Asymptotic Theory of Statistical Inference for Binary Time Series and Count Time Series and Its Applications

二値時系列及び計数時系列に対する統計的漸近理論の構築とその応用

February, 2021

Yuichi GOTO
後藤　佑一

Asymptotic Theory of Statistical Inference for Binary Time Series and Count Time Series and Its Applications

二値時系列及び計数時系列に対する統計的漸近理論の構築とその応用

February, 2021

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

Yuichi GOTO
後藤　佑一

Acknowledgements

First of all, I would like to my most profound appreciation to my supervisor Professor Masanobu Taniguchi of Waseda University for his guidance and strong encouragement. He gives me many valuable opportunities. The right opportunities to attend his weekly seminars for bachelor and master students when I was a senior student of the School of Education at Waseda University and to attend his weekly seminars for doctor students when I was a master student built my foundations of knowledge for statistics. The chances to make presentations at international symposiums organized by him and interact with the foremost statisticians. A golden opportunity to write a paper with Professor Marc Hallin at of Université Libre de Bruxelles broadened my horizons. Needless to say, I could not graduate from the graduate school of Waseda University in three years without his leading and persistent help. Again, I deeply grateful to him.

I would like to my sincere gratitude to Professor Yasutaka Shimizu and Professor Yoichi Nishiyama of Waseda University for their instructive comments on the earlier draft of this dissertation.

I am deeply indebted to Professor Marc Hallin of Université Libre de Bruxelles for his instructions on how to write an excellent academic paper and encouragement. His kindness and gentleness help me a lot to live abroad for the first time my life. He gave me the precious opportunities to discuss with Professor Davy Paindaveine, Professor Thomas Verdebout, and Professor David Preinerstorfer of Université Libre de Bruxelles and write a paper with Professor Holger Dette of Ruhr-Universitäet Bochum, Professor Stanislav Volgushev of the University of Toronto, and Dr. Tobias Kley of Georg-August-Universität Göttingen.

I would like to express my gratitude to Professor Stanislav Volgushev of the University of Toronto for teaching how to prove the validity of a uniform confidence band.

I would like to show my appreciation to Dr. Tobias Kley of Georg-August-Universität Göttingen for teaching how to write a right code on R.

I would like to thank Doctor Yan Liu of Waseda University for his insightful suggestions of research topics and papers and being concerned about me. I learned the importance of the amount of knowledge from him.

I am grateful to Doctor Fumiya Akashi of the University of Tokyo for his helpful advice. He is always kind and listens to my concerns. I learned lots of things including how to manage tasks and feelings from him.

I would also like to extend my gratitude to Doctor Hideaki Nagahata for his practical advice on making a good presentation and Doctor Kou Fujimori of Shinshu University for his comments on the functional spaces, the spatial statistics, and practical suggestions on my papers.

I am also grateful to Professor Haruyoshi Mita of Waseda University, who supervised me when I was a bachelor student. He strongly advised me to learn the measure theory and probability theory. Without his advice, I could not learn statistics smoothly.

I also wish to thank Professor Yasumasa Matsuda of Tohoku University for inviting me to the Data Science Workshop. Professor Matsuda, Professor Yuzo Hosoya, Professor Yoshihiro Yajima, and Professor Yoshimasa Uematsu of Tohoku University gave me instructive comments and valuable advice from the perspective of economics.

I also would like to thank Professor Junichi Hirukawa of Niigata University for inviting me to a symposium at Niigata University. Discussion with Professor Konstantinos Fokianos of the University of Cyprus was very informative.

I am grateful to Professor Hans Rudolf Künsch of ETH Zürich for kindly sharing the data set with me.

I am deeply grateful to Professor Christian Francq of ENSAE Paris for fruitful discussion and instructive comments.

I would like to express my thanks to Mr. Takeshi Tamaoka, the chief executive officer of Ananas Japan Co. Ltd, and Mr. Yuki Nakayasu, the chief executive officer of Minsetsu, Inc., for expanding my horizons and encouraging me to have big dreams. The discussions with them about stock prices and market capitalization help to know opinions from the practical point of view.

I sincerely thank all students of the Laboratory of Prof. Taniguchi, Doctor Tomoshige Yukita, and my friends, including Mr. Itsuki Watanabe and Mr. Soma Ohno, for casual conversations. These conversations were a good change of mind to me.

I must thank the funding support by Grand-in-Aid for JSPS Fellow (Grant Number JP201920060) and Grant-in-Aid for Scientific Research (S) (Grant Number JP18H05290, Masanobu Taniguchi).

I am also indebted to Ms. Miwa Oike for helping me going through several complicated procedures.

Last but not least, special thanks go to my family, especially my father and mother, for everything you have done for me and understanding me always. I can never thank you enough.

Yuichi Goto
Waseda University

Contents

List of Figures

List of Tables

List of Symbols

$\mathbb{N}$	The set of natural numbers
$\mathbb{Z}$	The set of integers
$\mathbb{R}^d$	The set of d-tuples of real numbers
A^T	the transpose of a matrix A
$A_n \to A \quad \text{in probability}$	A_n converges in probability to A
$A_n \Rightarrow A$	A_n converges in distribution to A
$A_n \to A \quad \text{a.s.}$	A_n converges almost surely to A
Z_t	stationary time series
$\gamma_Z(l)$	the autocovariance function of Z_t at lag l
$\rho_Z(l)$	the autocorrelation function of Z_t at lag l
$f_Z(\lambda)$	the spectral density function of Z_t
X_t	the binary time series from Z_t
D_k	the number of zero crossings at lag k
$\hat{\rho}_{ZC}(l)$	the zero crossings estimator for $\rho_Z(l)$
$\hat{\rho}_Z(\ell)$	the binary estimator for $\rho_Z(l)$
$\hat{\rho}_X(\ell)$	the sample autocorrelation at lag ℓ for $\rho_X(l)$
$\hat{f}_Z(\lambda)$	the kernel estimator based on the binary estimator
$\underline{f}_Z(\lambda)$	the smoothed periodogram
W_n	the window function
$w(x)$	the lag window function
$DM(\cdot,\cdot)$	the disparity measure
$D(\cdot,\cdot,\cdot)$	the classification statistic
Θ_k	the circular data
p_{circ}	a family of circular distributions of MA(p) type
${X_k}^j$	the binary series based on Θ_k with respect to the angular α_j
η_j	the j-th cosine moment
$\hat{\eta}_j$	the binary estimator for η_j
λ_t	the intensity function
$\tilde{\lambda}_t$	the approximation of λ_t

L_n^{j}	the log-likelihood function
$\tilde{L}_n^{\mathrm{j}}$	the approximation of L_n^{j} by $\tilde{\lambda}_t$
$\hat{\boldsymbol{\theta}}_n^{\mathrm{P}}$	the Poisson (conditional) QMLE
$\hat{\boldsymbol{\theta}}_{n,r}^{\mathrm{NB}}$	the negative binomial QMLE
$\hat{\boldsymbol{\theta}}_n^{\mathrm{E}}$	the exponential QMLE
$T_{\mathrm{KL,Wald}}^{j}$	the Wald type test statistic
$T_{\mathrm{DK,Wald}}^{j}$	the modified Wald type test statistic
T_{score}^{j}	the score based CUSUM Test statistic
T_{res}^{j}	the residual based CUSUM statistic
T_{res}^{j}	the residual based CUSUM statistic
$\Lambda(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n)$	the log-likelihood ratio
Δ_n	the central sequence
$\mathcal{I}(\boldsymbol{\theta}_0)$	the Fisher information matrix
$\phi_{n,h}$	the test function for a contiguous hypothesis
$c_{n,h}$	the critical value of the test $\phi_{n,h}$.

Chapter 1

Introduction

In this dissertation, we develop the asymptotic theory of statistics associated with binary series. We devote ourselves to independent and identically distributed (i.i.d.) series and stationary time series. A binary process is also called a clipped process and a zero-one valued process. It can be recognized (i) a process derived from an underlying latent process or (ii) a process whose marginal distribution is binomial.

In the former (i), Stieltjes (1889) gave a fundamental result. Let (G_1, G_2) be Gaussian process such that

$$\mathrm{E}\begin{pmatrix} G_1 \\ G_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathrm{E}(G_1, G_2)^T (G_1, G_2) := \begin{pmatrix} 1 & \rho_{G_1,G_2} \\ \rho_{G_1,G_2} & 1 \end{pmatrix}.$$

Then, the following relation holds

$$\rho_{G_1,G_2} = \sin\left(2\pi \mathrm{Cov}\{\mathbb{I}_{\{G_1>0\}}, \mathbb{I}_{\{G_2>0\}}\}\right).$$

This equation shows that binary series ($\mathbb{I}_{\{G_1>0\}}$ and $\mathbb{I}_{\{G_2>0\}}$) have information about the correlation of original series (G_1 and G_2).

Closely related to binary series, there is a concept called the number of *zero crossings* (ZC), which can be calculated from the clipped series. ZC are defined as the points where the sign changes in an observed time series. Rice (1944) investigated a trailblazing work of this field. In the context of continuous-time, Rice (1944) studied the relation between distributions of processes and expectations of the number of ZC. Afterwards, several authors studied the counterpart results (e.g.Ylvisaker (1965) for Gaussian processes, He and Kedem (1989) for ellipsoidal processes, Barnett and Kedem (1998) for Mixtures and Products of Gaussian processes, and Barnett and Kedem (1991a) for functions of Gaussian processes). Barnett (2001) gave a good review of these results. The relation is called *Rice's formula* for continuous cases or *cosine formula* for discrete cases.

Hence, the cosine formula allows us to estimate the autocorrelation of a broad class of stationary processes by using the ZC. The estimator of autocorrelations called a *ZC estimator*. Kedem (1980) showed the consistency and asymptotic normality of the ZC estimator for scalar Gaussian processes. Kedem (1994) developed the theory of the zero crossings in the filtered time series in discrete time. Recently, the level crossing and the categorical time series have been studied by many authors (see Blake and Lindsey (1973); Abrahams (1986); Kaufmann (1987); Fahrmeir and Kaufmann (1987); Fokianos and Kedem (1998, 2003); Kedem and Fokianos (2002)). There are a number of applications of ZC. For example, emotion recognition from brain signals (Petrantonakis and Hadjileontiadis (2010b,a)), speech discrimination (Panagiotakis and Tziritas (2005)), nondestructive testing of bounded metal adherents (Kedem (1994, p.7)), and tracking a vocal sound of a Humpback Whale (Kedem and Li (1989)), etc.

Going back to the story of zero-one valued processes, Buz and Litan (2012); Keenan (1982); Lomnicki and Zaremba (1955) studied the properties of binary time series. Estimation of the spectral density function based on binary series has been developed for stationary Gaussian processes. Hinich (1967) showed the consistency and derived the asymptotic variance of estimators of the spectral density for M-dependent stationary Gaussian processes. Brillinger (1968) derived the asymptotic normality for short memory stationary Gaussian processes.

So far, we know binary process for appropriate classes of stationary processes have the information of spectra of stationary process. However, the ordinary spectral density

$$f_Z(\lambda) := \frac{1}{2\pi} \sum_{\ell \in \mathbb{Z}} \mathrm{Cov}\{Z_t, Z_{t+\ell}\} \exp\{i\ell\lambda\}$$

is based on the second moment. For uncorrelated processes like GARCH process and QAR process (Koenker and Xiao (2006)) and processes with infinite variance, the methods based on the ordinal spectral density do not available. In order to overcome the defect, recently, several authors advocate new type of spectra. For strictly stationary process $\{Z_t\}$, Li (2008) proposed Laplace spectra defined by

$$f_{0,0}(\lambda) := \frac{1}{2\pi} \sum_{\ell \in \mathbb{Z}} \mathrm{Cov}\{\mathbb{I}_{\{Z_t \le 0\}}, \mathbb{I}_{\{Z_{t+\ell} \le 0\}}\} \exp\{i\ell\lambda\}.$$

This corresponds to spectrum based on binary series. Li (2012), Li (2014), and Hagemann (2011) studied quantile spectra, defined, for $x \in \mathbb{R}$, as

$$f_{x,1-x}(\lambda) := \frac{1}{2\pi} \sum_{\ell \in \mathbb{Z}} \mathrm{Cov}\{\mathbb{I}_{\{Z_t \le x\}}, \mathbb{I}_{\{Z_{t+\ell} \le 1-x\}}\} \exp\{i\ell\lambda\}.$$

More generally, Dette et al. (2015) introduced the Laplace spectral density kernel and the copula spectral density kernel, which are defined by, for strictly stationary process $\{Z_t\}$ with marginal distribution F, for any $(q_1, q_2) \in \mathbb{R}^2$ and $(\tau_1, \tau_2) \in (0,1)^2$,

$$f_{q_1,q_2}(\lambda) := \sum_{\ell\in\mathbb{Z}} \mathrm{Cov}\{\mathbb{I}_{\{Z_t \le q_1\}}, \mathbb{I}_{\{Z_{t+\ell} \le q_2\}}\} \exp\{i\ell\lambda\},$$
$$f_{\tau_1,\tau_2}(\lambda) := \sum_{\ell\in\mathbb{Z}} \mathrm{Cov}\{\mathbb{I}_{\{F(Z_t) \le \tau_1\}}, \mathbb{I}_{\{F(Z_{t+\ell}) \le \tau_2\}}\} \exp\{i\ell\lambda\},$$

respectively. See also Birr et al. (2017), Kley et al. (2016), Hong (2000), Lee and Rao (2012). Hong (1999) proposed the generalized spectral density defined, for $(u, v) \in \mathbb{R}^2$, as

$$f^G_{u,v}(\lambda) := \frac{1}{2\pi} \sum_{\ell\in\mathbb{Z}} \mathrm{Cov}\{e^{iZ_t u}, e^{iZ_{t+\ell} v}\} \exp\{i\ell\lambda\}.$$

As regards discriminant analysis for time series, statistical theory for discriminant analysis has been studied by many authors (see Anderson (1984), Johnson and Wichern (1988), Liggett Jr (1971), and Shumway and Unger (1974)). The quality of classification is measured by misclassification probability, that is, the probability of classifying the process into the incorrect category. For more delicate evaluation, we often consider the misclassification probability when the two categories are contiguous. For the nonparametric approach, Taniguchi and Kakizawa (2000) elucidated that, for the I-divergence measure, the misclassification probability tends to zero and evaluated when the categories are contiguous. Using the Chernoff disparity measure, Zhang and Taniguchi (1995) showed that misclassification probability tends to 0, and discussed robustness when a sharp peak contaminates the spectral density. Kakizawa (1996) introduced a more general disparity measure, which includes I-divergence and Chernoff divergence measure, and discussed the two properties. Sakiyama and Taniguchi (2004) showed the above for locally stationary processes.

However, these classification methods are based on a periodogram or smoothed periodogram, and properties of disparity measures are only discussed. On the other hand, Kedem and Slud (1982) introduced a discrimination method based on binary time series, and this was applied to emotion recognition from brain signals (Petrantonakis and Hadjileontiadis (2010a,b)), and speech discrimination Panagiotakis and Tziritas (2005). Bagnall and Janacek (2005) studied the advantages of using binary data for the classification of various time series models. They concluded that using binary data has three good points. First, the classification accuracy on binary data is not significantly less than that on the original data. Second, the accuracy is better than that on the original data when outliers contaminate the data. Third, we could analyze the time series data when the data is available only for binary time series. For example, rainfall data (rainy day= 1, dry day= 0) and binary self-assessment measurements of arthritis (good= 1, poor= 0) (see Fitzmaurice and Lipsitz (1995)). Although, theoretical properties of their classification method have not been exploited so far.

Regarding directional statistics, which is a field that deals with directional data, directional statistics date back to the 1950s. Fisher (1953) had considerable influence and appealed to the necessity of directional statistics. After that, many authors tackled the problem (see Mardia (1975); Watson (1983); Fisher et al. (1993)). In recent years, directional statistics have attracted attention because of Mardia and Jupp (2000). Many distributions on the circle have been developed (e.g., uniform, cardioid, wrapped Cauchy, von Mises distribution). These distributions are closely related to the spectral density functions in time series with complex-valued coefficients (see Taniguchi et al. (2020) for details). For example, the spectral density of the autoregressive model of order 1, that of the moving average model of order 1, and that of the autoregressive model of order 2 correspond to wrapped Cauchy distribution, cardioid distribution, and the more flexible distribution proposed by Kato and Jones (2013), respectively. Optimal estimation and testing based on local asymptotic normality for directional data are addressed in Ley and Verdebout (2017). However, estimation theory based on binary series in directional statistics has not yet been investigated.

In the latter (ii), the binary time series modeling has been developed recently (see Kedem and Fokianos (2002)). The categorical time series modeling is also studied by, for example, Fahrmeir and Kaufmann (1987); Fokianos and Kedem (1998, 2003); Kaufmann (1987); Kedem and Fokianos (2002). Binary and categorical time series can be regarded as a particular case of count time series.

Count time series appear in a variety of fields, for example, the number of patients with infectious diseases (Ferland et al. (2006), Pedeli et al. (2015)), that of transactions per minute for the stock (Fokianos et al. (2009)), that of corporate defaults (Agosto et al. (2016)), and that of earthquakes (Wang et al. (2014)).

The basic approach of the modeling of count time series is to use *generalized linear model* (GLM) advocated by McCullagh and Nelder (1989). GLM is constructed by three factors: a random component, a systematic component, and a link function. For the sake of explanation, we define a random variable R with some probability density function $p(x;\theta)$ and the expectation has the following structure

$$\varphi(\mathrm{E}(R)) = \mathbf{X}\boldsymbol{\beta},$$

where φ is some function, β and θ are unknown parameters, and $\boldsymbol{X}$ is a covariate. Then, the random component, the systematic component and the link function correspond to $p(x;\theta)$, $\mathbf{X}\boldsymbol{\beta}$, and φ, respectively. One often uses the exponential family as a random component. If we use $\mathbf{X}\boldsymbol{\beta}$ as the systematic component, then it is called a linear predictor. Time-varying parameter models classified into two classes, parameter-driven models and observation-driven models, by Cox (1981). For parameter driven model, we refer to Davis et al. (2000) and Davis and Wu (2009). We shall focus on observations driven models in this dissertation. For observation driven model, McKenzie (1985) and Al-Osh and Alzaid (1987) introduce first-order Integer Autoregressive (INAR) model, denoted by INAR(1). Alzaid and Al-Osh (1990) discussed INAR(p) model. Ferland et al. (2006), Fokianos et al. (2009), Fokianos and Tjøstheim (2011), and Wang et al. (2014) discussed Poisson integer-valued generalized autoregressive conditional heteroskedasticity (INGARCH) model. Davis et al. (2003) and Benjamin et al. (2003) scrutinize the model motivated by GLM. Poisson INGARCH model shows conditional equidispersion, whose conditional mean is equal to the conditional variance, and overdispersion property, whose variance is greater than the mean. Whereas many phenomena show underdispesion and strong overdispersion properties, authors examine distributions other than Poisson. Zhu (2011), Christou and Fokianos (2014), and Cui and Wang (2019) examined the negative binomial distribution which can capture strong overdispersion. Fokianos (2001), Heinen (2003), Zhu (2012a), Zhu (2012b) and Melo and Alencar (2020),Zhu (2012c), and Gorgi (2020) probe doubly truncated Poisson, double Poisson, generalized Poisson, Conway-Maxwell Poisson, zero inflated Poisson and zero-inflated negative binomial, Beta-negative binomial, respectively. Diop and Kengne (2020) introduced a piecewise stationary INGARCH model. The fundamental properties such that stationarity and ergodicity of the model are also explored by many researchers. Neumann (2011) showed a sufficient condition for a solution of nonlinear Poisson INGARCH(1,1) models to have stationarity, ergodicity, β-mixing property. Doukhan et al. (2013) clarified the existence of the stationary and ergodic solution of general nonlinear Poisson AR models with any finite moment under the contractive condition. Agosto et al. (2016) dealt with nonlinear Poisson INGARCH(p,q) models with exogenous variable. Davis and Liu (2016) introduced the one-parameter exponential family for count processes including Poisson, negative binomial, and Bernoulli distributions and proved stability of the models.

Structural break tests have been studied from the 1950s. Page (1955) gave a pioneer study and appealed to the importance of the detection of structural breaks. The cumulative sum (CUSUM) test for the parameters of time series models is proposed by Lee et al. (2003). For the integer-valued time series models, Franke et al. (2012) proposed the residual-based CUSUM test based on the conditional least square estimator for non-linear Poisson AR(1) models. Kang and Lee (2014) focused on Poisson nonlinear INGARCH(1,1) models and constructed the Wald type test and the normalized residual-based CUSUM test based on the conditional maximum likelihood estimator (CMLE). Lee et al. (2016) and Lee et al. (2018) studied for nonlinear zero-inflated generalized Poisson INGARCH(1,1) models and linear bivariate Poisson INGARCH(1,1) models, respectively. Lee and Lee (2019) compared several change tests include the Wald type, the score based, and the residual-based CUSUM tests. Doukhan and Kengne (2015) developed a change test for the general nonlinear Poisson AR models and investigated the asymptotics under the null and the alternative. Diop and Kengne (2017) studied the change detection problem for a one-parameter exponential INGARCH(1,1) models and constructed a consistent test. Hudecová et al. (2017) advocated test for structural change based on probability generating functions for INAR(1) model. Cui et al. (2020) established the statistical inference for a location of change point.

However, it is unrealistic to assume the knowledge of the underlying conditional distribution in practice. To the best of my knowledge, the change detection problem without conditional distribution assumption has not yet been investigated. Moreover, although several change test procedures for INGARCH (1,1) models and other simple models are well studied, those procedures for general non-linear Poisson AR models have not yet been discussed except for Doukhan and Kengne (2015). On the other hand, Ahmad and Francq (2016) dropped the assumption of conditional distribution and devoted to Poisson quasi maximum likelihood estimators (PQMLE) for the general nonlinear AR models. They showed the strong consistency and asymptotic normality (CAN) of PQMLE. Aknouche et al. (2018) and Aknouche and Francq (2020) proposed negative binomial (NB) QMLE and Exponential QMLE and showed CAN, respectively. Aknouche and Francq (2020) elucidated the sufficient condition for the count time series to have the properties of the strict stationarity, ergodicity, and β-mixing. The essential condition is called the stochastic-equal-mean order property and is satisfied by many time series models.

Local asymptotic normality (LAN) plays a vital role in optimal inference and testing. The concept is introduced by LeCam (1960). Once the LAN property holds, optimal tests and estimations can be constructed (LeCam (1960); Ibragimov and Khasminskii (1981); Taniguchi and Kakizawa (2000)). Several authors studied LAN property for various models (See Roussas (1972) for Markov process, Roussas (1979) for extension of Roussas (1972) to non-Markovian process, Swensen (1985) for AR (p) models with regression term, Hallin et al. (1985) for a hypothesis that the null and alternative is given by white noise and ARMA models, respectively, Kreiss (1987) for ARMA models, Kreiss (1990) for AR (∞) models, Hallin and Puri (1994) for regression models with ARMA disturbances, Garel and Hallin (1995) for multivariate version of Hallin and Puri (1994), Hallin et al. (1999) for regression models with long memory disturbances, Benghabrit and Hallin (1996b) and Benghabrit and Hallin (1996a) for a hypothesis that the null and alternative is given by AR models and bilinear models, respectively, Linton (1993) for ARCH models, Kato et al. (2006) for CHARN models, Dahlhaus (1996) for Gaussian locally stationary processes, Sakiyama and Taniguchi (2003) for optimal estimations based on Dahlhaus (1996), Hirukawa and Taniguchi (2006) for non-Gaussian locally stationary processes, Cutting et al. (2017) for a hypothesis that the null and alternative is given by uniform distributions and rotationally symmetric distributions on high dimensional spheres, respectively, Paindaveine et al. (2017) for rotational symmetric distributions when the concentration parameter converges to 0 as the sample size diverges.) Jeganathan (1995) made a general review including local asymptotically quadratic (LAQ) and local asymptotically mixed normal (LAMN). Cutting et al. (2017) for a hypothesis that the null and alternative is given by uniform distributions and rotationally symmetric distributions on high dimensional spheres, respectively.

The simultaneous equation system is one of the pivotal models in economics. The limited information maximum likelihood estimator (LIMLE) is introduced by Anderson et al. (1949). Theil (1953) and Basmann (1957) proposed two-stage least squares estimator (TSLSE). Both are best asymptotically normal estimators, that is, asymptotically efficient estimators. Theil (1961) advocated the k-class estimator, including the LIMLE and TSLSE as a special case, and showed the sufficient conditions that the k-class estimator is consistent and efficient. Fujikoshi et al. (1982) gave the asymptotic expansions of density functions for the LIML estimator and the TSLS estimator, and compared them. Takeuchi and Morimune (1985) showed that the bias-adjusted LIML is third-order efficient. As a result, LIML is superior to the bias-adjusted efficient estimator encompassing TSLS in the sense of concentration on the true parameter uniformly. Phillips (1989) proposed the concept of limiting Gaussian functional and dealt with partially identified structural equations.

Analysis of variance (ANOVA) is the statistical method to analyze the difference of more than groups, effects of factors, and interaction of the factors. Fisher (1918) gave a trailblazing work of analysis of variance (ANOVA) to analyze genetic research. There are many applications of ANOVA method to various fields, for example, genetics (Dickerson (1942) and Sprague and Tatum (1942)), biological experiments (Anderson, 1960), horticultural sampling (Sharpe and Van Middelem, 1955), agriculture (Talbot, 1984). Details of ANOVA can be seen in, for example, Searle et al. (1992) and Hirotsu (2017)

In Chapter 2, first, we elucidate the joint asymptotic distribution of the ZC estimator. Next, we show that the variance of the ZC estimator does not attain the Cramer-Rao lower bound (CRLB). However, it is shown that the ZC estimator has robustness when an outlier contaminates the process, In contrast with this, we observe that the quasi-maximum likelihood estimator (QMLE) attains the CRLB. However, we can see that QMLE is sensitive to outlier.

In Chapter 3, we introduce a discriminant analysis based on a binary time series. We show the consistency of our discrimination method and evaluate the misclassification probabilities for two contiguous categories. We also elucidate the robustness of our method against an outlier; the classical method is sensitive to the outlier. However, for I-divergence measure, our method is insensitive against an outlier.

Chapter 4 proposes a family of circular distributions of a moving average model of order p type and discusses estimation of trigonometric moments based on binary series. We derive an explicit form of the root n consistent estimator. Although the estimator based on clipped series does not attain Cramér–Rao lower bound, it enables us to construct an efficient estimator by the Newton–Raphson iterative method. We also show the robustness of the estimator when the probability density function is contaminated with noise. The finite sample performance of the proposed estimator is also investigated.

Chapter 5 tackles the change detection problem based on QMLEs under the general nonlinear AR model. We emphasize our model includes the nonlinear INGARCH(p,q) models and do not require to specify the conditional distribution. We advocate the Wald type, score-based, and residual-based CUSUM tests based on Kang and Lee (2014) and Lee and Lee (2019) and derive the limiting behavior for these statistics under the null. As a result, we obtain the distribution-free size α CUSUM test. However, since the asymptotics of these statistics under the alternative is difficult to show, we also propose the Wald type test statistics based on Doukhan and Kengne (2015) and establish the consistency of the test. From a mathematical point of view, we deal with the Skorokhod space $D([0,\infty),\mathbb{R}^d)$, and provide a clear proof by using the multidimensional martingale difference functional CLT.

Chapter 6 shows the LAN property for the curved normal families and the simultaneous equation systems. On the other hand, we reveal that one-way random ANOVA models do not have LAN property. We consider the two cases that variance of random effect belongs to the interior of parameter space and boundary of parameter space. In case of the former, the log-likelihood ratio converges to 0 or diverges. In case of the latter, the log-likelihood ratio has atypical limit distributions which depend on the contiguity orders. The contiguity order is also extraordinary. Consequently, we cannot use the optimal theory based on LAN property. Therefore, we show the test based on the log-likelihood ratio is asymptotically most powerful.

Figure 1.1 shows the relation of chapters. Chapters 2 and 3 are essentially based on the cosine formula. Chapter 4 develops the statistical theory by the cosine formula for circular data. Chapter 5 devotes the statistical theory for count time series. Count time series can be seen as a generalization of binary time series. Chapter 6 is independent of Chapters 2-5.

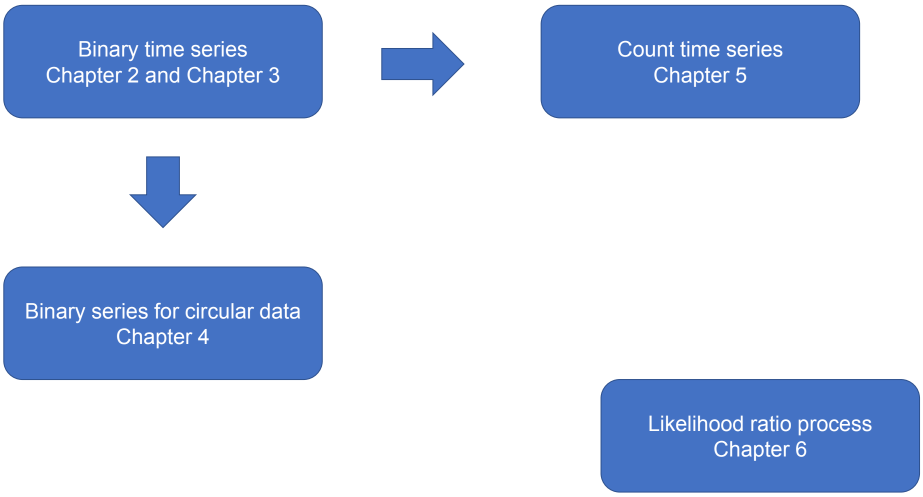

Figure 1.1: Chapter Relations

Chapter 2

Robustness of zero crossings estimator

In this Chapter, we examine the asymptotic theory of the estimator based on zero crossings (ZC), which is the number of zero crossings observed in a time series. As previously described in Introduction, the expected value of the ZC specifies the autocorrelations of the ellipsoidal processes. In Section 2.1, we derive the asymptotic distribution of the ZC estimator and compares the asymptotic variance of the ZC estimator with CRLB. Section 2.2 provides the robustness of ZC estimator against an outlier. In Section 2.3, we give the simulation study in order to know the influence of the outlier. We apply our estimator to the 91 monthly interest rates of an Austrian bank which contains three outliers in Section 2.4. This Chapter is based on Goto and Taniguchi (2019).

2.1 Asymptotic variance of ZC estimator

In this section, we define the ZC estimator, establish the asymptotic normality of the ZC estimator under the ellipsoidal assumption and elucidate the accuracy of the ZC. In the beginning, we introduce the ellipsoidal process.

Definition 2.1.1. A random vector $\mathbf{X} = (X_1, \cdots, X_n)'$ has an n-dimensional ellipsoidal distribution with location parameter μ, scale parameter Σ_n and functional parameter g, where $\mu \in \mathbb{R}^n$, Σ_n is positive definite matrix and g is a nonnegative continuous function on $[0, \infty)$ such that $\int_0^\infty t^{\frac{2}{n}-1} g(t)dt < \infty$, if its probability density function $p_{X_1,\cdots,X_n}$ is given by

$$p_{X_1,\cdots,X_n}(\boldsymbol{y}) := \frac{c_n}{\sqrt{|\Sigma_n|}} g\left((\boldsymbol{y} - \boldsymbol{\mu})\Sigma_n^{-1}(\boldsymbol{y} - \boldsymbol{\mu})'\right),$$

where $c_n = \Gamma(n/2)/(\pi^{n/2} \int_0^\infty t^{n/2-1} g(t)dt)$ and $\boldsymbol{y} = (y_1, \cdots, y_n)'$.

Gómez et al. (2003) is helpful as a reference for the ellipsoidal distribution.

Definition 2.1.2. A stochastic process $\{Z_t\}$ is called an ellipsoidal process if all the finite dimensional distributions are ellipsoidally distributed.

Kedem (1994, p.120) introduced an ellipsoidal process. The existence of a subclass of ellipsoidal process, at least, is stated in Tanaka and Shimizu (2001), which includes normal, logistic, Laplace, and double-exponential distributions. Furthermore, Kano (1994) showed that the ellipsoidal distribution whose any marginal distributions belong to an ellipsoidal family is a mixture of normal distribution. Thus, essentially, we deal with a process whose any finite distributions are a scale of mixture of normal.

Let $\{Z_t\}$ be an ellipsoidal m-dependent strictly stationary process with location parameter $\mathbf{0}$, scale parameter Σ, functional parameter g. Define a clipped time series $\{X_t\}$ from $\{Z_t\}$ by

$$X_t := \begin{cases} 1, & \text{if } Z_t \geq 0, \\ 0, & \text{if } Z_t < 0. \end{cases} \tag{2.1.1}$$

The number of zero-crossings D_k is defined by

$$\begin{aligned} D_k &:= \sum_{t=2}^{N_k} (X_{k(t-1)+1} - X_{k(t-2)+1})^2 \\ &= 2\sum_{t=1}^{N_k} (X_{k(t-1)+1} + X_{k(t-2)+1}) - 2\sum_{t=2}^{N_k} X_{k(t-1)+1} X_{k(t-2)+1} - X_1 - X_{k(N_k-1)+1}, \end{aligned}$$

where $N_k = [\frac{N-1}{k} - 1]$. Throughout this Chapter, we assume that g satisfies $\int_0^\infty t^{\frac{2}{n}} g(t)dt < \infty$ and Σ satisfies the following condition; for any random vector $(Z_{t_1}, \cdots, Z_{t_n})$, Σ has the form

$$\Sigma = \Sigma_{t_1,\cdots,t_n} = \frac{n\int_{[0,\infty)} t^{\frac{n}{2}-1} g(t)dt}{\int_{[0,\infty)} t^{\frac{n}{2}} g(t)dt} (\sigma_{t_i,t_j})_{i,j=1}^n,$$

where $\sigma_{s,t}$ satisfies $\sigma_{s,t} = \gamma_Z(s-t)$, which is the autocovariance function of $\{Z_t\}$ at lag $s-t$.
The condition for g implies $\mathrm{Var} Z_t < \infty$, the condition of Σ is technical assumption. It means that $\mathrm{Var}(Z_{t_1}, \cdots, Z_{t_n}) = (\gamma_Z(i-j))_{i,j=1}^n$.
We denote that the autocovariance function of X_t at lag l, and the autocorrelation function of X_t at lag l by $\gamma_X(l)$ and $\rho_X(l)$ respectively. Here we assume that the autocorrelations of $\{Z_t\}$ $\rho_Z(l)$, $l \in \mathbb{Z}$ are unknown, and estimate

$\rho_{\mathbf{Z}} = (\rho_Z(1), \cdots, \rho_Z(m))'$ by using ZC.

The following lemma is due to He and Kedem (1989) (see also Barnett and Kedem (1991b)).

Lemma 2.1.1. *If $\{Z_t\}$ is a strictly stationary ellipsoidal process with zero mean, finite variance and autocorrelation $\rho_Z(k)$. Then,*

$$\rho_Z(k) = \cos\left(\frac{\pi \mathrm{E} D_k}{N_k - 1}\right).$$

This equation is called as the *cosine formula*

Theorem 2.1.1. *Let $\{Z_t\}$ be a strictly stationary ellipsoidal m-dependent process with zero mean, and finite variance. Then,*

$$\frac{1}{\sqrt{N}}\begin{pmatrix} D_1 - \mathrm{E}D_1 \\ D_2 - \mathrm{E}D_2 \\ \vdots \\ D_m - \mathrm{E}D_m \end{pmatrix} \Rightarrow N(0, V)$$

where $V = (v_{s,t})_{s,t=1,\cdots,m}$ and $v_{s,t}$ is the limit value of $4/N \sum_{i=2}^{N_k} \sum_{j=2}^{N_s} [K_X(k, s(j-1) - k(i-2), s(j-2) - k(i-2)) + \gamma_X(k(i-1) - s(j-1))\gamma_X(k(i-2) - s(j-2)) + \gamma_X(k(i-2) - s(j-1))\gamma_X(k(i-1) - s(j-2))]$.

Now we proceed to estimate the unknown parameter vector $\rho_{\mathbf{Z}}$. The ZC estimator of $\rho_Z(k)$ is defined by

$$\hat{\rho}_Z(k) := \cos\left(\frac{\pi D_k}{N_k - 1}\right),$$

and we define $\hat{\boldsymbol{\rho}}_{\boldsymbol{ZC}} := (\hat{\rho}_{ZC}(1), \cdots, \hat{\rho}_{ZC}(m))'$ By employing the cosine formula and Theorem 2.1.1, the following Corollary is proved.

Corollary 2.1.1. *Let $\{Z_t\}$ be a strictly stationary ellipsoidal m-dependent process with zero mean, finite variance, autocorrelation $\rho_Z(k)$ and $\hat{\rho}_{ZC}(k)$ be the ZC estimator of $\rho_Z(k)$. Then,*

$$\sqrt{N}\begin{pmatrix} \hat{\rho}_{ZC}(1) - \rho_Z(1) \\ \hat{\rho}_{ZC}(2) - \rho_Z(2) \\ \vdots \\ \hat{\rho}_{ZC}(m) - \rho_Z(m) \end{pmatrix} \Rightarrow N(0, A'VA),$$

where $V = (v_{s,t})_{s,t=1,\cdots,m}$ and $v_{s,t}$ is the limit value of $4/N \sum_{i=2}^{N_k} \sum_{j=2}^{N_s} [K_X(k, s(j-1) - k(i-2), s(j-2) - k(i-2)) + \gamma_X(k(i-1) - s(j-1))\gamma_X(k(i-2) - s(j-2)) + \gamma_X(k(i-2) - s(j-1))\gamma_X(k(i-1) - s(j-2))]$ *and*

$A = \mathrm{diag}\left(\pi\sqrt{1 - \rho_Z(1)^2}, 2\pi\sqrt{1 - \rho_Z(2)^2}, \cdots, m\pi\sqrt{1 - \rho_Z(m)^2}\right)'$.

2.1.1 Evaluation of asymptotic variance of ZC estimator and CRLB

In this section we evaluate the Gaussian efficiency. We confine ourselves to the case of MA(1), so we assume $\rho_Z(l) = 0$, $|l| \geq 2$, and $\{Z_t\}$ is a Gaussian process with unit variance. This is a special case of $m-$dependent ellipsoidal process. In this case, the spectral density function f_Z of $\{Z_t\}$ is represented as

$$f_Z(\lambda) = \frac{1}{2\pi}(1 + 2\rho_Z(1)\cos\lambda). \tag{2.1.2}$$

It follows from Corollary 2.1.1 that

$$v_{1,1}^2 = -\frac{1}{16} - \frac{5}{4\pi^2}\sin^{-2}\rho_Z(1) - \frac{3}{4\pi}\sin^{-1}\rho_Z(1) + 2\mathrm{E}\{X_{t+2}X_{t+1}X_tX_{t-1}\}. \tag{2.1.3}$$

Proposition 2.1.1. *Let $\{Z_t\}$ be an MA(1) Gaussian stationary process, which is a special case of $m-$dependent ellipsoidal process, with mean zero, and spectral density function f, and let $\{X_t\}$ be the clipped time series in (2.1.1). Then,*

$$\begin{aligned}
&\mathrm{E}\{X_{t+2}X_{t+1}X_tX_{t-1}\}\\
=&\frac{1}{4\pi^2}\sum_{m_1,m_2,m_3=0}^{\infty}\frac{\rho_Z(1)^{2(m_1+m_2+m_3+1)}}{(2m_1+1)(2m_2)!(2m_3+1)}\\
&\times\frac{(2(m_1+m_2))!(2(m_2+m_3))!}{2^{m_1}(m_1)!2^{m_1+m_2}(m_1+m_2)!2^{m_2+m_3}(m_2+m_3)!2^{m_3}(m_3)!}\\
&+\frac{3}{8\pi}\sin^{-1}\rho_Z(1) + \frac{1}{16}.
\end{aligned}$$

From (2.1.3) and Proposition 2.1.1, we have

$$\begin{aligned}
&4\pi^2(1-\rho_Z(1)^2){v_{1,1}}^2\\
=&(1-\rho_Z(1)^2)(\frac{\pi^2}{4} - 5\sin^{-2}\rho_Z(1) + 2\sum_{m_1,m_2,m_3=0}^{\infty}\frac{\rho_Z(1)^{2(m_1+m_2+m_3+1)}}{(2m_1+1)(2m_2)!(2m_3+1)}\\
&\times\frac{(2(m_1+m_2))!(2(m_2+m_3))!}{2^{m_1}(m_1)!2^{m_1+m_2}(m_1+m_2)!2^{m_2+m_3}(m_2+m_3)!2^{m_3}(m_3)!}).
\end{aligned}$$

Next, we evaluate CRLB $1/I(\rho_Z(1))$, where $I(\rho_Z(1))$ is the Fisher information in stationary time series;

$$I(\rho_Z(1)) = \frac{1}{4\pi}\int_{[-\pi,\pi]}\left(\frac{\partial}{\partial\rho_Z(1)}\log f_Z(\lambda)\right)^2 d\lambda.$$

Note that $\rho_Z(1)$ satisfies $|\rho_Z(1)| \leq 1/2$ under the stationary condition.

Proposition 2.1.2. *For the spectral density f, represented as (2.1.2), we have*

$$I(\rho_Z(1)) = \begin{cases} 1 & (\rho_Z(1) = 0) \\ \frac{1}{2\rho_Z(1)^2}\{1 + \frac{8\rho_Z(1)^2-1}{(4\rho_Z(1)^2-1)^{\frac{2}{3}}}\} & (0 < |\rho_Z(1)| < \frac{1}{2}) \\ +\infty & (\rho_Z(1) = \pm\frac{1}{2}) \end{cases}.$$

2.1.2 Comparision of asymptotic variance of ZC estimator with CRLB

We have already calculated the asymptotic variance of ZC estimator and CRLB in Corollary 2.1.1 and Proposition 2.1.2 under Gaussian MA(1) model, which is a special case of $m-$dependent ellipsoidal process. So we compare the variance with $I(\rho_Z(1))$. We plotted the graphs of asymptotic variance of ZCE and CRLB where the variable on the horizontal axis is $\rho_Z(1)$ in Figures 2.1 and 2.2 respectively.

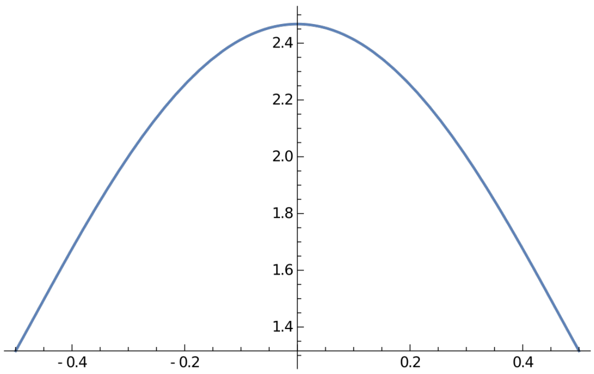

Figure 2.1: Asymptotic variance of ZCE

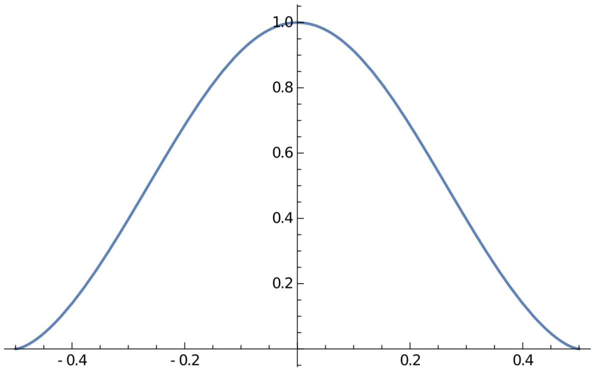

Figure 2.2: CRLB $I^{-1}(\rho_Z(1))$

$\rho_Z(1)$	$\phi(\rho_Z(1))$
0.	1.4674
0.1	1.5
0.2	1.56829
0.26	1.59816
0.27	1.60081
0.28	1.60261
0.29	1.60349
0.3	1.60338
0.31	1.60223
0.32	1.59997
0.33	1.59655
0.34	1.59192
0.35	1.58603
0.4	1.53621
0.5	1.31595

Table 2.1: Approximate values of $\phi(\rho_Z(1))$

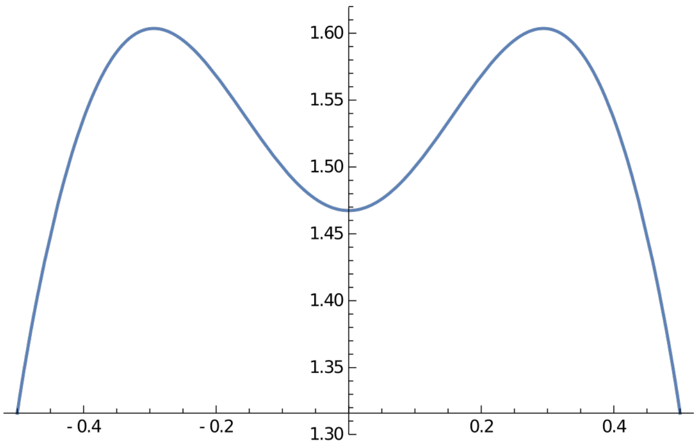

Figure 2.3: Values of function $\phi(\rho_Z(1))$

Define the function ϕ (= (*asymptotic variance of ZCE*) − (*CRLB*)) by

$$\phi(\rho_Z(1)) = 4\pi^2(1 - \rho_Z(1)^2)v_{1,1}^2 - I^{-1}(\rho_Z(1)).$$

Table 2.1 shows the value of $\rho_Z(1)$ and the approximation value $\phi(\rho_Z(1))$ corresponding to $\rho_Z(1)$. We plotted the graph of $\phi(\rho_Z(1))$ where the variable on the horizontal axis is $\rho_Z(1)$ in Figure 2.3. The asymptotic variance of the ZC estimator is the closest to the CRLB when $\rho_Z(1) = 0.5$ and the difference of the value between the asymptotic variance of the ZC estimator and CRLB is about 1.31595. Besides, we can show that $\phi(0) > 0$ mathematically. Thus we conclude the asymptotic variance of the ZC estimator does not attain the CRLB.

2.2 Outlier robustness of ZC estimator

In this section, we consider the case when $\{Z_t\}$ is contaminated by an outlier. Then, we elucidate that $\hat{\rho}_{ZC}$ is robust with respect to such an outlier but the quasi maximum likelihood estimator (QMLE) of $\rho_{\mathbf{Z}}$ is not so. Let $\{Z_t^s\}$ be a process whose the initial value is contaminated by an outlier, i.e. $Z_1^s = s$ and $Z_t^s = Z_t$ for $t \geq 2$.

2.2.1 Sensitivity of QMLE estimator

QMLE was studied by Hosoya and Taniguchi (1982). First, we fit a certain parametric spectral density f_θ to the spectral density of $\{Z_t\}$.

Let $T(\underline{I}_n)$ be a value that minimizes

$$DM(f_\theta, \underline{I}_n) = \int_{[-\pi,\, \pi]} \left[\log f_\theta(\lambda) + \frac{\underline{I}_n(\lambda)}{f_\theta(\lambda)} \right] d\lambda,$$

with respect to θ, where $\underline{I}_n(\lambda)$ is the periodogram of $\{Z_t\}$, given by

$$\underline{I}_n(\lambda) = \frac{1}{2\pi n} \{\sum_{t=1}^{n} Z_t e^{it\lambda}\}\{\sum_{t=1}^{n} Z_t e^{-it\lambda}\}.$$

We make the following assumptions on parameter space Θ:
The parameter space Θ is compact subset of $\mathbb{R}^m$, $\theta_1 \neq \theta_2$ implies $f_{\theta_1} \neq f_{\theta_2}$, $f_\theta(\lambda)$ is twice continuously differentable function of $\theta \in \Theta$ and continuous with respect to $\lambda \in [-\pi, \pi]$, there exist constants $c > 0$, C such that $c < f_\theta(\lambda) < C$, uniformly in $\theta \in \Theta$ and $\lambda \in [-\pi, \pi]$.
From Wold's decomposition (see Brockwell and Davis (1991, p.187) and the assumption of $m-$dependence of the process, we have $Z_t = \sum_{j=h}^{h+m} \phi_j e_{t-j}$ where $\{e_t\}$ is a $m-$dependent white noise sequence. Hosoya and Taniguchi (1982) does not impose distribution assumption, but require the below condition to $\{e_t\}$:
1) for each m,
$\mathrm{Var}\left[\mathrm{E}\left\{e(n)e(n+m)|\mathscr{B}(n-\tau)\right\} - \mathrm{E}\left\{e(n)e(n+m)\right\}\right] = O\left(\tau^{-2-\varepsilon}\right), \quad \varepsilon > 0,$
uniformly in n.
2)

$$\begin{aligned}&\mathrm{E}\Big|\mathrm{E}\{e(n_1)\,e(n_2)\,e(n_3)\,e(n_4)\,\big|\mathscr{B}(n_1-\tau)\} - \mathrm{E}\left\{e(n_1)\,e(n_2)\,e(n_3)\,e(n_4)\right\}\Big| \\ &= O\left(\tau^{-1-\eta}\right),\end{aligned}$$

uniformly in n_1, where $n_1 \leq n_2 \leq n_3 \leq n_4$ and $\eta > 0$.
3) $\sum_{j_1,j_2,j_3=-\infty}^{\infty} |K_e(j_1, j_2, j_3)| < \infty$,
where K_e is fourth order cumulant of $\{e_t\}$.
From $m-$dependency of $\{e_t\}$, the conditions 1), 2), and 3) are established.

The following lemma is essentially due to Hosoya and Taniguchi (1982).

Lemma 2.2.1. *Under the assumptions on parameter space, it holds that*

$$\frac{\sqrt{n}}{\gamma(0)}\left(T(\underline{I}_n) - T(f_Z)\right) \Rightarrow N(0, Q_{f_Z}^{-1} S_{f_Z} Q_{f_Z}^{-1}),$$

where

$$Q_{f_Z} = \int_{[-\pi,\pi]} \left[\frac{\partial^2}{\partial\theta^2} \frac{f_Z(\lambda)}{f_\theta(\lambda)} + \frac{\partial^2}{\partial\theta^2} \log f_\theta(\lambda) \right]_{\theta=T(f_Z)} d\lambda,$$

and $S_{f_Z} = (S_{f_Z}^{(i,j)})_{i,j=1,\cdots,n}$ *such that*

$$\begin{aligned} S_{f_Z}^{(i,j)} =& \frac{4\pi}{\gamma^2(0)} \int_{[-\pi,\pi]} \left[f_Z^2(\lambda) \left\{ \frac{\partial}{\partial \theta_i} \left(\frac{1}{f_\theta} \right) \right\} \left\{ \frac{\partial}{\partial \theta_j} \left(\frac{1}{f_\theta} \right) \right\} \right]_{\theta = T(f_Z)} d\lambda \\ &+ \frac{1}{(2\pi)^2 \gamma^2(0)} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left[\left(\frac{\partial}{\partial \theta_i} f_\theta(\lambda_1) \right) \left(\frac{\partial}{\partial \theta_j} f_\theta(\lambda_2) \right) \right]_{\theta = T(f_Z)} \\ &\times \sum_{t_1,t_2,t_3=-\infty}^{\infty} \exp\{-i(-\lambda_1 t_1 + \lambda_2 t_2 - \lambda_2 t_3)\} K_Z(t_1,t_2,t_3) d\lambda_1 d\lambda_2. \end{aligned}$$

For m-dependent process, the value $T(f_Z) \in \mathbb{R}^m$ coincides with ρ. Thus QMLE can be used for the autocorrelation estimation when we can obtain uncontaminated observations.

Next, we consider the case that we use observations contaminated by the outlier, that is, we estimate the autocorrelation by using $\{Z_t^s\}$.

Theorem 2.2.1. *For fixed n and given* $Z_1, Z_2, \cdots, Z_n$, *it holds that*

$$T(\underline{I}_n^s) \to \min_{\theta \in \Theta} \int_{[-\pi,\pi]} 1/f_\theta(\lambda) d\lambda \quad \text{as } s \to \infty, \tag{2.2.1}$$

where $\underline{I}_n^s$ *is the periodogram of* $\{Z_t^s\}$.

Remark 2.2.1. Kedem (1994) pointed out that if $Z_1 > 0$, the sample autocovariance tends to 0 as $s \to \infty$. Theorem 2.2.1 shows, if a process is contaminated by the outlier, QMLE tends to a non-zero constant (it is more general than the Kedem's result), i.e. the right hand side value of (2.2.1). Hence, QMLE does not work well. Thus we conclude that QMLE is sensitive to the outlier. Combined the above two results, we conclude that ZCE is better than QMLE if the process is contaminated by the outlier.

2.3 Simulation studies

In this section, we present the simulation studies. We compare the mean of ZCE, sample autocorrelation (SACF), and highly robust estimator (HRE), proposed by Ma and Genton (1998), for AR(1) (or MA(1)) model with for these with contaminated AR(1) (or MA(1)) model, i.e. the process is contaminated with outliers. HRE is a sort of the U statistics, defined by

$$\hat{\rho}_{HR}(l) = \frac{Q_{n-h}(u+v) - Q_{n-h}(u-v)}{Q_{n-h}(u+v) + Q_{n-h}(u-v)},$$

where u is a vector of the first $n-l$ observations of an observation sequence, v is a vector of the last $n-l$ observations of an observation sequence,

$$Q_n(z) = 2.219\{|Z_i - Z_j|; i < j\}_{(k)}, \quad k = \text{int}\left[\left(\binom{n}{2} + 2\right)/4\right] + 1,$$

int($\cdot$) denotes the integer part, and $\{|Z_i - Z_j|; i < j\}_{(k)}$ denotes kth order statistic of the set of all absolute differences $|Z_i - Z_j|$ for $i < j$. We consider AR(1) model and MA(1) defined by

$$Z_t = aZ_{t-1} + \epsilon_t, \quad Z_t = \epsilon_t + b\epsilon_{t-1}$$

where $\{\epsilon_t\} \overset{i.i.d.}{\sim} N(0,1)$, respectively.

The procedure is as follows: First, we generate 500 samples $Z_1, \cdots Z_{500}$ and computed the estimators, $\hat{\rho}_{ZC}(l)$,$\hat{\rho}_{SACF}(l) = \hat{\gamma}_{SACF}(l)/\hat{\gamma}_{SACF}(0)$, $\hat{\rho}_{HR}(l)$. Next, we replace Z_i with $Z_i + Mmax\{Z_1, \cdots Z_{500}\}$ for $i = 1, \cdots, 10$, and computed the three estimators, $\hat{\rho}^M_{ZC}(l)$,$\hat{\rho}^M_{SACF}(l) = \hat{\gamma}^M_{SACF}(l)/\hat{\gamma}^M_{SACF}(0)$, $\hat{\rho}^M_{HR}(l)$, where M is a constant. We iterated 500 times above procedure. Then, we compare the influence of the estimation value, the mean $\mathrm{E}(\hat{\rho}_j - \hat{\rho}^M_j) \approx 1/500 \sum_{i=1}^{500}(\hat{\rho}^{(i)}_j - \hat{\rho}^{M^{(i)}}_j)$, $j = ZCE, SACF, HRE$.

The results are in Table 2.2. Roughly speaking, the Tables shows that $\hat{\rho}_{ZC}(l)$ and $\hat{\rho}_{HR}(l)$ is not affected by the outliers. By contrast, $\hat{\rho}_{SACF}(l)$ is influenced by the outliers. We conclude that $\hat{\rho}_{ZC}(l)$ and $\hat{\rho}_{HR}(l)$ are insensitive to the outliers and $\hat{\rho}_{SACF}(l)$ is sensitive. More precisely, $\hat{\rho}_{ZC}(l)$ is a little bit more robust against outliers than $\hat{\rho}_{HR}(l)$.

Table 2.2: Influence of estimators in the mean value by the10 outliers

AR(1) model with $a = -0.6$				
time lag	M	$\hat{\rho}_{ZC}$	$\hat{\rho}_{HR}$	$\hat{\rho}_{SACF}$
1	1	-0.03	-0.03	-0.23
	2	-0.03	-0.03	-0.63
	4	-0.03	-0.02	-1.11
	8	-0.03	-0.02	-1.38
2	1	-0.02	-0.03	-0.07
	2	-0.02	-0.03	-0.19
	4	-0.02	-0.03	-0.33
	8	-0.02	-0.03	-0.41

MA(1) model with $b = 0.4$				
time lag	M	$\hat{\rho}_{ZC}$	$\hat{\rho}_{HR}$	$\hat{\rho}_{SACF}$
1	1	-0.02	-0.03	-0.08
	2	-0.02	-0.03	-0.23
	4	-0.02	-0.03	-0.41
	8	-0.02	-0.03	-0.51
2	1	-0.02	-0.03	-0.12
	2	-0.02	-0.03	-0.33
	4	-0.02	-0.03	-0.60
	8	-0.02	-0.03	-0.74

2.4 Real data analysis

In this section, we apply our estimator to the data of 91 monthly interest rates of an Austrian bank. The data is already analyzed by Kunsch (1984), Ma and Genton (1998), and many other authors. Figure 2.4 is the plot of the data. There are three doubtful points at number 18, 28, 29 whether these points are outliers or not. Kunsch (1984) replaced these points by 9.85, and discuss the influence which caused by the substitution. The influence of replacement is small, then we could be reliable the value of estimate.

We run $\hat{\rho}_{ZC}$ and $\hat{\rho}_{SAFC}$ on the original data and on the replaced data. Note that the mean value changes in the substitution. The result is in Table 2.3. Comparing the two tables, we can see that $\hat{\rho}_{ZC}$ is not influenced by the outliers at all, contrary to this, $\hat{\rho}_{SAFC}$ is affected by the outliers. Thus we conclude that $\hat{\rho}_{ZC}$ is more reliable than $\hat{\rho}_{SAFC}$. From the point of view of the interest rate of a bank, a high autocorrelation is reasonable. Because the interest rate of a bank is determined by reference to the past interest rate.

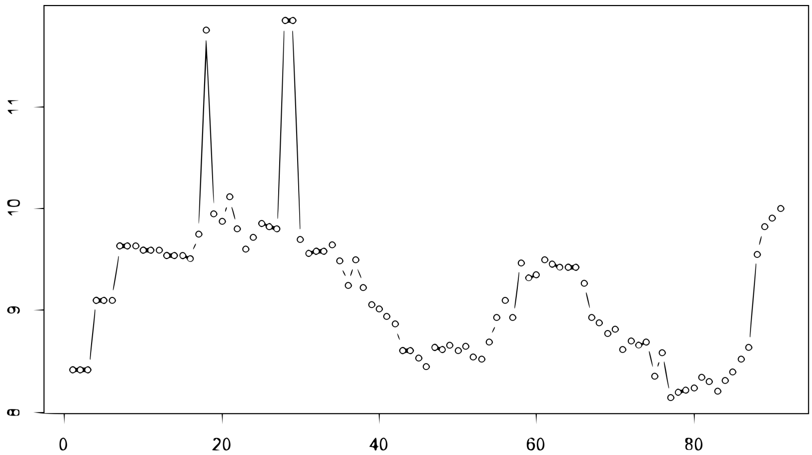

Figure 2.4: 91 monthly interest rates of an Austrian bank
F

Table 2.3: Estimated values of the autocorrelation

original data		
time lag	$\hat{\rho}_{ZC}$	$\hat{\rho}_{SAFC}$
1	0.98	0.79
2	0.94	0.62

replaced data		
time lag	$\hat{\rho}_{ZC}$	$\hat{\rho}_{SAFC}$
1	0.98	0.91
2	0.94	0.83

Chapter 3

Discriminant analysis based on binary time series

In Chapter 3, we discuss the discriminant analysis and propose a classification method based on binary time series for an ellipsoidal alpha mixing strictly stationary process. Assume that the observations are generated by time series which belongs to one of two categories described by different spectra. We propose a method to classify into the correct category with high probability. First, we will show that the misclassification probability tends to zero when the number of observation tends to infinity, that is, the consistency of our discrimination method. Further, we evaluate the asymptotic misclassification probability when the two categories are contiguous. Finally, we show that our classification method based on binary time series has good robustness properties when the process is contaminated by an outlier, that is, our classification method is insensitive to the outlier. However, the classical method based on smoothed periodogram is sensitive to outliers. We also deal with a practical case where the two categories are estimated from the training samples. For an electrocardiogram data set, we examine the robustness of our method when observations are contaminated with an outlier.

This chapter is organized as follows: Section 3.1 introduces some notations and assumptions. Section 3.2 provides the fundamental properties of our discriminant method. We show its consistency and evaluate the contiguous misclassification probability. In Section 3.3, we show robustness of our discriminant method against an outlier. In Section 3.4, unknown categories case is considered. We give a simulation study to see the robustness in Section 3.5. We apply our discriminant method to electrocardiogram (ECG) data in Section 3.6. This Chapter is based on Goto and Taniguchi (2020).

3.1 Settings

In this section, we discuss ellipsoidal α-mixing processes and introduce some relevant assumptions. First, we restate the definitions of the ellipsoidal distribution and the ellipsoidal process.

Definition 3.1.1. A random vector $\mathbf{Z} = (Z_1, \cdots, Z_n)'$ has an n-dimensional ellipsoidal distribution with a location parameter $\boldsymbol{\mu} \in \mathbb{R}^n$, scale parameter Σ_n and functional parameter g, where Σ_n is a positive definite matrix and g is a nonnegative continuous function on $[0, \infty)$ such that $\int_0^\infty t^{n/2-1} g(t) \mathrm{d}t < \infty$, if its probability density function $p_{Z_1,\cdots,Z_n}$ is given by

$$p_{Z_1,\cdots,Z_n}(\boldsymbol{y}) := \frac{c_n}{\sqrt{|\Sigma_n|}} g\left((\boldsymbol{y} - \boldsymbol{\mu})\Sigma_n^{-1}(\boldsymbol{y} - \boldsymbol{\mu})'\right),$$

where $c_n := \Gamma(n/2)/(\pi^{n/2} \int_0^\infty t^{n/2-1} g(t) \mathrm{d}t)$ and $\boldsymbol{y} = (y_1, \cdots, y_n)'$.

Gómez et al. (2003) is helpful as a reference for the ellipsoidal distribution.

Definition 3.1.2. A stochastic process $\{Z_t\}$ is called an ellipsoidal process if all the finite dimensional distributions are ellipsoidally distributed.

(Kedem, 1994, p.120) introduced an ellipsoidal process. The existence of a subclass of ellipsoidal processes, at least, is stated in Tanaka and Shimizu (2001), which includes normal, logistic, Laplace, and double-exponential distributions. Furthermore, Kano (1994) showed that the ellipsoidal distribution whose any marginal distributions belong to an ellipsoidal family is a mixture of normal distribution. Thus, essentially, we deal with a process whose any finite distributions are a scale of mixture of normal.

In this Chapter, we need the α-mixing assumption, which is associated with the dependence structure on the process.

Definition 3.1.3. A stochastic process $\{Z_t\}$ is called α-mixing or strongly mixing if, for each $k \in \mathbb{Z}$ and $n \in \mathbb{N}$, $A \in \mathcal{B}_{-\infty}^k$ and $B \in \mathcal{B}_{k+n}^\infty$ together imply

$$\sup_{k\in\mathbb{Z}, A\in\mathcal{B}_{-\infty}^k,\, B\in\mathcal{B}_{k+n}^\infty} |\mathrm{P}(AB) - \mathrm{P}(A)\mathrm{P}(B)| \le \alpha(n),$$

where $\mathcal{B}_a^b$ is the σ-field generated by $\{Z_t : a \le t \le b\}$ and $\alpha(\cdot)$ is a function, called α-mixing coefficients, satisfies $\alpha(n) \to 0$ as $n \to \infty$.

Let $\{Z_t\}$ be an ellipsoidal α-mixing strictly stationary process with location parameter $\mathbf{0}$, scale parameter Σ_n, functional parameter g, α-mixing coefficients $\alpha(\cdot)$ satisfying $\alpha(n) = O(1/n^{8+\delta})$ for some $\delta > 0$, and the second order spectral density function of $f_Z(\lambda)$. We define the clipped time series $\{X_t\}$ generated from $\{Z_t\}$ by

$$X_t := \begin{cases} 1, & \text{if } Z_t \geq 0, \\ 0, & \text{if } Z_t < 0. \end{cases}$$

The second order spectral density function of $\{X_t\}$ is denoted by $f_X(\lambda)$, and the autocovariance function of $\{Z_t\}$ at lag ℓ and that of $\{X_t\}$ are represented as $\gamma_Z(\ell)$ and $\gamma_X(\ell)$, respectively.

Throughout the Chapter, we assume that g satisfies $\int_0^\infty t^{n/2} g(t)\mathrm{d}t < \infty$, $\gamma_Z(0)$=1, and Σ_n satisfies the following conditions; for any random vector $(Z_{t_1}, \cdots, Z_{t_n})$,
Σ_n has the form

$$\Sigma_n := \Sigma_{t_1,\cdots,t_n} := \frac{n\int_{[0,\infty)} t^{\frac{n}{2}-1} g(t)dt}{\int_{[0,\infty)} t^{\frac{n}{2}} g(t)dt} (\sigma_{t_i,t_j})_{i,j=1}^n,$$

where $\sigma_{s,t} = \gamma_Z(s-t)$. Recall that cumulant of order ℓ of $(X_1, \cdots, X_\ell)$ is defined as

$$\mathrm{Cum}(X_1, \cdots, X_\ell) := \sum_{(\nu_1,\ldots,\nu_p)} (-1)^{p-1}(p-1)! \left(\mathrm{E}\prod_{j\in\nu_1} X_{\nu_1}\right) \cdots \left(\mathrm{E}\prod_{j\in\nu_p} X_{\nu_p}\right),$$

with the summation $\sum_{(\nu_1,\ldots,\nu_p)}$ extends over all partitions $(\nu_1, \ldots, \nu_p)$ of $\{1, 2, \cdots, \ell\}$ (see Brillinger (1981)). The order assumption of α-mixing coefficient implies that

(i)

$$f_Z(\lambda), f_X(\lambda) \in C^8[-\pi, \pi] \quad \text{(see (Ibragimov and Rozanov, 1978, p.181));}$$

(ii)

$$\sum_{\ell_1,\cdots,\ell_{k-1}=-\infty}^{\infty} |\mathrm{Cum}(X_t, X_{t+\ell_1}, \cdots, X_{t+\ell_{k-1}})| < \infty \quad \text{for } k = 2, \cdots, 8$$

(see Kley (2014));

(iii)

$$\hat{\gamma}_X(\ell) - \mathrm{E}\hat{\gamma}_X(\ell) = O_p(1/\sqrt{n}) \quad \text{uniformly in } \ell \text{ (see Robinson (1991)).}$$

The condition on g guarantees the existence of $\gamma_Z(0)$. The condition on the Σ is a technical assumption which ensures that $\mathrm{Var}(Z_{t_1}, \cdots, Z_{t_n}) = (\gamma_Z(i-j))_{i,j=1}^n$. $\gamma_Z(0)$=1 means $\gamma_Z(\ell) = \rho_Z(\ell)$ where $\rho_Z(\ell)$ is the autocorrelation function of $\{Z_t\}$ at lag ℓ.

Remark 3.1.1. Brillinger (1968) imposed the summability condition of the cumulant for the process $\{X_t\}$. Strictly stationarity and damping order of α-mixing coefficient are essential to describe the condition on $\{X_t\}$ in terms of that on $\{Z_t\}$. Also, the assumption for strictly stationarity is not restrictive. Giraitis et al. (2000) discuss the strict stationarity of a very general class of ARCH models. In this case we need stronger assumption for weakly stationarity.

Remark 3.1.2. In this dissertation, we restrict ourselves to an ellipsoidal process. For strictly stationary process, Li (2008) proposed Laplace spectra defined by

$$f_{0,0}(\lambda) := \frac{1}{2\pi} \sum_{\ell \in \mathbb{Z}} \mathrm{Cov}\{\mathbb{I}_{\{Z_t \le 0\}}, \mathbb{I}_{\{Z_{t+\ell} \le 0\}}\} \exp\{i\ell\lambda\}.$$

More generally, Dette et al. (2015) introduced Laplace spectral density kernel and copula spectral density kernel defined by, for strictly stationary process $\{Z_t\}$ with marginal distribution F, for any $(q_1, q_2) \in \mathbb{R}^2$ and $(\tau_1, \tau_2) \in (0, 1)^2$,

$$f_{q_1,q_2}(\lambda) := \sum_{\ell \in \mathbb{Z}} \mathrm{Cov}\{\mathbb{I}_{\{Z_t \le q_1\}}, \mathbb{I}_{\{Z_{t+\ell} \le q_2\}}\} \exp\{i\ell\lambda\},$$
$$f_{\tau_1,\tau_2}(\lambda) := \sum_{\ell \in \mathbb{Z}} \mathrm{Cov}\{\mathbb{I}_{\{F(Z_t) \le \tau_1\}}, \mathbb{I}_{\{F(Z_{t+\ell}) \le \tau_2\}}\} \exp\{i\ell\lambda\},$$

respectively.

The following lemma is due to He and Kedem (1989) (see also Barnett and Kedem (1991b)).

Lemma 3.1.1. *If $\{Z_t\}$ is a strictly stationary ellipsoidal process with zero mean, finite variance and autocorrelation $\rho_Z(\ell)$. Then,*

$$\rho_Z(\ell) = \sin\left(\frac{\pi}{2}\rho_X(\ell)\right),$$

where $\rho_X(\ell)$ is the autocorrelation function of $\{X_t\}$ at lag ℓ.

From the lemma, the sample version of the autocorrelations for $\{X_t\}$ and $\{Z_t\}$ at lag ℓ are given by

$$\hat{\rho}_X(\ell) := \frac{4}{n} \sum_{i=1}^{n-|\ell|} (X_i - \frac{1}{2})(X_{i+|\ell|} - \frac{1}{2}), \text{ and } \hat{\rho}_Z(\ell) = \sin(\frac{\pi}{2}\hat{\rho}_X(\ell)),$$

respectively.

We define the window function W_n by

$$W_n(\theta) := M \sum_{\nu=-\infty}^{\infty} W\left(M(\theta + 2\pi\nu)\right)$$

where $W(\theta)$ and M satisfy the following assumptions:

Assumption 3.1.1. $(W1)$ $W(\theta)$ is a real, bounded nonnegative, even function with

$$\int_{-\infty}^{\infty} W(\theta)\mathrm{d}s = 1, \int_{-\infty}^{\infty} \theta^2 W(\theta)\mathrm{d}s = k_2 > 0.$$

$(W2)$ $w(x) = \int_{-\infty}^{\infty} W(\theta)\exp(i\theta x)\mathrm{d}s$ satisfies $|w(x)| \leq \bar{w}(x)$, where $\bar{w}(x)$ is even, integrable, and monotonically decreasing on $[0, \infty)$.
(M) $M := M(n)$ satisfies the following relationship; $M/\sqrt{n} + n^{1/4}/M \to 0$ as n $\to \infty$.

Remark 3.1.3. The above assumptions are standard; $(W1)$ and $(W2)$ can be seen in Hannan (1970), Brillinger (1981), Taniguchi and Kakizawa (2000), and we can find (M) in Taniguchi (1987).

Let $\hat{f}_Z(\lambda)$ be a nonparametric spectral estimator given by

$$\hat{f}_Z(\lambda) := \int_{[-\pi,\pi]} W_n(\lambda - \mu)\hat{I}_n(\mu)\mathrm{d}\mu, \tag{3.1.1}$$

where

$$\hat{I}_n(\lambda) := \frac{1}{2\pi} \sum_{|\ell| \leq M} \hat{\rho}_Z(\ell)\exp(-i\ell\lambda).$$

3.2 Discriminant analysis

In this section, introducing a new discriminant method based on clipped time series, we elucidate its fundamental properties. We use a general disparity measure introduced by Kakizawa (1996), which includes I-divergence disparity measure and Chernoff disparity measure. Although Kakizawa (1996), Taniguchi and Kakizawa (2000), and Sakiyama and Taniguchi (2004) discussed the problem of discriminant analysis, the used spectra are based on $\{Z_t\}$, not clipped ones. Suppose that an observed stretch $\{Z_1, \cdots, Z_n\}$ (or only $\{X_1, \cdots, X_n\}$) is available. Then, we consider the problem classifying it into two categories Π_1, Π_2 described by spectra f_Z and g_Z:

$$\Pi_1 : f_Z(\lambda), \quad \Pi_2 : g_Z(\lambda).$$

Here we introduce a disparity measure $DM(f, g)$ by

$$DM(f_Z, g_Z) := \frac{1}{4\pi} \int_{[-\pi,\pi]} H\left(\frac{f_Z(\lambda)}{g_Z(\lambda)}\right) \mathrm{d}\lambda.$$

Remark 3.2.1. This class of disparity measure is sufficiently wide. Actually, if we take $H(x) = -log|x| + tr(x) - m$, then $DM(f, g)$ is I-divergence disparity measure (Whittle type disparity measure). If we take $H(x) = \log|\alpha x + (1-\alpha)| - \alpha \log|x|$, then $DM(f, g)$ is the Chernoff disparity measure.

We impose the following assumptions on $H(\cdot)$ and categories.

Assumption 3.2.1. $(H1)$ $H(\cdot)$ has a unique minimum at 1.
$(H2)$ $H(\cdot)$ is a holomorophic function in a neighborhood of 1.
$(H3)$ $H(1) = H'(1) = 0$, $H''(1) = c$.
(C) there exists a constant $d > 0$ such that $d \leq f_Z(\lambda),\ g_Z(\lambda)$.

$(H1)$ is for that $H(\cdot)$ leads to a distance between f and g, $(H2)$ is for the consistency, and $(H3)$ is for the evaluation of property when the two categories are the contiguous. (C) is not strong condition (e.g. Kakizawa (1997)).

Remark 3.2.2. Here, we deal with the divergence class of functionals of spectral density ratio $f_Z(\lambda)/g_Z(\lambda)$. Thus condition (C) is essential. If readers want to deal with no band-limited cases where the spectrum vanish over frequency subintervals, we can use the following important divergence measure;

$$DM(f_Z, g_Z) := \frac{1}{8\pi} \int_{[-\pi,\pi]} \{K(\lambda)\,(f_Z(\lambda) - g_Z(\lambda))\}^2 \mathrm{d}\lambda,$$

where K is a real function (see Taniguchi and Kakizawa (2000)).

For $\hat{f}_Z$ calculated from the clipped time series, we propose a classification statistic $D(\hat{f}_Z, f_Z, g_Z)$ by using above disparity measure

$$D(\hat{f}_Z, f_Z, g_Z) := DM(\hat{f}_Z, g_Z) - DM(\hat{f}_Z, f_Z),$$

and the classification rule is as follows:

If $D(\hat{f}_Z, f_Z, g_Z) > 0$ *(or* ≤ 0*), then* $\{Z_t\}$ *is classified into* Π_1 *(or* Π_2*),* (CR)

respectively. We show a basic property of the (CR) in view of misclassification probability.

Theorem 3.2.1. *Under Assumptions 3.1.1 and 3.2.1, we have*

$$\lim_{n\to\infty} \mathrm{P}(D < 0 \mid \Pi_1) = 0 \quad \text{and} \quad \lim_{n\to\infty} \mathrm{P}(D > 0 \mid \Pi_2) = 0.$$

Remark 3.2.3. This result shows the consistency of D. Many discrimination methods have this basic property (see Zhang and Taniguchi (1995), Kakizawa (1996)).

Thus, next, we evaluate a more delicate property of D. Consider the case when the two categories are contiguous:

$$\Pi_1 : f_Z(\lambda), \quad \Pi_2 : g_Z(\lambda) := f_Z(\lambda) + \frac{h(\lambda)}{\sqrt{n}},$$

where $h(\lambda) \in C^4([-\pi, \pi])$. Then we have,

Theorem 3.2.2. *Under Assumptions 3.1.1 and 3.2.1, it holds that*

$$\lim_{n\to\infty} \mathrm{P}(D < 0 \mid \Pi_1) = \lim_{n\to\infty} \mathrm{P}(D > 0 \mid \Pi_2) =$$
$$\Phi\left(-\frac{\int_{[-\pi,\pi]} h^2(\lambda) f_Z^{-2}(\lambda) d\lambda}{\sqrt{\sum_{t,\ell,r=-\infty}^{\infty} \tilde{f}(\ell)\tilde{f}(r)\sqrt{(1-\rho_Z^2(\ell))(1-\rho_Z^2(r))}Q(t,\ell,r)}}\right)$$

where $Q(t,\ell,r) := [16K_X(-\ell, t, t-r) + \{\rho_X(t)\rho_X(t+\ell-r) + \rho_X(t-r)\rho_X(t+\ell)\}]$ *and* $\tilde{f}(\ell) = 1/(2\pi)\int_{[-\pi,\pi]} h(\lambda) f_Z^{-1}(\lambda) e^{-i\ell\lambda} d\lambda.$

Remark 3.2.4. This result enables us to compare D with the other discrimination methods. However, generally, binary time series do not have this type of efficiency. Thus we discuss robustness against outliers in the next section.

3.3 Robustness against outlier

In this section, we deal with the case when the process of $\{Z_t\}$ is contaminated by an outlier. We assume that an observed stretch $\{Z_1, \cdots, Z_n\}$ is available, and show our discrimination method based on binary time series has a robustness against the outlier. However, the classical method based on the periodogram does not have this property. Let $\{Z_t^s\}$ be the process replaced Z_1 by $Z_1^s = s$.

First, we consider the classical method. For the original observations $\{Z_t\}$, the nonparametric estimator based on periodogram is defined by

$$\underline{f}_Z(\lambda) := \int_{[-\pi,\pi]} W_n(\lambda - \mu)\underline{I}_n(\mu)\mathrm{d}\mu, \tag{3.3.1}$$

where

$$\underline{I}_n(\lambda) := \frac{1}{2\pi n}\left|\sum_{\ell=1}^{n} Z_\ell \exp(-i\ell\lambda)\right|^2 .$$

Let DM be the I-divergence disparity measure for $\{Z_t\}$:

$$DM(f_Z, g_Z) := \frac{1}{4\pi}\int_{[-\pi,\pi]}\left(-\log\frac{|f(\lambda)|}{|g(\lambda)|} + \frac{f_Z(\lambda)}{g_Z(\lambda)} - 1\right)\mathrm{d}\lambda. \tag{3.3.2}$$

Then we have,

Theorem 3.3.1. *For fixed n and given $Z_2, \ldots, Z_n$, it holds that*

$$\lim_{s\to\infty} D(\underline{f}_{Z^s}, f_Z, g_Z) - D(\underline{f}_Z, f_Z, g_Z) = \begin{cases}\infty, & \int_{[-\pi,\pi]} \frac{1}{g_Z(\lambda)} - \frac{1}{f_Z(\lambda)}\mathrm{d}\lambda > 0\\ -\infty, & \int_{[-\pi,\pi]} \frac{1}{g_Z(\lambda)} - \frac{1}{f_Z(\lambda)}\mathrm{d}\lambda < 0.\end{cases}$$

Recall that $\hat{f}_Z$ is defined by (3.1.1), and that $\hat{f}_{Z^s}$ is the one replaced Z_t by Z_t^s from the clipped time series $\{X_t\}$. Then we have,

Theorem 3.3.2. *For fixed n and given $Z_2, \ldots, Z_n$, it holds that*

$$\lim_{s\to\infty} D(\hat{f}_{Z^s}, f_Z, g_Z) - D(\hat{f}_Z, f_Z, g_Z) = \begin{cases}0, & Z_1 \geq 0\\ c, & Z_1 < 0,\end{cases}$$

where c is a constant.

Remark 3.3.1. Theorems 3.3.1 and 3.3.2 show the classical method is sensitive to the outlier, but our new method is insensitive to it. Therefore, we conclude that our method has the robustness against the outlier.

3.4 Unknown category case

In this section, we deal with the practical case where the two spectra f_Z and g_Z, which describe two categories Π_1 and Π_2, are estimated from the training samples. Let $\{Z_{t,1}\}$ and $\{Z_{t,2}\}$ be the training samples of $\{Z_t\}$ under Π_1 and

Π_2, respectively. Assume that $\{Z_t\}$, $\{Z_{t,1}\}$, and $\{Z_{t,2}\}$ are mutually independent, and $\{Z_1, \cdots, Z_n\}$, $\{Z_{1,1}, \cdots, Z_{n,1}\}$, and $\{Z_{1,2}, \cdots, Z_{n,2}\}$ (or $\{X_1, \cdots, X_n\}$, $\{X_{1,1}, \cdots, X_{n,1}\}$, and $\{X_{1,2}, \cdots, X_{n,2}\}$, which are binary times series correspond to $\{Z_t\}$, $\{Z_{t,1}\}$, and $\{Z_{t,2}\}$, respectively) are available. We introduce nonparametric spectra estimator $\hat{f}^T_Z(\lambda)$ and $\hat{g}^T_Z(\lambda)$ in the same way as $\hat{f}_Z$. Then, we have the following theorem.

Theorem 3.4.1. *Under Assumptions 3.1.1 and 3.2.1, we have*

$$\lim_{n\to\infty} \mathrm{P}(D(\hat{f}_Z, \hat{f}^T_{Z_1}, \hat{g}^T_{Z_2}) < 0 \mid \Pi_1) = 0 \ \text{ and } \ \lim_{n\to\infty} \mathrm{P}(D(\hat{f}_Z, \hat{f}^T_{Z_1}, \hat{g}^T_{Z_2}) > 0 \mid \Pi_2) = 0.$$

Remark 3.4.1. The above result enables us to apply our method to practical situation.

Next, we elucidate that an outlier in the training samples deteriorates the classical method and, however, does not have a detrimental impact on our method. Let $\{Z_{t,1,s}\}$ and $\{Z_{t,2,s}\}$ be the process replaced $Z_{1,1}$ and $Z_{1,2}$ by s.

Theorem 3.4.2. *For fixed n and given $Z_2, \ldots, Z_n$, the following equations hold true:*

$$\text{(i)} \ \lim_{s\to\infty} D(\underline{f}_Z \text{ or } \underline{g}_Z, \underline{f}^T_{Z_{1,s}}, \underline{g}^T_{Z_2}) - D(\underline{f}_Z \text{ or } \underline{g}_Z, \underline{f}^T_{Z_1}, \underline{g}^T_{Z_2}) = -\infty$$
$$\text{(ii)} \ \lim_{s\to\infty} D(\underline{f}_Z \text{ or } \underline{g}_Z, \underline{f}^T_{Z_1}, \underline{g}^T_{Z_{2,s}}) - D(\underline{f}_Z \text{ or } \underline{g}_Z, \underline{f}^T_{Z_1}, \underline{g}^T_{Z_2}) = \infty$$

Theorem 3.4.3. *For fixed n and given $Z_2, \ldots, Z_n$, the following equations hold true:*

$$\text{(i)} \ \lim_{s\to\infty} D(\hat{f}_Z \text{ or } \hat{g}_Z, \hat{f}^T_{Z_{1,s}}, \hat{g}^T_{Z_2}) - D(\hat{f}_Z \text{ or } \hat{g}_Z, \hat{f}^T_{Z_1}, \hat{g}^T_{Z_2}) = \begin{cases} 0, & Z_{1,1} \geq 0 \\ c', & Z_{1,1} < 0 \end{cases},$$

where c' is a constant.

$$\text{(ii)} \ \lim_{s\to\infty} D(\hat{f}_Z \text{ or } \hat{g}_Z, \hat{f}^T_{Z_1}, \hat{g}^T_{Z_{2,s}}) - D(\hat{f}_Z \text{ or } \hat{g}_Z, \hat{f}^T_{Z_1}, \hat{g}^T_{Z_2}) = \begin{cases} 0, & Z_{1,2} \geq 0 \\ c'', & Z_{1,2} < 0 \end{cases},$$

where c'' is a constant.

Remark 3.4.2. These theorems show that if training sample $\{Z_{t,1}\}$ has a big outlier, then the classical method tends to classify the process into Π_2, and if $\{Z_{t,2}\}$ has a big outlier, then the classical method tends to classify the process into Π_1. However, our method has resistance to such a big outlier.

3.5 Simulation studies

This section provides simulation studies. We will see the robustness of the proposed method against the outlier. We generate $Z_1, \cdots, Z_n$ from the AR(1) model

$$Z_t = aZ_{t-1} + \epsilon_t,$$

where $\{\epsilon_t\} \overset{i.i.d.}{\sim} N(0,1)$. Let the two categories Π_1 and Π_2 be

$$\Pi_1 : \ f_Z(\lambda) := \frac{1}{2\pi|1 - a\exp(il\lambda)|^2}, \quad \Pi_2 : \ g_Z(\lambda) := \frac{1}{2\pi|1 - 0.2\exp(il\lambda)|^2}.$$

We choose a as $0.3, 0.5, 07$. Here, we use the I-divergence disparity measure, defined by (3.3.2), and Parzen's window function,

$$w(x) := \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{if } |x| < \frac{1}{2} \\ 2(1 - |x|)^3 & \text{if } \frac{1}{2} \leq |x| \leq 1 \\ 0 & \text{if } 1 < |x|. \end{cases}$$

The simulating procedure is as follows: first, we generate n (n = 100, 300, 500) samples $Z_1, \cdots, Z_n$ for each a = 0.3, 0.5, 0.7, and choose M (M = 5, 7, 8) corresponding to n ($n = 100, 300, 500$) respectively. Next, we replace Z_1 by $Z_1 = Z_1 - s \times \max\{Z_1, \cdots, Z_{500}\}$ for $s = 0, \pm 3, \pm 5$, then we computed the discriminant function $D(\hat{f}_{Z^s}, f_Z, g_Z)$ based on binary time series, and $D(\underline{f}_{Z^s}, f_Z, g_Z)$ based on the ordinary periodogram. Finally, we iterate 500 times and compute discriminant probability.

The results are shown in Table 3.1. Table 3.1 shows that the accuracies of our discriminant method are almost the same as that of the classical method for $s = 0$. For $s = \pm 3, \pm 5$, the detection probabilities of our method are still high. On the other hand, that of classical method are low. For $s = \pm 5$, all the correct classification probabilities of the classical method are lower than 50%. Thus, we conclude that our classification method is insensitive to the outlier. However, the classical one is sensitive. Comparing $s = 3$ and $s = -3$, these results have no significant difference. The same is true of $s = 5$ and $s = -5$.

3.6 Real data analysis

We apply our classification method to electrocardiogram (ECG) data. This dataset has two categories. One is the dataset recorded during normal heart beat, denoted by category A. The other is the dataset recorded during myocardial infraction, described as category B. This dataset (ECG200) was formatted by Olszewski

Table 3.1: Correct classification probabilities (%)

s	a	n	binary	periodogram
0	0.3	100	43.6	42.6
		300	60.0	63.2
		500	65.6	74.0
	0.5	100	67.8	72.8
		300	92.8	97.8
		500	96.8	99.8
	0.7	100	82.8	92.8
		300	99.6	100
		500	100	100
3	0.3	100	42.8	6.2
		300	59.0	20.4
		500	65.6	34.4
	0.5	100	66.0	6.0
		300	89.6	53.2
		500	96.8	86.6
	0.7	100	83.4	11.8
		300	99.6	85.8
		500	100	100
5	0.3	100	42.8	0.4
		300	59.0	0.4
		500	65.6	0.6
	0.5	100	66.0	0.2
		300	89.6	2.6
		500	96.8	14.6
	0.7	100	83.4	0.2
		300	99.6	7.8
		500	100	44.8

s	a	n	binary	periodogram
-3	0.3	100	42.8	6.4
		300	59.6	24.4
		500	65.4	37.6
	0.5	100	66.8	7.6
		300	90.4	53.8
		500	96.6	87.2
	0.7	100	81.3	10.6
		300	99.6	84.4
		500	100	100
-5	0.3	100	42.8	0.0
		300	59.6	1.4
		500	65.4	4.6
	0.5	100	66.8	0.0
		300	90.4	2.6
		500	96.6	19
	0.7	100	81.3	0.0
		300	99.6	8.2
		500	100	49.5

(2001) and is available from Dau et al. (2018). ECG200 consists of test data and training data, and the test data are composed of category A (36 time series, each length of series is 96) and category B (64 time series, each length of series is 96). Figures 3.1 and 3.2 show the plots of the data from category A and category B respectively. In this real data anlaysis, we will see whether the proposed method is valid for the real data or not and confirm the impact of the outliers.

The procedure is as follows: first, the series from categories A and B are denoted by $\{Z_{A,j}(t)\}$ and $\{Z_{B,i}(t)\}$ for $t = 1, \ldots, 96$, $j = 1, \cdots, 36$, and $i = 1, \cdots, 64$ respectively, and take difference $\{Z'_{A,j}(t+1)\} := \{Z_{A,j}(t+1) - Z_{A,j}(t)\}$ and $\{Z'_{B,i}(t+1)\} := \{Z_{B,i}(t+1) - Z_{B,i}(t)\}$ for $t = 1, \ldots, 95$ in order to obtain zero mean stationary sequences. Second, we estimate normalized spectral density ((spectral density)/(variance)) by nonparametric spectral estimator based on binary time series denoted by $\hat{f}_{A,j}(\lambda)$ and $\hat{f}_{B,i}(\lambda)$ and smoothed periodogram denoted by $\underline{f}_{A,j}(\lambda)$ and $\underline{f}_{B,i}(\lambda)$, respectively. Third, we compute the discriminant function $D(\hat{f}_{A,j'}(\lambda), \hat{f}_{A,j}(\lambda), \hat{f}_{B,i}(\lambda))$ and $D(\hat{f}_{B,i'}(\lambda), \hat{f}_{A,j}(\lambda), \hat{f}_{B,i}(\lambda))$ based on binary time series, and $D(\underline{f}_{A,j'}(\lambda), \underline{f}_{A,j}(\lambda), \underline{f}_{B,i}(\lambda))$ and $D(\underline{f}_{B,i'}(\lambda), \underline{f}_{A,j}(\lambda), \underline{f}_{B,i}(\lambda))$ based on the smoothed periodogram for $j, j' = 1, \cdots, 36$ and $i, i' = 1, \cdots, 64$ using I-divergence disparity measure. Finally, we compute the detection probability. Note that if the spectral density estimator takes negative values, the sequence is excluded from the dataset of both methods.

The detection probability of the proposed method is 60.25% and that of the classical method is 64.12%. These results shows the accuracy of both methods are almost the same. From the simulation study in Section 6, the results indicate the structures spectra of normal heart beat and myocardial infraction are different and these spectra possibly are not contiguous.

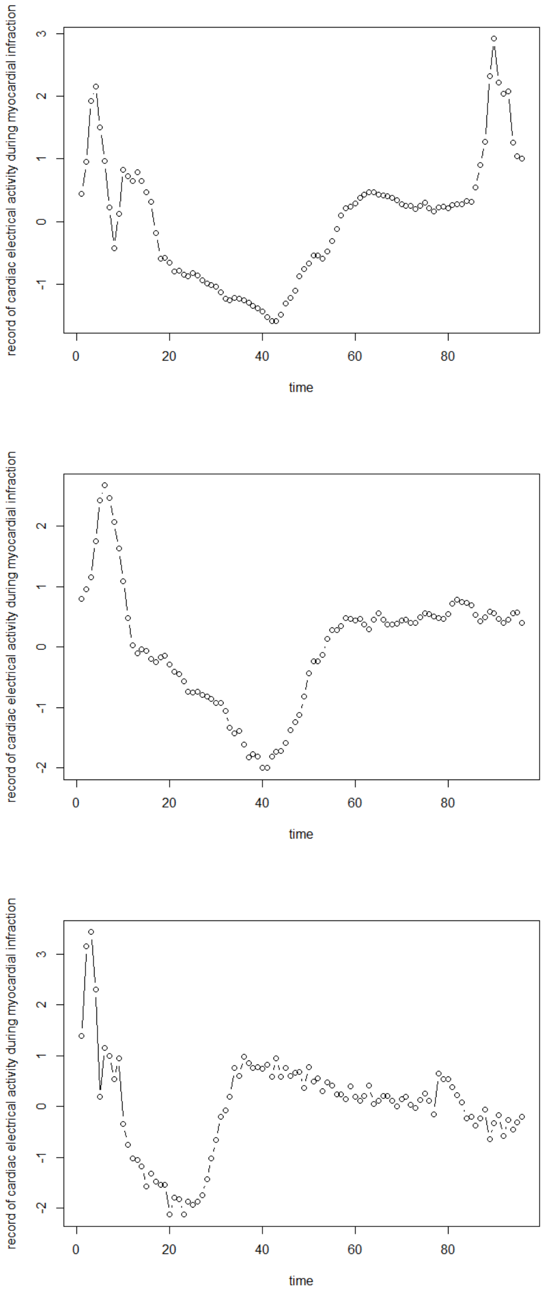

Figure 3.1: Plots of time series from category A

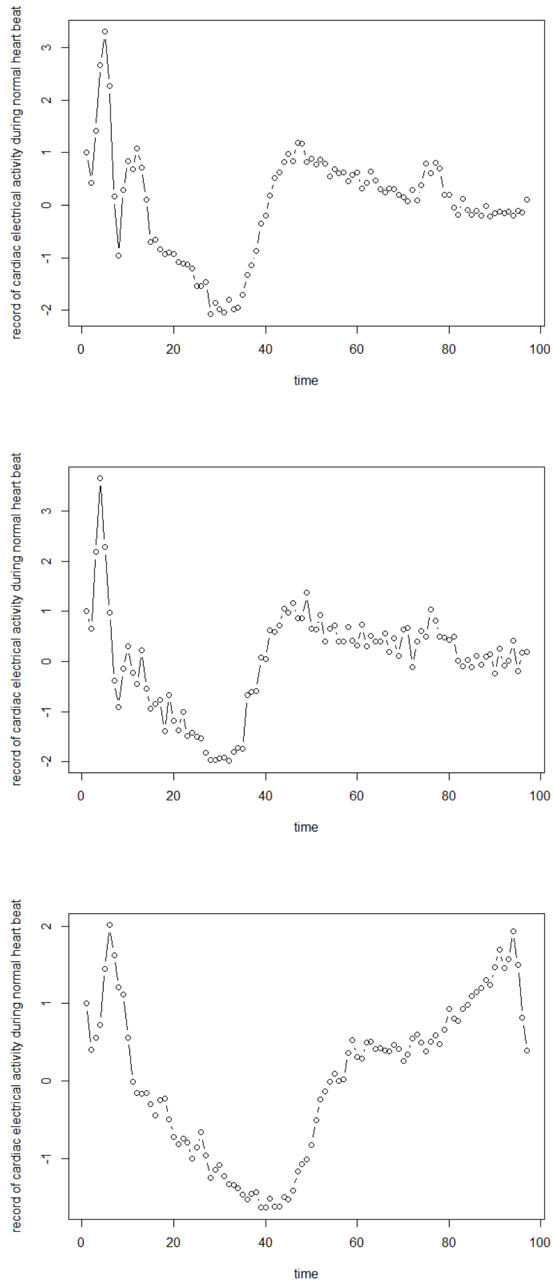

Figure 3.2: Plots of time series from category B

Next, we confirm Theorems 3.4.2 and 3.4.3. The procedure is as follows: First, the series from categories A and B are denoted by $\{Z_{A,j}(t)\}$ and $\{Z_{B,i}(t)\}$ for

$t = 1, \ldots, 96$, $j = 1, \cdots, 36$, and $i = 1, \cdots, 64$ respectively, and replace $\{Z_{A,1}(2)\}$ and $\{Z_{B,1}(2)\}$ as a and b, respectively (a and b are outliers). Second, we take difference $\{Z'_{A,j}(t+1)\}$ and $\{Z'_{B,i}(t+1)\}$ for $t = 1, \ldots, 95$. Third, we estimate normalized spectral densities $\hat{f}_{A,j}(\lambda)$, $\hat{f}_{B,i}(\lambda)$, $\underline{f}_{A,j}(\lambda)$, and $\underline{f}_{B,i}(\lambda)$. Finally, we compute the detection probabilities (I) $D(\hat{f}_{A,j}(\lambda), \hat{f}_{A,1}(\lambda), \hat{f}_{B,i}(\lambda))$ and $D(\underline{f}_{A,j}(\lambda), \underline{f}_{A,1}(\lambda), \underline{f}_{B,i}(\lambda))$ for $j = 2, \cdots, 36$, and $i = 1, \cdots, 64$ (the training sample of category A includes an outlier) respectively, and
(II) $D(\hat{f}_{A,j}(\lambda), \hat{f}_{A,j'}(\lambda), \hat{f}_{B,1}(\lambda))$, and $D(\underline{f}_{A,j}(\lambda), \underline{f}_{A,j'}(\lambda), \underline{f}_{B,1}(\lambda))$ for $j, j' = 1, \cdots, 36$ (the training sample of category B includes an outlier) respectively.

The result is in the Table 3.2. The case I in the Table 3.2 indicates if training sample $\{Z_{A,1}(t)\}$ has a big outlier, then the classical method tends to classify the process into Π_2. The case II in the Table 3.2 indicates if $\{Z_{B,1}(t)\}$ has a big outlier, then the classical method tends to classify the process into Π_1. On the other hand, the proposed method has resistance to such a big outlier (see Remark 3.4.2). Thus, we could confirm the results of Theorems 3.4.2 and 3.4.3 via ECG data. We conclude that the classification results of the classical method are unreliable and that of the proposed method are reliable when the observed time series have outliers

Table 3.2: Detection probabilities when the training sample is contaminated with an outlier

case I		
outlier	proposed method	classical method
$a = 0$	41.56%	55.17%
$a = 1$	50.06%	66.61%
$a = 2$	50.06%	26.71%
$a = 3$	50.06%	10.83%
$a = 5$	50.06%	3.33%
$a = 10$	50.06%	1.11%

case II		
outlier	proposed method	classical method
$b = 0$	59.21%	39.13%
$b = 1$	59.21%	44.22%
$b = 2$	65.45%	60.77%
$b = 3$	65.45%	74.40%
$b = 5$	65.45%	82.41%
$b = 10$	65.45%	90.32%

Chapter 4

Estimation of trigonometric moments for circular distribution of MA(p) type by using binary series

Directional statistics have received a great deal of interest in recent years, and a variety of distributions on the circle have been proposed. In this Chapter, we propose circular distributions of a moving average model of order p type which includes the cardioid distribution, and discuss estimation of trigonometric moments based on binary series. We give an explicit form of the root n consistent estimator based on clipped series, which enables us to construct an efficient estimator by the Newton–Raphson iterative method. We also show a robustness of the proposed estimator when the probability density function is contaminated with a noise term.

The Chapter is organized as follows: In Section 4.1, we introduce circular distributions of the moving average model of order p type and the estimator of trigonometric moments based on binary series for the proposed distribution. We show the asymptotic normality and compare the asymptotic variance with Cramér–Rao lower bound. In Section 4.2, we elucidate a robustness of the estimator when the probability density function is contaminated with noise. The finite sample performance of proposed estimator is investigated, and asymptotic normality of the proposed estimator is illustrated by computer simulation in Section 4.3. This Chapter is based on Goto (2020).

4.1 Settings and main result

In this section, we define a family of circular distributions of MA(p) type and propose a root n consistent estimator based on binary series. After that, we show the asymptotic normality and compare the asymptotic variance of the proposed estimator with Cramér–Rao lower bound.

Throughout this Chapter, we consider a family of circular distributions of MA(p) type whose probability density function is defined by

$$p_{\text{circ}}(\theta) = \frac{1}{2\pi(1+\phi_1^2+\cdots+\phi_p^2)}|\phi(e^{i\theta})|^2 \tag{4.1.1}$$

where $\phi(z) = 1+\phi_1 z+\phi_2 z^2+\cdots+\phi_p z^p$ and $\phi_j \in \mathbb{R}$ for any j.

Let $\{\Theta_k : k \in \mathbb{N}\}$ be independent random variables with a common circular distribution defined by (4.1.1). From the residue theorem and symmetry of (4.1.1), the j-th sine and cosine moments can be obtained as

$$\mathrm{E}\{\sin(j\Theta_k)\} = 0 \quad \text{for } j \in \mathbb{Z},$$
$$\mathrm{E}\{\cos(j\Theta_k)\} = \begin{cases} \frac{\phi_j+\phi_{j+1}\phi_1+\cdots+\phi_p\phi_{p-j}}{1+\phi_1^2+\cdots+\phi_p^2} & \text{for } |j| \le p, \\ 0 & \text{for } |j| \ge p+1, \end{cases}$$

respectively. Then, the mean resultant length and the mean direction of $\{\Theta_k : k \in \mathbb{N}\}$ can be obtained as

$$|\mathrm{E}\{e^{i\Theta_k}\}| = \left|\frac{\phi_1+\phi_2\phi_1+\cdots+\phi_p\phi_{p-1}}{1+\phi_1^2+\cdots+\phi_p^2}\right|,$$
$$\arg\mathrm{E}\{e^{i\Theta_k}\} = \begin{cases} 0 & \phi_1+\phi_2\phi_1+\cdots+\phi_p\phi_{p-1} > 0, \\ \pi & \phi_1+\phi_2\phi_1+\cdots+\phi_p\phi_{p-1} < 0, \\ \text{undefined} & \phi_1+\phi_2\phi_1+\cdots+\phi_p\phi_{p-1} = 0, \end{cases}$$

respectively. From Mardia and Jupp (2000, p.31), (4.1.1) can be written as

$$p_{\text{circ}}(\theta) = \frac{1}{2\pi}\left(1+\sum_{j=1}^{p}\eta_j\cos(j\theta)\right), \tag{4.1.2}$$

where $\eta_j = 2(\phi_j+\phi_{j+1}\phi_1+\cdots+\phi_p\phi_{p-j})/(1+\phi_1^2+\cdots+\phi_p^2)$. If we take $p=1$, (4.1.2) is the well-known cardioid distribution (see Mardia and Jupp (2000, p.45)). Clearly, if $\phi_j = 0$ for any $j \in \{1,\ldots,p\}$, (4.1.2) is a uniform distribution. The proposed model (4.1.1) is generally non-identifiable. Actually, for $p=2$ and $(\phi_1,\phi_2,\psi_1,\psi_2) := (0,-\frac{1}{2},\pm\sqrt{\frac{1}{2}},-1)$, we have $p(\theta;\phi_1,\phi_2) = p(\theta;\psi_1,\psi_2)$.

In this Chapter, we discuss the estimation problem of $\eta_1,\ldots,\eta_p$ of the proposed probability density function by using clipped series. Hereafter, we confine ourselves to the case that $(\phi_1,\ldots,\phi_p)$ satisfies $\phi_1+\phi_2\phi_1+\cdots+\phi_p\phi_{p-1} \ge 0$. Define $(\alpha_1,\alpha_2,\ldots,\alpha_p) \in \mathbb{R}^p$ such that $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_p < \pi$. For each α_j,

$j = 1, \ldots, p$, binary series $\{X_k^j\}$ are defined, for any $j = 1, \ldots, p$,

$$X_k{}^j = \begin{cases} 1 & -\alpha_j \le \Theta_k \le \alpha_j, \\ 0 & \text{otherwise.} \end{cases}$$

Applying the technique for the derivation of an orthant probability for normal distribution (see Kedem (1994, p.48)), we have the following equation

$$\begin{pmatrix} \mathrm{P}(-\alpha_1 \le \Theta_1 \le \alpha_1) \\ \mathrm{P}(-\alpha_2 \le \Theta_1 \le \alpha_2) \\ \vdots \\ \mathrm{P}(-\alpha_p \le \Theta_1 \le \alpha_p) \end{pmatrix} = \begin{pmatrix} \frac{\alpha_1}{\pi} \\ \frac{\alpha_2}{\pi} \\ \vdots \\ \frac{\alpha_p}{\pi} \end{pmatrix} + \frac{1}{2\pi} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pp} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_p \end{pmatrix},$$

where

$$\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pp} \end{pmatrix}$$
$$= \begin{pmatrix} \int_{-\alpha_1}^{\alpha_1} \cos\theta \mathrm{d}\theta & \int_{-\alpha_1}^{\alpha_1} \cos 2\theta \mathrm{d}\theta & \cdots & \int_{-\alpha_1}^{\alpha_1} \cos p\theta \mathrm{d}\theta \\ \int_{-\alpha_2}^{\alpha_2} \cos\theta \mathrm{d}\theta & \int_{-\alpha_2}^{\alpha_2} \cos 2\theta \mathrm{d}\theta & \cdots & \int_{-\alpha_2}^{\alpha_2} \cos p\theta \mathrm{d}\theta \\ \vdots & \vdots & \ddots & \vdots \\ \int_{-\alpha_p}^{\alpha_p} \cos\theta \mathrm{d}\theta & \int_{-\alpha_p}^{\alpha_p} \cos 2\theta \mathrm{d}\theta & \cdots & \int_{-\alpha_p}^{\alpha_p} \cos p\theta \mathrm{d}\theta \end{pmatrix}.$$

Here, we suppose the observed stretch $\{\Theta_1, \ldots, \Theta_n\}$ is available. We choose $(\alpha_1, \alpha_2, \ldots, \alpha_p) \in \mathbb{R}^p$ adequately so that $(b_{ij})_{i,j=1}^p$ is a nonsingular matrix, and substitute

$$\left(1/n \sum_{k=1}^n X_k{}^1, \ldots, 1/n \sum_{k=1}^n X_k{}^p \right)^{\mathrm{T}}$$

for

$$\left(\mathrm{P}(-\alpha_1 \le \Theta_1 \le \alpha_1), \ldots, \mathrm{P}(-\alpha_p \le \Theta_1 \le \alpha_p) \right)^{\mathrm{T}}.$$

Then, the binary estimator $\left(\hat{\eta}_1, \ldots, \hat{\eta}_p\right)^{\mathrm{T}}$ can be defined as

$$\begin{pmatrix} \hat{\eta}_1 \\ \hat{\eta}_2 \\ \vdots \\ \hat{\eta}_p \end{pmatrix} = 2\pi \begin{pmatrix} b^{11} & b^{12} & \cdots & b^{1p} \\ b^{21} & b^{22} & \cdots & b^{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b^{p1} & b^{p2} & \cdots & b^{pp} \end{pmatrix} \begin{pmatrix} \frac{1}{n} \sum_{k=1}^n X_k{}^1 - \frac{\alpha_1}{\pi} \\ \frac{1}{n} \sum_{k=1}^n X_k{}^2 - \frac{\alpha_2}{\pi} \\ \vdots \\ \frac{1}{n} \sum_{k=1}^n X_k{}^p - \frac{\alpha_p}{\pi} \end{pmatrix},$$

where $(b^{ij})_{i,j=1}^p$ is the inverse matrix of $(b_{ij})_{i,j=1}^p$.

Before we derive the asymptotic distribution of the proposed estimator, we give some examples that $(b_{ij})_{i,j=1}^p$ is a nonsingular matrix for specific models.

Example 4.1.1. MA(2) case: if we take $\alpha_1 = \frac{\pi}{4}$ and $\alpha_2 = \frac{\pi}{2}$, then

$$(b_{ij})_{i,j=1}^2 = \begin{pmatrix} \sqrt{2} & 1 \\ 2 & 0 \end{pmatrix}, \quad (b^{ij})_{i,j=1}^2 = \begin{pmatrix} 0 & \frac{1}{2} \\ 1 & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

Example 4.1.2. MA(3) case: if we take $\alpha_1 = \frac{\pi}{4}$, $\alpha_2 = \frac{\pi}{2}$, and $\alpha_3 = \frac{3\pi}{4}$, then

$$(b_{ij})_{i,j=1}^3 = \begin{pmatrix} \sqrt{2} & 1 & \frac{\sqrt{2}}{3} \\ 2 & 0 & -\frac{2}{3} \\ \sqrt{2} & -1 & \frac{\sqrt{2}}{3} \end{pmatrix}, \quad (b^{ij})_{i,j=1}^3 = \begin{pmatrix} \frac{1}{4\sqrt{2}} & \frac{1}{4} & \frac{1}{4\sqrt{2}} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ \frac{3}{4\sqrt{2}} & -\frac{3}{4} & \frac{1}{4\sqrt{2}} \end{pmatrix}.$$

The following theorem shows that the asymptotic normality of the proposed estimator.

Theorem 4.1.1. *It holds that*

$$\sqrt{n}\begin{pmatrix} \hat{\eta}_1 - \eta_1 \\ \hat{\eta}_2 - \eta_2 \\ \vdots \\ \hat{\eta}_p - \eta_p \end{pmatrix} \Rightarrow N(0, \boldsymbol{V}),$$

where $\boldsymbol{V} = (v_{ij})_{i,j=1}^p$ *and*

$$\begin{aligned} v_{ij} = \; & 4\pi^2 \textstyle\sum_{s,k=1}^p b^{is} b^{jk} \{\mathrm{P}(-\alpha_s \le \Theta_1 \le \alpha_s, -\alpha_k \le \Theta_1 \le \alpha_k) \\ & -\mathrm{P}(-\alpha_s \le \Theta_1 \le \alpha_s)\mathrm{P}(-\alpha_k \le \Theta_1 \le \alpha_k)\}. \end{aligned}$$

Next, we investigate whether our proposed method attains the Cramér–Rao lower bound or not. For simplicity, we confine ourselves to the case of circular distributions of MA(1) type.

Proposition 4.1.1. *The Cramér–Rao lower bound is given by*

$$\mathcal{I}^{-1}(\eta_1) = 1 - \eta_1^2 + \sqrt{1 - \eta_1^2}.$$

Proposition 4.1.1 enables us to compare the asymptotic variance of the proposed estimator with the Cramér–Rao lower bound. Thus, we have the following statement.

Remark 4.1.1. The Binary estimator is not efficient.

Actually, If we consider the case $\eta_1 = 1$, then it is easy to see that

$$(\text{Covariance of } \hat{\eta}_1) - \mathcal{I}^{-1}(\eta_1) > 0.$$

Remark 4.1.1 is not a preferable property of the estimator. However, from Hosoya and Taniguchi (1982, Theorem 5.1), we can construct an efficient estimator from $\hat{\eta}_1, \ldots, \hat{\eta}_p$ by the Newton–Raphson iterative method. In the next section, we show a robust property of the estimator when the true probability density function is contaminated.

4.2 Robustness of proposed estimator against contamination

In the previous section, we showed that proposed estimator is root n consistent, and it enable us to construct the efficient estimator by the Newton–Raphson iterative method. In this section, we show our estimator is robust when the true probability density function is contaminated with noise. Let $q_{\text{circ, contam}}(\cdot)$ be a contaminated probability density function defined, for $\theta \in [-\pi, \pi]$ and some $\beta \in (0, \pi/2)$, as

$$q_{\text{circ, contam}}(\theta) = \begin{cases} p_{\text{circ}}(\theta) & \text{if } -\pi + \beta \leq \theta \leq \pi - \beta, \\ c\xi(\theta) & \text{otherwise,} \end{cases}$$

where $p_{\text{circ}}(\theta)$ is defined by (4.1.1), $\xi(\theta)$ is a non-negative function with $\int_{\pi-\beta}^{\pi+\beta} \xi(\theta) \mathrm{d}\theta > 0$, c is some constant such that $q_{\text{circ, contam}}(\theta)$ is probability density function. In the above setting, $c\xi(\theta)$ corresponds to a noise. Assume that the process $\{\Theta_k : k \in \mathbb{N}\}$ is misspecified, that is, the true model of $\{\Theta_k : k \in \mathbb{N}\}$ comes from $q_{\text{circ, contam}}(\theta)$, but we fit the process to $p_{\text{circ}}(\theta)$.

Theorem 4.2.1. *If α_p and β satisfy $\alpha_p < \pi - \beta$, then the our estimator does not be influenced by the contamination.*

Thus, the proposed method is robust against the contamination of probability density.

4.3 Simulation studies

In this section, we study finite sample performance of the proposed method, and confirm the asymptotic normality of the proposed estimator based on binary process. In this simulation, the circular distributions of MA(1) and MA(2) types are discussed. First, we illustrate finite sample performance. The procedure is the following; first, we generate random variables $\{U_i : i = 1, \dots, n\}$ ($n = 100, 300, 500, 1000$), which follows i.i.d. standard uniform distribution. Next, we compute $\{\Theta_i = 1 \dots, n\} := \{F^{-1}(U_i) : i = 1, \dots, n\}$, where F^{-1} is the generalized inverse of a distribution function of (4.1.1), which follows the circular distribution of MA(p) type for $p = 1, 2$. Then, we calculate the proposed estimators of η_1 and η_2 for the each set of parameters $\phi_1 = 0.4, 0.7, -0.5$ and angulars $\alpha_1 = \pi/4, \pi/2, 3\pi/4$ for MA(1) type distribution, and $(\phi_1, \phi_2) = (0.7, 0.4), (1.0, 0.7), (0.9, -0.3)$ and $(\alpha_1, \alpha_2) = (\pi/4, \pi/2), (\pi/2, 3\pi/4)$ for MA(2) type distribution. We iterate 1000 times and calculate mean absolute error, defined as $\mathrm{MAE_j} := \sum_{k=1}^{1000} |\hat{\eta}_j^{(k)} - \eta_j|/n$ for $j = 1, 2$, where $\hat{\eta}_j^{(k)}$ is the estimator of η_j of k-th iteration. Next, we calculate, for $n = 1000$, $\{\sqrt{n}(\hat{\eta}_1^{(k)} - \eta_1);\ k = 1, \dots, 10000\}$ and $\{\sqrt{n}(\hat{\eta}_1^{(k)} - \eta_1), \sqrt{n}(\hat{\eta}_2^{(k)} - \eta_2);\ k = 1, \dots, 10000\}$ for circular distributions of MA(1) type with $\phi_1 = 0.7$ and MA(2) type with $(\phi_1, \phi_2) = (0.7, 0.4)$, respectively to confirm the asymptotic normality of the proposed estimator. Then, we give the Q-Q plots in Figures 4.1, 4.2, and 4.3. We also provide the Kolmogorov–Smirnov test of normality to check the asymptotic normality of the proposed estimator. The null hypothesis is that $\{\sqrt{n}(\hat{\eta}_1 - \eta_1)\}$ follows the normal distribution for large n. For $n = 100$, 1000, 10000, $\{\sqrt{n}(\hat{\eta}_1^{(k)} - \eta_1); k = 1, \dots, 100\}$ and $\{\sqrt{n}(\hat{\eta}_1^{(k)} - \eta_1), \sqrt{n}(\hat{\eta}_2^{(k)} - \eta_2);\ k = 1, \dots, 100\}$ are calculated for circular distributions of MA(1) type with $\phi_1 = 0.7$ and MA(2) type with $(\phi_1, \phi_2) = (0.7, 0.4)$. Then, we compute p-value by using R-function ks.test() when $\{\sqrt{n}(\hat{\eta}_1^{(k)} - \eta_1); k = 1, \dots, 100\}$ regarded as a set of i.i.d. observations with respect to k. Note that, from the definition of binary estimator, we possibly have the exact same value $\hat{\eta}_j^{(k)} = \hat{\eta}_j^{(k')}$ for some k and $k'(\neq k)$. Therefore, we added a small perturbation to $\{\sqrt{n}(\hat{\eta}_1^{(k)} - \eta_1); k = 1, \dots, 100\}$ by R function jitter() in order to compute p-value (see Robert et al. (2010, p.17-18)).

The results are shown in Tables 4.1 and 4.2 and Figures 4.1, 4.2, and 4.3. Tables 4.1 and 4.2 show the proposed estimator works well, and the mean absolute errors get smaller as the sample size increases. In Table 4.1, for $\phi_1 = 0.4$ and 0.7 in MA(1) type model, $\mathrm{MAE_1}$ is smallest when $\alpha_1 = 3\pi/4$ among $\alpha_1 = \pi/4,\ \pi/2,\ 3\pi/4$. On the other hand, for $\phi_1 = -0.5$ in MA(1) type model, $\mathrm{MAE_1}$ is smallest when $\alpha_1 = \pi/4$ among three angulars. It is because MA(1) model with $\phi_1 = -0.5$ has a mean direction π. The mean directions of the proposed model are 0 in the other cases. In Table 4.2, $\mathrm{MAE_1}$ are smaller than $\mathrm{MAE_2}$. For better estimation of ϕ_2, the set of angulars $(\pi/2, 3\pi/4)$ is better than $(\pi/4, \pi/2)$. Regarding to estimation of ϕ_1, both sets of angulars $(\pi/2, 3\pi/4)$ and $(\pi/4, \pi/2)$ provide almost the same $\mathrm{MAE_1}$. Figures 4.1 4.2, and 4.3 show that almost of all points are on the reference line, that is, we could confirm that our estimator has asymptotic normality. Moreover, for MA(1) model, the p-values of the KS test are obtained as 0.582, 0.987, 0.981 for n = 100, 1000 10000, respectively. For MA(2) model, the p-values of the KS test for $\{\sqrt{n}(\hat{\eta}_1^{(k)} - \eta_1); k = 1, \ldots, 100\}$ are obtained as 0.528, 0.507, 0.718 and that for $\{\sqrt{n}(\hat{\eta}_2^{(k)} - \eta_2); k = 1, \ldots, 100\}$ are obtained as 0.990, 0.799, 0.989 for n = 100, 1000 10000, respectively. As a result, it shows that we cannot reject the null hypothesis in all cases we investigated.

Table 4.1: MAE for circular distributions of MA(1) type

ϕ_1	α_1	n	MAE_1	ϕ_1	α_1	n	MAE_1
0.4	$\pi/4$	100	0.175	0.7	$\pi/4$	100	0.171
		300	0.103			300	0.108
		500	0.076			500	0.077
		1000	0.054			1000	0.055
	$\pi/2$	100	0.115		$\pi/2$	100	0.098
		300	0.065			300	0.060
		500	0.050			500	0.044
		1000	0.036			1000	0.032
	$3\pi/4$	100	0.106		$3\pi/4$	100	0.071
		300	0.059			300	0.038
		500	0.046			500	0.030
		1000	0.034			1000	0.022
-0.5	$\pi/4$	100	0.087				
		300	0.052				
		500	0.041				
		1000	0.029				
	$\pi/2$	100	0.113				
		300	0.065				
		500	0.049				
		1000	0.034				
	$3\pi/4$	100	0.180				
		300	0.103				
		500	0.077				
		1000	0.055				

Table 4.2: MAE for circular distributions of MA(2) type

(ϕ_1, ϕ_2)	(α_1, α_2)	n	MAE_1	MAE_2
(0.7,0.4)	$(\pi/4, \pi/2)$	100	0.081	0.215
		300	0.045	0.121
		500	0.036	0.099
		1000	0.026	0.069
	$(\pi/2, 3\pi/4)$	100	0.083	0.094
		300	0.046	0.055
		500	0.037	0.041
		1000	0.026	0.029
(1.0,0.7)	$(\pi/4, \pi/2)$	100	0.060	0.222
		300	0.035	0.129
		500	0.028	0.096
		1000	0.020	0.070
	$(\pi/2, 3\pi/4)$	100	0.064	0.070
		300	0.036	0.039
		500	0.027	0.032
		1000	0.019	0.021
(0.9,-0.3)	$(\pi/4, \pi/2)$	100	0.115	0.210
		300	0.066	0.125
		500	0.051	0.095
		1000	0.035	0.069
	$(\pi/2, 3\pi/4)$	100	0.111	0.155
		300	0.067	0.094
		500	0.052	0.075
		1000	0.036	0.052

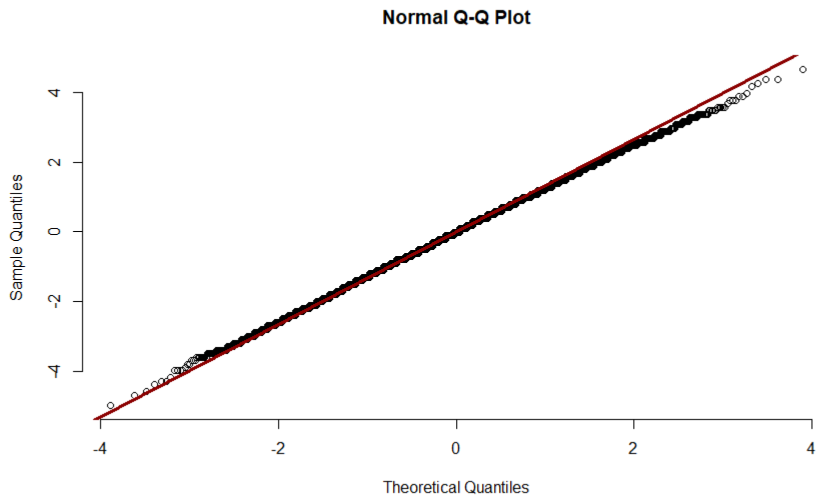

Figure 4.1: Q-Qplots of $\{\sqrt{n}(\hat{\eta}_1^{(k)} - \eta_1);\ k = 1, \ldots, 10000\}$ for a circular distribution of MA(1) type with $\phi_1 = 0.7$ for $n = 1000$.

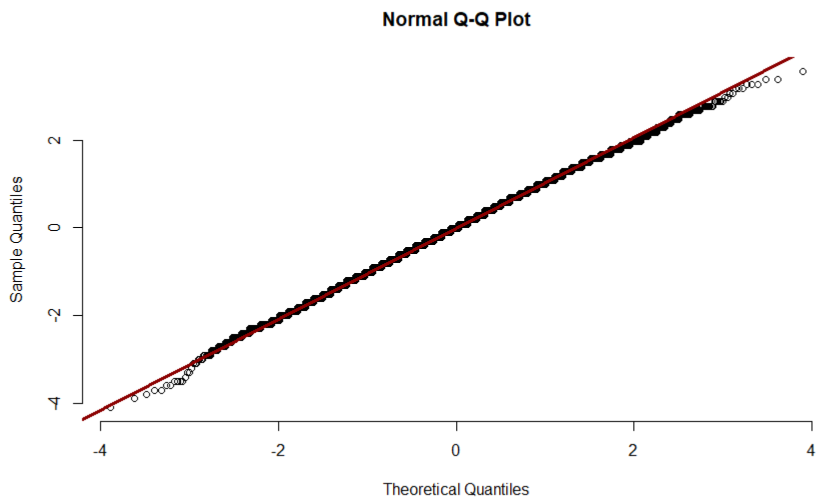

Figure 4.2: Q-Qplots of $\{\sqrt{n}(\hat{\eta}_1^{(k)} - \eta_1);\ k = 1, \ldots, 10000\}$ for a circular distribution of MA(2) type $(\phi_1, \phi_2) = (0.7, 0.4)$ for $n = 1000$.

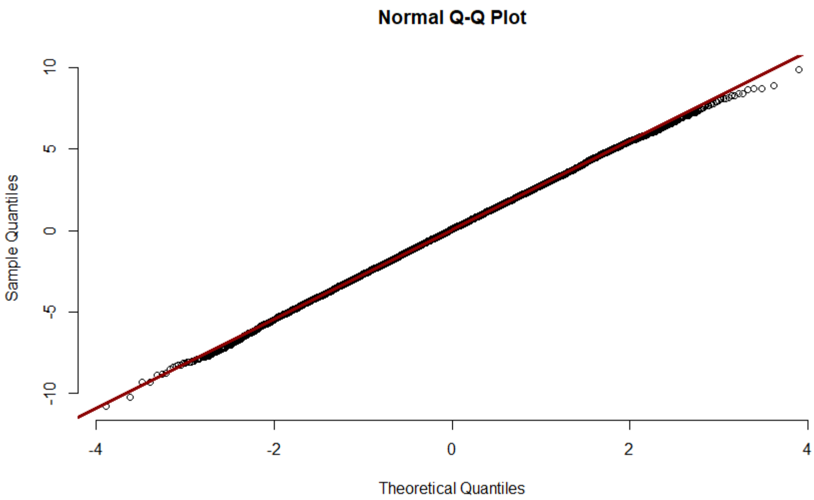

Figure 4.3: Q-Qplots of $\{\sqrt{n}(\hat{\eta}_2^{(k)} - \eta_2);\ k = 1, \ldots, 10000\}$ for a circular distribution of MA(2) type $(\phi_1, \phi_2) = (0.7, 0.4)$ for $n = 1000$.

Chapter 5

Distribution free tests for structural break of count time series

In this Chapter, testings for structural breaks of a count time series have been well studied under several distributional assumptions, including one-parameter exponential families and zero-inflated distributions. Here, we deal with the testing for parameter change for count time series whose intensity functions have nonlinear dependence structures without assuming any distributions. We derive the asymptotic null distribution for the Wald type, score based-, and residual-based CUSUM statistics and, consequently, obtain the distribution-free tests. We also show the test based on the modified Wald type statistic is consistent. A simulation study illustrates that the residual-based test outperforms the other proposed methods.

The Chapter is organized as follows: In Section 5.1, we introduce count processes with parametric intensities which have nonlinear dependence structures. Next, we define the Poisson, negative binomial, and exponential QMLEs and make assumptions. The strong consistency and asymptotic normality of QMLEs are reviewed. These results are shown by Ahmad and Francq (2016), Aknouche et al. (2018), and Aknouche and Francq (2020). In Section 5.2, we formulate the testing for structural breaks and define the Wald type, the modified Wald type, the score-based, and the residual-based CUSUM test statistics. We derive the limit laws of these statistics under the null hypothesis. We illustrate the finite sample performance in Section 5.3. The results reveal that the classical Wald type test is easily affected by the underlying conditional distribution. On the other hand, the proposed Wald tests work well. However, Wald type statistics show the size distortion because of the instability of PQMLE. In contrast, the score-based and the residual test provide good empirical size. Our simulation study suggests the residual-based test is superior to the score based test from the perspective of the power.

5.1 Quasi maximum likelihood estimators

In this section, the fundamental settings and the quasi maximum likelihood estimators (QMLEs) are introduced, which is investigated by Ahmad and Francq (2016), Aknouche et al. (2018), and Aknouche and Francq (2020).

Let $\{Z_t\}$ be a count time series or non-negative time series on the probability space $(\Omega, \mathcal{F}, P)$ with conditional expectation, for any $t \in \mathbb{Z}$

$$\mathrm{E}\left(Z_t|\mathcal{F}_{t-1}\right) := \lambda(Z_{t-1}, Z_{t-2}, \dots, ; \boldsymbol{\theta}_0)$$

where $\mathcal{F}_{t-1}$ is the σ-field generated by $\{Z_s, s \le t-1\}$, $\boldsymbol{\theta}_0$ is an unknown parameter on a parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^d$, and λ is a known measurable function on $[0,\infty)^\infty \times \boldsymbol{\Theta}$ to $(\delta, +\infty)$ for some $\delta > 0$.

For any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $t \in \mathbb{N} \cup \{0\}$, we define

$$\lambda_t(\boldsymbol{\theta}) := \lambda(Z_{t-1}, Z_{t-2}, \dots, ; \boldsymbol{\theta}), \quad \tilde{\lambda}_t(\boldsymbol{\theta}) := \lambda(Z_{t-1}, Z_{t-2}, \dots, Z_1, \mathbf{x_0}; \boldsymbol{\theta})$$

where $\mathbf{x_0} \in [0,\infty)^\infty$ be an initial parameter. The function $\tilde{\lambda}_t$ which can be calculate from observed process is proxy for λ_t which contains population values. Since we use specific models like the linear INGARCH(p,q) model as λ in practice, so $\mathbf{x_0} \in [0,\infty)^\infty$ reduces to a finite dimension vector. Moreover, the impact of the choice $\mathbf{x_0} \in [0,\infty)^\infty$ is asymptotically negligible as $n \to \infty$. Instead of the conditional distributional assumptions, we assume the stationarity and ergodicity of $\{Z_t\}$;

Assumption 5.1.1. **(A0)** $\{Z_t\}$ is strictly stationary and ergodic.

Remark 5.1.1. This assumption is satisfied by a broad class of the time series of counts. Aknouche and Francq (2020) showed sufficient conditions of Assumption (A0) for the non-linear INGARCH(p,q) model, and it includes the exponential family and the zero-inflated distributions.

The Poisson quasi (conditional) maximum likelihood estimator (PQMLE) is defined as follows;

$$\hat{\boldsymbol{\theta}}_n^{\mathrm{P}} := \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\max}\, \tilde{L}_n^{\mathrm{P}}, \quad \tilde{L}_n^{\mathrm{P}}(\boldsymbol{\theta}) := \frac{1}{n}\sum_{t=1}^{n} \tilde{\ell}_t^{\mathrm{P}}(\boldsymbol{\theta}), \quad \tilde{\ell}_t^{\mathrm{P}}(\boldsymbol{\theta}) := -\tilde{\lambda}_t(\boldsymbol{\theta}) + Z_t \log \tilde{\lambda}_t(\boldsymbol{\theta}).$$

Similarly, the negative binomial QMLE (NBQMLE) and the exponential QMLE (EQMLE) are defined below; for fixed $r > 0$,

$$\hat{\boldsymbol{\theta}}_{n,r}^{\mathrm{NB}} := \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\max}\, \tilde{L}_{n,r}^{\mathrm{NB}}, \quad \tilde{L}_{n,r}^{\mathrm{NB}}(\boldsymbol{\theta}) := \frac{1}{n}\sum_{t=1}^{n} \tilde{\ell}_{n,r}^{\mathrm{NB}}(\boldsymbol{\theta}),$$

$$\tilde{\ell}_{n,r}^{\mathrm{NB}}(\boldsymbol{\theta}) := r \log\left(\frac{r}{r + \tilde{\lambda}_t(\boldsymbol{\theta})}\right) + Z_t \log\left(\frac{\tilde{\lambda}_t(\boldsymbol{\theta})}{r + \tilde{\lambda}_t(\boldsymbol{\theta})}\right),$$

and

$$\hat{\boldsymbol{\theta}}_n^{\mathrm{E}} := \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \tilde{L}_n^{\mathrm{E}}, \quad \tilde{L}_n^{\mathrm{E}}(\boldsymbol{\theta}) := \frac{1}{n}\sum_{t=1}^{n} \tilde{\ell}_t^{\mathrm{E}}(\boldsymbol{\theta}), \quad \tilde{\ell}_t^{\mathrm{E}}(\boldsymbol{\theta}) := -\frac{Z_t}{\tilde{\lambda}_t(\boldsymbol{\theta})} - \log \tilde{\lambda}_t(\boldsymbol{\theta}),$$

respectively. We define ℓ_n^P, $\ell_{n,r}^{NB}$, and ℓ_n^E in the same way as $\tilde{\ell}_n^P$, $\tilde{\ell}_{n,r}^{NB}$, and $\tilde{\ell}_n^E$ using $\lambda_t(\boldsymbol{\theta})$ instead of $\tilde{\lambda}_t(\boldsymbol{\theta})$, respectively. Throughout this Chapter, $\|\cdot\|$ denotes ℓ_2 norm. We make the following assumptions;

Assumption 5.1.2. **(A1)** $\boldsymbol{\Theta}$ is a compact set and $\boldsymbol{\theta}_0$ belongs to the interior of $\boldsymbol{\Theta}$.

(A2) $\mathrm{E}Z_t^{1+\eta} < \infty$ for some $\eta < 0$.

(A3) $\lambda_t(\boldsymbol{\theta}) = \lambda_t(\boldsymbol{\theta}_0)$ a.s. if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ for any $t \in \mathbb{N}$.

(A4) For $a_t := \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\lambda_t(\boldsymbol{\theta}) - \tilde{\lambda}_t(\boldsymbol{\theta})\right|$, it holds that $a_t = o(1)$ and $a_t Z_t = o(1)$ a.s..

(A5) For any $t \in \mathbb{N}$, $\lambda_t(\boldsymbol{\theta})$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$.

(A6) If $\mathbf{s}^T \partial/\partial\boldsymbol{\theta}\lambda_t(\boldsymbol{\theta}_0) = 0$ a.s., then $\mathbf{s} = \mathbf{0}$.

(P7) For $b_t := \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\partial/\partial\boldsymbol{\theta}\left(\lambda_t(\boldsymbol{\theta}) - \tilde{\lambda}_t(\boldsymbol{\theta})\right)\right\|$ and

$$c_t := \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \max\left(\left\|\frac{1}{\lambda_t(\boldsymbol{\theta})}\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta})\right\|, \left\|\frac{1}{\tilde{\lambda}_t(\boldsymbol{\theta})}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\lambda}_t(\boldsymbol{\theta})\right\|\right),$$

there exists $\kappa > 1/2$ such that $b_t = O(t^{-\kappa})$, $b_t Z_t = O(t^{-\kappa})$, and $a_t c_t Z_t = O(t^{-\kappa})$ a.s..

(NB7) For $d_t := \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\lambda_t^2(\boldsymbol{\theta}) - \tilde{\lambda}_t^2(\boldsymbol{\theta})\right|$ and

$$e_t := \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \max\left(\left\|\frac{1}{\lambda_t(\boldsymbol{\theta})(r + \lambda_t(\boldsymbol{\theta}))}\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta})\right\|, \left\|\frac{1}{\tilde{\lambda}_t(\boldsymbol{\theta})(r + \tilde{\lambda}_t(\boldsymbol{\theta}))}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\lambda}_t(\boldsymbol{\theta})\right\|\right),$$

there exists $\kappa > 1/2$ such that $b_t = O(t^{-\kappa})$, $b_t Z_t = O(t^{-\kappa})$, $a_t e_t = O(t^{-\kappa})$, $a_t e_t Z_t = O(t^{-\kappa})$, and $d_t e_t Z_t = O(t^{-\kappa})$ a.s..

(E7) There exists $\kappa > 1/2$ such that $b_t = O(t^{-\kappa})$, $b_t Z_t = O(t^{-\kappa})$, $a_t c_t = O(t^{-\kappa})$ and $a_t c_t Z_t = O(t^{-\kappa})$ a.s..

(P8) For some neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ and any $i, j \in \{1, \ldots, d\}$, $\left|I_t^{\mathrm{P}}(\boldsymbol{\theta}_0)\right|^{1+\delta}$, $\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left|I_t^{\mathrm{P}}(\boldsymbol{\theta})\right|$, $\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left|J_t^{\mathrm{P}}(\boldsymbol{\theta})\right|$, and

$$\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left| \frac{(\lambda_t(\boldsymbol{\theta}_0) - \lambda_t(\boldsymbol{\theta}))}{\lambda_t(\boldsymbol{\theta})} \frac{\partial}{\partial \theta_i \partial \theta_j} \lambda_t(\boldsymbol{\theta}) \right|$$

are integrable for any $i, j \in \{1, \ldots, d\}$, where some $\delta > 0$,

$$I_t^{\mathrm{P}}(\boldsymbol{\theta}) := \frac{(Z_t - \lambda_t(\boldsymbol{\theta}))^2}{\lambda_t^2(\boldsymbol{\theta})} \frac{\partial}{\partial \theta_i} \lambda_t(\boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} \lambda_t(\boldsymbol{\theta})$$

and

$$J_t^{\mathrm{P}}(\boldsymbol{\theta}) := \frac{1}{\lambda_t(\boldsymbol{\theta})} \frac{\partial}{\partial \theta_i} \lambda_t(\boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} \lambda_t(\boldsymbol{\theta}).$$

In particular, we denote $I^{\mathrm{P}} := \mathrm{E} I_t^{\mathrm{P}}(\boldsymbol{\theta}_0) = \mathrm{E}\left(\frac{v_t(\boldsymbol{\theta})}{\lambda_t^2(\boldsymbol{\theta}_0)} \frac{\partial}{\partial \theta_i} \lambda_t(\boldsymbol{\theta}_0) \frac{\partial}{\partial \theta_j} \lambda_t(\boldsymbol{\theta}_0) \right)$, $J^{\mathrm{P}} := \mathrm{E} J_t^{\mathrm{P}}(\boldsymbol{\theta}_0)$ with the conditional variance $v_t(\boldsymbol{\theta}_0) := \mathrm{Var}(Z_t | \mathcal{F}_{t-1}) = \mathrm{E}(Z_t^2 | \mathcal{F}_{t-1}) - \lambda_t(\boldsymbol{\theta}_0)^2$.

(NB8) For some neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ and any $i, j \in \{1, \ldots, d\}$, $\left|I_t^{\mathrm{NB}}(\boldsymbol{\theta}_0)\right|^{1+\delta}$, $\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left|I_t^{\mathrm{NB}}(\boldsymbol{\theta})\right|$, $\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left|J_t^{\mathrm{NB}}(\boldsymbol{\theta})\right|$,

$$\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left| \left(\frac{r(\lambda_t(\boldsymbol{\theta}_0) - \lambda_t(\boldsymbol{\theta}))(2\lambda_t(\boldsymbol{\theta}) + r)}{\lambda_t^2(\boldsymbol{\theta})(r + \lambda_t(\boldsymbol{\theta}))^2} \right) \frac{\partial}{\partial \theta_i} \lambda_t(\boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} \lambda_t(\boldsymbol{\theta}) \right|,$$

$$\text{and } \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left| \frac{r(\lambda_t(\boldsymbol{\theta}_0) - \lambda_t(\boldsymbol{\theta}))}{\lambda_t(\boldsymbol{\theta})(r + \lambda_t(\boldsymbol{\theta}))} \frac{\partial}{\partial \theta_i \partial \theta_j} \lambda_t(\boldsymbol{\theta}) \right|$$

are integrable, where some $\delta > 0$,

$$I_t^{\mathrm{NB}}(\boldsymbol{\theta}) := \frac{r^2 (Z_t - \lambda_t(\boldsymbol{\theta}))^2}{\lambda_t^2(\boldsymbol{\theta})(r + \lambda_t(\boldsymbol{\theta}))^2} \frac{\partial}{\partial \theta_i} \lambda_t(\boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} \lambda_t(\boldsymbol{\theta}),$$

$$\text{and } J_t^{\mathrm{NB}}(\boldsymbol{\theta}) := \frac{r}{\lambda_t(\boldsymbol{\theta})(r + \lambda_t(\boldsymbol{\theta}))} \frac{\partial}{\partial \theta_i} \lambda_t(\boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} \lambda_t(\boldsymbol{\theta}).$$

In particular, we denote $I^{\mathrm{NB}} := \mathrm{E} I_t^{\mathrm{NB}}(\boldsymbol{\theta}_0)$, $J^{\mathrm{NB}} := \mathrm{E} J_t^{\mathrm{NB}}(\boldsymbol{\theta}_0)$.

(E8) For some neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ and any $i, j \in \{1, \ldots, d\}$, $\left|I_t^{\mathrm{E}}(\boldsymbol{\theta}_0)\right|^{1+\delta}$, $\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left|I_t^{\mathrm{E}}(\boldsymbol{\theta})\right|$, $\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left|J_t^{\mathrm{E}}(\boldsymbol{\theta})\right|$, and

$$\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left| \frac{\lambda_t(\boldsymbol{\theta}_0) - \lambda_t(\boldsymbol{\theta})}{\lambda_t^2(\boldsymbol{\theta})} \frac{\partial}{\partial \theta_i} \lambda_t(\boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} \lambda_t(\boldsymbol{\theta}) \right|$$

are integrable, where some $\delta > 0$,

$$I_t^{\mathrm{E}}(\boldsymbol{\theta}) := \frac{(Z_t - \lambda_t(\boldsymbol{\theta}))^2}{\lambda_t^4(\boldsymbol{\theta})} \frac{\partial}{\partial \theta_i} \lambda_t(\boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} \lambda_t(\boldsymbol{\theta}),$$

$$\text{and } J_t^{\mathrm{E}}(\boldsymbol{\theta}) := \frac{1}{\lambda_t^2(\boldsymbol{\theta})} \frac{\partial}{\partial \theta_i} \lambda_t(\boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} \lambda_t(\boldsymbol{\theta}).$$

In particular, we denote $I^{\mathrm{E}} := \mathrm{E} I_t^{\mathrm{E}}(\boldsymbol{\theta}_0)$, $J^{\mathrm{NB}} := \mathrm{E} J_t^{\mathrm{E}}(\boldsymbol{\theta}_0)$.

(A9) For some neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ and any $i, j \in \{1, \ldots, d\}$,

$$v_t^{1+\delta}(\boldsymbol{\theta}_0), \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left| \frac{\partial}{\partial \theta_i} \lambda_t(\boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} \lambda_t(\boldsymbol{\theta}) \right|, \text{ and } \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left| \frac{\partial}{\partial \theta_i} \lambda_t(\boldsymbol{\theta}) \right|$$

are integrable, where some $\delta > 0$.

Remark 5.1.2. (A1)-(A3) and (A5) are the fundamental assumptions to show the strong consistency and the asymptotic normality. (A4), (P7), (NB7), and (E7) are used to ensure that approximated processes based on $\tilde{\lambda}_t(\boldsymbol{\theta})$ converge to proper processes based on $\lambda_t(\boldsymbol{\theta})$. We can show that J^j is the invertible matrix by (A6). (P8), (NB8), and (E8) guarantee the existence of I^j and J^j and the almost surely convergences of $1/n \sum_{t=1}^n I_t^j(\hat{\boldsymbol{\theta}}_n)$ and $1/n \sum_{t=1}^n J_t^j(\hat{\boldsymbol{\theta}}_n)$ to I^j and J^j, respectively. (A9) is the moment conditions for the residual-based CUSUM test.

The strong consistency and the asymptotic normality is established by Ahmad and Francq (2016), Aknouche et al. (2018), and Aknouche and Francq (2020);

Lemma 5.1.1. *Assume that, for $j = \mathrm{P}, \mathrm{NB}$, or E, (A0)-(A6), (j7), and (j8). Then, it holds that,*

$$\hat{\boldsymbol{\theta}}_n^j \longrightarrow \boldsymbol{\theta}_0 \quad a.s.\ as\ n \longrightarrow \infty, \quad \sqrt{n}(\hat{\boldsymbol{\theta}}_n^j - \boldsymbol{\theta}_0) \Rightarrow N(\mathbf{0}, (J^j)^{-1} I^j (J^j)^{-1}) \quad as\ n \longrightarrow \infty.$$

Remark 5.1.3. If $Z_t | \mathcal{F}_{t-1} \sim \mathrm{Pois}(\lambda_t(\boldsymbol{\theta}_0))$, $\mathrm{NB}(r, r/(r + \lambda_t(\boldsymbol{\theta}_0)))$, or $\mathrm{Exp}(1/\lambda_t(\boldsymbol{\theta}_0))$, then $I^P = J^P$, $I^{NB} = J^{NE}$, or $I^E = J^E$, respectively. Hence, QMLE is efficient for each case.

5.2 Detection of structural breaks

In the previous section, we introduced QMLEs and confirmed the strong consistency and the asymptotic normality. In this section, we deal with tests for structural breaks, which is the main object in this Chapter. The null hypothesis H_0 is there is no change point of the true parameter, and the alternative H_1 is that the true parameter changes after an unknown point $\lfloor n\tau \rfloor$ where $\tau \in (0,1)$. More precisely, the null hypothesis is given by

$$H_0: \ \mathrm{E}\left(Z_t|\mathcal{F}_{t-1}\right) := \lambda(Z_{t-1}, Z_{t-2}, \ldots, ; \boldsymbol{\theta}_0) \quad \text{for any } t \in \{1, \ldots, n\},$$

and the alternative is, for some $\boldsymbol{\theta}_1 (\neq \boldsymbol{\theta}_0)$,

$$\begin{aligned} H_1: \mathrm{E}\left(Z_t|\mathcal{F}_{t-1}\right) &:= \lambda(Z_{t-1}, Z_{t-2}, \ldots, ; \boldsymbol{\theta}_0) \quad \text{for any } t \in \{1, \ldots, \lfloor n\tau \rfloor\}, \\ \mathrm{E}\left(Z'_t|\mathcal{F}'_{t-1}\right) &:= \lambda(Z'_{t-1}, Z'_{t-2}, \ldots, ; \boldsymbol{\theta}_1) \quad \text{for any } t \in \{\lfloor n\tau \rfloor + 1, \ldots, n\}, \end{aligned}$$

where $\mathcal{F}_{t-1}$ and $\mathcal{F}'_{t-1}$ are the σ-field generated by $\{Z_s, s \leq t-1\}$ and $\{Z'_s, s \leq t-1\}$, respectively.

5.2.1 Wald type CUSUM test

First, we propose the Wald type CUSUM test which is investigated by Kang and Lee (2014), Lee et al. (2016), and Lee et al. (2018). From the definition of the QMLEs and the Taylor's expansion, it follows that, for $j = \mathrm{P}, \mathrm{NB},$ or E,

$$\begin{aligned} \mathbf{0} &= \sqrt{n} \frac{\partial}{\partial \boldsymbol{\theta}} \tilde{L}_n^j(\hat{\boldsymbol{\theta}}_n^j) \\ &= \sqrt{n} \frac{\partial}{\partial \boldsymbol{\theta}} \tilde{L}_n^j(\boldsymbol{\theta}_0) + \frac{\partial}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \tilde{L}_n^j(\boldsymbol{\theta}_n^{j*}) \sqrt{n}(\hat{\boldsymbol{\theta}}_n^j - \boldsymbol{\theta}_0), \end{aligned}$$

where $\boldsymbol{\theta}_0 \leqslant \boldsymbol{\theta}_n^{j*} \leqslant \hat{\boldsymbol{\theta}}_n^j$. Then, we have

$$J^j \sqrt{n}(\hat{\boldsymbol{\theta}}_n^j - \boldsymbol{\theta}_0) = \sqrt{n} \frac{\partial}{\partial \boldsymbol{\theta}} \tilde{L}_n^j(\boldsymbol{\theta}_0) + \Delta_n^j,$$

where

$$\Delta_n^j := \begin{cases} -\left(J^j + \frac{\partial}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \tilde{L}_n^j(\boldsymbol{\theta}_n^{j*})\right) \left(\frac{\partial}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \tilde{L}_n^j(\boldsymbol{\theta}_n^{j*})\right)^{-1} \sqrt{n} \frac{\partial}{\partial \boldsymbol{\theta}} \tilde{L}_n^j(\boldsymbol{\theta}_0) & \text{if } \left(\frac{\partial}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \tilde{L}_n^j(\boldsymbol{\theta}_n^{j*})\right)^{-1} \text{ exists}, \\ \left(J^j + \frac{\partial}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \tilde{L}_n^j(\boldsymbol{\theta}_n^{j*})\right) \sqrt{n}(\hat{\boldsymbol{\theta}}_n^j - \boldsymbol{\theta}_0) & \text{otherwise.} \end{cases} \tag{5.2.1}$$

Since

$$\begin{aligned} &J^j \frac{\lfloor ns \rfloor}{\sqrt{n}} (\hat{\boldsymbol{\theta}}_{\lfloor ns \rfloor}^j - \hat{\boldsymbol{\theta}}_n^j) \\ =& \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor ns \rfloor} \frac{\partial}{\partial \boldsymbol{\theta}} \tilde{\ell}_t^j(\boldsymbol{\theta}_0) - \frac{\lfloor ns \rfloor}{n} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \tilde{\ell}_t^j(\boldsymbol{\theta}_0) + \sqrt{\frac{\lfloor ns \rfloor}{n}} \Delta_{\lfloor n\tau \rfloor}^j - \frac{\lfloor ns \rfloor}{n} \Delta_n^j, \end{aligned}$$

we know $J^j \lfloor ns \rfloor / \sqrt{n}(\hat{\boldsymbol{\theta}}^j_{\lfloor ns \rfloor} - \hat{\boldsymbol{\theta}}^j_n)$ takes the form of the CUSUM statistics. Under the assumption that I^j for $j = \mathrm{P}, \mathrm{NB}$, or E is positive definite matrix, let us define the Wald type test statistics as, for $j = \mathrm{P}, \mathrm{NB}$, or E,

$$T^j_{\mathrm{KL,Wald}} := \max_{1 \le k \le n} \frac{k^2}{n} (\hat{\boldsymbol{\theta}}^j_k - \hat{\boldsymbol{\theta}}^j_n)^T \hat{J}^j_{\mathrm{KL}} (\hat{I}^j_{\mathrm{KL}})^{-1} \hat{J}^j_{\mathrm{KL}} (\hat{\boldsymbol{\theta}}^j_k - \hat{\boldsymbol{\theta}}^j_n),$$

where

$$\hat{I}^{\mathrm{j}}_{\mathrm{KL}} := \frac{1}{n} \sum_{t=1}^{n} \tilde{I}^{\mathrm{j}}_t(\hat{\boldsymbol{\theta}}^{\mathrm{j}}_n), \quad \hat{J}^{\mathrm{j}}_{\mathrm{KL}} := \frac{1}{n} \sum_{t=1}^{n} \tilde{J}^{\mathrm{j}}_t(\hat{\boldsymbol{\theta}}^{\mathrm{j}}_n),$$

and $\tilde{I}^{\mathrm{j}}_t$ and $\tilde{J}^{\mathrm{j}}_t$ is defined in the same way as I^{j}_t and J^{j}_t using $\lambda_t(\boldsymbol{\theta})$ instead of $\tilde{\lambda}_t(\boldsymbol{\theta})$, respectively. Then, we have the following asymptotic null distribution.

Theorem 5.2.1. *Assume that, for* $j = \mathrm{P}, \mathrm{NB}$, *or* E, I^j *is positive definite matrix, (A0)-(A6), and (j7)-(j8). Then, under the null hypothesis H_0, it holds that*

$$T^j_{\mathrm{KL,Wald}} \Rightarrow \sup_{0 \le s \le 1} \|B^\circ_d(s)\|^2 \quad \text{as } n \to \infty,$$

where $B^\circ_d(s)$ is d-dimensional standard Brownian bridge.

From Theorem 5.2.1, we can construct the distribution-free CUSUM test based on $T^j_{\mathrm{KL,Wald}}$ which rejects the null hypothesis H_0 when $T^j_{\mathrm{KL,Wald}} \ge C$, where C is a critical value defined as follows; for significance level α such that $0 < \alpha < 1$, $\mathrm{P}\left(\sup_{0 \le s \le 1} \|B^\circ_d(s)\|^2 > C\right) = \alpha$. Then, this is asymptotic size α test. However, the consistency of this test is difficult to show since the asymptotics of $\hat{\boldsymbol{\theta}}^j_n$ has to under the alternative does not take into consideration. As an alternative, we propose Doukhan and Kengne (2015) type of statistics. Let us define the alternative statistics as follows; for $j = \mathrm{P}, \mathrm{NB}$, or E,

$$T^j_{\mathrm{DK,Wald}} := \max_{1 \le k \le n-1} \frac{k^2(n-k)^2}{n^3} (\hat{\boldsymbol{\theta}}^j_{1:k} - \hat{\boldsymbol{\theta}}^j_{k+1:n})^T \hat{J}^j_{\mathrm{DK}} (\hat{I}^j_{\mathrm{DK}})^{-1} \hat{J}^j_{\mathrm{DK}} (\hat{\boldsymbol{\theta}}^j_{1:k} - \hat{\boldsymbol{\theta}}^j_{k+1:n}),$$

where

$$\hat{I}^{\mathrm{j}}_{\mathrm{DK}} := \frac{1}{u_n} \sum_{t=1}^{u_n} \tilde{I}^{\mathrm{j}}_t(\hat{\boldsymbol{\theta}}^{\mathrm{j}}_{1:u_n}), \quad \hat{J}^{\mathrm{j}}_{\mathrm{DK}} := \frac{1}{u_n} \sum_{t=1}^{u_n} \tilde{J}^{\mathrm{j}}_t(\hat{\boldsymbol{\theta}}^{\mathrm{j}}_{1:u_n}),$$

u_n is an integer value sequence such that $u_n \to \infty$ and $u_n/n \to 0$ as $n \to \infty$, and $\hat{\boldsymbol{\theta}}^j_{a:b}$ is the jQMLE estimator based on $\{Z_a, \ldots, Z_b\}$ for $j = \mathrm{P}, \mathrm{NB}$, or E. Then, we have the asymptotic results.

Theorem 5.2.2. *Assume that, for* j = P, NB, *or* E, I^j *is positive definite matrix, (A0)-(A6), and* $(j7)$*-*$(j8)$*. Then, under the null hypothesis* H_0*, it holds that*

$$T^j_{\mathrm{DK,Wald}} \longrightarrow \sup_{0\le s\le 1} \|B^\circ_d(s)\|^2 \quad \text{as } n \to \infty,$$

where $B^\circ_d(s)$ *is* d*-dimensional standard Brownian bridge.*

From Theorem 5.2.2, we therefore obtain a distribution-free and asymptotic size α test if we reject H_0 when $T^j_{\mathrm{DK,Wald}} \ge C$. As expected, we obtain the consistency of the test.

Theorem 5.2.3. *Assume that, for* j = P, NB, *or* E, I^j *is positive definite matrix, (A0)-(A6), and* $(j7)$*-*$(j8)$*. Then, the test if we reject* H_0 *when* $T^j_{\mathrm{DK,Wald}} \ge C$*, where* C *is a critical value given by* $\mathrm{P}\left(\sup_{0\le s\le 1} \|B^\circ_d(s)\|^2 > C\right) = \alpha$ *is consistent.*

From Theorems 5.2.1-5.2.3, we conclude that we can construct the asymptotic size α CUSUM test based on $T^j_{\mathrm{KL,Wald}}$ and the asymptotic size α and consistent CUSUM test based on $T^j_{\mathrm{DK,Wald}}$. However, these Wald type tests often encounters size distortion (see Lee and Lee (2019)). Therefore, we examine more stable tests, that is, the score based and residual based CUSUM tests, which provide better size control than Wald type tests.

5.2.2 Score based CUSUM test

A score based statistics is studied by Berkes et al. (2004), Oh and Lee (2018), and Lee and Lee (2019). We define the alternative statistics as follows. For j = P, NB, or E,

$$T^j_{\mathrm{score}} := \max_{1\le k\le n} \frac{1}{n} \left(\sum_{t=1}^{k} \frac{\partial}{\partial \boldsymbol{\theta}} \tilde{\ell}^j_k(\hat{\boldsymbol{\theta}}^j_n) \right)^T (\hat{I}^j_{\mathrm{KL}})^{-1} \left(\sum_{t=1}^{k} \frac{\partial}{\partial \boldsymbol{\theta}} \tilde{\ell}^j_k(\hat{\boldsymbol{\theta}}^j_n) \right).$$

The reason why T^j_{score} is called the CUSUM test is

$$\max_{1\le k\le n} 1/\sqrt{n} \left(\sum_{t=1}^{k} \frac{\partial}{\partial \boldsymbol{\theta}} \tilde{\ell}^j_k(\hat{\boldsymbol{\theta}}^j_n) \right)$$

can be approximated by

$$\max_{1\le k\le n} \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{k} \frac{\partial}{\partial \boldsymbol{\theta}} \tilde{\ell}^j_t(\boldsymbol{\theta}_0) - \frac{k}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \tilde{\ell}^j_t(\boldsymbol{\theta}_0) \right),$$

which is proven in Lemma 7.4.3 in Section 6.

Theorem 5.2.4. *Assume that, for* j = P, NB, *or* E, I^j *is positive definite matrix, (A0)-(A6), and (j7)-(j8). Then, under the null hypothesis* H_0*, it holds that*

$$T^j_{\text{score}} \Rightarrow \sup_{0 \le s \le 1} \|B^\circ_d(s)\|^2 \quad \text{as } n \to \infty,$$

where $B^\circ_d(s)$ *is d-dimensional standard Brownian bridge.*

Similar to the above Wald type test, we obtain a distribution-free and asymptotic size α test which rejects the hypothesis H_0 whenever $T^j_{\text{score}} \ge C$ where C is a critical value defined by $\mathrm{P}\left(\sup_{0\le s\le 1} \|B^\circ_d(s)\|^2 > C\right) = \alpha$.

5.2.3 Residual based CUSUM test

Franke et al. (2012) and Kang and Lee (2014) discussed the residual based CUSUM test. Define $\epsilon_t := Z_t - \lambda(\boldsymbol{\theta}_0)$, $\tilde{\epsilon}_t(\hat{\boldsymbol{\theta}}^j_n) := Z_t - \tilde{\lambda}(\hat{\boldsymbol{\theta}}^j_n)$, and the residual based CUSUM statistics as

$$T^j_{\text{res}} := \max_{1\le k\le n} \frac{1}{\sqrt{\frac{1}{n}\sum_{t=1}^n \tilde{\epsilon}^2_t(\hat{\boldsymbol{\theta}}^j_n)}} \frac{1}{\sqrt{n}} \left| \sum_{t=1}^k \tilde{\epsilon}_t(\hat{\boldsymbol{\theta}}^j_n) - \frac{k}{n}\sum_{t=1}^n \tilde{\epsilon}_t(\hat{\boldsymbol{\theta}}^j_n) \right|.$$

Theorem 5.2.5. *Assume that, for* j = P, NB, *or* E, I^j *is positive definite matrix, (A0)-(A6), (j7)-(j8), and (A9). Then, it holds that, under* H_0*,*

$$T^j_{\text{res}} \Rightarrow \sup_{0\le s\le 1} |B^\circ_1(s)| \quad \text{as } n \to \infty.$$

Similarly, we obtain a distribution-free and asymptotic size α test rejects the hypothesis H_0 in favour of H_1 if $T^j_{\text{res}} \ge C'$, where C' is a critical value given by $\mathrm{P}\left(\sup_{0\le s\le 1} |B^\circ_1(s)| > C'\right) = \alpha$.

5.3 The simulation studies

In this section, we investigate the finite sample performance of the proposed tests. We use the linear INGARCH(1,1) model;

$$\lambda_t = \omega + \alpha Z_{t-1} + \beta\lambda_{t-1}.$$

The critical values are calculated by generating 10000 realizations of the standard Brownian bridge by the R function *BBridge* in the R package *sde* (Iacus, 2016) and these are 1.336995 and 3.011263 for $\max_{k=1,\dots,10000} |B^\circ_1(k/10000)|$ and $\max_{k=1,\dots,10000} \|B^\circ_3(k/10000)\|^2$ at a significance level of 0.05, respectively. The

INGARCH process is generated using the analytic mean of Z_t for the model as initial values of Z_0 and λ_0 by the R function *ingarch.mean* in the R package *tscount* (Liboschik et al., 2017).

Since PQMLE is not stable for small samples (see Figure 5.1), we used the modified statistics $T^{\mathrm{P}}_{\mathrm{KL,Wald2}}$ and $T^{\mathrm{P}}_{\mathrm{DK,Wald2}}$ which are defined as

$$T^{\mathrm{P}}_{\mathrm{KL,Wald2}} := \max_{v_n \le k \le n} \frac{k^2}{n}(\hat{\boldsymbol{\theta}}^{\mathrm{P}}_k - \hat{\boldsymbol{\theta}}^{\mathrm{P}}_n)^T \hat{J}^{\mathrm{P}}_{\mathrm{KL2}}(\hat{I}^{\mathrm{P}}_{\mathrm{KL}})^{-1}\hat{J}^{\mathrm{P}}_{\mathrm{KL}}(\hat{\boldsymbol{\theta}}^{\mathrm{P}}_k - \hat{\boldsymbol{\theta}}^{\mathrm{P}}_n),$$

$$T^{\mathrm{P}}_{\mathrm{DK,Wald2}} := \max_{v_n \le k \le n-v_n} \frac{k^2(n-k)^2}{n^3}(\hat{\boldsymbol{\theta}}^{\mathrm{P}}_{1:k} - \hat{\boldsymbol{\theta}}^{\mathrm{P}}_{k+1:n})^T \hat{J}^{\mathrm{P}}_{\mathrm{DK}}(\hat{I}^{\mathrm{P}}_{\mathrm{DK}})^{-1}\hat{J}^{\mathrm{P}}_{\mathrm{DK}}(\hat{\boldsymbol{\theta}}^{\mathrm{P}}_{1:k} - \hat{\boldsymbol{\theta}}^{\mathrm{P}}_{k+1:n}),$$

where $v_L = \lfloor (\log n)^2 \rfloor$, respectively. In this simulation, we also use the following test statistics

$$T^{\mathrm{P}}_{\mathrm{KL}} := \max_{v_L \le k \le n} \frac{k^2}{n}(\hat{\boldsymbol{\theta}}^{\mathrm{P}}_k - \hat{\boldsymbol{\theta}}^{\mathrm{P}}_n)^T (\hat{I}^{P}_{\mathrm{KL}})(\hat{\boldsymbol{\theta}}^{\mathrm{P}}_k - \hat{\boldsymbol{\theta}}^{\mathrm{P}}_n),$$

which is proposed by Kang and Lee (2014) under the conditional distribution of the process being Poisson, to investigate the impact of the misspecification.

The procedure is the following; First, We investigate the empirical sizes of the tests. we generate n ($n = 300, 600, 900$) samples of the Poisson or negative binomial INGARCH (1,1) with $r = 4$ for negative binomial distribution and $(\omega, \alpha, \beta) = (1, 0.3, 0.2)$. Next, we estimate the parameters $\hat{\boldsymbol{\theta}}^{\mathrm{P}}_{1:k}$ and $\hat{\boldsymbol{\theta}}^{\mathrm{P}}_{k+1:n}$ by PQMLE with the initial values $\sum_{t=1}^{k} Z_t/k$ and $\sum_{t=k+1}^{n} Z_t/(n-k)$ for $k = v_L, \ldots, n$ and $k = v_L, \ldots, n - v_L$, respectively. The optimization is obtained by use of the R package *constrOptim* given the gradient function and $\mathbf{0}$ as the initial value of the gradient. Then, we calculate the proposed test statistics, and replicate this procedure 200 times and calculate the rejection probabilities.

Second, we simulate the cases the parameters changes from
$(\omega, \alpha, \beta) = (1, 0.3, 0.2)$ to $(1, 0.3, 0.4)$ and $(1.5, 0.3, 0.2)$ at $\lfloor n/2 \rfloor + 1$ to study the empirical powers of the test statistics, respectively. The rest of the procedure is the same as the above null case.

The results are summarized in Table 5.1 and 5.2. Here, we denote Poisson distribution and negative binomial distribution with the parameter r as Pois(λ_t) and NB($\lambda_t, r/(r + \lambda_t)$), respectively. The classical statistics $T^{\mathrm{P}}_{\mathrm{KL}}$ are sensitive to the misspecification, the non-Poisson case. On the other hand, the alternative statistics $T^{\mathrm{P}}_{\mathrm{KL,Wald2}}$ and $T^{\mathrm{P}}_{\mathrm{DK,Wald2}}$ work well since we need not specify the underlying distribution. However, as Lee and Lee (2019) pointed out, these Wald type statistics show the severe size distortions in Table 5.1. Although the modified Wald statistics $T^{\mathrm{P}}_{\mathrm{DK,Wald2}}$ are mathematically tractable, it provides the worst size control. This can be explained from the perspective of the instability of PQMLE. We can confirm PQMLE based on small samples is not stable, and Wald type test statistics are calculated through PQMLEs which are based on small samples. These facts cause the size distortions.

In contrast, $T^{\mathrm{P}}_{\mathrm{score}}$ and $T^{\mathrm{P}}_{\mathrm{res}}$ are based on PQMLE which are calculated from full samples. Thus, these statistics achieve better size control. According to the empirical power, Table 5.2 shows the test based on the residual has better power than the test based on the score.

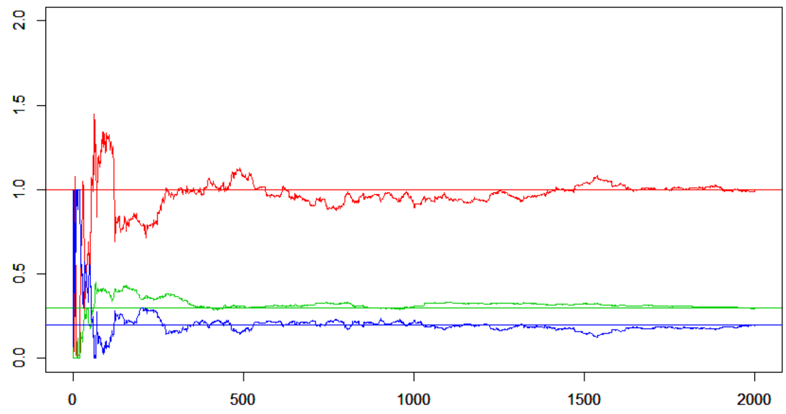

Figure 5.1: PQMLE based on $\{Z_1, \ldots, Z_k\}$ for $\lambda_t = 1 + 0.3Z_{t-1} + 0.2\lambda_{t-1}$; x-axis is the number of observations (k). y-axis is the estimated values ($\hat{\theta}^{\mathrm{P}}_{k,1}$,$\hat{\theta}^{\mathrm{P}}_{k,2}$,$\hat{\theta}^{\mathrm{P}}_{k,3}$) for (0.1,0.3,0.2).

Table 5.1: Empirical sizes at the nominal size $\alpha = 0.05$

$\lambda_t = 1 + 0.3Z_{t-1} + 0.2\lambda_{t-1}$ for all n						
Conditional distribution	sample size	$T^{\mathrm{P}}_{\mathrm{KL,Wald2}}$	$T^{\mathrm{P}}_{\mathrm{KL}}$	$T^{\mathrm{P}}_{\mathrm{DK,Wald2}}$	$T^{\mathrm{P}}_{\mathrm{score}}$	$T^{\mathrm{P}}_{\mathrm{res}}$
Pois(λ_t)	$n = 300$	0.160	0.145	0.240	0.025	0.025
	$n = 600$	0.160	0.170	0.215	0.040	0.035
	$n = 900$	0.095	0.080	0.220	0.020	0.020
NB(λ_t, $4/(4 + \lambda_t)$)	$n = 300$	0.175	0.970	0.305	0.035	0.045
	$n = 600$	0.185	0.980	0.290	0.055	0.065
	$n = 900$	0.155	0.995	0.260	0.020	0.020

Table 5.2: Empirical powers at the nominal size $\alpha = 0.05$

$\lambda_t = 1 + 0.3Z_{t-1} + 0.2\lambda_{t-1}$ for the first half of n $\lambda_t = 1 + 0.3Z_{t-1} + 0.4\lambda_{t-1}$ for the latter half of n						
Conditional distribution	sample size	$T^{\mathrm{P}}_{\mathrm{KL,Wald2}}$	$T^{\mathrm{P}}_{\mathrm{KL}}$	$T^{\mathrm{P}}_{\mathrm{DK,Wald2}}$	$T^{\mathrm{P}}_{\mathrm{score}}$	$T^{\mathrm{P}}_{\mathrm{res}}$
	$n = 300$	0.965	0.970	0.745	0.445	0.770
Pois(λ_t)	$n = 600$	1.000	1.000	0.935	0.965	0.995
	$n = 900$	1.000	1.000	0.985	1.000	1.000
	$n = 300$	0.795	1.000	0.610	0.255	0.345
NB($\lambda_t, 4/(4 + \lambda_t)$)	$n = 600$	0.890	1.000	0.660	0.575	0.800
	$n = 900$	0.995	1.000	0.895	0.885	0.975

$\lambda_t = 1 + 0.3Z_{t-1} + 0.2\lambda_{t-1}$ for the first half of n $\lambda_t = 1.5 + 0.3Z_{t-1} + 0.2\lambda_{t-1}$ for the latter half of n						
Conditional distribution	sample size	$T^{P}_{\mathrm{KL,Wald}}$	T^{P}_{KL}	$T^{P}_{\mathrm{DK,Wald}}$	T^{P}_{score}	T^{P}_{res}
	$n = 300$	0.840	0.840	0.840	0.540	0.755
Pois(λ_t)	$n = 600$	0.975	0.990	0.975	0.955	0.990
	$n = 900$	1.000	1.000	0.980	0.995	1.000
	$n = 300$	0.555	0.995	0.645	0.225	0.340
NB($\lambda_t, 4/(4 + \lambda_t)$)	$n = 600$	0.720	1.000	0.790	0.575	0.695
	$n = 900$	0.850	1.000	0.890	0.830	0.895

Chapter 6

Likelihood ratio processes under non-standard settings

Chapter 6 investigates the log-likelihood ratio for curved models and the one-way random effect ANOVA model. Local asymptotic normality (LAN) is the specific form of the asymptotic expansion of the log-likelihood ratio. Once we obtain the LAN results for models, optimal tests and estimations can be constructed by means of central sequences. More precisely, for example, the lower bound of loss functions can be obtained among regular estimators, and the power of the tests can be derived explicitly from the null distribution by the Le Cam's third lemma. Here, we show the simultaneous equation system, which plays a significant role in econometrics, has LAN property. Hence, we can construct an efficient estimator and an asymptotically maximin test. In contrast, we elucidate that the one-way ANOVA model does not have LAN property. The limiting behavior is out of the common. Thereby, we cannot use the existing theory to construct optimal statistical methods. By use of the Neyman–Pearson lemma, we show our test is the asymptotically most powerful.

The Chapter is organized as follows: In Section 6.1, we deal with the two types of curved normal families. The first one is normal distributions whose mean and variance are governed by the same parameter. The second one is the simultaneous equation system. The regression parameter of the reduced form of the system is endowed with a curved structure. For the two models, we show LAN property. In Section 6.2, we focus on one-way random effect ANOVA models and derive the limit distribution of the log-likelihood ratio when the variance of the random effect belongs to the boundary and interior of parameter space, respectively. After that, we show our test based on the log-likelihood ratio is asymptotically most powerful.

6.1 Local asymptotic normality for curved models

In this section, we investigate the local asymptotic normality (LAN) for curved models. Before getting to the main subject, we confirm the definition of LAN property. For a sequence of probability measures $\{P_{\boldsymbol{\theta}}^{(n)}\}$, $\{P_{\boldsymbol{\theta}}^{(n)}\}$ said to be local asymptotic normal (LAN) at $\boldsymbol{\theta}$ if the log-likelihood ratio admits the following expansion; For any vector $\boldsymbol{h}$ and any sequence of matrix $\{\boldsymbol{\tau}_n\}$ such that $\|\tau_n\| \to 0$ as $n \to \infty$, it holds that

$$\log \frac{\mathrm{dP}_{\boldsymbol{\theta}+\boldsymbol{\tau}_n \boldsymbol{h}}^{(n)}}{\mathrm{dP}_{\boldsymbol{\theta}}^{(n)}} = \boldsymbol{h}^T \boldsymbol{\Delta}_n - \frac{1}{2}\boldsymbol{h}^T \mathcal{I}(\boldsymbol{\theta})\boldsymbol{h} + o_p(1) \quad \text{as } n \to \infty,$$

where $\mathcal{I}(\boldsymbol{\theta})$ is the Fisher information matrix and

$$\boldsymbol{\Delta}_n \Rightarrow N(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta})) \quad \text{as } n \to \infty.$$

The random variable $\boldsymbol{\Delta}_n$ and the sequence $\boldsymbol{\tau}_n$ are called the central sequence and the contiguity order, respectively. Especially, $\mathcal{I}^{-1}(\boldsymbol{\theta})\boldsymbol{\Delta}_n$ is called the central sequence. As discussed at Introduction, many models including ARMA, ARCH, and CHARN hold LAN. Usually, the contiguity order is $1/\sqrt{n}$.

First, we consider the normal distribution $\mathrm{N}(\theta^\alpha, \theta^\beta)$, where $\theta > 0$ and $(\alpha, \beta) \in \mathbb{R}^2$, that is, mean and variance have the curved structure. We are interested in the hypothesis testing when parameter is contiguous;

$$\mathrm{H}_0^{(n)} : \theta = \theta_0 (> 0), \quad \mathrm{K}_0^{(n)} : \theta = \theta_n := \theta_0 + \frac{h}{\sqrt{n}} (> 0),$$

where $h > -\sqrt{n}\theta_0$. Assume that, for each n, $\{X_{ni}, i = 1, \ldots, n, n = 1, 2, \ldots\}$ is available and $\{X_{ni}\}$ is independent for each n and i. We denote, for each n, the sequence of hypothesis as $\mathrm{N}_{\theta_0}^{(n)}$ for $\{X_{ni}, i = 1, \ldots, n\}$ under $\mathrm{H}_0^{(n)}$ and $\mathrm{N}_{\theta_n}^{(n)}$ for $\{X_{ni}, i = 1, \ldots, n\}$ under $\mathrm{K}_0^{(n)}$. Then, the next theorem shows, under $\mathrm{H}_0^{(n)}$, the log-likelihood ratio $\Lambda_0(\theta_0, \theta_n) := \log \mathrm{dN}_{\theta_n}^{(n)} / \mathrm{dN}_{\theta_0}^{(n)}$ admits the asymptotic expansion.

Theorem 6.1.1. *Under the null hypothesis $\mathrm{H}_0^{(n)}$, $\{\mathrm{N}_{\theta_0}^{(n)}, \theta_0 > 0\}$ is local asymptotically normal, that is, for $h \in \mathbb{R}$ and sufficiently large n such that $h > -\sqrt{n}\theta_0$, it holds that*

$$\Lambda_0(\theta_0, \theta_n) = h\Delta_n^0 - \frac{1}{2}h^2 \mathcal{I}_0(\theta_0) + o_p(1) \quad \textit{as } n \to \infty,$$

where $\mathcal{I}_0(\theta_0)$ denotes the Fisher information defined as $\mathcal{I}_0(\theta_0) := \beta^2\theta_0^{-2}/2 + \alpha^2\theta_0^{2\alpha-\beta-2}$ and Δ_n^0 denotes the central sequence defined as

$$\Delta_n^0 := \begin{pmatrix} \frac{\sqrt{n}}{h}\left(\theta_n^{\alpha-\beta} - \theta_0^{\alpha-\beta}\right) & -\frac{\sqrt{n}}{2h}\left(\theta_n^{-\beta} - \theta_0^{-\beta}\right) \end{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} X_i - \theta_0^{a} & X_i^2 - \theta_0^{2\alpha} - \theta_0^{\beta} \end{pmatrix}^T,$$

$$= \left((\alpha - \beta)\theta_0^{\alpha-\beta-1} \quad \frac{\beta}{2}\theta_0^{-\beta-1} \right) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(X_i - \theta_0^{a} \quad X_i^2 - \theta_0^{2\alpha} - \theta_0^{\beta} \right)^T + o_p(1),$$

which converges to normal distribution with mean zero and variance $\mathcal{I}_0(\theta_0)$ as $n \to \infty$.

The LAN property of a family of the curved normal distribution is given in Theorem 6.1.1. However, one may think it is not natural to consider curved models. Therefore, we discuss the simultaneous equation system, which is naturally endowed with the curved structure. The model plays an essential role in econometrics; see Anderson et al. (1986), Hosoya et al. (1989), and the references therein.

Let a single structural equation be

$$\boldsymbol{Y}_1 = \boldsymbol{Y}_2\boldsymbol{\beta} + \boldsymbol{Z}_1\boldsymbol{\gamma} + \boldsymbol{\epsilon}$$

and the reduced form of the system of equation be

$$\left(\boldsymbol{Y}_1 \quad \boldsymbol{Y}_2\right) = \left(\boldsymbol{Z}_1 \quad \boldsymbol{Z}_2\right) \begin{pmatrix} \boldsymbol{\zeta}_{11} & \boldsymbol{\zeta}_{12} \\ \boldsymbol{\zeta}_{21} & \boldsymbol{\zeta}_{22} \end{pmatrix} + \left(\boldsymbol{\nu}_1 \quad \boldsymbol{\nu}_2\right)$$

where $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are $n \times 1$ and $n \times p_1$ matrices of endogenous variables, $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are $n \times p_2$ and $n \times p_3$ matrices of exogenous variables, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\zeta}_{11}$,$\boldsymbol{\zeta}_{21}$, $\boldsymbol{\zeta}_{12}$, and $\boldsymbol{\zeta}_{22}$ are $p_1 \times 1$, $p_2 \times 1$, $p_2 \times 1$, $p_3 \times 1$, $p_2 \times p_1$, $p_3 \times p_1$ matrices of coefficients, and $\boldsymbol{\epsilon}$, $\boldsymbol{\nu}_1$, and $\boldsymbol{\nu}_2$ are $n \times 1$, $n \times 1$, and $n \times p_1$ vectors of errors, respectively. Assume that $\{\boldsymbol{Y}_1 := (Y_{11}, \ldots, Y_{1n})'\}$, $\{\boldsymbol{Y}_2 := (\boldsymbol{Y}_{21}, \ldots, \boldsymbol{Y}_{2n})'; \boldsymbol{Y}_{2j} := (Y_{2j1}, \ldots, Y_{2jp_1}), j = 1, \ldots, n\}$, $\{\boldsymbol{Z}_1 := (\boldsymbol{Z}_{11}, \ldots, \boldsymbol{Z}_{1n})'; \boldsymbol{Z}_{1j} := (Z_{1j1}, \ldots, Z_{1jp_2}), j = 1, \ldots, n\}$, and $\{\boldsymbol{Z}_2 := (\boldsymbol{Z}_{21}, \ldots, \boldsymbol{Z}_{2n})'; \boldsymbol{Z}_{2j} := (Z_{2j1}, \ldots, Z_{2jp_3}), j = 1, \ldots, n\}$ are available.

We make the following assumption;

Assumption 6.1.1. *(S1) the ranks of $(\boldsymbol{\zeta}_{21}\ \boldsymbol{\zeta}_{22})$ and $\boldsymbol{\zeta}_{22}$ are p_1.*

(S2) the rows of $(\boldsymbol{\nu}_1\ \boldsymbol{\nu}_2)$ are independent of each other.

(S3) each row of $(\boldsymbol{\nu}_1\ \boldsymbol{\nu}_2)$ follows normal distribution with mean 0 *and covariance matrix*

$$\Omega := \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix}.$$

(S4) Ω is a non-singular matrix and the inverse matrix Ω^{-1} is written by

$$\Omega^{-1} := \begin{pmatrix} \omega^{11} & \omega^{12} \\ \omega^{21} & \omega^{22} \end{pmatrix}.$$

(S5)

$$\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\boldsymbol{Z}_{1i}{}^{T}\\ \boldsymbol{Z}_{2i}{}^{T}\end{pmatrix}\begin{pmatrix}\boldsymbol{Z}_{1i} & \boldsymbol{Z}_{2i}\end{pmatrix} = \boldsymbol{M} + o_p(1).$$

Then, under Assumption 6.1.1, it can be seen that $\boldsymbol{\epsilon} = \boldsymbol{\nu}_1 - \boldsymbol{\nu}_2\boldsymbol{\beta}, \boldsymbol{\gamma} = \boldsymbol{\zeta}_{11} - \boldsymbol{\zeta}_{12}\boldsymbol{\beta}$, $\boldsymbol{\zeta}_{21} = \boldsymbol{\zeta}_{22}\boldsymbol{\beta}$, the components of $\boldsymbol{\epsilon}$ are independent of each other and follow normal distribution with mean 0 and variance $\sigma^2_{\boldsymbol{\beta}} := \omega_{11} - 2\omega_{12}\boldsymbol{\beta} + \boldsymbol{\beta}'\omega_{22}\boldsymbol{\beta}$. Therefore, the reduced form of the system of equation can be written as

$$\begin{pmatrix}\boldsymbol{Y}_1 & \boldsymbol{Y}_2\end{pmatrix} = \begin{pmatrix}\boldsymbol{Z}_1 & \boldsymbol{Z}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{\zeta}_{12}\boldsymbol{\beta} + \boldsymbol{\gamma} & \boldsymbol{\zeta}_{12}\\ \boldsymbol{\zeta}_{22}\boldsymbol{\beta} & \boldsymbol{\zeta}_{22}\end{pmatrix} + \begin{pmatrix}\boldsymbol{\nu}_1 & \boldsymbol{\nu}_2\end{pmatrix}$$

Here, we suppose that we interested in $\boldsymbol{\theta} := (\boldsymbol{\beta}, \boldsymbol{\gamma})^T$. We consider the testing problem that the null hypothesis is

$$\mathrm{H}_1^{(n)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0 := \begin{pmatrix}\boldsymbol{\beta}_0\\ \boldsymbol{\gamma}_0\end{pmatrix},$$

and the alternative is

$$\mathrm{K}_1^{(n)} : \boldsymbol{\theta} = \boldsymbol{\theta}_n := \begin{pmatrix}\boldsymbol{\beta}_n\\ \boldsymbol{\gamma}_n\end{pmatrix}, := \begin{pmatrix}\boldsymbol{\beta}_0 + \frac{\boldsymbol{h}_1}{\sqrt{n}}\\ \boldsymbol{\gamma}_0 + \frac{\boldsymbol{h}_2}{\sqrt{n}}\end{pmatrix}$$

where $\boldsymbol{h} \in \mathbb{R}^{p_1}$. For each n, the sequence of hypotheses are denoted as $\mathrm{N}^{(n)}_{\boldsymbol{\theta}_0}$ for $(\boldsymbol{Y}_1\ \boldsymbol{Y}_2)$ under $\mathrm{H}_1^{(n)}$ and $\mathrm{N}^{(n)}_{\boldsymbol{\theta}_n}$ for $(\boldsymbol{Y}_1\ \boldsymbol{Y}_2)$ under $\mathrm{K}_1^{(n)}$, respectively. Then, we have the LAN theorem for the simultaneous equation system.

Theorem 6.1.2. *Under Assumption 6.1.1 and the null hypothesis* $\mathrm{H}_1^{(n)}$, $\{\mathrm{N}^{(n)}_{\boldsymbol{\theta}_0}\}$ *is local asymptotically normal, that is, for* $h \in \mathbb{R}$, *it holds that*

$$\begin{aligned}\Lambda_1(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) &:= \log \frac{d\mathrm{N}^{(n)}_{\boldsymbol{\theta}_n}}{d\mathrm{N}^{(n)}_{\boldsymbol{\theta}_0}}\\ &= \begin{pmatrix}\boldsymbol{h}_1{}^T & \boldsymbol{h}_2{}^T\end{pmatrix}\Delta_n^1 - \frac{1}{2}\begin{pmatrix}\boldsymbol{h}_1{}^T & \boldsymbol{h}_2{}^T\end{pmatrix}\mathcal{I}_1(\boldsymbol{\theta}_0)\begin{pmatrix}\boldsymbol{h}_1\\ \boldsymbol{h}_2\end{pmatrix} + o_p(1) \quad \text{as } n \to \infty,\end{aligned}$$

where

$$\mathcal{I}_1(\boldsymbol{\theta}_0) := \omega^{11}\begin{pmatrix}\boldsymbol{\zeta}_{12} & I_{p_2\times p_2}\\ \boldsymbol{\zeta}_{22} & \boldsymbol{0}_{p_3\times p_2}\end{pmatrix}^T \boldsymbol{M}\begin{pmatrix}\boldsymbol{\zeta}_{12} & I_{p_2\times p_2}\\ \boldsymbol{\zeta}_{22} & \boldsymbol{0}_{p_3\times p_2}\end{pmatrix}$$

is the Fisher information matrix of the following model with respect to $\boldsymbol{\theta}_0$;

$$\begin{pmatrix} Y_{1i} & \boldsymbol{Y}_{2i} \end{pmatrix} = \begin{pmatrix} \boldsymbol{W}_{1i} & \boldsymbol{W}_{2i} \end{pmatrix} \begin{pmatrix} \zeta_{12}\boldsymbol{\beta}_0 + \gamma_0 & \zeta_{12} \\ \zeta_{22}\boldsymbol{\beta}_0 & \zeta_{22} \end{pmatrix} + \begin{pmatrix} \nu_{1i} & \boldsymbol{\nu}_{2i} \end{pmatrix} \tag{6.1.1}$$

with $(\boldsymbol{W}_{1i}{}^T\ \boldsymbol{W}_{2i}{}^T)^T(\boldsymbol{W}_{1i}\ \boldsymbol{W}_{2i}) = \boldsymbol{M}$ *and*

$$\boldsymbol{\Delta}_n^1 := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} \zeta_{12} & I_{p_2 \times p_2} \\ \zeta_{22} & \mathbf{0}_{p_3 \times p_2} \end{pmatrix}^T \begin{pmatrix} \boldsymbol{Z}_{1i}^T \\ \boldsymbol{Z}_{2i}^T \end{pmatrix} (\omega^{11}\ \boldsymbol{\omega}^{12}) \begin{pmatrix} Y_{1i} - \boldsymbol{Z}_{1i}(\zeta_{12}\boldsymbol{\beta}_0 + \boldsymbol{\gamma}_0) - \boldsymbol{Z}_{2i}\zeta_{22}\boldsymbol{\beta}_0 \\ (\boldsymbol{Y}_{2i} - \boldsymbol{Z}_{1i}\zeta_{12} - \boldsymbol{Z}_{2i}\zeta_{22})^T \end{pmatrix}$$

which converges to normal distribution with mean zero variance $\mathcal{I}_1(\boldsymbol{\theta}_0)$.

Theorem 6.1.2 enables us to construct the local asymptotically maximin test for the hypothesis testing whose the null and alternative are given by $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.

See Taniguchi and Kakizawa (2000, Theorem 3.1.21, p.78) for details. As for estimation, we can see that the asymptotically centering estimator is efficient among the regular estimators in the sense of Taniguchi and Kakizawa (2000, Theorem 3.1.9, p.69).

$$\phi_{n,h} := \begin{cases} 1 & \Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) > c_{n,h}, \\ \gamma_{n,h} & \Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) = c_{n,h}, \\ 0 & \Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) < c_{n,h}, \end{cases}$$

6.2 Likelihood ratio process for random effect ANOVA model.

In this Section, we consider the one-way random effect ANOVA model; for $i = 1, \cdots, a$ and $j = 1, \cdots, n$,

$$Y_{ij} = \mu + \alpha_i + e_{ij}, \tag{6.2.1}$$

where $(\alpha_{i'}, e_{ij})^T$ follows the i.i.d. normal distribution with mean zero and variance $\begin{pmatrix} \sigma_\alpha^2 & 0 \\ 0 & \sigma_e^2 \end{pmatrix}$ for any $i, i' = 1, \cdots, a$ and $j = 1, \cdots, n$. Equivalently, we can rewrite the model in matrix form;

$$\boldsymbol{Y} = \mu \mathbf{1}_{an} + (I_a \otimes \mathbf{1}_n)\boldsymbol{\alpha} + \boldsymbol{e},$$

where $\boldsymbol{Y} := (Y_{11}, \cdots, Y_{1n}, Y_{21}, \cdots, Y_{a1}, \cdots, Y_{an})^T$,

$\boldsymbol{e} := (e_{11}, \cdots, e_{1n}, e_{21}, \cdots\cdots, e_{a1}, \cdots, e_{an})^T$, and $\boldsymbol{\alpha} := (\alpha_1, \cdots, \alpha_a)^T$. By noting that $\boldsymbol{Y} \sim N(\mu \mathbf{1}_{an}, I_a \otimes (\sigma_\alpha^2 J_n + \sigma_e^2 I_n))$ (see Searle et al. (1992, p.79)), the loglikelihood function is given by

$$\begin{aligned}\log \mathrm{dN}^{(n)}_{(\sigma_\alpha^2, \sigma_e^2, \mu)} :=& -\frac{an}{2}\log 2\pi - \frac{a(n-1)}{2}\log \sigma_e^2 - \frac{a}{2}\log(\sigma_e^2 + n\sigma_\alpha^2) \\ &- \frac{1}{2\sigma_e^2}\sum_{i=1}^{a}\sum_{j=1}^{n}(Y_{ij}-\mu)^2 + \frac{n^2\sigma_\alpha^2}{2\sigma_e^2(\sigma_e^2+n\sigma_\alpha^2)}\sum_{i=1}^{a}(\bar{Y}_{i\cdot}-\mu)^2,\end{aligned}$$

where $\bar{Y}_{i\cdot} := \sum_{j=1}^n Y_{ij}/n$. Hereafter, we denote that $\boldsymbol{\theta} := (\theta_1, \theta_2, \theta_3) := (\sigma_\alpha^2, \sigma_e^2, \mu)$.

6.2.1 When the variance of random effect belongs to the boundary of parameter space

In this subsection, we are interested in the following hypothesis;

$$\mathrm{H}_2^{(n)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0 := \begin{pmatrix} 0 \\ \theta_2 \\ \theta_3 \end{pmatrix}, \quad \mathrm{K}_2^{(n)} : \boldsymbol{\theta} = \boldsymbol{\theta}_n := \begin{pmatrix} \frac{h_1}{n^{k_1}} \\ \theta_2 + \frac{h_2}{n^{k_2}} \\ \theta_3 + \frac{h_3}{n^{k_3}} \end{pmatrix},$$

where $\theta_2 > 0$, $h_1 > 0$, and $h_2 > -n^{k_2}\theta_2$.

Theorem 6.2.1. *(i) Under the null hypothesis $\mathrm{H}_2^{(n)}$, the loglikelihood ratio has the following asymptotic expansion; for $h \in \mathbb{R}$ and sufficiently large n such that $h_2 > -n^{k_2}\theta_2$, it holds that*

$$\Lambda_2(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) := \log \frac{\mathrm{dN}^{(n)}_{\boldsymbol{\theta}_n}}{\mathrm{dN}^{(n)}_{\boldsymbol{\theta}_0}}$$
$$= \begin{cases} \begin{pmatrix} \frac{h_3\sqrt{\theta_2}}{\theta_2+h_1} & \frac{h_1}{2(\theta_2+h_1)} \end{pmatrix} \begin{pmatrix} g_1(\boldsymbol{T}_n) \\ g_2(\boldsymbol{T}_n) \end{pmatrix} - \frac{a}{2}\log(1+\frac{h_1}{\theta_2}) - \frac{ah_3^2}{2(\theta_2+h_1)} + o_P(1) & k_2 \geq 1,\ k_3 = \frac{1}{2},\ k_1 = 1, \\ \frac{h_3}{\sqrt{\theta_2}} g_1(\boldsymbol{T}_n) - \frac{ah_3^2}{2\theta_2} + o_P(1) & k_2 \geq 1,\ k_3 = \frac{1}{2},\ k_1 > 1, \\ \frac{h_1}{2(\theta_2+h_1)} g_2(\boldsymbol{T}_n) - \frac{a}{2}\log(1+\frac{h_1}{\theta_2}) + o_P(1) & k_2 \geq 1,\ k_3 > \frac{1}{2},\ k_1 = 1, \\ o_P(1) & k_2 \geq 1,\ k_3 > \frac{1}{2},\ k_1 > 1, \end{cases}$$

where

$$\boldsymbol{T}_n := \left(\frac{\sqrt{n}(\bar{Y}_{1\cdot}-\theta_3)}{\sqrt{\theta_2}}, \ldots, \frac{\sqrt{n}(\bar{Y}_{a\cdot}-\theta_3)}{\sqrt{\theta_2}}\right)^T,$$

$$\begin{pmatrix} g_1(x_1,\ldots,x_a) & g_2(x_1,\ldots,x_a) \end{pmatrix}^T := \begin{pmatrix} \sum_{i=1}^a x_i & \sum_{i=1}^a x_i^2, \end{pmatrix}^T$$

and

$$\boldsymbol{T}_n \Rightarrow N(\mathbf{0}, I_{a\times a}) \quad \text{as } n \to \infty \text{ under } \mathrm{H}_2^{(n)}.$$

(ii) *Under the null hypothesis* $\mathrm{H}_0^{(n)}$, *the fisher information of the model* (6.2.1) *is given by*

$$\mathcal{I}_2(\boldsymbol{\theta}_0) := \begin{pmatrix} \frac{1}{2(\theta_1+\theta_2)^2} & \frac{1}{2(\theta_1+\theta_2)^2} & 0 \\ \frac{1}{2(\theta_1+\theta_2)^2} & \frac{1}{2(\theta_1+\theta_2)^2} & 0 \\ 0 & 0 & \frac{1}{\theta_1+\theta_2} \end{pmatrix}.$$

Remark 6.2.1. For $\{k_2; 0 < k_2 < 1\}$ or $\{(k_1, k_2, k_3); k_2 \geq 1, k_3 < \frac{1}{2}, k_1 > k_3 + \frac{1}{2}\}$, $\Lambda_2(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n)$ shows indeterminate form asymptotically. For $\{(k_1, k_2, k_3); k_2 \geq 1, k_3 < \frac{1}{2}, k_1 \leq k_3 + \frac{1}{2}\}$ or $\{(k_1, k_2, k_3); k_2 \geq 1, k_3 \geq \frac{1}{2}, k_1 < 1\}$, $\Lambda_2(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n)$ tends to $-\infty$ as $n \longrightarrow \infty$.

Remark 6.2.2. Clearly, $g_1(\boldsymbol{T}_n)$ and $g_2(\boldsymbol{T}_n)$ converge in distribution to $N(0, a)$ and χ_a^2 as $n \longrightarrow \infty$, respectively.

Theorem 6.2.1 reveals that the asymptotic behavior of the likelihood ratio process for the random ANOVA model is atypical. The contiguity orders of this model are also unusual since these are different from the usual order $n^{1/2}$. The Fisher information matrix is singular. Obviously, we cannot apply the optimal theory based on LAN such as Le Cam's third lemma. In order to discuss optimality of the testing problem, we restrict ourselves to the following hypothesis;

$$\mathrm{H}_3^{(n)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0 := \begin{pmatrix} 0 \\ \theta_2 \\ \theta_3 \end{pmatrix}, \quad \mathrm{K}_3^{(n)} : \boldsymbol{\theta} = \boldsymbol{\theta}_n := \begin{pmatrix} \frac{h_1}{n^{k_1}} \\ \theta_2 \\ \theta_3 \end{pmatrix},$$

where $\theta_2 > 0$ and $h_1 > 0$. Thereupon, we can prove the next theorem.

Theorem 6.2.2. *Under the null hypothesis* $\mathrm{H}_3^{(n)}$, *it holds that*

$$\Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) := \log \frac{d\mathrm{N}_{\boldsymbol{\theta}_n}^{(n)}}{d\mathrm{N}_{\boldsymbol{\theta}_0}^{(n)}} = \begin{cases} -\frac{a}{2}\log(1 + \frac{h_1}{\theta_2}) + \frac{h_1}{2(\theta_2+h_1)} g_2(\boldsymbol{T}_n) + o_p(1) & k_1 = 1, \\ o_p(1) & k_1 > 1. \end{cases}$$

Remark 6.2.3. For $\{k_1; k_1 = 1\}$, $\Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n)$ tends to $-\infty$ as $n \to \infty$.

Theorem 6.2.2 shows that the random effect ANOVA model does also not have LAN property for the hypothesis $\mathrm{H}_3^{(n)}$. On the other hand, in this case, we can derive the asymptotic distribution of the log-likelihood ratio under the alternative $\mathrm{K}_3^{(n)}$ from the direct calculation.

Theorem 6.2.3. *Under the null hypothesis* $\mathrm{K}_3^{(n)}$*, it holds that*

$$\Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) = \begin{cases} -\frac{a}{2}\log(1+\frac{h_1}{\theta_2}) + \frac{h_1}{2\theta_2} g_2(\boldsymbol{T}_n') + o_p(1) & k_1 = 1, \\ o_p(1) & k_1 > 1, \end{cases}$$

as $n \longrightarrow \infty$*, where*

$$\boldsymbol{T}_n' := \left(\frac{\sqrt{n}(\bar{Y}_{1\cdot} - \theta_3)}{\sqrt{n\theta_1^{(n)} + \theta_2}}, \dots, \frac{\sqrt{n}(\bar{Y}_{a\cdot} - \theta_3)}{\sqrt{n\theta_1^{(n)} + \theta_2}} \right)^T$$

Remark 6.2.4. *For* $\{k_1; k_1 < 1\}$*,* $\Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n)$ *shows indeterminate form asymptotically.*

Next, we shall show the test derived from the log-likelihood is asymptotically most powerful (AMP). The definition of AMP test at asymptotic level α is given as follows ((Lehmann and Romano, 2005, Definition 13.3.1, p.541));

Definition 6.2.1. For the simple hypothesis $\theta = \theta_0$ against $\theta = \theta_n$, a sequence of test $\{\phi_{n,h}\}$ is asymptotically most powerful at asymptotic level α if $\limsup_n \mathrm{E}_{\theta_n}(\phi_{n,h}) \le \alpha$ and, for any test $\{\psi_{n,h}\}$ such that $\limsup_n \mathrm{E}_{\theta_n}(\psi_{n,h}) \le \alpha$,

$$\limsup_n \left(\mathrm{E}_{\theta_n}(\phi_{n,h}) - \mathrm{E}_{\theta_n}(\psi_{n,h}) \right) \ge 0.$$

We define the test function based on the log-likelihood, for α_n such that $\alpha_n \longrightarrow \alpha$ as $n \longrightarrow \infty$,

$$\phi_{n,h} := \begin{cases} 1 & \Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) > c_{n,h}, \\ \gamma_{n,h} & \Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) = c_{n,h}, \\ 0 & \Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) < c_{n,h}, \end{cases} \tag{6.2.2}$$

where the critical value $c_{n,h}$ and the constant $\gamma_{n,h}$ are determined by $\mathrm{E}_{\theta_0}(\phi_{n,h}) = \alpha_n$.

Then, the next theorem shows the asymptotic power of the proposed test (6.2.2) and our test is AMP.

Theorem 6.2.4. *For* $k_1 = 1$*, the following statements hold true;* (i) *It holds that*

$$c_{n,h} \to c := -\log\frac{a}{2}\log\left(1 + \frac{h_1}{\theta_2}\right) + \frac{h_1}{2(\theta_2 + h_1)}\chi_a^2[1-\alpha] \quad \textit{as } n \to \infty,$$

where $\chi_a^2[1-\alpha]$ *denotes the* $1-\alpha$ *quantile of chi-square with a degree of freedom.*

(ii) *The asymptotic power of the test is given by*

$$\lim_{n\to\infty} \mathrm{P}(\Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) \ge c_{n,h} | \mathrm{K}_3^{(n)}) = \mathrm{P}(\chi_a^2 \ge \frac{\theta_2}{(\theta_2 + h_1)}\chi_a^2[1-\alpha]).$$

(iii) *The test* $\{\phi_{n,h}\}$*, defined by* (6.2.2)*, is asymptotically most powerful at asymptotic level* α*.*

Theorem 6.2.4 shows our test is optimal in the sense of AMP. For a small perturbation h_1, the power is almost equal to significance level α. The optimal tests endowed with uniformity with respect to h_1 such as local asymptotically uniformly most powerful (LAUMP) and asymptotically uniformly most powerful (AUMP) are beyond the scope of the dissertation.

6.2.2 When the variance of random effect belongs to the interior of parameter space

In this subsection, we shall consider the case $\sigma_\alpha^2 > 0$. The null and alternative hypotheses are defined as

$$\mathrm{H}_4^{(n)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0 := \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}, \quad \mathrm{K}_4^{(n)} : \boldsymbol{\theta} = \boldsymbol{\theta}_n := \begin{pmatrix} \theta_1 + \frac{h_1}{n^{k_1}} \\ \theta_2 + \frac{h_2}{n^{k_2}} \\ \theta_3 + \frac{h_3}{n^{k_3}} \end{pmatrix},$$

where $\theta_1 > 0$, $\theta_2 > 0$, $h_1 > 0$, $h_1 > -n^{k_1}\theta_1$, and $h_2 > -n^{k_2}\theta_2$.

Theorem 6.2.5. *Under the null hypothesis $\mathrm{H}_4^{(n)}$, it holds that, for all $k_1 \geq 1$, $k_2 > 0$, and $k_3 > 0$,*

$$\Lambda_4(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) := \log \frac{\mathrm{dN}_{\boldsymbol{\theta}_n}^{(n)}}{\mathrm{dN}_{\boldsymbol{\theta}_0}^{(n)}} \longrightarrow 0 \quad \textit{in probability as } n \longrightarrow \infty.$$

Remark 6.2.5. For $\{k_1; k_1 < 1\}$, $\Lambda_4(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n)$ tends to indeterminate form as $n \longrightarrow \infty$.

We expect $\Lambda_4(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n)$ has LAN property since the parameters belongs to interior of the parameter spaces. However, Theorem 6.2.5 shows that the likelihood ratio converges to the degenerate distribution. This is also an unusual result.

Furthermore, we consider the case that $\sigma_\alpha^2 > 0$ is only perpetuated.

$$\mathrm{H}_5^{(n)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0 := \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}, \quad \mathrm{K}_5^{(n)} : \boldsymbol{\theta} = \boldsymbol{\theta}_n := \begin{pmatrix} \theta_1 + \frac{h_1}{n^{k_1}} \\ \theta_2 \\ \theta_3 \end{pmatrix},$$

where $\theta_1 > 0$, $\theta_2 > 0$, $h_1 > 0$, and $h_1 > -n^{k_1}\theta_1$. Then, we have the same result as Theorem 6.2.5.

Theorem 6.2.6. *Under the null hypothesis* $\mathrm{H}_5^{(n)}$*, it holds that, for all* $k_1 > 0$*,*

$$\Lambda_5(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) := \log \frac{\mathrm{dN}^{(n)}_{\boldsymbol{\theta}_n}}{\mathrm{dN}^{(n)}_{\boldsymbol{\theta}_0}} \to 0 \quad \textit{in probability as } n \to \infty.$$

Again, this theorem shows the erratic behavior of log likelihood ratio process.

Chapter 7

Proof chapter

In this Chapter, we give the proofs of the propositions and the theorems stated in Chapters 2, 3, 4, 5, and 6.

7.1 Proofs in Chapter 2

Proof of Theorem 2.1.1.
It is easy to see that$1/\sqrt{N}(D_k - \tilde{D}_k) = o_p(1)$, where $\tilde{D}_k = 2\sum_{t=2}^{n_k} X_{k(i-1)+1} - 2\sum_{t=2}^{n_k} X_{k(i-1)+1}X_{k(i-2)+1}$, hence it suffices to show that

$$\frac{1}{\sqrt{N}}\begin{pmatrix}\tilde{D}_1 - \mathrm{E}\tilde{D}_1\\ \tilde{D}_2 - \mathrm{E}\tilde{D}_2\\ \vdots \\ \tilde{D}_m - \mathrm{E}\tilde{D}_m\end{pmatrix} \Rightarrow N(0, V).$$

First we show the asymptotic variance of $1/\sqrt{N}(\tilde{D}_1 - \mathrm{E}\tilde{D}_1, \tilde{D}_2 - \mathrm{E}\tilde{D}_2, \ldots, \tilde{D}_m - \mathrm{E}\tilde{D}_m)'$ is given by V. We can see that

$$\begin{aligned}\frac{1}{N}\mathrm{Cov}(\tilde{D}_k, \tilde{D}_s) =& \frac{4}{N}\sum_{i=2}^{n_k}\sum_{j=2}^{n_s}[K_X(k, s(j-1) - k(i-2), s(j-2) - k(i-2))\\ &+ \gamma_X(k(i-1) - s(j-1))\gamma_X(k(i-2) - s(j-2))\\ &+ \gamma_X(k(i-2) - s(j-1))\gamma_X(k(i-1) - s(j-2))]\\ &+ \frac{4}{8\pi N}[\sin^{-1}\rho_z\{k(n_{k-1}) - s(n_{s-1})\} + \sin^{-1}\rho_z\{0\}\\ &- \sin^{-1}\rho_z\{s(n_{s-1})\} - \sin^{-1}\rho_z\{k(n_{k-1})\}],\end{aligned}$$

which tends to $v_{k,s}$ as $N \to \infty$.

Next, we show that the asymptotic distribution of $(1/\sqrt{N})\tilde{D}_k$. We define the random vector

$$\mathbf{Y_i} := \begin{pmatrix} X_{(i-1)+1}(1 - X_{(i-2)+1})\mathbb{I}_{\{(i-1)+1\leq N\}} \\ X_{2(i-1)+1}(1 - X_{2(i-2)+1})\mathbb{I}_{\{2(i-1)+1\leq N\}} \\ \vdots \\ X_{m(i-1)+1}(1 - X_{m(i-2)+1})\mathbb{I}_{\{m(i-1)+1\leq N\}} \end{pmatrix}.$$

From Cramer-Wold device, it is sufficient to show that, for arbitrary $\boldsymbol{\lambda} \in \mathbb{R}^m$, $(1/\sqrt{N})\boldsymbol{\lambda}'(\mathbf{Y_i} - \mathrm{E}\mathbf{Y_i}) \Rightarrow N(0, \boldsymbol{\lambda}'V\boldsymbol{\lambda})$. From the central limit theorem (see Brockwell and Davis (1991, p.213)) for m-dependent sequences and the fact that $(2/\sqrt{N})\sum_{i=2}^{N} \mathbf{Y'}_i = (\tilde{D}_1, \cdots, \tilde{D}_m)$, we get the desired result.

Next, we calculate $\mathrm{E}\{X_{t+2}X_{t+1}X_tX_{t-1}\}$. Let $H_a(x)$ be the function (see Baum (1957)) given by

$$\begin{aligned} H_a(x) &:= \begin{cases} -a & x < -a, \\ x & |x| \leq a, \\ a & a < x \end{cases} \\ &= \frac{2}{\pi}\int_0^\infty \frac{\sin au \sin xu}{u^2} du. \end{aligned} \tag{7.1.1}$$

The function H_a has the following relation:

$$\lim_{a\to 0} \frac{H_a(x)}{a} = \begin{cases} -1 & x < 0, \\ 0 & x = 0, \\ 1 & 0 < x. \end{cases} \tag{7.1.2}$$

Let h_m be the function defined by

$$h_m(x) = \frac{2}{\pi}\int_0^\infty u^{2m-1} \exp\frac{-u^2}{2} \sin(xu) du. \tag{7.1.3}$$

It holds that

$$\lim_{x\to 0} \frac{h_m(x)}{x} = \sqrt{\frac{2}{\pi}}\frac{(2m)!}{2^m m!}, \ for\ m = 0, 1, 2, \cdots. \tag{7.1.4}$$

Lemma 7.1.1. *Let $\{Z_t\}$ be a Gaussian stationary process with mean zero, and spectral density function f. Let $H_a(x)$ be the function defined by (7.1.1). Then the following statements hold:*

(i) $\mathrm{E}\{\mathrm{H}_{\mathrm{a}_1}(\mathrm{Z}_1)\mathrm{H}_{\mathrm{a}_2}(\mathrm{Z}_2)\mathrm{H}_{\mathrm{a}_3}(\mathrm{Z}_3)\mathrm{H}_{\mathrm{a}_4}(\mathrm{Z}_4)\}$

$$= \sum_{m_1,m_2,m_3=0}^{\infty} h_{m_1}(a_1)h_{m_1+m_2}(a_2)h_{m_2+m_3}(a_3)h_{m_3}(a_4)\frac{\rho_Z(1)^{2(m_1+m_2+m_3+1)}}{(2m_1+1)!(2m_2)!(2m_3+1)!},$$

(ii) $\mathrm{E}\{\mathrm{H}_{a_1}(\mathrm{Z}_1)\mathrm{H}_{a_2}(\mathrm{Z}_2)\mathrm{H}_{a_3}(\mathrm{Z}_3)\} = 0$,

(iii) $\mathrm{E}\{H_{a_1}(Z_1)H_{a_2}(Z_2)\} = \sum_{m=0}^{\infty} \frac{\rho_Z(1)^{2m+1}}{(2m+1)!} h_m(a_1)h_m(a_2)$,

(iv) $\mathrm{E}\{H_{a_1}(Z_1)\} = 0$.

Proof. (i) We first compute $\mathrm{E}\{\sin Z_1u_1 \sin Z_2u_2 \sin Z_3u_3 \sin Z_4u_4\}$.
Let Σ_4 be a variance-covariance matrix of (Z_1, Z_2, Z_3, Z_4), given by

$$\Sigma_4 = \begin{bmatrix} 1 & \rho_Z(1) & 0 & 0 \\ \rho_Z(1) & 1 & \rho_Z(1) & 0 \\ 0 & \rho_Z(1) & 1 & \rho_Z(1) \\ 0 & 0 & \rho_Z(1) & 1 \end{bmatrix}.$$

Recalling the characteristic function of multivariate normal distribution, we obtain

$$\begin{aligned}
&\mathrm{E}\{\sin(u_1Z_1)\sin(u_2Z_2)\sin(u_3Z_3)\sin(u_4Z_4)\}\\
=&(2i)^{-4}\mathrm{E}\{(e^{iu_1Z_1} - e^{-iu_1Z_1})(e^{iu_2Z_2} - e^{-iu_2Z_2})(e^{iu_3Z_3} - e^{-iu_3Z_3})(e^{iu_4Z_4} - e^{-iu_4Z_4})\}\\
=&2^{-4}\sum_{\epsilon_1,\epsilon_2,\epsilon_3,\epsilon_4=\pm1}(\epsilon_1\epsilon_2\epsilon_3\epsilon_4)\mathrm{E}\{\exp\{i(\epsilon_1u_1Z_1 + \epsilon_2u_2Z_2 + \epsilon_3u_3Z_3 + \epsilon_4u_4Z_4)\}\\
=&2^{-4}\sum_{\epsilon_1,\epsilon_2,\epsilon_3,\epsilon_4=\pm1}(\epsilon_1\epsilon_2\epsilon_3\epsilon_4)\exp\{-\frac{1}{2}(\epsilon_1u_1,\epsilon_2u_2,\epsilon_3u_3,\epsilon_4u_4)\Sigma_4(\epsilon_1u_1,\epsilon_2u_2,\epsilon_3u_3,\epsilon_4u_4)\}\\
=&2^{-4}\exp\{-\frac{1}{2}(u_1^2+u_2^2+u_3^2+u_4^2)\}\sum_{\epsilon_1,\epsilon_2,\epsilon_3,\epsilon_4=\pm1}(\epsilon_1\epsilon_2\epsilon_3\epsilon_4)\\
&\times\exp\{-\epsilon_1\epsilon_2u_1u_2\rho_Z(1) - \epsilon_2\epsilon_3u_2u_3\rho_Z(1) - \epsilon_3\epsilon_4u_3u_4\rho_Z(1)\}\\
=&2^{-4}\exp\{-\frac{1}{2}(u_1^2+u_2^2+u_3^2+u_4^2)\}2^3\{\sinh(u_1u_2\rho_Z(1) + u_2u_3\rho_Z(1))\\
&+\sinh(u_1u_2\rho_Z(1) - u_2u_3\rho_Z(1))\}\\
&\times\sinh(u_3u_4\rho_Z(1))\\
=&2^{-1}\exp\{-\frac{1}{2}(u_1^2+u_2^2+u_3^2+u_4^2)\}2\sinh u_1u_2\rho_Z(1)\cosh u_2u_3\rho_Z(1)\sinh u_3u_4\rho_Z(1)\\
=&\exp\{-\frac{1}{2}(u_1^2+u_2^2+u_3^2+u_4^2)\}\sum_{m_1,m_2,m_3=0}^{\infty}u_1u_2u_3u_4\rho_Z(1)^{2(m_1+m_2+m_3+1)}\\
&\times\frac{u_1^{2m_1}u_2^{2(m_1+m_2)}u_3^{2(m_2+m_3)}u_4^{2m_3}}{(2m_1+1)!(2m_2)!(2m_3+1)!}.
\end{aligned}$$

By (7.1.1) and (7.1.3), Fubini's theorem, the above calculation and term by term integration, we have

$\mathrm{E}\{H_{a_1}(Z_1)H_{a_2}(Z_2)H_{a_3}(Z_3)H_{a_4}(Z_4)\}$

$$
\begin{aligned}
&= \left(\frac{2}{\pi}\right)^4 \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \frac{\sin a_1u_1 \sin a_2u_2 \sin a_3u_3 \sin a_4u_4}{u_1^2u_2^2u_3^2u_4^2} \\
&\quad \times \mathrm{E}\left\{\sin Z_1u_1 \sin Z_2u_2 \sin Z_3u_3 \sin Z_4u_4\right\} du_1du_2du_3du_4 \\
&= \left(\frac{2}{\pi}\right)^4 \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \sum_{m_1,m_2,m_3=0}^{\infty} \frac{\sin a_1u_1 \sin a_2u_2 \sin a_3u_3 \sin a_4u_4}{u_1u_2u_3u_4} \\
&\quad \times \exp\{-\frac{1}{2}(u_1^2+u_2^2+u_3^2+u_4^2)\}\rho_Z(1)^{2(m_1+m_2+m_3+1)} \\
&\quad \times \frac{u_1^{2m_1}u_2^{2(m_1+m_2)}u_3^{2(m_2+m_3)}u_4^{2m_3}}{(2m_1+1)!(2m_2)!(2m_3+1)!} du_1du_2du_3du_4 \\
&= \sum_{m_1,m_2,m_3=0}^{\infty} h_{m_1}(a_1)h_{m_1+m_2}(a_2)h_{m_2+m_3}(a_3)h_{m_3}(a_4)\frac{\rho_Z(1)^{2(m_1+m_2+m_3+1)}}{(2m_1+1)!(2m_2)!(2m_3+1)!}.
\end{aligned}
$$

(ii) It suffices to show $\mathrm{E}\{\sin(u_1Z_1)\sin(u_2Z_2)\sin(u_3Z_3)\} = 0$. Let Σ_3 be the variance-covariance matrix of (Z_1, Z_2, Z_3), given by

$$
\Sigma_3 = \begin{bmatrix} 1 & \rho_Z(1) & 0 \\ \rho_Z(1) & 1 & \rho_Z(1) \\ 0 & \rho_Z(1) & 1 \end{bmatrix}.
$$

We have

$$
\begin{aligned}
&\mathrm{E}\{\sin(u_1Z_1)\sin(u_2Z_2)\sin(u_3Z_3)\} \\
&= (2i)^{-3}\mathrm{E}\{(e^{iu_1Z_1}-e^{-iu_1Z_1})(e^{iu_2Z_2}-e^{-iu_2Z_2})(e^{iu_3Z_3}-e^{-iu_3Z_3})\} \\
&= (2i)^{-3}\sum_{\epsilon_1=\pm1}\sum_{\epsilon_2=\pm1}\sum_{\epsilon_3=\pm1}(\epsilon_1\epsilon_2\epsilon_3)\mathrm{E}\{\exp\{i(\epsilon_1u_1Z_1+\epsilon_2u_2Z_2+\epsilon_3u_3Z_3\} \\
&= (2i)^{-3}\sum_{\epsilon_1=\pm1}\sum_{\epsilon_2=\pm1}\sum_{\epsilon_3=\pm1}(\epsilon_1\epsilon_2\epsilon_3)\exp\{-\frac{1}{2}(\epsilon_1u_1,\epsilon_2u_2,\epsilon_3u_3)\Sigma_3(\epsilon_1u_1,\epsilon_2u_2,\epsilon_3u_3)'\} \\
&= (2i)^{-3}\exp\{-\frac{1}{2}(u_1^2+u_2^2+u_3^2)\} \\
&\quad \times \sum_{\epsilon_1=\pm1}\sum_{\epsilon_2=\pm1}\sum_{\epsilon_3=\pm1}(\epsilon_1\epsilon_2\epsilon_3)\exp\{-\epsilon_1\epsilon_2u_1u_2\rho_Z(1)-\epsilon_2\epsilon_3u_2u_3\rho_Z(1)\} \\
&= 0.
\end{aligned}
$$

(iii) First, we observe

$$
\begin{aligned}
&\mathrm{E}\{sin(u_1Z_1)sin(u_2Z_2)\} \\
&= (2i)^{-2}\mathrm{E}\{(e^{iu_1Z_1}-e^{-iu_1Z_1})(e^{iu_2Z_2}-e^{-iu_2Z_2})\} \\
&= (2i)^{-2}\sum_{\epsilon_1=\pm1}\sum_{\epsilon_2=\pm1}(\epsilon_1\epsilon_2)\mathrm{E}\{\exp\{i(\epsilon_1u_1Z_1+\epsilon_2u_2Z_2)\}
\end{aligned}
$$

$$
\begin{aligned}
&= (2i)^{-2} \sum_{\epsilon_1=\pm1} \sum_{\epsilon_2=\pm1} (\epsilon_1\epsilon_2) \exp\{-\frac{1}{2}(\epsilon_1 u_1, \epsilon_2 u_2,)\Sigma_2(\epsilon_1 u_1, \epsilon_2 u_2,)'\} \\
&= (2i)^{-2} \exp\{-\frac{1}{2}(u_1^2 + u_2^2)\} \sum_{\epsilon_1=\pm1} \sum_{\epsilon_2=\pm1} (\epsilon_1\epsilon_2) \exp\{-\epsilon_1\epsilon_2 u_1 u_2 \rho_1\} \\
&= \exp\{-\frac{1}{2}(u_1^2 + u_2^2)\} \sinh u_1 u_2 \rho_Z(1) \\
&= \exp\{-\frac{1}{2}(u_1^2 + u_2^2)\} \sum_{m=0}^{\infty} \frac{(u_1u_2)^{2m}}{(2m+1)!} u_1 u_2 \rho_Z(1)^{2m+1}.
\end{aligned}
$$

By Fubini's theorem, it follows that

$$
\begin{aligned}
&\mathrm{E}\{H_{a_1}(Z_1)H_{a_2}(Z_2)\} \\
&= \left(\frac{2}{\pi}\right)^2 \mathrm{E}\left\{\int_0^\infty \int_0^\infty \frac{\sin a_1u_1 \sin a_2u_2}{u_1^2u_2^2} \sin Z_1u_1 \sin Z_2u_2 du_1du_2\right\} \\
&= \left(\frac{2}{\pi}\right)^2 \int_0^\infty \int_0^\infty \frac{\sin a_1u_1 \sin a_2u_2}{u_1^2u_2^2} \mathrm{E}\left\{\sin Z_1u_1 \sin Z_2u_2\right\} du_1du_2 \\
&= \left(\frac{2}{\pi}\right)^2 \int_0^\infty \int_0^\infty \frac{\sin a_1u_1 \sin a_2u_2}{u_1^2u_2^2} \exp\{-\frac{1}{2}(u_1^2 + u_2^2)\} \\
&\quad \times \sum_{m=0}^{\infty} \frac{(u_1u_2)^{2m}}{(2m+1)!} u_1u_2\rho_Z(1)^{2m+1} du_1du_2 \\
&= \sum_{m=0}^{\infty} \left(\frac{2}{\pi}\right)^2 \int_0^\infty \int_0^\infty \frac{\sin a_1u_1 \sin a_2u_2}{u_1u_2} \\
&\quad \times \exp\{-\frac{1}{2}(u_1^2 + u_2^2)\} \frac{(u_1u_2)^{2m}}{(2m+1)!} \rho_Z(1)^{2m+1} du_1du_2 \\
&= \sum_{m=0}^{\infty} \frac{\rho_Z(1)^{2m+1}}{(2m+1)!} h_m(a_1)h_m(a_2),
\end{aligned}
$$

which completes the proof of the assertion.

(iv) It suffices to show $\mathrm{E}\{sin(u_1Z_1)\} = 0$.

We obtain

$$
\begin{aligned}
&\mathrm{E}\{sin(u_1Z_1)\} \\
&= (2i)^{-1}\mathrm{E}\{(e^{iu_1Z_1} - e^{-iu_1Z_1})\} \\
&= (2i)^{-1} \sum_{\epsilon_1=\pm1} (\epsilon_1)\mathrm{E}\{\exp\{i(\epsilon_1u_1Z_1)\}
\end{aligned}
$$

$$
\begin{aligned}
&= (2i)^{-1} \sum_{\epsilon_1=\pm 1} (\epsilon_1) \exp\{-\frac{1}{2}(\epsilon_1 u_1)(\epsilon_1 u_1)'\} \\
&= (2i)^{-1} \exp\{-\frac{1}{2}u_1^2\} \sum_{\epsilon_1=\pm 1} \epsilon_1 \\
&= 0,
\end{aligned}
$$

which concludes the proof. □

Proof of Proposition 2.1.1.
From the definition of (7.1.1), then it follows that

$$
\left| \frac{\{H_{a_1}(Z_1)+a_1\}\{H_{a_2}(Z_2)+a_2\}\{H_{a_3}(Z_3)+a_3\}\{H_{a_4}(Z_4)+a_4\}}{a_1a_2a_3a_4} \right| \leq 2^4.
$$

Then, use of the dominated convergence theorem and (7.1.2) leads to

$$
\begin{aligned}
&\mathrm{E}\{X_{t+2}X_{t+1}X_tX_{t-1}\} \\
&= \frac{1}{16} \lim_{a_1,a_2,a_3,a_4 \to 0} \mathrm{E}\frac{\{H_{a_1}(Z_1)+a_1\}\{H_{a_2}(Z_2)+a_2\}\{H_{a_3}(Z_3)+a_3\}\{H_{a_4}(Z_4)+a_4\}}{a_1a_2a_3a_4}.
\end{aligned}
$$

Then from Lemma 7.1.1 and (7.1.4), we get the result. □

Proof of Proposition 2.1.2
From the definition of $I(\rho_Z(1))$ we have

$$
I(\rho_Z(1)) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\frac{2\cos\lambda}{1+2\rho_Z(1)\cos\lambda}\right)^2 d\lambda.
$$

First, for $\rho_Z(1) = 1$, $I(\rho_Z(1)) = 1$, by a straightforward calculation.
Second, for $|\rho_Z(1)| < \frac{1}{2}$, the residue theorem yields the assertion.
Third, for the case when $\rho_Z(1) = \frac{1}{2}$, we observe that

$$
\begin{aligned}
&\int_{-\pi}^{\pi}\left(\frac{\cos\lambda}{1+\cos\lambda}\right)^2 dx \\
=&2\int_{0}^{\pi}\left(\frac{\cos\lambda}{1+\cos\lambda}\right)^2 dx \\
=&2\int_{-1}^{1}\left(\frac{t}{1+t}\right)^2 \frac{1}{\sqrt{1-t^2}}dt \\
=&2\int_{0}^{2}\left(\frac{s-1}{s}\right)^2 \frac{1}{\sqrt{s(2-s)}}ds
\end{aligned}
$$

$$
\begin{aligned}
&=2\lim_{\epsilon\to 0}\int_{\epsilon}^{1}\left(\frac{s-1}{s}\right)^2\frac{1}{\sqrt{s(2-s)}}ds+2\lim_{\epsilon\to 0}\int_{1}^{2-\epsilon}\left(\frac{s-1}{s}\right)^2\frac{1}{\sqrt{s(2-s)}}ds\\
&\geq \lim_{\epsilon\to 0}\int_{\epsilon}^{1}\left(\frac{(s-1)^2}{s}\right)ds+2\lim_{\epsilon\to 0}\int_{1}^{2-\epsilon}\left(\frac{s-1}{s}\right)^2\frac{1}{\sqrt{s(2-s)}}ds\\
&\geq 2\lim_{\epsilon\to 0}\left[\log s-2s+\frac{1}{2}s^2\right]_{\epsilon}^{1}+2\lim_{\epsilon\to 0}\int_{1}^{2-\epsilon}\left(\frac{s-1}{s}\right)^2\frac{1}{\sqrt{s(2-s)}}ds\\
&=2(+\infty)+2\lim_{\epsilon\to 0}\int_{1}^{2-\epsilon}\left(\frac{s-1}{s}\right)^2\frac{1}{\sqrt{s(2-s)}}ds\\
&=+\infty.
\end{aligned}
$$

The remaining case $\rho_Z(1)=-\frac{1}{2}$ follows similarly. □

Proof of Theorem 2.2.1.
We can see that

$$
\begin{aligned}
&DM(f_\theta,\underline{I}_n^s)\\
&=\int_{[-\pi,\,\pi]}\left[\log f_\theta(\lambda)d\lambda+\frac{1}{2\pi n}\int_{[-\pi,\,\pi]}\frac{1}{f_\theta(\lambda)}\{\sum_{t=1}^{n}Z_t^s e^{it\lambda}\}\{\sum_{t=1}^{n}Z_t^s e^{-it\lambda}\}\right]d\lambda\\
&=\int_{[-\pi,\,\pi]}\left\{\log f_\theta(\lambda)+\frac{1}{2\pi n f_\theta(\lambda)}\sum_{t_1,t_2=2}^{n}Z_{t_1}Z_{t_2}e^{-i(t_1-t_2)\lambda}\right\}d\lambda\\
&\quad+\frac{2s}{2\pi n}\int_{[-\pi,\,\pi]}\frac{1}{f_\theta(\lambda)}\sum_{t}^{n}Z_t\cos\{(t-1)\lambda\}d\lambda+\frac{s^2}{2\pi n}\int_{[-\pi,\,\pi]}\frac{1}{f_\theta(\lambda)}d\lambda.
\end{aligned}
$$

For fixed n and sufficiently large s, the minimization of $D(f_\theta,\underline{I}_n^s)$ with respect to θ equal to the minimization of the main order term $s^2/(2\pi n)\int_{[-\pi,\,\pi]}1/f_\theta(\lambda)d\lambda$, so we get the result. □

7.2 Proofs in Chapter 3

The nonparametric spectral estimator $\hat{f}_Z(\lambda)$, defined by (3.1.1), can be written in the form

$$
\hat{f}_Z(\lambda)=\frac{1}{2\pi}\sum_{|\ell|\leq M}w\left(\frac{\ell}{M}\right)\hat{\rho}_Z(\ell)\exp(-i\ell\lambda).
$$

Expanding $\hat{\rho}_Z(\ell)$ around $\rho_Z(\ell)$, we obtain

$$\begin{aligned}\hat{\rho}_Z(\ell) =& \rho_Z(\ell) + \frac{\pi}{2}(\hat{\rho}_X(\ell) - \rho_X(\ell))\sqrt{1-\rho_Z^2(\ell)} \\ & - \frac{\pi^2}{8}(\hat{\rho}_X(\ell) - \rho_X(\ell))^2 \sin(\frac{\pi}{2}\rho_X^*(\ell)),\end{aligned}$$

where $\rho_X(\ell) \gtrless \rho_X^*(\ell) \gtrless \hat{\rho}_X(\ell)$. Write

$$\begin{aligned}\hat{f}_Z(\lambda) =& \frac{1}{2\pi}\sum_{|\ell|\le M} w\left(\frac{\ell}{M}\right)\rho_Z(\ell)\exp\left(-i\ell\lambda\right) \\ & + \frac{1}{4}\sum_{|\ell|\le M} w\left(\frac{\ell}{M}\right)(\hat{\rho}_X(\ell) - \rho_X(\ell))\sqrt{1-\rho_Z^2(\ell)}\exp\left(-i\ell\lambda\right) \\ & - \frac{\pi}{16}\sum_{|\ell|\le M} w\left(\frac{\ell}{M}\right)(\hat{\rho}_X(\ell) - \rho_X(\ell))^2 \sin(\frac{\pi}{2}\rho_X^*(\ell))\exp\left(-i\ell\lambda\right) \\ =: & f_n(\lambda) + \hat{g}(\lambda) + \hat{h}(\lambda).\end{aligned}$$

Lemma 7.2.1. *Under Assumptions 3.1.1 and 3.2.1, then*

$$\begin{aligned}& \lim_{n\to\infty}\frac{n}{M}\mathrm{E}\{\hat{f}_Z(\lambda) - f_Z(\lambda)\}^2 \\ & = \begin{cases} 4\pi^2\int_{-1}^{1}k^2(\theta)\mathrm{d}s f_X^2(\lambda) & (\lambda \neq 0, \pm\pi) \\ 8\pi^2\int_{-1}^{1}k^2(\theta)\mathrm{d}s f_X^2(\lambda) & (\lambda = 0, \pm\pi)\end{cases}.\end{aligned}$$

In order to prove Lemma 7.2.1, we wil use the following lemma is due to Leonov and Shiryaev (1959).

Lemma 7.2.2. *Consider a two-way array of random variables $X_{ij}, j = 1, \cdots, J_i, i = 1, \cdots, l$. Let*

$$Y_i = \prod_{j=i}^{J_i} X_{ij},\ i = 1, \cdots, l.$$

Then

$$\mathrm{Cum}(Y_1, \cdots, Y_l) = \sum_{\nu} \mathrm{Cum}\{X_{ij}; (i,j) \in \nu_1\}\cdots\mathrm{Cum}\{X_{ij}; (i,j)\in\nu_p\},$$

where the summation is taken over all indecomposable partitions $\nu = \nu_1 \cup \cdots \cup \nu_p$.

Proof of Lemma 7.2.1
Note that

$$\frac{n}{M}\mathrm{E}\{\hat{f}_Z(\lambda) - f_Z(\lambda)\}^2$$

$$=\frac{n}{M}(\mathrm{Var}(\hat{g}(\lambda)) + \mathrm{Var}(\hat{h}(\lambda)) + 2\mathrm{Cov}(\hat{g},\hat{h}))$$
$$+ \frac{n}{M}(f_n(\lambda) - f_Z(\lambda) + \mathrm{E}\hat{g}(\lambda) + \mathrm{E}\hat{h}(\lambda))^2.$$

First, we show that

$$\lim_{n\to\infty}\frac{n}{M}\mathrm{Cov}(\hat{g}(\lambda_1),\hat{g}(\lambda_2))$$
$$=\begin{cases}0 & (\lambda_1 \neq \pm\lambda_2)\\ 4\pi^2\int_{-1}^{1}k^2(\theta)\mathrm{d}s f_X^2(\lambda) & (\lambda_1=\lambda_2=\lambda\neq 0,\pm\pi)\\ 8\pi^2\int_{-1}^{1}k^2(\theta)\mathrm{d}s f_X^2(\lambda) & (\lambda_1=\lambda_2=\lambda=0,\pm\pi).\end{cases}$$

It is easy to see

$$\frac{n}{M}\mathrm{Cov}(\hat{g}(\lambda_1),\hat{g}(\lambda_2))$$
$$=\frac{1}{16M}\sum_{|\ell|,|r|\le M} w\left(\frac{\ell}{M}\right)w\left(\frac{r}{M}\right)\sqrt{1-\rho_Z^2(\ell)}\sqrt{1-\rho_Z^2(r)}$$
$$\times\exp\{i(r\lambda_2-\ell\lambda_1)\}n\mathrm{Cov}((\hat{\rho}_X(\ell)-\rho_X(\ell)),(\hat{\rho}_X(r)-\rho_X(r))),$$

and

$$n\mathrm{Cov}\{(\hat{\rho}_X(\ell)-\rho_X(\ell)),(\hat{\rho}_X(r)-\rho_X(r))\}$$
$$=\frac{16}{n}\sum_{i=\ell+1}^{n}\sum_{j=r+1}^{n}[K_X(-\ell,j-i,j-i-r)+\gamma_X(j-i)\gamma_X(j-i+\ell-r)$$
$$+\gamma_X(j-i-r)\gamma_X(j-i+\ell)].$$

Hence, we have

$$\frac{n}{M}\mathrm{Cov}(\hat{g}(\lambda_1),\hat{g}(\lambda_2))$$
$$=\frac{1}{M}\sum_{|\ell|,|r|\le M} w\left(\frac{\ell}{M}\right)w\left(\frac{r}{M}\right)\sqrt{1-\rho_Z^2(l)}\sqrt{1-\rho_Z^2(r)} \tag{7.2.1}$$
$$\times\exp\{i(r\lambda_2-\ell\lambda_1)\}\sum_{u=-\infty}^{\infty}[\{K_X(-\ell,u,u-r)$$
$$+\gamma_X(u)\gamma_X(u+\ell-r)+\gamma_X(u-r)\gamma_X(u+\ell)\}\phi_n(u,\ell,r)],$$

where for $r \le \ell$

$$\phi_n(u,\ell,r) := \begin{cases} 0 & (u \le -n+r) \\ 1-\frac{r-u}{n} & (-n+r \le u \le 0) \\ 1-\frac{r}{n} & (0 \le u \le \ell - r) \\ 1-\frac{\ell+u}{n} & (\ell - r \le u \le n-\ell) \\ 0 & (n-\ell \le u) \end{cases} .$$

(7.2.1) can be written approximately in the form

$$\begin{aligned} &\frac{n}{M}\mathrm{Cov}(\hat{g}(\lambda_1),\hat{g}(\lambda_2)) \\ =&\frac{1}{M}\sum_{|\ell|,|r|\le M} w\left(\frac{\ell}{M}\right) w\left(\frac{r}{M}\right) \exp\{i(r\lambda_2 - \ell\lambda_1)\} \\ &\times \sum_{u=-\infty}^{\infty} [\{K_X(-\ell,u,u-r) + \gamma_X(u)\gamma_X(u+\ell-r)+ \\ &\times \gamma_X(u-r)\gamma_X(u+\ell)\}\phi_n(u,\ell,r)] + o\left(\frac{M}{n}\right). \end{aligned} \tag{7.2.2}$$

The first term of (7.2.2) is of order $O(1/M)$. The second term of (7.2.2) converges to

$$\begin{cases} 0 & (\lambda_1 \ne \lambda_2) \\ 4\pi^2 \int_{-1}^{1} k^2(\theta)\mathrm{d}s f_X^2(\lambda) & (\lambda_1 = \lambda_2 = \lambda) \end{cases} \quad as\ n \to \infty.$$

The third term of (7.2.2) converges to

$$\begin{cases} 0 & (\lambda_1 + \lambda_2 \ne 0 \bmod 2\pi) \\ 4\pi^2 \int_{-1}^{1} k^2(\theta)\mathrm{d}s f_X^2(\lambda) & (\lambda_1 = -\lambda_2 = \lambda) \end{cases} \quad as\ n \to \infty$$

(see Hannan (1970)), therefore we get the result.

Next, we show that

$$\mathrm{Cov}(\hat{h}(\lambda_1),\hat{h}(\lambda_2)) = O\left(\frac{M^2}{n^2}\right).$$

For simplicity, we use the following notation:

$$A_\ell := \hat{\rho}_X(\ell) - \mathrm{E}\hat{\rho}_X(\ell), C_\ell := \mathrm{E}\hat{\rho}_X(\ell) - \rho_X(\ell).$$

Since there exists constant c such that $w(x) \le c$ uniformly in $x \in \mathbb{R}$, we have

$$
\begin{aligned}
&\mathrm{Cov}(\hat{h}(\lambda_1), \hat{h}(\lambda_2)) \\
\le& \frac{c\pi^2}{16^2} \sum_{|\ell|,|r| \le M} |\mathrm{Cov}((A_\ell + C_\ell)^2 \sin(\frac{\pi}{2}\rho_X^*(\ell)), (A_r + C_r)^2 \sin(\frac{\pi}{2}\rho_X^*(r)))| \\
=& \frac{c\pi^2}{16^2} \sum_{|\ell|,|r| \le M} |\mathrm{E}\{(A_\ell^2 + 2A_\ell C_\ell + C_\ell^2)(A_r^2 + 2A_r C_r + C_r^2) \\
&\times \sin(\frac{\pi}{2}\rho_X^*(\ell)) \sin(\frac{\pi}{2}\rho_X^*(r))\} \\
&- \mathrm{E}\{(A_\ell^2 + 2A_\ell C_\ell + C_\ell^2) \sin(\frac{\pi}{2}\rho_X^*(\ell))\} \\
&\times \mathrm{E}\{(A_r^2 + 2A_r C_r + C_r^2) \sin(\frac{\pi}{2}\rho_X^*(r))\}|.
\end{aligned}
$$

From the above discussion of $\mathrm{Var}(\hat{g}(\lambda))$, it follows that $\mathrm{E}A_\ell^2 = O(1/n)$.
We can see that

$$|C_\ell| = \frac{|\ell|}{n}\rho_X(\ell), \sin(\frac{\pi}{2}\rho_X^*(\ell)) \le 1 \;\text{ and } |A_\ell| \le 1,$$

therefore it is sufficient to show that $\sum_{|\ell|,|r| \le M} \mathrm{E}A_r^2 A_\ell^2 = O(M^2/n^2)$.
From Lemma 7.2.2, we have

$$
\begin{aligned}
&\sum_{|\ell|,|r| \le M} \mathrm{E}A_r^2 A_l^2 \\
=& \frac{4^4}{n^4} \sum_{|\ell|,|r| \le M} \sum_{\substack{i_1,i_2=\ell+1,\cdots,n \\ j_1,j_2=r+1,\cdots,n}} \mathrm{E}(Y_{i_1}Y_{i_1-\ell} - \mathrm{E}Y_{i_1}Y_{i_1-\ell})(Y_{i_2}Y_{i_2-\ell} - \mathrm{E}Y_{i_2}Y_{i_2-\ell}) \\
&\times (Y_{j_1}Y_{j_1-r} - \mathrm{E}Y_{j_1}Y_{j_1-r})(Y_{j_2}Y_{j_2-r} - \mathrm{E}Y_{j_2}Y_{j_2-r}) \\
=& \frac{4^2}{n^4} \sum_{|\ell|,|r| \le M} \sum_{\substack{i_1,i_2=\ell+1,\cdots,n \\ j_1,j_2=r+1,\cdots,n}} [\mathrm{Cum}\{Y_{i_1}, Y_{i_1-\ell}, Y_{i_2}, Y_{i_2-\ell}, Y_{j_1}, Y_{j_1-r}, Y_{j_2}, Y_{j_2-r}\} \\
&+ \sum \mathrm{Cum}\{6terms\}\mathrm{Cum}\{2terms\} + \sum \mathrm{Cum}\{5terms\}\mathrm{Cum}\{3terms\} \\
&+ \sum \mathrm{Cum}\{4terms\}\mathrm{Cum}\{4terms\} \\
&+ \sum \mathrm{Cum}\{4terms\}\mathrm{Cum}\{2terms\}\mathrm{Cum}\{2terms\} \\
&+ \sum \mathrm{Cum}\{3terms\}\mathrm{Cum}\{3terms\}\mathrm{Cum}\{2terms\}
\end{aligned}
$$

$$+\sum \text{Cum}\{2terms\}\text{Cum}\{2terms\}\text{Cum}\{2terms\}\text{Cum}\{2terms\}],$$

where the all summations are in an appropriate range. Because of summability of cumulant for the process $\{X_t\}$, we can see the each summation is of order $O(M^2/n^2)$, therefore we get the result.

Using the Cauchy-Schwarz inequality, we get

$$\text{Cov}(\hat{h}(\lambda_1), \hat{g}(\lambda_1)) = O\left(\left(\frac{M}{n}\right)^{\frac{3}{2}}\right).$$

Next, we show $f_z(\lambda) - f_n(\lambda) = O(1/M^2)$. We observe

$$\begin{aligned}
&M^2(f_Z(\lambda) - f_n(\lambda)) \\
=&\frac{1}{2\pi}\sum_{|\ell|\le M}\frac{1-w\left(\frac{\ell}{M}\right)}{\frac{\ell^2}{M^2}}\ell^2\rho_Z(\ell)\exp(-i\ell\lambda) + \frac{M^2}{2\pi}\sum_{|\ell|>M}\rho_Z(\ell)\exp(-i\ell\lambda) \\
\to&\frac{k_2}{4\pi}\sum_{\ell=-\infty}^{\infty}\ell^2\rho_Z(\ell)\exp\left(-i\ell\lambda\right) < \infty \quad (n\to\infty),
\end{aligned}$$

then we get the result.

Finally, we have

$$\text{E}\hat{g}(\lambda) = -\frac{1}{4n}\sum_{|\ell|\le M}\sqrt{1-\rho_Z^2(\ell)}w\left(\frac{\ell}{M}\right)\exp(-i\ell\lambda)\ell\rho_X(\ell) = O\left(\frac{1}{n}\right),$$

and $\text{E}\hat{h}(\lambda) = O\left(1/n^2\right) + O(M/n)$, Hence we have the desired result. □

Lemma 7.2.3. *Under Assumption 3.1.1, then*

$$\max_{\lambda\in[-\pi,\pi]}(\hat{f}_Z(\lambda) - f_Z(\lambda)) = o_p(1)$$

Proof. For any $\epsilon > 0, \eta > 0$

$$\begin{aligned}
&\text{P}\left(\max_{\lambda\in[-\pi,\pi]}|\hat{f}_Z(\lambda) - f_Z(\lambda)| > \eta\right) \\
\le&\text{P}\left(\max_{\lambda\in[-\pi,\pi]}|\hat{f}_Z(\lambda) - f_Z(\lambda)| > \eta, \frac{1}{\epsilon}\le M\le\epsilon\sqrt{n}\right) \qquad (7.2.3)
\end{aligned}$$

$$+\mathrm{P}\left(\epsilon M < 1\right) + \mathrm{P}\left(M > \epsilon\sqrt{n}\right).$$

by Assumption 3.1.1, the second term and third term of (7.2.3) tend to 0. Hence it is sufficient to show

$$\max_{\substack{1/\epsilon\le m\le\epsilon\sqrt{n}\\ \lambda\in[-\pi,\pi]}} |\hat{f}_Z^m(\lambda) - f_Z(\lambda)| = o_p(1),$$

where $\hat{f}_Z^m(\lambda) := \frac{1}{2\pi}\sum_{|\ell|\le m} w\left(\ell/m\right)\hat{\rho}_Z(\ell)\exp\left(-i\ell\lambda\right)$. Write as

$$\begin{aligned}\hat{f}_Z^m(\lambda) - f_Z(\lambda) =& \hat{f}_Z^m(\lambda) - \mathrm{E}\hat{f}_Z^m(\lambda)\\ &+ \int_{[-\pi,\pi]} W_n(\lambda-\mu)\{\mathrm{E}\hat{I}_n(\mu) - f_Z(\mu)\}\mathrm{d}x\\ &+ \int_{-\infty}^{\infty} W(\theta)\{f_Z(\lambda - \frac{\theta}{m}) - f_Z(\lambda)\}\mathrm{d}s\\ =& a_1 + a_2 + a_3,\end{aligned}$$

where

$$\begin{aligned}a_1 &:= \hat{f}_Z^m(\lambda) - \mathrm{E}\hat{f}_Z^m(\lambda),\\ a_2 &:= \int_{[-\pi,\pi]} W_n(\lambda-\mu)\{\mathrm{E}\hat{I}_n(\mu) - f_Z(\mu)\}\mathrm{d}x,\\ a_3 &:= \int_{-\infty}^{\infty} W(\theta)\{f_Z(\lambda - \frac{\theta}{m}) - f_Z(\lambda)\}\mathrm{d}s.\end{aligned}$$

First, we show $\max_{\substack{1/\epsilon\le m\le\epsilon\sqrt{n}\\ \lambda\in[-\pi,\pi]}} |a_1| = o_p(1)$. From $\hat{\rho}_X(\ell) - \rho_X(\ell) = O_p(1/\sqrt{n})$ uniformly in ℓ, therefore we obtain

$$\begin{aligned}\max_{\substack{1/\epsilon\le m\le\epsilon\sqrt{n}\\ \lambda\in[-\pi,\pi]}} |a_1| =& \max_{\substack{1/\epsilon\le m\le\epsilon\sqrt{n}\\ \lambda\in[-\pi,\pi]}} |\hat{g}(\lambda) - \mathrm{E}\hat{g}(\lambda) + \hat{h}(\lambda) - \mathrm{E}\hat{h}(\lambda)|\\ \le& \frac{1}{4}\sum_{|\ell|\le\epsilon\sqrt{n}} \bar{w}(\frac{\ell}{\epsilon\sqrt{n}})|\left(\hat{\rho}_X(\ell) - \mathrm{E}\hat{\rho}_X(\ell)\right)|\\ &+ \max_{\substack{1/\epsilon\le m\le\epsilon\sqrt{n}\\ \lambda\in[-\pi,\pi]}} \frac{\pi}{16}\sum_{|\ell|\le m} w(\frac{\ell}{m})\left(\hat{\rho}_X(\ell) - \rho_X(\ell)\right)^2\\ =& O_p\left(\frac{1}{\sqrt{n}} + \epsilon\int_{-1}^{1}\bar{w}(\theta)\mathrm{d}\theta\right) + o_p\left(1\right).\end{aligned}$$

Because ϵ is arbitrary, we get the statement. Next, we show $\max_{\substack{1/\epsilon \le m \le \epsilon\sqrt{n} \\ \lambda\in[-\pi,\pi]}} |a_2| = o_p(1)$. It is easy to see that

$$\max_{\substack{1/\epsilon \le m \le \epsilon\sqrt{n} \\ \lambda\in[-\pi,\pi]}} |a_2| \le \max_{\substack{1/\epsilon \le m \le \epsilon\sqrt{n} \\ \lambda\in[-\pi,\pi]}} |\mathrm{E}\hat{I}_n(\lambda) - f_Z(\lambda)| \int_{-\infty}^{\infty} W(\theta)\mathrm{d}s.$$

We can also show

$$\begin{aligned}\max_{\substack{1/\epsilon \le m \le \epsilon\sqrt{n} \\ \lambda\in[-\pi,\pi]}} |\mathrm{E}\hat{I}_n(\lambda) - f_Z(\lambda)| =& (\frac{1}{4} + \frac{\pi}{16}) \sum_{|\ell|\le\epsilon\sqrt{n}} \frac{|\ell|}{n} |\rho_X(\ell)| + \sum_{|\ell|>1/\epsilon} \rho_Z(\ell) \\ &+ \frac{\pi}{16} \sum_{|\ell|\le\epsilon\sqrt{n}} \mathrm{E}(\hat{\rho}_X(\ell) - \mathrm{E}\hat{\rho}_X(\ell))^2,\end{aligned}$$

which tends to zero as $n \to \infty$ and $\epsilon \to 0$.
Finally, we show $\max_{\substack{1/\epsilon \le m \le \epsilon\sqrt{n} \\ \lambda\in[-\pi,\pi]}} |a_3| = o_p(1)$. For any $\xi > 0$,

$$\begin{aligned}&\max_{\substack{1/\epsilon \le m \le \epsilon\sqrt{n} \\ \lambda\in[-\pi,\pi]}} |a_3| \\ \le& \max_{\substack{|\theta|\le\xi/\epsilon,\ \lambda\in[-\pi,\pi] \\ 1/\epsilon \le m \le \epsilon\sqrt{n}}} |f_Z(\lambda - \frac{\theta}{m}) - f_Z(\lambda)| \int_{-\infty}^{\infty} |W(\theta)|\mathrm{d}s \\ &+ \max_{\substack{1/\epsilon \le m \le \epsilon\sqrt{n} \\ \lambda\in[-\pi,\pi]}} \int_{|\theta|>\xi/\epsilon} |W(\theta)||f_Z(\lambda - \frac{\theta}{m}) - f_Z(\lambda)|\mathrm{d}\theta \\ &- \max_{\substack{|\theta|\le\xi/\epsilon,\ \lambda\in[-\pi,\pi] \\ 1/\epsilon \le m \le \epsilon\sqrt{n}}} |f_Z(\lambda - \frac{\theta}{m}) - f_Z(\lambda)| \int_{|\theta|>\xi/\epsilon} |W(\theta)|\mathrm{d}\theta \\ &+ \max_{\substack{1/\epsilon \le m \le \epsilon\sqrt{n} \\ \lambda\in[-\pi,\pi]}} \int_{|\theta|\le\xi/\epsilon} |W(\theta)||f_Z(\lambda - \frac{\theta}{m}) - f_Z(\lambda)|\mathrm{d}\theta \\ &- \max_{\substack{|\theta|\le\xi/\epsilon,\ \lambda\in[-\pi,\pi] \\ 1/\epsilon \le m \le \epsilon\sqrt{n}}} |f_Z(\lambda - \frac{\theta}{m}) - f_Z(\lambda)| \int_{|\theta|\le\xi/\epsilon} |W(\theta)|\mathrm{d}\theta \\ \le& \max_{\substack{|\theta|\le\xi/\epsilon,\ \lambda\in[-\pi,\pi] \\ 1/\epsilon \le m \le \epsilon\sqrt{n}}} |f_Z(\lambda - \frac{\theta}{m}) - f_Z(\lambda)| \int_{-\infty}^{\infty} |W(\theta)|\mathrm{d}s \qquad (7.2.4) \\ &+ 4 \max_{\lambda\in[-\pi,\pi]} |f_Z(\lambda)| \int_{|\theta|>\xi/\epsilon} |W(\theta)|\mathrm{d}\theta.\end{aligned}$$

Because f_Z is uniform continuous and $|\theta/m| < \xi$, the first term of the right hand side of (7.2.4) can be made arbitrarily small by choosing small ξ. The second term of the right hand side of (7.2.4) tends to zero as $\epsilon \to 0$, hence we have the desired result. □

Remark 7.2.1. *Robinson (1991) showed the consistency of smoothed periodogram uniformly in $\lambda \in [-\pi, \pi]$.*

Lemma 7.2.4. *Let $\{Z_t\}$ be a strictly stationary ellipsoidal α-mixing process with zero mean, finite variance, autocorrelation $\rho_Z(\ell)$ and and α-mixing coefficients $\alpha(\cdot)$ satisfying $\alpha(n) = O(1/n^{8+\delta})$ for some $\delta > 0$ and $\hat{\rho}_Z(\ell)$ be the binary estimator of $\rho_Z(\ell)$. Then,*

$$\sqrt{n}\begin{pmatrix} \hat{\rho}_Z(1) - \rho_Z(1) \\ \hat{\rho}_Z(2) - \rho_Z(2) \\ \vdots \\ \hat{\rho}_Z(m) - \rho_Z(m) \end{pmatrix} \Rightarrow N(0, A'VA)$$

where $V = (v_{\ell,r})_{\ell,r=1,\cdots,m}$,
$v_{\ell,r} = \sum_{u=-\infty}^{\infty}\{16K_X(-\ell, u, u-r) + \rho_X(u)\rho_X(u+\ell-r) + \rho_X(u-r)\rho_X(u+\ell)\}$ *and*
$A = \text{diag}\left(\pi\sqrt{1-\rho_1^2}, \pi\sqrt{1-\rho_2^2}, \cdots, \pi\sqrt{1-\rho_m^2}\right)'$.

From lemma 7.2.1, we can see the asymptotic variance is given by $A'VA$.
Then, we show that the asymptotic normality. We define the random vector

$$\mathbf{W_i} = \begin{pmatrix} Y_iY_{i-1}\mathbb{I}_{\{2\le i\}} \\ Y_iY_{i-2}\mathbb{I}_{\{3\le i\}} \\ \vdots \\ Y_iY_{i-m}\mathbb{I}_{\{m+1\le i\}} \end{pmatrix}.$$

From the Cramer-Wold device, it is sufficient to show that, for arbitrary $\boldsymbol{\lambda} \in \mathbb{R}^m$, $(4\sqrt{n})\sum_{i=1}^{n}\boldsymbol{\lambda}'(\mathbf{W_i} - \mathrm{E}\mathbf{W_i}) \Rightarrow N(0, \boldsymbol{\lambda}'V\boldsymbol{\lambda})$. We can see that $\boldsymbol{\lambda}'(\mathbf{W_i} - \mathrm{E}\mathbf{W_i})$ is α-mixing process with the same ϕ-mixing coefficients as $\{Z_t\}$. From the central limit theorem for mixing sequences (see Billinsley (1968)) and using Taylor expansion, we get the desired result. □

Let $H(\cdot,\cdot)$ be a function satisfying: (H1) (i)$H(\mathcal{Z}, \lambda)$ is defined on $\mathcal{D}\times[-\pi, \pi]$, where $\mathcal{D}$ is an open subset of $\mathbb{C}$ which contains the whole range of spectral density function of the process.
(ii) $H(\mathcal{Z}, \lambda)$ is holomorphic at $f_Z(\lambda)$.
(iii) For a real, positive definite function $\mathcal{Z} \in \mathcal{D}$, $H(\mathcal{Z}, \lambda)$ is real valued.

(H2) There exists a positive constant r such that (i) r is independent of λ,
(ii) for every $\lambda \in [-\pi, \pi]$, the ball $B_\lambda = \{\mathcal{Z} \in C^s : |\mathcal{Z} - f_Z(\lambda)| \leq r\}$ is contained in $\mathcal{D}$,
(iii) $sup_{\mathcal{Z} \in \partial C_\lambda} |H(\mathcal{Z}, \lambda) - H(f_Z(\lambda), \lambda)| \leq h(\lambda)$, where $\int_{[-\pi,\pi]} h(\lambda) \mathrm{d}\lambda < \infty$, and $\partial C_\lambda = \{\mathcal{Z} : \mathcal{Z} = f_Z(\lambda) + r \exp(i\theta), -\pi \leq \theta < \pi\}$.

(H3) The first derivative of $H(\mathcal{Z}, \lambda)$ respect to $\mathcal{Z}$, $H^{(1)}(\mathcal{Z}, \cdot)$, satisfies $H^{(1)}(f_Z(-\pi), -\pi) = H^{(1)}(f_Z(\pi), \pi)$.

Lemma 7.2.5. *Under Assumption 3.1.1 and (H1),(H2), we have*

$$\int_{[-\pi,\pi]} H\{\hat{f}_Z(\lambda), \lambda\} \mathrm{d}\lambda - \int_{[-\pi,\pi]} H\{f_Z(\lambda), \lambda\} \mathrm{d}\lambda = o_p(1). \tag{7.2.5}$$

Moreover, if $H(\cdot, \cdot)$ satisfies (H3) and (H4), then

$$\sqrt{n} \left\{ \int_{[-\pi,\pi]} H\{\hat{f}_Z(\lambda), \lambda\} \mathrm{d}\lambda - \int_{[-\pi,\pi]} H\{f_Z(\lambda), \lambda\} \mathrm{d}\lambda \right\} \Rightarrow N(0, w^2) \tag{7.2.6}$$

as $N \to \infty$, where $w^2 = \sum_{\ell,r=-\infty}^{\infty} \tilde{\phi}(\ell)\tilde{\phi}(r)\sqrt{1 - \rho_Z^2(\ell)}\sqrt{1 - \rho_Z^2(r)}$ $\times \sum_{t=-\infty}^{\infty} [16K_X(-\ell, t, t - r) + \{\rho_X(t)\rho_X(t + \ell - r) + \rho_X(t - r)\rho_X(t + \ell)\}]$ with $\tilde{\phi}(j) = 1/(2\pi) \int_{[-\pi,\pi]} H^{(1)}(f(\mu), \mu) \exp(-ij\mu) \mathrm{d}x$.

Proof. First, we show (7.2.5). From (H1), there exists an open neighborhood $U \subset D$ such that $H\{\mathcal{Z}, \lambda\}$, $\mathcal{Z} \in U$ is analytic. To begin with, we evaluate the remainder term when we expand $H\{\mathcal{Z}, \cdot\}$ as a Taylor expansion around $f(\lambda)$ under the condition $|\hat{f}_Z(\lambda) - f_Z(\lambda)| < \delta$, where δ is taken so that $\hat{f}(\lambda) \in U$. Then, the remainder term R is

$$\begin{aligned} R(\lambda) :=& H(\hat{f}_Z(\lambda), \lambda) - H(f_Z(\lambda), \lambda) - H^{(1)}(f_Z(\lambda), \lambda) \left(\hat{f}_Z(\lambda) - f_Z(\lambda)\right) \\ =& \frac{1}{2} H^{(2)}(\tilde{f}_Z(\lambda), \lambda) \left(\hat{f}_Z(\lambda) - f_Z(\lambda)\right)^2, \end{aligned}$$

where $\tilde{f}_Z(\lambda)$ is a function on U between $f_Z(\lambda)$ and $\hat{f}_Z(\lambda)$. By using Cauchy's integral formula for derivatives of $H(\tilde{f}_Z(\lambda), \lambda) - H(f_Z(\lambda), \lambda)$ and (H2), we observe that

$$|R(\lambda)| \leq a h(\lambda) |\hat{f}_Z(\lambda) - f_Z(\lambda)|^2 \qquad \left(|\hat{f}_Z(\lambda) - f_Z(\lambda)| \leq \delta\right),$$

where a is a constant. Moreover, by using Lemma 7.2.1 and employing Fubini's theorem, we can see, for arbitrary $\epsilon > 0$,

$$\mathrm{P}\Big(\left| \int_{[-\pi,\pi]} R(\lambda) \mathrm{d}\lambda \right| > \frac{M\epsilon}{N}\Big)$$

$$
\begin{aligned}
&\leq \frac{Na\int_{[-\infty,\infty]} h(\lambda)d\lambda}{M\epsilon} \mathrm{E}\left|\int_{[-\pi,\pi]} \{\hat{f}_Z(\lambda) - f_Z(\lambda)\}^2 \mathrm{d}\lambda\right| \\
&= \frac{a\int_{[-\infty,\infty]} h(\lambda)d\lambda}{\epsilon} \int_{[-\pi,\pi]} \frac{N}{M}\mathrm{E}\{\hat{f}_Z(\lambda) - f_Z(\lambda)\}^2 \mathrm{d}\lambda \\
&\quad \to \frac{8\pi^2 a\int_{[-\infty,\infty]} h(\lambda)d\lambda \int_{-1}^{1} k^2(\theta)\mathrm{d}s}{\epsilon} \int_{[-\pi,\pi]} f_X^2(\lambda)\mathrm{d}\lambda \\
&\quad \leq \frac{8\pi^2 a\int_{[-\infty,\infty]} h(\lambda)d\lambda \int_{-1}^{1} k^2(\theta)\mathrm{d}s}{\epsilon} \int_{[-\pi,\pi]} (\sum_{l=-\infty}^{\infty} |\rho_Z(l)|)^2 \mathrm{d}\lambda \\
&\quad < \infty \qquad \left(|\hat{f}_Z(\lambda) - f_Z(\lambda)| \leq \delta\right).
\end{aligned}
$$

From the above discussion and Lemma 7.2.3, for every $\epsilon > 0$, there exists constant c and $N \in \mathbb{N}$ such that for $n > N$,

$$
\begin{aligned}
&\mathrm{P}\left(\left|\int_{[-\pi,\pi]} R(\lambda)\mathrm{d}\lambda\right| > \frac{Mc}{n}\right) \\
\leq &\mathrm{P}(\frac{n}{M}\int_{[-\pi,\pi]} ah(\lambda)|\hat{f}_Z(\lambda) - f_Z(\lambda)|^2 \mathrm{d}\lambda \geq c) \\
&+ \mathrm{P}\left(\max_{\lambda\in[-\pi,\pi]} |\hat{f}_Z(\lambda) - f_Z(\lambda)| > \delta\right) \\
< &\epsilon.
\end{aligned}
$$

Then we have

$$
\begin{aligned}
&\int_{[-\pi,\pi]} H\{\hat{f}_Z(\lambda), \lambda\}\mathrm{d}\lambda - \int_{[-\pi,\pi]} H\{f_Z(\lambda), \lambda\}\mathrm{d}\lambda \\
= &\int_{[-\pi,\pi]} \{\hat{f}_Z(\lambda) - f_Z(\lambda)\}H^{(1)}(f_Z(\lambda), \lambda)\mathrm{d}\lambda + O_p\left(\frac{M}{n}\right) \\
\leq &\frac{1}{\delta}\int_{[-\pi,\pi]} h(\lambda)\mathrm{d}\lambda \max_{\lambda\in[-\pi,\pi]} |\hat{f}_Z(\lambda) - f_Z(\lambda)| + O_p\left(\frac{M}{n}\right) \\
= &o_p(1),
\end{aligned}
$$

which proves (7.2.5).

Next, we show (7.2.6). From above discussion, we have

$$
\begin{aligned}
&\sqrt{n}\left\{\int_{[-\pi,\pi]} H\{\hat{f}_Z(\lambda), \lambda\}\mathrm{d}\lambda - \int_{[-\pi,\pi]} H\{f_Z(\lambda), \lambda\}\mathrm{d}\lambda\right\} \\
= &\sqrt{n}\int_{[-\pi,\pi]} \{\hat{f}_Z(\lambda) - f_Z(\lambda)\}H^{(1)}(f_Z(\lambda), \lambda)\mathrm{d}\lambda + O_p\left(\frac{M}{\sqrt{n}}\right)
\end{aligned}
$$

$$
\begin{aligned}
=&\sqrt{n}\int_{[-\pi,\pi]} H^{(1)}(f_Z(\lambda),\lambda)\left\{\hat{I}_n(\lambda)-\mathrm{E}\hat{I}_n(\lambda)\right\}\mathrm{d}\lambda \\
&+\sqrt{n}\int_{[-\pi,\pi]} H^{(1)}(f_Z(\lambda),\lambda)\left\{\mathrm{E}\hat{I}_n(\lambda)-f_Z(\lambda)\right\}\mathrm{d}\lambda \\
&+\sqrt{n}\int_{[-\pi,\pi]}\left[H^{(1)}(f_Z(\lambda),\lambda)\left\{\int_{[-\pi,\pi]} f_Z(\mu)W_n(\lambda-\mu)\mathrm{d}x - f_Z(\lambda)\right\}\right]\mathrm{d}\lambda \\
&+\sqrt{n}\int_{[-\pi,\pi]}\left[H^{(1)}(f_Z(\lambda),\lambda)\int_{[-\pi,\pi]}\{\hat{I}_n(\mu)-f_Z(\mu)\}W_n(\lambda-\mu)\mathrm{d}x\right]\mathrm{d}\lambda \\
&-\sqrt{n}\int_{[-\pi,\pi]} H^{(1)}(f_Z(\lambda),\lambda)\left\{\hat{I}_n(\lambda)-f_Z(\lambda)\right\}\mathrm{d}\lambda + O_p\left(\frac{M}{\sqrt{n}}\right) \\
=&L_1+L_2+L_3+L_4+O_p\left(\frac{M}{\sqrt{n}}\right),
\end{aligned}
$$

where

$$
\begin{aligned}
L_1 :=&\sqrt{n}\int_{[-\pi,\pi]} H^{(1)}(f_Z(\lambda),\lambda)\left\{\hat{I}_n(\lambda)-\mathrm{E}\hat{I}_n(\lambda)\right\}\mathrm{d}\lambda, \\
L_2 :=&\sqrt{n}\int_{[-\pi,\pi]} H^{(1)}(f_Z(\lambda),\lambda)\left\{\mathrm{E}\hat{I}_n(\lambda)-f_Z(\lambda)\right\}\mathrm{d}\lambda, \\
L_3 :=&\sqrt{n}\int_{[-\pi,\pi]}\left[H^{(1)}(f_Z(\lambda),\lambda)\left\{\int_{[-\pi,\pi]} f(\mu)W_n(\lambda-\mu)\mathrm{d}x - f_Z(\lambda)\right\}\right]\mathrm{d}\lambda, \\
L_4 :=&\sqrt{n}\int_{[-\pi,\pi]}\left[H^{(1)}(f_Z(\lambda),\lambda)\int_{[-\pi,\pi]}\{\hat{I}_n(\mu)-f(\mu)\}W_n(\lambda-\mu)\mathrm{d}x\right]\mathrm{d}\lambda \\
&-\sqrt{n}\int_{[-\pi,\pi]} H^{(1)}(f_Z(\lambda),\lambda)\left\{\hat{I}_n(\lambda)-f_Z(\lambda)\right\}\mathrm{d}\lambda.
\end{aligned}
$$

(i) To begin with, we show $L_1 \Rightarrow N(0,w^2) \quad as\ N \to \infty.$

By Fejér's theorem, for any ϵ, there exists $L_0 \in \mathbb{N}$ such that the Cesaro sum, i.e.

$$
\phi_L(\lambda) := \frac{1}{2\pi}\sum_{j=-L+1}^{L-1}(1-\frac{|j|}{L})\tilde{\phi}(j)\exp(-ij\lambda),
$$

satisfies $\sup_{\lambda\in[-\pi,\pi]}|H^{(1)}(f_Z(\lambda),\lambda)-\phi_L(\lambda)| < \epsilon$ for $L > L_0$,
where $\tilde{\phi}(j) := 1/(2\pi)\int_{[-\pi,\pi]} H^{(1)}(f_Z(\mu),\mu)\exp(-ij\mu)\mathrm{d}x$.
Let $\delta_L(\lambda)$ be the function defined by

$$
\delta_L(\lambda) = H^{(1)}(f_Z(\lambda),\lambda) - \phi_L(\lambda).
$$

We have

$$
\begin{aligned}
&\mathrm{Var}\left[\sqrt{n}\int_{[-\pi,\pi]}\delta_L(\lambda)\hat{I}_n(\lambda)\mathrm{d}\lambda\right]\\
\leq&\sum_{\ell,r,t=-\infty}^{\infty}\left|\int_{[-\pi,\pi]}\delta_L(\lambda)\exp(-i\ell\lambda)\mathrm{d}\lambda\right|\left|\int_{[-\pi,\pi]}\delta_L(\lambda)\exp(-ir\lambda)\mathrm{d}\lambda\right|\\
&\times\sqrt{1-\rho_Z^2(\ell)}\sqrt{1-\rho_Z^2(r)}[16|K_X(-\ell,t,t-r)|\\
&+\{|\rho_Y(t)\rho_X(t+\ell-r)|+|\rho_X(t-r)\rho_X(t+\ell)|\}]+o(1)
\end{aligned}
$$

which tends to 0 as $L\to\infty$.

For each L, we observe

$$
\begin{aligned}
&\sqrt{n}\int_{[-\pi,\pi]}\phi_L(\lambda)\left\{\hat{I}_n(\lambda)-\mathrm{E}\hat{I}_n(\lambda)\right\}\mathrm{d}\lambda\\
=&\frac{1}{2\pi}\sum_{j=-L+1}^{L-1}(1-\frac{|j|}{L})\tilde{\phi}(j)\sqrt{n}(\hat{\rho}_Z(j)-\mathrm{E}\hat{\rho}_Z(j)).
\end{aligned}
$$

The asymptotic normality for follows from Bernstein's Lemma (Hannan (1970)), and the asymptotic variance is given by

$$
\begin{aligned}
&\lim_{L\to\infty}\sum_{\ell,r=-L+1}^{L-1}(1-\frac{|\ell|}{L})\tilde{\phi}(\ell)(1-\frac{|r|}{L})\tilde{\phi}(r)\sqrt{1-\rho_Z^2(\ell)}\sqrt{1-\rho_Z^2(r)}\\
&\times\sum_{t=-\infty}^{\infty}[16K_Y(-\ell,t,t-r)+\{\rho_X(t)\rho_X(t+\ell-r)\\
&+\rho_X(t-r)\rho_X(t+\ell)\}],
\end{aligned}
$$

which leads to the desired result.

(ii) Next, we show $L_2=o_p(1)\ as\ N\to\infty.$
From the continuity of $H^{(1)}$, we can see

$$
\begin{aligned}
&\sqrt{n}\int_{[-\pi,\pi]}H^{(1)}(f_Z(\lambda),\lambda)\left\{\mathrm{E}\hat{I}_n(\lambda)-f_Z(\lambda)\right\}\mathrm{d}\lambda\\
\leq&\max_{(y,x)\in Range(f)\times[-\pi,\pi]}|H^{(1)}(y,x)|\sqrt{n}\int_{[-\pi,\pi]}|\mathrm{E}\hat{I}_n(\lambda)-f_Z(\lambda)|\mathrm{d}\lambda\\
=&\max_{(y,x)\in Range(f)\times[-\pi,\pi]}|H^{(1)}(y,x)|\\
&\times\left(\sqrt{n}\sum_{|\ell|\leq M}|\mathrm{E}\hat{\rho}_Z(\ell)-\rho_Z(\ell)|+\sqrt{n}\sum_{|\ell|>M}|\rho_Z(\ell)|\right)
\end{aligned}
$$

$$
\begin{aligned}
\leq & \max_{(y,x)\in Range(f)\times[-\pi,\pi]} |H^{(1)}(y,x)| \\
& \times\left(\sqrt{n}\sum_{|\ell|\leq M}|\mathrm{E}\hat{\rho}_Z(\ell)-\rho_Z(\ell)| + \frac{\sqrt{n}}{M^2}\sum_{|\ell|>M}\ell^2|\rho_Z(\ell)|\right) \\
=& o_p(1).
\end{aligned}
$$

(iii) Next, we show $L_3 = o_p(1) \quad as\ N \to \infty$
We observe

$$
\begin{aligned}
& \sqrt{n}\left|\int_{[-\pi,\pi]} f_Z(\mu)W_n(\lambda-\mu)\mathrm{d}x - f_Z(\lambda)\right| \\
=& \frac{\sqrt{n}}{M^2}\frac{M^2}{2\pi}\left|\sum_{\ell=-\infty}^{\infty}\rho_Z(\ell)\exp(-i\ell\lambda)w\left(\frac{\ell}{M}\right) - \sum_{\ell=-\infty}^{\infty}\rho_Z(\ell)\exp(-i\ell\lambda)\right| \\
\leq& \frac{\sqrt{n}}{M^2}\frac{M^2}{2\pi}\left|\sum_{|\ell|\leq M}\rho_Z(\ell)\exp(-i\ell\lambda)\left\{1-w\left(\frac{\ell}{M}\right)\right\}\right| \\
& + \frac{\sqrt{n}}{M^2}\frac{M^2}{2\pi}\left|\sum_{\ell\geq M}\rho_Z(\ell)\exp(-i\ell\lambda)\right| \\
\leq& \frac{\sqrt{n}}{M^2}\frac{1}{2\pi}\sum_{|\ell|\leq M}|\ell|^2|\rho_Z(\ell)|\left|\frac{1-w\left(\frac{\ell}{M}\right)}{\frac{\ell^2}{M^2}}\right| + \frac{\sqrt{n}}{M^2}\frac{1}{2\pi}\sum_{|\ell|\geq M}|\ell|^2|\rho_Z(\ell)| \\
& \to 0\ as\ n\to\infty,
\end{aligned}
$$

we get the result.
(iv) Last of all, we show $L_4 = o_p(1) \quad as\ N \to \infty$.
Putting $M(\lambda - \mu + 2\pi\nu) = \eta$, we have

$$
\begin{aligned}
& \sqrt{n}\int_{[-\pi,\pi]}\left[H^{(1)}(f(\lambda),\lambda)\int_{[-\pi,\pi]}\{\hat{I}_n(\mu)-f(\mu)\}W_n(\lambda-\mu)\mathrm{d}x\right]\mathrm{d}\lambda \\
=& \sqrt{n}\int_{[-\pi,\pi]}\left[\int_{\infty}^{\infty}H^{(1)}(f_Z(\frac{\eta}{M}+\mu),\frac{\eta}{M}+\mu)W(\eta)\mathrm{d}\eta\right] \\
& \times\{\hat{I}_n(\mu)-f(\mu)\}\mathrm{d}\mu,
\end{aligned}
$$

where $\tilde{H}^{(1)}(\cdot,\cdot)$ is a periodic extension of $H^{(1)}(\cdot,\cdot)$. Then we have

$$
|\sqrt{n}\int_{[-\pi,\pi]}\left[H^{(1)}(f_Z(\lambda),\lambda)\int_{[-\pi,\pi]}\{\hat{I}_n(\mu)-f_Z(\mu)\}W_n(\lambda-\mu)\mathrm{d}x\right]\mathrm{d}\lambda
$$

$$
\begin{aligned}
&- \sqrt{n} \int_{[-\pi,\pi]} H^{(1)}(f(\lambda),\lambda) \left\{\hat{I}_n(\lambda) - f_Z(\lambda)\right\} \mathrm{d}\lambda| \\
&= \left| \sqrt{n} \int_{[-\pi,\pi]} A_M(\mu)\{\hat{I}_n(\mu) - f_Z(\mu)\} \mathrm{d}x \right|,
\end{aligned}
$$

where

$$
A_M(\mu) = \left[\int_{\infty}^{\infty} \tilde{H}^{(1)}(f_Z(\frac{\eta}{M} + \mu), \frac{\eta}{M} + \mu) W(\eta) \mathrm{d}\eta - H^{(1)}(f_Z(\mu),\mu) \right].
$$

By the above discussions,

$$
\begin{aligned}
&\mathrm{Var}\left[\sqrt{n} \int_{[-\pi,\pi]} A_M(\mu)\{\hat{I}_n(\mu) - f_Z(\mu)\} \mathrm{d}x \right] \\
&\leq \sum_{\ell,r,t=-\infty}^{\infty} \left| \int_{[-\pi,\pi]} A_M(\lambda) \exp(-i\ell\lambda) \mathrm{d}\lambda \right| \left| \int_{[-\pi,\pi]} A_M(\lambda) \exp(-ir\lambda) \mathrm{d}\lambda \right| \\
&\times \sqrt{1 - \rho_Z^2(\ell)} \sqrt{1 - \rho_Z^2(r)} [16 |K_X(-\ell, t, t-r)| \\
&+ \{ |\rho_X(t)\rho_X(t+\ell-r)| + |\rho_X(t-r)\rho_X(t+\ell)\} |] \\
&+ o(1),
\end{aligned}
$$

which tends to 0 as $N \to \infty$.

□

Remark 7.2.2. *We referred to Taniguchi and Kakizawa (2000) to prove the consistency, Hosoya and Taniguchi (1982) to evaluate L_1, L_2, Hannan (1970) to see $L_3 = o_p(1)$, Taniguchi (1987) and Taniguchi and Kakizawa (2000) to show $L_4 = o_p(1)$.*

Proof of Theorem 3.2.1
From Lemma 7.2.5, the result follows. □

Proof of Theorem3.2.2
Under Π_1, we can observe

$$
\begin{aligned}
&\int_{[-\pi,\pi]} H\left(\frac{\hat{f}_Z(\lambda)}{g_Z(\lambda)}\right) \mathrm{d}\lambda \\
&= \frac{c}{n} \int_{[-\pi,\pi]} \left(\frac{h(\lambda)}{f_Z(\lambda) + h(\lambda)/\sqrt{n}} \right)^2 \mathrm{d}\lambda + O\left(\frac{1}{n\sqrt{n}}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
&\int_{[-\pi,\pi]} H^{(1)}\left(\frac{\hat{f}_Z(\lambda)}{g_Z(\lambda)}\right)\exp(-i\ell\lambda)\mathrm{d}\lambda \\
=&\frac{c}{\sqrt{n}}\int_{[-\pi,\pi]}\left(\frac{h(\lambda)}{f_Z(\lambda)+h(\lambda)/\sqrt{n}}\right)\exp(-i\ell\lambda)\mathrm{d}\lambda + O\left(\frac{1}{n}\right).
\end{aligned}
$$

From a direct application of the above we can get the desired result. □

Proof of Theorem 3.3.1
From the definition of the nonparametric estimator (3.3.1), and I-divergence (3.3.2), we have

$$
\begin{aligned}
&D(\hat{f}_{Z^s}, f_Z, g_Z) - D(\hat{f}_Z, f_Z, g_Z) \\
=&\frac{1}{4\pi}\int_{[-\pi,\pi]}\left(\hat{f}_{Z^s}(\lambda)-\hat{f}_Z(\lambda)\right)\left(\frac{1}{g_Z(\lambda)}-\frac{1}{f_Z(\lambda)}\right)\mathrm{d}\lambda \\
=&\frac{Z_1^s}{8\pi^2 n}\left\{\omega(0)Z_1^s\int_{[-\pi,\pi]}\left(\frac{1}{g_Z(\lambda)}-\frac{1}{f_Z(\lambda)}\right)\mathrm{d}\lambda\right. \\
&\left.-\sum_{\substack{|\ell|\le M\\ \ell\neq 0}}\omega\left(\frac{\ell}{M}\right)Z_{1+\ell}\int_{[-\pi,\pi]}\left(\frac{1}{g_Z(\lambda)}-\frac{1}{f_Z(\lambda)}\right)\exp\{-i\ell\lambda\}\mathrm{d}\lambda\right\} \\
&+\frac{1}{8\pi^2 n}\omega(0)Z_1^2\int_{[-\pi,\pi]}\left(\frac{1}{g_Z(\lambda)}-\frac{1}{f_Z(\lambda)}\right)\mathrm{d}\lambda \\
&-\frac{1}{8\pi^2 n}\sum_{\substack{|\ell|\le M\\ \ell\neq 0}}\omega\left(\frac{\ell}{M}\right)Z_1 Z_{1+\ell}\int_{[-\pi,\pi]}\left(\frac{1}{g_Z(\lambda)}-\frac{1}{f_Z(\lambda)}\right)\exp\{-i\ell\lambda\}\mathrm{d}\lambda.
\end{aligned}
$$

Because the braces expression of the above is positive for sufficiently large s and the other terms are constant and independent of s, we get the desired result. □

Proof of Theorem 3.3.2
The proof is omitted. □

Proof of Theorem 3.4.1
First, we show

$$
\max_{\lambda\in[-\pi,\pi]}\left(\frac{\hat{f}_Z(\lambda)}{\hat{g}_Z^T(\lambda)}-\frac{f_Z(\lambda)}{g_Z(\lambda)}\right) = o_p(1).
$$

From Assumption 3.2.1 (C), there exist d such that $g(\lambda) - d > 0$. By Lemma

7.2.3, for any $\epsilon > 0$ and sufficiently large n,

$$
\begin{aligned}
&\mathrm{P}\left(\max_{\lambda\in[-\pi,\pi]}\left(\frac{\hat{f}_Z(\lambda)}{\hat{g}_Z^T(\lambda)}-\frac{f_Z(\lambda)}{g_Z(\lambda)}\right)>\epsilon\right)\\
=&\mathrm{P}\left(\max_{\lambda\in[-\pi,\pi]}\left(\frac{\hat{f}_Z(\lambda)}{\hat{g}_Z^T(\lambda)}-\frac{f_Z(\lambda)}{g_Z(\lambda)}\right)>\epsilon,\max_{\lambda\in[-\pi,\pi]}(\hat{g}_Z^T(\lambda)-g_Z(\lambda))>c\right)\\
&+\mathrm{P}\left(\max_{\lambda\in[-\pi,\pi]}\left(\frac{\hat{f}_Z(\lambda)}{\hat{g}_Z^T(\lambda)}-\frac{f_Z(\lambda)}{g_Z(\lambda)}\right)>\epsilon,\max_{\lambda\in[-\pi,\pi]}(\hat{g}_Z^T(\lambda)-g_Z(\lambda))\le c\right)\\
\le&\mathrm{P}\left(\max_{\lambda\in[-\pi,\pi]}(\hat{g}_Z^T(\lambda)-g_Z(\lambda))>c\right)\\
&+\mathrm{P}\left(\max_{\lambda\in[-\pi,\pi]}\left(\frac{g_Z(\lambda)(\hat{f}_Z(\lambda)-f_Z(\lambda))+f_Z(\lambda)(g_Z(\lambda)-g_Z^T(\lambda))}{(g(\lambda)-c)g(\lambda)}\right)>\epsilon\right)\\
&\to 0\quad n\to\infty.
\end{aligned}
$$

We can see the followings the same way as Lemma 7.2.5. Using continuity of $f(\lambda)$ and $g(\lambda)$, Lemma 7.2.1, and Lemma 7.2.3, for $\left(\frac{\hat{f}_Z(\lambda)}{\hat{g}_Z^T(\lambda)}-\frac{f_Z(\lambda)}{g_Z(\lambda)}\right)<\delta$, any $\epsilon>0$, and the above d,

$$
\begin{aligned}
&\mathrm{P}(\left|\int_{[-\pi,\pi]}R'(\lambda)\mathrm{d}\lambda\right|>\frac{M\epsilon}{N})\\
=&\mathrm{P}(\left|\int_{[-\pi,\pi]}R'(\lambda)\mathrm{d}\lambda\right|>\frac{M\epsilon}{N},\max_{\lambda\in[-\pi,\pi]}(\hat{g}_Z^T(\lambda)-g_Z(\lambda))>c)\\
&+\mathrm{P}(\left|\int_{[-\pi,\pi]}R'(\lambda)\mathrm{d}\lambda\right|>\frac{M\epsilon}{N},\max_{\lambda\in[-\pi,\pi]}(\hat{g}_Z^T(\lambda)-g_Z(\lambda))\le c)\\
\le&\mathrm{P}(\max_{\lambda\in[-\pi,\pi]}(\hat{g}_Z^T(\lambda)-g_Z(\lambda))>c)\\
&+\mathrm{P}(\left|\int_{[-\pi,\pi]}R'(\lambda)\mathrm{d}\lambda\right|>\frac{M\epsilon}{N},\max_{\lambda\in[-\pi,\pi]}(\hat{g}_Z^T(\lambda)-g_Z(\lambda))\le c)\\
\le&\epsilon+\frac{CN}{M}\mathrm{E}\left|\int_{[-\pi,\pi]}\left\{\left(\frac{\hat{f}_Z(\lambda)}{\hat{g}_Z^T(\lambda)}-\frac{f_Z(\lambda)}{g_Z(\lambda)}\right)\mathbb{I}_{\{\max_{\lambda\in[-\pi,\pi]}(\hat{g}^T(\lambda)-g_Z(\lambda))\le c)\}}\right\}^2\mathrm{d}\lambda\right|\\
=&\epsilon+\int_{[-\pi,\pi]}\frac{CN}{M}\mathrm{E}\left\{\left(\frac{\hat{f}_Z(\lambda)}{\hat{g}^T(\lambda)}-\frac{f_Z(\lambda)}{g_Z(\lambda)}\right)\mathbb{I}_{\{\max_{\lambda\in[-\pi,\pi]}(\hat{g}_Z^T(\lambda)-g_Z(\lambda))\le c)\}}\right\}^2\mathrm{d}\lambda\\
\le&\epsilon+\int_{[-\pi,\pi]}\frac{CN}{M}\mathrm{E}\left(\frac{g_Z(\lambda)(\hat{f}_Z(\lambda)-f_Z(\lambda))+f_Z(\lambda)(g_Z(\lambda)-g_Z^T(\lambda))}{(g_Z(\lambda)-c)g_Z(\lambda)}\right)^2\mathrm{d}\lambda
\end{aligned}
$$

$<\infty,$

where C is a some constant R' is the remainder term defined by

$$R'(\lambda) := H\left(\frac{\hat{f}_Z(\lambda)}{\hat{g}_Z^T(\lambda)}\right) - H\left(\frac{f_Z(\lambda)}{g_Z(\lambda)}\right) - H^{(1)}\left(\frac{f_Z(\lambda)}{g_Z(\lambda)}\right)\left(\frac{\hat{f}_Z(\lambda)}{\hat{g}^T(\lambda)} - \frac{f_Z(\lambda)}{g_Z(\lambda)}\right).$$

Then we have

$$\begin{aligned}
&\int_{[-\pi,\pi]} H\left(\frac{\hat{f}_Z(\lambda)}{\hat{g}_Z^T(\lambda)}\right)\mathrm{d}\lambda - \int_{[-\pi,\pi]} H^{(1)}\left(\frac{f_Z(\lambda)}{g_Z(\lambda)}\right)\mathrm{d}\lambda \\
&= \int_{[-\pi,\pi]} \left(\frac{\hat{f}_Z(\lambda)}{\hat{g}_Z^T(\lambda)} - \frac{f_Z(\lambda)}{g_Z(\lambda)}\right) H^{(1)}(f_Z(\lambda),\lambda)\mathrm{d}\lambda + O_p\left(\frac{M}{n}\right) \\
&\leq C \max_{\lambda\in[-\pi,\pi]} \left(\frac{\hat{f}_Z(\lambda)}{\hat{g}_Z^T(\lambda)} - \frac{f_Z(\lambda)}{g_Z(\lambda)}\right) + O_p\left(\frac{M}{n}\right) \\
&= o_p(1),
\end{aligned}$$

which leads to the result. □

Proof of Theorem 3.4.2
The proof is omitted. □

Proof of Theorem 3.4.3
The proof is omitted. □

7.3 Proofs in Chapter 4

In this section, we provide the proofs of Theorems 4.1.1 and 4.2.1 and Proposition 4.1.1.

Proof of Theorem 4.1.1. First, we show the binary estimator is centered. For each $j \in \{1,\ldots,p\}$,

$$\begin{aligned}
\mathrm{E}\{\sqrt{n}(\hat{\eta}_j - \eta_j)\} &= \sqrt{n}2\pi(b^{j1},\ldots,b^{jp})\begin{pmatrix} \frac{1}{n}\sum_{k=1}^n \mathrm{E}X_k{}^1 - \mathrm{P}(-\alpha_1 \leq \Theta_1 \leq \alpha_1) \\ \vdots \\ \frac{1}{n}\sum_{k=1}^n \mathrm{E}X_k{}^p - \mathrm{P}(-\alpha_p \leq \Theta_1 \leq \alpha_p) \end{pmatrix} \\
&= 0.
\end{aligned}$$

Next, we evaluate the variance of estimator. For $i, j \in \{1, \ldots, p\}$,

$$\begin{aligned}
&\mathrm{cum}\{\sqrt{n}(\hat{\eta}_i - \eta_i), \sqrt{n}(\hat{\eta}_j - \eta_j)\}\\
=&\frac{4\pi^2}{n} \sum_{s,k=1}^{p} b^{is} b^{jk} \sum_{v=1}^{n} \mathrm{cum}\{X_v{}^s, X_v{}^k\}\\
=&4\pi^2 \sum_{s,k=1}^{p} b^{is} b^{jk} \mathrm{cum}\{X_1^s, X_1^k\}.
\end{aligned}$$

Finally, we elucidate the L-th order cumulant ($L \geq 3$) of the binary estimator is of order $O(n^{-L/2+1})$. For $i_1, \ldots, i_L \in \{1, \ldots, p\}$,

$$\begin{aligned}
&\mathrm{cum}\{\sqrt{n}(\hat{\eta}_{i_1} - \eta_{i_1}), \ldots, \sqrt{n}(\hat{\eta}_{i_L} - \eta_{i_L})\}\\
=&n^{L/2}(2\pi)^L \sum_{s_1,\ldots,s_L=1}^{p} b^{i_1 s_1} \cdots b^{i_L s_l} \mathrm{cum}\{\frac{1}{n}\sum_{k=1}^{n} X_k{}^{s_1}, \ldots, \frac{1}{n}\sum_{k=1}^{n} X_k{}^{s_L}\}\\
=&n^{-L/2+1}(2\pi)^L \sum_{s_1,\ldots,s_L=1}^{p} b^{i_1 s_1} \cdots b^{i_L s_l} \mathrm{cum}\{X_1^{s_1}, \ldots, X_1^{s_L}\}\\
=&O(n^{-L/2+1}),
\end{aligned}$$

thus, we have the desired result. □

Proof of Proposition 4.1.1. It is sufficient to show the Fisher information $\mathcal{I}$, defined by

$$\mathcal{I}(\eta_1) = \int_{-\pi}^{\pi} \left(\frac{\partial}{\partial \eta_1} \log p_{\mathrm{circ}}(\theta) \right)^2 p_{\mathrm{circ}}(\theta) \mathrm{d}\theta,$$

becomes the following

$$\mathcal{I}(\eta_1) = \begin{cases} \frac{1}{2} & (\eta_1 = 0), \\ \frac{1}{\eta_1^2}\left(\frac{1}{\sqrt{1-\eta_1^2}} - 1\right) & (0 < |\eta_1| < 1), \\ \infty & (\eta_1 = \pm 1). \end{cases}$$

First, for $\eta_1 = 0$, by a straightforward calculation. Second, the residue theorem yields the assertion when η_1 satisfies $0 < |\eta_1| < 1$. Third, for $\eta_1 = \pm 1$, it is easy to see the integral diverges. □

Proof of Theorem 4.2.1. For any $a_j (< \pi - \beta)$, $j = 1, \ldots, p$, we have

$$\int_{-a_j}^{a_j} q_{\mathrm{circ,\ contam}}(\theta) \mathrm{d}\theta = \int_{-a_j}^{a_j} p_{\mathrm{circ}}(\theta) \mathrm{d}\theta,$$

from which the statement follows. □

7.4 Proofs in Chapter 5

Proof of Lemma 5.1.1
It is easy to see that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{L}_n^E(\boldsymbol{\theta})-\frac{\partial}{\partial\boldsymbol{\theta}}L_n^E(\boldsymbol{\theta})\right\|=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\frac{2}{\delta^2}Z_ta_tc_t+\frac{1}{\delta^2}Z_tb_t+\frac{1}{\delta}(b_t+a_tc_t)\right\},$$

which, from (E10) and Resnick (1999, p.198 exercise 6.16), converges to 0 a.s.. Since $\frac{\partial}{\partial\boldsymbol{\theta}}\ell_n^E(\boldsymbol{\theta})$ is the strictly stationary, ergodic, and martingale difference sequence, it holds that

$$\frac{\partial}{\partial\boldsymbol{\theta}}L_n^E(\boldsymbol{\theta})\Rightarrow N(0,I^E)$$

by the martingale central limit theorem Billinsley (1999, Theorem 18.1) and the Cramer–Wold devise. We can show $-\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{L}_n^E(\boldsymbol{\theta}_n^{E*})\to J$ a.s. in the same way as Ahmad and Francq (2016, Theorem 2.2). If

$$\mathbf{s}^TJ^E\mathbf{s}=\mathrm{E}\left(\frac{1}{\lambda_t^2(\boldsymbol{\theta}_0)}\left(\mathbf{s}^T\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_0)\right)\left(\mathbf{s}^T\frac{\partial}{\partial\boldsymbol{\theta}^T}\lambda_t(\boldsymbol{\theta}_0)\right)^T\right)=0$$

Then, $\mathbf{s}^T\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_0)=\mathbf{0}$ a.s. and, by (A9), $\mathbf{s}=\mathbf{0}$. Hence, J^E is non-singular matrix. For large n such that $-\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{L}_n^j(\boldsymbol{\theta}_n^{j*})$ is non-singular, we have

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\theta}}_n^E-\boldsymbol{\theta}_0)&=\left(\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{L}_n^E(\boldsymbol{\theta}_n^{E*})\right)^{-1}\sqrt{n}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{L}_n^E(\boldsymbol{\theta}_0)\\&\Rightarrow N(\mathbf{0},(J^E)^{-1}I^E(J^E)^{-1})\quad\text{as } n\to\infty.\end{aligned}$$

□

The essential tool to prove Theorem 5.2.1 is the multi-dimensional martingale functional central limit theorem (FCLT). Billinsley (1999, Theorem 18.2) shows the one-dimensional martingale FCLT on the Skorokhod space. However, the extension of the theorem to multi-dimension is not obvious. We define the functional space of càdlàg function on $[0,\infty)$ to $\mathbb{R}^d$. Let $D([0,\infty),\mathbb{R}^d):=\{x(t):[0,\infty)\to\mathbb{R}^d;\text{ right continuous with left limits everywhere}\}$ and, for any $x,y\in D([0,\infty),\mathbb{R}^d)$, $d_\infty(x,y):=\inf_{\lambda\in\Lambda}\max\left(\gamma(\lambda),\int_0^\infty e^{-u}d(x,y,\lambda,u)\mathrm{d}u\right)$ where Λ be the set of strictly increasing and Lipschitz continuous functions on $[0,\infty)$ to $[0,\infty)$ such that $\lambda(0)=0$, $\lim_{t\to\infty}\lambda(t)=\infty$,

$$\gamma(\lambda):=\sup_{0\le t<s}\log\left|\frac{\lambda(s)-\lambda(t)}{s-t}\right|<\infty,$$

$d(x,y,\lambda,u) := \sup_{0\le t} \min\{\|x(\min\{t,u\}) - y(\min\{\lambda(t),u\}))\|, 1\}$, $\|\cdot\|$ denotes ℓ_2 norm. Then $d_\infty(\cdot,\cdot)$ is norm (Ethier and Kurtz, 1986, p.118) and $(D([0,\infty),\mathbb{R}^d), d_\infty)$ is separable and complete (Ethier and Kurtz, 1986, Theorem 3.5.6).

The following multi-dimensional martingale FCLT is due to Ethier and Kurtz (1986, Theorem 3.5.6) and Whitt (2007, Theorem 2.1 and Section 5)

Lemma 7.4.1. *Let $\{\mathbf{m}_t \in \mathbb{R}^d : t \in \mathbb{Z}\}$ be a martingale difference sequence, i.e. $\mathrm{E}(\mathbf{m}_t|\mathcal{M}_{t-1}) = \mathbf{0}$ where $\mathcal{M}_{t-1}$ is the σ-field generated by $\{\mathbf{m}_s, s \le t-1\}$ and $\mathbf{M}_n(s) := \frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor ns\rfloor}\mathbf{m}_j$. If*

$$\lim_{n\to\infty}\frac{1}{\sqrt{n}}\mathrm{E}\left(\max_{1\le s\le n}\|\mathbf{m}_s\|\right) = 0,$$

and

$$\frac{1}{n}\sum_{j=1}^{n}\mathbf{m}_j\mathbf{m}_j^T \to \mathbf{\Sigma} \quad \text{in probability as } n\to\infty,$$

where $\mathbf{\Sigma}$ is positive definite. Then, it holds

$$\mathbf{\Sigma}^{-1/2}\mathbf{M}_n(s) \Rightarrow \mathbf{B}_d(s) \quad \text{in } (D([0,\infty),\mathbb{R}^d), d_\infty) \text{ as } n\to\infty,$$

where B_d is a d-dimensionalstandard Brownian motion (see Ethier and Kurtz (1986, p.276)).

The multi-dimensional martingale FCLT holds that for any ergodic, stationary, and martingale difference process $\{\boldsymbol{\xi}_t \in \mathbb{R}^d : t \in \mathbb{Z}\}$ with positive definite covariance matrix and $\mathrm{E}\|\boldsymbol{\xi}_1\|^{2+\delta} < \infty$ for some $\delta > 0$. Actually, the conditions of Lemma 7.4.1 can be checked easily; for any sequence $C_n > 0$ such that $C_n/n \to 0$ and $C_n \to \infty$ as $n \to \infty$,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\mathrm{E}\left(\max_{1\le j\le n}\|\boldsymbol{\xi}_j\|\right)\\
=&\sqrt{\frac{1}{n}\mathrm{E}\left(\max_{1\le j\le n}\|\boldsymbol{\xi}_j\|^2\mathbb{I}_{\{\|\boldsymbol{\xi}_j\|>C_n\}}\right)} + \sqrt{\frac{1}{n}\mathrm{E}\left(\max_{1\le j\le n}\|\boldsymbol{\xi}_j\|^2\mathbb{I}_{\{\|\boldsymbol{\xi}_j\|\le C_n\}}\right)}\\
\le&\sqrt{\frac{1}{n}\sum_{j=1}^{n}\mathrm{E}\left(\|\boldsymbol{\xi}_j\|^2\mathbb{I}_{\{\|\boldsymbol{\xi}_j\|>C_n\}}\right)} + \sqrt{\frac{C_n}{n}}\\
\le&\sqrt{\mathrm{E}\left(\|\boldsymbol{\xi}_j\|^2\mathbb{I}_{\{\|\boldsymbol{\xi}_j\|>C_n\}}\right)} + \sqrt{\frac{C_n}{n}}\\
\le&\frac{1}{C_n^\delta}\sqrt{\mathrm{E}\left(\|\boldsymbol{\xi}_j\|\right)^{2+\delta}} + \sqrt{\frac{C_n}{n}}
\end{aligned}$$

$$\to 0 \quad \text{as } n \to \infty.$$

The second condition follows from the ergodic theorem. Thus, it holds that

$$\left(\mathrm{E}\boldsymbol{\xi}_1\boldsymbol{\xi}_1^T\right)^{-1/2}\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor ns\rfloor}\boldsymbol{\xi}_j \Rightarrow \mathbf{B}_d(s) \quad \text{in } (D([0,\infty),\mathbb{R}^d), d_\infty) \text{ as } n \to \infty,$$

and

$$\left(\mathrm{E}\boldsymbol{\xi}_1\boldsymbol{\xi}_1^T\right)^{-1/2}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor ns\rfloor}\boldsymbol{\xi}_j - \frac{s}{\sqrt{n}}\sum_{j=1}^{n}\boldsymbol{\xi}_j\right)$$
$$\Rightarrow \mathbf{B}_d(s) - s\mathbf{B}_d(1) \quad \text{in } (D([0,\infty),\mathbb{R}^d), d_\infty) \text{ as } n \to \infty.$$

We use the following Lemma 7.4.2 to prove Thorem 5.2.1.

Lemma 7.4.2. *For $j = \mathrm{P}, \mathrm{NB}$, or E, we assume (A0)-(A6) and (j7)-(j8). Then, it holds that, under H_0,*

$$\max_{1\le k\le n}\sqrt{\frac{k}{n}}\left\|\Delta_k^j\right\| = o_p(1) \quad \text{as } n \to \infty.$$

Proof. We follow Kang and Lee (2014, Lemma 9)'s proof. By Assumption (j8), it can be shown that

$$-\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{L}_n^j(\boldsymbol{\theta}_n^{j^*}) \to J^j \quad \text{a.s. as } n \to \infty$$

in the same way as Ahmad and Francq (2016, Theorem 2.2). We can apply the Egorov's theorem, that is, for any $\epsilon > 0$, there exists some Borel set $A \in \mathcal{F}$ such that $\mathrm{P}(A) < \epsilon$ and

$$-\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{L}_n^j(\boldsymbol{\theta}_n^{j^*}) \to J^j \quad \text{uniformly on } \Omega\backslash A.$$

There exists N_1 such that, for any $n \ge N_1$,

$$\left|\det\left(J^j\right) - \det\left(-\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{L}_n^j(\boldsymbol{\theta}_n^{j^*})\right)\right| < \frac{1}{2}\det\left(J^j\right) \quad \text{on } \Omega\backslash A,$$

and then

$$\left|\det\left(-\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{L}_n^j(\boldsymbol{\theta}_n^{j^*})\right)\right| > \frac{1}{2}\det\left(J^j\right) \quad \text{on } \Omega\backslash A.$$

For any invertible matrix B_n and B such that $B_n \to B$ as $n \to \infty$, $\|B_n^{-1} - B^{-1}\| = \|A_n^{-1}(A_n - A)A^{-1}\| \le \|A_n^{-1}\|\|(A_n - A)\|\|A^{-1}\| \to 0$. Thus, there exists $N_2 \in \mathbb{N}$ such that, for any $n \ge N_2 \ge N_1$,

$$\left\| \left(-\frac{1}{n} \frac{\partial}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \tilde{L}_n^j(\boldsymbol{\theta}_n^{j*}) \right)^{-1} \right\| < \frac{3}{2} \left\| \left(J^j \right)^{-1} \right\| \quad \text{on } \Omega \backslash A.$$

For any $\epsilon > 0$,

$$\begin{aligned} &\mathrm{P}\left(\max_{1 \le k \le n} \sqrt{\frac{k}{n}} \left\| \Delta_k^j \right\| > \epsilon \right) \\ \le &\mathrm{P}\left(\max_{1 \le k \le N_2} \sqrt{\frac{k}{n}} \left\| \Delta_k^j \right\| > \epsilon, \Omega \backslash A \right) + \mathrm{P}\left(\max_{N_2+1 \le k \le n} \sqrt{\frac{k}{n}} \left\| \Delta_k^j \right\| > \epsilon, \Omega \backslash A \right) + \mathrm{P}(A) \end{aligned} \tag{7.4.1}$$

From the definition of the tightness in $\mathbb{R}$, there exists $\eta > 0$ such that

$$\mathrm{P}\left(\max_{1 \le k \le N_2} \sqrt{k} \left\| \Delta_k^j \right\| > \eta, \Omega \backslash A \right) < \epsilon.$$

The second term of (7.4.1) is asymptotically negligible along the line of Kang and Lee (2014, Lemma 9), which concludes the lemma. □

Proof of Theorem 5.2.1
We note that

$$\begin{aligned} &\sup_{0 \le s \le 1} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor ns \rfloor} \frac{\partial}{\partial \boldsymbol{\theta}} \tilde{\ell}_t^j(\boldsymbol{\theta}) - \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor ns \rfloor} \frac{\partial}{\partial \boldsymbol{\theta}} \ell_t^j(\boldsymbol{\theta}) \right\| \\ \le &\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \tilde{\ell}_t^j(\boldsymbol{\theta}) - \frac{\partial}{\partial \boldsymbol{\theta}} \ell_t^j(\boldsymbol{\theta}) \right\| \to 0 \quad \text{a.s. as } n \to \infty \end{aligned}$$

and

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}} \ell_t^P(\boldsymbol{\theta}_0) &= \left(\frac{Z_t}{\lambda_t} - 1 \right) \frac{\partial}{\partial \boldsymbol{\theta}} \lambda_t(\boldsymbol{\theta}_0), \\ \frac{\partial}{\partial \boldsymbol{\theta}} \ell_t^{NB}(\boldsymbol{\theta}_0) &= \frac{r(Z_t - \lambda_t)}{\lambda_t(\boldsymbol{\theta}_0)(r + \lambda_t(\boldsymbol{\theta}_0))} \frac{\partial}{\partial \boldsymbol{\theta}} \lambda_t(\boldsymbol{\theta}_0), \text{ and} \\ \frac{\partial}{\partial \boldsymbol{\theta}} \ell_t^E(\boldsymbol{\theta}_0) &= \frac{Z_t - \lambda_t}{\lambda_t^2} \frac{\partial}{\partial \boldsymbol{\theta}} \lambda_t(\boldsymbol{\theta}_0), \end{aligned}$$

which are strictly stationary, ergodic, and martingale difference sequences. The multi-dimensional martingale FCLT yields that

$$
\begin{aligned}
&(\hat{I}^j_{\mathrm{KL}})^{-1/2}\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor ns\rfloor}\frac{\partial}{\partial\boldsymbol{\theta}}\ell^j_t(\boldsymbol{\theta}_0)-(\hat{I}^j_{\mathrm{KL}})^{-1/2}\frac{\lfloor ns\rfloor}{n}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\ell^j_t(\boldsymbol{\theta}_0)\\
&=\left(I^j_{\mathrm{KL}}\right)^{-1/2}\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor ns\rfloor}\frac{\partial}{\partial\boldsymbol{\theta}}\ell^j_t(\boldsymbol{\theta}_0)-\left(I^j_{\mathrm{KL}}\right)^{-1/2}\frac{\lfloor ns\rfloor}{n}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\ell^j_t(\boldsymbol{\theta}_0)+o_p(1)\\
&\Rightarrow\mathbf{B}_d(s)-s\mathbf{B}_d(1)\quad\text{in }(D([0,\infty),\mathbb{R}^d),d_\infty)\text{ as }n\to\infty.
\end{aligned}
$$

Since Brownian motion has sample paths in $C([0,\infty),\mathbb{R}^d)$, we can apply the continuous mapping theorem (Ethier and Kurtz, 1986, 3.10.2). Hence, from Lemmas 7.4.1 and 7.4.2, the continuous mapping theorem, we obtain

$$
\begin{aligned}
T^j_{\mathrm{KL,Wald}}&=\sup_{0\le s\le 1}\left\|\frac{\lfloor ns\rfloor}{\sqrt{n}}(\hat{I}^j_{\mathrm{KL}})^{-1/2}\hat{J}^j_{\mathrm{KL}}(\hat{\boldsymbol{\theta}}^j_{\lfloor ns\rfloor}-\hat{\boldsymbol{\theta}}^j_n)\right\|^2\\
&=\sup_{0\le s\le 1}\left\|(I^j_{\mathrm{KL}})^{-1/2}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor ns\rfloor}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}^j_t(\boldsymbol{\theta}_0)-\frac{\lfloor ns\rfloor}{n}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}^j_t(\boldsymbol{\theta}_0)\right)\right\|^2+o_p(1)\\
&\Rightarrow\sup_{0\le s\le 1}\|B^\circ_d(s)\|^2\quad\text{as }n\to\infty,
\end{aligned}
$$

which completes Theorem 5.2.1. □

Proof of Theorem 5.2.2
We note that

$$
\begin{aligned}
&\frac{k(n-k)}{n\sqrt{n}}J^j(\hat{\boldsymbol{\theta}}^j_{1:k}-\hat{\boldsymbol{\theta}}^j_{k+1:n})\\
&=\frac{1}{\sqrt{n}}\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}^j_t(\boldsymbol{\theta}_0)-\frac{k}{n}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}^j_n(\boldsymbol{\theta}_0)+\frac{n-k}{n}\sqrt{\frac{k}{n}}\Delta^j_{1:k}-\frac{k}{n}\sqrt{\frac{n-k}{n}}\Delta^j_{k+1:n},
\end{aligned}
$$

where $\Delta^j_{a:b}$ is defined in a same way as (5.2.1) based on $\{Z_a,\dots,Z_b\}$. Then, we can show the asymptotic distribution of $T^j_{\mathrm{DK,Wald}}$ in a same way as Theorem 5.2.1. □
Proof of Theorem 5.2.3
It is easy to see that, under the alternative H_1,

$$
\begin{aligned}
&T^j_{\mathrm{DK,Wald}}\\
&\ge\frac{\lfloor n\tau\rfloor^2(n-\lfloor n\tau\rfloor)^2}{n^3}(\hat{\boldsymbol{\theta}}^j_{1:\lfloor n\tau\rfloor}-\hat{\boldsymbol{\theta}}^j_{\lfloor n\tau\rfloor+1:n})^T\hat{J}^j_{\mathrm{DK}}(\hat{I}^j_{\mathrm{DK}})^{-1}\hat{J}^j_{\mathrm{DK}}(\hat{\boldsymbol{\theta}}^j_{1:\lfloor n\tau\rfloor}-\hat{\boldsymbol{\theta}}^j_{\lfloor n\tau\rfloor+1:n})
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{n\tau^2(1-\tau)^2}{4}(\boldsymbol{\theta}_0-\boldsymbol{\theta}_1)^T J^j \left(I^j\right)^{-1} J^j(\boldsymbol{\theta}_0-\boldsymbol{\theta}_1)+o_p(1)\\
&\to\infty \quad \text{as } n\to\infty,
\end{aligned}
$$

which competes the proof. □

The next Lemma is needed to show Thorem 5.2.4.

Lemma 7.4.3. *For* $j = \mathrm{P}, \mathrm{NB}$, *or* E, *suppose that (A0)-(A6) and* $(j7)$-$(j8)$*. Then, it holds that, under* H_0,

$$
\begin{aligned}
&\max_{1\le k\le n}\frac{1}{\sqrt{n}}\left\|\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}_t^j(\hat{\boldsymbol{\theta}}_n^j)-\left(\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}_t^j(\boldsymbol{\theta}_0)-\frac{k}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}_t^j(\boldsymbol{\theta}_0)\right)\right\|\\
&=o_p(1) \quad \text{as } n\to\infty.
\end{aligned}
$$

Proof. By the Taylor's expansion, there exists $\boldsymbol{\theta}_n^{j^*}$ and $\boldsymbol{\theta}_n^{j^{**}}$ such that $\boldsymbol{\theta}_0 \leqslant \boldsymbol{\theta}_n^{j^*} \leqslant \hat{\boldsymbol{\theta}}_n^j$, $\boldsymbol{\theta}_0 \leqslant \boldsymbol{\theta}_n^{j^{**}} \leqslant \hat{\boldsymbol{\theta}}_n^j$,

$$
\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}_t^j(\hat{\boldsymbol{\theta}}_n^j)-\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}_t^j(\boldsymbol{\theta}_0)=\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{\ell}_t^j(\boldsymbol{\theta}_n^{j^*})\left(\hat{\boldsymbol{\theta}}_n^j-\boldsymbol{\theta}_0\right),
$$

and

$$
\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}_t^j(\hat{\boldsymbol{\theta}}_n^j)-\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}_t^j(\boldsymbol{\theta}_0)=\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{\ell}_t^j(\boldsymbol{\theta}_n^{j^{**}})\left(\hat{\boldsymbol{\theta}}_n^j-\boldsymbol{\theta}_0\right).
$$

Then, we have

$$
\begin{aligned}
&\max_{1\le k\le n}\frac{1}{\sqrt{n}}\left\|\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}_t^j(\hat{\boldsymbol{\theta}}_n^j)-\left(\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}_t^j(\boldsymbol{\theta}_0)-\frac{k}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}_t^j(\boldsymbol{\theta}_0)\right)\right\|\\
&=\max_{1\le k\le n}\frac{1}{n}\left\|\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{\ell}_t^j(\boldsymbol{\theta}_n^{j^*})\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n^j-\boldsymbol{\theta}_0\right)-\frac{k}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{\ell}_t^j(\boldsymbol{\theta}_n^{j^{**}})\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n^j-\boldsymbol{\theta}_0\right)\right\|\\
&\le\left\|\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n^j-\boldsymbol{\theta}_0\right)\right\|\max_{1\le k\le n}\frac{k}{n}\left\|\frac{1}{k}\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{\ell}_t^j(\boldsymbol{\theta}_n^{j^*})+J^j\right\|\\
&\quad+\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{\ell}_t^j(\boldsymbol{\theta}_n^{j^{**}})-J^j\right\|\left\|\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n^j-\boldsymbol{\theta}_0\right)\right\|.
\end{aligned}
$$

Since $\left\|\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n^j-\boldsymbol{\theta}_0\right)\right\|=O_p(1)$ and we can show that

$$
\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{\ell}_t^j(\boldsymbol{\theta}_n^{j^{**}})-J^j\right\|\to 0 \quad \text{a.s. as } n\to\infty \tag{7.4.2}
$$

in a same way as Ahmad and Francq (2016, Theorem 2.2), thus, the second term is asymptotically negligible. From (7.4.2), for any $\epsilon > 0$, there exists $N_4 > 0$ such that, for $k \geq N_4$,

$$\mathrm{P}\left(\left\|\frac{1}{k}\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{\ell}_t^j(\boldsymbol{\theta}_n^{j*}) + J^j\right\| > \epsilon\right) = 0.$$

Hence, it holds that

$$\begin{aligned}
&\max_{1\leq k\leq n}\frac{k}{n}\left\|\frac{1}{k}\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{\ell}_t^j(\boldsymbol{\theta}_n^{j*}) + J^j\right\| \\
&= \max_{1\leq k\leq N_4}\frac{k}{n}\left\|\frac{1}{k}\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{\ell}_t^j(\boldsymbol{\theta}_n^{j*}) + J^j\right\| + \max_{N_4\leq k\leq n}\frac{k}{n}\left\|\frac{1}{k}\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{\ell}_t^j(\boldsymbol{\theta}_n^{j*}) + J^j\right\| \\
&\leq \frac{N_4}{n}\sum_{t=1}^{N_4}\left\|\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{\ell}_t^j(\boldsymbol{\theta}_n^{j*})\right\| + \frac{N_4}{n}\left\|J^j\right\| + \max_{N_4\leq k\leq n}\left\|\frac{1}{k}\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{\ell}_t^j(\boldsymbol{\theta}_n^{j*}) + J^j\right\|,
\end{aligned}$$

which tends to 0 in probability since for any ϵ', there exists M such that

$$\mathrm{P}\left(\sum_{t=1}^{N_4}\left\|\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\tilde{\ell}_t^j(\boldsymbol{\theta}_n^{j*})\right\| > M\right) \leq \epsilon'.$$

□

Proof of Theorem 5.2.4
By employing Lemma 7.4.3, the ergodic theorem, and the multi-dimensional martingale FCLT, we can see that

$$\begin{aligned}
&(\hat{I}_{\mathrm{KL}}^j)^{-1/2}\frac{1}{\sqrt{n}}\left(\sum_{t=1}^{\lfloor ns\rfloor}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}_t^j(\hat{\boldsymbol{\theta}}_n^j)\right) \\
&=(\hat{I}_{\mathrm{KL}}^j)^{-1/2}\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor ns\rfloor}\frac{\partial}{\partial\boldsymbol{\theta}}\ell_t^j(\boldsymbol{\theta}_0) - (\hat{I}_{\mathrm{KL}}^j)^{-1/2}\frac{\lfloor ns\rfloor}{n}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\ell_t^j(\boldsymbol{\theta}_0) + o_p() \\
&= \left(I_{\mathrm{KL}}^j\right)^{-1/2}\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor ns\rfloor}\frac{\partial}{\partial\boldsymbol{\theta}}\ell_t^j(\boldsymbol{\theta}_0) - \left(I_{\mathrm{KL}}^j\right)^{-1/2}\frac{\lfloor ns\rfloor}{n}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\ell_t^j(\boldsymbol{\theta}_0) + o_p(1) \\
&\Rightarrow \mathbf{B}_d(s) - s\mathbf{B}_d(1) \quad \text{in } (D([0,\infty),\mathbb{R}^d), d_\infty) \text{ as } n\to\infty.
\end{aligned}$$

From the fact

$$\left|\sup_{0\leq s\leq 1}\left\|(\hat{I}_{\mathrm{KL}}^j)^{-1/2}\frac{1}{\sqrt{n}}\left(\sum_{t=1}^{\lfloor ns\rfloor}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}_t^j(\hat{\boldsymbol{\theta}}_n^j)\right)\right\| - \sup_{0\leq s\leq 1}\left\|(\hat{I}_{\mathrm{KL}}^j)^{-1/2}\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor ns\rfloor}\frac{\partial}{\partial\boldsymbol{\theta}}\ell_t^j(\hat{\boldsymbol{\theta}}_n^j)\right\|\right|$$

$$
\begin{aligned}
&\le \left\|(\hat{I}^j_{\mathrm{KL}})^{-1/2}\right\| \sup_{0\le s\le 1}\left\|\frac{1}{\sqrt{n}}\left(\sum_{t=1}^{\lfloor ns\rfloor}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}^j_t(\hat{\boldsymbol{\theta}}^j_n)\right)-\frac{1}{\sqrt{n}}\left(\sum_{t=1}^{\lfloor ns\rfloor}\frac{\partial}{\partial\boldsymbol{\theta}}\ell^j_t(\hat{\boldsymbol{\theta}}^j_n)\right)\right\| \\
&\le \left\|(\hat{I}^j_{\mathrm{KL}})^{-1/2}\right\| \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\|\left(\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}^j_t(\hat{\boldsymbol{\theta}}^j_n)\right)-\left(\frac{\partial}{\partial\boldsymbol{\theta}}\ell^j_t(\hat{\boldsymbol{\theta}}^j_n)\right)\right\| \to 0 \quad \text{a.s. } n\to\infty
\end{aligned}
$$

and the continuous mapping theorem (Ethier and Kurtz, 1986, 3.10.2), it holds that

$$
\sup_{0\le s\le 1}\left\|(\hat{I}^j_{\mathrm{KL}})^{-1/2}\frac{1}{\sqrt{n}}\left(\sum_{t=1}^{\lfloor ns\rfloor}\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{\ell}^j_t(\hat{\boldsymbol{\theta}}^j_n)\right)\right\|^2 \Rightarrow \sup_{0\le s\le 1}\left\|\mathbf{B}^\circ_d(s)\right\|^2 \quad \text{as } n\to\infty,
$$

which shows the result. □

Lemma 7.4.4. *If we assume that, for* $j=\mathrm{P},\mathrm{NB}$, *or* E, *(A0)-(A6), (j7)-(j8), and (A9). Then, it holds that, under* H_0,

$$
\max_{1\le k\le n}\frac{1}{\sqrt{n}}\left|\sum_{t=1}^{k}(\tilde{\epsilon}_t(\hat{\boldsymbol{\theta}}^j_n)-\epsilon_t)-\frac{k}{n}\sum_{t=1}^{n}(\tilde{\epsilon}_t(\hat{\boldsymbol{\theta}}^j_n)-\epsilon_t)\right| = o_p(1) \quad \textit{as } n\to\infty..
$$

Proof. By Taylor's expansion, we have

$$
\begin{aligned}
&\max_{1\le k\le n}\frac{1}{\sqrt{n}}\left|\sum_{t=1}^{k}(\tilde{\epsilon}_t(\hat{\boldsymbol{\theta}}^j_n)-\epsilon_t)-\frac{k}{n}\sum_{t=1}^{n}(\tilde{\epsilon}_t(\hat{\boldsymbol{\theta}}^j_n)-\epsilon_t)\right| \\
\le&\frac{2}{\sqrt{n}}\sum_{t=1}^{n}a_t+\max_{1\le k\le n}\frac{1}{\sqrt{n}}\left|\sum_{t=1}^{k}(\lambda_t(\hat{\boldsymbol{\theta}}^j_n)-\lambda_t(\boldsymbol{\theta}_0))-\frac{k}{n}\sum_{t=1}^{n}(\lambda_t(\hat{\boldsymbol{\theta}}^j_n)-\lambda_t(\boldsymbol{\theta}_0))\right| \\
\le&\frac{2}{\sqrt{n}}\sum_{t=1}^{n}a_t+\max_{1\le k\le n}\frac{1}{\sqrt{n}}\left|\sum_{t=1}^{k}(\hat{\boldsymbol{\theta}}^j_n-\boldsymbol{\theta}_0)^T\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_0)-\frac{k}{n}\sum_{t=1}^{n}(\hat{\boldsymbol{\theta}}^j_n-\boldsymbol{\theta}_0)^T\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_0)\right| \\
&+\max_{1\le k\le n}\frac{1}{\sqrt{n}}\left|\sum_{t=1}^{k}(\hat{\boldsymbol{\theta}}^j_n-\boldsymbol{\theta}_0)^T\left(\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}^{j*}_n)-\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_0)\right)\right. \\
&\left.-\frac{k}{n}\sum_{t=1}^{n}(\hat{\boldsymbol{\theta}}^j_n-\boldsymbol{\theta}_0)^T\left(\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}^{j*}_n)-\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_0)\right)\right| \\
\le&\frac{2}{\sqrt{n}}\sum_{t=1}^{n}a_t+\sqrt{n}\|\hat{\boldsymbol{\theta}}^j_n-\boldsymbol{\theta}_0\|\max_{1\le k\le n}\frac{k}{n}\left\|\frac{1}{k}\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_0)-\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_0)\right\| \\
&+\sqrt{n}\|\hat{\boldsymbol{\theta}}^j_n-\boldsymbol{\theta}_0\|\frac{2}{n}\sum_{t=1}^{n}\left\|\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}^{j*}_n)-\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_0)\right\|
\end{aligned}
$$

where $\boldsymbol{\theta}_0 \leqslant \boldsymbol{\theta}_n^{j*} \leqslant \hat{\boldsymbol{\theta}}_n^j$. Under the assumption, we can show that

$$\frac{2}{n}\sum_{t=1}^{n}\left\|\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_n^{j*}) - \frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_0)\right\| \longrightarrow 0 \quad \text{a.s. as } n \longrightarrow \infty$$

along the line of Ahmad and Francq (2016, Theorem 2.2), and, by the ergodic theorem, it is easy to see that

$$\max_{1\le k\le n}\frac{k}{n}\left\|\frac{1}{k}\sum_{t=1}^{k}\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_0) - \frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_0)\right\| \longrightarrow 0 \quad \text{a.s. as } n \longrightarrow \infty$$

in the same way as Lemma 7.4.2. Considering to $\sqrt{n}\|\hat{\boldsymbol{\theta}}_n^j - \boldsymbol{\theta}_0\| = O_p(1)$, the proof of Lemma 7.4.4 is complete. □

Lemma 7.4.5. *If we assume that, for* j = P, NB, *or* E, *(A0)-(A6), (j7)-(j8), and (A9), it holds that, under* H_0,

$$\frac{1}{n}\sum_{t=1}^{n}\tilde{\epsilon}_t^2(\hat{\boldsymbol{\theta}}_n^j) \longrightarrow \mathrm{E}\epsilon_t^2 \quad \textit{in probability as } n \longrightarrow \infty.$$

Proof. By Taylor's expansion, we obtain

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{t=1}^{n}(\hat{\epsilon}_t^2(\hat{\boldsymbol{\theta}}_n^j) - \epsilon_t^2)\right| \\
=&\left|\frac{1}{n}\sum_{t=1}^{n}(\hat{\epsilon}_t(\hat{\boldsymbol{\theta}}_n^j) - \epsilon_t)(\hat{\epsilon}_t(\hat{\boldsymbol{\theta}}_n^j) + \epsilon_t)\right| \\
\le&\frac{1}{n}\sum_{t=1}^{n}\left(a_t + \left|(\hat{\boldsymbol{\theta}}_n^j - \boldsymbol{\theta}_0)^T\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_n^*)\right|\right)\left(2|Z_t - \lambda_t(\boldsymbol{\theta}_0)| + a_t + \left|(\hat{\boldsymbol{\theta}}_n^j - \boldsymbol{\theta}_0)^T\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_n^*)\right|\right) \\
\le&\frac{1}{n}\sum_{t=1}^{n}\Bigg(a_t^2 + 2a_t|Z_t - \lambda_t(\boldsymbol{\theta}_0)| + 2a_t\left\|\hat{\boldsymbol{\theta}}_n^j - \boldsymbol{\theta}_0\right\|_{\ell_1}\left\|\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_n^*)\right\|_{\ell_1} \\
&+ 2|Z_t - \lambda_t(\boldsymbol{\theta}_0)|\left\|\hat{\boldsymbol{\theta}}_n^j - \boldsymbol{\theta}_0\right\|_{\ell_1}\left\|\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_n^*)\right\|_{\ell_1} + \left\|\hat{\boldsymbol{\theta}}_n^j - \boldsymbol{\theta}_0\right\|_{\ell_1}^2\left\|\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_n^*)\right\|_{\ell_1}^2\Bigg) \\
\le&\frac{1}{n}\sum_{t=1}^{n}a_t^2 + 2\sqrt{\frac{1}{n}\sum_{t=1}^{n}a_t^2}\sqrt{\frac{1}{n}\sum_{t=1}^{n}|Z_t - \lambda_t(\boldsymbol{\theta}_0)|^2} \\
&+ 2\left\|\hat{\boldsymbol{\theta}}_n^j - \boldsymbol{\theta}_0\right\|_{\ell_1}\sqrt{\frac{1}{n}\sum_{t=1}^{n}a_t^2}\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left\|\frac{\partial}{\partial\boldsymbol{\theta}}\lambda_t(\boldsymbol{\theta}_n^*)\right\|_{\ell_1}^2}
\end{aligned}$$

$$+ 2 \left\| \hat{\boldsymbol{\theta}}_n^j - \boldsymbol{\theta}_0 \right\|_{\ell_1} \sqrt{\frac{1}{n} \sum_{t=1}^{n} |Z_t - \lambda_t(\boldsymbol{\theta}_0)|^2} \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \lambda_t(\boldsymbol{\theta}_n^*) \right\|_{\ell_1}^2}$$
$$+ \left\| \hat{\boldsymbol{\theta}}_n^j - \boldsymbol{\theta}_0 \right\|_{\ell_1}^2 \frac{1}{n} \sum_{t=1}^{n} \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \lambda_t(\boldsymbol{\theta}_n^*) \right\|_{\ell_1}^2,$$

where $\boldsymbol{\theta}_0 \leqslant \boldsymbol{\theta}_n^{j*} \leqslant \hat{\boldsymbol{\theta}}_n^j$. By the ergodic theorem, we observe

$$\frac{1}{n} \sum_{t=1}^{n} |Z_t - \lambda_t(\boldsymbol{\theta}_0)|^2 \longrightarrow \mathrm{E}\nu_t(\boldsymbol{\theta}_0) \quad \text{a.s. as } n \to \infty \tag{7.4.3}$$

and, by Assumption (A9), we can show

$$\frac{1}{n} \sum_{t=1}^{n} \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \lambda_t(\boldsymbol{\theta}_n^*) \right\|_{\ell_1}^2 \longrightarrow \mathrm{E} \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \lambda_t(\boldsymbol{\theta}_0) \right\|_{\ell_1}^2 \quad \text{a.s. as } n \to \infty,$$

which shows that

$$\left| \frac{1}{n} \sum_{t=1}^{n} \hat{\epsilon}_t^2(\hat{\boldsymbol{\theta}}_n^j) - \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^2 \right| \longrightarrow 0 \quad \text{as } n \to \infty.$$

The conclusion of the Lemma 7.4.5 then follows from (7.4.3). □

Proof of Theorem 5.2.5
Since ϵ_t is the strictly stationary, ergodic, and martingale difference, we can apply the multi-dimensional martingale FCLT. From Lemmas 7.4.1, 7.4.4, and 7.4.5, we have

$$\begin{aligned}
&\max_{1 \le k \le n} \frac{1}{\sqrt{\frac{1}{n} \sum_{t=1}^{n} \hat{\epsilon}_t^2(\hat{\boldsymbol{\theta}}_n^j)}} \frac{1}{\sqrt{n}} \left| \sum_{t=1}^{k} \tilde{\epsilon}_t^j - \frac{k}{n} \sum_{t=1}^{n} \tilde{\epsilon}_t^j \right| \\
&= \sup_{0 \le s \le 1} \frac{1}{\sqrt{\mathrm{E}\nu_t(\boldsymbol{\theta}_0)}} \frac{1}{\sqrt{n}} \left| \sum_{t=1}^{\lfloor ns \rfloor} \epsilon_t - \frac{\lfloor ns \rfloor}{n} \sum_{t=1}^{n} \epsilon_t \right| + o_p(1) \\
&\Rightarrow \sup_{0 \le s \le 1} |B_1^{\circ}(s)| \quad \text{as } n \to \infty,
\end{aligned}$$

which proves the desired result. □

7.5 Proofs in Chapter 6

Proof of Theorem 6.1.1
Let $p_{\text{CG}}(\theta_0)$ be a probability density function of $N(\theta^\alpha, \theta^\beta)$. First observe that, for $X \sim N(\theta^\alpha, \theta^\beta)$,

$$\begin{aligned}\frac{\partial^2}{\partial\theta_0\partial\theta_0^T}\log p_{\text{CG}}(\theta_0) = &-\frac{\beta(\beta+1)}{2}\theta_0^{-\beta-2}X^2 + (\alpha-\beta)(\alpha-\beta-1)\theta_0^{\alpha-\beta-2}X \\ &-\frac{(2\alpha-\beta)(2\alpha-\beta-1)}{2}\theta_0^{2\alpha-\beta-2} + \frac{\beta}{2}\theta_0^{-2}\end{aligned}$$

and

$$\mathcal{I}(\theta_0) := -\mathrm{E}[\frac{\partial^2}{\partial\theta_0\partial\theta_0^T}\log p_{\text{CG}}(\theta_0)] = \frac{\beta^2}{2}\theta_0^{-2} + \alpha^2\theta_0^{2\alpha-\beta-2}.$$

A Simple algebra gives

$$\begin{aligned}\Lambda_0(\theta_0,\theta_n) = &\log\frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta_n^\beta}}\exp\{-\frac{(X_i-\theta_n^\alpha)^2}{2\theta_n^\beta}\}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta_0^\beta}}\exp\{-\frac{(X_i-\theta_0^\alpha)^2}{2\theta_0^\beta}\}} \\ =&h\begin{pmatrix}\frac{\sqrt{n}}{h}(\theta_n^{\alpha-\beta}-\theta_0^{\alpha-\beta}) \\ \frac{\sqrt{n}}{h}(\theta_n^{\alpha-\beta}-\theta_0^{\alpha-\beta})\end{pmatrix}^T\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{i=1}^n(X_i-\theta_0^\alpha) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n\{X_i^2-(\theta_0^{2\alpha}+\theta_0^\beta)\}\end{pmatrix} \\ &-n\Bigg\{\frac{\theta_0^{2\alpha}+\theta_0^\beta}{2}(\theta_n^{-\beta}-\theta_0^{-\beta}) - \theta_0^\alpha(\theta_n^{\alpha-\beta}-\theta_0^{\alpha-\beta}) \\ &+\frac{1}{2}(\theta_n^{2\alpha-\beta}-\theta_0^{2\alpha-\beta}) + \frac{\beta}{2}(\log\theta_n - \log\theta_0)\Bigg\}.\end{aligned}$$

By Taylor's expansion, we know that

$$\begin{aligned}\theta_n^{-\beta}-\theta_0^{-\beta} =& -\beta\theta_0^{-\beta-1}\frac{h}{\sqrt{n}} + \frac{1}{2}\beta(\beta+1)\theta_0^{-\beta-2}\frac{h^2}{n} + O(n^{-\frac{3}{2}}), \\ \theta_n^{\alpha-\beta}-\theta_0^{\alpha-\beta} =&(\alpha-\beta)\theta_0^{\alpha-\beta-1}\frac{h}{\sqrt{n}} + \frac{1}{2}(\alpha-\beta)(\alpha-\beta-1)\theta_0^{\alpha-\beta-2}\frac{h^2}{n} + O(n^{-\frac{3}{2}}), \\ \theta_n^{2\alpha-\beta}-\theta_0^{2\alpha-\beta} =&(2\alpha-\beta)\theta_0^{2\alpha-\beta-1}\frac{h}{\sqrt{n}} + \frac{1}{2}(2\alpha-\beta)(2\alpha-\beta-1)\theta_0^{2\alpha-\beta-2}\frac{h^2}{n} \\ &+ O(n^{-\frac{3}{2}}),\end{aligned}$$

$$\log\theta_n - \log\theta_0 = \theta_0^{-1}\frac{h}{\sqrt{n}} - \frac{1}{2}\theta_0^{-2}\frac{h^2}{n} + O(n^{-\frac{3}{2}}).$$

Then, it follows that

$$\begin{aligned} &-n\Big\{\frac{\theta_0^{2\alpha}+\theta_0^{\beta}}{2}(\theta_n^{-\beta}-\theta_0^{-\beta}) - \theta_0^{\alpha}(\theta_n^{\alpha-\beta}-\theta_0^{\alpha-\beta}) \\ &+\frac{1}{2}(\theta_n^{2\alpha-\beta}-\theta_0^{2\alpha-\beta}) + \frac{\beta}{2}(\log\theta_n - \log\theta_0)\Big\} \\ &\to -\frac{1}{2}h^2\mathcal{I}_0(\theta_0) \quad \text{as } n\to\infty, \end{aligned}$$

and

$$\begin{pmatrix}\frac{\sqrt{n}}{h}(\theta_n^{\alpha-\beta}-\theta_0^{\alpha-\beta}) & \frac{\sqrt{n}}{h}(\theta_n^{\alpha-\beta}-\theta_0^{\alpha-\beta})\end{pmatrix} \to \begin{pmatrix}(\alpha-\beta)\theta_0^{\alpha-\beta-1} & \frac{\beta}{2}\theta_0^{-\beta-1}\end{pmatrix} \quad \text{as } n\to\infty.$$

The central limit theorem gives that

$$\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{i=1}^n(X_i-\theta_0^{\alpha}) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n\{X_i^2-(\theta_0^{2\alpha}+\theta_0^{\beta})\}\end{pmatrix} \Rightarrow N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\theta_0^{\beta} & 2\theta_0^{\alpha+\beta} \\ 2\theta_0^{\alpha+\beta} & 2\theta_0^{2\beta}+4\theta_0^{2\alpha+\beta}\end{pmatrix}\right) \quad \text{as } n\to\infty.$$

Applying the Slutsky's lemma, we have the desired result. □

Proof of Theorem 6.1.2

First, we derive the Fisher information matrix. Let p_{SES} is a probability density function of the model (6.1.1). It is easy to see that

$$\begin{aligned} &\log p_{\mathrm{SES}}(\boldsymbol{\theta}_0) \\ =& -\frac{p+1}{2}\log 2\pi - \frac{1}{2}\log\det|\Omega| \\ &-\frac{1}{2}\begin{pmatrix}Y_{1i}-\boldsymbol{W}_{1i}(\boldsymbol{\zeta}_{12}\boldsymbol{\beta}_0+\boldsymbol{\gamma}_0)-\boldsymbol{W}_{2i}\boldsymbol{\zeta}_{22}\boldsymbol{\beta}_0 \\ (\boldsymbol{Y}_{2i}-\boldsymbol{W}_{1i}\boldsymbol{\zeta}_{12}-\boldsymbol{W}_{2i}\boldsymbol{\zeta}_{22})^T\end{pmatrix}^T \Omega^{-1} \\ &\times\begin{pmatrix}Y_{1i}-\boldsymbol{W}_{1i}(\boldsymbol{\zeta}_{12}\boldsymbol{\beta}_0+\boldsymbol{\gamma}_0)-\boldsymbol{W}_{2i}\boldsymbol{\zeta}_{22}\boldsymbol{\beta}_0 \\ (\boldsymbol{Y}_{2i}-\boldsymbol{W}_{1i}\boldsymbol{\zeta}_{12}-\boldsymbol{W}_{2i}\boldsymbol{\zeta}_{22})^T\end{pmatrix}, \end{aligned}$$

and

$$\begin{aligned} &-\frac{\partial^2}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0^T}\log p_{\mathrm{SES}}(\boldsymbol{\theta}_0) \\ &=\omega^{11}\begin{pmatrix}(\boldsymbol{Z}_{1i}\boldsymbol{\zeta}_{12}+\boldsymbol{Z}_{2i}\boldsymbol{\zeta}_{22})^T(\boldsymbol{Z}_{1i}\boldsymbol{\zeta}_{12}+\boldsymbol{Z}_{2i}\boldsymbol{\zeta}_{22}) & (\boldsymbol{Z}_{1i}\boldsymbol{\zeta}_{12}+\boldsymbol{Z}_{2i}\boldsymbol{\zeta}_{22})^T\boldsymbol{W}_{1i} \\ \boldsymbol{W}_{1i}^T(\boldsymbol{Z}_{1i}\boldsymbol{\zeta}_{12}+\boldsymbol{Z}_{2i}\boldsymbol{\zeta}_{22}) & \boldsymbol{W}_{1i}^T\boldsymbol{W}_{1i}\end{pmatrix}, \end{aligned}$$

which converges in probability to $\mathcal{I}_1(\boldsymbol{\theta}_0)$.

Second, we show the LAN property. Under the null hypothesis, we observe that

$$
\begin{aligned}
&\Lambda_1(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) \\
:=& -\frac{1}{2}\sum_{i=1}^{n}\begin{pmatrix} \nu_{1i} - \frac{1}{\sqrt{n}}(\boldsymbol{Z}_{1i}(\zeta_{12}\boldsymbol{h}_1 + \boldsymbol{h}_2) + \boldsymbol{Z}_{2i}\zeta_{22}\boldsymbol{h}_1) \\ \boldsymbol{\nu}_{2i}{}^T \end{pmatrix}^T \Omega^{-1} \\
&\times \begin{pmatrix} \nu_{1i} - \frac{1}{\sqrt{n}}(\boldsymbol{Z}_{1i}(\zeta_{12}\boldsymbol{h}_1 + \boldsymbol{h}_2) + \boldsymbol{Z}_{2i}\zeta_{22}\boldsymbol{h}_1) \\ \boldsymbol{\nu}_{2i}{}^T \end{pmatrix} + \frac{1}{2}\sum_{i=1}^{n}\begin{pmatrix} \nu_{1i} \\ \boldsymbol{\nu}_{2i}{}^T \end{pmatrix}^T \Omega^{-1}\begin{pmatrix} \nu_{1i} \\ \boldsymbol{\nu}_{2i}{}^T \end{pmatrix} \\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\boldsymbol{Z}_{1i}(\zeta_{12}\boldsymbol{h}_1 + \boldsymbol{h}_2) + \boldsymbol{Z}_{2i}\zeta_{22}\boldsymbol{h}_1\}^T \begin{pmatrix}\omega^{11} & \boldsymbol{\omega}^{12}\end{pmatrix}\begin{pmatrix} \nu_{1i} \\ \boldsymbol{\nu}_{2i}{}^T \end{pmatrix} \\
&- \frac{1}{2n}\sum_{i=1}^{n}\{\boldsymbol{Z}_{1i}(\zeta_{12}\boldsymbol{h}_1 + \boldsymbol{h}_2) + \boldsymbol{Z}_{2i}\zeta_{22}\boldsymbol{h}_1\}^T \omega^{11}\{\boldsymbol{Z}_{1i}(\zeta_{12}\boldsymbol{h}_1 + \boldsymbol{h}_2) + \boldsymbol{Z}_{2i}\zeta_{22}\boldsymbol{h}_1\} \\
=& \begin{pmatrix}\boldsymbol{h}_1{}^T & \boldsymbol{h}_2{}^T\end{pmatrix}\boldsymbol{\Delta}_n^1 - \frac{\omega^{11}}{2}\begin{pmatrix}\boldsymbol{h}_1{}^T & \boldsymbol{h}_2{}^T\end{pmatrix}\begin{pmatrix}\zeta_{12} & I_{p_2\times p_2} \\ \zeta_{22} & \boldsymbol{0}_{p_3\times p_2}\end{pmatrix}^T \\
&\times \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\boldsymbol{Z}_{1i}{}^T \\ \boldsymbol{Z}_{2i}{}^T\end{pmatrix}\begin{pmatrix}\boldsymbol{Z}_{1i} & \boldsymbol{Z}_{2i}\end{pmatrix}\begin{pmatrix}\zeta_{12} & I_{p_2\times p_2} \\ \zeta_{22} & \boldsymbol{0}_{p_3\times p_2}\end{pmatrix}\begin{pmatrix}\boldsymbol{h}_1 \\ \boldsymbol{h}_2\end{pmatrix}.
\end{aligned}
$$

By Assumption 6.1.1, it holds that

$$
(\omega^{11}\ \boldsymbol{\omega}^{12})\begin{pmatrix} Y_{1i} - \boldsymbol{Z}_{1i}(\zeta_{12}\boldsymbol{\beta}_0 + \boldsymbol{\gamma}_0) - \boldsymbol{Z}_{2i}\zeta_{22}\boldsymbol{\beta}_0 \\ (\boldsymbol{Y}_{2i} - \boldsymbol{Z}_{1i}\zeta_{12} - \boldsymbol{Z}_{2i}\zeta_{22})^T \end{pmatrix} \sim N(0, \omega^{11}),
$$

and this yields that $\boldsymbol{\Delta}_n^1 \Rightarrow N(\boldsymbol{0}, \mathcal{I}_1(\boldsymbol{\theta}_0))$ as $n \to \infty$. □

Proof of Theorem 6.2.1

Proof. (i) First observe that

$$
\begin{aligned}
&\frac{n(\theta_3^{(n)} - \theta_3)}{\theta_2^{(n)} + n\theta_1^{(n)}}\sum_i(\bar{y}_{i\cdot} - \theta_3) \\
&\Rightarrow \begin{cases} 0 & (k_3 \le \frac{1}{2},\ k_1 < \frac{1}{2} + k_3) \text{ or } (k_3 > \frac{1}{2}), \\ \frac{\sqrt{\theta_2}h_3}{h_1}g_1(\boldsymbol{T}_n) & k_3 < \frac{1}{2},\ k_1 = \frac{1}{2} + k_3, \\ \infty & k_3 < \frac{1}{2},\ k_1 > \frac{1}{2} + k_3, \\ \frac{h_3\sqrt{\theta_2}}{\theta_2 + h_1}g_1(\boldsymbol{T}_n) & k_3 = \frac{1}{2}, k_1 = 1, \\ \frac{h_3}{\sqrt{\theta_2}}g_1(\boldsymbol{T}_n) & k_3 = \frac{1}{2},\ k_1 > 1, \end{cases}
\end{aligned}
\tag{7.5.1}
$$

as $n \longrightarrow \infty$ and

$$
-\frac{an(\theta_3^{(n)}-\theta_3)^2}{2(\theta_2^{(n)}+n\theta_1^{(n)})} \Rightarrow \begin{cases} 0 & (k_3 \le \frac{1}{2},\ k_1 < 2k_3) \text{ or } (k_3 > \frac{1}{2}), \\ -\frac{ah_3^2}{2h_1} & k_3 < \frac{1}{2},\ k_1 = 2k_3, \\ -\infty & k_3 < \frac{1}{2},\ k_1 > 2k_3, \\ -\frac{ah_3^2}{2(\theta_2+h_1)} & k_3 = \frac{1}{2},\ k_1 = 1, \\ -\frac{ah_3^2}{2\theta_2} & k_3 = \frac{1}{2},\ k_1 > 1, \end{cases} \tag{7.5.2}
$$

as $n \longrightarrow \infty$.

From (7.5.1) and (7.5.2), it follows that

$$
\frac{n(\theta_3^{(n)}-\theta_3)}{\theta_2^{(n)}+n\theta_1^{(n)}} \sum_i (\bar{y}_{i\cdot}-\theta_3) - \frac{an(\theta_3^{(n)}-\theta_3)^2}{2(\theta_2^{(n)}+n\theta_1^{(n)})}
$$

$$
\Rightarrow \begin{cases} 0 & (k_3 \le \frac{1}{2},\ k_1 < 2k_3) \text{ or } (k_3 > \frac{1}{2}), \\ -\frac{ah_3^2}{2h_1} & k_3 < \frac{1}{2},\ k_1 = 2k_3, \\ -\infty & k_3 < \frac{1}{2},\ 2k_3 < k_1 \le k_3 + \frac{1}{2}, \\ \text{indeterminate form} & k_3 < \frac{1}{2},\ k_1 > k_3 + \frac{1}{2}, \\ \frac{h_3\sqrt{\theta_2}}{\theta_2+h_1} g_1(\boldsymbol{T}_n) - \frac{ah_3^2}{2(\theta_2+h_1)} & k_3 = \frac{1}{2},\ k_1 = 1, \\ \frac{h_3}{\sqrt{\theta_2}} g_1(\boldsymbol{T}_n) - \frac{ah_3^2}{2\theta_2} & k_3 = \frac{1}{2},\ k_1 > 1, \end{cases} \tag{7.5.3}
$$

as $n \longrightarrow \infty$.

Second, the Taylor's expansion yields that

$$
\begin{aligned}
& -\frac{a}{2}(n-1)\log\frac{\theta_2^{(n)}}{\theta_2} + \frac{\theta_2^{(n)}-\theta_2}{2\theta_2\theta_2^{(n)}} \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\cdot})^2 \\
= & -\frac{h_2 a}{2\theta_2}(1-\frac{1}{n})n^{1-k_2} + \frac{ah_2(1-\frac{1}{n})}{2(n^{k_2-1}\theta_2+\frac{h_2}{n})} \\
& + \frac{h_2\sqrt{2a(1-\frac{1}{n})}}{2(n^{k_2-\frac{1}{2}}\theta_2+\frac{h_2}{\sqrt{n}})} \left\{ \frac{\frac{1}{\theta_2}\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\cdot})^2 - a(n-1)}{\sqrt{2a(n-1)}} \right\} + O(n^{1-2k_2}) \\
\Rightarrow & \begin{cases} \text{indeterminate form} & 0 < k_2 < 1, \\ 0 & k_2 \ge 1, \end{cases}
\end{aligned} \tag{7.5.4}
$$

as $n \longrightarrow \infty$.

Finally, it holds that

$$
\begin{aligned}
&-\frac{a}{2}\log\frac{\theta_2^{(n)}+n\theta_1^{(n)}}{\theta_2}+\frac{n(\theta_2^{(n)}-\theta_2)+n^2\theta_1^{(n)}}{2\theta_2(\theta_2^{(n)}+n\theta_1^{(n)})}\sum_i(\bar{y}_{i\cdot}-\theta_3)^2 \\
&\Rightarrow \begin{cases} -\infty & 0<k_1<1, \\ \frac{h_1}{2(\theta_2+h_1)}g_2(\boldsymbol{T}_n)-\frac{a}{2}\log(1+\frac{h_1}{\theta_2}) & k_1=1, \\ 0 & k_1>1, \end{cases}
\end{aligned} \tag{7.5.5}
$$

as $n\longrightarrow\infty$. The conclusion of theorem then follows from (7.5.3), (7.5.4), and (7.5.5).

(ii) Let $p_{\mathrm{ANOVA}}(\boldsymbol{\theta})$ be a probability density function of (6.2.1). Since Y_{ij} follows the normal distribution with mean θ_3 and variance $\theta_1+\theta_2$, the loglikelihood function can obtained as

$$
\log p_{\mathrm{ANOVA}}(\boldsymbol{\theta})=-\frac{1}{2}\log 2\pi-\frac{1}{2}\log(\theta_1+\theta_2)-\frac{1}{2(\theta_1+\theta_2)}(y_{ij}-\theta_3)^2.
$$

A Simple algebra gives the desired result. □

Proof of Theorems 6.2.2 and 6.2.3

From the definition of log-likelihood function, we have

$$
\Lambda(\boldsymbol{\theta}_n,{}^1\boldsymbol{\theta})=-\frac{a}{2}\log\frac{\theta_2+n\theta_1^{(n)}}{\theta_2}+\frac{n^2\theta_1^{(n)}}{2\theta_2(\theta_2+n\theta_1^{(n)})}\sum_i(\bar{y}_{i\cdot}-\theta_3)^2.
$$

Under the null $\mathrm{H}_3^{(n)}$, it is easy to see that

$$
\Lambda(\boldsymbol{\theta}_n,{}^1\boldsymbol{\theta})=-\frac{a}{2}\log\frac{\theta_2+n\theta_1^{(n)}}{\theta_2}+\frac{n\theta_1^{(n)}}{2(\theta_2+n\theta_1^{(n)})}\left(\frac{n}{\theta_2}\sum_i(\bar{y}_{i\cdot}-\theta_3)^2\right),
$$

which shows the result of Theorem 6.2.2.

Similarly, we observe that, under the alternative $\mathrm{K}_3^{(n)}$, $n\sum_{i=1}^a(\bar{y}_{i\cdot}-\theta_3)^2/(\theta_2+n\theta_1^{(n)})$ follows χ_a^2 and

$$
\Lambda(\boldsymbol{\theta}_n,{}^1\boldsymbol{\theta})=-\frac{a}{2}\log\frac{\theta_2+n\theta_1^{(n)}}{\theta_2}+\frac{n\theta_1^{(n)}}{2\theta_2}\left(\frac{n}{(\theta_2+n\theta_1^{(n)})}\sum_i(\bar{y}_{i\cdot}-\theta_3)^2\right),
$$

which gives the result of Theorem 6.2.3. □

Proof of Theorem 6.2.4

We denote the asymptotic distribution of $\Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n)$ as $\mathfrak{L}_N$ for the null and $\mathfrak{L}_A$ for the alternative.

(i) From Theorem 6.2.2 and Problem 23.1 in Van der Vaart (2000, p.339), it can be seen that

$$\sup_{x \in \mathbb{R}} |\mathrm{P}(\Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) < x | \mathrm{H}_3^{(n)}) - \mathrm{P}(\mathfrak{L}_N < x)| \to 0 \quad \text{as } n \to \infty.$$

It implies that

$$|\mathrm{P}(\Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) \geq c_{n,h} | \mathrm{H}_3^{(n)}) - \mathrm{P}(\mathfrak{L}_N \geq c_{n,h})| \to 0 \quad \text{as } n \to \infty.$$

Hence, we obtain

$$\alpha = \mathrm{P}(\Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) \geq c_{n,h} | \mathrm{H}_3^{(n)}) = \lim_{n \to \infty} \mathrm{P}(\mathfrak{L}_N \geq c_{n,h}),$$

and consequently

$$c_{n,h} \to c = -\log \frac{a}{2} \log \left(1 + \frac{h_1}{\theta_2}\right) + \frac{h_1}{2(\theta_2 + h_1)} \chi_a^2[1 - \alpha] \quad \text{as } n \to \infty.$$

(ii) Theorems 6.2.3 and 6.2.4 (i) yields that

$$\Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) - c_{n,h} \Rightarrow \mathfrak{L}_A - c \quad \text{under } \mathrm{K}_3^{(n)}$$

and it gives that

$$\begin{aligned}\lim_{n \to \infty} \mathrm{P}(\Lambda_3(\boldsymbol{\theta}_0, \boldsymbol{\theta}_n) \geq c_{n,h} | \mathrm{K}_3^{(n)}) &= \mathrm{P}(\mathfrak{L}_N \geq c) \\ &= \mathrm{P}(\chi_a^2 \geq \frac{\theta_2}{(\theta_2 + h_1)} \chi_a^2[1 - \alpha]).\end{aligned}$$

(iii) By Neyman–Pearson lemma, for any $n \in \mathbb{N}$ and any test $\{\psi_{n,h}\}$ such that $\limsup_n \mathrm{E}_{\theta_n}(\psi_{n,h}) \leq \alpha$, it holds $\mathrm{E}_{\theta_n}(\phi_{n,h}) - \mathrm{E}_{\theta_n}(\psi_{n,h}) \geq 0$. Thus, the desired result holds. □

Proof of Theorem 6.2.5

From the fact that $\sum_{i=1}^{a} (\sqrt{n}(\bar{y}_{i\cdot} - \theta_3)/\sqrt{\theta_2 + n\theta_1})^2$ follows χ_a^2, simple algebra gives that

$$\begin{aligned}&\Lambda(\boldsymbol{\theta}_n, \boldsymbol{\theta}) \\ &= -\frac{a}{2}(n-1) \log \frac{\theta_2^{(n)}}{\theta_2} - \frac{a}{2} \log \frac{\theta_2^{(n)} + n\theta_1^{(n)}}{\theta_2 + n\theta_1} + \frac{\theta_2^{(n)} - \theta_2}{2\theta_2^{(n)}} \{\frac{1}{\theta_2} \sum_i \sum_j (y_{ij} - \bar{y}_{i\cdot})^2\}\end{aligned}$$

$$
\begin{aligned}
&+\frac{(\theta_2^{(n)}-\theta_2)+n(\theta_1^{(n)}-\theta_1)}{2(\theta_2^{(n)}+n\theta_1^{(n)})}\{\frac{n}{\theta_2+n\theta_1}\sum_i(\bar{y}_{i\cdot}-\theta_3)^2\} \\
&+\frac{\sqrt{n\theta_2+n^2\theta_1}(\theta_3^{(n)}-\theta_3)}{\theta_2^{(n)}+n\theta_1^{(n)}}\{\frac{\sqrt{n}}{\sqrt{\theta_2+n\theta_1}}\sum_i(\bar{y}_{i\cdot}-\theta_3)\}-\frac{an(\theta_3^{(n)}-\theta_3)^2}{2(\theta_2^{(n)}+n\theta_1^{(n)})} \\
=&-\frac{a}{2}(n-1)\log(1+\frac{h_2}{n^{k_2}\theta_2})+\frac{ah_2(1-\frac{1}{n})}{2(n^{k_2-1}\theta_2+\frac{h_2}{n})} \\
&+\frac{h_2\sqrt{2a(1-\frac{1}{n})}}{2(n^{k_2-\frac{1}{2}}\theta_2+\frac{h_2}{\sqrt{n}})}\left(\frac{\frac{1}{\theta_2}\sum_i\sum_j(y_{ij}-\bar{y}_{i\cdot})^2-a(n-1)}{\sqrt{2a(n-1)}}\right)+o_p(1),
\end{aligned}
$$

which shows the desired result. □

Proof of Theorem 6.2.6

In the same way to Theorem 6.2.5, it holds that

$$
\begin{aligned}
&\Lambda(\boldsymbol{\theta}_n,\boldsymbol{\theta}) \\
=&-\frac{a}{2}\log\frac{\theta_2+n\theta_1^{(n)}}{\theta_2+n\theta_1}+\frac{n(\theta_1^{(n)}-\theta_1)}{2(\theta_2+n\theta_1^{(n)})}\{\frac{n}{\theta_2+n\theta_1}\sum_i(\bar{y}_{i\cdot}-\theta_3)^2\} \\
=&-\frac{a}{2}\log(1+\frac{h_1}{n^{k_1-1}\theta_2+n^{k_1}\theta_1})+\frac{h_1}{2(n^{k_1-1}\theta_2+n^{k_1}\theta_1+h_1)}\sum_i(\frac{Y_i-\sqrt{n}\theta_3}{\sqrt{\theta_2+n\theta_1}})^2,
\end{aligned}
$$

which tends to 0 as $n\to\infty$. □

Bibliography

Abrahams, J. (1986). A survey of recent progress on level-crossing problems for random processes. In *Communications and Networks*. Blake, I.F. and Poor, H.V. (eds.) 6–25.

Agosto, A., Cavaliere, G., Kristensen, D., and Rahbek, A. (2016). Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX). *J. Empir. Finance* **38** 640–663.

Ahmad, A. and Francq, C. (2016). Poisson qmle of count time series models. *J. Time Ser. Anal.* **37** 291–314.

Aknouche, A., Bendjeddou, S., and Touche, N. (2018). Negative binomial quasi-likelihood inference for general integer-valued time series models. *J. Time Ser. Anal.* **39** 192–211.

Aknouche, A. and Francq, C. (2020). Count and duration time series with equal conditional stochastic and mean orders. *Econ. Theory*, p. 1–33.

Al-Osh, M. and Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR (1)) process. *J. Time Ser. Anal.* **8** 261–275.

Alzaid, A. and Al-Osh, M. (1990). An integer-valued pth-order autoregressive structure (INAR (p)) process. *J. Appl. Probab.* 314–324.

Anderson, R. (1960). Uses of variance component analysis in the interpretation of biological experiments. *Bull. Internat. Statist. Inst.,* **37** 71–90.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.

Anderson, T. W., Rubin, H. et al. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Ann. Math. Stat.* **20** 46–63.

Anderson, T. W., Kunitomo, N., and Morimune, K. (1986). Comparing single-equation estimators in a simultaneous equation system. *Econometric Theory* 1–32.

Bagnall, A. and Janacek, G. (2005). Clustering time series with clipped data. *Mach. Learn.* **58** 151–178.

Barnett, J. T. and Kedem, B. (1991a). Zero-crossing rates of functions of Gaussian processes. *IEEE Trans. Inf. Theory* **37** 1188–1194.

Barnett, J. T. and Kedem, B. (1991b). Zero-crossing rates of functions of Gaussian processes. *IEEE Trans. Inf. Theory* **37** 1188–1194.

Barnett, J. T. and Kedem, B. (1998). Zero-crossing rates of mixtures and products of Gaussian processes. *IEEE Trans. Inf. Theory* **44** 1672–1677.

Barnett, J. (2001). Zero-Crossings of Random Processes with Application to Estimation and Detection. In *Nonuniform Sampling*. Springer 393–437.

Basmann, R. L. (1957). A generalized classical method of linear estimation of coefficients in a structural equation. *Econometrica* 77–83.

Baum, R. (1957). The correlation function of smoothly limited gaussian noise. *IEEE Trans. Inform. Theory* **3** 193–197.

Benghabrit, Y. and Hallin, M. (1996a). Locally asymptotically optimal tests for autoregressive against bilinear serial dependence. *Stat. Sin.* 147–169.

Benghabrit, Y. and Hallin, M. (1996b). Rank-based tests for autoregressive against bilinear serial dependence. *J. Nonparametr. Stat.* **6** 253–272.

Benjamin, M. A., Rigby, R. A., and Stasinopoulos, D. M. (2003). Generalized autoregressive moving average models. *J. Amer. Stat. Assoc.* **98** 214–223.

Berkes, I., Horváth, L., and Kokoszka, P. (2004). Testing for parameter constancy in GARCH (p, q) models. *Stat. Probab. Lett.* **70** 263–273.

Billinsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley.

Billinsley, P. (1999). *Convergence of Probability Measures, 2nd ed.*. New York: Wiley.

Birr, S., Volgushev, S., Kley, T., Dette, H., and Hallin, M. (2017). Quantile spectral analysis for locally stationary time series. *J. Roy. Statist. Soc. Ser. B* **79** 1619–1643.

Blake, I. F. and Lindsey, W. C. (1973). Level-crossing problems for random processes. *IEEE Trans. Inform. Theory* **19** 295–315.

Brillinger, D. R. (1968). Estimation of the cross-spectrum of a stationary bivariate Gaussian process from its zeros. *J. R. Stat. Soc. Series B Stat. Methodol.* **30** 145–159.

Brillinger, D. R. (1981). *Time series: Data Analysis and Theory*. San Francisco: Holden-Day, expanded ed.

Brockwell, P. J. and Davis, R. A. (1991). *Time series: Theory and Methods*. New York: Springer-Verlag.

Buz, A. and Litan, C. (2012). What properties do clipped data inherit from the generating processes?. *Studia Universitatis Babes-Bolyai* **57** 85–96.

Christou, V. and Fokianos, K. (2014). Quasi-likelihood inference for negative binomial time series models. *J. Time Ser. Anal.* **35** 55–78.

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments. *Scand. Stat. Theory Appl.* 93–115.

Cui, Y. and Wang, X. (2019). Conditional maximum likelihood estimation in negative binomial INGARCH processes with known number of successes when the true parameter is at the boundary of the parameter space. *Commun. Stat. Theory Methods* **48** 3388–3401.

Cui, Y., Wu, R., and Zheng, Q. (2020). Estimation of change-point for a class of count time series models. *Scand. Stat. Theory Appl.*

Cutting, C., Paindaveine, D., and Verdebout, T. (2017). Testing uniformity on high-dimensional spheres against monotone rotationally symmetric alternatives. *Ann. Statist.* **45** 1024–1058.

Dahlhaus, R. (1996). Maximum likelihood estimation and model selection for locally stationary processes. *J. Nonparametr. Stat.* **6** 171–191.

Dau, H. A., Keogh, E., Kamgar, K., Yeh, C.-C. M., Zhu, Y., Gharghabi, S., Ratanamahatana, C. A., Chen, Y., Hu, B., Begum, N., Bagnall, A., Mueen, A., and Batista, G. (2018). The UCR Time Series Classification Archive. https://www.cs.ucr.edu/ eamonn/time_series_data_2018/.

Davis, R. A., Dunsmuir, W. T., and Streett, S. B. (2003). Observation-driven models for Poisson counts. *Biometrika* **90** 777–790.

Davis, R. A., Dunsmuir, W. T., and Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika* **87** 491–505.

Davis, R. A. and Liu, H. (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Stat. Sin.* 1673–1707.

Davis, R. A. and Wu, R. (2009). A negative binomial model for time series of counts. *Biometrika* **96** 735–749.

Dette, H., Hallin, M., Kley, T., and Volgushev, S. (2015). Of copulas, quantiles, ranks and spectra: An L_1-approach to spectral analysis. *Bernoulli* **21** 781–831.

Dickerson, G. E. (1942). Experimental design for testing inbred lines of swine. *J. Animal Sci.* **1** 326–341.

Diop, M. L. and Kengne, W. (2017). Testing parameter change in general integer-valued time series. *J. Time Ser. Anal.* **38** 880–894.

Diop, M. L. and Kengne, W. (2020). Piecewise autoregression for general integer-valued time series. *J. Stat. Plan. Inference*.

Doukhan, P., Fokianos, K., Tjøstheim, D. et al. (2013). Correction to " On weak dependence conditions for Poisson autoregressions " [Statist. Probab. Lett. 82 (2012) 942–948]. *Stat. Probab. Lett.* **83** 1926–1927.

Doukhan, P. and Kengne, W. (2015). Inference and testing for structural change in general poisson autoregressive models. *Electron. J. Stat.* **9** 1267–1314.

Ethier, S. N. and Kurtz, T. G. (1986). *Markov processes: characterization and convergence* **282**. John Wiley & Sons.

Fahrmeir, L. and Kaufmann, H. (1987). Regression models for non-stationary categorical time series. *J. Time Ser. Anal.* **8** 147–160.

Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued GARCH process. *J. Time Ser. Anal.* **27** 923–942.

Fisher, N. I., Lewis, T., and Embleton, B. J. (1993). *Statistical Analysis of Spherical Data*. Cambridge university press.

Fisher, R. A. (1918). 009: The Correlation Between Relatives on the Supposition of Mendelian Inheritance.. *Trans. Roy. Soc. Edinburgh* **52** 399–433.

Fisher, R. A. (1953). Dispersion on a sphere. *Proc. Roy. Soc. Lond. Ser. A* **217** 295–305.

Fitzmaurice, G. M. and Lipsitz, S. R. (1995). A model for binary time series data with serial odds ratio patterns. *J. R. Stat. Soc. Ser. C Appl. Stat.* **44** 51–61.

Fokianos, K. (2001). Truncated Poisson regression for time series of counts. *Scand. Stat. Theory Appl.* **28** 645–659.

Fokianos, K. and Kedem, B. (1998). Prediction and classification of non-stationary categorical time series. *J. Multivariate Anal.* **67** 277–296.

Fokianos, K. and Kedem, B. (2003). Regression theory for categorical time series. *Statist. Sci.* **18** 357–376.

Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009). Poisson autoregression. *J. Amer. Stat. Assoc.* **104** 1430–1439.

Fokianos, K. and Tjøstheim, D. (2011). Log-linear Poisson autoregression. *J. Multivar. Anal.* **102** 563–578.

Franke, J., Kirch, C., and Kamgaing, J. T. (2012). Changepoints in times series of counts. *J. Time Ser. Anal.* **33** 757–770.

Fujikoshi, Y., Morimune, K., Kunitomo, N., and Taniguchi, M. (1982). Asymptotic expansions of the distributions of the estimates of coefficients in a simultaneous equation system. *J. Econom.* **18** 191–205.

Garel, B. and Hallin, M. (1995). Local asymptotic normality of multivariate ARMA processes with a linear trend. *Ann. Inst. Stat. Math.* **47** 551–579.

Giraitis, L., Kokoszka, P., and Leipus, R. (2000). Stationary arch models: dependence structure and central limit theorem. *Econ. Theory* **16** 3–22.

Gómez, E., Gómez-Villegas, M. A., and Marín, J. M. (2003). A survey on continuous elliptical vector distributions. *Rev. Mat. Complut* **16** 345–361.

Gorgi, P. (2020). Beta-negative binomial auto-regressions for modelling integer-valued time series with extreme observations. *J. R. Stat. Soc. Series B Stat. Methodol.*

Goto, Y. (2020). Estimation of trigonometric moments for circular distribution of MA(p) type by using binary series. *Sci. Math. Jpn.* **e-2020 33** 2020–2 (in Editione Electronica).

Goto, Y. and Taniguchi, M. (2020). Discriminant analysis based on binary time series. *Metrika* **83** 569–595.

Goto, Y. and Taniguchi, M. (2019). Robustness of Zero Crossing Estimator. *J. Time Ser. Anal.* **40** 815–830.

Hagemann, A. (2011). Robust spectral analysis. *arXiv preprint arXiv:1111.1965*.

Hallin, M., Ingenbleek, J.-F., and Puri, M. L. (1985). Linear serial rank tests for randomness against ARMA alternatives. *Ann. Statist.* **13** 1156–181.

Hallin, M. and Puri, M. L. (1994). Aligned rank tests for linear models with autocorrelated error terms. *J. Multivar. Anal.* **50** 175–237.

Hallin, M., Taniguchi, M., Serroukh, A., and Choy, K. (1999). Local asymptotic normality for regression models with long-memory disturbance. *Ann. Statist.* **27** 2054–2080.

Hannan, E. J. (1970). *Multiple Time Series* **38**. John Wiley & Sons.

He, S. and Kedem, B. (1989). On the stieltjes-sheppard orthant probability formula. tech. rep. tr-89-69, Dept. Mathematics, Unvi. Maryland, College Park.

Heinen, A. (2003). Modelling time series count data: an autoregressive conditional Poisson model. *Munich Personal RePEc Archive*.

Hinich, M. (1967). Estimation of spectra after hard clipping of gaussian processes. *Technometrics* **9** 391–400.

Hirotsu, C. (2017). *Advanced analysis of variance* **384**. John Wiley & Sons.

Hirukawa, J. and Taniguchi, M. (2006). LAN theorem for non-Gaussian locally stationary processes and its applications. *J. Stat. Plan. Inference* **136** 640–688.

Hong, Y. (1999). Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. *J. Amer. Stat. Assoc.* **94** 1201–1220.

Hong, Y. (2000). Generalized spectral tests for serial dependence. *J. R. Stat. Soc. Series B Stat. Methodol.* **62** 557–574.

Hosoya, Y. and Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *Ann. Stat.* **10** 132–153.

Hosoya, Y., Tsukuda, Y., and Terui, N. (1989). Ancillarity and the limited information maximum-likelihood estimation of a structural equation in a simultaneous equation system. *Econ. Theory* 385–404.

Hudecová, Š., Hušková, M., and Meintanis, S. G. (2017). Tests for structural changes in time series of counts. *Scand. Stat. Theory Appl.* **44** 843–865.

Iacus, S. M. (2016). *sde: Simulation and Inference for Stochastic Differential Equations*, R package version 2.0.15, https://CRAN.R-project.org/package=sde.

Ibragimov, I. A. and Rozanov, Y. (1978). *Gaussian Random Processes* **9**. Springer Science & Business Media.

Ibragimov, I. and Khasminskii, R. (1981). Statistical Estimation: Asymptotic Theory. Springer, Berlin; New York.

Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econ. Theory* 818–887.

Johnson, R. and Wichern, D. (1988). *Applied Multivariable Statistical Analysis*. 2nd ed. Englewood Cliffs: Prentice-Hall.

Kakizawa, Y. (1996). Discriminant analysis for non-Gaussian vector stationary processes. *J. Nonparametr. Stat.* **7** 187–203.

Kakizawa, Y. (1997). Higher order asymptotic theory for discriminant analysis in gaussian stationary processes. *J. Jpn. Stat. Soc.* **27** 19–35.

Kang, J. and Lee, S. (2014). Parameter change test for poisson autoregressive models. *Scand. Stat. Theory Appl.* **41** 1136–1152.

Kano, Y. (1994). Consistency property of elliptic probability density functions. *J. Multivar. Anal* **51** 139–147.

Kato, H., Taniguchi, M., and Honda, M. (2006). Statistical analysis for multiplicatively modulated nonlinear autoregressive model and its applications to electrophysiological signal analysis in humans. *IEEE Trans. Signal Process.* **54** 3414–3425.

Kato, S. and Jones, M. (2013). An extended family of circular distributions related to wrapped Cauchy distributions via Brownian motion. *Bernoulli* **19** 154–171.

Kaufmann, H. (1987). Regression models for nonstationary categorical time series: asymptotic estimation theory. *Ann. Stat.* **15** 79–98.

Kedem, B. (1980). *Binary Time Series*. New York: Dekker.

Kedem, B. (1994). *Time Series Analysis by Higher Order Crossings*. New York: IEEE press.

Kedem, B. and Fokianos, K. (2002). *Regression models for time series analysis*. New York: Wiley.

Kedem, B. and Li, T. (1989). Higher order crossings from a parametric family of linear filters. tech. rep. tr-89-47, Dept. Mathematics, Unvi. Maryland, College Park.

Kedem, B. and Slud, E. (1982). Time series discrimination by higher order crossings. *Ann. Stat.* 786–794.

Keenan, D. M. (1982). A time series analysis of binary data. *J. Amer. Stat. Assoc.* **77** 816–821.

Kley, T. (2014). Quantile-based spectral analysis. Ph. D. *thesis*, Ruhr-Universitut Bochum.

Kley, T., Volgushev, S., Dette, H., and Hallin, M. (2016). Quantile spectral processes: Asymptotic analysis and inference. *Bernoulli* **22** 1770–1807.

Koenker, R. and Xiao, Z. (2006). Quantile autoregression. *J. Amer. Stat. Assoc.* **101** 980–990.

Kreiss, J.-P. (1987). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* 112–133.

Kreiss, J.-P. (1990). Local asymptotic normality for autoregression with infinite order. *J. Stat. Plan. Inference* **26** 185–219.

Kunsch, H. (1984). Infinitesimal robustness for autoregressive processes. *Ann. Stat.* 843–863.

LeCam, L. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.* **3** 37–98.

Lee, J. and Rao, S. S. (2012). The quantile spectral density and comparison based tests for nonlinear time series. *arXiv preprint arXiv:1112.2759*.

Lee, S., Ha, J., Na, O., and Na, S. (2003). The cusum test for parameter change in time series models. *Scand. Stat. Theory Appl.* **30** 781–796.

Lee, S., Lee, Y., and Chen, C. W. (2016). Parameter change test for zero-inflated generalized Poisson autoregressive models. *Statistics* **50** 540–557.

Lee, Y. and Lee, S. (2019). CUSUM test for general nonlinear integer-valued GARCH models: comparison study. *Ann. Inst. Stat. Math.* **71** 1033–1057.

Lee, Y., Lee, S., and Tjøstheim, D. (2018). Asymptotic normality and parameter change test for bivariate Poisson INGARCH models. *Test* **27** 52–69.

Lehmann, E. L. and Romano, J. P. (2005). *Testing statistical hypotheses (3rd ed.)*. Springer Science & Business Media.

Leonov, V. and Shiryaev, A. N. (1959). On a method of calculation of semi-invariants. *Theory Probab. Appl.* **4** 319–329.

Ley, C. and Verdebout, T. (2017). *Modern directional statistics*. CRC Press.

Li, T.-H. (2008). Laplace periodogram for time series analysis. *J. Amer. Stat. Assoc.* **103** 757–768.

Li, T.-H. (2012). Quantile periodograms. *J. Amer. Stat. Assoc.* **107** 765–776.

Li, T.-H. (2014). Quantile periodogram and time-dependent variance. *J. Time Ser. Anal.* **35** 322–340.

Liboschik, T., Fokianos, K., and Fried, R. (2017). tscount: An R package for analysis of count time series following generalized linear models. *J. Stat. Softw.* **82** 1–51.

Liggett Jr, W. (1971). On the asymptotic optimality of spectral analysis for testing hypotheses about time series. *Ann. Stat.* 1348–1358.

Linton, O. (1993). Adaptive estimation in ARCH models. *Econ. Theory* 539–569.

Lomnicki, Z. and Zaremba, S. (1955). Some applications of zero-one processes. *J. R. Stat. Soc. Series B Stat. Methodol.* **17** 243–255.

Ma, Y. and Genton, M. (1998). Highly robust estimation of the autocovariance function. *J. Time Ser. Anal.* **21** 663–684.

Mardia, K. V. (1975). Statistics of directional data. *J. R. Stat. Soc. Series B Stat. Methodol.* **37** 349–371.

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics* **494**. John Wiley & Sons.

McCullagh, P. and Nelder, J. (1989). Generalized Linear Models. *CRC Monographs on Statistics & Applied Probability, Springer Verlag, New York*.

McKenzie, E. (1985). Some simple models for discrete variate time series 1. *J. Am. Water Resour. Assoc.* **21** 645–650.

Melo, M. and Alencar, A. (2020). Conway–Maxwell–Poisson Autoregressive Moving Average Model for Equidispersed, Underdispersed, and Overdispersed Count Data. *J. Time Ser. Anal.*

Neumann, M. H. (2011). Absolute regularity and ergodicity of poisson count processes. *Bernoulli* **17** 1268–1284.

Oh, H. and Lee, S. (2018). On score vector-and residual-based CUSUM tests in ARMA–GARCH models. *Statist. Methods Appl.* **27** 385–406.

Olszewski, R. T. (2001). Generalized feature extraction for structural pattern recognition in time-series datas. Ph. D. *thesis*, Carnegie Mellon University.

Page, E. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42** 523–527.

Paindaveine, D., Verdebout, T. et al. (2017). Inference on the mode of weak directional signals: a Le Cam perspective on hypothesis testing near singularities. *Ann. Statist.* **45** 800–832.

Panagiotakis, C. and Tziritas, G. (2005). A speech/music discriminator based on rms and zero-crossings. *IEEE Trans. multimedia* **7** 155–166.

Pedeli, X., Davison, A. C., and Fokianos, K. (2015). Likelihood estimation for the INAR (p) model by saddlepoint approximation. *J. Amer. Stat. Assoc.* **110** 1229–1238.

Petrantonakis, P. C. and Hadjileontiadis, L. J. (2010a). Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis. *IEEE Trans. Affective Computing* **1** 81–97.

Petrantonakis, P. C. and Hadjileontiadis, L. J. (2010b). Emotion recognition from EEG using higher order crossings. *IEEE Trans. Inform. Tech. Bio* **14** 186–197.

Phillips, P. C. (1989). Partially identified econometric models. *Econometric Theory* **5** 181–240.

Resnick, S. I. (1999). *A probability path*. Springer.

Rice, S. O. (1944). Mathematical analysis of random noise. *Bell Syst. Tech. J.* **23** 282–332.

Robert, C. P., Casella, G., and Casella, G. (2010). *Introducing Monte Carlo Methods with R*. Springer.

Robinson, P. M. (1991). Automatic frequency domain inference on semiparametric and nonparametric models. *Econometrica* **59** 1329–1363.

Roussas, G. G. (1972). *Contiguity of probability measures: some applications in statistics* **63**. Cambridge university press.

Roussas, G. G. (1979). Asymptotic distribution of the log-likelihood function for stochastic processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **47** 31–46.

Sakiyama, K. and Taniguchi, M. (2003). Testing composite hypotheses for locally stationary processes. *J. Time Ser. Anal.* **24** 483–504.

Sakiyama, K. and Taniguchi, M. (2004). Discriminant analysis for locally stationary processes. *J. Multivariate Anal.* **90** 282–300.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance components* **391**. John Wiley & Sons.

Sharpe, R. and Van Middelem, C. (1955). Application of variance components to horticultural problems with special reference to a parathion residue study. *Proc. Amer. Soc. Horticultural Sci* **66** 415–420.

Shumway, R. and Unger, A. (1974). Linear discriminant functions for stationary time series. *J. Amer. Stat. Assoc.* **69** 948–956.

Sprague, G. F. and Tatum, L. A. (1942). General vs. specific combining ability in single crosses of corn 1. *Agron. J.* **34** 923–932.

Stieltjes, T. S. (1889). Extrait d'une lettre adréssee à M. Hermite. *Bull. Sci. Math. (Ser. 2)* **13** 170–172.

Swensen, A. R. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *J. Multivar. Anal.* **16** 54–70.

Takeuchi, K. and Morimune, K. (1985). Third-order efficiency of the extended maximum likelihood estimators in a simultaneous equation system. *Econometrica* 177–200.

Talbot, M. (1984). Yield variability of crop varieties in the UK. *J. Agric. Sci.* **102** 315–321.

Tanaka, M. and Shimizu, K. (2001). Discrete and continuous expectation formulae for level-crossings, upcrossings and excursions of ellipsoidal processes. *Stat. Probab. Lett.* **52** 225–232.

Taniguchi, M. (1987). Minimum contrast estimation for spectral densities of stationary processes. *J. R. Stat. Soc. Series B Stat. Methodol.* **49** 315–325.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. New York: Springer-Verlag.

Taniguchi, M., Kato, S., Ogata, H., and Pewsey, A. (2020). Models for circular data from time series spectra. *J. Time Ser. Anal.* **41** 808–829.

Theil, H. (1953). Repeated least squares applied to complete equation systems. *The Hague: central planning bureau*.

Theil, H. (1961). Economic forecasts and policy.

Van der Vaart, A. W. (2000). *Asymptotic statistics* **3**. Cambridge university press.

Wang, C., Liu, H., Yao, J.-F., Davis, R. A., and Li, W. K. (2014). Self-excited threshold Poisson autoregression. *J. Amer. Stat. Assoc.* **109** 777–787.

Watson, G. S. (1983). *Statistics on Spheres*. Wiley-Interscience.

Whitt, W. (2007). Proofs of the martingale FCLT. *Probability Surveys* **4** 268–302.

Ylvisaker, N. D. (1965). The expected number of zeros of a stationary gaussian process. *Ann. Stat.* **36** 1043–1046.

Zhang, G. and Taniguchi, M. (1995). Nonparametric approach for discriminant analysis in time series. *J. Nonparametr. Stat.* **5** 91–101.

Zhu, F. (2011). A negative binomial integer-valued GARCH model. *J. Time Ser. Anal.* **32** 54–67.

Zhu, F. (2012a). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *J. Math. Anal. Appl.* **389** 58–71.

Zhu, F. (2012b). Modeling time series of counts with COM-Poisson INGARCH models. *Math. Comput. Model* **56** 191–203.

Zhu, F. (2012c). Zero-inflated Poisson and negative binomial integer-valued GARCH models. *J. Stat. Plan. Inference* **142** 826–839.

List of Papers

[1] Y. Goto and M. Taniguchi (2019). Robustness of zero crossings estimator, *J. Time Ser. Anal.*, **40** 815-830

[2] Y. Goto and M. Taniguchi (2020). Discriminant analysis based on binary time series, *Metrika*, **83** 569-595

[3] Y. Goto (2020). Estimation of trigonometric moments for circular distribution of MA(p) type by using binary series, *Sci. Math. Jpn.*, e-2020 **33** 2020-2 (in Editione Electronica)