Frequency domain empirical likelihood method for stochastic processes with infinite variance and its application to discriminant analysis

無限分散確率過程モデルに対する 経験尤度法による周波数領域推定法の構築 及び判別解析への応用

December, 2014

Fumiya Akashi
明石　郁哉

Frequency domain empirical likelihood method for stochastic processes with infinite variance and its application to discriminant analysis

無限分散確率過程モデルに対する 経験尤度法による周波数領域推定法の構築 及び判別解析への応用

December, 2014

Waseda University

Graduate School of Fundamental Science and Engineering
Department of Pure and Applied Mathematics
Research on Finance, Time Series Analysis and Mathematical Statistics

Fumiya Akashi
明石　郁哉

Contents

1 Introduction **1**

2 Discriminant analysis of second-order stationary processes **4**
2.1 Introduction 4
2.2 Fundamental settings 5
2.3 Main results 12
2.4 Non-Gaussian robustness of classification statistics 13
2.5 Numerical examples 18
2.5.1 Example 1: AR(1) model 18
2.5.2 Example 2: ARMA(1,1) model 19

3 Empirical likelihood method for symmetric α-stale processes **25**
3.1 Introduction 25
3.2 Fundamental settings 26
3.3 Asymptotic theory for fundamental quantities 28
3.4 Main results 30
3.5 Numerical examples 32
3.5.1 Example 3: Stable ARMA(1,1) model 33
3.5.2 Example 4: Stable MA(100) model 34

4 Discriminant analysis of symmetric α-stale processes **44**
4.1 Introduction 44
4.2 Fundamental settings 44
4.3 Main results 46
4.4 Numerical examples 49
4.4.1 Example 5: Stable ARMA(1,1) model 49
4.4.2 Example 6: Stable MA(100) model 51

5 Generalized method of moments for symmetric α-stale processes **58**
5.1 Introduction 58
5.2 Fundamental settings 58
5.3 Main results 59
5.4 Numerical examples 62
5.4.1 Example 7: Stable AR(1) model 62
5.4.2 Example 8: Stable MA(1) model 63

6 Proofs **71**

6.1 Proofs of Chapter 2 . . . 71

6.2 Proofs of Chapter 3 . . . 75

6.3 Proofs of Chapter 4 . . . 86

6.4 Proofs of Chapter 5 . . . 89

Acknowledgments **93**

Bibliography **94**

List of papers **100**

1 Introduction

This dissertation discusses the nonparametric inference for time series models with infinite variance based on the frequency domain empirical likelihood method. In particular, we focus on the symmetric α-stable linear process, which is a linear process generated by independent and identically distributed (i.i.d.) symmetric α-stable random variables. Furthermore, we provide its application to the discriminant analysis and the generalized method of moments estimators for the stable process, and show the optimal properties of the empirical likelihood method.

Heavy-tailed data have been observed in variety of fields involving electrical engineering, hydrology, finance and physical systems in the last few decades, and there are a lot of papers discussing the estimation of tail index of the data (e.g., Drees, de Haan and Resnick [12], Mandelbrot [39], Resnick and Stărică [56]). When we analyze such time series data, the classical method of moments is not applicable since it is not natural to assume finite moments of underlying data generating distributions. It is also infeasible to work with the exact likelihood ratio of observations in practice, since the data usually have the dependence structure and the dimensionality of the data is often very large. To begin with, the true model is unknown in general. To overcome these hurdles, the empirical likelihood and frequency domain approach were proposed by many authors.

The empirical likelihood approach proposed by Owen [50] gives a convenient computational procedure of the approximation of the likelihood ratio without assuming that the data come from a known family of stochastic models. Owen [50] showed that the empirical likelihood ratio statistic converges in law to the chi-square distribution in i.i.d. and finite moment case. For dependent data, Monti [42] applied the empirical likelihood method to a stationary linear process. Subsequently, Ogata and Taniguchi [49] extended the method to multivariate second-order stationary processes, and constructed the method of nonparametric inference for important quantities of time series models even if the true model is unknown. On the other hand, there are a lot of papers discussing the large sample properties of a class of econometric estimators. The generalized method of moments (GMM) estimator is one of them, and is used when we focus on the estimation problem of over-identified unknown parameters of the model. Especially, Kakizawa [26] suggested the frequency domain over-identified model, and showed consistency and asymptotic normality of the frequency domain GMM estimator for second-order stationary processes.

Thus, the empirical likelihood and its extended versions are applied to various models, but most of them considered finite variance models. Meanwhile, the stable process is widely used to model real data more suitably than the second-order stationary processes. One of the biggest hurdle when we deal with the stable process is that the stable distribution does not have the finite variance and the probability density function in closed form in general. In such situation, the frequency domain approach is still useful. A series of works for the stable processes have been done by several authors. For example, Mikosch et al. [40] constructed the Whittle estimator for the stable autoregressive moving average (ARMA) models, and showed the consistency and limit distribution of the estimator. In addition, Klüppelberg and Mikosch [33, 34] studied the limit behavior of so-called self-normalized periodograms and the integrated version for the stable process. They elucidated the asymptotic properties of the statistics. In particular, it was shown that the limit distribution of the integrated self-normalized periodogram for the stable process is expressed as a sum of stable random variables. The results in Klüppelberg and Mikosch [34] are widely applicable to various problems such as classical testing and estimation problems.

Motivated the concepts above, in this dissertation we apply the frequency domain empirical likelihood method to the symmetric α-stable linear process, and show the optimal properties of the empirical likelihood via two applications: the discriminant analysis of time series models, and the frequency domain GMM.

The rest of this dissertation is organized as follows. Chapter 2 introduces fundamental settings and limit theorems of the empirical likelihood method for second-order stationary processes. We also construct a classification procedure of second-order stationary processes based on the empirical likelihood ratio statistic. It will be shown that the empirical likelihood ratio classification statistic has advantage in the sense of non-Gaussian robustness. In Chapter 3, we construct the nonparametric inference based on the empirical likelihood method for the stable process. The results for second-order stationary processes are nicely extended to the stable process in the chapter. Because of the infinite variance of the stable process, the extension is not straightforward and contains a lot of novel aspects. Chapter 4 provides the empirical likelihood ratio classification statistic for the stable process, and evaluates the delicate goodness of the classification statistic as well as the second-order stationary case. Through some simulation experiments, we observe that the empirical likelihood classification procedure has a potential for improving the goodness of the classification in practical situations. In Chapter 5, we apply the GMM for the stable process. GMM is one of the most famous tools in econometrics, and we extend the results in Kakizawa [26] to the stable process by the frequency domain approach. The limit distribution of the estimator is elucidated in the chapter, and we discuss the asymptotic optimality of the GMM estimators. In Chapter 6, we place rigorous proofs for theorems in this dissertation.

As for notations and symbols used in this dissertation, the set of all integers, non-negative integers and real numbers are denoted as $\mathbb{Z}$, $\mathbb{N}$ and $\mathbb{R}$, respectively. For any sequence of random vectors $\{A(t) : t \in \mathbb{Z}\}$, $A(t) \stackrel{\mathcal{P}}{\to} A$ and $A(t) \stackrel{\mathcal{L}}{\to} A$, respectively, denote the convergence to a random (or constant) vector A in probability and law. Especially, $p\text{-}\lim_{t\to\infty} A(t) = A$ implies $A(t) \stackrel{\mathcal{P}}{\to} A$ as $t \to \infty$. The transpose and conjugate transpose of matrix M are denoted by M' and M^*, and $\|M\|_E := \sqrt{\mathrm{tr}[M^*M]}$. For matrix-valued function $M(x) = (M(x)_{ij} : i, j = 1, \cdots, d)$, $\partial M(x)/\partial x$ denotes $(\partial M(x)_{ij}/\partial x : i, j = 1, \cdots, d)$. $\mathbb{I}\{\cdot\}$ denotes the indicator function of event $\{\cdot\}$. 0_i, $O_{j\times k}$ and $I_{l\times l}$ denote the i-dimensional zero vector, the $j \times k$ zero matrix and the $l \times l$ identity matrix, respectively. We denote the imaginary unit by i. δ_{ij} denotes the Kronecker delta.

2 Discriminant analysis of second-order stationary processes

2.1 Introduction

This chapter introduces the fundamental settings for the empirical likelihood method in time series analysis and constructs a classification procedure based on the empirical likelihood ratio statistic. Discriminant analysis is one of the most important topics in both i.i.d. case and time series analysis. Suppose that we observe a stretch $X^{(n)} = (X(1)', \cdots, X(n)')'$ (dn-vector) from a d-dimensional time series model, and we want to classify the observed stretch $X^{(n)}$ into one of two categories denoted by Π_1 and Π_2 with probability density functions (p.d.f.) $p_1(x)$ and $p_2(x)$, respectively. Usual discriminant procedure is to partition the dn-dimensional Euclidean space $\mathbb{R}^{dn}$ into two disjoint regions R_1 and R_2 such that if $X^{(n)}$ belongs to R_i ($i = 1$ and 2), then we assign $X^{(n)}$ into Π_i. It is known that the classification regions defined by

$$R_i = \left\{ x \in \mathbb{R}^{dn} : \frac{1}{n} \log \frac{p_i(x)}{p_j(x)} > 0 \right\}, \ (i, j) = (1, 2) \text{ and } (2, 1)$$

give the optimal classification regions in the sense that this classification procedure minimizes the quantity $\Pr(2|1) + \Pr(1|2)$.Here $\Pr(j|i)$ is the probability of misclassifying the observation from Π_i into Π_j. In our time series situation, however, the dimensionality n is often very large and the log-likelihood ratio has an intractable form. Therefore, some convenient computational procedures are important. There are various procedures in the frequency domain approach which are familiar with spectral analysis in stationary time series. In particular, it is known that the Kullback-Leibler information measure (Kullback and Leibler [35]) gives optimal time-frequency statistics for measuring a sort of distance between two time series models. As an example of the statistics based on the Kullback-Leibler information measure, Zhang and Taniguchi [68] adapted the Whittle likelihood ratio-type statistic as an approximation for the exact log-likelihood ratio. In the paper, they considered vector non-Gaussian stationary processes and showed that the Whittle likelihood ratio classification statistic is consistent classification criterion in the sense that the misclassification probabilities converge to 0 as $n \to \infty$. Zhang and Taniguchi [68] also discussed non-Gaussian robustness, which is an important concept when we deal with non-Gaussian processes. Many discriminant procedures and statistics have been introduced by several authors; see Kakizawa [24], Kakizawa, Shumway and Taniguchi [27], Zhang and Taniguchi [69], etc.

On the other hand, there has been a rich body of literature on novel idea of formulating versions of nonparametric likelihood in various settings of statistical inference in these few decades. One of them is the empirical likelihood method, which was introduced as a nonparametric method of inference based on a data-driven likelihood ratio function (Owen [50]) in i.i.d. case. Monti [42], Ogata and Taniguchi [49] applied the empirical likelihood method in frequency domain to second-order stationary processes. Especially, Ogata and Taniguchi [49] considered vector-valued linear processes, and showed that the limit distribution of the empirical likelihood ratio statistic is expressed as a quadratic form of standard normal random vectors under H: $\theta = \theta_0$, and showed that we can construct nonparametric inferences for $\theta_0 \in \mathbb{R}^p$, an important quantity of time series models. Various important indexes of time series models can be expressed as θ_0, and we shall give an example of θ_0 in Section 2.2. The advantage of this approach is that appropriate confidence regions of θ_0 are able to be constructed even if we do not know the true spectral density matrix of the process and the distribution of the innovation process.

By the motivation of the empirical likelihood approach, in this chapter we focus on the following two aims:

1. To apply the empirical likelihood method to classification problems.
2. To make a comparison between the misclassification probabilities by the empirical likelihood classification statistic and those by an existing method.

In Section 2.2, we state fundamental settings of the empirical likelihood method to time series models and construct the empirical likelihood-based discriminant procedure. Section 2.3 provides some of main results, which asserts fundamental goodness of the empirical likelihood ratio classification statistic. We also discuss delicate goodness of the statistic in line with Zhang and Taniguchi [68] in Section 2.4. Finally, the advantages of the empirical likelihood ratio classification statistic are shown in Section 2.5 via some numerical examples.

2.2 Fundamental settings

Suppose that $\{X(t) = (X_1(t), \cdots, X_d(t))' : t \in \mathbb{Z}\}$ is a non-Gaussian vector stationary process generated as

$$X(t) = \sum_{j=0}^{\infty} a(j)e(t-j), \tag{2.1}$$

where $\{a(t) : t \in \mathbb{N}\}$ is a sequence of $d \times d$ matrices with $a(0) = I_{d\times d}$ and $\{e(t) = (e_1(t), \cdots, e_d(t))' : t \in \mathbb{Z}\}$ is a d-dimensional fourth-order stationary white noise process; that is, $\{e(t) : t \in \mathbb{Z}\}$ satisfies $\mathrm{E}[e(t)] = 0_d$, $\mathrm{E}[e(t)e(s)'] = \delta_{ts}K$ with K, a $d \times d$ positive definite matrix, and

$$\mathrm{cum}\{e_k(t_1), e_l(t_2), e_u(t_3), e_v(t_4)\} = \begin{cases} \kappa^4_{kluv} & (t_1 = \cdots = t_4,\ |\kappa^4_{kluv}| < \infty) \\ 0 & (\text{otherwise}) \end{cases} .$$

These conditions are automatically satisfied when $\{e(t) : t \in \mathbb{Z}\}$ is a sequence of i.i.d. random vectors. If $\sum_{j=0}^{\infty} \mathrm{tr}\{a(j)Ka(j)'\} < \infty$ (this condition is assumed throughout this chapter), $\{X(t) : t \in \mathbb{Z}\}$ is a second-order stationary process and has a spectral density matrix represented by

$$f(\omega) = \frac{1}{2\pi} A(\omega) K A(\omega)^*, \quad A(\omega) = \sum_{j=0}^{\infty} a(j) \exp(\mathrm{i}j\omega).$$

As Ogata and Taniguchi [49], we make the following assumptions.

Assumption 2.1.

(i). $\{X(t) : t \in \mathbb{Z}\}$ *is strictly stationary and all its moments exist.*

(ii). *The joint cumulant*

$$Q^X_{\beta_1 \cdots \beta_k}(u_1, \cdots, u_{k-1}) = \mathrm{cum}\{X_{\beta_1}(t), X_{\beta_2}(t + u_1), \cdots, X_{\beta_k}(t + u_{k-1})\}$$

satisfies

$$\sum_{u_1, \cdots, u_{k-1} \in \mathbb{Z}} (1 + |u_i|) \left| Q^X_{\beta_1 \cdots \beta_k}(u_1, \cdots, u_{k-1}) \right| < \infty$$

for $i = 1, \cdots, k-1,\ \beta_1, \cdots, \beta_k \in \{1, \cdots, d\}$ *and any* $k = 2,\ 3,\ \cdots$.

Assumption 2.2. *For the sequence* $\{C_k : k = 1, 2, \cdots\}$ *defined by*

$$C_k = \sup_{\beta_1, \cdots, \beta_k} \sum_{u_1, \cdots, u_{k-1} \in \mathbb{Z}} \left| Q^X_{\beta_1 \cdots \beta_k}(u_1, \cdots, u_{k-1}) \right|,$$

it holds that

$$\sum_{k=1}^{\infty} \frac{C_k}{k!} z^k < \infty$$

for z *in a neighborhood of* 0.

Assumption 2.1 (ii) claims that the dependence between $X(t)$ and $X(t+l)$ becomes weaker as the time lag l becomes larger, so this assumption is quite natural. On the other hand, we need Assumption 2.2 in order to control the stochastic order of the residuals in the asymptotic expansion of the empirical likelihood ratio statistic defined later.

In this Chapter, let $\theta_0 = (\theta_0^1, \cdots, \theta_0^p)'$ be a solution to the p-equations

$$\frac{\partial}{\partial \theta} \int_{-\pi}^{\pi} \operatorname{tr}\left[g(\omega;\theta)^{-1} f(\omega)\right] d\omega \bigg|_{\theta=\theta_0} = 0_p, \tag{2.2}$$

where $g(\omega;\theta)$ is a $d \times d$ matrix-valued function, called a score function. We call this θ_0 a pivotal quantity of the process.

This setting is useful in many situations. Here we impose two concrete examples following Ogata and Taniguchi [49]; prediction and interpolation problems.

The first example is a prediction problem. Let us consider the h-step linear prediction of a scalar stationary process $\{X(t) : t \in \mathbb{Z}\}$, and suppose that we use the linear predictor

$$\widehat{X}(t) = \sum_{j=h}^{\infty} \phi_j(\theta) X(t-j)$$

to predict $X(t)$. The spectral representations of $X(t)$ and $\widehat{X}(t)$ are

$$X(t) = \int_{-\pi}^{\pi} \exp(-\mathrm{i}t\omega) d\zeta_X(\omega), \quad \widehat{X}(t) = \sum_{j=h}^{\infty} \phi_j(\theta) \int_{-\pi}^{\pi} \exp\{-\mathrm{i}(t-j)\omega\} d\zeta_X(\omega),$$

where $\{\zeta_X(\omega) : \omega \in [-\pi, \pi]\}$ is an orthogonal increment process satisfying

$$\mathrm{E}\left[d\zeta_X(\omega)\overline{d\zeta_X(\mu)}\right] = \begin{cases} f(\omega)d\omega & (\omega = \mu) \\ 0 & (\omega \neq \mu) \end{cases}.$$

Now, the best linear predictor minimizes the prediction error

$$\mathrm{E}\left[\left|X(t) - \widehat{X}(t)\right|^2\right] = \int_{-\pi}^{\pi} \left|1 - \sum_{j=h}^{\infty} \phi_j(\theta) \exp(\mathrm{i}j\omega)\right|^2 f(\omega) d\omega, \tag{2.3}$$

hence the best h-step linear predictor is given by $\sum_{j=h}^{\infty} \phi_j(\theta_0) X(t-j)$, where θ_0 minimizes (2.3). Comparing (2.3) with (2.2), if we set

$$g(\omega;\theta) = \left|1 - \sum_{j=h}^{\infty} \phi_j(\theta) \exp(\mathrm{i}j\omega)\right|^{-2},$$

this problem is exactly the same as that of seeking θ_0 in (2.2).

The second example is an interpolation problem. Suppose that we observe the entire time series except for $t = 0$, and estimate missing value $X(0)$ by $\{X(t) : t \neq 0\}$. If we use

$$\widehat{X}(0) = \sum_{j \neq 0} \phi_j^{(\text{int})}(\theta) X(j)$$

to interpolate $X(0)$, the interpolation error is given as

$$\int_{-\pi}^{\pi} \left| 1 - \sum_{j \neq 0} \phi_j^{(\text{int})}(\theta) \exp(\mathrm{i} j \omega) \right|^2 f(\omega) d\omega. \tag{2.4}$$

Therefore, the coefficients of the best interpolator are given as $\{\phi_j^{(\text{int})}(\theta_0) : j \neq 0\}$, where θ_0 is the minimizer of (2.4). Comparing (2.4) with (2.2), if we set

$$g(\omega; \theta) = \left| 1 - \sum_{j \neq 0} \phi_j^{(\text{int})}(\theta) \exp(\mathrm{i} j \omega) \right|^2,$$

this problem is exactly the same as that of seeking θ_0 in (2.2).

It should be noted that the score function $g(\omega; \theta)$ does not necessarily coincide with the true spectral density matrix $f(\omega)$, so this method is applicable to various models in time series analysis. Namely, we can consider estimation problems or hypothesis testings without assuming the true model to be a known parametric model.

To construct the method of nonparametric inference of θ_0, we next introduce the frequency domain empirical likelihood ratio statistic, which was originally introduced by Monti [42] and applied to multivariate non-Gaussian processes by Ogata and Taniguchi [49]. For an observed stretch $X(1), \cdots, X(n)$ from (2.1), let us define the periodogram matrix as

$$I_{n,X}(\omega) = \frac{1}{2\pi n} \left\{ \sum_{t=1}^{n} X(t) \exp(\mathrm{i} j \omega) \right\} \left\{ \sum_{t=1}^{n} X(t) \exp(\mathrm{i} j \omega) \right\}^{*}.$$

Then, the frequency domain empirical likelihood ratio statistic is defined as

$$r_n(\theta) = \sup \left\{ \prod_{t=1}^{n} n w_t : \sum_{t=1}^{n} w_t m(\lambda_t; \theta) = 0_p,\ \sum_{t=1}^{n} w_t = 1,\ 0 \leq w_t \leq 1 \right\}, \tag{2.5}$$

where

$$m(\omega;\theta) = \frac{\partial}{\partial\theta}\mathrm{tr}\left[g(\omega;\theta)^{-1}I_{n,X}(\omega)\right], \tag{2.6}$$

and $\lambda_t = 2\pi t/n$, $t = 1, \cdots, n$. Especially, we call this $m(\omega;\theta)$ an estimating function. This can be regarded as the nonparametric likelihood ratio between null hypothesis $H : \theta = \theta_0$ and alternative hypothesis $A : \theta \neq \theta_0$. Now, we introduce the motivation of the empirical likelihood ratio following Monti [42]. Let Y be a d-dimensional random variable with unknown distribution function $F_Y(y)$ and $y_n = (Y(1), \cdots, Y(n))'$ be i.i.d. observations from $F_Y(y)$. Suppose that we are interested in estimating the nonparametric quantity θ_0 defined as a solution to $\mathrm{E}_{F_Y}[\psi\{Y;\theta_0\}] = 0_p$, where ψ is a known function and E_F denotes the expectation under distribution F. In such case, the unknown parameter θ_0 is estimated by an M-estimator defined as a solution to

$$\sum_{t=1}^{n}\psi\{Y(t);\theta\} = 0_p.$$

On the other hand, if we want to test $H : \theta = \theta_0$ against $A : \theta \neq \theta_0$, it is known that the likelihood ratio test statistic has desirable properties in many ways. However, it is often infeasible to deal with the exact likelihood in practice. So, for given y_n, let us consider the empirical likelihood function

$$L(F_n|y_n) = \prod_{t=1}^{n} w_t,$$

where F_n belongs to $\mathcal{F}$, the family of nonparametric distribution functions defined as

$$\mathcal{F} = \left\{F_n(y) = \sum_{t=1}^{n} w_t\mathbb{I}\{Y(t) \leq y\} : \sum_{t=1}^{n} w_t = 1,\ 0 \leq w_t \leq 1\right\}.$$

Namely, $F_n(y)$ is a distribution function with point mass w_t on each $Y(t)$. Hence, given θ_0 and n, the profile empirical likelihood ratio is given by

$$\begin{aligned}
&\frac{\sup\{L(F_n|y_n) :\ H : \theta = \theta_0\}}{\sup\{L(F_n|y_n) :\ A : \theta \neq \theta_0\}} \\
&= \frac{\sup\left\{\prod_{t=1}^{n} w_t : \mathrm{E}_{F_n}[\psi\{Y;\theta_0\}] = 0_p,\ \sum_{t=1}^{n} w_t = 1,\ 0 \leq w_t \leq 1\right\}}{\prod_{t=1}^{n} n^{-1}} \\
&= \sup\left\{\prod_{t=1}^{n} nw_t : \sum_{t=1}^{n}\psi\{Y(t);\theta_0\} = 0_p,\ \sum_{t=1}^{n} w_t = 1,\ 0 \leq w_t \leq 1\right\}.
\end{aligned}$$

In time series analysis, the Whittle estimator is defined by the maximizer of

$$\sum_{t=1}^{n} \mathrm{tr}\left[g(\lambda_t;\theta)^{-1} I_{n,X}(\lambda_t)\right] d\omega,$$

so the corresponding estimating function is

$$\psi_t\left\{I_{n,X}(\lambda_t);\theta\right\} = \frac{\partial}{\partial\theta}\mathrm{tr}\left[g(\lambda_t;\theta)^{-1} I_{n,X}(\lambda_t)\right], \quad t = 1,\cdots,n.$$

Moreover, it is shown that $\{\mathrm{tr}[g(\lambda_t;\theta)^{-1} I_{n,X}(\lambda_t)] : t = 1,\cdots,n\}$ are asymptotically independent (e.g., Brillinger [3, 4]). Thus, the Whittle estimator has the interpretation of an M-estimator from approximately independent observations and we naturally set the estimating function $m(\lambda_t;\theta)$ as (2.6).

To evaluate the limit distribution of $r_n(\theta)$, we put some regularity conditions on the score function $g(\omega;\theta)$.

Assumption 2.3.

(i). *All components of $g(\omega;\theta)$ are continuously twice differentiable with respect to $\theta \in \Theta$, where Θ is a compact subset of $\mathbb{R}^p$.*

(ii). *$g(\omega;\theta)$ belongs to the parametric spectral family whose element is expressed as*

$$g(\omega;\theta) = B(\omega;\theta)\Omega B(\omega;\theta)^*,$$

where

$$B(\omega;\theta) = \sum_{j=0}^{\infty} b_j(\theta)\exp(\mathrm{i}j\omega),$$

$\{b_j(\theta) : j \in \mathbb{N}\}$ is a sequence of $d \times d$ matrices with $b_0(\theta) = I_{d\times d}$ and Ω is $d \times d$ positive definite which does not depend on θ.

(iii). *$\theta \neq \widetilde{\theta}$ implies $g(\omega;\theta) \neq g(\omega;\widetilde{\theta})$ on a set of positive Lebesgue measure.*

Then, Ogata and Taniguchi [49] showed the limit distribution of the empirical likelihood ratio statistic as follows.

Lemma 2.1 (Ogata and Taniguchi [49])**.** *Suppose that $\{X(t) : t \in \mathbb{Z}\}$ is generated as* (2.1) *and satisfies Assumptions 2.1 and 2.2. Furthermore, suppose that Assumption 2.3 holds. Then, under $H : \theta = \theta_0$,*

$$-2\log r_n(\theta_0) \xrightarrow{\mathcal{L}} N'\Sigma N$$

as $n \to \infty$ with N, a p-dimensional standard normal random vector and $\Sigma = W_2(\theta_0)^{1/2} W_1(\theta_0)^{-1} W_2(\theta_0)^{1/2}$. Here $W_1(\theta)$ and $W_2(\theta)$ are $p \times p$ matrices whose (i,j)th elements are expressed as

$$\begin{aligned} W_1(\theta)_{ij} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{tr} \left[\left\{ \frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^i} f(\omega) \right\} \left\{ \frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^j} f(\omega) \right\} \right] d\omega \\ &\quad + \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{tr} \left[\frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^i} f(\omega) \right] \mathrm{tr} \left[\frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^j} f(\omega) \right] d\omega, \end{aligned} \tag{2.7}$$

$$\begin{aligned} W_2(\theta)_{ij} &= \frac{1}{\pi} \int_{-\pi}^{\pi} \mathrm{tr} \left[\left\{ \frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^i} f(\omega) \right\} \left\{ \frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^j} f(\omega) \right\} \right] d\omega \\ &\quad + \frac{1}{4\pi^2} \sum_{k,l,u,v=1}^{d} \kappa^4_{kluv} \Gamma_i(\theta)_{kl} \Gamma_j(\theta)_{uv}, \end{aligned} \tag{2.8}$$

$\Gamma_i(\theta)_{kl}$ ($i = 1, \cdots, p$ and $k, l = 1, \cdots, d$) is the (k,l)th element of $d \times d$ matrix $\Gamma_i(\theta)$ which is defined as

$$\Gamma_i(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\omega)^* \frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^i} A(\omega) d\omega \tag{2.9}$$

and θ^i denotes the ith element of θ.

By this result, we can construct nonparametric inference for θ_0.

From here, we apply the empirical likelihood method to the discriminant analysis of time series models. Suppose that we observe $X^{(n)} = (X(1)', \cdots, X(n)')'$ from (2.1). In the framework of the discriminant analysis of time series models, we only know that the observed stretch is generated from either one of two categories described by hypotheses

$$\Pi_1 : f_1(\omega) \qquad \Pi_2 : f_2(\omega)$$

where $f_1(\omega)$ and $f_2(\omega)$ are the spectral density matrices associated with each category; that is, we assume that $\{X(t) : t \in \mathbb{Z}\}$ has the spectral density matrices $f_1(\omega)$ and $f_2(\omega)$ under Π_1 and Π_2, respectively. As we mentioned in Introduction of this chapter, we can construct the optimal classification procedure if the log-likelihood function is determined. However, it is usually impossible to deal with the exact likelihood based on $X^{(n)}$ when $\{X(t) : t \in \mathbb{Z}\}$ has dependence structure. So in this chapter we adopt

$$\mathrm{ELR}(\theta_1, \theta_2) = \frac{2}{n} \log \frac{r_n(\theta_1)}{r_n(\theta_2)}$$

as a classification statistic to investigate approximation of the log-likelihood ratio between categories described by two hypotheses

$$\Pi_1 : \theta_0 = \theta_1 \qquad \Pi_2 : \theta_0 = \theta_2.$$

Here $r_n(\theta)$ is the frequency domain empirical likelihood ratio statistic defined as (2.5) and θ_i is the pivotal quantity of the process under Π_i; that is, θ_i is the solution to

$$\frac{\partial}{\partial \theta} \int_{-\pi}^{\pi} \mathrm{tr}\left[g(\omega;\theta)^{-1} f(\omega)\right] d\omega \bigg|_{\theta=\theta_i} = 0_p$$

for $i = 1$ and 2. In other words, we suppose that $\{X(t) : t \in \mathbb{Z}\}$ has pivotal quantities θ_1 and θ_2 under Π_1 and Π_2, respectively, and if $\mathrm{ELR}(\theta_1, \theta_2) > 0$, we assign $X^{(n)}$ into category Π_1, otherwise we choose Π_2.

2.3 Main results

In this section, we evaluate the misclassification probabilities by the empirical likelihood ratio classification statistic. Hereafter, if misclassification probabilities by a classification statistic converge to zero asymptotically, we say that the classification statistic has consistency, or the statistic is consistent. To guarantee the consistency of our classification statistic $\mathrm{ELR}(\theta_1, \theta_2)$, we make the following assumption.

Assumption 2.4. *A $p \times p$ matrix*

$$\Bigg(\int_{-\pi}^{\pi} \mathrm{tr}\left[\left\{\frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^i} f(\omega)\right\}\left\{\frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^j} f(\omega)\right\}\right] d\omega + \int_{-\pi}^{\pi} \mathrm{tr}\left[\frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^i} f(\omega)\right] \mathrm{tr}\left[\frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^j} f(\omega)\right] d\omega : i,j = 1,\cdots,p\Bigg)$$

is positive definite for all $\theta \in \Theta$.

Now, let $\mathrm{Pr}^{(\mathrm{E})}(j|i)$ be the misclassification probability under Π_i when we use $\mathrm{ELR}(\theta_1, \theta_2)$; namely,

$$\mathrm{Pr}^{(\mathrm{E})}(2|1) = \mathrm{Pr}[\mathrm{ELR}(\theta_1,\theta_2) \le 0 | \text{under } \Pi_1],$$
$$\mathrm{Pr}^{(\mathrm{E})}(1|2) = \mathrm{Pr}[\mathrm{ELR}(\theta_1,\theta_2) > 0 | \text{under } \Pi_2].$$

Then, the following theorem shows that $\mathrm{ELR}(\theta_1, \theta_2)$ has fundamental goodness as a classification criterion.

Theorem 2.1. *Suppose that Assumptions 2.1-2.4 hold. Then,*

$$\lim_{n\to\infty} \Pr^{(\mathrm{E})}(2|1) = \lim_{n\to\infty} \Pr^{(\mathrm{E})}(1|2) = 0.$$

However, we can not evaluate the degree of goodness of ELR(θ_1, θ_2) under the situation where $\theta_1 \neq \theta_2$. So consider the situation where θ_2 is contiguous to θ_1. Now we set the pivotal quantities associated with categories as

$$\Pi_1 : \theta_0 = \theta_1 \qquad \Pi_2 : \theta_0 = \theta_{1n} \tag{2.10}$$

where $\theta_{1n} = \theta_1 + n^{-1/2}h$ and $h = (h_1, \cdots, h_p)'$ ($h_i \neq 0$ for all $i = 1, \cdots, p$). We can see the more delicate goodness of the classification statistic by evaluating $\Pr^{(\mathrm{E})}(j|i)$ under the contiguous condition (2.10). We place the following assumptions.

Assumption 2.5. *All components of $g(\omega;\theta)$ are continuously three times differentiable with respect to $\theta \in \Theta$.*

Assumption 2.6. *A $p\times 1$ vector $W_2(\theta_1)^{1/2}W_1(\theta_1)^{-1}F(\theta_1)h$ is not zero, where $W_1(\theta)$ and $W_2(\theta)$ are defined as* (2.7) *and* (2.8)*, respectively, and $F(\theta)$ is a $p \times p$ matrix whose (i,j)th element is expressed as*

$$F(\theta)_{ij} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\left[\frac{\partial^2 g(\omega;\theta)^{-1}}{\partial\theta^i\partial\theta^j} f(\omega)\right] d\omega. \tag{2.11}$$

Then, the misclassification probabilities under (2.10) are evaluated as follows.

Theorem 2.2. *Suppose that Assumptions 2.1-2.6 hold. Then, under the contiguous condition* (2.10)*,*

$$\lim_{n\to\infty} \Pr^{(\mathrm{E})}(2|1) = \lim_{n\to\infty} \Pr^{(\mathrm{E})}(1|2) = \Phi\left[-\frac{1}{2}\frac{h'F(\theta_1)W_1(\theta_1)^{-1}F(\theta_1)h}{\|W_2(\theta_1)^{1/2}W_1(\theta_1)^{-1}F(\theta_1)h\|_E}\right],$$

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

2.4 Non-Gaussian robustness of classification statistics

This section discusses non-Gaussian robustness of classification statistics under contiguous conditions, and compares the goodness of our classification statistic with an existing one. Zhang and Taniguchi [68] proposed the Whittle likelihood ratio type classification statistic

$$I(f_1, f_2) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left[\log\frac{\det[f_2(\omega)]}{\det[f_1(\omega)]} + \mathrm{tr}[I_{n,X}(\omega)\{f_2(\omega)^{-1} - f_1(\omega)^{-1}\}]\right] d\omega$$

for the classification problem described by two hypotheses:

$$\Pi_1 : f_1(\omega) \qquad \Pi_2 : f_2(\omega)$$

where $f_1(\omega)$ and $f_2(\omega)$ are $d \times d$ spectral density matrices associated with categories Π_1 and Π_2, respectively; namely, for $k = 1$ and 2, $f_k(\omega)$ is of the form

$$f_k(\omega) = \frac{1}{2\pi} A_k(\omega) K_k A_k^*(\omega),$$

where $A_k(\omega) = \sum_{j=0}^{\infty} a_k(j) \exp(\mathrm{i}j\omega)$, $\{a_k(j) : j \in \mathbb{N}\}$ is a sequence of $d \times d$ matrices satisfying

$$\sum_{j=0}^{\infty} \mathrm{tr}\{a_k(j) K_k a_k(j)'\} < \infty$$

and K_k is a $d \times d$ symmetric positive definite matrix. The statistic $I(f_1, f_2)$ can be regarded as the Whittle likelihood ratio between Π_1 and Π_2. Hereafter, $\mathrm{Pr}^{(\mathrm{W})}(j|i)$ denotes the misclassification probability of $I(f_1, f_2)$. Zhang and Taniguchi [68] showed that $I(f_1, f_2)$ is a consistent classification statistic. They also evaluated more delicate goodness of the statistic when Π_2 is contiguous to Π_1 as follows. Let the spectral density matrices associated with Π_1 and Π_2 be

$$\Pi_1 : f_1(\omega) = f(\omega|\eta) \qquad \Pi_2 : f_2(\omega) = f\left(\omega|\eta + n^{-1/2}\xi\right). \tag{2.12}$$

Here $f(\omega|\eta)$ is a parametric spectral density matrix of the form

$$f(\omega|\eta) = \frac{1}{2\pi} A(\omega|\eta) K(\eta) A(\omega|\eta)^*,$$

where

$$A(\omega|\eta) = \sum_{j=0}^{\infty} a(j|\eta) \exp(\mathrm{i}j\omega)$$

which depends on a q-dimensional parameter $\eta \in \mathbb{R}^q$ and $\xi = (\xi_1, \cdots, \xi_q)'$ ($\xi_i \neq 0$ for all $i = 1, \cdots, q$). We say that η is an innovation-free parameter if $K(\eta)$ is independent of η, since the matrices $K(\eta)$ is not affected by the coefficients $\{a(j|\eta) : j \in \mathbb{Z}\}$ but the variance of the innovation process $\{e(t) : t \in \mathbb{Z}\}$. In order to evaluate asymptotic misclassification probabilities, assume that all components of $f(\omega|\eta)$ are three times continuously differentiable with respect to η. Then Lemma 2.2 below describes the misclassification probabilities of $I(f_1, f_2)$ under the contiguous situation and gives sufficient condition for non-Gaussian robustness of the classification procedure.

Lemma 2.2 (Zhang and Taniguchi [68]). *Suppose that $\{X(t) : t \in \mathbb{Z}\}$ is the linear process with spectral density matrices $f_1(\omega)$ and $f_2(\omega)$ under Π_1 and Π_2, respectively, and satisfies Assumptions 2.1 and 2.2. Furthermore, assume that the minimum eigenvalue of $f(\omega|\eta)$ is bounded away from zero for all $\omega \in [-\pi, \pi]$.*

(i). *Under the contiguous condition* (2.12),

$$\lim_{n\to\infty} \Pr^{(\mathrm{W})}(2|1) = \lim_{n\to\infty} \Pr^{(\mathrm{W})}(1|2) = \Phi\left[-\frac{1}{2}\frac{I_{\mathrm{ZT}}(\eta)}{\sqrt{W_{\mathrm{ZT}}(\eta)}}\right],$$

where

$$I_{\mathrm{ZT}}(\eta) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \mathrm{tr}\left[\sum_{i=1}^{q} \xi_i \left\{\frac{\partial f(\omega|\eta)}{\partial \eta^i}\right\} f(\omega|\eta)^{-1}\right]^2 d\omega,$$

$$W_{\mathrm{ZT}}(\eta) = I_{\mathrm{ZT}}(\eta) + \frac{1}{16\pi^2}\sum_{k,l,u,v=1}^{d} \kappa^4_{kluv}\Delta_{kl}(\eta)\Delta_{uv}(\eta),$$

$\Delta_{kl}(\eta)$ is the (k,l)th element of $d \times d$ matrix $\Delta(\eta)$ which is defined as

$$\Delta(\eta) = \frac{1}{2\pi}\sum_{i=1}^{q} h_i \int_{-\pi}^{\pi} A(\omega|\eta)^* \left[f(\omega|\eta)^{-1}\left\{\frac{\partial f(\omega|\eta)}{\partial \eta^i}\right\} f(\omega|\eta)^{-1}\right] A(\omega|\eta) d\omega,$$

and η^i denotes the ith element of η.

(ii). *If η is innovation-free, then $I(f_1, f_2)$ is asymptotically independent of $\kappa^4 = \{\kappa^4_{kluv} : k, l, u, v = 1, \cdots, d\}$.*

(ii) of Lemma 2.2 says that $I(f_1, f_2)$ is not affected by the fourth-order cumulant of the process asymptotically if η is innovation-free. That is, the second assertion gives a sufficient condition that $I(f_1, f_2)$ has non-Gaussian robustness.

Next, we apply ELR to the classification problem (2.12). Let us define θ_{1n} by

$$\frac{\partial}{\partial \theta}\int_{-\pi}^{\pi} \mathrm{tr}\left[g(\omega;\theta)^{-1} f\left(\omega\left|\eta + n^{-1/2}\xi\right.\right)\right] d\omega\Bigg|_{\theta=\theta_{1n}} = 0_p,$$

and the misclassification probabilities are derived by evaluating the limit behavior of $\mathrm{ELR}(\theta_1, \theta_{1n})$. However, it is troublesome to deal with θ_{1n} directly, so we approximate θ_{1n} in terms of θ_1, $f(\omega|\eta)$ and $g(\omega;\theta)$. We impose the following assumption.

Assumption 2.7. *A $p \times p$ matrix*

$$\left(\int_{-\pi}^{\pi} \mathrm{tr} \left[\frac{\partial^2 g(\omega;\theta)^{-1}}{\partial\theta^i \partial\theta^j} f(\omega|\eta) \right] d\omega : i, j = 1, \cdots, p \right)$$

is nonsingular.

The Taylor expansion yields $\theta_{1n} = \theta_{1n}^* + O(n^{-1})$, where

$$\theta_{1n}^* = \theta_1 - \frac{1}{\sqrt{n}} F(\theta_1|\eta)^{-1} H(\theta_1|\eta)\xi,$$

$F(\theta|\eta)$ is a $p \times p$ matrix defined by replacing $f(\omega)$ in (2.11) with $f(\omega|\eta)$, and $H(\theta|\eta)$ is a $p \times q$ matrix which is defined in Theorem 2.3 below. The next lemma shows the disparity between $\mathrm{ELR}(\theta_1, \theta_{1n})$ and $\mathrm{ELR}(\theta_1, \theta_{1n}^*)$.

Lemma 2.3. *Suppose that all assumptions in Lemma 2.2, Assumptions 2.3-2.5 and 2.7 hold. Then,*

$$\mathrm{ELR}(\theta_1, \theta_{1n}) - \mathrm{ELR}(\theta_1, \theta_{1n}^*) = o_p\left(\frac{1}{n}\right).$$

As Assumption 2.6, we put the following assumption so that the limit distribution of $\mathrm{ELR}(\theta_1, \theta_2)$ does not degenerate.

Assumption 2.8. *A $p \times 1$ vector $W_2(\theta_1|\eta)^{1/2} W_1(\theta_1|\eta)^{-1} H(\theta_1|\eta)\xi$ is not zero, where $W_1(\theta|\eta)$ and $W_2(\theta|\eta)$ are $p \times p$ matrices; $W_1(\theta|\eta)$ is defined by replacing $f(\omega)$ in* (2.7) *with $f(\omega|\eta)$, and $W_2(\theta|\eta)$ is defined as*

$$\begin{aligned} W_2(\theta|\eta)_{ij} = &\frac{1}{\pi} \int_{-\pi}^{\pi} \mathrm{tr} \left[\left\{ \frac{\partial g(\omega;\theta)^{-1}}{\partial\theta^i} f(\omega|\eta) \right\} \left\{ \frac{\partial g(\omega;\theta)^{-1}}{\partial\theta^j} f(\omega|\eta) \right\} \right] d\omega \\ &+ \frac{1}{4\pi^2} \sum_{k,l,u,v=1}^{d} \kappa_{kluv}^4 \Gamma_i(\theta|\eta)_{kl} \Gamma_j(\theta|\eta)_{uv}, \end{aligned}$$

where $\Gamma_i(\theta|\eta)_{kl}$ $(i = 1, \cdots, p,\ k, l = 1, \cdots, d)$ is the (k, l)th element of $d \times d$ matrix $\Gamma_i(\theta|\eta)$ which is defined by replacing $A(\omega)$ in (2.9) *with $A(\omega|\eta)$. $H(\theta|\eta)$ is a $p \times q$ matrix whose (i, j)th element is expressed as*

$$H(\theta|\eta)_{ij} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{tr} \left[\frac{\partial g(\omega;\theta)^{-1}}{\partial\theta^i} \frac{\partial f(\omega|\eta)}{\partial\eta^j} \right] d\omega.$$

Then, Theorem 2.2 and Lemma 2.3 lead the following theorem, which describes more delicate goodness of $\mathrm{ELR}(\theta_1, \theta_{1n})$ in terms of misclassification probabilities.

Theorem 2.3. *Suppose that all assumptions in Lemma 2.2, Assumptions 2.3, 2.5, 2.7 and 2.8 hold. Furthermore, suppose that Assumption 2.4 holds for* $f(\omega) = f(\omega|\eta)$*.*

(i). *Under the contiguous condition* (2.12),

$$\lim_{n\to\infty} \Pr^{(\mathrm{E})}(2|1) = \lim_{n\to\infty} \Pr^{(\mathrm{E})}(1|2) = \Phi\left[-\frac{1}{2}\frac{\xi' H(\theta_1|\eta)' W_1(\theta_1|\eta)^{-1} H(\theta_1|\eta)\xi}{\|W_2(\theta_1|\eta)^{1/2} W_1(\theta_1|\eta)^{-1} H(\theta_1|\eta)\xi\|_E}\right].$$

(ii). *If* $B(\omega;\theta_1) \equiv A(\omega|\eta)$*, then* $\mathrm{ELR}(\theta_1,\theta_{1n})$ *is asymptotically independent of* κ^4*.*

(iii). *If* $\{X(t) : t \in \mathbb{Z}\}$ *is a scalar process, then* $\mathrm{ELR}(\theta_1,\theta_{1n})$ *is asymptotically independent of* κ^4*.*

Remark 2.1. *We observe an essential difference between the statistics* $\mathrm{ELR}(\theta_1,\theta_2)$ *and* $I(f_1,f_2)$ *from the statements (ii) and (iii) of Theorem 2.3. First,* ELR *is robust with respect to a change of the innovation variance. Second, let us consider a scalar process* $\{X(t) : t \in \mathbb{Z}\}$ *with innovation variance* $\sigma^2 > 0$ *and fourth-order innovation cumulant* $\kappa^4 > 0$*. Let us set* $f(\omega|\eta)$ *as*

$$f(\omega|\eta) = \frac{\sigma^2}{2\pi}\left|\sum_{j=0}^{\infty} a(j|\eta_{\mathrm{if}})\exp(\mathrm{i}j\omega)\right|^2, \quad \eta = (\eta_{\mathrm{if}}', \sigma)' = (\eta^1,\cdots,\eta^{q-1},\sigma)'.$$

Then, $\Delta(\eta)$ *in Lemma 2.2 becomes*

$$\Delta(\eta) = \frac{1}{\sigma^2}\left[\sum_{i=1}^{q-1}\xi_i\int_{-\pi}^{\pi}\left\{\frac{\partial f(\omega|\eta)}{\partial\eta^i}\right\}\frac{1}{f(\omega|\eta)}d\omega + \xi_q\int_{-\pi}^{\pi}\left\{\frac{\partial f(\omega|\eta)}{\partial\sigma}\right\}\frac{1}{f(\omega|\eta)}d\omega\right], \tag{2.13}$$

and it is not difficult to see that the integrations in the first summation of (2.13) *are all 0 (see Brockwell and Davis (1991)). On the other hand, the second term of* (2.13) *becomes* $4\pi\xi_q\sigma^{-3}$ *which does not vanish. So in this case, the misclassification probabilities by* $I(f_1,f_2)$ *always depend on non-Gaussianity of the process. Meanwhile,* $W_2(\theta_1|\eta)$ *is evaluated as*

$$W_2(\theta_1|\eta)_{ij} = \frac{1}{\pi}\int_{-\pi}^{\pi}\frac{\partial g(\omega;\theta)^{-1}}{\partial\theta^i}\frac{\partial g(\omega;\theta)^{-1}}{\partial\theta^j}\bigg|_{\theta=\theta_1} f(\omega)^2 d\omega$$
$$+ \frac{\kappa^4}{4\pi}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial g(\omega;\theta)^{-1}}{\partial\theta^i}\bigg|_{\theta=\theta_1} A(\omega|\eta)\overline{A(\omega|\eta)}d\omega\right\}$$

$$
\times \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \left. \frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^j} \right|_{\theta=\theta_1} A(\omega|\eta)\overline{A(\omega|\eta)} d\omega \right\}
$$

$$
= \frac{1}{\pi} \int_{-\pi}^{\pi} \left. \frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^i} \frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^j} \right|_{\theta=\theta_1} f(\omega)^2 d\omega
$$

$$
+ \frac{\kappa^4}{4\pi\sigma^4} \left\{ \int_{-\pi}^{\pi} \left. \frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^i} \right|_{\theta=\theta_1} f(\omega|\eta) d\omega \right\}
$$

$$
\times \left\{ \int_{-\pi}^{\pi} \left. \frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^j} \right|_{\theta=\theta_1} f(\omega|\eta) d\omega \right\}. \tag{2.14}
$$

Recalling the definition of the pivotal quantity, we can see that the second term in (2.14) *vanishes. That is, the misclassification probabilities by* ELR(θ_1, θ_{1n}) *do not depend on the fourth-order cumulant of the process for any* η_{if} *and* σ *asymptotically.*

2.5 Numerical examples

This section carries out numerical studies for Theorem 2.3. The misclassification probabilities both empirical and Whittle likelihood ratio classification statistics are compared. Consequently, the advantage of the empirical likelihood ratio classification statistic is elucidated.

2.5.1 Example 1: AR(1) model

First, let the spectral density functions associated with Π_1 and Π_2 be

$$
\Pi_1 : f_1(\omega) = f(\omega|\eta) \qquad \Pi_2 : f_2(\omega) = f\left(\omega \left| \eta + n^{-1/2}\xi\right.\right) \tag{2.15}
$$

where

$$
f(\omega|\eta) = \frac{\sigma^2}{2\pi} |1 - b\exp(i\omega)|^{-2}, \quad \eta = (b, \sigma)'
$$

and $\xi = (1, 1)'$. As seen in the previous section, If we use

$$
g(\omega;\theta) = |1 - \theta \exp(i\omega)|^{-2}
$$

as a score function, the contiguous hypothesis (2.15) is understood as

$$
\Pi_1 : \theta_0 = b \qquad \Pi_2 : \theta_0 = b - \frac{1}{\sqrt{n}} + O\left(\frac{1}{n}\right)
$$

in our framework. We calculate the theoretical misclassification probabilities by both $I(f_1, f_2)$ and ELR(θ_1, θ_2) in the following cases:

(i) $\{e(t) : t \in \mathbb{Z}\}$ are i.i.d. random variables with p.d.f.

$$p(x) = \frac{24\sqrt{3}}{\sigma\pi \left\{(x/\sigma)^2 + 3\right\}^3}.$$

(ii) $\{e(t) : t \in \mathbb{Z}\}$ are i.i.d. random variables with p.d.f.

$$p(x) = \begin{cases} \dfrac{2^{-\sigma^2/4}}{\Gamma(\sigma^2/4)} \left(x + \dfrac{\sigma^2}{2}\right)^{\sigma^2/4-1} \exp\left(-\dfrac{x}{2} - \dfrac{\sigma^2}{4}\right) & (x > -\sigma^2/2) \\ 0 & (\text{otherwise}) \end{cases}.$$

Case (i) can be regarded as $t(5)$-distributed random variables with scale σ (namely, $e(1) \sim_d \sqrt{3/5}\sigma Y$ with $Y \sim t(5)$), while case (ii) is shifted gamma distribution with shape and scale parameters $\sigma^2/4$ and 2, respectively. The gamma distribution is one of generalization of the exponential distribution, and the exponential distribution is often used to model lifetime distributions. So it is natural to consider case (ii) even if the distribution is asymmetric. Figures 2.1 and 2.2 show the theoretical limit misclassification probability $\Pr^{(\cdot)}(2|1)$ for $0 < b < 1$ and $\sigma = 0.5, 1$ in cases (i) and (ii).

Figures 2.1 and 2.2 are about here.

Since η is not innovation-free, the misclassification probabilities by $I(f_1, f_2)$ depend on the fourth-order cumulant κ ($6\sigma^4$ in case (i), and $24\sigma^4$ in case (ii)) of the process. On the other hand, by (iii) of Theorem 2.3, the asymptotics of $\mathrm{ELR}(\theta_1, \theta_2)$ are independent of non-Gaussianity of the process. Furthermore, ELR improves the delicate goodness of the classification in these cases.

2.5.2 Example 2: ARMA(1,1) model

We also check misclassification probabilities when a family of score functions does not contain the true spectral density function. We consider the ARMA type spectral density function defined as

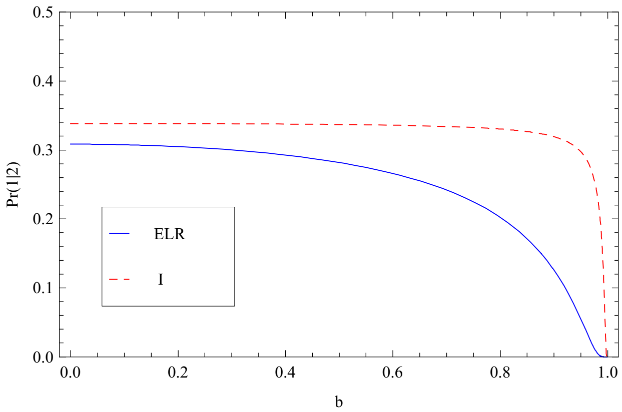
$$f(\omega|\eta) = \frac{\sigma^2}{2\pi} \left| \frac{1 + a\exp(i\omega)}{1 - b\exp(i\omega)} \right|^2, \quad \eta = (a, b, \sigma)'$$

and $\xi = (1, 1, 1)'$. For $|a|, |b| < 1$, Figures 2.3 and 2.4 show the regions where the theoretical limit misclassification probability by the empirical likelihood ratio statistic is smaller than that by the Whittle type statistic for $\sigma = 0.5, 1$. Namely, we plot the region where the empirical likelihood classification statistic shows better performance than classical one.

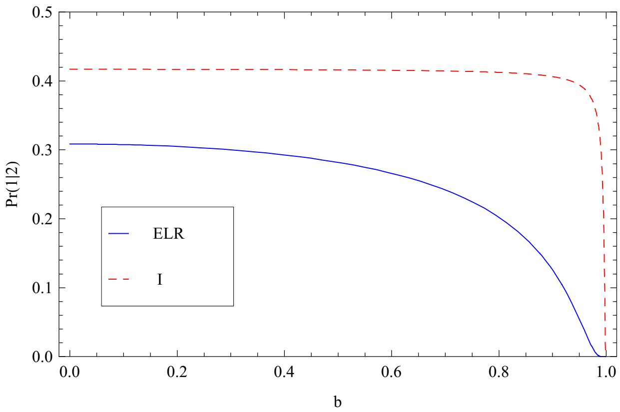
Figures 2.3 and 2.4 are about here.

It is seen that the smaller a becomes, the better performance $\mathrm{ELR}(\theta_1, \theta_2)$ shows, and that the empirical likelihood statistic is uniformly better than the classical one when $g(\omega; \theta_1) = f(\omega)$. As a becomes large, the difference between $g(\omega; \theta_1)$ and $f(\omega|\eta)$ tends to large. However, ELR improves the goodness of classification when the family of score function does not coincide with the true model.

Figure 2.1: Theoretical misclassification probabilities $\Pr^{(\cdot)}(2|1)$ ($\sigma = 0.5$)

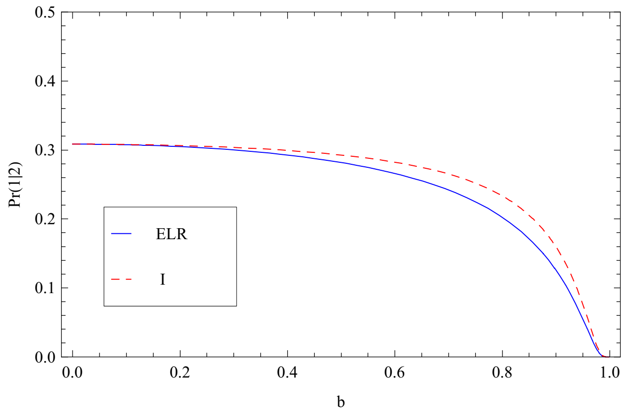


(a) Case (i)

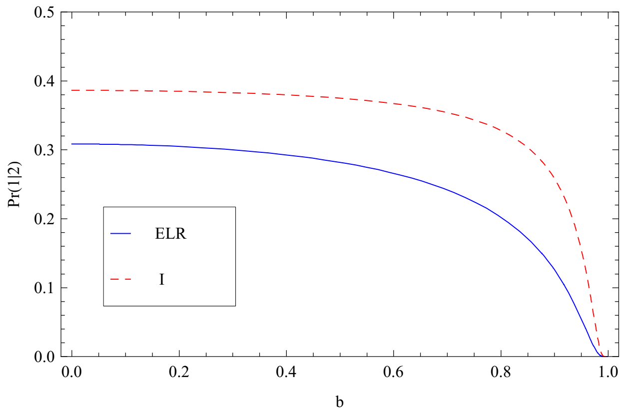


(b) Case (ii)

Figure 2.2: Theoretical misclassification probabilities $\Pr^{(\cdot)}(2|1)$ $(\sigma = 1)$

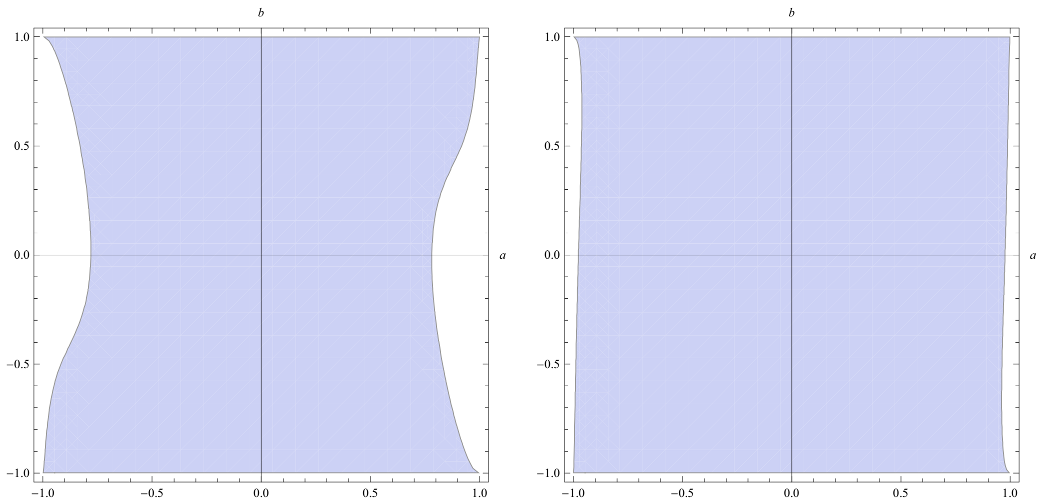


(a) Case (i)



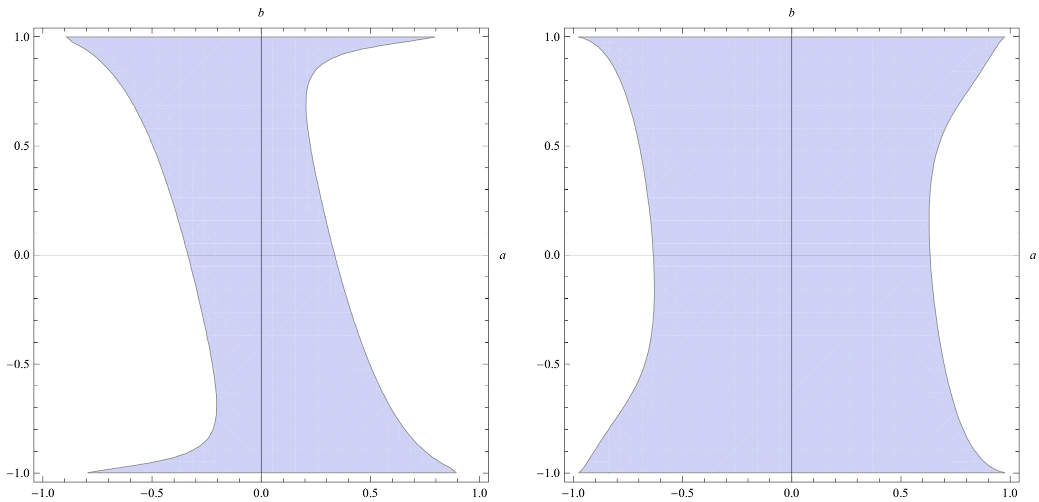
(b) Case (ii)

Figure 2.3: The region where ELR gives smaller misclassification probabilities than I ($\sigma = 0.5$)



(a) Case (i) (b) Case (ii)

Figure 2.4: The region where ELR gives smaller misclassification probabilities than I $(\sigma = 1)$



(a) Case (i)

(b) Case (ii)

3 Empirical likelihood method for symmetric α-stale processes

3.1 Introduction

This chapter extends the empirical likelihood approach to the infinite variance model. As mentioned in the previous chapter, there are ample results on the empirical likelihood approach for time series models with finite variance. However, in the last few decades, heavy-tailed data have been observed in a variety of fields. Nolan [47], Samorodnitsky and Taqqu [58] laid the foundation of the stable distributions, which has infinite variance. Furthermore, Mandelbrot [39] gave economic and financial examples which show that such data are poorly grasped by Gaussian models. Here we provide a concrete example of real data. Let us focus on the stock prices $P_{\rm H}(1), \cdots, P_{\rm H}(627)$ of Hewlett Packard company from January 1, 2012 to June 30, 2014. Figure 3.1 (a) shows daily log-stock returns $X(t) = \log\{P_{\rm H}(t+1)/P_{\rm H}(t)\}$ $(t = 1, \cdots, 626)$ of Hewlett Packard company. For the data, the value of $\widehat{\alpha}_{\rm Hill}^{-1}$ is shown in Figure 3.1 (b) in solid line, which is called Hill-plot (we shall provide the rigorous definition of Hill's estimator $\widehat{\alpha}_{\rm Hill}$ in Section 3.4). On the other hand, the dashed lines are also Hill-plot for i.i.d. α-stable random variables with various tail index α.

Figure 3.1 is about here.

It is known that Hill's estimator is a consistent estimator for the tail index α of the data. These figures imply that it is more suitable to suppose that this data is generated from a process with stable innovations rather than to assume that this data has finite variances (f l-plot, see Dress, de Haan and Resnick [12], Hall [17], Hsing [23], Resnick and Stărikă [56]).

To model such heavy-tailed data suitably, Section 3.2 introduces the linear process generated by stable innovations, called a symmetric α-stable linear process, and proposes natural extension of the frequency domain empirical likelihood ratio statistic to the stable process. Moreover, it is known that important statistics for stable processes behave very much like those for second-order stationary processes in a sense. Many authors studied the limit behavior of quadratic forms of an observed stretch from the stable processes (e.g., Davis and Resnick [11]) and its Fourier transforms (e.g., Klüppelberg and Mikosch [32, 34]). Especially, Klüppelberg and Mikosch [34] elucidated the limit distribution of the integrated self-normalized periodograms, which is one of fundamental tools for stable processes. We shall introduce notable results in Section 3.3. The limit distribution of the empirical likelihood ratio statistic itself is elucidated in Section 3.4. Moreover, we make some numerical examples, which compare the goodness of the empirical likelihood method with the classical sample autocorrelation method proposed by Davis and Resnick [11]. Section 3.5 give the results of the simulations.

3.2 Fundamental settings

Henceforth, we consider the symmetric α-stable linear process $\{X(t) : t \in \mathbb{Z}\}$ generated as

$$X(t) = \sum_{j=0}^{\infty} \psi_j Z(t-j), \tag{3.1}$$

where $\{Z(t);\ t \in \mathbb{Z}\}$ is a sequence of i.i.d. symmetric α-stable random variables with scale parameter $\sigma > 0$, and the characteristic function of $Z(1)$ is given as

$$\mathrm{E}\left[\exp\{\mathrm{i}uZ(1)\}\right] = \exp\{-\sigma|u|^{\alpha}\}, \quad u \in \mathbb{R}^1.$$

To guarantee the a.s. absolute convergence of the process, we make an assumption.

Assumption 3.1. *For some δ satisfying $0 < \delta < \min(1, \alpha)$,*

$$\sum_{j=0}^{\infty} j|\psi_j|^{\delta} < \infty.$$

Under Assumption 3.1, the series (3.1) converges almost surely. This is an easy consequence of the three-series theorem (c.f. Petrov [51]), and the process (3.1) has the normalized power transfer function

$$\widetilde{f}(\omega) = \frac{\left|\sum_{j=0}^{\infty} \psi(j)\exp(\mathrm{i}j\omega)\right|^2}{\sum_{j=0}^{\infty} \psi(j)^2}.$$

This transformation gives a representation of the stable process in frequency domain. Introducing the spectral restriction

$$\frac{\partial}{\partial\theta} \int_{-\pi}^{\pi} \frac{\widetilde{f}(\omega)}{g(\omega;\theta)} d\omega \bigg|_{\theta=\theta_0} = 0_p, \tag{3.2}$$

we define the pivotal quantity of the stable process. Here $g(\omega;\theta)$ is a scalar-valued score function. Bear in mind that the score function does not necessarily coincide with the true normalized power transfer function $\widetilde{f}(\omega)$, and we can choose various important quantities θ_0 by choosing appropriate $g(\omega;\theta)$ as well as the second-order stationary case. Using the method of self-normalizing, we introduce the frequency domain empirical likelihood ratio statistic for the stable process as

$$\widetilde{r}_n(\theta) = \sup\left\{\prod_{t=1}^{n} n w_t : \sum_{t=1}^{n} w_t\, \widetilde{m}(\lambda_t;\theta) = 0_p,\ \sum_{t=1}^{n} w_t = 1,\ 0 \le w_t \le 1\right\},$$

where $\widetilde{m}(\lambda_t;\theta)$ is the estimating function for the stable process which is defined as

$$\widetilde{m}(\omega;\theta) = \frac{\partial}{\partial\theta}\frac{\widetilde{I}_{n,X}(\omega)}{g(\omega;\theta)},$$

and $\widetilde{I}_{n,X}(\omega)$ is the self-normalized periodogram defined as

$$\widetilde{I}_{n,X}(\omega) = \frac{\left|\sum_{t=1}^{n} X(t)\exp(\mathrm{i}t\omega)\right|^2}{\sum_{t=1}^{n} X(t)^2}.$$

This is natural extension of the empirical likelihood ratio statistic in Ogata and Taniguchi [49] to the infinite variance case. The empirical likelihood approach is still useful when we deal with the stable process. For example, set

$$g(\omega;\theta) = \left|1 - \theta\exp(\mathrm{i}l\omega)\right|^{-2}.$$

Solving (3.2) for θ_0, we have

$$\theta_0 = \rho_X(l) = \frac{\sum_{j=0}^{\infty}\psi_j\psi_{j+l}}{\sum_{j=0}^{\infty}\psi_j^2}.$$

On the other hand, a sample autocorrelation function

$$\widehat{\rho}_{n,X}(l) = \frac{\sum_{t=1}^{n-l} X(t)X(t+l)}{\sum_{t=1}^{n} X(t)^2}, \quad l \in \mathbb{N}$$

for the stable process (3.1) is consistent to the autocorrelation function of the process; namely, for fixed l, $p\text{-}\lim_{n\to\infty}\widehat{\rho}_{n,X}(l) = \rho_X(l)$ (e.g., Davis and Resnick [11]).

3.3 Asymptotic theory for fundamental quantities

Current section gives limit theorems of fundamental quantities for both i.i.d. and dependent stable random variables. For any sequence of random variables $\{A(t) : t \in \mathbb{Z}\}$, let us define $\gamma_{n,A}^2$ and $\gamma_{n,A}(h)$ as

$$\gamma_{n,A}^2 = \frac{1}{n^{2/\alpha}} \sum_{t=1}^{n} A(t)^2,$$

$$\gamma_{n,A}(h) = \frac{1}{(n \log n)^{1/\alpha}} \sum_{t=1}^{n-|h|} A(t)A(t+|h|).$$

The following lemma is essentially due to Davis and Resnick [11].

Lemma 3.1 (Davis and Resnick [11]). *For fixed $h \in \mathbb{N}$,*

$$\left(\gamma_{n,Z}^2, \gamma_{n,Z}(1), \cdots, \gamma_{n,Z}(h)\right) \xrightarrow{\mathcal{L}} \left(S(0), S(1), \cdots, S(h)\right),$$

where $\{S(k) : k \in \mathbb{N}\}$ are independent stable random variables; $S(0)$ is a positive $\alpha/2$-stable random variable and $S(1), S(2), \cdots$ are identically distributed symmetric α-stable random variables.

It is shown that limit distributions of many important quantities are expressed by $\{S(k) : k \in \mathbb{N}\}$. Note that Davis and Resnick [11] did not mention the scale parameter of $S(k)$'s, so when we construct confidence intervals, we estimate the quantile of $S(k)$'s by Monte-Carlo simulation.

We next introduce another normalizing sequence

$$x_n = \left(\frac{n}{\log n}\right)^{1/\alpha}, \quad n = 2, 3, \cdots$$

in order to control the rate of convergence of the integrated self-normalized periodogram. It is well known that the self-normalized periodogram for stable processes behaves very much like the usual periodogram for second-order stationary processes (e.g., Klüppelberg and Mikosch [33, 34]). First, we give a representative example of an estimation problem in frequency domain; Whittle's approach for parametric stable process. Mikosch et al. [40] studied estimation of the following causal ARMA process:

$$X(t) + b_1 X(t-1) + \cdots + b_{q_2} X(t-q_2) = Z(t) + a_1 Z(t-1) + \cdots + a_{q_1} Z(t-q_1),$$

where $\{Z(t) : t \in \mathbb{Z}\}$ is a sequence of i.i.d. random variables in the domain of normal attraction of a symmetric α-stable random variable; that is, there exists a symmetric α-stable random variable Z such that

$$\frac{1}{n^{1/\alpha}} \sum_{t=1}^{n} Z(t) \xrightarrow{\mathcal{L}} Z$$

as $n \to \infty$. This is a generalization of the stable distribution, so of course, the stable ARMA model is included in their model. They considered to estimate the parameters of the process. Let $f_\beta(\omega)$ be the parametric power transfer function defined as

$$f_\beta(\omega) = \left| \frac{\sum_{j=0}^{q_1} a_j \exp(\mathrm{i}j\omega)}{\sum_{j=0}^{q_2} b_j \exp(\mathrm{i}j\omega)} \right|^2,$$

with $a_0 = b_0 = 1$, and consider the estimation problem of $\beta = (a_1, \cdots, a_{q_1}, b_1, \cdots, b_{q_2})'$ $\in \mathcal{B}$, where

$$\mathcal{B} = \{\beta \in \mathbb{R}^{q_1+q_2} : a_{q_1}, b_{q_2} \neq 0 \text{ and } a(z)b(z) \neq 0 \text{ for } |z| \leq 1\}$$

with $a(z) = \sum_{j=0}^{q_1} a_j z^j$ and $b(z) = \sum_{j=0}^{q_2} b_j z^j$. They defined the Whittle estimator of β by

$$\widehat{\beta}_n = \arg\min_{\beta \in \mathcal{B}} \int_{-\pi}^{\pi} \frac{\tilde{I}_{n,X}(\omega)}{f_\beta(\omega)} d\omega.$$

Then Mikosch et al. [40] showed that $\widehat{\beta}_n$ is a consistent estimator of the true parameter, and the limit distribution of $x_n(\widehat{\beta}_n - \beta)$ is represented as a sum of stable random variables. Thus Whittle's and frequency domain method has appropriate properties when we deal with the stable process as well as the second-order stationary case.

We next introduce fundamental lemmas for the self-normalized periodogram. Lemma 3.2 below is one of them, and is a multivariate extension of Proposition 3.5 of Klüppelberg and Mikosch [34].

Lemma 3.2. *Suppose that $\{X(t) : t \in \mathbb{Z}\}$ is generated as* (3.1) *satisfying Assumption 3.1, and $\phi_1(\omega), \cdots, \phi_p(\omega)$ be defined on $[-\pi, \pi]$ such that for all $i = 1, \cdots, p$, $\phi_i(\omega)\widetilde{f}(\omega)$ is continuous and it holds that*

$$\sum_{k=1}^{\infty} \left| \int_{-\pi}^{\pi} \sum_{i=1}^{p} \phi_i(\omega)\widetilde{f}(\omega) \cos(k\omega) d\omega \right|^{\mu} < \infty$$

for some $\mu \in (0, \min(1, \alpha))$. Furthermore, $\phi_i(\omega)$ satisfies

$$\int_{-\pi}^{\pi} \phi_i(\omega)\widetilde{f}(\omega)d\omega = 0$$

for all $i = 1, \cdots, p$. Then,

$$\left(x_n \int_{-\pi}^{\pi} \phi_i(\omega) \left\{\widetilde{I}_{n,X}(\omega) - U_n\widetilde{f}(\omega)\right\} d\omega : i = 1, \cdots, p\right)$$
$$\xrightarrow{\mathcal{L}} \left(2\sum_{l=1}^{\infty} \frac{S(l)}{S(0)} \int_{-\pi}^{\pi} \phi_i(\omega)\widetilde{f}(\omega)\cos(l\omega)d\omega : i = 1, \cdots, p\right)$$

for $\alpha \in (0, 2)$. Here

$$U_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\widetilde{I}_{n,X}(\mu)}{\widetilde{f}(\mu)}d\mu \tag{3.3}$$

and $\{S(k) : k \in \mathbb{N}\}$ is the same sequence of random variables as Lemma 3.1.

Lemma 3.2 is easily shown by their proposition and Cramér-Wold device, so we omit the proof here.

3.4 Main results

Now, we give the limit distribution of the frequency domain empirical likelihood ratio statistic and construct a nonparametric confidence region for the pivotal quantity of the stable process. The following assumption is considered to hold for Lemma 3.2.

Assumption 3.2. *There exists $\mu \in (0, \min(1, \alpha))$ such that*

$$\sum_{k=1}^{\infty} \left| \int_{-\pi}^{\pi} \sum_{i=1}^{p} \left.\frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^i}\right|_{\theta=\theta_0} \widetilde{f}(\omega)\cos(k\omega)d\omega \right|^{\mu} < \infty.$$

Roughly speaking, we show the limit distribution as follows; the Lagrangian argument gives

$$-2\log\widetilde{r}_n(\theta_0) = 2\sum_{t=1}^{n} \log\left\{1 + \tau_n'\widetilde{m}(\lambda_t;\theta_0)\right\}, \tag{3.4}$$

where $\tau_n \in \mathbb{R}^p$ is the Lagrange multiplier which is defined as the solution to p-restrictions

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\widetilde{m}(\lambda_t;\theta_0)}{1+\tau_n'\widetilde{m}(\lambda_t;\theta_0)} = 0_p.$$

The residual term in the Taylor expansion of (3.4) contains the maximum of the self-normalized periodogram, whose behavior is differ from that of second-order stationary case. So using the results of Mikosch, Resnick, and Samorodnitsky [41] we control the residual term, and obtain the asymptotic expansion

$$\begin{aligned}
&-\frac{2x_n^2}{n}\log\widetilde{r}_n(\theta_0)\\
&=\left\{\frac{x_n}{2\pi}\int_{-\pi}^{\pi}\widetilde{m}(\omega;\theta_0)d\omega\right\}'\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\widetilde{m}(\omega;\theta_0)\widetilde{m}(\omega;\theta_0)'d\omega\right\}^{-1}\left\{\frac{x_n}{2\pi}\int_{-\pi}^{\pi}\widetilde{m}(\omega;\theta_0)d\omega\right\}\\
&\quad+O_p\left\{\frac{(\log n)^{2-1/\alpha}}{n^{1/\alpha}}\right\} \qquad (3.5)
\end{aligned}$$

under $H:\theta=\theta_0$. Applying Lemma 3.2 to (3.5), we derive the limit distribution of the statistic as follows.

Theorem 3.1. *Suppose that $\{X(t):t\in\mathbb{Z}\}$ is generated as (3.1) with $\alpha\in[1,2)$, and Assumptions 2.3, 3.1 and 3.2 hold. Then, under $H:\theta=\theta_0$,*

$$-\frac{2x_n^2}{n}\log\widetilde{r}_n(\theta_0)\xrightarrow{\mathcal{L}}\widetilde{V}(\theta_0)'\widetilde{W}(\theta_0)^{-1}\widetilde{V}(\theta_0) \qquad (3.6)$$

as $n\to\infty$. Here $\widetilde{V}(\theta)$ and $\widetilde{W}(\theta)$ are $p\times 1$ random vector and $p\times p$ constant matrix, respectively, whose ith and (i,j)th elements are expressed as

$$\begin{aligned}
\widetilde{V}(\theta)_i &= \frac{1}{\pi}\sum_{l=1}^{\infty}\frac{S(l)}{S(0)}\left\{\int_{-\pi}^{\pi}\frac{\partial g(\omega;\theta)^{-1}}{\partial\theta^i}\widetilde{f}(\omega)\cos(l\omega)d\omega\right\}, \qquad (3.7)\\
\widetilde{W}(\theta)_{ij} &= \frac{1}{\pi}\int_{-\pi}^{\pi}\frac{\partial g(\omega;\theta)^{-1}}{\partial\theta^i}\frac{\partial g(\omega;\theta)^{-1}}{\partial\theta^j}\widetilde{f}(\omega)^2d\omega
\end{aligned}$$

and $\{S(k):k\in\mathbb{N}\}$ is the same sequence of random variables as Lemma 3.1.

Remark 3.1. *Generally, we can define the stable process for $\alpha\in(0,2]$. However, we assume that $\alpha\in[1,2)$ to guarantee probability convergence of residual terms appearing in proofs of the theorem. This restriction is not quite strict, since the process (3.1) with $\alpha\in[1,2)$ still does not have the finite second moment.*

Remark 3.2. *The limit distribution* (3.6) *depends on the characteristic exponent* α *and unknown normalized power transfer function* $\widetilde{f}(\omega)$. *We can construct appropriate consistent estimators of them. As mentioned in the previous section, Hill's estimator*

$$\widehat{\alpha}_{\text{Hill}} = \left\{ \frac{1}{k} \sum_{t=1}^{k} \log \frac{|X|_{(t)}}{|X|_{(k+1)}} \right\}^{-1}$$

is a consistent estimator of α, *where* $|X|_{(1)} > \cdots > |X|_{(n)}$ *is the order statistics of* $|X(1)|, \cdots, |X(n)|$ *and* $k = k(n)$ *is an integer satisfying some conditions (e.g. Resnick and Stărică [56]). Next, it is known that the smoothed self-normalized periodogram by an appropriate weighting function* $w_n(\cdot)$ *is consistent to the normalized power transfer function. That is,*

$$\widetilde{J}_{n,X}(\omega) = \sum_{|k| \le m} w_n(k) \widetilde{I}_{n,X}(\omega + \lambda_k) \xrightarrow{\mathcal{P}} \widetilde{f}(\omega)$$

for any $\omega \in [-\pi, \pi]$ *(Klüppelberg and Mikosch [33], Theorem 4.1), where the integer* $m = m(n)$ *satisfies* $m \to \infty$ *and* $m/n \to 0$ *as* $n \to \infty$. *One possible choice of the weighting function* $w_n(\cdot)$ *and* $m = m(n)$ *are* $w_n(k) = (2m+1)^{-1}$ *and* $m = [\sqrt{n}]$ *(*$[x]$ *denotes the integer part of* x*). Then, by Slutsky's lemma and continuous mapping theorem, we obtain consistent estimator* $\widehat{W}(\theta)$ *of* $\widetilde{W}(\theta)$. *So if we choose a proper threshold value* γ_p, *which is the p-percentile corresponding to* $\widetilde{V}(\theta_0)'\widetilde{W}(\theta_0)\widetilde{V}(\theta_0)$, *then* C_p *defined as*

$$C_p = \left\{ \theta \in \Theta : -\frac{2x_n^2}{n} \log \widetilde{r}_n(\theta) < \gamma_p \right\} \tag{3.8}$$

becomes an approximate $p/100$ *level confidence region of* θ_0.

3.5 Numerical examples

In this section, we carry out some simulation studies for Theorems 3.1. We focus on the autocorrelation of the stable process,

$$\rho_X(h) = p\text{-}\lim_{n\to\infty} \frac{\sum_{t=1}^{n-h} X(t)X(t+h)}{\sum_{t=1}^{n} X(t)^2}. \tag{3.9}$$

If we set the score function as $g(\omega;\theta) = |1 - \theta \exp(\mathrm{i}h\omega)|^{-2}$, we obtain

$$\theta_0 = \rho_X(h) = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+h}}{\sum_{j=0}^{\infty} \psi_j^2}.$$

On the other hand, from Davis and Resnick [11], the right-hand side limit of (3.9) exists, and is equal to this θ_0. So it is natural that we define the estimating function $\widetilde{m}(\omega;\theta)$ by this $g(\omega;\theta)$ to estimate $\rho_X(h)$.

The autocorrelation can be estimated by the sample autocorrelation (SAC) method as well. From Davis and Resnick [11], for fixed $l \in \mathbb{N}$,

$$x_n\{\widehat{\rho}_{n,X}(l) - \rho_X(l)\} \xrightarrow{\mathcal{L}} \frac{S(1)}{S(0)} \left\{ \sum_{j=1}^{\infty} |\rho_X(l+j) + \rho_X(l-j) - 2\rho_X(j)\rho_X(l)|^{\alpha} \right\}^{1/\alpha},$$

where $S(0)$ and $S(1)$ are the same random as in Lemma 3.1. Under this setting, we construct confidence intervals of $\theta_0 = \rho_X(1)$ and $\rho_X(2)$ by calculating $\widetilde{r}_n(\theta)$ at numerous point over $(-1, 1)$, and compare confidence intervals constructed by the empirical likelihood method with the SAC method.

3.5.1 Example 3: Stable ARMA(1,1) model

Suppose that the observations $X(1), \cdots, X(n)$ are generated from the following scalar-valued stable ARMA(1,1) model

$$X(t) = bX(t-1) + Z(t) + aZ(t-1), \tag{3.10}$$

where $\{Z(t);\ t \in \mathbb{Z}\}$ is a sequence of i.i.d. symmetric α-stable random variables with scale parameter $\sigma = 1$, and coefficients a and b satisfy $|a|, |b| < 1$. By simple calculation, the normalized power transfer function of the process (3.10) is given as

$$\widetilde{f}(\omega) = \frac{(1+a^2) + 2a\cos(\omega)}{(1+b^2) - 2b\cos(\omega)} \frac{1-b^2}{1+a^2+2ab}.$$

First, we generate 512 samples from (3.10) with $\alpha = 1.5$. Note that in this case, the characteristic exponent $\alpha = 1.5$ and the normalized power transfer $\widetilde{f}(\omega)$ is known. Even though we now neither α nor $\widetilde{f}(\omega)$, we can calculate the consistent estimator $\widetilde{J}_{n,X}(\omega)$ of $\widetilde{f}(\omega)$ and obtain a consistent estimator of $\widetilde{W}(\theta_0)$ as mentioned in Remark 3.2. We also use the Monte-Carlo simulation to calculate γ_{90} which is the 90 percentile of $\widetilde{V}(\theta_0)'\widetilde{W}(\theta_0)^{-1}\widetilde{V}(\theta_0)$. From the reproductive property of stable distribution, the limit distribution reduces to

$$\left\{\frac{S(1)}{S(0)}\right\}^2 \left\{ \sum_{h=1}^{\infty} \left| \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{\partial g(\omega;\theta)^{-1}}{\partial \theta} \widetilde{f}(\omega) \cos(h\omega) d\omega \right|^{\alpha} \right\}^{2/\alpha} \frac{1}{\widetilde{W}(\theta)},$$

so we generate $S(1)/S(0)$ for 10^5 times, and construct an approximate 90% confidence interval of θ_0 as (3.8) (also, see Figure 3.2 below).

Figure 3.2 is about here.

In the tables below, we show values of lower and upper endpoints of confidence intervals and their length by the empirical likelihood (E.L.) and SAC for various a and b.

Table 3.1 is about here.

From Table 3.1, it is seen that the empirical likelihood ratio confidence intervals work as well as the classical one. In particular, almost all widths of confidence intervals for $\rho_X(2)$ are shorter than those of classical one; that is, the empirical likelihood method gives better inference.

We also focus on the two-sided coverage errors to evaluate the performances of the intervals. Let θ^L and θ^U be the lower and upper endpoints of a confidence interval, respectively. The two-sided coverage error is given by

$$\left|\Pr\left[\left\{\theta_0 < \theta^L\right\} \cup \left\{\theta^U < \theta_0\right\}\right] - 0.1\right|.$$

In this time, we calculate the empirical coverage errors of the confidence intervals constructed by both methods by 1000 times of Monte Carlo simulations. Namely, we made 1000 confidence intervals (θ_l^L, θ_l^U), $l = 1, \cdots, 1000$, independently, and calculate the quantity

$$\left|\frac{\sum_{l=1}^{1000} \mathbb{I}\left\{\theta_0 \notin (\theta_l^L, \theta_l^U)\right\}}{1000} - 0.1\right|$$

for given α, a and b. Empirical coverage errors are shown in Table 3.2.

Table 3.2 is about here.

From this table, the empirical likelihood method gives more accurate approximation of theoretical confidence intervals than the existing method in many cases. We also observe that for small α, the empirical likelihood ratio confidence intervals uniformly improve the goodness of inference. When α is near 2, the empirical likelihood method also betters the confidence intervals in some cases. Therefore, it is worth considering the empirical likelihood method, not only the sample autocorrelation method.

3.5.2 Example 4: Stable MA(100) model

Next, we consider the following stable MA(100) model

$$X(t) = \sum_{j=0}^{100} \psi_j Z(t-j),$$

where coefficients $\{\psi_j;\ j \in \mathbb{N}\}$ are defined as

$$\psi_j = \begin{cases} 1 & (j = 0) \\ b^j/j & (1 \le j \le 100) \\ 0 & (\text{otherwise}) \end{cases} . \tag{3.11}$$

Since this process can not be expressed as parametric AR or ARMA models with small dimension, it is suitable to apply the empirical likelihood approach to estimate pivotal unknown quantities. Table 3.3 shows the values of θ_0 and confidence intervals by the empirical likelihood method and the SAC method for $b = 0.1, 0.5,\ 0.9$.

Table 3.3 is about here.

By this simulation, it is shown that the lengths of intervals by the empirical likelihood method are shorter than those by the SAC method in many cases.

Next, we fix $b = 0.5$, and construct confidence intervals in the cases where $\alpha = 1.0$ (Cauchy), 1.5 and 1.9 (near Gaussian).

Table 3.4 is about here.

It is seen that the lengths of confidence intervals depend on the characteristic exponent α. In particular, the empirical likelihood method provides better inferences than the SAC method when α is nearly 1.

Moreover, we investigate the length of intervals when $b = 0.5$ and $\alpha = 1.5$ for small samples. Table 3.5 shows the result for $n = 64$ and 128.

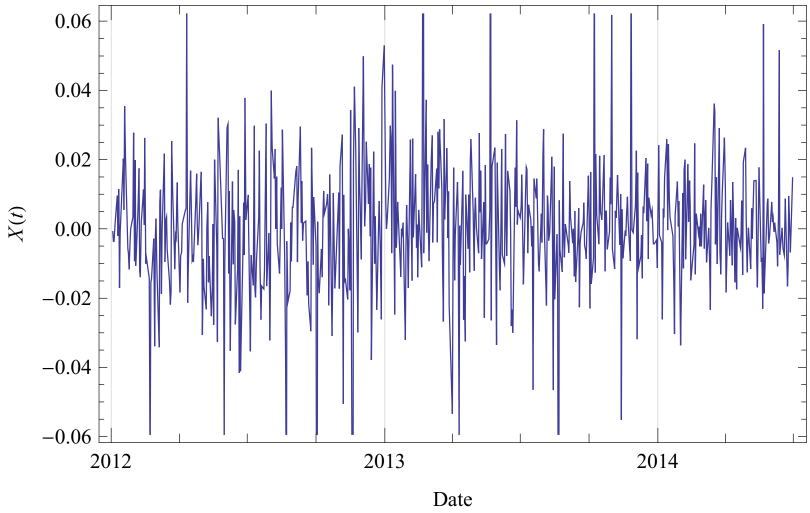
Table 3.5 is about here.

Even though sample size is small, the empirical likelihood method also works well. The empirical coverage errors are also calculated as well as Example 3. Tables 3.6 and 3.7 show the results for $n = 512, 128, 64$, $\alpha = 1.0, 1.5, 1.9$.

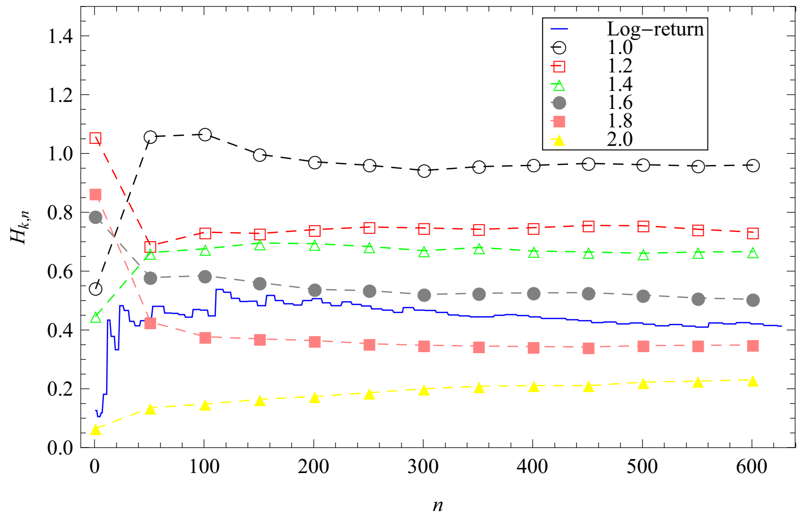
Tables 3.6 and 3.7 are about here.

As the overall tendency, the coverage errors decrease as the sample size increases, and the empirical likelihood method often gives smaller coverage errors for heavy-tailed data.

Figure 3.1: The log-stock return process of Hewlett Packard company (from Jan. 1, 2012 to Jun. 30, 2014) and its hill-plot



(a) Log-stock return $X(t) = \log\{P_{\mathrm{H}}(t+1)/P_{\mathrm{H}}(t)\}$



(b) Hill-plot for $X(t)$

Figure 3.2: The value of the empirical likelihood ratio statistic and threshold value γ_{90} for ARMA(1,1) process with $n = 512$, $(a, b) = (0.1, 0.5)$

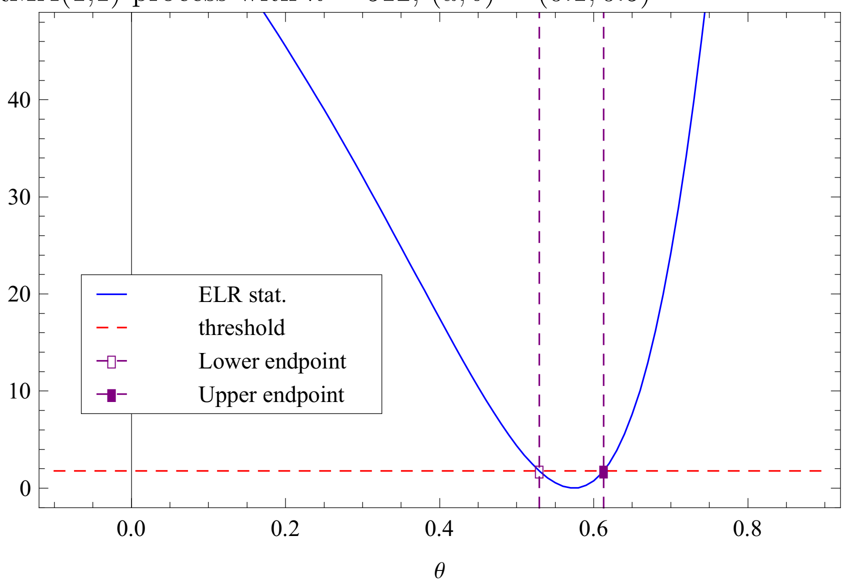


Table 3.1: Lower and upper endpoints of 90% confidence intervals (and length) for $\rho_X(1)$ and $\rho_X(2)$ of the stable ARMA(1,1) process ($n = 512$ and $\alpha = 1.5$)

(a, b)	$\rho_X(1)$	E.L.			SAC		
(0.1,0.1)	0.19612	0.11567	0.22948	(0.11381)	0.11418	0.23339	(0.11921)
(0.1,0.5)	0.56757	0.50321	0.61447	(0.11126)	0.51950	0.62920	(0.10970)
(0.1,0.9)	0.91597	0.86102	0.94252	(0.08150)	0.88139	0.95046	(0.06907)
(0.5,0.1)	0.46667	0.39893	0.48799	(0.08905)	0.39509	0.49486	(0.09977)
(0.5,0.5)	0.71429	0.65515	0.73834	(0.08318)	0.66082	0.74184	(0.08102)
(0.5,0.9)	0.94419	0.91659	0.95666	(0.04007)	0.91939	0.96619	(0.04680)
(0.9,0.1)	0.54774	0.52937	0.61254	(0.08317)	0.52807	0.61874	(0.09067)
(0.9,0.5)	0.74908	0.72851	0.79695	(0.06844)	0.73010	0.80329	(0.07320)
(0.9,0.9)	0.94985	0.90036	0.95909	(0.05873)	0.92307	0.96530	(0.04224)

(a, b)	$\rho_X(2)$	E.L.			SAC		
(0.1,0.1)	0.01961	0.00175	0.10645	(0.10470)	−0.01163	0.12079	(0.13241)
(0.1,0.5)	0.28378	0.22514	0.36719	(0.14205)	0.21626	0.38543	(0.16917)
(0.1,0.9)	0.82437	0.70627	0.83112	(0.12485)	0.71932	0.86186	(0.14254)
(0.5,0.1)	0.04667	−0.02411	0.11606	(0.14017)	−0.03346	0.12968	(0.16314)
(0.5,0.5)	0.35714	0.22549	0.38510	(0.15961)	0.22468	0.39921	(0.17453)
(0.5,0.9)	0.84977	0.80698	0.90784	(0.10086)	0.81277	0.93776	(0.12499)
(0.9,0.1)	0.05477	0.04592	0.20407	(0.15815)	0.04074	0.21413	(0.17339)
(0.9,0.5)	0.37454	0.22142	0.39407	(0.17265)	0.22705	0.40219	(0.17515)
(0.9,0.9)	0.85487	0.78321	0.91030	(0.12709)	0.81162	0.93298	(0.12135)

Table 3.2: Empirical coverage errors of confidence intervals for $\rho_X(1)$ and $\rho_X(2)$ of the stable ARMA(1,1) process ($n = 512$)

		$\alpha = 1.0$		$\alpha = 1.5$		$\alpha = 1.9$	
(a, b)	$\rho_X(1)$	E.L.	SAC	E.L.	SAC	E.L.	SAC
(0.1,0.1)	0.19612	0.000	0.034	0.024	0.019	0.005	0.029
(0.1,0.5)	0.56757	0.004	0.053	0.030	0.043	0.035	0.015
(0.1,0.9)	0.91597	0.025	0.087	0.023	0.063	0.035	0.000
(0.5,0.1)	0.46667	0.019	0.051	0.025	0.040	0.027	0.022
(0.5,0.5)	0.71429	0.016	0.061	0.022	0.040	0.037	0.033
(0.5,0.9)	0.94419	0.035	0.083	0.025	0.042	0.030	0.023
(0.9,0.1)	0.54774	0.000	0.057	0.026	0.037	0.006	0.015
(0.9,0.5)	0.74908	0.028	0.068	0.034	0.033	0.043	0.016
(0.9,0.9)	0.94985	0.033	0.082	0.030	0.063	0.043	0.020

		$\alpha = 1.0$		$\alpha = 1.5$		$\alpha = 1.9$	
(a, b)	$\rho_X(2)$	E.L.	SAC	E.L.	SAC	E.L.	SAC
(0.1,0.1)	0.01961	0.015	0.038	0.011	0.008	0.020	0.022
(0.1,0.5)	0.28378	0.015	0.057	0.007	0.040	0.041	0.032
(0.1,0.9)	0.82437	0.033	0.087	0.042	0.059	0.018	0.025
(0.5,0.1)	0.04667	0.015	0.054	0.029	0.025	0.009	0.030
(0.5,0.5)	0.35714	0.014	0.068	0.045	0.044	0.024	0.001
(0.5,0.9)	0.84977	0.025	0.081	0.034	0.058	0.040	0.027
(0.9,0.1)	0.05477	0.016	0.050	0.021	0.016	0.014	0.025
(0.9,0.5)	0.37454	0.008	0.063	0.011	0.045	0.048	0.023
(0.9,0.9)	0.85487	0.037	0.088	0.027	0.058	0.047	0.027

Table 3.3: Lower and upper endpoints of 90% confidence intervals (and length) for $\rho_X(1)$ and $\rho_X(2)$ of the stable MA(100) process ($n = 512$ and $\alpha = 1.5$)

b	$\rho_X(1)$	E.L.			SAC		
0.1	0.09950	0.03043	0.13768	(0.10725)	0.03981	0.15934	(0.11953)
0.5	0.44845	0.42595	0.53784	(0.11189)	0.39401	0.50302	(0.10901)
0.9	0.69180	0.65973	0.75848	(0.09875)	0.64246	0.74126	(0.09880)

b	$\rho_X(2)$	E.L.			SAC		
0.1	0.00498	−0.03597	0.08471	(0.12069)	−0.05681	0.06692	(0.12372)
0.5	0.11683	0.04761	0.17770	(0.13009)	0.03682	0.19702	(0.16021)
0.9	0.36034	0.26524	0.46044	(0.19520)	0.26110	0.45982	(0.19872)

Table 3.4: Lower and upper endpoints of 90% confidence intervals (and length) for $\rho_X(1)$ and $\rho_X(2)$ of the stable MA(100) process ($n = 512$ and $b = 0.5$)

$\rho_X(1) = 0.44845$						
α	E.L.			SAC		
1.0	0.40860	0.49536	(0.08676)	0.40110	0.49506	(0.09396)
1.5	0.43015	0.53169	(0.10154)	0.39401	0.50302	(0.10901)
1.9	0.35614	0.46635	(0.11021)	0.39368	0.50314	(0.10947)

$\rho_X(2) = 0.11683$						
α	E.L.			SAC		
1.0	0.04222	0.18020	(0.13799)	0.04100	0.19146	(0.15046)
1.5	0.00671	0.15346	(0.14675)	0.03682	0.19702	(0.16021)
1.9	0.07135	0.22026	(0.14891)	0.03895	0.19460	(0.15565)

Table 3.5: Lower and upper endpoints of 90% confidence intervals (and length) for $\rho_X(1)$ and $\rho_X(2)$ of the stable MA(100) process ($b = 0.5$, $\alpha = 1.5$)

$\rho_X(1) = 0.44845$						
n	E.L.			SAC		
64	0.34256	0.69135	(0.34878)	0.28226	0.61503	(0.33277)
128	0.31261	0.51944	(0.20683)	0.33242	0.56474	(0.23232)

$\rho_X(2) = 0.11683$						
n	E.L.			SAC		
64	−0.10756	0.24402	(0.35159)	−0.12741	0.36163	(0.48904)
128	−0.18279	0.11484	(0.29763)	−0.05368	0.28774	(0.34142)

Table 3.6: Empirical coverage errors of confidence intervals for $\rho_X(1)$ of the stable MA(100) process

($n = 512$)

		$\alpha = 1.0$		$\alpha = 1.5$		$\alpha = 1.9$	
b	$\rho_X(1)$	E.L.	SAC	E.L.	SAC	E.L.	SAC
0.1	0.09950	0.024	0.018	0.001	0.017	0.016	0.030
0.5	0.44845	0.012	0.038	0.008	0.012	0.034	0.034
0.9	0.69180	0.053	0.068	0.015	0.056	0.030	0.030

($n = 128$)

		$\alpha = 1.0$		$\alpha = 1.5$		$\alpha = 1.9$	
b	$\rho_X(1)$	E.L.	SAC	E.L.	SAC	E.L.	SAC
0.1	0.09950	0.036	0.046	0.008	0.002	0.086	0.062
0.5	0.44845	0.061	0.070	0.011	0.025	0.068	0.050
0.9	0.69180	0.080	0.090	0.003	0.040	0.070	0.074

($n = 64$)

		$\alpha = 1.0$		$\alpha = 1.5$		$\alpha = 1.9$	
b	$\rho_X(1)$	E.L.	SAC	E.L.	SAC	E.L.	SAC
0.1	0.09950	0.039	0.072	0.005	0.009	0.091	0.082
0.5	0.44845	0.078	0.072	0.005	0.044	0.096	0.107
0.9	0.69180	0.080	0.085	0.001	0.048	0.137	0.072

Table 3.7: Empirical coverage errors of confidence intervals for $\rho_X(2)$ of the stable MA(100) process

$(n = 512)$

		$\alpha = 1.0$		$\alpha = 1.5$		$\alpha = 1.9$	
b	$\rho_X(2)$	E.L.	SAC	E.L.	SAC	E.L.	SAC
0.1	0.00498	0.014	0.018	0.002	0.024	0.042	0.026
0.5	0.11683	0.025	0.048	0.009	0.026	0.024	0.019
0.9	0.36034	0.057	0.070	0.016	0.036	0.013	0.020

$(n = 128)$

		$\alpha = 1.0$		$\alpha = 1.5$		$\alpha = 1.9$	
b	$\rho_X(2)$	E.L.	SAC	E.L.	SAC	E.L.	SAC
0.1	0.00498	0.032	0.046	0.018	0.000	0.079	0.030
0.5	0.11683	0.052	0.062	0.005	0.035	0.091	0.064
0.9	0.36034	0.080	0.078	0.028	0.031	0.086	0.055

$(n = 64)$

		$\alpha = 1.0$		$\alpha = 1.5$		$\alpha = 1.9$	
b	$\rho_X(2)$	E.L.	SAC	E.L.	SAC	E.L.	SAC
0.1	0.00498	0.059	0.050	0.005	0.006	0.089	0.091
0.5	0.11683	0.070	0.078	0.008	0.030	0.122	0.087
0.9	0.36034	0.087	0.095	0.004	0.045	0.140	0.093

4 Discriminant analysis of symmetric α-stale processes

4.1 Introduction

We so far showed that the empirical likelihood method is applicable to the discriminant analysis (Chapter 2), and the method works well in the case of the stable process (Chapter 3) without assuming the structure of the true model to be known. In this chapter, we integrate the results in Chapter 3 with 2; that is, we construct the empirical likelihood-based classification procedure for the class of the symmetric α-stable processes. As seen in Chapter 3, there are a lot of heavy-tailed data in practice, so it is to be desired that the results in Chapter 2 are extended to the case of the symmetric α-stable process. In the stable case, Nishikawa and Taniguchi [46] considered the Whittle likelihood ratio type statistic for classification problems, and obtained satisfactory results. Thus, the likelihood ratio (or its approximation) seems to allow us to construct appropriate classification procedures even if observations have heavy tail.

We provide the empirical likelihood classification statistic for the stable process in Section 4.2. In Section 4.3, we show that our method still has fundamental goodness as a classification procedure. We also elucidate delicate goodness of our statistic in terms of misclassification probabilities under the contiguous condition. It is noteworthy that, in the stable case, the misclassification probabilities under the contiguous condition have different behavior from the second-order stationary case. Section 4.4 gives numerical examples under the situation where we know neither true normalized power transfer functions nor pivotal quantities of categories. This setting resembles the practical situation, and we observe that the empirical likelihood classification statistic works better than the existing one.

4.2 Fundamental settings

Let $\{X(t) : t \in \mathbb{Z}\}$ be the symmetric α-stable linear process (3.1). As the second-order stationary case, we consider the problem of classifying an observed stretch $X^{(n)} = \{X(1), \cdots, X(n)\}$ into one of categories described by two hypotheses:

$$\Pi_1 : \theta_0 = \theta_1 \qquad \Pi_2 : \theta_0 = \theta_2.$$

Namely, the symmetric α-stable linear process (3.1) has p-dimensional pivotal quantities θ_1 and θ_2 under Π_1 and Π_2, respectively. In other words, the normalized power transfer function $\widetilde{f}(\omega)$ of (3.1) satisfies

$$\frac{\partial}{\partial\theta}\int_{-\pi}^{\pi}\frac{\widetilde{f}(\omega)}{g(\omega;\theta)}d\omega\bigg|_{\theta=\theta_i}=0_p$$

under Π_i, $i=1$ and 2.

Remark 4.1. *As seen in Figure* 3.1, *we can estimate that Hewlett Packard company's log-stock return has heavy-tailed distribution. Analogously, when we apply the same method as the case of Hewlett Packard to the stock prices of IBM (P_{I}) and Ford (P_{F}), it seems that IBM and Ford's log-stock return processes are also heavy-tailed data (Figures 4.1 and 4.2).*

Figures 4.1 and 4.2 are about here.

If we use $g(\omega;\theta)=|1-\theta\exp(\mathrm{i}l\omega)|^{-2}$ as a score function, the pivotal quantity of the process (3.1) *is given as*

$$\theta_0=\frac{\sum_{j=0}^{\infty}\psi_j\psi_{j+l}}{\sum_{j=0}^{\infty}\psi_j^2},$$

and $\widehat{\rho}_{n,X}(l)$ is a consistent estimator of θ_0. Table 4.1 *below shows the values of $\widehat{\rho}_{n,X}(1)$ and $\widehat{\rho}_{n,X}(2)$ for the log-stock return processes of Hewlett Packard, IBM and Ford.*

Table 4.1 is about here.

It seems that there is disparities between two of the pivotal quantities of the log-stock return processes of the companies. In our framework, we utilize such difference of the categories for the discriminant analysis. In particular, the classification problem of α-stable processes has a big potential which is applicable to various fields involving electrical engineering, hydrology, finance, physical systems, and so on.

We need to investigate the log-likelihood ratio (or its approximation) between Π_1 and Π_2 to carry out the discriminant analysis, so we adopt

$$\widetilde{\mathrm{ELR}}(\theta_1,\theta_2)=\frac{2}{n}\log\frac{\widetilde{r}_n(\theta_1)}{\widetilde{r}_n(\theta_2)}$$

as a classification statistic.

4.3 Main results

This section gives asymptotics of misclassification probabilities by the empirical likelihood ratio classification statistic for the stable process. The following assumption is needed to null the misclassification probabilities asymptotically.

Assumption 4.1. *A $p \times p$ matrix*

$$\int_{-\pi}^{\pi} \frac{\partial g(\omega;\theta)^{-1}}{\partial\theta} \frac{\partial g(\omega;\theta)^{-1}}{\partial\theta'} \widetilde{f}(\omega)^2 d\omega$$

is positive definite for all $\theta \in \Theta$.

We obtain the consistency of the classification from the following theorem.

Theorem 4.1. *Suppose that Assumptions 2.3, 3.1, 3.2 and 4.1 hold. Then, for $0 < \alpha < 2$,*

$$\lim_{n\to\infty} \Pr{}^{(\mathrm{E})}(1|2) = \lim_{n\to\infty} \Pr{}^{(\mathrm{E})}(2|1) = 0.$$

Thus the empirical likelihood ratio classification procedure also has fundamental goodness even if the process has the infinite variance. Next, we evaluate the limit behavior of misclassification probabilities under the contiguous condition

$$\Pi_1 : \theta_0 = \theta_1 \qquad \Pi_2 : \theta_0 = \widetilde{\theta}_{1n} \tag{4.1}$$

where $\widetilde{\theta}_{1n} = \theta_1 + x_n^{-1}h$ and $h \in \mathbb{R}^p$.

Remark 4.2. *The reason why we use such $\widetilde{\theta}_{1n}$ is that the rate of convergence of the integrated self-normalized periodograms is different from that of the second-order stationary case (see Lemma 3.2). Suppose that we choose b_n ($\to 0$ as $n \to \infty$) such that $a_n/b_n \to 0$, $a_n = x_n^{-1}$, and consider the contiguous condition*

$$\Pi_1 : \theta_0 = \theta_1 \qquad \Pi_2 : \theta_0 = \theta_1 + b_n h \tag{4.2}$$

instead of (4.1)*. Then, it is shown that* $\Pr^{(\mathrm{E})}(2|1)$ *and* $\Pr^{(\mathrm{E})}(1|2)$ *converge to 0 under* (4.2)*. On the other hand, if we choose c_n ($\to 0$ as $n \to \infty$) such that $a_n/c_n \to \infty$, both* $\Pr^{(\mathrm{E})}(2|1)$ *and* $\Pr^{(\mathrm{E})}(1|2)$ *converge to* $1/2$*. Therefore, we can not compare the goodness of the classification by the empirical likelihood approach with the existing method, which is introduced in the next section, in terms of the misclassification probabilities if we choose different rate from x_n.*

To elucidate the more delicate goodness of our classification procedure, we assume the following.

Assumption 4.2. *A $p \times 1$ vector $\widetilde{W}(\theta_1)^{-1}\widetilde{F}(\theta)h$ is not zero, where $\widetilde{W}(\theta)$ is the same $p \times p$ matrix as Theorem 3.1 and*

$$\widetilde{F}(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial^2 g(\omega;\theta)^{-1}}{\partial\theta\partial\theta'}\widetilde{f}(\omega)d\omega.$$

Consequently, the delicate goodness of our statistic for the stable process is evaluated as follows.

Theorem 4.2. *Suppose that all assumptions in Theorem 4.1, Assumptions 2.5 and 4.2 hold. Then, under the contiguous condition* (4.1)*,*

$$\begin{aligned}\lim_{n\to\infty} \Pr{}^{(\mathrm{E})}(2|1) &= \lim_{n\to\infty} \Pr{}^{(\mathrm{E})}(1|2)\\ &= \Pr\left[2h'\widetilde{F}(\theta)\widetilde{W}(\theta_1)^{-1}\widetilde{V}(\theta_1) \leq -h'\widetilde{F}(\theta_1)\widetilde{W}(\theta_1)^{-1}\widetilde{F}(\theta_1)h\right]\end{aligned}$$

for $1 \leq \alpha < 2$, where $\widetilde{V}(\theta)$ is the same $p \times 1$ random vector as Theorem 3.1.

When we deal with the stable processes, the delicate misclassification probabilities show different behavior from the second-order stationary case. We also apply $\widetilde{\mathrm{ELR}}$ to the situation of Nishikawa and Taniguchi [46]. As well as the second-order stationary case, the Whittle likelihood ratio classification statistic is defined as

$$\widetilde{I}(\widetilde{f}_1,\widetilde{f}_2) = \int_{-\pi}^{\pi}\left[\log\frac{\widetilde{f}_2(\omega)}{\widetilde{f}_1(\omega)} + \left\{\frac{1}{\widetilde{f}_2(\omega)} - \frac{1}{\widetilde{f}_1(\omega)}\right\}\widetilde{I}_{n,X}(\omega)\right]d\omega,$$

where $\widetilde{f}_i(\omega)$ is the normalized power transfer function which describes the ith category. Furthermore, the delicate goodness of $\widetilde{I}(\widetilde{f}_1,\widetilde{f}_2)$ is evaluated as follows; let the normalized power transfer functions associated with categories be

$$\Pi_1 : \widetilde{f}_1(\omega) = \widetilde{f}(\omega|\eta) \qquad \Pi_2 : \widetilde{f}_2(\omega) = \widetilde{f}\left(\omega\left|\eta + x_n^{-1}\xi\right.\right) \tag{4.3}$$

where $\widetilde{f}(\omega|\eta)$ is a parametric normalized power transfer function and $\eta, \xi \in \mathbb{R}^q$. Hereafter, $\widetilde{f}(\omega|\eta)$ is assumed to be continuously three times differentiable with respect to η and $\widetilde{f}(\omega|\eta) > 0$ for all ω. Then, the misclassification probabilities by the Whittle likelihood ratio classification statistics are given as follows.

Lemma 4.1 (Nishikawa and Taniguchi [46])**.** *Suppose that Assumption 3.1 holds, and for some $\delta \in (0, \min(1,\alpha))$, the followings hold:*

$$\sum_{k=1}^{q}\sum_{l=1}^{\infty}\left|\int_{-\pi}^{\pi}\frac{\partial\log\widetilde{f}(\omega|\eta)}{\partial\eta^k}\cos(l\omega)d\omega\right|^{\delta} < \infty,$$

$$\sum_{k=1}^{q}\sum_{l=1}^{q}\sum_{l=1}^{\infty}\left|\int_{-\pi}^{\pi}\left\{\frac{\partial^2\widetilde{f}(\omega|\eta)}{\partial\eta^k\partial\eta^l}\right\}\widetilde{f}(\omega|\eta)\cos(l\omega)d\omega\right|^{\delta}<\infty.$$

Then, under the contiguous condition (4.3),

$$\lim_{n\to\infty}\Pr{}^{(\mathrm{W})}(2|1)=\lim_{n\to\infty}\Pr{}^{(\mathrm{W})}(1|2)=\Pr\left[\frac{Z(1)}{Y}\le-\frac{\xi'\mathcal{F}(\eta)\xi}{4\left\{C_\alpha\sum_{k=1}^{\infty}|\mathcal{E}_k(\eta)'\xi|^\alpha\right\}^{1/\alpha}}\right].$$

Here

$$\begin{aligned}\mathcal{F}(\eta)&=\int_{-\pi}^{\pi}\frac{\partial\log\widetilde{f}(\omega|\eta)}{\partial\eta}\frac{\partial\log\widetilde{f}(\omega|\eta)}{\partial\eta'}d\omega,\\ \mathcal{E}_k(\eta)&=\int_{-\pi}^{\pi}\frac{\partial\log\widetilde{f}(\omega|\eta)}{\partial\eta}\{\cos(k\omega)-\rho_X(k)\}d\omega,\\ C_\alpha&=\begin{cases}\dfrac{(1-\alpha)\sigma}{2\Gamma(2\alpha)\cos(\pi\alpha/2)} & (\alpha\ne1)\\ \dfrac{\sigma}{\pi} & (\alpha=1)\end{cases}\end{aligned}$$

and Y *is an* $\alpha/2$*-stable random variable which is independent of* $\{Z(t):t\in\mathbb{Z}\}$ *and has Laplace transform*

$$\mathrm{E}\left[\exp(-rY)\right]=\exp\left(-\sigma K_\alpha r^{\alpha/2}\right),$$

where $K_\alpha=\mathrm{E}[|N|^{\alpha/2}]$ *for an* $N(0,2)$*-random variable.*

Next, we consider $\Pr^{(\mathrm{E})}(j|i)$ under (4.3). The following assumption is needed for the evaluation.

Assumption 4.3. *A* $p\times1$ *vector* $\widetilde{W}(\theta_1|\eta)^{-1}\widetilde{H}(\theta_1|\eta)\xi$ *is not zero, where* $\widetilde{W}(\theta|\eta)$ *is defined by replacing* $\widetilde{f}(\omega)$ *of* $\widetilde{W}(\theta)$ *in Theorem 3.1 with* $\widetilde{f}(\omega|\eta)$ *and*

$$\widetilde{H}(\theta|\eta)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial g(\omega;\theta)^{-1}}{\partial\theta}\frac{\partial\widetilde{f}(\omega|\eta)}{\partial\eta'}d\omega.$$

Theorem 4.3. *Suppose that Assumptions 2.5, 4.3 and all assumptions in Theorem 4.1 for* $\widetilde{f}(\omega)=\widetilde{f}(\omega|\eta)$ *hold. Furthermore, Assumption 2.7 holds for* $f(\omega|\eta)=\widetilde{f}(\omega|\eta)$*. Then, under the contiguous condition* (4.3),

$$\lim_{n\to\infty}\Pr{}^{(\mathrm{E})}(2|1)=\lim_{n\to\infty}\Pr{}^{(\mathrm{E})}(1|2)$$

$$= \Pr\left[2\xi'\widetilde{H}(\theta_1|\eta)'\widetilde{W}(\theta_1|\eta)^{-1}\widetilde{V}(\theta_1|\eta) \le -\xi'\widetilde{H}(\theta_1|\eta)'\widetilde{W}(\theta_1|\eta)^{-1}\widetilde{H}(\theta_1|\eta)\xi\right]$$

for $1 \le \alpha < 2$*, where* $\widetilde{V}(\theta|\eta)$ *is defined by replacing* $\widetilde{f}(\omega)$ *of* $\widetilde{V}(\theta)$ *in Theorem 3.1 with* $\widetilde{f}(\omega|\eta)$*.*

Remark 4.3. *We can write*

$$\begin{aligned}
&2\xi'\widetilde{H}(\theta_1|\eta)'\widetilde{W}(\theta_1|\eta)^{-1}\widetilde{V}(\theta_1|\eta)\\
&= \sum_{i=1}^{p} L_i \widetilde{V}(\theta_1|\eta)_i\\
&= \sum_{l=1}^{\infty} \frac{S(l)}{S(0)} \left\{ \sum_{i=1}^{p} \frac{L_i}{\pi} \int_{-\pi}^{\pi} \left.\frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^i}\right|_{\theta=\theta_1} \widetilde{f}(\omega)\cos(l\omega)d\omega \right\},
\end{aligned}$$

and this random variable bas the same distribution as

$$\left\{ \sum_{l=1}^{\infty} \left| \sum_{i=1}^{p} \frac{L_i}{\pi} \int_{-\pi}^{\pi} \left.\frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^i}\right|_{\theta=\theta_1} \widetilde{f}(\omega)\cos(l\omega)d\omega \right|^{\alpha} \right\}^{1/\alpha} \frac{S(1)}{S(0)}, \qquad (4.4)$$

where L_i *is the ith element of of* $2\xi'\widetilde{H}(\theta_1|\eta)'\widetilde{W}(\theta_1|\eta)^{-1}$*.* (4.4) *implies that the random variable* $2\xi'\widetilde{H}(\theta_1|\eta)'\widetilde{W}(\theta_1|\eta)^{-1}\widetilde{V}(\theta_1|\eta)$ *is symmetric around 0. Moreover, the quantity* $-\xi'\widetilde{H}(\theta_1|\eta)'\widetilde{W}(\theta_1|\eta)^{-1}\widetilde{H}(\theta_1|\eta)\xi$ *is negative. So the misclassification probability by the empirical likelihood ratio classification statistic is smaller than* $1/2$*.*

4.4 Numerical examples

In this section, we carry out simulation studies under practical situations; that is, we assume that the true normalized power transfer functions and pivotal quantities associated with categories are unknown. In such case, we estimate the normalized power transfer function using the average sample normalized power transfer function for each category, and calculate empirical detection probabilities of plug-in statistics. As the parametric normalized power transfer function, we consider ARMA(1,1) and MA(100) type ones in succession.

4.4.1 Example 5: Stable ARMA(1,1) model

We consider the classification problem of ARMA(1,1) process, and let the normalized power transfer functions associated with Π_1 and Π_2 be

$$\Pi_1 : \widetilde{f}_1(\omega) = \widetilde{f}(\omega|\eta) \qquad \Pi_2 : \widetilde{f}_2(\omega) = \widetilde{f}\left(\omega\left|\eta + x_n^{-1}\xi\right.\right)$$

where the parametric normalized power transfer function is defined as

$$\widetilde{f}(\omega|\eta) = \frac{(1+a^2) + 2a\cos(\omega)}{(1+b^2) - 2b\cos(\omega)} \frac{1-b^2}{1+a^2+2ab}, \quad \eta = (a, b)'$$

and $\xi = (1, 1)'$. In practice, however, it is not natural to suppose that we know the true normalized power transfer functions. So suppose that we have 10 training samples with length $n = 512$ from each category in advance, and estimate the group normalized power transfer function using the average sample normalized power transfer function for each group. First, we calculate the smoothed self-normalized power transfer function $\widehat{f}_i^{(l)}(\omega)$ with $w_n(k) = (2m-1)^{-1}$ and $m = [\sqrt{n}]$ (see Remark 3.2) from $X_i^{(l)}(1), \cdots, X_i^{(l)}(n)$ (the lth observed stretch in category Π_i). Then, the group normalized power transfer function

$$\widehat{f}_i(\omega) = \frac{1}{10} \sum_{l=1}^{10} \widehat{f}_i^{(l)}(\omega), \;\; i = 1 \text{ and } 2,$$

is a consistent estimator of $\widetilde{f}_i(\omega)$ for each $\omega \in [-\pi, \pi]$. So for another observed stretch $X(1), \cdots, X(n)$, we use the plug-in version $\widetilde{I}(\widehat{f}_1, \widehat{f}_2)$ as the classification statistic. On the other hand, the pivotal quantity θ_0 coincides with $\rho_X(h)$ when we use $g(\omega; \theta) = |1 - \exp(\mathrm{i}h\omega)|^{-2}$ as a score function. To estimate the pivotal quantities under each category, we can use

$$\widehat{\theta}_i = \frac{1}{10} \sum_{l=1}^{10} \widehat{\theta}_i^{(l)}, \quad i = 1 \text{ and } 2,$$

where

$$\widehat{\theta}_i^{(l)} = \frac{\sum_{t=1}^{n-1} X_i^{(l)}(t) X_i^{(l)}(t+h)}{\sum_{t=1}^{n} X_i^{(l)}(t)^2},$$

which is the sample autocorrelation for the lth realization in category Π_i. Then, we can classify another observed stretch $X(1), \cdots, X(n)$ into one of Π_1 and Π_2 by the plug-in version $\widetilde{\mathrm{ELR}}(\widehat{\theta}_1, \widehat{\theta}_2)$.

When we use the empirical likelihood ratio classification statistic, we can flexibly choose pivotal quantities or score functions to focus on. In this section, we use the following three score functions:

(i). $g(\omega; \theta) = |1 - \theta \exp(\mathrm{i}\omega)|^{-2}$

(ii). $g(\omega;\theta) = |1 - \theta \exp(2\mathrm{i}\omega)|^{-2}$

(iii). $g(\omega;\theta) = \left(\left|1 - \theta^1 \exp(\mathrm{i}\omega)\right|^2 + \left|1 - \theta^2 \exp(2\mathrm{i}\omega)\right|^2\right)^{-1}$

and the corresponding pivotal quantities are given as

(i). $\theta_0 = \rho_X(1)$

(ii). $\theta_0 = \rho_X(2)$

(iii). $\theta_0 = (\rho_X(1), \rho_X(2))'$.

Under the setting above, we use the Monte Carlo simulation to calculate detection probabilities by both classification statistics for $n = 512$. First, using the training sets

$$\left\{X_i^{(l)}(t) : X_i^{(l)}(t) \text{ is generated from } \Pi_i,\ t = 1, \cdots, n\right\},\ l = 1, \cdots, 10,\ i = 1 \text{ and } 2,$$

we calculate $\widehat{f}_1(\omega)$, $\widehat{f}_2(\omega)$, $\widehat{\theta}_1$ and $\widehat{\theta}_2$ for fixed a, b and α. Second, we generate a test set $\{X(t) : t = 1, \cdots, n\}$ from Π_1, and calculate $\widetilde{I}(\widehat{f}_1, \widehat{f}_2)$ and $\widetilde{\mathrm{ELR}}(\widehat{\theta}_1, \widehat{\theta}_2)$. Iterating this procedure for 1000 times, we obtain the frequencies of $\widetilde{I}(\widehat{f}_1, \widehat{f}_2) > 0$ and $\widetilde{\mathrm{ELR}}(\widehat{\theta}_1, \widehat{\theta}_2) > 0$. Tables 4.2-4.4 show the empirical detection probabilities of $\widetilde{I}$ and $\widetilde{\mathrm{ELR}}$ for various a, b and α.

Tables 4.2-4.4 are about here.

The tables above imply that the empirical likelihood ratio classification statistic works better than the Whittle likelihood ratio type. Furthermore, except for a few cases, the empirical likelihood classification statistics give higher detection probabilities for small α (Table 4.2), and when the process has a near unit root, it is better to consider our approach (see the case when $b = 0.9$ in Tables 4.2-4.4).

4.4.2 Example 6: Stable MA(100) model

Next, let the parametric normalized power transfer function be

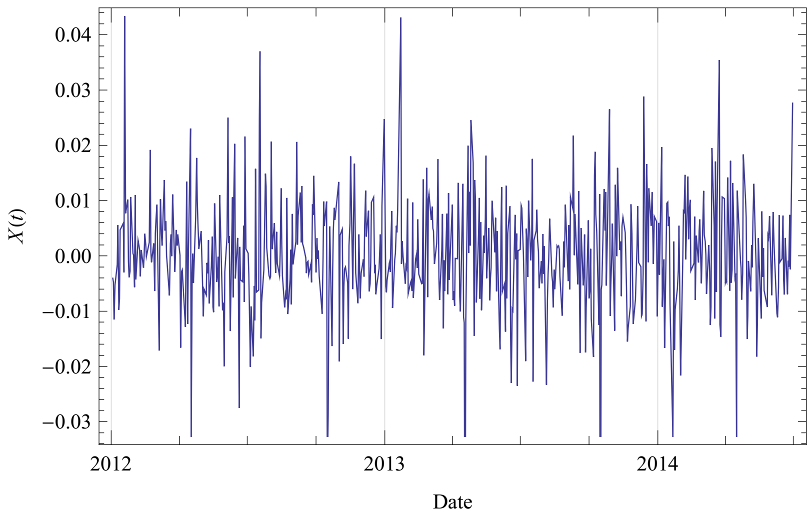
$$\widetilde{f}(\omega|\eta) = \frac{\left|\sum_{j=0}^{100} \psi_j(b) \exp(\mathrm{i}j\omega)\right|^2}{\sum_{j=0}^{100} \psi_j(b)^2}, \quad \eta = b$$

and the coefficients $\{\psi_j(b) : j \in \mathbb{N}\}$ is defined as (3.11) (and we do not omit argument $\eta = b$ to emphasize that b is the contiguous parameter). In this case, the normalized power transfer function associated with the contiguous hypothesis Π_2 is $f(\omega|\eta + x_n^{-1})$, and we carry out the same procedure as Example 5.

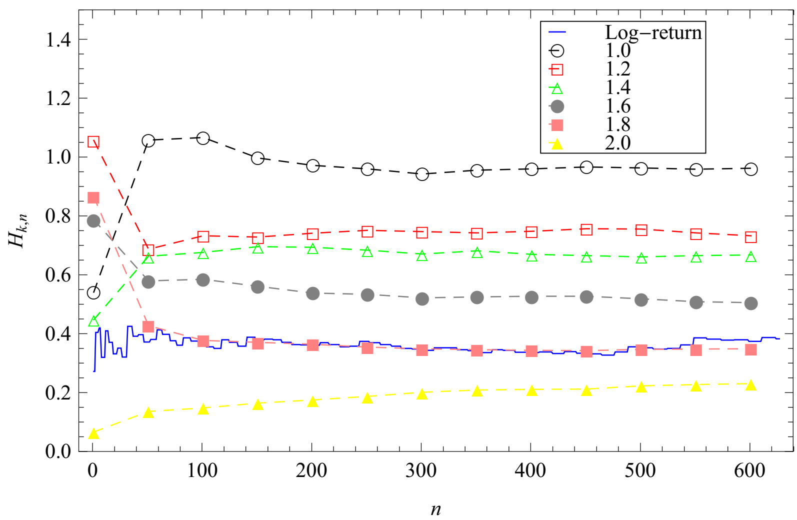
Tables 4.5-4.7 are about here.

The empirical likelihood method performs uniformly better than the existing method in this case. Moreover, it seems to be not always appropriate to use much information on the process. We sometimes obtain higher detection probabilities when we use either $\rho_X(1)$ or $\rho_X(2)$ than both $\rho_X(1)$ and $\rho_X(2)$. Thus, it is significant to focus on various important quantities of the process when we carry out the discriminant analysis.

Figure 4.1: The log-stock return process of IBM (from Jan. 1, 2012 to Jun. 30, 2014) and its hill-plot

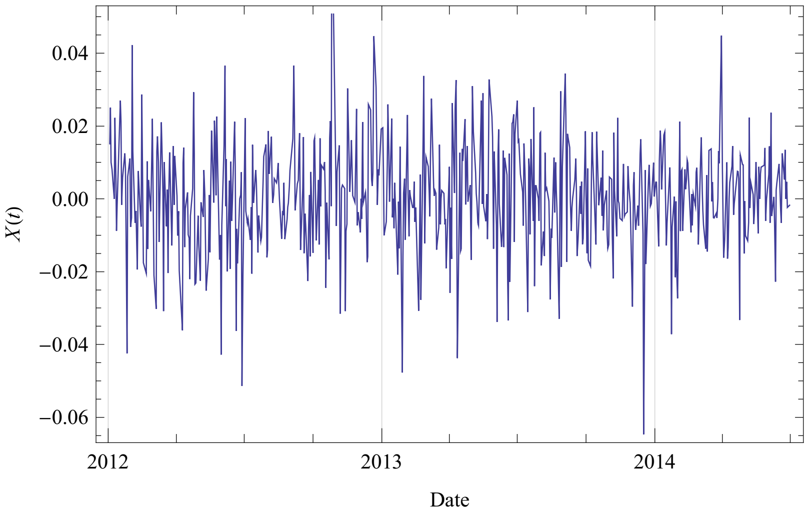


(a) Log-stock return $X(t) = \log\{P_{\mathrm{I}}(t+1)/P_{\mathrm{I}}(t)\}$

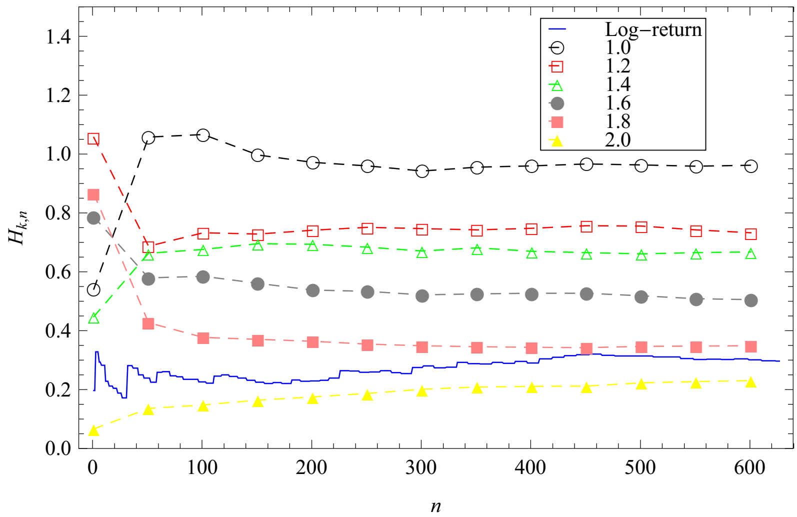


(b) Hill-plot for $X(t)$

Figure 4.2: The log-stock return process of Ford (from Jan. 1, 2012 to Jun. 30, 2014) and its hill-plot



(a) Log-stock return $X(t) = \log\{P_{\mathrm{F}}(t+1)/P_{\mathrm{F}}(t)\}$



(b) Hill-plot for $X(t)$

Table 4.1: $\widehat{\rho}_{n,X}(1)$ and $\widehat{\rho}_{n,X}(2)$ of Hewlett Packard, IBM and Ford's log-stock return processes

	Hewlett Packard	IBM	Ford
$\widehat{\rho}_{n,X}(1)$	-0.0305907	0.0462158	0.1227030
$\widehat{\rho}_{n,X}(2)$	0.0393583	-0.0060251	-0.0573039

Table 4.2: Empirical detection probabilities by $\widetilde{\text{ELR}}(\widehat{\theta}_1, \widehat{\theta}_2)$ and $\widetilde{I}(\widehat{f}_1, \widehat{f}_2)$ for stable ARMA(1,1) processes ($\alpha = 1.0$)

	$\widetilde{\text{ELR}}$			$\widetilde{I}$
(a, b)	(i)	(ii)	(iii)	
(0.1,0.1)	0.949	0.866	0.947	0.973
(0.1,0.5)	0.966	0.946	0.970	0.949
(0.1,0.9)	0.987	0.943	0.999	0.977
(0.5,0.1)	0.924	0.736	0.959	0.826
(0.5,0.5)	0.912	0.891	0.919	0.841
(0.5,0.9)	0.947	0.972	0.897	0.829
(0.9,0.1)	0.922	0.863	0.943	0.620
(0.9,0.5)	0.688	0.912	0.963	0.973
(0.9,0.9)	0.990	0.985	0.993	0.174

Table 4.3: Empirical detection probabilities by $\widetilde{\text{ELR}}(\widehat{\theta}_1, \widehat{\theta}_2)$ and $\widetilde{I}(\widehat{f}_1, \widehat{f}_2)$ for stable ARMA(1,1) processes ($\alpha = 1.5$)

	$\widetilde{\text{ELR}}$			$\widetilde{I}$
(a, b)	(i)	(ii)	(iii)	
(0.1,0.1)	0.897	0.511	0.916	0.892
(0.1,0.5)	0.916	0.681	0.973	0.878
(0.1,0.9)	0.971	0.939	0.981	0.936
(0.5,0.1)	0.877	0.782	0.896	0.907
(0.5,0.5)	0.688	0.625	0.798	0.808
(0.5,0.9)	0.379	0.232	0.992	0.638
(0.9,0.1)	0.641	0.666	0.556	0.790
(0.9,0.5)	0.644	0.626	0.647	0.208
(0.9,0.9)	0.304	0.203	0.958	0.302

Table 4.4: Empirical detection probabilities by $\widetilde{\text{ELR}}(\widehat{\theta}_1, \widehat{\theta}_2)$ and $\widetilde{I}(\widehat{f}_1, \widehat{f}_2)$ for stable ARMA(1,1) processes ($\alpha = 1.9$)

	$\widetilde{\text{ELR}}$			$\widetilde{I}$
(a, b)	(i)	(ii)	(iii)	
(0.1,0.1)	0.780	0.487	0.815	0.832
(0.1,0.5)	0.886	0.701	0.906	0.894
(0.1,0.9)	0.709	0.429	0.999	0.893
(0.5,0.1)	0.676	0.645	0.662	0.734
(0.5,0.5)	0.637	0.560	0.760	0.679
(0.5,0.9)	0.597	0.387	0.998	0.881
(0.9,0.1)	0.534	0.518	0.587	0.72
(0.9,0.5)	0.484	0.490	0.497	0.222
(0.9,0.9)	0.592	0.444	0.996	0.805

Table 4.5: Empirical detection probabilities by $\widetilde{\mathrm{ELR}}(\widehat{\theta}_1, \widehat{\theta}_2)$ and $\widetilde{I}(\widehat{f}_1, \widehat{f}_2)$ for stable MA(100) processes ($\alpha = 1.0$)

	$\widetilde{\mathrm{ELR}}$			$\widetilde{I}$
b	(i)	(ii)	(iii)	
0.1	0.916	0.661	0.913	0.872
0.5	0.854	0.720	0.890	0.835
0.9	0.594	0.416	0.572	0.479

Table 4.6: Empirical detection probabilities by $\widetilde{\mathrm{ELR}}(\widehat{\theta}_1, \widehat{\theta}_2)$ and $\widetilde{I}(\widehat{f}_1, \widehat{f}_2)$ for stable MA(100) processes ($\alpha = 1.5$)

	$\widetilde{\mathrm{ELR}}$			$\widetilde{I}$
b	(i)	(ii)	(iii)	
0.1	0.678	0.593	0.679	0.667
0.5	0.790	0.510	0.841	0.831
0.9	0.670	0.563	0.674	0.667

Table 4.7: Empirical detection probabilities by $\widetilde{\mathrm{ELR}}(\widehat{\theta}_1, \widehat{\theta}_2)$ and $\widetilde{I}(\widehat{f}_1, \widehat{f}_2)$ for stable MA(100) processes ($\alpha = 1.9$)

	$\widetilde{\mathrm{ELR}}$			$\widetilde{I}$
b	(i)	(ii)	(iii)	
0.1	0.770	0.511	0.798	0.761
0.5	0.670	0.558	0.647	0.664
0.9	0.554	0.486	0.446	0.538

5 Generalized method of moments for symmetric α-stale processes

5.1 Introduction

The GMM is one of the most popular tools in econometrics, and the method has been applied in variety of fields. In i.i.d. case, Newey and McFadden [44] gave a unified view for the estimation problems based on the class of extremum estimators which contains the GMM estimators and the maximum likelihood estimators as special cases. For dependent data, Hansen [17] introduced the GMM estimator for time series models with the finite second moment, and consider the estimation problem of parameters of over-identified moment restriction models in time domain. Subsequently, Kakizawa [26] extended Hansen's approach to frequency domain. Kakizawa [26] worked with the over-identified spectral restriction model and proposed the frequency domain GMM estimator. In the paper, the consistency and asymptotic normality of the estimator were shown. Furthermore, Kakizawa [26] gave the optimal weighting matrix of the GMM estimator, and showed that we can construct an asymptotic optimal estimator without assuming that the true model is known.

One of the advantages of considering the spectral restriction is that we can naturally extend results for second-order stationary processes toward stable processes, as seen in the previous chapters. In Section 5.2, we define the frequency domain GMM estimator for the stable process as Kakizawa [26] did for second-order stationary processes. We show the consistency and asymptotic distribution of the GMM estimator in Section 5.3. It is shown that the limit distribution is expressed as a sum of stable random variables, and depends on a weighting matrix, which defines the objective function of the estimator. The optimality of the GMM estimator is also discussed based on a sort of covariance matrix. Section 5.4 gives numerical studies, and we observe the effect of the weighting matrices.

5.2 Fundamental settings

Current chapter considers the following over-identified spectral restriction model

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} G(\omega;\theta_0)\widetilde{f}(\omega)d\omega = 0_m, \tag{5.1}$$

where $G(\omega;\theta) = (G_1(\omega;\theta),\cdots,G_m(\omega;\theta))'$ is an $m\times 1$ vector-valued function, called a score function hereafter. In particular, θ_0 in (5.1) is called an over-identified pivotal quantity of the process, because the number of restriction m can be greater than the dimension p of θ. Given the over-identified spectral restriction (5.1) and an observed stretch $X(1), \cdots, X(n)$ from the stable process (3.1), the frequency domain GMM estimator is defined as

$$\widehat{\theta}_{\mathrm{GMM}} = \arg\max_{\theta \in \Theta} q_n(\theta),$$

where

$$q_n(\theta) = -\widehat{R}_n(\theta)' \widehat{W}_n \widehat{R}_n(\theta),$$
$$\widehat{R}_n(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\omega; \theta) \widetilde{I}_{n,X}(\omega) d\omega$$

and $\widehat{W}_n$ is an $m \times m$ nonnegative definite matrix, called a weighting matrix. Note that $\widehat{W}_n$ can be a sequence of random matrices, and we assume the following condition on the weighting matrix.

Assumption 5.1. $\widehat{W}_n \xrightarrow{\mathcal{P}} W$, *where* W *is an* $m \times m$ *positive definite matrix.*

Remark 5.1. *In the proof for Theorem 3.1, we make use of the following asymptotic expansion of the empirical likelihood ratio statistic:*

$$\frac{2}{n} \log \widetilde{r}_n(\theta_0) = -\widehat{R}_n(\theta_0)' \widetilde{S}_n(\theta_0)^{-1} \widehat{R}_n(\theta_0) + o_p(1),$$

and it is shown that

$$\widetilde{S}_n(\theta_0)^{-1} \xrightarrow{\mathcal{P}} \left[\frac{1}{\pi} \int_{-\pi}^{\pi} G(\omega; \theta_0)' G(\omega; \theta_0) \widetilde{f}(\omega)^2 d\omega \right]^{-1} = W_{\mathrm{EL}} \quad (say).$$

From this expansion, we can regard W_{EL} *as the weighting matrix for the maximum empirical likelihood estimator as GMM.*

5.3 Main results

Now let us see the limit behavior of the frequency domain GMM estimator to the stable process. To simplify notation, define an $m \times p$ matrix

$$Q(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial G(\omega; \theta)}{\partial \theta'} \widetilde{f}(\omega) d\omega.$$

In addition, the conditions summarized in the next assumption are required.

Assumption 5.2.

(i). *There is a unique solution* $\theta_0 \in \mathrm{Int}(\Theta)$ *to* (5.1), *where* Θ *is a compact subset of* $\mathbb{R}^p$.

(ii). $G(-\omega;\theta) = G(\omega;\theta)$ *for all* $\theta \in \Theta$.

(iii). $G(\omega;\theta)$ *is continuously differentiable with respect to* $\theta \in \Theta$ *for all* $\omega \in [-\pi,\pi]$.

(iv). $Q(\theta_0)$ *has full-column rank.*

(v). *For each* $i = 1,2$ *and* $j = 1,\cdots,m$, *there exist* $c < \infty$ *and* $\beta > 0$ *such that*

$$\sup_{\omega\in[-\pi,\pi]} \sup_{\substack{\theta,\widetilde{\theta}\in\Theta,\\ \theta\neq\widetilde{\theta}}} \frac{\left\| h_j^{(i)}(\omega;\theta) - h_j^{(i)}(\omega;\widetilde{\theta}) \right\|_E}{\left\| \theta - \widetilde{\theta} \right\|_E^{\beta}} \leq c,$$

where $h_j^{(1)}(\omega;\theta) = G_j(\omega;\theta)$ *and* $h_j^{(2)}(\omega;\theta) = \partial G_j(\omega;\theta)/\partial\theta$.

(vi). *There exists* $\mu \in (0,\alpha)$ *such that*

$$\sum_{k=1}^{\infty} \left| \int_{-\pi}^{\pi} \sum_{i=1}^{m} G_i(\omega;\theta_0) \widetilde{f}(\omega) \cos(k\omega) d\omega \right|^{\mu} < \infty.$$

(i) and (iv) of Assumption 5.2 are needed so that the limit distribution of the GMM estimator has an appropriate property. We shall give a counter example of Theorem 5.1 below when θ_0 belongs to the boundary of Θ. Especially, the latter condition is required to guarantee that the limit distribution does not degenerate. Conditions (ii) and (vi) are imposed to use Lemma 3.2. Moreover, we put condition (v) to ensure the consistency of the estimator. Then, we can show the limit distribution of the frequency domain GMM estimator for the stable process as follows.

Theorem 5.1. *Suppose that* $\{X(t) : t \in \mathbb{Z}\}$ *is generated as* (3.1) *with* $\alpha \in (0,2)$, *and Assumptions 3.1, 5.1 and 5.2 hold.*

(i). $\widehat{\theta}_{\mathrm{GMM}} \xrightarrow{\mathcal{P}} \theta_0$ *as* $n \to \infty$.

(ii). $x_n \left(\widehat{\theta}_{\mathrm{GMM}} - \theta_0 \right) \xrightarrow{\mathcal{L}} L(W)\widetilde{V}(\theta_0)$ *as* $n \to \infty$. *Here* $L(W)$ *is a* $p \times m$ *matrix defined as* $L(W) = \{Q(\theta_0)'WQ(\theta_0)\}^{-1} Q(\theta_0)'W$ *and* $\widetilde{V}(\theta)$ *is an* $m \times 1$ *random vector defined by replacing* $\partial g(\omega;\theta)^{-1}/\partial\theta^i$ *in* (3.7) *with* $G_i(\omega;\theta)$.

Remark 5.2. *The condition (i) of Assumption 5.1 is essential to guarantee that the limit distribution of the GMM estimator becomes symmetric. We impose a representative counter example here. Suppose that* $\{X(t) : t \in \mathbb{Z}\}$ *is generated by stable MA(1) model* $X(t) = Z(t) + aZ(t-1)$ *with* $a \in \mathbb{R}^1$*. Then, the autocorrelations are given as* $\rho_X(1) = a/(1+a^2)$ *and* $\rho_X(h) = 0$ *for any* $h \geq 2$*. Therefore, one can estimate* $\theta_0 = \rho_X(1)$ *by the GMM estimator based on the score function*

$$G(\omega;\theta) = (\theta - \cos(\omega), \cos(l\omega))'$$

with $l \geq 2$ *and the identity weighting matrix* $I_{2\times 2}$*. If we set* $\Theta = [0,1]$ *and* $a = 0$*, then* $\theta_0 = 0 \notin \mathrm{Int}(\Theta)$*. A simple calculation yields that the GMM estimator is evaluated as* $\widehat{\theta}_{\mathrm{GMM}} = \max\{\widehat{\rho}_{n,X}(1), 0\}$*, and the limit of the distribution function is given as*

$$\lim_{n\to\infty} \Pr\left\{x_n\left(\widehat{\theta}_{\mathrm{GMM}} - \theta_0\right) \leq y\right\} = \begin{cases} 0 & (y < 0) \\ 1/2 & (y = 0) \\ \Pr\{-S(1)/S(0) \leq y\} & (y > 0) \end{cases}.$$

Thus, the limit distribution of $x_n(\widehat{\theta}_{\mathrm{GMM}} - \theta_0)$ *has a spike at the origin, and the conclusion (ii) in Theorem 5.1 fails. The general theory for the boundary case is quite complicated, and we omit to state the details here.*

We next focus on the asymptotic optimality of $\widehat{\theta}_{\mathrm{GMM}}$ based on a sort of covariance matrix. Set $\widetilde{\Omega}(W) = L(W)\widetilde{\Sigma}L(W)'$, where

$$\widetilde{\Sigma} = \sum_{l=1}^{\infty} v(l)v(l)',$$
$$v(l) = \frac{1}{\pi}\int_{-\pi}^{\pi} G(\omega;\theta_0)\widetilde{f}(\omega)\cos(l\omega)d\omega. \tag{5.2}$$

The matrix $\widetilde{\Omega}(W)$ has the following interpretation; let $\{U_i : i = 1, 2, \cdots\}$ be the sequence of i.i.d. random variables with zero mean and $\mathrm{E}[U_1^2] = 1$, and U be an $m \times 1$ random vector defined as $U = \sum_{l=1}^{\infty} v(l)U_l$. Then, the covariance matrix of $L(W)U$ coincides with $\widetilde{\Omega}(W)$ defined above. Comparing U with $\widetilde{V}(\theta_0)$, this matrix can be regarded as a sort of the covariance matrix of $L(W)\widetilde{V}(\theta_0)$. The lower bound of $\widetilde{\Omega}(W)$ is then given by the following theorem.

Theorem 5.2. *Suppose that all assumptions in Theorem 5.1 hold. Then,*

(i). $\widetilde{\Sigma}^{-1} = W_{\mathrm{EL}}$.

(ii). *For any* W, $\widetilde{\Omega}(W) - \widetilde{\Omega}(W_{\mathrm{EL}})$ *is nonnegative definite.*

Recalling Remark 5.1, Theorem 5.2 describes the asymptotic optimality of the empirical likelihood method in the class of GMM estimators.

5.4 Numerical examples

This section provides numerical studies, and make a comparison between some GMM estimators. We follow the examples introduced by Kakizawa [26]: the over-identified AR(1) and MA(1) models.

5.4.1 Example 7: Stable AR(1) model

Let $\{X(t) : t \in \mathbb{Z}\}$ be a stable AR(1) process with normalized power transfer function

$$\widetilde{f}(\omega) = \frac{1 - b^2}{|1 - b\exp(\mathrm{i}\omega)|^2}, \quad |b| < 1.$$

The autocorrelation of the process at lag l is given by $\rho_X(l) = b^l$, and this model satisfies the over-identified restriction (5.1) with the score function

$$G(\omega; \theta) = (\theta - \cos(\omega), \cdots, \theta^m - \cos(m\omega))'. \tag{5.3}$$

In this case, we estimate $\theta_0 = \rho_X(1)$ based on the over-identified score function (5.3).

As we discuss in the previous section, the optimal weighting matrix is given as

$$W_{\mathrm{EL}} = \left[\frac{1}{\pi}\int_{-\pi}^{\pi} G(\omega; \theta_0) G(\omega; \theta_0)' \widetilde{f}(\omega)^2 d\omega\right]^{-1}.$$

In practice, however, we know neither true $\widetilde{f}(\omega)$ and θ_0, so we should estimate them from observations. The consistent estimator of W_{EL} is given as

$$\widehat{W}_{\mathrm{EL}} = \left[\frac{2}{n}\sum_{t=1}^{n} G(\lambda_t; \widehat{\theta}^{(1)}_{\mathrm{GMM}}) G(\lambda_t; \widehat{\theta}^{(1)}_{\mathrm{GMM}})' \widetilde{J}_{n,X}(\lambda_t)^2\right]^{-1},$$

where $\widetilde{J}_{n,X}(\omega)$ is the smoothed self-normalized periodogram with the same wighting function as Example 5 and 6. $\widehat{\theta}^{(1)}_{\mathrm{GMM}}$ is the GMM estimator with the identity weighting matrix. That is, we can construct the 2-step GMM estimator, which is asymptotically equivalent to the GMM estimator with the theoretical optimal weighting matrix W_{EL}.

Hereafter, use the following three weighting matrices:

(i). $W_1 = I_{m\times m}$

(ii). $W_2 = W_{\mathrm{EL}}$

(iii). $W_3 = \widehat{W}_{\mathrm{EL}}$

and the GMM estimator based on W_i is denoted by $\widehat{\theta}^{(i)}_{\mathrm{GMM}}$ $(i = 1, 2, 3)$. In particular, we adopt the notations $\widehat{\theta}_{\mathrm{EL}} = \widehat{\theta}^{(2)}_{\mathrm{GMM}}$ and $\widehat{\theta}^{*}_{\mathrm{EL}} = \widehat{\theta}^{(3)}_{\mathrm{GMM}}$.

With the settings above, we generate 1000 samples from the stable AR(1) model of size n for various b, α and n, and estimate $\theta_0 = \rho_X(1)$ by the GMM estimators. Then, we obtain the empirical bias

$$\mathrm{EB}^{(i)} = \frac{1}{1000} \sum_{j=1}^{1000} B_j^{(i)}$$

and the root mean square error

$$\mathrm{RMSE}^{(i)} = \frac{1}{1000} \sum_{j=1}^{1000} \left| B_j^{(i)} \right|$$

for $i = 1, 2, 3$, where $B_j^{(i)}$ is equal to the quantity $\widehat{\theta}^{(i)}_{\mathrm{GMM}} - \theta_0$ for the jth sample. Adding to the three estimators, we also calculate the empirical bias and root mean square error for the sample autocorrelation $\widehat{\rho}_{n,X}(1)$. Tables 5.1-5.3 show $\mathrm{EB}^{(i)}$ and $\mathrm{RMSE}^{(i)}$ of the four estimators for $n = 128$ and 512. Tables 5.1, 5.2 and 5.3 correspond to the case where $m = 2, 5$ and 8, respectively.

Tables 5.1-5.3 are about here.

5.4.2 Example 8: Stable MA(1) model

Next, we consider to estimate $\theta_0 = \rho_X(1)$ of a stable MA(1) process with normalized power transfer function

$$f(\omega) = \frac{|1 + a \exp(\mathrm{i}\omega)|^2}{1 + a^2}.$$

Since the autocorrelation at lag h of the MA(1) process is zero for any $h \geq 2$, the model satisfies the over-identified restriction (5.1) with the score function

$$G(\omega; \theta) = (\theta - \cos(\omega), \cos(2\omega), \cdots, \cos(m\omega))'.$$

By the same procedure as Example 7, we calculate the empirical biases and root mean square errors of the GMM estimators. Note that $\widehat{\theta}^{(1)}_{\mathrm{GMM}}$ coincides with the sample autocorrelation at lag 1 in this case. In fact, it is easily seen that

$$\widehat{R}_n(\theta) = \begin{pmatrix} \theta - \widehat{\rho}_{n,X}(1) \\ \widehat{\rho}_{n,X}(2) \\ \vdots \\ \widehat{\rho}_{n,X}(m) \end{pmatrix},$$

so the objective function based on $I_{m\times m}$ is given as

$$q_n(\theta) = -\widehat{R}_n(\theta)' I_{m\times m} \widehat{R}_n(\theta)$$
$$= -\{\theta - \widehat{\rho}_{n,X}(1)\}^2 - \sum_{i=2}^{m} \widehat{\rho}_{n,X}(i)^2. \qquad (5.4)$$

Since the second term of (5.4) is independent of θ, the maximizer is $\widehat{\theta}^{(1)}_{\mathrm{GMM}} = \widehat{\rho}_{n,X}(1)$. We omit the results for $\widehat{\rho}_{n,X}(1)$ in the following tables.

Tables 5.4-5.6 are about here.

From Tables 5.1-5.3, it is seen that the sample autocorrelation method has advantage in the sense of the root mean square errors when we work with AR(1) model and α is large. However, these results do not imply that the GMM estimator is worth than $\widehat{\rho}_{n,X}(1)$, since the class of GMM estimators contains $\widehat{\rho}_{n,X}(1)$ as a special case; the GMM estimator with $G(\omega, \theta) = \theta - \cos(\omega)$ (and $m = p = 1$). On the other hand, If we focus on the empirical bias, the inference by the GMM estimator with optimal weighting matrix (or its estimator) is better than those by the sample autocorrelation method when we deal with AR(1) model. In the case of MA(1) model, none of the estimators give the smallest empirical bias uniformly. However, the GMM estimator based on the empirical likelihood weight gives smaller root mean square errors than $\widehat{\rho}_{n,X}(1)$ uniformly (Tables 5.4-5.6).

In both cases, the root mean square errors decrease as the index α decrease. These phenomena are quite natural, since the larger α becomes, the slower the divergence of x_n becomes.

Lastly, we consider the results in view of the number of restrictions m. We observe that the root mean square errors of $\widehat{\theta}_{\mathrm{EL}}$ and $\widehat{\theta}^{*}_{\mathrm{EL}}$ tend to decrease as m increases, while the increment of m does not seem to affect the goodness of $\widehat{\rho}_{n,X}(1)$ and $\widehat{\theta}^{(1)}_{\mathrm{GMM}}$. Thus, from the numerical examples, we can see that the GMM estimator with optimal weighting matrix has appropriate properties.

Table 5.1: Empirical biases (and root mean square errors) of $\widehat{\rho}_{n,X}(1)$ and $\widehat{\theta}^{(i)}_{\mathrm{GMM}}$ ($i = 1, 2, 3$) for $\rho_X(1)$ of the stable AR(1) model

$m = 2$, $n = 128$									
α	b	$\widehat{\rho}_{n,X}(1)$		$\widehat{\theta}^{(1)}_{\mathrm{GMM}}$		$\widehat{\theta}_{\mathrm{EL}}$		$\widehat{\theta}^{*}_{\mathrm{EL}}$	
	0.1	−0.00135	(0.01536)	−0.00109	(0.01615)	−0.00091	(0.01480)	−0.00045	(0.01660)
0.5	0.5	−0.00879	(0.01990)	−0.00990	(0.02282)	−0.00863	(0.01986)	−0.00838	(0.02060)
	0.9	−0.01788	(0.02493)	−0.01869	(0.02584)	−0.01716	(0.02420)	−0.01749	(0.02459)
	0.1	−0.00155	(0.03792)	−0.00051	(0.03950)	−0.00096	(0.03756)	0.00032	(0.03974)
1.0	0.5	−0.01000	(0.03458)	−0.01078	(0.03852)	−0.01000	(0.03508)	−0.00862	(0.03519)
	0.9	−0.01887	(0.03017)	−0.01995	(0.03128)	−0.01842	(0.02981)	−0.01846	(0.02979)
	0.1	−0.00260	(0.05550)	−0.00252	(0.05625)	−0.00276	(0.05520)	−0.00155	(0.05664)
1.5	0.5	−0.01022	(0.05250)	−0.01134	(0.05846)	−0.01001	(0.05322)	−0.00769	(0.05370)
	0.9	−0.02072	(0.03409)	−0.02148	(0.03491)	−0.02161	(0.03501)	−0.02006	(0.03352)
	0.1	−0.00727	(0.06799)	−0.00738	(0.06899)	−0.00700	(0.06825)	−0.00667	(0.06965)
1.9	0.5	−0.01269	(0.06074)	−0.01385	(0.06510)	−0.01330	(0.06189)	−0.00917	(0.06141)
	0.9	−0.02159	(0.03572)	−0.02217	(0.03622)	−0.02242	(0.03650)	−0.02095	(0.03523)
$m = 2$, $n = 512$									
α	b	$\widehat{\rho}_{n,X}(1)$		$\widehat{\theta}^{(1)}_{\mathrm{GMM}}$		$\widehat{\theta}_{\mathrm{EL}}$		$\widehat{\theta}^{*}_{\mathrm{EL}}$	
	0.1	−0.00011	(0.00343)	0.00014	(0.00398)	−0.00004	(0.00334)	0.00003	(0.00366)
0.5	0.5	−0.00248	(0.00665)	−0.00305	(0.00764)	−0.00263	(0.00656)	−0.00273	(0.00679)
	0.9	−0.00796	(0.01093)	−0.00864	(0.01164)	−0.00704	(0.01000)	−0.00791	(0.01095)
	0.1	−0.00157	(0.01564)	−0.00130	(0.01617)	−0.00142	(0.01530)	−0.00145	(0.01591)
1.0	0.5	−0.00153	(0.01408)	−0.00187	(0.01612)	−0.00148	(0.01429)	−0.00146	(0.01452)
	0.9	−0.00546	(0.01287)	−0.00551	(0.01301)	−0.00558	(0.01304)	−0.00535	(0.01290)
	0.1	0.00026	(0.02670)	0.00016	(0.02724)	−0.00001	(0.02677)	0.00026	(0.02675)
1.5	0.5	−0.00180	(0.02299)	−0.00259	(0.02578)	−0.00163	(0.02308)	−0.00175	(0.02328)
	0.9	−0.00454	(0.01410)	−0.00492	(0.01438)	−0.00449	(0.01403)	−0.00447	(0.01414)
	0.1	−0.00262	(0.03379)	−0.00209	(0.03495)	−0.00239	(0.03387)	−0.00250	(0.03393)
1.9	0.5	−0.00282	(0.02923)	−0.00325	(0.03213)	−0.00276	(0.02926)	−0.00242	(0.02925)
	0.9	−0.00598	(0.01630)	−0.00611	(0.01675)	−0.00591	(0.01627)	−0.00560	(0.01641)

Table 5.2: Empirical biases (and root mean square errors) of $\widehat{\rho}_{n,X}(1)$ and $\widehat{\theta}^{(i)}_{\mathrm{GMM}}$ ($i = 1,2,3$) for $\rho_X(1)$ of the stable AR(1) model

$m=5$, $n=128$									
α	b	$\widehat{\rho}_{n,X}(1)$		$\widehat{\theta}^{(1)}_{\mathrm{GMM}}$		$\widehat{\theta}_{\mathrm{EL}}$		$\widehat{\theta}^{*}_{\mathrm{EL}}$	
	0.1	−0.00244	(0.01085)	−0.00125	(0.01300)	−0.00182	(0.01082)	−0.00058	(0.01309)
0.5	0.5	−0.00860	(0.01777)	−0.01075	(0.02758)	−0.00753	(0.01735)	−0.00673	(0.02067)
	0.9	−0.01646	(0.02237)	−0.02064	(0.02724)	−0.01514	(0.02121)	−0.01578	(0.02235)
	0.1	−0.00032	(0.03543)	0.00005	(0.03698)	−0.00038	(0.03475)	0.00115	(0.03825)
1.0	0.5	−0.00833	(0.03745)	−0.01014	(0.05050)	−0.00795	(0.03740)	−0.00379	(0.04085)
	0.9	−0.01712	(0.02731)	−0.02070	(0.03111)	−0.01710	(0.02718)	−0.01603	(0.02706)
	0.1	−0.00409	(0.05475)	−0.00307	(0.05660)	−0.00429	(0.05516)	−0.00175	(0.05858)
1.5	0.5	−0.01078	(0.05078)	−0.01662	(0.06623)	−0.00989	(0.05062)	−0.00459	(0.05268)
	0.9	−0.02072	(0.03523)	−0.02337	(0.03831)	−0.02140	(0.03582)	−0.01879	(0.03467)
	0.1	0.00134	(0.06611)	0.00208	(0.06821)	0.00172	(0.06670)	0.00552	(0.07054)
1.9	0.5	−0.01628	(0.06061)	−0.02081	(0.07918)	−0.01544	(0.06089)	−0.00740	(0.06269)
	0.9	−0.01863	(0.03339)	−0.02170	(0.03770)	−0.01890	(0.03373)	−0.01655	(0.03303)

$m=5$, $n=512$									
α	b	$\widehat{\rho}_{n,X}(1)$		$\widehat{\theta}^{(1)}_{\mathrm{GMM}}$		$\widehat{\theta}_{\mathrm{EL}}$		$\widehat{\theta}^{*}_{\mathrm{EL}}$	
	0.1	−0.00041	(0.00438)	0.00001	(0.00529)	−0.00013	(0.00439)	0.00010	(0.00506)
0.5	0.5	−0.00207	(0.00567)	−0.00263	(0.00967)	−0.00182	(0.00552)	−0.00192	(0.00670)
	0.9	−0.00515	(0.00825)	−0.00623	(0.00972)	−0.00483	(0.00800)	−0.00458	(0.00810)
	0.1	−0.00078	(0.01405)	−0.00046	(0.01500)	−0.00068	(0.01401)	−0.00029	(0.01484)
1.0	0.5	−0.00110	(0.01624)	−0.00103	(0.02265)	−0.00117	(0.01611)	−0.00054	(0.01749)
	0.9	−0.00596	(0.01242)	−0.00642	(0.01348)	−0.00618	(0.01267)	−0.00539	(0.01263)
	0.1	−0.00120	(0.02583)	−0.00097	(0.02701)	−0.00104	(0.02574)	−0.00110	(0.02695)
1.5	0.5	−0.00344	(0.02359)	−0.00323	(0.03168)	−0.00330	(0.02372)	−0.00201	(0.02458)
	0.9	−0.00457	(0.01396)	−0.00524	(0.01534)	−0.00463	(0.01423)	−0.00337	(0.01377)
	0.1	0.00044	(0.03319)	0.00115	(0.03413)	0.00058	(0.03336)	0.00070	(0.03419)
1.9	0.5	−0.00400	(0.02961)	−0.00594	(0.04018)	−0.00381	(0.02971)	−0.00323	(0.03032)
	0.9	−0.00485	(0.01596)	−0.00537	(0.01716)	−0.00491	(0.01605)	−0.00356	(0.01592)

Table 5.3: Empirical biases (and root mean square errors) of $\widehat{\rho}_{n,X}(1)$ and $\widehat{\theta}^{(i)}_{\mathrm{GMM}}$ ($i = 1, 2, 3$) for $\rho_X(1)$ of the stable AR(1) model

$m = 8$, $n = 128$									
α	b	$\widehat{\rho}_{n,X}(1)$		$\widehat{\theta}^{(1)}_{\mathrm{GMM}}$		$\widehat{\theta}_{\mathrm{EL}}$		$\widehat{\theta}^{*}_{\mathrm{EL}}$	
	0.1	−0.00040	(0.01173)	0.00069	(0.01377)	0.00034	(0.01168)	0.00170	(0.01427)
0.5	0.5	−0.00816	(0.02287)	−0.00980	(0.03459)	−0.00780	(0.02270)	−0.00582	(0.02595)
	0.9	−0.01664	(0.02365)	−0.02378	(0.03174)	−0.01447	(0.02166)	−0.01734	(0.02485)
	0.1	−0.00048	(0.03571)	0.00124	(0.03857)	0.00019	(0.03586)	0.00297	(0.03941)
1.0	0.5	−0.00726	(0.03660)	−0.00639	(0.05194)	−0.00740	(0.03719)	−0.00231	(0.03961)
	0.9	−0.01771	(0.02935)	−0.02204	(0.03458)	−0.01819	(0.02966)	−0.01747	(0.02962)
	0.1	−0.00157	(0.05389)	−0.00092	(0.05491)	−0.00185	(0.05409)	0.00031	(0.05781)
1.5	0.5	−0.01324	(0.04993)	−0.01454	(0.06777)	−0.01343	(0.05092)	−0.00473	(0.05134)
	0.9	−0.01842	(0.03130)	−0.02399	(0.03838)	−0.01832	(0.03137)	−0.01817	(0.03213)
	0.1	−0.00072	(0.06698)	0.00000	(0.06855)	−0.00104	(0.06734)	0.00258	(0.07145)
1.9	0.5	−0.00943	(0.05940)	−0.01163	(0.07826)	−0.00969	(0.05997)	−0.00031	(0.06111)
	0.9	−0.02126	(0.03477)	−0.02618	(0.04178)	−0.02163	(0.03524)	−0.02063	(0.03545)

$m = 8$, $n = 512$									
α	b	$\widehat{\rho}_{n,X}(1)$		$\widehat{\theta}^{(1)}_{\mathrm{GMM}}$		$\widehat{\theta}_{\mathrm{EL}}$		$\widehat{\theta}^{*}_{\mathrm{EL}}$	
	0.1	−0.00043	(0.00378)	−0.00010	(0.00452)	−0.00021	(0.00373)	−0.00002	(0.00447)
0.5	0.5	−0.00206	(0.00737)	−0.00237	(0.01173)	−0.00201	(0.00739)	−0.00121	(0.00855)
	0.9	−0.00513	(0.00863)	−0.00673	(0.01094)	−0.00462	(0.00822)	−0.00515	(0.00929)
	0.1	−0.00025	(0.01192)	−0.00012	(0.01304)	−0.00017	(0.01182)	0.00011	(0.01283)
1.0	0.5	−0.00349	(0.01506)	−0.00314	(0.02435)	−0.00338	(0.01510)	−0.00215	(0.01649)
	0.9	−0.00404	(0.01091)	−0.00604	(0.01428)	−0.00351	(0.01060)	−0.00304	(0.01096)
	0.1	−0.00090	(0.02677)	−0.00062	(0.02765)	−0.00071	(0.02664)	−0.00044	(0.02861)
1.5	0.5	−0.00444	(0.02570)	−0.00277	(0.03607)	−0.00429	(0.02614)	−0.00208	(0.02715)
	0.9	−0.00477	(0.01509)	−0.00560	(0.01686)	−0.00493	(0.01535)	−0.00326	(0.01478)
	0.1	0.00236	(0.03485)	0.00293	(0.03586)	0.00243	(0.03505)	0.00297	(0.03659)
1.9	0.5	−0.00097	(0.02931)	0.00070	(0.03953)	−0.00106	(0.02960)	0.00158	(0.03117)
	0.9	−0.00395	(0.01497)	−0.00500	(0.01762)	−0.00401	(0.01531)	−0.00223	(0.01500)

Table 5.4: Empirical biases (and root mean square errors) of $\widehat{\theta}^{(i)}_{\mathrm{GMM}}$ $(i = 1, 2, 3)$ for $\rho_X(1)$ of the stable MA(1) model

$m = 2,\ n = 128$							
α	a	$\widehat{\theta}^{(1)}_{\mathrm{GMM}}$		$\widehat{\theta}_{\mathrm{EL}}$		$\widehat{\theta}^{*}_{\mathrm{EL}}$	
	0.1	−0.00031	(0.01520)	−0.00054	(0.01495)	−0.00157	(0.01512)
0.5	0.5	−0.00895	(0.02004)	−0.00941	(0.01651)	−0.01060	(0.01701)
	0.9	−0.00762	(0.01836)	−0.00779	(0.01224)	−0.00808	(0.01319)
	0.1	0.00040	(0.03452)	−0.00021	(0.03346)	−0.00234	(0.03337)
1.0	0.5	−0.00841	(0.03439)	−0.00827	(0.02534)	−0.01010	(0.02593)
	0.9	−0.01163	(0.03556)	−0.01049	(0.02179)	−0.01130	(0.02327)
	0.1	−0.00262	(0.05231)	−0.00190	(0.05150)	−0.00438	(0.05163)
1.5	0.5	−0.00784	(0.04730)	−0.00864	(0.03386)	−0.01145	(0.03496)
	0.9	−0.01045	(0.04678)	−0.00724	(0.02491)	−0.00784	(0.02611)
	0.1	0.00048	(0.06765)	−0.00044	(0.06676)	−0.00358	(0.06562)
1.9	0.5	−0.01420	(0.05526)	−0.00928	(0.03838)	−0.01321	(0.04023)
	0.9	−0.01093	(0.05134)	−0.00917	(0.02699)	−0.00979	(0.02784)

$m = 2,\ n = 512$							
α	a	$\widehat{\theta}^{(1)}_{\mathrm{GMM}}$		$\widehat{\theta}_{\mathrm{EL}}$		$\widehat{\theta}^{*}_{\mathrm{EL}}$	
	0.1	−0.00066	(0.00370)	−0.00062	(0.00365)	−0.00091	(0.00367)
0.5	0.5	−0.00310	(0.00579)	−0.00273	(0.00467)	−0.00337	(0.00501)
	0.9	−0.00446	(0.00786)	−0.00446	(0.00596)	−0.00471	(0.00632)
	0.1	−0.00134	(0.01210)	−0.00137	(0.01201)	−0.00172	(0.01207)
1.0	0.5	−0.00192	(0.01325)	−0.00260	(0.00914)	−0.00318	(0.00924)
	0.9	−0.00263	(0.01366)	−0.00278	(0.00835)	−0.00295	(0.00864)
	0.1	−0.00006	(0.02517)	−0.00029	(0.02493)	−0.00114	(0.02487)
1.5	0.5	−0.00186	(0.02189)	−0.00099	(0.01514)	−0.00178	(0.01533)
	0.9	−0.00143	(0.02076)	−0.00208	(0.01075)	−0.00214	(0.01089)
	0.1	−0.00199	(0.03518)	−0.00214	(0.03474)	−0.00316	(0.03460)
1.9	0.5	−0.00277	(0.02689)	−0.00206	(0.01867)	−0.00330	(0.01895)
	0.9	−0.00162	(0.02364)	−0.00197	(0.01232)	−0.00214	(0.01234)

Table 5.5: Empirical biases (and root mean square errors) of $\widehat{\theta}^{(i)}_{\mathrm{GMM}}$ ($i = 1, 2, 3$) for $\rho_X(1)$ of the stable MA(1) model

$m = 5,\ n = 128$							
α	a	$\widehat{\theta}^{(1)}_{\mathrm{GMM}}$		$\widehat{\theta}_{\mathrm{EL}}$		$\widehat{\theta}^{*}_{\mathrm{EL}}$	
	0.1	−0.00181	(0.01572)	−0.00160	(0.01552)	−0.00227	(0.01567)
0.5	0.5	−0.00965	(0.02084)	−0.00882	(0.01496)	−0.01099	(0.01576)
	0.9	−0.00799	(0.01906)	−0.00847	(0.01061)	−0.00852	(0.01162)
	0.1	−0.00297	(0.03482)	−0.00288	(0.03447)	−0.00441	(0.03562)
1.0	0.5	−0.01161	(0.03525)	−0.00955	(0.02305)	−0.01211	(0.02377)
	0.9	−0.01100	(0.03525)	−0.00820	(0.01234)	−0.00893	(0.01449)
	0.1	−0.00115	(0.05603)	−0.00131	(0.05498)	−0.00334	(0.05656)
1.5	0.5	−0.01215	(0.04876)	−0.00632	(0.02782)	−0.01111	(0.03034)
	0.9	−0.01169	(0.04664)	−0.00921	(0.01388)	−0.00950	(0.01639)
	0.1	−0.00868	(0.06796)	−0.00824	(0.06680)	−0.01031	(0.06836)
1.9	0.5	−0.00762	(0.05391)	−0.00635	(0.02965)	−0.01116	(0.03247)
	0.9	−0.00863	(0.04923)	−0.00776	(0.01192)	−0.00775	(0.01451)

$m = 5,\ n = 512$							
α	a	$\widehat{\theta}^{(1)}_{\mathrm{GMM}}$		$\widehat{\theta}_{\mathrm{EL}}$		$\widehat{\theta}^{*}_{\mathrm{EL}}$	
	0.1	−0.00032	(0.00513)	−0.00035	(0.00513)	−0.00064	(0.00528)
0.5	0.5	−0.00284	(0.00654)	−0.00331	(0.00522)	−0.00366	(0.00546)
	0.9	−0.00179	(0.00516)	−0.00200	(0.00279)	−0.00230	(0.00302)
	0.1	0.00267	(0.01507)	0.00225	(0.01452)	0.00170	(0.01458)
1.0	0.5	−0.00066	(0.01213)	−0.00051	(0.00712)	−0.00183	(0.00725)
	0.9	−0.00333	(0.01358)	−0.00146	(0.00337)	−0.00181	(0.00406)
	0.1	−0.00305	(0.02611)	−0.00261	(0.02569)	−0.00339	(0.02642)
1.5	0.5	−0.00138	(0.02155)	−0.00160	(0.01227)	−0.00336	(0.01285)
	0.9	−0.00108	(0.02019)	−0.00139	(0.00435)	−0.00163	(0.00458)
	0.1	0.00010	(0.03335)	−0.00008	(0.03290)	−0.00136	(0.03358)
1.9	0.5	−0.00314	(0.02709)	−0.00229	(0.01426)	−0.00441	(0.01491)
	0.9	−0.00273	(0.02449)	−0.00152	(0.00460)	−0.00182	(0.00495)

Table 5.6: Empirical biases (and root mean square errors) of $\widehat{\theta}^{(i)}_{\mathrm{GMM}}$ ($i = 1, 2, 3$) for $\rho_X(1)$ of the stable MA(1) model

$m = 8,\ n = 128$							
α	a	$\widehat{\theta}^{(1)}_{\mathrm{GMM}}$		$\widehat{\theta}_{\mathrm{EL}}$		$\widehat{\theta}^{*}_{\mathrm{EL}}$	
	0.1	−0.00253	(0.01321)	−0.00239	(0.01319)	−0.00255	(0.01322)
0.5	0.5	−0.00489	(0.01542)	−0.00453	(0.01093)	−0.00548	(0.01077)
	0.9	−0.01131	(0.02151)	−0.01077	(0.01210)	−0.01197	(0.01396)
	0.1	−0.00221	(0.03341)	−0.00238	(0.03310)	−0.00332	(0.03334)
1.0	0.5	−0.01007	(0.03675)	−0.00914	(0.02321)	−0.01184	(0.02422)
	0.9	−0.01185	(0.03507)	−0.00894	(0.01131)	−0.00918	(0.01343)
	0.1	−0.00348	(0.05477)	−0.00342	(0.05342)	−0.00550	(0.05484)
1.5	0.5	−0.00596	(0.04603)	−0.00684	(0.02548)	−0.01016	(0.02751)
	0.9	−0.00835	(0.04400)	−0.00670	(0.00933)	−0.00698	(0.01215)
	0.1	−0.00738	(0.06414)	−0.00719	(0.06308)	−0.01037	(0.06493)
1.9	0.5	−0.01244	(0.05272)	−0.00737	(0.02817)	−0.01165	(0.03123)
	0.9	−0.00791	(0.04736)	−0.00618	(0.00842)	−0.00631	(0.01093)

$m = 8,\ n = 512$							
α	a	$\widehat{\theta}^{(1)}_{\mathrm{GMM}}$		$\widehat{\theta}_{\mathrm{EL}}$		$\widehat{\theta}^{*}_{\mathrm{EL}}$	
	0.1	−0.00045	(0.00386)	−0.00036	(0.00379)	−0.00046	(0.00398)
0.5	0.5	−0.00166	(0.00482)	−0.00170	(0.00358)	−0.00207	(0.00366)
	0.9	−0.00280	(0.00687)	−0.00285	(0.00339)	−0.00272	(0.00373)
	0.1	−0.00074	(0.01374)	−0.00073	(0.01373)	−0.00134	(0.01449)
1.0	0.5	−0.00083	(0.01273)	−0.00096	(0.00733)	−0.00207	(0.00745)
	0.9	−0.00232	(0.01254)	−0.00184	(0.00311)	−0.00177	(0.00336)
	0.1	−0.00140	(0.02626)	−0.00166	(0.02558)	−0.00246	(0.02565)
1.5	0.5	−0.00171	(0.02255)	−0.00136	(0.01211)	−0.00282	(0.01263)
	0.9	−0.00322	(0.02029)	−0.00194	(0.00346)	−0.00197	(0.00384)
	0.1	−0.00131	(0.03345)	−0.00146	(0.03292)	−0.00202	(0.03450)
1.9	0.5	−0.00273	(0.02695)	−0.00181	(0.01465)	−0.00386	(0.01528)
	0.9	−0.00201	(0.02357)	−0.00165	(0.00308)	−0.00162	(0.00350)

6 Proofs

6.1 Proofs of Chapter 2

Set

$$P_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} m(\lambda_t;\theta),$$
$$S_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} m(\lambda_t;\theta)m(\lambda_t;\theta)'$$

and we denote the ith element of $P_n(\theta)$ and the (i,j)th element of $S_n(\theta)$ as $P_n(\theta)_i$ and $S_n(\theta)_{ij}$, respectively.

Lemma 6.1. *Suppose that all assumptions in Theorem 2.1 hold. Then, for* $(i,j) = (1,2)$ *and* $(2,1)$,

$$P_n(\theta_i) \xrightarrow{\mathcal{P}} \begin{cases} 0_p & \textit{under } \Pi_i \\ D(\theta_i) & \textit{under } \Pi_j \end{cases}$$

and

$$S_n(\theta) \xrightarrow{\mathcal{P}} W_1(\theta)$$

for all $\theta \in \Theta$, *where* $D(\theta)$ *is a* $p \times 1$ *vector whose* i*th element is*

$$D(\theta)_i = \frac{1}{2\pi}\int_{-\pi}^{\pi} \operatorname{tr}\left[\frac{\partial g(\omega;\theta)^{-1}}{\partial\theta^i} f(\omega)\right] d\omega.$$

Proof. We prove only for $(i,j) = (1,2)$. Under Π_1, the kth element of $P_n(\theta_1)$ is evaluated as

$$\begin{aligned} P_n(\theta_1)_k &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \operatorname{tr}\left[\left.\frac{\partial g(\omega;\theta)^{-1}}{\partial\theta^k}\right|_{\theta=\theta_1} I_{n,X}(\omega)\right] d\omega + O_p\left(\frac{1}{n}\right) \\ &\xrightarrow{\mathcal{P}} \frac{1}{2\pi}\int_{-\pi}^{\pi} \operatorname{tr}\left[\left.\frac{\partial g(\omega;\theta)^{-1}}{\partial\theta^k}\right|_{\theta=\theta_1} f(\omega)\right] d\omega \quad \text{(e.g., Brillinger [4])} \\ &= \begin{cases} 0 & \text{under } \Pi_1 \\ D(\theta_1)_k & \text{under } \Pi_2 \end{cases}, \end{aligned}$$

and in Section 5 of Ogata and Taniguchi [49], the assertion on $S_n(\theta)$ was shown. □

Lemma 6.2. *Suppose that all assumptions in Theorem 2.1 hold. Then,*

$$\mathrm{ELR}(\theta_1, \theta_2) \xrightarrow{\mathcal{P}} \begin{cases} D(\theta_2)'W_1(\theta_2)D(\theta_2) & \textit{under } \Pi_1 \\ -D(\theta_1)'W_1(\theta_1)D(\theta_1) & \textit{under } \Pi_2 \end{cases}.$$

Proof. From Section 5 of Ogata and Taniguchi [49], we can see that $\mathrm{ELR}(\theta_1, \theta_2)$ admits, as $n \to \infty$, the asymptotic representation

$$\mathrm{ELR}(\theta_1, \theta_2) = -P_n(\theta_1)'S_n(\theta_1)^{-1}P_n(\theta_1) + P_n(\theta_2)'S_n(\theta_2)^{-1}P_n(\theta_2) + O_p\left(\frac{\log n}{n^{3/2}}\right). \tag{6.1}$$

Using this representation and Lemma 6.1, we obtain the desired result. □

Proof of Theorem 2.1. Lemma 6.2 implies that the empirical likelihood classification statistic converges to a positive and negative constant under Π_1 and Π_2, respectively, and this implies the result. □

Proof of Theorem 2.2. We modify $P_n(\theta)$ as

$$\sqrt{n}P_n(\theta) = C_n(\theta) + D_n(\theta) + O_p\left(\frac{1}{\sqrt{n}}\right), \tag{6.2}$$

where

$$C_n(\theta) = \frac{\sqrt{n}}{2\pi}\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta}\mathrm{tr}\left[g(\omega;\theta)^{-1}\left\{I_{n,X}(\omega) - f(\omega)\right\}\right]d\omega,$$

$$D_n(\theta) = \frac{\sqrt{n}}{2\pi}\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta}\mathrm{tr}\left[g(\omega;\theta)^{-1}f(\omega)\right]d\omega.$$

First, we consider the asymptotic behavior of $C_n(\theta_1)$ and $C_n(\theta_{1n})$ under Π_1. From Brillinger [4],

$$C_n(\theta_1) \xrightarrow{\mathcal{L}} N(0_p, W_2(\theta_1)) \tag{6.3}$$

under Π_1, and by the Taylor expansion, we have

$$C_n(\theta_{1n}) = C_n(\theta_1) + O_p\left(\frac{1}{\sqrt{n}}\right). \tag{6.4}$$

F lling that under Π_1,

$$\frac{\partial}{\partial\theta}\int_{-\pi}^{\pi}\mathrm{tr}[g(\omega;\theta)^{-1}f(\omega)]d\omega\bigg|_{\theta=\theta_1} = 0_p,$$

we have

$$D_n(\theta_1) = 0_p \quad \text{and} \quad D_n(\theta_{1n}) = F(\theta_1)\, h + O\left(\frac{1}{\sqrt{n}}\right). \tag{6.5}$$

From (6.1)-(6.5),

$$\begin{aligned}
&\frac{n}{2}\mathrm{ELR}(\theta_1, \theta_{1n}) \\
&= h'F(\theta_1)W_1(\theta_1)^{-1}C_n(\theta_1) + \frac{1}{2}h'F(\theta_1)W_1(\theta_1)^{-1}F(\theta_1)h + O_p\left(\frac{\log n}{\sqrt{n}}\right) \\
&\xrightarrow{\mathcal{L}} h'F(\theta_1)W_1(\theta_1)^{-1}W_2(\theta_1)^{1/2}N + \frac{1}{2}h'F(\theta_1)W_1(\theta_1)^{-1}F(\theta_1)h,
\end{aligned} \tag{6.6}$$

where N is a p-dimensional standard normal random vector. Obviously, the first term of (6.6) is a normal random variable with mean 0 and variance

$$\left\| W_2(\theta_1)^{1/2} W_1(\theta_1)^{-1} F(\theta_1) h \right\|_E^2 .$$

Therefore, the misclassification probability is

$$\begin{aligned}
\Pr{}^{(\mathrm{E})}(2|1) &= \Pr\left[\mathrm{ELR}(\theta_1, \theta_{1n}) < 0 \middle| \text{ under } \Pi_1\right] \\
&\to \Pr\left[[N(0,1)\text{-random variable}] < -\frac{1}{2}\frac{h'F(\theta_1)W_1(\theta_1)^{-1}F(\theta_1)h}{\|W_2(\theta_1)^{1/2}W_1(\theta_1)^{-1}F(\theta_1)h\|_E}\right] \\
&= \Phi\left[-\frac{1}{2}\frac{h'F(\theta_1)W_1(\theta_1)^{-1}F(\theta_1)h}{\|W_2(\theta_1)^{1/2}W_1(\theta_1)^{-1}F(\theta_1)h\|_E}\right].
\end{aligned}$$

□

Proof of Lemma 2.3. First of all, we prove the following asymptotics:

$$\left.\frac{\partial \mathrm{ELR}(\theta_1, \theta)}{\partial \theta}\right|_{\theta = \theta^*_{1n}} \xrightarrow{\mathcal{P}} 0_p \tag{6.7}$$

under Π_1 as $n \to \infty$. As in the proof of Theorem 2.2, we can easily see that

$$\begin{aligned}
P_n(\theta^*_{1n})_i &= P_n(\theta_1)_i + O_p\left(\frac{1}{\sqrt{n}}\right) \\
&= O_p\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}$$

$$\left.\frac{\partial P_n(\theta)_i}{\partial \theta^j}\right|_{\theta=\theta_{1n}^*} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\left[\left.\frac{\partial^2 g(\omega;\theta)^{-1}}{\partial\theta^i\partial\theta^j}\right|_{\theta=\theta_1} I_{n,X}(\omega)\right] d\omega + O_p\left(\frac{1}{\sqrt{n}}\right)$$
$$= F(\theta_1) + O_p\left(\frac{1}{\sqrt{n}}\right)$$

and the same procedure as Section 5 of Ogata and Taniguchi [49], we have

$$S(\theta_{1n}^*)_{ij} = S_n(\theta_1)_{ij} + O_p\left(\frac{1}{\sqrt{n}}\right)$$
$$= O_p(1)$$

and

$$\left.\frac{\partial S_n(\theta)_{ij}}{\partial \theta^k}\right|_{\theta=\theta_{1n}^*} = \left.\frac{\partial S_n(\theta)_{ij}}{\partial \theta^k}\right|_{\theta=\theta_1} + O_p\left(\frac{1}{\sqrt{n}}\right)$$
$$= O_p(1)$$

for all $i, j, k \in \{1, \cdots, p\}$. Therefore,

$$\left.\frac{\partial \mathrm{ELR}(\theta_1,\theta)}{\partial \theta^i}\right|_{\theta=\theta_{1n}^*} = \left.\frac{\partial P_n(\theta)' S_n(\theta)^{-1} P_n(\theta)}{\partial \theta^i}\right|_{\theta=\theta_{1n}^*}$$
$$= 2\left(\left.\frac{\partial P_n(\theta)}{\partial \theta^i}\right|_{\theta=\theta_{1n}^*}\right)' S_n(\theta_{1n}^*)^{-1} P_n(\theta_{1n}^*)$$
$$- P_n(\theta_{1n}^*)' S_n(\theta_{1n}^*)^{-1}\left(\left.\frac{\partial S_n(\theta)}{\partial \theta^i}\right|_{\theta=\theta_{1n}^*}\right) S_n(\theta_{1n}^*)^{-1} P_n(\theta_{1n}^*)$$
$$= O_p\left(\frac{1}{\sqrt{n}}\right),$$

so we can see the convergence (6.7). By this result and the relationship $\theta_{1n} - \theta_{1n}^* = O(n^{-1})$, we can also see that

$$n\left\{\mathrm{ELR}(\theta_1,\theta_{1n}) - \mathrm{ELR}(\theta_1,\theta_{1n}^*)\right\} = o_p(1).$$

□

Proof of Theorem 2.3.
(i) Lemma 2.3 implies that we can work with $n\mathrm{ELR}(\theta_1, \theta_{1n}^*)$ instead of $n\mathrm{ELR}(\theta_1, \theta_{1n})$ in results of type (6.6). Therefore, regarding $-F(\theta_1|\eta)^{-1}H(\theta_1|\eta)\xi$ as h and using

Theorem 2.2, we obtain the desired result.

(ii) If $B(\omega;\theta_1) \equiv A(\omega|\eta)$, it follows that

$$
\begin{aligned}
&\int_{-\pi}^{\pi} A(\omega|\eta)^* \left.\frac{\partial g(\omega;\theta)^{-1}}{\partial\theta^i}\right|_{\theta=\theta_1} A(\omega|\eta) d\omega \\
&= \int_{-\pi}^{\pi} \left[A(\omega|\eta)^* \left\{ \left.\frac{\partial B(\omega;\theta)^{*-1}}{\partial\theta^i}\right|_{\theta=\theta_1} \right\} \Omega^{-1} B(\omega;\theta_1)^{-1} A(\omega|\eta) \right. \\
&\qquad \left. + A(\omega|\eta)^* B(\omega;\theta_1)^{*-1} \Omega^{-1} \left\{ \left.\frac{\partial B(\omega;\theta)^{-1}}{\partial\theta^i}\right|_{\theta=\theta_1} \right\} A(\omega|\eta) \right] d\omega \\
&= \int_{-\pi}^{\pi} \left[A(\omega|\eta)^* B(\omega;\theta_1)^{*-1} \left\{ \left.\frac{\partial B(\omega;\theta)^{*}}{\partial\theta^i}\right|_{\theta=\theta_1} \right\} B(\omega;\theta_1)^{*-1} \Omega^{-1} \right. \\
&\qquad \left. + \Omega^{-1} B(\omega;\theta_1)^{-1} \left\{ \left.\frac{\partial B(\omega;\theta)}{\partial\theta^i}\right|_{\theta=\theta_1} \right\} B(\omega;\theta_1)^{-1} A(\omega|\eta) \right] d\omega \\
&= \int_{-\pi}^{\pi} \left\{ \left.\frac{\partial B(\omega;\theta)^{*}}{\partial\theta^i}\right|_{\theta=\theta_1} \right\} B(\omega;\theta_1)^{*-1} d\omega \Omega^{-1} \qquad (6.8) \\
&\qquad + \Omega^{-1} \int_{-\pi}^{\pi} B(\omega;\theta_1)^{-1} \left\{ \left.\frac{\partial B(\omega;\theta)}{\partial\theta^i}\right|_{\theta=\theta_1} \right\} d\omega \qquad (6.9) \\
&= O_{d\times d},
\end{aligned}
$$

since the integrands in (6.8) and (6.9) are, respectively, expressed as linear combinations of $\{\exp(-\mathrm{i}j\omega) : j = 1, 2, \cdots\}$ and $\{\exp(\mathrm{i}j\omega) : j = 1, 2, \cdots\}$. This implies that $\Gamma_i(\theta_1|\eta)$ vanishes for all $i = 1, \cdots, p$ and the asymptotic misclassification probabilities are then independent of the fourth-order cumulant of the process.

(iii) See Remark 2.1 (also refer Remark 3.2 of Hosoya and Taniguchi [22]). □

6.2 Proofs of Chapter 3

As well as the previous section, we set

$$\widetilde{P}_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \widetilde{m}(\lambda_t;\theta),$$

$$\widetilde{S}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \widetilde{m}(\lambda_t;\theta)\widetilde{m}(\lambda_t;\theta)'$$

and the ith and the (i,j)th components of $\widetilde{P}_n(\theta)$ and $\widetilde{S}_n(\theta)$ are denoted by $\widetilde{P}_n(\theta)_i$ and $\widetilde{S}_n(\theta)_{ij}$, respectively. We start with some auxiliary results. For any sequence of random variables $\{A(t) : t = 1,\cdots,n\}$, define

$$\widetilde{A}(t) = \frac{A(t)}{\{\sum_{s=1}^{n} A(s)^2\}^{1/2}}$$

and

$$T_{n,A}(\omega) = 2\sum_{h=1}^{n-1} \widehat{\rho}_{n,A}(h)\cos(h\omega).$$

Lemma 6.3.

$$\mathrm{E}\left[T_{n,Z}(\omega)\right] = 0, \quad \mathrm{E}\left[T_{n,Z}(\omega)^2\right] \to \begin{cases} 1 & (\omega \not\equiv 0 \mod \pi) \\ 2 & (\omega \equiv 0 \mod \pi) \end{cases}$$

as $n \to \infty$ uniformly in $\alpha \in (0,2]$ and $\sigma > 0$.

Proof. By symmetry and boundedness of $\widetilde{Z}(t)$'s, $\mathrm{E}[\widetilde{Z}(1)]$ exists and is equal to 0. Furthermore, from the definition of $\widetilde{Z}(t)$'s, we can see that $\sum_{t=1}^{n} \widetilde{Z}(t)^2 = 1$ almost surely, so $\mathrm{E}[\widetilde{Z}(1)^2] = 1/n$. Using Chebyshev's inequality, we can see

$$\Pr\left\{\left|\widetilde{Z}(1)\right| < \epsilon^{-1/2}n^{-1/2}\right\} > 1-\epsilon$$

for any $\epsilon > 0$. This inequality means $\sqrt{n}\widetilde{Z}(1)^2$ is $O_p(n^{-1/2})$, hence $\sqrt{n}\widetilde{Z}(1)^2$ converges to 0 in distribution uniformly in $\alpha \in (0,2]$. Therefore, by Lévy's continuity theorem and Taylor's theorem, there exists a constant c such that

$$\begin{aligned}\mathrm{E}\exp\left\{\mathrm{i}\xi\sqrt{n}\widetilde{Z}(1)^2\right\} &= 1 - \frac{\xi^2}{2}n\mathrm{E}\left[\widetilde{Z}(1)^4\right] + \frac{\xi^3\sin(\xi c)}{6}n^{3/2}\mathrm{E}\left[\widetilde{Z}(1)^6\right] \\ &\quad + \mathrm{i}\mathrm{Im}\left(\mathrm{E}\left[\exp\left\{\mathrm{i}\xi\sqrt{n}\widetilde{Z}(1)^2\right\}\right]\right) \\ &\to 1\end{aligned}$$

uniformly in $\xi \in \mathbb{R}$, where $\mathrm{Im}(z)$ denotes the imaginary part of a complex number z. So we can conclude that $n\mathrm{E}[\widetilde{Z}(1)^4]$ converges to 0 as n tends to ∞. We also find that $n(n-1)\mathrm{E}[\widetilde{Z}(1)^2\widetilde{Z}(2)^2]$ converges to 1 by taking expectations on both sides of following identical equation:

$$1=\sum_{t=1}^{n}\widetilde{Z}(t)^4+\sum_{t\neq s}\widetilde{Z}(t)^2\widetilde{Z}(s)^2. \tag{6.10}$$

Remembering the facts above, let us evaluate the expectations. First, from symmetry of $\widetilde{Z}(t)$'s, it is easy to see that $\mathrm{E}\left[T_{n,Z}(\omega)\right]$ is exactly equal to 0. Next, we expand $T_{n,Z}(\omega)^2$ and obtain that

$$\mathrm{E}\left[T_{n,Z}(\omega)^2\right]=\mathrm{E}\left[\widetilde{Z}(1)^2\widetilde{Z}(2)^2\right]\left\{n(n-1)+2n\sum_{h=1}^{n-1}\cos(2h\omega)-2\sum_{h=1}^{n-1}h\cos(2h\omega)\right\} \tag{6.11}$$

The first term of (6.11) converges to 1 as $n\to\infty$. Suppose that $\omega\equiv 0 \mod \pi$, then

$$\mathrm{E}\left[\widetilde{Z}(1)^2\widetilde{Z}(2)^2\right]\left\{2n\sum_{h=1}^{n-1}\cos(2h\omega)-2\sum_{h=1}^{n-1}h\cos(2h\omega)\right\}=n(n-1)\mathrm{E}\left[\widetilde{Z}(1)^2\widetilde{Z}(2)^2\right]\to 1.$$

Next, for $\omega\not\equiv 0 \mod \pi$, the following two identical equations hold;

$$\sum_{h=1}^{n-1}\cos(2h\omega)=\frac{\cos\{2(n-1)\omega\}+\cos(2\omega)-\cos(2n\omega)}{2\{1-\cos(2\omega)\}},$$
$$\sum_{h=1}^{n-1}h\cos(2h\omega)=\frac{n\cos\{2(n-1)\omega\}-(n-1)\cos(2n\omega)-1}{2\{1-\cos(2\omega)\}}.$$

Using these equations, we obtain that

$$2n\mathrm{E}\left[\widetilde{Z}(1)^2\widetilde{Z}(2)^2\right]\sum_{h=1}^{n-1}\cos(2h\omega)-2\mathrm{E}\left[\widetilde{Z}(1)^2\widetilde{Z}(2)^2\right]\sum_{h=1}^{n-1}h\cos(2h\omega)\to 0.$$

Hence we get the desired result. □

Lemma 6.4. *Under Assumption 2.3,*

$$\sum_{k\neq l}\mathrm{Cov}\left\{\widetilde{I}_{n,Z}(\lambda_k)^2,\widetilde{I}_{n,Z}(\lambda_l)^2\right\}=O(n),$$

where $\widetilde{I}_{n,Z}(\omega)$ is the self-normalized periodogram for $Z(1),\cdots,Z(n)$.

Proof. From Brillinger [4],

$$\mathrm{Cov}\left\{\widetilde{I}_{n,Z}(\lambda_k)^2, \widetilde{I}_{n,Z}(\lambda_l)^2\right\} = \sum_{\nu:p=1}^{8}\prod_{j=1}^{p}\mathrm{cum}\left\{\widetilde{d}_{n,Z}(\lambda_{k_j}) : k_j \in \nu_j\right\},$$

where $\widetilde{d}_{n,Z}(\omega) = \sum_{t=1}^{n}\widetilde{Z}(t)\exp(\mathrm{i}t\omega)$ and the summation is taken over all indecomposable partitions $\nu = \nu_1 \cup \cdots \cup \nu_p$, $p = 1, \cdots, 8$, of a table

$$\left\{\begin{array}{cccc} (k, & k, & -k, & -k) \\ (l, & l, & -l, & -l) \end{array}\right\} \tag{6.12}$$

(see Brillinger [4]). Note that

$$\mathrm{cum}\left\{\widetilde{d}_{n,Z}(\lambda_{k_1}), \cdots, \widetilde{d}_{n,Z}(\lambda_{k_m})\right\} = 0$$

for odd m. Let us consider following five partitions;

$$\begin{array}{ll} p=1, & \{(k,k,-k,-k,l,l,-l,-l)\}, \\ p=2, & \{(k,-k,l,-l),(k,-k,l,-l)\}, \\ & \{(k,-k),(k,-k,l,l,-l,-l)\}, \\ & \{(l,-l),(k,k,-k,-k,l,-l)\}, \\ p=3, & \{(k,-k),(l,-l),(k,-k,l,-l)\}. \end{array} \tag{6.13}$$

First, we show that with different k and l in ν,

$$\sum_{k\neq l}\sum_{\nu':p=1}^{8}\prod_{j=1}^{p}\mathrm{cum}\left\{\widetilde{d}_{n,Z}(\lambda_{k_j}) : k_j \in \nu_j\right\} = O(n). \tag{6.14}$$

for indecomposable decompositions $\nu' = \nu \setminus (6.13)$. However, the proof for (6.14) contains lengthy and complex algebra, so we confine to giving a representative example here. Let us consider partitions for $p = 4$. We can evaluate the second-order cumulant as

$$\begin{aligned}\mathrm{cum}\left\{\widetilde{d}_{n,Z}(\lambda_k), \widetilde{d}_{n,Z}(\lambda_l)\right\} &= \mathrm{E}\left[\widetilde{Z}(1)^2\right]\sum_{t=1}^{n}\exp\{\mathrm{i}t(\lambda_k - \lambda_l)\} \\ &= \frac{1}{n}\sum_{t=1}^{n}\exp\left\{\mathrm{i}t\frac{2\pi(k-l)}{n}\right\} \\ &= \left\{\begin{array}{ll} 1 & (k-l \equiv 0 \mod n) \\ 0 & (k-l \not\equiv 0 \mod n) \end{array}\right.,\end{aligned}$$

therefore

$$\begin{aligned}&\operatorname{cum}\left\{\widetilde{d}_{n,Z}(\lambda_{k_1}),\widetilde{d}_{n,Z}(\lambda_{k_2})\right\}\cdots\operatorname{cum}\left\{\widetilde{d}_{n,Z}(\lambda_{k_7}),\widetilde{d}_{n,Z}(\lambda_{k_8})\right\}\\&\quad=\begin{cases}1 & (k_1-k_2,\cdots,k_7-k_8\equiv 0 \mod n)\\ 0 & (\text{otherwise})\end{cases}\end{aligned}$$

So when $p=4$, we obtain

$$\sum_{k\neq l}\prod_{j=1}^{p}\operatorname{cum}\left\{\widetilde{d}_{n,Z}(\lambda_{k_j}):k_j\in\nu_j\right\}=O(n)$$

for any indecomposable partition of (6.12). Similarly, we can check (6.14) for $p=2$ and 3. Note that we can neglect the terms in p-decompositions with $p>5$, since such term always contains the odd-order cumulants of $\widetilde{d}_{n,Z}(\omega)$'s. Next, we need to check the cumulants on partitions (6.13). For simplicity, we introduce generic residual terms $R_n^{(1)}(k,l)$, $\cdots$, $R_n^{(4)}(k,l)$ such that

$$\sum_{k\neq l}R_n^{(i)}(k,l)^j=O(n)$$

for $j=1$,2 and $i=1,2,3,4$. A simple example of $R_n^{(i)}(k,l)$ is given as

$$R_n^{(i)}(k,l)=\begin{cases}{}^{\exists}(\text{constant}) & (k-l\equiv 0 \mod n)\\ 0 & (k-l\not\equiv 0 \mod n)\end{cases},$$

and these will appear when we expand the cumulants concerned. The fourth-order joint cumulant on $(\lambda_k,-\lambda_k,\lambda_l,-\lambda_l)$ is represented as

$$\begin{aligned}&\operatorname{cum}\left\{\widetilde{d}_{n,Z}(\lambda_k),\widetilde{d}_{n,Z}(-\lambda_k),\widetilde{d}_{n,Z}(\lambda_l),\widetilde{d}_{n,Z}(-\lambda_l)\right\}\\&\quad=n\mathrm{E}\left[\widetilde{Z}(1)^4\right]+n(n-1)\mathrm{E}\left[\widetilde{Z}(1)^2\widetilde{Z}(2)^2\right]-1+R_n^{(1)}(k,l).\end{aligned}\tag{6.15}$$

From (6.10), (6.15) becomes

$$\operatorname{cum}\left\{\widetilde{d}_{n,Z}(\lambda_k),\widetilde{d}_{n,Z}(-\lambda_k),\widetilde{d}_{n,Z}(\lambda_l),\widetilde{d}_{n,Z}(-\lambda_l)\right\}=R_n^{(1)}(k,l).$$

By the same argument as above, and using identical equations

$$\sum_{t=1}^{n}\widetilde{Z}(t)^2=1,$$

$$\sum_{t,s}^{*} \widetilde{Z}(t)^2\widetilde{Z}(s)^2 = \left\{\sum_{t=1}^{n} \widetilde{Z}(t)^2\right\}\left\{\sum_{t,s}^{*} \widetilde{Z}(t)^2\widetilde{Z}(s)^2\right\}$$
$$= 2\sum_{t,s}^{*} \widetilde{Z}(t)^4\widetilde{Z}(s)^2 + \sum_{t,s,u}^{*} \widetilde{Z}(t)^2\widetilde{Z}(s)^2\widetilde{Z}(u)^2,$$
$$\sum_{t=1}^{n} \widetilde{Z}(t)^4 = \left\{\sum_{t=1}^{n} \widetilde{Z}(t)^2\right\}\left\{\sum_{t=1}^{n} \widetilde{Z}(t)^4\right\} = \sum_{t,s}^{*} \widetilde{Z}(t)^4\widetilde{Z}(s)^2 + \sum_{t=1}^{n} \widetilde{Z}(t)^6$$

and

$$1 = \sum_{t=1}^{n} \widetilde{Z}(t)^8 + 4\sum_{t,s}^{*} \widetilde{Z}(t)^6\widetilde{Z}(s)^2 + 3\sum_{t,s}^{*} \widetilde{Z}(t)^4\widetilde{Z}(s)^4$$
$$+ 6\sum_{t,s,u}^{*} \widetilde{Z}(t)^4\widetilde{Z}(s)^2\widetilde{Z}(u)^2 + \sum_{t,s,u,v}^{*} \widetilde{Z}(t)^2\widetilde{Z}(s)^2\widetilde{Z}(u)^2\widetilde{Z}(v)^2, \tag{6.16}$$

we obtain that

$$\mathrm{cum}\left\{\widetilde{d}_{n,Z}(\lambda_k), \widetilde{d}_{n,Z}(-\lambda_k), \widetilde{d}_{n,Z}(\lambda_l), \widetilde{d}_{n,Z}(\lambda_l), \widetilde{d}_{n,Z}(-\lambda_l), \widetilde{d}_{n,Z}(-\lambda_l)\right\}$$
$$= R_n^{(2)}(k,l),$$
$$\mathrm{cum}\left\{\widetilde{d}_{n,Z}(\lambda_k), \cdots, \widetilde{d}_{n,Z}(-\lambda_l)\right\} \quad \text{(the eighth-order joint cumulant)}$$
$$= 2n^2\mathrm{E}\left[\widetilde{Z}(1)^4\widetilde{Z}(2)^4\right] - 6n^3\mathrm{E}\left[\widetilde{Z}(1)^4\widetilde{Z}(2)^2\widetilde{Z}(3)^2\right]$$
$$+ n^4\mathrm{E}\left[\widetilde{Z}(1)^2\widetilde{Z}(2)^2\widetilde{Z}(3)^2\widetilde{Z}(4)^2\right] - \left\{n^2\mathrm{E}\left[\widetilde{Z}(1)^2\widetilde{Z}(2)^2\right]\right\}^2 + R_n^{(3)}(k,l), \tag{6.17}$$

where $\sum^{*}_{t_1,\cdots,t_m}$ is a summation taken over all $t_1, \cdots, t_m$ are different from each other. According to the same argument as Lemma 6.1, the first and second terms in (6.17) converge to 0 as $n \to \infty$, and the fourth term converges to 1.

Finally, from (6.16), the third term converges to 1. Hence the eighth-order joint cumulant becomes

$$\mathrm{cum}\left\{\widetilde{d}_{n,Z}(\lambda_k), \cdots, \widetilde{d}_{n,Z}(-\lambda_l)\right\} = R_n^{(4)}(k,l).$$

Finally, we have

$$\sum_{k\neq l}\sum_{\nu:p=1}^{8}\prod_{j=1}^{p} \mathrm{cum}\left\{\widetilde{d}_{n,Z}(\lambda_{k_j}) : k_j \in \nu_j\right\} = O(n).$$

□

Lemma 6.5. *Under Assumptions 2.3 and 3.1,*

$$\widetilde{S}_n(\theta_0) \xrightarrow{\mathcal{P}} \widetilde{W}(\theta_0)$$

as $n \to \infty$. *Here* $\widetilde{W}(\theta)$ *is defined in Theorem 3.1.*

Proof. We first make use of the decomposition of the periodogram in Klüppelberg and Mikosch [33] as follows;

$$\begin{aligned} \widetilde{I}_{n,X}(\omega)^2 &= \widetilde{f}(\omega)^2 \widetilde{I}_{n,Z}(\omega)^2 + o_p(1) \\ &= \widetilde{f}(\omega)^2 \left\{ 1 + 2\sum_{h=1}^{n-1} \widehat{\rho}_{n,Z}(h)\cos(h\omega) \right\}^2 + o_p(1) \\ &= \widetilde{f}(\omega)^2 \left\{ 1 + 2T_{n,Z}(\omega) + T_{n,Z}(\omega)^2 \right\} + o_p(1). \end{aligned} \tag{6.18}$$

Then from Lemma 6.3, we obtain that

$$\begin{aligned} \mathrm{E}\left[\widetilde{S}_n(\theta_0)\right] &= \frac{1}{n}\sum_{t=1}^{n} \left.\frac{\partial g(\lambda_t;\theta)^{-1}}{\partial \theta}\frac{\partial g(\lambda_t;\theta)^{-1}}{\partial \theta'}\right|_{\theta=\theta_0} \mathrm{E}\left[\widetilde{I}_{n,X}(\lambda_t)^2\right] \\ &\to \frac{1}{2\pi}\int_{-\pi}^{\pi} \left.\frac{\partial g(\omega;\theta)^{-1}}{\partial \theta}\frac{\partial g(\omega;\theta)^{-1}}{\partial \theta'}\right|_{\theta=\theta_0} 2\widetilde{f}(\omega)^2 d\omega = \widetilde{W}(\theta_0). \end{aligned}$$

Moreover, from Lemma 6.4, Assumption 2.3 and (6.18), if we define

$$h_{ij}(\omega) = \left.\frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^i}\frac{\partial g(\omega;\theta)^{-1}}{\partial \theta^j}\right|_{\theta=\theta_0} \widetilde{f}(\omega)^2,$$

then

$$\begin{aligned} \mathrm{Cov}\left\{\widetilde{S}_n(\theta_0)_{ij}, \widetilde{S}_n(\theta_0)_{kl}\right\} &= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n} h_{ij}(\lambda_t)h_{kl}(\lambda_s)\mathrm{Cov}\left\{\widetilde{I}_{n,Z}(\lambda_t)^2, \widetilde{I}_{n,Z}(\lambda_s)^2\right\} \\ &= \frac{1}{n^2}\sum_{t=1}^{n} h_{ij}(\lambda_t)h_{kl}(\lambda_t)\mathrm{Var}\left[\widetilde{I}_{n,Z}(\lambda_t)^2\right] \\ &\quad + \frac{1}{n^2}\sum_{t\neq s} h_{ij}(\lambda_t)h_{kl}(\lambda_s)\mathrm{Cov}\left\{\widetilde{I}_{n,Z}(\lambda_t)^2, \widetilde{I}_{n,Z}(\lambda_s)^2\right\} + o(1) \\ &\to 0 \end{aligned}$$

for any $i, j, k, l \in \{1, \cdots, p\}$. These facts imply the convergence of $\widetilde{S}_n(\theta_0)$ in probability. □

Proof of Theorem 3.1. By Lagrange's multiplier method, $w_1, \cdots, w_n$ which maximize the objective function in $\widetilde{r}_n(\theta_0)$ are given as

$$w_t = \frac{1}{n} \frac{1}{1 + \tau_n' \widetilde{m}(\lambda_t; \theta_0)}, \quad t = 1, \cdots, n,$$

where $\tau_n \in \mathbb{R}^p$ is the Lagrange multiplier which is defined as the solution to p-restrictions

$$\frac{1}{n} \sum_{t=1}^{n} \frac{\widetilde{m}(\lambda_t; \theta_0)}{1 + \tau_n' \widetilde{m}(\lambda_t; \theta_0)} = 0_p. \tag{6.19}$$

Note that every w_t is nonnegative, so the quantity $1 + \tau_n' \widetilde{m}(\lambda_t; \theta_0)$ should be nonnegative as well for $t = 1, \cdots, n$. First of all, let us derive the stochastic order of τ_n. Set $Y_t = \tau_n' \widetilde{m}(\lambda_t; \theta_0)$ and from (6.19),

$$\begin{aligned}
0_p &= \frac{1}{n} \sum_{t=1}^{n} \frac{\widetilde{m}(\lambda_t; \theta_0)}{1 + Y_t} \\
&= \frac{1}{n} \sum_{t=1}^{n} \left\{ 1 - Y_t + \frac{Y_t^2}{1 + Y_t} \right\} \widetilde{m}(\lambda_t; \theta_0) \\
&= \widetilde{P}_n(\theta_0) - \widetilde{S}_n(\theta_0) \tau_n + \frac{1}{n} \sum_{t=1}^{n} \frac{\widetilde{m}(\lambda_t; \theta_0) Y_t^2}{1 + Y_t}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\tau_n &= \widetilde{S}_n(\theta_0)^{-1} \left\{ \widetilde{P}_n(\theta_0) + \frac{1}{n} \sum_{t=1}^{n} \frac{\widetilde{m}(\lambda_t; \theta_0) Y_t^2}{1 + Y_t} \right\} \\
&= \widetilde{S}_n(\theta_0)^{-1} \widetilde{P}_n(\theta_0) + \zeta_n \quad \text{(say)}.
\end{aligned} \tag{6.20}$$

Now, we introduce $M_n = \max_{1 \le k \le n} \|\widetilde{m}(\lambda_k; \theta_0)\|_E$. Noting Assumption 2.3, there exists $c_1 < \infty$ such that the stochastic order of M_n is given by

$$\begin{aligned}
M_n &= \max_{1 \le t \le n} \left\| \left. \frac{\partial g(\lambda_t; \theta)^{-1}}{\partial \theta} \right|_{\theta = \theta_0} \widetilde{I}_{n,X}(\lambda_t) \right\|_E \\
&\le \max_{\omega \in [-\pi, \pi]} \left\| \left. \frac{\partial g(\omega; \theta)^{-1}}{\partial \theta} \right|_{\theta = \theta_0} \right\|_E \max_{\omega \in [-\pi, \pi]} \left| \widetilde{I}_{n,X}(\omega) \right|
\end{aligned}$$

$$
\begin{aligned}
&= \max_{\omega\in[-\pi,\pi]} \left\| \left. \frac{\partial g(\omega;\theta)^{-1}}{\partial\theta} \right|_{\theta=\theta_0} \right\|_E \max_{\omega\in[-\pi,\pi]} \left|\widetilde{f}(\omega)\right| \frac{\max_{\omega\in[-\pi,\pi]} \left|\widetilde{I}_{n,X}(\omega)\right|}{\max_{\omega\in[-\pi,\pi]} \left|\widetilde{f}(\omega)\right|} \\
&\le \max_{\omega\in[-\pi,\pi]} \left\| \left. \frac{\partial g(\omega;\theta)^{-1}}{\partial\theta} \right|_{\theta=\theta_0} \right\|_E \max_{\omega\in[-\pi,\pi]} \left|\widetilde{f}(\omega)\right| \max_{\omega\in[-\pi,\pi]} \left| \frac{\widetilde{I}_{n,X}(\omega)}{\widetilde{f}(\omega)} \right| \\
&= c_1 \max_{\omega\in[-\pi,\pi]} \left| \frac{\widetilde{I}_{n,X}(\omega)}{\widetilde{f}(\omega)} \right|.
\end{aligned}
$$

It is not difficult to check that Assumption 3.1 is sufficient condition for Corollary 3.3 of Mikosch, Resnick, and Samorodnitsky [41], so we have $M_n = O_p(\beta_n^2)$, where

$$
\beta_n = \begin{cases} (\log n)^{1-1/\alpha} & (1 < \alpha < 2) \\ \log\log n & (\alpha = 1) \end{cases}.
$$

Henceforth, we confine α within $1 < \alpha < 2$. When $\alpha = 1$, the same argument as follows will succeed.

We next show the rate of convergence of ζ_n in (6.20). There exists a unit vector v in $\mathbb{R}^p$ such that $\tau_n = \|\tau_n\|_E v$ and we can see that

$$
\begin{aligned}
\|\tau_n\|_E v'\widetilde{S}_n(\theta_0)v &= \|\tau_n\|_E u' \left\{ \frac{1}{n}\sum_{t=1}^n \widetilde{m}(\lambda_t;\theta_0)\widetilde{m}(\lambda_t;\theta_0)' \right\} v \\
&= \|\tau_n\|_E v' \left\{ \frac{1}{n}\sum_{t=1}^n \frac{\widetilde{m}(\lambda_t;\theta_0)\widetilde{m}(\lambda_t;\theta_0)'}{1+Y_t}(1+Y_t) \right\} v \\
&\le \|\tau_n\|_E v' \left\{ \frac{1}{n}\sum_{t=1}^n \frac{\widetilde{m}(\lambda_t;\theta_0)\widetilde{m}(\lambda_t;\theta_0)'}{1+Y_t}(1+\|\tau_n\|_E M_n) \right\} v \\
&= v' \left\{ \frac{1}{n}\sum_{t=1}^n \frac{\widetilde{m}(\lambda_t;\theta_0)\widetilde{m}(\lambda_t;\theta_0)'}{1+Y_t}(1+\|\tau_n\|_E M_n) \right\} \tau_n \\
&= v' \left\{ \frac{1}{n}\sum_{t=1}^n \frac{Y_t}{1+Y_t}\widetilde{m}(\lambda_t;\theta_0)(1+\|\tau_n\|_E M_n) \right\} \\
&= v' \left\{ \frac{1}{n}\sum_{t=1}^n \left\{ 1 - \frac{1}{1+Y_t} \right\} \widetilde{m}(\lambda_t;\theta_0)(1+\|\tau_n\|_E M_n) \right\} \qquad (6.21)
\end{aligned}
$$

Recalling (6.19), (6.21) finally becomes

$$
\|\tau_n\|_E \left\{ v'\widetilde{S}_n(\theta_0)v - v'M_n\widetilde{P}_n(\theta_0) \right\} \le v'\widetilde{P}_n(\theta_0). \qquad (6.22)
$$

Lemma P5.1 of Brillinger [4] allows us to write $x_n \widetilde{P}_n(\theta_0)$ as

$$x_n \widetilde{P}_n(\theta_0) = \frac{x_n}{2\pi} \int_{-\pi}^{\pi} \left. \frac{\partial g(\omega;\theta)}{\partial \theta} \right|_{\theta=\theta_0} \left\{ \widetilde{I}_{n,X}(\omega) - U_n \widetilde{f}(\omega) \right\} d\omega + O_p\left(\frac{x_n}{n}\right),$$

where U_n is defined as (3.3). Then, by Lemma 3.2 we have

$$x_n \widetilde{P}_n(\theta_0) \xrightarrow{\mathcal{L}} \widetilde{V}(\theta_0) \tag{6.23}$$

for $\alpha \in [1,2)$ as $n \to \infty$, where $\widetilde{V}(\theta)$ is defined in Theorem 3.1. So from (6.22) we obtain

$$O_p(\|\tau_n\|_E) \left[O_p(1) - O_p\left\{ (\log n)^{2-2/\alpha} \right\} O_p(x_n^{-1}) \right] \le O_p(x_n^{-1}). \tag{6.24}$$

Noting that $(\log n)^{2-2/\alpha} x_n^{-1} \to 0$ as $n \to \infty$, (6.24) implies that

$$O_p(\|\tau_n\|_E) \le O_p(x_n^{-1}) \tag{6.25}$$

asymptotically. On the other hand,

$$\begin{aligned}
\frac{1}{n} \sum_{t=1}^{n} \|\widetilde{m}(\lambda_t;\theta_0)\|_E^3 &= \frac{1}{n} \sum_{t=1}^{n} \|\widetilde{m}(\lambda_t;\theta_0)\|_E \|\widetilde{m}(\lambda_t;\theta_0)\|_E^2 \\
&\le \frac{1}{n} \sum_{t=1}^{n} M_n \widetilde{m}(\lambda_t;\theta_0)' \widetilde{m}(\lambda_t;\theta_0) \\
&= M_n \mathrm{tr}\left[\widetilde{S}_n(\theta_0) \right] \\
&= O_p\left\{ (\log n)^{2-2/\alpha} \right\}.
\end{aligned} \tag{6.26}$$

It is easily shown that ζ_n in (6.20) satisfies

$$\|\zeta_n\|_E \le \frac{1}{n} \sum_{t=1}^{n} \|\widetilde{m}(\lambda_t;\theta_0)\|_E^3 \|\tau_n\|_E^2 |1 + Y_t|^{-1}.$$

Thus, from (6.25) and (6.26) we have

$$O_p(\|x_n \zeta_n\|_E) = O_p\left\{ \frac{(\log n)^{2-1/\alpha}}{n^{1/\alpha}} \right\} \xrightarrow{\mathcal{P}} 0.$$

Now let us show the convergence of the empirical likelihood ratio statistic. Under $H : \theta = \theta_0$, the empirical likelihood ratio statistic is expanded as

$$-2\frac{x_n^2}{n} \log \widetilde{r}_n(\theta_0) = -2\frac{x_n^2}{n} \sum_{t=1}^{n} \log n w_t$$

$$
\begin{aligned}
&= 2\frac{x_n^2}{n}\sum_{t=1}^{n}\log(1+Y_t) \\
&= 2\frac{x_n^2}{n}\sum_{t=1}^{n}Y_t - \frac{x_n^2}{n}\sum_{t=1}^{n}Y_t^2 + 2\frac{x_n^2}{n}\sum_{t=1}^{n}O_p(Y_t^3),
\end{aligned}
$$

where

$$
\begin{aligned}
2\frac{x_n^2}{n}\sum_{t=1}^{n}Y_t &= 2\frac{x_n^2}{n}\sum_{t=1}^{n}\tau_n'\widetilde{m}(\lambda_t;\theta_0) \\
&= 2\frac{x_n^2}{n}\left\{\widetilde{S}_n(\theta_0)^{-1}\widetilde{P}_n(\theta_0)+\zeta_n\right\}'\sum_{t=1}^{n}\widetilde{m}(\lambda_t;\theta_0) \\
&= 2x_n^2\left\{\widetilde{P}_n(\theta_0)'\widetilde{S}_n(\theta_0)^{-1}+\zeta_n{}'\right\}\widetilde{P}_n(\theta_0) \\
&= 2\left\{x_n\widetilde{P}_n(\theta_0)\right\}'\widetilde{S}_n(\theta_0)^{-1}\left\{x_n\widetilde{P}_n(\theta_0)\right\}+2(x_n\zeta_n)'\left\{x_n\widetilde{P}_n(\theta_0)\right\},
\end{aligned}
$$

$$
\begin{aligned}
\frac{x_n^2}{n}\sum_{t=1}^{n}Y_t^2 &= \frac{x_n^2}{n}\sum_{t=1}^{n}\{\tau_n'\widetilde{m}(\lambda_t;\theta_0)\}^2 \\
&= x_n^2\tau_n'\widetilde{S}_n(\theta_0)\tau_n \\
&= x_n^2\left\{\widetilde{P}_n(\theta_0)'\widetilde{S}_n(\theta_0)^{-1}+\zeta_n{}'\right\}\widetilde{S}_n(\theta_0)\left\{\widetilde{S}_n(\theta_0)^{-1}\widetilde{P}_n(\theta_0)+\zeta_n\right\} \\
&= \left\{x_n\widetilde{P}_n(\theta_0)\right\}'\widetilde{S}_n(\theta_0)^{-1}\left\{x_n\widetilde{P}_n(\theta_0)\right\} \\
&\quad + (x_n\zeta_n)'\,\widetilde{S}_n(\theta_0)\,(x_n\zeta_n)+2\,(x_n\zeta_n)'\left\{x_n\widetilde{P}_n(\theta_0)\right\},
\end{aligned}
$$

and there exists finite c_2 such that

$$
\begin{aligned}
\frac{x_n^2}{n}\left|\sum_{t=1}^{n}O_p(Y_t^3)\right| &\le \frac{x_n^2}{n}c_2\sum_{t=1}^{n}|Y_t|^3 \\
&\le \frac{x_n^2}{n}c_2\|\tau_n\|_E^3\sum_{t=1}^{n}\|\widetilde{m}(\lambda_t;\theta_0)\|_E^3 \\
&= O_p\left\{\frac{(\log n)^{2-1/\alpha}}{n^{1/\alpha}}\right\} \\
&\xrightarrow{\mathcal{P}} 0
\end{aligned}
$$

as $n \to \infty$. As a consequence of Lemma 6.5 and (6.23),

$$-\frac{2x_n^2}{n} \log \widetilde{r}_n(\theta_0) = \left\{ x_n \widetilde{P}_n(\theta_0) \right\}' \widetilde{S}_n(\theta_0)^{-1} \left\{ x_n \widetilde{P}_n(\theta_0) \right\} + o_p(1) \qquad (6.27)$$
$$\xrightarrow{\mathcal{L}} \widetilde{V}(\theta_0)' \widetilde{W}(\theta_0)^{-1} \widetilde{V}(\theta_0)$$

as $n \to \infty$ for $\alpha \in [1, 2)$. □

6.3 Proofs of Chapter 4

Proof of Theorem 4.1. Under the assumptions of Theorem 4.1, we have

$$\widetilde{P}_n(\theta) \xrightarrow{\mathcal{P}} \widetilde{D}(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial g(\omega; \theta)^{-1}}{\partial \theta} \widetilde{f}(\omega) d\omega.$$

Recalling the expansion (6.27), we can also see that

$$\widetilde{\mathrm{ELR}}(\theta_1, \theta_2) \xrightarrow{\mathcal{P}} \begin{cases} \widetilde{D}(\theta_2)' \widetilde{W}(\theta_2) \widetilde{D}(\theta_2) > 0 & \text{under } \Pi_1 \\ -\widetilde{D}(\theta_1)' \widetilde{W}(\theta_1) \widetilde{D}(\theta_1) < 0 & \text{under } \Pi_2 \end{cases} \qquad (6.28)$$

as $n \to \infty$. (6.28) implies that $\widetilde{\mathrm{ELR}}(\theta_1, \theta_2)$ converges to a positive and negative constant in probability under Π_1 and Π_2, respectively. Then, the misclassification probabilities by $\widetilde{\mathrm{ELR}}(\theta_1, \theta_2)$ converge to zero as $n \to \infty$. □

Proof of Theorem 4.2. From (6.27),

$$x_n^2 \widetilde{\mathrm{ELR}}(\theta_1, \widetilde{\theta}_{1n}) = -\left\{ x_n \widetilde{P}_n(\theta_1) \right\}' \widetilde{S}_n(\theta_1)^{-1} \left\{ x_n \widetilde{P}_n(\theta_1) \right\}$$
$$+ \left\{ x_n \widetilde{P}_n(\widetilde{\theta}_{1n}) \right\}' \widetilde{S}_n(\widetilde{\theta}_{1n})^{-1} \left\{ x_n \widetilde{P}_n(\widetilde{\theta}_{1n}) \right\} + o_p(1). \qquad (6.29)$$

Now, write

$$\partial g(\omega; \theta) = \frac{\partial g(\omega; \theta)^{-1}}{\partial \theta} = (\partial_1 g(\omega; \theta), \cdots, \partial_p g(\omega; \theta))',$$
$$\partial^2 g(\omega; \theta) = \left(\frac{\partial^2 g(\omega; \theta)^{-1}}{\partial \theta^i \partial \theta^j} : i, j = 1, \cdots, p \right).$$

Then, we decompose $x_n \widetilde{P}_n(\theta)$ as

$$x_n \widetilde{P}_n(\theta) = \frac{x_n}{2\pi} \int_{-\pi}^{\pi} \partial g(\omega; \theta) \widetilde{I}_{n,X}(\omega) d\omega + R_n(\alpha)$$

$$= \widetilde{C}_n(\theta) + \widetilde{D}_n(\theta) + R_n(\alpha), \quad \text{(say)}$$

where

$$\widetilde{C}_n(\theta) = \frac{x_n}{2\pi} \int_{-\pi}^{\pi} \partial g(\omega;\theta) \left\{ \widetilde{I}_{n,X}(\omega) - U_n \widetilde{f}(\omega) \right\} d\omega,$$

$$\widetilde{D}_n(\theta) = \frac{x_n U_n}{2\pi} \int_{-\pi}^{\pi} \partial g(\omega;\theta) \widetilde{f}(\omega) d\omega$$

and

$$R_n(\alpha) = x_n \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \partial g(\omega;\theta) \widetilde{I}_{n,X}(\omega) d\omega - \frac{1}{n} \sum_{t=1}^{n} \partial g(\lambda_t;\theta) \widetilde{I}_{n,X}(\lambda_t) \right) = O_p\left(\frac{x_n}{n} \right).$$

Noting that under Π_1,

$$\widetilde{D}_n(\theta_1) = \frac{x_n U_n}{2\pi} \int_{-\pi}^{\pi} \left. \frac{\partial g(\omega;\theta)^{-1}}{\partial \theta} \right|_{\theta=\theta_1} \widetilde{f}(\omega) d\omega = 0_p$$

almost surely, we have

$$\begin{aligned}
(6.29) = & - \widetilde{C}_n(\theta_1)' \widetilde{S}_n(\theta_1)^{-1} \widetilde{C}_n(\theta_1) & (6.30)\\
& + \widetilde{C}_n(\widetilde{\theta}_{1n})' \widetilde{S}_n(\widetilde{\theta}_{1n})^{-1} \widetilde{C}_n(\widetilde{\theta}_{1n}) & (6.31)\\
& + 2\widetilde{D}_n(\widetilde{\theta}_{1n})' \widetilde{S}_n(\widetilde{\theta}_{1n})^{-1} \widetilde{C}_n(\widetilde{\theta}_{1n}) & (6.32)\\
& + \widetilde{D}_n(\widetilde{\theta}_{1n})' \widetilde{S}_n(\widetilde{\theta}_{1n})^{-1} \widetilde{D}_n(\widetilde{\theta}_{1n}) & (6.33)\\
& + o_p(1)
\end{aligned}$$

for $1 \leq \alpha < 2$. From Lemma 3.2,

$$\widetilde{C}_n(\theta_1) \xrightarrow{\mathcal{L}} \widetilde{V}(\theta_1),$$

where $\widetilde{V}(\theta)$ is the same random vector as Theorem 3.1. Using Taylor's theorem, we have

$$\partial g(\omega;\widetilde{\theta}_{1n}) = \partial g(\omega;\theta_1) + x_n^{-1} \partial^2 g(\omega;\theta_1) h + O(x_n^{-2}).$$

Hence,

$$\begin{aligned}
\widetilde{C}_n(\theta_{1n}) - \widetilde{C}_n(\theta_1) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \partial^2 g(\omega;\theta_1) h \left\{ \widetilde{I}_{n,X}(\omega) - \widetilde{f}(\omega) U_n \right\} d\omega + O_p\left(x_n^{-1} \right)\\
&\xrightarrow{\mathcal{P}} 0_p.
\end{aligned}$$

On the other hand, it is seen that

$$\underset{n\to\infty}{p\text{-lim}}\, \widetilde{S}_n(\theta_1) = \underset{n\to\infty}{p\text{-lim}}\, \widetilde{S}_n(\widetilde{\theta}_{1n}) = \widetilde{W}(\theta_1).$$

Therefore, by Slutsky's lemma and the continuous mapping theorem, (6.30)+(6.31) $\xrightarrow{\mathcal{P}}$ 0. Next, we evaluate $\widetilde{D}_n(\widetilde{\theta}_{1n})$. Note that $p\text{-lim}_{n\to\infty} U_n = 1$ (e.g., Klüppelberg and Mikosch [34]), then, under Π_1,

$$\begin{aligned}
\widetilde{D}_n(\widetilde{\theta}_{1n}) &= \frac{1+o_p(1)}{2\pi} x_n \int_{-\pi}^{\pi} \partial g(\omega;\widetilde{\theta}_{1n}) \widetilde{f}(\omega) d\omega \\
&= \frac{1+o_p(1)}{2\pi} \left\{ x_n \int_{-\pi}^{\pi} \partial g(\omega;\theta_1) \widetilde{f}(\omega) d\omega + \int_{-\pi}^{\pi} \partial^2 g(\omega;\theta_1) h \widetilde{f}(\omega) d\omega + O\left(x_n^{-1}\right) \right\} \\
&= \frac{1+o_p(1)}{2\pi} \int_{-\pi}^{\pi} \partial^2 g(\omega;\theta_1) h \widetilde{f}(\omega) d\omega + O_p\left(x_n^{-1}\right). \\
&= \{1+o_p(1)\} \widetilde{F}(\theta_1) h + O_p\left(x_n^{-1}\right).
\end{aligned}$$

The facts above lead

$$\begin{aligned}
(6.32) &\xrightarrow{\mathcal{L}} 2h'\widetilde{F}(\theta_1)\widetilde{W}(\theta_1)^{-1}\widetilde{V}(\theta_1), \\
(6.33) &\xrightarrow{\mathcal{P}} h'\widetilde{F}(\theta_1)\widetilde{W}(\theta_1)^{-1}\widetilde{F}(\theta_1)h.
\end{aligned}$$

So the asymptotic misclassification probability under the contiguous condition (4.1) is evaluated as

$$\begin{aligned}
&\lim_{n\to\infty} \mathrm{Pr}^{(\mathrm{E})}(2|1) \\
&= \lim_{n\to\infty} \Pr\left[\widetilde{\mathrm{ELR}}(\theta_1, \theta_{1n}) \le 0 \,\middle|\, \text{under } \Pi_1 \right] \\
&= \Pr\left[2h'\widetilde{F}(\theta_1)\widetilde{W}(\theta_1)^{-1}\widetilde{V}(\theta_1) + h'\widetilde{F}(\theta_1)\widetilde{W}(\theta_1)^{-1}\widetilde{F}(\theta_1)h \le 0 \right].
\end{aligned}$$

Thus, we obtain the desired result. □

Proof of Theorem 4.3. Using the same procedure as in the proof of Theorem 2.3, it is shown that

$$\left. \frac{\partial \widetilde{\mathrm{ELR}}(\theta_1, \theta)}{\partial \theta} \right|_{\theta = \widetilde{\theta}^*_{1n}} \xrightarrow{\mathcal{P}} 0$$

under Π_1 as $n \to \infty$, where $\widetilde{\theta}_{1n}^*$ is defined as $\widetilde{\theta}_{1n}^* = \theta_1 - x_n^{-1}\widetilde{F}(\theta_1|\eta)^{-1}\widetilde{H}(\theta_1|\eta)\xi$. On the other hand, by Taylor's theorem we can also see that there exists $c \in (0,1)$ such that

$$\begin{aligned} &x_n^2\widetilde{\mathrm{ELR}}(\theta_1, \widetilde{\theta}_{1n}) \\ &= x_n^2\widetilde{\mathrm{ELR}}(\theta_1, \widetilde{\theta}_{1n}^*) + x_n^2(\widetilde{\theta}_{1n} - \widetilde{\theta}_{1n}^*)' \frac{\partial \widetilde{\mathrm{ELR}}(\theta_1, \theta)}{\partial \theta}\bigg|_{\theta = \widetilde{\theta}_{1n}^* + c(\widetilde{\theta}_{1n} - \widetilde{\theta}_{1n}^*)}. \end{aligned}$$

Therefore, from the relationship $\|\widetilde{\theta}_{1n} - \widetilde{\theta}_{1n}^*\|_E = O\left(x_n^{-2}\right)$, we have

$$x_n^2 \left\{ \widetilde{\mathrm{ELR}}(\theta_1, \widetilde{\theta}_{1n}) - \widetilde{\mathrm{ELR}}(\theta_1, \widetilde{\theta}_{1n}^*) \right\} = o_p(1).$$

This implies that we can work with $x_n^2\widetilde{\mathrm{ELR}}(\theta_1, \widetilde{\theta}_{1n}^*)$ instead of $x_n^2\widetilde{\mathrm{ELR}}(\theta_1, \widetilde{\theta}_{1n})$ asymptotically in results of type Theorem 4.2. Therefore, using Theorem 4.2, the assertion is proved. □

6.4 Proofs of Chapter 5

First, we introduce a condition, which is Assumption 3A in Newey [43].

Assumption 6.1 (Newey [43]). *Θ is a metric space with a metric $d(\cdot,\cdot)$ and there exists B_n and $h : [0,\infty) \to [0,\infty)$ such that $B_n = O_p(1)$, $h(0) = 0$, h is continuous at zero and for all θ, $\widetilde{\theta} \in \Theta$, $n = 1, 2, \cdots$,*

$$\left\| \widehat{R}_n(\theta) - \widehat{R}_n(\widetilde{\theta}) \right\|_E \le B_n h \left\{ d\left(\theta, \widetilde{\theta}\right) \right\}.$$

We first show that $\{\widehat{R}_n(\theta) : n = 1, 2, \cdots\}$ satisfies Assumption 6.1. From Assumption 5.2, there exists c and β such that

$$\left\| \widehat{R}_n(\theta) - \widehat{R}_n(\widetilde{\theta}) \right\|_E \le c \left\| \theta - \widetilde{\theta} \right\|_E^\beta$$

for θ and $\widetilde{\theta} \in \Theta$. So Assumption 6.1 is satisfied with $B_n = 1$, $h(x) = cx^\beta$ and $d(\theta, \widetilde{\theta}) = \|\theta - \widetilde{\theta}\|_E$. On the other hand, we can see the pointwise convergence $\widehat{R}_n(\theta) - R(\theta) = o_p(1)$ for each θ, where

$$R(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\omega; \theta)\widetilde{f}(\omega) d\omega$$

and then the limit behavior of $\sup_{\theta \in \Theta} \|\widehat{R}_n(\theta) - R(\theta)\|_E$ is given by Lemma 6.6 below, which is Corollary 2.2 of Newey [43].

Lemma 6.6 (Newey [43]). *Suppose that all assumptions in Theorem 5.1 and Assumption 6.1 hold. Then,*

$$\sup_{\theta\in\Theta}\left\|\widehat{R}_n(\theta)-R(\theta)\right\|_E = o_p(1).$$

Proof of Theorem 5.1. We follow the similar arguments as Theorem 2.1 and 2.6 of Newey and McFadden [44].
(i) Define $q_0(\theta) = -R(\theta)'WR(\theta)$. By the triangle and Cauchy-Schwartz inequality,

$$\begin{aligned}
|q_n(\theta)-q_0(\theta)| \le & \left|\left\{\widehat{R}_n(\theta)-R(\theta)\right\}'\widehat{W}_n\left\{\widehat{R}_n(\theta)-R(\theta)\right\}\right| \\
&+\left|R(\theta)'\left(\widehat{W}_n+\widehat{W}_n'\right)\left\{\widehat{R}_n(\theta)-R(\theta)\right\}\right| \\
&+\left|R(\theta)'\left(\widehat{W}_n-W\right)R(\theta)\right| \\
\le & \left\|\widehat{R}_n(\theta)-R(\theta)\right\|_E^2\left\|\widehat{W}_n\right\|_E \\
&+2\left\|R(\theta)\right\|_E\left\|\widehat{R}_n(\theta)-R(\theta)\right\|_E\left\|\widehat{W}_n\right\|_E \\
&+\left\|R(\theta)\right\|_E^2\left\|\widehat{W}_n-W\right\|_E .
\end{aligned} \tag{6.34}$$

From Lemma 6.6, the right hand side of (6.34) converges to zero as $n\to\infty$ uniformly in $\theta\in\Theta$. Therefore the following inequalities hold with probability approaching one (w.p.a.1) for any $\epsilon>0$;

$$q_0(\widehat{\theta}_{\mathrm{GMM}}) > q_n(\widehat{\theta}_{\mathrm{GMM}}) - \frac{\epsilon}{3}, \tag{6.35}$$

$$q_n(\widehat{\theta}_{\mathrm{GMM}}) > q_n(\theta_0) - \frac{\epsilon}{3}, \tag{6.36}$$

$$q_n(\theta_0) > q_0(\theta_0) - \frac{\epsilon}{3}. \tag{6.37}$$

Here (6.35) and (6.37) follow from the relation $\sup_{\theta\in\Theta}|q_n(\theta)-q_0(\theta)| = o_p(1)$, while (6.36) follows from the definition of the GMM estimator. Therefore, w.p.a.1,

$$q_0(\widehat{\theta}_{\mathrm{GMM}}) > q_0(\theta_0) - \epsilon. \tag{6.38}$$

Let Θ_0 be any open subset of Θ containing θ_0. Since $\Theta\setminus\Theta_0$ is compact, there exists $\check{\theta}\in\Theta\setminus\Theta_0$ such that

$$q_0(\check{\theta}) = \sup_{\theta\in\Theta\setminus\Theta_0} q_0(\theta) < \sup_{\theta\in\Theta} q_0(\theta) = q_0(\theta_0).$$

Therefore, substituting $\epsilon = q_0(\theta_0) - q_0(\check{\theta})$ into (6.38), we can see that

$$q_0(\widehat{\theta}_{\mathrm{GMM}}) > \sup_{\theta \in \Theta \setminus \Theta_0} q_0(\theta)$$

w.p.a.1. That is, $\Pr\{\widehat{\theta}_{\mathrm{GMM}} \in \Theta_0\} \to 1$ as $n \to \infty$, and this is equivalent to $\widehat{\theta}_{\mathrm{GMM}} \xrightarrow{\mathcal{P}} \theta_0$.

(ii) $\widehat{\theta}_{\mathrm{GMM}}$ satisfies p equations

$$\frac{1}{2} \frac{\partial q_n(\theta)}{\partial \theta}\bigg|_{\theta = \widehat{\theta}_{\mathrm{GMM}}} = \widehat{Q}_n(\widehat{\theta}_{\mathrm{GMM}})' \widehat{W}_n \widehat{R}_n(\widehat{\theta}_{\mathrm{GMM}}) = 0_p, \tag{6.39}$$

where $\widehat{Q}_n(\theta)$ is an $m \times p$ random matrix which is defined as

$$\widehat{Q}_n(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial G(\omega; \theta)}{\partial \theta'} \widetilde{I}_{n,X}(\omega) d\omega.$$

Expanding $\widehat{R}_n(\widehat{\theta}_{\mathrm{GMM}})$ in (6.39) around θ_0, we have

$$0_p = \widehat{Q}_n(\widehat{\theta}_{\mathrm{GMM}})' \widehat{W}_n \left\{ \widehat{R}(\theta_0) + \widehat{Q}_n(\overline{\theta}_n) \left(\widehat{\theta}_{\mathrm{GMM}} - \theta_0 \right) \right\}, \tag{6.40}$$

where $\overline{\theta}_n$ is a mean value vector $\delta\widehat{\theta}_{\mathrm{GMM}} + (1-\delta)\theta_0$ with some $\delta \in (0,1)$. Let 1^* be the indicator function for the event that $\widehat{Q}_n(\widehat{\theta}_{\mathrm{GMM}})' \widehat{W}_n \widehat{Q}_n(\overline{\theta}_n)$ is a nonsingular matrix. Then, (6.40) implies that

$$x_n \left(\widehat{\theta}_{\mathrm{GMM}} - \theta_0 \right) = -1^* \cdot \widehat{H}_n^{-1} \widehat{Q}_n(\widehat{\theta}_{\mathrm{GMM}})' \widehat{W}_n \left\{ x_n \widehat{R}_n(\theta_0) \right\} + (1 - 1^*) x_n \left(\widehat{\theta}_{\mathrm{GMM}} - \theta_0 \right),$$

where $\widehat{H}_n = \widehat{Q}_n(\widehat{\theta}_{\mathrm{GMM}})' \widehat{W}_n \widehat{Q}_n(\overline{\theta}_n)$. (6.40) also asserts that $x_n(\widehat{\theta}_{\mathrm{GMM}} - \theta_0)$ is bounded, because it will be shown that $x_n \widehat{Q}_n(\widehat{\theta}_{\mathrm{GMM}})' \widehat{W}_n \widehat{R}(\theta_0)$ is $O_p(1)$ and $\widehat{H}_n$ converges to non-zero constant matrix. Since $\widehat{\theta}_{\mathrm{GMM}} \xrightarrow{\mathcal{P}} \theta_0$, we have $\overline{\theta}_n \xrightarrow{\mathcal{P}} \theta_0$. On the other hand, we have $\sup_{\theta \in \Theta} |\widehat{Q}_n(\theta) - Q(\theta)| = o_p(1)$ by the same argument as we did for $\widehat{R}(\theta)$. Therefore,

$$\begin{aligned} \left| \widehat{Q}_n(\widehat{\theta}_{\mathrm{GMM}}) - Q(\theta_0) \right| &\le \left| \widehat{Q}_n(\widehat{\theta}_{\mathrm{GMM}}) - Q(\widehat{\theta}_{\mathrm{GMM}}) \right| + \left| Q(\widehat{\theta}_{\mathrm{GMM}}) - Q(\theta_0) \right| \\ &\le \sup_{\theta \in \Theta} \left| \widehat{Q}_n(\theta) - Q(\theta) \right| + o_p(1) \\ &= o_p(1). \end{aligned}$$

Thus, $\widehat{Q}_n(\overline{\theta}_n)$ and $\widehat{Q}_n(\widehat{\theta}_{\text{GMM}})$ converge to $Q(\theta_0)$ in probability. Recalling that $\widehat{W}_n \xrightarrow{\mathcal{P}} W$, with W being positive definite, and that $Q(\theta_0)$ has full-column rank, we have $1^* \xrightarrow{\mathcal{P}} 1$. It is seen that

$$-\widehat{H}_n^{-1}\widehat{Q}_n(\widehat{\theta}_{\text{GMM}})'\widehat{W}_n\left\{x_n\widehat{R}_n(\theta_0)\right\} \xrightarrow{\mathcal{L}} H^{-1}Q(\theta_0)'W\widetilde{V}(\theta_0)$$

by Lemma 3.2 and symmetry of $S(l)$'s. Moreover, $(1-1^*)x_n(\widehat{\theta}_{\text{GMM}}-\theta_0) \xrightarrow{\mathcal{P}} 0$ by $1^* \xrightarrow{\mathcal{P}} 1$ and boundedness of $x_n(\widehat{\theta}_{\text{GMM}}-\theta_0)$. Therefore, we obtain that

$$x_n(\widehat{\theta}_{\text{GMM}}-\theta_0) \xrightarrow{\mathcal{L}} \{Q(\theta_0)'WQ(\theta_0)\}^{-1}Q(\theta_0)'W\widetilde{V}(\theta_0)$$

as $n \to \infty$. □

Proof of Theorem 5.2.
(i) Under Assumption 5.2, $\phi_i(\omega) = G_i(\omega;\theta_0)\widetilde{f}(\omega)$ $(i = 1,\cdots,m)$ has the following series representation:

$$\phi_i(\omega) = \sum_{k=1}^{\infty} v_i(k)\cos(k\omega),$$

where $v_i(k)$ is the ith element of the vector $v(k)$ defined as (5.2). Then, the (i,j)th element of W_{EL}^{-1} is evaluated as

$$\begin{aligned}
W_{\text{EL}}^{ij} &= \frac{1}{\pi}\int_{-\pi}^{\pi} G_i(\omega;\theta_0)G_j(\omega;\theta_0)\widetilde{f}(\omega)^2 d\omega \\
&= \frac{1}{\pi}\int_{-\pi}^{\pi} \phi_i(\omega)\phi_j(\omega)d\omega \\
&= \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} v_i(k)v_j(l)\frac{1}{\pi}\int_{-\pi}^{\pi}\cos(k\omega)\cos(l\omega)d\omega \\
&= \sum_{l=1}^{\infty} v_i(l)v_j(l). \qquad (6.41)
\end{aligned}$$

(6.41) is the (i,j)th element of $\widetilde{\Sigma}$, hence the proof is completed.
(ii) Writing $Q = Q(\theta_0)$, we can modify $\widetilde{\Omega}(W) - \widetilde{\Omega}(W_{\text{EL}})$ as $\widetilde{\Omega}(W) - \widetilde{\Omega}(W_{\text{EL}}) = D'\widetilde{\Sigma}^{-1}D$, where

$$D = \left\{\widetilde{\Sigma} - Q\left(Q'\widetilde{\Sigma}^{-1}Q\right)^{-1}Q'\right\}WQ(Q'WQ)^{-1}$$

and this implies that $\widetilde{\Omega}(W) - \widetilde{\Omega}(W_{\text{EL}})$ is nonnegative definite.

□

Acknowledgments

The author would like to express his deepest appreciation to his supervisor Professor Masanobu Taniguchi of Waseda University for successive guidance, advices and wholehearted encouragement. Professor Taniguchi supervised the author during his bachelor, master and doctoral courses, and led him to study of fundamental statistical inference and time series analysis. Furthermore, the author leaned from Professor Taniguchi not only academic matter but also the way how to conduct himself as a researcher. Again, I would like to appreciate his six years' leading.

The author acknowledges Professor Takeru Suzuki, Associate Professor Yasutaka Shimizu and Professor Ritei Shibata of Waseda University for their constructive feedback and critical readings of earlier drafts of this dissertation. Especially, the author learned applications of the wavelet analysis from Professor Suzuki in his master course. Also, advice and comments given by Professor Suzuki, Professor Shimizu and Professor Shibata in seminars and academic meetings have been a great help for the author, and their lectures certainly helped this dissertation to be superioir.

The author would like to express his gratitude to Associate Professor Tomoyuki Amano of Wakayama University who took care of our seminar as an adviser when the author was a bachelor student. He also introduced the author how to use a lot of PC softwares; R, S-PLUS, LaTeX, and so on.

The author would like to express sincere thanks to the students of Professor Taniguchi laboratory who worked with him for their continuing discussions and comments in studies. Moreover, thanks are extended to all the other staffs of Graduate School of Fundamental Science and Engineering, Waseda University for their various helps.

Finally, the author thanks with his whole heart to his parents for their tremendous moral support.

Bibliography

[1] Andrews, D. W. (1999). Estimation when a parameter is on a boundary. *Econometrica*, 67(6):1341–1383.

[2] Bartolucci, F. (2007). A penalized version of the empirical likelihood ratio for the population mean. *Statistics & probability letters*, 77(1):104–110.

[3] Brillinger, D. R. (1969). Asymptotic properties of spectral estimates of second order. *Biometrika*, 56(2):375.

[4] Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*, volume 36. SIAM.

[5] Capon, J. (1965). An asymptotic simultaneous diagonalization procedure for pattern recognition. *Information and control*, 8(3):264–281.

[6] Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3):305–334.

[7] Chan, N. H. and Liu, L. (2010). Bartlett correctability of empirical likelihood for time series. *J. Japan Statist. Soc*, 40(2):221–238.

[8] Chandra, T. K. and Mukerjee, R. (1991). Bartlett-type modification for Rao's efficient score statistic. *Journal of Multivariate Analysis*, 36(1):103–112.

[9] Davis, R. and Resnick, S. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *The Annals of Probability*, pages 179–195.

[10] Davis, R. and Resnick, S. (1985). More limit theory for the sample correlation function of moving averages. *Stochastic Processes and their Applications*, 20(2):257–279.

[11] Davis, R. and Resnick, S. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *The Annals of Statistics*, pages 533–558.

[12] Drees, H., Haan, d. , L., and Resnick, S. How to make a hill plot. Technical report, Cornell University Operations Research and Industrial Engineering, (1998).

[13] Dzhaparidze, K. O. and Kotz, S. (1986). *Parameter estimation and hypothesis testing in spectral analysis of stationary time series*. Springer-Verlag New York.

[14] Hájek, J., Šidák, Z., and Sen, P. K. (1967). *Theory of rank tests*. Academic press New York.

[15] Hall, P. (1990). Pseudo-likelihood theory for empirical likelihood. *The Annals of Statistics*, pages 121–140.

[16] Hannan, E. J. (1970). *Multiple time series*, volume 38. Wiley. com.

[17] Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054.

[18] Heyde, C. C. (1997). *Quasi-likelihood and its application: a general approach to optimal parameter estimation*. Springer.

[19] Hirukawa, J. (2006). Cluster analysis for non-Gaussian locally stationary processes. *International Journal of Theoretical and Applied Finance*, 9(01):113–132.

[20] Hosoya, Y. (1978). Discrete-time stable processes and their certain properties. *The Annals of Probability*, pages 94–105.

[21] Hosoya, Y. (1982). Harmonizable stable processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 60(4):517–533.

[22] Hosoya, Y. and Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *The Annals of Statistics*, pages 132–153.

[23] Hsing, T. (1991). On tail index estimation using dependent data. *The Annals of Statistics*, 19(3):1547–1569.

[24] Kakizawa, Y. (1996). Discriminant analysis for non-Gaussian vector stationary processes. *Journal of Nonparametric Statistics*, 7(2):187–203.

[25] Kakizawa, Y. (1997). Higher-order Bartlett-type adjustment. *Journal of statistical planning and inference*, 65(2):269–280.

[26] Kakizawa, Y. (2013). Frequency domain generalized empirical likelihood method. *Journal of Time Series Analysis*.

[27] Kakizawa, Y., Shumway, R. H., and Taniguchi, M. (1998). Discrimination and clustering for multivariate time series. *Journal of the American Statistical Association*, 93(441):328–340.

[28] Kitamura, Y. (2001). Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica*, 69(6):1661–1672.

[29] Kitamura, Y. (2006). Empirical likelihood methods in econometrics: Theory and practice.

[30] Kitamura, Y. et al. (1997). Empirical likelihood methods with weakly dependent processes. *The Annals of Statistics*, 25(5):2084–2102.

[31] Kitamura, Y., Santos, A., and Shaikh, A. M. (2012). On the asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica*, 80(1):413–423.

[32] Klüppelberg, C. and Mikosch, T. (1993). Spectral estimates and stable processes. *Stochastic processes and their applications*, 47(2):323–344.

[33] Klüppelberg, C. and Mikosch, T. (1994). Some limit theory for the self-normalised periodogram of stable processes. *Scandinavian journal of statistics*, pages 485–491.

[34] Kluppelberg, C. and Mikosch, T. (1996). The integrated periodogram for stable processes. *The Annals of Statistics*, pages 1855–1879.

[35] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

[36] Lahiri, S. N. and Mukhopadhyay, S. (2012). A penalized empirical likelihood method in high dimensions. *The Annals of Statistics*, 40(5):2511–2540.

[37] Liggett Jr, W. (1971). On the asymptotic optimality of spectral analysis for testing hypotheses about time series. *The Annals of Mathematical Statistics*, pages 1348–1358.

[38] Logan, B. F., Mallows, C., Rice, S., and Shepp, L. (1973). Limit distributions of self-normalized sums. *The Annals of Probability*, 1(5):788–809.

[39] Mandelbrot, B. B. New methods in statistical economics. In *Fractals and Scaling in Finance*, pages 79–104. Springer, (1997).

[40] Mikosch, T., Gadrich, T., Kluppelberg, C., and Adler, R. J. (1995). Parameter estimation for ARMA models with infinite variance innovations. *The Annals of Statistics*, pages 305–326.

[41] Mikosch, T., Resnick, S., and Samorodnitsky, G. (2000). The maximum of the periodogram for a heavy-tailed sequence. *Annals of probability*, pages 885–908.

[42] Monti, A. C. (1997). Empirical likelihood confidence regions in time series models. *Biometrika*, 84(2):395–405.

[43] Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica*, 59(4):1161–1167.

[44] Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.

[45] Newey, W. K. and Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, 72(1):219–255.

[46] Nishikawa, M. and Taniguchi, M. (2009). Discriminant analysis for dynamics of stable processes. *Statistical Methodology*, 6(1):82–96.

[47] Nolan, J. P. (2015). *Stable Distributions - Models for Heavy Tailed Data*. Birkhauser, Boston, In progress, Chapter 1 online at academic2.american.edu/∼jpnolan.

[48] Ogata, H. and Taniguchi, M. (2009). Cressie-Read power-divergence statistics for non-Gaussian vector stationary processes. *Scandinavian Journal of Statistics*, 36(1):141–156.

[49] Ogata, H. and Taniguchi, M. (2010). An empirical likelihood approach for non-Gaussian vector stationary processes and its application to minimum contrast estimation. *Australian & New Zealand Journal of Statistics*, 52(4):451–468.

[50] Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249.

[51] Petrov, V. V. (1975). *Sums of independent random variables*. Berlin.

[52] Pratt, J. W. and Gibbons, J. D. Asymptotic relative efficiency. In *Concepts of Nonparametric Theory*, pages 345–424. Springer, (1981).

[53] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, pages 300–325.

[54] Rao, C. R. and Mukerjee, R. (1995). Comparison of bartlett-type adjustments for the efficient score statistic. *Journal of statistical planning and inference*, 46(2):137–146.

[55] Rao, C. R. and Mukerjee, R. (1997). Comparison of lr, score, and wald tests in a non-iid setting. *Journal of Multivariate Analysis*, 60(1):99–110.

[56] Resnick, S. and Stărică, C. (1997). Asymptotic behavior of hill's estimator for autoregressive data. *Communications in statistics. Stochastic models*, 13(4):703–721.

[57] Sakiyama, K. and Taniguchi, M. (2004). Discriminant analysis for locally stationary processes. *Journal of Multivariate Analysis*, 90(2):282–300.

[58] Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable non-Gaussian random processes: stochastic models with infinite variance*, volume 1. CRC Press.

[59] Shumway, R. (1982). Discriminant analysis for time series. *Handbook of statistics*, 2:1–46.

[60] Shumway, R. and Unger, A. (1974). Linear discriminant functions for stationary time series. *Journal of the American Statistical Association*, 69(348):948–956.

[61] Strasser, H. (1985). *Mathematical theory of statistics: statistical experiments and asymptotic decision theory*, volume 7. Walter de Gruyter.

[62] Tamaki, K. (2007). Power properties of empirical likelihood for stationary processes. *Scientiae Mathematicae Japonicae*, 66(3):359.

[63] Tang, C. Y. and Leng, C. (2010). Penalized high-dimensional empirical likelihood. *Biometrika*, 97(4):905–920.

[64] Taniguchi, M. (1980). On estimation of the integrals of certain functions of spectral density. *Journal of Applied Probability*, pages 73–83.

[65] Taniguchi, M. (1991). *Higher order asymptotic theory for time series analysis.* Springer.

[66] Taniguchi, M. (1991). Third-order asymptomic properties of a class of test statistics under a local alternative. *Journal of Multivariate Analysis*, 37(2):223–238.

[67] Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic theory of statistical inference for time series.* Springer New York.

[68] Zhang, G. and Taniguchi, M. (1994). Discriminant analysis for stationary vector time series. *Journal of Time Series Analysis*, 15(1):117–126.

[69] Zhang, G. and Taniguchi, M. (1995). Nonparametric approach for discriminant analysis in time series. *Journaltitle of Nonparametric Statistics*, 5(1):91–101.

List of papers

1. Akashi, F. (2013). An empirical likelihood approach for discriminant analysis of non-Gaussian vector stationary linear processes. *Scientiae Mathematicae Japonicae Online*, e-2013:645-660.

2. Akashi, F., Liu, Y. and Taniguchi, M. (2014). An empirical likelihood approach for symmetric α-stable processes. *To appear in Bernoulli.*

3. Akashi, F. (2014). Empirical likelihood approach toward discriminant analysis for dynamics of stable processes. *Statistical Methodology* **19** 25-43.