On Bartlett Correction of Empirical Likelihood in Time Series

CHEN, Kun

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy
in
Statistics

The Chinese University of Hong Kong

July 2014

Thesis/Assessment Committee

Professor WONG Hoi Ying (Chair)

Professor CHAN Ngai Hang (Thesis Supervisor)

Professor YAU Chun Yip (Thesis Supervisor)

Professor Tony Sit (Committee Member)

Professor PENG Liang (External Examiner)

Abstract

Empirical likelihood (EL) is a nonparametric likelihood approach for parameter estimation and hypothesis testing. A desirable feature of the EL method is that it allows Bartlett correction, which is a simple statistical adjustment on the test statistic to construct confidence regions with improved coverage accuracies. Previous studies have demonstrated the Bartlett correctability of EL for independent and identically distributed data and Gaussian short-memory time series. However, it is still unknown whether EL is Bartlett correctable for long-memory or non-Gaussian distributed time series. In this thesis, we establish the validity of the Edgeworth expansion for the signed root empirical log-likelihood ratio statistic to prove that EL is Bartlett correctable for Gaussian long-memory and non-Gaussian short-memory time series. For Gaussian long-memory time series, the Edgeworth expansion admits an irregular form with a power series of order $\log^3 n/\sqrt{n}$. Based on the expansion, the coverage error of the EL confidence region can be reduced from $O(\log^6 n/n)$ to $O(\log^3 n/n)$. For non-Gaussian short-memory time series, by carefully calculating the higher-order cumulants of the signed root empirical log-likelihood ratio statistic, the valid Edgeworth expansion can be established as a power series of order $O(n^{-1/2})$. Based on the expansion, the coverage error of the EL confidence region can be reduced from $O(n^{-1})$ to $O(n^{-2})$ using the Bartlett correction technique.

摘要

经验似然是一种做参数估计和假设检验的非参似然方法。经验似然的一个非常好的性质是巴特莱特纠正性。巴特莱特纠正性是对检验函数做一个简单的调整，使之建立的置信区域拥有更好的覆盖概率。以前的研究证明了对于独立同分布的数据和高斯分布的短记忆时间序列，经验似然可以被巴特莱特纠正。但是，对长记忆或非高斯分布的时间序列数据，经验似然是否拥有巴特莱特纠正性仍然未知。在这篇论文里，通过建立经验似然比平方根统计量的埃奇沃思展开表达式，我们建立了对高斯长记忆时间序列数据和非高斯短记忆时间序列数据，经验似然方法的巴特莱特纠正性。对于高斯长记忆时间序列，经验似然比平方根的埃奇沃思展开式是 $\log^3 n/\sqrt{n}$的幂级数。根据这个展开式，巴特莱特纠正方法使得经验似然置信区间的误差阶由 $O(\log^6 n/n)$ 减少到 $O(\log^3 n/n)$。对于非高斯短记忆时间序列数据，我们必须考虑离散傅里叶变换的高阶累积量。对于非高斯时间序列数据，通过计算经验似然比单位根统计量的高阶累积量，埃奇沃思展开式是$n^{-1/2}$ 的幂级数。根据这个展开式，巴特莱特纠正方法使得经验似然置信区域的误差阶由$O(n^{-1})$ 减少到$O(n^{-2})$。

Acknowledgments

I would like to thank my supervisor Prof. Chan, Ngai Hang and my co-supervisor Prof. Yau, Chun Yip for their invaluable guidance and assistance, without which the completion of this thesis would not have been possible. I am also sincerely grateful for their teaching, encouragement, and patience throughout the researching and writing of this thesis, which helped me to overcome so many obstacles.

I would like to thank my committee members, including Profs. Wong, Hoi Ying and Tony Sit, who provided so many insightful suggestions and comments on my research; and Prof. Peng Liang, whose work demonstrates the importance of empirical likelihood. I also acknowledge the staff and my classmates at the Department of Statistics, CUHK, for their tireless discussions with me on the essence of statistics from both theorectical and practical perspectives.

I am also grateful to my academic friends from the Department of Systems Engineering and Engineering Management, CUHK, who shared research topics and possible applications of statistics to engineering fields. Last, but not least, I would like to thank my parents, who have listened to my constant complaints, tried to understand my suffering and shared my happiest moments.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Empirical likelihood (EL) is a nonparametric likelihood approach for statistical inference introduced by Owen (1988, 1990, 2001). The EL function is formulated by appointing each observed datum with a given probability and maximizing the product probability under some constraints. One key property of the EL method for independent and identically distributed (i.i.d.) data is called "self-studentization" (see Hall and La Scala (1990)), meaning that the EL function automatically converges to a chi-squared distributed random variable without assuming any joint distribution of the data. Based on the chi-squared limiting distribution, confidence regions and hypothesis tests can be performed easily. Compared to other nonparametric methods such as Bootstrap, the EL method is an attractive alternative as it does not require resampling.

Most of the literature has shown that the EL method performs well in a variety of settings. By incorporating different constraints, the EL method, for i.i.d. data, has been extended to make inferences on estimating equations, linear regressions, and quantiles of distribution (see, for example, Qin and Lawless (1994), Owen (1991), Chen (1993)). Recently, many authors (see, for example, Hjort et al. (2009), Chen et al. (2009)) have investigated the EL method's performance in high-dimensional settings. The EL method has also been extended to accommodate serial correlation. Kitamura (1997) studied the blockwise EL method for weakly dependent time series. Chuang and Chan (2002) used the EL method with martingale estimating equations for unstable autoregressive models. Monti (1997) used a periodoram-based estimating function to apply the EL method formulated by Qin and Lawless (1994) to stationary short-memory time series (SMTS). It was later extended for use in a general setting with both SMTS and long-memory time series (LMTS) by Nordman and Lahiri (2006). For Gaussian SMTS, the periodogram ordinates are asymptotically independent (see Brillinger (1981)), such that the periodogram-based estimating functions are asymptotically independent. By mimicking independence of the estimating functions, the EL method also admits the "self-studentization" property of the time series.

Various approaches have been proposed to improve the accuracy of EL methods. Bartlett correction is one of the popular improvement techniques. It is a second-order improvement method that adjusts the empirical log-likelihood ratio statistic using a simple factor. Hall and La Scala (1990) demonstrated the Bartlett correction for EL inference on population mean, and Bartlett correction for smooth functions of means is established by DiCiccio, Hall and Romano (1991). Chen and Cui (2007) studied the Bartlett correctability of EL with over-identified estimating equations (i.e., the number of estimating functions is larger than that of parameters) in econometrics.

Nevertheless, the establishment of Bartlett correction for time series has received little consideration. Chan and Liu (2010) first developed Bartlett correction for Gaussian SMTS. However, in practice, many time series do not follow Gaussian distribution; in fact, some may even exhibit long-range dependence. In this thesis, we establish the Bartlett correctability of the EL method for Gaussian LMTS models and non-Gaussian SMTS models.

For Gaussian LMTS, the main problem is that the periodogram ordinates near the origin are no longer asymptotically independent (see, for example, Hurvich and Beltrao (1993), Robinson (1995)). Given this problem, the periodogram-based EL may fail to be Bartlett correctable. Specifically, the dependence introduces a bias of order $O(\log^3 n/n)$ for Gaussian LMTS, instead of order $O(n^{-1})$ for Gaussian SMTS. To tackle this, we establish an irregular Edgeworth expansion for the signed root empirical log-likelihood ratio function with a power series of $\log^3 n/\sqrt{n}$. Based on such expansion, the coverage error of the Bartlett corrected EL can be reduced from order $O(\log^6 n/n)$ to $O(\log^3 n/n)$.

It is well known that the higher-order cumulants of periodogram ordinates can be decomposed into products with the higher-order cumulants of the discrete Fourier transform (DFT), which are negligible for Gaussian processes (see, for example, Brillinger (1981), Priestley (1981)). For non-Gaussian SMTS, the main problem is that the higher order cumulants of DFT are no longer negligible. Thus, the periodogram-based EL method may fail to be Bartlett correctable. We show that for a stationary linear non-Gaussian SMTS, the cumulants of DFT with an even order larger than four or an odd order decay to zero with sufficiently small rates, and thus can be neglected. Therefore, only the fourth-order cumulant of the DFT requires consideration. Surprisingly, some calculations reveal that the non-negligible fourth-order cumulant can be canceled in the third-and the fourth-order cumulants of the signed root decomposition. This property ensures that the Edgeworth expansion is a power series of order $n^{-1/2}$. Based on the expansion, the coverage error of the EL confidence interval can still be reduced from $O(n^{-1})$ to $O(n^{-2})$ by Bartlett correction. We also conduct simulation studies to demonstrate the effectiveness of the periodogram-based EL calibration method when the underlying process exhibits both short- and long-range dependence, possibly non-Gaussian distributed.

In Chapter 2, we begin with a basic setting comprising EL with i.i.d. data to infer from the population mean, the smooth functions of means, and the parameters described by generalized estimating equations. Then, we review the EL method with dependent data.

In Chapter 3, we review the Bartlett correction for the EL method. We introduce some fundamental tools such as signed root empirical log-likelihood ratio statistic and the Edgeworth expansion technique, which are required to prove Bartlett correction. Establishing the validity of the Edgeworth expansion for the probability density function (p.d.f.) of the signed root empirical log-likelihood ratio statistic allows us to evaluate the coverage errors of EL confidence regions. The coverage errors of Bartlett-corrected EL confidence regions can also be derived using a similar procedure.

In Chapter 4, we use the periodogram-based EL for Gaussian LMTS to establish the valid Edgeworth expansion for the p.d.f. of the signed root empirical log-likelihood ratio statistic and its Bartlett corrected counterpart. This exploration provides some tools for further development on Bartlett correction with non-Gaussian time series.

In Chapter 5, we evaluate the higher-order cumulants of the DFT for non-Gaussian time series are non-negligible to establish a valid Edgeworth expansion for the signed root empirical log-likelihood ratio statistic. We show that the existence of higher-order cumulants of the DFT does not affect the Bartlett correctability of EL for non-Gaussian SMTS.

Chapter 6 concludes and explores possible future research areas.

Chapter 2

Empirical Likelihood Background

The empirical likelihood (EL) method is a nonparametric likelihood-based method applicable to versatile areas of statistics and econometrics. The idea of EL originated from Thomas and Prentice's (1980) work in survival analysis, and the formal methodology was developed by Owen (1988, 1990 and 1991). In this chapter, we review the EL approach for both independent and dependent data.

2.1 Empirical Likelihood for Independent Data

The EL method can be viewed as an extension of the parametric likelihood method in a nonparametric setting. Assume that $X_1, X_2, \ldots, X_n$ are $\mathbb{R}^d$ independent and identically distributed (i.i.d.) random vectors with a known non-singular covariance matrix $\mathrm{Var}(X_i)$. We consider constructing a confidence region for the unknown population mean $\mathrm{E}(X_i) = \mu \in \mathbb{R}^d$. If X_i follows a distribution with a probability density function (p.d.f.) of $f(x; \mu)$, the parametric likelihood ratio statistic for testing the null hypothesis $H_0 : \mu = \mu_0 \in \mathbb{R}^k$, $k < d$, versus $H_1 : \mu \neq \mu_0$ is given by

$$\frac{L(\mu_0)}{\max_\mu L(\mu)},$$

where $L(\mu) = \prod_{i=1}^{n} f(X_i, \mu)$ is the likelihood function for the mean. If the likelihood ratio function is small, the null hypothesis is rejected. Let $l(\mu) = \log L(\mu)$ be the log-likelihood function, and the log-likelihood ratio test statistic is defined by

$$LR(\mu_0) = 2\{l(\hat{\mu}) - l(\mu_0)\},$$

where $\hat{\mu} = \text{argmax}_\mu \, l(\mu)$ is the unconstrained maximum likelihood estimator (MLE). The parametric Wilks's theorem (see, Wilks (1938)) states that

$$LR(\mu_0) = 2\{l(\hat{\mu}) - l(\mu_0)\} \Rightarrow \chi^2_{d-k}, \quad \text{as} \quad n \to \infty, \tag{2.1}$$

where χ^2_d is a chi-squared distributed random variable with degrees of freedom d. Therefore, the threshold to reject μ_0 can be found from the critical value of χ^2_d.

The EL method replaces $L(\mu)$ with the nonparametric likelihood function supported by the observed data. Assume that X_i follows a common cumulative distribution function (c.d.f.) $F(x) = P(X \leq x) = \sum_{i=1}^{n} p_i \mathbf{1}_{\{X_i \leq x\}}$ and $F(x_-) = P(X < x) = \sum_{i=1}^{n} p_i \mathbf{1}_{\{X_i < x\}}$, where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function. Assigning a probability p_i to each point X_i, such that $\sum_{i=1}^{n} p_i = 1$, the nonparametric likelihood function is given by

$$L(F) = \prod_{i=1}^{n} (F(X_i) - F(X_{i^-})) = \prod_{i=1}^{n} p_i,$$

which is maximized when F is the empirical c.d.f., $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}}$ (i.e., $p_i = n^{-1}$). Analog to $LR(\mu_0)$, the nonparametric likelihood ratio function is defined by

$$\frac{L(F)}{L(F_n)} = \prod_{i=1}^{n} np_i.$$

The EL ratio function incorporates the information from the mean by the constrained maximization,

$$\mathcal{R}_n(\mu) = \max_{p_i} \left\{ \prod_{i=1}^{n} np_i \mid \sum_{i=1}^{n} p_i (X_i - \mu) = 0_d, \;\; \sum_{i=1}^{n} p_i = 1, \;\; p_i \geq 0 \right\},$$

where 0_d is the d-dimensional zero vector. Clearly, analogous to the parametric likelihood ratio test, the small values of $\mathcal{R}_n(\mu)$ reject the null hypothesis.

Analogous to (2.1), Owen (1988, 1990) proved the nonparametric Wilks's theorem for the empirical log-likelihood (log-EL) ratio function $-2\log\mathcal{R}(\mu_0)$ at the true value μ_0. That is,

$$-2\log\mathcal{R}_n(\mu_0) \Rightarrow \chi^2_d, \quad \text{as} \;\; n\to\infty. \tag{2.2}$$

Based on (2.2), the approximate confidence region for the mean at a confidence level $100(1-\alpha)\%$ is given by

$$\mathbf{I}_{n,1-\alpha} = \left\{\mu \mid -2\log\mathcal{R}_n(\mu) \le \chi^2_{d,1-\alpha}\right\},$$

where $\chi^2_{d,1-\alpha}$ is the upper α critical value of χ^2_d.

For example, the left plot of Figure 2.1 shows the empirical distribution of the simulated random variable following univariate standard normal distribution $N(0,1)$, with a sample size of $n=200$. The right plot of Figure 2.1 shows the log-EL ratio statistic at various mean values. The horizontal line shows the threshold with a value of $\chi^2_{1,0.95}$, indicating that the log-EL ratio function is below the threshold in the neighborhood of the mean zero.

The EL method can be extended to make inferences on the smooth function of means (see Hall (1992)), i.e., $\theta = h(\mu)$, where $h(\cdot)$ is a smooth function and μ is the mean. For example, the variance $\sigma^2 = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$ can be seen as the smooth function of the mean vector $(\mathrm{E}(X), \mathrm{E}(X^2))$. Then, the EL ratio function for the variance is

$$\mathcal{R}_n(\sigma^2) = \max_{p_i}\left\{\prod_{i=1}^n np_i \mid \sum_{i=1}^n p_i(X_i-\overline{X})^2 = \sigma^2,\; \sum_{i=1}^n p_i = 1,\; p_i\ge 0\right\}.$$

For example, we consider constructing a confidence interval for the volatility of the S&P 500 index in the financial stock market. In Figure 2.2, the daily return of the S&P 500 index for $n=256$ trading days and QQ-plot are plotted. The daily return of financial asset is defined as $r_t = \log P_t - \log P_{t-1}$, where P_t

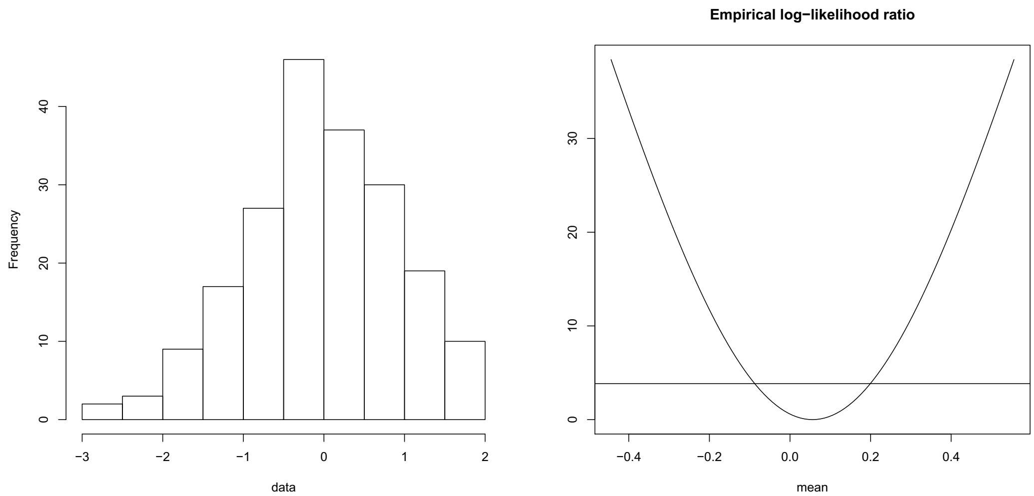

Figure 2.1: The left plot is 200 standard normal distributed data and the right plot is the empirical log-likelihood ratio function for the data.

represents the asset price. The volatility of an financial asset is defined as the sample standard deviation $(n^{-1}\sum_{t=1}^{n}(r_t-\overline{r})^2)^{-1/2}$. For the S&P 500 index, the annual volatility is $\hat{\sigma}=(\frac{256}{n}\sum_{t=1}^{n}(r_t-\overline{r}))^{-1/2}=0.7386$.

For simplicity, we assume that the returns are i.i.d.. The basic asymptotic normal theory tells us that

$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \Rightarrow \chi^2_{n-1}, \quad as \;\; n\to\infty. \tag{2.3}$$

Based on (2.3), the $(1-\alpha)100\%$ confidence interval for the volatility can be constructed through

$$\chi^2_{n-1,\alpha/2} < \frac{(n-1)\hat{\sigma}^2}{\sigma^2} < \chi^2_{n-1,1-\alpha/2}.$$

Here, for $\alpha = 0.05$, Table 2.1 compares the confidence interval formed by asymptotic normal theory and the EL method. It should be noted that the confidence interval based on the normal method is narrower than that based on the EL method. However, as the QQ-plot suggests that the distribution of the return is heavy-tailed, the EL method is more trust-worthy.

Qin and Lawless (1994) linked EL with generalized estimating equations, which provides a flexible way to incorporate the parameter information. For

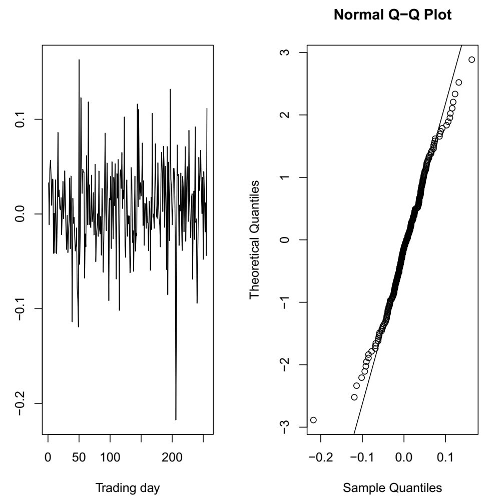

Figure 2.2: The left plot shows the S&P 500 daily return for 256 trading days. The right plot shows the QQ plot of the data, which indicates that the distribution of the return is heavy-tailed.

Method	Lower	Upper
Empirical Likelihood	0.6395	0.8275
Normal Theory	0.6797	0.8088

Table 2.1: 95% confidence interval for the volatility of S& P 500 index σ.

i.i.d. random vector $X \in \mathbb{R}^d$, and $\theta \in \Theta \subset \mathbb{R}^p$, the generalized estimating equation (GEE) $m(X, \theta): \mathbb{R}^d \times \Theta \to \mathbb{R}^k$ satisfies

$$\mathrm{E}(m(X, \theta_0)) = 0, \tag{2.4}$$

at true parameter value θ_0. The GEE estimator $\hat{\theta}$ is the solution of

$$\frac{1}{n}\sum_{i=1}^{n} m(X_i, \hat{\theta}) = 0. \tag{2.5}$$

Different estimating equations lead to different parameter estimators. For example, if we take $m(X, \theta) = X - \theta$, then (2.5) gives $\hat{\theta} = \overline{X}$; if we take $m(X, \theta) = 1_{\{X \in A\}} - \theta$, then (2.5) gives $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \in A\}}$. In particular, when $A = (-\infty, \theta]$ and $m(X, \theta) = 1_{\{X \in A\}} - \alpha$, then $\hat{\theta}$ is the sample quantile estimator. The EL ratio function is constructed by profiling the probability under GEE constraints,

$$\mathcal{R}_n(\theta) = \max_{p_i} \left\{ \prod_{i=1}^{n} np_i \mid \sum_{i=1}^{n} p_i m(X_i, \theta) = 0, \;\; \sum_{i=1}^{n} p_i = 1, \; p_i \geq 0 \right\}. \tag{2.6}$$

Qin and Lawless (1994) proved that

$$-2 \log \mathcal{R}_n(\theta_0) \Rightarrow \chi^2_k, \qquad \text{as} \quad n \to \infty. \tag{2.7}$$

For example, we construct the EL ration function for the quantile with read data in Owen (2001). Define the quantile $q(p)$ as $P(X \leq q(p)) = p$. In this case, the EL ratio function for $q(p)$ is

$$\mathcal{R}_n(q(p)) = \left(\frac{p}{\hat{p}}\right)^{n\hat{p}} \left(\frac{1-p}{1-\hat{p}}\right)^{n(1-\hat{p})},$$

where $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq q(p)\}}$. The data are pounds of milk produced by 22 dairy cows from Table 3.2 in Owen (2001). For $p = 0.5$, the sample median is $q(0.5) = 3527$. Also, the 85% sample quantile is $q(0.85) = 4628$ and the 10% sample quantile is $q(0.1) = 1932$.

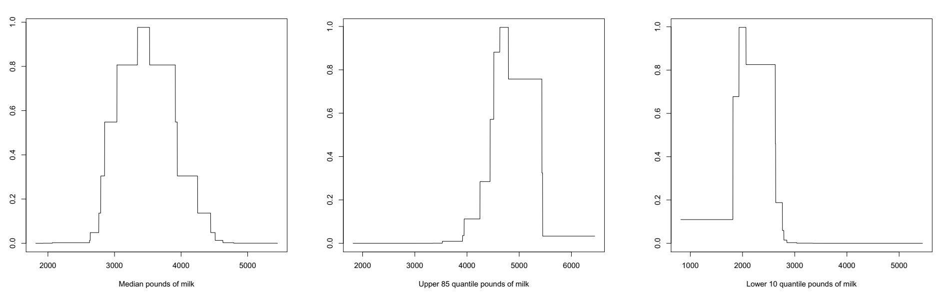

Figure 2.3: The empirical likelihood ratio function for the median, upper 85% and lower 10% quantiles of pounds of milk.

Estimating equations can also be defined through parametric likelihoods. Suppose $X_1, \ldots, X_n$ are i.i.d. random variables with density function $f(X, \theta)$. Consider the estimating function

$$m(X, \theta) = \frac{\partial}{\partial \theta} \log f(X, \theta) = \frac{g(X, \theta)}{f(X, \theta)}, \tag{2.8}$$

where $g(X,\theta) = \frac{\partial}{\partial\theta} f(X,\theta)$ is the first derivative of density function. In this case, $m(X,\theta)$ satisfies (2.4) and is called the score function. The parametric maximum likelihood estimator (MLE) is obtained by solving equation (2.5) with score function (2.8). The nonparametric Wilks's theorem for EL with independent score functions is a typical example in Qin and Lawless (1994).

Many authors have applied EL to make inferences from independent data in numerous areas. For example, Chen and Hall (1993) introduced EL to quantile estimation using a kernel smoothing technique, while Hjort et al. (2009) and Chen et al. (2009) developed EL for high-dimension data analysis and Kitamura (2001) used EL in econometric moment restriction testing problems.

2.2 Empirical Likelihood for Dependent Data

Recall that one key property of the EL method is "self-studentization". However, this property may fail for dependent data. To illustrate the problem, consider the example of the stationary time series in Kitamura (1997). Consider constructing a confidence region for the mean $\mu = \mathrm{E}(X_t)$ with stationary time series $X_1, \ldots, X_n \in \mathbb{R}^d$. Suppose that we treat X_i as independent and define the EL ratio function as

$$\mathcal{R}_n(\mu) = \max_{p_i} \left\{ \prod_{i=1}^{n} np_i \mid \sum_{i=1}^{n} p_i(X_i - \mu) = 0_d,\ \sum_{i=1}^{n} p_i = 1,\ p_i \geq 0 \right\}.$$

Under mild conditions, the log-EL ratio statistic evaluated at the true value μ_0 satisfies

$$-2\log \mathcal{R}_n(\mu_0) = n(\overline{X} - \mu_0)'\hat{\Sigma}^{-1}(\overline{X} - \mu_0) + o_p(1),$$

where $\hat{\Sigma} = n^{-1}\sum_{i=1}^{n}(X_i - \mu_0)(X_i - \mu_0)'$. If X_i are i.i.d., then $-2\log \mathcal{R}_n(\mu_0)$ has an asymptotic chi-squared distribution. However, for stationary time series, the covariance matrix $\hat{\Sigma}$ converges to $\mathrm{Var}(X_i)$ in probability, rather than the desired limit $\sum_{-\infty}^{\infty} \text{Cov}(X_i, X_{i-j})$. In this case, the log-EL function does not converge to the desired chi-squared limit, and the EL method fails.

2.2.1 Martingale Estimating Equations

A way to remedy the failure of the EL method in dependent processes is to establish a model (e.g., an autoregressive moving average (ARMA) model) to remove the dependence so that the residuals become independent. Suppose that a score function m: $\mathbb{R}^{d(q+1)} \times \Theta \to \mathbb{R}^k$ can be specified such that

$$m(X_t, \ldots, X_{t-q}; \theta_0) = \epsilon_t \,,$$

where $\{\epsilon_t\}$ is an i.i.d. (or more generally, a martingale difference sequence) process. Note that $m(X_t, \ldots, X_{t-q}; \theta_0)$ satisfies equation (2.4). For example, we specify a causal real-valued AR(q) model for X_t. That is, we can express X_t as $X_t = \phi_1 X_{t-1} + \cdots + \phi_q X_{t-q} + e_t$, where $\{e_t\}$ are i.i.d. with finite variance $\text{Var}(e_t)$. Inferring from the AR parameters $\theta = \{\phi_1, \ldots, \phi_q\} \in \Theta \subset \mathbb{R}^q$, the score function is

$$m(X_t, \ldots, X_{t-q}; \theta) = \{X_t - \sum_{i=1}^{q} \phi_i X_{t-i}\}(X_{t-1}, \ldots, X_{t-q})' \in \mathbb{R}^q.$$

It can be easily shown that $m(X, \theta)$, for $X = (X_t, \ldots, X_{t-q}, \theta)$ is a martingale difference sequence and satisfies equation (2.4). Based on this score function, the EL ratio statistic is

$$\mathcal{R}_n(\theta) = \max_{p_i} \left\{ \prod_{i=1}^{n} np_i \mid \sum_{i=q+1}^{n} p_i m(X_i, \ldots, X_{i-q}; \theta) = 0_q, \sum_{i=1}^{n} p_i = 1, p_i \geq 0 \right\}.$$

Chuang and Chan (2002) showed that when the roots of characteristic functions lie outside the unit circle (i.e., unstable AR process), profiling the $n - q$ data results that

$$-2 \log \mathcal{R}_n(\theta_0) \Rightarrow \chi_q^2, \quad \text{as} \quad n \to \infty.$$

The martingale estimating equation method is more natural than Kitamura's (1997) blockwise method, which is introduced in the next section, as it does not need to study the block size to reduce the dependence.

2.2.2 Blockwise Empirical Likelihood

Kitamura (1997) proposed a blockwise EL method as a way to capture the dependence of the underlying process. He noted that constructing a parametric model to remove the dependence is often too restrictive, and that the result may be sensitive to the specification of the unknown dependence structure. Thus, he proposed a blockwise EL method that preserves the dependence among the neighboring data and adjusts the log-EL ratio function to provide a valid confidence region.

Assume that $\{X_1, \ldots, X_n\} \in \mathbb{R}^d$ is a stationary realization with estimating functions $m(X,\theta): \mathbb{R}^d \times \Theta \subset \mathbb{R}^p \to \mathbb{R}^k$. Let the block length $l_n = l$ be an integer sequence satisfying $l^{-1} + l/n \to 0$ as $n \to \infty$; that is, the block length increases as the sample size increases, but the rate is slower. In this case, one block of data is $(X_t, \ldots, X_{t+l-1})$, for $t = l, \ldots, N = n - l + 1$. Instead of constructing EL with independent estimating functions, the blockwise EL is constructed by the block $\Phi_i(\theta) = \frac{1}{l}\sum_{j=i}^{i+l-1} m(X_i, \theta)$, for $i = 1, \ldots, N$. The blockwise EL ratio function is

$$\mathcal{R}_N(\theta) = \max_{p_i}\left\{\prod_{i=1}^{N} Np_i \mid \sum_{i=1}^{N} p_i\Phi_i(\theta) = 0_k.\ \sum_{i=1}^{N} p_i = 1,\ p_i \geq 0\right\}.$$

Here, the block data average $\Phi_i(\theta)$ can be treated as independent.

If the series $\{X_t\}$ satisfies the strong mixing condition: $\alpha_X(k) \to 0$, as $k \to \infty$, where $\alpha_X(k) = \sup_{A,B}|P(A\cap B) - P(A)P(B)|$, $A \in \mathcal{F}_{-\infty}^{0}$, $B \in \mathcal{F}_k^{\infty}$, and $\mathcal{F}_m^n = \sigma(X_i, m \leq i \leq n)$; $\sum_{k=1}^{\infty}\alpha_X(k)^{1-1/c} < \infty$ for $c > 1$; $l^{-1} + l^2/n \to 0$, as $n \to \infty$, then the log-EL ratio function admits the self-studentization property

$$-2l^{-1}\log\mathcal{R}_N(\theta_0) \Rightarrow \chi_k^2, \quad \text{as} \quad n \to \infty.$$

In particular, $-2\log\mathcal{R}_N(\theta_0)$ with $l=1$ reproduces the result of (2.7) for i.i.d. data.

One extension of the blockwise EL method is the tapered blockwise EL method, which introduces data tapers to define a smoothed blockwise EL. Inferring from population mean μ, the block estimating functions can be constructed as $\Phi_i(\mu) = \sum_{j=i}^{i+l-1}(X_j-\mu)/l$ as usual. Define the length l data taper sequence $w_l(1),\ldots,w_l(l)\in[0,1]$, and $\|w_l\|_1 \equiv \sum_{j=1}^{l} w_l(j)$. The tapered blockwise EL method replaces $\Phi_i(\mu)$ by tapered block estimating functions

$$T_i(\mu) \equiv \sum_{j=1}^{l} w_l(j)(X_{i+j-1}-\mu)/\|w_l\|_1, \tag{2.9}$$

for $i=1,\ldots,N$. The sequence of weights are formulated as

$$w_l(j) \equiv w\left(\frac{j-0.5}{l}\right), \quad j=1,\ldots,l,$$

where $w(t): \mathbb{R}\to[0,1]$, and $w(t)=0$, when $t\notin[0,1]$. To make the edges of the data tapers downweight the block data, $w(t)$ is symmetric about $\frac{1}{2}$, nondecreasing for $t\in[0,1/2]$. Downweighting the block data at the edges can reduce the dependence of the block data averages.

For example, we may use the trapezoidal taper $w_{trap}(t) = 2t\mathbf{1}_{[0,1/2]}(t) + 2(1-t)\mathbf{1}_{[1/2,1]}(t)$ or the cosine-bell taper $w_{cos}(t) = \frac{1-\cos(2\pi t)}{2}\mathbf{1}_{[0,1]}(t)$. Figure 2.4 plots the two data tapers, and shows that both data tapers achieve maximum at $t=1/2$ and decrease to zero at $t=0$ or $t=1$. If $w(t)=\mathbf{1}_{[0,1]}(t)$, the tapered block estimating function reduces to an un-tapered block estimating function.

More generally, the tapered blockwise EL ratio function for the smooth function of means is

$$\mathcal{R}_N(\theta) = \max_{p_i}\left\{\prod_{i=1}^{N} Np_i \mid h(\sum_{i=1}^{N} p_iT_i)=\theta.\ \sum_{i=1}^{N} p_i=1,\ p_i\geq 0\right\},$$

where $T_i=\sum_{j=i}^{l} w_l(j)X_{i+j-1}/\|w_l\|$. Under conditions that are similar to those of the blockwise EL method, Nordman (2009) proved the self-studentization

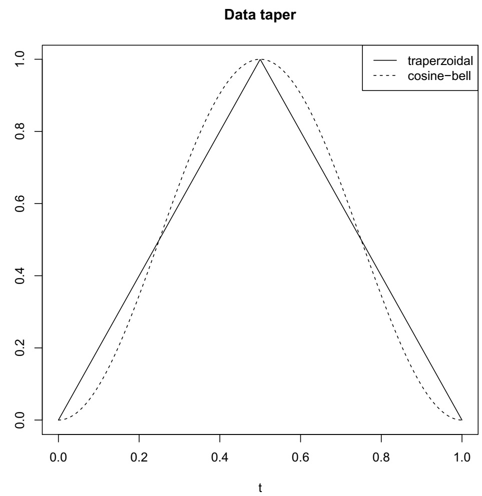

Figure 2.4: Data tapers for blockwise empirical likelihood.

property for the tapered blockwise EL,

$$-2a_N \log \mathcal{R}_N(\theta_0) \Rightarrow \chi^2_k, \quad \text{as} \quad n \to \infty, \tag{2.10}$$

where $a_N = \sum_{j=1}^{l} w_l(j)^2/(\sum_{j=1}^{l} w_l(j))^2$. In equation (2.10), the adjustment sequence a_n accounts for both overlapping blocks and the data taper. One attractive property of the tapered blockwise EL is its smaller coverage errors of EL confidence regions. For the blockwise EL, the coverage error is of an order $O(n^{-1/3})$, when the optimal rate of l is $O(n^{1/3})$. For the tapered blockwise EL, the coverage error is of an order $O(n^{-9/20})$, when the optimal rate of l is $O(n^{1/5})$.

2.2.3 Frequency Domain Empirical Likelihood

Another way to make inferences from dependent data using the EL method is to do the formulation in a frequency domain. Unlike the blockwise EL method, which uses the data blocks to capture the dependence, the frequency domain EL method uses the discrete Fourier transform (DFT) to weaken the dependence. It is well known that, under some regularity conditions, the DFTs at different frequencies are asymptotically uncorrelated (see Brillinger (1981), Priestley (1981)). By mimicking the periodogram ordinates as independent data, we can construct the approximate independent estimating functions. Incorporating these functions as EL constraints, we can formulate a frequency domain EL function for dependent processes. Monti (1997) first established the periodogram-based EL function for short-memory time series. Nordman and Lahiri (2006) extended Monti's formulation to a general setting for both SMTS and LMTS.

To introduce Monti's formulation, we consider a stationary linear process $\{X_t\}$ satisfying

$$X_t = \sum_{j=0}^{\infty} a_j(\theta)\epsilon_{t-j}, \tag{2.11}$$

where $\{\epsilon_t\}$ is an i.i.d. innovation process with a mean of zero and a finite variance of $\sigma_\epsilon^2 < \infty$. Let $\gamma_\theta(k) = \mathrm{Cov}(X_t, X_{t+k})$ be the autocovariance function at lag k. If the autocovariance function (ACVF) is summable, i.e., $\sum_{k=-\infty}^{\infty} \gamma_\theta(k) < \infty$, we call $\{X_t\}$ a SMTS. Assume further that the spectral density function of X_t given by

$$f(\omega, \theta) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_\theta(k) e^{-ik\omega}, \quad \omega \in \Pi = [-\pi, \pi], \tag{2.12}$$

where $i = \sqrt{-1}$, has a second-order continuous derivative on Π. Denote the normalized DFT of sample $\{X_1, \ldots, X_n\}$ as $J_n(\omega) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t e^{-it\omega}$, then the periodogram is $I_n(\omega) = J_n(\omega)J_n(-\omega)$. Monti formulated the EL function, for $\theta \in \mathbb{R}^p$ and $\omega_j = 2\pi j/n$ for $j = 1, \ldots, n$, with score function

$$m_j(\theta) = \frac{\partial \log\{f(\omega_j, \theta)\}}{\partial\theta} \left\{ \frac{I_n(\omega_j)}{f(\omega_j, \theta)} - 1 \right\}, \tag{2.13}$$

which is the score function of the Whittle likelihood (see Whittle (1953))

$$WL(\theta) = -\sum_{j=1}^{n} \log\{f(\omega_j, \theta)\} - \sum_{j=1}^{n} \frac{I_n(\omega_j)}{f(\omega_j, \theta)}.$$

Because the periodogram ordinate $I_n(\omega_j)$ is asymptotically i.i.d. under some conditions (see Brillinger (1981)), $m_j(\theta)$ is asymptotically i.i.d.. Based on

$m_j(\theta)$, the Whittle-type periodogram-based EL ratio function can be constructed as

$$\mathcal{R}_{n,\mathcal{F}}(\theta) = \max_{p_j} \left\{ \prod_{j=1}^{n} np_j \mid \sum_{j=1}^{n} p_j m_j(\theta) = 0, \ \sum_{j=1}^{n} p_j = 1, \ p_j \geq 0 \right\}. \qquad (2.14)$$

We use a standard Lagrange multiplier argument and obtain

$$-2 \log \mathcal{R}_{n,\mathcal{F}}(\theta) = 2 \sum_{j=1}^{n} \log(1 + t(\theta)' m_j(\theta)),$$

where $t(\theta)$ is the solution of

$$\sum_{j=1}^{n} \frac{m_j(\theta)}{1 + t(\theta)' m_j(\theta)} = 0_p.$$

It should be noted that, because

$$\mathrm{E}(m_j(\theta_0)) = O(n^{-1}),$$

the Whittle-type periodogram-based score function $m_j(\theta)$ fails to satisfy equation (2.4). Despite the order n^{-1} bias problem, Monti (1997) established the asymptotic chi-squared limit of the periodogram-based log-EL ratio function,

$$-2 \log \mathcal{R}_{n,\mathcal{F}}(\theta_0) \Rightarrow \chi_p^2, \quad \text{as} \quad n \to \infty.$$

Based on the chi-squared limit, one can construct confidence region and test hypothesis for SMTS.

Nordman and Lahiri (2006) formulated the frequency domain EL function in a more general setting. Assume that the parameter of interest θ is defined through estimating equation

$$\int_{-\pi}^{\pi} \Phi(\omega, \theta_0) f(\omega)\, d\omega = \phi_0 \in \mathbb{R}^k, \qquad (2.15)$$

for some known vector ϕ_0 and $\Phi(\omega, \theta): \mathbb{R} \times \Theta \to \mathbb{R}^k$. Typically, the constants ϕ_0 should equal 0_k. However, in special cases, ϕ_0 allows for other values.

The general frequency domain EL method can be combined with the Whittle estimation. Consider that $f(\omega, \theta)$ belongs to a parametric family of spectral densities $\mathcal{F} \equiv \{f(\omega,\theta), \theta \in \Theta\}$. The Whittle estimation aims to find θ_0, which minimizes $WL(\theta) = \int_{-\pi}^{\pi} \{\log f(\omega,\theta) + f(\omega)/f(\omega,\theta)\}\, d\omega$, where $f(\omega)$ is the true spectral density function. Consider a particular parametrization,

$$f(\omega,\theta) = \frac{\sigma_\epsilon^2}{2\pi} k_\beta(\omega), \quad \omega \in \Pi,$$

where $\theta = (\sigma_\epsilon^2, \beta)' \in (0,\infty) \times \mathbb{R}^{p-1}$ and k_β is the density kernel on Π. Note that k_β satisfies Kolmogorov's formula $\int_{-\pi}^{\pi} \log k_\beta(\omega)\, d\omega = 0$. Then, the minimizer θ_0 of $WL(\theta)$ solves

$$\int_{-\pi}^{\pi} \frac{\partial}{\partial\beta} k_{\beta_0}^{-1}(\omega) f(\omega)\, d\omega = 0_{p-1}, \quad \text{and} \quad \int_{-\pi}^{\pi} \frac{f(\omega)}{f(\omega,\theta_0)}\, d\omega = 2\pi.$$

In this case, (2.15) holds with

$$\phi_0 = (\pi, 0, \ldots, 0)' \in \mathbb{R}^p,$$

and

$$\Phi(\omega,\theta) = (f^{-1}(\omega,\theta), \frac{\partial}{\partial\beta} k_\beta^{-1}(\omega))' \in \mathbb{R}^p.$$

For inference on β, one should choose $\tilde{\Phi}(\omega,\theta) = \frac{\partial}{\partial\beta} k_\beta^{-1}(\omega) \in \mathbb{R}^{p-1}$, which satisfies (2.15) with $\phi_0 = 0_{p-1}$.

The general frequency domain EL ratio statistic with estimating equation (2.15) is

$$\mathcal{R}_n(\theta) = \max_{p_i} \left\{ \prod_{i=1}^{n} np_i \mid \sum_{i=1}^{n} p_i \Phi(\omega_i,\theta) I_n(\omega_i) = \phi_0,\ \sum_{i=1}^{n} p_i = 1,\ p_i \geq 0 \right\}. \tag{2.16}$$

The nonparametric Wilks's theorem holds for the general frequency domain EL.

Theorem: Suppose that $\{X_t\}$ is a linear process defined in (2.11) with $\mathrm{E}(\epsilon_t^2) > 0$, $\mathrm{E}(\epsilon_t^8) < \infty$ and $\sum_{j=-\infty}^{\infty} |a_j(\theta)| < \infty$. Suppose that $\int_{-\pi}^{\pi} \Phi(\omega,\theta_0) f(\omega)\, d\omega = \phi_0$ holds; and each component of $\Phi(\cdot,\theta_0)$ is a Lipschitz continuous function of an order greater than 1/2 on $[-\pi,\pi]$; and the $k \times k$ matrix $\int_{-\pi}^{\pi} f^2(\omega)\Phi(\omega,\theta_0)\Phi(\omega,\theta_0)'\, d\omega$ is full rank. If $\phi_0 = 0_k$, then

$$-2\log \mathcal{R}_n(\theta_0) \Rightarrow \chi_k^2 \quad \text{as} \quad n \to \infty.$$

Also, Nordman and Lahiri proposed a model-based version of the general frequency domain EL method, when the true spectral density $f(\omega)$ is assumed to lie in a parametric family $\mathcal{F} \equiv \{f(\cdot,\theta) : \theta \in \Theta\}$. The frequency domain EL ratio statistic is

$$\mathcal{R}_{n,\mathcal{F}}(\theta) = \max_{p_i} \left\{ \prod_{i=1}^{n} np_i \mid \sum_{i=1}^{n} p_i \Phi(\omega_i,\theta)(I_n(\omega_i) - f(\omega_i,\theta)) = 0,\ \ \sum_{i=1}^{n} p_i = 1,\ p_i \geq 0 \right\}.$$

Under some conditions, $-2\log \mathcal{R}_{n,\mathcal{F}}(\theta)$ has asymptotic chi-squared limiting distribution.

Theorem: In addition of assumptions on the theorem for $\mathcal{R}_n(\theta)$, suppose that $f(\cdot) = f(\cdot,\theta_0) \in \mathcal{F}$; and each component of $f(\cdot,\theta_0)m(\cdot,\theta_0)$ is a Lipschitz continuous function of an order greater than $1/2$ on $[-\pi,\pi]$. Then, if $\phi_0 = 0_k$ in (2.15),

$$-2\log \mathcal{R}_{n,\mathcal{F}}(\theta_0) \Rightarrow \chi^2_k, \qquad \text{as} \quad n \to \infty. \tag{2.17}$$

Furthermore, (2.17) still holds even for $\phi_0 \neq 0_k$, if $\kappa_{4,\epsilon} \equiv \mathrm{E}(\epsilon_t^4) - 3(\mathrm{E}(\epsilon_t^2))^2 = 0$.

One difference between $\mathcal{R}_n(\theta)$ and $\mathcal{R}_{n,\mathcal{F}}(\theta)$ is that the formulation of $\mathcal{R}_{n,\mathcal{F}}(\theta_0)$ requires the true spectral density $f(\omega)$ to belong to a model class $f(\omega) = f(\omega,\theta_0) \in \mathcal{F}$. In contrast, $\mathcal{R}_n(\theta)$ only requires the spectral moment condition (2.15). The other difference is that the chi-squared limit of $\mathcal{R}_{n,\mathcal{F}}(\theta)$ still holds when $\phi_0 \neq 0_k$. In this case, we require that $\kappa_{4,\epsilon} = 0$ (i.e. $\{\epsilon_t\}$ is a Gaussian process).

Monti's Whittle-type periodogram-based EL method is linked to the model-based version of the frequency domain EL method through the score functions

$$m_j(\theta) = \Phi(\omega_j,\theta)(I_n(\omega_j) - f(\omega_j,\theta)),$$

where $\Phi(\omega_j,\theta) = \frac{\partial}{\partial\theta} f^{-1}(\omega_j,\theta)$.

Treating the periodogram collections $\{I_n(\omega_j) : j \in \mathbb{S}_n\}$ as approximately independent in the bootstrap context is discussed in Hurvich and Zeger (1987), and Kreiss and Paparoditis (2011). However, the dependence among the periodogram collection creates problems in estimation when applied to the Whittle score function (see Dahlhaus and Janas (1996)). Specifically, the dependence between different periodogram ordinates for non-Gaussian processes leads to non-negligible terms of order $O(n^{-1})$, which are negligible for Gaussian processes. Based on a different score function, Ogata (2005) considered EL for non-Gaussian stationary processes. In Chapters 4 and 5, our discussion of the Bartlett correction mainly relies on the formulation of the Whittle-type periodogram-based EL ratio function defined in (2.14).

Chapter 3

Bartlett Correction Background

As noted in Chapter 2, the key property of the empirical likelihood (EL) method is "self-studentization", meaning that the test statistic formulated by the empirical log-likelihood (log-EL) ratio function automatically converges to a chi-squared distribution without assuming any joint distribution of the data. Based on this property, the confidence region can be constructed using the critical value of the chi-squared distribution. Therefore, a natural question arises: is the chi-squared approximation good enough, or can we correct the critical values and get a more accurate approximation?

In the following sections, we review the Bartlett correction method under the parametric and nonparametric likelihood (i.e., EL) settings. Section 3.1 presents the method for three types of statistics: the log-likelihood ratio (LR) test, the score (S) test, and the Wald (W) test. Section 3.2 reviews the Bartlett correction for EL with independent and identically distributed (i.i.d.) data.

3.1 Bartlett Correction for Parametric Likelihood

3.1.1 Bartlett Correction for the Log-Likelihood Ratio Statistic

Recall that the log-likelihood ratio statistic for $\theta \in \Theta \subset \mathbb{R}^p$ is

$$LR(\theta_0) = 2\{l(\hat{\theta}) - l(\theta_0)\},$$

where $l(\theta) = \log f(X_1, \ldots, X_n; \theta)$ denotes the log-likelihood; $f(\cdot)$ is the probability density function (p.d.f.); and $\hat{\theta} = \text{argmax}_\theta\, l(\theta)$ is the unconstrained maximum likelihood estimator (MLE). Under the null hypothesis H_0: $\theta = \theta_0 \in \mathbb{R}^k$, $k \le p$, the parametric Wilks's theorem shows that

$$LR(\theta_0) = 2\{l(\hat{\theta}) - l(\theta_0)\} \Rightarrow \chi^2_{p-k}, \quad \text{as} \quad n \to \infty. \tag{3.1}$$

Given some higher-order asymptotic expansion techniques, it is not difficult to derive

$$P\{LR(\theta_0) \le x\} = P\{\chi^2_q \le x\}\{1 + O(n^{-1})\}. \tag{3.2}$$

Bartlett (1937) first proposed a method to improve the approximation (3.2) so that a faster convergence rate can be obtained. To introduce this method, given

$$\mathrm{E}(LR(\theta_0)) = q(1 + b/n + O(n^{-2})),$$

where b is a constant that can be consistently estimated, and $q = p - k$, we can scale $LR(\theta_0)$ by $1 + b/n$, such that $\mathrm{E}(LR(\theta_0)/(1 + b/n))$ approximates more accurately to the mean of χ^2_q. This simple adjustment is called a "Bartlett correction", and b here is called the "Bartlett correction factor". After the adjustment, the cumulative distribution function (c.d.f.) of

$LR^*(\theta_0) = LR(\theta_0)/(1+b/n)$ converges faster than the c.d.f. of $LR(\theta)$,

$$P\{LR^*(\theta_0) \leq x\} = P\{\chi_q^2 \leq x\}\{1 + O(n^{-2})\}.$$

By showing that the higher-order cumulants of the derivatives of log-likelihood functions are more equivalent to those of the chi-squared distribution, Lawley (1985) first formally proved that the LR method is Bartlett correctable. Beale (1960) interpreted the Bartlett correction factor in relation to the curvature of the surface in normal regression models, which was extended to non-normal models by McCullagh and Cox (1986).

An important extension of Lawley's result was developed by Hayakawa (1977), who obtained the asymptotic expansion for the distribution of LR under H_0 against a composite alternative hypothesis H_1. Specifically, he showed that

$$\begin{aligned} P\{LR(\theta_0) \leq x\} &= F_q(x) + \frac{1}{24n}[A_2 F_{q+4}(x) - (2A_2 - A_1)F_{q+2}(x) \\ &\quad + (A_2 - A_1)F_q(x)] + O(n^{-3/2}), \end{aligned} \tag{3.3}$$

where $F_q(x) = P(\chi_q^2 \leq x)$. A_1 and A_2 are functions of the cumulants of derivatives of the log-likelihood function. The error $O(n^{-3/2})$ in (3.3) is always $O(n^{-2})$ (see Barndorff-Nielson and Hall (1988)). For general statistics, $A_2 = 0$ and thus the Bartlett correction factor admits a simple form $b = A_1/(12q)$. Formula (3.3) is widely applicable to models with nuisance parameters, and to inference with independent but not identically distributed data.

The Bartlett correction for likelihood ratio test statistics has been extended in many ways. Cordeiro (1983, 1987) derived a closed-form of the Bartlett correction factor in generalized linear models. Attfield (1991, 1995) applied Bartlett correction to the LR test for homoskedasticity in linear models and systems of equations, respectively. DiCiccio (1984) proved the Bartlett correction for the signed root log-likelihood ratio statistic $LR^{1/2}$. McCullagh and Cox (1986) expressed the Bartlett correction factor as the invariant combinations of cumulants of the first two derivatives of the log-likelihood and gave it a geometric interpretation. Bickel and Ghosh (1990) considered Bartlett correction in a Bayesian framework. Specifically, the posterior distribution of the LR statistic converges to that of the referenced chi-squared distribution with error $O(n^{-1})$, and is reduced to order $O(n^{-2})$ by Bartlett correction. Based on the marginal posterior p.d.f. given by Tierney et al. (1989), DiCiccio and Stern (1993) derived an explicit formula for the Bartlett correction factor in a general Bayesian framework.

Finally, it is not generally true that the LR method is Bartlett correctable with discrete data. Frydenberg and Jensen (1989) presented extensive numerical results showing that Bartlett correction does not always produce an order $O(n^{-2})$ error for lattice distributed data. In addition, Bartlett correctability does not generally hold for the Score and Wald statistics.

3.1.2 Bartlett-Type Correction for the Score Statistic

Assume that $X_1, \ldots, X_n$ are i.i.d., and define $U(\theta) = \nabla_\theta l(\theta) = [\partial l(\theta)/\partial\theta_j]$ for $j = 1, \ldots, p$, where $\theta \in \mathbb{R}^p$, as the first derivative of a log-likelihood; $\mathrm{Var}(U(\theta)) = \mathrm{E}(U(\theta)U(\theta)') = \mathrm{E}(-\nabla_\theta\nabla_\theta' l(\theta)) = \{\mathrm{E}(-\partial^2 l(\theta)/\partial\theta_i\partial\theta_j)\} = \mathcal{I}(\theta)$ for $i, j = 1, \ldots, p$, as the Fisher information matrix. The Rao score statistic for θ is

$$S(\theta) = U'(\theta)\mathcal{I}^{-1}(\theta)U(\theta).$$

It is well known that, under H_0: $\theta = \theta_0 \in \mathbb{R}^k$, for $k \leq p$,

$$S(\theta_0) = U'(\theta_0)\mathcal{I}^{-1}(\theta_0)U(\theta_0) \Rightarrow \chi^2_{p-k}, \quad \text{as} \quad n \to \infty.$$

Harris (1985) derived an asymptotic expansion for the null distribution of $S(\theta)$ in the presence of nuisance parameters,

$$\begin{aligned} P(S(\theta_0) \leq x) &= F_q(x) + \frac{1}{24n}[A_3F_{q+6}(x) + (A_2 - 3A_3)F_{q+4}(x) + (3A_3 - 2A_2 \\ &\quad + A_1)F_{q+2}(x) + (A_2 - A_1 + A_3)F_q(x)] + o(n^{-1}), \end{aligned} \qquad (3.4)$$

where $F_q(x) = P(\chi_q^2 \leq x)$, for $q = p - k$, and A_1, A_2 and A_3 are functions of higher cumulants of log-likelihood derivatives. Obviously, equation (3.4) implies that it is impossible to scale $S(\theta_0)$ by a linear transformation, which corrects all of the cumulants of $S(\theta_0)$ to smaller orders. To improve the chi-squared approximation in (3.4), Cox (1988) first proposed a new correction method, which was later generalized by Cordeiro and Ferrari (1991). They showed that the corrected score statistic is

$$S^*(\theta) = S(\theta)\left\{1 - \frac{1}{n}\sum_{j=1}^{3} r_j S^{j-1}(\theta)\right\}, \qquad (3.5)$$

where $r_1 = (A_1 - A_2 + A_3)/(12q)$, $r_2 = (A_2 - 2A_3)/(12q(q+2))$ and $r_3 = A_3/(12q(q+2)(q+4))$. Under regularity conditions,

$$P(S^*(\theta_0) \leq x) \quad = \quad F_q(x) + o(n^{-1}).$$

The non-linear correction method in (3.5) is called the "Bartlett-type correction". In conclusion, the Bartlett-type correction for the score statistic reduces the convergence rate from $O(n^{-1})$ to $o(n^{-1})$.

3.1.3 Bartlett-Type Correction for the Wald Statistic

The Wald statistic under H_0: $\theta = \theta_0 \in \mathbb{R}^k$, for $k < p$, is

$$W(\theta_0) = (\hat{\theta} - \theta_0)'\mathcal{I}(\theta_0)(\hat{\theta} - \theta_0),$$

where $\hat{\theta}$ is unconstrained MLE. Under some regularity conditions,

$$W(\theta_0) = (\hat{\theta} - \theta_0)'\mathcal{I}(\theta_0)(\hat{\theta} - \theta_0) \Rightarrow \chi^2_{p-k}, \quad \text{as} \quad n \to \infty.$$

Phillips and Park (1988) obtained an Edgeworth expansion for $W(\theta_0)$,

$$\begin{aligned} P(W(\theta_0) \leq x) \quad = \quad & F_q(x) + \frac{1}{n}[a_3 F_{q+6}(x) + a_2 F_{q+4}(x) \\ & + a_1 F_{q+2}(x) + a_0 F_q(x) + b_0 f_q(x)] + o(n^{-1}), \end{aligned} \qquad (3.6)$$

where $f_q(x) = \partial F_q(x)/\partial x$; and a_js are functions of the higher-order cumulants of the derivatives of the log-likelihood function. Note that expansion (3.6) is different from expansion (3.3) due to the non-negligible term $f_q(\cdot)$. But Phillips and Park showed that, for general statistics, this term equals zero.

Analogous to the score statistic, the Bartlett-type correction for the Wald statistic is

$$W^*(\theta_0) = W(\theta_0)\left\{1 - \frac{1}{n}\sum_{j=0}^{3} r_j W^{j-1}(\theta_0)\right\}, \tag{3.7}$$

where r_j are functions of a_j in (3.6). Under regularity conditions,

$$P(W^*(\theta_0) \leq x) = F_q(x) + o(n^{-1}).$$

In conclusion, the convergence rate of the score statistic to the chi-squared limit is reduced from $O(n^{-1})$ to $o(n^{-1})$ by the Bartlett-type correction technique.

3.1.4 Numerical Studies on Bartlett Correction for Parametric Likelihood

In this simulation study, we perform a finite sample comparison of two types of parametric likelihood test statistics: the log-likelihood ratio (LR) statistic and the score (S) statistic. We simulate the sample $X_1, \ldots, X_n$ from univariate distribution $N(\mu, \sigma^2)$, with the sample sizes $n = 20$, $n = 50$ and $n = 100$. The nominal type-I error rates are $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$, respectively. We are interested in testing for variance H_0: $\sigma^2 = \sigma_0^2 \in \mathbb{R}$, against a two-sided alternative. The true values of the mean and the variance are $\mu_0 = 0$ and $\sigma_0^2 = 1$, respectively.

For the LR test statistic, Hayakawa (1977) gave the Edgeworth expansion under H_0,

$$P(LR(\sigma_0^2) \leq x) = F_f(x) + \frac{1}{24n} d(2d^2 + 3d - 1)[F_{f+2}(x) - F_f(x)] + o(1/n),$$

where d is the difference in dimension between the null and alternative distribution and $f = d(d+1)/2$. For a real-valued process $\{X_t\}$, the Bartlett correction factor is $b = 1/3$.

n		$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
20	LR	0.8711	0.9343	0.9852
	Bart. LR	0.8722	0.937	0.9857
50	LR	0.8822	0.9368	0.9867
	Bart. LR	0.8826	0.9372	0.9871
100	LR	0.8875	0.952	0.988
	Bart. LR	0.895	0.9514	0.988

Table 3.1: Coverage accuracy of LR test for normal distribution variance σ^2, replications = $10,000$.

Table 3.1 shows that the coverage probabilities of LR and Bartlett-corrected LR increase with the sample size. For different sample size and confidence levels $(1-\alpha)100\%$, the Bartlett correction successfully increases the coverage probabilities for the LR test statistic. The same conclusion about the Bartlett-type correctability for the score test statistic is confirmed by the finite sample coverage probabilities in Table 3.2.

For the score test, the corrected statistic for σ^2 using (3.5) is

$$S^*(\sigma_0^2) = S(\sigma_0^2)\left\{1 - \frac{1}{n}\sum_{j=1}^{3} r_j S^{j-1}(\sigma_0^2)\right\}.$$

For the i.i.d. data from the univariate normal distribution $N(\mu, \sigma^2)$ with $q = 1$, one can obtain that $A_1 = -6$, $A_2 = 12$ and $A_3 = 40$. Thus, the polynomials r_j in (3.5) are $r_1 = 11/6$, $r_2 = -17/9$ and $r_3 = 2/9$.

n		$\alpha = 0.15$	$\alpha = 0.1$	$\alpha = 0.05$
20	S	0.8574	0.9158	0.9906
	Bart. S	0.8474	0.9014	0.9482
100	S	0.853	0.9064	0.9558
	Bart. S	0.816	0.898	0.9534

Table 3.2: Coverage accuracy of score test for normal distribution univariate variance σ^2, replications $= 5,000$.

3.2 Bartlett Correction for Empirical Likelihood with Independent Data

Introduced in Section 2, the parameter of interest θ can be defined through a smooth function of means $\theta = h(\mu) \in \mathbb{R}^d$, where $h(\cdot)$ is a smooth function. For inference on θ, the log-EL ratio function admits the nonparametric Wilks's theorem, i.e., for $\theta = \theta_0$

$$-2\log \mathcal{R}_n(\theta_0) \Rightarrow \chi^2_d, \quad \text{as} \quad n \to \infty,$$

where $\mathcal{R}_n(\theta_0)$ is defined as

$$\mathcal{R}_n(\theta_0) = \max_{p_i} \left\{ \prod_{i=1}^n np_i \mid h\left(\sum_{i=1}^n p_i X_i\right) = \theta, \ \sum_{i=1}^n p_i = 1, \ p_i \geq 0 \right\}.$$

Under regularity conditions plus the Cramér's condition

$$\limsup_{\tau \to \infty} |\mathrm{E}(\exp(i\tau X_t))| < \infty,$$

it follows that

$$P(-2\log \mathcal{R}_n(\theta_0) \leq \chi^2_{d,1-\alpha}) = 1 - \alpha + O(n^{-1}). \tag{3.8}$$

If the Cramér's condition does not hold, the coverage error in (3.8) should be $O(n^{-1/2})$ (see Owen (2001)). When

$$\mathrm{E}(-2\log \mathcal{R}_n(\theta_0)) = 1 - \frac{b}{n} + o(n^{-1}),$$

The Bartlett correction of EL replaces $-2\log\mathcal{R}_n(\theta)$ by $(1+b/n)-2\log\mathcal{R}_n(\theta)$ or equivalently, replaces the critical value $\chi^2_{d,1-\alpha}$ by $(1-b/n)\chi^2_{d,1-\alpha}$. Here, b is the Bartlett correction factor. Then, the Bartlett corrected EL confidence region becomes

$$\mathbf{I}_{n,1-\alpha} = \left\{\theta \mid -2\log\mathcal{R}_n(\theta) \le (1-\frac{b}{n})\chi^2_{d,1-\alpha}\right\}.$$

DiCiccio, Hall and Romano (1991) surprisingly showed that the coverage probability of $\mathbf{I}_{n,1-\alpha}$ approximates more accurately to nominal level $1-\alpha$. Generally, b is unknown and can be consistently estimated. One can use the plug-in estimator $\hat{b}_n$, which leads to an $\sqrt{n}$-consistent estimate of b. Based on $\hat{b}_n$, the Bartlett-corrected EL confidence region becomes

$$\mathbf{I}^*_{n,1-\alpha} = \left\{\theta \mid -2\log\mathcal{R}_n(\theta) \le (1-\frac{\hat{b}_n}{n})\chi^2_{d,1-\alpha}\right\}.$$

In this case,

$$P(\theta \in \mathbf{I}_{n,1-\alpha}) = P(\theta \in \mathbf{I}^*_{n,1-\alpha}) = 1-\alpha+O(n^{-2}).$$

Specifically, inferring from mean $\theta=\mu\in\mathbb{R}$, the Bartlett correction factor is

$$b = \frac{1}{2}\frac{\mu_4}{\mu_2^2} - \frac{1}{3}\frac{\mu_3^2}{\mu_2^3} = \frac{\kappa+3}{2} - \frac{\gamma^2}{3},$$

where $\mu_k = \mathrm{E}(X-\mathrm{E}(X))^k$ and γ and κ are the skewness and the excess kurtosis, respectively. The Bartlett-corrected EL confidence interval for the mean is

$$\mathbf{I}_{n,1-\alpha} = \left\{\mu \mid -2\log\mathcal{R}_n(\mu) \le (1-\frac{b}{n})\chi^2_{1,1-\alpha}\right\}.$$

In the simulation study of the Bartlett correction for EL inference on the univariate mean, we generate $X \overset{i.i.d.}{\sim} N(\mu,\sigma^2)$, with $\mu_0=0$ and $\sigma_0^2=1$. The results are based on different sample sizes $n=20$ and $n=50$, with the nominal type-I error rates are $\alpha=0.01$, $\alpha=0.05$ and $\alpha=0.1$.

The finite sample results of Table 3.3 clearly show that Bartlett correction successfully improves the coverage accuracies of EL inference on mean.

n		$\alpha = 0.15$	$\alpha = 0.1$	$\alpha = 0.05$
20	EL	0.8374	0.8882	0.8792
	Bart. EL	0.8564	0.9026	0.8918
50	EL	0.846	0.8937	0.9446
	Bart. EL	0.8488	0.8999	0.9482

Table 3.3: Coverage accuracy of EL test for normal distribution univariate mean μ, replications $= 5,000$.

The theorectical proofs of Bartlett correction for EL mainly base on the Edgeworth expansion technique. The validity of Edgeworth expansion is established in Bhattacharya and Ghosh (1978) by assuming

$$\mathrm{E}(|X|^{j+2}) < \infty, \qquad \text{and} \quad \limsup_{|t|\to\infty} |\varphi(t)| < 1.$$

The former condition ensures finite moment existence and the latter restriction is Cramér's condition. For certain estimates $\hat{\theta}$ and the true parameter value θ_0, the distribution function of $\sqrt{n}(\hat{\theta}-\theta_0)/\sigma$ can be expanded as a power series of $n^{-1/2}$ (see Hall (1992)), i.e.,

$$P(\sqrt{n}(\hat{\theta}-\theta_0)/\sigma \leq x) = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + \cdots + n^{-j/2}p_j(x)\phi(x) + \cdots, \tag{3.9}$$

where $\phi(x)$ is p.d.f. of standard normal distribution and $\Phi(x) = \int_{-\infty}^{x} \phi(t)\, dt$. The expansion (3.9) is called Edgeworth expansion. The coefficients of Edgeworth polynomials $p_j(x)$ depend on the cumulants of $\sqrt{n}(\hat{\theta}-\theta)/\sigma$. In addition, $p_j(x)$ is of degree $3j-1$, even/odd when j is odd/even.

For example, consider inference on the population mean, when $\theta = \mu$. Given the i.i.d. real-valued random variables $X_1, \ldots, X_n$ with known variance $\sigma^2 = \mathrm{Var}(X_i)$, not necessarily normal distributed, one may use $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$. By the central limit theorem, $S_n = \sqrt{n}(\hat{\theta}-\theta)/\sigma \Rightarrow N(0,1)$. Based on the asymptotic normal distribution, the confidence interval with nominal level $1-\alpha$

is

$$\left\{\theta \mid \theta \in (\hat{\theta} - n^{-1/2}\sigma N_{1-\alpha},\ \hat{\theta} + n^{-1/2}\sigma N_{1-\alpha})\right\},$$

where $N_{1-\alpha}$ is the upper α critical value of the standard normal distribution. The quality of the normal approximation can be described through the characteristic function (ch.f.). Given $Y_i = (X_i - \mu)/\sigma$ and $S_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$, it follows that

$$\mathrm{E}(e^{itS_n}) = \varphi_n(t) \to \mathrm{E}(e^{itN}) = e^{-t^2/2}, \quad \text{as} \quad n \to \infty.$$

Here, $\varphi_n(t) = [\varphi(t/\sqrt{n})]^n$, where $\varphi(t)$ is the ch.f. of Y_i. The ch.f. of S_n can thus be expanded as

$$\varphi_n(t) = e^{-t^2/2}\left\{1 + n^{-1/2}r_1(it) + n^{-1}r_2(it) + \cdots + n^{-j/2}r_j(it) + \cdots\right\}, \quad (3.10)$$

where $r_j(\cdot)$ is a polynomial with degree $3j$, and even/odd when j is even/odd. Taking $\kappa_j(X)$ as the j-th cumulant of the random variable X, we have $\kappa_1(Y_i) = 0$ and $\kappa_2(Y_i) = 1$. After some algebra, it can be derived that $r_1(x) = \frac{1}{6}\kappa_3(Y)x^3$, and $r_2(x) = \frac{1}{24}\kappa_4(Y)x^4 + \frac{1}{72}\kappa_3(Y)^2x^6$ in (3.10). By definition

$$\varphi_n(t) = \int_{-\infty}^{\infty} e^{itx}\, dP(S_n \le x)$$

and formula (3.10), $P(S_n \le x)$ can be expanded as

$$P(S_n \le x) = \Phi(x) + n^{-1/2}R_1(x) + n^{-1}R_2(x) + \cdots + n^{-j/2}R_j(x) + \cdots, \quad (3.11)$$

where $R_j(x) = p_j(x)\phi(x)$ and

$$\int_{-\infty}^{\infty} e^{itx}\, dR_j(x) = r_j(it)e^{-t^2/2}.$$

Using the well known formula

$$\int_{-\infty}^{\infty} e^{itx}\, d[r_j(-\partial/\partial x)\Phi(x)] = r_j(it)e^{-t^2/2},$$

it follows that

$$R_j(x) = r_j(-\partial/\partial x)\Phi(x).$$

In particular, for $j = 1, 2$,

$$R_1(x) = -\frac{1}{6}\kappa_3(Y)(x^2 - 1)\phi(x), \tag{3.12}$$

and

$$R_2(x) = -x\left[\frac{1}{24}\kappa_4(Y)(x^2 - 3) + \frac{1}{72}\kappa_3(Y)^2(x^4 - 10x^2 + 15)\right]\phi(x). \tag{3.13}$$

Then, the valid Edgeworth expansion to order $O(n^{-3/2})$ is

$$P(S_n \le x) = \Phi(x) + n^{-1/2}R_1(x) + n^{-1}R_2(x) + O(n^{-3/2}),$$

where $R_1(x)$ and $R_2(x)$ are given in (3.12) and (3.13).

Based on the Edgeworth expansion technique, Bartlett correction can be applied to improve the EL method with independent estimating functions. Given the i.i.d. real-valued estimating functions $m(X, \theta)$: $\mathbb{R}^d \times \Theta \to \mathbb{R}$, the EL ratio function is

$$\mathcal{R}_n(\theta) = \max_{p_i}\left\{\prod_{i=1}^n np_i \mid \sum_{i=1}^n p_i m(X_i, \theta) = 0,\ \ \sum_{i=1}^n p_i = 1,\ p_i \ge 0\right\}.$$

Under regularity conditions, $\mathrm{E}(-2\log\mathcal{R}_n(\theta_0)) = 1 + b/n + O(n^{-2})$ for some constant b. Following the previous argument, the Bartlett-corrected EL confidence interval is given by

$$\mathbf{I}_{n,1-\alpha} = \left\{\theta \mid -2\log\mathcal{R}_n(\theta_0) \le \left(1 + \frac{b}{n}\right)\chi^2_{1,1-\alpha}\right\}.$$

One step in the proof is the formulation of the signed root empirical log-likelihood ratio statistic. Define the signed root empirical log-likelihood ratio $SR = R_1 + R_2 + R_3$ as

$$-2\log\mathcal{R}_n(\theta_0) = nSR^2 + O_p(n^{-3/2}) = n(R_1 + R_2 + R_3)^2 + O_p(n^{-3/2}),$$

where $R_1 = O_p(n^{-1/2})$, $R_2 = O_p(n^{-1})$ and $R_3 = O_p(n^{-3/2})$. The specific expressions of R_j can be found in Zhang (1996). The idea is that the square root of an asymptotic chi-squared function should converge to the normal distribution. Here, SR can be treated as $\hat{\theta} - \theta_0$ in the classic parametric setting. The normalized variable $\sqrt{n}SR/\mathrm{Var}(\sqrt{n}SR)$ converges to $N(0,1)$ in distribution, and by tedious calculations of the higher-order cumulants of $\sqrt{n}SR$, the valid Edgeworth expansion for the p.d.f. $\pi(x)$ of $\sqrt{n}SR$ can be established as

$$\pi(x) = \phi(x)+n^{-1/2}r_1(x)\phi(x)+n^{-1}r_2(x)\phi(x)+n^{-3/2}r_3(x)\phi(x)+O(n^{-2}), \quad (3.14)$$

where r_1 and r_3 are odd polynomials and r_2 is an even polynomial of degree 2. The coefficients of r_j involve higher-order cumulants of $m(X_i,\theta)$. Unlike general statistics, r_2 does not involve terms of degrees 4 and 6. The feature results from the fact that $\kappa_3(\sqrt{n}SR) = O(n^{-3/2})$ and $\kappa_4(\sqrt{n}SR) = O(n^{-2})$.

Furthermore, it can be shown that

$$\mathrm{E}(-2\log \mathcal{R}_n(\theta_0)) = 1 + \frac{b}{n} + O(n^{-3/2}), \quad (3.15)$$

where b is a function of higher-order cumulants of $m(X_i,\theta)$. Equation (3.15) implies that scaling the log-EL ratio statistic by the mean can improve the coverage accuracy. As DiCiccio, Hall and Romano (1991) noted, the EL method is Bartlett correctable because $r_2(x)$ in (3.14) is of degree 2. In particular, $r_2(x) = \frac{b}{2}(x^2-1)$. With this form of $r_2(x)$, terms of order n^{-1} in (3.14) can be removed through the simple adjustment. Let $c_\alpha = \chi^2_{1,1-\alpha}$, $c'_\alpha = c_\alpha(1+b/n)$, and $g(\cdot)$ denotes the p.d.f. of the χ^2_1 distribution. Applying the Edgeworth expansion (3.14) gives us

$$\begin{aligned} P(\theta_0 \in \mathbf{I}_{n,1-\alpha}) &= P(-2\log\mathcal{R}_n(\theta_0) \le c'_\alpha) = P((\sqrt{n}SR)^2 + O_p(n^{-3/2}) \le c'_\alpha) \\ &= \int_{-\sqrt{c'_\alpha}}^{\sqrt{c'_\alpha}} \phi(x)\,dx + \int_{-\sqrt{c'_\alpha}}^{\sqrt{c'_\alpha}} \left\{n^{-1/2}r_1(x)\phi(x) + n^{-1}r_2(x)\phi(x)\right. \\ &\quad \left. +n^{-3/2}r_3(x)\phi(x)\right\}\,dx + O(n^{-2}) \\ &= 1-\alpha + \frac{b}{n}c_\alpha g(c_\alpha) - \frac{b}{n}c_\alpha g(c_\alpha) + O(n^{-2}) \\ &= 1-\alpha + O(n^{-2}). \end{aligned}$$

Thus, the Bartlett correction method reduces the coverage error of the EL confidence interval from $O(n^{-1})$ to $O(n^{-2})$. Clearly, if either term of degree 4 or 6 in r_2 does not vanish, then terms of order n^{-1} cannot be removed. This is why Bartlett correction works for EL, but not for general statistics. More essentially, the sufficient fast decay rates $\kappa_3(SR) = O(n^{-3})$ and $\kappa_4(SR) = O(n^{-4})$ ensure that the terms of orders 4 and 6 disappear. Note that this conclusion corresponds with terms of orders 3 and 5, which vanish in (3.11).

3.3 Bartlett Correction for Empirical Likelihood with Gaussian Short-Memory Time Series

Consider a stationary linear process $\{X_t; t \in \mathbb{Z}\}$ satisfying

$$X_t = \sum_{j=0}^{\infty} a_j(\theta)\epsilon_{t-j},$$

where $\{\epsilon_t\}$ is an i.i.d. innovation process with a mean of zero and a finite variance of $\sigma_\epsilon^2 < \infty$. Let $\gamma_\theta(k) = \mathrm{Cov}(X_t, X_{t+k})$ be the autocovariance function (ACVF), where $\theta \in \mathbb{R}^p$ is the parameter of interest. If the ACVFs are summable, i.e., $\sum_{k=-\infty}^{\infty} \gamma_\theta(k) < \infty$, we call $\{X_t\}$ a SMTS (see Priestley (1981)). Assume further that the spectral density function $f(\omega, \theta)$ defined by (2.12) has a continuous second-order derivative on $\Pi = [-\pi, \pi]$.

Stationary ARMA models have been known to belong to SMTS. For example, we simulate a time series $\{X_t\}$ from an ARMA(1,1) model with a length of $n = 200$,

$$X_t = 0.1X_{t-1} + \epsilon_t + 0.7\epsilon_{t-1}, \quad \text{for} \quad \epsilon_t \overset{i.i.d.}{\sim} N(0,1). \tag{3.16}$$

We plot the data and their ACF in Figure 3.1. The ACF decays very quickly and lies in the asymptotic convergence band within 1 lag.

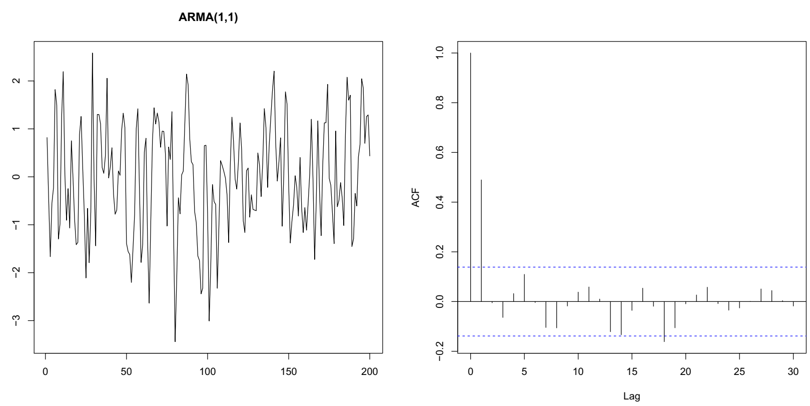

Figure 3.1: The left plot shows a 200 length ARMA(1,1) time series and the right plot shows the ACF for the data.

The spectral density function for Gaussian ARMA(1,1) models with various coefficients is shown in Figure 3.2. The spectral density function is bounded above and below from zero.

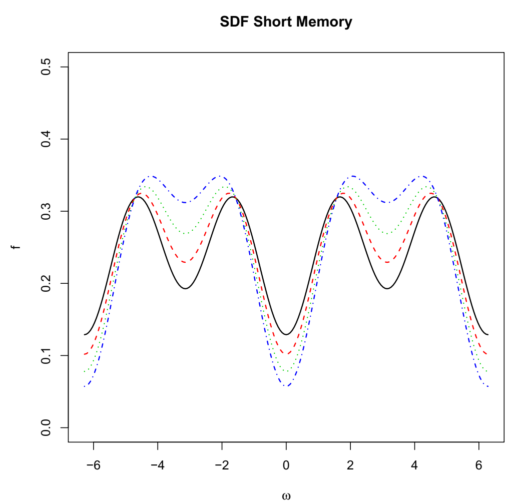

Figure 3.2: The spectral density function of ARMA(1,1) model.

The formulation of Bartlett correction for EL with time series data relies on the Whittle-type periodogram-based estimating functions introduced in Section 2.2.3. In this case, the log-EL ratio function is

$$\mathcal{R}_{n,\mathcal{F}}(\theta) = \max_{p_j} \left\{ \prod_{j=1}^{n} np_j \mid \sum_{j=1}^{n} p_j m_j(\theta) = 0,\ \ \sum_{j=1}^{n} p_j = 1,\ p_j \geq 0 \right\},$$

where $m_j(\theta)$ is given by equation (2.13).

From Theorem 4.3.2 in Brillinger (2001), it is known for SMTS that

$$\mathrm{cum}(J_n(\omega_1), J_n(\omega_2)) = \begin{cases} f(\omega_1, \theta) + O(n^{-1}), & \omega_1 + \omega_2 \equiv 0(\mathrm{mod}\ 2\pi), \\ O(n^{-1}), & \text{otherwise.} \end{cases} \tag{3.17}$$

Based on these two equations, it follows that

$$\mathrm{E}(I_n(\omega_j)) = f(\omega_j, \theta_0) - \frac{c}{n} + o(n^{-2}), \tag{3.18}$$

where

$$c = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \mid j \mid \gamma(j) e^{-ij\omega_j}.$$

In addition,

$$\begin{aligned} \mathrm{cum}(I_n(\omega_1), I_n(\omega_2)) &= \mathrm{cum}^2(J_n(\omega_1), J_n(-\omega_2)) + \mathrm{cum}^2(J_n(\omega_1), J_n(\omega_2)) \\ &\quad + \mathrm{cum}(J_n(\omega_1), J_n(-\omega_1), J_n(\omega_2), J_n(-\omega_2)). \end{aligned} \tag{3.19}$$

For Gaussian processes, the fourth-order cumulant is zero. By equations (3.17) and (3.19), the periodogram ordinates are asymptotically independent for Gaussian SMTS. Due to the n^{-1} bias in (3.18),

$$\mathrm{E}\left(\overline{m}\right) = \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n} m_j(\theta_0)\right] \sim Kn^{-1}. \tag{3.20}$$

Based on the asymptotic independent periodogram, Chan and Liu (2010) established the stochastic expansion for the signed root empirical log-likelihood ratio function SR,

$$-2\log \mathcal{R}_{n,\mathcal{F}}(\theta_0) = nSR^2 + O_p(n^{-3/2}) = n(R_1 + R_2 + R_3)^2 + O_p(n^{-3/2}). \tag{3.21}$$

For Gaussian SMTS, it still holds that $R_j = O_p(n^{-j/2})$ for $j = 1, 2, 3$. Here, R_j are functions of higher-order moments of score functions $m_j(\theta_0)$. Whether the periodogram-based EL is Bartlett correctable relies on the order n^{-1} bias of periodogram ordinates in (3.18), because the bias may prevent $\kappa_3(SR)$ and $\kappa_4(SR)$ from achieving the desired orders $O(n^{-3})$ and $O(n^{-4})$, respectively. After some tedious calculations, Chan and Liu established the stochastic expansion of SR and derived its first four cumulants. Based on the higher-order cumulants, they established the valid Edgeworth expansion using the standard argument in Section 3.2,

$$\begin{aligned} P(\sqrt{n}SR \le x) \quad = \quad & \Phi(x) - \phi(x)\left\{\frac{C_{11}}{\sqrt{n}} + \frac{C_{12}}{n} + \frac{1}{2}\left(\frac{C_{22}}{\sqrt{n}} + \frac{C_{23}}{n} + \frac{C_{11}^2}{n}\right) r_1(x)\right. \\ & + \left(\frac{C_{31}}{6\sqrt{n}} + \frac{C_{32}}{6n} + \frac{C_{11}C_{22}}{2n}\right) r_2(x) \\ & + \left(\frac{C_{41}}{24\sqrt{n}} + \frac{C_{22}^2}{8n} + \frac{C_{11}C_{31}}{6n}\right) r_3(x) \\ & \left. + \left(\frac{C_{22}C_{31}}{12n}\right) r_4(x) + \left(\frac{C_{31}^2}{72n}\right) r_5(x)\right\} \\ & + o(n^{-1}), \end{aligned}$$

where $r_j(x)$, $j = 1, \ldots, 5$ are Hermite polynomials, the form of which can be found in Hall (1992), and C_{ij}s are functions of the first four cumulants of $m_j(\theta_0)$. Despite the non-negligible n^{-1} bias in (3.20), Chan and Liu showed that $C_{12} = C_{22} = C_{31} = C_{32} = C_{41} = 0$, such that $r_2(x), \ldots, r_5(x)$ vanish. Thus, the valid Edgeworth expansion for the distribution function of $\sqrt{n}SR$ simplifies to

$$P(\sqrt{n}SR \le x) = \Phi(x) - \phi(x)\left\{\frac{C_{11}}{\sqrt{n}} + \frac{1}{2}\left(\frac{C_{23} + C_{11}^2}{n}\right) r_1(x)\right\} + o_p(n^{-1}). \tag{3.22}$$

Moreover,

$$\mathrm{E}(-2\log \mathcal{R}_n(\theta_0)) = 1 + \frac{b}{n} + o(n^{-1}),$$

where $b = C_{23}+C_{11}^2$. In essence, the orders $\kappa_3(\sqrt{n}SR) = o(n^{-1})$ and $\kappa_4(\sqrt{n}SR) =$

$o(n^{-1})$ guarantee the Bartlett correctability of EL with Gaussian SMTS. Applying the standard procedure, it follows that

$$P(-2\log\mathcal{R}_n(\theta_0) \leq \chi^2_{1,1-\alpha}(1+b/n)) = 1-\alpha+o(n^{-1}),$$

indicating that the Bartlett correction reduces the coverage errors of the EL confidence intervals from order $O(n^{-1})$ to $o(n^{-1})$.

Chapter 4

Bartlett Correction for EL with Gaussian Long-Memory Time Series

As mentioned in Section 3.2, Chan and Liu (2010) considered the Bartlett correctability of the EL method for Gaussian weakly dependent processes. However, it is still unknown whether EL is Bartlett correctable for strongly dependent processes. The strong dependence phenomenon is important because it has been widely observed in various fields such as astronomy, chemistry, economics, engineering, physics and statistics. In this chapter, we prove the Bartlett correctablity of EL for Gaussian long-memory time series (LMTS). In the following, we formulate the argument based on Monti's Whittle-type periodogram-based EL ratio function $-2\log\mathcal{R}_{n,\mathcal{F}}(\theta)$ defined in (2.14), and denote $l(\theta) = -2\log\mathcal{R}_{n,\mathcal{F}}(\theta)$ for simplicity.

4.1 Introduction

As noted in Chapter 3, one attractive feature of the EL method is Bartlett correctability, which means that a simple adjustment to the log-EL ratio function improves the approximation to the chi-squared limit. Subsequently, the Bartlett-corrected confidence regions achieve better coverage accuracies. We define the coverage error for true value $\theta_0 \in \mathbb{R}^p$ as

$$E_\alpha = P(l(\theta_0) \leq \chi^2_{p,1-\alpha}) - (1-\alpha).$$

Zhang (1996) showed that for EL with independent real-valued estimating functions, E_α can be reduced from order $O(n^{-1})$ to $O(n^{-2})$ using the Bartlett correction technique. Chan and Liu (2010) used the Whittle-type periodogram-based EL to show that E_α can be reduced from order $O(n^{-1})$ to $o(n^{-1})$ for Gaussian short-memory time series (SMTS).

It is unclear, however, whether Bartlett correction is applicable to LMTS. Hurvich and Beltrao (1993) and Robinson (1995) proved that the periodogram ordinates of LMTS are asymptotically dependent for frequencies near the origin. That means that formula (3.17) does not hold for LMTS. Therefore, the proof of the Bartlett correction of EL for weakly dependent processes, which relies on the asymptotic independence of periodograms, cannot be directly generalized to LMTS. In this paper, we establish the validity of the Edgeworth expansion for Gaussian LMTS to make EL moderately Bartlett correctable, in the sense that, E_α is reduced from order $\log^6 n/n$ to $\log^3 n/n$. Although we only establish the Bartlett correctability of EL for Gaussian distributed time series, this exploration provides a fundamental step in further research on non-Gaussian cases. Moreover, our simulation results demonstrate that the performance of Bartlett-corrected EL is better than that of the Bartlett-corrected version of the Whittle likelihood in Gaussian autoregressive fractionally integrated moving average (ARFIMA) models, which justifies the usefulness of Bartlett correction for EL with Gaussian LMTS.

This chapter is organized as follows. Section 4.2 reviews LMTS models and the Bartlett correction of EL for i.i.d. observations and Gaussian SMTS. In Section 4.3, we establish the validity of Edgeworth expansion, which provides a fundamental tool for the main results. Section 4.4 presents simulation studies that demonstrate the good finite sample performance of Bartlett correction in ARFIMA models. Furthermore, we study the coverage errors of both the Whittle likelihood ratio statistic and the Bartlett-corrected counterpart. Proofs of technical results are given in Section 4.5.

4.2 Review of Long-Memory Time Series

Consider a weakly stationary real-valued process $\{X_t\}$ with a mean of zero and a spectral density function of

$$f(\omega, \theta) \sim K\omega^{-\theta} \qquad \text{as} \;\; \omega \to 0^+, \tag{4.1}$$

where $K > 0$ and $\theta \in \mathbb{R}$. The parameter θ is known as the memory parameter. The process X_t is said to have short memory when $\theta = 0$, long memory when $\theta \in (0, 1)$ and negative memory when $\theta \in (-1, 0)$. This model includes two widely used long memory parametric models: the ARFIMA model (Granger and Joyeux, 1980; Hosking, 1981), where the fractional parameter d is defined by $d = \theta/2$, and the fractional Gaussian noise (Mandelbrot and Van Ness, 1968) model, in which the self-similar parameter H satisfies $H = (\theta + 1)/2$. For details, see Beran (1994).

In the following Figure (4.1), we plot the time series of the Nile's minimum river level from year 600 to year 1300, and the autocorrelation function (ACF). It is noted that the ACF decays polynomially slow, and its quantity lies outside the asymptotic convergence band even at lag 70.

Figure 4.2 plots the simulated spectral density functions of Gaussian ARFIMA$(0, d, 0)$ models with different memory parameters $\theta = 2d$. In this case, the spectral

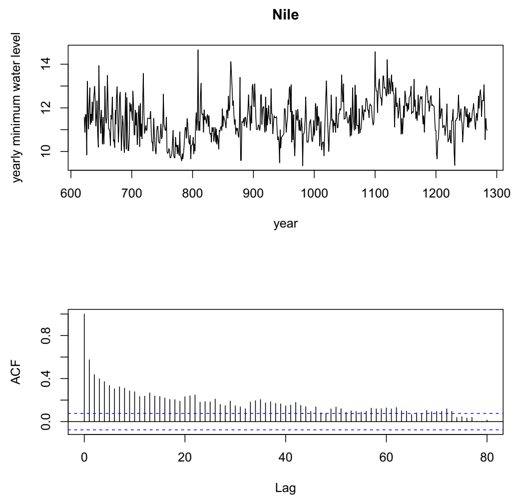

Figure 4.1: Autocorrelation function of Gaussian long-memory time series.

density function admits an expression

$$f(\omega, \theta) = \frac{1}{2\pi} \frac{1}{|1 - \exp(-i\omega)|^{\theta}}.$$

Note that $f(\omega, \theta)$ is unbounded but integrable at the origin. Due to this singu-

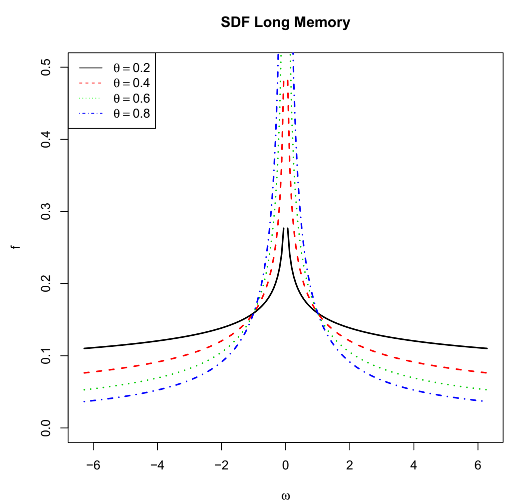

Figure 4.2: The spectral density function of Gaussian long-memory time series.

larity, periodogram ordinates are asymptotically biased estimators for spectral density at low frequencies (see Lemma 4.9). Different periodogram ordinates are also asymptotically correlated when the frequencies tend toward zero (see

Lemma 4.9). However, when separated by a fixed distance, periodogram ordinates are asymptotically uncorrelated (see Lemmas 4.7 and 4.8).

Recall that the key property in proving the Bartlett correction for EL with Gaussian SMTS is the asymptotic independent $I_n(\omega_j)$ distributed as $\frac{1}{2}f(\omega_j,\theta)\chi_1^2$ with rate n^{-1}. The n^{-1} rate causes the bias of the Whittle-type score function $m_j(\theta)$ in (2.13) with magnitude $O(n^{-1})$. For LMTS, however, (3.18) does not hold, i.e., $\mathrm{E}(\overline{m})$ does not converge to zero at rate n^{-1}. In Lemma 1 below, we obtain the order $O(\log^3 n/n)$ magnitude of $\mathrm{E}(\overline{m})$ for LMTS by carefully bounding the summation in $\overline{m}$ over different frequency ranges through the entire collection. Given the larger bias involving $\log n$, an irregular form of the Edgeworth expansion is established to show a "slight" Bartlett correctability in the next section.

4.3 Main Results

Before establishing the validity of Edgeworth expansion and Bartlett correction, we impose the following assumptions.

Assumptions.

1. $\{X_t\}$ is a real-valued linear weakly stationary process satisfying

$$X_t = \sum_{u=0}^{\infty} a_u \epsilon_{t-u},$$

where $a_0 = 1$ and $\sum_{u=0}^{\infty} a_u^2 < \infty$. The noise process $\{\epsilon_t\}$ is a sequence of Gaussian i.i.d. random variables with $\mathrm{E}(\epsilon_t) = 0$ and finite known innovation variance $\mathrm{E}(\epsilon_t^2) = \sigma_\epsilon^2$.

2. The spectral density function of $\{X_t\}$ is given by

$$f(\omega,\theta) = \frac{\sigma_\epsilon^2}{\omega^{2d}} f^*(\omega),$$

where $d = \theta/2 \in (0, \frac{1}{2})$ is the parameter of interest and $f^*(\omega)$ is an even, positive, continuous function on $[-\pi, \pi]$, bounded above and away from zero. In addition, we assume that the true spectral density ($\sqrt{-1} = i$),

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \text{Cov}(X_t, X_{t+k}) \exp(-i\omega k) = f(\omega, \theta).$$

The Gaussian assumption is used to deduce bounds for higher-order moments of $\overline{m}$ (see Lemma 4.12). This condition rules out distributions supported on lattice, and implies the Cramér's condition $\limsup_{\tau\to\infty} |\text{E}(\exp(i\tau X_t))| < \infty$, which is necessary to establish a valid Edgeworth expansion establishment. Assumption 2 ensures the integrability of the spectral density function and the existence of a positive definite autocovariance function for $\{X_t\}$. By the approximation $\omega^{2d} \sim |1 - \exp(-i\omega)|^{2d}$ as $\omega \to 0$, Assumption 2 can be applied to ARFIMA models with the spectral density $\frac{\sigma_\epsilon^2}{2\pi} \frac{1}{|1-\exp(-i\omega)|^{2d}} \tilde{f}(\omega)$, as $\tilde{f}(\omega)$ is bounded above and away from zero for all ω.

Bartlett correction for Gaussian LMTS

For Gaussian LMTS, the profile EL ratio function for the parameter of interest θ_0 is constructed as

$$\mathcal{R}_n(\theta) = \max_{p_j} \left\{ \prod_{j=1}^{n} np_j \mid \sum_{j=1}^{n} p_j m_j(\theta) = 0, \ \sum_{j=1}^{n} p_j = 1, \ p_j \geq 0 \right\}.$$

The Lagrange multiplier argument leads to the log-EL ratio statistic as

$$l(\theta) = -2 \log \mathcal{R}_n(\theta) = 2 \sum_{j=1}^{n} \log(1 + t m_j(\theta)), \tag{4.2}$$

where t is the solution of equation

$$\frac{1}{n} \sum_{j=1}^{n} \frac{m_j(\theta)}{1 + t m_j(\theta)} = 0. \tag{4.3}$$

To study the Bartlett correction of EL for Gaussian LMTS, we first establish the stochastic expansion of $l(\theta_0)$. We begin with Lemma 4.1, which evaluates the bias of the estimating equation. For simplicity, define $m_j \equiv m_j(\theta_0)$

and for $k = 2, 3, 4$, let

$$\lambda_k = \mathrm{E}\left(\frac{1}{n}\sum_{j=1}^{n} m_j^k\right) \quad \text{and} \quad \Delta_k = \frac{1}{n}\sum_{j=1}^{n}(m_j^k - \lambda_k). \tag{4.4}$$

Lemma 4.1 Under Assumptions 1-2, we have

$$\mathrm{E}\left(\overline{m}\right) = O\left(\frac{\log^3 n}{n}\right), \quad \text{and} \quad \mathrm{Var}\left(\overline{m}\right) = O\left(\frac{1}{n}\right). \tag{4.5}$$

Unlike the bias in (3.20) of order n^{-1}, the larger bias in (4.5) for LMTS results from the dependence of periodogram ordinates at frequencies near origin. However, if the integers j in Fourier frequency $\omega_j = 2\pi j/n$ are allowed to increase at a comparable rate with the sample size, i.e., $j/n = j(n)/n \in (0, 1]$, then the periodogram ordinates are independently chi-squared distributed (see Lemmas 4.7 and 4.8). Hence, after some tedious calculations, variance of the estimating function is bounded above by an order n^{-1} quantity. By applying Chebyshev's inequality to (4.5), we have $\overline{m} = O_p(n^{-1/2})$. Together with Lemma 4.12 in the Appendix, we establish the stochastic expansion for the periodogram-based log-EL ratio,

$$\begin{aligned}\frac{1}{n}l(\theta_0) &= \frac{\overline{m}^2}{\lambda_2} - \frac{\overline{m}^2\Delta_2}{\lambda_2^2} + \frac{2}{3}\frac{\lambda_3\overline{m}^3}{\lambda_2^3} + \frac{\overline{m}^2\Delta_2^2}{\lambda_2^3} + \frac{2}{3}\frac{\overline{m}^3\Delta_3}{\lambda_2^3} - 2\frac{\lambda_3\overline{m}^3\Delta_2}{\lambda_2^4} \\ &\quad + \frac{\lambda_3^2\overline{m}^4}{\lambda_2^5} - \frac{1}{2}\frac{\lambda_4\overline{m}^4}{\lambda_2^4} + O_p\left(n^{-\frac{5}{2}}\right).\end{aligned} \tag{4.6}$$

The details to derive this formula are given in the proof of Theorem 4.3. Based on (4.6), the signed root empirical log-likelihood ratio $SR = R_1 + R_2 + R_3$, where $R_j = O_p(n^{-j/2})$, can be derived as follows. If we collect the terms of order $O_p(n^{-1})$ in (4.6) and compare them to R_1^2, we have

$$R_1 = \frac{\overline{m}}{\sqrt{\lambda_2}}.$$

If we collect the terms of order $O_p(n^{-3/2})$ and compare them to $2R_1R_2$, we have

$$R_2 = \frac{1}{3}\frac{\lambda_3\overline{m}^2}{\lambda_2^{5/2}} - \frac{1}{2}\frac{\overline{m}\Delta_2}{\lambda_2^{3/2}}.$$

Finally, if we collect the terms of order $O_p(n^{-2})$ and compare them to $2R_1R_3 + R_2^2$, we have

$$R_3 = \frac{3}{8}\frac{\overline{m}\Delta_2^2}{\lambda_2^{5/2}} + \frac{1}{3}\frac{\overline{m}^2\Delta_3}{\lambda_2^{5/2}} - \frac{5}{6}\frac{\lambda_3\overline{m}^2\Delta_2}{\lambda_2^{7/2}} + \frac{4}{9}\frac{\lambda_3^2\overline{m}^3}{\lambda_2^{9/2}} - \frac{1}{4}\frac{\lambda_4\overline{m}^3}{\lambda_2^{7/2}}.$$

Using the above R_j forms, Lemma 4.2 gives the asymptotic expansions on the cumulants of $\sqrt{n}SR$.

Lemma 4.2 Let k_j, $j = 1, \ldots, 4$, be the first four cumulants of $\sqrt{n}SR$. The asymptotic expansion for k_j is given by

$$k_1 = k_{1,1}\frac{\log^3 n}{\sqrt{n}} + k_{1,2}\frac{1}{\sqrt{n}} + k_{1,3}\frac{1}{n} + O\left(\frac{\log^6 n}{n^{3/2}}\right), \quad (4.7)$$

$$k_2 = k_{2,1} + k_{2,2}\frac{\log^6 n}{n} + k_{2,3}\frac{\log^3 n}{n} + k_{2,4}\frac{1}{n} + O\left(\frac{\log^9 n}{n^2}\right), \quad (4.8)$$

$$k_3 = k_{3,1}\frac{1}{\sqrt{n}} + k_{3,2}\frac{1}{n} + O\left(\frac{\log^9 n}{n^{3/2}}\right), \quad (4.9)$$

$$k_4 = k_{4,1}\frac{1}{n} + O\left(\frac{\log^{12} n}{n^2}\right), \quad (4.10)$$

where the coefficients in the asymptotic expansion satisfy

$$k_{1,1} = \frac{1}{\sqrt{\lambda_2}}\frac{n}{\log^3 n}\mathrm{cum}\left(\overline{m}\right), \quad (4.11)$$

$$k_{1,2} = \frac{n}{3}\frac{\lambda_3}{\lambda_2^{5/2}}\rho_{11} - \frac{n}{2}\frac{1}{\lambda_2^{3/2}}\rho_{12}, \quad (4.12)$$

$$k_{2,2} = -\frac{2}{3}\frac{\lambda_3}{\lambda_2^3}\frac{n^2}{\log^6 n}\mathrm{cum}^2\left(\overline{m}\right),$$

$$k_{2,3} = \frac{n^2}{\log^3 n}\left(\frac{4}{3}\frac{\lambda_3}{\lambda_2^3}\rho_{11} - \frac{1}{\lambda_2^2}\rho_{12}\right)\mathrm{cum}(\overline{m})$$

$$k_{2,4} = -\frac{n^2}{\lambda_2^2}\rho_{112} + n^2\frac{2}{3}\frac{\lambda_3}{\lambda_2^3}\rho_{111} + n^2\frac{7}{4}\frac{1}{\lambda_2^3}\rho_{12}^2 - n^2\frac{17}{3}\frac{\lambda_3}{\lambda_2^4}\rho_{11}\rho_{12} + n^2\frac{1}{\lambda_2^3}\rho_{11}\rho_{22} + n^2\frac{2}{\lambda_2^3}\rho_{11}\rho_{13} + n^2\left(\frac{26}{9}\frac{\lambda_3^2}{\lambda_2^5} - \frac{3}{2}\frac{\lambda_4}{\lambda_2^4}\right)\rho_{11}^2,$$

$$k_{2,1} = 1, \quad k_{1,3} = k_{3,1} = k_{3,2} = k_{4,1} = 0,$$

and ρ_{uv} and ρ_{uvw} are defined as

$$\rho_{uv} = \mathrm{cum}\left(\frac{1}{n}\sum_{j=1}^{n} m_j^u, \frac{1}{n}\sum_{j=1}^{n} m_j^v\right),$$
$$\rho_{uvw} = \mathrm{cum}\left(\frac{1}{n}\sum_{j=1}^{n} m_j^u, \frac{1}{n}\sum_{j=1}^{n} m_j^v, \frac{1}{n}\sum_{j=1}^{n} m_j^w\right).$$

Note that $k_{u,v}$s are bounded by some constants, for $u, v \in \mathbb{S}_4 = \{1, 2, 3, 4\}$. Then, the cumulants' expansions (5.5)-(4.10) lead to the coefficients of polynomials in the Edgeworth expansion for $\sqrt{n}SR$. Given the larger bias in (4.5), the expansion has an irregular form with a decreasing power series of order $\log^3 n/\sqrt{n}$ instead of order $\sqrt{n}$ in the weakly dependent processes (3.14).

Theorem 4.3 Under Assumptions 1-2, the p.d.f. $\pi(x)$ of $\sqrt{n}SR$ admits a valid Edgeworth expansion

$$\pi(x) = \phi(x) + \frac{r_1(x)\log^3 n}{\sqrt{n}}\phi(x) + \frac{r_2(x)\log^6 n}{n}\phi(x) + O\left(\frac{\log^9 n}{n^{3/2}}\right), \qquad (4.13)$$

where

$$r_1(x) = \frac{\sqrt{n}}{\log^3 n}\left\{k_1 x + \frac{1}{6}k_3(x^3 - 3x)\right\},$$
$$r_2(x) = \frac{n}{\log^6 n}\frac{1}{2}(k_2 - 1 + k_1^2)(x^2 - 1),$$

and r_1, r_2 are bounded above and below.

Given the particular form of Edgeworth expansion in (4.13), calculating the coverage error of EL is equivalent to calculating the integral of the density expansion of the signed root decomposition. The decreasing series of power $\log^3/\sqrt{n}$ makes the coverage error larger than its conventional counterpart for i.i.d. data.

Theorem 4.4 If Assumptions 1-2 hold, then

$$P(l(\theta_0) \le c_\alpha) = 1 - \alpha + O\left(\frac{\log^6 n}{n}\right).$$

Theorem 4.4 states that the coverage error of order $\log^6 n/n$ is larger for LMTS, which is larger than that for i.i.d. data or weakly dependent time series. The expectation of the periodogram-based log-EL ratio in this case becomes

$$\begin{aligned} \mathrm{E}(l(\theta_0)) &= \mathrm{E}(\sqrt{n}SR)^2 + O(n^{-3/2}) = \int_{-\infty}^{\infty} x^2\pi(x)\,dx + O(n^{-3/2}) \\ &= \int_{-\infty}^{\infty} \{x^2\phi(x) + x^2 r_1(x)\phi(x)\log^3 n/\sqrt{n} + x^2 r_2(x)\phi(x)\log^6 n/n\}\,dx \\ &\quad + O(\log^9 n/n^{3/2}) = 1 + b\log^6 n/n + O(\log^9 n/n^{3/2}), \end{aligned}$$

where $b = \int_{-\infty}^{\infty} x^2 r_2(x)\phi(x)\,dx$. The feature that r_2 has no term of degree 4 or 6 prompts us to scale $l(\theta_0)$ by $1+b\log^6 n/n$ for a more accurate approximation. In contrast to the weakly dependent series, the additional terms $k_{1,1}$ and $k_{1,2}$ in (5.5) in LMTS prevent the coverage error from reducing to n^{-2} via Bartlett correction. However, this scaling adjustment can remove terms involving $k_{2,1}$ and $k_{2,2}$ in (5.6) such that a "slight" Bartlett correction (i.e., from $O(\log^6 n/n)$ to $O(\log^3 n/n)$) can be still achieved.

Theorem 4.5 Define $c^*_\alpha = (1 + b\log^6 n/n)c_\alpha$. Under Assumptions 1-2, it follows that

$$P\left(l(\theta_0) \le c^*_\alpha\right) = 1 - \alpha + O\left(\frac{\log^3 n}{n}\right),$$

where $b = k^2_{1,1} + k_{2,2}$, and $k_{1,1}$, $k_{2,2}$ are given by (4.11) and (4.12).

In practice, b is unknown, which can be estimated by the Bootstrap method in Monti (1997) from the data. We mention the procedure for the sake of completeness. First, for each series, we calculate the normalized periodogram $\{I_n(\omega_j)/f(\omega_j,\theta)\}$. Because $I_n(\pi+\lambda) = I_n(\pi-\lambda)$, we can restrict our attention to the frequencies ω_j for $j = 1, 2, \ldots, [(n-1)/2]$. Defining $T = [(n-1)/2]$, we calculate

$$y_j = \frac{I_n(\omega_j)}{f(\omega_j, \hat{\theta})} \Bigg/ \frac{1}{T}\sum_{l=1}^{T} \frac{I_n(\omega_l)}{f(\omega_l, \hat{\theta})},$$

where $\hat{\theta}$ is a consistent estimator of θ. Let F_T be the empirical distribution function that assigns mass T^{-1} to each y_j. A bootstrap sample $(y_1^a, y_2^a, \ldots, y_T^a)$ can be obtained by resampling from F_T with replacement. Then, using $I_n^a(\omega_j) = y_j^a f(\omega_j, \hat{\theta})$, we get the resampled periodogram $I_n^a(\omega_1), I_n^a(\omega_2), \ldots, I_n^a(\omega_T)$. Using the resampled periodogram, we compute the periodogram-based log-EL ratio $\{\hat{l}(\hat{\theta}_T)\}$. The resampling procedure is repeated B times to obtain a new set $\{\hat{l}_a(\hat{\theta}_T)\}$, $a = 1, 2, \ldots, B$. Finally, we estimate the unknown factor $\hat{b}$ by

$$\frac{1}{B}\sum_{a=1}^{B} \hat{l}_a(\hat{\theta}_T) = 1 + \frac{\hat{b}\log^6 T}{T}.$$

Consequently, the Bartlett-corrected confidence interval is given by

$$\left\{\theta \in \Theta \mid l(\theta) \leq \chi^2_{1,1-\alpha}\left(1 + \frac{\hat{b}\log^6 T}{T}\right)\right\}.$$

4.4 Simulation Studies

In this section, we perform Monte Carlo experiments to demonstrate the Bartlett correction of EL for LMTS models. A simple LMTS model, ARFIMA $(0, d, 0)$, is used. Also, we compare the performance of the Whittle likelihood ratio test and Bartlett corrected test, under ARFIMA models. All of the simulations are conducted using R version 2.15.1.

The ARFIMA (p, d, q) process X_t with memory parameter d is given by

$$\phi(B)X_t = (1-B)^{-d}\theta(B)\epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} N(0,1),$$

where $N(0, 1)$ denotes the standard normal distribution with a mean of zero and a variance of one. $\theta(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$ and $\phi(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$.

Recall that the asymptotic $1-\alpha$ confidence interval for $d = \theta/2$ is given by $\mathbf{I}_\alpha(d) = \{d \mid l(d) \le c_\alpha\}$ and the Bartlett-corrected confidence interval is $\mathbf{I}'_\alpha(d) = \{d \mid l(d) \le c^*_\alpha\}$. To construct $\mathbf{I}_\alpha(d)$ and $\mathbf{I}'_\alpha(d)$, we substitute various values of d into the log-EL ratio and compare them with the critical value c_α. The simulations are conducted for $d_0 = 0.1,\ 0.2,\ 0.3,\ 0.4$. In Table 4.1, we compare E_α and E'_α for sample sizes $n = 200,\ 1,000$ and $1,500$. In each case, 1,000 replications are drawn. In the procedure for Bootstrap sampling, we adopt the Whittle maximum likelihood estimator as the consistent estimator and set the resampling replications B to be 500.

We use the coverage error to evaluate the performance of the asymptotic distribution confidence intervals. Let $d_0, d_{[\alpha/2]}$ and $d_{[1-\alpha/2]}$ be the true values of the parameter, the lower and the upper endpoints of the confidence interval, respectively. The one- and two-sided coverage errors are defined by

$$\left|P\{d_0 < d_{[\alpha/2]}\} - \alpha/2\right| + \left|P\{d_0 > d_{[1-\alpha/2]}\} - \alpha/2\right|,$$

and

$$\left|P\{(d_0 < d_{[\alpha/2]}) \cup (d_0 > d_{[1-\alpha/2]})\} - \alpha\right|.$$

	Two-sided coverage error				One-sided coverage error			
	d=0.1	d=0.2	d=0.3	d=0.4	d=0.1	d=0.2	d=0.3	d=0.4
	$n = 200$				$n = 200$			
EL	0.0469	0.0392	0.0302	0.0393	0.0469	0.0392	0.0302	0.0419
Bart. EL	0.0227	0.0169	0.0067	0.0527	0.0227	0.0169	0.0112	0.0717
	$n = 1,000$				$n = 1,000$			
EL	0.036	0.01	0.008	0.02	0.036	0.028	0.03	0.02
Bart. EL	0.0223	0.008	0.003	0.023	0.0223	0.024	0.027	0.023
	$n = 1,500$				$n = 1,500$			
EL	0.028	0.003	0.006	0.025	0.028	0.007	0.008	0.025
Bart. EL	0.021	0.001	0.002	0.022	0.021	0.007	0.004	0.022

Table 4.1: Coverage errors of EL and Bartlett-corrected EL confidence intervals for ARFIMA $(0, d, 0)$ models, replications $= 1,000$.

The simulations show that the coverage accuracy of both confidence intervals is higher for larger n, which supports both statistics tending toward a χ_1^2 variate. In addition, except for $d = 0.4$ and $n \leq 1,000$, Bartlett-corrected intervals have smaller coverage errors than the non-Bartlett-corrected counterparts. Figures 4.3-4.5 present the QQ plot between the log-EL ratio and the Bartlett-corrected log-EL ratio and the χ_1^2 random variable. The closer are the lines to the 45° straight line, the more accurate are the corresponding asymptotic distributions. For large n, the asymptotic accuracy of both the EL and the Bartlett corrected EL are very similar, thus we only show the cases with small sample sizes. It can be seen that the Bartlett correction does improve the approximation.

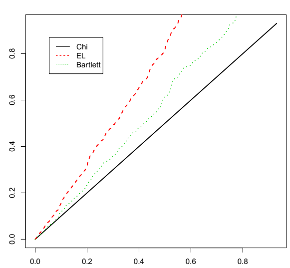

Figure 4.3: $n = 50$ and $H_0 : d = 0.1$.

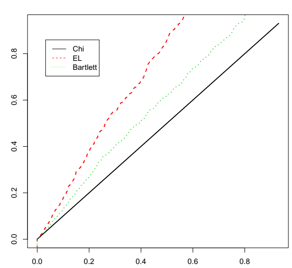

Figure 4.4: $n = 50$ and $H_0 : d = 0.2$.

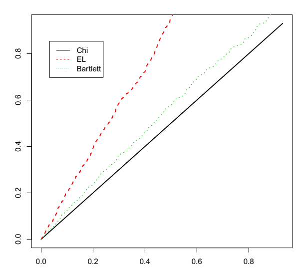

Figure 4.5: $n = 25$ and $H_0 : d = 0.1$.

Alternatively, we may use the Whittle likelihood ratio test to construct the confidence interval, because Hosoya (1997) proved for LMTS that

$$-2\{WL(\theta_0) - WL(\hat{\theta})\} \xrightarrow{d} \chi_1^2,$$

where $\hat{\theta}$ is a consistent estimator of θ. However, Bartlett correction does not work for Whittle likelihood ratio statistics in finite sample performance. To show this, we further compare the coverage errors of the Whittle likelihood ratio and Bartlett-corrected tests with sizes $n = 50$, 200 and 500. In calculating the Bartlett correction factor step, we also adopt the Whittle estimator as the consistent estimator, and the Bootstrap iteration is set to 500 times.

	Two-sided coverage error				One-sided coverage error			
	d=0.1	d=0.2	d=0.3	d=0.4	d=0.1	d=0.2	d=0.3	d=0.4
	$n = 200$				$n = 200$			
EL	0.011	0.006	0.007	0.019	0.011	0.006	0.007	0.019
Bart. EL	0.034	0.051	0.053	0.057	0.041	0.051	0.053	0.057
	$n = 1,000$				$n = 1,000$			
EL	0.003	0.002	0.001	0.012	0.03	0.002	0.003	0.012
Bart. EL	0.007	0.01	0.006	0.015	0.03	0.04	0.044	0.033

Table 4.2: Coverage errors of Whittle and Bartlett-corrected confidence intervals for ARFIMA $(0, d, 0)$ models, replications $= 1,000$.

The simulation results in Table 4.2 show that the Whittle likelihood ratio converges to a chi-squared random variable as the sample size increases, but, the Bartlett-corrected coverage error does not converge to zero in general. However, this technique fails to improve the conventional Whittle likelihood ratio test in almost all cases. This feature provides our periodogram-based EL superior in real application.

4.5 Proof of Theorems

Proof of Theorem 4.3. To study Edgeworth expansion of density for signed root empirical log-likelihood ratio, we must develop a stochastic expansion for $l(\theta_0)$. Applying the Taylor expansion to (4.3), we get

$$\frac{1}{n}\sum_{j=1}^{n} m_j(1 - tm_j + (tm_j)^2 + \cdots) = 0.$$

Solving for t, it follows that

$$t = \frac{\overline{m}}{\lambda_2} - \frac{\overline{m}\Delta_2}{\lambda_2^2} + \frac{\lambda_3\overline{m}^2}{\lambda_2^3} + \frac{\overline{m}\Delta_2^2}{\lambda_2^3} - 3\frac{\lambda_3\overline{m}^2\Delta_2}{\lambda_2^4} + 2\frac{\lambda_3^2\overline{m}^3}{\lambda_2^5} + \frac{\overline{m}^2\Delta_3}{\lambda_2^3} - \frac{\lambda_4\overline{m}^3}{\lambda_2^4} + O_p(n^{-2}).$$

Note that $t = O_p(n^{-1/2})$. Substituting t into (5.2), we have the stochastic expansion (4.6). Then, R_js and asymptotic expansions of k_js are obtained as discussed in Section 3. The characteristic function of $\sqrt{n}SR$ is thus given by

$$\begin{aligned}\psi_n(x) &= \exp\left\{k_1(ix) + \frac{1}{2}k_2(ix)^2 + O\left(\frac{\log^9 n}{n^{3/2}}\right)\right\} \\ &= e^{-\frac{x^2}{2}}\left\{1 + k_1(ix) + \frac{1}{2}(k_2 - 1 + k_1^2)(ix)^2\right\} + O\left(\frac{\log^9 n}{n^{3/2}}\right).\end{aligned}$$

Applying the Fourier inversion formula to $\psi_n(x) = \int_{-\infty}^{\infty} e^{i\tau x}\pi(x)\,d\tau$, $\pi(x)$ admits the Edgeworth expansion (4.13). This completes the proof of Theorem 4.3. □

Proof of Theorem 4.4. From the Edgeworth expansion (4.13) of $\pi(x)$, we have

$$\begin{aligned}P(l(\theta_0) \le c_\alpha) &= P(nSR^2 + O_p(n^{-3/2}) \le c_\alpha) \\ &= \int_{-\sqrt{c_\alpha}}^{\sqrt{c_\alpha}} \phi(x)\,dx + \int_{-\sqrt{c_\alpha}}^{\sqrt{c_\alpha}} k_1 H_1(x)\phi(x)\,dx \\ &\quad + \int_{-\sqrt{c_\alpha}}^{\sqrt{c_\alpha}} \left(\frac{1}{2}(k_2 - 1 + k_1^2)H_2(x)\right)\phi(x)\,dx + O\left(\frac{\log^9 n}{n^{3/2}}\right) \\ &= 1 - \alpha + O\left(\frac{\log^6 n}{n}\right),\end{aligned}$$

where $H_j(x)$, $j = 1, \ldots, 6$ are Hermite polynomials (see Hall (1992)). The order $O(\log^6 n/n)$ of the error term comes from $k_2 - 1 + k_1^2$. This completes the proof of Theorem 4.4. □

Proof of Theorem 4.5. The proof relies on the Edgeworth expansion of density for the corrected signed root empirical log-likelihood ratio, i.e., $SR^* = SR(1 - \frac{b\log^6 n}{2n})$, which requires the asymptotic expansion of the cumulants of $\sqrt{n}SR^*$. Scaling the log-EL ratio by its mean, we have

$$\begin{aligned}l(\theta_0)/(1 + b\log^6 n/n) &= \left\{\sqrt{n}SR(1 - \frac{b\log^6 n}{2n})\right\}^2 + O_p(n^{-3/2}) \\ &= (\sqrt{n}SR^*)^2 + O_p(n^{-3/2}).\end{aligned}$$

Using this equation, we can deduce the asymptotic expansion of k_j^*, $j =$

$1, \ldots, 4$, which are the first four cumulants of $\sqrt{n}SR^*$,

$$k_1^* = \frac{k_{1,1}\log^3 n}{\sqrt{n}} + \frac{k_{1,2}}{n^{1/2}} + O\left(\frac{\log^9 n}{n^{3/2}}\right), \qquad k_2^* = 1 - \frac{k_{1,1}^2 \log^6 n}{n} + \frac{k_{2,3}\log^3 n}{n} + O\left(\frac{1}{n}\right),$$

$$k_3^* = \frac{k_{3,1}}{\sqrt{n}} + O\left(\frac{\log^6 n}{n^{3/2}}\right), \qquad k_4^* = \frac{k_{4,1}}{n} + O\left(\frac{\log^{12} n}{n^2}\right).$$

Thus, the p.d.f. $\pi^*(x)$ of $\sqrt{n}SR^*$ admits an Edgeworth expansion

$$\pi^*(x) = \phi(x) + \frac{r_1^*(x)\log^3 n}{\sqrt{n}}\phi(x) + \frac{r_2^*(x)\log^6 n}{n}\phi(x) + O\left(\frac{\log^9 n}{n^{3/2}}\right),$$

where

$$\begin{aligned} r_1^*(x) &= \frac{\sqrt{n}}{\log^3 n}\left[k_1^* H_1(x) + \frac{1}{6}k_3^* H_3(x)\right], \\ r_2^*(x) &= \frac{n}{\log^6 n}\left[\frac{1}{2}(k_2^* - 1 + (k_1^*)^2)H_2(x) + (\frac{k_4^*}{24} + \frac{k_1^* k_3^*}{6})H_4(x) + \frac{(k_3^*)^2}{72}H_6(x)\right]. \end{aligned}$$

The coefficients of Hermite polynomials in $r_2^*(x)$ satisfy

$$\begin{aligned} k_2^* - 1 + (k_1^*)^2 &= (2k_{1,1}k_{1,2} + k_{2,3})\frac{\log^3 n}{n} + O\left(\frac{1}{n}\right), \\ \frac{1}{6}k_1^* k_3^* + \frac{1}{24}k_4^* &= \frac{k_{1,1}k_{3,1}}{6}\frac{\log^3 n}{n} + \frac{k_{4,1}}{24}\frac{1}{n} + O\left(\frac{\log^9 n}{n^{3/2}}\right), \\ \frac{1}{72}(k_3^*)^2 &= \frac{k_{3,1}^2}{72}\frac{1}{n} + O\left(\frac{\log^6 n}{n^2}\right). \end{aligned}$$

In the Supplementary Materials we show that $k_{3,1} = k_{4,1} = 0$. Hence, $\frac{1}{6}k_1^* k_3^* + \frac{1}{24}k_4^*$ and $\frac{1}{72}(k_3^*)^2$ are bounded by $O(\log^3 n/n)$. Consequently, after a standard argument, the coverage probability follows

$$P(l(\theta_0) \leq c_\alpha^*) = 1 - \alpha + O\left(\frac{\log^3 n}{n}\right),$$

completing the proof of Theorem 4.5. □

Technical Lemmas for Cumulant Calculations

Let $\delta \in (0, 1)$ and $\epsilon > 0$ be generic constants. Define the discrete Fourier transform (DFT) $J_n(\omega_j)$ as

$$J_n(\omega_j) = \frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n} X_t e^{-it\omega_j}.$$

Note that $I_n(\omega_j) = J_n(\omega_j)J_n(\omega_{-j})$ and $J_n(\omega_j)/\sqrt{f(\omega_j)}$ is referred as the normalized DFT. To calculate $\sum_{j=1}^{n} R(I_n(\omega_j)/f(\omega_j))$ for some functions R, we must study the limiting covariance of the normalized DFT for the whole frequency domain. Define $\Lambda = \{j \in \mathbb{S}_n : \delta n \le j \le n\}$, then the ω_j in this region are bounded away from zero. The whole frequency region on the plane $(0, 2\pi) \times (0, 2\pi)$ can be partitioned into four complementary regions

$$\Lambda_1 = \{(j,k): \ j,k \in \Lambda; \ j \ne k \ ; \ |\omega_j - \omega_k| \le \epsilon\},$$

$$\Lambda_2 = \{(j,k): \ j,k \in \Lambda; \ \epsilon < |\omega_j - \omega_k| < 2\pi\},$$

$$\Lambda_3 = \{(j,k): \ j \in \mathbb{S}_n, k \in \Lambda; |\omega_j| \le \epsilon\} \cup \{(j,k): \ k \in \mathbb{S}_n, j \in \Lambda; |\omega_k| \le \epsilon\},$$

$$\Lambda_4 = \{(j,k): \ j,k \in \mathbb{S}_n; \ |\omega_j| \le \epsilon; \ |\omega_k| \le \epsilon; \ j \ne k\}.$$

Proof of Lemma 4.1. For any constants K and any small $\delta > 0$, the expectation of $\overline{m}$ are divided into three regions as

$$\begin{aligned}\mathrm{E}(\overline{m}) &= \frac{1}{n}E\left(\sum_{j=1}^{[\log\log n]} m_j + \sum_{j=[\log\log n]+1}^{[\delta n]} m_j + \sum_{j=[\delta n]+1}^{n} m_j\right) \\ &\le K\frac{\log^{1+\delta} n}{n} + \frac{K}{n}\sum_{j=[\log\log n]+1}^{[\delta n]} \frac{\log j}{j}\log\left(\frac{j}{n}\right) \\ &\quad +\frac{K}{n}\sum_{j=[\delta n]+1}^{n}\frac{\log n}{n}, \end{aligned} \tag{4.14}$$

where $[x]$ is the largest integer that is less than or equals x. For the first term of (4.14), $m_j = O(\log n)$ because $\frac{\partial}{\partial\theta}\log f(\omega_j) \sim \log n$, and $I_n(\omega_j)/f(\omega_j) - 1 = O(1)$ by Lemma 4.9. Given

$$\sum_{j=1}^{n}\frac{\log j}{j} = \frac{1}{2}\log^2 n + O\left(\frac{\log^2 n}{n}\right)$$

and Lemma 4.10, the second term in (4.14) is of order $O(\log^3 n/n)$. Additionally, using Lemma 4.7, the last term is of order $O(\log n/n)$. Summing these parts, $\mathrm{E}(\overline{m}) = O(\log^3 n/n)$. For the variance of $\overline{m}$, note that

$$\mathrm{Var}(\overline{m}) = P_1 + P_2,$$

where

$$P_1 = \frac{1}{n^2}\sum_{j=1}^{n} \mathrm{Var}(m_j), \qquad P_2 = \frac{1}{n^2}\sum_{j\neq k} \mathrm{Cov}(m_j, m_k).$$

Then by Lemmas 4.7 and 4.8,

$$\begin{aligned} P_1 &= \frac{1}{n^2}\sum_{j=1}^{n}(\frac{\partial}{\partial\theta}\log f(\omega_j))^2\left[\mathrm{E}^2\left(\frac{J_n(\omega_j)J_n(\omega_{-j})}{\sqrt{f(\omega_j)f(\omega_{-j})}}\right) + \mathrm{E}\left(\frac{J_n^2(\omega_j)}{f(\omega_j)}\right)\mathrm{E}\left(\frac{J_n^2(\omega_{-j})}{f(\omega_{-j})}\right)\right] \\ &\leq \frac{K}{n^2}\left[\sum_{j=1}^{[\log\log n]}\log^2 n + \sum_{j=[\log\log n]+1}^{[\delta n]}(\log j - \log n)^2\left(1+\frac{\log j}{j}\right)^2\right. \\ &\quad \left. + \sum_{j=[\delta n]+1}^{n}\left(1 + O(\frac{\log n}{n})\right)^2\right] = O(n^{-1}), \end{aligned}$$

and

$$\begin{aligned} P_2 &= \frac{1}{n^2}\sum_{j\neq k}(\frac{\partial}{\partial\theta}\log f(\omega_j))(\frac{\partial}{\partial\theta}\log f(\omega_k))\left[\mathrm{E}\left(\frac{J_n(\omega_j)J_n(\omega_{-k})}{\sqrt{f(\omega_j)f(\omega_{-k})}}\right)\mathrm{E}\left(\frac{J_n(\omega_{-j})J_n(\omega_k)}{\sqrt{f(\omega_{-j})f(\omega_k)}}\right)\right. \\ &\quad \left. + \mathrm{E}\left(\frac{J_n(\omega_j)J_n(\omega_k)}{\sqrt{f(\omega_j)f(\omega_k)}}\right)\mathrm{E}\left(\frac{J_n(\omega_{-j})J_n(\omega_{-k})}{\sqrt{f(\omega_{-j})f(\omega_{-k})}}\right)\right] = \frac{1}{n^2}\left[\sum_{|\omega_j-\omega_k|\leq\epsilon} + \sum_{|\omega_j-\omega_k|\geq\epsilon}\right]. \end{aligned}$$

Considering the two parts separately, it follows that

$$\begin{aligned} \frac{1}{n^2}\sum_{|\omega_j-\omega_k|\leq\epsilon} &= \frac{1}{n^2}\left\{\sum_{j,k\in\Lambda_4} + \sum_{j,k\in\Lambda_1}\right\} \\ &\leq \frac{K}{n^2}\left\{\sum_{1\leq k\leq[\log\log n]\leq j\leq\delta n}\log n(\log j - \log n)(\frac{\log j}{k})^2\right. \\ &\quad \left. + \sum_{j=[\log\log n]+1}^{[\delta n]}\sum_{k=[\log\log n]}^{j}\log^2 n(\frac{\log j}{k})^2 + (1-\delta)^2 n^2(\frac{\log^2 n}{n^2})\right\} \\ &= o(n^{-1}), \end{aligned}$$

and

$$\begin{aligned}\frac{1}{n^2}\sum_{|\omega_j-\omega_k|\geq\epsilon} &= \frac{1}{n^2}\left\{\sum_{j,k\in\Lambda_2}+\sum_{j,k\in\Lambda_3}\right\}\leq\frac{K}{n^2}\left\{\sum_{j=1}^{[\delta n]}(\log j-\log n)^2(1-\delta)n(\frac{\log^2 n}{n^2})\right.\\ &\left.+\sum_{k=1}^{[\delta n]}(\log k-\log n)^2(1-\delta)n(\frac{\log^2 n}{n^2})+(1-\delta)^2n^2(\frac{\log^2 n}{n^2})\right\}\\ &= o(n^{-1}).\end{aligned}$$

Thus, $P_2 = o(n^{-1})$ and $\mathrm{Var}(\overline{m}) = P_1 + P_2 = O(n^{-1})$. □

Next we state some technical Lemmas for the cumulant expansions. The proofs can be found in the Supplementary Materials.

Lemma 4.6

$$\frac{1}{n}\int_{-\pi}^{\pi}\left|\frac{\sin(n\mu/2)}{\sin(\mu/2)}\right|d\mu\sim\frac{1}{\pi}\frac{\log n}{n}\qquad\text{as}\quad n\to\infty. \tag{4.15}$$

Lemma 4.7 For any sequences of integers $j = j(n)$ with $j \in \Lambda$, we have

$$\mathrm{E}\left(\frac{J_n(\omega_j)J_n(\omega_{-j})}{\sqrt{f(\omega_j)f(\omega_{-j})}}\right)=1+O\left(\frac{\log n}{n}\right),\qquad\text{and}\qquad \mathrm{E}\left(\frac{J_n^2(\omega_j)}{f(\omega_j)}\right)=O\left(\frac{\log n}{n}\right).$$

Lemma 4.8 For any two sequences of integers $j = j(n)$ and $k = k(n)$ such that $\{j,k\} \in \Lambda_1 \cup \Lambda_2 \cup \Lambda_3$, we have

$$\mathrm{E}\left(\frac{J_n(\omega_j)J_n(\omega_{-k})}{\sqrt{f(\omega_j)f(\omega_{-k})}}\right)=O\left(\frac{\log n}{n}\right),\quad\text{and}\quad \mathrm{E}\left(\frac{J_n(\omega_j)J_n(\omega_k)}{\sqrt{f(\omega_j)f(\omega_k)}}\right)=O\left(\frac{\log n}{n}\right).$$

Lemmas 4.9 and 4.10 describe the different behavior of the expectation of a product of DFTs under Fourier frequencies ω_j with fixed j and slowly increasing j, respectively.

Lemma 4.9 (P.M. Robinson 1995)

For $0 < |d| < \frac{1}{2}$ and any fixed integers $j \neq k$, $b_d = 2\Gamma(1-2d)\sin \pi d$, we have

$$\lim_{n\to\infty} \mathrm{E}\left(\frac{I_n(\omega_j)}{f(\omega_j)}\right) = \frac{b_d |j|^{2d}}{(2\pi)^{1-2d}} \left\{ 4\int_0^1 u^{2d-1}(u-1)\sin^2(\pi j u)\,du + \frac{1}{d(2d+1)} \right\},$$

and

$$\lim_{n\to\infty} \mathrm{E}\left(\frac{J_n(\omega_j)J_n(\omega_k)}{\sqrt{f(\omega_j)f(\omega_k)}}\right) = P_d(j,k),$$

where

$$P_d(j,k) = \frac{-2b_d\,|jk|^d}{(2\pi)^{1-2d}(j+k)} \int_0^1 u^{2d-1}\{\sin(2\pi j u) + \sin(2\pi k u)\}\,du.$$

In particular, if the white noise process is Gaussian and $j \pm k \neq 0$, then

$$\lim_{n\to\infty} \mathrm{E}\left(\frac{I_n(\omega_j)}{f(\omega_j)}\right) = 2P_d^2(j,j),$$

$$\lim_{n\to\infty} \mathrm{Cov}\left(\frac{I_n(\omega_j)}{f(\omega_j)}, \frac{I_n(\omega_k)}{f(\omega_k)}\right) = P_d^2(j,k) + P_d^2(j,-k).$$

Lemma 4.10 (P.M. Robinson, 1995)

Under Assumptions 1-2, for sequences of positive integers j, k that satisfy $K\log\log n < k < j < \delta n$, we have

$$\mathrm{E}\left(\frac{J_n(\omega_j)J_n(\omega_{-j})}{\sqrt{f(\omega_j)f(\omega_{-j})}}\right) = 1 + O\left(\frac{\log j}{j}\right),$$

$$\mathrm{E}\left(\frac{J_n(\omega_j)J_n(\omega_j)}{\sqrt{f(\omega_j)f(\omega_j)}}\right) = O\left(\frac{\log j}{j}\right),$$

$$\mathrm{E}\left(\frac{J_n(\omega_j)J_n(\omega_k)}{\sqrt{f(\omega_j)f(\omega_k)}}\right) = O\left(\frac{\log j}{k}\right),$$

$\mathrm{E}\left(\frac{J_n(\omega_j)J_n(\omega_{-k})}{\sqrt{f(\omega_j)f(\omega_{-k})}}\right) = O(\frac{\log j}{k}).$

The following Lemma gives the lower bound for the covariance of DFT at the conjugate frequency in the first case of Lemma 4.10.

Lemma 4.11 Under Assumptions 1-2, for sequences of positive integers j, satisfying $0 < K \log\log n < j < \delta n$, we have

$$\mathrm{E}\left(\frac{J_n(\omega_j)J_n(\omega_{-j})}{\sqrt{f(\omega_j)f(\omega_{-j})}} - 1\right) \geq K\frac{1}{j}. \tag{4.16}$$

Remark: The side lobes of the Fejér kernel in the range $[2\pi/n, \infty)$ makes it difficult to evaluate the exact magnitude of the integral in $[2\pi/n, \omega_j]$ when ω_j is not a fixed constant. Using the properties of the Dirichlet kernel, Robinson (1995) only derived an upper bound for the bias $\mathrm{E}(I_n(\omega_j)/f(\omega_j)) - 1$, and the order $O(\log^3 n/n)$ in (4.5) is actually an upper bound of the $\mathrm{E}(\overline{m})$. Using order j^{-1} in (4.16), we can find the lower bound $O(\log^2 n/n)$ for $\mathrm{E}(\overline{m})$. It follows that the lower bound of the coverage error is reduced from order $O_p(\log^4 n/n)$ to $O_p(\log^2 n/n)$ using the Bartlett correction. This argument justifies the slight Bartlett correction in improving the coverage accuracy even for LMTS.

The following Lemma provides the key order magnitude to derive the asymptotic expansion of t in Lemma 4.3 and $l(\theta_0)$ in (4.6).

Lemma 4.12 Under Assumptions 1-2, we have

$$\lambda_k = O(1), \qquad \Delta_k = O_p\left(\frac{1}{\sqrt{n}}\right), \quad \text{for} \quad k = 2, 3, 4,$$

where λ_k and Δ_k are defined in (4.4).

Proof of Lemma 4.12. Under the Gaussian assumption, the cumulants of a normalized DFT with an order 3 or higher vanish, so it suffices to consider the products of second-order cumulants. Together with the cumulant expansion formula in Brillinger (1981), λ_k admits the following asymptotic expansion:

$$\lambda_2 = \frac{1}{n}\sum_{j=1}^{n}\left[\mathrm{cum}(m_j, m_j) + \mathrm{cum}^2(m_j)\right] = O(1),$$

$$\lambda_3 = \frac{1}{n}\sum_{j=1}^{n}\left[\mathrm{cum}(m_j, m_j, m_j) + 3\mathrm{cum}(m_j, m_j)\mathrm{cum}(m_j) + \mathrm{cum}^3(m_j)\right] = O(1),$$

$$\lambda_4 = \frac{1}{n}\sum_{j=1}^{n}[\mathrm{cum}(m_j, m_j, m_j, m_j) + 4\mathrm{cum}(m_j, m_j, m_j)\mathrm{cum}(m_j) + 3\mathrm{cum}^2(m_j, m_j) + 6\mathrm{cum}(m_j, m_j)\mathrm{cum}^2(m_j) + \mathrm{cum}^4(m_j)\big] = O(1).$$

The calculations for the variance of higher moments (i.e. $\frac{1}{n}\sum_j m_j^k$, $k = 2, 3, 4$) can be handled similarly as in the proof of Lemma 4.1, although the process is more tedious. □

Proof of Lemma 4.2

Based on Lemmas 4.7 and 4.8, the cumulants of $\sqrt{n}SR$ are derived to establish its Edgeworth expansion. We only derive the first four cumulants because the higher-order cumulants have smaller orders and can be neglected. From the definition of cumulants, we have

$$k_1 = \mathrm{cum}(\sqrt{n}(R_1 + R_2 + R_3)), \tag{4.17}$$

$$k_2 = \mathrm{cum}(nSR^2) - \mathrm{cum}^2(\sqrt{n}(R_1 + R_2 + R_3)), \tag{4.18}$$

$$k_3 = \mathrm{cum}(\sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_1) + 3\mathrm{cum}(\sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_2) + O\left(\frac{\log^3 n}{n^{3/2}}\right), \tag{4.19}$$

$$k_4 = \mathrm{cum}(\sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_1) + 4\mathrm{cum}(\sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_2) + 4\mathrm{cum}(\sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_3) + 6\mathrm{cum}(\sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_2, \sqrt{n}R_2) + O\left(\frac{\log^4 n}{n^2}\right). \tag{4.20}$$

Consider k_1 in (4.17),

$$
\begin{aligned}
k_1 &= \mathrm{E}(\sqrt{n}(R_1+R_2+R_3)) \\
&= \mathrm{cum}\left(\frac{\overline{m}}{\sqrt{\lambda_2}}\right) + \mathrm{cum}\left(\frac{1}{3}\frac{\lambda_3\overline{m}^2}{\lambda_2^{5/2}} - \frac{1}{2}\frac{\overline{m}\Delta_2}{\lambda_2^{3/2}}\right) \qquad (4.21) \\
&\quad + \mathrm{cum}\left(\frac{3}{8}\frac{\overline{m}\Delta_2^2}{\lambda_2^{5/2}} + \frac{1}{3}\frac{\overline{m}^2\Delta_3}{\lambda_2^{5/2}} - \frac{5}{6}\frac{\lambda_3\overline{m}^2\Delta_2}{\lambda_2^{7/2}} + \frac{4}{9}\frac{\lambda_3^2\overline{m}^3}{\lambda_2^{9/2}} - \frac{1}{4}\frac{\lambda_4\overline{m}^3}{\lambda_2^{7/2}}\right) \qquad (4.22)
\end{aligned}
$$

By Lemma 4.1, the bounds of the terms in (4.21) can be directly derived, i.e.,

$$
\begin{aligned}
\mathrm{cum}(\overline{m}) &= O\left(\frac{\log^3 n}{n}\right) + O\left(n^{-1}\right), \\
\mathrm{cum}(\overline{m}^2) &= \mathrm{cum}\left(\frac{1}{n}\sum_{j=1}^{n} m_j, \frac{1}{n}\sum_{k=1}^{n} m_k\right) + \mathrm{cum}^2\left(\frac{1}{n}\sum_{j=1}^{n} m_j\right) \\
&= O\left(n^{-1}\right) + O\left(\frac{\log^3 n}{n}\right)^2, \\
\mathrm{cum}(\overline{m}\Delta_2) &= \mathrm{cum}\left(\frac{1}{n}\sum_{j=1}^{n} m_j, \frac{1}{n}\sum_{k=1}^{n} m_k^2\right) = O(n^{-1}).
\end{aligned}
$$

In the Gaussian case, the higher-order cumulant

$$
\mathrm{cum}(I_n(\omega_j)/f(\omega_j), I_n(\omega_k)/f(\omega_k), I_n(\omega_k)/f(\omega_k))
$$

can be decomposed into products of the second-order cumulants of normalized DFT, in which cumulants of an order 3 and higher-order vanish. To be specific, we have

$$
\begin{aligned}
&\mathrm{cum}\left(\frac{I_n(\omega_j)}{f(\omega_j)}, \frac{I_n(\omega_k)}{f(\omega_k)}, \frac{I_n(\omega_k)}{f(\omega_k)}\right) \\
&= 2\mathrm{cum}^2\left(\frac{J_n(\omega_j)}{\sqrt{f(\omega_j)}}, \frac{J_n(\omega_{-k})}{\sqrt{f(\omega_{-k})}}\right)\mathrm{cum}\left(\frac{J_n(\omega_k)}{\sqrt{f(\omega_k)}}, \frac{J_n(\omega_{-k})}{\sqrt{f(\omega_{-k})}}\right).
\end{aligned}
$$

Together with Lemmas 4.7, 4.8, 4.9 and 4.10, the terms in (4.22) are of the following orders

$$\begin{aligned}
\mathrm{cum}(\overline{m}\Delta_2^2) &= \mathrm{cum}\left(\frac{1}{n}\sum_{i=1}^{n} m_i, \frac{1}{n}\sum_{j=1}^{n} m_j^2, \frac{1}{n}\sum_{k=1}^{n} m_k^2\right) \\
&\quad + \mathrm{cum}\left(\frac{1}{n}\sum_{j=1}^{n} m_j\right) \mathrm{cum}\left(\frac{1}{n}\sum_{j=1}^{n} m_j^2, \frac{1}{n}\sum_{k=1}^{n} m_k^2\right) \\
&= O\left(\frac{\log^3 n}{n^2}\right) + O\left(\frac{1}{n^2}\right),
\end{aligned}$$

$$\begin{aligned}
\mathrm{cum}(\overline{m}^2\Delta_3) &= \mathrm{cum}\left(\frac{1}{n}\sum_{i=1}^{n} m_i, \frac{1}{n}\sum_{j=1}^{n} m_j, \frac{1}{n}\sum_{k=1}^{n} m_k^3\right) \\
&\quad + 2\mathrm{cum}\left(\frac{1}{n}\sum_{j=1}^{n} m_j\right) \mathrm{cum}\left(\frac{1}{n}\sum_{j=1}^{n} m_j, \frac{1}{n}\sum_{k=1}^{n} m_k^3\right) \\
&= O\left(\frac{\log^3 n}{n^2}\right) + O\left(\frac{1}{n^2}\right),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{cum}(\overline{m}^2\Delta_2) &= \mathrm{cum}\left(\frac{1}{n}\sum_{i=1}^{n} m_i, \frac{1}{n}\sum_{j=1}^{n} m_j, \frac{1}{n}\sum_{k=1}^{n} m_k^2\right) \\
&\quad + 2\mathrm{cum}\left(\frac{1}{n}\sum_{j=1}^{n} m_j, \frac{1}{n}\sum_{k=1}^{n} m_k^2\right) \mathrm{cum}\left(\frac{1}{n}\sum_{j=1}^{n} m_j\right) \\
&= O\left(\frac{1}{n^2}\right) + O\left(\frac{\log^3 n}{n^2}\right),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{cum}(\overline{m}^3) &= \mathrm{cum}\left(\frac{1}{n}\sum_{i=1}^{n} m_i, \frac{1}{n}\sum_{j=1}^{n} m_j, \frac{1}{n}\sum_{k=1}^{n} m_k\right) \\
&\quad + 3\mathrm{cum}\left(\frac{1}{n}\sum_{j=1}^{n} m_j, \frac{1}{n}\sum_{k=1}^{n} m_k\right) \mathrm{cum}\left(\frac{1}{n}\sum_{j=1}^{n} m_j\right) + \mathrm{cum}^3\left(\frac{1}{n}\sum_{j=1}^{n} m_j\right) \\
&= O\left(\frac{1}{n^2}\right) + O\left(\frac{\log^3 n}{n^2}\right) + O\left(\frac{\log^3 n}{n}\right)^3.
\end{aligned}$$

This leads to the asymptotic expansion of k_1 in (5.5). The expansion of k_2 can be computed similarly. For k_3 and k_4, define

$$\begin{aligned}\rho_3 &= \sum_{j=1}^{n}(\frac{\partial}{\partial\theta}\log f(\omega_j))^3 \text{cum}\left(\frac{I_n(\omega_j)}{f(\omega_j)}, \frac{I_n(\omega_j)}{f(\omega_j)}, \frac{I_n(\omega_j)}{f(\omega_j)}\right), \\ \rho_4 &= \sum_{j=1}^{n}(\frac{\partial}{\partial\theta}\log f(\omega_j))^4 \text{cum}\left(\frac{I_n(\omega_j)}{f(\omega_j)}, \frac{I_n(\omega_j)}{f(\omega_j)}, \frac{I_n(\omega_j)}{f(\omega_j)}, \frac{I_n(\omega_j)}{f(\omega_j)}\right).\end{aligned}$$

Direct calculations give us

$$\begin{aligned}\text{cum}(\sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_1) &= \frac{2}{\lambda_2^{3/2}}\rho_3 + o(n^{-1}), \\ \text{cum}(\sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_2) &= -\frac{2}{3}\frac{1}{\lambda_2^{3/2}}\rho_3 + o(n^{-1}).\end{aligned}$$

Substituting the above into (5.9), we have

$$k_{3,1} = \sqrt{n}\left\{\frac{2}{\lambda_2^{3/2}}\rho_3 - 3\frac{2}{3}\frac{1}{\lambda_2^{3/2}}\rho_3\right\} = 0.$$

The calculations for the fourth-order cumulants are similar. Again, direct calculations give us

$$\begin{aligned}\text{cum}(\sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_1) &= \frac{1}{n^2}\frac{\rho_4}{\lambda_2^2} + o(n^{-1}), \\ \text{cum}(\sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_2) &= \frac{1}{2n}\frac{\lambda_3^2}{\lambda_2^3} - \frac{3}{2n^2}\frac{\rho_4}{\lambda_2^2} + O\left(\frac{\log^3 n}{n^3}\right), \\ \text{cum}(\sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_3) &= -\frac{1}{12n}\frac{\lambda_3^2}{\lambda_2^3} + \frac{1}{2n^2}\frac{\rho_4}{\lambda_2^2} + o(n^{-1}) + O\left(\frac{\log^3 n}{n^4}\right), \\ \text{cum}(\sqrt{n}R_1, \sqrt{n}R_1, \sqrt{n}R_2, \sqrt{n}R_2) &= -\frac{5}{18n}\frac{\lambda_3^2}{\lambda_2^3} + \frac{1}{2n^2}\frac{\rho_4}{\lambda_2^2} + o(n^{-1}) + O\left(\frac{\log^3 n}{n^4}\right).\end{aligned}$$

Substituting the above into (4.20), we have

$$k_{4,1} = \frac{1}{n}\frac{\rho_4}{\lambda_2^2} + 4\frac{1}{2}\frac{\lambda_3^2}{\lambda_2^3} - \frac{3}{2n}\frac{\rho_4}{\lambda_2^2} - 4\frac{1}{12}\frac{\lambda_3^2}{\lambda_2^3} + \frac{1}{2n}\frac{\rho_4}{\lambda_2^2} - 6\frac{5}{18}\frac{\lambda_3^2}{\lambda_2^3} + \frac{1}{2n}\frac{\rho_4}{\lambda_2^2} = 0.$$

In conclusion, we have $k_{3,1} = k_{4,1} = 0$. □

Proof of Lemma 4.6. Because the function $\text{cosec}(\mu) - \mu^{-1}$ is bounded on $(0, \frac{\pi}{2})$, it follows that

$$
\begin{aligned}
&\int_{-\pi}^{\pi} \left| \frac{\sin(n\mu/2)}{\sin(\mu/2)} \right| d\mu \\
= \;& 4 \int_{0}^{\pi} \left| \frac{\sin(n\mu/2)}{\mu} \right| d\mu + O(1) \\
= \;& 4 \left(\int_{0}^{\pi} \frac{\sin \mu}{\mu} \, d\mu + \int_{0}^{\pi/n} \sin(n\mu/2) \left\{ \sum_{k=1}^{n-1} \frac{1}{\mu + k\pi/n} \right\} d\mu \right) + O(1).
\end{aligned}
$$

The sum in the braces has lower- and upper-bound $\frac{n}{\pi} \sum_{j=1}^{n-1} \frac{1}{j}$ and $\frac{n}{\pi} \sum_{j=2}^{n} \frac{1}{j}$, respectively, which equals $\frac{n}{\pi}[\log n + O(1)]$. The proof is completed by noting that, from Brillinger (1981),

$$
\int_{0}^{\pi/n} \sin(n\mu) \, d\mu = 2/n \qquad \text{and} \qquad \int_{0}^{\pi} \frac{\sin \mu}{\mu} \, d\mu = O(1).
$$

□

Proof of Lemma 4.7. Without loss of generality, assume that $\omega_j \leq \frac{\pi}{2}$ for sufficiently large n. Note that

$$
\mathrm{E}(I(\omega_j) - f(\omega_j)) = \mathrm{E}(J_n(\omega_j) J_n(\omega_{-j}) - f(\omega_j)) = \int_{-\pi}^{\pi} F_n(\omega_j - \mu)(f(\mu) - f(\omega_j)) \, d\mu,
$$

where $F_n(\mu)$ is the Fejér kernel $F_n(\mu) = \frac{1}{2\pi n} \frac{\sin^2(n\mu/2)}{\sin^2(\mu/2)}$. Because $f(\mu) = \frac{1}{\mu^{2d}} f^*(\mu)$, where $f^*(\mu)$ is a even function bounded above and below from zero, we have

$$
\frac{1}{f(\omega_j)} \int_{-\pi}^{\pi} F_n(\omega_j - \mu)(f(\mu) - f(\omega_j)) \, d\mu = \frac{K}{2\pi n} \int_{-\pi}^{\pi} \frac{\sin^2 n(\omega_j - \mu)/2}{\sin^2(\omega_j - \mu)/2} (f(\mu) - f(\omega_j)) \, d\mu.
$$

The idea is to decompose the range $[-\pi, \pi]$ into five parts and establish a bound for each part,

$$
\frac{K}{n} \int_{-\pi}^{\pi} = \frac{K}{n} \left\{ \int_{-\pi}^{\omega_j - \epsilon} + \int_{\omega_j - \epsilon}^{\omega_j - \frac{1}{n}} + \int_{\omega_j - \frac{1}{n}}^{\omega_j + \frac{1}{n}} + \int_{\omega_j + \frac{1}{n}}^{\omega_j + \epsilon} + \int_{\omega_j + \epsilon}^{\pi} \right\}. \tag{4.23}
$$

The first part of (4.23) is bounded by

$$
\begin{aligned}
\frac{K}{n} \int_{-\pi}^{\omega_j - \epsilon} \;&\leq\; \frac{K}{n} \left(\max_{\mu \in [-\pi,\; \omega_j - \epsilon]} \frac{\sin^2 n(\omega_j - \mu)/2}{\sin^2(\omega_j - \mu)/2} \right) \int_{-\pi}^{\omega_j - \epsilon} |f(\mu) - f(\omega_j)| \, d\mu \\
&\leq\; \frac{K}{n} \int_{-\pi}^{\omega_j - \epsilon} \left(\left| \mu^{-2d} \right| + |f(\omega_j)| \right) d\mu = O(n^{-1}).
\end{aligned}
$$

The second part is bounded above by

$$\begin{aligned}\frac{K}{n}\int_{\omega_j-\epsilon}^{\omega_j-\frac{1}{n}} &\leq \frac{K}{n}\int_{\omega_j-\epsilon}^{\omega_j-\frac{1}{n}}\left(\frac{\sin^2 n(\omega_j-\mu)/2}{\sin^2(\omega_j-\mu)/2}\max_{\mu}\frac{\partial}{\partial\mu}f(\mu)|\mu-\omega_j|\right)d\mu \\ &\leq \frac{K}{n}\int_{-\epsilon}^{-\frac{1}{n}}\frac{\sin^2(n\lambda/2)}{\sin^2(\lambda/2)}|\lambda|\,d\lambda \leq \frac{K}{n}\int_{-\epsilon}^{-\frac{1}{n}}\frac{1}{\lambda}\,d\lambda = O\left(\frac{\log n}{n}\right),\end{aligned}$$

where the third inequality follows from the Zygmund (1977) that

$$\frac{1}{n}\frac{\sin^2(n\lambda/2)}{\sin^2(\lambda/2)} = O((n\lambda^2)^{-1}), \quad 0<|\lambda|<\pi.$$

The third part in (4.23) has an order of

$$\begin{aligned}\frac{K}{n}\int_{\omega_j-\frac{1}{n}}^{\omega_j+\frac{1}{n}} &\leq \frac{K}{n}\left(\max_{\mu\in[\omega_j-\frac{1}{n},\,\omega_j+\frac{1}{n}]}\frac{\partial}{\partial\mu}f(\mu)\right)\int_{\omega_j-\frac{1}{n}}^{\omega_j+\frac{1}{n}}\left(\frac{\sin^2 n(\omega_j-\mu)/2}{\sin^2(\omega_j-\mu)/2}\right)|\mu-\omega_j|\,d\mu \\ &\leq \frac{K}{n}\int_{-\frac{1}{n}}^{\frac{1}{n}}\left(\frac{\sin^2 n\lambda/2}{\sin^2\lambda/2}\right)|\lambda|\,d\lambda \leq \frac{K}{n^2}\int_{-\pi}^{\pi}\frac{\sin^2(n\lambda/2)}{\sin^2(\lambda/2)}\,d\lambda = O(n^{-1}).\end{aligned}$$

The last step is obtained from

$$\int_{-\pi}^{\pi}\left[\frac{\sin(n\lambda/2)}{\sin(\lambda/2)}\right]^2 d\lambda = 2\pi n.$$

The fourth and fifth terms are bounded by $O(\log n/n)$ and $\frac{K}{n}\int_{\omega_j+\epsilon}^{\pi} = O\left(n^{-1}\right)$, respectively. The proof for $\mathrm{E}(J_n(\omega_j)J_n(\omega_j)/\sqrt{f(\omega_j)f(\omega_j)})$ is similar. □

Proof of Lemma 4.8. Without loss of generality, assuming that $0<j\leq k<n$, we consider three situations (a) $j,k\in\Lambda_1$, (b) $j,k\in\Lambda_2$ and (c) $j,k\in\Lambda_3$. We prove that for each case, the covariances of the normalized DFT at different Fourier frequencies are bounded by a term with order $\log n/n$. First, define

$$F_{j,k} = \frac{1}{\sqrt{f(\omega_j)f(\omega_{-k})}}\frac{1}{2\pi n}\frac{\sin n(\omega_j-\mu)/2}{\sin(\omega_j-\mu)/2}\frac{\sin n(\omega_k-\mu)/2}{\sin(\omega_k-\mu)/2}.$$

For (a), note that by using $\int_{-\pi}^{\pi}F_{j,-k}(\mu)\,d\mu = 0$ for $j,k\in\Lambda_1$, we have

$$\mathrm{E}\left(\frac{J_n(\omega_j)J_n(\omega_k)}{\sqrt{f(\omega_j)f(\omega_k)}}\right) = \int_{-\pi}^{\pi}\left[F_{j,-k}(\mu)(f(\mu)-f(\omega_k))\right]d\mu, \tag{4.24}$$

The integral in (4.24) is divided into seven parts

$$\int_{-\pi}^{\pi} = \int_{-\pi}^{-\omega_k-\epsilon} + \int_{-\omega_k-\epsilon}^{-\omega_k+\epsilon} + \int_{-\omega_k+\epsilon}^{-\frac{1}{n}} + \int_{-\frac{1}{n}}^{\frac{1}{n}} + \int_{\frac{1}{n}}^{\omega_j-\epsilon} + \int_{\omega_j-\epsilon}^{\omega_j+\epsilon} + \int_{\omega_j+\epsilon}^{\pi}. \tag{4.25}$$

The terms $\int_{-\omega_k-\epsilon}^{-\omega_k+\epsilon} + \int_{\omega_j-\epsilon}^{\omega_j+\epsilon}$ in (4.25) are bounded by

$$\begin{aligned}
& \frac{K}{n}\left\{\max_{\mu\in[-\omega_k-\epsilon,\ -\omega_k+\epsilon]}\left(\frac{1}{|2\sin(\mu/2)|^{2d}}\frac{|\sin n(\omega_j-\mu)/2|}{|\sin(2\pi\omega_j-\mu)/2|}\right)\left(\int_{-\omega_k-\epsilon}^{-\omega_k+\epsilon}+\int_{\omega_j-\epsilon}^{\omega_j+\epsilon}\right)\right.\\
& \left.+\max_{\mu\in[\omega_j-\epsilon,\ \omega_j+\epsilon]}\left(\frac{1}{|2\sin(\mu/2)|^{2d}}\frac{|\sin n(\omega_k+\mu)/2|}{|\sin(\omega_k+\mu)/2|}\right)\left(\int_{-\omega_k-\epsilon}^{-\omega_k+\epsilon}+\int_{\omega_j-\epsilon}^{\omega_j+\epsilon}\right)\right\}\\
\leq\ & \frac{K}{n}\int_{-\pi}^{\pi}\left|\frac{\sin(n\mu/2)}{\sin(\mu/2)}\right| d\mu = O\left(\frac{\log n}{n}\right),
\end{aligned}$$

where the last step follows from Lemma 4.6. Also, $\int_{-\pi}^{-\omega_k-\epsilon} + \int_{\omega_j+\epsilon}^{\pi}$ are bounded by $O(n^{-1})$. Finally,

$$\begin{aligned}
& \frac{K}{2\pi n}\left(\int_{-\frac{1}{n}}^{\frac{1}{n}}+\int_{\frac{1}{n}}^{\omega_j-\epsilon}+\int_{-\omega_k+\epsilon}^{-\frac{1}{n}}\right)\\
\leq\ & \frac{K}{n}\int_{-\frac{1}{n}}^{\frac{1}{n}}\left(|\mu^{-2d}|+|\omega_j|^{-2d}\right) d\mu\\
& +\left(\max_{\mu\in[\frac{1}{n},\omega_j-\epsilon]}\frac{\sin n(\omega_j-\mu)/2}{\sin(\omega_j-\mu)/2}\right)\int_{\frac{1}{n}}^{\omega_j-\epsilon}\left(|\mu^{-2d}|+|\omega_j|^{-2d}\right) d\mu\\
& +\left(\max_{\mu\in[-\omega_k+\epsilon,-\frac{1}{n}]}\frac{\sin n(\omega_k+\mu)/2}{\sin(\omega_k+\mu)/2}\right)\int_{-\omega_k+\epsilon}^{-\frac{1}{n}}\left(|\mu^{-2d}|+|\omega_j|^{-2d}\right) d\mu\\
=\ & O(n^{-1}).
\end{aligned}$$

Next, to prove $\mathrm{E}(J_n(\omega_j)J_n(\omega_{-k})/\sqrt{f(\omega_j)f(\omega_{-k})}) = O(\log n/n)$ for (a), note that

$$\begin{aligned}
& \mathrm{E}\left(\frac{J_n(\omega_j)J_n(\omega_{-k})}{\sqrt{f(\omega_j)f(\omega_{-k})}}\right)\\
=\ & \int_{(\omega_j+\omega_k)/2}^{2\omega_j}\left(f(\mu)-f(\omega_j)\right)F_{j,k}(\mu)\, d\mu\\
& +\int_{\omega_k/2}^{(\omega_j+\omega_k)/2}\left(f(\mu)-f(\omega_k)\right)F_{j,k}(\mu)\, d\mu\\
& -\left(f(\omega_j)-f(\omega_k)\right)\int_{\omega_k/2}^{(\omega_j+\omega_k)/2}F_{j,k}(\mu)\, d\mu\\
& +\left(\int_{2\omega_j}^{\pi}+\int_{-\pi}^{\omega_k/2}\right)\left(f(\mu)-f(\omega_j)\right)F_{j,k}(\mu)\, d\mu. \qquad (4.26)
\end{aligned}$$

For the first part, as

$$\left|\frac{\sin n(\omega_j-\mu)/2}{\sin(\omega_j-\mu)/2}\right| \le \left|\frac{2}{\omega_j-\mu}\right|, \quad 0<|\omega_j-\mu|<\pi,$$

it follows that

$$\begin{aligned}
&\int_{(\omega_j+\omega_k)/2}^{2\omega_j} (f(\mu)-f(\omega_j))\, F_{j,k}(\mu)\, d\mu \\
\le\ & \frac{K}{2\pi n}\left(\max_{(\omega_j+\omega_k)/2\le\mu\le 2\omega_j} \frac{\partial}{\partial\mu} f(\mu)\right) \int_{(\omega_j+\omega_k)/2}^{2\omega_j} \left|\frac{\sin n(\omega_k-\mu)/2}{\sin(\omega_k-\mu)/2}\right| d\mu \\
\le\ & \frac{K}{2\pi n}\int_{-\pi}^{\pi}\left|\frac{\sin(n\lambda/2)}{\sin(\lambda/2)}\right| d\lambda = O\left(\frac{\log n}{n}\right).
\end{aligned}$$

For the second part, we have

$$\int_{\omega_k/2}^{(\omega_j+\omega_k)/2} (f(\mu)-f(\omega_k))\, F_{j,k}(\mu)\, d\mu = O\left(\frac{\log n}{n}\right).$$

The third part is bounded by

$$\begin{aligned}
& (\omega_j-\omega_k)\left(\max_{\omega_k\le\mu\le\omega_j}\frac{\partial}{\partial\mu}f(\mu)\right)\frac{K}{2\pi n}\int_{\omega_k/2}^{(\omega_j+\omega_k)/2}\left(\frac{2}{|\omega_j-\mu|}\,\frac{\sin n(\omega_k-\mu)/2}{\sin(\omega_k-\mu)/2}\right) d\mu \\
\le\ & (\omega_j-\omega_k)\left(\max_{\omega_k/2\le\mu\le(\omega_j+\omega_k)/2}\frac{1}{|\omega_j-\mu|}\right)\frac{K}{2\pi n}\int_{\omega_k/2}^{(\omega_j+\omega_k)/2}\left|\frac{\sin n(\omega_k-\mu)/2}{\sin(\omega_k-\mu)/2}\right| d\mu \\
=\ & O\left(\frac{\log n}{n}\right).
\end{aligned}$$

For the last part, we have

$$\begin{aligned}
&\int_{-\pi}^{\omega_k/2} (f(\mu)-f(\omega_j))\, F_{j,k}(\mu)\, d\mu \\
\le\ & \frac{K}{2\pi n}\left(\max_{-\epsilon\le\mu\le\epsilon}\frac{\sin n(\omega_j-\mu)/2}{\sin(\omega_j-\mu)/2}\frac{\sin n(\omega_k-\mu)/2}{\sin(\omega_k-\mu)/2}\right)\int_{-\epsilon}^{\epsilon}(|\mu|^{-2d}+|\omega_j|^{-2d})\, d\mu \\
=\ & O(n^{-1})
\end{aligned}$$

and

$$\int_{2\omega_j}^{\pi} (f(\mu)-f(\omega_j))\, F_{j,k}(\mu)\, d\mu = O(n^{-1}).$$

Thus, for (a), we have

$$\mathrm{E}\left(\frac{J_n(\omega_j)J_n(\omega_k)}{\sqrt{f(\omega_j)f(\omega_k)}}\right) = O\left(\frac{\log n}{n}\right) \quad \text{and} \quad \mathrm{E}\left(\frac{J_n(\omega_j)J_n(\omega_{-k})}{\sqrt{f(\omega_j)f(\omega_{-k})}}\right) = O\left(\frac{\log n}{n}\right).$$

To address the argument that b) is similar to the proof of a), we omit the details. For (c), note that

$$\begin{aligned} \mathrm{E}\left(\frac{J_n(\omega_j)J_n(\omega_k)}{\sqrt{f(\omega_j)f(\omega_k)}}\right) &\leq K\omega_k^{2d}\int_{-\pi}^{\pi}\frac{F_{jk}(\mu)}{|2\sin(\mu/2)|^{2d}}\,d\mu \\ &\leq \frac{K}{2\pi n}\frac{|2\pi k|^{2d}}{n^{2d}}\left[\int_{-\pi}^{-\epsilon}+\int_{-\epsilon}^{\epsilon}+\int_{\epsilon}^{\omega_j-\epsilon}+\int_{\omega_j-\epsilon}^{\omega_j+\epsilon}+\int_{\omega_j+\epsilon}^{\pi}\right]. \end{aligned}$$

The second term is bounded by

$$\frac{K}{2\pi n}\frac{|2\pi k|^{2d}}{n^{2d}}\max_{\mu\in[-\epsilon,\epsilon]}\left(\frac{\sin n(\omega_j-\mu)/2}{\sin(\omega_j-\mu)/2}\right)\int_{-\pi}^{\pi}\left|\frac{\sin n(\frac{2\pi k}{n}+\mu)/2}{\sin(\frac{2\pi k}{n}+\mu)/2}\frac{1}{|2\sin(\mu/2)|^{2d}}\right|d\mu$$
$$= O\left(n^{-1}\right).$$

The calculations for the other terms are similar, and the bounds of the terms are of order $\log n/n$. It is also true that $\mathrm{E}(J_n(\omega_j)J_n(\omega_{-k})/\sqrt{f(\omega_j)f(\omega_{-k})}) = O(\log n/n)$. Combining $(a)-(c)$, Lemma 4.8 follows. □

Proof of Lemma 4.11. Because

$$\mathrm{E}\left(J_n(\omega_j)J_n(\omega_{-j})-f(\omega_j)\right)=\int_{-\pi}^{\pi}F_n(\omega-\omega_j)(f(\omega)-f(\omega_j))\,d\omega,$$

where $F_n(\omega)$ is the Fejér kernel, it suffices to show that

$$\int_{-\pi}^{\pi}F_n(\omega-\omega_j)(f(\omega)-f(\omega_j))\,d\omega \geq \frac{K}{j}\omega_j^{-2d}. \tag{4.27}$$

Decompose the integral on the left side of (4.27) into five parts

$$\int_{-\pi}^{\pi}=\int_{\epsilon}^{\pi}+\int_{2\omega_j}^{\epsilon}+\int_{-2\omega_j}^{2\omega_j}+\int_{-\epsilon}^{-2\omega_j}+\int_{-\pi}^{-\epsilon}, \tag{4.28}$$

where for sufficiently large n, $2\omega_j<\epsilon<\pi/2$. The first term of (4.28) is bounded below by

$$\begin{aligned} \int_{\epsilon}^{\pi} &\geq \min_{\epsilon\leq\omega\leq\pi}\{f(\omega)-f(\omega_j)\}\int_{\epsilon}^{\pi}\frac{1}{2\pi n}\frac{\sin^2 n(\omega-\omega_j)/2}{\sin^2(\omega-\omega_j)/2}\,d\omega \\ &\geq \frac{K}{2\pi n}\int_{\epsilon-\omega_j}^{\pi-\omega_j}\frac{\sin^2 n\omega/2}{\sin^2\omega/2}\,d\omega=\frac{K}{n}. \end{aligned}$$

For small ϵ, let $\delta_\epsilon = \left[\frac{\epsilon - 2\omega_j}{n}\right]$, then using the property of the Fejér kernel (Priestley, 1981), the second term of (4.28) is bounded below by

$$\begin{aligned}
\left|\int_{2\omega_j}^{\epsilon}\right| &\geq \min_{2\omega_j \leq \omega \leq \epsilon}\left(\frac{\partial}{\partial\omega} f(\omega)\right) \int_{2\omega_j}^{\epsilon} \frac{1}{2\pi n} \frac{\sin^2 n(\omega-\omega_j)/2}{\sin^2(\omega-\omega_j)/2}(\omega-\omega_j)\, d\omega \\
&\geq \frac{K\epsilon^{-1-2d}}{n} \sum_{k=0}^{n-1} \int_{\omega_j+\delta_\epsilon k}^{\omega_j+\delta_\epsilon(k+1)} \frac{\sin^2 n\omega/2}{\sin^2 \omega/2}\omega\, d\omega \\
&\geq \frac{K}{n}\sum_{k=0}^{n-1}\left(\min_{\omega_j+\delta_\epsilon k \leq \omega \leq \omega_j+\delta_\epsilon(k+1)} \omega\right) \int_{\omega_j+\delta_\epsilon k}^{\omega_j+\delta_\epsilon(k+1)} \frac{\sin^2 n\omega/2}{\sin^2 \omega/2}\, d\omega \\
&\geq \sum_{k=0}^{n-1}\frac{K}{n^2}(\omega_j+\delta_\epsilon k) = K\left\{\frac{\omega_j}{n} + \frac{\epsilon - 2\omega_j}{n}\right\} \geq \frac{K}{n}.
\end{aligned}$$

Using an identical argument, we have $\int_{-\pi}^{-\epsilon} \geq K/n$ and $\int_{-\epsilon}^{-2\omega_j} \geq K/n$. For $\int_{-2\omega_j}^{2\omega_j}$ in (4.28), we have

$$\int_{-2\omega_j}^{2\omega_j} = \int_{-2\omega_j}^{\omega_j - 2\pi/n} + \int_{\omega_j-2\pi/n}^{\omega_j+2\pi/n} + \int_{\omega_j+2\pi/n}^{2\omega_j} = L_1 + L_2 + L_3$$

Note that L_3 is bounded below by

$$\begin{aligned}
& \min_{\omega_j+2\pi/n \leq \omega \leq 2\omega_j}\left(\frac{\partial}{\partial\omega} f(\omega)\right) \int_{\omega_j+2\pi/n}^{2\omega_j} \frac{1}{2\pi n}\frac{\sin^2 n(\omega-\omega_j)/2}{\sin^2(\omega-\omega_j)/2}(\omega-\omega_j)\, d\omega \\
\geq\ & K\omega_j^{-(1+2d)} \int_{2\pi/n}^{\omega_j} \frac{1}{2\pi n}\frac{\sin^2 n\omega/2}{\sin^2\omega/2}\omega\, d\omega \\
\geq\ & K\omega_j^{-(1+2d)}\left(\min_{2\pi/n \leq \omega \leq \omega_j} \omega\right) \int_{2\pi/n}^{\omega_j} \frac{1}{2\pi n}\frac{\sin^2 n\omega/2}{\sin^2\omega/2}\, d\omega \\
=\ & K\omega_j^{-(1+2d)}\frac{1}{n}\int_{2\pi/n}^{\omega_j} \frac{1}{2\pi n}\frac{\sin^2 n\omega/2}{\sin^2\omega/2}\, d\omega.
\end{aligned}$$

Because the integral in the last equation is bounded above by small $\epsilon > 0$ from Section 6.1 in Priestley (1981), the lower bound of L_3 is no more than $K\frac{1}{j}\omega_j^{-2d}$. By symmetry, it is also true that the lower bound of L_1 is no more than $K\frac{1}{j}\omega_j^{-2d}$. Next,

$$\begin{aligned}
L_2 &= \left(\int_{-2\pi/n}^{-1/n} + \int_{-1/n}^{1/n} + \int_{1/n}^{2\pi/n}\right)\frac{1}{2\pi n}\frac{\sin^2 n\omega/2}{\sin^2\omega/2}\left\{(\omega+\omega_j)^{-2d} - \omega_j^{-2d}\right\} d\omega \\
&= L_{21} + L_{22} + L_{23}.
\end{aligned}$$

Note that

$$L_{22} \geq \omega_j^{-1-2d} \int_{-1/n}^{1/n} \frac{1}{2\pi n} \frac{\sin^2 n\omega/2}{\sin^2 \omega/2} \omega \, d\omega = 0.$$

For L_{23},

$$L_{23} \geq \left\{ \min_{\omega \in [1/n, 2\pi/n]} \min_{\xi \in [\omega_j, \omega_j + \omega]} \xi^{-1-2d} \omega \right\} \int_{1/n}^{2\pi/n} \frac{1}{2\pi n} \frac{\sin^2 n\omega/2}{\sin^2 \omega/2} \, d\omega \geq K \frac{1}{j} \omega_j^{-2d}.$$

We also have $L_{21} \geq K \frac{1}{j} \omega_j^{-2d}$. From the above analysis, the division $\int_{-2\omega_j}^{2\omega_j}$ is bounded below by $K \frac{1}{j} \omega_j^{-2d}$. Then, (4.16) is established by collecting the lower bounds of each part in (4.28).

Chapter 5

Bartlett Correction for EL with non-Gaussian Short-Memory Time Series

In practice, the Gaussian assumption can be too restrictive or even unrealistic. For example, in financial markets, when using an AR(1) model $\log P_t = \phi_1 \log P_{t-1} + \epsilon_t$ to model the log-price of a security, a skewed t-distribution or a stable Paretian distribution with heavy-tails is often used to describe the error process $\{\epsilon_t\}$ (see Fama (1965) and Mandelbrot (1963)). It is therefore meaningful to develop an EL method for statistical inference problems with non-Gaussian time series. The non-Gaussian distribution of financial data, seen in Intel's stock return, can be observed in the density estimator plot in Figure 5.1. The Leptokurtic shape indicates heavy-tailed data.

It is, however, still unclear whether EL is Bartlett correctable for non-Gaussian time series. Following the standard argument with i.i.d. data in Diciccio, Hall and Romano (1991), we must characterize the third- and the

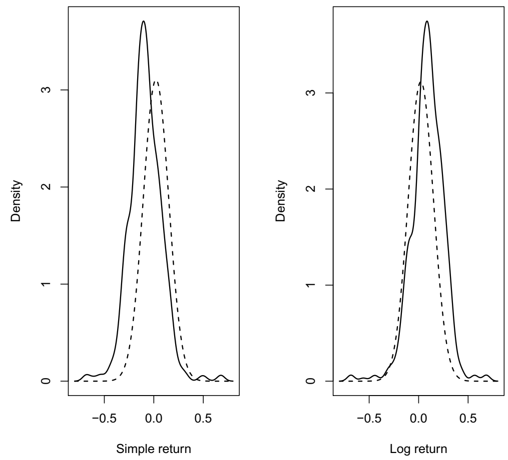

Figure 5.1: Comparison of empirical and normal densities for monthly returns of Intel's stock from January 1973 to December 2008. Normal density, shown by the dashed line, used the sample mean and standard deviation of return.

fourth-order cumulants of the signed root empirical log-likelihood ratio, denoted by SR, and show that these cumulants decay to zero at fast rates of n^{-3} and n^{-4}, respectively. In establishing the formula, the coefficients of Edgeworth polynomials are functionals of the first four cumulants of SR. If we use the same periodogram-based EL version as Monti (1997), then higher-order cumulants of SR involve higher-order cumulants of peridogram ordinates for the underlying process, e.g., $\log P_t$ in the AR(1) model. For a Gaussian process, the cumulants of the periodogram ordinates can be decomposed into products of the second-order cumulants of the discrete Fourier transform (DFT), and higher-order cumulants of the DFT are negligible. Under these circumstances, calculations on the higher-order cumulants of SR can be derived, but, this property no longer holds for non-Gaussian processes. Higher-order cumulants of the DFT cannot be ignored. To circumvent this difficulty, we show that the sixth and eighth cumulants of the DFT are of orders n^{-2} and n^{-3}, and they vanish in the third- and fourth-order cumulants of SR, respectively. It is also shown that although the fourth cumulant of DFT with order n^{-1} accounts for the second-order cumulant of SR, with a known variance of innovation, this non-zero quantity is canceled in the third- and the fourth-order cumulants of SR. To sum up, the higher-order cumulants of the DFT do not affect the Bartlett correctability of EL for non-Gaussian dependent processes.

This chapter is organized as follows. Section 5.1 reviews the Bartlett correction of EL for Gaussian weakly dependent time series. In Section 5.2, we derive the asymptotic expansion of the signed root empirical log-likelihood ratio and establish the validity of Edgeworth expansion under non-Gaussian assumption. Finally, Section 5.3 presents simulation studies demonstrating the satisfactory finite sample performance of the Bartlett correction of AR models with non-Gaussian noise. Proofs of the technical results are given in Section 5.4.

5.1 Bartlett Correction for Time Series

As introduced in Sections 2.2 and 3.3, let $\kappa_r(J_j)$ be the r-th cumulant of J_j, and $\kappa_{\epsilon,r}$ be the r-th cumulant of ϵ_t. The Whittle likelihood, defined by

$$\sum_{j=1}^{n}\left\{\log f(\omega_j,\theta)+\frac{I_n(\omega_j)}{f(\omega_j,\theta)}\right\}, \tag{5.1}$$

is an approximation to the Gaussian log-likelihood function. Then, the score function of the Whittle likelihood m_j is the first derivative of (5.1) and can be written as $m_j = g(\omega_j,\theta)(I_n(\omega_j)/f(\omega_j,\theta)-1)$, where $g(\omega,\theta)=\frac{\partial}{\partial\theta}\log f(\omega,\theta)$. The corresponding Whittle estimator is defined as the solution of $\sum_{j=1}^{n} m_j = 0$. Hosoya (1974, 1997) establishes the limit theory for the Whittle estimator for possible non-Gaussian short-memory processes. However, using the Whittle estimator to construct confidence regions involves an asymptotic covariance matrix estimation (see Hosoya (1997)). Using the EL avoids this difficulty. Based on the periodogram-based Whittle estimating function m_j, the profile

EL function is given by

$$\mathcal{R}_n(\theta) = \max_{p_j} \left\{ \prod_{j=1}^{n} np_j \mid \sum_{j=1}^{n} p_j m_j = 0,\ \sum_{j=1}^{n} p_j = 1,\ p_j \geq 0 \right\}.$$

Then, the log-EL ratio for the problem of testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ becomes

$$l(\theta) = 2\sum_{j=1}^{n} \mathcal{R}_n(\theta). \tag{5.2}$$

For Gaussian processes, $\kappa_r(J_j) = 0$ for $r \geq 3$ such that $I_n(\omega_j)$ are asymptotically independent and unbiased estimators of $f(\omega_j, \theta)$. Based on this result, one can show that $\mathrm{E}(l(\theta_0)) = 1 + b/n + O(n^{-2})$, for some constant b. The Bartlett correction of EL means dividing $l(\theta_0)$ by the quantity $1 + b/n$, which scales the statistic to equate the mean of chi-squared distribution more accurately (see Bartlett (1937)). For this reason, the Bartlett corrected EL confidence interval is given by

$$\mathbf{I}'_{n,1-\alpha} = \left\{ \theta \mid l(\theta) \leq \left(1 + \frac{b}{n}\right) \chi^2_{1,1-\alpha} \right\}.$$

The coverage error defined by $\Pr(\theta_0 \in \mathbf{I}'_{\alpha}) - (1-\alpha)$ is reduced from order $O(n^{-1})$ to $o(n^{-1})$ for Gaussian SMTS (see Chan and Liu (2010)), and from $O(\log^6 n/n)$ to $O(\log^3 n/n)$ for Gaussian LMTS with a different correction factor (see Chapter 4).

To establish the Bartlett correctability of EL for non-Gaussian SMTS, we use the signed root empirical log-likelihood ratio approach proposed by Diciccio, Hall and Romano (1991). If $l(\theta_0) \xrightarrow{d} \chi^2_k$, then we may write $l(\theta_0) = W'W$, where W is asymptotically normal. Expanding a power series in $n^{-1/2}$ of W, i.e. $W = \sqrt{n}SR + O_p(n^{-3/2})$, where $SR = R_1 + R_2 + R_3$, where $R_j = O_p(n^{-j/2})$, SR is called the signed root empirical log-likelihood ratio statistic and $l(\theta_0) = (\sqrt{n}SR)^2 + O_p(n^{-3/2})$. In particular, the density function $\pi(x)$ of $\sqrt{n}SR$ admits the Edgeworth expansion

$$\pi(x) = \phi(x) + \frac{r_1(x)\phi(x)}{\sqrt{n}} + \frac{r_2(x)\phi(x)}{n} + \frac{r_3(x)\phi(x)}{n^{3/2}} + O(n^{-2}), \tag{5.3}$$

where r_j is a polynomial with a degree of at most $3j$ and is an odd or even function based on whether j is odd or even (see Hall (1992)). Here, $\phi(x)$ is the probability density function (p.d.f.) of a standard normal random variable. In essence, Bartlett correction works for EL, but not in general for bootstrap, because the fourth- and sixth-order polynomials in $r_2(x)$ vanish for EL, and not in other cases. This is a consequence of the fact that the third and fourth cumulants of $\sqrt{n}SR$ decay to zero with rates of $O(n^{-3/2})$ and $O(n^{-2})$, respectively. For Gaussian processes, the cumulants attain these rates because $\kappa_r(J_j) = 0$ for $r \geq 3$, and terms with products of $\mathrm{cum}(J_j, J_j)$ and $\mathrm{cum}(J_j, J_{-j})$ of smaller rates n^{-1} are canceled in the cumulants expansion (see Chan and Liu (2010) and Chapter 4). However, for non-Gaussian processes, $\kappa_r(J_j)$ is non-negligible and it is unknown whether terms with non-zero $\kappa_r(J_j)$ can be canceled. Due to the existence of higher order cumulants, orders of the third and the fourth cumulants of SR may become larger and the associate degrees 4 and 6 terms in $r_2(x)$ may be non-vanishing. To deal with these problems, we need to evaluate the higher-order cumulants of DFT as functionals of higher order cumulants of innovation and their effects on the third and the fourth cumulants of SR.

5.2 Main Results

Before establishing the coverage errors of EL and the Bartlett correctability, we impose some assumptions on the time series under consideration.

3.1 Regularity Conditions (RC).

a. The linear process $\{X_t\}$ has a representation $X_t = \sum_{j=0}^{\infty} a_j(\beta)\epsilon_{t-j}$, where $\{\epsilon_t\}$ is an independent innovation process with a known variance of σ_ϵ^2. In addition, ϵ_t has a finite sixteenth-order cumulant, i.e. $\kappa_{\epsilon,16} < \infty$. $\sum_{j=0}^{\infty} |a_j(\beta)|^2 <$

∞, where $a_j(\beta) = 0$ for $j < 0$ and $\theta = (\beta, \sigma_\epsilon^2)' \in \Theta \subset \mathbb{R}^2$, where the parameter space Θ is a compact set.

b. The spectral density function for process $\{X_t\}$ in a is given by

$$f(\omega, \theta) = \frac{\sigma_\epsilon^2}{2\pi}|a_\beta(\omega)|^2, \quad \text{for} \quad \omega \in \Pi = [-\pi, \pi],$$

where $a_\beta(\omega) = \sum_{j\in\mathbb{Z}} a_j(\beta)e^{ij\omega}$, with $i = \sqrt{-1}$. It is also assumed that $f(\omega, \theta)$ has continuous second-order derivatives in domain Π.

c. $\{X_t\}$ satisfies the Cramér's condition, i.e. $\limsup_{\tau\to\infty} |\mathrm{E}(\exp(i\tau X_t))| < \infty$.

Remark on the Assumptions: One important requirement in our assumptions is the known variance of innovation processes σ_ϵ^2, that is, for the parametric model class of spectral densities for $\{X_t\}$,

$$\mathcal{F} \equiv \left\{ f(\omega, \theta) = \frac{\sigma_\epsilon^2}{2\pi} f^*(\omega, \beta) \right\},$$

we only consider inference on parameters described by the kernel $f^*(\omega, \beta)$, and treat σ_ϵ^2 as a nuisance parameter. If σ_ϵ^2 is unknown, then the asymptotic chi-squared distribution of periodogram-based EL may generally fail for non-Gaussian processes. This phenomenon is also implied in Nordman and Lahiri (2006). Thus, in the following, we denote $\theta = \beta$ as the parameter of interest. Condition a requires that the innovation process has a sixteenth-order cumulant, which is stronger than the eighth-order cumulant of ϵ_t in Nordman and Lahiri (2006). In developing the Bartlett correctability of EL with i.i.d. data, Diccicio, Hall and Romano (1991) assumed sufficiently many moments of the underlying distribution. The moment condition is necessary when higher order asymptotics of EL are studied. Condition b restricts our discussion within SMTS. and Condition c is a regular condition to establish the validity of Edgeworth expansion.

3.2 Edgeworth Expansion

For SMTS, it is shown in Brillinger (1981) that

$$\mathrm{E}(I_n(\omega_j)) = f(\omega_j, \theta_0) + \frac{b_I(\omega_j)}{n} + o(n^{-1}),$$

where $b_I(\omega) = -\frac{1}{2\pi}\sum_{u=-\infty}^{\infty} |u|\gamma_{\theta_0}(u)\exp(-iu\omega)$ is the bias term. We have

$$\mathrm{E}(\overline{m}) = \frac{b_m}{n} + o(n^{-1}),$$

where $\overline{m} = \frac{1}{n}\sum_{j=1}^{n} m_j$ and $b_m = \frac{1}{n}\sum_{j=1}^{n} g(\omega_j, \theta_0) f(\omega_j, \theta_0) b_I(\omega_j)$. Let the higher-order moments and centered moments, respectively, be

$$\lambda_k = \mathrm{E}\left(\frac{1}{n}\sum_{j=1}^{n} m_j^k\right) \quad \text{and} \quad \Delta_k = \frac{1}{n}\sum_{j=1}^{n} m_j^k - \lambda_k, \quad \text{for} \quad k = 2, 3, 4.$$

To derive the asymptotic expansion of $l(\theta_0)$, we first establish the orders of λ_k and Δ_k in the next Lemma, which also provides the order of magnitude to derive an explicit form of $\sqrt{n}SR$.

Lemma 5.1 Under **RC**, we have

$$\lambda_k = O(1), \qquad \text{and} \qquad \Delta_k = O_p\left(\frac{1}{\sqrt{n}}\right), \quad \text{for} \quad k = 2, 3, 4.$$

Proof. The proof is given in the Appendix. ■

As in Zhang (1996) and Chapter 4, and given Lemma 5.1, expanding $l(\theta_0)$ in (5.2) to a term of order $O_p(n^{-3/2})$ and equating the stochastic expansion to nSR^2 gives us $SR = R_1 + R_2 + R_3 + O_p(n^{-2})$, where

$$\begin{aligned} R_1 &= \frac{\overline{m}}{\sqrt{\lambda_2}}, \\ R_2 &= \frac{1}{3}\frac{\lambda_3 \overline{m}^2}{\lambda_2^{5/2}} - \frac{1}{2}\frac{\overline{m}\Delta_2}{\lambda_2^{3/2}}, \\ R_3 &= \frac{3}{8}\frac{\overline{m}\Delta_2^2}{\lambda_2^{5/2}} + \frac{1}{3}\frac{\overline{m}^2\Delta_3}{\lambda_2^{5/2}} - \frac{5}{6}\frac{\lambda_3\overline{m}^2\Delta_2}{\lambda_2^{7/2}} + \frac{4}{9}\frac{\lambda_3^2\overline{m}^3}{\lambda_2^{9/2}} - \frac{1}{4}\frac{\lambda_4\overline{m}^3}{\lambda_2^{7/2}}. \end{aligned} \tag{5.4}$$

Calculating the first four cumulants of singed root decomposition is equivalent to calculating the higher-order cumulants of the Whittle-type score function. The next lemma provides a general rule to deal with such problems for non-Gaussian SMTS. The idea is to decompose the higher-order cumulants of the periodogram-based score function into products of the higher-order cumulants of the DFT. This result shows that the term with $\kappa_{\epsilon,4}$ is significant, even though it was negligible for Gaussian processes.

Lemma 5.2 Under **RC**, for partition $s \subset \{1, \ldots, 8\}$, and $j \neq k \in \{1, \ldots, n\}$, we have

$$\begin{aligned} & \frac{1}{n^2} \sum_{j \neq k} \prod_{s} \mathrm{cum}\{m_{1_1}, \ldots, m_{\nu_1}; \nu_1 \in s_1\} \cdots \mathrm{cum}\{m_{1_p}, \ldots, m_{\nu_p}; \nu_p \in s_p\} \\ = & \ \frac{C}{n} \frac{\kappa_{\epsilon,4}}{\sigma_\epsilon^4} \iint_{\Pi^2} g(\omega, \theta)^a g(\lambda, \theta)^b \, d\omega \, d\lambda + O(n^{-2}), \end{aligned}$$

for $1_1, \ldots, \nu_p \in \{j, k\}$, and the product is over all indecomposable partitions $s = s_1 \cup \cdots \cup s_p$, where C is a generic constant corresponding to the cumulant decomposition of score functions m_j and m_k. For each partition, $\nu_p \geq 2$, ν_i are not all equal, and

$$a = \sum_{i=1}^{p} \sum_{\alpha=1}^{\nu} \mathbf{1}_{\{\alpha_i = j\}} \quad \text{and} \quad b = \sum_{i=1}^{p} \sum_{\alpha=1}^{\nu} \mathbf{1}_{\{\alpha_i = k\}},$$

where $\mathbf{1}_A = 1$ for true event A and $\mathbf{1}_A = 0$ otherwise.

Proof. The proof is given in Section 5.4. ■

Based on the formula for SR in (5.4) and the higher-order cumulants rule of periodogram-based score functions in Lemma 5.2, we obtain the cumulants' expansions of $\sqrt{n} SR$ in the next Lemma, where the detailed calculations are given in the Appendix. Using these established cumulants' expansions, the standard procedure (see Hall (1992)) leads to the Edgeworth expansion for non-Gaussian SMTS.

Lemma 5.3 Let $\kappa_j(\sqrt{n}SR)$, $j = 1, 2, 3, 4$, be the first four cumulants of the signed root empirical log-likelihood ratio. Under **RC** and (5.4), $\kappa_j(\sqrt{n}SR)$ has the asymptotic expansion

$$\begin{aligned}\kappa_1(\sqrt{n}SR) &= \frac{k_{1,1}}{\sqrt{n}} + \frac{k_{1,2}}{n} + \frac{k_{1,3}}{n^{3/2}} + O(n^{-2}),\\ \kappa_2(\sqrt{n}SR) &= 1 + \frac{k_{2,1}}{\sqrt{n}} + \frac{k_{2,2}}{n} + \frac{k_{2,3}}{n^{3/2}} + O(n^{-2}),\\ \kappa_3(\sqrt{n}SR) &= O(n^{-3/2}),\ \kappa_4(\sqrt{n}SR) = O(n^{-2}),\end{aligned}$$

where

$$k_{1,1} = \frac{1}{\lambda_2^{1/2}}\frac{1}{2\pi}\int_{\Pi} b_I(\omega)g(\omega,\theta)f(\omega,\theta)^{-1}\,d\omega - \frac{2}{3}\frac{\lambda_3}{\lambda_2^{3/2}}, \tag{5.5}$$

$$k_{2,2} = 4\frac{\lambda_3^2}{\lambda_2^3} - \frac{29}{18}\frac{\lambda_4}{\lambda_2^2} + \frac{26}{9}\frac{\lambda_2\lambda_4}{\lambda_3^2} - \frac{\kappa_{\epsilon,4}}{\sigma_\epsilon^4}, \tag{5.6}$$

and $k_{1,2} = k_{1,3} = k_{2,1} = k_{2,3} = 0$.

Proof. The proof is given in Section 5.4. ■

Under condition **RC**, the Edgeworth polynomials $r_1(x)$ and $r_2(x)$ in (5.3) admit the forms

$$\begin{aligned} r_1(x) &= \sqrt{n}\{\kappa_1(\sqrt{n}SR)x + \frac{1}{6}\kappa_3(\sqrt{n}SR)(x^3-3x)\} = k_{1,1}x + O(n^{-1}),\\ r_2(x) &= \frac{n}{2}\{\kappa_2(\sqrt{n}SR) - 1 + \kappa_1(\sqrt{n}SR)^2\}\{x^2-1\}\\ &= \frac{1}{2}(k_{1,1}^2 + k_{2,2})(x^2-1) + O(n^{-1}),\end{aligned}$$

and $r_3(x)$ is an odd polynomial with a degree of no more than 9. For non-Gaussian SMTS, $r_1(x)$ and $r_2(x)$ are different from their Gaussian counterparts because the terms $k_{1,1}$ and $k_{2,2}$ in (5.5) and (5.6) involve non-zero $\kappa_{\epsilon,4}$. In particular, it is proven that with known σ_ϵ^2, terms with $\kappa_{\epsilon,4}$ can be canceled in $\kappa_3(\sqrt{n}SR)$ and $\kappa_4(\sqrt{n}SR)$, and terms with $\kappa_{\epsilon,r}$ are of orders n^{-2} when r is odd, or r is even and larger than 6. Therefore, $\kappa_3(\sqrt{n}SR)$ and $\kappa_4(\sqrt{n}SR)$ involving $\kappa_{\epsilon,r}$ for $r \geq 3$ decay fast enough to ensure the degree 2 of $r_2(x)$. With the particular form of $r_2(x)$, where fourth- and sixth-degrees of the polynomials vanish, the coverage error of EL is of an order n^{-1}, which can be reduced to n^{-2} by the Bartlett correction technique.

Theorem 5.4 Under **RC**, based on the Edgeworth expansion (5.3), for sufficiently large n,

$$P(l(\theta_0) \leq \chi^2_{1,1-\alpha}) = 1 - \alpha + O(n^{-1}).$$

Proof. The procedure is the same as that in Chapter 4 and the proof is omitted. ■

Along with the discussion on the Bartlett correctability of EL, the Edgeworth expansion for non-Gaussian processes admits a particular form (5.3), in which $r_2(x)$ is a degree 2 polynomial, such that we can scale the chi-squared critical value by the mean of the log-EL ratio to achieve better accuracy. In addition, it should be noted that the Bartlett correction factor for a non-Gaussian process is no longer the same as that for a Gaussian process.

Theorem 5.5 Under **RC**, EL is Bartlett correctable,

$$P(l(\theta_0) \leq (1 + b/n)\chi^2_{1,1-\alpha}) = 1 - \alpha + O(n^{-2}),$$

where $b = k_{1,1}^2 + k_{2,2}$, and $k_{1,1}, k_{2,2}$ are given by (5.5) and (5.6).

Proof. Based on the cumulant expansions of Lemma 5.3, we have

$$\mathrm{E}(l(\theta_0)) = \mathrm{E}(nSR^2) = \kappa_2(\sqrt{n}SR) + \kappa_1(\sqrt{n}SR)^2 = 1 + \frac{k_{2,2} + k_{1,1}}{n} + O(n^{-2}).$$

As introduced in Section 2, the mean of $l(\theta_0)/(1 + b/n)$ approximates more accurately to the mean of a chi-squared random variable with 1 degree of freedom. Because dividing $l(\theta_0)$ by $1+b/n$ is equivalent to multiplying $\sqrt{n}SR$ by $1-b/2n$, we use the conventional approach to derive the coverage error of the corrected statistic $\sqrt{n}SR(1-b/2n)$. For the Bartlett-corrected signed root empirical log-likelihood ratio $SR^* = SR(1-b/2n)$, it can be shown that

$$\begin{aligned}\kappa_1(\sqrt{n}SR^*) &= \frac{k_{1,1}}{\sqrt{n}} - \frac{k_{1,1}^3 + k_{1,1}k_{2,2}}{2n^{3/2}} + O(n^{-2}),\\ \kappa_2(\sqrt{n}SR^*) &= 1 - \frac{k_{1,1}^2}{n} + O(n^{-2}).\end{aligned}$$

In this case, the Edgeworth expansion of the p.d.f. $\pi^*(x)$ of $\sqrt{n}SR^*$ is

$$\begin{aligned}\pi^*(x) &= \phi(x) + \frac{r_1^*(x)\phi(x)}{\sqrt{n}} + \frac{r_2^*(x)\phi(x)}{n} + \frac{r_3^*(x)\phi(x)}{n^{3/2}} + O(n^{-2})\\ &= \phi(x) + \{\kappa_1(\sqrt{n}SR^*) + \frac{1}{6}\kappa_3(\sqrt{n}SR^*)(x^3-3x)\}\phi(x)\\ &\quad + \frac{1}{2}\{\kappa_2(\sqrt{n}SR^*) - 1 + \kappa_1(\sqrt{n}SR^*)^2\}(x^2-1)\phi(x)\\ &\quad + \frac{r_3^*(x)\phi(x)}{n^{3/2}} + O(n^{-2}).\end{aligned}$$

Here, $k_2(\sqrt{n}SR^*) - 1 + k_1(\sqrt{n}SR^*)^2 = 1 - \frac{k_{1,1}^2}{n} - 1 + \frac{k_{1,1}^2}{n} = 0$, so $r_2^*(x)$ is of order n^{-1}. Applying the preceding Edgeworth expansion gives us

$$\begin{aligned}&P(l(\theta_0) \le (1+b/n)\chi^2_{1,1-\alpha}) = P(nSR^2 \le (1+b/n)\chi^2_{1,1-\alpha})\\ =& \int_{-\sqrt{\chi^2_{1,1-\alpha}}}^{\sqrt{\chi^2_{1,1-\alpha}}} \phi(x)\,dx + \int_{-\sqrt{\chi^2_{1,1-\alpha}}}^{\sqrt{\chi^2_{1,1-\alpha}}} \left\{\frac{r_1^*(x)}{\sqrt{n}} + \frac{r_2^*(x)}{n} + \frac{r_3^*(x)}{n^{3/2}}\right\}\phi(x)\,dx + O(n^{-2})\\ =& 1 - \alpha + O(n^{-2}).\end{aligned}$$

Integrals of the orders $n^{-1/2}$ and $n^{-3/2}$ equal zero due to the oddness of polynomials $r_1^*(x)$ and $r_3^*(x)$. ■

5.3 Simulation Study

In this section, an Monte Carlo simulation is conducted to assess the finite sample performance of Bartlett correctability in improving the coverage accuracy of an EL confidence interval for non-Gaussian SMTS. Simple AR(1) models are used with innovation processes following non-Gaussian distribution: t_5, t_8, Exp(1) and χ^2_5. All of the simulations are carried out in R version 2.15.2. We consider the AR(1) model

$$(1-\phi B)X_t = \epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} (0, \sigma^2_\epsilon),$$

where B is the back-shift operator, i.e., $BX_t = X_{t-1}$, and ϵ_t has a mean of zero and a finite known variance of σ^2_ϵ. Note that σ^2_ϵ differs according to different noise generation mechanisms. The true values for the AR parameter ϕ are 0.3 and 0.6. Let ϕ_0, $\phi_{\alpha/2}$ and $\phi_{1-\alpha/2}$ be the true values of the AR parameter, the lower and the upper endpoints of the confidence interval, respectively. In Table 5.1, we study the coverage error of the 95% confidence interval (i.e. $\alpha = 0.05$)

$$\left| P\{(\phi_0 < \phi_{[\alpha/2]}) \cup (\phi_0 > \phi_{[1-\alpha/2]})\} - \alpha \right|,$$

for different sample sizes $n = 200$, 400, 600, 800. In each case, 1,000 replications are drawn. To calculate the critical value of Bartlett-corrected EL, the Bartlett correction factor $1+b/n$ is approximated using the Bootstrap method. The detailed procedure is the same as Monti (1997). In the Bootstrap sampling, the Whittle maximum likelihood estimator is adopted as a consistent estimator and the resampling replication B is set to be 500.

Table 5.1 shows that Bartlett correction significantly reduces the coverage error for each white noise distribution. As the sample size increases, the coverage error for both the EL confidence interval and Bartlett-corrected confidence interval decreases. Even for the possible heavy-tailed distribution t_5 with excess kurtosis 6, Bartlett correction successfully improves the coverage accuracy of the EL confidence intervals.

AR(1) Model									
		$\phi = 0.3$				$\phi = 0.6$			
n		t_5	t_8	Exp(1)	χ^2_5	t_5	t_8	Exp(1)	χ^2_5
200	EL	0.011	0.021	0.028	0.015	0.04	0.081	0.012	0.042
	Bart. EL	0.001	0.02	0.019	0.01	0.025	0.061	0.009	0.016
400	EL	0.044	0.035	0.019	0.006	0.016	0.029	0.013	0.021
	Bart. EL	0.034	0.031	0.014	0.003	0.009	0.019	0.011	0.015
600	EL	0.034	0.028	0.01	0.004	0.011	0.018	0.014	0.011
	Bart. EL	0.03	0.024	0.006	0.002	0.004	0.014	0.005	0.005
800	EL	0.021	0.019	0.002	0.002	0.015	0.004	0.004	0.007
	Bart. EL	0.017	0.014	0.001	0.001	0.01	0.002	0.003	0.004

Table 5.1: Coverage errors of confidence intervals for AR(1) models, replications = $1,000$.

5.4 Lemmas and Proofs

In the following, let $\kappa(j,k) = \text{Cov}(J_j, J_k)$, $\kappa(j,k,l,m) = \text{cum}(J_j, J_k, J_l, J_m)$ and κ with different indices as higher-order cumulants of the DFT. For example, $\kappa(j,-j,j,-j,k,-k,k,-k)$ denotes the eighth-order cumulant of J_j and J_k; $\kappa_r(J_j)$ denotes the r-th cumulant of J_j.

Proof of Lemma 5.2:

To illustrate the proof, consider specific examples where $sp = 4$, $a = 2$ and $b = 2$. There are four possibilities,

$$\frac{1}{n^2}\sum_{j\neq k}\text{cum}(m_j, m_j, m_k, m_k), \qquad \frac{1}{n^2}\sum_{j\neq k}\text{cum}^2(m_j, m_k)$$
$$\frac{1}{n^2}\sum_{j\neq k}\text{cum}(m_j)\text{cum}(m_j, m_k, m_k),$$
$$\frac{1}{n^2}\sum_{j\neq k}\text{cum}(m_j)\text{cum}(m_k)\text{cum}(m_j, m_k). \tag{5.7}$$

For the first term in (5.7), we have

$$\begin{aligned} &\frac{1}{n^2}\sum_{j\neq k}\mathrm{cum}(m_j,m_j,m_k,m_k) \\ = &\frac{1}{n^2}\sum_{j\neq k}\frac{g(\omega_j,\theta)^2 g(\omega_k,\theta)^2}{f(\omega_j,\theta)^2 f(\omega_k,\theta)^2}\mathrm{cum}(I_n(\omega_j),I_n(\omega_j),I_n(\omega_k),I_n(\omega_k)). \quad (5.8)\end{aligned}$$

Using the cumulant decomposition principle in Brillinger (1981),

$$\mathrm{cum}(I_n(\omega_j),I_n(\omega_j),I_n(\omega_k),I_n(\omega_k)) = \sum_{\boldsymbol{\nu}:p=1}^{8}\prod_{j=1}^{p}\mathrm{cum}\{J_{k_j};k_j\in\nu_j\},$$

where the summation is taken over all indecomposable partitions $\boldsymbol{\nu}=\nu_1\cup\cdots\cup\nu_p, p=1,\ldots,8$. By the decomposition of the cumulants' principle introduced in Brillinger (1981), we first need to consider the following partitions, where in each partition, the DFTs are taken at conjugate frequencies;

$$\begin{aligned} p=1, &\quad (j,j,-j,-j,k,k,-k,-k), \\ p=2, &\quad (j,-j)\cup(j,-j,k,k,-k,-k), \\ &\quad (j,-j,k,-k)\cup(j,-j,k,-k), \\ p=3, &\quad (j,-j)\cup(k,-k)\cup(j,-j,k,-k). \end{aligned} \quad (5.9)$$

For the partition $p = 1$ in (5.9), under **RC**, it follows that

$$\begin{aligned}
&\kappa(j, -j, j, -j, k, -k, k, -k) \\
= &\frac{\kappa_{\epsilon,8}}{(2\pi n)^4} \sum_{t_1,\ldots,t_4=1}^{n} \sum_{s_1,\ldots,s_4=1}^{n} \prod_{\alpha=1}^{2} \prod_{\beta=3}^{4} \cos(t_\alpha - s_\alpha)\omega_j \cos(t_\beta - s_\beta)\omega_k \\
&\times \sum_{p=-\max(t_1,\ldots,s_4)}^{\infty} \prod_{l=1}^{4} a_{t_l+p} a_{s_l+p} \\
= &\frac{\kappa_{\epsilon,8}}{(2\pi n)^4} \sum_{r_1,\ldots,r_4=-(n-1)}^{n-1} \prod_{\alpha=1}^{2} \prod_{\beta=3}^{4} \cos r_\alpha \omega_j \cos r_\beta \omega_k \\
&\times \sum_{s_1 \in S_{r_1},\ldots,s_4 \in S_{r_4}} \sum_{p=-\max(s_1+r_1,\ldots,s_4+r_4)}^{\infty} \prod_{l=1}^{4} a_{p+s_l+r_l} a_{p+s_l} \\
= &\frac{\kappa_{\epsilon,8}}{(2\pi n)^4} \sum_{r_1,\ldots,r_4=-(n-1)}^{n-1} \prod_{\alpha=1}^{2} \prod_{\beta=3}^{4} \cos r_\alpha \omega_j \cos r_\beta \omega_k \sum_{s_1 \in S_{r_1},\ldots,s_4 \in S_{r_4}} \sum_{p=-\infty}^{\infty} \prod_{l=1}^{4} a_{p+s_l+r_l} a_{p+s_l} \\
= &\frac{\kappa_{\epsilon,8}}{(2\pi n)^4} \sum_{r_1,\ldots,r_4=-(n-1)}^{n-1} \prod_{\alpha=1}^{2} \prod_{\beta=3}^{4} \cos r_\alpha \omega_j \cos r_\beta \omega_k \\
&\times \sum_{s_1 \in S_{r_1},\ldots,s_4 \in S_{r_4}} \sum_{q=-\infty}^{\infty} a_{q+r_1} a_q \prod_{l'=2}^{4} a_{q+r_{l'}+s_{l'}-s_1} a_{q+s_{l'}-s_1},
\end{aligned}$$

where $a_j = 0$ for $j < 0$, $S_r = \{1, \ldots, n - r\}$ for $r \geq 0$ and $S_r = \{1 - r, \ldots, n\}$ for $r \leq 0$. Denoting $u_i = s_{i+1} - s_1$ for $i = 1, 2, 3$, using

$$\gamma_\theta(u) = \mathrm{Cov}\,(X_t, X_{t+u}) = \mathrm{Cov}(\sum_{j=0}^{\infty} a_j \epsilon_{t-j}, \sum_{k=0}^{\infty} a_k \epsilon_{t+u-k}) = \sum_{j=0}^{\infty} a_j a_{j+u} \sigma_\epsilon^2,$$

and Cesaro averages, we have

$$\begin{aligned}
&\frac{1}{n^4} \sum_{s_1,\ldots,s_4} \sum_{q=-\infty}^{\infty} a_{q+r_1} a_q \prod_{l'=2}^{4} a_{q+r_{l'}+s_{l'}-s_1} a_{q+s_{l'}-s_1} \\
= &\frac{1}{n^3} \left\{ \sum_q (1 - \frac{|r_1|}{n}) a_{q+r_1} a_q \right\} \left\{ \prod_{l=1}^{3} \sum_{u_l} (1 - \frac{|r_{l+1}|}{n}) a_{q+u_l+r_{l+1}} a_{q+u_l} \right\} \\
= &\frac{\prod_{l=1}^{4} \gamma_\theta(r_l)}{n^3 \sigma_\epsilon^8} + O(n^{-4}).
\end{aligned}$$

Therefore, by $f(\omega, \theta) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \gamma_\theta(u) \cos(u\omega)$, we have

$$\kappa(j, -j, j, -j, k, -k, k, -k) = \frac{\kappa_{\epsilon,8}}{n^3 \sigma_\epsilon^8} f(\omega_j, \theta)^2 f(\omega_k, \theta)^2 + O(n^{-4}).$$

Applying a similar procedure to $\kappa(j,-j,k,-k)$ and $\kappa(j,-j,j,-j,k,-k)$ in $p=2$ and $p=3$ in (5.9), it follows that

$$\kappa(j,-j,k,-k) = \frac{\kappa_{\epsilon,4}}{n\sigma_\epsilon^4} f(\omega_j,\theta) f(\omega_k,\theta) + O(n^{-3}), \tag{5.10}$$

and

$$\kappa(j,-j,j,-j,k,-k) = \frac{\kappa_{\epsilon,6}}{n^2\sigma_\epsilon^6} f(\omega_j,\theta)^2 f(\omega_k,\theta) + O(n^{-3}).$$

It should be noted that the order $O(n^{-1})$ of $\kappa(j,-j,k,-k)$ is larger than the orders $O(n^{-2})$ and $O(n^{-3})$ of $\kappa(j,-j,j,-j,k,-k)$ and $\kappa(j,-j,j,-j,k,-k,k,-k)$, respectively. By this property, the partition problem of fourth order periodogram cumulants is simplified. In this case, we only need to consider the partition product of the second-order cumulant of the DFT $\kappa_2(J_j)$ and at most one fourth-order cumulant of the DFT at different frequencies, i.e. $\kappa(j,-j,k,-k)$. Also for partitions $\boldsymbol{\nu}'=\boldsymbol{\nu}\ \backslash(5.9)$,

$$\sum_{j\neq k} \sum_{\boldsymbol{\nu}':p=1}^{8} \prod_{j=1}^{p} \mathrm{cum}\{J_{k_j}; k_j \in \nu_j\} = O(n).$$

Hence, using (5.10), the fact that $\kappa(j,-j) = f(\omega_j,\theta) + O(n^{-1})$, and the decomposition formula in Brillinger (1981), it follows that

$$\begin{aligned} \mathrm{cum}(I_n(\omega_j), I_n(\omega_j), I_n(\omega_k), I_n(\omega_k)) &= 4\kappa(j,-j)\kappa(k,-k)\kappa(j,-j,k,-k) + O(n^{-2}) \\ &= \frac{4\kappa_{\epsilon,4}}{n\sigma_\epsilon^4} f(\omega_j,\theta)^2 f(\omega_k,\theta)^2 + O(n^{-2}). \quad (5.11) \end{aligned}$$

Substituting (5.11) into (5.8), we have

$$\begin{aligned} & \frac{1}{n^2}\sum_{j\neq k} \mathrm{cum}(m_j,m_j,m_k,m_k) \quad (5.12) \\ = & \frac{1}{n^3}\frac{4\kappa_{\epsilon,4}}{\sigma_\epsilon^4}\sum_{j\neq k} g(\omega_j,\theta)^2 g(\omega_k,\theta)^2 + O(n^{-2}) \\ = & \frac{1}{n}\frac{\kappa_{\epsilon,4}}{\pi^2\sigma_\epsilon^4}\iint_{\Pi^2} g(\omega,\theta)^2 g(\lambda,\theta)^2\, d\omega\, d\lambda + O(n^{-2}). \quad (5.13) \end{aligned}$$

For the second term in (5.7), applying (5.10) to

$$\begin{aligned} & \mathrm{cum}(m_j, m_k) \\ = & \frac{g(\omega_j,\theta)g(\omega_k,\theta)}{f(\omega_j,\theta)f(\omega_k,\theta)}\mathrm{cum}(I_n(\omega_j), I_n(\omega_k)) \\ = & \frac{g(\omega_j,\theta)g(\omega_k,\theta)}{f(\omega_j,\theta)f(\omega_k,\theta)}[\kappa(j,-k)\kappa(-j,k)+\kappa(j,k)\kappa(-j,-k)+\kappa(j,-j,k,-k)] \\ = & \frac{g(\omega_j,\theta)g(\omega_k,\theta)}{f(\omega_j,\theta)f(\omega_k,\theta)}\frac{\kappa_{\epsilon,4}}{n\sigma_\epsilon^4}+O(n^{-2}), \end{aligned}$$

we have

$$\frac{1}{n^2}\sum_{j\neq k}\mathrm{cum}^2(m_j, m_k) = \frac{1}{n^2}\frac{\kappa_{\epsilon,4}^2}{\sigma_\epsilon^8}\iint_{\Pi^2} g(\omega,\theta)^2 g(\lambda,\theta)^2\, d\omega\, d\lambda + O(n^{-3}). \quad (5.14)$$

Given that $\mathrm{cum}(m_j) = O(n^{-1})$ and $\mathrm{cum}(m_j, m_k, m_k) = O(n^{-1})$, we have

$$\frac{1}{n^2}\sum_{j\neq k}\mathrm{cum}(m_j)\mathrm{cum}(m_j, m_k, m_k) = O(n^{-2}). \quad (5.15)$$

Similarly, because $\mathrm{cum}(m_j, m_k) = O(n^{-1})$, we have

$$\frac{1}{n^2}\sum_{j\neq k}\mathrm{cum}(m_j)\mathrm{cum}(m_k)\mathrm{cum}(m_j, m_k) = O(n^{-3}). \quad (5.16)$$

Combining formulas (5.12), (5.14), (5.15) and (5.16), it is concluded that for $a = 2$ and $b = 2$,

$$\begin{aligned} & \frac{1}{n^2}\sum_{j\neq k}\mathrm{cum}(m_{1_1},\ldots,m_{s_1})\cdots\mathrm{cum}(m_{1_p},\ldots,m_{s_p}) \\ = & C\frac{\kappa_{\epsilon,4}}{n\sigma_\epsilon^4}\iint_{\Pi^2} g(\omega,\theta)^2 g(\lambda,\theta)^2\, d\omega,\, d\lambda + O(n^{-2}). \end{aligned}$$

□

Proof of Lemma 5.1:

First, we show an important formula (5.18) for the cumulants of the DFT at conjugate frequencies, which are widely used in the following proof. Using the Kolmogorov's formula $(2\pi)^{-1}\int_\Pi \log f(\omega,\theta)\, d\omega = \log(\sigma_\epsilon^2/(2\pi))$, and the fact that $\log f(\omega, \theta)$ is twice differentiable for $\theta \in \Theta$, the true value θ_0 with known variance σ_ϵ^2 is determined by the equation

$$\int_\Pi \frac{\partial}{\partial \theta} \log f(\omega, \theta_0)\, d\omega = 0. \tag{5.17}$$

By the same argument of $\kappa(j, -j, j, -j, k, -k, k, -k)$ in Lemma 5.2, equation (5.17) gives a set of cumulants' equations of the DFT at the conjugate frequencies,

$$\left| \kappa_p(J_j) - \frac{\kappa_{\epsilon,p}}{n^{p/2-1}\sigma_\epsilon^p} f(\omega_j, \theta_0)^{p/2} \right| = o(1), \quad \text{for} \quad p = 2, 4, 6, 8. \tag{5.18}$$

In particular, $\kappa_2(J_j) = f(\omega_j, \theta_0) + O(n^{-1})$. We only show the proof for λ_4 because cases for λ_2 and λ_3 can be derived similarly. For λ_4, we argue that

$$\lambda_4 = \frac{1}{n}\sum_{j=1}^n \mathrm{E}(m_j^4) = \frac{9}{2\pi}\int_\Pi g(\omega, \theta)^4\, d\omega + \frac{1}{n}\frac{(c_{2,2,4}+6)\kappa_{\epsilon,4}}{2\pi\sigma_\epsilon^4}\int_\Pi g(\omega, \theta)^4\, d\omega + O(n^{-2}), \tag{5.19}$$

where $c_{2,2,4}$ denotes the number of combinations with respect to $\kappa_2(J_j)^2\kappa_4(J_j)$. To see this,

$$\begin{aligned} \lambda_4 &= \frac{1}{n}\sum_{j=1}^n [\mathrm{cum}(m_j, m_j, m_j, m_j) + 3\mathrm{cum}^2(m_j, m_j) + 4\mathrm{cum}(m_j)\mathrm{cum}(m_j, m_j, m_j) \\ &\quad +6\mathrm{cum}^2(m_j)\mathrm{cum}(m_j, m_j) + \mathrm{cum}^4(m_j)]. \end{aligned} \tag{5.20}$$

Because $\mathrm{cum}(m_j) = O(n^{-1})$ and $\mathrm{cum}(m_j, m_j, m_j) = O(1)$, then

$$n^{-1}\sum_{j=1}^n \mathrm{cum}(m_j, m_j, m_j)\mathrm{cum}(m_j)$$

is of order $O(n^{-1})$, which is smaller than $O(1)$ of $\frac{1}{n}\sum_{j=1}^n \mathrm{cum}(m_j, m_j, m_j, m_j)$ as shown below. Similarly, the other two terms $\frac{1}{n}\sum_{j=1}^n \mathrm{cum}^2(m_j)\mathrm{cum}(m_j, m_j)$ and $\frac{1}{n}\sum_{j=1}^n \mathrm{cum}^2(m_j)\mathrm{cum}(m_j, m_j)$ have even smaller orders, so we do not consider them in the expression of λ_4. In the calculation of $\mathrm{Var}(\Delta_r)$, terms with an order smaller than $O(n^{-2})$ are not explicitly written. For the first term in

(5.20), as

$$\begin{aligned} & \frac{1}{n}\sum_{j=1}^{n}\mathrm{cum}(m_j, m_j, m_j, m_j) \\ = & \frac{1}{n}\sum_{j=1}^{n}\frac{g(\omega_j,\theta)^4}{f(\omega_j,\theta)^4}\mathrm{cum}\{I_n(\omega_j), I_n(\omega_j), I_n(\omega_j), I_n(\omega_j)\}, \end{aligned} \tag{5.21}$$

and

$$\begin{aligned} & \mathrm{cum}\{I_n(\omega_j), I_n(\omega_j), I_n(\omega_j), I_n(\omega_j)\} \\ = & \kappa_8(J_j) + c_{4,4}\kappa_4(J_j)^2 + c_{2,6}\kappa_2(J_j)\kappa_6(J_j) \\ & + c_{2,2,4}\kappa_2(J_j)^2\kappa_4(J_j) + 6\kappa_2(J_j)^4, \end{aligned} \tag{5.22}$$

where $c_{4,4}$ and $c_{2,6}$ are number of combinations corresponding to the $\kappa_4(J_j)^2$, $\kappa_2(J_j)\kappa_6(J_j)$, together with (5.18), (5.22) becomes

$$\frac{\kappa_{\epsilon,8}}{n^3\sigma_\epsilon^8}f(\omega_j,\theta)^4+\frac{c_{4,4}\kappa_{\epsilon,4}^2}{n^2\sigma_\epsilon^8}f(\omega_j,\theta)^4+\frac{c_{2,6}\kappa_{\epsilon,6}}{n^2\sigma_\epsilon^6}f(\omega_j,\theta)^4+\frac{c_{2,2,4}\kappa_{\epsilon,4}}{n\sigma_\epsilon^4}f(\omega_j,\theta)^4+6f(\omega_j,\theta)^4+o(1).$$

Finally, substituting the above formula for (5.21), we have

$$\frac{1}{n}\sum_{j=1}^{n}\mathrm{cum}(m_j, m_j, m_j, m_j) = \frac{3}{\pi}\int_\Pi g(\omega,\theta)^4\,d\omega+\frac{1}{n}\frac{c_{2,2,4}\kappa_{\epsilon,4}}{2\pi\sigma_\epsilon^4}\int_\Pi g(\omega,\theta)^4\,d\omega+O(n^{-2}). \tag{5.23}$$

For the other term involving $\mathrm{cum}(m_j, m_j)$ in (5.20), we have

$$\frac{3}{n}\sum_{j=1}^{n}\mathrm{cum}^2(m_j, m_j) = \frac{3}{n}\sum_{j=1}^{n}\frac{g(\omega_j,\theta)^4}{f(\omega_j,\theta)^4}\{\kappa_2^4(J_j)+2\kappa_4(J_j)\kappa_2^2(J_j)+R_n(j)\},$$

where $R_n(j)$ denotes the residual terms with non-conjugate frequencies. Hence, because $\sum_j R_n(j) = o(1)$, it follows that

$$\frac{3}{n}\sum_{j=1}^{n}\mathrm{cum}^2(m_j, m_j) = \frac{3}{2\pi}\int_\Pi g(\omega,\theta)^4\,d\omega + \frac{1}{n}\frac{3\kappa_{\epsilon,4}}{\pi\sigma_\epsilon^4}\int_\Pi g(\omega,\theta)^4\,d\omega + o(n^{-1}). \tag{5.24}$$

Combining (5.23) and (5.24), equation (5.19) follows.

Similar to the argument of λ_4, we conclude that

$$\lambda_2 = \frac{1}{2\pi}\int_{\Pi} g(\omega,\theta)^2\, d\omega + O(n^{-1}), \tag{5.25}$$

$$\lambda_3 = \frac{1}{\pi}\int_{\Pi} g(\omega,\theta)^3\, d\omega + O(n^{-1}). \tag{5.26}$$

Next, we consider the centered higher-order moments. For Δ_2,

$$\mathrm{Var}(\Delta_2) = \frac{1}{n^2}\sum_j\sum_k [\mathrm{cum}(m_j,m_j,m_k,m_k) + 2\mathrm{cum}^2(m_j,m_k)] + O(n^{-2}).$$

Note that by applying (5.12) in Lemma 5.2 and (5.23) to the first term, we have

$$\begin{aligned}
&\frac{1}{n^2}\sum_j\sum_k \mathrm{cum}(m_j,m_j,m_k,m_k) \\
&= \frac{1}{n^2}\sum_j \mathrm{cum}(m_j,m_j,m_j,m_j) + \frac{1}{n^2}\sum_{j\neq k}\mathrm{cum}(m_j,m_j,m_k,m_k) \\
&= \frac{3}{n\pi}\int_{\Pi} g(\omega,\theta)^4\, d\omega + \frac{1}{n}\frac{4\kappa_\epsilon, 4}{(2\pi)^2\sigma_\epsilon^4}\iint_{\Pi^2} g(\omega,\theta)^2 g(\lambda,\theta)^2\, d\omega\, d\lambda \\
&\quad + O(n^{-2}),
\end{aligned} \tag{5.27}$$

where the second equality is derived from the results of (5.23) and (5.12).

It is also true that

$$\frac{2}{n^2}\sum_j\sum_k \mathrm{cum}^2(m_j,m_k) = \frac{1}{n\pi}\int_{\Pi} g(\omega,\theta)^4\, d\omega + O(n^{-2}). \tag{5.28}$$

Combing (5.27) and (5.28), with (5.25) and (5.19), it is concluded that

$$\begin{aligned}
\mathrm{Var}(\Delta_2) &= \frac{1}{n}\frac{4}{\pi}\int_{\Pi} g(\omega,\theta)^4\, d\omega + \frac{1}{n}\frac{4\kappa_{\epsilon,4}}{\sigma_\epsilon^4}\frac{1}{(2\pi)^2}\left(\int_{\Pi} g(\omega,\theta)^2\, d\omega\right)^2 + O(n^{-2}) \\
&= \frac{1}{n}\frac{8}{9}\lambda_4 + \frac{1}{n}\frac{4\kappa_{\epsilon,4}}{\sigma_\epsilon^4}\lambda_2^2 + O(n^{-2}).
\end{aligned}$$

Next, we prove that $\Delta_3 = O_p(n^{-1/2})$.

$$\mathrm{Var}(\Delta_3) = \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{cum}(m_j^3, m_j^3) + \frac{1}{n^2}\sum_{j\neq k}\mathrm{cum}(m_j^3, m_k^3). \tag{5.29}$$

For the first term of the score function at the same frequencies in (5.29), we have

$$\begin{aligned}\frac{1}{n^2}\sum_{j=1}^{n}\mathrm{cum}(m_j^3, m_j^3) &= \frac{1}{n^2}[\mathrm{cum}(\underbrace{m_j, \ldots, m_j}_{6}) \\ &+15\mathrm{cum}(m_j, m_j)\mathrm{cum}(m_j, m_j, m_j, m_j) \\ &+9\mathrm{cum}^2(m_j, m_j, m_j) + 15\mathrm{cum}^3(m_j, m_j)] + O(n^{-2}).\end{aligned}$$

Because

$$\begin{aligned}\frac{1}{n^2}\sum_{j}\mathrm{cum}(\underbrace{m_j, \ldots, m_j}_{6}) &= \frac{1}{n^2}\sum_{j} g(\omega_j, \theta)^6 f(\omega_j, \theta)^{-6}\mathrm{cum}(\underbrace{I_n(\omega_j), \ldots, I_n(\omega_j)}_{6}) \\ &= \frac{c_{2,2,2,2,2,2}}{2\pi n}\int_{\Pi} g(\omega, \theta)^6\, d\omega + O(n^{-2}),\end{aligned}$$

where $c_{2,2,2,2,2,2}$ denotes the number of combination corresponding to $\kappa_2(J_j)^6$. By tedious calculation, we have

$$\frac{15}{n^2}\sum_{j}\mathrm{cum}(m_j, m_j)\mathrm{cum}(m_j, m_j, m_j, m_j) = \frac{45}{\pi n}\int_{\Pi} g(\omega, \theta)^6\, d\omega + O(n^{-2}),$$

$$\frac{9}{n^2}\sum_{j}\mathrm{cum}^2(m_j, m_j, m_j) = \frac{18}{\pi n}\int_{\Pi} g(\omega, \theta)^6\, d\omega + O(n^{-2}),$$

and

$$\frac{15}{n^2}\sum_{j}\mathrm{cum}^3(m_j, m_j) = \frac{15}{2\pi n}\int_{\Pi} g(\omega, \theta)^6\, d\omega + O(n^{-2}).$$

Summing up these equations, we have

$$\frac{1}{n^2}\sum_{j=1}^{n}\mathrm{cum}(m_j^3, m_j^3) = \frac{c_{2,2,2,2,2,2} + 141}{2\pi n}\int_{\Pi} g(\omega, \theta)^6\, d\omega + O(n^{-2}).$$

For the second term of the score function at different frequencies in (5.29), we have

$$\begin{aligned}&\frac{1}{n^2}\sum_{j\neq k}\mathrm{cum}(m_j^3, m_k^3) \\ =\ &\frac{1}{n^2}\sum_{j\neq k}[\mathrm{cum}(m_j, m_j, m_j, m_k, m_k, m_k) \\ &+3\mathrm{cum}(m_j, m_j)\mathrm{cum}(m_j, m_k, m_k, m_k) + 3\mathrm{cum}(m_k, m_k)\mathrm{cum}(m_j, m_j, m_j, m_k) \\ &+9\mathrm{cum}(m_j, m_j)\mathrm{cum}(m_k, m_k)\mathrm{cum}(m_j, m_k) + 6\mathrm{cum}^3(m_j, m_k)]. \qquad (5.30)\end{aligned}$$

Given that

$$\begin{aligned}&\frac{1}{n^2}\sum_{j\neq k}\mathrm{cum}(m_j,m_j,m_j,m_k,m_k,m_k)\\ =&\ \frac{1}{n^3}\frac{c_{2,2,2,2,4}\kappa_{\epsilon,4}}{\sigma_\epsilon^4}\sum_{j\neq k}g(\omega_j,\theta)^3 g(\omega_k,\theta)^3+O(n^{-2}),\end{aligned}$$

$$\begin{aligned}&\frac{3}{n^2}\sum_{j\neq k}\mathrm{cum}(m_j,m_j)\mathrm{cum}(m_j,m_k,m_k,m_k)\\ =&\ \frac{1}{n^3}\frac{18\kappa_{\epsilon,4}}{\sigma_\epsilon^4}\sum_{j\neq k}g(\omega_j,\theta)^3 g(\omega_k,\theta)^3+O(n^{-2}),\end{aligned}$$

and

$$\begin{aligned}&\frac{9}{n^2}\sum_{j\neq k}\mathrm{cum}(m_j,m_j)\mathrm{cum}(m_k,m_k)\mathrm{cum}(m_j,m_k)\\ =&\ \frac{1}{n^3}\frac{9\kappa_{\epsilon,4}}{\sigma_\epsilon^4}\sum_{j\neq k}g(\omega_j,\theta)^3 g(\omega_k,\theta)^3+O(n^{-2}),\end{aligned}$$

substituting these equalities into (5.30), we have

$$\frac{1}{n^2}\sum_{j\neq k}\mathrm{cum}(m_j^3,m_k^3)=\frac{1}{n^3}\frac{(c_{2,2,2,2,4}+45)\kappa_{\epsilon,4}}{\sigma_\epsilon^4}\sum_{j\neq k}g(\omega_j,\theta)^3 g(\omega_k,\theta)^3+O(n^{-2}).$$

Next, we prove that

$$\Delta_4=\frac{1}{n}\sum_{j=1}^n m_j^4-\mathrm{E}\left(\frac{1}{n}\sum_{j=1}^n m_j^4\right)=O_p(n^{-1/2}).$$

Because

$$\mathrm{Var}(\Delta_4)=\frac{1}{n^2}\sum_j\mathrm{Var}(m_j^4)+\frac{1}{n^2}\sum_{j\neq k}\mathrm{Cov}(m_j^4,m_k^4),$$

we consider the two terms on the right side individually. For the first term, we have

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^n\mathrm{Var}\left(m_j^4\right)\ =&\ \frac{1}{n}\sum_{j=1}^n[\mathrm{cum}(\underbrace{m_j,\ldots,m_j}_{8})+28\mathrm{cum}(m_j,m_j)\mathrm{cum}(\underbrace{m_j,\ldots,m_j}_{6})\\ &+56\mathrm{cum}(m_j,m_j,m_j)\mathrm{cum}(\underbrace{m_j,\ldots,m_j}_{5})+42\mathrm{cum}^4(m_j,m_j,m_j,m_j)\\ &+90\mathrm{cum}^2(m_j,m_j)\mathrm{cum}(m_j,m_j,m_j,m_j)+96\mathrm{cum}^4(m_j,m_j)\\ &+280\mathrm{cum}(m_j,m_j)\mathrm{cum}^2(m_j,m_j,m_j)]+O(n^{-2}).\end{aligned}\qquad(5.31)$$

By Lemma 5.2, we have

$$\mathrm{cum}(\underbrace{m_j,\ldots,m_j}_{8}) = c_{\underbrace{2,\ldots,2}_{8}} g(\omega_j,\theta)^8 + O(n^{-1}),$$

$$28\mathrm{cum}(m_j,m_j)\mathrm{cum}(\underbrace{m_j,\ldots,m_j}_{6}) = 28c_{\underbrace{2,\ldots,2}_{6}} g(\omega_j,\theta)^8 + O(n^{-1}),$$

$$56\mathrm{cum}(m_j,m_j,m_j)\mathrm{cum}(\underbrace{m_j,\ldots,m_j}_{5}) = 112c_{\underbrace{2,\ldots,2}_{5}} g(\omega_j,\theta)^8 + O(n^{-1}),$$

$$42\mathrm{cum}^4(m_j,m_j,m_j,m_j) = 1512 g(\omega_j,\theta)^8 + O(n^{-1}),$$

$$90\mathrm{cum}^2(m_j,m_j)\mathrm{cum}(m_j,m_j,m_j,m_j) = 540 g(\omega_j,\theta)^8 + O(n^{-1}),$$

and finally

$$280\mathrm{cum}(m_j,m_j)\mathrm{cum}^2(m_j,m_j,m_j) = 1120 g(\omega_j,\theta)^8 + O(n^{-1}).$$

Substituting the above formulas into (5.31), it follows that

$$\begin{aligned}
&\frac{1}{n^2}\sum_{j=1}^{n}\mathrm{Var}(m_j^4) \\
= \; & \frac{1}{2\pi n}\{c_{\underbrace{2,\ldots,2}_{8}} + 28c_{\underbrace{2,\ldots,2}_{6}} + 112c_{\underbrace{2,\ldots,2}_{5}} + 3268\}\int_{\Pi} g(\omega,\theta)^8\, d\omega + O(n^{-2}).
\end{aligned}$$

With a similar but more tedious argument of cumulants, we have

$$\begin{aligned}
&\frac{1}{n^2}\sum_{j\neq k}\mathrm{Cov}(m_j^4,m_k^4) \\
= \; & \frac{1}{n^2}\sum_{j\neq k}[\mathrm{cum}(m_j,m_j,m_j,m_j,m_k,m_k,m_k,m_k) \\
&+6\mathrm{cum}(m_j,m_j)\mathrm{cum}(m_j,m_j,m_k,m_k,m_k,m_k) \\
&+6\mathrm{cum}(m_k,m_k)\mathrm{cum}(m_j,m_j,m_j,m_j,m_k,m_k) \\
&+4\mathrm{cum}(m_j,m_j,m_j)\mathrm{cum}(m_j,m_k,m_k,m_k,m_k) \\
&+4\mathrm{cum}(m_k,m_k,m_k)\mathrm{cum}(m_j,m_j,m_j,m_j,m_k) \\
&+36\mathrm{cum}(m_j,m_j)\mathrm{cum}(m_k,m_k)\mathrm{cum}(m_j,m_j,m_k,m_k) \\
&+24\mathrm{cum}(m_j,m_j)\mathrm{cum}(m_k,m_k,m_k)\mathrm{cum}(m_j,m_j,m_k) \\
&+24\mathrm{cum}(m_k,m_k)\mathrm{cum}(m_j,m_j,m_j)\mathrm{cum}(m_j,m_k,m_k)] + O(n^{-2}).
\end{aligned}$$

Because the respective terms in the above expansion satisfy

$$\frac{1}{n^2}\sum_{j\neq k}\text{cum}(m_j,m_j,m_j,m_j,m_k,m_k,m_k,m_k) = c_{\underbrace{2,\ldots,2}_{6},4}\frac{\kappa_{\epsilon,4}}{n\sigma_\epsilon^4}\frac{\lambda_4^2}{81}+O(n^{-2}),$$

$$\frac{6}{n^2}\sum_{j\neq k}\text{cum}(m_j,m_j)\text{cum}(m_j,m_j,m_k,m_k,m_k,m_k) = c_{2,2,2,2,4}\frac{\kappa_{\epsilon,4}}{n\sigma_\epsilon^4}\frac{2\lambda_4^2}{27}+O(n^{-2}),$$

$$\frac{4}{n^2}\sum_{j\neq k}\text{cum}(m_j,m_j,m_j)\text{cum}(m_j,m_k,m_k,m_k,m_k) = c_{2,2,2,4}\frac{\kappa_{\epsilon,4}}{n\sigma_\epsilon^4}\frac{8\lambda_4^2}{81}+O(n^{-2}),$$

$$\frac{36}{n^2}\sum_{j\neq k}\text{cum}(m_j,m_j)\text{cum}(m_k,m_k)\text{cum}(m_j,m_j,m_k,m_k) = \frac{16\lambda_4^2}{9}\frac{\kappa_{\epsilon,4}}{n\sigma_\epsilon^4}+O(n^{-2}),$$

it follows that

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{j\neq k}\text{cum}(m_j^4,m_k^4)-\frac{\tilde{k}_4\kappa_{\epsilon,4}\lambda_4^2}{81\sigma_\epsilon^4}\right| = O\left(n^{-1}\right),$$

where $\tilde{k}_4 = c_{2,2,2,2,2,2,4}+12c_{2,2,2,2,4}+16c_{2,2,2,4}+336$. Therefore, using Chebyshev's inequality, $\Delta_4 = O_p(n^{-1/2})$. □

Proof of Lemma 5.3:

Based on the formula of SR in (5.4), we obtain the first-order cumulant expansion

$$\text{cum}(R_1+R_2+R_3) = \frac{1}{n}\frac{1}{\lambda_2^{1/2}}\frac{1}{2\pi}\int_\Pi b_I(\omega)g(\omega,\theta)f(\omega,\theta)^{-1}\,d\omega-\frac{1}{n}\frac{2}{3}\frac{\lambda_3}{\lambda_2^{3/2}}+O(n^{-2}).$$

Note that

$$\begin{aligned}\text{cum}(R_1+R_2+R_3,R_1+R_2+R_3) &= \text{cum}(R_1,R_1)+2\text{cum}(R_1,R_2)+2\text{cum}(R_1,R_3)\\ &\quad+\text{cum}(R_2,R_2)+O(n^{-3}).\end{aligned} \tag{5.32}$$

Similar to the calculations of λ_r and Δ_r, we have

$$\text{cum}(R_1,R_1) = 1+O(n^{-3}),$$

$$\text{cum}(R_1,R_2) = \frac{1}{n^2}\frac{\lambda_3^2}{\lambda_2^3}-\frac{1}{n^2}\frac{9}{2}\frac{\lambda_4}{\lambda_2^2}-\frac{1}{n^2}\frac{\kappa_{\epsilon,4}}{\sigma_\epsilon^4}+O(n^{-3}),$$

$$\text{cum}(R_1,R_3) = \frac{1}{n^2}\frac{7}{12}\frac{\lambda_4}{\lambda_2^2}+\frac{1}{n^2}\frac{275}{72}\frac{\lambda_3^2}{\lambda_2^3}+\frac{1}{n^2}\frac{3}{8}\frac{\kappa_{\epsilon,4}}{\sigma_\epsilon^4}+O(n^{-3}),$$

and

$$\mathrm{cum}(R_2, R_2) = \frac{1}{n^2}\frac{2}{9}\frac{\lambda_3^2}{\lambda_2^3} + \frac{1}{n^2}\frac{1}{9}\frac{\lambda_2\lambda_4}{\lambda_3^2} + \frac{1}{n^2}\frac{1}{8}\frac{\kappa_{\epsilon,4}}{\sigma_\epsilon^4} + O(n^{-3}),$$

Substituting these equations into (5.32), we have

$$\mathrm{cum}(R_1 + R_2 + R_3, R_1 + R_2 + R_3) = \frac{1}{n} + \frac{k_{2,2}}{n^2} + O(n^{-3}),$$

where by the formulas (5.25), (5.26) and (5.19) for λ_r defined in Section 3.2, $r = 2, 3, 4$,

$$\begin{aligned}
k_{2,2} &= \frac{\lambda_3}{3\pi\lambda_2^3}\int_\Pi g(\omega,\theta)^3\,d\omega - \frac{1}{2\pi\lambda_2^2}\int_\Pi g(\omega,\theta)^4\,d\omega \\
&\quad + \frac{13}{2\pi^2\lambda_3^2}\iint_{\Pi^2} g(\omega,\theta)^2 g(\lambda,\theta)^4\,d\omega\,d\lambda \\
&\quad + \frac{1}{\pi^2\lambda_2^3}\left(\int_\Pi g(\omega,\theta)^3\,d\omega\right)^2 - \frac{17\lambda_3}{6\pi^2\lambda_2^4}\iint_{\Pi^2} g(\omega,\theta)^2 g(\lambda,\theta)^3\,d\omega\,d\lambda \\
&\quad + \left\{-\frac{\kappa_{\epsilon,4}}{2\pi^2\sigma_\epsilon^4\lambda_2^2} + \frac{37\lambda_3^2}{18\pi^2\lambda_2^5} - \frac{3\lambda_4}{8\pi^2\lambda_2^4}\right\}\left(\int_\Pi g(\omega,\theta)^2\,d\omega\right)^2 \\
&\quad + \frac{7\kappa_{\epsilon,4}}{32\pi^3\sigma_\epsilon^4\lambda_2^3}\left(\int_\Pi g(\omega,\theta)^2\,d\omega\right)^3, \\
&= 4\frac{\lambda_3^2}{\lambda_2^3} - \frac{29}{18}\frac{\lambda_4}{\lambda_2^2} + \frac{26}{9}\frac{\lambda_2\lambda_4}{\lambda_3^2} - \frac{\kappa_{\epsilon,4}}{\sigma_\epsilon^4}.
\end{aligned}$$

Using the same higher-order expansion formula of the signed root empirical log-likelihood ratio in Chapter 4, we have

$$\mathrm{cum}(SR, SR, SR) = \mathrm{cum}(R_1, R_1, R_1) + 3\mathrm{cum}(R_1, R_1, R_2) + O(n^{-3}). \quad (5.33)$$

For the two terms in (5.33),

$$\begin{aligned}
\mathrm{cum}(R_1, R_1, R_1) &= \frac{1}{n^2}\frac{\lambda_3}{\lambda_2^{3/2}} + O(n^{-3}), \\
3\mathrm{cum}(R_1, R_1, R_2) &= \frac{1}{n^2}\frac{\lambda_3}{\lambda_2^{7/2}}\mathrm{cum}(\overline{m}, \overline{m}, \overline{m}^2) - \frac{1}{n^2}\frac{3}{2}\frac{1}{\lambda_2^{5/2}}\mathrm{cum}(\overline{m}, \overline{m}, \overline{m}\Delta_2) \\
&= \frac{1}{n^2}\left\{\frac{2\lambda_3}{\lambda_2^{3/2}} - \frac{3\lambda_3}{\lambda_2^{3/2}}\right\} + O(n^{-3}) = -\frac{1}{n^2}\frac{\lambda_3}{\lambda_2^{3/2}} + O(n^{-3}).
\end{aligned}$$

It follows that

$$\text{cum}(SR, SR, SR) = \frac{1}{n^2}\frac{\lambda_3}{\lambda_2^{3/2}} - \frac{1}{n^2}\frac{\lambda_3}{\lambda_2^{3/2}} + O(n^{-3}) = O(n^{-3}).$$

Similarly,

$$\begin{aligned}\text{cum}(SR, SR, SR, SR) &= \text{cum}(R_1, R_1, R_1, R_1) + 4\text{cum}(R_1, R_1, R_1, R_2)\\ &\quad +4\text{cum}(R_1, R_1, R_1, R_3) + 6\text{cum}(R_1, R_1, R_2, R_2)\\ &\quad +O(n^{-4}), \end{aligned} \tag{5.34}$$

and for the four terms in (5.34), we have

$$\text{cum}(R_1, R_1, R_1, R_1) = \frac{1}{n^3}\frac{2}{3}\frac{\lambda_4}{\lambda_2^2} + \frac{1}{n^3}\frac{12\kappa_{\epsilon,4}}{\sigma_\epsilon^4} + O(n^{-4}),$$

$$\begin{aligned}\text{cum}(R_1, R_1, R_1, R_2) &= \frac{1}{3}\frac{\lambda_3}{\lambda_2^4}\text{cum}(\overline{m}, \overline{m}, \overline{m}, \overline{m}^2) - \frac{1}{2}\frac{1}{\lambda_2^3}\text{cum}(\overline{m}, \overline{m}, \overline{m}, \overline{m}\Delta_2)\\ &= \frac{1}{n^3}\frac{\lambda_3^2}{2\lambda_2^3} - \frac{1}{n^3}\frac{4}{3}\frac{\lambda_4}{\lambda_2^2} - \frac{1}{n^3}\frac{6\kappa_{\epsilon,4}}{\sigma_\epsilon^4} + O(n^{-4}),\end{aligned}$$

$$\begin{aligned}\text{cum}(R_1, R_1, R_1, R_3) &= \frac{3}{8}\frac{1}{\lambda_2^4}\text{cum}(\overline{m}, \overline{m}, \overline{m}, \overline{m}\Delta_2) + \frac{1}{3}\frac{1}{\lambda_2^4}\text{cum}(\overline{m}, \overline{m}, \overline{m}, \overline{m}^2\Delta_3)\\ &\quad -\frac{5}{6}\frac{\lambda_3}{\lambda_2^5}\text{cum}(\overline{m}, \overline{m}, \overline{m}, \overline{m}^2\Delta_2)\\ &\quad +\left\{\frac{4}{9}\frac{\lambda_3^2}{\lambda_2^6} - \frac{1}{4}\frac{\lambda_4}{\lambda_2^5}\right\}\text{cum}(\overline{m}, \overline{m}, \overline{m}, \overline{m}^3)\\ &= -\frac{1}{n^3}\frac{1}{12}\frac{\lambda_3^2}{\lambda_2^3} + \frac{1}{n^3}\frac{1}{2}\frac{\lambda_4}{\lambda_2^2} + O(n^{-4}),\end{aligned}$$

and

$$\begin{aligned}\text{cum}(R_1, R_1, R_2, R_2) &= \frac{1}{9}\frac{\lambda_3^2}{\lambda_2^6}\text{cum}(\overline{m}, \overline{m}, \overline{m}^2, \overline{m}^2) - \frac{1}{3}\frac{\lambda_3}{\lambda_2^5}\text{cum}(\overline{m}, \overline{m}, \overline{m}^2, \overline{m}\Delta_2)\\ &\quad +\frac{1}{4}\frac{1}{\lambda_2^4}\text{cum}(\overline{m}, \overline{m}, \overline{m}\Delta_2, \overline{m}\Delta_2)\\ &= -\frac{1}{n^3}\frac{5}{18}\frac{\lambda_3^2}{\lambda_2^3} + \frac{1}{n^3}\frac{4}{9}\frac{\lambda_4}{\lambda_2^2} + \frac{1}{n^3}\frac{2\kappa_{\epsilon,4}}{\sigma_\epsilon^4} + O(n^{-4}).\end{aligned}$$

Substituting the above expansion multiplied by their coefficients into (5.34), it follows that

$$\begin{aligned}\mathrm{cum}(SR, SR, SR, SR) &= \frac{1}{n^3}\left\{\frac{2}{3}\frac{\lambda_4}{\lambda_2^2} + \frac{12\kappa_{\epsilon,4}}{\sigma_\epsilon^4} - \frac{16}{3}\frac{\lambda_4}{\lambda_2^2} - \frac{24\kappa_{\epsilon,4}}{\sigma_\epsilon^4} + 2\frac{\lambda_3^2}{\lambda_2^3} - \frac{1}{3}\frac{\lambda_3^2}{\lambda_2^3}\right. \\ &\left. +2\frac{\lambda_4}{\lambda_2^2} - \frac{5}{3}\frac{\lambda_3^2}{\lambda_2^3} + \frac{8}{3}\frac{\lambda_4}{\lambda_2^2} + \frac{12\kappa_{\epsilon,4}}{\sigma_\epsilon^4}\right\} + O(n^{-4}) = O(n^{-4}).\end{aligned}$$

□

Chapter 6

Conclusions and Further Research

6.1 Conclusions

Empirical likelihood (EL) is a nonparametric likelihood method, with a key property of "self-studentization", meaning that it automatically constructs confidence regions based on the asymptotic chi-squared limiting distribution without assuming any joint distribution of the data. EL has been shown to have many advantages over other methods (see, for example, Owen (2001) and Kitamura (2006)). Among these benefits, Bartlett correction is one attractive property, as it allows the construction of the confidence region with a smaller coverage error.

In this thesis, we discuss the Bartlett correction of EL for time series with Whittle-type periodogram-based score functions. By establishing the valid Edgeworth expansion of the signed root empirical log-likelihood ratio statistic with an irregular form, we proved that the coverage error of periodogram-based EL for Gaussian LMTS can be reduced from order $O(\log n^6/n)$ to $O(\log n^3/n)$, instead of from $O(n^{-1})$ to $O(n^{-2})$ in the i.i.d. setting using the Bartlett correction technique.

We also extend Bartlett correction for EL to non-Gaussian time series data. By tedious calculation, we find that the non-negligible higher-order cumulants can be removed in the higher-order cumulants of the signed root empirical log-likelihood ratio statistic for non-Gaussian short-memory time series (SMTS). Therefore, the valid Edgeworth expansion can also be established as a power series of $n^{-1/2}$. Based on the Edgeworth expansion, we prove the Bartlett correctability of EL for non-Gaussian SMTS. The coverage error is reduced from order $O(n^{-1})$ to order $O(n^{-2})$.

6.2 Further Research

These results are obtained only for one-dimensional data with scalar parameters. Future work could extend the Bartlett corretability of EL to inferences made from multivariate observations with multivariate parameters. However, this work requires the formidable calculation of asymptotic expansion involving extremely complex tensors.

In our work, we only consider the Bartlett correction for non-Gaussian time series with finite moment conditions. However, as demonstrated by Fama (1965) and Mandelbrot (1963), asset returns in financial markets always exhibit heavy-tailed phenomena. In this case, the moment condition fails and we may need to adjust the log-EL ratio to admit a proper "self-studentization". It may be possible that the log-EL ratio function does not have a chi-squared limit, as for unstable AR(p) models in Chuang and Chan (2002). Furthermore, it is of great interest to develop the EL method and Bartlett correction for heavy-tailed data.

Another extension of EL method is to consider other dependence structures, such as spatial dependence. One main theorectical difference between spatial and time series data is that the former may be irregularly spaced and they may need to consider various sampling schemes, such as stochastic spatial locations sampling and infill sampling. It is still an open question whether the EL method admits "self-studentization" under a different sampling scheme.

Bibliography

[1] Anderson, T.W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.

[2] Attfield, C.L.F. (1991). A Bartlett adjustment to the likelihood ratio test for homoskedasticity in the linear model. *Economics Letters* **37,** 119–123.

[3] Attfield, C.L.F. (1995). A Bartlett adjustment to the likelihood ratio test for a system of equations. *Journal of Econometrics* **66,** 207–223.

[4] Bartlett, M.S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society A* **160,** 268–282.

[5] Barndorff-Nielsen, O.E. and Hall, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika* **75,** 374–378.

[6] Beale, E.M.L. (1960). Confidence regions in non-linear estimation. *Journal of the Royal Statistical Society B* **26,** 41–88.

[7] Bhattacharya, R.N. and Ghosh, J.K. (1974). On the validity of the formal Edgeworth expansion. *Annals of Statistics* **6,** 434–451.

[8] Brillinger, D.R. (1981). *Time Series: Data Analysis and Theory.* Holden-Day, San Francisco.

[9] Chambers, J.M. (1967). On methods of asymptotic approximation for multivariate distributions. *Biometrika* **54,** 367–383.

[10] Chan, N.H. and Ling, S. (2006). Empirical likelihood for GARCH models. *Econometric Theory* **22,** 402–428.

[11] Chan, N.H. and Liu, L. (2010). Bartlett correctability of empirical likelihood in time series. *Journal of the Japanese Statistical Society* **40,** 1–5.

[12] Chen, X.S. and Cui, H. (2007). On the second order properties of empirical likelihood with moment restrictions. *Journal of Econometrics* **141,** 492–516.

[13] Chen, X.S. and Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Annals of Statistics* **21,** 1166–1181.

[14] Chen, X.S., Peng, L. and Qin, Y. (2009). Empirical likelihood methods for high dimension. *Biometrika* **96,** 711–722.

[15] Chuang, C.S. and Chan, N.H. (2002). Empirical likelihood for autoregressive models with applications to unstable time series. *Statistica Sinica* **12,** 387–407.

[16] Cordeiro, G.M. (1983). Improved likelihood ratio statistics for generalized linear models. *Biometrika* **74,** 265–274.

[17] Cordeiro, G.M. (1987). On the correction to the likelihood ratio statistics. *Journal of the Royal Statistical Society B* **45,** 404–413.

[18] Cordeiro, G.M. and Ferrari, S.L.P. (1991). A modified score test statistic having chi-squared distribution to order n^{-1}. *Biometrika* **78,** 573–582.

[19] Cox, D.R. (1988). Some aspects of conditional and asymptotic inference: a review. *Sankhyā A* **50,** 314–337.

[20] DiCiccio, T.J. (1984). On parameter transformations and interval estimation. *Biometrika* **71,** 477–485.

[21] DiCiccio, T.J., Hall, P. and Romano, J.P. (1991). Empirical likelihood is Bartlett correctable. *Annals of Statistics* **19,** 1053–1061.

[22] DiCiccio, T.J. and Romano, J.P. (1989). On adjustments based on the signed root of the empirical likelihood ratio statistics. *Biometrika* **76,** 447–456.

[23] DiCiccio, T.J. and Stern, S.E. (1993). On Bartlett adjustments for approximate Bayesian inference. *Biometrika* **80,** 731–740.

[24] DiCiccio, T.J. and Stern, S.E. (1994). Frequentist and Bayesian Bartlett correction to test statistics based on adjusted profile likelihood. *Journal of the Royal Statistical Society B* **56,** 397–408.

[25] Durbin, J. (1980). Approximations for densities of sufficient estimators. *Biometrika* **67**, 311–333.

[26] Fama, E.F. (1965). The behavior of stock-market prices. *Journal of Business* **38,** 34–105.

[27] Frydenberg and Jensen (1989). Is the "improved likelihood statistic" really improved in the discrete case? *Biometrika* **76**, 655–661.

[28] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer, New York.

[29] Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *Internat. Statist. Rev.* **58,** 109–127.

[30] Hayakawa, T. (1977). The likelihood ratio criterion and the asymptotic expansion of its distribution. *Annals of the Institute of Statistical Mathematics A* **29,** 359–378.

[31] Hjort, N.L., McKeague, I.W., Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *Annals of Statistics* **37,** 1079–1111.

[32] Hosoya, Y. (1974). Estimation problems on stationary time series models. Ph.D. dissertation, Dept. Statistics, Yale Univ.

[33] Hosoya Y. (1997). A limit theory for long-range dependence and statistical inference on related models. *Annals of Statistics* **25,** 105–137.

[34] Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Annals of Statistics* **25,** 2084–2102.

[35] Kitamura, Y. (2001). Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica* **69,** 1661–1672.

[36] Kreiss, J.-P. and Paparoditis, E. (2011). Bootstrap for dependent data: a review (with discussion). *Journal of the Korean Statistical Society* **40,** 357–395.

[37] Künsch, H.R. (1989). The jackknife and bootstrap for general stationary observations. *Annals of Statistics* **17,** 1217–1261.

[38] Lawley, D.N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **71,** 233–244.

[39] Mandelbrot, B. (1963) New methods in statistical economies. *Journal of Political Economy* **71,** 421–440.

[40] McCullagh, P. and Cox, D.R. (1986). Invariants and the likelihood ratio statistic. *Annals of Statistics* **14,** 1419–1430.

[41] Monti, A.C. (1997). Empirical likelihood confidence regions in time series models. *Biometrika* **84,** 395–405.

[42] Nordman, D.J. (2009). Tapered empirical likelihood for time series data in time- and frequency-domains. *Biometrika* **96,** 119–132.

[43] Nordman, D.J. and Lahiri, S.N. (2006). A frequency domain empirical likelihood for short- and long-range dependence. *Annals of Statistics* **34,** 3019–3050.

[44] Nordman, D.J. and Lahiri, S.N. (2013). A review of empirical likelihood methods for time series. *Journal of Statistical Planning and Inference* In press.

[45] Ogata, H. (2005). Empirical likelihood approach for non-Gaussian stationary processes. *Scientiae Mathematicae Japonicae* **62,** 429–438.

[46] Owen, A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75,** 237–249.

[47] Owen, A.B. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics* **18,** 90–120.

[48] Owen, A.B. (2001). *Empirical likelihood.* Chapman & Hall, New York.

[49] Paparoditis, E. and Politis, D.N. (2001). Tapered block bootstrap. *Biometrika.* **88,** 1105–1119.

[50] Phillips, P.C.B. and Park, J.Y. (1988). On the formulation of Wald tests of nonlinear restrictions. *Econometrica.* **56,** 1065–1083.

[51] Priestley, M.B. (1981). *Spectral Analysis and Time Series.* Academic Press.

[52] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating functions. *Annals of Statistics* **22,** 300–325.

[53] Qin, J. and Zhang, B. (2007). Empirical-likelihood-based inference in missing response problems and its application in observational studies. *Journal of the Royal Statistical Society B* **69,** 101–122.

[54] Tierney, L., Kass, R. E. and Kadane, J. B. (1989). Approximate marginal densities of nonlinear functions. *Biometrika* **76,** 425–33.

[55] Thomas, D.R. and Grunkemeier, G.L. (1975). Confidence interval estimation of survival probabilities for censored data. *Journal of the American Statistical Association* **70,** 865–871.

[56] Welch, P.D. (1967). The use of the Fast Fourier Transform for estimation of spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **15,** 423–434.

[57] Whittle, P. (1953). Estimation and information in stationary time series. *Archiv. Math.* **2,** 423–434.

[58] Wilks, S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics* **9,** 60–62.

[59] Yau, C.Y. (2010). Empirical likelihood in long-memory time series models. *Journal of Time Series Analysis* **33,** 269–275.

[60] Zhang, B. (1996). On the accuracy of empirical likelihood confidence intervals for M-Functionals. *Journal of Nonparametric Statistics* **6,** 311–321.

[61] Zygmund, A. (1977). *Trigonometric Series.* Cambridge Univ. Press.