Waseda University

Doctoral Dissertation

presented to the Department of Pure and Applied Mathematics

Asymptotic Theory for Non-standard Estimating Function and Self-normalized Method in Time Series Analysis

時系列解析におけるノンスタンダード推定関数法と
自己基準化法に対する漸近理論

Author:
Yan Liu

Supervisor:
Dr. Masanobu Taniguchi

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

December, 2014
Academic Year 2014/2015

WASEDA UNIVERSITY

DOCTORAL DISSERTATION

Asymptotic Theory for Non-standard Estimating Function and Self-normalized Method in Time Series Analysis

時系列解析におけるノンスタンダード推定関数法と自己基準化法に対する漸近理論

December, 2014

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

Yan LIU

劉 言

Declaration of Authorship

I, Yan Liu, declare that this thesis titled, 'Asymptotic Theory for Non-standard Estimating Function and Self-normalized Method in Time Series Analysis' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ____________________

Date: 18 November 2014

"Far better an approximate answer to the right question, than the exact answer to the wrong question, which can always be made precise."

John Tukey

"Advice is what we ask for when we already know the answer but wish we didn't."

Erica Jong

WASEDA UNIVERSITY

Abstract

Graduate School of Fundamental Science and Engineering
Department of Pure and Applied Mathematics

Doctor of Philosophy

Asymptotic Theory for Non-standard Estimating Function and Self-normalized Method in Time Series Analysis

by Yan Liu

In this dissertation, we develop asymptotic theory for non-standard estimating function and self-normalized method in time series analysis. Time series analysis can be divided into two parts: time domain analysis and frequency domain analysis. In time domain analysis, we extend the estimating function to be non-differentiable one for the scale parameter and the coefficients in the model. The idea is also applied to generalize the definition of periodogram. In frequency domain analysis, we explore the parameter estimation based on minimum contrast function with exotic disparity. We also give a counterexample which shows that the functional based on periodogram for estimation is not always asymptotically normal. As an extension, we derive the asymptotic distribution of empirical likelihood ratio statistics based on the Whittle disparity for the time series models with infinite variance. We complement the statistical inference for pivotal quantity based on empirical likelihood method by estimating the tail index with self-normalized method.

Acknowledgements

First, I would like to express my most heartfelt gratitude to my supervisor, Professor Masanobu Taniguchi, for five whole years with his constant advice and encouragement, no doubt, the best fortune in my future research. Needless to say, he shared with me countless ideas in time series analysis and the orientation of my research. Besides, I am grateful to his guidance on how to explain my study in a seminar, how to ask a question to overseas distinguished researchers, how to present the product of my research in an academic meeting, how to summarize my results in a paper and even how to write a proposal for funding, and furthermore all the chances he gave. It is the only regretful thing for me that I have not attended his lecture. All things, however, I learned from his daily advice and research seminar are irreplaceable, and must be more than a lecture.

I would also like to express my deep gratitude to Professor Takeru Suzuki, who introduced me to the field of probability and statistics. In addition, he taught me the stochastic process, wavelet analysis and the asymptotic theory for statistics. The books he recommended to me are still my most favorite ones.

I am also grateful to Professor Yasutaka Shimizu for his critical questions on my dissertation and to Professor Ritei Shibata for his comments on the appearance and the logical order of my dissertation. My dissertation, of course, is improved by their precious time.

Some of my special thanks go to Professor Yu Hayakawa in School of International Liberal Studies in Waseda University. She gave me opportunities to study the situation on the life insurance in Japan with Professor Paul Yip in the University of Hong Kong, taxation statistics with Professor Maji Rhee in Waseda University and work on the numerical studies of warranty with her. My other special thanks go to Professor Fusa Katada, the director of Center for English Language Education in Waseda University. I studied the contingency table in the practical situations with her. She spared much time for improvement in my English and consultation on my overseas research.

My research has also benefitted from much advice from the attitude of mind to the statistical methods by many researchers. In no particular order, I would like to thank Professor Yoshihide Kakizawa, Professor Takayuki Shiohama, Professor Kiyoshi Inoue, Professor Junichi Hirukawa, Professor Hiroshi Shiraishi, Professor Yoichi Miyata, Professor Hiroaki Ogata, Professor Tomoyuki Amano, Professor Hiroyuki Taniai. I would also like to thank Professor Yoichi Nishiyama for his lecture on martingale theory and priceless notes on mathematics, Professor Hironori Fujisawa for his insightful explanation on the method of divergence, Professor Kenichiro Shimatani for his experience on the model selection by AIC, Professor Toshihiro Abe for his advice on the functional analysis, Professor Taiji Suzuki for his comment on the multicollinearity of the criterion based on the l_1-norm, Professor Masaaki Fukasawa for his talk on the Whittle likelihood for high-frequency data, and furthermore Professor Takeaki Kariya for his general notes on statistics.

I am also indebted to the researchers visiting Waseda University. I learned the methods on the infinite variance processes from Professor Claudia Klüppelberg, Professor Thomas Mikosch and Professor Richard A. Davis, the perspective of the quantile regression from Professor Roger Koenker, the stochastic dominance and smoothed quantile regression from Professor Yoon-Jae Whang, the local linear quantile regression from Professor Ngai-hang Chan and bayesian inference and DIC from Professor Cathy W. S. Chen. I received thoughtful questions from all of them for my future work. I also learned the origin of the empirical likelihood method in time series analysis from Professor Anna Clara Monti, the theory for quasi-martingale and cube root asymptotics from Professor Prakasa Rao and the estimation based on the diffusion process from Professor Ilia Negri. The most memorable visiting researcher for me is Professor Thomas Diccio, who gave me a set of very nice and considerate presents for my daily study.

I would also like to express my sincere gratitude to my colleague, Fumiya Akashi, for his valuable and helpful advice for my study. The collaborative work with him has encouraged my great growth in my academic study. Not only my research, he also willingly helped me over and over again in my daily life. I would also thank Yujiro Ishimura, Kenta Hamada, Tsuyoshi Kato, Hideaki Nagahata, Yoshihiro Suto, Junichi Yokozuka, Hiroyuki Ando, Tokihiro Watanabe and many others for academic discussion, which much influenced myself in an invisible form. Moreover, I am very grateful to Hiroaki Odashima, who persuaded me to come to this academic world. He did not only always give me a good incentive for study, but also openly showed me all the material he had.

Needless to say, I really appreciate the funding support by the Early Bird Program Fellowships from Waseda University in 2013 and Grand-in-Aid for JSPS Fellows: 26·7404 from 2014. The appreciation can be naturally extended to all the other staffs in Department of Pure and Applied Mathematics for their devotion. Especially, I would like to thank Professor Yoshio Yamada, who taught me the first mathematics class in university: Calculus, and Theory of Ordinary Differential Equations, and still have concerns for me.

Finally and mostly, there are not enough words to express my gratitude to both my parents. Even if I could find a remarkable dictionary, there would not be enough room in this dissertation. *Thank you!*

Contents

List of Figures

List of Tables

Abbreviations

i.i.d.	**i**ndependent and **i**dentically **d**istributed (as)
MA	**M**oving **A**verage model
AR	**A**uto**R**egressive model
ARMA	**A**uto**R**egressive **M**oving **A**verage model
S.V.	**S**tochastic **V**olatility model
RMSE	**R**oot **M**ean **S**quare **Error**

Notations

A_j	the jth element of vector A
A_{ij}	the (i,j)th element of matrix A
A^T	the transpose of matrix A
$\bar{A}$	the matrix with complex conjugated entries of matrix A
A^*	the conjugate transpose of matrix A
A_n	the sample of random variables A
$\mathbb{1}(A)$	the indicator function for the set A
$\hat{a}$, $\hat{a}_n$	the estimator of a
α	tail index (or characteristic exponent)
$B(\omega)$	Brownian motion
$\mathcal{B}(t)$	the σ-field generated by $\{\epsilon(n); n \leq t\}$
$b(j)$	AR(∞) coefficient
$\boldsymbol{\beta}_0$	the regression parameter
$C(m)$	the sample covariance
$\Gamma(x)$	the gamma function
$\gamma(k)$	the autocovariance function of the stochastic process
$D(f,g)$	disparity between the functions f and g
$\delta(i,j)$	the Kronecker delta
E	expectation operator
$\{e(t)\}$	the residual process of the stochastic process
$\{\epsilon(t)\}$	innovation process
$\{e(t;\boldsymbol{\eta})\}$	parametrized residual process
$F(\mu)$	the spectral distribution function
$F_t(x)$	the marginal distribution of $X(t)$
$F_{st}(x_1,x_2)$	the bivariate marginal distribution of $(X(s),X(t))$
$\mathcal{F}$	the family of all spectral densities
$\mathcal{F}(\Theta)$	the subfamily of spectral densities indexed by Θ
$f(\omega)$	the spectral density of the stochastic process
$f_{\boldsymbol{\theta}}(\omega)$, $f(\omega;\boldsymbol{\theta})$	parametrized spectral density of the stochastic process
$G(j)$	coefficient in linear process

$g(x)$	the density function of stable distribution
$g(\boldsymbol{\theta})$	the subgradient function of $\rho(\boldsymbol{\theta})$
$\boldsymbol{g}(\omega)$	the power transfer function
H	Hessian matrix
$\boldsymbol{\eta}$	parameters for AR(∞) coefficients
I_d	the d-dimensional identity matrix
$I_{n,A}(\omega)$	periodogram for the process $\{A(t)\}$
$J^{(n)}(\lambda)$	the empirical special distribution
$K(\cdot)$	contrast function
κ_4	the fourth order cumulant of $X(t)$
$L_n(\lambda)$	the Laplace periodogram
Λ	set of frequency parameters
λ_j	the Fourier frequencies $2\pi j/n \in [-\pi, \pi]$
$\boldsymbol{m}(\lambda_t; \boldsymbol{\theta})$	estimating function in the empirical likelihood ratio
μ, $\mu(t)$	the population mean of the stochastic process
$P(A)$	probability of the event A
p	number of parameters
$\rho(\boldsymbol{\theta})$	convex function defined for the estimation
ρ_a	the partial derivative function with respective to a
$Q(\boldsymbol{\theta})$	the objective function for estimation
$Q_n(\lambda)$	quantile periodogram
$Q_\epsilon(t_1, t_2, t_3)$	the joint fourth order cumulant of the process $\{\epsilon(t)\}$
$\tilde{Q}_\epsilon(\omega_1, \omega_2, \omega_3)$	the fourth order spectral density
$R(\boldsymbol{\theta})$	the empirical likelihood ratio
σ^2	the variance of innovation process
ς	the true scale parameter
T	functional defined on $\mathcal{F}(\Theta)$
τ	the true quantile
V	covariance matrix of the score function
$\{X(t)\}$	the stochastic process in observation
$\xi(\boldsymbol{\theta})$	the derivative of $\rho(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$
$\Psi(j)$	coefficient matrix in linear process
Θ	set of all parameters
θ_1	scale parameter
$\boldsymbol{\theta}^0$	the true parameter
$\hat{\boldsymbol{\theta}}_n$	estimator for the parameter
ζ	misestimated parameter

Chapter 1

Introduction

This doctoral thesis develops the statistical inference theory including non-standard estimating function and self-normalized method in time series analysis. Generally, the statistical inference consists of point estimation and hypothesis testing on independent identically distributed samples. In this dissertation, we relax the restriction of the condition of independence but add the stationary property to the assumption. For second order stationary time series models, they can always be represented by linear models from Wold's decomposition theorem. Therefore, we focus on the asymptotic theory of the statistical inference in time series analysis, which furthermore can be divided into two parts: time domain approach and frequency domain approach.

This dissertation consists of 7 chapters. All chapters are self-contained. The list of chapters is given below:

Chapter 1 Introduction

Chapter 2 M-estimation in time series analysis

Chapter 3 Asymptotic properties of generalized spectral via M-estimators

Chapter 4 Parameter estimation based on minimum contrast estimators

Chapter 5 Quantile estimation in frequency domain

Chapter 6 Empirical likelihood method for time series with infinite variance

Chapter 7 Tail index estimation

As the structure of this dissertation, we start the argument from time domain approach and turn our mind to the frequency domain approach step by step. Empirical likelihood method, a nonparametric method involving estimating functions, is preceded by the parametric approach mentioned in the sequel of M-estimation in time domain and minimum contrast estimation in frequency domain. We give the main scope of this dissertation below.

In Chapter 2, we derive the asymptotic distribution of M-estimators for time series models in time domain. It is remarkable that we do not only consider the problem for the regular case in Huber's sense, but also more generally, derive the limiting results from the viewpoint of Lehmann. The M-estimation is a general parametric approach, which includes several optimization methods like maximum likelihood method or minimal distance method. Asymptotic normality of the estimation is given without the smoothness around the neighborhood of true parameter. As a special case, maximum likelihood estimator for Gaussian process is also included in our result. Not only location parameter, we also include the scale parameter in our estimating functions. We can simultaneously estimate the (absolute) moment of innovation process as we estimate the coefficients of the model. Taking the Koenker and Bassett's check function, a non-standard estimating function, we obtain the asymptotic results of estimation for coefficients and the absolute first moment of innovation process.

In Chapter 3, we investigate the Laplace periodogram and further the quantile periodogram, which are recently proposed from their robust characteristics. In several equivalent definitions of the quantile periodogram, we define the quantity as finite Fourier transformation of a plug-in statistic defined by the estimating function corresponding to the quantile. We used a general form for the estimating function and derived the asymptotic distribution of the generalized periodograms with different frequencies are independently exponential.

In Chapter 4, we obtain asymptotic properties of parameter estimation based on minimum contrast estimators in frequency domain. The difference with previous work is that we do not use the direct location or scale disparity for minimum contrast estimation. The approach we took is to extend the form of prediction error and interpolation error in a consistent way and then obtain an exotic disparity, which is not familiar for estimation. Asymptotic normality is derived from the integration functional of periodogram and we found the case corresponding to minimizing the prediction error is the best in the sense of efficiency of the estimation.

In Chapter 5, we expand the idea of quantile in time domain to frequency domain. The estimating function based on the integration functional of periodogram is usually believed to lead an asymptotic result of normal. As for the quantile estimation in frequency domain, we found it is a counterexample to the belief and derived the asymptotic distribution consisting of a sandwich form made by a normal distributed random variable inserted between two exponential random variables. For estimation, if we smooth the periodogram by some window functions, then asymptotic normality of quantile estimation in frequency domain is still obtained.

In Chapter 6, we employ the empirical likelihood approach to test a hypothesis containing a pivotal quantity observed in a heavy tailed process. The estimating function involved in the empirical likelihood statistic is defined by the form of prediction error, which is known as the best from Chapter 4. The asymptotic distribution can be simply summarized by a squared form of a ratio of two type stable distributed random variables with different exponent of α and $\alpha/2$. If we know α well, then the empirical likelihood statistic gives a well-behaved confidence domain for the pivotal quantity.

Chapter 7 is a complement of Chapter 6 for estimating the exponent of innovation process. We call the exponent by tail index in a general way. If a sequence of random variables belongs to the domain of attraction α, then we can calculate the moments of the limiting distribution of the self-normalized random variables by the formula we proposed in this chapter. As a result, we can use the moment estimator to estimate the innovation process from its simplicity. The result shows its stability around an important bound 2 even for dependent random variables.

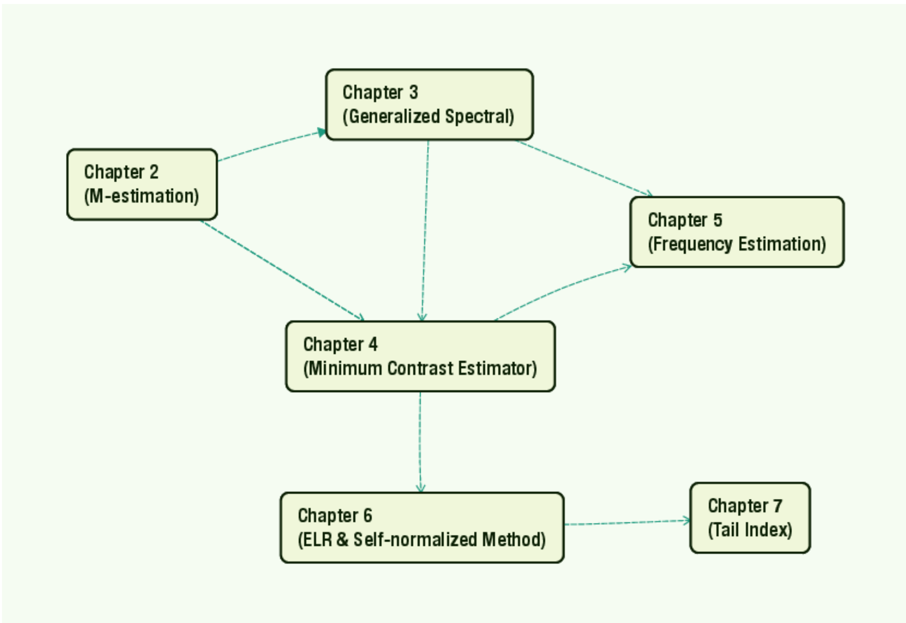


FIGURE 1.1: Dependence chart of all chapters.

Figure 1.1 shows the relationship between all subsequent chapters. We treat a general approach of estimating function in Chapter 2 in time domain. This idea of estimating function can be generalized to the concept of the spectral density given in Chapter 3. This is the first step toward the frequency domain approach. On the other hand, the estimating function studied in frequency domain is used in integration functional form, and usually called "minimum contrast estimator", which is given in Chapter 4. We give a special aspect of estimation by integration functional around the frequency estimation of the spectral distribution function in Chapter 5. In Chapter 6, we use the most efficient form of the minimum contrast estimator as estimating function in empirical likelihood ratio and show asymptotic distribution of the statistics by self-normalized method. Lastly, we investigate the moments of the limiting distribution of the self-normalized random variables and link the result to the estimation of tail index.

Chapter 2

M-estimation in Time Series Analysis

2.1 Introduction

In Chapter 2, we review a general approach of estimating functions, M-estimation, before going into non-standard case. Historically, Hodges and Lehmann (1963) proposed a lemma on the asymptotic normality for the estimation of location of independent identically distributed (i.i.d.) samples. Huber (1964) generalized the idea into M-estimation concept and investigated the robustness of the estimators.

In time series settings, there are two types of M-estimation, one is done in time domain and the other is in frequency domain. M-estimation in time domain aims at the maximum likelihood approach and eventually settles on the quasi-likelihood method. On the other hand, M-estimation in frequency domain aims at the search of the structure of the model. In Chapter 2, we work on M-estimation in time domain, but first give a short review on the frequency domain for the comparison in the following. For the regularity, we always assume the considered process $\{X(t)\}$ is second order stationary. Under the condition, the process $\{X(t)\}$ has its own spectral density $f(\lambda)$ in the frequency domain. The parameter estimation is based on the periodogram

$$I_{n,X}(\omega) = \frac{1}{2\pi n}\Big\{\sum_{t=1}^{n} X(t)e^{it\omega}\Big\}\Big\{\sum_{t=1}^{n} X(t)e^{it\omega}\Big\}^*,$$

which is defined on the observed stretch $\{X(1), \ldots, X(n)\}$. The Whittle estimator, the origin of semiparametric estimation, is defined as the minimizer of

$$D(f_{\boldsymbol{\theta}}, I_{n,X}) = \int_{-\pi}^{\pi} [\log\det\{f_{\boldsymbol{\theta}}(\omega)\}] + \mathrm{tr}\{I_{n,X}(\omega)f_{\boldsymbol{\theta}}(\omega)^{-1}\}d\omega.$$

The asymptotic normality of the Whittle estimator under regular conditions was shown in Dzhaparidze (1971), Dunsmuir (1979), Hannan (1973b) and Hosoya and Taniguchi (1982). For semiparametric estimation, see Beran (1978).

In this chapter, we suppose the process $\{X(t)\}$ has a unique one-sided autoregressive (AR) representation in time domain as

$$\sum_{j=0}^{\infty} b(j)(X(t-j)-\mu) = \epsilon(t),$$

with $b(0) = 1$. The process is second order stationary with independent $\epsilon(t)$ identically distributed as $(0, \sigma^2)$, and may has some "nice" structures for dependent data, such as ergodicity and mixing conditions. For example, the asymptotics of the process under mixing conditions is well investigated in Ibragimov and Linnik (1971).

Even if the the process is second order stationary nonlinear, the process can be decomposed into a linear part and a deterministic part (Wold's decomposition), and in turn it has AR(∞) representation since the process is invertible (See Brockwell and Davis (1991), p.90).

For M-estimation, the problem may be generalized as follows. We assume that the AR(∞) representation of the model is characterized by a finite dimensional vector $\boldsymbol{\theta} = (\theta_1, \boldsymbol{\eta}^T)^T$. That is to say, the model is defined by

$$\sum_{j=0}^{\infty} b(j;\boldsymbol{\eta})X(t-j) = \epsilon(t),$$

where $b(0) = 1$ and $\{\epsilon(t)\}$ is i.i.d. $(0, \theta_1)$. The parameter θ_1 denotes the scale parameter of the model, and the remaining parameters $(\theta_2, \ldots, \theta_p)^T$ is denoted by $\boldsymbol{\eta}$. The true parameter is represented by $\boldsymbol{\theta}^0$.

Define the residual process $e(t;\boldsymbol{\eta})$ and its corresponding form in the following way:

$$e(t;\boldsymbol{\eta}) = \sum_{j=0}^{t-1} b(j;\boldsymbol{\eta})X(t-j), \qquad r(t;\boldsymbol{\theta}) = \theta_1^{-1/2} e(t;\boldsymbol{\eta}), \tag{2.1.1}$$

$$\epsilon(t;\boldsymbol{\eta}) = \sum_{j=0}^{\infty} b(j;\boldsymbol{\eta})X(t-j), \qquad v(t;\boldsymbol{\theta}) = \theta_1^{-1/2} \epsilon(t;\boldsymbol{\eta}). \tag{2.1.2}$$

The approximate maximum likelihood estimator is equivalent to finding the solution of

$$\sum_{t=2}^{n} r(t;\boldsymbol{\theta})\partial_{\boldsymbol{\theta}} r(t;\boldsymbol{\theta}) - c(\boldsymbol{\theta}) = \mathbf{0},$$

which is considered in Beran (1994). They extend the estimator to a class of

Z-estimators and investigate the robustness under Gaussian long-memory situation.

We look into the properties of M-estimators in both non-differentiable and differentiable cases of the objective functions. Even we showed the asymptotic normality of M-estimator with non-differentiable objective function under some regular conditions, the asymptotic variance matrix of the estimator is not explicitly obtained. For this reason, we give the asymptotic variance matrix of the estimator under the differentiable conditions. The objective function is considered to be convex, so the corresponding M-estimator includes LAD estimators (see Niemiro (1992)). We also extend the result to $\mathrm{M_m}$-estimator since the class is much richer. The class includes Oja's median and even Hodges-Lehmann's estimators of location (see Bose (1998)). The proof is similar to the method for U statistics in depend case, which is in order examined by Hoeffding (1948), Sen (1972), Yoshihara (1976) and Denker and Keller (1983).

This chapter is organized as follows. In section 2.2, we review the sufficient conditions for asymptotic normality of M-estimators. Without the condition of differentiability for the convex objective function, we derive asymptotic normality of M-estimators in time series settings under the new class of conditions in 2.3. Also, the asymptotic result is given in the detailed way if the objective function is differentiable in Section 2.4. In section 2.5, we extend the result to $\mathrm{M_m}$-estimators. Section 2.6 contains two important cases of the inference in time series analysis as examples of the main result. We give some numerical results in Section 2.7.

2.2 Asymptotic Normality of M-estimators

First we revisit the work of Niemiro (1992) and Hodges and Lehmann (1963) in this section.

Suppose ϵ, $\epsilon(1)$, …, $\epsilon(n)$ are i.i.d. random variables. Let $\rho(\boldsymbol{\theta}, \epsilon)$ be a real function defined for $\boldsymbol{\theta} \in \mathbb{R}^p$ and $\boldsymbol{g}(\boldsymbol{\theta}, \epsilon)$ be a subgradient of $\rho(\boldsymbol{\theta}, \epsilon)$. Define

$$Q(\boldsymbol{\theta}) = E\rho(\boldsymbol{\theta}, \epsilon).$$

The empirical analog of $Q(\boldsymbol{\theta})$ is defined by

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \rho(\boldsymbol{\theta}, \epsilon(i)).$$

Remark 2.2.1. The definition of $Q_n(\boldsymbol{\theta})$ is optimal in the regular class of estimating functions for i.i.d. samples. For details, see Godambe (1960), Godambe and Thompson (1978) and Godambe and Thompson (1984).

Denote the corresponding score function $\boldsymbol{\xi}_n(\boldsymbol{\theta})$. Usually, $\boldsymbol{\xi}_n(\boldsymbol{\theta})$ is considered as $\nabla Q_n(\boldsymbol{\theta})$, where ∇ is an operation which means the differentiation with respect to $\boldsymbol{\theta}$. In the case of nonexistence of unique $\hat{\boldsymbol{\theta}}_n$, set the ith element $\hat{\theta}_{ni}$ of minimizer $\hat{\boldsymbol{\theta}}_n$ satisfying

$$\hat{\theta}_{ni} = \alpha_i \theta^*_{ni} + (1-\alpha_i)\theta^{**}_{ni},$$

for $0 \leq \alpha_i \leq 1$ $(i = 1, \ldots, p)$, where

$$\begin{aligned} \theta^*_{ni} &= \sup\{r : \xi_{ni}(r) \geq 0\}, \\ \theta^{**}_{ni} &= \inf\{r : \xi_{ni}(r) \leq 0\}. \end{aligned}$$

Assumption 2.2.2 (Niemiro (1992))**.**

(i) $\rho(\boldsymbol{\theta}, \epsilon)$ is convex with respect to $\boldsymbol{\theta}$ for each fixed ϵ.

(ii) $Q(\boldsymbol{\theta})$ is well defined, that is, the expectation exists and is finite for all $\boldsymbol{\theta}$.

(iii) $\boldsymbol{\theta}^0$ satisfying $Q(\boldsymbol{\theta}^0) = \min_{\boldsymbol{\theta}} Q(\boldsymbol{\theta})$ exists and is unique.

(iv) $E|\boldsymbol{g}(\boldsymbol{\theta}, \epsilon)|^2 < \infty$ for each $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}^0$.

(v) $Q(\boldsymbol{\theta})$ is twice differentiable at $\boldsymbol{\theta}^0$ and $\nabla^2 Q(\boldsymbol{\theta}^0)$ is positive definite.

Assumption 2.2.3 (Hodges and Lehmann (1963), Inagaki and Kondo (1980))**.**

(i) $\boldsymbol{\xi}_n(\boldsymbol{\theta})$ is a non-decreasing function of every element of $\boldsymbol{\theta}$.

(ii) For any vector valued $\boldsymbol{u}$, it holds that

$$\sqrt{n}(\boldsymbol{\xi}_n(\boldsymbol{\theta}^0 + \boldsymbol{u}/\sqrt{n}) - \boldsymbol{\xi}_n(\boldsymbol{\theta}^0)) \xrightarrow{\mathcal{P}} H^T \boldsymbol{u},$$

where H is a positive definite matrix.

(iii) $\sqrt{n}\boldsymbol{\xi}_n(\boldsymbol{\theta}^0) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, V)$.

Lemma 2.2.4. *Under each of Assumption 2.2.2 or Assumption 2.2.3, it holds that*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, H^{-1} V H^{-1}),$$

where $H = \nabla^2 Q(\boldsymbol{\theta}^0)$ *and* $V = \operatorname{Var} \boldsymbol{g}(\boldsymbol{\theta}^0)$.

Proposition 2.2.5. *If* $\rho(\boldsymbol{\theta}, \epsilon)$ *is differentiable with respect to* $\boldsymbol{\theta}$ *in a neighborhood of* $\boldsymbol{\theta}^0$*, then Assumption 2.2.2 implies Assumption 2.2.3.*

Proof. From Assumption 2.2.2 (i) and (ii), we see that $\boldsymbol{\xi}_n(\boldsymbol{\theta})$ is a non-decreasing function of all elements of $\boldsymbol{\theta}$. Next, according to Niemiro's proof, we see, for each $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}^0$,

$$Q_n(\boldsymbol{\theta}+\frac{\boldsymbol{u}}{\sqrt{n}})-Q_n(\boldsymbol{\theta})-\frac{\boldsymbol{u}^T}{\sqrt{n}}\nabla Q_n(\boldsymbol{\theta})-Q(\frac{\boldsymbol{u}}{\sqrt{n}},\boldsymbol{\theta})\xrightarrow{\mathcal{P}}0. \tag{2.2.1}$$

Differentiate the equation with respect to $\boldsymbol{\theta}$ above and substitute $\boldsymbol{\theta}=\boldsymbol{\theta}^0$, and

$$\boldsymbol{\xi}_n(\boldsymbol{\theta}^0+\frac{\boldsymbol{u}}{\sqrt{n}})-\boldsymbol{\xi}_n(\boldsymbol{\theta}^0)-\frac{\boldsymbol{u}^T}{\sqrt{n}}\nabla^2 Q_n(\boldsymbol{\theta}^0)\xrightarrow{\mathcal{P}}0.$$

From Assumption 2.2.2 (v), it holds that

$$\nabla^2 Q_n(\boldsymbol{\theta}^0)\xrightarrow{\mathcal{P}}\nabla^2 Q(\boldsymbol{\theta}^0),$$

and by Theorem 4.1 of Billingsley (1968), we obtain the desirable result:

$$\sqrt{n}(\boldsymbol{\xi}_n(\boldsymbol{\theta}^0+\boldsymbol{u}/\sqrt{n})-\boldsymbol{\xi}_n(\boldsymbol{\theta}^0))\xrightarrow{\mathcal{P}}\nabla^2 Q(\boldsymbol{\theta}^0)^T\boldsymbol{u}.$$

Under Assumption 2.2.2 (iv), we have the last result,

$$\sqrt{n}\boldsymbol{\xi}_n(\boldsymbol{\theta}^0)\xrightarrow{\mathcal{L}}\mathcal{N}(0,V).$$

□

Remark 2.2.6. We mainly show the result by (2.2.1) in the subsequent section. It is sufficient for the asymptotics. (See Niemiro (1992).)

2.3 Estimation in Linear Time Series Models

For time series model, we consider $\rho(\theta_1^{-1/2},e(t;\boldsymbol{\eta}))$, which sometimes is written as $\rho(\boldsymbol{\theta})$ for short. In this section, we only assume $\rho(\boldsymbol{\theta})$ is convex with respect to each parameter. Generally, consider $\boldsymbol{\theta}=(\theta_1,\boldsymbol{\eta}^T)^T$ in a compact set $\Theta\subset\mathbb{R}^m$, where $\boldsymbol{\eta}=(\theta_2,\dots,\theta_p)^T$. For simplicity, we write

$$Q(\boldsymbol{\theta})=E\rho(\theta_1^{-1/2},e(t;\boldsymbol{\eta})). \tag{2.3.1}$$

The true value of $\boldsymbol{\theta}$, represented by $\boldsymbol{\theta}^0$, is defined by

$$Q(\boldsymbol{\theta}^0)=\min_{\boldsymbol{\theta}\in\Theta}Q(\boldsymbol{\theta}). \tag{2.3.2}$$

The sample version corresponding to the objective function is

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\rho(\theta_1^{-1/2}, e(t;\boldsymbol{\eta})),$$

and

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\Theta} Q_n(\boldsymbol{\theta}). \tag{2.3.3}$$

Assumption 2.3.1.

(i) $\rho(\boldsymbol{\theta})$ is convex with respect to $\boldsymbol{\theta}$.

(ii) $Q(\boldsymbol{\theta})$ is well defined.

(iii) $\boldsymbol{\theta}^0$ satisfying (2.3.1) exists and is unique.

Since $\rho(\boldsymbol{\theta})$ is a convex function, there exists a subgradient of $\rho(\boldsymbol{\theta})$, which is denoted by $\boldsymbol{g}(\boldsymbol{\theta})$ satisfying

$$\rho(\boldsymbol{\alpha}) + (\boldsymbol{\beta}-\boldsymbol{\alpha})^T\boldsymbol{g}(\boldsymbol{\alpha}) \le \rho(\boldsymbol{\beta})$$

for all $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^d$. Without loss of generality, we consider the case $\boldsymbol{\theta}^0 = \mathbf{0}$ and $Q(\mathbf{0}) = \mathbf{0}$. It is easy to see that

$$\begin{aligned}\boldsymbol{\alpha}^T\boldsymbol{g}(\mathbf{0}) \le \rho(\boldsymbol{\alpha}) - \rho(\mathbf{0}) \le \boldsymbol{\alpha}^T\boldsymbol{g}(\boldsymbol{\alpha}),\\ 0 \le \rho(\boldsymbol{\alpha}) - \rho(\mathbf{0}) - \boldsymbol{\alpha}^T\boldsymbol{g}(\mathbf{0}) \le \boldsymbol{\alpha}^T(\boldsymbol{g}(\boldsymbol{\alpha}) - \boldsymbol{g}(\mathbf{0})).\end{aligned}$$

For $\boldsymbol{\alpha} = (\alpha_1, \boldsymbol{\alpha}_2^T)^T$, we use the following symbols for the simplicity of the notations:

$$\begin{aligned}\rho(t;\frac{\boldsymbol{\alpha}}{\sqrt{n}}) &:= \rho((\boldsymbol{\theta}_1^0 + \frac{\alpha_1}{\sqrt{n}})^{-1/2}, e(t;\boldsymbol{\eta}^0 + \frac{\boldsymbol{\alpha}_2}{\sqrt{n}})),\\ \rho(t;\mathbf{0}) &:= \rho((\boldsymbol{\theta}_1^0)^{-1/2}, e(t;\boldsymbol{\eta}^0)) = \rho((\boldsymbol{\theta}_1^0)^{-1/2}, \epsilon).\end{aligned}$$

$\boldsymbol{g}(t;\cdot)$ is also defined in the same way. At neighborhood of $\boldsymbol{\theta}^0$, we have, for each t,

$$0 \le \rho(t;\frac{\boldsymbol{\alpha}}{\sqrt{n}}) - \rho(t;\mathbf{0}) - \frac{\boldsymbol{\alpha}^T}{\sqrt{n}}\boldsymbol{g}(t;\mathbf{0}) \le \frac{\boldsymbol{\alpha}^T}{\sqrt{n}}(\boldsymbol{g}(t;\frac{\boldsymbol{\alpha}}{\sqrt{n}}) - \boldsymbol{g}(t;\mathbf{0})). \tag{2.3.4}$$

Assumption 2.3.2.

(i) For all i $(1 \le i \le p)$ and $\boldsymbol{\alpha}$ in a neighborhood of $\boldsymbol{\theta}^0$,

$$\sum_{k=1}^{\infty} E\big|\big(g(1;\boldsymbol{\alpha})_i - g(1;\mathbf{0})_i\big)\big(g(k;\boldsymbol{\alpha})_i - g(k;\mathbf{0})_i\big)\big| < \infty.$$

(ii) $Q(\boldsymbol{\theta})$ is twice differentiable at $\boldsymbol{\theta}^0$ and $\nabla^2 Q(\boldsymbol{\theta}^0)$ is positive definite.

Remark 2.3.3. Assumption 2.3.2.(i) is a condition to control the correlation of the score function.

Theorem 2.3.4. *Let $\hat{\boldsymbol{\theta}}_n$ be defined in* (2.3.3) *under Assumptions 2.3.1 and 2.3.2. Then we obtain*

(i) *$\hat{\boldsymbol{\theta}}_n$ converges to the true value $\boldsymbol{\theta}^0$ in probability as $n \to \infty$.*

(ii) *$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0)$ has a joint asymptotic normal distribution whose mean is 0 and the asymptotic covariance matrix is give by $H^{-1}VH^{-1}$, where*

$$\begin{aligned} H &= \nabla^2 Q(\boldsymbol{\theta}^0), \\ V &= \mathrm{Var}\, \boldsymbol{g}(\boldsymbol{\theta}^0). \end{aligned}$$

Proof. First, we show (2.2.1) holds in this case. For fixed $\boldsymbol{\alpha}$, define Y_{nt} as

$$Y_{nt} = \rho(t; \frac{\boldsymbol{\alpha}}{\sqrt{n}}) - \rho(t; \mathbf{0}) - \frac{\boldsymbol{\alpha}^T}{\sqrt{n}} \boldsymbol{g}(t; \mathbf{0}).$$

Then it holds that

$$EY_{nt} = Q(\frac{\boldsymbol{\alpha}}{\sqrt{n}}), \quad \sum_{t=1}^n Y_{nt} = \sum_{t=1}^n \rho(t; \frac{\boldsymbol{\alpha}}{\sqrt{n}}) - \rho(t; \mathbf{0}) - \frac{\boldsymbol{\alpha}^T}{\sqrt{n}} \boldsymbol{g}(t; \mathbf{0}).$$

From (2.3.4), we have

$$\begin{aligned} \mathrm{Var} \sum_{t=1}^n Y_{nt} &\leq E\Big(\sum_{t=1}^n Y_{nt}\Big)^2 \\ &= E \sum_{t=1}^n \sum_{s=1}^n \Big\{\rho(t; \frac{\boldsymbol{\alpha}}{\sqrt{n}}) - \rho(t; \mathbf{0}) - \frac{\boldsymbol{\alpha}^T}{\sqrt{n}} \boldsymbol{g}(t; \mathbf{0}))\Big\} \\ &\qquad \times \Big\{\rho(s; \frac{\boldsymbol{\alpha}}{\sqrt{n}}) - \rho(s; \mathbf{0}) - \frac{\boldsymbol{\alpha}^T}{\sqrt{n}} \boldsymbol{g}(s; \mathbf{0}))\Big\} \\ &\leq E \sum_{t=1}^n \sum_{s=1}^n \frac{1}{n} \boldsymbol{\alpha}^T (\boldsymbol{g}(t; \frac{\boldsymbol{\alpha}}{\sqrt{n}}) - \boldsymbol{g}(t; \mathbf{0}))(\boldsymbol{g}(s; \frac{\boldsymbol{\alpha}}{\sqrt{n}}) - \boldsymbol{g}(s; \mathbf{0}))^T \boldsymbol{\alpha} \\ &\leq 2E \sum_{k=0}^{n-1} \frac{n-k}{n} \boldsymbol{\alpha}^T (\boldsymbol{g}(1; \frac{\boldsymbol{\alpha}}{\sqrt{n}}) - \boldsymbol{g}(1; \mathbf{0}))(\boldsymbol{g}(k+1; \frac{\boldsymbol{\alpha}}{\sqrt{n}}) - \boldsymbol{g}(k+1; \mathbf{0}))^T \boldsymbol{\alpha}. \end{aligned} \tag{2.3.5}$$

Also, for any k $(1 \leq k \leq n-1)$, $\boldsymbol{g}(k; \frac{\boldsymbol{\alpha}}{\sqrt{n}}) - \boldsymbol{g}(k; \mathbf{0}) \geq \mathbf{0}$ almost surely, and since its expectation, which is bounded by $2E \sum_{i=1}^p \alpha_i^2 \sum_{k=1}^\infty |\boldsymbol{g}(1; \boldsymbol{\alpha})_i - \boldsymbol{g}(1; \mathbf{0})_i|$

$\times|\boldsymbol{g}(k;\boldsymbol{\alpha})_i - \boldsymbol{g}(k;\mathbf{0})_i|$, goes to $\mathbf{0}$ by monotone convergence theorem, we obtain

$$\boldsymbol{g}(k;\frac{\boldsymbol{\alpha}}{\sqrt{n}}) - \boldsymbol{g}(k;\mathbf{0}) \xrightarrow{a.s.} \mathbf{0}.$$

Again, by monotone convergence theorem, the right hand side of (2.3.5) converges to 0. Therefore, by Chebyshev's inequality, we have

$$\sum_{t=1}^{n} Y_{nt} - nEY_{nt} \xrightarrow{\mathcal{P}} 0,$$

where $nEY_{nt} \to \nabla^2 Q(\boldsymbol{\theta}^0)$ by the Talor expansion. Asymptotic normality of $n^{-1/2}\sum \boldsymbol{g}(t;\boldsymbol{\theta}^0)$ follows from the classic central limit theorem under the Assumption 3.2.(ii). □

Remark 2.3.5. Note that Y_{nt} is not independent, which is different from Niemiro (1992).

2.4 Estimation with Differentiable Objective Functions

As mentioned in Basawa (1985), a general method of establishing the asymptotic normality of M-estimators is to use martingale theory, although there are several ways like assuming some mixing conditions in time series context. In this section, we suppose the objective function satisfies the differentiable conditions and therefore the result can be further shown easily and concretely by the martingale difference central limit theorem. The central limit theorem for $\hat{\boldsymbol{\theta}}_n$ is proved by Hodges-Lehmann's criteria, which has been given in Section 2.2. Consider $\rho(\theta_1^{-1/2}, e(t;\boldsymbol{\eta})) \equiv \rho(x,y)$. For simplicity of notation, we write $\rho(t;\boldsymbol{\theta}) = \rho(\theta_1^{-1/2}, e(t;\boldsymbol{\eta}))$ where $\boldsymbol{\theta} = (\theta_1, \boldsymbol{\eta}^T)^T \in \Theta \subset \mathbb{R}^m$: compact if it is not necessary to think θ_1 and $\boldsymbol{\eta}$ in a separate way. Here θ_1 is a scale parameter and $\boldsymbol{\eta} = (\theta_2, \ldots, \theta_p)^T$.

The objective function is denoted by

$$Q(\boldsymbol{\theta}) = E\rho(\theta_1^{-1/2}, e(t;\boldsymbol{\eta})).$$

The true value of $\boldsymbol{\theta}$, represented by $\boldsymbol{\theta}^0$, is defined by

$$Q(\boldsymbol{\theta}^0) = \min_{\boldsymbol{\theta}\in\Theta} Q(\boldsymbol{\theta}). \tag{2.4.1}$$

On account of the simplicity of notation, we use ς and $\boldsymbol{\eta}^0$ for the true value of θ_1 and $\boldsymbol{\eta}$ separately. The sample version corresponding to the objective function is

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \rho(\theta_1^{-1/2}, e(t;\boldsymbol{\eta})),$$

and the estimator is defined by

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\Theta} Q_n(\boldsymbol{\theta}). \tag{2.4.2}$$

Let ρ_x, ρ_y be the partial derivative of $\rho(x,y)$ with respect to x and y, $\mathcal{B}(t)$ be the σ-field generated by the set of random variables $\{X(n);\, n \le t\}$. As seen in the definition of $\rho(\theta_1^{-1/2}, e(t;\boldsymbol{\eta}))$, it is $\mathcal{B}(t)$-measurable. Also, by (2.1.1), we can see that $\frac{\partial}{\partial \boldsymbol{\eta}} e(t;\boldsymbol{\eta})$ is $\mathcal{B}(t-1)$-measurable.

Assumption 2.4.1.

(i) Let $\rho(\boldsymbol{\theta})$ be a measurable convex function with respect to $\boldsymbol{\theta}$ from $\mathbb{R}\times\mathbb{R}^{p-1}$ to $\mathbb{R}$.

(ii) $Q(\boldsymbol{\theta})$ is well defined.

(iii) $\boldsymbol{\theta}^0$ satisfying (2.4.1) exists and is unique.

(iv) $E\rho_x(\boldsymbol{\theta}^0) = 0$ and $E\rho_y(\boldsymbol{\theta}^0) = 0$.

(v) $E\rho(\boldsymbol{\theta}^0)^2 < \infty$, $E\rho_y(\boldsymbol{\theta}^0)^2 < \infty$, $E(\frac{\partial^2}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T} e(t;\boldsymbol{\eta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0})^T(\frac{\partial^2}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T} e(t;\boldsymbol{\eta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0})$ and
$E(\frac{\partial}{\partial\boldsymbol{\eta}} e(t;\boldsymbol{\eta})\frac{\partial}{\partial\boldsymbol{\eta}^T} e(t;\boldsymbol{\eta}))^T(\frac{\partial}{\partial\boldsymbol{\eta}} e(t;\boldsymbol{\eta})\frac{\partial}{\partial\boldsymbol{\eta}^T} e(t;\boldsymbol{\eta}))$ exist.

(vi) Define $\tilde{\rho}(t;\boldsymbol{\theta}^0) \equiv \rho_y(t;\boldsymbol{\theta}^0)\frac{\partial}{\partial\boldsymbol{\eta}} e(t;\boldsymbol{\eta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$. To show central limit theorem for the martingale difference sequence $\{\tilde{\rho}(t;\boldsymbol{\theta}^0), \mathcal{B}(t)\}$, we suppose the Lindeberg condition, by Euclidean norm $\|\cdot\|$ and indicator function $1(\,\cdot\,)$,

$$\frac{1}{n}\sum_{t=1}^{n} E(\|\tilde{\rho}(t;\boldsymbol{\theta}^0)\|^2 1(\|\tilde{\rho}(t;\boldsymbol{\theta}^0)\| \ge \epsilon)) \to 0,$$
$$\frac{1}{n}\sum_{t=1}^{n} E(\tilde{\rho}(t;\boldsymbol{\theta}^0)\tilde{\rho}(t;\boldsymbol{\theta}^0)^T \Big| \mathcal{B}(t-1)) \xrightarrow{\mathcal{P}} S,$$

where
$$S = E\tilde{\rho}(t;\boldsymbol{\theta}^0)\tilde{\rho}(t;\boldsymbol{\theta}^0)^T.$$

Remark 2.4.2. As a result of Assumption 2.4.1.(i), $\rho_x(\boldsymbol{\theta})$ and $\rho_y(\boldsymbol{\theta})$ are also functions from $\mathbb{R}\times\mathbb{R}^{p-1}$ to $\mathbb{R}$.

Remark 2.4.3. In the L^2 theory, set $\rho : \mathbb{R} \times \mathbb{R}^{p-1} \mapsto \mathbb{R}$ as

$$\rho(x, y) = (xy)^2 - 2\log x. \tag{2.4.3}$$

In this case, suppose σ^2 is the variance of the innovation process and then the true value of θ_1 is given by $\theta_1 = \sigma^2$, that is, $\theta_1^{-1/2} = \sigma^{-1}$. The asymptotic result of Theorem 2.4.6 given below is the same as that for the Whittle estimator in the frequency domain in the case of second-order stationary process.

Remark 2.4.4. θ_1 can be a scale parameter different from the variance of the innovation process. As an example, let the objective function be defined as follows:

$$\rho(x, y) = (xy)^k - \frac{k}{k-2} x^{k-2}.$$

Then it is easy to see that ς is $E\epsilon(t)^k$. Note that in most cases in time series analysis, $E\epsilon(t)$ is assumed to be 0 or the symmetricity of $\epsilon(t)$ is assumed. As an alternative, θ_1 can be defined by

$$\rho(x, y) = (x|y|)^k - \frac{k}{k-2} x^{k-2},$$

then ς is $E|\epsilon(t)|^k$.

Remark 2.4.5. Since the random structure of $\rho_x(\boldsymbol{\theta})$ is the same as $\rho(\boldsymbol{\theta})$, it is sufficient to only suppose $E\rho(\boldsymbol{\theta}^0)^2 < \infty$.

The asymptotic result of M-estimation in time domain is given in the following theorem.

Theorem 2.4.6. *Let $\hat{\boldsymbol{\theta}}_n$ be defined by (2.4.2) under Assumption 2.4.1. Then we obtain*

(i) *$\hat{\boldsymbol{\theta}}_n$ converges to the true value $\boldsymbol{\theta}^0$ in probability as $n \to \infty$.*

(ii) *$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0)$ have a joint normal distribution asymptotically whose mean is $\mathbf{0}$ and the asymptotic covariance matrix is give by $H^{-1}VH^{-1}$, where*

$$\begin{aligned}
H &= \begin{pmatrix} \frac{1}{2}\varsigma^{-3} E\rho_{xx}(\boldsymbol{\theta}^0) & 0 \\ 0 & E\rho_{yy}(\boldsymbol{\theta}^0) E \frac{\partial}{\partial \boldsymbol{\eta}} e(t;\boldsymbol{\eta}) \frac{\partial}{\partial \boldsymbol{\eta}'} e(t;\boldsymbol{\eta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \end{pmatrix}, \\
V &= \begin{pmatrix} E\rho_x(\boldsymbol{\theta}^0)^2 & 0 \\ 0 & E\rho_y(\boldsymbol{\theta}^0)^2 E \frac{\partial}{\partial \boldsymbol{\eta}} e(t;\boldsymbol{\eta}) \frac{\partial}{\partial \boldsymbol{\eta}'} e(t;\boldsymbol{\eta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \end{pmatrix}.
\end{aligned}$$

In the case of Gaussian process with the objective function (2.4.3), the covariance matrix is

$$V = 2D^{-1},$$

where

$$D_{ij} = (2\pi)^{-1}\Big\{\int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_i}\log f(\lambda)\frac{\partial}{\partial\theta_j}\log f(\lambda)\,dx\Big\}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0},$$

and $f(\lambda)$ is the spectral density of the model.

Proof. We will show Assumption 2.4.1 satisfies Assumption 2.2.3. From the definition,

$$\boldsymbol{\xi}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}(-\frac{1}{2}\theta_1^{-3/2}\rho_x(\boldsymbol{\theta}), \rho_y(\boldsymbol{\theta})\frac{\partial}{\partial\boldsymbol{\eta}^T}e(t;\boldsymbol{\eta}))^T.$$

(i) If ρ is convex, then its derivative $\boldsymbol{g}$ is a non-decreasing function in each argument.

(ii) Stochastic expansion of $\boldsymbol{\xi}_n(\boldsymbol{\theta})$ yields

$$\sqrt{n}(\boldsymbol{\xi}_n(\boldsymbol{\theta}^0 + \boldsymbol{u}/\sqrt{n}) - \boldsymbol{\xi}_n(\boldsymbol{\theta}^0)) = \frac{1}{n}\sum_{t=1}^{n}\nabla\boldsymbol{\xi}_n(\boldsymbol{\theta}^0)^T\boldsymbol{u} + o_p(n^{-3/2}).$$

The $(1,1)$-element of $\nabla\boldsymbol{\xi}_n(\boldsymbol{\theta})$ is

$$\nabla\boldsymbol{\xi}_n(\boldsymbol{\theta}^0)_{11} = \frac{1}{n}\sum_{t=1}^{n}(-\frac{1}{2}\varsigma^{-3/2}\rho_{xx}(t;\boldsymbol{\theta}^0) + \frac{3}{4}\varsigma^{-5/2}\rho_x(t;\boldsymbol{\theta}^0)).$$

Since $E\rho_x(\boldsymbol{\theta}^0) = 0$ and $\rho_x(\boldsymbol{\theta}^0)$ is $\mathcal{B}(t)$-measurable, $\{\rho_x(t;\boldsymbol{\theta}^0)\}$ is i.i.d. sequence with mean 0 and finite variance. Thus it holds that

$$\nabla\boldsymbol{\xi}_n(\boldsymbol{\theta}^0)_{11} \xrightarrow{\mathcal{P}} -\frac{1}{2}\varsigma^{-3/2}E\rho_{xx}(\boldsymbol{\theta}^0),$$

since

$$\sum_{t=1}^{n}\frac{3}{4}\varsigma^{-5/2}\rho_x(t;\boldsymbol{\theta}^0) \xrightarrow{\mathcal{P}} 0.$$

Similarly, we have (i,j)-element of $\nabla\boldsymbol{\xi}_n(\boldsymbol{\theta})$ $(i \geq 2, j \geq 2)$,

$$\nabla\boldsymbol{\xi}_n(\boldsymbol{\theta}^0)_{ij} = \frac{1}{n}\sum_{t=1}^{n}\rho_{yy}(t;\boldsymbol{\theta}^0)\frac{\partial}{\partial\boldsymbol{\eta}}e(t;\boldsymbol{\eta})\frac{\partial}{\partial\boldsymbol{\eta}^T}e(t;\boldsymbol{\eta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} + \frac{1}{n}\sum_{t=1}^{n}\rho_y(t;\boldsymbol{\theta}^0)\frac{\partial^2}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}e(t;\boldsymbol{\eta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}.$$

The last term in the right hand side of the equation above forms a martingale since $b(0) = 1$ implies that $\frac{\partial}{\partial\boldsymbol{\eta}}e(t;\boldsymbol{\eta})$ and $\frac{\partial^2}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}e(t;\boldsymbol{\eta})$ is $\mathcal{B}(t-1)$-measurable. Noting that

$$E\rho_y(\boldsymbol{\theta}^0)\frac{\partial^2}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}e(t;\boldsymbol{\eta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = E\Big(E\rho_y(\boldsymbol{\theta}^0)(\frac{\partial^2}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}e(t;\boldsymbol{\eta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0})\;\Big|\;\mathcal{B}(t-1)\Big)$$

$$= E\frac{\partial^2}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}e(t;\boldsymbol{\eta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}E\rho_y(\boldsymbol{\theta}^0) = 0,$$

under Assumption 2.4.1.(v), we obtain

$$\nabla\boldsymbol{\xi}_n(\boldsymbol{\theta}^0)_{ij} \xrightarrow{\mathcal{P}} E\rho_{yy}(\boldsymbol{\theta}^0)E\frac{\partial}{\partial\boldsymbol{\eta}}e(t;\boldsymbol{\eta})\frac{\partial}{\partial\boldsymbol{\eta}'}e(t;\boldsymbol{\eta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}.$$

By Chebyshev's inequality, the last term is shown to converge to 0 in probability.

(iii) Consider

$$\sqrt{n}\boldsymbol{\xi}_n(\boldsymbol{\theta}^0) = n^{-1/2}\sum_{t=1}^{n}\Big(-\frac{1}{2}\varsigma^{-3/2}\rho_x(\boldsymbol{\theta}^0), \rho_y(\boldsymbol{\theta}^0)\frac{\partial}{\partial\boldsymbol{\eta}^T}e(t;\boldsymbol{\eta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\Big)^T.$$

As seen in (ii), $\{\boldsymbol{\xi}_n(\boldsymbol{\theta}^0), \mathcal{B}(n)\}$ is a martingale with respect to $\mathcal{B}(n)$. Under Assumption 2.4.1, we have

$$\sqrt{n}\boldsymbol{\xi}_n(\boldsymbol{\theta}^0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V),$$

where

$$V = \begin{pmatrix} E\rho_x(\boldsymbol{\theta}^0)^2 & 0 \\ 0 & E\rho_y(\boldsymbol{\theta}^0)^2 E\frac{\partial}{\partial\boldsymbol{\eta}}e(t;\boldsymbol{\eta})\frac{\partial}{\partial\boldsymbol{\eta}^T}e(t;\boldsymbol{\eta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \end{pmatrix}.$$

As a result, the asymptotic normality for $\hat{\boldsymbol{\theta}}$ is shown and the asymptotic variance is given by $H^{-1}VH^{-1}$.

□

2.5 Asymptotic Results of $\mathrm{M_m}$-Estimators

In this section, we give conditions for asymptotic normality of $\mathrm{M_m}$-estimators. Let the objective function be defined by

$$Q(\boldsymbol{\theta}) = E\rho(\theta_1^{-1/2}, e(1;\boldsymbol{\eta}), \ldots, e(m;\boldsymbol{\eta})),$$

with the true value defined by

$$Q(\boldsymbol{\theta}^0) = \min_{\boldsymbol{\theta}\in\Theta} Q(\boldsymbol{\theta}). \tag{2.5.1}$$

The sample analog of $Q_n(\boldsymbol{\theta})$ is defined by

$$Q_n(\boldsymbol{\theta}) = \binom{n}{m}^{-1} \sum_{1\le i_1<\cdots<i_m\le n} \rho(\theta_1^{-1/2}, e(i_1;\boldsymbol{\eta}), \ldots, e(i_m;\boldsymbol{\eta})),$$

and the estimator is defined by

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\Theta} Q_n(\boldsymbol{\theta}). \tag{2.5.2}$$

Assumption 2.5.1.

(i) $\rho(\boldsymbol{\theta})$ is convex with respect to $\boldsymbol{\theta}$ and symmetric in each component $e(i;\boldsymbol{\eta})$.

(ii) $Q(\boldsymbol{\theta})$ is well defined.

(iii) $\boldsymbol{\theta}^0$ satisfying (2.5.1) exists and is unique.

(iv) For all i $(1 \le i \le p)$ and $\boldsymbol{\alpha}$ in a neighborhood of $\boldsymbol{\theta}^0$,

$$\sum_{k=1}^{\infty} E\big|\big(g(1;\boldsymbol{\alpha})_i - g(1;\mathbf{0})_i\big)\big(g(k;\boldsymbol{\alpha})_i - g(k;\mathbf{0})_i\big)\big| < \infty.$$

(v) $Q(\boldsymbol{\theta})$ is twice differentiable at $\boldsymbol{\theta}^0$ and $\nabla^2 Q(\boldsymbol{\theta}^0)$ is positive definite.

Theorem 2.5.2. *Let $\hat{\boldsymbol{\theta}}_n$ be defined by (2.5.2) under Assumption 2.5.1. Then we obtain*

(i) *$\hat{\boldsymbol{\theta}}_n$ converges to the true value $\boldsymbol{\theta}^0$ in probability as $n \to \infty$.*

(ii) *$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0)$ asymptotically has a joint normal distribution whose mean is $\mathbf{0}$ and the asymptotic covariance matrix is give by $m^2 H^{-1} V H^{-1}$, where*

$$\begin{aligned} H &= \nabla^2 Q(\boldsymbol{\theta}^0) \\ V &= \mathrm{Var}\, \boldsymbol{g}(\boldsymbol{\theta}^0). \end{aligned}$$

Proof. Let J denote the set of all m element subsets of $\{1,\ldots,n\}$. For any $j = \{i_1,\ldots,i_m\} \in J$, let Y_j be the random vector $(e(i_1),\ldots,e(i_m))$. Accordingly, the notation $\rho(t;\frac{\boldsymbol{\alpha}}{\sqrt{n}})$ and $\rho(t;\mathbf{0})$ are changed in the following way: for $\boldsymbol{\alpha} = (\alpha_1, \boldsymbol{\alpha}_2)$,

$$\begin{aligned} \rho(j;\frac{\boldsymbol{\alpha}}{\sqrt{n}}) &:= \rho((\varsigma + \frac{\alpha_1}{\sqrt{n}})^{-1/2}, Y_j(\boldsymbol{\eta}^0 + \frac{\boldsymbol{\alpha}_2}{\sqrt{n}})), \\ \rho(j;\mathbf{0}) &:= \rho(\varsigma^{-1/2}, Y_j(\boldsymbol{\eta}^0)). \end{aligned}$$

For any fixed $\boldsymbol{\alpha}$ and j, define

$$Z_{nj} = \rho(j;\frac{\boldsymbol{\alpha}}{\sqrt{n}}) - \rho(j;\mathbf{0}) - \frac{\boldsymbol{\alpha}^T}{\sqrt{n}}\boldsymbol{g}(j;\mathbf{0}).$$

Note that $EZ_{nj} = Q(\frac{\boldsymbol{\alpha}}{\sqrt{n}})$. For the same reason, we have

$$\mathrm{Var}\sum Z_{nj} \xrightarrow{\mathcal{P}} 0.$$

Since $\mathrm{Var}(n\binom{n}{m}^{-1}\sum Z_{nj}) \leq m^2 \mathrm{Var}\sum Z_{nj}$,

$$n\binom{n}{m}^{-1}\sum Z_{nj} - nEZ_{nj} \xrightarrow{\mathcal{P}} 0.$$

Thus the result of Theorem 2.5.2 depends on the asymptotics of $\sqrt{n}\binom{n}{m}^{-1}\sum \boldsymbol{g}(j;\mathbf{0})$. Regard $\boldsymbol{g}(j;\mathbf{0})$ as a kernel of U-statistics, we define the degenerate kernel of $\boldsymbol{g}(\mathbf{0})$ by

$$\boldsymbol{g}_0^c(x_1,\ldots,x_c) = \sum_{r=0}^{c}\binom{c}{r}(-1)^{c-r}\int\cdots\int_{\mathbb{R}^{m-r}} \boldsymbol{g}(\boldsymbol{\theta}^0, x_1,\ldots,x_m)\prod_{i=r+1}^{m} dF(x_i).$$

Suppose U_n is generated by $\boldsymbol{g}(\mathbf{0})$ and U_n^c is generated by $\boldsymbol{g}_0^c$, we have, by Hoeffding's projection, that

$$U_n = \sum_{c=1}^{m}\binom{m}{c}U_n^c = \frac{m}{n}\sum_{t=1}^{n}\boldsymbol{g}(t;\mathbf{0}) + \boldsymbol{R}_n,$$

where $\boldsymbol{R}_n \xrightarrow{\mathcal{P}} \mathbf{0}$. The conclusion is completed by the asymptotic normality of $n^{-1/2}\sum_{t=1}^{n}\boldsymbol{g}(t;\mathbf{0})$. $\square$

2.6 Examples

Suppose the second order stationary process $\{X(t)\}$ is generated by the model

$$X(t) - \beta_0 X(t-1) = \epsilon(t),$$

where $\epsilon(t) \sim$ i.i.d. $(0,\sigma^2)$. For the estimation of β_0, take

$$b(j;\boldsymbol{\eta}) = \begin{cases} \beta & j = 1, \\ 0 & j \geq 2. \end{cases}$$

2.6.1 Asymptotics of L^2 theory in AR(1) case

From Remark 2.4.3, the objective function in L^2 theory is given by

$$Q_n(\boldsymbol{\theta}) = \log\theta_1 + \frac{1}{n}\sum_{t=2}^{n}\theta_1^{-1}(X(t) - \beta X(t-1))^2 + o_p(1).$$

Note that the objective function is asymptotically equivalent to Whittle estimator. Also, the estimator is a modification of least square estimation since the scale parameter is estimated simultaneously. With

$$\rho_x(\boldsymbol{\theta}^0) = \frac{2\epsilon^2}{\sigma} - 2\sigma, \qquad \rho_y(\boldsymbol{\theta}^0) = \frac{2\epsilon}{\sigma^2},$$
$$\rho_{xx}(\boldsymbol{\theta}^0) = 2\sigma^2 + 2\epsilon^2, \qquad \rho_{yy}(\boldsymbol{\theta}^0) = \frac{2}{\sigma^2},$$

by Theorem 2.4.6 we obtain

$$H = \begin{pmatrix} 2\sigma^{-1} & 0 \\ 0 & 2(1-\beta_0^2)^{-1} \end{pmatrix},$$
$$V = \begin{pmatrix} 4\sigma^{-2}(\mu_4 - \sigma^4) & 0 \\ 0 & 4(1-\beta_0^2)^{-1} \end{pmatrix},$$

where μ_4 is the fourth moment of $\epsilon(t)$. As a result, it holds that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathrm{diag}(\kappa_4 - \sigma^4, 1 - \beta_0^2)).$$

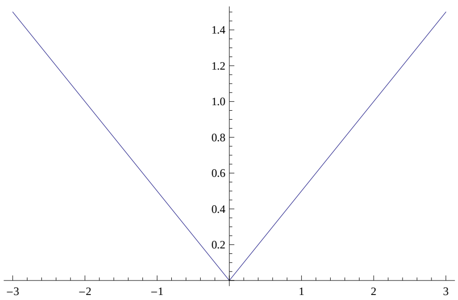
In the Gaussian case, $\kappa_4 - \sigma^4 = 2\sigma^4$.

2.6.2 Asymptotics of L^1 theory in AR(1) case

One parametrization for L^1 case is to set $k = 1$ in Remark 2.4.4. In this subsection, we are interested in another parametrization with convex objective function defined by Koenker and Bassett (1978)'s check function $\rho^\tau(u)$. The check function $\rho^\tau(u)$ is given by

$$\rho^\tau(u) = u(\tau - \mathbb{1}(u < 0)),$$

where $\mathbb{1}(\cdot)$ is the indicator function. Two examples with $\tau = 0.5$ and $\tau = 0.1$ are shown in the following figures.



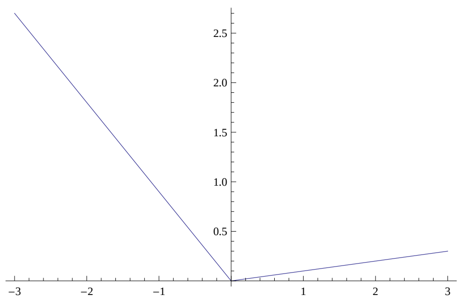


FIGURE 2.1: Koenker and Bassett (1978)'s check function with $\tau = 0.5$(left) and $\tau = 0.1$(right).

As seen from two figures, the objective function is not differentiable around the origin. This is why we have to explore the non-standard case in Section 2.3. A straightforward understanding for the function is to regarded it as a weighting function on the observations $\{X(1), \dots X(n)\}$. The median is corresponding to the symmetric objective function and the lower quantile is corresponding to a weighting function with larger weight on the smaller samples.

With estimation of the scale parameter in the linear model, suppose $\rho(x, y)$ is defined by

$$\rho(x, y) = \rho^\tau(xy) + x^{-1}.$$

Then the true parameters are given by

$$\varsigma = \frac{1}{2}E|\epsilon|, \quad b(j; \boldsymbol{\eta}^0)(\boldsymbol{\eta}^0) = \beta_0\delta(j, 1) \quad \text{for } j \geq 1,$$

from the result that $E\epsilon(t)\mathbb{1}(\epsilon(t) > 0) = \frac{1}{2}E|\epsilon|$ and $EX(t-1)\mathbb{1}(e(t;\boldsymbol{\eta}) < 0) \neq 0$ if $b(j; \boldsymbol{\eta}) \neq \beta_0$. Interestingly, the true parameters do not depend on τ, even τ is included in the check function.

To generalize the result, we suppose a misspecification case, that is to say, we suppose that

$$\zeta = P(\epsilon < 0),$$

where ζ is different from τ in the check function. Then we obtain

$$\begin{aligned} \rho_x(\boldsymbol{\theta}^0) &= \epsilon(\tau - 1(\epsilon < 0)) - \varsigma, & \rho_{xx}(\boldsymbol{\theta}^0) &= 2\varsigma^{3/2}, \\ \rho_y(\boldsymbol{\theta}^0) &= \varsigma(\tau - 1(\epsilon < 0)), & \rho_{yy}(\boldsymbol{\theta}^0) &= \delta(\epsilon), \end{aligned}$$

where $\delta(\cdot)$ is the Dirac delta function. In conclusion, the asymptotic variance of the M-estimator defined by $\rho^\tau(u)$ is given by

$$H^{-1}VH^{-1} = \begin{pmatrix} \frac{1}{4}\varsigma^{-3}(\tau^2\varsigma^2 + (1-2\tau)a) & 0 \\ 0 & \frac{\tau^2 - 2\tau\zeta + \zeta}{\sigma^2} f(0)^{-2}(1 - \beta_0^2) \end{pmatrix},$$

where $a = E\epsilon^2 1(\epsilon < 0)$. As a special case $\zeta = \tau$, the variance of the coefficient parameter is given by

$$\frac{\tau(1-\tau)}{\sigma^2} f(0)^{-2}(1 - \beta_0^2).$$

2.7 Numerical Results

We carry out the numerical experiments for the estimation of coefficient parameters in AR(1) models, i.e., for different β_0,

$$X(t) - \beta_0 X(t-1) = \epsilon(t), \quad \epsilon(t) \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2).$$

The estimations are repeated 100 times on 1000 samples generated from the AR(1) model. We give the average of the estimated coefficient for the model in the left figure and simultaneously the root mean square error (RMSE) in the right figure. All cases are classified by the true parameter β_0.

(1) $\beta_0 = 0.1$

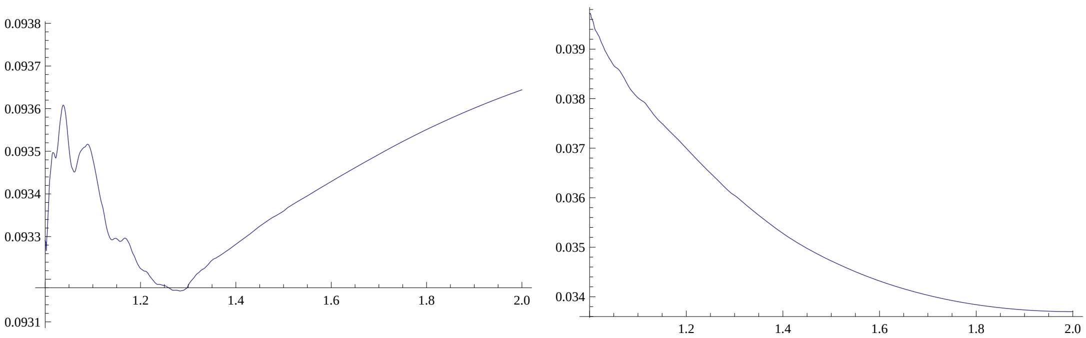


FIGURE 2.2: $\hat{\beta}_0$ (left) and RMSE (right) by L^t estimates ($1 \le t \le 2$).

(2) $\beta_0 = 0.3$

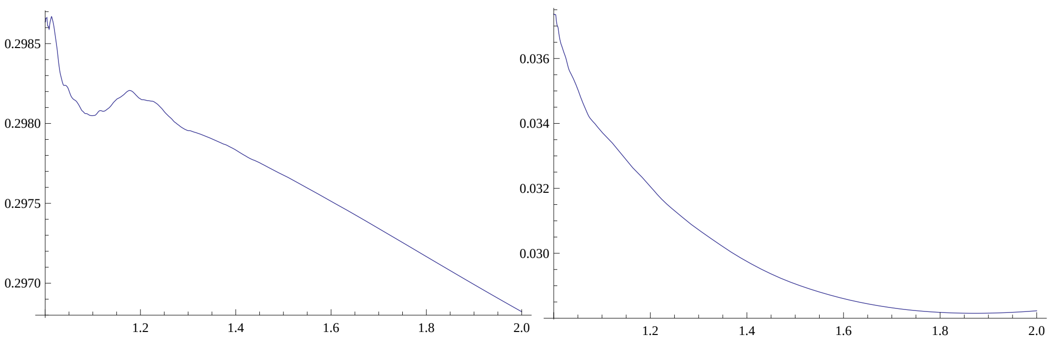


FIGURE 2.3: $\hat{\beta}_0$ (left) and RMSE (right) by L^t estimates ($1 \le t \le 2$).

(3) $\beta_0 = 0.5$

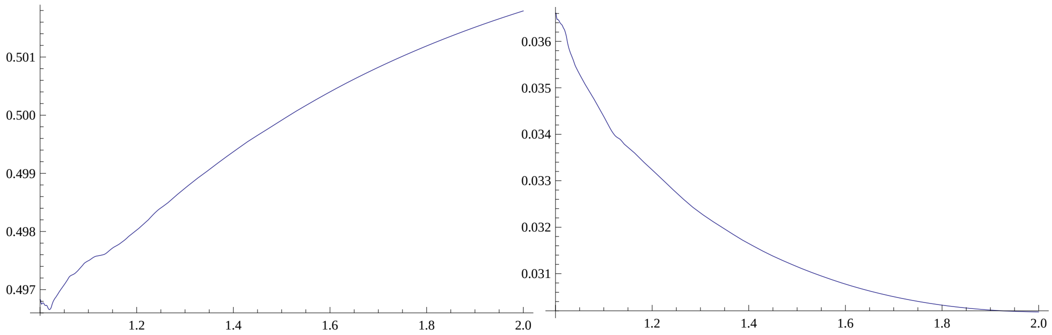


FIGURE 2.4: $\hat{\beta}_0$ (left) and RMSE (right) by L^t estimates ($1 \leq t \leq 2$).

(4) $\beta_0 = 0.7$

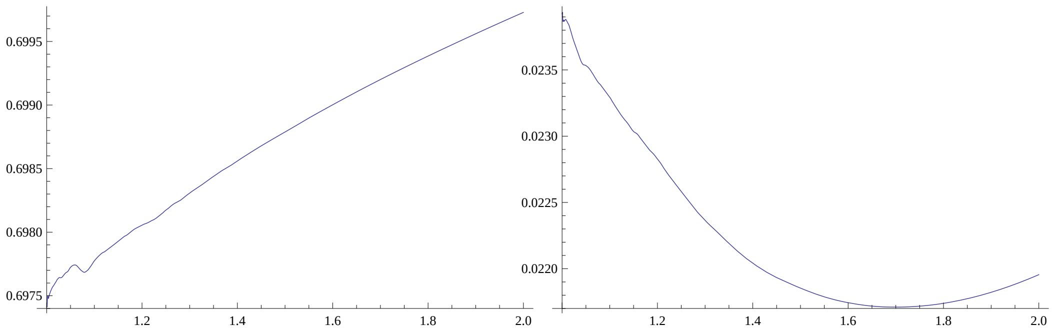


FIGURE 2.5: $\hat{\beta}_0$ (left) and RMSE (right) by L^t estimates ($1 \leq t \leq 2$).

(5) $\beta_0 = 0.9$

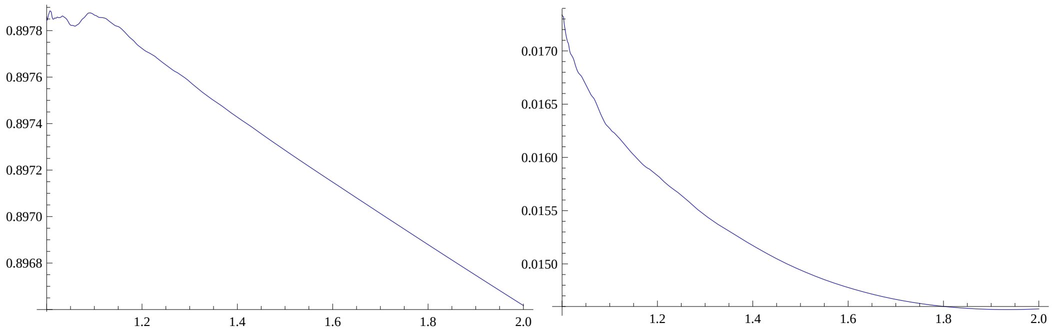


FIGURE 2.6: $\hat{\beta}_0$ (left) and RMSE (right) by L^t estimates ($1 \leq t \leq 2$).

Chapter 3

Asymptotic Properties of Generalized Spectral via M-estimators

3.1 Introduction

In chapter 3, we develop asymptotic theory for the generalized spectral, which provides a characterization of the intrinsic properties in the time series $\{X(t)\}$. As an example, second order stationary process $\{X(t)\}$ has a characteristic of serial autocovariances independent of any time points but only depend on the interval between two time points. The spectral density function of the process then gives a Fourier transformation of all serial autocovariances. Not restricting to the serial autocovariances, the spectral density function can be always suppose to be a Fourier transformation of serial intrinsic properties of the process. Nowadays, quantile and copula-related spectral density function is proposed in many different literature.

For the estimation of the spectral density, periodogram plays a crucial role in time series analysis. Li (2008) proposed Laplace periodogram for estimation for zero-crossing spectral. The robust aspects of Laplace periodogram in the case of heavy tailed models and nonlinear time series models are also shown in numerical way in the paper. The sequential paper Li (2012) has generated the Laplace periodogram to the quantile periodogram and investigated the asymptotic properties of them. On the other hand, Hagemann (2011) also examine the quantile periodogram by another definition. The properties of

(i) summarizing the cyclical behavior of time series,

(ii) capturing systematic changes in the impact of cycles,

(iii) complementing the defect of consideration on autocovariance function,

possessed by the quantile periodogram, are given in the paper. Furthermore, several generalizations for the periodogram and investigation of asymptotic properties have been much studied recently. (See Kley et al. (2014) and Skowronek et al. (2014)).

In the following sections, we give the definition of the quantile periodogram and generalize the idea via M-estimators, which we mentioned in Chapter 2. In Section 3.2, we review the origin of the Laplace periodogram and its extension to the quantile periodogram. The quantile periodogram has an equivalent definition for calculation in statistics, which is given in Section 3.3. We extend the idea to the generalized periodogram in Section 3.4. For the proof of Theorem 3.4.3, we review the concept of stochastic equicontinuity in Section 3.5 and give the proof in Section 3.6. The notations and symbols used in Chapter 3 are listed in the following: $\mathbb{1}(\cdot)$ denotes the indicator function; e denotes the Napier's constant; I_d denotes the d-dimensional identity matrix; $\nu_n(\cdot)$ denotes the empirical process; $\xrightarrow{\mathcal{P}}$ and $\xrightarrow{\mathcal{L}}$ denote the convergence in probability and the convergence in law, respectively.

3.2 Quantile Periodogram

For a discrete stochastic process $\{X(t); t \in \mathbb{Z}\}$, the Laplace periodogram $L_n(\lambda_j)$ is defined based on its observation stretch $\{X(t); t = 1, \dots, n\}$ by

$$L_n(\lambda_j) = \frac{n}{4} \|\hat{\boldsymbol{\beta}}_n(\lambda_j)\|_2^2, \tag{3.2.1}$$

where

$$\hat{\boldsymbol{\beta}}_n(\lambda_j) := \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^2} \sum_{t=1}^{n} |X(t) - \boldsymbol{c}_t^T(\lambda_j)\boldsymbol{\beta}|, \tag{3.2.2}$$

and

$$\boldsymbol{c}_t(\lambda_j) = (\cos(t\,\lambda_j), \sin(t\,\lambda_j))^T.$$

Here, λ_j is supposed to be

$$\frac{2\pi j}{n}, \quad j = 0, \pm 1, \dots, \pm(n-1). \tag{3.2.3}$$

The definition of the Laplace periodogram by (3.2.1) and (3.2.2) is a direct extension of the ordinary periodogram from the property that the ordinary periodogram is decomposed by

$$I_{n,X}(\lambda_j) = |d_{n,X}(\lambda_j)|^2, \quad d_{n,X}(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X(t) e^{it\lambda_j},$$

where $d_{n,X}(\lambda_j)$ can be regarded as a regression on the observation stretch $\{X(1), \ldots, X(n)\}$.

The Laplace periodogram is an original idea of the quantile periodogram. It is easy to see that the change in the definition of the Laplace periodogram from the ordinary periodogram is that the distance to be minimized between the observation and trigonometric regressor varied from l_2-norm to l_1-norm. Not only the symmetric l_1-norm, the distance can be also defined by Koenker and Bassett's check function $\rho_\tau(u)$, i.e.,

$$\rho_\tau(u) = u(\tau - \mathbb{1}(u < 0)).$$

The figures for this function are given in Figure 2.1 in Chapter 2 for reference. The quantile periodogram is defined on the check function $\rho_\tau(u)$ by

$$Q_n(\lambda_j) = \frac{n}{4} \|\hat{\boldsymbol{\beta}}_n(\lambda_j)\|_2^2, \tag{3.2.4}$$

where

$$\hat{\boldsymbol{\beta}}_n(\lambda_j) := \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^2} \sum_{t=1}^{n} \rho_\tau(X(t) - \lambda - \boldsymbol{c}_t^T(\lambda_j)\boldsymbol{\beta}),$$

and

$$\boldsymbol{c}_t(\lambda_j) = (\cos(t\,\lambda_j), \sin(t\,\lambda_j))^T.$$

Asymptotic properties of the quantile periodogram are investigated in Li (2008, 2012) in a general way by so called quantile regression Lemma. For the introduction of Lemma, we first impose assumptions on the process

$$X(t) = \mu(t) + \epsilon(t), \tag{3.2.5}$$

where $\mu(t)$ is a deterministic process and $\epsilon(t)$ is a random process whose marginal distribution is given by $F_t(x)$ and bivariate marginal distribution given by $F_{st}(x_1, x_2)$. The regression estimator $\hat{\boldsymbol{\beta}}_n$ is given by

$$\hat{\boldsymbol{\beta}}_n := \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{t=1}^{n} \rho_\tau(X(t) - \boldsymbol{c}_t^T \boldsymbol{\beta}). \tag{3.2.6}$$

Several notations in Assumption 3.2.1 and Lemma 3.2.2 below are given by

$$
\begin{aligned}
f_t(x) &:= F_t'(x), \\
\delta_t &:= \boldsymbol{c}_t^T \boldsymbol{\beta}_0 - \mu(t), \\
\gamma_{st}(x_1, x_2) &:= F_{st}(x_1, x_2) - F_s(x_1)F_t(x_2), \\
H_n &:= n^{-1} \sum_{t=1}^{n} f_t(\delta_t) \boldsymbol{c}_t \boldsymbol{c}_t^T, \\
V_n &:= n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} \gamma_{st}(x_1, x_2) \boldsymbol{c}_t \boldsymbol{c}_s^T.
\end{aligned}
$$

As similar to Serfling (1980), the notation for asymptotic normality for d dimensional random variables $\{\boldsymbol{X}(t); t = 1, \ldots, n\}$ in Serfling's sense is given by

$$
\boldsymbol{X}_n \xrightarrow{\mathcal{L}} \mathcal{AN}(\boldsymbol{\mu}_n, \Sigma_n),
$$

which actually makes sense in mathematics by

$$
\Sigma_n^{-1}(\boldsymbol{X}_n - \boldsymbol{\mu}_n) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_d).
$$

Assumption 3.2.1. For (3.2.5), suppose the following assumptions hold.

(i) $f_t(x)$ exists for all $x \in \mathbb{R}$ and $f_t(\delta_t) = O(1)$ uniformly.

(ii) For $d > 0$ and $u_0 > 0$,

$$
F_t(u + \delta_t) - F_t(\delta_t) = f_t(\delta_t)u + O(u^{d+1}) \quad \text{uniformly for } |u| \le u_0.
$$

(iii) For any $n \in \mathbb{N}$, there exists a matrix H such that $H_n \ge H$.

(iv) For any $n \in \mathbb{N}$, there exists a matrix V such that $V_n \ge V$.

(v) $\{X(t)\}$ is an m-dependence process or a linear process of the form $\sum_{l=-\infty}^{\infty} \phi(l)e(t-l)$, where $\{e_t\}$ is an i.i.d. random sequence with $E|e(t)| < \infty$ and $\{\phi(l)\}$ is an absolutely summable deterministic sequence such that $\sum_{|l|>n^r} \phi(l) = o(n^{-1})$ as $n \to \infty$ for some constant $r \in [0, 1/4)$.

Quantile regression lemma is given as follows.

Lemma 3.2.2 (Li (2012), Quantile regression lemma)**.** *Let $\{X(t)\}$ be a random process given by* (3.2.5)*. Suppose Assumption 3.2.1 holds. Then we obtain*

$$
\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{L}} \mathcal{AN}(\boldsymbol{\mu}_n, \Sigma_n),
$$

where

$$\begin{aligned}\boldsymbol{\mu}_n &= H_n^{-1}\boldsymbol{h}_n, \quad \boldsymbol{h}_n = n^{-1/2}\sum_{t=1}^{n}(\tau - F_t(\delta_t))\boldsymbol{c}_t,\\ \Sigma_n &= H_n^{-1}V_nH_n^{-1}.\end{aligned}$$

To see the result is broad enough, we give an example, which is also given in Li (2008).

Example 1. Suppose the model is

$$X(t) = \boldsymbol{c}_t^T\boldsymbol{\beta}_0 + \epsilon(t), \quad \epsilon(t) \sim \text{i.i.d. } (0, \sigma^2),$$

$\hat{\boldsymbol{\beta}}_n$ is given by (3.2.6) and the design matrix is $D_n = n^{-1}\sum_{t=1}^{n}\boldsymbol{c}_t\boldsymbol{c}_t^T$. Under Assumption 3.2.1,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{L}} \mathcal{AN}(\mathbf{0}, \Sigma_n),$$

where $\Sigma_n = \tau(1-\tau)f(0)^{-2}D_n$.

Proof. From $\delta_t = 0$, $H_n = f(0)D_n$ and $V_n = \tau(1-\tau)D_n$, the conclusion holds. $\square$

To derive the distribution of the quantile periodogram $Q_n(\lambda_j)$ by Lemma 3.2.2, the stochastic process $\{X(t)\}$ involved has to be stationary in level-crossings. That is to say, if we consider the binary level-crossing process $\{\mathbb{1}(X(t)) \le \xi\}$, then it must be stationary in the sense that there exists a sequence $\{\gamma(k)\}$ such that

$$\gamma(t-s) = P\{(X(t)-\lambda)(X(s)-\lambda) < 0\} \quad \text{for all } s, t \in \mathbb{Z}. \tag{3.2.7}$$

This assumption seems strong but it is not so. In fact, if the process is strict stationary, then (3.2.7) always holds.

Under this setting, we suppose the level-crossing process $\{\mathbb{1}(X(t) < \lambda)\}$ is stationary with mean τ and autocovariance $\tau(1-\tau) - (1/2)\gamma(k)$. Corresponding to this level-crossing process $\{(\mathbb{1}(X(t) < \lambda) - \tau)/\sqrt{\tau(1-\tau)}\}$, we have the quantile spectral density defined by

$$S(\omega) = \sum_{k=-\infty}^{\infty}\left\{1 - \frac{1}{2\tau(1-\tau)}\gamma(k)\right\}\exp(ik\omega). \tag{3.2.8}$$

We impose the following assumptions which are sufficient for Assumption 3.2.1.

Assumption 3.2.3. For the process $\{X(t); t \in \mathbb{Z}\}$, we suppose

(i) $F_t(\xi) = \tau$ and $f_t(\xi) = \iota > 0$ for all $t \in \mathbb{Z}$.

(ii) For $d > 0$ and $u_0 > 0$,

$$F_t(u + \delta) - F_t(\delta) = f_t(\delta)u + O(u^{d+1}) \quad \text{uniformly for } |u| \le u_0.$$

(iii) $\{X(t)\}$ is stationary in ξ-level crossings and the quantile spectral density is well defined, that is,

$$\sum_{k=-\infty}^{\infty} \left|1 - \frac{1}{2\tau(1-\tau)}\gamma(k)\right| < \infty.$$

(iv) $\{X(t)\}$ is an m-dependence process or a linear process of the form $\sum_{l=-\infty}^{\infty} \phi(l)e(t-l)$, where $\{e_t\}$ is an i.i.d. random sequence with $E|e(t)| < \infty$ and $\{\phi(l)\}$ is an absolutely summable deterministic sequence such that $\sum_{|l|>n^r} \phi(l) = o(n^{-1})$ as $n \to \infty$ for some constant $r \in [0, 1/4)$.

Theorem 3.2.4 (Li (2012))**.** *Let $\{X(t)\}$ satisfy Assumption 3.2.3. If the quantile periodogram is given by* (3.2.4)*, then as $n \to \infty$, the joint distribution of the quantile periodogram has asymptotics*

$$(Q_n(\lambda_1), Q_n(\lambda_2), \ldots, Q_n(\lambda_l) \xrightarrow{\mathcal{L}} \frac{1}{2}\frac{\tau(1-\tau)}{\iota^2}\Big(S(\lambda_1), S(\lambda_2), \ldots, S(\lambda_l)\Big)\chi_2^2,$$

where $S(\omega)$ is given by (3.2.8)*.*

Proof. Suppose $\hat{\boldsymbol{\beta}}_n(j)$ is corresponding to the element $\boldsymbol{c}_t(\lambda_j)$ in $\boldsymbol{c}_t$, where $\boldsymbol{c}_t(\lambda_j) = (\cos(\lambda_j), \sin(\lambda_j))^T$. As a direct result of (3.2.3), we obtain that

$$\sum_{t=1}^{n} e^{it\lambda_k} = \begin{cases} n & \text{for } k = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2.9}$$

From (3.2.9), it is not difficult to see that

$$n^{-1}\sum_{t=1}^{n} \boldsymbol{c}_t(\lambda_j)\boldsymbol{c}_t(\lambda_k)^T = \frac{1}{2}\delta(j,k)I_2,$$

and by trigonometric formula

$$n^{-1}\sum_{t=\max(1,1+s)}^{\min(n,n+s)} \boldsymbol{c}_t(\lambda_j)\boldsymbol{c}(\lambda_k)_{t-s}^T = \frac{1}{2}\delta(j,k)\begin{pmatrix}\cos(s\lambda_j) & -\sin(s\lambda_j)\\ \sin(s\lambda_j) & \cos(s\lambda_j)\end{pmatrix} + o(1).$$

In view of $\delta_t = \xi$ and $F_t(\xi) = \tau$, we obtain

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n(j) - \boldsymbol{\beta}_0(j)) \xrightarrow{\mathcal{L}} \mathcal{AN}(\mathbf{0}, \Sigma_n),$$

where Σ_n is given by

$$\Sigma_n = \frac{1}{2}\frac{\tau(1-\tau)}{\iota^2}S(\lambda_j)I_2,$$

and $\mathrm{Cov}(\sqrt{n}(\hat{\boldsymbol{\beta}}_n(j) - \boldsymbol{\beta}_0(j)), \sqrt{n}(\hat{\boldsymbol{\beta}}_n(k) - \boldsymbol{\beta}_0(k))) = 0$ if $k \neq j$. Therefore, the conclusion follows. □

Remark 3.2.5. The problem with the quantile periodogram proposed by Li (2008, 2012) is that if we use the quantile periodogram for hypothesis testing, we have to estimate $\iota = f_t(\xi)$. This estimation may lead a different asymptotic distribution of the quantile periodogram.

3.3 Alternative Definition of Quantile Periodogram

To avoid the problem in Remark 3.2.5, Hagemann (2011) proposed another definition of the quantile periodogram, which does not include the term of $f_t(\xi)$. As what we mentioned before, the τ-th sample quantile can be obtained by minimizing Koenker and Bassett's check function, that is,

$$\hat{\xi}_n = \min_{x\in\mathbb{R}} \sum_{t=1}^{n} \rho_\tau(X(t) - x),$$

where

$$\rho_\tau(u) = u(\tau - \mathbb{1}(u < 0)).$$

The score function of Koenker and Bassett's check function $\psi_\tau(x)$ is defined by

$$\psi_\tau(x) = \tau - \mathbb{1}(u < 0).$$

With this notation, the alternative definition of the quantile periodogram is given by

$$Q_n(\lambda_j) = \frac{1}{2\pi}|n^{-1/2}\sum_{t=1}^{n}\psi(t;\hat{\xi}_n)e^{it\lambda_j}|^2, \tag{3.3.1}$$

where $\psi(t;\hat{\xi}_n) = \psi_\tau(X(t) - \hat{\xi}_n)$. This definition is corresponding to Li's quantile periodogram without the multiplier ι, which is seen from the stochastic equicontinuity in Section 3.5.

Remark 3.3.1. As (3.6.1) below holds, the definition of the quantile periodogram

$$S_n(\lambda) = \frac{1}{2\pi n}\Big|\sum_{t=1}^{n} g(t;\hat{\theta}_n)e^{it\lambda}\Big|^2$$

is equivalent to (3.2.4) without the constant ι. This trait leads the definition of the quantile periodogram introduced by Hagemann (2011) to an easy handling in practice.

3.4 Generalized Periodogram

From the idea of the definition in (3.3.1), we give an extension of periodogram via the subgradient function $g(\theta)$ of M-estimators. Suppose we have to minimize a convex function $\rho(\theta)$ and $g(\theta)$ is the subgradient function of $\rho(\theta)$. Here we only suppose the parameter θ is 1-dimension and $g(\theta)$ is a real-valued function.

Remark 3.4.1. If we suppose $\boldsymbol{g}(\boldsymbol{\theta})$ is p-dimension, then asymptotic distribution of the generalized periodogram is well known to be a complex Wishart distribution in general. The story goes beyond this dissertation.

To be clear, suppose $\rho(X(t);\theta)$ is denoted by $\rho(t;\theta)$ and the estimate $\hat{\theta}_n$ is defined by

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n} \rho(t;\theta).$$

Define the generalized periodogram by

$$S_n(\lambda) = \frac{1}{2\pi n}\Big|\sum_{t=1}^{n} g(t;\hat{\theta}_n)e^{it\lambda}\Big|^2, \tag{3.4.1}$$

where $g(t;\theta)$ is a subgradient function of the corresponding objective function $\rho(t;\theta)$.

As what we mentioned in Section 3.1, we have to suppose a serial intrinsic properties for the time series for estimation. That is to say, the process $\{X(t); t\in\mathbb{Z}\}$ is stationary in the sense of a parameter θ. The serial intrinsic properties may be independent identically distributed, strict stationary, second order stationary, ξ-level crossing or even zero-crossing. That is to say, we can define the subgradient function g to correspond to the property and the assumption of stationarity guarantees that there exist serial functions $\gamma_g(j)$ such that

$$\gamma_g(s-t) = Eg(s;\theta_0)g(t;\theta^0) \quad \text{for all } s,t\in\mathbb{Z}. \tag{3.4.2}$$

The serial function $\gamma_g(j)$ then can be Fourier transformed and the spectral density function is

$$f_g(\lambda) = \frac{1}{2\pi}\sum_{j\in\mathbb{Z}} \gamma_g(j)e^{ij\lambda}.$$

Assumption 3.4.2.

(i) $\rho(\theta)$ is convex with respect to θ.

(ii) $Q(\theta)$ is well defined.

(iii) θ^0 satisfying (2.3.1) exists and is unique.

(iv) For α in a neighborhood of θ^0,

$$\sum_{k=1}^{\infty} E\big|\big(g(1;\alpha) - g(1;\theta^0)\big)\big(g(k;\alpha) - g(k;\theta^0)\big)\big| < \infty.$$

(v) $Q(\theta)$ is twice differentiable at θ^0 and $\nabla^2 Q(\theta^0)$ is positive definite.

(vi) $\{X(t)\}$ is stationary in the sense of (3.4.2) and

$$\sum_{j\in\mathbb{Z}} |\gamma_g(j)| < \infty.$$

(vii) The subgradient function $g(\cdot)$ is Lipschitz continuous in the argument of the function of random variables $X(t)$.

Theorem 3.4.3. *Let $\{X(t)\}$ satisfy Assumption 3.4.2. If the generalized periodogram is given by* (3.4.1)*, then as $n \to \infty$, the joint distribution of the generalized periodograms*

$$\big(S_n(\lambda_1), S_n(\lambda_2), \dots, S_n(\lambda_l)\big), \quad |l| \le n-1,$$

asymptotically converge to independent exponential distributed random variables with mean $f_g(\lambda)$.

We will skip the proof of Theorem 3.4.3 for a while, since we need a concept called "stochastic equicontinuity" for the proof. Before looking at the concept, we leave a remark for Theorem 3.4.3.

Although the convergence of (3.4.1) is shown to convergence to an exponential distributed random variable with mean $f_g(\lambda)$, it does not mean the generalized periodogram is a consistent statistic for spectral density. In general, we have to smooth the periodogram to obtain an consistent nonparametric estimator of the spectral density. As for the smoothing, we will mention it in Chapter 5.

3.5 Stochastic Equicontinuity

To show the equivalence of the definition of the quantile periodogram proposed by Li and Hagemann, and further Theorem 3.4.3, we have to show the stochastic equicontinuity for the general subgradient function g by bracketing conditions.

Suppose we bracket the subgradient function g by g_i, $(i = 1, \dots, n)$. The class for the parametric function g is defined by

$$\mathcal{F} = \{g; \sup_x |g(x)| < \infty\},$$

and the bracketing number $N(\delta)$ for the class equals the smallest value of N for which it holds that for any $g \in \mathcal{F}$, there exists an $i \in \{1, 2, \dots, N\}$ such that

$$|g - g_i| \leq b_i,$$

where $\rho(b_i) \leq \delta$ for any $i = 1, \dots, N$. The distance function ρ is generally defined by L_2-norm, that is,

$$\rho(g) = \sup_{i=1,\dots,n} \|g(X_i)\|_2.$$

The argument of empirical process is usually done with the assumption that random variables are i.i.d. or at most, in dependent case, strongly mixing. See (Andrews and Pollard (1994)). For the context of time series model, we have to impose another condition, the geometric moment contraction property, for the process, although the condition is not stronger nor weaker than the strongly mixing condition. The geometric moment contraction is first given in Hsing and Wu (2004). Before introducing the condition, we first suppose the time series $\{X(t); t = 1, \dots, n\}$ is generated by a system,

$$X(i) = F(\dots, \epsilon(i-1), \epsilon(i)),$$

where $\epsilon(i)$ are i.i.d. random variables. For this process, we make a copy of the process by

$$X^*(i) = F(\dots, \epsilon^*(-1), \epsilon^*(0), \dots, \epsilon(i)),$$

where $\epsilon^*(k)$ is a copy of $\epsilon(k)$. The geometric moment contraction condition is given by

$$\|X(n) - X^*(n)\|_\alpha = O(r(\alpha)^n)$$

for some $\alpha > 0$ and $0 < r(\alpha) < 1$. As mentioned in Hagemann (2014), the models of stationary ARMA, ARCH, GARCH, and so on satisfy this condition.

Assumption 3.5.1. We impose two assumptions to the function class $\mathcal{F}$:

(i) For the subgradient function g and b_i $(i = 1, \dots, n)$, the Lipschitz continuity holds, that is, for some $p, q > 0$ and $L > 0$,

$$\begin{aligned}
\|g(X(n)) - g(X^*(n))\|_p &\leq L\|X(n) - X(n)^*\|_q, \\
\|b_k(X(n)) - b_k(X^*(n))\|_p &\leq L\|X(n) - X(n)^*\|_q.
\end{aligned}$$

(ii) The bracketing number $N(\delta)$ satisfies

$$\int_0^1 x^{-\gamma/(2+\gamma)} N(x, \mathcal{F})^{1/q} dx < \infty$$

for some $\gamma > 0$ and an even number $q \geq 2$.

Under Assumption 3.5.1, Hagemann (2014) gave a crucial result for the stochastic equicontinuity for the subgradient function g. From the problem of measurablily, suppose P^* and E^* are outer probability and outer expectation. ν_n represents the empirical process, i.e.,

$$\nu_n g(\cdot) := n^{-1/2} \sum_{t=1}^{n} \big(g(t;\cdot) - Eg(t;\cdot)\big).$$

Lemma 3.5.2 (Hagemann (2014))**.** *Suppose Assumption 3.5.1 holds. Then we obtain that for any $\epsilon > 0$, there exists a $\delta > 0$ such that*

$$\limsup_{n\to\infty} E^* \Big(\sup_{f,g \in \mathcal{F}, \rho(f-g) < \delta} |\nu_n(f-g)| \Big)^2 < \epsilon.$$

Proof. See Hagemann (2014) and Andrews and Pollard (1994). □

3.6 Proof of Theorem 3.4.3

We show Theorem 3.4.3 in this section. Before the proof, we first give a remark relating the method used in the proof.

Remark 3.6.1. As mentioned in Andrews and Pollard (1994), the function ρ defined in the proof can be seemed as a new distance on Θ, that is, $\rho : \Theta \to \mathbb{R}$. Note that the estimator $\hat{\theta}_n$ is then plugged into the distance later on.

Proof. To treat the plugin type statistics, we approximate (3.4.1) by the statistics evaluated at the true value θ^0. In fact, for $k = 1, \ldots, l$,

$$S_n(\lambda_k) = \frac{1}{2\pi} \Big| \nu_n\{g(\hat{\theta}_n) - g(\theta_0)\} e^{-it\lambda_k} + \nu_n g(\theta_0) e^{-it\lambda_k} \Big|^2.$$

What we have to show is that

(i) the first term satisfies

$$\nu_n\{g(\hat{\theta}_n) - g(\theta_0)\} \to_p 0, \tag{3.6.1}$$

(ii) and the second term satisfies

$$\frac{1}{2\pi} \Big| \nu_n g(\theta_0) e^{-it\lambda_k} \Big|^2 \xrightarrow{\mathcal{L}} \mathscr{E}_k. \tag{3.6.2}$$

Assertion (3.6.1) follows the stochastic equicontinuity. Actually, suppose the maximal diameter of Θ is δ and set the grid on Θ where $\rho(\cdot)$ is set to be L_1-norm. As the parameter θ is 1-dimension, suppose the grid is defined by

$$\theta_0 - \delta = \theta_1 < \cdots < \theta_{N+1} = \theta_0 + \delta.$$

Suppose the functions b_i is defined by

$$b_i = g(\theta_{i+1}) - g(\theta_i), \quad i = 1, \ldots, N.$$

Then it is easy to see that

$$\rho(b_i) = \|g(\theta_{i+1}) - g(\theta_i)\| \leq \{ER(\theta^*)\}(\theta_{i+1} - \theta_i), \quad \theta_{i+1} \leq \theta^* \leq \theta_i,$$

where $R(\theta^*) > 0$ almost surely from Assumption 3.4.2 (v). Define $L = ER(\theta^*)$. In view of $L > 0$, we obtain the Lipschitz property for $\rho(b_i)$ and consequently the bracketing number $N(\delta)$ is $O(\delta^{-1})$ from Andrews and Pollard (1994). If we set $q = 2$, then Assumption 3.5.1 (ii) is satisfied. Assumption 3.5.1 is guaranteed by Assumption 3.4.2 (vii). Therefore we obtain

$$\limsup_{n\to\infty} E^*\Big(\sup_{f,g\in\mathcal{F},\rho(f-g)<\delta} |\nu_n(f-g)|\Big)^2 < \epsilon. \tag{3.6.3}$$

Also note that under Assumption 3.4.2 (i) - (iv), we have

$$\hat{\theta}_n \xrightarrow{\mathcal{P}} \theta_0,$$

and therefore by continuous mapping theorem,

$$\rho(g(\hat{\theta}_n) - g(\theta_0)) \xrightarrow{\mathcal{P}} 0.$$

Consequently, for any $\eta > 0$, $\delta > 0$ and $\epsilon_1 > 0$,

$$\begin{aligned}
P(|\nu_n\{g(\hat{\theta}_n) - g(\theta_0)\}| > \eta) &< P(|\nu_n\{g(\hat{\theta}_n) - g(\theta_0)\}| > \eta, \rho(g(\hat{\theta}_n) - g(\theta_0)) < \delta) + \epsilon_1 \\
&\leq P^*\Big(\sup_{g\in\mathcal{F},\rho(g-g(\theta_0)))<\delta} |\nu_n(g(\hat{\theta}_n) - g(\theta_0))| > \eta\Big) \\
&\leq \eta^{-2} E^*\Big(\sup_{g\in\mathcal{F},\rho(g-g(\theta_0)))<\delta} |\nu_n(g(\hat{\theta}_n) - g(\theta_0))|\Big)
\end{aligned}$$

According to (3.6.3), we obtain

$$\limsup_{n\to\infty} P(|\nu_n\{g(\hat{\theta}_n) - g(\theta_0)\}| > \eta) < \epsilon.$$

For proof of the second assertion (3.6.2), note that

$$n^{-1/2}\sum_{t=1}^{n} g_t(\theta_0)\cos(t\lambda_k), n^{-1/2}\sum_{t=1}^{n} g_t(\theta_0)\sin(t\lambda_k)$$

are both asymptotically uncorrelated normal and accordingly, (3.6.2) holds. $\square$

Chapter 4

Parameter Estimation based on Minimum Contrast Estimators

4.1 Introduction

In this chapter, we are interested in parameter estimation by an estimation procedure in frequency domain. The concept of the spectral density function for a stationary linear process has a long history. Whittle (1952) systematically investigated the parameter estimation by the means of the spectral density for the first time after he found that it was difficult to derive the inverse matrix of parametrized variance matrix of Gaussian stationary process explicitely. He proposed the method to approximate the matrix by the spectral density and next estimate the parameters in the spectral density. In the case, Whittle suggested minimizing the functional $\int_{-\pi}^{\pi} I_{n,X}(\omega)/f_{\boldsymbol{\theta}}(\omega)d\omega$ for the Gaussian stationary process $\{X(t);\, t \in \mathbb{Z}\}$.

The method has been generalized to be a minimization problem of a certain criterion (or disparity measure) $D(f_{\boldsymbol{\theta}}, \hat{g}_n)$ in Taniguchi (1979). Here, the nonparametric estimator $\hat{g}_n$ is substituted for the periodogram $I_{n,X}(\omega)$ because of non-consistency in the case of nonlinear integral functional of $g(\omega)$ in the criterion. The approach minimizing criterion D between the parametrized spectral density and the nonparametric estimator is called *minimum contrast estimation*. The asymptotic properties of the approach for time series case have been considered for a long time. Taniguchi (1981a) proposed $D(f_{\boldsymbol{\theta}}, \hat{g}_n) = \int_{-\pi}^{\pi}[\Phi(f_{\boldsymbol{\theta}}(\omega))^2 - 2\Phi(f_{\boldsymbol{\theta}}(\omega))\Phi(\hat{g}_n(\omega))]d\omega$ as the disparity measure between $f_{\boldsymbol{\theta}}(\omega)$ and $\hat{g}_n(\omega)$ with a bijective function $\Phi(\cdot)$, and showed the asymptotic properties of the estimator. As another direction, Taniguchi (1987) proposed $D(f_{\boldsymbol{\theta}}, \hat{g}_n) = \int_{-\pi}^{\pi} K(f_{\boldsymbol{\theta}}(\omega)/\hat{g}_n(\omega))$

$d\omega$ as the disparity measure with a sufficiently smooth contrast function $K(\cdot)$ whose minima exists uniquely at 1, and showed the asymptotic normality based on the contrast function $K(\cdot)$.

We mainly focus on the minimum contrast estimators based on the exotic disparity. The disparity includes the integration functional form of prediction error and interpolation error up to some constant multiples. For consistency, there is another functional contained in the disparity for adjustment. In chapter 4, we show that the estimator is consistent under some appropriate conditions. The estimator also has some good properties like asymptotic normality and robustness against the fourth order cumulant. However, only the case $\alpha = -1$ leads to the efficient result, that is, the asymptotic variance of the estimation is the inverse of fisher information in time series analysis. We will give some comments on this method.

Chapter 4 is organized as follows. We review all the ideas for parameter estimation based on minimum contrast estimators as historical remarks, and investigate the properties of the exotic disparity, which is proposed in this dissertation in Section 4.2. In Section 4.3, we define the estimator based on the disparity and develop the asymptotic theory based on the exotic disparity for the linear process. The robustness and efficiency of the estimator are given in Section 4.4. The numerical results for the estimators based on the exotic disparity are given in Section 4.5. The notations and symbols used in Chapter 4 are listed in the following: the constants ϵ, δ and C denote some real numbers which vary from context; ∂_i denote the differentiation with respect to ith element of $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$; A_j and A_{ij} denote the jth and the (i, j)th element of corresponding vector and matrix; $\mathrm{cum}(X_1, \dots, X_n)$ denotes the cumulant of the random variables $\{X_1, \dots, X_n\}$; $\mathbb{1}(\cdot)$ denotes the indicator function; e denotes the Napier's constant; $\mathcal{B}(t)$ denotes the σ-field generated by the uncorrelated process $\{\epsilon(s)\}_{s=-\infty}^{t}$; the L_2-norm $\|f\|^2$ is defined by $\int_{-\pi}^{\pi} |f(\omega)|^2 d\omega$; Correspondingly, we say $f_n \xrightarrow{L^2} f$ if $\|g_n - g\|_2 \to 0$ as $n \to \infty$; $\xrightarrow{\mathcal{P}}$ and $\xrightarrow{\mathcal{L}}$ denote the convergence in probability and the convergence in law, respectively.

4.2 Classification of Disparity Measure

Suppose $\{X_t; t \in \mathbb{Z}\}$ is a stationary process with mean zero, which is generated by

$$X(t) = \sum_{j=0}^{\infty} G(j)\epsilon(t-j), \quad t \in \mathbb{Z},$$

where $E\epsilon(t) = 0$ and $E\epsilon(s)\epsilon(t) = \delta(s,t)\sigma^2$ with $\sigma^2 > 0$. Furthermore, suppose that $\{\epsilon(t)\}$ is fourth order stationary, and $Q_\epsilon(t_1, t_2, t_3)$ is the joint fourth order cumulant of $\epsilon(t), \epsilon(t+t_1), \epsilon(t+t_2), \epsilon(t+t_3)$.

Let $\mathcal{F}$ denote the family of all spectral densities with respect to the Lebesgue measure on $[-\pi, \pi]$. More specifically, we define $\mathcal{F}$ as

$$\mathcal{F} = \left\{ g : g(\omega) = \frac{\sigma^2}{2\pi} \Big| \sum_{j=0}^{\infty} G(j) e^{-ij\omega} \Big|^2 \right\}.$$

Not so strongly, we suppose the following Assumption 4.2.1.

Assumption 4.2.1. For all $|z| \le 1$, there exist $C < \infty$ and $\delta > 0$ such that

(i) $\sum_{j=0}^{\infty} (1 + j^2) |G(j)| \le C$,

(ii) $\left| \sum_{j=0}^{\infty} G(j) z^j \right| \ge \delta$,

(iii) $\sum_{t_1, t_2, t_3 = -\infty}^{\infty} |Q_\epsilon(t_1, t_2, t_3)| < \infty$.

The absolute summability of $Q_\epsilon(t_1, t_2, t_3)$ guarantee the existence of a fourth-order spectral density $\tilde{Q}_\epsilon(\omega_1, \omega_2, \omega_3)$ such that

$$\tilde{Q}_\epsilon(\omega_1, \omega_2, \omega_3) = \left(\frac{1}{2\pi}\right)^3 \sum_{t_1, t_2, t_3 = -\infty}^{\infty} Q_\epsilon(t_1, t_2, t_3) e^{-i(\omega_1 t_1 + \omega_2 t_2 + \omega_3 t_3)}.$$

The parametric family of the spectral densities with respect to $\mathcal{F}$ is given by

$$\mathcal{F}(\Theta) = \{ f_{\boldsymbol{\theta}}(\omega) \in \mathcal{F}; \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p \}.$$

A functional T on all spectral densities $\mathcal{F}$ is defined by

$$D(f_{T(g)}, g) = \min_{t \in \Theta} D(f_t, g), \quad \text{for every } g \in \mathcal{F}.$$

For $g(\omega)$, we suppose $\hat{g}_n(\omega)$ is a consistent estimator for it. It is always possible to construct a consistent estimator of $\Phi(g(\omega))$ if $\Phi(\cdot)$ is known under some regularity conditions. See Taniguchi (1980).

4.2.1 Location disparity

First, we give the well-known location disparity defined in Taniguchi (1981a). The criterion is given by

$$D_l(f_{\boldsymbol{\theta}}, \hat{g}_n) = \int_{-\pi}^{\pi} \Phi(f_{\boldsymbol{\theta}}(\omega))^2 - 2\Phi(f_{\boldsymbol{\theta}}(\omega))\Phi(\hat{g}_n(\omega)) d\omega.$$

For example, the appropriate bijective function $\Phi(\cdot)$ can be chose as

(i) $\Phi(x) = \log x$;

(ii) $\Phi(x) = 1$.

The first choice of the function $\Phi(\cdot)$ is based on the exponential model, given in Bloomfield (1973). The motivation of the idea is that the discrete periodogram $I_{n,X}(\lambda_k)$ for every different λ_k is asymptotically exponential distributed. This is what we have seen in Chapter 3. The second choice, the best choice in the sense of asymptotic efficiency, is given in Theorem 4 in Taniguchi (1981a). The disparity is also directly corresponding to the scale disparity given below.

4.2.2 Scale disparity

Next, we give another generalization of criterion, scale disparity, which is first studied in Taniguchi (1987). The discussion on the higher order asymptotics of the disparity is given in Taniguchi et al. (2003). The criterion is given by

$$D_s(f_{\boldsymbol{\theta}}, \hat{g}_n) = \int_{-\pi}^{\pi} K(f_{\boldsymbol{\theta}}(\omega)/\hat{g}_n(\omega))d\omega,$$

where K is sufficiently smooth with its minimum at 1. Without loss of generality, we also can consider some function $\tilde{K}$ that

$$\tilde{K}\Big(\frac{f_{\boldsymbol{\theta}}(\omega)}{\hat{g}_n(\omega)} - 1\Big) = K(f_{\boldsymbol{\theta}}(\omega)/\hat{g}_n(\omega)),$$

where the minima of $\tilde{K}$ is 0. The examples of K are

(i) $K(x) = \log x + 1/x$;

(ii) $K(x) = -\log x + x$;

(iii) $K(x) = (\log x)^2$;

(iv) $K(x) = (x^\alpha - 1)^2, \quad \alpha \neq 0$;

(v) $K(x) = x \log x - x$;

(vi) $K(x) = \log\{(1-\alpha) + \alpha x\} - \alpha \log x, \quad \alpha \in (0, 1)$.

The scale disparity, in fact, is much broad in concept. The choice (i) and (iii) are considered in the location case. The choice (vi) is given in Albrecht (1984) as an α-entropy criterion for a Gaussian process. The case is robust with respect to the peak, which is discussed in Zhang and Taniguchi (1995).

4.2.3 Exotic disparity

In this chapter, we mainly focus on the new disparity, called *exotic disparity* in the context, and look into the asymptotic properties of the criterion. For α and $\beta \in \mathbb{R}$, the exotic disparity is given by

$$D_e(f_{\boldsymbol{\theta}}, I_{n,X}) = \int_{-\pi}^{\pi} a(\boldsymbol{\theta}) f_{\boldsymbol{\theta}}(\omega)^{\alpha} I_{n,X}(\omega) d\omega, \quad \alpha \neq 0, \tag{4.2.1}$$

where $a(\boldsymbol{\theta})$ is given by

$$a(\boldsymbol{\theta}) = \Big(\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}(\omega)^{\alpha+1} d\omega\Big)^{\beta}.$$

This disparity is not included in either location disparity or scale disparity since the parametrized spectral density and the true density would not be homogenous in most cases. However, the definition of the disparity is motivated by the following two equivalent examples up to a constant multiple:

(i) the Whittle disparity is given when $\alpha = -1$;

(ii) the error based on interpolation is given when $\alpha = -2$ and $\beta = -2$.

Without confusion, we only write $D(\cdot,\cdot)$ for $D_e(\cdot,\cdot)$ in the following.

The formulation of the disparity does not guarantee any fundamental property for it. At first, we have to consider the problem of definition of the functional $T(g)$ for the disparity. From now on, we suppose $f_{\boldsymbol{\theta}}$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$ and simplify the notation for differentiation as follows: Let ∂_i denotes $\partial/\partial\theta_i$,

$$\begin{aligned}
A_1(\boldsymbol{\theta}) &= \int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}^{\alpha+1}(\lambda) d\lambda, \\
A_2(\boldsymbol{\theta})_i &= \int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}^{\alpha}(\lambda) \partial_i f_{\boldsymbol{\theta}}(\lambda) d\lambda, \\
A_3(\boldsymbol{\theta})_{ij} &= \int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}^{\alpha-1}(\lambda) \partial_i f_{\boldsymbol{\theta}}(\lambda) \partial_j f_{\boldsymbol{\theta}}(\lambda) d\lambda, \\
B_1(\boldsymbol{\theta})_i &= f_{\boldsymbol{\theta}}^{\alpha-1}(\omega) \partial_i f_{\boldsymbol{\theta}}(\omega), \\
B_2(\boldsymbol{\theta}) &= f_{\boldsymbol{\theta}}^{\alpha}(\omega), \\
C_1(\boldsymbol{\theta}) &= \beta \left(\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}^{\alpha+1}(\lambda) d\lambda\right)^{\beta-1}.
\end{aligned}$$

We give two properties for reference hereafter. Before looking at the first property, we review Kolmogorov's Formula:

Lemma 4.2.2 (Kolmogorov's Formula)**.** *The one step mean square prediction error of the stationary process* $\{X_t\}$ *is*

$$\sigma^2 = 2\pi \exp\Big\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f(\omega)d\omega\Big\}.$$

If the parameter $\boldsymbol{\theta}$ *is innovation free, then for all* $i = 1, \dots, p$*,*

$$\partial_i \int_{-\pi}^{\pi}\log f_{\boldsymbol{\theta}}(\omega)d\omega = 0.$$

The first property is about the information of the extreme value of the exotic disparity.

Lemma 4.2.3. *For the exotic disparity with either the case of*

(i) $\alpha \neq -1$ *and* $\beta = -\frac{\alpha}{\alpha+1}$ *or*

(ii) $\alpha = -1$ *and* $\boldsymbol{\theta}$ *is innovation free,*

it holds that

$$\partial_i D(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}^0})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = 0, \quad \textit{for all } 1 \leq i \leq p.$$

Proof. It is easy to see that if $\alpha \neq -1$, then

$$\partial_i D(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}^0})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = (\alpha+1)C_1(\boldsymbol{\theta}^0)A_1(\boldsymbol{\theta}^0)A_2(\boldsymbol{\theta}^0)_i + \beta^{-1}\alpha C_1(\boldsymbol{\theta}^0)A_1(\boldsymbol{\theta}^0)A_2(\boldsymbol{\theta}^0)_i. \tag{4.2.2}$$

The conclusion follows $\beta = -\frac{\alpha}{\alpha+1}$. If $\alpha = -1$, then the result is just an extension of Kolmogorov's formula. □

To see next property, we need Hölder's inequality for $p > 0$ $(p \neq 1)$. Suppose p' satisfy

$$\frac{1}{p} + \frac{1}{p'} = 1.$$

Lemma 4.2.4 (Hewitt and Stromberg (1975))**.** *Suppose* $f \in L_p$*,* $g \in L_{p'}$*.*

(i) *If* $p > 1$*, then*

$$\|fg\|_1 \leq \|f\|_p\|g\|_{p'}.$$

(ii) *If* $0 < p < 1$ *and further suppose* $f \in L_p^+$ *and* $g \in L_{p'}^+$*, then*

$$\|fg\|_1 \geq \|f\|_p\|g\|_{p'}.$$

The equality holds if and only if

$$|f|^p = C|g|^{p'}, \quad a.e.$$

Remark 4.2.5. Note that if $0 < p < 1$, then $p' < 0$ and vice versa. That is to say, (ii) is equivalent to

(ii)' if $p < 0$ and $f \in L_p^+$ and $g \in L_{p'}^+$, then

$$\|fg\|_1 \geq \|f\|_p \|g\|_{p'}.$$

Assumption 4.2.6.

(i) $\alpha = -1$ and $\boldsymbol{\theta}$ is innovation free or $\beta = -\frac{\alpha}{\alpha+1}$ $(\alpha \neq -1)$.

(ii) Θ is a compact subset of $\mathbb{R}^p$.

(iii) If $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, then $f_{\boldsymbol{\theta}_1} \neq f_{\boldsymbol{\theta}_2}$ on a set of positive Lebesgue measure.

The second property is that $\boldsymbol{\theta}^0$ can be the extreme value for the exotic disparity in the regular situation, even if the fitted spectral density $f_{\boldsymbol{\theta}}$ is not differentiable.

Lemma 4.2.7. *Under Assumption 4.2.6, we have the following results:*

(i) *If $\alpha > 0$, then $\boldsymbol{\theta}^0$ maximize the exotic disparity $D(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}^0})$ for any $f_{\boldsymbol{\theta}^0} \in \mathcal{F}(\Theta)$.*

(ii) *If $\alpha < 0$, then $\boldsymbol{\theta}^0$ minimize the exotic disparity $D(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}^0})$ for any $f_{\boldsymbol{\theta}^0} \in \mathcal{F}(\Theta)$.*

Proof. First, suppose $\alpha > 0$. The exotic disparity (4.2.6) can be rewritten by

$$D(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}^0}) = \frac{\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}(\omega)^\alpha f_{\boldsymbol{\theta}^0}(\omega) d\omega}{(\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}(\omega)^{\alpha+1} d\omega)^{\frac{\alpha}{\alpha+1}}}.$$

From Lemma 4.2.4, the numerator then satisfies

$$\begin{aligned}
\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}(\omega)^\alpha f_{\boldsymbol{\theta}^0}(\omega) d\omega &\leq \left(\int_{-\pi}^{\pi} \{f_{\boldsymbol{\theta}}(\omega)^\alpha\}^{\frac{\alpha+1}{\alpha}} d\omega\right)^{\frac{\alpha}{\alpha+1}} \left(\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}^0}(\omega)^{\alpha+1} d\omega\right)^{\frac{1}{\alpha+1}} \\
&= \left(\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}}(\omega)^{\alpha+1} d\omega\right)^{\frac{\alpha}{\alpha+1}} \left(\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}^0}(\omega)^{\alpha+1} d\omega\right)^{\frac{1}{\alpha+1}}.
\end{aligned}$$

Therefore,

$$D(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}^0}) \leq \left(\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}^0}(\omega)^{\alpha+1} d\omega\right)^{\frac{1}{\alpha+1}}.$$

The equality holds only when $f_{\boldsymbol{\theta}} = f_{\boldsymbol{\theta}^0}$ almost everywhere. From Assumptions 4.2.6 (ii) and (iii), the conclusion holds.

On the other hand, if $\alpha < 0$, then there are three cases (a) $-1 < \alpha < 0$, (b) $\alpha < -1$ and (c) $\alpha = -1$ must be considered. However, it is easy to see that both first two cases are corresponding to the case (ii) and (ii)' in Hölder's inequality since if $-1 < \alpha < 0$ then $(\alpha + 1)/\alpha < 0$ and if $\alpha < -1$ then $0 < (\alpha + 1)/\alpha < 1$. As a result, we obtain

$$D(f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}^0}) \geq \Big(\int_{-\pi}^{\pi} f_{\boldsymbol{\theta}^0}(\omega)^{\alpha+1} d\omega\Big)^{\frac{1}{\alpha+1}},$$

with a minima from Assumption 4.2.6. For the case (c), the disparity is corresponding to the predictor error. There is a lower bound for the disparity. (See Proposition 10.8.1 in Brockwell and Davis (1991).) □

Before considering the asymptotic properties of the disparity, we finally give a stronger result with stronger assumptions, supposing that the fitted spectral density is twice continuously differentiable, to understand the exotic disparity well. To see the result, we have to first look at the following Lemma, which is a generalization of the Cauchy-Bunyakovsky inequality, first given in Grenander and Rosenblatt (1957) in the context of time series analysis and then the paper by Kholevo (1969) later on.

Lemma 4.2.8 (Grenander and Rosenblatt (1957), Kholevo (1969))**.** *Let $A(\omega)$, $B(\omega)$ be $r \times s$ matrix-valued functions, and $g(\omega)$ a positive function on $\omega \in [-\pi, \pi]$. If*

$$\Big\{\int_{-\pi}^{\pi} B(\omega)B(\omega)^T g(\omega)^{-1} d\omega\Big\}^{-1}$$

exists, the following inequality

$$\begin{aligned}&\int_{-\pi}^{\pi} A(\omega)A(\omega)^T g(\omega) d\omega \\ &\geq \Big\{\int_{-\pi}^{\pi} A(\omega)B(\omega)^T d\omega\Big\}\Big\{\int_{-\pi}^{\pi} B(\omega)B(\omega)^T g(\omega)^{-1} d\omega\Big\}^{-1}\Big\{\int_{-\pi}^{\pi} A(\omega)B(\omega)^T d\omega\Big\}^T\end{aligned} \tag{4.2.3}$$

holds. In (4.2.3), the equality holds if and only if there exists a constant matrix C such that

$$g(\omega)A(\omega) + CB(\omega) = O, \quad a.e.\ \omega \in [-\pi, \pi]. \tag{4.2.4}$$

Lemma 4.2.9. *Suppose $g \in \mathcal{F}(\Theta)$, that is, the spectral density g can be represented by p parameters. Under Assumption 4.2.6, if $\alpha > 0$, then the exotic disparity $D(f_{\boldsymbol{\theta}}, g)$ is convex upward with respect to $\boldsymbol{\theta}$. Inversely, if $\alpha < 0$, then the exotic disparity $D(f_{\boldsymbol{\theta}}, g)$ is convex downward with respect to $\boldsymbol{\theta}$.*

Proof. Suppose $\alpha \neq -1$ and $\beta = -\frac{\alpha}{\alpha+1}$. Then

$$\partial_i D(f_{\boldsymbol{\theta}}, g) = (\alpha+1)C_1(\boldsymbol{\theta})\Big\{A_1(\boldsymbol{\theta})\int_{-\pi}^{\pi} B_1(\boldsymbol{\theta})_i g(\omega) d\omega - A_2(\boldsymbol{\theta})\int_{-\pi}^{\pi} B_2(\boldsymbol{\theta})_i g(\omega) d\omega\Big\}$$

If $g = f_{\boldsymbol{\theta}}$, then it is easy to see that $\partial_i D(f_{\boldsymbol{\theta}}, g) = 0$ for any $i = 1, \dots, p$. Considering twice derivative of $D(f_{\boldsymbol{\theta}}, g)$, we have

$$\partial_i \partial_j D(f_{\boldsymbol{\theta}}, g) = (\alpha + 1) C_1(\boldsymbol{\theta}) \Big(A_1(\boldsymbol{\theta}) A_3(\boldsymbol{\theta})_{ij} - A_2(\boldsymbol{\theta})_i A_2(\boldsymbol{\theta})_j \Big)$$

if $g = f_{\boldsymbol{\theta}}$. Regarding $A(\omega)$, $B(\omega)$, $g(\omega)$ in Lemma 4.2.8 as

$$\begin{aligned} A(\omega) &= f_{\boldsymbol{\theta}}^{\alpha/2}(\omega) \\ B(\omega) &= f_{\boldsymbol{\theta}}^{\alpha/2}(\omega) \partial_i f_{\boldsymbol{\theta}}(\omega) \\ g(\omega) &= f_{\boldsymbol{\theta}}(\omega), \end{aligned}$$

Then we can see that the matrix $A_1(\boldsymbol{\theta}) A_3(\boldsymbol{\theta})_{ij} - A_2(\boldsymbol{\theta})_i A_2(\boldsymbol{\theta})_j$ is positive definite. Since $(\alpha+1)\beta = -\alpha$, the conclusion of the convexity of the exotic disparity holds. The convexity in the case of $\alpha = -1$ is also easy to show. □

Returning to Lemma 4.2.7, it is shown that the true value is a maxima or a minima $\boldsymbol{\theta} \in \Theta$ for the criterion $D(f_{\boldsymbol{\theta}}, g)$. The possibility of being a maxima or a minima is symmetric with respect to $\alpha = 0$. To keep uniformity of the context, we suppose $\alpha < 0$. That is, the functional T has the same definition as

$$D(f_{T(g)}, g) = \min_{t \in \Theta} D(f_t, g), \quad \text{for every } g \in \mathcal{F}. \tag{4.2.5}$$

We also suppose if $g \in \mathcal{F}(\Theta)$,

$$\boldsymbol{\theta}^0 = T(g). \tag{4.2.6}$$

4.3 Estimation Theory Based on Exotic Disparity

In this section, we investigate asymptotic behavior of the parameter estimation based on exotic disparity. The result directly follows the well-known asymptotic properties of smoothed periodogram. For simplicity, we suppose $p = 1$, that is, $\theta \in \Theta \subset \mathbb{R}$ from now on. The case of general p is only a technical problem. $\alpha < 0$ is also assumed in this section.

For the linear process $\{X(t); t \in \mathbb{Z}\}$, $I_{n,X}(\omega)$ denotes the periodogram constructed from a partial realization $\{X(1), \dots, X(n)\}$, that is,

$$I_{n,X}(\omega) = \frac{1}{2\pi n} \Big| \sum_{t=1}^{n} X(t) e^{it\omega} \Big|^2, \quad -\pi \le \omega \le \pi.$$

For simplicity, we sometimes use

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} D(f_\theta, I_{n,X}) \tag{4.3.1}$$

Theorem 4.3.1. *Under Assumptions 4.2.1 and 4.2.6, we have the following results.*

(i) *For every $g \in \mathcal{F}$, there exists a value $T(g) \in \Theta$ satisfying (4.2.5).*

(ii) *If $T(g)$ is unique and if $g_n \xrightarrow{L^2} g$, then $T(g_n) \to T(g)$ as $n \to \infty$.*

(iii) *$T(f_\theta) = \theta$ for every $\theta \in \Theta$.*

Proof. (i) Define $h(\theta)$ by $h(\theta) = D(f_\theta, g)$. If the continuity of $h(\theta)$ in $\theta \in \Theta$ is shown, then the existence of $T(g)$ follows the compactness of Θ. From Assumption 4.2.1 and Lemma 4.2.4,

$$h(\theta) \leq \left(\int_{-\pi}^{\pi} g(\omega)^{\alpha+1} d\omega\right)^{\frac{1}{\alpha+1}} \leq C.$$

By Lebesgue's dominated convergence theorem,

$$|h(\theta_n) - h(\theta)| \leq \left|\int_{-\pi}^{\pi} (a(\theta_n) f_{\theta_n}(\omega)^\alpha - a(\theta) f_\theta(\omega)^\alpha) g(\omega) d\omega\right| \to 0$$

for any convergence sequence $\{\theta_n \in \Theta; \theta_n \to \theta\}$, which the continuity of $h(\theta)$ follows.

(ii) Similarly, suppose $h_n(\theta) = D(f_\theta, g_n)$. Then

$$\begin{aligned}
\lim_{n\to\infty} \sup_{\theta \in \Theta} |h_n(\theta) - h(\theta)| &= \lim_{n\to\infty} \sup_{\theta \in \Theta} \left|\int_{-\pi}^{\pi} (a(\theta) f_\theta(\omega)^\alpha)(g_n(\omega) - g(\omega)) d\omega\right| \\
&\leq \lim_{n\to\infty} \sup_{\theta \in \Theta} \left|\int_{-\pi}^{\pi} (a(\theta) f_\theta(\omega)^\alpha)^2 d\omega \right.\\
&\qquad \left. \times \int_{-\pi}^{\pi} (g_n(\omega) - g(\omega))^2 d\omega\right|^{1/2} \\
&\leq C \lim_{n\to\infty} \sup_{\theta \in \Theta} \left|\int_{-\pi}^{\pi} (g_n(\omega) - g(\omega))^2 d\omega\right|^{1/2} \\
&= 0,
\end{aligned}$$

by $g_n \xrightarrow{L^2} g$. As a result, we obtain

$$\begin{aligned}
|h_n(T(g_n)) - h(T(g))| &\to 0, \\
|h_n(T(g_n)) - h(T(g_n))| &\to 0,
\end{aligned}$$

and therefore

$$h(T(g_n)) \to h(T(g)).$$

The conclusion follows the uniqueness of $T(g)$.

(iii) This is equivalent to Lemma 4.2.7 (ii), which we have shown before.

□

Assumption 4.3.2. The spectral density $f_\theta(\omega)$ is three times continuously differentiable with respect to θ and the second derivative $\frac{\partial^2}{\partial\theta^2} f_\theta(\omega)$ is continuous in ω.

Theorem 4.3.3. *Suppose that $T(g)$ exists uniquely and lies in* $\mathrm{Int}(\Theta)$. *For every spectral density sequence $\{g_n\}$ satisfying $g_n \xrightarrow{L^2} g$, we have*

$$T(g_n) = T(g) - \int_{-\pi}^{\pi} \rho(\omega)(g_n(\omega) - g(\omega))d\omega,$$

where

$$\rho(\omega) = \Big(A_1(\theta^0)A_3(\theta^0) - A_2(\theta^0)^2\Big)^{-1}\Big(A_1(\theta^0)B_1(\theta^0) - A_2(\theta^0)B_2(\theta^0)\Big).$$

Proof. In view of $T(g) \in \mathrm{Int}(\Theta)$, we have

$$\begin{aligned} D(f_\theta, g_n)\Big|_{\theta=T(g_n)} &= 0, \\ D(f_\theta, g)\Big|_{\theta=T(g)} &= 0. \end{aligned}$$

Then there exists a θ^* such that $T(g) \le \theta^* \le T(g_n)$ and

$$\begin{aligned} T(g_n) - T(g) = \Big\{(\alpha+1)C_1(\theta)\Big(A_1(\theta)A_3(\theta)_{ij} - A_2(\theta)_i A_2(\theta)_j\Big)\Big|_{\theta=\theta^*}\Big\}^{-1} \\ \int_{-\pi}^{\pi} A_1(\theta^0)B_1(\theta^0) - A_2(\theta^0)B_2(\theta^0)(g_n - g)d\omega. \end{aligned}$$

From the fact that

$$\begin{aligned} (\alpha+1)C_1(\theta)\Big(A_1(\theta)A_3(\theta)_{ij} - A_2(\theta)_i A_2(\theta)_j\Big)\Big|_{\theta=\theta^*} \\ - \Big((\alpha+1)C_1(\theta^0)(A_1(\theta^0)A_3(\theta^0) - A_2(\theta^0)^2\Big) \end{aligned}$$

is bounded by $C|T(g_n) - T(g)|$, the conclusion follows. □

To see asymptotic properties of the estimation based on the exotic disparity, we first impose assumptions on innovation process for limit theorem of integration functional of periodogram. Assumption 4.3.4 is much weaker since the conditions guarantee the approximation of martingale sequence for innovation process based on martingale central limit theorem. Denote by $\mathcal{B}(t)$ the σ-field generated by $\{\epsilon(n); n \le t\}$.

Assumption 4.3.4.

(i) For each nonnegative integer m and $\eta_1 > 0$,

$$\mathrm{Var}[E\{\epsilon(t)\epsilon(t+m)|\mathcal{B}(t-\tau)\} - \delta(m,0)\sigma^2] = O(\tau^{-2-\eta_1})$$

uniformly in t.

(ii) For $\eta_2 > 0$,

$$E|E\{\epsilon(t_1)\epsilon(t_2)\epsilon(t_3)\epsilon(t_4)|\mathcal{B}(t_1-\tau)\} - E\{\epsilon(t_1)\epsilon(t_2)\epsilon(t_3)\epsilon(t_4)\}| = O(\tau^{-1-\eta_2}),$$

uniformly in t_1, where $t_1 \le t_2 \le t_3 \le t_4$.

(iii) For any $\eta_3 > 0$ and for any integer $L \ge 0$, there exists $B_\rho > 0$ such that

$$E[T(n,s)^2 \mathbb{1}\{T(n,s) > B_{\eta_3}\}] < \eta_3$$

uniformly in n and s, where

$$T(n,s) = n^{-1/2} \sum_{r=0}^{L} \Big\{ \sum_{t=1}^{n} \epsilon(t+s)\epsilon(t+s+r) - \sigma^2\delta(0,r) \Big\}^2$$

Lemma 4.3.5 (Hosoya and Taniguchi (1982), Taniguchi and Kakizawa (2000))**.** *Suppose $\psi(\omega)$ is a $p\times 1$ vector-valued symmetric continuous function on $[-\pi,\pi]$. Under Assumption 4.3.4, we have*

(a) *the consistency*

$$\int_{-\pi}^{\pi} \psi(\omega) I_{n,X}(\omega) d\omega \xrightarrow{\mathcal{P}} \int_{-\pi}^{\pi} \psi(\omega) f_{\theta^0}(\omega) d\omega,$$

(b) *and asymptotic normality that*

$$\sqrt{n} \int_{-\pi}^{\pi} \psi(\omega)\{I_{n,X}(\omega) - f_{\theta^0}(\omega)\} d\omega \xrightarrow{\mathcal{L}} \mathcal{N}(0,V),$$

where

$$V = 4\pi \int_{-\pi}^{\pi} \psi(\omega)\psi(\omega)^T f_{\theta^0}(\omega)^2 d\omega + 2\pi \iint_{-\pi}^{\pi} \psi(\omega_1)\psi(\omega_2)^T \tilde{Q}_X(-\omega_1,\omega_2,-\omega_2) d\omega_1 d\omega_2.$$

Theorem 4.3.6. *Under Assumptions 4.2.1, 4.2.6, 4.3.2 and 4.3.4, if $T(g)$ exists uniquely in* $\mathrm{Int}(\Theta)$*, then for the estimator $\hat{\theta}_n$ defined by* (4.3.1)*, it holds that*

$$
\begin{aligned}
&\text{(a)} \quad \hat{\theta}_n \xrightarrow{p} \theta^0, \\
&\text{(b)} \quad \sqrt{n}(\hat{\theta}_n - \theta^0) \xrightarrow{d} \mathcal{N}(0, H(\theta^0)^{-1} V(\theta^0) H(\theta^0)^{-1}),
\end{aligned}
$$

where

$$
\begin{aligned}
H(\theta^0) &= \left(\int_{-\pi}^{\pi} f_\theta^\alpha(\lambda) \partial f_\theta(\lambda) d\lambda\right)^2 - \int_{-\pi}^{\pi} f_\theta^{\alpha+1}(\lambda) d\lambda \int_{-\pi}^{\pi} f_\theta^{\alpha-1}(\lambda) \Big(\partial f_\theta(\lambda)\Big)^2 d\lambda \Bigg|_{\theta=\theta^0}, \\
V(\theta^0) &= \Bigg[4\pi \int_{-\pi}^{\pi} \Big(f_\theta^\alpha(\omega) \partial f_\theta(\omega) \int_{-\pi}^{\pi} f_\theta^{\alpha+1}(\lambda) d\lambda - f_\theta^{\alpha+1}(\omega) \int_{-\pi}^{\pi} f_\theta^\alpha(\lambda) \partial f_\theta(\lambda) d\lambda\Big)^2 d\omega \\
&\quad + 2\pi \iint_{-\pi}^{\pi} \Big(f_\theta^{\alpha-1}(\omega_1) \partial f_\theta(\omega_1) \int_{-\pi}^{\pi} f_\theta^{\alpha+1}(\lambda) d\lambda - f_\theta^\alpha(\omega_1) \int_{-\pi}^{\pi} f_\theta^\alpha(\lambda) \partial f_\theta(\lambda) d\lambda\Big) \\
&\qquad\qquad \times \Big(f_\theta^{\alpha-1}(\omega_2) \partial f_\theta(\omega_2) \int_{-\pi}^{\pi} f_\theta^{\alpha+1}(\lambda) d\lambda - f_\theta^\alpha(\omega_2) \int_{-\pi}^{\pi} f_\theta^\alpha(\lambda) \partial f_\theta(\lambda) d\lambda\Big) \\
&\qquad\qquad\qquad \times \tilde{Q}_X(-\omega_1, \omega_2, -\omega_2) d\omega_1 d\omega_2\Bigg] \Bigg|_{\theta=\theta^0}.
\end{aligned}
$$

Proof. In view of (4.3.1), it is equivalent to consider $\hat{\theta}_n$ satisfies

$$
\partial D(f_\theta, I_{n,X}) \Big|_{\theta=\hat{\theta}_n} = 0.
$$

The result that $\hat{\theta}_n \xrightarrow{\mathcal{P}} \theta^0$ follows that for any $\theta \in \Theta$ compact.

$$
\partial D(f_\theta, I_{n,X}) \xrightarrow{\mathcal{P}} \partial D(f_\theta, f_{\theta^0}),
$$

which is guaranteed by Lemma 4.3.5 (a). Differentiating the disparity (4.2.6) with respect to θ, then we have

$$
\partial D(f_\theta, I_{n,X}) = C_1(\theta) \int_{-\pi}^{\pi} (A_1(\theta) B_1(\theta) - A_2(\theta) B_2(\theta)) I_{n,X}(\omega) d\omega.
$$

Note that $\partial D(f_\theta, f_{\theta^0})\Big|_{\theta=\theta^0} = 0$. Asymptotic normality follows that by Lemma 4.3.5 (b),

$$
\begin{aligned}
\partial D(f_\theta, I_{n,X}) \Big|_{\theta=\theta^0} &= C_1(\theta^0) \int_{-\pi}^{\pi} (A_1(\theta^0) B_1(\theta^0) - A_2(\theta^0) B_2(\theta^0)) \\
&\qquad\qquad \times \big(I_{n,X}(\omega) - f_{\theta^0}(\omega)\big) d\omega \\
&\xrightarrow{\mathcal{L}} \mathcal{N}\left(0, C_1(\theta^0)^2 V(\theta^0)\right).
\end{aligned}
$$

Noting that $\partial D(f_\theta, f_{\theta^0})\Big|_{\theta=\theta^0} = 0$ again, we see that

$$\left|\partial^2 D(f_\theta, I_{n,X}) - C_1(\theta)\partial\Big(\int_{-\pi}^{\pi}(A_1(\theta)B_1(\theta) - A_2(\theta)B_2(\theta))I_{n,X}(\omega)d\omega\Big)\Big|_{\theta=\theta^0}\right| \xrightarrow{\mathcal{P}} 0.$$

We also have

$$\begin{aligned}\int_{-\pi}^{\pi} B_1(\theta)f_\theta(\omega)d\omega\,\partial A_1(\theta) - \int_{-\pi}^{\pi} A_2(\theta)f_\theta(\omega)d\omega\,\partial B_2(\theta) &= \Big(A_2(\theta)\Big)^2\\ \int_{-\pi}^{\pi} A_1(\theta)f_\theta(\omega)d\omega\,\partial B_1(\theta) - \int_{-\pi}^{\pi} B_2(\theta)f_\theta(\omega)d\omega\,\partial A_2(\theta) &= -A_1(\theta)A_3(\theta),\end{aligned}$$

and therefore

$$C_1(\theta)\partial\Big(\int_{-\pi}^{\pi}(A_1(\theta)B_1(\theta) - A_2(\theta)B_2(\theta))I_{n,X}(\omega)d\omega\Big)\Big|_{\theta=\theta^0} \xrightarrow{\mathcal{P}} C_1(\theta^0)H(\theta^0).$$

As a result, we obtain

$$\partial^2 D(f_\theta, I_{n,X})\Big|_{\theta=\theta^0} \xrightarrow{\mathcal{P}} C_1(\theta^0)H(\theta^0).$$

Canceling $C_1(\theta^0)$, the desirable result is obtained. □

If we impose a simple but stronger assumption, asymptotic distribution of $\hat{\theta}_n$ will be much easier. The assumption is given below.

Assumption 4.3.7. The fourth order cumulant of $\epsilon(t)$ satisfies

$$\mathrm{cum}\{\epsilon(t_1), \epsilon(t_2), \epsilon(t_3), \epsilon(t_4)\} = \begin{cases}\kappa_4 & \text{if } t_1 = t_2 = t_3 = t_4,\\ 0 & \text{otherwise.}\end{cases}$$

Before simplifying the asymptotic distribution, we first give a result on the fourth-order spectral density of the process $\{X(t)\}$.

Lemma 4.3.8 (Hosoya and Taniguchi (1982) Lemma A2.1)**.** *If* $\sum_{j=0}^{\infty} G(j)^2 < \infty$ *and* $\sum_{t_1,t_2,t_3=-\infty}^{\infty}|Q_\epsilon(t_1,t_2,t_3)| \leq \infty$*, then the process* $\{X(t)\}$ *has a fourth-order spectral density* $\tilde{Q}_X(\omega_1, \omega_2, \omega_3)$ *such that*

$$\tilde{Q}_X(\omega_1, \omega_2, \omega_3) = k(\omega_1 + \omega_2 + \omega_3)k(-\omega_1)k(-\omega_2)k(-\omega_3),$$

where $k(\omega) = \sum_{j=0}^{\infty} G(j)e^{ij\omega}$.

Corollary 4.3.9. Under Assumption 4.2.1,

$$\tilde{Q}_X(-\omega_1, \omega_2, -\omega_2) = \sigma^{-4} f_{\theta^0}(\omega_1) f_{\theta^0}(\omega_2).$$

Theorem 4.3.10. *Suppose Assumptions 4.2.1, 4.2.6, 4.3.2 and 4.3.7 hold. If $g = f_{\theta^0}$, then asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta^0)$ is given by $H(\theta^0)^{-1}\tilde{V}(\theta^0)H(\theta^0)^{-1}$, where*

$$\tilde{V}(\theta^0) = \Bigg[4\pi \int_{-\pi}^{\pi} \Big(f_\theta^\alpha(\omega)\partial f_\theta(\omega) \int_{-\pi}^{\pi} f_\theta^{\alpha+1}(\lambda) d\lambda \\ - f_\theta^{\alpha+1}(\omega) \int_{-\pi}^{\pi} f_\theta^\alpha(\lambda)\partial f_\theta(\lambda) d\lambda\Big)^2 d\omega\Bigg]\Bigg|_{\theta=\theta^0}.$$

Proof. The result directly follows the equation (4.2.2), that is,

$$\partial D(f_\theta, f_{\theta^0}) = 0.$$

As a result, the last term in the asymptotic variance is 0. □

4.4 Robustness and Efficiency of $T(g)$

4.4.1 Robustness of $T(g)$

First, we can see that the minimum contrast estimation based on exotic disparity is robust against the fourth cumulant. If the process $\{X(t)\}$ considered is 1-dimension, then the property holds even if $f(\omega) \neq f_\theta(\omega)$ from Theorem 4.3.10. Recently, the author found there has been a large amount of discussion on the robustness of the minimum contrast estimation based on the exotic disparity from different points of view. We do not plan to step into the robustness in the paper. See Basu et al. (1998), Fujisawa and Eguchi (2008) and Kanamori and Fujisawa (2014).

4.4.2 Efficiency of $T(g)$

In this section, we focus on the asymptotic variance of the estimators $\hat{\theta}_n$ generated by different α. It is well known in time series analysis that the Fisher information matrix for Gaussian process is asymptotically given by

$$\mathcal{F}(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} f_\theta^{-2}(\omega)(\partial f_\theta(\omega))^2 d\omega.$$

The estimator $\hat{\theta}_n$ attaining the Cramer-Rao lower bound, that is, the inverse matrix of Fisher information matrix $\mathcal{F}(\theta)^{-1}$, is called *asymptotically efficient*. The lower bound of the asymptotic variance is again, shown by Lemma 4.2.8.

Theorem 4.4.1. *Suppose Assumptions 4.2.1, 4.2.6, 4.3.2 and 4.3.7 hold. The inequality always holds that*

$$H(\theta^0)^{-1}\tilde{V}(\theta^0)H(\theta^0)^{-1} \geq \mathcal{F}(\theta^0)^{-1}. \tag{4.4.1}$$

The equality holds when $\alpha = -1$ or the spectral density does not depend on ω.

Proof. Define

$$\begin{aligned}
A(\omega) &= A_1(\theta^0)B_1(\theta^0) - A_2(\theta^0)B_2(\theta^0),\\
B(\omega) &= \partial f_\theta(\omega)\Big|_{\theta=\theta^0},\\
g(\omega) &= f_\theta^2(\omega)\Big|_{\theta=\theta^0}.
\end{aligned}$$

Then (4.4.1) holds from Lemma 4.2.8. According to (4.2.4), the equality holds when

$$\int_{-\pi}^{\pi} f_\theta^{\alpha+1}(\lambda)d\lambda\, f_\theta^{\alpha+1}(\omega)\partial f_\theta(\omega) - \int_{-\pi}^{\pi} f_\theta^{\alpha}(\lambda)\partial f_\theta(\lambda)d\lambda\, f_\theta^{\alpha+2}(\omega) - C\partial f_\theta(\omega)\Big|_{\theta=\theta^0} = 0 \tag{4.4.2}$$

with a constant c. Note that if $\alpha = -1$, then the left hand side of (4.4.2) is

$$2\pi\partial f_\theta(\omega) - \int_{-\pi}^{\pi} f_\theta^{-1}(\lambda)\partial f_\theta(\lambda)d\lambda\, f_\theta(\omega) - C\partial f_\theta(\omega)\Big|_{\theta=\theta^0}.$$

However, Lemma 4.2.2 tells us

$$\int_{-\pi}^{\pi} \log f_\theta(\lambda)d\lambda\Big|_{\theta=\theta^0} = \log\frac{\sigma^2}{2\pi},$$

which is followed by

$$\int_{-\pi}^{\pi} f_\theta^{-1}(\lambda)\partial f_\theta(\lambda)d\lambda\Big|_{\theta=\theta^0} = 0.$$

If we choose $c = 2\pi$, then the equality holds. If $\alpha \neq -1$, then (4.4.2) does not hold generally. It is easy to see that (4.2.4) holds if the spectral density does not depend on ω. □

Without obvious cases, Theorem 4.4.1 shows that if $\alpha \neq 1$, the exotic disparity is not desirable from the viewpoint of asymptotic efficiency. However, if we want to solve an estimating function, it is usually done by Newton-Raphson method. (See Hamilton (1994)). An non-iterative efficient estimating method for different parametric spectral density $f_\theta(\omega)$ is given in Taniguchi (1987). From the robustness of the exotic disparity, we can regard the estimate $\hat{\theta}_n$ as the initial value $\tilde{\theta}$ and follow the procedure

$$\hat{\theta}_e = \tilde{\theta} - \left[\frac{\partial^2}{\partial\theta\partial\theta^T} D_{-1}(f_\theta, I_{n,X})^{-1}\right]\left[\frac{\partial}{\partial\theta} D_{-1}(f_\theta, I_{n,X}),\right]$$

where $D_{-1}(f_\theta, I_{n,X})$ is corresponding to the case $\alpha = -1$. This procedure gives an efficient estimate.

4.5 Numerical Results

Consider a AR(1) model defined by

$$X(t) - \theta^0 X(t-1) = \epsilon(t), \quad \epsilon(t) \sim \text{i.i.d.}, \tag{4.5.1}$$

where the innovation process is assumed to be i.i.d. random variables. The distributions of random variables are assumed to be normal distribution, Laplace distribution, Student's t distribution with degrees of freedom 1, 2, 3, 4. We are interested in estimating the coefficient θ^0 in the model (4.5.1), although the model with Student's t distributions here are beyond our results in the previous section since they do not have enough moments.

We generate the model with the coefficients of $\theta^0 = 0.1, 0.3, 0.5, 0.7, 0.9$. The estimators for the coefficients are given by the procedure (4.3.1) with $\alpha = -3, -2, -1$. The initial value in the model is assumed to be 0 in the simulation. Since the model will be influenced by the choice of the initial value, we set a warming sample zone for 30 samples, that is to say the initial value 0 is not contained in our samples for estimation. Without the initial 30 samples, we use another 30 samples generated from the model (4.5.1) for estimation. The estimation is repeated for 100 times.

The tables with odd numbers are the mean of 100 estimations. On the other hand, the tables with even numbers are the root mean squared error (RMSE) for the estimation.

Obviously from all tables, we can see the efficiency of the estimation with exotic disparity with $\alpha = -1$. Furthermore, the choice of $\alpha = -1$ leads to a estimation robust to the unit root process and non-stationary process from the results of Tables 4.7 and 4.9.

TABLE 4.1: Mean of Estimators for $\theta^0 = 0.1$ with $\alpha = -3, -2, -1$.

Mean	$\alpha = -3$	$\alpha = -2$	$\alpha = -1$
Normal	-0.013	0.011	0.056
Laplace	-0.054	0.034	0.060
t(1)	0.027	-0.031	0.052
t(2)	-0.080	0.018	0.043
t(3)	-0.014	0.003	0.079
t(4)	-0.088	-0.0845	0.043

TABLE 4.2: RMSE of Estimators for $\theta^0 = 0.1$ with $\alpha = -3, -2, -1$.

RMSE	$\alpha = -3$	$\alpha = -2$	$\alpha = -1$
Normal	0.957	0.477	0.171
Laplace	1.005	0.538	0.162
t(1)	0.626	0.574	0.134
t(2)	1.006	0.446	0.179
t(3)	1.219	0.634	0.192
t(4)	0.893	0.773	0.205

TABLE 4.3: Mean of Estimators for $\theta^0 = 0.3$ with $\alpha = -3, -2, -1$.

Mean	$\alpha = -3$	$\alpha = -2$	$\alpha = -1$
Normal	0.483	0.290	0.220
Laplace	0.492	0.267	0.209
t(1)	0.469	0.235	0.225
t(2)	0.360	0.288	0.204
t(3)	0.387	0.363	0.225
t(4)	0.427	0.302	0.218

TABLE 4.4: RMSE of Estimators for $\theta^0 = 0.3$ with $\alpha = -3, -2, -1$.

RMSE	$\alpha = -3$	$\alpha = -2$	$\alpha = -1$
Normal	0.726	0.353	0.186
Laplace	0.746	0.506	0.188
t(1)	0.758	0.163	0.151
t(2)	0.766	0.388	0.179
t(3)	0.676	0.494	0.181
t(4)	0.739	0.564	0.202

TABLE 4.5: Mean of Estimators for $\theta^0 = 0.5$ with $\alpha = -3, -2, -1$.

Mean	$\alpha = -3$	$\alpha = -2$	$\alpha = -1$
Normal	0.852	0.681	0.398
Laplace	0.887	0.742	0.422
t(1)	0.816	0.675	0.405
t(2)	0.697	0.777	0.408
t(3)	0.828	0.634	0.412
t(4)	0.946	0.793	0.412

TABLE 4.6: RMSE of Estimators for $\theta^0 = 0.5$ with $\alpha = -3, -2, -1$.

RMSE	$\alpha = -3$	$\alpha = -2$	$\alpha = -1$
Normal	0.809	0.610	0.187
Laplace	0.852	0.641	0.175
t(1)	0.812	0.620	0.186
t(2)	0.521	0.691	0.181
t(3)	0.697	0.504	0.178
t(4)	0.861	0.680	0.180

TABLE 4.7: Mean of Estimators for $\theta^0 = 0.7$ with $\alpha = -3, -2, -1$.

Mean	$\alpha = -3$	$\alpha = -2$	$\alpha = -1$
Normal	0.981	1.057	0.569
Laplace	0.954	0.981	0.566
t(1)	1.04	1.252	0.598
t(2)	1.056	1.165	0.587
t(3)	1.005	1.053	0.564
t(4)	1.025	0.957	0.564

TABLE 4.8: RMSE of Estimators for $\theta^0 = 0.7$ with $\alpha = -3, -2, -1$.

RMSE	$\alpha = -3$	$\alpha = -2$	$\alpha = -1$
Normal	0.684	0.655	0.206
Laplace	0.572	0.570	0.198
t(1)	0.580	0.731	0.149
t(2)	0.599	0.686	0.173
t(3)	0.608	0.624	0.209
t(4)	0.699	0.565	0.213

TABLE 4.9: Mean of Estimators for $\theta^0 = 0.9$ with $\alpha = -3, -2, -1$.

Mean	$\alpha = -3$	$\alpha = -2$	$\alpha = -1$
Normal	1.043	1.167	0.716
Laplace	1.106	1.150	0.743
t(1)	1.062	1.058	0.733
t(2)	1.086	1.087	0.750
t(3)	1.093	1.128	0.718
t(4)	1.046	1.105	0.735

TABLE 4.10: RMSE of Estimators for $\theta^0 = 0.9$ with $\alpha = -3, -2, -1$.

RMSE	$\alpha = -3$	$\alpha = -2$	$\alpha = -1$
Normal	0.419	0.433	0.218
Laplace	0.475	0.419	0.201
t(1)	0.397	0.359	0.221
t(2)	0.372	0.326	0.190
t(3)	0.439	0.412	0.214
t(4)	0.464	0.411	0.211

Chapter 5

Quantile Estimation in Frequency Domain

5.1 Introduction

Nowadays, the quantile-based estimation becomes a notable method in statistics for its robustness against the moments of random variables. In this chapter, we extend the idea of quantile method in time domain to that in frequency domain. As the spectral distribution function for real-valued stationary process is also a well-behaved monotone function bounded by the autocovariance function of the process, the objective function for the quantile estimator in time domain can be naturally extended into frequency domain.

In the context of time series analysis, Whittle (1952) mentioned that "the search for periodicities" constituted the whole of time series theory. He proposed an estimation method based on a nonlinear model driven by a simple harmonic component. After the work, to estimate the frequency has been a remarkable statistical analysis. A sequential literature by Whittle (1952), Walker (1971), Hannan (1973a), Rice and Rosenblatt (1988) and Quinn and Thomson (1991) investigated the method proposed by Whittle (1952) and pointed out the misunderstandings in Whittle (1952) respectively. The noise structure included in the model is also generalized from white noise to the stationary process. The main result in those works revealed the properties of the periodogram and showed that the convergence factor of the estimator for the frequencies is $n^{3/2}$, which is different from the well known order $n^{1/2}$, although the asymptotic distribution of the method is Gaussian.

Quinn and Hannan (2001) reviewed all the results above and proposed an alternative approach based on an iterative ARMA method. In reality, they found that the nonlinear model for $Y(t)$ with a peculiar frequency structure plus stationary process $X(t)$, i.e.,

$$Y(t) = A\cos(\lambda t + \phi) + X(t)$$

can be rewritten, by the trigonometric relation, as

$$Y(t) - \beta Y(t-1) + Y(t-2) = X(t) - \alpha X(t-1) + X(t-2),$$

where $\alpha = \beta$ depend on the peculiar frequency. The method is to estimate β for given α and substitute β for α. The procedure repeats until α and β are sufficiently close.

Different from all the methods above, Koenker and Bassett (1978)'s check function for estimating quantile can be applied to the periodogram of the stationary process to estimate frequencies. In view of correspondence between the spectral density function and the periodogram for the stationary process, we first directly apply the objective function to the integration functional with the bare periodogram. Asymptotic normality of the estimator was expected from the result by Hosoya (1989) on the bracketing condition in frequency case. The approach for estimating quantiles in frequency domain certainly has the consistency for the true value. However, asymptotic normality of the quantile estimator based on the bare periodogram does not hold, which is obviously different from the quantile estimation theory in time domain. We give the results on the asymptotic properties of the estimator. The modified estimator for asymptotic normality of the estimation, smoothing the bare periodogram in other words, will also be provided. From the history of the search for periodicities, we also give numerical results in the nonlinear time series model, although the model is beyond the scope of this doctoral thesis.

In Chapter 5, we show the asymptotic results for our estimator of frequencies based on the quantile method in frequency domain. In Section 5.2, we define the quantile in frequency domain, according to the quantile defined in time domain. For the asymptotics, we review some crucial results on spectral distribution functions and derive the asymptotic properties of the perturbation of the functional of the periodogram in Section 5.3. In Section 5.4, we explore the asymptotic properties of the estimator of frequencies based on the quantile method in frequency domain. We give the improvement of the estimator in Section 5.5 since it is not asymptotic normal. The numerical results are given in Section 5.6. The notations and symbols used in this section are listed in the following: $\mathrm{cum}(X_1, \dots, X_n)$ denotes the cumulant of the random variables

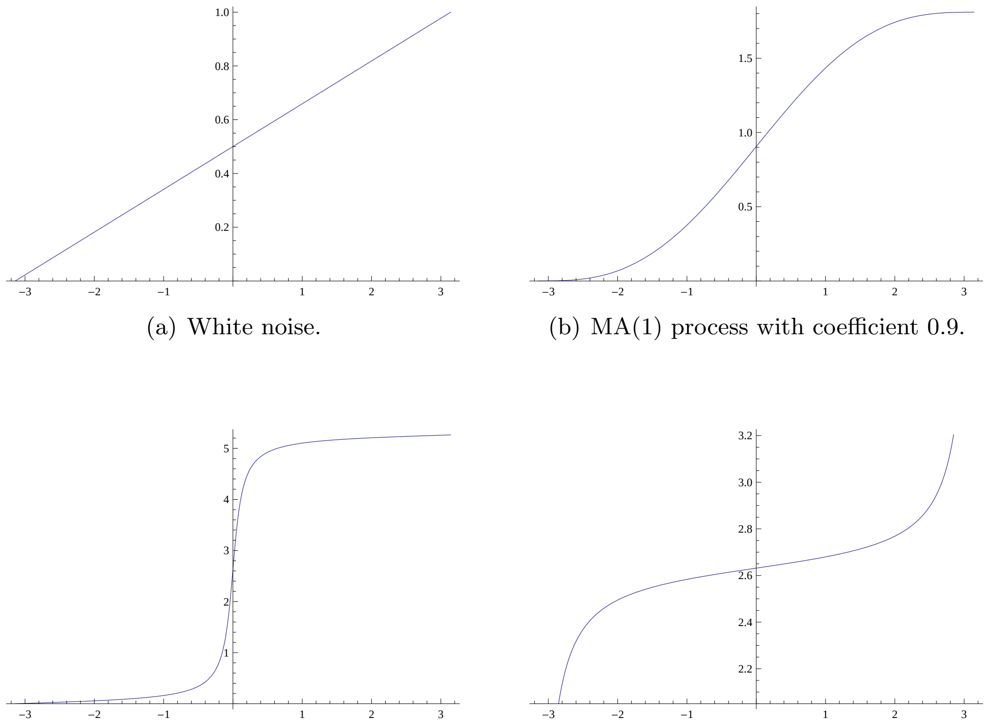
$\{X_1, \ldots, X_n\}$; $\mathbb{1}(\cdot)$ denotes the indicator function; e denotes the Napier's constant; $\mathcal{B}(t)$ denotes the σ-field generated by the uncorrelated process $\{\epsilon(s)\}_{s=-\infty}^{t}$ in the section; I_d denotes the d-dimensional identity matrix; $\xrightarrow{\mathcal{P}}$ and $\xrightarrow{\mathcal{L}}$ denote the convergence in probability and the convergence in law, respectively.

5.2 Quantiles in Frequency Domain

Suppose $\{X(t)\,;\, t \in \mathbb{Z}\}$ is a second order stationary process. The observation stretch for the process is defined by $\{X(t)\,;\, 1 \leq t \leq n\}$. From Herglotz's theorem, there exists a right continuous distribution function $F(\mu)$ for the autocovariance function $\gamma(k)$ of the process such that

$$\gamma(k) = \int_{-\pi}^{\pi} e^{ik\mu} F(d\mu), \quad (k \in \mathbb{Z}).$$

The function $F(\mu)$ is called the spectral distribution function. The structure of the second order stationary process can be discriminated by their own spectral distribution function $F(\mu)$.



(a) White noise. (b) MA(1) process with coefficient 0.9.

(c) AR(1) process with coefficient 0.9. (d) AR(1) process with coefficient -0.9.

FIGURE 5.1: Spectral distributions of second order stationary processes.

Here, we give 4 figures of spectral distribution functions of second order stationary processes with standard normal innovation. For simplicity from now on, write $R(0) = \Sigma_X$. Suppose the ψth quantile λ for the distribution function

$F(\mu)$ over $0 \le \psi \le \Sigma_X$ defined by

$$\lambda = F^{-1}(\psi) = \inf\{\mu\,;F(\mu) \ge \psi\}. \tag{5.2.1}$$

Equivalently, for $0 \le p \le 1$,

$$\lambda = \inf\{\mu\,;F(\mu)\Sigma_X^{-1} \ge p\}.$$

In the following, we propose the estimator for the frequency λ and discuss the remarkable asymptotic properties of the estimation.

In the estimation for quantiles in time domain, the check function $\rho_\tau(u)$ defined in the following way is usually used (e.g. Koenker and Bassett (1978), Koenker (2005)):

$$\rho_\tau(u) = u(\tau - \mathbb{1}(u < 0)). \tag{5.2.2}$$

The graphs of the functions (5.2.2) with different τ are shown in Figure 2.1. We minimize this check function to estimate the τth quantile of a distribution function for a random variable.

The idea can be naturally extended into frequency domain, i.e., define $\hat\lambda_n$ for λ by

$$\hat\lambda_n = \arg\min_\theta \int_{-\pi}^{\pi} \rho_p(\omega - \theta) I_{n,X}(\omega) d\omega. \tag{5.2.3}$$

Here, the periodogram $I_{n,X}(\omega)$ based on $\{X(t)\,; 1 \le t \le n\}$ is defined by

$$I_{n,X}(\omega) = \frac{1}{2\pi n}\Big|\sum_{j=1}^{n} X(j)e^{ij\omega}\Big|^2. \tag{5.2.4}$$

As an result, it is easy to check that for the spectral density $f(\omega)$, the minimizer of

$$\int_{-\pi}^{\pi} \rho_p(\omega - \theta) f(\omega) d\omega$$

is corresponding to the ψth quantile λ of the distribution function $F(\omega)$.

5.3 Some Results on Empirical Spectral Distributions

In this section, we review some results on empirical spectral distributions. Suppose $\{X(t)\}$ is second order stationary and has a spectral density $f(\omega)$, that is to say that the spectral distribution function is absolutely continuous with respect to the Lebesgue measure. Correspondingly, suppose the kth autocovariance function $\gamma(k)$ defined by

$$\gamma(k) = \int_{-\pi}^{\pi} e^{-ik\omega} f(\omega) d\omega.$$

It is a usual method that we use to estimate parameters by using the form

$$J(A) = \int_{-\pi}^{\pi} A(\omega) f(\omega) d\omega.$$

For estimating $J(A)$, we use a discrete statistic or equivalently a continuous statistic

$$\begin{aligned} J^{(n)}(A) &= \frac{2\pi}{n} \sum_{t=1}^{n-1} A\Big(\frac{2\pi t}{n}\Big) I_{n,X}\Big(\frac{2\pi t}{n}\Big) \\ &= \int_{-\pi}^{\pi} A(\omega) I_{n,X}(\omega) d\omega. \end{aligned}$$

As for this statistic, the asymptotic properties are given in the following Lemma.

Assumption 5.3.1.

(i) For each nonnegative integer m and $\eta_1 > 0$,

$$\mathrm{Var}[E\{\epsilon(t)\epsilon(t+m)|\mathcal{B}(t-\tau)\} - \delta(m,0)\sigma^2] = O(\tau^{-2-\eta_1})$$

uniformly in t.

(ii) For $\eta_2 > 0$,

$$E|E\{\epsilon(t_1)\epsilon(t_2)\epsilon(t_3)\epsilon(t_4)|\mathcal{B}(t_1-\tau)\} - E\{\epsilon(t_1)\epsilon(t_2)\epsilon(t_3)\epsilon(t_4)\}| = O(\tau^{-1-\eta_2}),$$

uniformly in t_1, where $t_1 \le t_2 \le t_3 \le t_4$.

Lemma 5.3.2 (Brillinger (2001), Theorem 7.6.1)**.** *Suppose $\{X(t)\}$ is a real valued stationary process satisfying Assumption 5.3.1. For any bounded variation function $A(\omega)$ on $[-\pi, \pi]$,*

$$\begin{aligned} EJ^{(n)}(A) &= \frac{2\pi}{n} \sum_{t=1}^{n-1} A\Big(\frac{2\pi t}{n}\Big) f\Big(\frac{2\pi t}{n}\Big) + o(1) \\ &= \int_{-\pi}^{\pi} A(\omega) f(\omega) d\omega + o(1). \qquad (5.3.1) \end{aligned}$$

Furthermore, we have

$$\mathrm{Cov}(J^{(n)}(A), J^{(n)}(A)) = \frac{2\pi}{n}\Big(\int_{-\pi}^{\pi} A(\omega)\overline{A(\omega)}f(\omega)^2 d\omega + \int_{-\pi}^{\pi} A(\omega)\overline{A(-\omega)}f(\omega)^2 d\omega\Big)$$
$$+ \frac{2\pi}{n}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} A(\omega_1)A(-\omega_2)\tilde{Q}_X(\omega_1,\omega_2,-\omega_2)d\omega_1 d\omega_2 + o(n^{-1}). \quad (5.3.2)$$

Lemma 5.3.2 gives a good grasp of the estimation of empirical spectral distribution. In fact, suppose

$$B_\lambda(\omega) = \begin{cases} 1 & -\pi \le \omega \le \lambda, \\ 0 & \text{otherwise,} \end{cases}$$

then $J^{(n)}(B_\lambda)$ is the empirical spectral distribution, that is,

$$J^{(n)}(\lambda) \equiv J^{(n)}(B_\lambda) = \frac{2\pi}{n}\sum_{0<2\pi t/n\le\lambda} I_{n,X}\Big(\frac{2\pi t}{n}\Big).$$

Since $B_\lambda(\omega)$ is obviously bounded variation, we have from (5.3.1) and (5.3.2), for $0 \le \lambda, \mu \le 0$,

$$EJ^{(n)}(B_\lambda) \to J(\lambda) \equiv \int_{\pi}^{\lambda} f(\omega)d\omega,$$

and

$$n\mathrm{Cov}(J^{(n)}(\lambda), J^{(n)}(\mu)) =$$
$$2\pi\Big(\int_0^{\min(\lambda,\mu)} f(\omega)^2 d\omega + \int_0^{\lambda}\int_0^{\mu}\tilde{Q}_X(\omega_1,\omega_2,-\omega_2)d\omega_1 d\omega_2\Big).$$

Another appealing result on the empirical spectral distribution is that it can be embedded in the space of càdlàg functions. Let $D_c[-\pi,0]$ denote the space of complex càdlàg functions. There is a metric that makes the space complete and separable. We suppose the space is equipped with the metric and then the topology follows.

Assumption 5.3.3. For any $\eta_3 > 0$ and for any integer $L \ge 0$, there exists $B_\rho > 0$ such that

$$E[T(n,s)^2 \mathbb{1}\{T(n,s) > B_{\eta_3}\}] < \eta_3$$

uniformly in n and s, where

$$T(n,s) = n^{-1/2}\sum_{r=0}^{L}\Big\{\sum_{t=1}^{n}\epsilon(t+s)\epsilon(t+s+r) - \sigma^2\delta(0,r)\Big\}^2$$

With the same definition, we have the following result:

Lemma 5.3.4 (Brillinger (2001), Theorem 7.6.3). *Suppose $\{X(t)\}$ is a real-valued stationary process satisfying Assumptions 5.3.1 and 5.3.3. The sequence of empirical spectral distributions $\sqrt{n}(J^{(n)}(\lambda) - J(\lambda))$ on $[-\pi, 0]$ converges in distribution to a Gaussian process $\{G(\lambda)\}$ on $[-\pi, 0]$ with mean 0 and covariance*

$$\mathrm{Cov}(G(\lambda), G(\mu)) = 2\pi\Big(\int_0^{\min(\lambda,\mu)} f(\omega)^2 d\omega + \int_0^{\lambda}\int_0^{\mu} \tilde{Q}_X(\omega_1, \omega_2, -\omega_2) d\omega_1 d\omega_2\Big)$$

for $-\pi \leq \lambda, \mu \leq 0$.

The result is powerful since asymptotic normality for the empirical spectral distribution $\sqrt{n}(J^{(n)}(\lambda) - J(\lambda))$ holds simultaneously, and naturally the distribution of any proposed test statistics for Gaussian process follows. If the process $\{X(t)\}$ is Gaussian, then the integration of the fourth order spectral is 0 and then the Gaussian process $G(\lambda)$ has independent increments

$$\mathrm{Cov}(G(\lambda_1) - G(\lambda_2), G(\mu_1) - G(\mu_2)) = 0$$

for $-\pi \leq \lambda_1 \leq \lambda_2 \leq \mu_1 \leq \mu_2 \leq 0$. Furthermore, the Gaussian process is exactly Brownian motion and the process is given by

$$\sqrt{2\pi}\int_0^{\lambda} f(\omega) dB(\omega).$$

As an application of the results in this section, we would like to consider the distribution of

$$T_n(\lambda) \equiv a_n \int_{\lambda}^{\lambda + a_n^{-1}} I_{n,X}(\omega) d\omega, \tag{5.3.3}$$

where $\{a_n\}$ is a monotone increasing process with $a_n \to \infty$. Note that if $a_n \to 0$, then the quantity is a form of smoothing the periodogram, so we eliminate the case. It is not difficult to expect the quantity can be approximated by $I_{n,X}(\lambda)$ without any influence by the factor a_n. The expectation, however, is wrong. We will investigate the case in a general way.

Divide $T_n(\lambda)$ by

$$a_n \int_0^{\lambda + a_n^{-1}} I_{n,X}(\omega) d\omega - a_n \int_0^{\lambda} I_{n,X}(\omega) d\omega.$$

The variance structure of two parts is given by

$$\begin{aligned}
&\mathrm{Var}\Big(a_n \int_0^{\lambda + a_n^{-1}} I_{n,X}(\omega) d\omega\Big) \\
&\quad = \frac{a_n^2}{n} 2\pi\Big(\int_0^{\lambda + a_n^{-1}} f(\omega)^2 d\omega + \int_0^{\lambda + a_n^{-1}}\int_0^{\lambda + a_n^{-1}} \tilde{Q}_X(\omega_1, \omega_2, -\omega_2) d\omega_1 d\omega_2\Big),
\end{aligned}$$

$$\operatorname{Var}\Big(a_n \int_0^\lambda I_{n,X}(\omega)d\omega\Big)$$
$$= \frac{a_n^2}{n} 2\pi \Big(\int_0^\lambda f(\omega)^2 d\omega + \int_0^\lambda \int_0^\lambda \tilde{Q}_X(\omega_1, \omega_2, -\omega_2) d\omega_1 d\omega_2 \Big),$$

and

$$\operatorname{Cov}\Big(a_n \int_0^{\lambda + a_n^{-1}} I_{n,X}(\omega)d\omega, a_n \int_0^\lambda I_{n,X}(\omega)d\omega\Big)$$
$$= \frac{a_n^2}{n} 2\pi \Big(\int_0^\lambda f(\omega)^2 d\omega + \int_0^\lambda \int_0^{\lambda + a_n^{-1}} \tilde{Q}_X(\omega_1, \omega_2, -\omega_2) d\omega_1 d\omega_2 \Big).$$

As a result, the variance of $T_n(\lambda)$ is given by

$$\begin{aligned} \operatorname{Var}(T_n(\lambda)) = \frac{a_n^2}{n} 2\pi \Big(&\int_\lambda^{\lambda + a_n^{-1}} \int_0^{\lambda + a_n^{-1}} \tilde{Q}_X(\omega_1, \omega_2, -\omega_2) d\omega_1 d\omega_2 \\ &+ \int_\lambda^{\lambda + a_n^{-1}} f(\omega)^2 d\omega - \int_0^\lambda \int_\lambda^{\lambda + a_n^{-1}} \tilde{Q}_X(\omega_1, \omega_2, -\omega_2) d\omega_1 d\omega_2 \Big). \end{aligned} \quad (5.3.4)$$

We can see the result from (5.3.4) by cases:

(i) if $a_n = n^\beta$ where $\beta < 1$, then the limiting variance of $T_n(\lambda)$ is

$$\operatorname{Var}(T_n(\lambda)) \to 0,$$

(ii) if $a_n = n^\beta$ where $\beta > 1$, then the limiting variance of $T_n(\lambda)$ is

$$\operatorname{Var}(T_n(\lambda)) \to \infty,$$

(iii) if $a_n = n^\beta$ where $\beta = 1$, then the limiting variance of $T_n(\lambda)$ is

$$\operatorname{Var}(T_n(\lambda)) \to 2\pi f(\lambda)^2.$$

We remove the case of $\beta > 1$ and the reason is given below the next theorem. The results are summarized as follows.

Theorem 5.3.5. *Suppose $\{X(t)\}$ is a real valued stationary process satisfying Assumption 5.3.1. Let $T_n(\lambda)$ be*

$$T_n(\lambda) = n^\beta \int_\lambda^{\lambda + n^{-\beta}} I_{n,X}(\omega) d\omega.$$

Then the asymptotic variance of $T_n(\lambda)$ is given by

$$\lim_{n\to\infty} \mathrm{Var}(T_n(\lambda)) = \begin{cases} 0, & \text{if } \beta < 1, \\ 2\pi f(\lambda)^2, & \text{if } \beta = 1. \end{cases}$$

In the case of i.i.d. random variables, the empirical distribution function can be approximated by the distribution function and the results do not depend on the order of factors. The result in Theorem 5.3.5 may seem surprising at first glance. However, we can explain the result in a heuristic way.

Returning back to the definition of $T_n(\lambda)$. The quantity

$$\int_{\lambda}^{\lambda+n^{-\beta}} I_{n,X}(\omega)d\omega$$

is actually a discrete statistic

$$\frac{2\pi}{n} \sum_{\lambda\le 2\pi s/n\le\lambda+n^{-\beta}} I_{n,X}(\frac{2\pi s}{n}).$$

See the number of periodograms $I_{n,X}(\lambda_s)$ with different frequencies, we can find that it depends on the order of $n^{-\beta}$. If $\beta < 1$, then more and more periodograms will be involved in the summation. Conversely, if $\beta > 1$, then the interval for the frequency will be smaller and smaller. The variance depends on the definition of the periodogram and the frequency is rational or not. Only the case $\beta = 1$ keeps the same order and therefore only one periodogram $I_{n,X}(\lambda)$ in the summation.

Thus in the case of $\beta < 1$, the case corresponds to the smoothing method of the periodogram for the process and asymptotic variance will be 0. On the other hand, if $\beta = 1$, approximation by $I_{n,X}(\lambda)$ is reasonable.

5.4 Estimation of Quantiles in Frequency Domain

Consider the samples $\{X(t)\}_{t=1}^n$ are generated by a symmetric real-valued process

$$X(t) = \sum_{j=0}^{\infty} G(j)\epsilon(t-j), \tag{5.4.1}$$

where the process $\{\epsilon(t)\}$ is a uncorrelated fourth-order stationary process, i.e.,

$$\begin{aligned} E\epsilon(t) &= 0, \\ E\epsilon(t)\epsilon(s) &= \sigma^2\delta(t,s), \end{aligned}$$

with $\sum_{t_1,t_2,t_3=-\infty}^{\infty}|Q_\epsilon(t_1,t_2,t_3)| < \infty$. The coefficients $G(j)$ are supposed to be

$$\sum_{j=0}^{\infty} G(j)^2 < \infty, \tag{5.4.2}$$

where $A(0) = 1$. Under the condition (5.4.2), the process $\{X(t)\}$ has a spectral density function

$$f_X(\omega) = \frac{\sigma^2}{2\pi} d(\omega) d(\omega)^*, \tag{5.4.3}$$

where $d(\omega) = \sum_{j=0}^{\infty} G(j) e^{i\omega j}$.

In this section, we afresh assumptions for simplicity of representation in the following results.

Assumption 5.4.1.

(i) For each nonnegative integer m and $\eta_1 > 0$,

$$\mathrm{Var}[E\{\epsilon(t)\epsilon(t+m)|\mathcal{B}(t-\tau)\} - \delta(m,0)\sigma^2] = O(\tau^{-2-\eta_1})$$

uniformly in t.

(ii) For $\eta_2 > 0$,

$$E|E\{\epsilon(t_1)\epsilon(t_2)\epsilon(t_3)\epsilon(t_4)|\mathcal{B}(t_1-\tau)\} - E\{\epsilon(t_1)\epsilon(t_2)\epsilon(t_3)\epsilon(t_4)\}| = O(\tau^{-1-\eta_2}),$$

uniformly in t_1, where $t_1 \le t_2 \le t_3 \le t_4$.

(iii) for some $\eta > 0$, the spectral density f satisfies

$$\sup_{|\lambda|<\epsilon} \int_{-\pi}^{\pi} \mathrm{tr}[\,\{\, f(\omega) - f(\omega-\lambda)\,\}\{\, f(\omega) - f(\omega-\lambda)\,\}^*\,]d\omega = O(\epsilon^\eta),$$

as $\epsilon \to 0$.

Based on observation stretch $(X(1),\dots,X(n))$, define the periodogram $I_{n,X}(\omega)$ by (5.2.4).

Lemma 5.4.2 (Hosoya (1989))**.** *Under Assumption 5.4.1, we have*

(i) *for any square-integrable Hermitian matrix-valued function $\phi(\omega)$,*

$$\int_{-\pi}^{\pi} \mathrm{tr}\{\,(I_{n,X}(\omega) - EI_{n,X}(\omega))\phi(\omega)\,\}d\omega \xrightarrow{\mathcal{P}} 0.$$

(ii) *for $\eta > 0$ defined in Assumption 5.4.1,*

$$\int_{-\pi}^{\pi} |EI_{n,X}(\omega) - f(\omega)|^2 d\omega = O(n^{-\eta}).$$

In this section, we are interested in the parameter λ, the ψth quantile of the spectral distribution $F(\omega)$, and derive the asymptotic properties of the estimation based on the quantile method. Suppose the parameter space is defined by $\Lambda = [-\pi, \pi]$ and λ is in the interior of Λ. Drawing the motivation in independent and identical distribution case in Koenker (2005), define $S(\theta)$ by

$$S(\theta) = \int_{-\pi}^{\pi} \rho_p(\omega - \mu) f(\omega) d\omega, \tag{5.4.4}$$

where

$$\rho_p(u) = u(p - \mathbb{1}(u < 0)).$$

For estimation, suppose the objective function $S_n(\theta)$ is given by

$$S_n(\theta) = \int_{-\pi}^{\pi} \rho_p(\omega - \theta) I_{n,X}(\omega) d\omega.$$

Hence, we can construct the estimator $\hat{\lambda}_n$ for λ by

$$\hat{\lambda}_n = \arg\min_{\theta \in \Lambda} S_n(\theta). \tag{5.4.5}$$

Theorem 5.4.3. *Suppose $\{X(t)\}_{t=1}^n$ are generated by (5.4.1) and the ψth quantile λ of the spectral distribution function of $\{X(t)\}$ is defined by (5.2.1). If Assumption 5.4.1 holds and $\hat{\lambda}_n$ is defined by (5.4.5), then we obtain*

$$\hat{\lambda}_n \to \lambda.$$

Proof. First, we confirm the uniqueness of λ. From the equivalent definition that λ minimizes (5.4.4), we first differentiate the estimating function (5.4.4) with respect to θ,

$$\frac{\partial}{\partial \theta} S(\theta) = F(\theta) - p\Sigma_X.$$

There exists only one zero of $\frac{\partial}{\partial \theta} S(\theta) = 0$ from the monotonicity of the spectral distribution $F(\theta)$, which the uniqueness of λ follows. Let m be the minimum of $S(\theta)$. Next, the convexity of $S_n(\theta)$ is shown by the positiveness of the second derivative of $S_n(\theta)$, i.e.,

$$\frac{\partial^2}{\partial \theta^2} S_n(\theta) = I_{n,X}(\theta) > 0 \quad a.s.$$

Also, the pointwise convergence of $S_n(\theta)$ is obtained from the definition of $S(\theta)$, since for each $\theta \in \Lambda$,

$$|S_n(\theta) - S(\theta)| \leq \Big| \int_{-\pi}^{\pi} \rho_p(\omega - \theta)(I_{n,X}(\omega) - EI_{n,X}(\omega)) d\omega \Big|$$

$$+\left|\int_{-\pi}^{\pi}\rho_p(\omega-\theta)(EI_{n,X}(\omega)-f(\omega))d\omega\right|,$$

where the right hand side converges to 0 in probability by Lemma 5.4.2.

By the Convexity Lemma in Pollard (1991),

$$\sup_{\lambda\in K}|S_n(\theta)-S(\theta)| \xrightarrow{\mathcal{P}} 0, \tag{5.4.6}$$

for any compact subset $K\subset\Lambda$.

Let $B(\lambda)$ be any open neighborhood of λ. From the uniqueness of zero of $S(\theta)$, there exists an $\epsilon>0$ such that $\inf_{\mu\in\Lambda/B(\lambda)}|S(\mu)|>m+\epsilon$. Thus, with probability tending to 1,

$$\inf_{\mu\in\Lambda/B(\lambda)} S_n(\mu) \geq \inf_{\mu\in\Lambda/B(\lambda)} S(\mu) - \sup_{\mu\in\Lambda/B(\lambda)}|S(\mu)-S_n(\mu)| > m,$$

where it is implied by (5.4.6) that the second term can be chosen arbitrarily small. The conclusion follows that with probability tending to 1, $S_n(\hat{\lambda}_n)\leq m-\epsilon^*$ by the pointwise convergence of $S_n(\lambda)$ in probability. □

To investigate the asymptotic distribution of the estimator $\hat{\lambda}_n$, we follow the argument of the empirical process. First, we review the general result. If the objective function is defined in the way of empirical process, i.e.,

$$S_n(\delta)=\frac{a_n}{\sqrt{n}}\sum_{i=1}^{n}\rho(\theta-a_n^{-1}\delta),$$

where ρ is a convex function, which is minimized by the minimizer $\hat{\delta}_n=a_n(\hat{\theta}_n-\theta)$ of the process $S_n(\delta)$, then the argument to derive the asymptotic distribution of the estimator is mainly based on the result that if $S_n(\delta)\xrightarrow{\mathcal{L}}S(\delta)$, then

$$\hat{\delta}_n=a_n(\hat{\theta}-\theta)\xrightarrow{\mathcal{L}}\delta\equiv\arg\min S(\delta),$$

which is studied in Hjort and Pollard (1993) and Geyer (1996). We need a Lindeberg-type condition to guarantee the asymptotic normality for the convergence in distribution of the process.

Assumption 5.4.4 (Lindeberg type condition)**.** For any $\eta_3>0$ and for any integer $L\geq 0$, there exists $B_{\eta_3}>0$ such that

$$E[T(n,s)^2\mathbb{1}\{T(n,s)>B_{\eta_3}\}]<\eta_3$$

uniformly in n and s, where

$$T(n,s) = n^{-1/2} \sum_{r=0}^{L} \Big\{ \sum_{t=1}^{n} \epsilon(t+s)\epsilon(t+s+r) - \sigma^2 \delta(0,r) \Big\}^2.$$

The problem with the asymptotics of the estimator is its convergence order since the behavior of (5.3.3) depends heavily on the choice of the order a_n. The answer of this question is $a_n = \sqrt{n}$. We will give the proof below the following theorem, and other orders can be found not adequate following the same argument in the proof.

Theorem 5.4.5. *Suppose $\{X(t)\}_{t=1}^{n}$ are generated by (5.4.1) and the ψth quantile λ of the spectral distribution function of $\{X(t)\}$ is defined by (5.2.1). If Assumptions 5.4.1 and 5.4.4 hold, and $\hat{\lambda}_n$ is defined by (5.2.3), then for $-\pi \le \lambda \le 0$,*

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \to_d \mathcal{N}(0, \Sigma),$$

where the asymptotic variance is given by

$$\Sigma = \mathscr{E}^{-2}\sigma^2,$$

with $\mathscr{E}$, an exponential random variable with mean $f(\lambda)$ and

$$\begin{aligned} \sigma^2 = 4\pi p^2 \int_{-\pi}^{\pi} f(\omega)^2 d\omega + 2\pi(1-4p) \int_{-\pi}^{\lambda} f(\omega)^2 d\omega \\ + 2\pi \iint_{-\pi}^{\pi} (p - \mathbb{1}(\omega_1 < \lambda))(p - \mathbb{1}(\omega_2 < \lambda)) Q_X(\omega_1, -\omega_2, \omega_2) d\omega_1 d\omega_2. \end{aligned}$$

We remark that the asymptotic distribution is normal given a exponential distributed random variable independent of the normal distribution. This is not expected only from the quantile estimation in time domain.

Proof. The proof depends on the theory of empirical process (see van der Vaart and Wellner (1996)), where we only have to show the discretized case. First, consider the process

$$M_n(\delta) = n\big\{S_n(\lambda + \frac{\delta}{\sqrt{n}}) - S_n(\lambda)\big\},$$

which is minimized by $\sqrt{n}(\hat{\lambda}_n - \lambda)$ from the definition. To treat $\rho_p(u)$ explicitly in the proof, divide $\rho_p(u)$ by Knight's identity (see Knight (1998)) as

$$M_n(\delta) \quad = \quad -\delta\sqrt{n} \left\{ \int_{-\pi}^{\pi} (p - \mathbb{1}(\omega < \lambda))(I_{n,X}(\omega) - f(\omega)) d\omega \right\}$$

$$
\begin{aligned}
&+\int_{-\pi}^{\pi}\int_{0}^{\delta/\sqrt{n}} n\left(\mathbb{1}(\omega \le \lambda + s) - \mathbb{1}(\omega \le \lambda)I_{n,X}(\omega)dsd\omega\right.\\
=\;& M_{n1}(\delta) + M_{n2}(\delta), \quad \text{(say)}.
\end{aligned}
$$

By the central limit theorem in the frequency domain,

$$M_{n1}(\delta) \xrightarrow{\mathcal{L}} -\delta\mathcal{N}(0, \sigma^2),$$

where

$$
\begin{aligned}
\sigma^2 = \pi p^2 \int_{-\pi}^{\pi} f(\omega)^2 d\omega &+ 2\pi(1-4p)\int_{-\pi}^{\lambda} f(\omega)^2 d\omega \\
&+ 2\pi \iint_{-\pi}^{\pi} (p - \mathbb{1}(\omega_1 < \lambda))(p - \mathbb{1}(\omega_2 < \lambda))Q_X(\omega_1, -\omega_2, \omega_2)d\omega_1 d\omega_2.
\end{aligned}
$$

From Theorem 5.3.5, in view of

$$n\int_{\lambda}^{\lambda+n^{-1}} I_{n,X}(\omega)d\omega = I_{n,X}(\lambda), \tag{5.4.7}$$

we will use this order for evaluating $M_{n2}(\delta)$. The second term $M_{n2}(\delta)$ can be evaluated by

$$
\begin{aligned}
M_{n2}(\delta) &= \int_{0}^{\delta/\sqrt{n}}\int_{\lambda}^{\lambda+s} n\, I_{n,X}(\omega)d\omega ds \\
&= \frac{1}{n}\int_{0}^{\delta\sqrt{n}}\left(n\int_{\lambda}^{\lambda+t/n} I_{n,X}(\omega)d\omega\right)dt \\
&= \frac{1}{n}\int_{0}^{\delta\sqrt{n}} tI_{n,X}(\lambda)dt \\
&= \frac{1}{2}I_{n,X}(\lambda)\delta^2.
\end{aligned}
$$

This term actually does not converge in probability, but has an asymptotic exponential distribution which has mean $f(\lambda)$ $(= \mathscr{E}$, say$)$. If we show that

$$\begin{pmatrix} M_{n1}(\delta) \\ M_{n2}(\delta) \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} \mathcal{N} \\ \mathscr{E} \end{pmatrix},$$

then by continuous mapping theorem, we obtain

$$M_n(\delta) \xrightarrow{\mathcal{L}} M(\delta) = -\delta\mathcal{N} + \frac{1}{2}\delta\mathscr{E}^2,$$

which is minimized by $\delta = \mathscr{E}^{-1}\mathcal{N}$. As a consequence,

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \to_d \mathcal{N}(0, U).$$

□

From here, we show the convergence in distribution of the joint random vectors

$$\begin{pmatrix} M_{n1}(\delta) \\ M_{n2}(\delta) \end{pmatrix}.$$

Denote $C(m) = \sum_{s=1}^{n-m} X(s)X(s+m)$ as the sample covariance.

Lemma 5.4.6 (Hosoya and Taniguchi (1982), Lemma 2.2)**.** *Under Assumptions 5.4.1 and 5.4.4, Then $n^{-1/2}(C(m)-\gamma(m))$ $(m = 1, \dots, l)$ have a joint asymptotic normal distribution with covariance V, where $V_{m_1m_2}$ is given by*

$$\begin{aligned} V_{m_1m_2} = 2\pi \int_{-\pi}^{\pi} & f^2(\omega)\{\exp(-i(m_2 - m_1)\omega) + \exp(i(m_2 + m_1)\omega)\}d\omega \\ &+ (2\pi)^{-2}\kappa_4 \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \exp(im_1\omega_1 + im_2\omega_2)Q(\omega_1, -\omega_2, \omega_2)d\omega_1 d\omega_2, \end{aligned}$$

where $Q(\omega_1, -\omega_2, \omega_2)$ is the fourth order spectral density of the process.

For next result, we have to extend the domain of periodogram on the lattice as in Brockwell and Davis (1991). That is to say, for any $\omega \in [-\pi, \pi]$, define the periodogram $I_{n,X}(\omega)$ discretely by $I_{n,X}(\omega_k)$, where ω_k is defined as the closest frequency of the multiple of $2\pi/n$. It is easy to see that

$$|I_{n,X}(\omega) - I_{n,X}(\omega_k)| = o_p(1).$$

Note that the general result is studied in Chapter 3. For asymptotic distribution of $I_{n,X}(\omega_k)$, we only have to apply Theorem 3.4.3 to the second order stationary process $\{X(t); t \in \mathbb{Z}\}$.

Lemma 5.4.7 (Brockwell and Davis (1991))**.** *If $-\pi < \omega_k < 0$ then the random vector*

$$n^{-1/2}\Big(\sum_{t=1}^{n} X(t)\cos(\omega_k t), \sum_{t=1}^{n} X(t)\sin(\omega_k t)\Big)'$$

has a joint asymptotic normal distribution with the covariance matrix $1/2\,\sigma I_2$.

Combing these two lemmas, we have the following result.

Lemma 5.4.8. *Under Assumptions 5.4.1 and 5.4.4, the asymptotic joint distribution of the sample covariances and the trigonometric transform of samples is* *given by*

$$n^{-1/2}\begin{pmatrix} C(1)-\gamma(1) \\ \vdots \\ C(l)-\gamma(l) \\ \sum_{t=1}^{n} X(t)\cos(\lambda t) \\ \sum_{t=1}^{n} X(t)\sin(\lambda t) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \begin{pmatrix} V & 0 & 0 \\ 0 & \frac{1}{2}\sigma^2 & 0 \\ 0 & 0 & \frac{1}{2}\sigma^2 \end{pmatrix}). \tag{5.4.8}$$

Proof. The statement will be shown by Cramér-Wold device. Suppose $t = (t_1, \dots, t_{L+2})$ and denote the left hand side of (5.4.8) by S. This time,

$$n^{-1}\sum_{t=1}^{n} E((t^T S)^2 \mathbb{1}(|t^T S| > n^{1/2}\epsilon)) \to 0,$$

by the symmetricity of the process. Also, since the process is symmetric, for any $1 \le m \le L$,

$$\begin{aligned} \mathrm{Cov}(n^{-1/2}C(m), n^{-1/2}\sum_{t=1}^{n} X(t)\cos(\lambda t)) \\ = n^{-1}\sum_{s=1}^{n-m}\sum_{t=1}^{n} \mathrm{cum}(X(s), X(s+m), X(t)) = 0, \end{aligned}$$

By Lindeberg's central limit theorem, the conclusion holds. □

5.5 Improvement of Quantile Estimation

As what we have seen in the previous section, the asymptotic distribution is peculiar while it acts like a sandwich form. Simultaneously, it is obviously possible to improve the asymptotic distribution of the estimator. The characteristic property of the distribution of (5.4.7) can be made converge in probability by the method of smoothing or sometimes called tapering. Instead of bare periodogram, we propose smoothed periodogram defined by

$$\hat{f}(\lambda) = \frac{1}{2\pi}\sum_{|k|\le m_n} A_n(k) I_{n,X}(\omega_{j+k}). \tag{5.5.1}$$

Assumptions on $A_n(k)$ are given as follows.

Assumption 5.5.1. Let $A_n(k)$ satisfy

(i) $m \to \infty$ and $m/n \to 0$ as $n \to \infty$.

(ii) For all k, $A_n(k) = A_n(-k)$ and $A_n(k) \ge 0$.

(iii) $\sum_{|k|\le m_n} A_n(k) = 1$.

(iv) $\sum_{|k|\le m_n} A_n(k)^2 \to 0$ as $n \to \infty$.

Lemma 5.5.2 (Brockwell and Davis (1991), Theorem 10.4.1)**.** *Suppose* $\sum_{t=-\infty}^{\infty} |j|^{1/2}|G(j)| < \infty$*. If* $\hat{f}(\lambda)$ *is defined by* (5.5.1) *under Assumption 5.5.1, we obtain for* $\lambda, \omega \in [0, \pi]$,

(i) $\lim_{n\to\infty} \hat{f}(\lambda) \to f(\lambda)$,

(ii) *the covariance is given by*

$$\lim_{n\to\infty}\Big(\sum_{|k|\le m} A_n(k)^2\Big)^{-1}\mathrm{Cov}(\hat{f}(\lambda), \hat{f}(\omega)) = \begin{cases} 2f(\lambda)^2 & \textit{if } \lambda = \omega = 0 \textit{ or } \pi, \\ f(\lambda)^2 & \textit{if } 0 < \lambda = \omega < \pi, \\ 0 & \textit{if } \lambda \neq \omega. \end{cases}$$

Lemma 5.5.2 shows the consistency of the smoothed periodogram. The discrete form (5.5.1) can be rewritten in a integration form by

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \phi(\omega) I_{n,X}(\omega) d\omega,$$

for some symmetric continuous function $\phi(\omega)$. Then define the estimator $\hat{\lambda}_n^*$ by

$$\hat{\lambda}_n^* = \arg\min_{\theta\in\Lambda} \int_{-\pi}^{\pi} \rho_p(\omega - \theta)\phi(\omega) I_{n,X}(\omega) d\omega. \tag{5.5.2}$$

The result in the previous section then can be improved by the following result.

Theorem 5.5.3. *Suppose* $\{X(t)\}_{t=1}^n$ *are generated by* (5.4.1) *and the* ψ*th quantile* λ *of the spectral distribution function of* $\{X(t)\}$ *is defined by* (5.2.1)*. If Assumptions 5.4.1 and 5.4.4 hold, and* $\hat{\lambda}_n^*$ *is defined by* (5.5.2)*, then for* $-\pi \le \lambda \le 0$,

$$\sqrt{n}(\hat{\lambda}_n^* - \lambda) \to_d \mathcal{N}(0, \Sigma),$$

where

$$\Sigma = f(\lambda)^{-2}\sigma^2,$$

and

$$\begin{aligned}\sigma^2 = \pi p^2 \int_{-\pi}^{\pi} \phi(\omega)^2 f(\omega)^2 d\omega + 2\pi(1-4p)\int_{-\pi}^{\lambda} \phi(\omega)^2 f(\omega)^2 d\omega \\ + 2\pi \iint_{-\pi}^{\pi} (p - \mathbb{1}(\omega_1 < \lambda))(p - \mathbb{1}(\omega_2 > -\lambda)) \\ \times \phi(\omega_1)\phi(\omega_2) Q_X(\omega_1, \omega_2, -\omega_2) d\omega_1 d\omega_2.\end{aligned}$$

5.6 Numerical Results

In this section, we give the numerical results of the estimated quantiles of the spectral distribution we mentioned in Section 5.2. The estimator $\hat{\lambda}_n$ is given by (5.2.3), and the size of samples generated from the process is supposed to be 30.

5.6.1 stationary case

First, we give numerical results on the second order stationary process given in Section 5.2, that is white Gaussian noise model, MA(1) Gaussian process with coefficient 0.9, AR(1) Gaussian process with coefficient 0.9 and AR(1) Gaussian process with coefficient -0.9. The structures of these four models are obviously different. Suppose we want to specify the structure of the model by the method of estimating the frequencies.

TABLE 5.1: Estimated quantiles $\hat{\lambda}_n$ of the spectral distribution.

p	White noise	MA(1)	AR(1) with 0.9	AR(1) with -0.9
0.5	0.000	0.000	0.000	0.000
0.6	0.305	0.211	0.026	2.940
0.7	1.187	0.576	0.055	3.030
0.8	1.564	0.891	0.092	3.074
0.9	2.093	1.235	0.190	3.109
1.0	3.142	3.142	3.142	3.142

It is easy to see that the results are corresponding to the figure of the distribution function in Section 5.2. However, the result is not so desirable since the estimated frequencies are not equidistant in the white noise.

As the results in the general asymptotic theory, if the estimator is asymptotically normal, then the estimate result will be improved by the increasing number of samples. However, as what we have shown in Section 5.4, the estimator (5.2.3) is based on the bare periodogram and therefore is not asymptotically normal. We next give the results in the white noise case with different numbers of samples to see the phenomenon.

TABLE 5.2: Estimated quantiles $\hat{\lambda}_n$ in the case of white noise with different numbers of samples.

$p\backslash\ n$	30	50	100	200
0.5	0.000	0.000	0.000	0.000
0.6	0.305	0.663	0.366	0.745
0.7	1.187	1.226	0.966	1.260
0.8	1.564	1.990	1.602	1.881
0.9	2.093	2.440	2.251	2.334
1.0	3.142	3.142	3.142	3.142

From Table 2, we can see the accuracy is not obviously improved by the increase of the number of samples. This result supports the asymptotic results given in Theorem 5.4.5 in Section 5.4.

5.6.2 nonlinear time series model

At last, we want to search the frequencies in the nonlinear time series models given in the historical consideration. With the same settings of $X(t)$ given in 5.6.1, we add a harmonic component m_t in the model with $\omega_0 = \pi/2$, i.e.

$$Y_t = 1/2\cos(\omega_0\, t) + \sin(\omega_0\, t) + X(t), \quad m_t = 1/2\cos(\omega_0\, t) + \sin(\omega_0\, t).$$

As already known, the periodogram has a large change at the certain frequency $\omega_0 = \pi/2$, which we add in the harmonic component in this example. Compared with the example in Section 5.6.1, we can see that the estimated quintiles are pulled around to the certain frequency ω_0.

TABLE 5.3: Estimated quantiles $\hat{\lambda}_n$ of the spectral distribution.

p	White noise	MA(1)	AR(1) with 0.9	AR(1) with -0.9
0.5	0.000	0.000	0.000	0.000
0.6	1.399	0.412	0.030	2.610
0.7	1.513	0.789	0.065	3.014
0.8	1.577	1.254	0.116	3.066
0.9	1.679	1.582	1.152	3.106
1.0	3.142	3.142	3.142	3.142

If we have a purpose to estimate the certain frequency supposed in the nonlinear time series model, then we have to narrow the span in the ratio p and find the greatest increase in the frequency. We will skip the discussion on the method since it is beyond the scope of this doctoral thesis.

Chapter 6

Empirical Likelihood Method for Time Series with Infinite Variance

6.1 Introduction

Recently, the multivariate data with infinite variance appear in various fields like finance, economics and hydrology. To model these phenomena, one choice is to apply generalized linear process with stable innovations to the case.

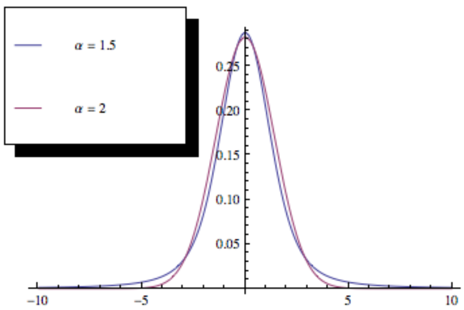


FIGURE 6.1: Stable distribution ($\alpha = 1.5$) and normal distribution ($\alpha = 2$).

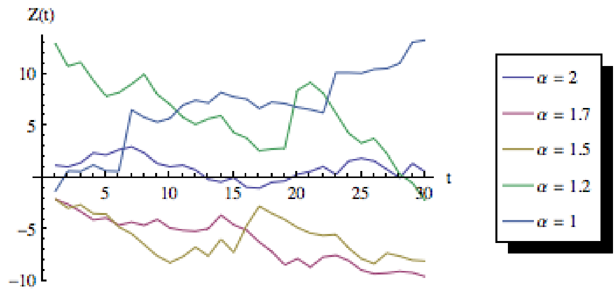


FIGURE 6.2: AR(1) model driven by Stable distribution ($\alpha = 2, 1.7, 1.5, 1.2, 1$).

Figure 6.1 shows the probability density of both stable distribution and normal distribution ($\alpha = 1.5$ and $\alpha = 2$ with $\sigma = 1$). Figure 6.2 shows a linear process driven by the stable innovation ($\alpha = 2, 1.7, 1.5, 1.2, 1$) with $\sigma = 1$. Although the difference between the probability density functions is not seemingly so much, it is easy to see that the linear process with smaller index waves more dynamically in its range. In fact, only the case of $\alpha = 2$ has finite variance, and the others do not. Accordingly, a process under regular conditions with finite moments is much more different from the stable one.

As for stable random variables, Feller (1971) gives an overview of them from several different points of view. The pioneer works on the dependent case, especially for 1-dimensional linear process, are given by (a) Davis and Resnick (1985), (b) Davis and Resnick (1985) and (c) Davis and Resnick (1986). They studied the asymptotic distribution of the partial sum of the product of two stable random variables, and then the asymptotic distribution of the sample autocorrelation function (ACF). Applying the result on the ACF, Klüppelberg and Mikosch (1993, 1994) and Klüppelberg and Mikosch (1996) proposed the self-normalized periodogram in frequency domain for estimation and hypothesis testing of the process with infinite variance. They gave some asymptotic properties of the self-normalized periodogram and proved the convergence of the integral functional containing it.

As we consider a linear time series model, not only the infinite variance model, it is always found that the exact likelihood of the samples cannot be specified. For second order stationary process, we usually suppose the model is Gaussian and apply the Whittle likelihood since the likelihood corresponds to the quasi-maximum likelihood under the assumption of Gaussian. The assumption, however, is too strong. As in Chapter 4, the Whittle likelihood is shown to be the most efficient statistic in the many classes of integration functional type statistics and therefore is recommendable. To deal with the problem of the unknown likelihood, the nonparametric methods are always considered in statistics. In the context of time series analysis, rank based statistics are considered by Hallin et al. (1985, 1987) and Hallin and Puri (1991). Beyond the serial sample covariances, they focused on the linear serial rank statistics and compared the asymptotic efficiency in the class. On the other hand, asymptotic properties of empirical likelihood method for i.i.d. samples have been developed in successive works by Owen (1990, 1988). The empirical likelihood statistic has attraction for statistician since it is bartlett correctable, which is shown in DiCiccio et al. (1991), while its competitive method "bootstrap" is not. The method then is generalized to the dependent case by Kitamura et al. (1997) with clockwise empirical likelihood in time domain and by Monti (1997) with quasi-likelihood in frequency domain, whose idea is originated from Monti and Ronchetti (1993). Ogata and Taniguchi (2010) also employed the empirical likelihood method to construct confidence region for multivariate linear processes when innovations have finite moments.

In this chapter, we apply the empirical likelihood ratio statistic to the linear time series model with infinite variance and give asymptotic distribution of the statistic under the hypothesis. The estimating function contained in the empirical likelihood ratio statistic is based on the Whittle likelihood, which is shown to be most efficient in finite variance case in Chapter 4. In Section 6.2, we give the definitions of vector α-stable process and the empirical likelihood ratio statistic and assumptions in this chapter. In Section 6.3, we look into the asymptotic properties of the functional of the periodogram of the considered process and derive the asymptotic distribution of the empirical likelihood ratio statistic under the hypothesis based on the previous results. The procedure to construct the confidence interval for the pivotal quantity contained in the empirical likelihood ratio statistic and the numerical results are given in Section 6.4. We show Theorems 6.3.1 and 6.3.2 in Section 6.5 with some remarks given in Section 6.6.

6.2 Vector α-stable Processes and Preliminaries

We consider a d-dimensional vector-valued linear process $\{\boldsymbol{X}(t); t \in \mathbb{Z}\}$ generated by

$$\boldsymbol{X}(t) = \sum_{j=0}^{\infty} \Psi(j)\boldsymbol{Z}(t-j), \tag{6.2.1}$$

where $\Psi(j)$ are $d \times d$ real-valued matrices, and $\{\boldsymbol{Z}(t)\}$ is an independently and identically distributed sequence of α-stable random vectors with symmetric independent elements. For linear model (6.2.1), we define the true power transfer function $\boldsymbol{g}(\omega)$ by

$$\boldsymbol{g}(\omega) = \Psi(\omega)\Psi(\omega)^*,$$

where $\Psi(\omega) = \sum_{j=0}^{\infty} \Psi(j)e^{ij\omega}$.

First, we give a brief review on the notations in this chapter. we use Bold letters to represent vectors or matrices. For an element in vectors or matrices, we use underscript to represent it. For instance, A_j denotes the jth entry in the vector $\boldsymbol{A}$, where A_{ij} denotes the element lying in the ith row and jth column of the matrix $\boldsymbol{A}$. The argument about the frequency domain makes us unable to escape from the complex numbers. We generally use $\bar{A}$ to denote the complex conjugate of A regardless of whether A is a complex number or a complex matrix.

Next, characters are defined as follows. Note that we use $\omega \in [-\pi, \pi]$ for continuous case, and $\lambda_t = \frac{2\pi t}{n} \in (-\pi, \pi]$ for discrete case. For any random vector $\boldsymbol{A}$, the sample autocovariance and the periodogram matrices are defined by

$$\hat{\Gamma}_{n,A}(h) = n^{-2/\alpha} \sum_{t=1}^{n-|h|} \boldsymbol{A}(t)\boldsymbol{A}(t+h)^T,$$

$$\boldsymbol{I}_{n,A}(\omega) = d_{n,A}(\omega) d_{n,A}(\omega)^*, \quad d_{n,A}(\omega) = n^{-1/\alpha} \sum_{t=1}^{n} \boldsymbol{A}(t) e^{i\omega t}.$$

Also, for the precise convergence order in the linear stable process case, suppose

$$\begin{aligned} x_n &= \left(\frac{n}{\log n}\right)^{1/\alpha}, \\ y_n &= \Big(n \log n\Big)^{1/\alpha}. \end{aligned}$$

There are two norms used in this paper. One is the Euclidean norm, which is denoted by $\| \boldsymbol{A} \|_E = \sqrt{\operatorname{tr}(\boldsymbol{A}^* \boldsymbol{A})}$. Secondly, based on the observed stretch $\{\boldsymbol{X}(t), 1 \le t \le n\}$, the self-normalized term, denoted by $\|Z\|_N$, is defined as follows:

$$\|Z\|_N \equiv \sqrt{\sum_{t=1}^{n} \sum_{i=1}^{d} Z(t)_i^2}.$$

It is well known that $Z(t)_i^2$ is in the domain of attraction of a stable limit with the exponent $\alpha/2$, and the linear transformation of stable distribution with nonrandom scale is also stable with the same characteristic exponent. Thus the sum $\sum_{i=1}^{d} Z(t)_i^2$ is also in the domain of attraction of a stable limit with the exponent $\alpha/2$. The normalized form of vectors is written by

$$\tilde{Z}(t)_i = \frac{Z(t)_i}{\|Z\|_N}, \quad i = 1, \ldots, d.$$

Suppose we have to test H: $\boldsymbol{\theta} = \boldsymbol{\theta}^0$, where $\boldsymbol{\theta}^0$ is defined by the solution of the estimating function

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\pi}^{\pi} \operatorname{tr}\left[\, \{\boldsymbol{f}(\omega; \boldsymbol{\theta})\}^{-1} \boldsymbol{g}(\omega) \,\right] d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \boldsymbol{0}. \tag{6.2.2}$$

The condition of equation (6.2.2) is not restrictive since we can use the constrained family, for example, to derive the autocorrelation, the interpolation or the prediction of the process (6.2.1).

The empirical likelihood ratio function introduced for the problem of testing H: $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ is defined by

$$R(\boldsymbol{\theta}) = \max_{\omega_1,\ldots,\omega_n} \left\{ \prod_{t=1}^{n} n\omega_t \, ; \, \sum_{t=1}^{n} \omega_t \, \boldsymbol{m}(\lambda_t; \boldsymbol{\theta}) = \mathbf{0}, \sum_{t=1}^{n} \omega_t = 1, 0 \le \omega_t \le 1 \ , \ \forall t \right\}, \tag{6.2.3}$$

which is based on a likelihood ratio between $\prod_{t=1}^{n} \omega_t$ and $\prod_{t=1}^{n} n^{-1}$. Here, $\boldsymbol{m}(\lambda_t; \boldsymbol{\theta})$ is an estimating function defined to correspond to the pivotal quantity $\boldsymbol{\theta}^0$. Note that we have shown the distribution of $I_{n,X}(\lambda_t)$, and therefore the distribution of $\boldsymbol{m}(\lambda_t; \boldsymbol{\theta})$, is asymptotically independent identically distributed from the results in Chapter 3. Thus the empirical likelihood ratio statistic $R(\boldsymbol{\theta})$ is a well-defined ratio statistic if we use the discretized Whittle likelihood

$$\boldsymbol{m}(\lambda_t; \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{tr}\{\boldsymbol{f}(\lambda_t; \boldsymbol{\theta})^{-1} \boldsymbol{I}_{n,X}(\lambda_t)\}. \tag{6.2.4}$$

For the brevity in the following section, we define moment functions of the estimating function $\boldsymbol{P}_n(\boldsymbol{\theta})$ and $\boldsymbol{S}_n(\boldsymbol{\theta})$ as follows:

$$\begin{aligned} \boldsymbol{P}_n(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{m}(\lambda_t; \boldsymbol{\theta}) \\ \boldsymbol{S}_n(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{m}(\lambda_t; \boldsymbol{\theta}) \boldsymbol{m}(\lambda_t; \boldsymbol{\theta})'. \end{aligned}$$

Assumptions through this chapter are given below:

Assumption 6.2.1. Assume that $\boldsymbol{X}(t)$ is generated by (6.2.1) where for $0 < \delta < 1$,

$$\sum_{j=0}^{\infty} j |\Psi(j)_{kl}|^{\delta} < \infty, \quad \text{for } k, l = 1, 2, \ldots, d. \tag{6.2.5}$$

Remark 6.2.2. Under this assumption, (6.2.1) is well-defined. See Brockwell and Davis (1991) or Petrov (1975).

Define the family $\mathcal{F}(\Theta)$ of the parametrized power transfer function by

$$\mathcal{F}(\Theta) = \left\{ \boldsymbol{f}(\omega; \boldsymbol{\theta}) \middle| \boldsymbol{f}(\omega; \boldsymbol{\theta}) = \left(\sum_{j=0}^{\infty} \Xi(j; \boldsymbol{\theta}) e^{ij\omega} \right) \left(\sum_{j=0}^{\infty} \Xi(j; \boldsymbol{\theta}) e^{ij\omega} \right)^{*}, \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p \right\}.$$

Assumption 6.2.3.

(i) Θ is a compact subset of $\mathbb{R}^p$ and $\boldsymbol{f}(\omega; \boldsymbol{\theta})$ has an parametrized representation as an element of $\mathcal{F}(\Theta)$.

(ii) For any $\boldsymbol{\theta} \in \mathrm{Int}(\Theta)$, $\boldsymbol{f}(\omega; \boldsymbol{\theta}) \in \mathcal{F}(\Theta)$ is continuously twice differentiable with respect to $\boldsymbol{\theta}$.

(iii) There exists a unique $\boldsymbol{\theta}^0 \in \Theta$ satisfying (6.2.2).

The assumption below guarantees the convergence of the functional of periodogram by inequality of an application of Theorem 3.1 in Rosinski and Woyczynski (1987).

Assumption 6.2.4. For some $\mu \in (0, \alpha)$ and all $k = 1, \cdots, p$,

$$\sum_{t=1}^{\infty} \left\| \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_k} \Psi(\omega)^* \boldsymbol{f}(\omega; \boldsymbol{\theta}) \Psi(\omega) e^{it\omega} d\omega \right\|_E^{\mu} < \infty.$$

6.3 Asymptotic Distribution of Empirical Likelihood Ratio

First the limit of functional form of periodogram is shown in the following theorem, which is a generalization of 1-dimensional result. Note that the discretized form of (6.2.4) only holds when $\alpha \geq 1$. Accordingly, we assume $1 \leq \alpha < 2$ for the empirical likelihood ratio statistic in this chapter.

Theorem 6.3.1. *Let $(\boldsymbol{X}(t))_{t \in \mathbb{Z}}$ be a linear process as defined in (6.2.1) with coefficient matrices $(\Psi(j))_{j \in \mathbb{Z}}$ satisfying (6.2.5) and suppose that $\alpha \in (0, 2)$. Furthermore, let $\phi_k(\omega), j = 1, \ldots, d$, be $d \times d$ matrix-valued 2π-periodic continuous function with $\phi_k(\omega) = \phi_k(\omega)^*$ such that the Fourier coefficients of $\Psi(\cdot)\phi_k(\cdot)\Psi(\cdot)^*$ are absolutely summable and*

$$\sum_{t=1}^{\infty} \left\| \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_k} \Psi(\omega)^* \phi_k(\omega) \Psi(\omega) e^{it\omega} d\omega \right\|_E^{\mu} < \infty$$

for some $\mu \in (0, \alpha)$ and all $k = 1, \cdots, p$. Then

$$(n^{-2/\alpha} \|Z\|_N^2, x_n \int_{-\pi}^{\pi} \mathrm{tr} \left[\{ \boldsymbol{I}_{n,X}(\omega) - \Psi(\omega) \hat{\Gamma}_{n,Z}(0) \Psi(\omega)^* \} \phi_k(\omega) \right] d\omega)$$
$$\xrightarrow{\mathcal{L}} (\, S_{\alpha/2}, \sum_{i,j=1}^{d} \sum_{h=1}^{\infty} S(h)_{ij} \int_{-\pi}^{\pi} (A(\omega) + \overline{A(\omega)})_{ij} \, d\omega \,),$$

where

$$A(\omega) = \Psi(\omega)^* \phi_k(\omega) \Psi(\omega) e^{ih\omega},$$

and $S(h)_{ij}$ is the (i, j)-component of the limit stable random matrix $S(h)$, where

$$x_n \hat{\Gamma}_{n,Z}(h) \Rightarrow S(h) \quad \text{for } h = 1, 2, \ldots.$$

The scale parameter of $S_{\alpha/2}$ is given by $(\sigma^\alpha K_\alpha \cos(\pi\alpha/4))^{2/\alpha}$, where $K_\alpha = E|N|^\alpha$ for an normal random variable $\mathcal{N}(0, 2)$, while the scale parameters of *the elements in* $S(h)$ *are given by* C_α, *i.e.,*

$$C_\alpha = \begin{cases} \frac{2(1-\alpha)\sigma}{\Gamma(2-\alpha)cos(\pi\alpha/2)} & \text{if } \alpha \neq 1, \\ \frac{4}{\pi}, & \text{if } \alpha = 1. \end{cases}$$

Theorem 6.3.1 gives a joint distribution of the random variables used for the self-normalized periodogram. From the result, we can then derive asymptotic distribution of the logarithm of the empirical likelihood ratio statistic $R(\boldsymbol{\theta}^0)$ under the hypothesis H: $\boldsymbol{\theta} = \boldsymbol{\theta}^0$. As mentioned in the beginning of the section, only $\alpha \geq 1$ can be applied to the empirical likelihood ratio statistic.

Theorem 6.3.2. *Let* $(\boldsymbol{X}(t))_{t\in\mathbb{Z}}$ *be a linear process as defined in* (6.2.1) *with coefficient matrices* $(\Psi(j))_{j\in\mathbb{Z}}$ *satisfying* (6.2.5) *and suppose that* $\alpha \in [1, 2)$. *Under Assumptions 6.2.3 and 6.2.4, if*

$$\frac{\partial}{\partial\boldsymbol{\theta}} \int_{-\pi}^{\pi} \Psi(\omega)^* \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1} \Psi(\omega)\, d\omega \,\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \mathbf{0}, \tag{6.3.1}$$

we have

$$-2\frac{x_n^2}{n} \log R(\boldsymbol{\theta}^0) \xrightarrow{\mathcal{L}} \boldsymbol{V}^T \boldsymbol{W}^{-1} \boldsymbol{V} \quad \textit{under } H\text{: } \boldsymbol{\theta} = \boldsymbol{\theta}^0, \tag{6.3.2}$$

where

$$\boldsymbol{V} = \frac{1}{2\pi} \sum_{i,j=1}^{d} \sum_{h=1}^{\infty} \frac{S(h)_{ij}}{S_{\alpha/2}} \begin{pmatrix} \int_{-\pi}^{\pi} (B_1(\omega) + \overline{B_1(\omega)})_{ij} d\omega \\ \int_{-\pi}^{\pi} (B_2(\omega) + \overline{B_2(\omega)})_{ij} d\omega \\ \vdots \\ \int_{-\pi}^{\pi} (B_q(\omega) + \overline{B_q(\omega)})_{ij} d\omega, \end{pmatrix}$$

with

$$B_k(\omega) = \Psi(\omega)^* \frac{\partial}{\partial\theta_k} \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1} \Psi(\omega) \quad k = 1, \ldots, q,$$

and the component of $\boldsymbol{W}$ *is expressed as*

$$\begin{aligned} W_{ab} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Bigg(&\mathrm{tr} \left[\tilde{\boldsymbol{g}}(\omega) \frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \tilde{\boldsymbol{g}}(\omega) \frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \right] \\ &+\mathrm{tr} \left[\tilde{\boldsymbol{g}}(\omega) \frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \right] \mathrm{tr} \left[\tilde{\boldsymbol{g}}(\omega) \frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \right] \Bigg) d\omega, \end{aligned}$$

where $\tilde{\boldsymbol{g}}(\omega)$ *is defined as*

$$\tilde{\boldsymbol{g}}(\omega) = \Psi(\omega) \Sigma_{\tilde{Z}} \Psi(\omega)^*.$$

Remark 6.3.3. The condition (6.3.1) is a restriction on the parametrization of the pivotal quantity. However, in the case of 1-dimension, the condition is always satisfied. This is a big difference from the result of the scalar process.

Remark 6.3.4. The condition (6.3.1) seems restrictive but it has some interesting aspects. The appealing example is to consider whether the wave structures of the spectral density function between all components are "close" to each other. Consider a 2-dimensional case and assume the true power transfer function $\boldsymbol{g}(\omega)$ is given by

$$\boldsymbol{g}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \tilde{\boldsymbol{R}}(k) e^{-ik\omega}.$$

If we have to test the null hypothesis

$$H: \quad \tilde{\boldsymbol{R}}(k) = \theta^0 \tilde{\boldsymbol{R}}(j) \quad \text{or} \quad \tilde{\boldsymbol{R}}(k) = \theta^0 \tilde{\boldsymbol{R}}(j)' \quad \text{for some } k \text{ and } j,$$

then there exists a considerable way to choose an estimating function. Let the estimating function $\boldsymbol{m}(\lambda_t; \boldsymbol{\theta})$ be defined by an inverse correlation function $\boldsymbol{f}(\lambda_t; \boldsymbol{\theta})^{-1}$, which was first introduced in Cleveland (1972), and deeply discussed by Bhansali (1980). Suppose

$$\boldsymbol{f}(\omega; \theta)^{-1} = (e^{k\omega} + e^{-k\omega}) \begin{pmatrix} \theta & 0 \\ 0 & \theta \end{pmatrix} + (e^{j\omega} + e^{-j\omega}) \begin{pmatrix} \frac{1}{2}\theta^2 & 0 \\ 0 & \frac{1}{2}\theta^2 \end{pmatrix},$$

then under the hypothesis, we obtain

$$\frac{\partial}{\partial \theta} \int_{-\pi}^{\pi} \Psi(\omega)^* \boldsymbol{f}(\omega; \boldsymbol{\theta})^{-1} \Psi(\omega) \, d\omega \bigg|_{\theta = \theta^0} = \boldsymbol{0},$$

which satisfies the condition (6.3.1) in Theorem 6.3.2.

Corollary 6.3.5. With the same assumptions and condition (6.3.1), if the process (6.2.1) has the innovation process whose marginal distributions are all the same, then we can simplify the equation (6.3.2) by

$$-2\frac{x_n^2}{n} \log R(\boldsymbol{\theta}^0) \xrightarrow{\mathcal{L}} \boldsymbol{V}^T \boldsymbol{W}^{-1} \boldsymbol{V} \quad \text{under } H\colon\ \boldsymbol{\theta} = \boldsymbol{\theta}^0,$$

where V is defined above but the components of $\boldsymbol{W}$ can be expressed as

$$\begin{aligned} W_{ab} = \frac{1}{2\pi d^2} \int_{-\pi}^{\pi} \Bigg(&\mathrm{tr} \left[\boldsymbol{g}(\omega) \frac{\partial \boldsymbol{f}(\omega; \boldsymbol{\theta})^{-1}}{\partial \theta_a} \bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}^0} \boldsymbol{g}(\omega) \frac{\partial \boldsymbol{f}(\omega; \boldsymbol{\theta})^{-1}}{\partial \theta_b} \bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}^0} \right] \\ &+ \mathrm{tr} \left[\boldsymbol{g}(\omega) \frac{\partial \boldsymbol{f}(\omega; \boldsymbol{\theta})^{-1}}{\partial \theta_a} \bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}^0} \right] \mathrm{tr} \left[\boldsymbol{g}(\omega) \frac{\partial \boldsymbol{f}(\omega; \boldsymbol{\theta})^{-1}}{\partial \theta_b} \bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}^0} \right] \Bigg) d\omega. \end{aligned}$$

6.4 Numerical Results

First, we mention the procedure to construct the confidence region for the pivotal quantity $\boldsymbol{\theta}$. It is easy to see that from (6.3.2), the logarithm of the empirical likelihood ratio $\log R(\boldsymbol{\theta})$ is bounded above. That is to say, the asymptotic distribution of $-2(x_n^2/n)\log R(\boldsymbol{\theta})$ is bounded below. For this reason, the confidence region is constructed by

$$C_q = \left\{\boldsymbol{\theta} \in \Theta;\ -2\frac{x_n^2}{n}\log R(\boldsymbol{\theta}) < u_q\right\},$$

where u_q is corresponding to the qth quantile of the distribution of $\boldsymbol{V}^T\boldsymbol{W}^{-1}\boldsymbol{V}$.

Suppose that the observations $(\boldsymbol{X}(1), \cdots, \boldsymbol{X}(300))$ are generated from the 2-dimensional VAR(1) model:

$$\boldsymbol{X}(t) + A\boldsymbol{X}(t-1) = \boldsymbol{Z}(t),$$

where the marginal distributions of $\{\boldsymbol{Z}(t)\}$ are assumed to be i.i.d. symmetric 1.5-stable variables with scale 1 for simplicity. The true coefficient matrix is given by

$$A = \begin{pmatrix} 0.7 & \theta^0 \\ 0.1 & 0.5 \end{pmatrix}.$$

Like other methods, the parametrization for the power transfer function $\boldsymbol{f}(\omega;\boldsymbol{\theta})$ is possible. First, we define the fitted power transfer function $\boldsymbol{f}(\omega;\boldsymbol{\theta})$ corresponding to the estimating function as

$$\boldsymbol{f}(\omega;\boldsymbol{\theta}) = (I - B_\theta e^{i\omega})^{-1}(I - B_\theta e^{i\omega})^{-1*}, \quad \text{where} \quad B_\theta = \begin{pmatrix} 0.7 & \theta \\ 0.1 & 0.5 \end{pmatrix}.$$

The numerical results in this case are given in Table 6.1.

Table 6.1: 90% confidence intervals and lengths for true parameter.

	θ^0	Confidence Interval	Length
case 1.	0	(-0.1045, 0.0767)	0.1812
case 2.	0.3	(0.2546, 0.3387)	0.0840
case 3.	0.6	(0.5197, 0.6929)	0.1732

The figure corresponding to the confidence interval is given below. As for Figure 6.3, the blue line shows the behavior of $-2(x_n^2/n)\log R(\theta)$ with respect to the pivotal quantity θ. The interval of the red line intercepted by the blue line is the confidence interval for the pivotal quantity.

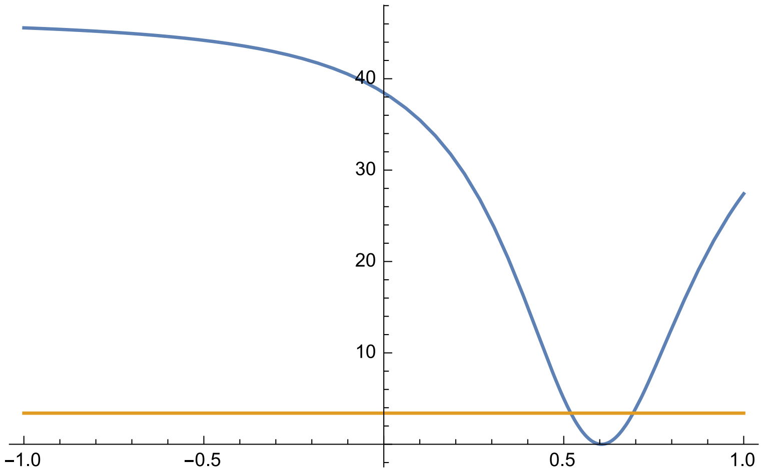


FIGURE 6.3: Confidence interval in case 3.

Next, we examine the (1,1)-component of autocorrelation (See Brockwell and Davis (1991)), which is defined as

$$\rho_{11}(l) = \gamma_{11}(l)/\gamma_{11}(0), \quad l = 0, 1, \dots .$$

This time, we fixed the matrix A to be

$$A = \begin{pmatrix} 0.7 & 0.5 \\ 0.1 & 0.5 \end{pmatrix}.$$

The estimation of this quantity is equivalent to fit the power transfer function

$$(I - B_\theta e^{-il\omega})(I - B_\theta e^{il\omega})^{-1*},$$

where B_θ has the form

$$\begin{pmatrix} \theta & b \\ 0 & c \end{pmatrix}, \quad \text{where } b \text{ and } c \text{ are arbitrary constants. (See Appendix in this Chapter.)}$$

The numerical results are given in Table 6.2.

TABLE 6.2: 90% confidence intervals and lengths for true parameter.

	l	θ^0	Confidence Interval	Length
case 4.	2	0.700	(0.5246, 0.8012)	0.2765
case 5.	3	0.590	(0.2930, 0.6077)	0.3147
case 6.	4	0.498	(0.0523, 0.5864)	0.5341

We also considered the case of estimation for pseudo true value. Consider the observations $(\boldsymbol{X}(1), \cdots, \boldsymbol{X}(300))$ are generated from the 2-dimensional

VMA(100) model with innovations $\{\boldsymbol{Z}(t);\ t \in \mathbb{Z}\}$ whose marginal distributions are i.i.d. symmetric 1.5-stable random variables with scale 1, and the coefficient matrices $A(j)$, $j = 1, \ldots, 100$ are assumed to be

$$A(j) = \begin{pmatrix} 0.7^j & j^{-2}b^j \\ 0 & 0.5^j \end{pmatrix}.$$

Suppose we use the following power transfer function $\boldsymbol{f}(\omega; \theta)$ defined by

$$\boldsymbol{f}(\omega;\theta) = (I - B_\theta \exp(i\omega))^{-1}(I - B_\theta \exp(i\omega))^{-1*}, \quad \text{where} \quad B_\theta = \begin{pmatrix} 0.5 & \theta \\ 0.4 & 0.2 \end{pmatrix},$$

for estimation. The true coefficient b of the linear model, the pseudo true value θ^0 and the confidence intervals with their lengths are reported in the following table.

TABLE 6.3: 90% confidence intervals and lengths for pseudo true parameter.

	b	$\theta^0 \approx$	Confidence Interval	Length
case 7.	0	0.0000	(−0.1685, 0.1690)	0.3375
case 8.	0.3	0.1755	(0.0467, 0.3208)	0.2741
case 9.	0.6	0.3669	(0.2601, 0.4920)	0.2320
case 10.	0.9	0.5787	(0.5046, 0.6641)	0.1596

To confirm the adequacy of the approach, we evaluated the coverage error of the empirical likelihood ratio statistics with 1000 iterations. Generally, the coverage error is considered as a sampling error away from the theoretical probability. More specifically, we generated 1000 VMA(100) processes and calculated the empirical likelihood ratio statistic $-2(x_n^2/n)\log R(\theta^0)_i$ for ith process. The coverage error is considered as the empirical probability that

$$\frac{1}{1000}\sum_{i=1}^{1000} \mathbb{1}\Big(-2\frac{x_n^2}{n}\log R(\theta^0)_i \geq u_q\Big).$$

From Table 6.4 below, the coverage error becomes worse as the pseudo true value gets larger. Looking back again from Table 6.3, we can find that the confidence intervals correspondingly becomes smaller.

TABLE 6.4: Coverage errors of confidence intervals in case 7, 8, 9 ,10.

	Coverage Error
case 7.	0.011
case 8.	0.027
case 9.	0.032
case 10.	0.049

As conclusion, it can be seen that the confidence interval for the pivotal quantity, made from the empirical likelihood ratio statistics, is not meaningless. Furthermore, the interval is narrow enough for a good understanding of the pivotal quantity $\boldsymbol{\theta}^0$.

6.5 Proof of Theorems 6.3.1 and 6.3.2

This section is devoted to show Theorems 6.3.1 and 6.3.2. First, we derive the asymptotics of $\boldsymbol{P}_n(\boldsymbol{\theta}^0)$ and $\boldsymbol{S}_n(\boldsymbol{\theta}^0)$.

Lemma 6.5.1. *Suppose $\{X(t)\}_{t=0}^{\infty}$ is generated by (6.2.1) satisfying (6.2.5). Then*

$$\boldsymbol{I}_{n,X}(\omega) = \Psi(\omega)\boldsymbol{I}_{n,Z}(\omega)\Psi(\omega)^* + \boldsymbol{R}_n(\omega).$$

If $\phi(\omega)$ is a $d \times d$ matrix-valued continuous function on $[-\pi, \pi]$, then

$$x_n \int_{-\pi}^{\pi} \mathrm{tr}[\, \boldsymbol{R}_n(\omega)\phi(\omega)\,] d\omega \xrightarrow{\mathcal{P}} 0.$$

Proof. We follow the proof of the univariate case in Mikosch et al. (1995).

$$\begin{aligned} d_{n,X}(\omega) &= n^{-1/\alpha}\sum_{t=1}^{n}\boldsymbol{X}(t)e^{i\omega t} = n^{-1/\alpha}\sum_{t=1}^{n} e^{i\omega t}(\sum_{j=0}^{\infty}\Psi(j)\boldsymbol{Z}(t-j)) \\ &= \Psi(\omega)d_{n,Z}(\omega) + n^{-1/\alpha}\sum_{j=0}^{\infty}\Psi(j)e^{ij\omega}\mathbf{Y}_{n,j}(\omega), \\ &= \boldsymbol{J}_{n,Z}(\omega) + n^{-1/\alpha}\mathbf{Y}_n(\omega) \quad \text{(say)}, \end{aligned}$$

where

$$\mathbf{Y}_{n,j}(\omega) = \sum_{t=1-j}^{n-j}\boldsymbol{Z}(t)e^{i\omega t} - \sum_{t=1}^{n}\boldsymbol{Z}(t)e^{i\omega t}.$$

Then we have

$$\boldsymbol{R}_n(\omega) = n^{-1/\alpha}\mathbf{Y}_n(\omega)\boldsymbol{J}_{n,Z}(\omega)^* + n^{-1/\alpha}\boldsymbol{J}_{n,Z}(\omega)\mathbf{Y}_n(\omega)^* + n^{-2/\alpha}\mathbf{Y}_n(\omega)\mathbf{Y}_n(\omega)^*.$$

$\sum_{j=0}^{\infty} \Psi(j)e^{ij\omega} \leq \sum_{j=0}^{\infty}\|\Psi(j)\| < \infty$, so that $\|\Psi(\omega)\|$ is stochastically bounded. Since every element of $\boldsymbol{Z}(t)$ is in the domain of attraction of a stable law with a parameter α, $\boldsymbol{J}_{n,Z}(\omega)$ is also stochastically bounded. As results in the proof of lemma 6.2 in Mikosch et al. (1995), we know that for each $l \in 1, 2, \ldots, d$,

$$\sum_{j=0}^{\infty} \Psi(j)_{kl} e^{ij\omega} Y_{n,j}(\omega)_l = O_p(1)$$

and

$$\int_{-\pi}^{\pi} n^{-2/\alpha} |\sum_{j=0}^{\infty} \Psi(j)_{kl} e^{ij\omega} Y_{n,j}(\omega)_l|^2 d\omega = o_p(x_n^{-2}).$$

Combining these two results, it is easy to see that $\mathbf{Y}_n(\omega) = O_p(1)$, and by the boundedness of $\phi(\omega)$, the residual term

$$\begin{aligned}
& x_n \Big| \int_{-\pi}^{\pi} \mathrm{tr}[\boldsymbol{R}_n(\omega)\phi(\omega)]\, d\omega \Big| \\
\leq\ & x_n \int_{-\pi}^{\pi} |\mathrm{tr}[\, \boldsymbol{R}_n(\omega)\phi(\omega)\,]|\, d\omega \\
\leq\ & x_n \int_{-\pi}^{\pi} \|\boldsymbol{R}_n(\omega)\|_E \|\phi(\omega)\|_E\, d\omega \\
\leq\ & c_1 x_n \int_{-\pi}^{\pi} \|n^{-1/\alpha} \mathbf{Y}_n(\omega) \boldsymbol{J}_{n,Z}(\omega)^*\|_E \\
& + \|n^{-1/\alpha} \boldsymbol{J}_{n,Z}(\omega) \mathbf{Y}_n(\omega)^*\|_E + \|n^{-2/\alpha} \mathbf{Y}_{n,Z}(\omega) \mathbf{Y}_{n,Z}(\omega)^*)\|_E\, d\omega \\
\leq\ & c_2 x_n \Big\{ \Big(\int_{-\pi}^{\pi} \|\boldsymbol{I}_{n,Z}(\omega)\|_E^2\, d\omega \Big)^{1/2} \Big(\int_{-\pi}^{\pi} n^{-2/\alpha} \|\mathbf{Y}_n(\omega)\|_E^2\, d\omega \Big)^{1/2} \\
& \quad + \int_{-\pi}^{\pi} n^{-2/\alpha} \|\mathbf{Y}_n(\omega)\|_E^2\, d\omega \Big\}. \\
\xrightarrow{\mathcal{P}}\ & 0.
\end{aligned}$$

This is what we have to show. □

Before looking into the asymptotics of $\boldsymbol{P}_n(\boldsymbol{\theta}^0)$, we have to show the existence of the limit matrix of the autocovariance matrix in distribution. If the components of the vector $\boldsymbol{Z}$ are mutually independent, then we have the lemma due to Davis et al. (1986) by applying continuous mapping theorem.

As what we mentioned in Section 6.2, $y_n = (n \log n)^{1/\alpha}$. It is obvious that $Z(1)_k$'s satisfy followings:

$$P(|Z(1)_i| > x) = x^{-\alpha} L(x), \quad i = 1, 2, \ldots, d \tag{6.5.1}$$

with $\alpha > 0$ and $L(x)$ a slowly varying function at ∞ and

$$\frac{P(Z(1)_i > x)}{P(|Z(1)_i| > x)} \to w, \quad \frac{P(Z(1)_i < -x)}{P(|Z(1)_i| > x)} \to v \tag{6.5.2}$$

as $x \to \infty$, $0 \le w \le 1$ and $v = 1 - w$.

Lemma 6.5.2. *Let $\{\boldsymbol{Z}(t)\}$ be a sequence of iid random vectors satisfying (6.5.1) and (6.5.2) with $0 < \alpha < 2$ and $E|Z(1)_i|^\alpha = \infty$ for all $i = 1, 2, \ldots, d$. Then*

$$\left(n^{-2/\alpha}\sum_{t=1}^{n}\boldsymbol{Z}(t)\boldsymbol{Z}(t)', y_n^{-1}\sum_{t=1}^{n}\boldsymbol{Z}(t)\boldsymbol{Z}(t+1)', \ldots, y_n^{-1}\sum_{t=1}^{n}\boldsymbol{Z}(t)\boldsymbol{Z}(t+h)'\right) \Rightarrow (S(0), S(1), \ldots, S(h)),$$

where $S(0), S(1), \ldots, S(h)$ are independent stable random matrices; the components of $S(0)$ are all positive with index $\alpha/2$, and $S(1), \ldots, S(h)$ are identically distributed with index α.

Proof. See Davis et al. (1986). □

Proof of Theorem 6.3.1. From Lemma 6.5.2, we can see that

$$\left(n^{-2/\alpha}\hat{\Gamma}_{n,Z}(0), y_n^{-1}\hat{\Gamma}_{n,Z}(k), k = 1 \ldots, h\right) \Rightarrow (S(0), S(1), \ldots, S(h)).$$

Note that $\operatorname{tr}\hat{\Gamma}_{n,Z}(0) = \|Z\|_N^2$, according to the continuous mapping theorem, the statement holds true if we show

$$x_n\int_{-\pi}^{\pi}\operatorname{tr}\left[\{\boldsymbol{I}_{n,X}(\omega) - \Psi(\omega)\hat{\Gamma}_{n,Z}(0)\Psi(\omega)^*\}\phi_k(\omega)\right]d\omega \xrightarrow{\mathcal{L}} \sum_{i,j=1}^{d}\sum_{h=1}^{\infty}S(h)_{ij}\int_{-\pi}^{\pi}(A(\omega) + \overline{A(\omega)})_{ij}\,d\omega. \tag{6.5.3}$$

From Lemma 6.5.1 and Lemma 6.6.1 in Appendix of this chapter,

$$\begin{aligned}
&x_n\int_{-\pi}^{\pi}\operatorname{tr}\Big[\{\boldsymbol{I}_{n,X}(\omega) - \Psi(\omega)\hat{\Gamma}_{n,Z}(0)\Psi(\omega)^*\}\phi_k(\omega)\Big]d\omega\\
=\;& x_n\int_{-\pi}^{\pi}\operatorname{tr}\left[\{\Psi(\omega)\boldsymbol{I}_{n,Z}(\omega)\Psi(\omega)^* - \Psi(\omega)\hat{\Gamma}_{n,Z}(0)\Psi(\omega)^* + \boldsymbol{R}(\omega)\}\phi_k(\omega)\right]d\omega\\
=\;& x_n\int_{-\pi}^{\pi}\operatorname{tr}\left[\{\Psi(\omega)(\boldsymbol{I}_{n,Z}(\omega) - \hat{\Gamma}_{n,Z}(0))\Psi(\omega)^*\}\phi_k(\omega)\right]d\omega\\
&\qquad + x_n\int_{-\pi}^{\pi}\operatorname{tr}\left[\boldsymbol{R}(\omega)\phi_k(\omega)\right]d\omega\\
=\;& x_n\int_{-\pi}^{\pi}\operatorname{tr}\left[\left(\sum_{h=1}^{n-1}\hat{\Gamma}_{n,Z}(h)e^{-ih\omega}\right)\Psi(\omega)^*\phi_k(\omega)\Psi(\omega)\right]d\omega
\end{aligned}$$

$$
\begin{aligned}
& + x_n \int_{-\pi}^{\pi} \mathrm{tr}\left[\left(\sum_{h=1}^{n-1} \hat{\Gamma}_{n,Z}(h)' e^{ih\omega} \right) \Psi(\omega)^* \phi_k(\omega) \Psi(\omega) \right] d\omega + o_p(1) \\
= \quad & x_n \left(\int_{-\pi}^{\pi} \sum_{i,j=1}^{d} \left[\sum_{h=1}^{n-1} \hat{\Gamma}_{n,Z}(h) \right]_{ij} \left[e^{-ih\omega} \Psi(-\omega)^* \phi_k(-\omega) \Psi(-\omega) \right]_{ij} d\omega \right. \\
& \left. + \int_{-\pi}^{\pi} \sum_{i,j=1}^{d} \left[\sum_{h=1}^{n-1} \hat{\Gamma}_{n,Z}(h) \right]_{ij} \left[e^{ih\omega} \Psi(\omega)^* \phi_k(\omega) \Psi(\omega) \right]_{ij} d\omega \right) \\
\xrightarrow{\mathcal{L}} \quad & \sum_{i,j=1}^{d} \sum_{h=1}^{\infty} S(h)_{ij} \int_{-\pi}^{\pi} (A(\omega) + \overline{A(\omega)})_{ij} \, d\omega.
\end{aligned}
$$

Thus the assertion of Lemma 6.5.2 is shown by continuous mapping theorem. □

Remark 6.5.3. The last result of convergence is due to Lemma 6.6.1, which guarantees the tightness of the convergence.

Remark 6.5.4. The assumption of independence on the components of $\boldsymbol{Z}(t)$ is for simplicity and for simulation. The condition of regular variation on the vector case is crucial for the convergence of $\boldsymbol{Z}(t)$ with some other technical conditions. For detail, we recommend to refer to Bartkiewicz et al. (2011).

From the definition, we have

$$
\sum_{t=1}^{n} \sum_{i=1}^{d} \tilde{Z}(t)_i^2 = 1 \quad \text{almost surely},
$$

which shows the second moment of $\tilde{\boldsymbol{Z}}(t)$ is finite. By the properties that the components of vectors are mutually independent and they are symmetry around 0, we assume generally

$$
E\left[\tilde{Z}(t)_i \tilde{Z}(s)_j \right] = \Sigma_{\tilde{Z}} = \begin{cases} \frac{\sigma_{ij}}{n}, & \text{if } t = s, \\ 0 & \text{if } t \neq s. \end{cases} \tag{6.5.4}
$$

This is not a special case since we have the following example:

Example 2 (Case that the correlation between all elements of $\boldsymbol{Z}(t)$ is 1)**.** Assume that the marginal distributions $Z(t)_j$ of $\boldsymbol{Z}(t)$ are independent symmetric α-stable distributions with different scales σ_j. Then since the sum of all marginal distribution $\sum_{j=1}^{n} Z(t)_j$ has the same distribution, we can see that

$$
E \sum_{i=1}^{d} \tilde{Z}(t)_i^2 = \frac{1}{n}.
$$

Also, according to the different scale, we can write $Z(t)_j = \sigma_j Z'(t)_j$ where all $Z'(t)_j$ are stable with scale 1. Then we have

$$E\sum_{i=1}^{d}\tilde{Z}(t)_i^2 = E\sum_{i=1}^{d}\sigma_i^2\tilde{Z}'(t)_i = \frac{1}{n},$$

which is followed by

$$E\,\tilde{Z}'(t)_i = \frac{1}{n}\left(\sum_{i=1}^{d}\sigma_i^2\right)^{-1}.$$

Accordingly, we have

$$E(\tilde{Z}(t)_i\tilde{Z}(t)_j) = \frac{1}{n}\frac{\sigma_i\sigma_j}{\sum_{i=1}^{d}\sigma_i^2}.$$

The representation (6.5.4) is just a generalization of this idea.

Lemma 6.5.5. *Assume the covariance matrix of self-normalized process $\{\tilde{Z}\}$ is given by $\Sigma_{\tilde{Z}}$. If $\alpha \in [1,2)$, then*

$$(n^{-2/\alpha}\|Z\|_N^2)^{-2}\boldsymbol{S}_n(\boldsymbol{\theta}^0) \xrightarrow{\mathcal{P}} \boldsymbol{W},$$

where the (a,b)-component of $\boldsymbol{W}$ satisfies

$$W_{ab} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\mathrm{tr}\left[\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]\right.$$
$$\left.+\mathrm{tr}\left[\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]\mathrm{tr}\left[\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]\right)d\omega,$$

where $\tilde{\boldsymbol{g}}(\omega)$ is defined as

$$\tilde{\boldsymbol{g}}(\omega) = \Psi(\omega)\Sigma_{\tilde{Z}}\Psi(\omega)^*.$$

Proof. Apply the decomposition in Lemma 6.5.1 again, we have

$$\boldsymbol{I}_{n,X}(\omega) = \Psi(\omega)\boldsymbol{I}_{n,Z}(\omega)\Psi(\omega)^* + \boldsymbol{R}_n(\omega).$$

Using self-normalized form, we can see that

$$\boldsymbol{I}_{n,\tilde{X}}(\omega) \equiv (n^{-2/\alpha}\|Z\|_N^2)^{-2}\boldsymbol{I}_{n,X}(\omega) = \Psi(\omega)\boldsymbol{I}_{n,\tilde{Z}}(\omega)\Psi(\omega)^* + \boldsymbol{R}_n(\omega).$$

Taking the expectation of the product of periodogram of $\tilde{\boldsymbol{Z}}$, we obtain

$$E(I_{n,\tilde{Z}}(\lambda_1)_{pq}I_{n,\tilde{Z}}(\lambda_2)_{rs})$$

$$= E\left(\sum_{m,l,k,j} \tilde{Z}_p(m)\tilde{Z}_q(l)\tilde{Z}_r(k)\tilde{Z}_s(j)\exp\{\,i((j-k)\lambda_1-(l-m)\lambda_2)t\}\right)$$

$$= \begin{cases} \sigma_{pq}\sigma_{rs}+\sigma_{pr}\sigma_{qs}+o_p(1) & \text{if } \lambda_1=\lambda_2, \\ \sigma_{pq}\sigma_{rs}+\sigma_{ps}\sigma_{qr}+o_p(1) & \text{if } \lambda_1=-\lambda_2. \end{cases}$$

Therefore, if we write $g(\lambda)_{ab} = (\sum_{j=0}^{n}\Psi(j)e^{-ij\lambda})_{ab}$, then

$$\begin{aligned} &\lim_{n\to\infty} E(I_{n,\tilde{X}}(\lambda_t)_{pq}I_{n,\tilde{X}}(\lambda_t)_{rs}) \\ =& \sum_{k,l,m,n} g(\lambda_t)_{pk}\overline{g}_{ql}(\lambda_t)\overline{g}_{rm}(\lambda_t)g_{sn}(\lambda_t)(\sigma_{pq}\sigma_{rs}+\sigma_{pr}\sigma_{qs}) \\ =& \tilde{\boldsymbol{g}}(\omega)_{pq}\tilde{\boldsymbol{g}}(\omega)_{rs}+\tilde{\boldsymbol{g}}(\omega)_{pr}\tilde{\boldsymbol{g}}(\omega)_{qs}. \end{aligned}$$

If $\alpha\in[1,2)$, we can write $\boldsymbol{S}_n(\boldsymbol{\theta}^0)$ in the integration form, i.e.

$$\begin{aligned} E[\,\boldsymbol{S}_n(\boldsymbol{\theta}^0)_{ab}\,] =& \frac{1}{2\pi} \\ &\times\Bigg\{\int_{-\pi}^{\pi}\sum_{\beta_1,\beta_2,\beta_3,\beta_4=1}^{d}\tilde{\boldsymbol{g}}(\omega)_{\beta_1\beta_2}\tilde{\boldsymbol{g}}(\omega)_{\beta_3\beta_4}\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{\beta_2\beta_1}}{\partial\theta_a}\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{\beta_4\beta_3}}{\partial\theta_b}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}d\omega \\ &+\int_{-\pi}^{\pi}\sum_{\beta_1,\beta_2,\beta_3,\beta_4=1}^{d}\tilde{\boldsymbol{g}}(\omega)_{\beta_1\beta_3}\tilde{\boldsymbol{g}}(\omega)_{\beta_2\beta_4}\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{\beta_2\beta_1}}{\partial\theta_a}\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{\beta_4\beta_3}}{\partial\theta_b}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}d\omega\Bigg\}. \end{aligned}$$

In other words,

$$\begin{aligned} W_{ab} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\Bigg(&\mathrm{tr}\left[\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right] \\ &+\mathrm{tr}\left[\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]\mathrm{tr}\left[\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]\Bigg)d\omega. \end{aligned}$$

The convergence in probability is guaranteed by the result that

$$\sum_{k\neq l}\mathrm{Cov}(I_{n,\tilde{Z}}(\lambda_k)^2_{pq}, I_{n,\tilde{Z}}(\lambda_l)^2_{rs}) = O(n).$$

□

Corollary 6.5.6. If all elements of $\boldsymbol{Z}(t)$ are i.i.d symmetric α stable random variables, then

$$(n^{-2/\alpha}\|Z\|_N^2)^{-2}\boldsymbol{S}_n(\boldsymbol{\theta}^0)\xrightarrow{\mathcal{P}}\boldsymbol{W},$$

where the (a,b)-component of $\boldsymbol{W}$ is

$$W_{ab} = \frac{1}{2\pi d^2}\int_{-\pi}^{\pi}\Bigg(\mathrm{tr}\left[\boldsymbol{g}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\boldsymbol{g}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]$$

$$+\mathrm{tr}\left[\boldsymbol{g}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]\mathrm{tr}\left[\boldsymbol{g}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}\right]\Bigg) d\omega.$$

Proof of Theorem 6.3.2. First, we will derive the asymptotic distribution of the empirical likelihood ratio. For convenience, we set $\boldsymbol{p} = (p_1, \ldots, p_n)$. Introducing Lagrange multiplier $L(\boldsymbol{p}, \boldsymbol{\phi}, k)$,

$$L(\boldsymbol{p}, \boldsymbol{\phi}, k) = \sum_{t=1}^{n} \log(np_t) - n\boldsymbol{\phi}^T \sum_{t=1}^{n} p_t \boldsymbol{m}(\lambda_t; \boldsymbol{\theta}^0) + k\left(\sum_{t=1}^{n} p_t - 1\right).$$

Differentiating $L(\boldsymbol{p}, \boldsymbol{\phi}, k)$ with respect to all parameters, we have equations:

$$p_t = \frac{1}{n}\frac{1}{1 + \boldsymbol{\phi}^T \boldsymbol{m}(\lambda_t; \boldsymbol{\theta}^0)},$$

where $\boldsymbol{\phi}$ satisfies

$$\frac{1}{n}\sum_{t=1}^{n} \frac{\boldsymbol{m}(\lambda_t; \boldsymbol{\theta}^0)}{1 + \phi^T \boldsymbol{m}(\lambda_t; \boldsymbol{\theta}^0)} = \mathbf{0}. \tag{6.5.5}$$

If we write

$$Y_t = \boldsymbol{\phi}^T \boldsymbol{m}(\lambda_t; \boldsymbol{\theta}^0), \tag{6.5.6}$$

then we have

$$np_t = (1 + Y_t)^{-1},$$

and from (6.5.5)

$$\boldsymbol{\phi} = \boldsymbol{S}_n(\boldsymbol{\theta}^0)^{-1}\left\{\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{m}(\lambda_t; \boldsymbol{\theta}^0)\right\} + \boldsymbol{\epsilon}, \tag{6.5.7}$$

where we obtain the expression

$$\boldsymbol{\epsilon} = \frac{1}{n}\sum_{t=1}^{n} \frac{\boldsymbol{m}(\lambda_t; \boldsymbol{\theta}^0) Y_t^2}{1 + Y_t}, \tag{6.5.8}$$

since it is easy to see that

$$\frac{1}{1 + Y_t} = 1 - Y_t + \frac{Y_t^2}{1 + Y_t}.$$

Thus the empirical likelihood ratio can be decomposed like

$$-2\log R(\boldsymbol{\theta}^0) \quad = \quad 2\sum_{t=1}^{n} \log(1 + Y_t)$$

$$\begin{aligned} &= 2\sum_{t=1}^{n} Y_t - \sum_{t=1}^{n} Y_t^2 + \sum_{t=1}^{n} O_p(Y_t^3) \\ &= n\boldsymbol{P}_n(\boldsymbol{\theta}^0)^T \boldsymbol{S}_n(\boldsymbol{\theta}^0)^{-1} \boldsymbol{P}_n(\boldsymbol{\theta}^0) - n\boldsymbol{\epsilon}^T \boldsymbol{S}_n(\boldsymbol{\theta}^0)\boldsymbol{\epsilon} + \sum_{t=1}^{n} O_p(Y_t^3). \end{aligned} \tag{6.5.9}$$

To guarantee the order of convergence, we must control the order of $\boldsymbol{\epsilon}$ and Y_t^3 simultaneously. Fortunately, we can keep them converge to 0 in probability in the stable case as well as in the usual regular case.

To give a clear overview, we define an order in the probability order notation. If $O_p(A_n)/O_p(B_n) \to o_p(1)$, then we say A_n convergences to 0 in probability faster than B_n. It is denoted by

$$\min(O_p(A_n), O_p(B_n)) = O_p(A_n) \quad \text{or} \quad O_p(A_n) \leq O_p(B_n).$$

If $O_p(A_n)/O_p(B_n) \to O_p(1)$, then we say that A_n is equivalent to B_n in the order sense.

Define $Z_n = \max_{1\leq i\leq p} \max_{1\leq t\leq n} \boldsymbol{m}(\lambda_t; \boldsymbol{\theta}^0)$. Then it is easy to see that

$$Z_n \leq \max_{1\leq t\leq n} \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{tr}\{\boldsymbol{f}(\lambda_t; \boldsymbol{\theta}^0)^{-1} \boldsymbol{I}_{n,X}(\lambda_t)\} \right\| \leq C \sup_{\substack{i,j \\ \omega\in[-\pi,\pi]}} |\boldsymbol{I}_{n,X}(\omega)_{ij}|.$$

Then from (6.5.7), we have

$$O_p(\|\boldsymbol{\phi}\|)(O_p(\|\boldsymbol{S}_n(\boldsymbol{\theta}^0)\|) - O_p(Z_n)O_p(\|\boldsymbol{P}_n(\boldsymbol{\theta}^0)\|)) \leq O_p(\|\boldsymbol{P}_n(\boldsymbol{\theta}^0)\|).$$

From (6.5.8), this leads us to the order of $\boldsymbol{\epsilon}$,

$$\begin{aligned} \|\boldsymbol{\epsilon}\| &\leq \|Z_n\| \|\boldsymbol{S}_n(\boldsymbol{\theta}^0)\| \|\boldsymbol{\phi}\|^2 |1 + Y_t|^{-1} = O_p(\|Z_n\|) O_p(\|\boldsymbol{S}_n(\boldsymbol{\theta}^0)\|) O_p(\|\boldsymbol{\phi}\|^2) \\ &= \min(O_p(Z_n) O_p(\|\boldsymbol{S}_n(\boldsymbol{\theta}^0)^{-1}\|) O_p(\|\boldsymbol{P}_n(\boldsymbol{\theta}^0)\|^2), O_p(Z_n^{-1}) O_p(\boldsymbol{S}_n(\boldsymbol{\theta}^0))). \end{aligned}$$

In the regular case, we can see that

$$O_p(\|\boldsymbol{S}_n(\boldsymbol{\theta}^0)\|) = O_p(\|\boldsymbol{S}_n(\boldsymbol{\theta}^0)\|^{-1}) = O_p(1).$$

Then the expression can be considered by

$$\begin{aligned} &\min(O_p(Z_n) O_p(\|\boldsymbol{S}_n(\boldsymbol{\theta}^0)^{-1}\|) O_p(\|\boldsymbol{P}_n(\boldsymbol{\theta}^0)\|^2), O_p(Z_n^{-1}) O_p(\boldsymbol{S}_n(\boldsymbol{\theta}^0))) \\ &\qquad = O_p(Z_n) O_p(\|\boldsymbol{P}_n(\boldsymbol{\theta}^0)\|^2) \end{aligned}$$

from LIL and CLT. Thus the order of $\|\boldsymbol{\epsilon}\|$ is

$$\|\boldsymbol{\epsilon}\| = O_p(n^{-1}\log n),$$

because of

$$\begin{aligned} O_p(Z_n) &= O_p(\log n), \\ O_p(\|\boldsymbol{P}_n(\boldsymbol{\theta}^0)\|) &= O_p(n^{-1/2}). \end{aligned}$$

In the stable case, defining the statistic well seems a little crucial. In this paper, the order of the statistic $\boldsymbol{S}_n(\boldsymbol{\theta}^0)$ is $O_p(1)$. As shown in Mikosch and Samorodnitsky (2000), if $\alpha \neq 1$,

$$\begin{aligned} O_p(Z_n) &= O_p((\log n)^{2-2/\alpha}) \\ O_p(\|\boldsymbol{P}_n(\boldsymbol{\theta}^0)\|) &= O_p((\log n/n)^{1/\alpha}), \end{aligned}$$

which is followed by

$$\|\boldsymbol{\phi}\| = O_p((\log n/n)^{1/\alpha}).$$

Therefore

$$\|\boldsymbol{\epsilon}\| = O_p((\log n)^2 n^{-2/\alpha}).$$

In both cases, we can simplify the notation, that is,

$$\|\boldsymbol{\epsilon}\| = o_p(1).$$

Last, we investigate the third term in (6.5.9). From (6.5.6), it is easy to see that

$$|Y_t|^3 \leq \|\phi\|^3 \|\boldsymbol{m}(\lambda_t; \boldsymbol{\theta}^0)\|^3 = O_p((\log n)^{2+1/\alpha} n^{-3/\alpha}).$$

Now, multiplying true order x_n^2/n to the empirical likelihood ratio in (6.5.9), the orders of the last two terms in the right hand side are $O_p((\log n)^{4-2/\alpha} n^{-2/\alpha})$ and $O_p((\log)^{2-1/\alpha} n{-}1/\alpha)$ respectively, and thus $o_p(1)$.

Apply Theorem 6.3.1 to $x_n \boldsymbol{P}_n(\boldsymbol{\theta}^0)$, we can see that

$$\begin{aligned} x_n \boldsymbol{P}_n(\boldsymbol{\theta}^0) &= \frac{x_n}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{tr}\left[\, \boldsymbol{f}(\omega; \boldsymbol{\theta})^{-1} \boldsymbol{I}_{n,X}(\omega) \,\right] d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \\ &= \frac{x_n}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{tr}\left[\, \boldsymbol{f}(\omega; \boldsymbol{\theta})^{-1} \{\boldsymbol{I}_{n,X}(\omega) - \Psi(\omega)(\hat{\Gamma}_{n,Z}(0))\Psi(\omega)^*\} \,\right] d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \\ &= \frac{x_n}{2\pi} \end{aligned}$$

$$\times \begin{pmatrix} \int_{-\pi}^{\pi} \mathrm{tr}\left[\frac{\partial}{\partial\theta_1} \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}\{\boldsymbol{I}_{n,X}(\omega) - \Psi(\omega)(\hat{\Gamma}_{n,Z}(0))\Psi(\omega)^*\}\right] d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \\ \int_{-\pi}^{\pi} \mathrm{tr}\left[\frac{\partial}{\partial\theta_2} \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}\{\boldsymbol{I}_{n,X}(\omega) - \Psi(\omega)(\hat{\Gamma}_{n,Z}(0))\Psi(\omega)^*\}\right] d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \\ \vdots \\ \int_{-\pi}^{\pi} \mathrm{tr}\left[\frac{\partial}{\partial\theta_q} \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}\{\boldsymbol{I}_{n,X}(\omega) - \Psi(\omega)(\hat{\Gamma}_{n,Z}(0))\Psi(\omega)^*\}\right] d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \end{pmatrix}$$

$$\xrightarrow{\mathcal{L}} \frac{1}{2\pi}\sum_{i,j=1}^{d}\sum_{h=1}^{\infty} S(h)_{ij} \begin{pmatrix} \int_{-\pi}^{\pi}(B_1(\omega)+\overline{B_1(\omega)})_{ij}d\omega \\ \int_{-\pi}^{\pi}(B_2(\omega)+\overline{B_2(\omega)})_{ij}d\omega \\ \vdots \\ \int_{-\pi}^{\pi}(B_q(\omega)+\overline{B_q(\omega)})_{ij}d\omega. \end{pmatrix}$$

where

$$B_k(\omega) = \Psi(\omega)^* \frac{\partial}{\partial\theta_k}\boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}\Psi(\omega) \quad k = 1,\dots,q.$$

Remember that

$$n^{-2/\alpha}\|Z\|_N^2 \xrightarrow{\mathcal{L}} S_{\alpha/2},$$

and then the asymptotic distribution is obtained by

$$\frac{x_n \boldsymbol{P}_n(\boldsymbol{\theta}^0)}{n^{-2/\alpha}\|Z\|_N^2} \xrightarrow{\mathcal{L}} \frac{1}{2\pi}\sum_{i,j=1}^{d}\sum_{h=1}^{\infty} \frac{S(h)_{ij}}{S_{\alpha/2}} \begin{pmatrix} \int_{-\pi}^{\pi}(B_1(\omega)+\overline{B_1(\omega)})_{ij}d\omega \\ \int_{-\pi}^{\pi}(B_2(\omega)+\overline{B_2(\omega)})_{ij}d\omega \\ \cdots \\ \int_{-\pi}^{\pi}(B_q(\omega)+\overline{B_q(\omega)})_{ij}d\omega \end{pmatrix}.$$

Thus the limit of $-2\frac{x_n^2}{n}\log R(\boldsymbol{\theta})$ is

$$-2\frac{x_n^2}{n}\log R(\boldsymbol{\theta}^0) \xrightarrow{\mathcal{L}} \boldsymbol{V}'\boldsymbol{W}^{-1}\boldsymbol{V},$$

where

$$\boldsymbol{V} = \frac{1}{2\pi}\sum_{i,j=1}^{d}\sum_{h=1}^{\infty} \frac{S(h)_{ij}}{S_{\alpha/2}} \begin{pmatrix} \int_{-\pi}^{\pi}(B_1(\omega)+\overline{B_1(\omega)})_{ij}d\omega \\ \int_{-\pi}^{\pi}(B_2(\omega)+\overline{B_2(\omega)})_{ij}d\omega \\ \cdots \\ \int_{-\pi}^{\pi}(B_q(\omega)+\overline{B_q(\omega)})_{ij}d\omega, \end{pmatrix}$$

and the (a, b)-element of $\boldsymbol{W}$ is given in Theorem 6.3.1. □

6.6 Appendix

6.6.1 Tightness

Lemma 6.6.1 (Klüppelberg and Mikosch (1996))**.** *Suppose $\{Z_t\}_{1\le t\le n}$ is a sequence of iid symmetric α-stable random variables for $\alpha \in (0,2)$. Let f_t be real* *numbers such that*

$$\sum_{t=-\infty}^{\infty} |f_t|^{\mu} < \infty$$

for some $\mu < \alpha$. *If* $f_0 = 0$, *then*

$$(\gamma_{n,Z}^2, y_n^{-1} \sum_{1 \le t,s \le n} f_{t-s} Z_t Z_s) \to_d (Y_0, Z_1 (\sum_{t=1}^{\infty} |f_t + f_{-t}|^{\alpha})^{1/\alpha}).$$

If $f_0 \neq 0$, *then*

$$n^{-2/\alpha} \sum_{1 \le t,s \le n} f_{t-s} Z_t Z_s \to_d f_0 Y_0.$$

6.6.2 Regularly varying tail

Definition 6.6.2. A distribution F has exponential tails with rate $\alpha > 0$, if

$$\lim_{t \to \infty} \frac{\bar{F}(t-u)}{\bar{F}(t)} = e^{\alpha u} \quad \text{for all real } u.$$

It is denoted by $F \in \mathcal{L}_{\alpha}$.

The definition is equivalent to the definition of regularly varying tail, if one put $\log t$ and $\log u$ into the definition above. To guarantee $\|Z\|_N^2$ is in the domain of attraction of a stable limit with $\alpha/2$, we have the Theorem below.

Theorem 6.6.3 (Embrechts and Goldie (1980), Theorem 3)**.** *If both* $F \in \mathcal{L}_{\alpha}$, $G \in \mathcal{L}_{\alpha}$, *then* $H = F * G \in \mathcal{L}_{\alpha}$.

Remark 6.6.4. Regularly varying tail is necessary and sufficient condition for a sequence of i.i.d random variables or random vectors in the domain of attraction of a stable law.

6.6.3 Estimation of autocorrelation

To estimate $\Gamma(0)^{-1}\Gamma(j)$, we can see the problem as a fitting problem, i.e. fit the spectral whose inverse matrix is

$$(I - \Theta e^{-ij\omega})(I - \Theta' e^{ij\omega}).$$

Then the pivotal value satisfies

$$\frac{\partial}{\partial \Theta} \int_{-\pi}^{\pi} (I - \Theta e^{-ij\omega})(I - \Theta' e^{ij\omega}) \boldsymbol{g}(\omega)\, d\omega |_{\boldsymbol{\theta} = \boldsymbol{\theta}^0} = \mathbf{0}.$$

In fact, using formula

$$\begin{aligned}
\frac{\partial}{\partial\Theta}\mathrm{tr}(\Theta\boldsymbol{g}(\omega)) &= \Theta'\boldsymbol{g}(\omega), \\
\frac{\partial}{\partial\Theta}\mathrm{tr}(\Theta\Theta'\boldsymbol{g}(\omega)) &= \boldsymbol{g}(\omega)\Theta + \boldsymbol{g}(\omega)\Theta',
\end{aligned}$$

we have

$$\Gamma(0)\,\Theta_0 = \Gamma(j)$$

by Herglotz's spectral representation theorem. That is,

$$\Gamma(j) = \int_{-\pi}^{\pi} e^{ij\omega}\, dF(\omega).$$

Thus we have

$$\Theta_0 = \Gamma(0)^{-1}\Gamma(j).$$

Also, we can extend this full autocorrelation case to any one element case in the same way. (See Section 6.4 above.)

6.6.4 Program

In this subsection, we give the program code of Mathematica we used for the numerical results in Section 6.4. For practical usage, the sample size T, the length p of VMA process, the dimension d, the true matrix A and the parametrized matrix B has to be determined in advance.

```
MakeA[a_, b_, c_,
   p_] := (A = Table[{{a^i, b^i/i^2}, {0, c^i}}, {i, 1, p}]);
VMA[p_] := (z = {RandomVariate[StableDistribution[1, 1.5, 0, 0, 1],
      2]};
For[i = 1, i < T + p, i++,
    z = Append[z,
      RandomVariate[StableDistribution[1, 1.5, 0, 0, 1], 2]]];
obs = Table[
     z[[p + j]] + Total[Table[A[[i]].z[[p + j - i]], {i, 1, p}]], {j,
      1, T}]);
filter[P__] := -{-IdentityMatrix[d], P}
InvSpectral[x_, P__] := (G = filter[P];
Transpose[Total[MapThread[Times, {G, ec[-x, Length[G]]}]]].Total[
    MapThread[Times, {G, ec[x, Length[G]]}]])
ec[w_, n_] := Table[E^(I (s - 1) w), {s, 1, n}]
SpectralMA[x_, P_] := (G = Prepend[P, IdentityMatrix[2]];
```

```
Total[MapThread[Times, {G, ec[x, Length[G]]}]].Transpose[
    Total[MapThread[Times, {G, ec[-x, Length[G]]}]]])
syu[x_, P__] := (G = Prepend[P, IdentityMatrix[2]];
Total[MapThread[Times, {G, ec[x, Length[G]]}]]);
e[r_, n_] := Table[E^(2 Pi I (r - 1) (s - 1)/T), {s, 1, n}]
Pdg[x_, r_, a_] :=
 T^(-2/a) Outer[Times, e[r, T].N[x], Conjugate[e[r, T].N[x]]]
fitar[P__, r_] := (F = filter[P];
Transpose[
    Total[MapThread[Times, {F, Conjugate[e[r, Length[F]]]}]]].Total[
    MapThread[Times, {F, e[r, Length[F]]}]])
M[x_, P_, r_, a_] := D[Tr[fitar[P, r].Pdg[x, r, a]], t1]
MList[x_, P_, tr_, a_] :=
  Chop[Table[M[x, P, r, a], {r, 1, T}] /. {t1 -> tr}];
ko[a_] := If[a == 1, 2/Pi, (1 - a)/(Gamma[2 - a] Cos[Pi a/2])];
Y0[a_] := StableDistribution[a/2, 1, 0, (ko[a/2])^(-2/a)];
Z0[a_] := StableDistribution[a, 0, 0, (ko[a])^(-1/a)];
Quant = Quantile[(RandomVariate[Z0[1.5], 10000]/
      RandomVariate[Y0[1.5], 10000])^2, 0.90];
tr = Rationalize[
  Chop[t1 /. (FindRoot[
       D[Integrate[
          Tr[InvSpectral[x, B].SpectralMA[x, A]], {x, -Pi, Pi}],
         t1] == 0, {t1, 0.701}])[[1]]]];
w0 = Simplify[(D[InvSpectral[x, B], t1] /. {t1 -> tr})].SpectralMA[x,
     A];
w = Integrate[Tr[w0.w0] + Tr[w0] Tr[w0], {x, -Pi, Pi}]/(2 Pi);
v := Total[Total[w0]]
VB[a_] := (k = 1; S = -1; s = 1;
While[(s > S/1000 || s == 0) && k < 20, S = S + s;
s = Abs[Integrate[v Cos[k x], {x, -Pi, Pi}]]^(a); Print[s]; k++]; S)
coef[a_] :=
 Abs[Total[
    Flatten[Chop[
      D[Integrate[
         ConjugateTranspose[syu[x, A]].InvSpectral[x, B].syu[x,
           A], {x, -Pi, Pi}], t1] /. {t1 -> tr}], 1]]]^a
V[a_] := (VB[a] + coef[a])^(1/a)/Pi
Val = V[1.5];
```

```
confi = Quant Val^2/w;
low = Chop[
   t /. FindRoot[(T/Log[T])^(2/
           1.5) ((Total[MList[X, B, t, 1.5]]/T)^2/(Total[
             MList[X, B, t, 1.5]^2]/T)) == confi, {t, tr - 0.2}][[
     1]]];
up = Chop[
   t /. FindRoot[(T/Log[T])^(2/
           1.5) ((Total[MList[X, B, t, 1.5]]/T)^2/(Total[
             MList[X, B, t, 1.5]^2]/T)) == confi, {t, tr + 0.2}][[
     1]]];
```

Chapter 7

Tail Index Estimation

7.1 Introduction and preliminaries

The self-normalized method has been focused on in these two decades, and many interesting results are obtained. (See Griffin and Mason (1991), Klüppelberg and Mikosch (1996), Logan et al. (1973), Peña et al. (2009).) We center on this method since there are many random variables without sufficient order of moments for estimation. As an example, any random variable belongs to the domain of attraction of a stable law (not normal) has not well-defined variance. In Chapter , We give a general and explicit formula for the moments of the limiting distribution of symmetric self-normalized sum of i.i.d. random variables, which belong to the domain of attraction of a stable law. The result shows that the finite order moments for symmetric self-normalized sums are always finite. As an application, tail index can be estimated through our results by using moment estimators. In Section 7.1, we introduce the limiting distribution of the self-normalized sums with its characteristic function. In Section 7.2, we give the formula of the limiting moments of the symmetric self-normalized sums. The proof of the main result is given in Section 7.3. The numerical results and Mathematica code are given in Section 7.4.

Consider a sequence $\{X(i)\}_{i=1,\dots,n}$ that $X(i)$'s are assumed to be independent and identically distributed and belong to the domain of attraction of a stable law G, the parameter of attracting stable law G is denoted by α. More specifically, we assume that the density function g of the stable distribution G satisfies

$$x^{\alpha+1}g(x) \to r, \quad x^{\alpha+1}g(-x) \to l,$$

where $0 < \alpha < 2$, $r + l > 0$. Also U_n and V_n^2 are defined as

$$U_n = \frac{X(1) + \cdots + X(n)}{n^{1/\alpha}}$$

and

$$V_n^2 = \frac{|X(1)|^2 + \cdots + |X(n)|^2}{n^{2/\alpha}}.$$

To have the limiting distribution of $S_n(2)$ $(= U_n/V_n)$ exist, we further assume that

$$EX(i) = 0 \quad \text{if } 1 < \alpha < 2.$$

The limiting distribution of $S_n(2)$ is denoted by $S(2)$.

It is shown in Logan et al. (1973) that if $\alpha \neq 1$, the moments of $S(2)$ can be derived from

$$\frac{1}{\pi}\int_0^\infty \varphi(t)e^{-st}\,dt = \int_0^\infty e^{-s^2t^2/2}\mathcal{D}(t)\,dt, \tag{7.1.1}$$

where

$$\varphi(t) = Ee^{iS(2)t} = \lim_{n\to\infty} Ee^{iS_n(2)t},$$

the characteristic function of the limiting distribution of $S(2)$, and

$$\mathcal{D}(t) = (1-\alpha)(2\pi^{-3})^{1/2}\frac{rD_{\alpha-2}(-it) + lD_{\alpha-2}(it)}{rD_\alpha(-it) + lD_\alpha(it)},$$

$D_\nu(z)$ $(z \in \mathbb{C})$ is the parabolic cylinder functions. (See Magnus and Oberhettinger (1954).) Here are two important properties of parabolic cylinder functions for calculation.

$$\frac{d}{dz}D_\nu(z) - \frac{z}{2}D_\nu(z) + D_{\nu+1}(z) = 0; \tag{7.1.2}$$

$$\frac{d}{dz}D_\nu(z) + \frac{z}{2}D_\nu(z) - \nu D_{\nu-1}(z) = 0. \tag{7.1.3}$$

The calculation of the moments depends on the expansion of (7.1.1). We first decompose the left hand side into the form of power series.

$$\begin{aligned}
\frac{1}{\pi}\int_0^\infty \varphi(t)e^{-st}\,dt &= \frac{1}{\pi}\int_0^\infty \sum_{k=0}^\infty \frac{\varphi^{(k)}(0)}{k!}t^k e^{-st}\,dt \\
&= \sum_{k=0}^\infty \frac{\varphi^{(k)}(0)}{k!}\frac{1}{\pi}\int_0^\infty t^k e^{-st}\,dt \\
&= \sum_{k=0}^\infty \frac{1}{\pi}\varphi^{(k)}(0)s^{-k-1}.
\end{aligned} \tag{7.1.4}$$

Secondly, the right hand side of (7.1.1) can be written as follows.

$$\begin{aligned}
\int_0^\infty e^{-s^2t^2/2}\mathcal{D}(t)dt &= \sum_{k=0}^\infty \frac{\mathcal{D}^{(k)}(0)}{k!}\int_0^\infty e^{-s^2t^2/2}t^k\,dt \\
&= \sum_{k=0}^\infty \frac{\mathcal{D}^{(k)}(0)}{k!}\int_0^\infty e^{-u}\left(\frac{2u}{s^2}\right)^{\frac{k}{2}} \cdot \frac{1}{s\sqrt{2u}}\,du
\end{aligned} \tag{7.1.5}$$

$$\text{(change the variable } u = \frac{s^2t^2}{2})$$

$$= \sum_{k=0}^{\infty} \frac{\mathcal{D}^{(k)}(0)}{k!} 2^{\frac{k-1}{2}} \Gamma(\frac{k+1}{2}) s^{-k-1}. \tag{7.1.6}$$

Equating coefficients of like powers of s^{-1} in (7.1.4) and (7.1.5), we can see that

$$E(S(2)^k) = i^k \varphi^{(k)}(0) = \frac{\mathcal{D}^{(k)}(0)}{k!} 2^{\frac{k-1}{2}} \Gamma(\frac{k+1}{2})\pi. \tag{7.1.7}$$

The main purpose of this paper is to derive a general and explicit formula to calculate the moments of the distribution $S(2)$ when $r = l$.

7.2 Asymptotic Moments of Self-normalized Sums

We assume $X(i)$'s are symmetric, i.e., $r = l$, then $\mathcal{D}(t)$ becomes

$$\mathcal{D}(t) = (1-\alpha)(2\pi^{-3})^{1/2} \frac{D_{\alpha-2}(-it) + D_{\alpha-2}(it)}{D_\alpha(-it) + D_\alpha(it)} \equiv (1-\alpha)(2\pi^{-3})^{1/2}\mathcal{A}(t).$$

For symmetry case, we simplify the notation of the limiting distribution $S(2)$ by S.

Theorem 7.2.1. *Let S be defined above. Then for any $m = 1, 2, \ldots$, $E(S^{2m-1}) = 0$ and*

$$E(S^{2m}) = \frac{(2m-1)!!}{(2m)!} 2\{(D_{\alpha-2}^{(2m)}(0) - D_\alpha^{(2m)}(0)) \\ - (1-\alpha) \sum_{k=0}^{m-1} \frac{(-1)^{m-k}\mathcal{A}^{(2k)}(0)}{(2(m-k))!!(2k)!} D_\alpha^{2(m-k)}(0)\},$$

where $\mathcal{A}^{(2k)}(0)$ satisfies $\mathcal{A}(0) = D_{\alpha-2}(0)/D_\alpha(0)$ and

$$\mathcal{A}^{(2k)}(0) = \frac{(-1)^k}{1-\alpha} 2(D_{\alpha-2}^{(2k)}(0) - D_\alpha^{(2k)}(0)) \\ - 2\sum_{l=0}^{k-1} \frac{\mathcal{A}^{(2l)}(0)}{(2(k-l))!!(2l)!} (-1)^{(k-l)} D_\alpha^{2(k-l)}(0)\}.$$

Furthermore, suppose

$$D_\nu^{(2k)}(0) = \eta^0(k) + \sum_{j=1}^{k} \eta^j(k)\nu^j,$$

then $\eta^j(k)$ *satisfies*

$$\begin{cases} \eta^j(k) = -\sum_{t=0}^{k-j} \binom{k-t}{j} \nu^{k-t}(k-1) + \frac{2k-1}{2}\eta^j(k-1) \quad \text{for } j \geq 0; \\ \nu^1(k) = k! \left((-1)^k + \sum_{l=1}^{k} \frac{(2l-1)!!(-1)^{k-l}}{2^l l!} \right); \\ \nu^j(k) = -\sum_{l=1}^{k-j+2} \frac{(2k-1)!!}{(2k-(2l-1))!!2^{l-1}} \nu^{j-1}(k-l) - k\nu^j(k-1) \quad \text{for } j \geq 2; \\ \nu^1(0) = 1,\ \nu^j(0) = 0 \ \text{ for } j \geq 2; \quad \nu^0(k) = 0 \ \text{ for any } k \geq 0; \\ \eta^j(k) = 0 \ \text{ for } j > k; \quad \eta^k(k) = (-1)^k \ \text{ for any } k \geq 0. \end{cases}$$

Corollary 7.2.2. The finite order moments for self-normalized sum of i.i.d. random variables in the domain of attraction of a stable law are always finite.

7.3 Proof of Theorem 7.2.1

We show Theorem 7.2.1 in this section. Set

$$A_\nu(t) = D_\nu(-it) + D_\nu(it),$$

then we obtain

$$\mathcal{A}(t) = \frac{D_{\alpha-2}(-it) + D_{\alpha-2}(it)}{D_\alpha(-it) + D_\alpha(it)} = \frac{A_{\alpha-2}(t)}{A_\alpha(t)}.$$

From (7.1.7), the moment of the limiting distribution can be simply written as

$$E(S^k) = \frac{(k-1)!!}{k!} i^k (1-\alpha)\mathcal{A}^{(k)}(0). \tag{7.3.1}$$

Note that $\frac{d}{dt}D_\nu(-it) = -i\frac{d}{dz}D_\nu(z)\Big|_{z=-it}$ and $\frac{d}{dt}D_\nu(it) = i\frac{d}{dz}D_\nu(z)\Big|_{z=it}$, it is obvious that

$$A_k(\nu) \equiv \frac{d^k}{dt^k}A_\nu(t)\Big|_{t=0} = \begin{cases} 0 & \text{if k is odd;} \\ (-1)^{k/2}2D_\nu^{(k)}(0), & \text{if k is even,} \end{cases}$$

where $D_\nu^{(k)}(0) = \frac{d^k}{dz^k}D_\nu(z)\Big|_{z=0}$. To prove the first statement, we use a recursive formula for nth derivative.

Lemma 7.3.1 (Xenophontos (2007))**.**

$$\left(\frac{u(x)}{v(x)}\right)^{(n)} = \frac{1}{v(x)} \left(u^{(n)}(x) - n! \sum_{j=1}^{n} \frac{v(x)^{(n+1-j)}}{(n+1-j)!(j-1)!} \left(\frac{u(x)}{v(x)}\right)^{(j-1)} \right).$$

Applying the formula to the case that $u(x) = A_{\alpha-2}(x)$ and $v(x) = A_\alpha(x)$, we have

$$\mathcal{A}^{(2k)}(0) = \frac{1}{1-\alpha}(A_{2k}(\alpha-2) - A_{2k}(\alpha)) - \sum_{l=0}^{k-1}\binom{2k}{2l}A_{2(k-l)}(\alpha)\mathcal{A}^{(2l)}(0).$$

The first result is straightforward from (7.3.1).

Next, we show the second half of Theorem 7.2.1. Differentiating (7.1.2) and (7.1.3) iteratively, we have

$$\begin{aligned} D_\nu^{(k)}(z) - \frac{z}{2}D_\nu^{(k-1)}(z) - \frac{k-1}{2}D_\nu^{(k-2)}(z) + D_{\nu+1}^{(k-1)}(z) = 0; \\ D_\nu^{(k)}(z) + \frac{z}{2}D_\nu^{(k-1)}(z) + \frac{k-1}{2}D_\nu^{(k-2)}(z) - \nu D_{\nu-1}^{(k-1)}(z) = 0. \end{aligned}$$

Thus $D_\nu^{(k)}(0)$ can be derived from

$$\begin{aligned} D_\nu^{(k)}(0) &= \frac{k-1}{2}D_\nu^{(k-2)}(0) - D_{\nu+1}^{(k-1)}(0); \\ D_\nu^{(k)}(0) &= -\frac{k-1}{2}D_\nu^{(k-2)}(0) + \nu D_{\nu-1}^{(k-1)}(0). \end{aligned}$$

In the case when k is odd, rewrite $2k+1$ as k, then

$$\begin{aligned} D_\nu^{(2k+1)}(0) &= -kD_\nu^{(2k-1)}(0) + \nu(\frac{2k-1}{2}D_{\nu-1}^{(2k-2)}(0) - D_\nu^{(2k-1)}(0)) \\ &= -kD_\nu^{(2k-1)}(0) \\ &\quad -\nu\sum_{l=1}^{k}\frac{(2k-1)!!}{(2k-2l+1)!!2^{l-1}}D_\nu^{(2k-2l+1)}(0) + \nu\frac{(2k-1)!!}{2^k}D_{\nu-1}(0). \end{aligned}$$

This is a recurrence formula for $D_\nu^{(2k+1)}(0)$. If we can expand it, then it must be the product of a polynomial of ν and $D_{\nu-1}(0)$. Let $\nu^j(k)$ denote the coefficient of ν^j in the case of $(2k+1)$th derivative.

For the initial values, we can see that $\nu^1(0) = 1$, $\nu^j(0) = 0$ for all $j \geq 2$ and $\nu^0(k) = 0$ for all $k \geq 0$ from (7.1.3). After some painful calculation, we have

$$\begin{aligned} \nu^1(k) &= k!\left((-1)^k + \sum_{l=1}^{k}\frac{(2l-1)!!(-1)^{k-l}}{2^l l!}\right); \\ \nu^j(k) &= -\sum_{l=1}^{k-j+2}\frac{(2k-1)!!}{(2k-(2l-1))!!2^{l-1}}\nu^{j-1}(k-l) - k\nu^j(k-1) \quad \text{for } j \geq 2. \end{aligned} \tag{7.3.2}$$

From the recurrence formula, one can see that the highest degree of the polynomial is $k+1$, which can be shown by the induction. Using this property reversely, one also can see that $\nu^j(k) = 0$ for any j and k satisfying $j \geq k+2$.

Corollary 7.3.2.

$$\nu^{k+1}(k) = (-1)^k, \quad \nu^k(k) = (-1)^k \frac{k}{2}. \tag{7.3.3}$$

Proof. Applying this result to (7.3.2),

$$\nu^{k+1}(k) = -\nu^k(k-1)$$

holds, and since the initial value $\nu^1(0) = 1$, we have

$$\nu^{k+1}(k) = (-1)^k. \tag{7.3.4}$$

Also applying the result to $\nu^k(k)$,

$$\nu^k(k) = -\nu^{k-1}(k-1) - \frac{1}{2},$$

which implies

$$\nu^k(k) = (-1)^k \frac{k}{2},$$

since $\nu^0(0) = 0$. $\square$

On the other hand, when k is even, rewrite $2k+2$ as k and we have

$$D_\nu^{(2k+2)}(0) \quad = \quad \frac{2k+1}{2} D_\nu^{(2k)}(0) - D_{\nu+1}^{(2k+1)}(0).$$

Here, let $\eta^j(k)$ denote the coefficient of ν^j in the case of $2k$th derivative. Then we have

$$\eta^j(k) = -\sum_{t=0}^{k-j} \binom{k-t}{j} \nu^{k-t}(k-1) + \frac{2k-1}{2} \eta^j(k-1) \quad \text{for } j \geq 0.$$

From (7.3.4), $\eta^j(k) = 0$ if $j > k$ and $\eta^k(k) = -\nu^k(k-1) = (-1)^k$. $\square$

7.4 Examples

7.4.1 Mathematica code

This section provides Mathematica code. The functions $f(j,k)$ and $g(j,k)$ denote the function $\nu^j(k)$ and $\eta^j(k)$ in the previous section, respectively. The function $A(n,a)$ is corresponding to $1/2\ A_{2n}(\alpha)$, while $CA(n,a)$ represents the function $\mathcal{A}^{(2n)}(0)$ above. Lastly, function $M(n,a)$ indicates the $2n$th moment of the limit distribution S.

```
f[1, k_] :=
  k! ((-1)^k + Sum[(2 l - 1)!! (-1)^(k - l)/(2^l l!), {l, 1, k}]);
f[j_, k_] :=
  If[j == 0, 0,
   If[k <= -1, 0,
    If[k > -1 && j >= 2,
      -(2 k - 1)!! Sum[
        f[j - 1, k - l]/((2 k - (2 l - 1))!! 2^(l - 1)),
        {l, 1, k - j + 2}] - k f[j, k - 1]]]];
g[j_, k_] :=
  If[j > k, 0,
   If[j == k, (-1)^k,
    If[k > 0, -Sum[
        Binomial[k - t, j] f[k - t, k - 1], {t, 0, k - j}]
        + (2 k - 1) g[j, k - 1]/2, 0]]];
A[n_, a_] := (-1)^n g[0, n] + (-1)^n Sum[g[t, n] a^t, {t, 1, n}];
CA[n_, a_] :=
  If[n > 0, (A[n, a - 2] - A[n, a])/(1 - a) - Sum[CA[t, a]
      Binomial[2 n, 2 t] A[n - t, a], {t, 1, n - 1}], 0];
M[n_, a_] :=
  Simplify[(2 n - 1)!!/(2 n)! (-1)^n (1 - a) CA[n, a], a > 0];
```

7.4.2 Some results and knowledge

Using the code above, we obtain the general result for the moments of symmetric self-normalized moments and some special cases of $\alpha = 0.5$, $\alpha = 1.5$ and $\alpha = 2$.

TABLE 7.1: $2k$th moments of symmetric self-normalized sum for the case of $\alpha = 0.5, 1.5, 2$.

k	$E(S^{2k})$	$\alpha = 0.5$	$\alpha = 1.5$	$\alpha = 2$
1	1	1	1	1
2	$1 + \alpha$	1.5	2.5	3
3	$1 + 3\alpha + 2\alpha^2$	3	10	15
4	$1/3(3 + 20\alpha + 34\alpha^2 + 17\alpha^3)$	7.875	55.625	105
5	$1/3(3 + 40\alpha + 130\alpha^2 + 155\alpha^3 + 62\alpha^4)$	26.25	397.5	945

When $\alpha = 2$, the limiting distribution S is standard normal distribution. From Table 7.1, we can see the result is corresponding to the moments we can obtain from other methods.

7.4.3 Tail index estimation

Hill's estimator is proposed to be an estimator for tail index. It is defined as

$$\hat{\alpha}_H = \Big(\frac{1}{k}\sum_{j=1}^{k}\log X_{n,n-j+1} - \log X_{n,n-k}\Big)^{-1},$$

where $X_{n,1} \leq \cdots \leq X_{n,n}$ are the order statistics of $X(1), \ldots, X(n)$. (See Hall (1982).) As an alternative to it, we can apply the result above to the derivation of the tail index of the random variables in the domain of stable distribution. This is achieved by moment estimators after calculating the asymptotic moments for the self-normalized sums.

As an example, we use the fourth moment estimator of the self-normalized sums, which is denoted by

$$\hat{\alpha} = \frac{1}{K}\sum_{i=1}^{K} S_i^4 - 1, \tag{7.4.1}$$

where K is the number of blocks, for the self-normalized sums, of the original samples.

We compare the performance of (7.4.1) with Hill's estimator by means of Monte Carlo experiments. All numerical results in this paper are based on 250 simulations. The sample sizes are 200 and 2000. The latter size is typical for current financial data, and the former is relatively small for the observation of the behaviors of the estimates in small samples. In each case, we evaluate the performance of the estimators on the basis of i.i.d. random variables and dependent ones. The number in the parentheses after the name of the distribution is the tail index of random variables.

For dependent case, we follow the examples in Danielsson et al. (2001). The MA process $Y(t) = X(t) + X(t-1)$, where the $X(t)$ are i.i.d. stable with the tail index α is considered. The other stochastic process $Y(t)$, stochastic volatility, is defined as follows:

$$\begin{aligned}
&Y(t) = U(t)X(t)H(t),\\
&U(t) \quad \text{i.i.d. discrete uniform on } -1 \text{ and } 1,\\
&X(t) = \sqrt{57/Z(t)}, \quad Z(t) \sim \text{i.i.d. } \chi_1^2,\\
&H(t) = 0.1Q(t) + 0.9H(t-1), \quad Q(t) \sim \text{i.i.d. } \mathcal{N}(0,1).
\end{aligned}$$

This process is denoted by S.V.(1). Its marginal distribution has a Student-t with 1 degree of freedom.

For Hill's estimator, it is known that $k = O(n^{2/3})$ is optimal (Hall (1982), Resnick and Stărică (1998)). However, k is not specified since it is sensitive to the sample size and the assumed model. For simplicity, we use $k = \lceil n^{2/3} \rceil$ in all simulations, where $\lceil \cdot \rceil$ is the ceiling function. On the other hand, let T be the number of samples for a self-normalized sum. To guarantee K is large enough, we use bootstrap samples for the self-normalized sums. For each distribution, we report the true value of α, the optimal T for the estimation of α, the mean and the root mean squared error (RMSE) of each estimator in the case that $K = 5n$.

TABLE 7.2: Monte Carlo experiment with $n = 200$.

Distribution	α	T	Mean ($\hat{\alpha}_H$)	Mean ($\hat{\alpha}$)	RMSE ($\hat{\alpha}_H$)	RMSE ($\hat{\alpha}$)
Stable(0.5)	0.5	8	0.468	0.506	0.082	0.092
Stable(1)	1	10	0.988	0.996	0.169	0.186
Stable(1.5)	1.5	12	1.690	1.475	0.366	0.334
Stable(2)	2	16	2.574	1.995	0.710	0.516
Student(0.5)	0.5	5	0.527	0.493	0.093	0.082
Student(1.5)	1.5	19	1.290	1.493	0.296	0.366
MA(0.5)	0.5	6	0.489	0.511	0.014	0.010
MA(1)	1	8	1.019	1.007	0.045	0.053
MA(1.5)	1.5	9	1.690	1.492	0.172	0.200
MA(2)	2	12	2.631	2.016	0.608	0.433
S.V.(1)	1	5	1.617	1.020	0.461	0.021

From Table 7.2, the moment estimator of the self-normalized sums can be sufficiently accurate even in the small sample cases. The RMSEs of both Hill's estimator and the moment estimator become larger as the tail index increases. The difference between two estimators is that the RMSE of Hill's estimator becomes large more sharply than that of the moment estimator as the tail index grows larger than 1 in the stable case.

TABLE 7.3: Monte Carlo experiment with $n = 2000$.

Distribution	α	T	Mean ($\hat{\alpha}_H$)	Mean ($\hat{\alpha}$)	RMSE ($\hat{\alpha}_H$)	RMSE ($\hat{\alpha}$)
Stable(0.5)	0.5	23	0.484	0.500	0.040	0.037
Stable(1)	1	34	0.999	1.001	0.079	0.078
Stable(1.5)	1.5	43	1.753	1.505	0.292	0.148
Stable(2)	2	46	3.877	2.005	1.898	0.170
Student(0.5)	0.5	11	0.502	0.502	0.039	0.030
Student(1.5)	1.5	79	1.414	1.505	0.139	0.208
MA(0.5)	0.5	17	0.485	0.502	0.003	0.002
MA(1)	1	23	1.006	1.002	0.010	0.012
MA(1.5)	1.5	28	1.792	1.500	0.129	0.024
MA(2)	2	35	3.919	1.996	3.767	0.055
S.V.(1)	1	5	1.900	0.969	0.836	0.003

As shown in Table 7.3, we can see that the moment estimators attain to the true tail index if we choose a proper T for each case. The RMSEs of two estimators are lower for the larger sample size, while the comparison between two estimators is almost similar to the sample size $n = 200$. However, we also find that the behavior of the moment estimators in the student case and the S.V. case are a little different from the stable case and MA case from both tables. The representation of the tail of t-distribution is more complicate than that of stable, since there is a second term in the representation of the former. Nevertheless, the moment estimator of self-normalized sums performs well in the estimation of the tail index.

Bibliography

[1] Albrecht, V. (1984). On the convergence rate of probability of error in Bayesian discrimination between two Gaussian processes. in *Proc. of the Third Prague symposium on asymptotic statistics.*

[2] Anderson, T. (1971). *The Statistical Analysis of Time Series.* J. Wiley and sons.

[3] Andrews, D. W. and Pollard, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review/Revue Internationale de Statistique* **62** 119–132.

[4] Bartkiewicz, K., Jakubowski, A., Mikosch, T., and Wintenberger, O. (2011). Stable limits for sums of dependent infinite variance random variables. *Probability theory and related fields* **150** 337–372.

[5] Basawa, I. (1985). Neyman-Le Cam tests based on estimating functions. in *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer* **2** 811–825.

[6] Basu, A., Harris, I. R., Hjort, N. L., and Jones, M. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika* **85** 549–559.

[7] Beran, J. (1994). On a class of M-estimators for Gaussian long-memory models. *Biometrika* **81** 755–766.

[8] Beran, R. (1978). An efficient and robust adaptive estimator of location. *The Annals of Statistics* **6** 292–313, March.

[9] Bernstein, S. (1927). Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Mathematische Annalen* **97** 1–59.

[10] Bhansali, R. J. (1980). Autoregressive and window estimates of the inverse correlation function. *Biometrika* **67** 551–566.

[11] Bickel, P., Klaassen, C., Ya'acov Ritov, J., and Wellner, J. (1998). *Efficient and Adaptive Estimation for Semiparametric Models.* **1** Springer New York.

[12] Billingsley, P. (1968). *Convergence of Probability Measures.* **493** Wiley-Interscience.

[13] Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* **60** 217–226.

[14] Bose, A. (1998). Bahadur representation of M_m estimates. *The Annals of Statistics* **26** 771–777.

[15] Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory.* **36** Siam.

[16] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Springer-Verlag.

[17] Can, S., Mikosch, T., and Samorodnitsky, G. (2010). Weak convergence of the function-indexed integrated periodogram for infinite variance processes. *Bernoulli* **16** 995–1015.

[18] Cleveland, W. S. (1972). The inverse autocorrelations of a time series and their applications. *Technometrics* **14** 277–293.

[19] Danielsson, J., de Haan, L., Peng, L., and de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate analysis* **76** 226–248.

[20] Davis, R. and Resnick, S. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *The Annals of Probability* **13** 179–195.

[21] Davis, R. and Resnick, S. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *The Annals of Statistics* **14** 533–558.

[22] Davis, R., Marengo, J., and Resnick, S. (1986). Extremal properties of a class of multivariate moving averages. *Department of Statistics, Colorado State University.*

[23] Davis, R. and Hsing, T. (1995). Point process and partial sum convergence for weakly dependent random variables with infinite variance. *The Annals of Probability* **23** 879–917.

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

[25] Denker, M. and Keller, G. (1983). On U-statistics and v. mises' statistics for weakly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **64** 505–522.

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

[27] Dunsmuir, W. (1979). A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise. *The Annals of Statistics* **7** 490–506.

[28] Dzhaparidze, K. O. (1971). On methods for obtaining asymptotically efficient spectral parameter estimates for a stationary Gaussian process with rational spectral density. *Theory of Probability & Its Applications* **16** 550–554.

[29] Dzhaparidze, K. (1986). *Parameter Estimation and Hypotheses Testing in Spectral Analysis of Stationary Time Series.* Springer.

[30] Embrechts, P. and Goldie, C. M. (1980). On closure and factorization properties of subexponential and related distributions. *Journal of the Australian Mathematical Society (Series A)* **29** 243–256.

[31] Fabian, V. and Hannan, J. (1982). On estimation and adaptive estimation for locally asymptotically normal families. *Probability Theory and Related Fields* **59** 459–478.

[32] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications.* **2** John Wiley & Sons.

[33] Fujisawa, H. and Eguchi, S. (2008). Robust parameter estimation with a small bias against heavy contamination. *Journal of Multivariate Analysis* **99** 2053–2081.

[34] Fuller, W. A. (1996). *Introduction to Statistical Time Series.* **230** John Wiley & Sons.

[35] Garel, B. and Hallin, M. (1995). Local asymptotic normality of multivariate ARMA processes with a linear trend. *Annals of the Institute of Statistical Mathematics* **47** 551–579.

[36] Geyer, C. J. (1996). On the asymptotics of convex stochastic optimization. *Unpublished manuscript* **37**.

[37] Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics* **31** 1208–1211, December.

[38] Godambe, V. P. and Thompson, M. E. (1978). Some aspects of the theory of estimating equations. *Journal of Statistical Planning and Inference* **2** 95–104.

[39] Godambe, V. P. and Thompson, M. E. (1984). Robust estimation through estimating equations. *Biometrika* **71** 115–125, April.

[40] Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series.* Johm Wiley & Sons, New York.

[41] Griffin, P. S. and Mason, D. M. (1991). On the asymptotic normality of self-normalized sums. in *Mathematical Proceedings of the Cambridge Philosophical Society* **109** 597–610, Cambridge Univ Press.

[42] Hagemann, A. (2011). Robust spectral analysis. *Unpublished manuscipt, Department of Exonomics, University of Illinois, arXiv preprint arXiv:1111.1965.*

[43] Hagemann, A. (2014). Stochastic equicontinuity in nonlinear time series models. *The Econometrics Journal* **17** 188–196.

[44] Hájek, J., Šidák, Z., and Sen, P. (1967). *Theory of Rank Tests.* Academic press New York.

[45] Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society. Series B (Methodological)* 37–42.

[46] Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic press New York.

[47] Hallin, M., Ingenbleek, J., and Puri, M. (1985). Linear serial rank tests for randomness against ARMA alternatives. *The Annals of Statistics* **13** 1156–1181.

[48] Hallin, M., Ingenbleek, J., and Puri, M. (1987). Linear and quadratic serial rank tests for randomness against serial dependence. *Journal of time series analysis* **8** 409–424.

[49] Hallin, M. and Puri, M. (1991). Rank tests for time series analysis a survey.

[50] Hallin, M. and Werker, B. (1999). Optimal testing for semi-parametric AR models-from Gaussian lagrange multipliers to autoregression rank scores and adaptive tests. *statistics textbooks and monographs* **158** 295–350.

[51] Hallin, M. and Werker, B. (2003). Semi-parametric efficiency, distribution-freeness and invariance. *Bernoulli* **9** 137–165.

[52] Hamilton, J. D. (1994). *Time Series Analysis.* **2** Princeton university press Princeton.

[53] Hannan, E. J. (1973a). The asymptotic theory of linear time-series models. *Journal of Applied Probability* **10** 130–145.

[54] Hannan, E. J. (1973b). The estimation of frequency. *Journal of Applied probability* **10** 510–519.

[55] Hannan, E. (1970). *Multiple Time Series.* John Wiley & Sons.

[56] Hewitt, E. and Stromberg, K. (1975). *Real and Abstract Analysis. A Modern Treatment of the Theory of Functions of a Real Variable. Third printing.* **25** Splinger-Verlag.

[57] Hjort, N. and Pollard, D. (1993). Asymptotics for minimisers of convex processes. *Unpublished manuscript.*

[58] Hodges, J. L. and Lehmann, E. L. (1963). Estimates of location based on rank tests. *The Annals of Mathematical Statistics* **34** 598–611.

[59] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* **19** 293–325.

[60] Hosoya, Y. (1989). The bracketing condition for limit theorems on stationary linear processes. *The Annals of Statistics* **17** 401–418.

[61] Hosoya, Y. and Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *The Annals of Statistics* **10** 132–153.

[62] Hsing, T. and Wu, W. B. (2004). On weighted U-statistics for stationary processes. *Annals of Probability* **32** 1600–1631.

[63] Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics* **35** 73–101.

[64] Huber, P. (2009). *Robust Statistics*. Wiley, New York.

[65] Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.

[66] Inagaki, N. and Kondo, M. (1980). Several estimators of the autocorrelation based on limiter estimating functions for a stationary Gaussian process. *Journal of the Japan Statistical Society* **10** 1–15.

[67] Kanamori, T. and Fujisawa, H. (2014). Affine invariant divergences associated with proper composite scoring rules and their applications. *Bernoulli* **20** 2278–2304.

[68] Kholevo, A. (1969). On estimates of regression coefficients. *Theory of Probability & Its Applications* **14** 79–104.

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

[70] Kley, T., Volgushev, S., Dette, H., and Hallin, M. (2014). Quantile spectral processes: Asymptotic analysis and inference. *arXiv preprint arXiv:1401.8104*.

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

[72] Klüppelberg, C. and Mikosch, T. (1994). Some limit theory for the self-normalized periodogram of stable processes. *Scandinavian Journal of Statistics* **21** 485-491.

[73] Klüppelberg, C. and Mikosch, T. (1996). The integrated periodogram for stable processes. *The Annals of Statistics* **24** 1855–1879.

[74] Knight, K. (1998). Limiting distributions for l_1 regression estimators under general conditions. *The Annals of Statistics* **26** 755–770.

[75] Koenker, R. (2005). *Quantile Regression*. Cambridge university press.

[76] Koenker, R. and Bassett, J. G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society* **46** 33–50.

[77] Kreiss, J. (1987). On adaptive estimation in stationary ARMA processes. *The Annals of Statistics* **15** 112–133.

[78] Lehmann, E. and Romano, J. (2005). *Testing Statistical Hypotheses*. Springer.

[79] Li, T.-H. (2008). Laplace periodogram for time series analysis. *Journal of the American Statistical Association* **103** 757–768.

[80] Li, T.-H. (2012). Quantile periodograms. *Journal of the American Statistical Association* **107** 765–776.

[81] Liu, Y. (2013a). Asymptotic moments of symmetric self-normalized sums. *Scientiae Mathematicae Japonicae Online* **26** 561–569.

[82] Liu, Y. (2013b). Nonparametric methods in time series analysis. *Master's thesis* 1–101.

[83] Liu, Y. (2014). Asymptotics for M-estimators in time series. *Advances in Science, Technology and Environmentology* **B10** 63–77.

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

[85] Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Cambridge Univ Press.

[86] Magnus, W. and Oberhettinger, F. (1954). *Formulas and Theorems for the Functions of Mathematical Physics*. Chelsea Pub. Co.

[87] Meerschaert, M. and Scheffler, H. (2001). *Limit Distributions for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice*. **321** Wiley-Interscience.

[88] Mikosch, R. S., T. and Samorodnitsky, G. (2000). The maximum of the periodogram for a heavy-tailed sequence. *The Annals of Probability* **28** 885–908.

[89] Mikosch, T., Gadrich, T., Kluppelberg, C., and Adler, R. (1995). Parameter estimation for ARMA models with infinite variance innovations. *The Annals of Statistics* **23** 305–326.

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

[91] Monti, A. C. and Ronchetti, E. (1993). On the relationship between empirical likelihood and empirical saddlepoint approximation for multivariate M-estimators. *Biometrika* **80** 329–338.

[92] Niemiro, W. (1992). Asymptotics for M-estimators defined by convex minimization. *The Annals of Statistics* **20** 1514–1533.

[93] Nolan, J. (2012). *Stable Distributions - Models for Heavy Tailed Data*. Birkhauser.

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

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

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

[97] Peña, V. H., Lai, T. L., and Shao, Q.-M. (2009). *Self-normalized processes: Limit theory and Statistical Applications*. Springer.

[98] Petrov, V. V. (1975). *Sums of Independent Random Variables*. Springer, Berlin.

[99] Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7** 186–199.

[100] Quinn, B. G. and Thomson, P. J. (1991). Estimating the frequency of a periodic function. *Biometrika* **78** 65.

[101] Quinn, B. G. and Hannan, E. J. (2001). *The Estimation and Tracking of Frequency*. **9** Cambridge University Press.

[102] Resnick, S. (2007). *Extreme Values, Regular Variation, and Point Processes*. Springer.

[103] Resnick, S. and Stărică, C. (1998). Tail index estimation for dependent data. *The Annals of Applied Probability* **8** 1156–1183.

[104] Rice, J. A. and Rosenblatt, M. (1988). On frequency estimation. *Biometrika* **75** 477–484, September.

[105] Rozanov, Y. A. (1969). On a new class of statistical estimates. *Soviet-Japanese Symposium, Theory of Probability* 239–252.

[106] Rozanov, Y. A. (1972). Some approximation problems in the theory of stationary processes. *Journal of Multivariate Analysis* **2** 135–144.

[107] Samoradnitsky, G. and Taqqu, M. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall/CRC.

[108] Sen, P. K. (1972). Limiting behavior of regular functionals of empirical distributions for stationary*-mixing processes. *Probability Theory and Related Fields* **25** 71–82.

[109] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* **162** Wiley. com.

[110] Skowronek, S., Volgushev, S., Kley, T., Dette, H., and Hallin, M. (2014). Quantile spectral analysis for locally stationary time series. *arXiv preprint arXiv:1404.4605.*

[111] Strang, G. (2005). *Linear Algebra and Its Applications.* Cengage Learning; 4th edition.

[112] Swensen, A. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *Journal of Multivariate Analysis* **16** 54–70.

[113] Taniguchi, M. and Hirukawa, J. (2011). Generalized information criterion. *Journal of Time Series Analysis* **33** 287–297.

[114] Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series.* New York: Springer-Verlag.

[115] Taniguchi, M. (1979). On estimation of parameters of Gaussian stationary processes. *Journal of Applied Probability* **16** 575–591.

[116] Taniguchi, M. (1980). On estimation of the integrals of certain functions of spectral density. *Journal of Applied Probability* **17** 73–83.

[117] Taniguchi, M. (1981a). An estimation procedure of parameters of a certain spectral density model. *Journal of the Royal Statistical Society. Series B* **43** 34–40.

[118] Taniguchi, M. (1981b). Robust regression and interpolation for time series. *Journal of Time Series Analysis* **2** 53–62.

[119] Taniguchi, M. (1987). Minimum contrast estimation for spectral densities of stationary processes. *Journal of the Royal Statistical Society. Series B* **49** 315–325.

[120] Taniguchi, M., van Garderen, K. J., and Puri, M. L. (2003). Higher order asymptotic theory for minimum contrast estimators of spectral parameters of stationary processes. *Econometric theory* **19** 984–1007.

[121] van der Vaart, A. W. (2000). *Asymptotic Statistics.* **3** Cambridge university press.

[122] van der Vaart and Wellner, J. A. (1996). *Weak Convergence.* Springer.

[123] Walker, A. M. (1971). On the estimation of a harmonic component in a time series with stationary independent residuals. *Biometrika* **58** 21.

[124] Whittle, P. (1952). The simultaneous estimation of a time series harmonic components and covariance structure. *Trabajos de estadística* **3** 43–57.

[125] Whittle, P. (1952). Some results in time series analysis. *Scandinavian Actuarial Journal* **1952** 48–60.

[126] Xenophontos, C. (2007). A formula for the nth derivative of the quotient of two functions.

[127] Yoshihara, K.-i. (1976). Limiting behavior of U-statistics for stationary, absolutely regular processes. *Probability Theory and Related Fields* **35** 237–252.

[128] Zhang, G. and Taniguchi, M. (1995). Nonparametric approach for discriminant analysis in time series. *Journaltitle of Nonparametric Statistics* **5** 91–101.

List of Papers

Main Papers

1. Liu, Y. (2013) "Asymptotic moments of symmetric self-normalized sums", *Scienticae Mathematicae Japanicae Online*, Vol. 26, pp. 561-569.
2. Liu, Y. (2014) "Asymptotics for M-estimators in time series", *Advances in Science, Technology and Environmentoloty*, Vol. B10, pp. 63-77.

Sub-paper

1. Akashi, F., Liu, Y., and Taniguchi, M. (2014) "An empirical likelihood approach for symmetric α-stable processes", *to appear in Bernoulli*

   Contributions: Section 4, Part of Section 5, Supplement to "An empirical likelihood approach for symmetric α-stable processes."