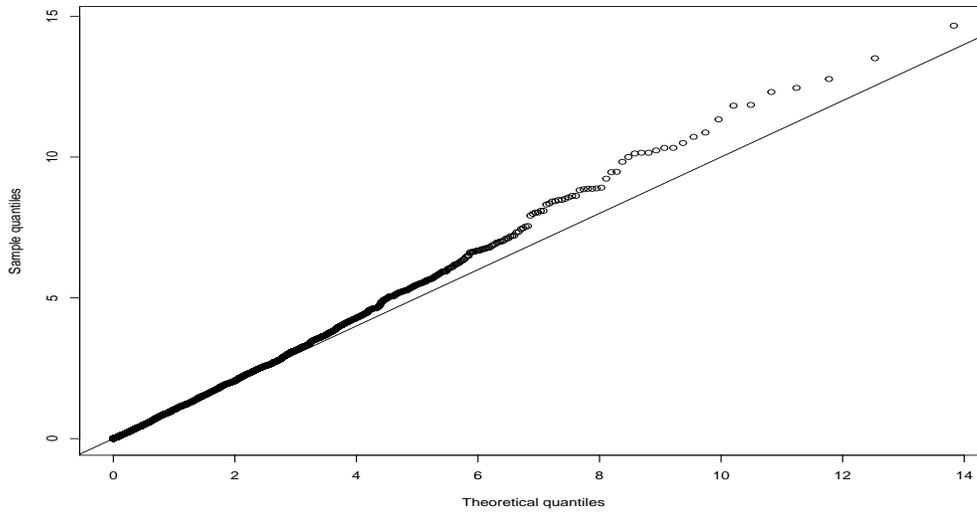


Figure 3.2: Quantile-quantile plot of data generated from χ_4^2 .

3.2.2 Simulation study

In this section, we use simulations to explore the finite sample performance of the AEL method on the Sharpe ratio compared with other popular methods in terms of coverage probability. The comparison is performed for sample sizes $n = 20, 50, 200, 500$ at nominal values 0.9, 0.95. Each coverage probability is obtained from 5000 simulations. The data are generated from the normal distribution with mean $\mu = 1$ and standard deviation $\sigma = 0.5$, t -distribution and the chi-square distributions with various degrees of freedom.

The methods under comparison are the following: the Jobson and Korkie (1981)'s method (JK), the Mertens (2002)'s method (Mertens), the usual EL inferential method (EL), the method applying the delta method on the asymptotic distribution of the EL estimator of the mean and standard deviation (Delta), and the proposed method (AEL) with the adjustment level $a_n = 0.5 \log n$. Jobson and Korkie (1981) assumed that the data are from a normal distribution. By applying the delta method to approximate the mean and variance of the Sharpe ratio, confidence interval for the Sharpe ratio can then be approximated by the Central Limit Theorem. Mertens (2002) used the skewness and kurtosis to give an adjusted approximation of the variance of the Sharpe ratio derived in Jobson and Korkie (1981) and again obtained the confidence interval of the Sharpe ratio from the Central Limit Theorem.

For the EL method, whenever the convex hull problem occurs for a set of simulated data, we use the convention to set the value of the profile log-EL function as negative infinity. The approach denoted by “Delta” is similar to JK but based on the EL. The details of the “Delta” method are as following.

Let $\theta = (\mu, \sigma)^\tau$ with the maximum empirical likelihood estimate $\tilde{\theta} = (\tilde{\mu}, \tilde{\sigma})^\tau$. Set up estimating equations for θ as

$$g(x, \theta) = (X - \mu, (X - \mu)^2 - \sigma^2)^\tau.$$

We have

$$\frac{\partial g(x, \theta)}{\partial \theta} = \begin{pmatrix} -1 & -2(x - \mu) \\ 0 & -2\sigma \end{pmatrix}.$$

Qin & Lawless (1994) (Theorem 1. on page 306) showed that

$$\sqrt{n}(\tilde{\theta} - \theta_0) \longrightarrow N(0, V),$$

where

$$V = \left[\left(E \frac{\partial g}{\partial \theta} \right)^\tau (E g g^\tau)^{-1} \left(E \frac{\partial g}{\partial \theta} \right) \right]^{-1}.$$

They further commented that the asymptotic variance V could be consistently estimated by

$$\left[\left(\sum_{i=1}^n \tilde{p}_i \frac{\partial g(x_i, \tilde{\theta})}{\partial \theta} \right)^\tau \left(\sum_{i=1}^n \tilde{p}_i g(x_i, \tilde{\theta}) g^\tau(x_i, \tilde{\theta}) \right)^{-1} \left(\sum_{i=1}^n \tilde{p}_i \frac{\partial g(x_i, \tilde{\theta})}{\partial \theta} \right) \right]^{-1},$$

or by the sample expression with the \tilde{p}_i 's replaced by n^{-1} .

Let $sr = h(\theta) = \mu/\sigma$. Then

$$\frac{\partial h(\theta)}{\partial \theta} = \left(\frac{1}{\sigma}, -\frac{\mu}{\sigma^2} \right)^\tau.$$

The delta method implies

$$\sqrt{n}(\tilde{sr} - sr_0) \longrightarrow N(0, H),$$

where

$$H = \left(\frac{\partial h(\theta)}{\partial \theta} \right)^\tau V \left(\frac{\partial h(\theta)}{\partial \theta} \right) = \left(\frac{\partial h}{\partial \theta} \right)^\tau \left[\left(E \frac{\partial g}{\partial \theta} \right)^\tau (E g g^\tau)^{-1} \left(E \frac{\partial g}{\partial \theta} \right) \right]^{-1} \left(\frac{\partial h}{\partial \theta} \right).$$

Therefore, we can estimate H by

$$\left(\frac{\partial h(\tilde{\theta})}{\partial \theta} \right)^\tau \left[\left(\sum_{i=1}^n \tilde{p}_i \frac{\partial g(x_i, \tilde{\theta})}{\partial \theta} \right)^\tau \left(\sum_{i=1}^n \tilde{p}_i g(x_i, \tilde{\theta}) g^\tau(x_i, \tilde{\theta}) \right)^{-1} \left(\sum_{i=1}^n \tilde{p}_i \frac{\partial g(x_i, \tilde{\theta})}{\partial \theta} \right) \right]^{-1} \left(\frac{\partial h(\tilde{\theta})}{\partial \theta} \right),$$

where each \tilde{p}_i 's can be replaced by n^{-1} . Note that since there are two estimating equations for two parameters (μ, σ) , $\tilde{p}_i = n^{-1}$ for all i in the just-determined case. This implies that

$$\tilde{\mu} = \bar{x} = \sum_{i=1}^n x_i/n \quad \text{and} \quad \tilde{\sigma} = \left(\sum_{i=1}^n (x_i - \bar{x})^2/n \right)^{1/2}.$$

Since there is no convex hull problem in the just-determined case, we can apply the delta method to the EL asymptotic distribution of mean and standard deviation. The simulated coverage probabilities of the above methods are summarized in Table 3.2.

$1 - \alpha$	Method	$n = 20$	$n = 50$	$n = 200$	$n = 500$
$N(1, 0.25)$					
0.9	JK	0.8956	0.9022	0.8968	0.9060
	Mertens	0.8258	0.8694	0.8894	0.9004
	EL	0.8210	0.8760	0.8906	0.9040
	AEL	0.8486	0.8874	0.8942	0.9058
	Delta	0.8428	0.8840	0.8896	0.8976
0.95	JK	0.9460	0.9488	0.9488	0.9522
	Mertens	0.8926	0.9270	0.9414	0.9494
	EL	0.8762	0.9214	0.9440	0.9514
	AEL	0.8980	0.9312	0.9466	0.9534
	Delta	0.9054	0.9334	0.9408	0.9476
t_3					
0.9	JK	0.8960	0.9004	0.9030	0.9040
	Mertens	0.8390	0.8646	0.8782	0.8890
	EL	0.8428	0.8738	0.8884	0.8946
	AEL	0.8794	0.8896	0.8944	0.8976
	Delta	0.8240	0.8586	0.8766	0.8858
0.95	JK	0.9494	0.9538	0.9516	0.9550
	Mertens	0.9028	0.9144	0.9326	0.9442
	EL	0.9042	0.9268	0.9438	0.9508
	AEL	0.9340	0.9402	0.9466	0.9514
	Delta	0.8910	0.9092	0.9318	0.9372
t_6					
0.9	JK	0.8982	0.8984	0.8934	0.8976
	Mertens	0.8738	0.8840	0.8900	0.8954
	EL	0.8700	0.8860	0.8928	0.8962
	AEL	0.8986	0.9018	0.8966	0.8978
	Delta	0.8634	0.8828	0.8936	0.9008

Table 3.2: Coverage probabilities of the Sharpe ratio

$1 - \alpha$	Method	$n = 20$	$n = 50$	$n = 200$	$n = 500$
t_6					
0.95	JK	0.9504	0.9482	0.9466	0.9476
	Mertens	0.9246	0.9364	0.9426	0.9458
	EL	0.9240	0.9394	0.9438	0.9466
	AEL	0.9466	0.9470	0.9480	0.9481
	Delta	0.9214	0.9330	0.9460	0.9494
χ_4^2					
0.9	JK	0.9640	0.9532	0.9476	0.9474
	Mertens	0.8048	0.8474	0.8676	0.8938
	EL	0.7800	0.8352	0.8626	0.8942
	AEL	0.8216	0.8536	0.8664	0.8952
	Delta	0.8072	0.8354	0.8660	0.8914
0.95	JK	0.9872	0.9808	0.9808	0.9774
	Mertens	0.8780	0.9072	0.9278	0.9422
	EL	0.8562	0.8972	0.9194	0.9418
	AEL	0.8924	0.9126	0.9228	0.9430
	Delta	0.8872	0.9046	0.9252	0.9388
χ_6^2					
0.9	JK	0.9466	0.9476	0.9392	0.9414
	Mertens	0.8048	0.8460	0.8760	0.8916
	EL	0.7996	0.8392	0.8728	0.8904
	AEL	0.8346	0.8568	0.8780	0.8926
	Delta	0.8236	0.8524	0.8766	0.8846
0.95	JK	0.9796	0.9776	0.9758	0.9754
	Mertens	0.8780	0.9168	0.9352	0.9412
	EL	0.8626	0.9032	0.9284	0.9402
	AEL	0.8894	0.9156	0.9326	0.9410
	Delta	0.8916	0.9170	0.9336	0.9432

Table 3.2: Coverage probabilities of the Sharpe ratio (continued)

From Table 3.2, we can see that the AEL method has the most robust performance for various underlying population distributions. The AEL method always has significant better performance over the EL method in terms of coverage probability. When the data is normal distributed, the JK method performs the best while when the data comes from a skewed distribution, the JK method performs poorly. For normal data with small sample size, the AEL has slightly less coverage probabilities than the JK method, while for normal data with sample size larger than 50 and data from various t distributions, the AEL has comparable performance with the JK method. For all other situations, the AEL method significantly outperforms all other methods, especially for cases with small sample sizes.

3.2.3 Real data analysis

The data we consider is the Nasdaq GS return of the Apple Inc. from October 03, 2017 to December 12, 2017 (<https://finance.yahoo.com/quote/AAPL/>). The return is evaluated from the close price of the current day compared with the close price of the previous day. There are 50 trading days during the period considered. We use the yearly return rate of the 5-year bonds, which is 2.116%, as the yearly risk-free return. Therefore, the daily risk-free return rate used in the analysis is $0.02116/252 = 8.397 \times 10^{-5}$. For the return data of size 50, the

Durbin-Watson test statistic is 1.58, and the p -value for a two-sided test is 0.1285, which does not show any significant evidence of serial correlation. The qqplot of the returns in Figure 3.3 reveals some skewness of the data. The confidence intervals of the Sharpe ratio for the Apple Inc. return data produced by different methods are listed in Table 3.3. For JK and Mertens methods, the point estimates are the value of sr that corresponding to the 50% quantile of the standard normal limiting distribution of their test statistics. The estimates of the Delta, EL and AEL methods are the value of the maximum EL and AEL estimates, respectively.

From Table 3.3, we see that since JK and Mertens methods are moment-based methods, both their estimates are the same as the sample Sharpe ratio. The Delta, EL and AEL methods are empirical-likelihood-based methods so the corresponding estimates are different from the previous two approaches. We observe that there is some difference in the confidence intervals for various approaches. Note that the data has some skewness as shown in Figure 3.3. Based on the observation from our simulation studies, the skewness will affect the JK method but not the rest of the three methods. The confidence interval based on our proposed AEL method is more robust and trustworthy.

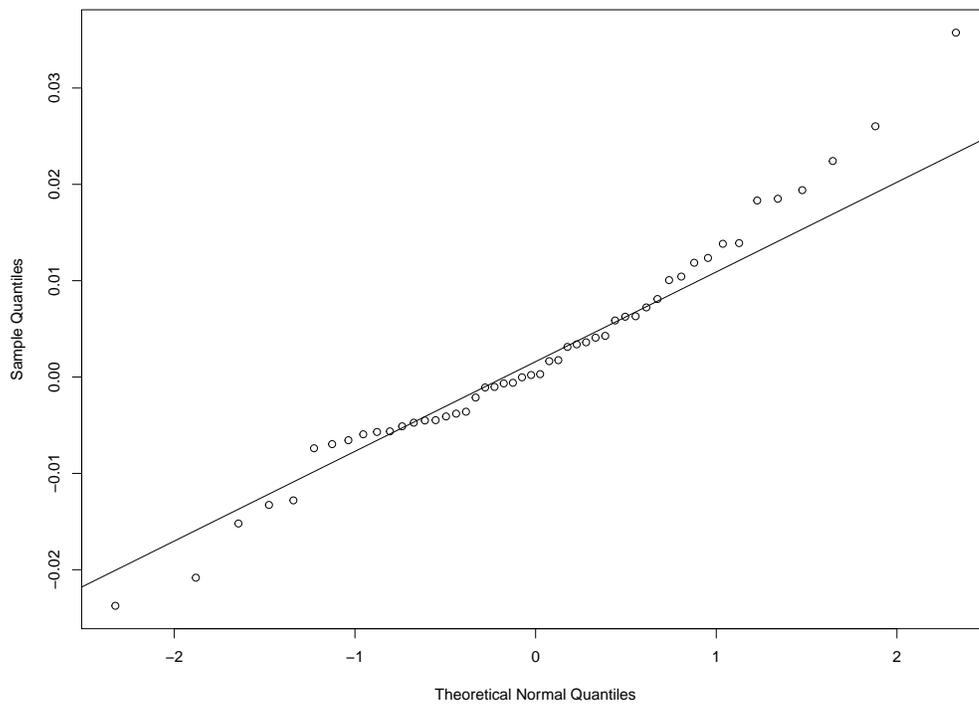


Figure 3.3: Quantile-quantile plot of Apple Inc. return data

$1 - \alpha$	Method	Estimate	Lower Bound	Upper Bound
0.9	JK	0.1907	-0.0441	0.4254
	Mertens	0.1907	-0.0350	0.4163
	Delta	0.1926	-0.0329	0.4181
	EL	0.1926	-0.0376	0.4140
	AEL	0.1926	-0.0479	0.4241
0.95	JK	0.1907	-0.0890	0.4703
	Mertens	0.1907	-0.0783	0.4596
	Delta	0.1926	-0.0761	0.4613
	EL	0.1926	-0.0827	0.4558
	AEL	0.1926	-0.0949	0.4683

Table 3.3: Confidence Intervals of the Sharpe ratio for Apple Inc. return data

3.3 Unbounded confidence region problem

One side effect of the AEL is that we may overshoot. That is, the confidence region may become the whole parameter space. This undesirable finite-sample property was first noticed and proved by Emerson & Owen (2009). Chen & Huang (2013) studied the finite-sample properties of the AEL method. They find that the AEL ratio function decreases when the level of adjustment a_n increases. Thus, the AEL confidence region has higher coverage probabilities when the level of adjustment increases. Recall that the adjustment a_n must be positive and has order $o_p(n^{2/3})$. As a_n increases from 0, the AEL confidence region continuously expands when the confidence level is fixed.

In order to solve the unbounded confidence region problem, Chen & Huang (2013) introduced the idea to modify a_n into a function of the sample and the true value of the parameter θ . To be specific, the adjustment level can be substituted by $a_n = aK(t)$, where $a > 0$ is related to n whereas t is related to the sample and some assumed value of θ , such as $t = |\hat{\theta} - \theta|$, where $\hat{\theta}$ is a consistent estimator of θ . Here $K(t)$ is a function satisfying the following conditions:

1. $0 < K(t) \leq 1$ for all t .

2. $K(t)$ strictly increases to 1 as t decreases to 0.

3. $K(t)$ strictly decreases to 0 as t increases to ∞ .

The choice of $K(t)$ can be quite versatile. Chen & Huang (2013) recommended functions such as $K(t) = (1 + t^{1+\alpha})^{-1}$ and $\exp\{-t^\alpha\}$ for some $\alpha > 0$. Note that as long as $a = o_p(n)$, Theorem 2.3.1 will hold if we replace a_n with $aK(t)$ given as above.

The introduction of the $K(t)$ function essentially links the level of adjustment with the size of some sample related values such as $|\hat{\theta} - \theta|$. This idea comes from Emerson and Owen (2009) in order to justify the unbounded confidence region problem. Let the parameter under inference be the population mean μ . Then the following proposition gives an upper bound to the AEL ratio statistic $W(\mu) = 2[l_{AEL}(\tilde{\mu}) - l_{AEL}(\mu)]$ defined in Section 2.2. The proof can be found in Emerson and Owen (2009), who gave the following result.

Proposition 3.3.1. of Emerson and Owen (2009). *For any finite sample size n and a fixed a_n , the AEL ratio statistic $W(\mu) = 2[l_{AEL}(\tilde{\mu}) - l_{AEL}(\mu)]$ is bounded above by:*

$$W(\mu) \leq B(n, a_n) = -2 \left[n \log \left(\frac{(n+1)a_n}{n(a_n+1)} \right) + \log \left(\frac{n+1}{a_n+1} \right) \right].$$

Proposition 3.3.1 indicates that for the inference of μ , when the sample size n is small and when the dimension p of μ is large, we may face the problem that the confidence region

built by the AEL method may include the whole parameter space if we set $a_n = 0.5 \log n$ or any sample-size-determined adjustment levels. In order to solve the unbounded confidence region problem for the inference of the population mean μ , Chen & Huang (2013) replaced a_n with a $K(t)$ function defined above and set

$$t = \|\bar{x}_n - \mu\|,$$

where $\|\cdot\|$ was introduced by Hotelling (1931) such that

$$\|\bar{x}_n - \mu\|^2 = (\bar{x}_n - \mu)^T S_n^{-1} (\bar{x}_n - \mu),$$

here S_n is the sample variance-covariance matrix.

Chen & Huang (2013) further proved that for any $a = o_p(n^{2/3})$, the modified adjustment level $a_n = aK(t)$ guarantees that the AEL ratio statistic $W(\mu)$ can no longer be bounded above. This indicates that the confidence region constructed for μ will not be the whole parameter space.

Similar unbounded confidence region problem can also occur for the AEL inference on the Sharpe ratio sr . We have run 5000 repetitions with sample sizes $n = 10, 20, 50$ for data generated from $N(1, 0.25)$, χ_4^2 , χ_6^2 , t_3 and t_6 distributions, respectively. The corresponding 95% and 99% CIs for sr are computed and results are shown in Table 3.4. We compare the

following two methods: (1) the AEL method (AEL) with $a_n = 0.5 \log n$; (2) the modified AEL method (AEL*) with $a_n = aK(t)$ where $a = 0.5 \log n$ and $K(t) = (1 + 0.1t^2)^{-1}$. We set $t = \hat{sr} - sr = \bar{x}/s - sr$, where \bar{x} and s be the sample mean and sample standard deviation of the data, respectively.

From Table 3.4, we can observe that for various underlying population distributions, the AEL confidence intervals for sr are often unbounded when the sample size is small enough and the nominal value is 0.99. This undermines the reporting of the median length in addition to the average length. Our results indicate that the AEL* confidence intervals are always bounded and they have reasonable median lengths.

^aCoverage probability (median length, average length).

Data	CI	Method	$n = 10$	$n = 20$	$n = 50$
$N(1, 0.5)$	0.95	AEL	0.831 (3.226 , 5.150) ^a	0.898 (1.631 , 1.702)	0.931 (0.985 , 1.001)
		AEL*	0.824 (2.544 , 2.674)	0.898 (1.608 , 1.669)	0.931 (0.983 , 1.000)
	0.99	AEL	0.942 (39.59 , ∞)	0.948 (2.412 , 2.593)	0.978 (1.311 , 1.332)
		AEL*	0.909 (3.918 , 4.084)	0.947 (2.235 , 2.310)	0.977 (1.307 , 1.327)
χ_4^2	0.95	AEL	0.859 (2.346 , 5.352)	0.892 (1.010 , 1.043)	0.913 (0.594 , 0.606)
		AEL*	0.855 (1.761 , 1.841)	0.891 (1.001 , 1.029)	0.913 (0.593 , 0.605)
	0.99	AEL	0.965 (21.69 , 23.31)	0.963 (1.563 , 2.657)	0.968 (0.789 , 0.803)
		AEL*	0.946 (3.229 , 3.239)	0.962 (1.470 , 1.515)	0.968 (0.788 , 0.801)
χ_6^2	0.95	AEL	0.843 (2.716 , 5.141)	0.889 (1.211 , 1.254)	0.916 (0.710 , 0.724)
		AEL*	0.838 (2.042 , 2.138)	0.888 (1.198 , 1.233)	0.915 (0.709 , 0.724)
	0.99	AEL	0.961 (21.38 , 22.55)	0.957 (1.843 , 2.564)	0.976 (0.943 , 0.960)
		AEL*	0.934 (3.455 , 3.513)	0.956 (1.719 , 1.772)	0.976 (0.941 , 0.958)
t_3	0.95	AEL	0.937 (1.867 , 8.221)	0.934 (0.970 , 1.183)	0.940 (0.567 , 0.556)
		AEL*	0.933 (1.708 , 1.804)	0.933 (0.966 , 0.967)	0.940 (0.567 , 0.556)
	0.99	AEL	0.997 (52.11 , ∞)	0.984 (1.402 , 5.739)	0.987 (0.746 , 0.843)
		AEL*	0.992 (3.894 , 4.303)	0.983 (1.370 , 1.515)	0.987 (0.745 , 0.745)
t_6	0.95	AEL	0.939 (1.852 , 5.383)	0.947 (0.991 , 0.998)	0.951 (0.578 , 0.575)
		AEL*	0.934 (1.720 , 1.773)	0.946 (0.987 , 0.985)	0.951 (0.577 , 0.574)
	0.99	AEL	0.996 (54.98 , ∞)	0.989 (1.417 , 2.423)	0.993 (0.763 , 0.773)
		AEL*	0.991 (3.712 , 3.970)	0.989 (1.388 , 1.415)	0.993 (0.763 , 0.759)

Table 3.4: Coverage probability and median/mean length.

4 Future Works

In Chapter 1, we have discussed the accurate parametric method for the inference of a p -dimensional parameter vector θ with a scalar parameter of interest. Recent development shows that higher order likelihood inference can be performed for a vector parameter of interest. In particular, Davison et al. (2014) discussed the problem in the context of the exponential family distributions and developed a directional test with accuracy $O(n^{-\frac{3}{2}})$ for a d -dimensional parameter vector of interest $\psi(\theta)$ with $d \leq p$.

Similar to the accurate parametric method of Chapter 1, the inference for the vector parameter of interest requires an explicit likelihood function and an expression of θ in the form of canonical parameters. Davison et al. (2014) proposed that one can set up proper sufficient statistics for the nuisance parameters and use them as conditions to establish the plane L^0 of basic density for the inference of $\psi(\theta)$, which essentially eliminates the nuisance parameters by conditioning. Focusing on the L^0 , directional departures can be

easily expressed in the form of canonical parameters and accurate tests can be based on the saddlepoint approximation of the aforementioned likelihood function.

Wong & Zhang (2017) applied Davison et al. (2014)'s approach to the test of homogeneity of the inverse Gaussian scale-like parameters. The inverse Gaussian distribution, parameterized by a mean parameter $\mu > 0$ and a scale-like parameter $\lambda > 0$, is widely used to model positive right-skewed population. Wong & Zhang (2017) studied the test for k independent inverse Gaussian distributions under the hypotheses

$$H_0 : \lambda_1 = \dots = \lambda_k \quad \text{v.s.} \quad H_a : \lambda_i \neq \lambda_j \text{ for some } i \neq j.$$

The method developed in Davison et al. (2014) were used to achieve higher order asymptotic results.

Shi & Wong (2018) accessed the test for homogeneity of multiple parameters with a different approach. They proposed to test the coefficients of variation. With normality assumption, Shi & Wong (2018) used Bartlett-correction on the log-likelihood ratio test to achieve higher order accuracy, where the Bartlett-correction was obtained numerically.

Following Davison et al. (2014), we plan to extend the higher order likelihood inference for one scalar parameter of interest to a vector parameter of interest. For example, when Analysis of Variance (ANOVA) is used to test differences among multiple population means,

the canonical parameters and likelihood functions are readily obtainable with normality assumption.

In Chapter 2, we extended the AEL method to obtain inference for a parameter of interest in the presence of nuisance parameters. The advantage of the proposed method is that it does not rely on the distributional assumption of the data. In particular, we applied the proposed method to obtain inference for the Sharpe ratio. Simulation results show that the proposed method has the most robust performance for different underlying population distributions. In addition, when the data are from a skewed distribution, the proposed method outperforms all other existing methods.

For the real data analysis on the Sharpe ratio, the time-series properties of investment strategies can have a nontrivial impact on the Sharpe ratio estimator. Empirical likelihood was motivated by independent and identically distributed data. When dealing with dependent data, we need to account for the dependency structure in constructing confidence regions for the parameter of interest. In general, the approach to handle dependent data within the EL framework is parallel to the methods based on parametric likelihood. The extension of our approach for dependent data is valuable and interesting. We will consider it in future research.

One side effect of the AEL is that the confidence region may become the whole parameter space. Chen & Huang (2013) proposed a modified adjustment level and showed that the modified AEL confidence region is always bounded for the population mean case. In Section 3.3, we applied the method of Chen & Huang (2013) to obtain inference for the Sharpe ratio. Our simulation results indicate that the unbounded confidence region problem could be resolved by using modified adjustment level. It is desired to develop a general theory for arbitrary parameter of interest, and also for the situation with nuisance parameters. Therefore, a good future work is to study the unbounded confidence region problem for a general parameter under AEL inference.

Bibliography

- [1] Agarwal V. and Naik N. Y. (2004). Risk and portfolio decisions involving hedge funds. *Review of Financial Studies*, 17(1), 63-98.
- [2] Alkema L., Raftery A.E., Gerland P., Clark S.J., Pelletier F., Buettner T. and Heilig G. (2012). Probabilistic projections of the total fertility rate for all countries. *Demography*, 48, 815-839.
- [3] Anis M.Z. (2008). Estimating the mean of normal distribution with known coefficient of variation. *American Journal of Mathematical and Management Sciences*, 28, 469-487.
- [4] Banik S. and Kibria B. (2017). Size and power probabilities of some test statistics for testing the population correlation coefficient ρ . *Journal of Biometrics & Biostatistics*, 8(3), DOI: 10.4172/2155-6180.1000353.

- [5] Barnard J., McCulloch R. and Meng X. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistics Sinica*, 10, 1281-1311.
- [6] Barndorff-Nielsen O.E. (1986). Inference on full and partial parameters, based on the standardized signed log-likelihood ratio. *Biometrika*, 73, 307-322.
- [7] Barndorff-Nielsen O.E. (1991). Modified signed log-likelihood ratio. *Biometrika*, 78, 557-563.
- [8] Bartolucci F. (2007). A penalized version of the empirical likelihood ratio for the population mean. *Statistics and Probability Letters*, 77, 104-110.
- [9] Berger J.O. and Sun D. (2008). Objective priors for the bivariate normal model. *Annals of Statistics*, 36, 963-982.
- [10] Bhat K. and Rao K.A. (2007). On tests for a normal mean with known coefficient of variation. *International Statistical Review*, 75, 170-182.
- [11] Brazauskas V. and Ghorai J. (2007). Estimating the common parameter of normal models with known coefficients of variation: a sensitivity study of asymptotically efficient estimators. *Journal of Statistical Computation and Simulation*, 77, 663-681.

- [12] Chaturvedi A. and Tomer S.K. (2003). Three-stage and “accelerated” sequential procedures for the mean of a normal population with known coefficient of variation. *Statistics*, 1, 51-64.
- [13] Chen J., Variyath A. M. and Abraham B. (2008). Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics*, 17(2), 426-443.
- [14] Chen J. and Huang Y. (2013). Finite-sample properties of the adjusted empirical likelihood. *Journal of Nonparametric Statistics*, 25, 147-159.
- [15] Chen J. and Liu L. (2012). Adjusted empirical likelihood with high-order one sided coverage precision. *Statistics and its Interface*, 5, 281-292.
- [16] Chen J. and Liu Y. (2010) Adjusted empirical likelihood with high-order precision. *The Annals of Statistics*, 38, 1341-1361.
- [17] Chen J. and Qin J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, 80, 107-116.
- [18] Chen S. and Cui H. (2006). On Bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika*, 93, 215-220.

- [19] Cox D.R. and Hinkley D.V. (1974). *Theoretical Statistics*. Chapman and Hall, London.
- [20] Davison A.C., Fraser D.A.S., Reid N. and Sartori N. (2014). Accurate directional inference for vector parameters in linear exponential families. *Journal of American Statistical Association*, 109, 302-314.
- [21] Diccio T., Hall P. and Romano J. (1991). Empirical likelihood is Bartlett-correctable. *The Annals of Statistics*, 19, 1053-1061.
- [22] Dogmaksoy N. and Schme J. (1993). Comparisons of approximate confidence intervals for distributions used in life-data analysis. *Technometrics*, 35, 175-184.
- [23] Efron B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *Annals of Statistics*, 3, 1189-1242.
- [24] Emerson S. (2009). Small sample performance and calibration of the empirical likelihood method. *Dissertation*.
- [25] Emerson S. and Owen A. (2009). Calibration of the empirical likelihood method for a vector mean. *Electronic Journal of Statistics*, 3, 1161-1192.
- [26] Fama E. (1965). The behavior of stock-market prices. *Journal of Business*, 38, 34-105.

- [27] Fisher R.A. (1915). Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika*, 10, 507-521.
- [28] Fosdick B.K. and Raftery, A.E. (2012). Estimating the correlation in bivariate normal data with known variances and small sample sizes. *The American Statisticians*, 66, 34-41.
- [29] Fraser D.A.S. (1990). Tail probabilities from observed likelihoods. *Biometrika*, 77, 65-76.
- [30] Fraser D.A.S., Reid N. and Wu J. (1999). A simple general formula for tail probabilities for frequentist and bayesian inference. *Biometrika*, 86, 249-264.
- [31] Fraser D.A.S. and Reid N. (1995). Ancillaries and third order significance. *Utilitas Mathematica*, 7, 33-55.
- [32] Fraser D.A.S., Reid N. and Wong A. (1997). Simple and accurate inference for the mean parameter of the gamma model. *Canadian Journal of Statistics*, 25, 91-99.
- [33] Fu Y., Wang H. and Wong A. (2013). Inference for the normal mean with known coefficient of variation. *Open Journal of Statistics*, 3(6), 45-51.

- [34] Fu Y., Wang H. and Wong A. (2013) Small sample inference for the correlation in bivariate normal with known variances. *Far East Journal of Theoretical Statistics*, 45, 147-164.
- [35] Fu Y., Wang H. and Wong A. (2018). Adjusted empirical likelihood method in the presence of nuisance parameters with application to the sharpe ratio. *Entropy*, 20, 316.
- [36] Gajjar A.V. and Subrahmaniam K. (1978). On the sample correlation coefficient in the truncated bivariate normal population. *Communications in Statistics - Simulation and Computation*, 7, 455-477.
- [37] Gallier J. (2011). *Geometric Methods and Applications For Computer Science and Engineering*. Springer, New York.
- [38] Ghosh M., Mukherjee B., Santra U. and Kim D. (2010). Bayesian and likelihood-based inference for the bivariate normal correlation coefficient. *Journal of Statistical Planning and Inference*, 140, 1410-1416.
- [39] Gleser L.J. and Healy J.D. (1976). Estimating the mean of normal distribution with known coefficient of variation. *Journal of the American Statistical Association*, 71, 977-981.

- [40] Gorsuch R.L. and Lehmann C.S. (2010). Correlation coefficient mean bias and confidence interval distortions. *Journal of Methods and Measurement in the Social Science*, 1, 52-65.
- [41] Guo H. and Pal N. (2003). On a normal mean with known coefficient of variation. *Calcutta Statistical Association Bulletin*, 54, 17-29.
- [42] Hall P. and La Scala B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review*, 58(2), 109-127.
- [43] Hjort N. L., Mckeague I. W. and Keilegom I. V. (2009). Extending the scope of empirical likelihood. *The Annals of Statistics*, 37, 1079-1111.
- [44] Hinkley D.V. (1977). Conditional inference about a normal mean with known coefficient of variation. *Biometrika*, 64, 105-108.
- [45] Hotelling H. (1931). The generalization of student's ratio. *The Annals of Mathematical Statistics*, 4(1), 199-218.
- [46] Ingersoll J., Spiegel M. and Goetzmann W. (2007). Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies*, 20(5), 1503-1546.

- [47] Jobson J.D. and Korkie B.M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *The Journal of Finance*, 36(4), 889-908.
- [48] Khan R.A. (1968). A note on estimating the mean of a normal distribution with known coefficient of variation. *Journal of the American Statistical Association*, 63, 1039-1041.
- [49] Khan R.A. (2013). A remark on estimating the mean of a normal distribution with known coefficient of variation. *Statistics*, doi:10.1080/02331888.2013.809722.
- [50] Kuk A.Y.C. and Mak T.K. (1989). Median estimation in the presence of auxiliary information. *Journal of Royal Statistical Society Series B*, 51, 261-269.
- [51] Lazar N. A. and Mykland P.A. (1999). Empirical likelihood in the presence of nuisance parameters. *Biometrika*, 86, 203-211.
- [52] Lahiri S. and Mukhopadhyay S. (2012). A penalized empirical likelihood method in high dimensions. *The Annals of Statistics*, 40, 2511-2540.
- [53] Leland H. E. (1999). Beyond mean-variance: risk and performance measurement in a nonsymmetric world. *Finance Analysts Journal*, 1, 27-36.

- [54] Li X., Chen J., Wu Y. and Tu D. (2011). Constructing nonparametric likelihood confidence regions with high-order precisions. *Statistical Sinica*, 21, 1767-1783.
- [55] Liechty J.C., Liechty M. and Muller, R. (2004). Bayesian correlation estimation. *Biometrika*, 91, 1-14.
- [56] Liu Y. and Chen J. (2010). Adjusted empirical likelihood with high-order precision. *The Annals of Statistics*, 38, 1341-1362.
- [57] Liu Y., Rekkas M. and Wong A. (2012). Inference for the Sharpe ratio using a likelihood-based approach. *Journal of Probability Statistics*. Volume 2012.
- [58] Liu Y. and Yu C.W. (2010). Bartlett correctable two-sample adjusted empirical likelihood. *Journal of Multivariate Analysis*, 101, 1701-1711.
- [59] Liu Y., Zou C. and Zhang R. (2007). Empirical likelihood for the two-sample mean problem. *Statistics and Probability Letters*, 78, 548-556.
- [60] Lock R. H., Lock P. F., Morgan K. L., Lock E. F. and Lock D. F. (2013). *Statistics: Unlocking the Power of Data*. John Wiley & Sons.

- [61] Mertens E. (2002). Comments on variance of the IID estimator in Lo. Technical report, Working Paper University of Basel, Department of Finance, 2002. <http://www.elmarmertens.com/research/discussion/soprano01.pdf>.
- [62] Niwitpong S. (2012). Confidence intervals for the normal mean with known coefficient of variation. World Academy of Science, Engineering and Technology, 69, 677-680.
- [63] Owen A. (1988). Empirical likelihood ratio confidence intervals for a single functional. Biometrika, 75, 237-249.
- [64] Owen A. (2001). Empirical Likelihood. Chapman and Hall, Boca Raton etc.
- [65] Owen A. (1991). Empirical likelihood for linear models. The Annals of Statistics, 19, 1725-1747.
- [66] Owen A. (1990). Empirical likelihood ratio confidence regions. The Annals of Statistics, 18, 90-120.
- [67] Pearson K. (1920). Note on the history of correlation. Biometrika, 13, 25-45.
- [68] Pierro M. D. and Mosevich J. (2011). Effects of skewness and kurtosis on portfolio rankings. Quantitative Finance, 11(10), 1449-1453.

- [69] Pu X., Chen J. and Wang L. (2015). Resampling calibrated adjusted empirical likelihood. *Canadian Journal of Statistics*, 43, 42-59.
- [70] Qin J. and Lawless J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22, 300-325.
- [71] Rahman N. A. (1968). *A course in theoretical statistics*. Charles Griffin and Company.
- [72] Rao C. R. (1947). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 44, 50-57.
- [73] Rao C. R. (1973). *Linear Statistical Inference and Its Applications*. Wiley, New York.
- [74] Reid N. (1996). Likelihood and higher-order approximations to tail areas: a review and annotated bibliography. *Canadian Journal of Statistics*, 24, 141-166.
- [75] Rezaei S., Tahmasbi R. and Mahmoodi M. (2010). Estimation of $P(Y < X)$ for generalized Pareto distribution. *Journal of Statistical Planning and Inference*, 140, 480-494.

- [76] Rodgers J.L. and Nicewander W.A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42, 59-66.
- [77] Samuelson P. (1970). The fundamental approximation theorem of portfolio analysis in terms of means, variances, and higher moments. *The Review of Economic Studies*, 37(4), 537-542.
- [78] Scott R. and Horvath P. (1980). On the direction of preference for moments of higher order than variance. *Journal of Finance*, 35, 915-919.
- [79] Searls D.T. (1964). The utilization of a known coefficient of variation in the estimation procedure. *Journal of the American Statistical Association*, 59, 1225-1226.
- [80] Sen A.R. (1979). Relative efficiency of estimators of the mean of a normal distribution when coefficient of variation is known. *Biometrical Journal*, 21, 131-137.
- [81] Severeni T. (2000). *Likelihood Methods in Statistics*. Oxford University Press, New York.
- [82] Shao J. (2009). *Mathematical statistics*. Springer.
- [83] Sharpe W. F. (1966). Mutual fund performance. *Journal of Business*, 39, 119-138.

- [84] Shi X. and Wong A. (2018) Accurate tests for the equality of coefficients of variation. *Journal of statistical computation and simulation*, 88, 3529-3543.
- [85] Singh R. (1998). Sequential estimation of the mean of normal population with known coefficient of variation. *Metron*, 56, 73-90.
- [86] Soofi E. S. and Gokhale D.V. (1991). Minimum discrimination information estimator of the mean with known coefficient of variation. *Computational Statistics and Data Analysis*, 11, 165-177.
- [87] Spruill N.L. and Gastwirth J.L. (1982). On the estimation of the correlation coefficient from grouped data. *Journal of the American Statistical Association*, 77, 614-620.
- [88] Srisodaphol W. and Tongmol N. (2012). Improved estimators of the mean of a normal distribution with a known coefficient of variation. *Journal of Probability and Statistics*, doi: 10.1155/2012/807045.
- [89] Thomas D.R. and Grunkemeier G.L. (1975). Confidence interval estimation of survival probabilities for censored data. *Journal of American Statistical Association*, 70, 865-871.

- [90] Tsao M. and Wu F. (2013). Empirical likelihood on the full parameter space. *The Annals of Statistics*, 41, 2176-2196.
- [91] Tsao M. (2013). Extending the empirical likelihood by domain expansion. *The Canadian Journal of Statistics*, 41, 257-274.
- [92] Vexler A., Liu S., Kang L. and Hutson A.D. (2016). Modifications of the empirical likelihood interval estimation with improved coverage probabilities. *Communications in Statistics Simulation and Computation*, 38, 2171-2183.
- [93] Wang L. (2016). Bartlett-corrected two-sample adjusted empirical likelihood via resampling. *Communications in Statistics - Theory and Methods*, 24, 10941-10952.
- [94] Wald A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54, 426-483.
- [95] Wilks S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1), 60-62.
- [96] Wong A. and Zhang S. (2017) A directional approach for testing homogeneity of inverse Gaussian scale-like parameters. *Biostat Biometrics Open Acc J*, 3(2).

- [97] Yu W., Barmi H.E. and Ying Z. (2011). Restricted one way analysis of variance using the empirical likelihood ratio test. *Journal of Multivariate Analysis*, 102, 629-640.
- [98] Yu W., Zhao Z. and Zheng M. (2012). Empirical likelihood methods based on influence functions. *Statistics and Its Interface*, 5, 355-366.
- [99] Zakamouline V. and Koekebakker S. (2009). Portfolio performance evaluation with generalized Sharpe ratios: beyond the mean and variance. *Journal of Banking and Finance*, 33(7), 1242-1254.