ABSTRACT

Title of Dissertation: STATISTICAL DATA FUSION WITH DENSITY RATIO MODEL AND EXTENSION TO RESIDUAL COHERENCE

Xuze Zhang
Doctor of Philosophy, 2024

Dissertation Directed by: Professor Benjamin Kedem
Department of Mathematics

Nowadays, the statistical analysis of data from diverse sources has become more prevalent. The Density Ratio Model (DRM) is one of the methods for fusing and analyzing such data. The population distributions of different samples can be estimated based on fused data, which leads to more precise estimates of the probability distributions. These probability distributions are related by assuming the ratios of their probability density functions (PDFs) follow a parametric form. In the previous works, this parametric form is assumed to be uniform for all ratios. In Chapter 1, an extension is made to allow this parametric form to vary for different ratios. Two methods of determining the parametric form for each ratio are developed based on asymptotic test and Akaike Information Criterion (AIC). This extended DRM is applied to Radon concentration and Pertussis rates to demonstrate the use of this extension in univariate case and multivariate case, respectively.

The above analysis is made possible when data in each sample are independent and identically distributed (IID). However, in many cases, statistical analysis is entailed for time series in which data appear to be sequentially dependent. In Chapter 2, an extension is made for DRM to account for weakly dependent data, which allows us to investigate the structure of multiple time series on the strength of each other. It is shown that the IID assumption can be replaced by proper stationarity, mixing and moment conditions. This extended DRM is applied to the analysis of air quality data which are recorded in chronological order.

As mentioned above, DRM is suitable for the situation where we investigate a single time series based on multiple alternative ones. These time series are assumed to be mutually independent. However, in time series analysis, it is often of interest to detect linear and nonlinear dependence between different time series. In such dependent scenario, coherence is a common tool to measure the linear dependence between two time series, and residual coherence is used to detect a possible quadratic relationship. In Chapter 3, we extend the notion of residual coherence and develop statistical tests for detecting linear and nonlinear associations between time series. These tests are applied to the analysis of brain functional connectivity data.

STATISTICAL DATA FUSION WITH DENSITY RATIO MODEL AND EXTENSION TO RESIDUAL COHERENCE

by

Xuze Zhang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2024

Advisory Committee:
Professor Benjamin Kedem, Chair/Advisor
Professor Xin He
Professor Partha Lahiri
Professor Doron Levy
Professor Vince Lyzinski

Foreword

Xuze Zhang, made substantial contributions to the relevant aspects of the jointly authored work included in the dissertation. This pertains to the following papers:

Goto, Y., Zhang, X., Kedem, B., & Chen, S. (2023). Residual Spectrum Applied in Brain Functional Connectivity. Manuscript submitted for publication. arXiv preprint arXiv:2305.19461.

Zhang, X., Pyne, S., & Kedem, B. (2021). Multivariate Tail Probabilities: Predicting Regional Pertussis Cases in Washington State. *Entropy, 23*(6), 675.

Zhang, X., & Kedem, B. (2021). Extended residual coherence with a financial application. *Statistics in Transition New Series, 22*(2), 1-14.

Zhang, X., Pyne, S., & Kedem, B. (2020). Model selection in radon data fusion. *Statistics in Transition, New Series*, 167-174.

Zhang, X., Pyne, S., & Kedem, B. (2020). Estimation of residential radon concentration in Pennsylvania counties by data fusion. *Applied Stochastic Models in Business and Industry, 36*(6), 1094-1110.

Acknowledgments

I would like to express my sincere gratitude to all the people who helped me during my graduate study and dissertation research.

First, I would like to thank my advisor, Professor Benjamin Kedem for his guidance in my graduate study. His guidance has always given me inspiration and motivation for my dissertation research. He has always been available for offering help and advice. Without his support, I could not have finished my dissertation. It is my pleasure to work under the guidance of such a knowledgeable and supportive advisor. Professor Kedem, thank you for everything that you have done for me!

Moreover, I would like to thank Professor Paul Smith, Professor Eric Slud, Professor Partha Lahiri and Professor Lizhen Lin for helping me with my study and career advancement. Owning to their advices and help, I am able to pursue my PhD degree and further my career.

I would also like to thank my co-authors of the publications during my graduate study. I appreciate the opportunity that Professor Saumyadipta Pyne gave me to work on the analysis of radon concentration and pertussis occurrence, and the support he provided during our collaboration. I want to thank Dr. Yuichi Goto whom I worked with on the residual coherence project. He is insightful and knowledgeable, and I learned a lot from him during our discussions.

My friends and classmates has offered me help and support during my graduate study. I would like to express my thanks to Dr. Yifan Yang, Dr. Yiran Zhang, Dr. Yunjiang Ge and Dr.

Nathan Yu.

I would also acknowledge help and support from staff members in the math department. Especially, I want express my gratitude and appreciation to former and current staff members, Cristina Garcia, Stephanie Padgett, Trystan Denhard and Jemma Natanson.

In addition, I want thank the members of my dissertation committee, Professor Xin He, Professor Partha Lahiri, Professor Doron Levy and Professor Vince Lyzinski for their support.

Finally, I want to thank my parents and my aunt for their support. They have always helped me with my difficulties and believed in me. Without their support, I could not have finished my graduate study. I want to express my deepest gratitude to them for their guidance and trust.

Table of Contents

Foreword ii

Acknowledgements iii

Table of Contents v

List of Tables vii

List of Figures ix

List of Abbreviations x

Chapter 1: Density Ratio Model with Variable Tilts 1
- 1.1 Introduction . . . 1
- 1.2 A Density Ratio Model with Variable Tilts . . . 6
  - 1.2.1 Estimation . . . 7
  - 1.2.2 Asymptotic Behavior of $\widetilde{\boldsymbol{\theta}}$ . . . 10
  - 1.2.3 Asymptotic Behavior of $\widetilde{G}$ . . . 21
  - 1.2.4 Consistent Estimators of Asymptotic Variances . . . 33
- 1.3 Simulation . . . 39
- 1.4 Tilt Selection . . . 42
- 1.5 Estimation of Residential Radon Concentration in Pennsylvania Counties . . . 43
  - 1.5.1 Beaver County . . . 44
  - 1.5.2 Forest County . . . 48
- 1.6 Estimation of Regional Pertussis Rates in Washington State . . . 50

Chapter 2: A Density Ratio Model with Weakly Dependent Data 55
- 2.1 Introduction . . . 55
- 2.2 A Density Ratio Model with Weakly Dependent Data . . . 58
  - 2.2.1 Estimation . . . 60
  - 2.2.2 Asymptotic Behavior of $\widetilde{\boldsymbol{\theta}}$ . . . 61
  - 2.2.3 Asymptotic Behavior of $\widetilde{G}$ . . . 71
  - 2.2.4 Consistent Estimators of Asymptotic Variances . . . 93
- 2.3 Simulation . . . 103
  - 2.3.1 Confidence Interval for $\boldsymbol{\theta}$ . . . 104
  - 2.3.2 Confidence Interval for G . . . 109
  - 2.3.3 Testing for Equality of Distributions . . . 113

2.4 Analysis of Air Monitoring Data 116

Chapter 3: Extension to Residual Coherence 120
3.1 Introduction 120
3.2 A Multi-Input Single-Output Model 123
3.2.1 Orthogonal Projection 123
3.2.2 Spectral Representation of Orthogonal Processes 125
3.2.3 Testing for Input Effect 131
3.3 Nonlinearity and Residual Coherence 146
3.4 Simulation 149
3.5 Analysis of Brain Functional Connectivity Data 151
3.5.1 Data Processing 152
3.5.2 Testing for Quadratic Association 153
3.5.3 A Forward Selection Method 156

Bibliography 161

List of Tables

1.1 MIAE and MISE for $\widetilde{G}_v$, $\widetilde{G}_u$ and $\widehat{G}$. 40
1.2 Average of threshold probability estimates, 95% CIs, lengths of CIs, coverage rates of CIs for thresholds $T = 0.1, 0.5, 1, 2, 3$. "-" indicates that some of the CIs generated can not be constructed since the estimates are 0. 42
1.3 A Comparison between $\widetilde{G}_v$, $\widetilde{G}_u$ and $\widehat{G}$ based on samples of six periods in Beaver county. Threshold probability estimates, 95% CIs and lengths of CIs are computed for thresholds $T = 5, 10, 50, 100, 200$. "-" indicates that some of the CIs generated can not be constructed since the estimates are 0. 47
1.4 A Comparison between $\widetilde{G}_v$, $\widetilde{G}_u$ and $\widehat{G}$ based on samples from Beaver, Lawrence, Butler, Allegheny and Washington. Threshold probability estimates, 95% CIs and lengths of CIs are computed for thresholds $T = 5, 10, 50, 100, 200$. "-" indicates that some of the CIs generated can not be constructed since the estimates are 0. . 50
1.5 AIC values of models based on different choices of $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ in the Forest case. A hyphen "-" indicates that $\boldsymbol{h}_k(x) = 0$ and therefore g_0 and g_k are identical for $k = 1, 2$. 51
1.6 Comparison between $\widetilde{G}_v$, $\widetilde{G}_u$ and $\widehat{G}$ based on samples from Forest, Warren and Elk. Threshold probability estimates, 95% CIs and lengths of CIs are computed for thresholds $T = 5, 10, 50, 100, 200$. 52
1.7 Summary statistics of pertussis rates in Clark, Cowlitz, Pierce, King, Skagit and Whatcom. 52
1.8 Comparison between $\widetilde{P}_v$, $\widetilde{P}_u$ and $\widehat{P}$ based on samples from (Clark, Cowlitz), (Pierce, King) and (Skagit, Whatcom). t_1 and t_2 represent Clark and Cowlitz, respectively. Point estimates, 95% CIs and lengths of CIs are computed for the selected threshold probabilities. "-" indicates that some of the CIs generated can not be constructed since the estimates are 0. 54

2.1 Coverage rates of CIs for β_1 and β_2 in Case 1 and 2. 105
2.2 Coverage rates of CIs for β_1 and β_2 in Case 3-6. A hyphen "-" indicates that the corresponding parameter is not in the model. 107
2.3 Coverage rates of CIs for β_1 β_2 and β_3 in Case 7 and 8. 108
2.4 Coverage rates of CIs for β_1 and β_2 in Case 9 and 10. 109
2.5 Coverage rates of CIs for threshold probability $G(T)$ with $T = 3, 4, 5, 6, 7$. 111
2.6 Coverage rates of CIs for the reference threshold probabilities $P_1 = P(X_{t-1}^{(0)} \leq 1, X_t^{(0)} > 1)$, $P_2 = P(X_{t-1}^{(0)} \leq 0, X_t^{(0)} \leq 1)$, $P_3 = P(X_{t-1}^{(0)} > 1, X_t^{(0)} \leq -1)$, $P_4 = P(X_{t-1}^{(0)} > 2, X_t^{(0)} > 1)$ and $P_5 = P(X_{t-1}^{(0)} > 0, X_t^{(0)} \leq 0)$. 113

2.7 Rejection rates of tests for equality of distributions based on DRM, DRMWD and K-S test. For Case 1, Case 2 and the univariate case in Case 3, the rejection rate represents the estimated Type-I error. For the bivariate case in Case 3 and Case 4, the rejection rate represents the estimated power. 117
2.8 The information of selected sites. 117
2.9 The result of the test for equality of distributions of the two time periods. Each entry indicates whether the corresponding null hypothesis is rejected or not at the significance level of 0.05. 119
2.10 Estimates of selected reference threshold probabilities and the corresponding 95% CIs. 119

3.1 The models and tests used for the 11 simulated cases. 151
3.2 Rejection rates of 11 simulated cases. For Case 1, 2, 5, 6, 11, the rejection rate represents the estimated Type-I error. For Case 3, 4, 7, 8, 9, 10, the rejection rate represents the estimated power. 152

List of Figures

1.1 Histograms of Beaver residential radon concentration in six periods. 44
1.2 Histograms of Beaver residential radon concentration in six periods truncated at 40 pCi/L. 45
1.3 Histograms of residential radon concentration in Beaver, Lawrence, Butler, Allegheny and Washington. 48
1.4 Histograms of residential radon concentration in Beaver, Lawrence, Butler, Allegheny and Washington truncated at 40 pCi/L. 49
1.5 Histograms and truncated histograms (truncated at 40 pCi/L) of residential radon concentration in Forest, Warren and Elk. 51
1.6 Histograms of pertussis rates in Clark, Cowlitz, Pierce, King, Skagit and Whatcom. 53

3.1 Test result of $\{X_{H,i}(t)\}$, $i = 1, 2, 3$. Blue indicates significance. 155
3.2 Test result of $\{X_{P,i}(t)\}$, $i = 1, 2, 3$. Red indicates significance. 155
3.3 Colored blocks obtained by adding up the result of Figure 3.1 and 3.2. 155
3.6 GRC_1 of subregion 2-18. The vacancy indicates the corresponding subregion is insignificant or has already been included as an input. 157
3.7 GRC_2 of subregion 2-18. The vacancy indicates the corresponding subregion is insignificant or has already been included as an input. 157
3.8 GRC_3 of subregion 2-18. The vacancy indicates the corresponding subregion is insignificant or has already been included as an input. 157
3.9 GRC_1 of subregion 2-18. The vacancy indicates the corresponding subregion is insignificant or has already been included as an input. 158
3.10 GRC_2 of subregion 2-18. The vacancy indicates the corresponding subregion is insignificant or has already been included as an input. 158
3.4 Colored blocks based on testing for linear association. 159
3.5 Colored blocks based on testing for lagged processes with lags $u = 0, \ldots, 4$. . . . 160

List of Abbreviations

AIC	Akaike Information Criterion
ARMA	Autoregressive Moving Average
BOLD	Blood oxygen level dependent
CDF	Cumulative distribution function
CI	Confidence interval
CLT	Central Limit Theorem
DRM	Density Ratio Model
DRMWD	Density Ratio Model with weakly dependent data
EPA	Environmental Protection Agency
GLM	Generalized Linear Model
GRC	Generalized Residual Coherence
IID	Independent and identically distributed
K-S	Kolmogorov-Smirnov
MC	Monte Carlo
MIAE	Mean integrated absolute error
MISE	Mean integrated squared errror
NO_2	Nitrogen dioxide
OLS	Ordinary Least Squares
OSF	Out of Sample Fusion
PA	Pennsylvania
pCi/L	Picocuries per liter
PDF	Probability density function

ROSF Repeated Out of Sample Fusion

SLLN Strong Law of Large Number

Chapter 1: Density Ratio Model with Variable Tilts

1.1 Introduction

The ideal of DRM is to combine independent random samples based on the ratios of their densities and make inferences based on the combined sample.

An early DRM work can be traced back to [59] in the context of biased sampling. Two independent random samples, $\boldsymbol{X}^{(0)} = (X_1^{(0)}, \cdots, X_{n_0}^{(0)}) \sim G$ and $\boldsymbol{X}^{(1)} = (X_1^{(1)}, \cdots, X_{n_1}^{(1)}) \sim G_1$, satisfy

$$dG_1(x) = \frac{x}{\int_{-\infty}^{\infty} y dG(y)} dG(x),$$

and G, G_1 are estimated by maximizing the nonparametric likelihood

$$L(G) = \prod_{i=1}^{n} p_i \prod_{j=1}^{n_1} \frac{X_j^{(1)}}{\sum_{i=1}^{n} t_i p_i}$$

with constraint $\sum_{i=1}^{n} p_i = 1$ where $\boldsymbol{t} = (\boldsymbol{X}^{(0)}, \boldsymbol{X}^{(1)})$, $n = n_0 + n_1$ and $p_i = dG(t_i)$.

[60] extended this model to $m+1$ samples with known nonnegative weight functions $w_k(\cdot)$, $k = 1, \cdots, m$, such that

$$dG_k(x) = \frac{w_k(x)}{\int_{-\infty}^{\infty} w_k(y) dG(y)} dG(x),$$

and the corresponding nonparametric likelihood is

$$L(G) = \prod_{i=1}^{n} p_i \prod_{k=1}^{m} \prod_{j=1}^{n_k} \frac{w_k(X_j^{(k)})}{\sum_{i=1}^{n} w_k(t_i) p_i}$$

with the constraint with constraint $\sum_{i=1}^{n} p_i = 1$. The consistency and the asymptotic results are established in [23].

A semiparametric estimator of $g_k = dG_k$ based on the density ratio structure is proposed by [14] such that

$$\widetilde{g}_k(x) = \widetilde{g}(x) \exp\left(\widehat{\beta}_0 + \widehat{\boldsymbol{\beta}}_1^\top \boldsymbol{T}(x)\right),$$

where g and g_k's are the densities from an exponential family, $\widetilde{g}$ is the kernel density estimator of g, $\widehat{\beta}_0$ and $\widehat{\boldsymbol{\beta}}_1$ are the maximum likelihood estimators and $\boldsymbol{T}(x)$ is the sufficient statistic. Therefore, $\widetilde{g}_k$ consists of both parametric estimators and nonparametric estimators. This result shows a vital point that exponential family satisfies the density ratio structure automatically.

[54] developed the two-sample DRM in the case-control study starting from the standard logistic regression model

$$P(Y = 1|\boldsymbol{x}) = \frac{\exp\left(\alpha^* + \boldsymbol{\beta}^\top \boldsymbol{x}\right)}{1 + \exp\left(\alpha^* + \boldsymbol{\beta}^\top \boldsymbol{x}\right)}.$$

Based on this logistic regression model,

$$\begin{aligned}\frac{dF(\boldsymbol{x}|Y = 1)}{dF(\boldsymbol{x}|Y = 0)} &= \frac{P(Y = 0)}{P(Y = 1)} \exp\left(\alpha^* + \boldsymbol{\beta}^\top \boldsymbol{x}\right) \\ &= \exp\left(\alpha + \boldsymbol{\beta}^\top \boldsymbol{x}\right)\end{aligned}$$

for $\alpha = \alpha^* + \log P(Y = 0) - \log P(Y = 1)$. If we consider the observations of the covariates can be categorized based on the value of Y, then we have two independent random samples $\boldsymbol{X}^{(0)} \sim G(\boldsymbol{x}) = F(\boldsymbol{x}|Y = 1)$ and $\boldsymbol{X}^{(1)} \sim G_1(\boldsymbol{x}) = F(\boldsymbol{x}|Y = 0)$ satisfying

$$dG_1(\boldsymbol{x}) = \exp{(\alpha + \boldsymbol{\beta}^\top \boldsymbol{x})} dG(\boldsymbol{x}),$$

and G, $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta})^\top$ can be estimated by maximizing the empirical likelihood

$$L(\boldsymbol{\theta}, G) = \prod_{i=1}^{n} p_i \prod_{j=1}^{n_1} w(X_j^{(1)}; \boldsymbol{\theta}),$$

where $w(x; \boldsymbol{\theta}) = \exp{(\alpha + \boldsymbol{\beta}^\top \boldsymbol{x})}$ satisfies

$$\sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i [w(t_i; \boldsymbol{\theta}) - 1] = 0,$$

and the estimation procedure follows [53]. The asymptotic results of the estimators $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{G}$, and a goodness-of-fit test for the DRM are discussed in [54]. An alternative goodness-of-fit test based on the asymptotic normality of $\widetilde{G}$ is proposed by [72] and [76].

The model is extended with $w(x; \boldsymbol{\theta}) = \exp{(\alpha + \boldsymbol{\beta}^\top \boldsymbol{h}(x))}$ in [74]. A more general form of $w(x; \boldsymbol{\theta})$ is discussed in [22] and [73] and it is shown that the empirical likelihood estimator may not be attainable in a more general setting with $w(x; \boldsymbol{\theta}) = \exp{(\alpha + s(x; \boldsymbol{\beta}))}$ while $w(x; \boldsymbol{\theta}) = \exp{(\alpha + \beta h(x))}$ is satisfactory.

An extension is made for the multi-sample case with the likelihood

$$L(\boldsymbol{\theta}, G) = \prod_{i=1}^{n} p_i \prod_{k=1}^{m} \prod_{j=1}^{n_k} w_k(X_j^{(k)}; \boldsymbol{\theta})$$

subject to

$$\sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i[w_k(t_i; \boldsymbol{\theta}) - 1] = 0, \; k = 1, \ldots, m,$$

where $w_k(x; \boldsymbol{\theta}) = \exp\left(\alpha_k + \boldsymbol{\beta}_k^\top \boldsymbol{h}(x)\right)$ and $\boldsymbol{h}(x)$ is referred as the tilt function is made in [21] with the asymptotic results of $\widetilde{\boldsymbol{\theta}}$. A more detailed dicussion of the asymptotic results of $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{G}$ can be found in [47] and [32]. If $\boldsymbol{\theta}$ is not constrained by

$$\sum_{i=1}^{n} p_i[w_k(t_i; \boldsymbol{\theta}) - 1] = 0 \; k = 1, \ldots, m,$$

then the likelihood can be considered as

$$L(\boldsymbol{\theta}, G) = \prod_{i=1}^{n} p_i \prod_{k=1}^{m} \prod_{j=1}^{n_k} \frac{w_k(X_j^{(k)}; \boldsymbol{\theta})}{\int w_k(\boldsymbol{y}; \boldsymbol{\theta}) dG(\boldsymbol{y})}$$

subject to $\sum_{i=1}^{n} p_i = 1$ only, which resembles (2.2) in [22]. The extension to the multivariate case and a graphical goodness-of-fit checking method are established in [62] and [63]. In such case, $\boldsymbol{h}(\boldsymbol{x})$ is considered as a multivariate function.

Additionally, various related research works have been done, and are briefly summarized here. A kernel density estimator $\widetilde{g}$ based on the $\widetilde{G}$ is examined in [18], and a goodness-of-fit test was developed by [68] based on the Hellinger distance between kernel density estimators $\widetilde{g}$ and $\widehat{g}$, where $\widehat{g}$ is the kernel density estimator obtained based on the empirical distribution $\widehat{G}$. The effect of misspecification of tilt functions is discussed in [19] and the selection of tilt functions from the Box-Cox family based on information criterions is dicussed in [20]. The analysis of the quantile process $\sqrt{n}(\widetilde{G}^{-1} - G^{-1})$ can be found in [75] and [7]. A more recent examination of quantile tests and efficiency can be found in [71] and [69]. Also, a data-adaptive method of determining the tilt functions using functional principal component analysis is proposed by [70] and [69]. [11] applied the DRM to the spectral distribution to measure extremal denpendence. A Bayesian extension of the DRM can be found in [12] and [32]. In [66] and [41], the DRM is used as a clustering method to detect homogeneous distributions.

It is worth emphasizing that a series of research works related to DRM has been conducted for the estimation of small tail probability. When the reference sample does not contain any observation that exceeds a certain threshold T but the interval estimation of $1 - G(T)$ is of interest, [81] proposed to combine the reference sample with an artificial sample that contains observations that exceed T, and use DRM to produce a confidence interval (CI) for $1 - G(T)$. This method is referred as Out of Sample Fusion (OSF). [30] applied this method to the mortality surveillance. [37] considered a Monte Carlo (MC) method that performs OSF iteratively. Such method is referred as Repeated Out of Sample Fusion (ROSF). [51] and [38] constructed a CI for $1 - G(T)$ based on the maximal upper bounds of CIs obtained by ROSF. An algorithm based on all upper bounds of ROSF CIs is proposed in [65] and further examined by [36], [39] and [40].

The extension made in this chapter is that $w_k(\boldsymbol{x};\boldsymbol{\theta}) = \exp\left(\alpha_k + \boldsymbol{\beta}_k^\top \boldsymbol{h}_k(\boldsymbol{x})\right)$ is defined with different $\boldsymbol{h}_k(\boldsymbol{x})$'s instead of just one uniformly fixed tilt $\boldsymbol{h}(\boldsymbol{x})$. $\boldsymbol{h}_k(\boldsymbol{x})$'s are referred as variable tilts. In Section 1.2, a DRM with variable tilts is constructed with estimation procedure and asymptotic result provided. In Section 1.3, a simulation is performed to compare the estimators of the reference CDF obtained by DRM with variable tilts, DRM with a uniform tilt and empirical distribution, respectively. In Section 1.4, two methods of selecting variable tilts are introduced of which one is based on asymptotic test and the other is based on AIC. In Section 1.5, the DRM with variable tilts is applied to estimating residential radon concentration in Pennsylvania counties. In these two sections, two tilt selection methods are demonstrated. In Section 1.6, the regional pertussis rates in Washington State are estimated by this extended DRM, which shows the application of this model to a multivariate case.

1.2 A Density Ratio Model with Variable Tilts

Let

$$\boldsymbol{X}^{(k)} = (\boldsymbol{X}_1^{(k)}, \ldots, \boldsymbol{X}_{n_k}^{(k)}), \;\; k = 0, \ldots, m,$$

be $m+1$ independent random samples with size $n_0, \ldots, n_m$, respectively. Each observation is a d-dimensional random vector such that

$$\boldsymbol{X}_j^{(k)} = (X_{j,1}^{(k)}, \ldots, X_{j,d}^{(k)})^\top$$

for $j = 1, \ldots, n_k$, $k = 0, \ldots, m$. Let $\boldsymbol{X}^{(0)}$ denotes the reference sample. Combine the data and denote the combined sample as

$$\boldsymbol{t} = (\boldsymbol{X}^{(0)}, \ldots, \boldsymbol{X}^{(m)})$$

with size $n = \sum_{k=0}^{m} n_k$. Suppose $\boldsymbol{X}^{(k)}$ comes from a population with a PDF g_k for $k = 0, \ldots, m$, and they have the density ratio structure as

$$\frac{g_k(\boldsymbol{x})}{g_0(\boldsymbol{x})} = \exp\left(\alpha_k + \boldsymbol{\beta}_k^\top \boldsymbol{h}_k(\boldsymbol{x})\right), \;\; k = 1, \ldots, m. \tag{1.1}$$

Denote $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)^\top$, $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kp_k})^\top$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_m^\top)^\top$, $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$ and $\boldsymbol{h}_k(\cdot) = (h_{k1}(\cdot), \ldots, h_{kp_k}(\cdot))^\top$. Denote the reference cumulative distribution function (CDF) as G and let $w_0(\cdot; \boldsymbol{\theta}) = 1$, $w_k(\cdot; \boldsymbol{\theta}) = \exp(\alpha_k + \boldsymbol{\beta}_k^\top \boldsymbol{h}_k(\cdot))$ for $k = 1, \ldots, m$. We make the following assumptions for the model and tilt functions.

Assumption 1.1. $|h_{lp}(\boldsymbol{X}_j^{(k)})|$ *satisfies*

$$E\left|h_{lp}(\boldsymbol{X}_j^{(k)})\right|^3 < \infty$$

for $k = 0, \ldots, m$, $l = 1, \ldots, m$ *and* $p = 1, \ldots, p_l$.

Assumption 1.2. *The DRM (1.1) is non-degenerate, that is, all parameters are non-zero and tilt functions are not linearly dependent.*

1.2.1 Estimation

Denote all observations in the combined sample as $\boldsymbol{t} = (\boldsymbol{t}_1, ..., \boldsymbol{t}_n)$. Let $p_i = dG(\boldsymbol{t}_i)$ for $i = 1, \ldots, n$. We obtain the estimators $\widetilde{\boldsymbol{\alpha}}$, $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{p}_i$'s by maximizing

$$L(\boldsymbol{\theta}, G) = \prod_{i=1}^{n} p_i \prod_{k=1}^{m} \prod_{j=1}^{n_k} w_k(\boldsymbol{X}_j^{(k)}; \boldsymbol{\theta}) \tag{1.2}$$

subject to

$$\sum_{i=1}^{n} p_i = 1, \qquad \sum_{i=1}^{n} p_i[w_k(\boldsymbol{t}_i; \boldsymbol{\theta}) - 1] = 0, \qquad k = 1, \ldots, m. \tag{1.3}$$

Then the objective function with Lagrange multipliers $\lambda_0, \ldots, \lambda_m$ becomes

$$\log L(\boldsymbol{\theta}, G) - \lambda_0(1 - \sum_{i=1}^{n} p_i) - \cdots - \lambda_m \sum_{i=1}^{n} p_i[w_m(\boldsymbol{t}_i; \boldsymbol{\theta}) - 1].$$

Differentiate the objective function with respect to p_i for $i = 1, \ldots, n$, and obtain

$$1 + \lambda_0 p_i - \lambda_1 p_i[w_1(\boldsymbol{t}_i; \boldsymbol{\theta}) - 1] - \cdots - \lambda_m p_i[w_m(\boldsymbol{t}_i; \boldsymbol{\theta}) - 1] = 0. \tag{1.4}$$

Sum over i, then $\lambda_0 = -n$. Replace λ_0 with $-n$ in (1.4),

$$p_i = \frac{1}{n + \lambda_1[w_1(\boldsymbol{t}_i; \boldsymbol{\theta}) - 1] + \cdots + \lambda_m[w_m(\boldsymbol{t}_i; \boldsymbol{\theta}) - 1]} \tag{1.5}$$

for $i = 1, \ldots, n$. Substitute (1.5) into $\log L(\boldsymbol{\theta}, G)$,

$$\begin{aligned}\log L(\boldsymbol{\theta}, G) = &- \sum_{i=1}^{n} \log\left(n + \lambda_1[w_1(\boldsymbol{t}_i; \boldsymbol{\theta}) - 1] + \cdots + \lambda_m[w_m(\boldsymbol{t}_i; \boldsymbol{\theta}) - 1]\right) \\ &+ \sum_{k=1}^{m}(n_k \alpha_k + \sum_{j=1}^{n_k} \boldsymbol{\beta}_k^\top \boldsymbol{h}_k(\boldsymbol{X}_j^{(k)})).\end{aligned} \tag{1.6}$$

Differentiate $\log L(\boldsymbol{\theta}, G)$ with respect to α_k for $k = 1, \ldots, m$,

$$\begin{aligned}\frac{\partial \log L(\boldsymbol{\theta}, G)}{\partial \alpha_k} &= - \sum_{i=1}^{n} \frac{\lambda_k w_k(\boldsymbol{t}_i; \boldsymbol{\theta})}{n + \lambda_1[w_1(\boldsymbol{t}_i; \boldsymbol{\theta}) - 1] + \cdots + \lambda_m[w_m(\boldsymbol{t}_i; \boldsymbol{\theta}) - 1]} + n_k \\ &= - \sum_{i=1}^{n} \lambda_k p_i w_k(\boldsymbol{t}_i; \boldsymbol{\theta}) + n_k \\ &= -\lambda_k + n_k.\end{aligned}$$

Therefore, $\lambda_k = n_k$ is obtained by equating $\frac{\partial \log L(\boldsymbol{\theta}, G)}{\partial \alpha_k} = 0$ for $k = 1, \ldots, m$. Substitute λ_k's into (1.5) and denote $\rho_k = \frac{n_k}{n_0}$ for $k = 0, \ldots, m$, p_i has the expression

$$p_i = \frac{1}{n_0 \sum_{k=0}^{m} \rho_k w_k(\boldsymbol{t}_i; \boldsymbol{\theta})}$$

Substitute p_i's into (1.6), $\log L(\boldsymbol{\theta}, G)$ reduces to a function of $\boldsymbol{\theta}$ only.

$$\begin{aligned} l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}, G) = &-\sum_{i=1}^{n} \log\left(n_0 \sum_{k=0}^{m} \rho_k w_k(\boldsymbol{t}_i; \boldsymbol{\theta})\right) \\ &+ \sum_{k=1}^{m} \left(n_k \alpha_k + \sum_{j=1}^{n_k} \boldsymbol{\beta}_k^\top \boldsymbol{h}_k(\boldsymbol{X}_j^{(k)})\right). \end{aligned}$$

Then the estimators $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\alpha}}^\top, \widetilde{\boldsymbol{\beta}}^\top)^\top$ are obtained via solving the system of equations

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \alpha_k} &= -\sum_{i=1}^{n} \frac{\rho_k w_k(\boldsymbol{t}_i; \boldsymbol{\theta})}{\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{t}_i; \boldsymbol{\theta})} + n_k = 0, \\ \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_k} &= -\sum_{i=1}^{n} \frac{\rho_k w_k(\boldsymbol{t}_i; \boldsymbol{\theta}) \boldsymbol{h}_k(\boldsymbol{t}_i)}{\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{t}_i; \boldsymbol{\theta})} + \sum_{j=1}^{n_k} \boldsymbol{h}_k(\boldsymbol{X}_j^{(k)}) = \boldsymbol{0}. \end{aligned}$$

Subsequently, the estimators $\tilde{p}_i$ is obtained for $i = 1, \ldots, n$ such that

$$\widetilde{p}_i = \frac{1}{n_0 \sum_{k=0}^{m} \rho_k w_k(\boldsymbol{t}_i; \widetilde{\boldsymbol{\theta}})}.$$

Thus, the estimator of the reference CDF G is

$$\widetilde{G}(\boldsymbol{x}; \widetilde{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \widetilde{p}_i I_{\prod_{j=1}^{d}(-\infty, x_d]}(\boldsymbol{t}_i) \tag{1.7}$$

where

$$\boldsymbol{t}_i = (t_{i1}, \ldots, t_{id})^\top$$

when t_i has dimension d, and

$$I_{\prod_{j=1}^{d}(-\infty, x_d]}(\boldsymbol{t}_i) = I[t_{i1} \leq x_1, \ldots, t_{id} \leq x_d].$$

1.2.2 Asymptotic Behavior of $\widetilde{\boldsymbol{\theta}}$

In this section, we establish the consistency and asymptotic normality of $\widetilde{\boldsymbol{\theta}}$. First, the strong consistency of score and hessian is established. Next, we obtain the asymptotic normality of the score, and we use it together with the negative definiteness of the hessian to obtain the strong consistency and asymptotic normality of $\widetilde{\boldsymbol{\theta}}$.

1.2.2.1 Strong Consistency of Score and Hessian

Consider that ρ_k's are fixed as $n \to \infty$. Let $\boldsymbol{\theta}_0$ be the true parameter vector. The expectations of the first order derivatives for $k = 1, \ldots, m$, are obtained by

$$\begin{aligned} E\left(\left.\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) &= -\sum_{i=1}^{n} E\left(\frac{\rho_k w_k(\boldsymbol{t}_i; \boldsymbol{\theta}_0)}{\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{t}_i; \boldsymbol{\theta}_0)} \right) + n_k \\ &= -\sum_{h=0}^{m} n_h \int \frac{\rho_k w_k(\boldsymbol{y}; \boldsymbol{\theta}_0)}{\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{y}; \boldsymbol{\theta}_0)} w_h(\boldsymbol{y}; \boldsymbol{\theta}_0) dG(\boldsymbol{y}) + n_k \\ &= 0, \end{aligned}$$

$$
\begin{aligned}
E\left(\left.\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) &= -\sum_{i=1}^{n} E\left(\frac{\rho_k w_k(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{t}_i)}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta}_0)}\right) + \sum_{j=1}^{n_k} E\boldsymbol{h}_k(\boldsymbol{X}_j^{(k)}) \\
&= -\sum_{h=0}^{m} n_h \int \frac{\rho_k w_k(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{y})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)} w_h(\boldsymbol{y};\boldsymbol{\theta}_0)dG(\boldsymbol{y}) \\
&\quad + n_k \int w_k(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{y})dG(\boldsymbol{y}) \\
&= \boldsymbol{0}.
\end{aligned}
$$

By Strong Law of Large Number (SLLN) and Assumption 1.1, for $k = 1, \ldots, m$,

$$
\begin{aligned}
\frac{1}{n}\left.\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= -\frac{n_0}{n}\sum_{h=0}^{m}\rho_h \frac{1}{n_h}\sum_{l=1}^{n_h}\frac{\rho_k w_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)} + \frac{\rho_k}{\sum_{j=0}^{m}\rho_j} \\
&\overset{a.s.}{\to} -\frac{1}{\sum_{j=0}^{m}\rho_j}\sum_{h=0}^{m}\rho_h \int \frac{\rho_k w_k(\boldsymbol{y};\boldsymbol{\theta}_0)}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)} w_h(\boldsymbol{y};\boldsymbol{\theta}_0)dG(\boldsymbol{y}) + \frac{\rho_k}{\sum_{j=0}^{m}\rho_j} \\
&= 0, \\
\frac{1}{n}\left.\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= -\frac{n_0}{n}\sum_{h=0}^{m}\rho_h \frac{1}{n_h}\sum_{l=1}^{n_h}\frac{\rho_k w_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{X}_l^{(h)})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)} + \frac{n_k}{n}\frac{1}{n_k}\sum_{j=1}^{n_k}\boldsymbol{h}_k(\boldsymbol{X}_j^{(k)}) \\
&\overset{a.s.}{\to} -\frac{1}{\sum_{j=0}^{m}\rho_j}\sum_{h=0}^{m}\rho_h \int \frac{\rho_k w_k(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{y})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)} w_h(\boldsymbol{y};\boldsymbol{\theta}_0)dG(\boldsymbol{y}) \\
&\quad + \frac{\rho_k}{\sum_{j=0}^{m}\rho_j}\int w_k(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{y})dG(\boldsymbol{y}) \\
&= \boldsymbol{0}.
\end{aligned}
$$

By SLLN and Assumption 1.1, for $k, k' = 1, \ldots, m$, $k \neq k'$, as $n \to \infty$,

$$
\begin{aligned}
&-\frac{1}{n}\left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \alpha_k^2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
=&\frac{1}{n}\sum_{i=1}^{n}\frac{\left[\sum_{j\neq k,j=0}^{m}\rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\right]\rho_k w_k(\boldsymbol{t}_i;\boldsymbol{\theta}_0)}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\right]^2} \\
=&\sum_{h=0}^{m}\frac{n_h}{n}\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\frac{\left[\sum_{j\neq k,j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\right]\rho_k w_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\right]^2}\right) \\
\xrightarrow{a.s.}&\frac{\rho_k}{\sum_{j=0}^{m}\rho_j}-\frac{\rho_k^2}{\sum_{j=0}^{m}\rho_j}\int\frac{w_k^2(\boldsymbol{y};\boldsymbol{\theta}_0)}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)}dG(\boldsymbol{y}), \\
&-\frac{1}{n}\left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \alpha_k\partial\alpha_{k'}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
=&-\frac{1}{n}\sum_{i=1}^{n}\frac{\rho_k w_k(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\rho_{k'}w_{k'}(\boldsymbol{t}_i;\boldsymbol{\theta}_0)}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\right]^2} \\
=&-\sum_{h=0}^{m}\frac{n_h}{n}\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\frac{\rho_k w_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\rho_{k'}w_{k'}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\right]^2}\right) \\
\xrightarrow{a.s.}&-\frac{\rho_k\rho_{k'}}{\sum_{j=0}^{m}\rho_j}\int\frac{w_k(\boldsymbol{y};\boldsymbol{\theta}_0)w_{k'}(\boldsymbol{y};\boldsymbol{\theta}_0)}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)}dG(\boldsymbol{y}), \\
&-\frac{1}{n}\left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \alpha_k\partial\boldsymbol{\beta}_k^\top}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
=&\frac{1}{n}\sum_{i=1}^{n}\frac{\left[\sum_{j\neq k,j=0}^{m}\rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\right]\rho_k w_k(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\boldsymbol{h}_k^\top(\boldsymbol{t}_i)}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\right]^2} \\
=&\sum_{h=0}^{m}\frac{n_h}{n}\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\frac{\left[\sum_{j\neq k,j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\right]\rho_k w_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\boldsymbol{h}_k^\top(\boldsymbol{X}_l^{(h)})}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\right]^2}\right) \\
\xrightarrow{a.s.}&\frac{\rho_k}{\sum_{j=0}^{m}\rho_j}\int w_k(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{h}_k^\top(\boldsymbol{y})dG(\boldsymbol{y})-\frac{\rho_k^2}{\sum_{j=0}^{m}\rho_j}\int\frac{w_k^2(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{h}_k^\top(\boldsymbol{y})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)}dG(\boldsymbol{y}),
\end{aligned}
$$

$$
\begin{aligned}
& -\frac{1}{n}\left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \alpha_k \partial \boldsymbol{\beta}_{k'}^{\top}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
=& -\frac{1}{n}\sum_{i=1}^{n}\frac{\rho_k w_k(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\rho_{k'} w_{k'}(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\boldsymbol{h}_{k'}^{\top}(\boldsymbol{t}_i)}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\right]^2} \\
=& -\sum_{h=0}^{m}\frac{n_h}{n}\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\frac{\rho_k w_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\rho_{k'} w_{k'}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\boldsymbol{h}_{k'}^{\top}(\boldsymbol{X}_l^{(h)})}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\right]^2}\right) \\
\overset{a.s.}{\to}& -\frac{\rho_k\rho_{k'}}{\sum_{j=0}^{m}\rho_j}\int\frac{w_k(\boldsymbol{y};\boldsymbol{\theta}_0)w_{k'}(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{h}_{k'}^{\top}(\boldsymbol{y})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)}dG(\boldsymbol{y}), \\
& -\frac{1}{n}\left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_k \partial \boldsymbol{\beta}_{k}^{\top}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
=& \frac{1}{n}\sum_{i=1}^{n}\frac{\left[\sum_{j\neq k,j=0}^{m}\rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\right]\rho_k w_k(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{t}_i)\boldsymbol{h}_{k}^{\top}(\boldsymbol{t}_i)}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\right]^2} \\
=& \sum_{h=0}^{m}\frac{n_h}{n}\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\frac{\left[\sum_{j\neq k,j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\right]\rho_k w_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{X}_l^{(h)})\boldsymbol{h}_{k}^{\top}(\boldsymbol{X}_l^{(h)})}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\right]^2}\right) \\
\overset{a.s.}{\to}& \frac{\rho_k}{\sum_{j=0}^{m}\rho_j}\int w_k(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{y})\boldsymbol{h}_{k}^{\top}(\boldsymbol{y})dG(\boldsymbol{y})-\frac{\rho_k^2}{\sum_{j=0}^{m}\rho_j}\int\frac{w_k^2(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{y})\boldsymbol{h}_{k}^{\top}(\boldsymbol{y})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)}dG(\boldsymbol{y}), \\
& -\frac{1}{n}\left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_k \partial \boldsymbol{\beta}_{k'}^{\top}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
=& -\frac{1}{n}\sum_{i=1}^{n}\frac{\rho_k w_k(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\rho_{k'} w_{k'}(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{t}_i)\boldsymbol{h}_{k'}^{\top}(\boldsymbol{t}_i)}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\right]^2} \\
=& -\sum_{h=0}^{m}\frac{n_h}{n}\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\frac{\rho_k w_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\rho_{k'} w_{k'}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{X}_l^{(h)})\boldsymbol{h}_{k'}^{\top}(\boldsymbol{X}_l^{(h)})}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\right]^2}\right) \\
\overset{a.s.}{\to}& -\frac{\rho_k\rho_{k'}}{\sum_{j=0}^{m}\rho_j}\int\frac{w_k(\boldsymbol{y};\boldsymbol{\theta}_0)w_{k'}(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{y})\boldsymbol{h}_{k'}^{\top}(\boldsymbol{y})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)}dG(\boldsymbol{y}).
\end{aligned}
$$

Define the following quantities for $k, k' = 1, \ldots, m$,

$$A_{kk'}(\boldsymbol{\theta}) = \int \frac{w_k(\boldsymbol{y};\boldsymbol{\theta})w_{k'}(\boldsymbol{y};\boldsymbol{\theta})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{y};\boldsymbol{\theta})} dG(\boldsymbol{y}),$$

$$\boldsymbol{B}_{kk'}(\boldsymbol{\theta}) = \int \frac{w_k(\boldsymbol{y};\boldsymbol{\theta})w_{k'}(\boldsymbol{y};\boldsymbol{\theta})\boldsymbol{h}_{k'}^{\top}(\boldsymbol{y})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{y};\boldsymbol{\theta})} dG(\boldsymbol{y}),$$

$$\boldsymbol{C}_{kk'}(\boldsymbol{\theta}) = \int \frac{w_k(\boldsymbol{y};\boldsymbol{\theta})w_{k'}(\boldsymbol{y};\boldsymbol{\theta})\boldsymbol{h}_k(\boldsymbol{y})\boldsymbol{h}_{k'}^{\top}(\boldsymbol{y})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{y};\boldsymbol{\theta})} dG(\boldsymbol{y}),$$

$$\boldsymbol{E}_k(\boldsymbol{\theta}) = \int w_k(\boldsymbol{y};\boldsymbol{\theta})\boldsymbol{h}_k(\boldsymbol{y}) dG(\boldsymbol{y}),$$

$$\overline{\boldsymbol{E}}_k(\boldsymbol{\theta}) = \int w_k(\boldsymbol{y};\boldsymbol{\theta})\boldsymbol{h}_k(\boldsymbol{y})\boldsymbol{h}_k^{\top}(\boldsymbol{y}) dG(\boldsymbol{y}),$$

$$\boldsymbol{V}_k(\boldsymbol{\theta}) = \overline{\boldsymbol{E}}_k(\boldsymbol{\theta}) - \boldsymbol{E}_k(\boldsymbol{\theta})\boldsymbol{E}_k^{\top}(\boldsymbol{\theta}),$$

$$\boldsymbol{A}(\boldsymbol{\theta}) = (A_{ij}(\boldsymbol{\theta}))_{m\times m}, \qquad \boldsymbol{B}(\boldsymbol{\theta}) = (\boldsymbol{B}_{ij}(\boldsymbol{\theta}))_{m\times r}, \qquad \boldsymbol{C}(\boldsymbol{\theta}) = (\boldsymbol{C}_{ij}(\boldsymbol{\theta}))_{r\times r},$$

$$\boldsymbol{\rho} = \begin{bmatrix} \rho_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \rho_m \end{bmatrix}_{m\times m}, \qquad \overline{\boldsymbol{\rho}} = \begin{bmatrix} \rho_1\boldsymbol{I}_{r_1} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \rho_m\boldsymbol{I}_{r_m} \end{bmatrix}_{r\times r},$$

$$\boldsymbol{E}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{E}_1(\boldsymbol{\theta}) & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{E}_m(\boldsymbol{\theta}) \end{bmatrix}_{r\times m}, \qquad \overline{\boldsymbol{E}}(\boldsymbol{\theta}) = \begin{bmatrix} \overline{\boldsymbol{E}}_1(\boldsymbol{\theta}) & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \overline{\boldsymbol{E}}_m(\boldsymbol{\theta}), \end{bmatrix}_{r\times r}$$

$$\boldsymbol{V}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{V}_1(\boldsymbol{\theta}) & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{V}_m(\boldsymbol{\theta}) \end{bmatrix}_{r\times r},$$

where r_k is the dimension of $\boldsymbol{h}_k$, and $r = \sum_{k=1}^{m} r_k$. Therefore,

$$-\frac{1}{n}\left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \stackrel{a.s.}{\to} \boldsymbol{S}(\boldsymbol{\theta}_0) = \frac{1}{\sum_{k=0}^{m}\rho_k}\begin{bmatrix} \boldsymbol{S}_{11}(\boldsymbol{\theta}_0) & \boldsymbol{S}_{12}(\boldsymbol{\theta}_0) \\ \boldsymbol{S}_{12}^\top(\boldsymbol{\theta}_0) & \boldsymbol{S}_{22}(\boldsymbol{\theta}_0) \end{bmatrix}, \tag{1.8}$$

where

$$\begin{aligned}
\boldsymbol{S}_{11}(\boldsymbol{\theta}_0) &= \boldsymbol{\rho} - \boldsymbol{\rho}\boldsymbol{A}(\boldsymbol{\theta}_0)\boldsymbol{\rho}, \\
\boldsymbol{S}_{12}(\boldsymbol{\theta}_0) &= \boldsymbol{\rho}\boldsymbol{E}^\top(\boldsymbol{\theta}_0) - \boldsymbol{\rho}\boldsymbol{B}(\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}}, \\
\boldsymbol{S}_{22}(\boldsymbol{\theta}_0) &= \overline{\boldsymbol{\rho}}\overline{\boldsymbol{E}}(\boldsymbol{\theta}_0) - \overline{\boldsymbol{\rho}}\boldsymbol{C}(\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}}.
\end{aligned}$$

1.2.2.2 Asymptotic Normality of Score

Now, we obtain the expression of $\mathrm{Var}\left(\frac{1}{\sqrt{n}}\left.\frac{\partial l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)$. For $k, k' = 1, \ldots, m$ and $k \neq k'$,

$$\begin{aligned}
&\mathrm{Var}\left(\frac{1}{\sqrt{n}}\left.\frac{\partial l(\boldsymbol{\theta})}{\partial\alpha_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) \\
=&\frac{\rho_k^2}{\sum_{j=0}^{m}\rho_j}\left(A_{kk}(\boldsymbol{\theta}_0) - \sum_{j=1}^{m}\rho_j A_{kj}^2(\boldsymbol{\theta}_0) - \left(1 - \sum_{j=1}^{m}\rho_j A_{kj}(\boldsymbol{\theta}_0)\right)^2\right), \\
&\mathrm{Cov}\left(\frac{1}{\sqrt{n}}\left.\frac{\partial l(\boldsymbol{\theta})}{\partial\alpha_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \frac{1}{\sqrt{n}}\left.\frac{\partial l(\boldsymbol{\theta})}{\partial\alpha_{k'}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) \\
=&\frac{\rho_k\rho_{k'}}{\sum_{j=0}^{m}\rho_j}\left(A_{kk'}(\boldsymbol{\theta}_0) - \sum_{j=1}^{m}\rho_j A_{kj}(\boldsymbol{\theta}_0)A_{k'j}(\boldsymbol{\theta}_0) - \left(1 - \sum_{j=1}^{m}\rho_j A_{kj}(\boldsymbol{\theta}_0)\right)\left(1 - \sum_{j=1}^{m}\rho_j A_{k'j}(\boldsymbol{\theta}_0)\right)\right),
\end{aligned}$$

$$
\begin{aligned}
&\mathrm{Cov}\left(\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha_k}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_k}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\\
=&\frac{\rho_k^2}{\sum_{j=0}^m \rho_j}\left(A_{kk}(\boldsymbol{\theta}_0)\boldsymbol{E}_k(\boldsymbol{\theta}_0) - \sum_{j=1}^m \rho_j A_{kj}(\boldsymbol{\theta}_0)\boldsymbol{B}_{jk}^\top(\boldsymbol{\theta}_0)\right.\\
&\left.-\left(1-\sum_{j=1}^m \rho_j A_{kj}(\boldsymbol{\theta}_0)\right)\left(\boldsymbol{E}_k(\boldsymbol{\theta}_0) - \sum_{j=1}^m \rho_j \boldsymbol{B}_{jk}^\top(\boldsymbol{\theta}_0)\right)\right),\\
&\mathrm{Cov}\left(\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha_k}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_{k'}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\\
=&\frac{\rho_k\rho_{k'}}{\sum_{j=0}^m \rho_j}\left(A_{kk'}(\boldsymbol{\theta}_0)\boldsymbol{E}_{k'}(\boldsymbol{\theta}_0) - \sum_{j=1}^m \rho_j A_{kj}(\boldsymbol{\theta}_0)\boldsymbol{B}_{jk'}^\top(\boldsymbol{\theta}_0)\right.\\
&\left.-\left(1-\sum_{j=1}^m \rho_j A_{kj}(\boldsymbol{\theta}_0)\right)\left(\boldsymbol{E}_{k'}(\boldsymbol{\theta}_0) - \sum_{j=1}^m \rho_j \boldsymbol{B}_{jk'}^\top(\boldsymbol{\theta}_0)\right)\right),\\
&\mathrm{Var}\left(\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_k}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\\
=&\frac{\rho_k^2}{\sum_{j=0}^m \rho_j}\left(-\boldsymbol{C}_{kk}(\boldsymbol{\theta}_0) - \sum_{j=1}^m \rho_j \boldsymbol{B}_{jk}^\top(\boldsymbol{\theta}_0)\boldsymbol{B}_{jk}(\boldsymbol{\theta}_0) - \left(\boldsymbol{E}_k(\boldsymbol{\theta}_0) - \sum_{j=1}^m \rho_j \boldsymbol{B}_{jk}^\top(\boldsymbol{\theta}_0)\right)\right.\\
&\left.\times\left(\boldsymbol{E}_k^\top(\boldsymbol{\theta}_0) - \sum_{j=1}^m \rho_j \boldsymbol{B}_{jk}(\boldsymbol{\theta}_0)\right) + 2\boldsymbol{B}_{kk}^\top(\boldsymbol{\theta}_0)\boldsymbol{E}_k^\top(\boldsymbol{\theta}_0)\right) + \frac{\rho_k}{\sum_{j=0}^m \rho_j}\boldsymbol{V}_k(\boldsymbol{\theta}_0),\\
&\mathrm{Cov}\left(\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_k}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_{k'}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\\
=&\frac{\rho_k\rho_{k'}}{\sum_{j=0}^m \rho_j}\left(-\boldsymbol{C}_{kk'}(\boldsymbol{\theta}_0) - \sum_{j=1}^m \rho_j \boldsymbol{B}_{jk}^\top(\boldsymbol{\theta}_0)\boldsymbol{B}_{jk'}(\boldsymbol{\theta}_0) - \left(\boldsymbol{E}_k(\boldsymbol{\theta}_0) - \sum_{j=1}^m \rho_j \boldsymbol{B}_{jk}^\top(\boldsymbol{\theta}_0)\right)\right.\\
&\left.\times\left(\boldsymbol{E}_{k'}^\top(\boldsymbol{\theta}_0) - \sum_{j=1}^m \rho_j \boldsymbol{B}_{jk'}(\boldsymbol{\theta}_0)\right) + \boldsymbol{B}_{kk'}^\top(\boldsymbol{\theta}_0)\boldsymbol{E}_k^\top(\boldsymbol{\theta}_0) + \boldsymbol{B}_{kk'}^\top(\boldsymbol{\theta}_0)\boldsymbol{E}_{k'}^\top(\boldsymbol{\theta}_0)\right).
\end{aligned}
$$

Let $\Lambda(\boldsymbol{\theta}_0) = \mathrm{Var}\left(\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)$ and $\Lambda(\boldsymbol{\theta}_0) = \frac{1}{\sum_{k=0}^m \rho_k}\begin{bmatrix}\boldsymbol{\Lambda}_{11}(\boldsymbol{\theta}_0) & \boldsymbol{\Lambda}_{12}(\boldsymbol{\theta}_0)\\ \boldsymbol{\Lambda}_{12}^\top(\boldsymbol{\theta}_0) & \boldsymbol{\Lambda}_{22}(\boldsymbol{\theta}_0)\end{bmatrix}$, then

$$
\begin{aligned}
\boldsymbol{\Lambda}_{11}(\boldsymbol{\theta}_0) =&\boldsymbol{\rho}(\boldsymbol{A}(\boldsymbol{\theta}_0) - \boldsymbol{A}(\boldsymbol{\theta}_0)\boldsymbol{\rho}\boldsymbol{A}(\boldsymbol{\theta}_0) - (\boldsymbol{I}_m - \boldsymbol{A}(\boldsymbol{\theta}_0)\boldsymbol{\rho})\boldsymbol{J}_m(\boldsymbol{I}_m - \boldsymbol{\rho}\boldsymbol{A}(\boldsymbol{\theta}_0)))\boldsymbol{\rho} \\
=&\boldsymbol{S}_{11}(\boldsymbol{\theta}_0) - \boldsymbol{S}_{11}(\boldsymbol{\theta}_0)(\boldsymbol{J}_m + \boldsymbol{\rho}^{-1})\boldsymbol{S}_{11}(\boldsymbol{\theta}_0), \\
\boldsymbol{\Lambda}_{12}(\boldsymbol{\theta}_0) =&\boldsymbol{\rho}(\boldsymbol{A}(\boldsymbol{\theta}_0)\boldsymbol{E}^\top(\boldsymbol{\theta}_0) - \boldsymbol{A}(\boldsymbol{\theta}_0)\boldsymbol{\rho}\boldsymbol{B}(\boldsymbol{\theta}_0) - (\boldsymbol{I}_m - \boldsymbol{A}(\boldsymbol{\theta}_0)\boldsymbol{\rho})\boldsymbol{J}_m(\boldsymbol{E}^\top(\boldsymbol{\theta}_0) - \boldsymbol{\rho}\boldsymbol{B}(\boldsymbol{\theta}_0)))\overline{\boldsymbol{\rho}} \\
=&\boldsymbol{S}_{12}(\boldsymbol{\theta}_0) - \boldsymbol{S}_{11}(\boldsymbol{\theta}_0)(\boldsymbol{J}_m + \boldsymbol{\rho}^{-1})\boldsymbol{S}_{12}(\boldsymbol{\theta}_0), \\
\boldsymbol{\Lambda}_{22}(\boldsymbol{\theta}_0) =&\overline{\boldsymbol{\rho}}(-\boldsymbol{C}(\boldsymbol{\theta}_0) - \boldsymbol{B}^\top(\boldsymbol{\theta}_0)\boldsymbol{\rho}\boldsymbol{B}(\boldsymbol{\theta}_0) - (\boldsymbol{E}(\boldsymbol{\theta}_0) - \boldsymbol{B}^\top(\boldsymbol{\theta}_0)\boldsymbol{\rho})\boldsymbol{J}_m(\boldsymbol{E}^\top(\boldsymbol{\theta}_0) - \boldsymbol{\rho}\boldsymbol{B}(\boldsymbol{\theta}_0)) \\
&+ \boldsymbol{B}^\top(\boldsymbol{\theta}_0)\boldsymbol{E}^\top(\boldsymbol{\theta}_0) + \boldsymbol{E}(\boldsymbol{\theta}_0)\boldsymbol{B}(\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}} + \overline{\boldsymbol{\rho}}(\overline{\boldsymbol{E}}(\boldsymbol{\theta}_0) - \boldsymbol{E}(\boldsymbol{\theta}_0)\boldsymbol{E}^\top(\boldsymbol{\theta}_0)) \\
=&\boldsymbol{S}_{22}(\boldsymbol{\theta}_0) - \boldsymbol{S}_{12}^\top(\boldsymbol{\theta}_0)(\boldsymbol{J}_m + \boldsymbol{\rho}^{-1})\boldsymbol{S}_{12}(\boldsymbol{\theta}_0).
\end{aligned}
$$

By Central Limit Theorem (CLT),

$$
\frac{1}{\sqrt{n}} \left.\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Lambda}(\boldsymbol{\theta}_0)). \tag{1.9}
$$

1.2.2.3 Strong Consistency of $\widetilde{\boldsymbol{\theta}}$

Denote

$$
\begin{aligned}
l_1(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \log \sum_{k=0}^{m} \rho_k w_k(\boldsymbol{t}_i; \boldsymbol{\theta}), \\
l_2(\boldsymbol{\theta}) &= \sum_{k=1}^{m} (n_k \alpha_k + \sum_{j=1}^{n_k} \boldsymbol{\beta}_k^\top \boldsymbol{h}_k(\boldsymbol{X}_j^{(k)})),
\end{aligned}
$$

and they satisfy $l(\boldsymbol{\theta}) = -n \log n_0 - l_1(\boldsymbol{\theta}) + l_2(\boldsymbol{\theta})$. Moreover, $l_1(\boldsymbol{\theta})$ satisfies that

$$
\frac{\partial^2 l_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}. \tag{1.10}
$$

According to [52] and [47], the strong consistency of $\widetilde{\boldsymbol{\theta}}$ can be established based on the positive definiteness of $\frac{\partial^2 l_1(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$ when the density ratios have the form $w_k(\boldsymbol{x};\boldsymbol{\theta}) = \exp\left(\alpha_k + \boldsymbol{\beta}_k^\top \boldsymbol{x}\right)$ and $w_k(x;\boldsymbol{\theta}) = \exp\left(\alpha_k + \boldsymbol{\beta}_k^\top \boldsymbol{h}(x)\right)$, respectively. Both [52] and [47] claim the positive definiteness of $\frac{\partial^2 l_1(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$ can be guaranteed by the non-degeneracy of the DRM. Here, we propose a way to verify the positive definiteness of $\frac{\partial^2 l_1(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$. Since

$$l_1(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \sum_{k=0}^{m} \rho_k w_k(\boldsymbol{t}_i;\boldsymbol{\theta}),$$

it is sufficient to show that $\frac{\partial^2 l_{1,i}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$ is positive definite where

$$l_{1,i}(\boldsymbol{\theta}) = \log \sum_{k=0}^{m} \rho_k w_k(\boldsymbol{t}_i;\boldsymbol{\theta}).$$

Also, since this is for arbitrary i, we may simplify the problem as

$$l^*(\boldsymbol{x};\boldsymbol{\theta}) = \log \sum_{k=0}^{m} \rho_k w_k(\boldsymbol{x};\boldsymbol{\theta}),$$

and the notations as

$$l^*(\boldsymbol{x};\boldsymbol{\theta}) = l^*, \quad w_k(\boldsymbol{x};\boldsymbol{\theta}) = w_k, \quad h_{kp}(\boldsymbol{x}) = h_{kp},$$

for $k = 1, \ldots, m$, and $p = 1, \ldots, p_k$.

Let $\boldsymbol{M} = (M_0, \ldots, M_m)^\top$ be Multinomial$(1, (P_0, \ldots, P_m)^\top)$ where

$$P_k = \frac{\rho_k w_k}{\sum_{j=0}^{m} \rho_j w_j}.$$

Then denote $\boldsymbol{M}_{-1} = (M_1, \ldots, M_m)^\top$. Let $\boldsymbol{H}_k$ be a vector of independent variables such that

$$\boldsymbol{H}_k = (H_{k1}, \ldots, H_{kp_k})^\top, \;\; E\boldsymbol{H}_k = (h_{k1}, \ldots, h_{kp_k})^\top,$$

for $k = 1, \ldots, m$. Also, assume that $\boldsymbol{M}$ and $\boldsymbol{H}_k$'s are independent. Then

$$\begin{aligned}
\frac{\partial^2 l^*}{\partial \alpha_k^2} &= \frac{\left(\sum_{j\neq k, j=0}^{m} \rho_j w_j\right) \rho_k w_k}{\left(\sum_{j=0}^{m} \rho_j w_j\right)^2} = P_k(1-P_k) \\
\frac{\partial^2 l^*}{\partial \alpha_k \partial \alpha_{k'}} &= -\frac{\rho_k w_k \rho_{k'} w_{k'}}{\left(\sum_{j=0}^{m} \rho_j w_j\right)^2} = -P_k P_{k'} \\
\frac{\partial^2 l^*}{\partial \alpha_k \partial \boldsymbol{\beta}_k} &= \frac{\left(\sum_{j\neq k, j=0}^{m} \rho_j w_j\right) \rho_k w_k \boldsymbol{h}_k}{\left(\sum_{j=0}^{m} \rho_j w_j\right)^2} = P_k(1-P_k) E\boldsymbol{H}_k \\
\frac{\partial^2 l^*}{\partial \alpha_k \partial \boldsymbol{\beta}_{k'}} &= -\frac{\rho_k w_k \rho_{k'} w_{k'} \boldsymbol{h}_{k'}}{\left(\sum_{j=0}^{m} \rho_j w_j\right)^2} = -P_k P_{k'} E\boldsymbol{H}_{k'} \\
\frac{\partial^2 l^*}{\partial \boldsymbol{\beta}_k \partial \boldsymbol{\beta}_k^\top} &= \frac{\left(\sum_{j\neq k, j=0}^{m} \rho_j w_j\right) \rho_k w_k \boldsymbol{h}_k \boldsymbol{h}_k^\top}{\left(\sum_{j=0}^{m} \rho_j w_j\right)^2} = P_k(1-P_k) E\boldsymbol{H}_k E\boldsymbol{H}_k^\top \\
\frac{\partial^2 l^*}{\partial \boldsymbol{\beta}_k \partial \boldsymbol{\beta}_{k'}^\top} &= -\frac{\rho_k w_k \rho_{k'} w_{k'} \boldsymbol{h}_k \boldsymbol{h}_{k'}^\top}{\left(\sum_{j=0}^{m} \rho_j w_j\right)^2} = -P_k P_{k'} E\boldsymbol{H}_k E\boldsymbol{H}_{k'}^\top
\end{aligned}$$

for $k, k' = 1, \ldots, m$, $k \neq k'$. Therefore,

$$\frac{\partial^2 l^*}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \mathrm{Var}\begin{bmatrix} \boldsymbol{M}_{-1} \\ \boldsymbol{M}_{-1} \odot \boldsymbol{H}_1 \\ \vdots \\ \boldsymbol{M}_{-1} \odot \boldsymbol{H}_m \end{bmatrix}$$

is positive definite based on Assumption 1.2.

$\forall\, 0 < \varepsilon < \frac{1}{2}$, $\forall \boldsymbol{\theta}^* \in \partial B(\boldsymbol{\theta}_0, n^{\varepsilon - \frac{1}{2}})$, expand $l(\boldsymbol{\theta}^*)$ at $\boldsymbol{\theta}_0$,

$$l(\boldsymbol{\theta}^*) = l(\boldsymbol{\theta}_0) + (\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)^\top \left.\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \frac{1}{2}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)^\top \left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{**}} (\boldsymbol{\theta}^* - \boldsymbol{\theta}_0),$$

where $\boldsymbol{\theta}^{**} \in B(\boldsymbol{\theta}_0, ||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||)$. By (1.9),

$$(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)^\top \left.\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \sim O_p(n^{\varepsilon}).$$

By (1.8) and (1.10),

$$\frac{1}{2}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)^\top \left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{**}} (\boldsymbol{\theta}^* - \boldsymbol{\theta}_0) \sim O_p(n^{2\varepsilon}),$$

and it is negative. For large n, almost surely,

$$l(\boldsymbol{\theta}^*) < l(\boldsymbol{\theta}_0).$$

By the continuity of l in $\boldsymbol{\theta}$, l is maximized within $B(\boldsymbol{\theta}_0, n^{\varepsilon - \frac{1}{2}})$, and subsequently, the strong consistency of the estimator $\widetilde{\boldsymbol{\theta}}$ is established.

1.2.2.4 Asymptotic Normality of $\widetilde{\boldsymbol{\theta}}$

Denote

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) = \boldsymbol{S}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{\Lambda}(\boldsymbol{\theta}_0)\boldsymbol{S}^{-1}(\boldsymbol{\theta}_0) = \boldsymbol{S}^{-1}(\boldsymbol{\theta}_0) - \sum_{j=0}^{m} \rho_j \begin{bmatrix} \boldsymbol{J}_m + \boldsymbol{\rho}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{1.11}$$

Expand $\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ at $\boldsymbol{\theta}_0$ and plug in $\widetilde{\boldsymbol{\theta}}$,

$$\left.\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \left.\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p(||\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0||),$$

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\left(\frac{1}{n} \left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right)^{-1} \left(\frac{1}{\sqrt{n}} \left.\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) + o_p(1). \tag{1.12}$$

And since $-\frac{1}{n}\left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{a.s.} \boldsymbol{S}(\boldsymbol{\theta}_0)$, by Slutsky's theorem,

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)). \tag{1.13}$$

Hence, the asymptotic normality of $\widetilde{\boldsymbol{\theta}}$ is established.

1.2.3 Asymptotic Behavior of $\widetilde{G}$

In this section, it is aimed to show that $\sqrt{n}(\widetilde{G} - G)$ converges weakly to a zero-mean Gaussian process in $D(-\infty, \infty)^d$. To prove the weak convergence of $\sqrt{n}(\widetilde{G} - G)$, it is sufficient to show the weak convergence of $\sqrt{n}(\widetilde{G} - \widehat{G})$ since the weak convergence of the empirical process $\sqrt{n}(\widehat{G} - G)$ is ensured by the classical result. The following definitions are taken from [58], which will be used for establishing the weak convergence of $\sqrt{n}(\widetilde{G} - G)$.

Definition 1.1. A collection $\mathcal{F}$ of measurable functions $f : \mathcal{X} \to \mathbb{R}$ is called a *Glivenko-Cantelli* *class* if

$$\sup_{f\in\mathcal{F}} |\mathbb{P}_n f - Pf| \to 0,$$

where $\mathbb{P}_n$ is the empirical measure of a random sample $X_1, \ldots, X_n$ on a measurable space $(\mathcal{X}, \mathcal{A})$ and $\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$in outer probability or outer almost surely.

Definition 1.2. A collection $\mathcal{F}$ of measurable functions $f : \mathcal{X} \to \mathbb{R}$ satisfying

$$\sup_{f\in\mathcal{F}} |f(x) - Pf| < \infty, \qquad \forall\, x,$$

is called a *Donsker class* if

$$\sqrt{n}(\mathbb{P}_n - P) \xrightarrow{d} \mathbb{W}, \qquad \text{in } l^\infty(\mathcal{F}),$$

where $\mathbb{W}$ is a tight borel measurable element in $l^\infty(\mathcal{F})$.

Definition 1.3. A class of measurable functions satisfies *uniform entropy condition* if

$$\int_0^\infty \sup_Q \sqrt{\log N\Big(\varepsilon||F||_{L_2(Q)}, \mathcal{F}, L_2(Q)\Big)}d\varepsilon < \infty,$$

where Q is a probability measure, F is the envelope function of $\mathcal{F}$ i.e. $|f| \leq F\ \forall f \in \mathcal{F}$ and $N\Big(\varepsilon||F||_{L_2(Q)}, \mathcal{F}, L_2(Q)\Big)$ represents the number of covers with radius ε under $L_2(Q)$ norm required for $\mathcal{F}$.

Then we prove the following theorem using the above definitions.

Theorem 1.1. *Let $X_1, \cdots, X_n \sim G$ be a random sample and f be a square integrable function with respect to G. Then $\{I_{\prod_{j=1}^{d}(-\infty,x_d]}(\cdot)\} \cdot f$ is a Donsker class.*

Proof. By Example 2.5.4 in [58], the class of indicator functions satisfy uniform entropy condition. f is a single function so that it also satisfies the uniform entropy condition. Let $F = 1$ be the envelope function for $\{I_{\prod_{j=1}^{d}(-\infty, x_d]}(\cdot)\}$. Since $\int f^2 dG < \infty$, then $\{I_{\prod_{j=1}^{d}(-\infty, x_d]}(\cdot)\} \cdot f$ is a Donsker class by Example 2.10.23 in [58]. $\square$

1.2.3.1 An Approximation of $\widetilde{G}$

Let

$$
\begin{aligned}
A_k(\boldsymbol{x};\boldsymbol{\theta}) &= \int \frac{w_k(\boldsymbol{y};\boldsymbol{\theta}) I_{\prod_{j=1}^{d}(-\infty, x_d]}(\boldsymbol{y})}{\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta})} dG(\boldsymbol{y}),\\
\boldsymbol{B}_k(\boldsymbol{x};\boldsymbol{\theta}) &= \int \frac{w_k(\boldsymbol{y};\boldsymbol{\theta}) \boldsymbol{h}_k(\boldsymbol{y}) I_{\prod_{j=1}^{d}(-\infty, x_d]}(\boldsymbol{y})}{\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta})} dG(\boldsymbol{y}),\\
\overline{\boldsymbol{A}}(\boldsymbol{x};\boldsymbol{\theta}) &= (A_1(\boldsymbol{x};\boldsymbol{\theta}), \cdots, A_m(\boldsymbol{x};\boldsymbol{\theta}))^\top,\\
\overline{\boldsymbol{B}}(\boldsymbol{x};\boldsymbol{\theta}) &= (\boldsymbol{B}_1^\top(\boldsymbol{x};\boldsymbol{\theta}), \cdots, \boldsymbol{B}_m^\top(\boldsymbol{x};\boldsymbol{\theta}))^\top,\\
H_1(\boldsymbol{x};\boldsymbol{\theta}) &= \frac{1}{n_0} \sum_{i=1}^{n} \frac{I_{\prod_{j=1}^{d}(-\infty, x_d]}(\boldsymbol{t}_i)}{\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta})},\\
H_2(\boldsymbol{x};\boldsymbol{\theta}) &= \frac{1}{n} (\overline{\boldsymbol{A}}^\top(\boldsymbol{x};\boldsymbol{\theta})\boldsymbol{\rho}, \overline{\boldsymbol{B}}^\top(\boldsymbol{x};\boldsymbol{\theta})\overline{\boldsymbol{\rho}}) \boldsymbol{S}^{-1}(\boldsymbol{\theta}) \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.
\end{aligned}
$$

Differentiate $H_1(\boldsymbol{x};\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$,

$$\left.\frac{\partial H_1(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = -\frac{1}{n_0}\begin{bmatrix} \sum_{i=1}^{n} \frac{\rho_1 w_1(\boldsymbol{t}_i;\boldsymbol{\theta}_0) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{t}_i)}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\right]^2} \\ \vdots \\ \sum_{i=1}^{n} \frac{\rho_m w_m(\boldsymbol{t}_i;\boldsymbol{\theta}_0) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{t}_i)}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\right]^2} \\ \sum_{i=1}^{n} \frac{\rho_1 w_1(t_i;\boldsymbol{\theta}_0)\boldsymbol{h}_1(\boldsymbol{t}_i) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{t}_i)}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\right]^2} \\ \vdots \\ \sum_{i=1}^{n} \frac{\rho_m w_m(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\boldsymbol{h}_m(t_i) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{t}_i)}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\right]^2} \end{bmatrix}.$$

Then,

$$E\left(\left.\frac{\partial H_1(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) = -\frac{1}{n_0}\begin{bmatrix} \sum_{k=1}^{m}\sum_{l=1}^{n_k} E\left(\frac{\rho_1 w_1(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{X}_l^{(k)})}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)\right]^2}\right) \\ \vdots \\ \sum_{k=1}^{m}\sum_{l=1}^{n_k} E\left(\frac{\rho_1 w_1(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{X}_l^{(k)})}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)\right]^2}\right) \\ \sum_{k=1}^{m}\sum_{l=1}^{n_k} E\left(\frac{\rho_1 w_1(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)\boldsymbol{h}_1(X_l^{(k)}) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{X}_l^{(k)})}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)\right]^2}\right) \\ \vdots \\ \sum_{k=1}^{m}\sum_{l=1}^{n_k} E\left(\frac{\rho_m w_m(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)\boldsymbol{h}_m(\boldsymbol{X}_l^{(k)}) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{X}_l^{(k)})}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)\right]^2}\right) \end{bmatrix}$$

$$
= -\begin{bmatrix}
\sum_{k=1}^{m} \rho_k \int \frac{\rho_1 w_1(\boldsymbol{y};\boldsymbol{\theta}_0) w_k(\boldsymbol{y};\boldsymbol{\theta}_0) I_{\prod_{j=1}^{d}(-\infty, x_d]}(\boldsymbol{y})}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)\right]^2} dG(\boldsymbol{y}) \\
\vdots \\
\sum_{k=1}^{m} \rho_k \int \frac{\rho_m w_m(\boldsymbol{y};\boldsymbol{\theta}_0) w_k(\boldsymbol{y};\boldsymbol{\theta}_0) I_{\prod_{j=1}^{d}(-\infty, x_d]}(\boldsymbol{y})}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)\right]^2} dG(\boldsymbol{y}) \\
\sum_{k=1}^{m} \rho_k \int \frac{\rho_1 w_1(\boldsymbol{y};\boldsymbol{\theta}_0) w_k(\boldsymbol{y};\boldsymbol{\theta}_0) \boldsymbol{h}_1(y) I_{\prod_{j=1}^{d}(-\infty, x_d]}(\boldsymbol{y})}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)\right]^2} dG(\boldsymbol{y}) \\
\vdots \\
\sum_{k=1}^{m} \rho_k \int \frac{\rho_m w_m(\boldsymbol{y};\boldsymbol{\theta}_0) w_k(\boldsymbol{y};\boldsymbol{\theta}_0) \boldsymbol{h}_m(\boldsymbol{y}) I_{\prod_{j=1}^{d}(-\infty, x_d]}(\boldsymbol{y})}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)\right]^2} dG(\boldsymbol{y})
\end{bmatrix}
$$

$$
= -\begin{bmatrix}
\rho_1 A_1(\boldsymbol{x};\boldsymbol{\theta}_0) \\
\vdots \\
\rho_m A_m(\boldsymbol{x};\boldsymbol{\theta}_0) \\
\rho_1 \boldsymbol{B}_1(\boldsymbol{x};\boldsymbol{\theta}_0) \\
\vdots \\
\rho_m \boldsymbol{B}_m(\boldsymbol{x};\boldsymbol{\theta}_0)
\end{bmatrix}
$$

$$
= -(\overline{\boldsymbol{A}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\rho}, \overline{\boldsymbol{B}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}})^\top,
$$

and by SLLN and Assumption 1.1,

$$
\left.\frac{\partial H_1(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = -\begin{bmatrix}
\sum_{k=1}^{m} \rho_k \left(\frac{1}{n_k} \sum_{l=1}^{n_k} \frac{\rho_1 w_1(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0) I_{\prod_{j=1}^{d}(-\infty, x_d]}(\boldsymbol{X}_l^{(k)})}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)\right]^2} \right) \\
\vdots \\
\sum_{k=1}^{m} \rho_k \left(\frac{1}{n_k} \sum_{l=1}^{n_k} \frac{\rho_1 w_1(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0) I_{\prod_{j=1}^{d}(-\infty, x_d]}(\boldsymbol{X}_l^{(k)})}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)\right]^2} \right) \\
\sum_{k=1}^{m} \rho_k \left(\frac{1}{n_k} \sum_{l=1}^{n_k} \frac{\rho_1 w_1(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0) \boldsymbol{h}_1(\boldsymbol{X}_l^{(k)}) I_{\prod_{j=1}^{d}(-\infty, x_d]}(\boldsymbol{X}_l^{(k)})}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)\right]^2} \right) \\
\vdots \\
\sum_{k=1}^{m} \rho_k \left(\frac{1}{n_k} \sum_{l=1}^{n_k} \frac{\rho_m w_m(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0) \boldsymbol{h}_m(\boldsymbol{X}_l^{(k)}) I_{\prod_{j=1}^{d}(-\infty, x_d]}(\boldsymbol{X}_l^{(k)})}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)\right]^2} \right)
\end{bmatrix}
$$

$$
\xrightarrow{a.s.} -\begin{bmatrix} \sum_{k=1}^{m} \rho_k \int \frac{\rho_1 w_1(\boldsymbol{y};\boldsymbol{\theta}_0) w_k(\boldsymbol{y};\boldsymbol{\theta}_0) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{y})}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)\right]^2} dG(\boldsymbol{y}) \\ \vdots \\ \sum_{k=1}^{m} \rho_k \int \frac{\rho_m w_m(\boldsymbol{y};\boldsymbol{\theta}_0) w_k(\boldsymbol{y};\boldsymbol{\theta}_0) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{y})}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)\right]^2} dG(\boldsymbol{y}) \\ \sum_{k=1}^{m} \rho_k \int \frac{\rho_1 w_1(\boldsymbol{y};\boldsymbol{\theta}_0) w_k(\boldsymbol{y};\boldsymbol{\theta}_0) \boldsymbol{h}_1(\boldsymbol{y}) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{y})}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)\right]^2} dG(\boldsymbol{y}) \\ \vdots \\ \sum_{k=1}^{m} \rho_k \int \frac{\rho_m w_m(\boldsymbol{y};\boldsymbol{\theta}_0) w_k(\boldsymbol{y};\boldsymbol{\theta}_0) \boldsymbol{h}_m(\boldsymbol{y}) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{y})}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)\right]^2} dG(\boldsymbol{y}) \end{bmatrix}
$$

$$
= E\left(\left. \frac{\partial H_1(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right).
$$

Note that $\frac{\rho_k w_k(\boldsymbol{y};\boldsymbol{\theta}_0)}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)\right]^2} \leq 1$. Then by Assumption 1.1 and Theorem 1.1, $\frac{\rho_k w_k(\boldsymbol{y};\boldsymbol{\theta}_0)}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)\right]^2} \cdot \{I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{y})\}$ and $\frac{\rho_k w_k(\boldsymbol{y};\boldsymbol{\theta}_0) h_k(\boldsymbol{y})}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)\right]^2} \cdot \{I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{y})\}$ are Donsker classes $\forall\ k$ and hence Glivenko classes by [58] (p. 82).

Thus,

$$
\sup_{\boldsymbol{x}} \left|\left| \left. \frac{\partial H_1(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} - E\left(\left. \frac{\partial H_1(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) \right|\right| \xrightarrow{a.s.} 0. \tag{1.14}
$$

Observe that $\widetilde{G}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}}) = H_1(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$ and expand $\widetilde{G}$ at $\boldsymbol{\theta}_0$,

$$
\begin{aligned}
\widetilde{G}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}}) =& H_1(\boldsymbol{x};\boldsymbol{\theta}_0) + \left(\left. \frac{\partial H_1(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} - E\left(\left. \frac{\partial H_1(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) \right)^\top (\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\
&+ E\left(\left. \frac{\partial H_1(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right)^\top \left(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 - \frac{1}{n} \boldsymbol{S}^{-1}(\boldsymbol{\theta}_0) \left. \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) \\
&+ E\left(\left. \frac{\partial H_1(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right)^\top \left(\frac{1}{n} \boldsymbol{S}^{-1}(\boldsymbol{\theta}_0) \left. \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) + o_p(||\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0||).
\end{aligned}
$$

By (1.12), $o_p(||\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0||) \sim o_p\left(\frac{1}{\sqrt{n}}\right)$.

By (1.12)(1.14), $\left(\left. \frac{\partial H_1(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} - E\left(\left. \frac{\partial H_1(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) \right)^\top (\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sim o_p\left(\frac{1}{\sqrt{n}}\right)$.

Notice that $E\left(\frac{\partial H_1(\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) = -(\overline{\boldsymbol{A}}^\top(\boldsymbol{x};\boldsymbol{\theta})\boldsymbol{\rho}, \overline{\boldsymbol{B}}^\top(\boldsymbol{x};\boldsymbol{\theta})\overline{\boldsymbol{\rho}})^\top$ is uniformly bounded in $\boldsymbol{x}$, then by (1.12),

$$E\left(\frac{\partial H_1(\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)^\top \left(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 - \frac{1}{n}\boldsymbol{S}^{-1}(\boldsymbol{\theta}_0)\,\frac{\partial l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) \sim o_p\left(\frac{1}{\sqrt{n}}\right).$$

Therefore,

$$\widetilde{G}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}}) = H_1(\boldsymbol{x};\boldsymbol{\theta}_0) - H_2(\boldsymbol{x};\boldsymbol{\theta}_0) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

so that $\sqrt{n}(\widetilde{G} - \widehat{G})$ can be approximated by $\sqrt{n}(H_1 - H_2 - \widehat{G})$ uniformly in $\boldsymbol{x}$.

1.2.3.2 Convergence of the Finite Dimension Distribution of $\sqrt{n}(H_1 - H_2 - \widehat{G})$

By the similar derivation in the proof of the Lemma 3.5 in [47],

$$\begin{aligned}
&E\left(H_1(\boldsymbol{x};\boldsymbol{\theta}_0) - \widehat{G}(\boldsymbol{x})\right) = \sum_{k=0}^m \rho_k \int \frac{w_k(\boldsymbol{y};\boldsymbol{\theta}_0) I_{\prod_{j=1}^d(-\infty,x_d]}(\boldsymbol{y})}{\sum_{j=0}^m \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)} dG(\boldsymbol{y}) - G(\boldsymbol{x}) = 0,\\
&E\left(H_2(\boldsymbol{x};\boldsymbol{\theta}_0)\right) = \frac{1}{n}(\overline{\boldsymbol{A}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\rho}, \overline{\boldsymbol{B}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}})\boldsymbol{S}^{-1}(\boldsymbol{\theta}_0) E\left(\frac{\partial l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) = 0,\\
&\mathrm{Cov}\left(\sqrt{n}(H_1(\boldsymbol{x};\boldsymbol{\theta}_0) - \widehat{G}(\boldsymbol{x})), \sqrt{n}(H_1(\boldsymbol{y};\boldsymbol{\theta}_0) - \widehat{G}(\boldsymbol{y}))\right)\\
=&\left(\sum_{k=0}^m \rho_k\right)\left(\sum_{k=1}^m \rho_k A_k(\boldsymbol{x}\wedge\boldsymbol{y};\boldsymbol{\theta}_0) - \sum_{k=1}^m \rho_k A_k(\boldsymbol{x};\boldsymbol{\theta}_0)A_k(\boldsymbol{y};\boldsymbol{\theta}_0)\right.\\
&\left.- \left(\sum_{k=1}^m \rho_k A_k(\boldsymbol{x};\boldsymbol{\theta}_0)\right)\left(\sum_{k=1}^m \rho_k A_k(\boldsymbol{y};\boldsymbol{\theta}_0)\right)\right),\\
&\mathrm{Cov}\left(\sqrt{n}(H_1(\boldsymbol{x};\boldsymbol{\theta}_0) - \widehat{G}(\boldsymbol{x})), \sqrt{n}H_2(\boldsymbol{y};\boldsymbol{\theta}_0)\right)\\
=&(\overline{\boldsymbol{A}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\rho}, \overline{\boldsymbol{B}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}})\boldsymbol{S}^{-1}(\boldsymbol{\theta}_0)(\overline{\boldsymbol{A}}^\top(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{\rho}, \overline{\boldsymbol{B}}^\top(\boldsymbol{y};\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}})^\top\\
&- \left(\sum_{k=0}^m \rho_k\right)\left(\sum_{k=1}^m \rho_k A_k(\boldsymbol{x};\boldsymbol{\theta}_0)A_k(\boldsymbol{y};\boldsymbol{\theta}_0) + \left(\sum_{k=1}^m \rho_k A_k(\boldsymbol{x};\boldsymbol{\theta}_0)\right)\left(\sum_{k=1}^m \rho_k A_k(\boldsymbol{y};\boldsymbol{\theta}_0)\right)\right),
\end{aligned}$$

$$
\begin{aligned}
&\operatorname{Cov}\left(\sqrt{n}H_2(\boldsymbol{x};\boldsymbol{\theta}_0),\sqrt{n}H_2(\boldsymbol{y};\boldsymbol{\theta}_0)\right)\\
=&(\overline{\boldsymbol{A}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\rho},\overline{\boldsymbol{B}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}})\boldsymbol{S}^{-1}(\boldsymbol{\theta}_0)(\overline{\boldsymbol{A}}^\top(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{\rho},\overline{\boldsymbol{B}}^\top(\boldsymbol{y};\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}})^\top\\
&-\left(\sum_{k=0}^m\rho_k\right)\left(\sum_{k=1}^m\rho_kA_k(\boldsymbol{x};\boldsymbol{\theta}_0)A_k(\boldsymbol{y};\boldsymbol{\theta}_0)+\left(\sum_{k=1}^m\rho_kA_k(\boldsymbol{x};\boldsymbol{\theta}_0)\right)\left(\sum_{k=1}^m\rho_kA_k(\boldsymbol{y};\boldsymbol{\theta}_0)\right)\right),
\end{aligned}
$$

where

$$
\boldsymbol{x}\wedge\boldsymbol{y}=(x_1\wedge y_1,\ldots,x_d\wedge y_d)^\top.
$$

By CLT, $\forall\ k$ and $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k$,

$$
\sqrt{n}(H_1(\boldsymbol{x}_1;\boldsymbol{\theta_0})-H_2(\boldsymbol{x}_1;\boldsymbol{\theta_0})-\widehat{G}(\boldsymbol{x}_1),\ldots,H_1(\boldsymbol{x}_k;\boldsymbol{\theta_0})-H_2(\boldsymbol{x}_k;\boldsymbol{\theta_0})-\widehat{G}(\boldsymbol{x}_k))^\top\xrightarrow{d}N(\mathbf{0},\boldsymbol{\Delta}),
$$

where $\boldsymbol{\Delta}$ is the covariance matrix depending on the covariance function

$$
\begin{aligned}
&\operatorname{Cov}(\sqrt{n}(H_1(\boldsymbol{x};\boldsymbol{\theta}_0)-H_2(\boldsymbol{x};\boldsymbol{\theta}_0)-\widehat{G}(\boldsymbol{x})),\sqrt{n}(H_1(\boldsymbol{y};\boldsymbol{\theta}_0)-H_2(\boldsymbol{y};\boldsymbol{\theta}_0)-\widehat{G}(\boldsymbol{y})))\\
=&\left(\sum_{j=0}^m\rho_j\right)\left(\sum_{j=1}^m\rho_jA_j(\boldsymbol{x}\wedge\boldsymbol{y};\boldsymbol{\theta}_0)\right)\\
&-(\overline{\boldsymbol{A}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\rho},\overline{\boldsymbol{B}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}})\boldsymbol{S}^{-1}(\boldsymbol{\theta}_0)(\overline{\boldsymbol{A}}^\top(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{\rho},\overline{\boldsymbol{B}}^\top(\boldsymbol{y};\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}})^\top.
\end{aligned}
\tag{1.15}
$$

1.2.3.3 Weak Convergence of $\sqrt{n}(H_1-\widehat{G})$

Follow the proof of Lemma 3.6 in [47], it can be shown that

$$
\begin{aligned}
&H_1(\boldsymbol{x};\boldsymbol{\theta}_0) - \widehat{G}(\boldsymbol{x}) \\
=&\frac{1}{n_0}\sum_{k=0}^{m}\sum_{l=1}^{n_k}\frac{I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{X}_l^{(k)})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)} - \frac{1}{n_0}\sum_{l=1}^{n_0} I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{X}_l^{(0)}) \\
=&\frac{1}{n_0}\sum_{k=1}^{m}\sum_{l=1}^{n_k}\frac{I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{X}_l^{(k)})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)} \\
&-\left(\frac{1}{n_0}\sum_{l=1}^{n_0}\frac{\left(\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)-1\right) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{X}_l^{(0)})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)}\right) \\
=&\sum_{k=1}^{m}\left(\frac{1}{n_k}\sum_{l=1}^{n_k}\frac{\rho_k I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{X}_l^{(k)})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)} - \rho_k A_k(\boldsymbol{x};\boldsymbol{\theta}_0)\right) \\
&-\left(\frac{1}{n_0}\sum_{l=1}^{n_0}\frac{\left(\sum_{j=1}^{m}\rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)\right) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{X}_l^{(0)})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)} - \sum_{k=1}^{m}\rho_k A_k(\boldsymbol{x};\boldsymbol{\theta}_0)\right),
\end{aligned}
$$

and

$$
E\left(\frac{\rho_k I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{X}_l^{(k)})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)}\right) = \rho_k A_k(\boldsymbol{x};\boldsymbol{\theta}_0),
$$

$$
E\left(\frac{\left(\sum_{j=1}^{m}\rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)\right) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{X}_l^{(0)})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)}\right) = \sum_{k=1}^{m}\rho_k A_k(\boldsymbol{x};\boldsymbol{\theta}_0).
$$

By Theorem 1.1, $\{I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{y})\}\cdot\frac{\rho_k}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)}$ and $\{I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{y})\}\cdot\frac{\sum_{j=1}^{m}\rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)}$ are Donsker classes. Therefore, $\sqrt{n}(H_1 - \widehat{G})$ converges weakly to a zero-mean Guassian process.

1.2.3.4 Tightness of $\sqrt{n}H_2$

To establish the weak convergence of $\sqrt{n}(\widetilde{G}-\widehat{G})$, it is left to show the tightness of $\sqrt{n}H_2$. We first prove the following theorem.

Theorem 1.2. *Let $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ be a d-dimensional random vector that has a density, and*

$$f(\boldsymbol{x}) = Eg(\boldsymbol{X}) I_{\prod_{j=1}^d (-\infty, x_j]}(\boldsymbol{X})$$

for some g such that $Eg^2(\boldsymbol{X}) < \infty$. Let Y_n be a sequence of random variables such that $Y_n \xrightarrow{d} Y$ as $n \to \infty$ and $EY^2 < \infty$. Then the process $f(\boldsymbol{x})Y_n$ is asymptotically tight.

Proof. Define a metric $\rho(\boldsymbol{x}, \boldsymbol{y}) = \sum_{j=1}^d |x_j - y_j|$.

$$\begin{aligned} |f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq& E\left(|g(\boldsymbol{X})||I_{\prod_{j=1}^d (-\infty, x_j]}(\boldsymbol{X}) - I_{\prod_{j=1}^d (-\infty, y_j]}(\boldsymbol{X})|\right) \\ \leq& E\left(|g(\boldsymbol{X})| \sum_{j=1}^d I_{(x_j \wedge y_j, x_j \vee y_j]}(X_j)\right) \\ \leq& \sum_{j=1}^d \left(Eg^2(\boldsymbol{X})\right)^{\frac{1}{2}} \left(P(X_j \in (x_j \wedge y_j, x_j \vee y_j])\right)^{\frac{1}{2}}. \end{aligned}$$

Since $\boldsymbol{X}$ has a density, then the marginal CDFs are uniformly continuous and thus $\forall\ K > 0$, $\exists\ \delta > 0$ such that

$$\begin{aligned} \sup_{\rho(\boldsymbol{x}, \boldsymbol{y}) < \delta} |f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq& \sum_{j=1}^d \left(Eg^2(\boldsymbol{X})\right)^{\frac{1}{2}} \sup_{\rho(\boldsymbol{x}, \boldsymbol{y}) < \delta} \left(P(X_j \in (x_j \wedge y_j, x_j \vee y_j])\right)^{\frac{1}{2}} \\ \leq& \sum_{j=1}^d \left(Eg^2(\boldsymbol{X})\right)^{\frac{1}{2}} \sup_{|x_j - y_j| < \delta} \left(P(X_j \in (x_j \wedge y_j, x_j \vee y_j])\right)^{\frac{1}{2}} \\ <& K. \end{aligned}$$

Then $\forall\, \varepsilon, \eta > 0$, take $K < \frac{\eta\varepsilon}{E|Y|}$, $\exists\, \delta > 0$ such that

$$\begin{aligned}
&\limsup_{n\to\infty} P\left(\sup_{\rho(\boldsymbol{x},\boldsymbol{y})<\delta} |f(\boldsymbol{x})Y_n - f(\boldsymbol{y})Y_n| > \varepsilon\right)\\
=&P\left(\sup_{\rho(\boldsymbol{x},\boldsymbol{y})<\delta} |f(\boldsymbol{x}) - f(\boldsymbol{y})||Y| > \varepsilon\right)\\
\leq&\frac{E|Y|\sup_{\rho(\boldsymbol{x},\boldsymbol{y})<\delta}|f(\boldsymbol{x}) - f(\boldsymbol{y})|}{\varepsilon}\\
<&\eta.
\end{aligned}$$

Additionally, $\forall\, \eta > 0$, take $\varepsilon > \frac{E|g(\boldsymbol{X})|E|Y|}{\eta}$,

$$\limsup_{n\to\infty} P\left(|f(\boldsymbol{x})Y_n| > \varepsilon\right) = P\left(|f(\boldsymbol{x})Y| > \varepsilon\right) \leq \frac{E|g(\boldsymbol{X})|E|Y|}{\varepsilon} < \eta.$$

Hence, $f(\boldsymbol{x})Y_n$ is asymptotically tight by Theorem 1.5.7 in [58]. □

Note that $A_k(\boldsymbol{x};\boldsymbol{\theta}_0)$ and the elements of $\boldsymbol{B}_k(\boldsymbol{x};\boldsymbol{\theta}_0)$ satisfy the conditions of $f(\boldsymbol{x})$ in Theorem 1.2. By (1.9), $\frac{1}{\sqrt{n}}\left.\frac{\partial l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ converges weakly to a Guassian random variable. Thus, based on the expression of $\sqrt{n}H_2$,

$$\sqrt{n}H_2(\boldsymbol{x};\boldsymbol{\theta}_0) = (\overline{\boldsymbol{A}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\rho}, \overline{\boldsymbol{B}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}})\boldsymbol{S}^{-1}(\boldsymbol{\theta}_0)\left(\frac{1}{\sqrt{n}}\left.\frac{\partial l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right),$$

by Theorem 1.2, $\sqrt{n}H_2$ is asymptotically tight.

1.2.3.5 Weak Convergence of $\sqrt{n}(\widetilde{G} - G)$

Based on the results in Section 1.2.3.2, 1.2.3.3, and 1.2.3.4, by Theorem 1.5.4 in [58], $\sqrt{n}(H_1 - H_2 - \widehat{G})$ converges weakly to a zero-mean Guassian process with convariance function

(1.15), which implies the weak convergence of $\sqrt{n}(\widetilde{G}-\widehat{G})$. By classical result, the empirical process $\sqrt{n}(\widehat{G}-G)$ converges weakly to a zero-mean Guassian process so that the weak convergence of $\sqrt{n}(\widetilde{G}-G)$ is established. It is left to show the covariance function of the limit Guassian process. By the same derivation as the proof of Theorem 3.9 in [47],

$$
\begin{aligned}
&\operatorname{Cov}\left(\sqrt{n}(H_1(\boldsymbol{x};\boldsymbol{\theta}_0)-\widehat{G}(\boldsymbol{x})), \sqrt{n}\widehat{G}(\boldsymbol{y})\right)\\
=&\left(\sum_{j=0}^{m}\rho_j\right)\left(\sum_{j=1}^{m}\rho_j A_j(\boldsymbol{x};\boldsymbol{\theta}_0)\right)G(\boldsymbol{y})-\left(\sum_{j=0}^{m}\rho_j\right)\left(\sum_{j=1}^{m}\rho_j A_j(\boldsymbol{x}\wedge\boldsymbol{y};\boldsymbol{\theta}_0)\right),\\
&\operatorname{Cov}\left(\sqrt{n}H_2(\boldsymbol{x};\boldsymbol{\theta}_0), \sqrt{n}\widehat{G}(\boldsymbol{y})\right)\\
=&\left(\sum_{j=0}^{m}\rho_j\right)\left(\sum_{j=1}^{m}\rho_j A_j(\boldsymbol{x};\boldsymbol{\theta}_0)\right)G(\boldsymbol{y})\\
&-(\overline{\boldsymbol{A}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\rho},\overline{\boldsymbol{B}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}})\boldsymbol{S}^{-1}(\boldsymbol{\theta}_0)(\bar{\boldsymbol{A}}^\top(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{\rho},\overline{\boldsymbol{B}}^\top(\boldsymbol{y};\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}})^\top,\\
&\operatorname{Cov}\left(\sqrt{n}\widehat{G}(\boldsymbol{x}), \sqrt{n}\widehat{G}(\boldsymbol{y})\right)\\
=&\left(\sum_{j=0}^{m}\rho_j\right)(G(\boldsymbol{x}\wedge\boldsymbol{y})-G(\boldsymbol{x})G(\boldsymbol{y})).
\end{aligned}
$$

Together with (1.15), we obtain

$$
\begin{aligned}
&\operatorname{Cov}(\sqrt{n}(\widetilde{G}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})-G(\boldsymbol{x})), \sqrt{n}(\widetilde{G}(\boldsymbol{y};\widetilde{\boldsymbol{\theta}})-G(\boldsymbol{y})))\\
=&\left(\sum_{j=0}^{m}\rho_j\right)\left(G(\boldsymbol{x}\wedge\boldsymbol{y})-G(\boldsymbol{x})G(\boldsymbol{y})-\sum_{j=1}^{m}\rho_j A_j(\boldsymbol{x}\wedge\boldsymbol{y};\boldsymbol{\theta}_0)\right)\\
&+(\overline{\boldsymbol{A}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\rho},\overline{\boldsymbol{B}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}})\boldsymbol{S}^{-1}(\boldsymbol{\theta}_0)(\overline{\boldsymbol{A}}^\top(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{\rho},\overline{\boldsymbol{B}}^\top(\boldsymbol{y};\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}})^\top.
\end{aligned}
$$

Subsequently,

$$\sqrt{n}(\widetilde{G}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})-G(\boldsymbol{x}))\xrightarrow{d}N(0,\sigma(\boldsymbol{x};\boldsymbol{\theta}_0)),$$

as $n \to \infty$, where

$$\sigma(\boldsymbol{x};\boldsymbol{\theta}_0) = \left(\sum_{j=0}^{m}\rho_j\right)\left(G(\boldsymbol{x}) - G^2(\boldsymbol{x}) - \sum_{j=1}^{m}\rho_j A_j(\boldsymbol{x};\boldsymbol{\theta}_0)\right) + (\overline{\boldsymbol{A}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\rho}, \overline{\boldsymbol{B}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}})\boldsymbol{S}^{-1}(\boldsymbol{\theta}_0)(\overline{\boldsymbol{A}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\rho}, \overline{\boldsymbol{B}}^\top(\boldsymbol{x};\boldsymbol{\theta}_0)\overline{\boldsymbol{\rho}})^\top. \tag{1.16}$$

Remark: Since $\widetilde{G}$ can be approximated by $H_1 - H_2$, then one may show the weak convergence of $\sqrt{n}(\widetilde{G} - G)$ by establishing the weak convergence of $\sqrt{n}(H_1 - H_2 - G)$ directly. However, the limiting distribution of $\sqrt{n}(\widetilde{G} - \widehat{G})$ is of interest for other research purposes, for example, checking goodness-of-fit. Hence, we follow the steps in [47].

1.2.4 Consistent Estimators of Asymptotic Variances

In this section, the consistency of the plug-in estimators of asymptotic variances of $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{G}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$ are established.

1.2.4.1 Consistency of Plug-in Estimator $\widetilde{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{\theta}})$

Based on the expression of $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ (1.11), to show the consistency of $\widetilde{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{\theta}})$, it is sufficient to show that $\widetilde{\boldsymbol{S}}(\widetilde{\boldsymbol{\theta}})$ is consistent. We will use one entry of $\boldsymbol{S}(\boldsymbol{\theta}_0)$ as an example to show the consistency. The rest of the entries can be proved in the same manner. We first prove the following theorem.

Theorem 1.3. *Suppose $\widetilde{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$, and f satisfies that $\forall$ $\boldsymbol{x}$, $\boldsymbol{\theta}$,*

$$\left\|\frac{\partial f(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\| \le g(\boldsymbol{x}),$$

and

$$\frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{X}_i) \xrightarrow{p} Eg(\boldsymbol{X}_1) < \infty,$$

as $n \to \infty$ *for a sequence of identically distributed random variables* $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$*. Then*

$$\frac{1}{n}\sum_{i=1}^{n} \left| f(\boldsymbol{X}_i; \widetilde{\boldsymbol{\theta}}) - f(\boldsymbol{X}_i; \boldsymbol{\theta}_0) \right| \xrightarrow{p} 0,$$

as $n \to \infty$*.*

Proof. By Mean Value Theorem, $\exists\, \boldsymbol{\theta}_i^* \in B(\boldsymbol{\theta}_0, \|\boldsymbol{\theta}_0 - \widetilde{\boldsymbol{\theta}}\|)$ for $i = 1, \ldots, n$,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \left| f(\boldsymbol{X}_i; \widetilde{\boldsymbol{\theta}}) - f(\boldsymbol{X}_i; \boldsymbol{\theta}_0) \right| &= \frac{1}{n}\sum_{i=1}^{n} \left| (\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \left.\frac{\partial f(\boldsymbol{X}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta} = \boldsymbol{\theta}_i^*} \right| \\ &\leq \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{X}_i) \|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \\ &\xrightarrow{p} 0, \end{aligned}$$

as $n \to \infty$. □

Consider the entry of $\boldsymbol{S}(\boldsymbol{\theta}_0)$ corresponding to β_{kp} and $\beta_{k'p'}$,

$$E\left(-\frac{1}{n} \left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta_{kp} \partial \beta_{k'p'}}\right|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} \right) = -\frac{\rho_k \rho_{k'}}{\sum_{j=0}^{m} \rho_j} \int \frac{w_k(\boldsymbol{y}; \boldsymbol{\theta}_0) w_{k'}(\boldsymbol{y}; \boldsymbol{\theta}_0) h_{kp}(\boldsymbol{y}) h_{k'p'}(\boldsymbol{y})}{\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{y}; \boldsymbol{\theta}_0)} dG(\boldsymbol{y}).$$

The corresponding estimator is

$$
\begin{aligned}
-\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\beta_{kp}\partial\beta_{k'p'}}\bigg|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} &= -\frac{1}{n}\sum_{i=1}^{n}\frac{\rho_k w_k(\boldsymbol{t}_i;\widetilde{\boldsymbol{\theta}})\rho_{k'}w_{k'}(\boldsymbol{t}_i;\widetilde{\boldsymbol{\theta}})h_{kp}(\boldsymbol{t}_i)h_{k'p'}(\boldsymbol{t}_i)}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{t}_i;\widetilde{\boldsymbol{\theta}})\right]^2} \\
&= -\sum_{h=0}^{m}\frac{n_h}{n}\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\frac{\rho_k w_k(\boldsymbol{X}_l^{(h)};\widetilde{\boldsymbol{\theta}})\rho_{k'}w_{k'}(\boldsymbol{X}_l^{(h)};\widetilde{\boldsymbol{\theta}})h_{kp}(\boldsymbol{X}_l^{(h)})h_{k'p'}(\boldsymbol{X}_l^{(h)})}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(h)};\widetilde{\boldsymbol{\theta}})\right]^2}\right).
\end{aligned}
$$

Thus, it is sufficient to show

$$
\begin{aligned}
\frac{1}{n_h}\sum_{l=1}^{n_h}\Bigg|&\frac{\rho_k w_k(\boldsymbol{X}_l^{(h)};\widetilde{\boldsymbol{\theta}})\rho_{k'}w_{k'}(\boldsymbol{X}_l^{(h)};\widetilde{\boldsymbol{\theta}})h_{kp}(\boldsymbol{X}_l^{(h)})h_{k'p'}(\boldsymbol{X}_l^{(h)})}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(h)};\widetilde{\boldsymbol{\theta}})\right]^2} \\
&-\frac{\rho_k w_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\rho_{k'}w_{k'}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)h_{kp}(\boldsymbol{X}_l^{(h)})h_{k'p'}(\boldsymbol{X}_l^{(h)})}{\left[\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\right]^2}\Bigg| \xrightarrow{p} 0,
\end{aligned}
$$

as $n_h \to \infty$.

Recall that $w_k(\boldsymbol{x};\boldsymbol{\theta}) = \exp\left(\alpha_k + \boldsymbol{\beta}_k^\top \boldsymbol{h}_k(\boldsymbol{x})\right)$, and hence $\forall\, k$

$$
\begin{aligned}
\left\|\frac{\partial \log w_k(\boldsymbol{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\| &= \sqrt{1+\sum_{p=1}^{p_k}|h_{kp}(\boldsymbol{x})|^2} \\
&\le 1+\sum_{p=1}^{p_k}|h_{kp}(\boldsymbol{x})| \\
&\le 1+\sum_{k=1}^{m}\sum_{p=1}^{p_k}|h_{kp}(\boldsymbol{x})|.
\end{aligned}
$$

Then, by the Triangle Inequality,

$$\begin{aligned}
&\left\| \frac{\partial \frac{w_k(\boldsymbol{x};\boldsymbol{\theta})w_{k'}(\boldsymbol{x};\boldsymbol{\theta})}{\left[\sum_{j=0}^m \rho_j w_j(\boldsymbol{x};\boldsymbol{\theta})\right]^2}}{\partial \boldsymbol{\theta}} \right\| \\
=&\left\| \frac{w_k(\boldsymbol{x};\boldsymbol{\theta})w_{k'}(\boldsymbol{x};\boldsymbol{\theta})\left[\frac{\partial \log w_k(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial \log w_{k'}(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]}{\left[\sum_{j=0}^m \rho_j w_j(\boldsymbol{x};\boldsymbol{\theta})\right]^2} \right. \\
&\left. - \frac{2w_k(\boldsymbol{x};\boldsymbol{\theta})w_{k'}(\boldsymbol{x};\boldsymbol{\theta})\left[\sum_{j=0}^m \rho_j w_j(\boldsymbol{x};\boldsymbol{\theta})\frac{\partial \log w_j(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]}{\left[\sum_{j=0}^m \rho_j w_j(\boldsymbol{x};\boldsymbol{\theta})\right]^3} \right\| \\
\le&\frac{w_k(\boldsymbol{x};\boldsymbol{\theta})w_{k'}(\boldsymbol{x};\boldsymbol{\theta})\left[\left\|\frac{\partial \log w_k(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\| + \left\|\frac{\partial \log w_{k'}(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\|\right]}{\left[\sum_{j=0}^m \rho_j w_j(\boldsymbol{x};\boldsymbol{\theta})\right]^2} \\
&+ \frac{2w_k(\boldsymbol{x};\boldsymbol{\theta})w_{k'}(\boldsymbol{x};\boldsymbol{\theta})\left[\sum_{j=0}^m w_j(\boldsymbol{x};\boldsymbol{\theta})\left\|\frac{\partial \log w_j(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\|\right]}{\left[\sum_{j=0}^m \rho_j w_j(\boldsymbol{x};\boldsymbol{\theta})\right]^3} \\
\le&\frac{4w_k(\boldsymbol{x};\boldsymbol{\theta})w_{k'}(\boldsymbol{x};\boldsymbol{\theta})\left[1+\sum_{k=1}^m \sum_{p=1}^{p_k} |h_{kp}(\boldsymbol{x})|\right]}{\left[\sum_{j=0}^m \rho_j w_j(\boldsymbol{x};\boldsymbol{\theta})\right]^2}.
\end{aligned}$$

Note that

$$\frac{\rho_k w_k(\boldsymbol{x};\boldsymbol{\theta})\rho_{k'} w_{k'}(\boldsymbol{x};\boldsymbol{\theta})}{\left[\sum_{j=0}^m \rho_j w_j(\boldsymbol{x};\boldsymbol{\theta})\right]^2} \le 1,$$

and therefore

$$\begin{aligned}
&\frac{1}{n_h}\sum_{l=1}^{n_h}\left| \frac{\rho_k w_k(\boldsymbol{X}_l^{(h)};\widetilde{\boldsymbol{\theta}})\rho_{k'} w_{k'}(\boldsymbol{X}_l^{(h)};\widetilde{\boldsymbol{\theta}})h_{kp}(\boldsymbol{X}_l^{(h)})h_{k'p'}(\boldsymbol{X}_l^{(h)})}{\left[\sum_{j=0}^m \rho_j w_j(\boldsymbol{X}_l^{(h)};\widetilde{\boldsymbol{\theta}})\right]^2} \right. \\
&\left. - \frac{\rho_k w_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\rho_{k'} w_{k'}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)h_{kp}(\boldsymbol{X}_l^{(h)})h_{k'p'}(\boldsymbol{X}_l^{(h)})}{\left[\sum_{j=0}^m \rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\right]^2} \right| \\
\le&\frac{1}{n_h}\sum_{l=1}^{n_h} 4\left[1+\sum_{k=1}^m\sum_{p=1}^{p_k}\left|h_{kp}(\boldsymbol{X}_l^{(h)})\right|\right]\left|h_{kp}(\boldsymbol{X}_l^{(h)})h_{k'p'}(\boldsymbol{X}_l^{(h)})\right| \|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\| \xrightarrow{p} 0,
\end{aligned}$$

as $n_h \to \infty$, by Assumption 1.1 and Theorem 1.3. Therefore,

$$-\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\beta_{kp}\partial\beta_{k'p'}}\bigg|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \xrightarrow{p} -\frac{\rho_k\rho_{k'}}{\sum_{j=0}^m \rho_j}\int \frac{w_k(\boldsymbol{y};\boldsymbol{\theta}_0)w_{k'}(\boldsymbol{y};\boldsymbol{\theta}_0)h_{kp}(\boldsymbol{y})h_{k'p'}(\boldsymbol{y})}{\sum_{j=0}^m \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)}dG(\boldsymbol{y}),$$

as $n \to \infty$. Similarly, the convergence of the rest of entries can be established, and hence

$$\widetilde{\boldsymbol{S}}(\widetilde{\boldsymbol{\theta}}) \xrightarrow{p} \boldsymbol{S}(\boldsymbol{\theta}_0),$$

as $n \to \infty$.

1.2.4.2 Consistency of Plug-in Estimator $\widetilde{\sigma}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$

Based on the expression of $\sigma(\boldsymbol{x};\boldsymbol{\theta}_0)$ (1.16), since the consistency of $\widetilde{\boldsymbol{S}}(\widetilde{\boldsymbol{\theta}})$ has been established, it is sufficient to show the consistency of the plug-in estimators of $A_k(\boldsymbol{x};\boldsymbol{\theta}_0)$ and $\boldsymbol{B}_k(\boldsymbol{x};\boldsymbol{\theta}_0)$, which are $\widetilde{A}_k(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$ and $\widetilde{\boldsymbol{B}}_k(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$, respectively. We prove the consistency of $\widetilde{\boldsymbol{B}}_k(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$, and that of $\widetilde{A}_k(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$ can be proved in the same manner.

First, we show the convergence of $\widetilde{\boldsymbol{B}}_k(\boldsymbol{x};\boldsymbol{\theta}_0)$. By SLLN and Assumption 1.1,

$$\begin{aligned}
\widetilde{\boldsymbol{B}}_k(\boldsymbol{x};\boldsymbol{\theta}_0) &= \sum_{i=1}^n p_i \frac{w_k(\boldsymbol{t}_i;\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{t}_i)I_{\prod_{j=1}^d(-\infty,x_d]}(\boldsymbol{t}_i)}{\sum_{j=0}^m \rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta}_0)} \\
&= \sum_{h=0}^m \rho_h \frac{1}{n_h}\sum_{l=1}^{n_h} \frac{w_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{X}_l^{(h)})I_{\prod_{j=1}^d(-\infty,x_d]}(\boldsymbol{X}_l^{(h)})}{\left[\sum_{j=0}^m \rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\right]^2} \\
&\xrightarrow{a.s.} \sum_{h=0}^m \rho_h \int \frac{w_k(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{y})I_{\prod_{j=1}^d(-\infty,x_d]}(\boldsymbol{y})}{\left[\sum_{j=0}^m \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)\right]^2} w_h(\boldsymbol{y};\boldsymbol{\theta}_0)dG(\boldsymbol{y}) \\
&= \int \frac{w_k(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{h}_k(\boldsymbol{y})I_{\prod_{j=1}^d(-\infty,x_d]}(\boldsymbol{y})}{\sum_{j=0}^m \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta}_0)}dG(\boldsymbol{y}) = \boldsymbol{B}_k(\boldsymbol{x};\boldsymbol{\theta}_0).
\end{aligned}$$

Next, $\widetilde{\boldsymbol{B}}_k(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$ is given by

$$\widetilde{\boldsymbol{B}}_k(\boldsymbol{x};\widetilde{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \widetilde{p}_i \frac{w_k(\boldsymbol{t}_i;\widetilde{\boldsymbol{\theta}})\boldsymbol{h}_k(\boldsymbol{t}_i) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{t}_i)}{\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{t}_i;\widetilde{\boldsymbol{\theta}})}$$
$$= \sum_{h=0}^{m} \rho_h \frac{1}{n_h} \sum_{l=1}^{n_h} \frac{w_k(\boldsymbol{X}_l^{(h)};\widetilde{\boldsymbol{\theta}})\boldsymbol{h}_k(\boldsymbol{X}_l^{(h)}) I_{\prod_{j=1}^{d}(-\infty,x_d]}(\boldsymbol{X}_l^{(h)})}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{X}_l^{(h)};\widetilde{\boldsymbol{\theta}})\right]^2},$$

and by Triangle Inequality,

$$\left\| \frac{\partial \frac{w_k(\boldsymbol{x};\boldsymbol{\theta})}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{x};\boldsymbol{\theta})\right]^2}}{\partial \boldsymbol{\theta}} \right\| = \left\| \frac{w_k(\boldsymbol{x};\boldsymbol{\theta}) \frac{\partial \log w_k(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{x};\boldsymbol{\theta})\right]^2} - \frac{2 w_k(\boldsymbol{x};\boldsymbol{\theta}) \left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{x};\boldsymbol{\theta}) \frac{\partial \log w_j(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{x};\boldsymbol{\theta})\right]^3} \right\|$$
$$\leq \frac{w_k(\boldsymbol{x};\boldsymbol{\theta}) \left\| \frac{\partial \log w_k(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{x};\boldsymbol{\theta})\right]^2} + \frac{2 w_k(\boldsymbol{x};\boldsymbol{\theta}) \left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{x};\boldsymbol{\theta}) \left\| \frac{\partial \log w_j(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|\right]}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{x};\boldsymbol{\theta})\right]^3}$$
$$\leq \frac{3 w_k(\boldsymbol{x};\boldsymbol{\theta}) \left[1 + \sum_{k=1}^{m} \sum_{p=1}^{p_k} |h_{kp}(\boldsymbol{x})|\right]}{\left[\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{x};\boldsymbol{\theta})\right]^2}$$
$$\leq 3 \frac{1}{\rho_k} \left[1 + \sum_{k=1}^{m} \sum_{p=1}^{p_k} |h_{kp}(\boldsymbol{x})|\right].$$

Then, by Assumption 1.1 and Theorem 1.3, as $n \to \infty$,

$$\widetilde{\boldsymbol{B}}_k(\boldsymbol{x};\widetilde{\boldsymbol{\theta}}) - \widetilde{\boldsymbol{B}}_k(\boldsymbol{x};\boldsymbol{\theta}_0) \xrightarrow{p} 0,$$

so that

$$\widetilde{\boldsymbol{B}}_k(\boldsymbol{x};\widetilde{\boldsymbol{\theta}}) \xrightarrow{p} \boldsymbol{B}_k(\boldsymbol{x};\boldsymbol{\theta}_0),$$

and the consistency of $\widetilde{A}_k(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$ can be shown by the same way. Therefore, the consistency of $\widetilde{\sigma}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$ is established.

1.3 Simulation

In this section, based on the simulation result, it is illustrated that the estimate of the reference CDF obtained by a DRM with variable tilts is more precise than the estimate obtained by a DRM with a uniform tilt and empirical estimate in terms of mean integrated absolute error (MIAE) and mean integrated squared error (MISE),

$$\text{MIAE} = E\int \left|\widehat{f}(x) - f(x)\right| dx,$$
$$\text{MISE} = E\int \left(\widehat{f}(x) - f(x)\right)^2 dx.$$

Here, $\widehat{f}$ represents the three different estimates $\widetilde{G}_u$, $\widetilde{G}_v$ and $\widehat{G}$ where $\widetilde{G}_u$ is the estimate obtained with the uniform tilt and $\widetilde{G}_v$ is that obtained with the variable tilts, while $\widehat{G}$ is the empirical estimate. f represents the true reference CDF G.

Consider three random samples $\boldsymbol{X}^{(0)} \sim Exp(2)$, $\boldsymbol{X}^{(1)} \sim Gamma(2,2)$ and $\boldsymbol{X}^{(2)} \sim Lognormal(1,1)$ with size n_0, n_1 and n_2, respectively. They follow the true density ratio structure

$$\frac{g_1(x)}{g_0(x)} = \exp(\log 2 + \log(x)),$$
$$\frac{g_2(x)}{g_0(x)} = \exp(-\frac{1}{2} - \log 2\sqrt{2\pi} + 2x - \frac{1}{2}(\log(x))^2).$$

In this case, the uniform tilt is considered to be $\boldsymbol{h}_1(x) = \boldsymbol{h}_2(x) = (x, \log(x), (\log(x))^2)^\top$, which is used to obtain the estimate $\widetilde{G}_u$. The variable tilts are $\boldsymbol{h}_1(x) = \log(x)$ and $\boldsymbol{h}_2(x) = (x, (\log(x))^2)^\top$, which are used to obtain the estimate $\widetilde{G}_v$.

The simulation is performed by following the steps:

1. Generate random samples

$$\boldsymbol{X}^{(0)} \sim Exp(2), \quad \boldsymbol{X}^{(1)} \sim Gamma(2,2), \quad \boldsymbol{X}^{(2)} \sim Lognormal(1,1),$$

   with size n_0, n_1 and n_2.

2. Obtain the estimates $\widetilde{G}_v$, $\widetilde{G}_u$ and $\widehat{G}$.

3. Repeat steps 1 and 2 for I times.

4. Approximate MIAE and MISE by

$$\widehat{\text{MIAE}} = \frac{1}{I}\sum_{i=1}^{I}\delta\sum_{j=1}^{J}\left|\hat{f}^{(i)}(M_1+j\delta) - f^{(i)}(M_1+j\delta)\right|,$$
$$\widehat{\text{MISE}} = \frac{1}{I}\sum_{i=1}^{I}\delta\sum_{j=1}^{J}\left(\hat{f}^{(i)}(M_1+j\delta) - f^{(i)}(M_1+j\delta)\right)^2,$$

   where (M_1, M_2) is the region we integrate over that satisfies $\int_{-\infty}^{\infty} f(x)dx \approx \int_{M_1}^{M_2} f(x)dx$ and $\delta = \frac{M_2-M_1}{J}$.

Set $I = 1000$, $n_0 = n_1 = n_2 = 200$, $(M_1, M_2) = (0, 10)$ and $J = 1000$. The results shown in Table 1.1 indicate that the DRM with variable tilts gives a better estimate of the reference CDF judging by both measures MIAE and MISE.

Table 1.1: MIAE and MISE for $\widetilde{G}_v$, $\widetilde{G}_u$ and $\widehat{G}$.

Estimate	$\widetilde{G}_v$	$\widetilde{G}_u$	$\widehat{G}$
MIAE	3.273×10^{-2}	3.791×10^{-2}	4.358×10^{-2}
MISE	9.137×10^{-4}	1.066×10^{-3}	1.275×10^{-3}

Another comparison is made via the 95% CIs for threshold probability $1 - G(T)$ with

$T = 0.1, 0.5, 1, 2, 3$. Here we simulate the 95% CI for the threshold probability $1 - G(T)$ based on $\widetilde{G}_v$, $\widetilde{G}_u$ and $\widehat{G}$, respectively. The simulation is performed by the following steps:

1. Generate random samples

$$\boldsymbol{X}^{(0)} \sim Exp(2), \quad \boldsymbol{X}^{(1)} \sim Gamma(2, 2), \quad \boldsymbol{X}^{(2)} \sim Lognormal(1, 1),$$

   with size n_0, n_1 and n_2.

2. Obtain point estimates and 95% CIs for threshold probability $1 - G(T)$ based on the three methods.

3. Repeat step 1 and 2 for I times.

4. Calculate the average of estimates, confidence limits, lengths of CIs. Also obtain the coverage rates of CIs.

Take $I = 1000$ and $n_0 = n_1 = n_2 = 200$. The results are shown by Table 1.2. Based on the lengths of the CIs, $\widetilde{G}_v$ is more precise than $\widetilde{G}_u$ and $\widehat{G}$ since the corresponding CIs are the shortest for each T. Also, the results reveal an advantage of combining multiple samples that it allows us to estimate outside of the range of the reference sample. The empirical CDF cannot provide an estimate for $T = 2, 3$ while this is available using DRM. Additionally, both DRM estimates are more precise than the empirical estimate in terms of the lengths of CIs, which accords with the conclusion in [81]. Furthermore, the coverage rates of DRM with variable tilts are closed to 0.95 and have a slower decay than the coverage rates of the other two models when the threshold becomes larger.

Table 1.2: Average of threshold probability estimates, 95% CIs, lengths of CIs, coverage rates of CIs for thresholds $T = 0.1, 0.5, 1, 2, 3$. "-" indicates that some of the CIs generated can not be constructed since the estimates are 0.

T	$1 - \widetilde{G}_v(T)$	95% CI	CI length	Coverage rate
0.1	0.8189	$(0.7677, 0.8700)$	0.1023	0.951
0.5	0.3654	$(0.3049, 0.4259)$	0.1210	0.947
1.0	0.1336	$(0.0974, 0.1698)$	0.0724	0.936
2.0	0.0182	$(0.0091, 0.0272)$	0.0181	0.930
3.0	0.0025	$(0.0004, 0.0046)$	0.0042	0.900
T	$1 - \widetilde{G}_u(T)$	95% CI	CI length	Coverage rate
0.1	0.8176	$(0.7653, 0.8699)$	0.1046	0.953
0.5	0.3677	$(0.3051, 0.4302)$	0.1251	0.946
1.0	0.1355	$(0.0941, 0.1770)$	0.0828	0.937
2.0	0.0177	$(0.0039, 0.0314)$	0.0275	0.870
3.0	0.0024	$(-0.0013, 0.0061)$	0.0074	0.771
T	$1 - \widehat{G}(T)$	95% CI	CI length	Coverage rate
0.1	0.8183	$(0.7651, 0.8715)$	0.1064	0.965
0.5	0.3664	$(0.2998, 0.4330)$	0.1332	0.946
1.0	0.1347	$(0.0877, 0.1818)$	0.0941	0.940
2.0	-	-	-	0.878
3.0	-	-	-	0.386

1.4 Tilt Selection

Since model (1.1) accommodates different tilts, the selection of the tilts between every pair $(\boldsymbol{X}^{(0)}, \boldsymbol{X}^{(k)})$ should be addressed.

One way is to employ significance tests based on the asymptotic normality of the estimators (1.13), and then construct the DRM for all samples using the selected tilts. Construct a two-sample DRM for the pair $(\boldsymbol{X}^{(0)}, \boldsymbol{X}^{(k)})$ based on a possibly redundant tilt $\boldsymbol{h}$ with dimension p, and obtain the estimated parameters $\widetilde{\beta}_j^{(k)}$'s and their estimated standard deviations $\widetilde{\sigma}_{\widetilde{\beta}_j^{(k)}}$. Test $H_0 : \beta_j^{(k)} = 0$ for $j = 1, \cdots, p$ by a Z-test with the test statistic

$$Z = \frac{\widetilde{\beta}_j^{(k)}}{\widetilde{\sigma}_{\widetilde{\beta}_j^{(k)}}}. \tag{1.17}$$

If $\beta_j^{(k)}$ is insignificant, the corresponding term is eliminated from the tilt function and then a reduced tilt function $\boldsymbol{h}_k$ can be obtained. Finally, use all reduced tilts $\boldsymbol{h}_k$'s to construct a DRM for all samples by (1.1).

Another way of tilt selection is to use a model selection criterion as suggested in [20]. Instead of combining pairs, we construct DRMs with different combinations of tilt functions for all samples. The possible choices of tilt functions are the redundant tilt $\boldsymbol{h}$ and all of its reduced forms. Then, the optimal DRM is selected from 2^{pm} possible models using the AIC

$$-2\log L(\widetilde{\boldsymbol{\theta}}, \widetilde{G}) + 2q, \tag{1.18}$$

where q is the number of free parameters in the model. Here, the likelihood used for the AIC is the empirical likelihood.

1.5 Estimation of Residential Radon Concentration in Pennsylvania Counties

In this section, we demonstrate the use of DRM in the estimation of Radon concentration, which is largely based on [78] and [79]. According to [1], Radon-222, or just radon, is a colorless, odorless, radioactive noble gas that stems from the decay series of radium-226. It is widely distributed in and naturally released from soils and rocks. The indoor radon has been determined to be the second leading cause of lung cancer by National Academy of Sciences. Environmental Protection Agency (EPA) estimated that 13.4% lung cancer deaths nationally in 1995 are radon related, and anticipated that indoor radon would be a significant contributor of lung cancer deaths annually. In United States, radon concentration in air is measured in picocuries per liter (pCi/L). EPA recommended to control the indoor radon concentration between 2 pCi/L and 4 pCi/L.

According to Pennsylvania (PA) Department of Environmental Protection [2], residential radon problem is serious in PA, and approximately 40% of the homes have radon concentration above 4 pCi/L. Zip code level residential radon concentration data are collected from PA counties in six time periods. We examine two counties, Beaver and Forest, to demonstrate the two tilt selection methods in Section 1.4.

1.5.1 Beaver County

The six periods are 1989-1993, 1994-1998, 1999-2003, 2004-2008, 2009-2013, 2014-2017. We consider the sample of 2014-2017 as the reference sample $\boldsymbol{X}^{(0)}$. $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(5)}$, are the samples of 1989-1993, 1994-1998, 1999-2003, 2004-2008, 2009-2013, respectively. The histograms of the six samples are plotted in Figure 1.1.

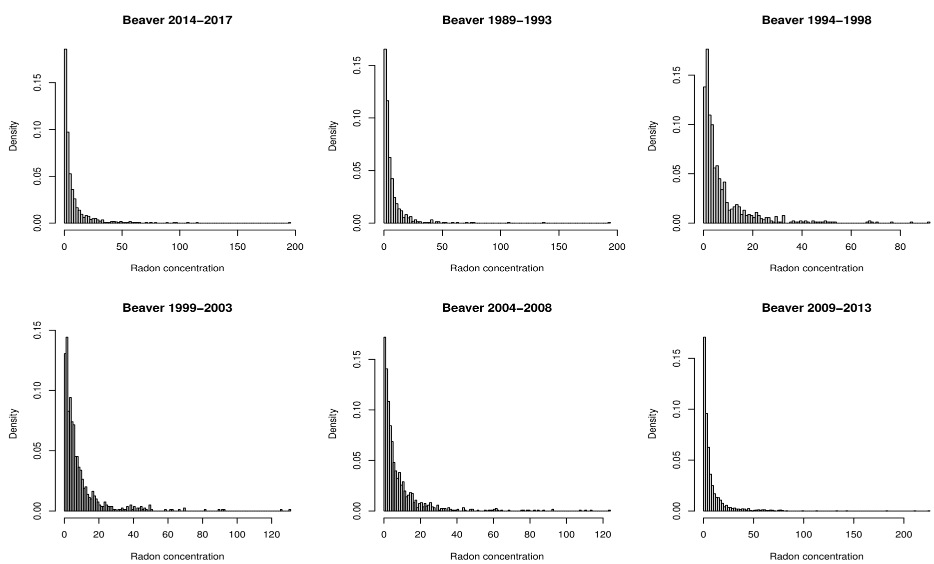

Figure 1.1: Histograms of Beaver residential radon concentration in six periods.

In order to depict the pattern of radon concentration closed to 0, we also plot the histograms truncated at 40 pCi/L in Figure 1.2

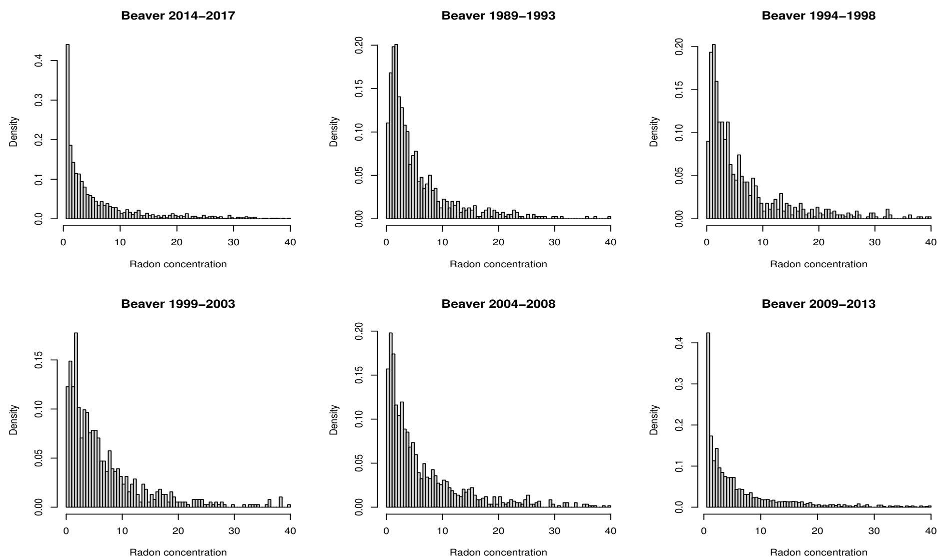

Figure 1.2: Histograms of Beaver residential radon concentration in six periods truncated at 40 pCi/L.

From Figure 1.1 and 1.2, we observe that data are positive and the distributions are right skewed. Gamma and Lognormal are two distributions frequently used to describe such data. Thus, we shall start with the uniform tilt $\boldsymbol{h}(x) = (x, \log(x), (\log(x))^2)^\top$, a tilt that represents the density ratio of Gamma and Lognormal densities.

Construct a two-sample DRM for each of the five pairs, $(\boldsymbol{X}^{(0)}, \boldsymbol{X}^{(1)})$,..., $(\boldsymbol{X}^{(0)}, \boldsymbol{X}^{(5)})$, using the uniform tilt $\boldsymbol{h}(x)$. Then test the parameters as introduced in Section 1.4 at significance level of 0.05. Remove the insignificant terms in the tilt, we obtain the reduced tilts such that

$$\boldsymbol{h}_1(x) = (x, \log(x), (\log(x))^2)^\top, \;\; \boldsymbol{h}_2(x) = \log(x),$$

$$\boldsymbol{h}_3(x) = (x, (\log(x))^2)^\top, \qquad\qquad \boldsymbol{h}_4(x) = \boldsymbol{h}_5(x) = 0.$$

Finally, construct a DRM with the variable tilts $\boldsymbol{h}_1(x), \ldots, \boldsymbol{h}_5(x)$, and obtain 95% CI for threshold probability $1 - G(T)$ with $T = 5, 10, 50, 100, 200$. The DRM estimator of G is denoted as $\widetilde{G}_v$. For comparison, we computed 95% CIs for these threshold probabilities using a DRM with the uniform tilt, that is,

$$\boldsymbol{h}_1(x) = \boldsymbol{h}_2(x) = \boldsymbol{h}_3(x) = \boldsymbol{h}_4(x) = \boldsymbol{h}_5(x) = \boldsymbol{h}_(x).$$

The corresponding estimator of G is denoted as $\widetilde{G}_u$. We also obtain the empirical estimator $\widehat{G}$ based on the reference sample, and computed the CIs for the threshold probabilities. From Table 1.3, it is observed that for each T, the shortest CI corresponds to the DRM with variable tilts. The CIs corresponds to the two DRMs are both shorter than that corresponds to the empirical distribution. Also, since no observation exceeds 200 in the reference sample, then the empirical distribution cannot provide an interval estimation for $1 - G(200)$. However, the DRMs can give such interval estimation in that there are observations that exceed 200 in the rest of samples. Indeed, these conclusions matches the ones obtained by the simulation in Section 1.3.

The above analysis is to estimate the distribution of residential radon concentration in Beaver county in one period based on the fused sample from all periods. One can also fix the time period, and estimate the radon concentration in Beaver county based on its neighboring counties. The four neighboring counties of Beaver are Lawrence, Butler, Allegheny and Washington. Let $\boldsymbol{X}^{(0)}, \ldots, \boldsymbol{X}^{(4)}$, be the samples from Beaver, Lawrence, Butler, Allegheny and Washington, respectively, during 2014-2017. Based the histograms in Figure 1.3 and the truncated ones in Figure 1.4, the samples from the five counties exhibit the same right skewed characteristic as the samples obtained during different periods in Beaver. Therefore, we still start with the uniform tilt

Table 1.3: A Comparison between $\widetilde{G}_v$, $\widetilde{G}_u$ and $\widehat{G}$ based on samples of six periods in Beaver county. Threshold probability estimates, 95% CIs and lengths of CIs are computed for thresholds $T = 5, 10, 50, 100, 200$. "-" indicates that some of the CIs generated can not be constructed since the estimates are 0.

T	$1 - \widetilde{G}_v(T)$	95% CI	CI length
5	0.3956	(0.3827, 0.4085)	0.0258
10	0.2132	(0.2024, 0.2239)	0.0214
50	0.0208	(0.0173, 0.0244)	0.0071
100	0.0027	(0.0015, 0.0040)	0.0026
200	0.0003	(−0.0001, 0.0006)	0.0008
T	$1 - \widetilde{G}_u(T)$	95% CI	CI length
5	0.3764	(0.3554, 0.3974)	0.0420
10	0.2019	(0.1846, 0.2192)	0.0347
50	0.0211	(0.0152, 0.0270)	0.0119
100	0.0030	(0.0009, 0.0051)	0.0042
200	0.0003	(−0.0002, 0.0008)	0.0010
T	$1 - \widehat{G}(T)$	95% CI	CI length
5	0.3764	(0.3528, 0.3999)	0.0471
10	0.2054	(0.1858, 0.2250)	0.0393
50	0.0197	(0.0129, 0.0264)	0.0135
100	0.0025	(0.0001, 0.0049)	0.0048
200	-	-	-

$\boldsymbol{h}(x) = (x, \log(x), (\log(x))^2)^\top$ and obtain the reduced ones by the significance tests such that

$$\boldsymbol{h}_1(x) = (\log(x), (\log(x))^2)^\top, \ \boldsymbol{h}_2(x) = (x, (\log(x))^2)^\top,$$

$$\boldsymbol{h}_3(x) = (x, (\log(x))^2)^\top, \qquad \boldsymbol{h}_4(x) = (x, \log(x), (\log(x))^2)^\top.$$

Again, obtain the point and interval estimates for the threshold probability $1 - G(T)$ with $T = 5, 10, 50, 100, 200$ as the case of fusing samples from six periods in Beaver. As shown in Table 1.4, the DRM with variable tilts produces a shorter CI for $T = 5, 10, 50$ while the DRM with a uniform tilt has a shorter CI for $T = 100, 200$. Also, the lengths of CIs obtained from the two DRMs are closed to each other for all thresholds. This is because the tilts are only reduced to

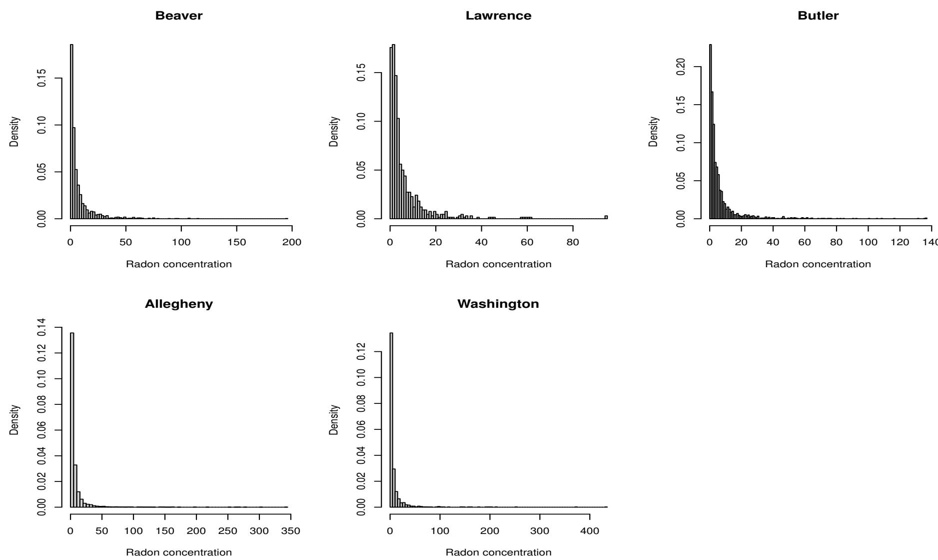

Figure 1.3: Histograms of residential radon concentration in Beaver, Lawrence, Butler, Allegheny and Washington.

a moderate extent. Only one term has been removed from $\boldsymbol{h}_1(x)$, $\boldsymbol{h}_2(x)$, and $\boldsymbol{h}_3(x)$, respectively. In the previous case, both $\boldsymbol{h}_4(x)$ and $\boldsymbol{h}_5(x)$ are reduced to 0, which makes a clear distinction between the two DRMs. Comparing to the empirical distribution, we can still draw the conclusion that the DRMs are better regrading the lengths of CIs and the ability to provide interval estimation for a relatively high threshold.

1.5.2 Forest County

For Forest county, there are only 47 observations over the six periods. Thus, in this case, we use the observations obtained from all time periods. Two neighboring counties of Forest, Warren and Elk, are selected, and they have 837 and 1191 observations over the six periods. Let $\boldsymbol{X}^{(0)}$, $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$, be the samples from Forest, Warren and Elk, respectively.

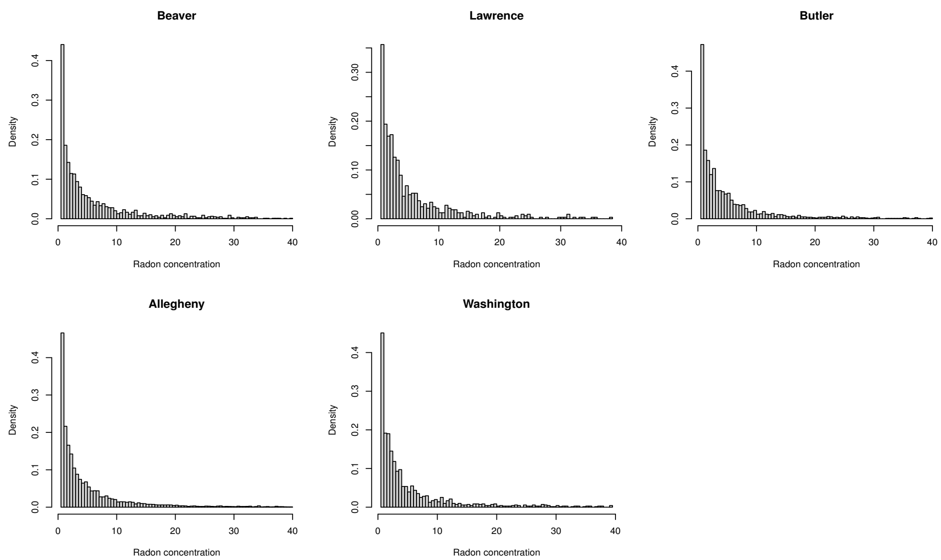

Figure 1.4: Histograms of residential radon concentration in Beaver, Lawrence, Butler, Allegheny and Washington truncated at 40 pCi/L.

Based on the histograms and the truncated ones in Figure 1.5, we still let

$$\boldsymbol{h}(x) = (x, \log(x), (\log(x))^2)^\top$$

to be the uniform tilt. Initially, we set $\boldsymbol{h}_1(x) = \boldsymbol{h}_2(x) = \boldsymbol{h}(x)$, and then obtain AIC values for DRMs with $\boldsymbol{h}_1(x)$ and $\boldsymbol{h}_2(x)$ taking their all possible reduced forms. The result in Table 1.5 indicates that smallest AIC value of 31677.07 is achieved by the DRM with tilts $\boldsymbol{h}_1(x) = (x, \log^2(x))^\top$ and $\boldsymbol{h}_2(x) = x$.

Then use these tilts as variable tilts, compare $\widetilde{G}_v$, $\widetilde{G}_u$ and $\widehat{G}$ as we did in the Beaver case. As given by Table 1.6, the DRM with variable tilts gives the best interval estimates regarding to lengths of CIs.

Table 1.4: A Comparison between $\widetilde{G}_v$, $\widetilde{G}_u$ and $\widehat{G}$ based on samples from Beaver, Lawrence, Butler, Allegheny and Washington. Threshold probability estimates, 95% CIs and lengths of CIs are computed for thresholds $T = 5, 10, 50, 100, 200$. "-" indicates that some of the CIs generated can not be constructed since the estimates are 0.

T	$1 - \widetilde{G}_v(T)$	95% CI	CI length
5	0.3810	$(0.3607, 0.4013)$	0.0406
10	0.2056	$(0.1883, 0.2228)$	0.0345
50	0.0205	$(0.0150, 0.0260)$	0.0110
100	0.0028	$(0.0010, 0.0047)$	0.0037
200	0.0001	$(-0.0001, 0.0003)$	0.0004
T	$1 - \widetilde{G}_u(T)$	95% CI	CI length
5	0.3802	$(0.3593, 0.4011)$	0.0418
10	0.2079	$(0.1903, 0.2255)$	0.0352
50	0.0201	$(0.0145, 0.0258)$	0.0113
100	0.0024	$(0.0007, 0.0042)$	0.0035
200	0.0001	$(-0.0001, 0.0002)$	0.0003
T	$1 - \widehat{G}(T)$	95% CI	CI length
5	0.3764	$(0.3528, 0.3999)$	0.0471
10	0.2054	$(0.1858, 0.2250)$	0.0393
50	0.0197	$(0.0129, 0.0264)$	0.0135
100	0.0025	$(0.0001, 0.0049)$	0.0048
200	-	-	-

1.6 Estimation of Regional Pertussis Rates in Washington State

In this section, we construct a multivariate DRM with variable tilts for regional pertussis rates. Pertussis, also known as whooping cough, is a highly contagious respiratory disease caused by bacterium Bordetella pertussis. Pertussis data from 1997 to 2018 are collected from annual communicable disease reports provided by Washington State Department of Heath. The reports contain county-level annual pertussis cases and rates (per 100,000 population). In [80], a DRM is applied to the analysis of regional pertussis cases. Since the number of cases varies drastically between the counties, then a DRM can provide interval estimation for a wider range of threshold. However, in such case, the number of cases are considered as a continuous random variable. To avoid this issue, we consider examining the regional pertussis rates instead of cases.

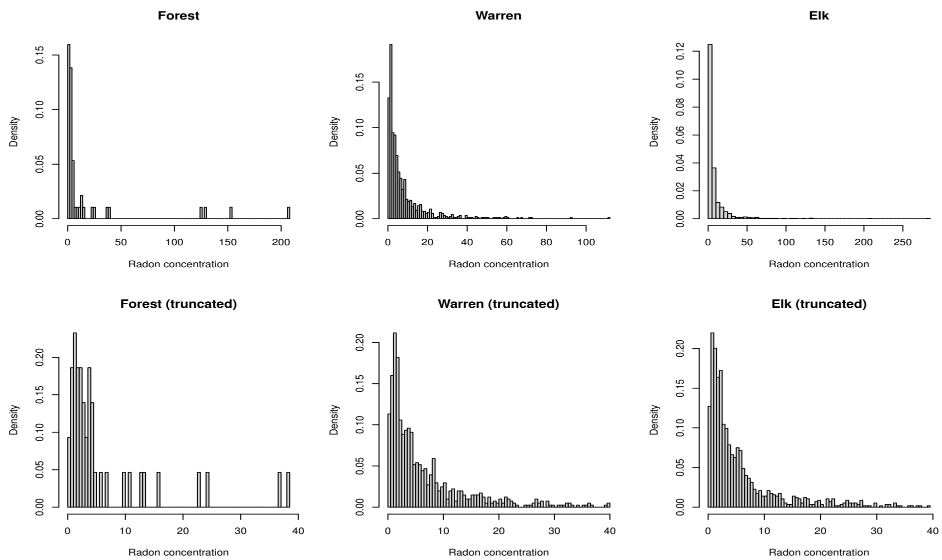

Figure 1.5: Histograms and truncated histograms (truncated at 40 pCi/L) of residential radon concentration in Forest, Warren and Elk.

Table 1.5: AIC values of models based on different choices of $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ in the Forest case. A hyphen "-" indicates that $\boldsymbol{h}_k(x) = 0$ and therefore g_0 and g_k are identical for $k = 1, 2$.

AIC $\boldsymbol{h}_1$ / $\boldsymbol{h}_2$	-	x	$\log(x)$	$\log^2(x)$	$(x, \log(x))$	$(x, \log^2(x))$	$(\log(x), \log^2(x))$	$(x, \log(x), \log^2(x))$
-	31696.52	31697.86	31694.68	31697.54	31686.85	31682.73	31694.35	31684.24
x	31698.24	31691.11	31695.63	31699.20	31680.96	31677.07	31696.32	31678.58
$\log(x)$	31693.46	31685.55	31695.07	31692.86	31687.35	31683.05	31694.81	31684.70
$\log^2(x)$	31695.67	31680.36	31696.67	31694.28	31680.14	31680.10	31691.31	31681.62
$(x, \log(x))$	31693.43	31684.21	31695.04	31694.63	31682.37	31679.01	31696.63	31680.02
$(x, \log^2(x))$	31693.13	31682.36	31695.03	31691.36	31681.38	31678.75	31690.98	31680.26
$(\log(x), \log^2(x))$	31695.11	31681.91	31696.71	31693.58	31680.03	31682.06	31691.40	31681.93
$(x, \log(x), \log^2(x))$	31694.44	31683.83	31696.05	31692.66	31680.67	31680.48	31690.13	31682.01

First, we need to insert the missing values of pertussis rates. The annual pertussis rates are not calculated when the number of cases are less than 5. Since the population does not vary in short time, the number of cases and rates in the year before and after are used to calculate the missing rates. Let c_t and r_t be the number of cases and rates in year t. If r_t is missing, then it is calculate by

$$r_t = \frac{c_t}{2}\left(\frac{r_{t-1}}{c_{t-1}} + \frac{r_{t+1}}{c_{t+1}}\right).$$

Let $\boldsymbol{X}^{(0)}$, $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$, be three bivariate samples from three regions: 0-(Clark, Cowlitz)

Table 1.6: Comparison between $\widetilde{G}_v$, $\widetilde{G}_u$ and $\widehat{G}$ based on samples from Forest, Warren and Elk. Threshold probability estimates, 95% CIs and lengths of CIs are computed for thresholds $T = 5, 10, 50, 100, 200$.

T	$1 - \widetilde{G}_v(T)$	95% CI	CI length
5	0.4447	$(0.3773, 0.5121)$	0.1349
10	0.2790	$(0.2004, 0.3577)$	0.1573
50	0.0915	$(0.0201, 0.1629)$	0.1429
100	0.0548	$(-0.0041, 0.1138)$	0.1178
200	0.0264	$(-0.0135, 0.0662)$	0.0798
T	$1 - \widetilde{G}_u(T)$	95% CI	CI length
5	0.4082	$(0.2881, 0.5284)$	0.2403
10	0.2565	$(0.1452, 0.3679)$	0.2227
50	0.0914	$(0.0179, 0.1649)$	0.1471
100	0.0580	$(-0.0038, 0.1198)$	0.1235
200	0.0296	$(-0.0155, 0.0746)$	0.0901
T	$1 - \widehat{G}(T)$	95% CI	CI length
5	0.3191	$(0.1859, 0.4524)$	0.2665
10	0.2553	$(0.1307, 0.3800)$	0.2493
50	0.0851	$(0.0053, 0.1649)$	0.1595
100	0.0851	$(0.0053, 0.1649)$	0.1595
200	0.0213	$(-0.0200, 0.0625)$	0.0825

(reference), 1-(Pierce, King) and 2-(Skagit, Whatcom). Thus, for example, $\boldsymbol{X}_j^{(0)}$ is a bivariate vector of which the first entry is the pertussis rates in Clark, and the second entry is that in Cowlitz. j goes from 1 to 22, representing the rates from 1997 to 2018. We consider that the neighboring counties are dependent while the three separate regions are independent. The summary statistics of the pertussis rates in the six counties are shown in Table 1.7.

Table 1.7: Summary statistics of pertussis rates in Clark, Cowlitz, Pierce, King, Skagit and Whatcom.

County	Clark	Cowlitz	Pierce	King	Skagit	Whatcom
Min.	0.9	0.0	2.9	2.0	1.0	5.0
Q1	5.5	3.2	7.1	5.6	4.1	13.0
Median	8.5	7.8	10.1	7.2	8.1	25.4
Mean	16.2	18.9	14.5	10.3	34.3	36.4
Q3	19.7	22.2	14.1	11.0	14.8	33.1
Max.	75.6	100.6	96.9	40.1	473.9	170.9

As shown in Figure 1.6, the histograms of the pertussis rates in the six counties appear to be right skewed so that the uniform tilt is chosen as $\boldsymbol{h}(\boldsymbol{x}) = (x_1, x_2, \log(x_1), \log(x_2 + 0.1))^\top$. Note that one observation in Cowlitz is 0 so that we need to use $\log(x_2 + 0.1)$ instead of $\log(x_2)$.

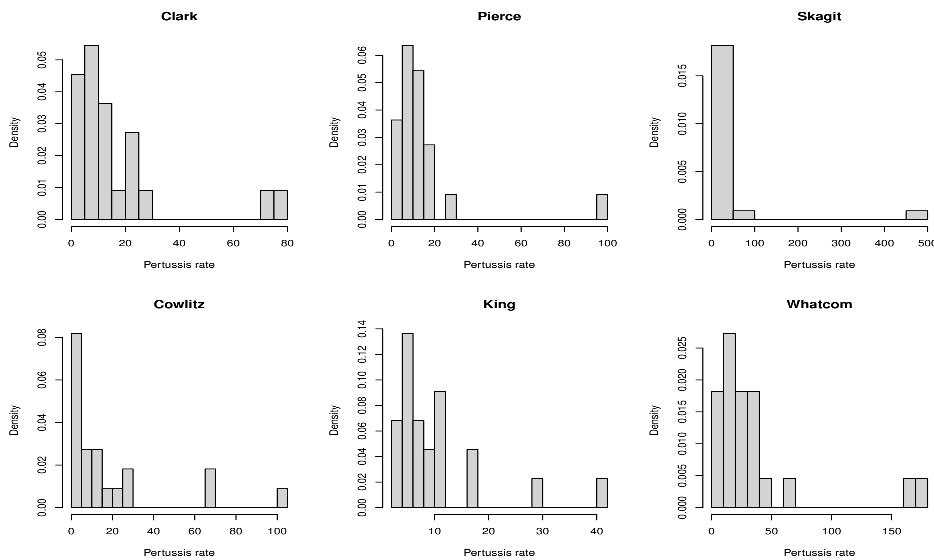

Figure 1.6: Histograms of pertussis rates in Clark, Cowlitz, Pierce, King, Skagit and Whatcom.

Based on the significance tests, the selected tilts are

$$\begin{aligned}\boldsymbol{h}_1(\boldsymbol{x}) =& (x_2, \log(x_2 + 0.1))^\top, \\ \boldsymbol{h}_2(\boldsymbol{x}) =& (x_1, x_2, \log(x_1), \log(x_2 + 0.1))^\top.\end{aligned}$$

Based on the summary statistics in Table 1.7, bivariate threshold probabilities are selected to compare the point and interval estimates of bivariate threshold probabilities obtained by the DRM with variable tilts ($\widetilde{P}_v$), the DRM with the uniform tilts ($\widetilde{P}_u$), and the empirical distribution ($\widehat{P}$). From Table 1.8, it is observed that the DRM with variable tilts yields the shortest CI for each selected threshold probability. When the reference sample does not contain any observation that can be used for constructing a CI for a threshold probability, the DRMs can still provide a CI as long as there are such observations in the fused sample.

Table 1.8: Comparison between $\widetilde{P}_v$, $\widetilde{P}_u$ and $\widehat{P}$ based on samples from (Clark, Cowlitz), (Pierce, King) and (Skagit, Whatcom). t_1 and t_2 represent Clark and Cowlitz, respectively. Point estimates, 95% CIs and lengths of CIs are computed for the selected threshold probabilities. "-" indicates that some of the CIs generated can not be constructed since the estimates are 0.

Threshold probability	$\widetilde{P}_v$	95% CI	CI length
$P(t_1 \leq 5.5, t_2 \leq 3.2)$	0.1450	$(0.0147, 0.2754)$	0.2607
$P(t_1 \leq 5.5, t_2 > 22.2)$	0.0119	$(-0.0083, 0.0321)$	0.0405
$P(t_1 \leq 8.5, t_2 > 7.8)$	0.1152	$(0.0263, 0.2040)$	0.1777
$P(t_1 > 16.2, t_2 \leq 18.9)$	0.0752	$(0.0071, 0.1433)$	0.1362
$P(t_1 > 75.6, t_2 > 100.6)$	0.0024	$(-0.0168, 0.0217)$	0.0385
Threshold probability	$\widetilde{P}_u$	95% CI	CI length
$P(t_1 \leq 5.5, t_2 \leq 3.2)$	0.1486	$(0.0157, 0.2815)$	0.2658
$P(t_1 \leq 5.5, t_2 > 22.2)$	0.0136	$(-0.0105, 0.0377)$	0.0482
$P(t_1 \leq 8.5, t_2 > 7.8)$	0.1240	$(0.0215, 0.2265)$	0.2050
$P(t_1 > 16.2, t_2 \leq 18.9)$	0.0686	$(-0.0028, 0.1399)$	0.1427
$P(t_1 > 75.6, t_2 > 100.6)$	0.0026	$(-0.0176, 0.0228)$	0.0404
Threshold probability	$\widehat{P}$	95% CI	CI length
$P(t_1 \leq 5.5, t_2 \leq 3.2)$	0.1364	$(-0.0070, 0.2798)$	0.2868
$P(t_1 \leq 5.5, t_2 > 22.2)$	-	-	-
$P(t_1 \leq 8.5, t_2 > 7.8)$	0.1364	$(-0.0070, 0.2798)$	0.2868
$P(t_1 > 16.2, t_2 \leq 18.9)$	0.0455	$(-0.0416, 0.1325)$	0.1741
$P(t_1 > 75.6, t_2 > 100.6)$	-	-	-

Chapter 2: A Density Ratio Model with Weakly Dependent Data

2.1 Introduction

In Chapter 1, we have thoroughly reviewed the history and relevant research works of DRM, and examined the proposed extension, a DRM with variable tilts. A basic assumption of DRM is that observations in each sample are IID. This is a natural assumption for empirical likelihood based methods. It is also fundamental for establishing asymptotic properties of DRM estimators $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{G}$. The former relies on SLLN and CLT for which IID assumption is a sufficient condition, and the latter depends on the empirical process defined by IID observations.

However, in application, it is often of interest to apply a DRM to data that have more complicated structures. It should be noted that a series of research works including [25], [47], [35] and [34], has been focusing on fusing residuals from a system of time series models by a DRM to provide predictions for the corresponding time series. As noted by [34], the residuals are dependent in general so that such method should be carried out with further means of alleviating the dependence. One of the ways to mitigate the dependence proposed in [34] is to sample from the residuals. In [65], a similar sampling idea has been implemented directly on time series instead of residuals from fitted time series models. The samples are used to test the homogeneity of their sources based on a DRM, and the result is shown to be satisfactory with a properly selected tilt function. Nevertheless, two issues still entail a further investigation. One is that a considerable amount of information is not utilized when we only work on the sampled data. The other is that the sampling scheme may not be effective to attenuate the dependence for certain types of time series. For example, it is less likely to reduce the dependence by sampling for periodical time series.

To resolve the issues mentioned above, it is tempting to relax the IID assumption to accommodate DRM to time series while maintaining asymptotic properties for the estimators. To avoid making the problem too broad, we focus on a particular yet widely used type of stationary time series, following the strong mixing condition. The strong mixing, or α-mixing condition was first introduced by [57]. The strong mixing coefficient is used to measure the dependence based on σ-algebras. Throughout this chapter, the definition of the strong mixing coefficient follows [55].

Definition 2.1. *For two σ-algebras, $\mathcal{A}$ and $\mathcal{B}$, the strong mixing coefficient $\alpha(\mathcal{A}, \mathcal{B})$ is defined by*

$$\alpha(\mathcal{A}, \mathcal{B}) = 2 \sup_{\substack{A \in \mathcal{A}, \\ B \in \mathcal{B}}} |P(A \cap B) - P(A)P(B)|.$$

Definition 2.2. *The strong mixing condition is satisfied by a strictly stationary time series $\boldsymbol{X}_t$ if the strong mixing coefficient*

$$\alpha_{\boldsymbol{X}}(n) = 2 \sup_{\substack{A \in \sigma(\boldsymbol{X}_i, i \leq j), \\ B \in \sigma(\boldsymbol{X}_i, i \geq j+n)}} |P(A \cap B) - P(A)P(B)| \to 0,$$

as $n \to \infty$.

Other mixing conditions such as β-mixing introduced by [61] and ρ-mixing introduced by [46], are proposed as alternative ways to characterize weak dependence based on σ-algebras. An extensive survey of various mixing conditions can be found in [4]. Alternative weak dependence conditions based on covariance can be found in [13].

Mixing conditions are commonly accompanied by strictly stationary condition since the mixing coefficient is defined based on σ-algebras. Observed from Definition 2.2, strictly stationary condition ensures that $\alpha_{\boldsymbol{X}}(n)$ is invariant with respect to j. Moreover, strictly stationary condition is also imperative to DRM since density ratio structures are assumed to be invariant over time. Strictly stationary and mixing conditions are also adopted in various empirical methods for time series. To name a few, [45] developed a blockwise empirical likelihood for strictly stationary and strong mixing time series. [8] introduced a goodness-of-fit test of parametric regression model against nonparametric alternatives based on empirical likelihood in the case where both responses and covariates are strictly stationary and strong mixing processes. [16] applied empirical likelihood to strictly stationary processes satisfying β-mixing condition. A broader review of empirical methods for time series can be referred to [50].

Thus, under that data are strictly stationary and strong mixing time series, we reexamine the DRM (1.1) in Chapter 1 with additional conditions to ensure the asymptotic properties of DRM estimators $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{G}$. In Section 2.2, we construct the DRM with observations assumed to be strictly stationary and strong mixing sequences. We shall refer this model as a DRM with weakly dependent data (DRMWD). Asymptotic properties of the estimators $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{G}$ are derived under our assumptions on the data and the model. In Section 2.3, we use simulation to verify the asymptotic properties of the estimators, and show the improvement made with dependence taken into consideration. In Section 2.4, the DRMWD is applied to air monitoring data to detect the structural changes over different time periods, and estimate the joint distribution of consecutive observations.

2.2 A Density Ratio Model with Weakly Dependent Data

The DRMWD resembles model (1.1) in Section 1.2 but additional assumptions are entailed. Thus, we shall reconstruct the model and supplement the assumptions as we proceed.

Consider $\{\boldsymbol{X}_j^{(k)}\}$, $j = 1, \dots, n_k$, $k = 0, \dots, m$, are $m+1$ d-dimensional processes. Denote the processes as

$$\boldsymbol{X}^{(k)} = (\boldsymbol{X}_1^{(k)}, \dots, \boldsymbol{X}_{n_k}^{(k)}), \quad k = 0, \dots, m,$$

where

$$\boldsymbol{X}_j^{(k)} = (X_{j,1}^{(k)}, \dots, X_{j,d}^{(k)})^\top, \quad j = 1, \dots, n_k.$$

Assumption 2.1. $\{\boldsymbol{X}_j^{(k)}\}$ *is a d-dimensional strictly stationary strong mixing process with mixing coefficient $\alpha_{\boldsymbol{X}^{(k)}}(n) \leq c_k(n+1)^{-a_k}$ for some $c_k \geq 1$ and $a_k > 1$, $k = 0, \dots, m$. $\{\boldsymbol{X}_j^{(k)}\}$'s are mutually independent processes. The CDF of $\boldsymbol{X}_j^{(k)}$ is absolutely continuous, $k = 0, \dots, m$, $j = 1, \dots, n_k$.*

Let $\boldsymbol{X}^{(0)}$ be the reference process, and G_k be the CDF for $\boldsymbol{X}_j^{(k)}$, $j = 1, \dots, n_k$, $k = 0, \dots, m$. Also, denote $G_{k,i}$ as the ith marginal CDF of G_k for $i = 1, \dots, d$, $k = 0, \dots, m$.

Based on Assumption 2.1, each process is strictly stationary and strong mixing with the strong mixing coefficient approaching 0 algebraically as $n \to \infty$. $a_k > 1$ guarantees that the mixing coefficients are summable. The absolute continuity of the CDF is an underlying assumptions in the IID case since all observations are assumed to have a density. The independence between the processes follows the independence between random samples in the IID case.

Combine the $m+1$ processes and denote it as

$$\boldsymbol{t} = (\boldsymbol{X}^{(0)}, \ldots, \boldsymbol{X}^{(m)})$$

with size $n = \sum_{k=0}^{m} n_k$. Let $g_0, \ldots, g_m$, be the PDFs that correspond to $G_0, \ldots, G_m$, respectively. Assume that the PDFs follow the density ratio structure with variable tilts

$$\frac{g_k(\boldsymbol{x})}{g_0(\boldsymbol{x})} = \exp\left(\alpha_k + \boldsymbol{\beta}_k^\top \boldsymbol{h}_k(\boldsymbol{x})\right), \quad k = 1, \ldots, m. \tag{2.1}$$

Denote $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)^\top$, $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kp_k})^\top$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_m^\top)^\top$, $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$ and $\boldsymbol{h}_k(\cdot) = (h_{k1}(\cdot), \ldots, h_{kp_k}(\cdot))^\top$. Denote the reference CDF $G_0 = G$, and let $w_0(\cdot; \boldsymbol{\theta}) = 1$, $w_k(\cdot; \boldsymbol{\theta}) = \exp\left(\alpha_k + \boldsymbol{\beta}_k^\top \boldsymbol{h}_k(\cdot)\right)$ for $k = 1, \ldots, m$.

Assumption 2.2. $|h_{lp}(\boldsymbol{X}_j^{(k)})|$ *satisfies*

$$E|h_{lp}(\boldsymbol{X}_j^{(k)})|^{m_k} < \infty$$

for some $m_k > \frac{4a_k}{a_k - 1}$, $k = 0, \ldots, m$, $j = 1, \ldots, n_k$, $l = 1, \ldots, m$, $p = 1, \ldots, p_l$.

Assumption 2.3. *The DRM (2.1) is non-degenerate, that is, all parameters are non-zero and tilt functions are not linearly dependent.*

Assumption 2.2 is a modification of Assumption 1.1 to account for the weak dependence in the processes. The relationship between a_k and m_k reveals the trade off between dependence and moments, the existence of higher moments is entailed for a more dependent process. Moreover, summation and product of $|h_{lp}(\boldsymbol{X}_j^{(k)})|$ for different l, p can also be ensured to have continuous

CDFs by Theorem 3 (p.134) in [56]. Assumption 2.3 is identical to Assumption 1.2.

2.2.1 Estimation

The empirical likelihood and the estimation procedure follows Section 1.2.1. First, denote all observations in the combined sample as $\boldsymbol{t} = (\boldsymbol{t}_1, ..., \boldsymbol{t}_n)$. Let $p_i = dG(\boldsymbol{t}_i)$ for $i = 1, \ldots, n$. Write the empirical likelihood as

$$L(\boldsymbol{\theta}, G) = \prod_{i=1}^{n} p_i \prod_{k=1}^{m} \prod_{j=1}^{n_k} w_k(\boldsymbol{X}_j^{(k)}; \boldsymbol{\theta}),$$

and the constraints as

$$\sum_{i=1}^{n} p_i = 1, \qquad \sum_{i=1}^{n} p_i[w_k(\boldsymbol{t}_i; \boldsymbol{\theta}) - 1] = 0, \qquad k = 1, \ldots, m.$$

Use profiling to obtain

$$p_i = \frac{1}{n_0 \sum_{k=0}^{m} \rho_k w_k(\boldsymbol{t}_i; \boldsymbol{\theta})}$$

for $i = 1, \ldots, n$, in which $\rho_k = \frac{n_k}{n_0}$ for $k = 0, \ldots, m$. Then obtain $\widetilde{\boldsymbol{\theta}}$ by

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha_k} = -\sum_{i=1}^{n} \frac{\rho_k w_k(\boldsymbol{t}_i; \boldsymbol{\theta})}{\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{t}_i; \boldsymbol{\theta})} + n_k = 0,$$
$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_k} = -\sum_{i=1}^{n} \frac{\rho_k w_k(\boldsymbol{t}_i; \boldsymbol{\theta}) \boldsymbol{h}_k(\boldsymbol{t}_i)}{\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{t}_i; \boldsymbol{\theta})} + \sum_{l=1}^{n_k} \boldsymbol{h}_k(\boldsymbol{X}_l^{(k)}) = \boldsymbol{0},$$

where

$$l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}, G) = -\sum_{i=1}^{n} \log\left(n_0 \sum_{k=0}^{m} \rho_k w_k(\boldsymbol{t}_i; \boldsymbol{\theta})\right) + \sum_{k=1}^{m}\left(n_k \alpha_k + \sum_{j=1}^{n_k} \boldsymbol{\beta}_k^\top \boldsymbol{h}_k(\boldsymbol{X}_j^{(k)})\right).$$

Finally, obtain

$$\widetilde{p}_i = \frac{1}{n_0 \sum_{k=0}^{m} \rho_k w_k(\boldsymbol{t}_i; \widetilde{\boldsymbol{\theta}})}$$

for $i = 1, \ldots, n$, and

$$\widetilde{G}(\boldsymbol{x}; \widetilde{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \widetilde{p}_i I_{\prod_{j=1}^{d}(-\infty, x_d]}(\boldsymbol{t}_i).$$

2.2.2 Asymptotic Behavior of $\widetilde{\boldsymbol{\theta}}$

Following Section 1.2.2, we first establish the strong consistency of $\frac{1}{n}\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ and $-\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ in Section 2.2.2.1. Then, obtain the asymptotic normality of $\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ in Section 2.2.2.2. Finally, we show the strong consistency and asymptotic normality of $\widetilde{\boldsymbol{\theta}}$ in Section 2.2.2.3, following Section 1.2.2.3 and 1.2.2.4.

2.2.2.1 Weak Consistency of Score and Hessian

In Section 1.2.2.1, the strong consistency of $\frac{1}{n}\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ and $-\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is established based on SLLN for IID observations. Though the independence assumption no longer holds in this case, SLLN can still be established based on strictly stationary condition and ergodicity. The following lemma and theorem are stated here to simplify citing while we establish the strong consistency of $\frac{1}{n}\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ and $-\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$.

Lemma 2.1. *If $\{\boldsymbol{X}_t\}$ is a d-dimensional strictly stationary and strong mixing process and $f : \mathbb{R}^d \to \mathbb{R}$ is a Borel function, then $\{f(\boldsymbol{X}_t)\}$ is also a strictly stationary and strong mixing process with $\alpha_{f\circ\boldsymbol{X}}(n) \leq \alpha_{\boldsymbol{X}}(n)$.*

Proof. Since f is a Borel function, then

$$\{f^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\} \subseteq \mathcal{B}(\mathbb{R}^d).$$

Subsequently,

$$\begin{aligned}\sigma(f(\boldsymbol{X}_t)) =&\sigma(\{\boldsymbol{X}_t^{-1} \circ f^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\}) \\ \subseteq&\sigma(\{\boldsymbol{X}_t^{-1}(B) : B \in \mathcal{B}(\mathbb{R}^d)\}) = \sigma(\boldsymbol{X}_t),\end{aligned}$$

and then $\forall\, T \subseteq \mathbb{Z}$,

$$\sigma(f(\boldsymbol{X}_t), t \in T) =\sigma\left(\bigcup_{t\in T} \sigma(f(\boldsymbol{X}_t))\right) \subseteq \sigma\left(\bigcup_{t\in T} \sigma(\boldsymbol{X}_t)\right) = \sigma(\boldsymbol{X}_t, t \in T).$$

Therefore,

$$\sup_{\substack{A\in\sigma(f(\boldsymbol{X}_t),t\leq j),\\ B\in\sigma(f(\boldsymbol{X}_t),t\geq j+n)}} |P(A\cap B) - P(A)P(B)| \leq \sup_{\substack{A\in\sigma(\boldsymbol{X}_t,t\leq j),\\ B\in\sigma(\boldsymbol{X}_t,t\geq j+n)}} |P(A\cap B) - P(A)P(B)|,$$

and the mixing condition is proved. The strictly stationary condition can be proved by $\forall\ k \in$

$\mathbb{Z}^+$, $\forall\, t_1, \ldots, t_k \in \mathbb{Z}$, $\forall\, s \in \mathbb{Z}$, $\forall\, A_1, \ldots, A_k \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} P(f(\boldsymbol{X}_{t_1}) \in A_1, \ldots, f(\boldsymbol{X}_{t_k}) \in A_k) =& P(\boldsymbol{X}_{t_1} \in f^{-1}(A_1), \ldots, \boldsymbol{X}_{t_k} \in f^{-1}(A_k)) \\ =& P(\boldsymbol{X}_{t_1+s} \in f^{-1}(A_1), \ldots, \boldsymbol{X}_{t_k+s} \in f^{-1}(A_k)) \\ =& P(f(\boldsymbol{X}_{t_1+s}) \in A_1, \ldots, f(\boldsymbol{X}_{t_k+s}) \in A_k). \end{aligned}$$

□

Theorem 2.1. *If $\{\boldsymbol{X}_t\}$ is a d-dimensional strictly stationary and strong mixing process and $f : \mathbb{R}^d \to \mathbb{R}$ satisfies*

$$E|f(\boldsymbol{X}_t)| < \infty,$$

then

$$\frac{1}{n}\sum_{t=1}^{n} f(\boldsymbol{X}_t) \overset{a.s.}{\to} Ef(\boldsymbol{X}_1).$$

Proof. By Lemma 2.1, $\{f(\boldsymbol{X}_t)\}$ is strictly stationary and strong mixing. As noted by [28] (pp. 488-489), for strictly stationary processes, the strong mixing condition implies ergodicity. Thus, the almost sure convergence is proved by Theorem 5.5 in [28]. □

Recall the expression of $\frac{1}{n}\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ from Section 1.2.2.1,

$$\begin{aligned} \frac{1}{n}\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha_k} &= -\frac{n_0}{n}\sum_{h=0}^{m} \rho_h \frac{1}{n_h}\sum_{l=1}^{n_h} \frac{\rho_k w_k(\boldsymbol{X}_l^{(h)}; \boldsymbol{\theta})}{\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{X}_l^{(h)}; \boldsymbol{\theta})} + \frac{\rho_k}{\sum_{j=0}^{m} \rho_j}, \\ \frac{1}{n}\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_k} &= -\frac{n_0}{n}\sum_{h=0}^{m} \rho_h \frac{1}{n_h}\sum_{l=1}^{n_h} \frac{\rho_k w_k(\boldsymbol{X}_l^{(h)}; \boldsymbol{\theta}) \boldsymbol{h}_k(\boldsymbol{X}_l^{(h)})}{\sum_{j=0}^{m} \rho_j w_j(\boldsymbol{X}_l^{(h)}; \boldsymbol{\theta})} + \frac{n_k}{n}\frac{1}{n_k}\sum_{l=1}^{n_k} \boldsymbol{h}_k(\boldsymbol{X}_l^{(k)}). \end{aligned}$$

By Theorem 2.1 and Assumption 2.2,

$$\frac{1}{n}\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \overset{a.s.}{\to} \mathbf{0}, \tag{2.2}$$

as $n \to \infty$, following the same derivation in Section 1.2.2.1. Similarly, the almost sure convergence (1.8) can be established once again

$$-\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \overset{a.s.}{\to} \boldsymbol{S}(\boldsymbol{\theta}), \tag{2.3}$$

as $n \to \infty$ where $\boldsymbol{S}(\boldsymbol{\theta})$ is given by Section 1.2.2.1.

2.2.2.2 Asymptotic Normality of Score

Denote $\varphi_k(\cdot;\boldsymbol{\theta}) = \frac{\rho_k w_k(\cdot;\boldsymbol{\theta})}{\sum_{j=0}^m \rho_j w_j(\cdot;\boldsymbol{\theta})}$ for $k = 1, \ldots, m$, then the expression of $\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ can be written as

$$\begin{aligned}
\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha_k} =& \sum_{h=0}^m \sqrt{\frac{n_h}{n}} \left(\frac{1}{\sqrt{n_h}} \sum_{l=1}^{n_h} \left(\frac{\rho_k}{\sum_{j=0}^m \rho_j} - \frac{\rho_k w_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})}{\sum_{j=0}^m \rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})} \right) \right) \\
=& \sum_{h=0}^m \sqrt{\frac{\rho_h}{\sum_{j=0}^m \rho_j}} \left(\frac{1}{\sqrt{n_h}} \sum_{l=1}^{n_h} \left(\frac{\rho_k}{\sum_{j=0}^m \rho_j} - \varphi_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}) \right) \right), \\
\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \beta_{kp}} =& \sqrt{\frac{n_k}{n}} \left(\frac{1}{\sqrt{n_k}} \sum_{l=1}^{n_k} \frac{\sum_{j\neq k, j=0}^m \rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta})}{\sum_{j=0}^m \rho_j w_j(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta})} h_{kp}(\boldsymbol{X}_l^{(k)}) \right) \\
&- \sum_{h\neq k, h=0}^m \sqrt{\frac{n_h}{n}} \left(\frac{1}{\sqrt{n_h}} \sum_{l=1}^{n_h} \frac{\rho_k w_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})}{\sum_{j=0}^m \rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})} h_{kp}(\boldsymbol{X}_l^{(h)}) \right) \\
=& \sqrt{\frac{\rho_k}{\sum_{j=0}^m \rho_j}} \left(\frac{1}{\sqrt{n_k}} \sum_{l=1}^{n_k} \left(1 - \varphi_k(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta})\right) h_{kp}(\boldsymbol{X}_l^{(k)}) \right) \\
&+ \sum_{h\neq k, h=0}^m \sqrt{\frac{\rho_h}{\sum_{j=0}^m \rho_j}} \left(\frac{1}{\sqrt{n_h}} \sum_{l=1}^{n_h} \left(-\varphi_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}) h_{kp}(\boldsymbol{X}_l^{(h)}) \right) \right),
\end{aligned}$$

for $k = 1, \ldots, m$, and $p = 1, \ldots, p_k$.

Assumption 2.1 and 2.2 satisfy the condition of Theorem 18.5.3 in [26] and hence Theorem 18.5.3 can be applied to $\frac{1}{\sqrt{n}} \frac{\partial l(\boldsymbol{\theta})}{\partial \alpha_k}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ and $\frac{1}{\sqrt{n}} \frac{\partial l(\boldsymbol{\theta})}{\partial \beta_{kp}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$. The Theorem 18.5.3 in [26] is stated as a univariate result and the multivariate case is given by the following theorem which is an immediate result of Theorem 18.5.3 in [26].

Theorem 2.2. *$\{\boldsymbol{X}_t\}$ is a d-dimensional strictly stationary and strong mixing process. Let $\boldsymbol{g} = (g_1, \ldots, g_p)^\top : \mathbb{R}^d \to \mathbb{R}^p$ be a nondegenerate function, that is,*

$$\forall\, \boldsymbol{x},\ \sum_{i=1}^{p} c_i g_i(\boldsymbol{x}) = 0 \ \Rightarrow\ \forall\, i = 1, \ldots, p,\ c_i = 0.$$

Suppose the strong mixing coefficient $\alpha_{\boldsymbol{X}}(n) \leq c(n+1)^{-a}$ for some $c \geq 1$ and $a > 1$, and

$$E|g_i(\boldsymbol{X}_t)|^m < \infty$$

for some $m > \frac{2a}{a-1}$, $\forall\, i = 1, \ldots, p$, then

$$\sqrt{n}\left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{g}(\boldsymbol{X}_t) - E\boldsymbol{g}(\boldsymbol{X}_1)\right) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{V}),$$

as $n \to \infty$, where $\boldsymbol{V}$ is a $p \times p$ matrix with

$$\boldsymbol{V}_{ij} = 2\pi f_{ij}(0)$$

for $i, j = 1, \ldots, p$, and $f_{ij}(0)$ is the cross spectral density of $\{g_i(\boldsymbol{X}_t)\}$ and $\{g_j(\boldsymbol{X}_t)\}$ evaluated at 0.

Proof. $\forall\, \boldsymbol{l} = (l_1, \ldots, l_p)^\top$ such that $\|\boldsymbol{l}\| \neq 0$,

$$\begin{aligned}
\frac{1}{n}\mathrm{Var}\left(\sum_{t=1}^{n} \boldsymbol{l}^\top \boldsymbol{g}(\boldsymbol{X}_t)\right) =& \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\mathrm{Cov}(\boldsymbol{l}^\top \boldsymbol{g}(\boldsymbol{X}_t), \boldsymbol{l}^\top \boldsymbol{g}(\boldsymbol{X}_s)) \\
=& \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{i=1}^{p}\sum_{j=1}^{p} l_i l_j \mathrm{Cov}(g_i(\boldsymbol{X}_t), g_j(\boldsymbol{X}_s)) \\
=& \sum_{i=1}^{p}\sum_{j=1}^{p} l_i l_j \left(\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\mathrm{Cov}(g_i(\boldsymbol{X}_t), g_j(\boldsymbol{X}_s))\right) \\
\to& \sum_{i=1}^{p}\sum_{j=1}^{p} l_i l_j \left(2\pi f_{ij}(0)\right) = \boldsymbol{l}^\top \boldsymbol{V} \boldsymbol{l},
\end{aligned}$$

as $n \to \infty$. By Lemma 2.1 and Theorem 18.5.3 in [26],

$$\sqrt{n}\left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{l}^\top \boldsymbol{g}(\boldsymbol{X}_t) - \boldsymbol{l}^\top E\boldsymbol{g}(\boldsymbol{X}_1)\right) \xrightarrow{d} N(0, \boldsymbol{l}^\top \boldsymbol{V} \boldsymbol{l}),$$

as $n \to \infty$. Since $\boldsymbol{l}$ is arbitrary, then

$$\sqrt{n}\left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{g}(\boldsymbol{X}_t) - E\boldsymbol{g}(\boldsymbol{X}_1)\right) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{V}),$$

as $n \to \infty$. □

By Theorem 2.2 and Assumption 2.1-2.3,

$$\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\mathcal{V}}(\boldsymbol{\theta}_0)), \tag{2.4}$$

as $n \to \infty$ for some $\boldsymbol{\mathcal{V}}(\boldsymbol{\theta}_0)$. It remains to show the entries of $\boldsymbol{\mathcal{V}}(\boldsymbol{\theta}_0)$.

$$
\begin{aligned}
&\operatorname{Cov}\left(\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha_k}, \frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha_{k'}}\right)\\
=&\sum_{h=0}^{m}\frac{\rho_h}{\sum_{j=0}^{m}\rho_j}\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\sum_{l'=1}^{n_h}\operatorname{Cov}\left(\frac{\rho_k}{\sum_{j=0}^{m}\rho_j}-\varphi_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}), \frac{\rho_{k'}}{\sum_{j=0}^{m}\rho_j}-\varphi_{k'}(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta})\right)\right)\\
=&\sum_{h=0}^{m}\frac{\rho_h}{\sum_{j=0}^{m}\rho_j}\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\sum_{l'=1}^{n_h}\operatorname{Cov}\left(\varphi_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}), \varphi_{k'}(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta})\right)\right)\\
\to&\frac{2\pi\sum_{h=0}^{m}\rho_h f_{\alpha_k,\alpha_{k'},h}(0;\boldsymbol{\theta})}{\sum_{j=0}^{m}\rho_j},
\end{aligned}
$$

as $n \to \infty$ for $k, k' = 1, \ldots, m$ and $f_{\alpha_k,\alpha_{k'},h}$ is the cross spectral density of $\{\varphi_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})\}$ and $\{\varphi_{k'}(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta})\}$.

$$
\begin{aligned}
&\operatorname{Cov}\left(\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha_k}, \frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \beta_{k'p}}\right)\\
=&\frac{\rho_{k'}}{\sum_{j=0}^{m}\rho_j}\left(\frac{1}{n_{k'}}\sum_{l=1}^{n_{k'}}\sum_{l'=1}^{n_{k'}}\operatorname{Cov}\left(\varphi_k(\boldsymbol{X}_l^{(k')};\boldsymbol{\theta}), (\varphi_{k'}(\boldsymbol{X}_{l'}^{(k')};\boldsymbol{\theta})-1)h_{k'p}(\boldsymbol{X}_{l'}^{(k')})\right)\right)\\
&+\sum_{h\neq k', h=0}^{m}\frac{\rho_h}{\sum_{j=0}^{m}\rho_j}\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\sum_{l'=1}^{n_h}\operatorname{Cov}\left(\varphi_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}), \varphi_{k'}(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta})h_{k'p}(\boldsymbol{X}_{l'}^{(h)})\right)\right)\\
\to&\frac{2\pi\sum_{h=0}^{m}\rho_h f_{\alpha_k,\beta_{k'p},h}(0;\boldsymbol{\theta})}{\sum_{j=0}^{m}\rho_j},
\end{aligned}
$$

as $n \to \infty$ for $k, k' = 1, \ldots, m$, $p = 1, \ldots, p_{k'}$. $f_{\alpha_k,\beta_{k'p},k'}$ is the cross spectral density of $\{\varphi_k(\boldsymbol{X}_l^{(k')};\boldsymbol{\theta})\}$ and $\{(\varphi_{k'}(\boldsymbol{X}_{l'}^{(k')};\boldsymbol{\theta})-1)h_{k'p}(\boldsymbol{X}_{l'}^{(k')})\}$ while $f_{\alpha_k,\beta_{k'p},h}$ is the cross spectral density of $\{\varphi_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})\}$ and $\{\varphi_{k'}(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta})h_{k'p}(\boldsymbol{X}_{l'}^{(h)})\}$ for $h \neq k', h = 0, \ldots, m$. Note that

$$
f_{\alpha_k,\beta_{k'p},h}(0;\boldsymbol{\theta}) = f_{\beta_{k'p},\alpha_k,h}(0;\boldsymbol{\theta})
$$

for $k, k' = 1, \ldots, m$, $p = 1, \ldots, p_{k'}$, $h = 0, \ldots, m$.

$$
\begin{aligned}
&\mathrm{Cov}\left(\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \beta_{kp}}, \frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \beta_{kp'}}\right) \\
=&\frac{\rho_k}{\sum_{j=0}^m \rho_j}\left(\frac{1}{n_k}\sum_{l=1}^{n_k}\sum_{l'=1}^{n_k}\mathrm{Cov}\left((\varphi_k(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta})-1)h_{kp}(\boldsymbol{X}_l^{(k)}), (\varphi_k(\boldsymbol{X}_{l'}^{(k)};\boldsymbol{\theta})-1)h_{kp'}(\boldsymbol{X}_{l'}^{(k)})\right)\right) \\
&+\sum_{h\neq k, h=0}^{m}\frac{\rho_h}{\sum_{j=0}^m \rho_j}\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\sum_{l'=1}^{n_h}\mathrm{Cov}\left(\varphi_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})h_{kp}(\boldsymbol{X}_l^{(h)}), \varphi_k(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta})h_{kp'}(\boldsymbol{X}_{l'}^{(h)})\right)\right) \\
\to&\frac{2\pi\sum_{h=0}^m \rho_h f_{\beta_{kp},\beta_{kp'},h}(0;\boldsymbol{\theta})}{\sum_{j=0}^m \rho_j},
\end{aligned}
$$

as $n \to \infty$ for $k = 1, \ldots, m$, $p, p' = 1, \ldots, p_k$. $f_{\beta_{kp},\beta_{kp'},k}$ is the cross spectral density of $\{(\varphi_k(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}) - 1)h_{kp}(\boldsymbol{X}_l^{(k)})\}$ and $\{(\varphi_k(\boldsymbol{X}_{l'}^{(k)};\boldsymbol{\theta}) - 1)h_{kp'}(\boldsymbol{X}_{l'}^{(k)})\}$ while $f_{\beta_{kp},\beta_{kp'},h}$ is the cross spectral density of $\{\varphi_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})h_{kp}(\boldsymbol{X}_l^{(h)})\}$ and $\{\varphi_k(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta})h_{kp'}(\boldsymbol{X}_{l'}^{(h)})\}$ for $h \neq k, h = 0, \ldots, m$.

$$
\begin{aligned}
&\mathrm{Cov}\left(\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \beta_{kp}}, \frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial \beta_{k'p'}}\right) \\
=&\frac{\rho_k}{\sum_{j=0}^m \rho_j}\left(\frac{1}{n_k}\sum_{l=1}^{n_k}\sum_{l'=1}^{n_k}\mathrm{Cov}\left((\varphi_k(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta})-1)h_{kp}(\boldsymbol{X}_l^{(k)}), \varphi_{k'}(\boldsymbol{X}_{l'}^{(k)};\boldsymbol{\theta})h_{k'p'}(\boldsymbol{X}_{l'}^{(k)})\right)\right) \\
&+\frac{\rho_{k'}}{\sum_{j=0}^m \rho_j}\left(\frac{1}{n_{k'}}\sum_{l=1}^{n_{k'}}\sum_{l'=1}^{n_{k'}}\mathrm{Cov}\left(\varphi_k(\boldsymbol{X}_l^{(k')};\boldsymbol{\theta})h_{kp}(\boldsymbol{X}_l^{(k')}), (\varphi_{k'}(\boldsymbol{X}_{l'}^{(k')};\boldsymbol{\theta})-1)h_{k'p'}(\boldsymbol{X}_{l'}^{(k')})\right)\right) \\
&+\sum_{h\neq k,k', h=0}^{m}\frac{\rho_h}{\sum_{j=0}^m \rho_j}\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\sum_{l'=1}^{n_h}\mathrm{Cov}\left(\varphi_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})h_{kp}(\boldsymbol{X}_l^{(h)}), \varphi_{k'}(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta})h_{k'p'}(\boldsymbol{X}_{l'}^{(h)})\right)\right) \\
\to&\frac{2\pi\sum_{h=0}^m \rho_h f_{\beta_{kp},\beta_{k'p'},h}(0;\boldsymbol{\theta})}{\sum_{j=0}^m \rho_j},
\end{aligned}
$$

as $n \to \infty$ for $k \neq k'$, $k, k' = 1, \ldots, m$, $p = 1, \ldots, p_k$, and $p_{k'} = 1, \ldots, p_{k'}$. $f_{\beta_{kp},\beta_{k'p'},k}$ is the cross spectral density of $\{(\varphi_k(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}) - 1)h_{kp}(\boldsymbol{X}_l^{(k)})\}$ and $\{\varphi_{k'}(\boldsymbol{X}_{l'}^{(k)};\boldsymbol{\theta})h_{k'p'}(\boldsymbol{X}_{l'}^{(k)})\}$. $f_{\beta_{kp},\beta_{k'p'},k'}$ is the cross spectral density of $\{\varphi_k(\boldsymbol{X}_l^{(k')};\boldsymbol{\theta})h_{kp}(\boldsymbol{X}_l^{(k')})\}$ and $\{(\varphi_{k'}(\boldsymbol{X}_{l'}^{(k')};\boldsymbol{\theta}) - 1)h_{k'p'}(\boldsymbol{X}_{l'}^{(k')})\}$;

$f_{\beta_{kp},\beta_{k'p'},h}$ is the cross spectral density of $\{\varphi_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})h_{kp}(\boldsymbol{X}_l^{(h)})\}$ and $\{\varphi_{k'}(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta})h_{k'p'}(\boldsymbol{X}_{l'}^{(h)})\}$ for $h \neq k, k'$, $h = 0, \ldots, m$.

Then $\boldsymbol{\mathcal{V}}(\boldsymbol{\theta}_0)$ is given by

$$\boldsymbol{\mathcal{V}}(\boldsymbol{\theta}_0) = \frac{2\pi}{\sum_{j=0}^{m} \rho_j} \sum_{h=0}^{m} \rho_h \boldsymbol{\mathcal{V}}_h(\boldsymbol{\theta}_0),$$

where

$$\boldsymbol{\mathcal{V}}_h(\boldsymbol{\theta}_0) = \begin{bmatrix} \boldsymbol{\mathcal{V}}_{11,h}(\boldsymbol{\theta}_0) & \boldsymbol{\mathcal{V}}_{12,h}(\boldsymbol{\theta}_0) \\ \boldsymbol{\mathcal{V}}_{12,h}^\top(\boldsymbol{\theta}_0) & \boldsymbol{\mathcal{V}}_{22,h}(\boldsymbol{\theta}_0) \end{bmatrix},$$

and

$$\boldsymbol{\mathcal{V}}_{11,h}(\boldsymbol{\theta}_0) = \begin{bmatrix} f_{\alpha_1,\alpha_1,h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\alpha_1,\alpha_m,h}(0;\boldsymbol{\theta}_0) \\ \vdots & \ddots & \vdots \\ f_{\alpha_m,\alpha_1,h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\alpha_m,\alpha_m,h}(0;\boldsymbol{\theta}_0) \end{bmatrix},$$

$$\begin{aligned} &\boldsymbol{\mathcal{V}}_{12,h}(\boldsymbol{\theta}_0) \\ &= \begin{bmatrix} f_{\alpha_1,\beta_{11},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\alpha_1,\beta_{1p_1},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\alpha_1,\beta_{m1},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\alpha_1,\beta_{mp_m},h}(0;\boldsymbol{\theta}_0) \\ \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ f_{\alpha_m,\beta_{11},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\alpha_m,\beta_{1p_1},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\alpha_m,\beta_{m1},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\alpha_m,\beta_{mp_m},h}(0;\boldsymbol{\theta}_0) \end{bmatrix}, \end{aligned}$$

$$\boldsymbol{\mathcal{V}}_{22,h}(\boldsymbol{\theta}_0) = \begin{bmatrix} f_{\beta_{11},\beta_{11},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\beta_{11},\beta_{1p_1},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\beta_{11},\beta_{m1},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\beta_{11},\beta_{mp_m},h}(0;\boldsymbol{\theta}_0) \\ \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ f_{\beta_{1p_1},\beta_{11},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\beta_{1p_1},\beta_{1p_1},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\beta_{1p_1},\beta_{m1},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\beta_{1p_1},\beta_{mp_m},h}(0;\boldsymbol{\theta}_0) \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ f_{\beta_{m1},\beta_{11},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\beta_{m1},\beta_{1p_1},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\beta_{m1},\beta_{m1},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\beta_{m1},\beta_{mp_m},h}(0;\boldsymbol{\theta}_0) \\ \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ f_{\beta_{mp_m},\beta_{11},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\beta_{mp_m},\beta_{1p_1},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\beta_{mp_m},\beta_{m1},h}(0;\boldsymbol{\theta}_0) & \cdots & f_{\beta_{mp_m},\beta_{mp_m},h}(0;\boldsymbol{\theta}_0) \end{bmatrix}.$$

2.2.2.3 Strong Consistency and Asymptotic Normality of $\widetilde{\boldsymbol{\theta}}$

The formula of $\left.\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is the same as that in the IID case so that it is still negative definite according to Section 1.2.2.3. Also, the strong consistency of $\left.\frac{1}{n}\frac{\partial l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is established once again by (2.2). Thus, the strong consistency of $\widetilde{\boldsymbol{\theta}}$ is established following Section 1.2.2.3. Then we can obtain (1.12) again

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\left(\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right)^{-1} \left(\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) + o_p(1),$$

and subsequently obtain

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\mathcal{U}}(\boldsymbol{\theta}_0)), \tag{2.5}$$

as $n \to \infty$, where $\boldsymbol{\mathcal{U}}(\boldsymbol{\theta}_0) = \boldsymbol{S}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{\mathcal{V}}(\boldsymbol{\theta}_0)\boldsymbol{S}^{-1}(\boldsymbol{\theta}_0)$.

2.2.3 Asymptotic Behavior of $\widetilde{G}$

In this section, we show the weak convergence of $\sqrt{n}(\widetilde{G} - G)$ under Assumption 2.1-2.3. The weak convergence of $\sqrt{n}(\widehat{G} - G)$ is established by Theorem 7.3 in [55]. Thus, it is sufficient to show the weak convergence of $\sqrt{n}(\widetilde{G} - \widehat{G})$, which resembles the IID case in Chapter 1. In the IID case, we first approximate $\widetilde{G}$ by $H_1 - H_2$. Next, the weak convergence of finite dimension distribution of $\sqrt{n}(H_1 - H_2 - \widehat{G})$ is established. Then, the weak convergence of $\sqrt{n}(H_1 - \widehat{G})$ and the tightness of $\sqrt{n}H_2$ are proved. Finally, based on these results, the weak convergence of $\sqrt{n}(\widetilde{G} - \widehat{G})$ is verified. Also, as stated by the remark in Section 1.2.3.5, an alternative way is to examine $\sqrt{n}(H_1 - H_2 - G)$ directly to show the weak convergence of $\sqrt{n}(\widetilde{G} - G)$. But since it is also mentioned in the remark that the limiting distribution of $\sqrt{n}(\widetilde{G} - \widehat{G})$ may be of interest, then we shall follow the above procedure to show the weak convergence of $\sqrt{n}(\widetilde{G} - \widehat{G})$ and $\sqrt{n}(\widetilde{G} - G)$.

In this case, the weak convergence of finite dimension distribution of $\sqrt{n}(H_1 - H_2 - \widehat{G})$ can be shown based on Theorem 2.2 and the tightness of $\sqrt{n}H_2$ can be verified based on Theorem 1.2. Both the approximation of $\widetilde{G}$ and the weak convergence of $\sqrt{n}(H_1 - \widehat{G})$ are based on Theorem 1.1 so that we shall first prove this theorem in the weakly dependent setting.

Theorem 2.3. *Let $\{\boldsymbol{X}_t\}$ be a d-dimensional strictly stationary and strong mixing process with the strong mixing coefficient $\alpha_{\boldsymbol{X}}(n) \leq c(n+1)^{-a}$ for some $c \geq 1$ and $a > 1$. Assume that $\boldsymbol{X}_t$ has an absolutely continuous CDF. Let $g : \mathbb{R}^d \to \mathbb{R}$ be a function satisfying*

$$E|g(\boldsymbol{X}_t)|^m < \infty$$

for some $m > \frac{2a}{a-1}$. Then the process

$$Z_n(\boldsymbol{x}) = \sqrt{n}\left(\frac{1}{n}\sum_{t=1}^{n} g(\boldsymbol{X}_t) I_{\prod_{j=1}^{d}[-\infty, x_j]}(\boldsymbol{X}_t) - Eg(\boldsymbol{X}_1) I_{\prod_{j=1}^{d}[-\infty, x_j]}(\boldsymbol{X}_1)\right)$$

converges weakly to a zero-mean Gaussian process.

Proof. This proof largely follows the proofs of Theorem 7.2 and Theorem 7.3 in [55]. Through out the proof, we use P^* and E^* for outer measure and outer expectation, respectively. The weak convergence of finite dimension distribution of $Z_n(\boldsymbol{x})$ can be proved directly by Theorem 2.2. If we can show the tightness of $Z_n(\boldsymbol{x})$, then the weak convergence is proved by Theorem 1.5.4 in [58]. Moreover, by Theorem 2.2,

$$\sqrt{n}\left(\frac{1}{n}\sum_{t=1}^{n} g(\boldsymbol{X}_t) - Eg(\boldsymbol{X}_1)\right) \xrightarrow{d} Z_g \sim N(0, 2\pi f(0)),$$

where $f(0)$ is the spectral density of $g(\boldsymbol{X}_t)$ evaluated at 0. Then, $\forall\, \eta > 0$, take $\varepsilon > \frac{E|Z_g|}{\eta}$, by Theorem 2.2,

$$\begin{aligned}\limsup_{n\to\infty} P(|Z_n(\boldsymbol{x})| > \varepsilon) \leq & \limsup_{n\to\infty} P\left(\left|\sqrt{n}\left(\frac{1}{n}\sum_{t=1}^{n} g(\boldsymbol{X}_t) - Eg(\boldsymbol{X}_1)\right)\right| > \varepsilon\right) \\ = & P(|Z_g| > \varepsilon) \leq \frac{E|Z_g|}{\varepsilon} < \eta.\end{aligned}$$

Thus, according to Theorem 1.5.7 in [58], it is left to show $\forall\, \varepsilon, \eta > 0$, $\exists\, \delta > 0$,

$$\limsup_{n\to\infty} P^*\left(\sup_{d(\boldsymbol{x},\boldsymbol{y})<\delta} |Z_n(\boldsymbol{x}) - Z_n(\boldsymbol{y})| > \varepsilon\right) < \eta$$

with $d(\boldsymbol{x}, \boldsymbol{y}) = \sup_{j=1,\ldots,d} |F_j(x_j) - F_j(y_j)|$ to complete the proof. Then it is sufficient to show

$$\lim_{\delta \to 0} \limsup_{n \to \infty} E^* \left(\sup_{d(\boldsymbol{x}, \boldsymbol{y}) < \delta} |Z_n(\boldsymbol{x}) - Z_n(\boldsymbol{y})| \right) = 0. \tag{2.6}$$

By the continuity of the CDF of $\boldsymbol{X}_t$,

$$\begin{aligned}
Z_n(\boldsymbol{x}) =& \sqrt{n} \left(\frac{1}{n} \sum_{t=1}^{n} g(\boldsymbol{X}_t) I[X_{t,1} \le x_1, \ldots, X_{t,d} \le x_d] \right. \\
& \left. - Eg(\boldsymbol{X}_1) I[X_{1,1} \le x_1, \ldots, X_{1,d} \le x_d] \right) \\
=& \sqrt{n} \left(\frac{1}{n} \sum_{t=1}^{n} g(\boldsymbol{X}_t) I[F_1(X_{t,1}) \le F_1(x_1), \ldots, F_d(X_{t,d}) \le F_d(x_d)] \right. \\
& \left. - Eg(\boldsymbol{X}_1) I[F_1(X_{1,1}) \le F_1(x_1), \ldots, F_d(X_{1,d}) \le F_d(x_d)] \right) \\
=& \sqrt{n} \left(\frac{1}{n} \sum_{t=1}^{n} g([F_1^{-}(U_{t,1}), \ldots, F_d^{-}(U_{t,d})]) I[U_{t,1} \le F_1(x_1), \ldots, U_{t,d} \le F_d(x_d)] \right. \\
& \left. - Eg([F_1^{-}(U_{1,1}), \ldots, F_d^{-}(U_{1,d})]) I[U_{1,1} \le F_1(x_1), \ldots, U_{1,d} \le F_d(x_d)] \right),
\end{aligned}$$

where $F_1, \ldots, F_d$ are marginal CDFs of $\boldsymbol{X}_t$, and $\boldsymbol{U}_t = (U_{t,1}, \ldots, U_{t,d})^\top$ has uniform marginals. Following [55], it is sufficient to show (2.6) when $\boldsymbol{X}_t$ has uniform marginals and the metric $d(\boldsymbol{x}, \boldsymbol{y}) = \sup_{j=1,\ldots,d} |x_j - y_j|$.

Let x be approximated by the base 2 such that $x = \sum_{k=1}^{K} b_k(x) 2^{-k} + r_K(x)$ with $b_k(x) = 0, 1$, and $0 \le r_K(x) \le 2^{-K}$. Let $B_{K,j}(x) = \sum_{k=1}^{K} b_{k,j}(x) 2^{-k}$ be the approximation of jth marginal

x_j, and $\boldsymbol{B}_K(\boldsymbol{x}) = (B_{K,1}(x_1), \ldots, B_{K,d}(x_d))^\top$. Then

$$\lim_{\delta\to 0}\limsup_{n\to\infty} E^*\left(\sup_{d(\boldsymbol{x},\boldsymbol{y})<\delta}|Z_n(\boldsymbol{x}) - Z_n(\boldsymbol{y})|\right)$$
$$= \lim_{K\to\infty}\limsup_{n\to\infty} E^*\left(\sup_{\boldsymbol{x}\in[0,1]^d}|Z_n(\boldsymbol{x}) - Z_n(\boldsymbol{B}_K(\boldsymbol{x}))|\right).$$

Let $\mathbb{Z}_n$ be the empirical measure of $\{\boldsymbol{X}_t\}$ so that $\forall\ A \in \mathcal{B}(\mathbb{R}^d)$,

$$\mathbb{Z}_n(|g|I_A) = \sqrt{n}\left(\frac{1}{n}\sum_{t=1}^{n}|g(\boldsymbol{X}_t)|I_A(\boldsymbol{X}_t) - E|g(\boldsymbol{X}_1)|I_A(\boldsymbol{X}_1)\right).$$

Let $\boldsymbol{L} = (L_1, \ldots, L_d)^\top$, and $\mathcal{D}_{\boldsymbol{L}}$ be the class of dyadic boxes $\prod_{j=1}^{d}((k_j-1)2^{-L_j}, k_j 2^{-L_j}]$, $k_j = 1, \ldots, 2^{L_j}$ (Notation 7.1 in [55]).

Let $\{\varepsilon_S\}_{S\in\mathcal{D}_{\boldsymbol{L}}}$ be IID random variables such that $P(\varepsilon_S = 1) = P(\varepsilon_S = -1) = \frac{1}{2}$ and they are independent of $\{\boldsymbol{X}_t\}$. Without the loss of generality, assume that $L_1 = \max(\boldsymbol{L})$. Let $M \in \mathbb{Z}\cap[1, L_1]$, $h \in \mathbb{Z}\cap[1, 2^M]$, and

$$\mathcal{D}_{\boldsymbol{L}}^h = \left\{\prod_{j=1}^{d}((k_j-1)2^{-L_j}, k_j 2^{-L_j}] : k_1 \in \{(h-1)2^{L_1-M}, \ldots, h2^{L_1-M}\}\right\}.$$

Denote K_x as the set of indices of dyadic boxes satisfying $|\mathbb{Z}_n(|g|I_S)| \geq x$ such that

$$K_x^h = \left\{\boldsymbol{k} : |\mathbb{Z}_n(|g|I_{\prod_{j=1}^{d}((k_j-1)2^{-L_j}, k_j 2^{-L_j}]})| \geq x, k_1 \in \{(h-1)2^{L_1-M}, \ldots, h2^{L_1-M}\}\right\},$$

and hence

$$P\left(\sup_{S\in\mathcal{D}_{\boldsymbol{L}}^h}|\mathbb{Z}_n(|g|I_S)| \geq x\right) = P(K_x^h \neq \emptyset).$$

Note that $\sup_{S\in\mathcal{D}_{\boldsymbol{L}}^h}|\mathbb{Z}_n(|g|I_S)|$ is a supreme over finite dyadic boxes so that $\{\sup_{S\in\mathcal{D}_{\boldsymbol{L}}^h}|\mathbb{Z}_n(|g|I_S)|\geq x\}$ is measurable. Let S_x^h represent the dyadic box with smallest indices (in each dimension) in K_x^h, then

$$
\begin{aligned}
&P\left(\left|\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h}\varepsilon_S\mathbb{Z}_n(|g|I_S)\right|\geq x\right)\\
\geq& P\left(\left|\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h,S\neq S_x^h}\varepsilon_S\mathbb{Z}_n(|g|I_S)+\varepsilon_{S_x^h}\mathbb{Z}_n(|g|I_{S_x^h})\right|\geq x\cap K_x^h\neq\emptyset\right)\\
=&P(\varepsilon_{S_x^h}=-1)P(K_x^h\neq\emptyset)P\left(\left|\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h,S\neq S_x^h}\varepsilon_S\mathbb{Z}_n(|g|I_S)+\varepsilon_{S_x^h}\mathbb{Z}_n(|g|I_{S_x^h})\right|\geq x\middle|K_x^h\neq\emptyset,\varepsilon_{S_x^h}=-1\right)\\
&+P(\varepsilon_{S_x^h}=1)P(K_x^h\neq\emptyset)P\left(\left|\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h,S\neq S_x^h}\varepsilon_S\mathbb{Z}_n(|g|I_S)+\varepsilon_{S_x^h}\mathbb{Z}_n(|g|I_{S_x^h})\right|\geq x\middle|K_x^h\neq\emptyset,\varepsilon_{S_x^h}=1\right)\\
=&\frac{1}{2}P(K_x^h\neq\emptyset)\left[P\left(\left|\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h,S\neq S_x^h}\varepsilon_S\mathbb{Z}_n(|g|I_S)-\mathbb{Z}_n(|g|I_{S_x^h})\right|\geq x\middle|K_x^h\neq\emptyset,\varepsilon_{S_x^h}=-1\right)\right.\\
&\left.+P\left(\left|\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h,S\neq S_x^h}\varepsilon_S\mathbb{Z}_n(|g|I_S)+\mathbb{Z}_n(|g|I_{S_x^h})\right|\geq x\middle|K_x^h\neq\emptyset,\varepsilon_{S_x^h}=1\right)\right]\\
=&\frac{1}{2}P(K_x^h\neq\emptyset)\left[P\left(\left|\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h,S\neq S_x^h}\varepsilon_S\mathbb{Z}_n(|g|I_S)-\mathbb{Z}_n(|g|I_{S_x^h})\right|\geq x\middle|K_x^h\neq\emptyset\right)\right.\\
&\left.+P\left(\left|\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h,S\neq S_x^h}\varepsilon_S\mathbb{Z}_n(|g|I_S)+\mathbb{Z}_n(|g|I_{S_x^h})\right|\geq x\middle|K_x^h\neq\emptyset\right)\right].
\end{aligned}
$$

Notice that $K_x^h\neq\emptyset\iff|\mathbb{Z}_n(|g|I_{S_x^h})|\geq x$ so that at least one of

$$
\begin{aligned}
\left|\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h,S\neq S_x^h}\varepsilon_S\mathbb{Z}_n(|g|I_S)-\mathbb{Z}_n(|g|I_{S_x^h})\right|\geq& x,\\
\left|\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h,S\neq S_x^h}\varepsilon_S\mathbb{Z}_n(|g|I_S)+\mathbb{Z}_n(|g|I_{S_x^h})\right|\geq& x,
\end{aligned}
$$

holds. Hence,

$$\begin{aligned}&\frac{1}{2}P(K_x^h \neq \emptyset)\left[P\left(\left|\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h, S\neq S_x}\varepsilon_S\mathbb{Z}_n(|g|I_S)+\mathbb{Z}_n(|g|I_{S_x^h})\right|\geq x\middle|K_x^h\neq\emptyset\right)\right.\\&\left.+P\left(\left|\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h, S\neq S_x^h}\varepsilon_S\mathbb{Z}_n(|g|I_S)-\mathbb{Z}_n(|g|I_{S_x^h})\right|\geq x\middle|K_x^h\neq\emptyset\right)\right]\\\geq&\frac{1}{2}P(K_x^h\neq\emptyset),\end{aligned}$$

and therefore

$$P\left(\sup_{S\in\mathcal{D}_{\boldsymbol{L}}^h}|\mathbb{Z}_n(|g|I_S)|\geq x\right)\leq 2P\left(\left|\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h}\varepsilon_S\mathbb{Z}_n(|g|I_S)\right|\geq x\right).$$

Thus,

$$\begin{aligned}P\left(\sup_{S\in\mathcal{D}_{\boldsymbol{L}}}|\mathbb{Z}_n(|g|I_S)|\geq x\right)=&P\left(\bigcup_{h=1}^{2^M}\sup_{S\in\mathcal{D}_{\boldsymbol{L}}^h}|\mathbb{Z}_n(|g|I_S)|\geq x\right)\\\leq&\sum_{h=1}^{2^M}P\left(\sup_{S\in\mathcal{D}_{\boldsymbol{L}}^h}|\mathbb{Z}_n(|g|I_S)|\geq x\right)\\\leq&2\sum_{h=1}^{2^M}P\left(\left|\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h}\varepsilon_S\mathbb{Z}_n(|g|I_S)\right|\geq x\right).\end{aligned}$$

Let

$$Y_t=\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h}\varepsilon_S|g(\boldsymbol{X}_t)|I[\boldsymbol{X}_t\in S],$$

and $Y_t^\dagger$ be the centered Y_t such that

$$Y_t^\dagger=\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h}\varepsilon_S(|g(\boldsymbol{X}_t)|I[\boldsymbol{X}_t\in S]-E|g(\boldsymbol{X}_t)|I[\boldsymbol{X}_t\in S]).$$

Then,

$$\frac{1}{\sqrt{n}}Y_t^{\dagger} = \sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h} \varepsilon_S \mathbb{Z}_n(|g|I_S),$$

and

$$\begin{aligned} s_n^2 =& \sum_{t=1}^{n}\sum_{s=1}^{n} |\mathrm{Cov}(Y_t^{\dagger}, Y_s^{\dagger}|\boldsymbol{\varepsilon})| \\ =& \sum_{t=1}^{n}\sum_{s=1}^{n} |\mathrm{Cov}(Y_t, Y_s|\boldsymbol{\varepsilon})| \\ \leq& 4\sum_{l=1}^{n}\int_0^1 (\alpha_Y^{-1}(u)\wedge n) Q_Y^2(u)du, \end{aligned}$$

by Corollary 1.1 in [55], where $Q_Y(u)$ is the quantile function defined as

$$Q_Y(u) = \inf\{s : P(|Y_1| > s) \leq u\},$$

and

$$\alpha_Y^{-1}(u) = \sum_{i=0}^{\infty} I[u \leq \alpha_Y(i)].$$

Since the marginals of $\boldsymbol{X}_t$ are assumed to be $U[0,1]$, then

$$\begin{aligned} P(|Y_t| > 0|\boldsymbol{\varepsilon}) =& P(|g(\boldsymbol{X}_t)| \neq 0 \cap X_{t,1} \in ((k_1-1)2^{-M}, k_1 2^{-M}]) \\ \leq& P(X_{t,1} \in ((k_1-1)2^{-M}, k_1 2^{-M}]) \\ =& 2^{-M}, \end{aligned}$$

and hence $Q_Y(u) = 0$ for $u \geq 2^{-M}$.

$$\begin{aligned}
& 4\sum_{t=1}^{n}\int_0^1 (\alpha_Y^{-1}(u) \wedge n) Q_Y^2(u)du \\
& \leq 4n\int_0^1 \alpha_Y^{-1}(u)Q_Y^2(u)du \\
& = 4n\int_0^1 \left(\lim_{k\to\infty}\sum_{i=0}^{k} I[u \leq \alpha_Y(i)]\right) Q_Y^2(u)du \\
\text{(Fatou's Lemma)} \quad & \leq 4n\sum_{i=0}^{\infty}\int_0^1 I[u \leq \alpha_Y(i)]Q_Y^2(u)du \\
& = 4n\sum_{i=0}^{\infty}\int_0^{2^{-M}} I[u \leq \alpha_Y(i)]Q_Y^2(u)du \\
\text{(Hölder's Inequality)} \quad & \leq 4n\sum_{i=0}^{\infty}\left(\int_0^{2^{-M}} I[u \leq \alpha_Y(i)]du\right)^{\frac{m-2}{m}} \left(\int_0^{2^{-M}} Q_Y^m(u)du\right)^{\frac{2}{m}} \\
& = 4n\sum_{i=0}^{\infty}\left(\min\left(2^{-M}, \alpha_Y(i)\right)\right)^{\frac{m-2}{m}} \left(\int_0^1 Q_Y^m(u)du\right)^{\frac{2}{m}} \\
\text{(Lemma 2.1)} \quad & \leq 4n(E|Y_1|^m)^{\frac{2}{m}}\sum_{i=0}^{\infty}\left(\min\left(2^{-M}, \alpha_{\boldsymbol{X}}(i)\right)\right)^{\frac{m-2}{m}} \\
(c \geq 1) \quad & \leq 4n(E|g(\boldsymbol{X}_1)|^m)^{\frac{2}{m}} c\left(\sum_{i=0}^{\lfloor 2^{\frac{M}{a}}\rfloor - 1} 2^{-\frac{M(m-2)}{m}} + \sum_{\lfloor 2^{\frac{M}{a}}\rfloor}^{\infty} (i+1)^{-\frac{a(m-2)}{m}}\right).
\end{aligned}$$

Since $m > \frac{2a}{a-1}$, then $\frac{a(m-2)}{m} > 1$ and hence

$$\sum_{\lfloor 2^{\frac{M}{a}}\rfloor + 1}^{\infty} i^{-\frac{a(m-2)}{m}} < \infty.$$

Therefore, $\exists\, C_0 \geq 1$,

$$s_n^2 \leq C_0 n 2^{M\left(\frac{1}{a} - \frac{m-2}{m}\right)}.$$

Also, by Lemma 2.1,

$$\begin{aligned}\alpha_Y^{-1}(u) &= \sum_{i=0}^{\infty} I[u \le \alpha_Y^{-1}(i)] \\ &\le \sum_{i=0}^{\infty} I[u \le c(i+1)^{-a}] \\ &= \sum_{i=1}^{\infty} I[c^{\frac{1}{a}} u^{-\frac{1}{a}} \ge i] \\ &\le c^{\frac{1}{a}} u^{-\frac{1}{a}}.\end{aligned}$$

Moreover,

$$\begin{aligned}Q_Y(u) &= \inf\{s : P(|Y_t| > s) \le u\} \\ &\le \inf\{s : s^{-m} E|Y_t|^m \le u\} \\ &= (E|Y_1|^m)^{\frac{1}{m}} u^{-\frac{1}{m}}.\end{aligned}$$

Then

$$R(u) = \alpha_Y^{-1}(u) Q(u) \le c^{\frac{1}{a}} (E|Y_1|^m)^{\frac{1}{m}} u^{-\frac{a+m}{am}} = b(u).$$

$R(u)$ is decreasing so that

$$H(u) = R^{-1}(u) \le b^{-1}(u) = c^{\frac{m}{a+m}} (E|Y_1|^m)^{\frac{a}{a+m}} u^{-\frac{am}{a+m}}.$$

Thus, the quantity $\lambda^{-1}\int_0^{H(\frac{\lambda}{r})} Q_Y(u)du$ in (6.5) of [55] satisfies that $\forall\ \lambda > 0,\ r \geq 1$,

$$\begin{aligned}\lambda^{-1}\int_0^{H(\frac{\lambda}{r})} Q_Y(u)du \leq &\lambda^{-1}\int_0^{c^{\frac{m}{a+m}}(E|Y_1|^m)^{\frac{a}{a+m}}\left(\frac{r}{\lambda}\right)^{\frac{am}{a+m}}} (E|Y_1|^m)^{\frac{1}{m}} u^{-\frac{1}{m}} du \\ =&\lambda^{-1}(E|Y_1|^m)^{\frac{1}{m}}\frac{m}{m-1}\left(c^{\frac{m}{a+m}}(E|Y_1|^m)^{\frac{a}{a+m}}\left(\frac{r}{\lambda}\right)^{\frac{am}{a+m}}\right)^{\frac{m-1}{m}} \\ =&(E|Y_1|^m)^{\frac{a+1}{a+m}}\frac{m}{m-1}c^{\frac{m-1}{a+m}}r^{-1}\left(\frac{r}{\lambda}\right)^{\frac{(a+1)m}{a+m}} \\ \leq&C_1 r^{-1}\left(\frac{r}{\lambda}\right)^{\frac{(a+1)m}{a+m}}\end{aligned}$$

for some $C_1 \geq 1$.

Then, by Theorem 6.2 in [55],

$$\begin{aligned}&P\left(\left|\sum_{S\in\mathcal{D}_{\boldsymbol{L}}^h}\varepsilon_S\mathbb{Z}_n(|g|I_S)\right| \geq 4\lambda\middle|\boldsymbol{\varepsilon}\right) \\ =&P\left(\frac{1}{\sqrt{n}}\left|\sum_{t=1}^n Y_t^\dagger\right| \geq 4\lambda\middle|\boldsymbol{\varepsilon}\right) \\ =&P\left(\left|\sum_{t=1}^n Y_t^\dagger\right| \geq 4\lambda\sqrt{n}\middle|\boldsymbol{\varepsilon}\right) \\ \leq&4\left(1+\frac{(\lambda\sqrt{n})^2}{rs_n^2}\right)^{-\frac{r}{2}} + 4nC_1r^{-1}\left(\frac{r}{\lambda\sqrt{n}}\right)^{\frac{(a+1)m}{a+m}} \\ \leq&4\left(\frac{\lambda^2 n}{rs_n^2}\right)^{-\frac{r}{2}} + 4nC_1r^{-1}\left(\frac{r}{\lambda\sqrt{n}}\right)^{\frac{(a+1)m}{a+m}} \\ \leq&4\lambda^{-r}r^{\frac{r}{2}}C_0^{\frac{r}{2}}2^{M\left(\frac{1}{a}-\frac{m-2}{m}\right)\frac{r}{2}} + 4C_1 r^{\frac{a(m-1)}{a+m}}\lambda^{-\frac{m(a+1)}{a+m}}n^{\frac{2a+m-am}{2(a+m)}}.\end{aligned}$$

Hence,

$$
\begin{aligned}
&P\left(\left|\sum_{S\in\mathcal{D}^h_{\boldsymbol{L}}}\varepsilon_S\mathbb{Z}_n(|g|I_S)\right|\geq 4\lambda\right)\\
=&\sum_{\boldsymbol{\varepsilon}}P\left(\left|\sum_{S\in\mathcal{D}^h_{\boldsymbol{L}}}\varepsilon_S\mathbb{Z}_n(|g|I_S)\right|\geq 4\lambda\middle|\boldsymbol{\varepsilon}\right)p(\boldsymbol{\varepsilon})\\
\leq&4\lambda^{-r}r^{\frac{r}{2}}C_0^{\frac{r}{2}}2^{M\left(\frac{1}{a}-\frac{m-2}{m}\right)\frac{r}{2}}+4C_1r^{\frac{a(m-1)}{a+m}}\lambda^{-\frac{m(a+1)}{a+m}}n^{\frac{2a+m-am}{2(a+m)}},
\end{aligned}
$$

and subsequently,

$$
\begin{aligned}
&P\left(\sup_{S\in\mathcal{D}_{\boldsymbol{L}}}|\mathbb{Z}_n(|g|I_S)|\geq 4\lambda\right)\\
\leq&2\cdot 2^M\left(4\lambda^{-r}r^{\frac{r}{2}}C_0^{\frac{r}{2}}2^{M\left(\frac{1}{a}-\frac{m-2}{m}\right)\frac{r}{2}}+4C_1r^{\frac{a(m-1)}{a+m}}\lambda^{-\frac{m(a+1)}{a+m}}n^{\frac{2a+m-am}{2(a+m)}}\right)\\
\leq&C_2\lambda^{-r}2^{M\left(\frac{1}{a}-\frac{m-2}{m}\right)\frac{r}{2}+M}+C_32^M\lambda^{-\frac{m(a+1)}{a+m}}n^{\frac{2a+m-am}{2(a+m)}}\\
\leq&\min\left(1,C_2\lambda^{-r}2^{M\left(\frac{1}{a}-\frac{m-2}{m}\right)\frac{r}{2}+M}+C_32^M\lambda^{-\frac{m(a+1)}{a+m}}n^{\frac{2a+m-am}{2(a+m)}}\right)\\
\leq&\min\left(1,C_2\lambda^{-r}2^{M\left(\frac{1}{a}-\frac{m-2}{m}\right)\frac{r}{2}+M}\right)+\min\left(1,C_32^M\lambda^{-\frac{m(a+1)}{a+m}}n^{\frac{2a+m-am}{2(a+m)}}\right)\\
\leq&C_2\min\left(1,\lambda^{-r}2^{M\left(\frac{1}{a}-\frac{m-2}{m}\right)\frac{r}{2}+M}\right)+C_3\min\left(1,2^M\lambda^{-\frac{m(a+1)}{a+m}}n^{\frac{2a+m-am}{2(a+m)}}\right),
\end{aligned}
$$

where $C_2 = 8r^{\frac{r}{2}}C_0^{\frac{r}{2}} \geq 1$ and $C_3 = 8C_1r^{\frac{a(m-1)}{a+m}} \geq 1$. Since this holds for all $r \geq 1$, take $r > \frac{4}{\frac{m-2}{m}-\frac{1}{a}}$, then $2^{M\left(\frac{1}{a}-\frac{m-2}{m}\right)\frac{r}{2}+M} \leq 2^{-M}$.Therefore,

$$
P\left(\sup_{S\in\mathcal{D}_{\boldsymbol{L}}}|\mathbb{Z}_n(|g|I_S)|\geq 4\lambda\right)\leq C_2\min\left(1,\lambda^{-r}2^{-M}\right)+C_3\min\left(1,2^M\lambda^{-\frac{m(a+1)}{a+m}}n^{\frac{2a+m-am}{2(a+m)}}\right).
$$

By Fubini's Theorem,

$$
\begin{aligned}
&E\left(\sup_{S\in\mathcal{D}_{\boldsymbol{L}}} |\mathbb{Z}_n(|g|I_S)|\right)\\
=&\int_0^\infty EI[\sup_{S\in\mathcal{D}_{\boldsymbol{L}}} |\mathbb{Z}_n(|g|I_S)| \geq t]dt\\
=&4\int_0^\infty P(\sup_{S\in\mathcal{D}_{\boldsymbol{L}}} |\mathbb{Z}_n(|g|I_S)| \geq 4\lambda)d\lambda\\
\leq&4\left(C_2\int_0^\infty \min(1,\lambda^{-r}2^{-M})d\lambda + C_3\int_0^\infty \min\left(1, 2^M\lambda^{-\frac{m(a+1)}{a+m}} n^{\frac{2a+m-am}{2(a+m)}}\right)d\lambda\right).
\end{aligned}
$$

Calculate the two integrals,

$$
\begin{aligned}
\int_0^\infty \min(1,\lambda^{-r}2^{-M})d\lambda =& \int_0^{2^{-\frac{M}{r}}} d\lambda + \int_{2^{-\frac{M}{r}}}^\infty \lambda^{-r}2^{-M}d\lambda\\
=&\frac{r}{r-1}2^{-\frac{M}{r}},
\end{aligned}
$$

and

$$
\begin{aligned}
&\int_0^\infty \min\left(1, 2^M\lambda^{-\frac{m(a+1)}{a+m}} n^{\frac{2a+m-am}{2(a+m)}}\right)d\lambda\\
=&\int_0^{2^{\frac{M(a+m)}{m(a+1)}} n^{\frac{2a+m-am}{2m(a+1)}}} d\lambda + \int_{2^{\frac{M(a+m)}{m(a+1)}} n^{\frac{2a+m-am}{2m(a+1)}}}^\infty 2^M\lambda^{-\frac{m(a+1)}{a+m}} n^{\frac{2a+m-am}{2(a+m)}} d\lambda\\
=&\frac{m(a+1)}{(m-1)a}2^{\frac{M(a+m)}{m(a+1)}} n^{\frac{2a+m-am}{2m(a+1)}}.
\end{aligned}
$$

Hence,

$$
E\left(\sup_{S\in\mathcal{D}_{\boldsymbol{L}}} |\mathbb{Z}_n(|g|I_S)|\right) \leq \frac{4r}{r-1}C_2 2^{-\frac{M}{r}} + \frac{4m(a+1)}{(m-1)a}C_3 2^{\frac{M(a+m)}{m(a+1)}} n^{\frac{2a+m-am}{2m(a+1)}}.
$$

Take $M = \lceil \frac{L_1}{r} \rceil$ so that $\frac{L_1}{r} \leq M < \frac{L_1}{r} + 1$, then $2^{-\frac{M}{r}} \leq 2^{-\frac{L_1}{r^2}}$ and $2^{\frac{M(a+m)}{m(a+1)}} \leq 2^{\left(\frac{L_1}{r}+1\right)\frac{a+m}{m(a+1)}}$.

Also, for sufficiently large n such that $n > 2^{L_1-1}$, $n^{\frac{2a+m-am}{2m(a+1)}} < 2^{(L_1-1)\frac{2a+m-am}{2m(a+1)}}$. Hence,

$$E\left(\sup_{S\in\mathcal{D}_{\boldsymbol{L}}} |\mathbb{Z}_n(|g|I_S)|\right) \leq \frac{4r}{r-1}C_2 2^{-\frac{L_1}{r^2}} + \frac{4m(a+1)}{(m-1)a}C_3 2^{\frac{1}{2}} 2^{L_1\left(\frac{1}{r}\frac{a+m}{m(a+1)}+\frac{2a+m-am}{2m(a+1)}\right)}.$$

Take

$$r > \max\left(\frac{4}{\frac{m-2}{m}-\frac{1}{a}}, \frac{-(a+m)-\sqrt{(a+m)^2-2m(a+1)(2a+m-am)}}{2a+m-am}\right),$$

then

$$(2a+m-am)r^2 + 2(a+m)r + 2m(a+1) < 0$$
$$\frac{1}{r}\frac{a+m}{m(a+1)} + \frac{2a+m-am}{2m(a+1)} < -\frac{1}{r^2}$$
$$2^{\frac{L_1}{r}\frac{a+m}{m(a+1)}+L_1\frac{2a+m-am}{2m(a+1)}} < 2^{-\frac{L_1}{r^2}}.$$

Let $C_4 = \frac{4r}{r-1}C_2 + \frac{4m(a+1)}{(m-1)a}C_3 2^{\frac{1}{2}}$, then

$$E\left(\sup_{S\in\mathcal{D}_{\boldsymbol{L}}} |\mathbb{Z}_n(|g|I_S)|\right) < C_4 2^{-\frac{L_1}{r^2}} = C_4 2^{-\frac{\max \boldsymbol{L}}{r^2}}.$$

Take N such that $2^{N-1} < n \leq 2^N$,

$$\sup_{\boldsymbol{x}\in[0,1]^d} |Z_n(\boldsymbol{x}) - Z_n(\boldsymbol{B}_K(\boldsymbol{x}))| \leq \sup_{\boldsymbol{x}\in[0,1]^d} |Z_n(\boldsymbol{x}) - Z_n(\boldsymbol{B}_N(\boldsymbol{x}))|$$
$$+ \sup_{\boldsymbol{x}\in[0,1]^d} |Z_n(\boldsymbol{B}_N(\boldsymbol{x})) - Z_n(\boldsymbol{B}_K(\boldsymbol{x}))|.$$

$$\sup_{\boldsymbol{x}\in[0,1]^d} |Z_n(\boldsymbol{x}) - Z_n(\boldsymbol{B}_N(\boldsymbol{x}))| \leq \frac{1}{\sqrt{n}} \sup_{\boldsymbol{x}\in[0,1]^d} \left| \sum_{t=1}^{n} g(\boldsymbol{X}_t)(I_{\prod_{j=1}^d (0,x_j]}(\boldsymbol{X}_t) - I_{\prod_{j=1}^d (0,B_{N,j}(x_j)]}(\boldsymbol{X}_t)) \right|$$
$$+ \sqrt{n} \sup_{\boldsymbol{x}\in[0,1]^d} \left| Eg(\boldsymbol{X}_1)(I_{\prod_{j=1}^d (0,x_j]}(\boldsymbol{X}_1) - I_{\prod_{j=1}^d (0,B_{N,j}(x_j)]}(\boldsymbol{X}_1)) \right|.$$

Note that $B_{N,j}(x_j) \leq x_j \leq B_{N,j}(x_j) + 2^{-N}$ for $j = 1, \ldots, d$, by definition.

$$\begin{aligned}
&\sqrt{n} \sup_{\boldsymbol{x}\in[0,1]^d} \left| Eg(\boldsymbol{X}_1)(I_{\prod_{j=1}^d (0,x_j]}(\boldsymbol{X}_1) - I_{\prod_{j=1}^d (0,B_{N,j}(x_j)]}(\boldsymbol{X}_1)) \right| \\
&\leq \sqrt{n} \sup_{\boldsymbol{x}\in[0,1]^d} \sum_{j=1}^{d} E|g(\boldsymbol{X}_1)| I[B_{N,j}(x_j) < X_{1,j} \leq x_j] \\
&\leq \sqrt{n} \sup_{\boldsymbol{x}\in[0,1]^d} \sum_{j=1}^{d} \left(E|g(\boldsymbol{X}_1)|^m\right)^{\frac{1}{m}} \left(P(B_{N,j}(x_j) < X_{1,j} \leq x_j)\right)^{\frac{m-1}{m}} \\
(X_{1,j} \sim U[0,1]) \quad &\leq \sqrt{n} d \left(E|g(\boldsymbol{X}_1)|^m\right)^{\frac{1}{m}} 2^{-\frac{N(m-1)}{m}} \\
&\leq \sqrt{n} d \left(E|g(\boldsymbol{X}_1)|^m\right)^{\frac{1}{m}} n^{-\frac{m-1}{m}} \\
&\leq d \left(E|g(\boldsymbol{X}_1)|^m\right)^{\frac{1}{m}} n^{\frac{2-m}{2m}} \\
(m > 2) \quad &\sim o(1).
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sup_{\boldsymbol{x}\in[0,1]^d} \left| \sum_{t=1}^{n} g(\boldsymbol{X}_t)(I_{\prod_{j=1}^d (0,x_j]}(\boldsymbol{X}_t) - I_{\prod_{j=1}^d (0,B_{N,j}(x_j)]}(\boldsymbol{X}_t)) \right| \\
&\leq \frac{1}{\sqrt{n}} \sup_{\boldsymbol{x}\in[0,1]^d} \sum_{t=1}^{n} |g(\boldsymbol{X}_t)| \left(\sum_{i=1}^{d} I_{S_i}(\boldsymbol{X}_t) \right) \\
&\leq \sum_{i=1}^{d} \sup_{x_i\in[0,1]} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |g(\boldsymbol{X}_t)| I_{S_i}(\boldsymbol{X}_t) \\
&\leq \sum_{i=1}^{d} \sup_{x_i\in[0,1]} \sqrt{n} \left| \frac{1}{n} \sum_{t=1}^{n} |g(\boldsymbol{X}_t)| I_{S_i}(\boldsymbol{X}_t) - E|g(\boldsymbol{X}_1)| I_{S_i}(\boldsymbol{X}_1) \right| + \sqrt{n} \sum_{i=1}^{d} \sup_{x_i\in[0,1]} E|g(\boldsymbol{X}_1)| I_{S_i}(\boldsymbol{X}_1),
\end{aligned}$$

where

$$S_i = \prod_{j\neq i, j=1}^{d} (0, B_{N,j}(x_j) + 2^{-N}] \times (B_{N,i}(x_i), B_{N,i}(x_i) + 2^{-N}].$$

S_i can be considered as a union of dyadic boxes where $L_i = N$ and $L_j \leq N$ for $j \neq i$. Note that the dyadic boxes have distinct $\boldsymbol{L}$, so that the number of all dyadic boxes that satisfying $L_i = N$ and $L_j \leq N$ for $j \neq i$ is $(N+1)^{d-1}$. Thus, $\max \boldsymbol{L} = N$ and

$$\begin{aligned}
&E^* \left(\sup_{x_i \in [0,1]} \sqrt{n} \left| \frac{1}{n} \sum_{t=1}^{n} |g(\boldsymbol{X}_t)| I_{S_i}(\boldsymbol{X}_t) - E|g(\boldsymbol{X}_1)| I_{S_i}(\boldsymbol{X}_1) \right| \right) \\
\leq &(N+1)^{d-1} E \left(\sup_{S \in \mathcal{D}_{\boldsymbol{L}}} |\mathbb{Z}_n(|g| I_S)| \right) \\
< &C_4 (N+1)^{d-1} 2^{-\frac{N}{r^2}} \\
\sim &o(1).
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\sqrt{n} \sum_{i=1}^{d} \sup_{x_i \in [0,1]} E|g(\boldsymbol{X}_1)| I_{S_i}(\boldsymbol{X}_1) \\
= &\sqrt{n} \sum_{i=1}^{d} \sup_{x_i \in [0,1]} E|g(\boldsymbol{X}_1)| I[B_{N,i}(x_i) < X_{1,i} \leq B_{N,i}(x_i) + 2^{-N}] \\
\leq &\sqrt{n} d \left(E|g(\boldsymbol{X}_1)|^m \right)^{\frac{1}{m}} 2^{-\frac{N(m-1)}{m}} \\
\sim &o(1).
\end{aligned}$$

Thus,

$$E^* \left(\sup_{\boldsymbol{x} \in [0,1]^d} |Z_n(\boldsymbol{x}) - Z_n(\boldsymbol{B}_N(\boldsymbol{x}))| \right) \sim o(1).$$

Then, following [55] (p.124),

$$E^*\left(\sup_{\boldsymbol{x}\in[0,1]^d}|Z_n(B_N(\boldsymbol{x}))-Z_n(B_K(\boldsymbol{x}))|\right)\leq\sum_{\boldsymbol{L}:\max\boldsymbol{L}=K+1,\ldots,N}E\left(\sup_{S\in\mathcal{D}_{\boldsymbol{L}}}|\mathbb{Z}_n(fI_S)|\right)$$
$$\leq C_4\sum_{J=K+1}^{N}[(J+1)^d-J^d]2^{-\frac{J}{r2}}.$$

Note that $\sum_{J=1}^{\infty}[(J+1)^d-J^d]2^{-\frac{J}{r2}}<\infty$ and hence $\sum_{J=K+1}^{\infty}[(J+1)^d-J^d]2^{-\frac{J}{r2}}\to 0$ as $K\to\infty$. Thus,

$$\lim_{K\to\infty}\limsup_{n\to\infty}E^*\left(\sup_{\boldsymbol{x}\in[0,1]^d}|Z_n(B_N(\boldsymbol{x}))-Z_n(B_K(\boldsymbol{x}))|\right)=0.$$

Subsequently,

$$\lim_{K\to\infty}\limsup_{n\to\infty}E^*\left(\sup_{\boldsymbol{x}\in[0,1]^d}|Z_n(\boldsymbol{x})-Z_n(B_K(\boldsymbol{x}))|\right)=0,$$

which completes the proof. □

2.2.3.1 An Approximation of $\tilde{G}$

Following Section 1.2.3.1, let

$$
\begin{aligned}
A_k(\boldsymbol{x};\boldsymbol{\theta}) &= \int \frac{w_k(\boldsymbol{y};\boldsymbol{\theta}) I_{\prod_{j=1}^d(-\infty,x_d]}(\boldsymbol{y})}{\sum_{j=0}^m \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta})} dG(\boldsymbol{y}),\\
\boldsymbol{B}_k(\boldsymbol{x};\boldsymbol{\theta}) &= \int \frac{w_k(\boldsymbol{y};\boldsymbol{\theta}) \boldsymbol{h}_k(\boldsymbol{y}) I_{\prod_{j=1}^d(-\infty,x_d]}(\boldsymbol{y})}{\sum_{j=0}^m \rho_j w_j(\boldsymbol{y};\boldsymbol{\theta})} dG(\boldsymbol{y}),\\
\overline{\boldsymbol{A}}(\boldsymbol{x};\boldsymbol{\theta}) &= (A_1(\boldsymbol{x};\boldsymbol{\theta}), \cdots, A_m(\boldsymbol{x};\boldsymbol{\theta}))^\top,\\
\overline{\boldsymbol{B}}(\boldsymbol{x};\boldsymbol{\theta}) &= (\boldsymbol{B}_1^\top(\boldsymbol{x};\boldsymbol{\theta}), \cdots, \boldsymbol{B}_m^\top(\boldsymbol{x};\boldsymbol{\theta}))^\top,\\
H_1(\boldsymbol{x};\boldsymbol{\theta}) &= \frac{1}{n_0}\sum_{i=1}^n \frac{I_{\prod_{j=1}^d(-\infty,x_d]}(\boldsymbol{t}_i)}{\sum_{j=0}^m \rho_j w_j(\boldsymbol{t}_i;\boldsymbol{\theta})},\\
H_2(\boldsymbol{x};\boldsymbol{\theta}) &= \frac{1}{n}(\overline{\boldsymbol{A}}^\top(\boldsymbol{x};\boldsymbol{\theta})\boldsymbol{\rho}, \overline{\boldsymbol{B}}^\top(\boldsymbol{x};\boldsymbol{\theta})\overline{\boldsymbol{\rho}})\boldsymbol{S}^{-1}(\boldsymbol{\theta})\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.
\end{aligned}
$$

Replace Assumption 1.1 by Assumption 2.1 and 2.2, and Theorem 1.1 by Theorem 2.3. The approximation

$$
\widetilde{G}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}}) = H_1(\boldsymbol{x};\boldsymbol{\theta}_0) - H_2(\boldsymbol{x};\boldsymbol{\theta}_0) + o_p\left(\frac{1}{\sqrt{n}}\right) \tag{2.7}
$$

is obtained once again using the same derivation in Section 1.2.3.1.

2.2.3.2 Weak Convergence of $\sqrt{n}(\widetilde{G} - \widehat{G})$

Based on (2.7), it is sufficient to show the weak convergence of $\sqrt{n}(H_1 - H_2 - \widehat{G})$. We first show the convergence of the finite dimension distribution of $\sqrt{n}(H_1 - H_2 - \widehat{G})$. $\forall \boldsymbol{x} \in \mathbb{R}^d$, denote $\gamma_{\boldsymbol{x}}(\cdot;\boldsymbol{\theta}) = \frac{I_{\prod_{j=1}^{d}(-\infty,x_j]}(\cdot)}{\sum_{j=0}^{m}\rho_j w_j(\cdot;\boldsymbol{\theta})}$ and $\gamma^*_{\boldsymbol{x}}(\cdot;\boldsymbol{\theta}) = \frac{\left[\sum_{j=1}^{m}\rho_j w_j(\cdot;\boldsymbol{\theta})\right] I_{\prod_{j=1}^{d}(-\infty,x_j]}(\cdot)}{\sum_{j=0}^{m}\rho_j w_j(\cdot;\boldsymbol{\theta})}$. Then,

$$
\begin{aligned}
&H_1(\boldsymbol{x};\boldsymbol{\theta}_0) - \widehat{G}(\boldsymbol{x}) \\
=&\frac{1}{n_0}\sum_{h=1}^{m}\sum_{l=1}^{n_h}\frac{I_{\prod_{j=1}^{d}(-\infty,x_j]}(\boldsymbol{X}_l^{(h)})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)} - \frac{1}{n_0}\sum_{l=1}^{n_0}\frac{\left[\sum_{j=1}^{m}\rho_j w_j(\boldsymbol{X}_l^{(0)};\boldsymbol{\theta}_0)\right] I_{\prod_{j=1}^{d}(-\infty,x_j]}(\boldsymbol{X}_l^{(0)})}{\sum_{j=0}^{m}\rho_j w_j(\boldsymbol{X}_l^{(0)};\boldsymbol{\theta}_0)} \\
=&\sum_{h=1}^{m}\frac{\rho_h}{n_h}\sum_{l=1}^{n_h}\gamma_{\boldsymbol{x}}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0) - \frac{1}{n_0}\sum_{l=1}^{n_0}\gamma^*_{\boldsymbol{x}}(\boldsymbol{X}_l^{(0)};\boldsymbol{\theta}).
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}(\sqrt{n}(H_1(\boldsymbol{x};\boldsymbol{\theta}_0) - \widehat{G}(\boldsymbol{x})), \sqrt{n}(H_1(\boldsymbol{y};\boldsymbol{\theta}_0) - \widehat{G}(\boldsymbol{y}))) \\
=&\left(\sum_{j=0}^{m}\rho_j\right)\sum_{h=1}^{m}\rho_h\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\sum_{l'=1}^{n_h}\mathrm{Cov}\left(\gamma_{\boldsymbol{x}}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0), \gamma_{\boldsymbol{y}}(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta}_0)\right)\right) \\
&+\left(\sum_{j=0}^{m}\rho_j\right)\frac{1}{n_0}\sum_{l=1}^{n_h}\sum_{l'=1}^{n_h}\mathrm{Cov}\left(\gamma^*_{\boldsymbol{x}}(\boldsymbol{X}_l^{(0)};\boldsymbol{\theta}_0), \gamma^*_{\boldsymbol{y}}(\boldsymbol{X}_{l'}^{(0)};\boldsymbol{\theta}_0)\right) \\
\to&2\pi\left(\sum_{j=0}^{m}\rho_j\right)\left(\sum_{h=1}^{m}\rho_h f_{\gamma_{\boldsymbol{x}},\gamma_{\boldsymbol{y}},h}(0;\boldsymbol{\theta}_0) + f_{\gamma^*_{\boldsymbol{x}},\gamma^*_{\boldsymbol{y}},0}(0;\boldsymbol{\theta}_0)\right),
\end{aligned}
$$

as $n \to \infty$ where $f_{\gamma^*_{\boldsymbol{x}},\gamma^*_{\boldsymbol{y}},0}$ is the cross spectral density of $\{\gamma^*_{\boldsymbol{x}}(\boldsymbol{X}_l^{(0)};\boldsymbol{\theta}_0)\}$ and $\{\gamma^*_{\boldsymbol{y}}(\boldsymbol{X}_{l'}^{(0)};\boldsymbol{\theta}_0)\}$, and $f_{\gamma_{\boldsymbol{x}},\gamma_{\boldsymbol{y}},h}$ is the cross spectral density of $\{\gamma_{\boldsymbol{x}}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\}$ and $\{\gamma_{\boldsymbol{y}}(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta}_0)\}$ for $h = 1, \ldots, m$.

Denote $\boldsymbol{K}(\boldsymbol{x};\boldsymbol{\theta}) = (\overline{\boldsymbol{A}}^{\top}(\boldsymbol{x};\boldsymbol{\theta})\boldsymbol{\rho}, \overline{\boldsymbol{B}}^{\top}(\boldsymbol{x};\boldsymbol{\theta})\overline{\boldsymbol{\rho}})\boldsymbol{S}^{-1}(\boldsymbol{\theta})$. Then,

$$
H_2(\boldsymbol{x};\boldsymbol{\theta}) = \frac{1}{n}\boldsymbol{K}(\boldsymbol{x};\boldsymbol{\theta})\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},
$$

so that

$$
\begin{aligned}
&\mathrm{Cov}(\sqrt{n}H_2(\boldsymbol{x};\boldsymbol{\theta}_0),\sqrt{n}H_2(\boldsymbol{y};\boldsymbol{\theta}_0))\\
=&\boldsymbol{K}(\boldsymbol{x};\boldsymbol{\theta}_0)\mathrm{Var}\left(\frac{1}{\sqrt{n}}\left.\frac{\partial l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\boldsymbol{K}^\top(\boldsymbol{y};\boldsymbol{\theta}_0)\\
\to&\boldsymbol{K}(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\mathcal{V}}(\boldsymbol{\theta}_0)\boldsymbol{K}^\top(\boldsymbol{y};\boldsymbol{\theta}_0),
\end{aligned}
$$

as $n\to\infty$. Moreover,

$$
\begin{aligned}
&\mathrm{Cov}\left(\sqrt{n}(H_1(\boldsymbol{x};\boldsymbol{\theta}_0)-\widehat{G}(\boldsymbol{x})),\sqrt{n}H_2(\boldsymbol{y};\boldsymbol{\theta}_0)\right)\\
=&\boldsymbol{K}(\boldsymbol{y};\boldsymbol{\theta}_0)\mathrm{Cov}\left(H_1(\boldsymbol{x};\boldsymbol{\theta}_0)-\widehat{G}(\boldsymbol{x}),\left.\frac{\partial l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right).
\end{aligned}
$$

Additionally,

$$
\begin{aligned}
&\mathrm{Cov}\left(H_1(\boldsymbol{x};\boldsymbol{\theta}_0)-\widehat{G}(\boldsymbol{x}),\left.\frac{\partial l(\boldsymbol{\theta})}{\partial\alpha_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\\
=&\sum_{h=1}^{m}\rho_h\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\sum_{l'=1}^{n_h}\mathrm{Cov}\left(\gamma_{\boldsymbol{x}}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0),-\varphi_k(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta}_0)\right)\right)\\
&-\frac{1}{n_0}\sum_{l=1}^{n_0}\sum_{l'=1}^{n_0}\mathrm{Cov}\left(\gamma_{\boldsymbol{x}}^*(\boldsymbol{X}_l^{(0)};\boldsymbol{\theta}_0),-\varphi_k(\boldsymbol{X}_{l'}^{(0)};\boldsymbol{\theta}_0)\right)\\
\to&2\pi\left(\sum_{h=1}^{m}\rho_h f_{\gamma_{\boldsymbol{x}},\alpha_k,h}(0;\boldsymbol{\theta}_0)-f_{\gamma_{\boldsymbol{x}}^*,\alpha_k,0}(0;\boldsymbol{\theta}_0)\right),
\end{aligned}
$$

as $n\to\infty$ for $k=1,\dots,m$. $f_{\gamma_{\boldsymbol{x}}^*,\alpha_k,0}$ is the cross spectral density of $\{\gamma_{\boldsymbol{x}}^*(\boldsymbol{X}_l^{(0)};\boldsymbol{\theta}_0)\}$ and $\{-\varphi_k(\boldsymbol{X}_{l'}^{(0)};\boldsymbol{\theta}_0)\}$, and $f_{\gamma_{\boldsymbol{x}},\alpha_k,h}$ is the cross spectral density of $\{\gamma_{\boldsymbol{x}}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\}$ and $\{-\varphi_k(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta}_0)\}$ for $h=1,\dots,m$.

$$
\begin{aligned}
&\mathrm{Cov}\left(H_1(\boldsymbol{x};\boldsymbol{\theta}_0)-\widehat{G}(\boldsymbol{x}),\frac{\partial l(\boldsymbol{\theta}_0)}{\partial \beta_{kp}}\right)\\
=&\rho_k\left(\frac{1}{n_k}\sum_{l=1}^{n_k}\sum_{l'=1}^{n_k}\mathrm{Cov}\left(\gamma_{\boldsymbol{x}}(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0),(1-\varphi_k(\boldsymbol{X}_{l'}^{(k)};\boldsymbol{\theta}_0))h_{kp}(\boldsymbol{X}_{l'}^{(k)})\right)\right)\\
&+\sum_{h\neq k,h=1}^{m}\rho_h\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\sum_{l'=1}^{n_h}\mathrm{Cov}\left(\gamma_{\boldsymbol{x}}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0),-\varphi_k(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta}_0)h_{kp}(\boldsymbol{X}_{l'}^{(h)})\right)\right)\\
&-\frac{1}{n_0}\sum_{l=1}^{n_0}\sum_{l'}^{n_0}\mathrm{Cov}\left(\gamma_{\boldsymbol{x}}^*(\boldsymbol{X}_l^{(0)};\boldsymbol{\theta}_0),-\varphi_k(\boldsymbol{X}_{l'}^{(0)};\boldsymbol{\theta}_0)h_{kp}(\boldsymbol{X}_{l'}^{(0)})\right)\\
\to&2\pi\left(\sum_{h=1}^{m}\rho_h f_{\gamma_{\boldsymbol{x}},\beta_{kp},h}(0;\boldsymbol{\theta}_0)-f_{\gamma_{\boldsymbol{x}}^*,\beta_{kp},0}(0;\boldsymbol{\theta}_0)\right),
\end{aligned}
$$

as $n\to\infty$ for $k=1,\dots,m, p=1,\dots,p_k$. $f_{\gamma_{\boldsymbol{x}}^*,\beta_{kp},0}$ is the cross spectral density of $\{\gamma_{\boldsymbol{x}}^*(\boldsymbol{X}_l^{(0)};\boldsymbol{\theta}_0)\}$ and $\{-\varphi_k(\boldsymbol{X}_{l'}^{(0)};\boldsymbol{\theta}_0)h_{kp}(\boldsymbol{X}_{l'}^{(0)})\}$, $f_{\gamma_{\boldsymbol{x}},\beta_{kp},k}$ is the cross spectral density of $\{\gamma_{\boldsymbol{x}}(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0)\}$ and $\{(1-\varphi_k(\boldsymbol{X}_{l'}^{(k)};\boldsymbol{\theta}_0))h_{kp}(\boldsymbol{X}_{l'}^{(k)})\}$ and $f_{\gamma_{\boldsymbol{x}},\beta_{kp},h}$ is the cross spectral density of $\{\gamma_{\boldsymbol{x}}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0)\}$ and $\{-\varphi_k(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta}_0)h_{kp}(\boldsymbol{X}_{l'}^{(h)})\}$ for $h\neq k$ and $h=1,\dots,m$.

Let $\boldsymbol{\mathcal{W}}^*(\boldsymbol{x};\boldsymbol{\theta}_0)$ be the limit of $\mathrm{Cov}\left(H_1(\boldsymbol{x};\boldsymbol{\theta}_0)-\widehat{G}(\boldsymbol{x}),\left.\frac{\partial l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)$, then

$$
\begin{aligned}
&\mathrm{Cov}\left(\sqrt{n}(H_1(\boldsymbol{x};\boldsymbol{\theta}_0)-H_2(\boldsymbol{x};\boldsymbol{\theta}_0)-\widehat{G}(\boldsymbol{x})),\sqrt{n}(H_1(\boldsymbol{y};\boldsymbol{\theta}_0)-H_2(\boldsymbol{y};\boldsymbol{\theta}_0)-\widehat{G}(\boldsymbol{y}))\right)\\
\to&2\pi\left(\sum_{j=0}^{m}\rho_j\right)\left(\sum_{h=1}^{m}\rho_h f_{\gamma_{\boldsymbol{x}},\gamma_{\boldsymbol{y}},h}(0;\boldsymbol{\theta}_0)+f_{\gamma_{\boldsymbol{x}}^*,\gamma_{\boldsymbol{y}}^*,0}(0;\boldsymbol{\theta}_0)\right)-\boldsymbol{K}(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{\mathcal{W}}^*(\boldsymbol{x};\boldsymbol{\theta}_0)\\
&-\boldsymbol{K}(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\mathcal{W}}^*(\boldsymbol{y};\boldsymbol{\theta}_0)+\boldsymbol{K}(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\mathcal{V}}(\boldsymbol{\theta}_0)\boldsymbol{K}^\top(\boldsymbol{y};\boldsymbol{\theta}_0),
\end{aligned}
\tag{2.8}
$$

as $n\to\infty$.

$\forall\ k$ and $\boldsymbol{x}_1,\dots,\boldsymbol{x}_k$, by Theorem 2.2,

$$
\sqrt{n}(H_1(\boldsymbol{x}_1;\boldsymbol{\theta_0})-H_2(\boldsymbol{x}_1;\boldsymbol{\theta_0})-\widehat{G}(\boldsymbol{x}_1),\dots,H_1(\boldsymbol{x}_k;\boldsymbol{\theta_0})-H_2(\boldsymbol{x}_k;\boldsymbol{\theta_0})-\widehat{G}(\boldsymbol{x}_k))^\top\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{\Delta}),
$$

where $\mathbf{\Delta}$ is the covariance matrix depending on the limit of the covariance function in (2.8). Moreover, following Section 1.2.3.3, by Theorem 2.3, the weak convergence of $\sqrt{n}(H_1 - \widehat{G})$ is established. By Theorem 1.2, the tightness of $\sqrt{n}H_2$ is proved, following Section 1.2.3.4. Thus, Theorem 1.5.4 in [58], $\sqrt{n}(\widetilde{G} - \widehat{G})$ converges weakly to a zero-mean Gaussian process with the covariance function as the limit in (2.8).

2.2.3.3 Weak Convergence of $\sqrt{n}(\widetilde{G} - G)$

The weak convergence $\sqrt{n}(\widetilde{G} - G)$ is an immediate result of the weak convergence of $\sqrt{n}(\widetilde{G} - \widehat{G})$ and $\sqrt{n}(\widehat{G} - G)$. It remains to show the limiting covariance function.

$$
\begin{aligned}
&\operatorname{Cov}(\sqrt{n}H_1(\boldsymbol{x};\boldsymbol{\theta}_0), \sqrt{n}H_1(\boldsymbol{y};\boldsymbol{\theta}_0)) \\
=&\left(\sum_{j=0}^{m}\rho_j\right)\sum_{h=0}^{m}\rho_h\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\sum_{l'=1}^{n_h}\operatorname{Cov}\left(\gamma_{\boldsymbol{x}}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0), \gamma_{\boldsymbol{y}}(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta}_0)\right)\right) \\
\to&2\pi\left(\sum_{j=0}^{m}\rho_j\right)\sum_{h=0}^{m}\rho_h f_{\gamma_{\boldsymbol{x}},\gamma_{\boldsymbol{y}},h}(0;\boldsymbol{\theta}_0),
\end{aligned}
$$

as $n \to \infty$.

$$
\begin{aligned}
&\operatorname{Cov}\left(H_1(\boldsymbol{x};\boldsymbol{\theta}_0), \left.\frac{\partial l(\boldsymbol{\theta})}{\partial \alpha_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) \\
=&\sum_{h=0}^{m}\rho_h\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\sum_{l'=1}^{n_h}\operatorname{Cov}\left(\gamma_{\boldsymbol{x}}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0), -\varphi_k(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta}_0)\right)\right) \\
\to&2\pi\sum_{h=0}^{m}\rho_h f_{\gamma_{\boldsymbol{x}},\alpha_k,h}(0;\boldsymbol{\theta}_0),
\end{aligned}
$$

as $n \to \infty$ for $k = 1, \ldots, m$.

$$
\begin{aligned}
&\mathrm{Cov}\left(H_1(\boldsymbol{x};\boldsymbol{\theta}_0), \left.\frac{\partial l(\boldsymbol{\theta})}{\partial \beta_{kp}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)\\
=&\rho_k\left(\frac{1}{n_k}\sum_{l=1}^{n_k}\sum_{l'=1}^{n_k}\mathrm{Cov}\left(\gamma_{\boldsymbol{x}}(\boldsymbol{X}_l^{(k)};\boldsymbol{\theta}_0), (1-\varphi_k(\boldsymbol{X}_{l'}^{(k)};\boldsymbol{\theta}_0))h_{kp}(\boldsymbol{X}_{l'}^{(k)})\right)\right)\\
&+\sum_{h\neq k, h=0}^{m}\rho_h\left(\frac{1}{n_h}\sum_{l=1}^{n_h}\sum_{l'=1}^{n_h}\mathrm{Cov}\left(\gamma_{\boldsymbol{x}}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}_0), -\varphi_k(\boldsymbol{X}_{l'}^{(h)};\boldsymbol{\theta}_0)h_{kp}(\boldsymbol{X}_{l'}^{(h)})\right)\right)\\
\to&2\pi\sum_{h=0}^{m}\rho_h f_{\gamma_{\boldsymbol{x}},\beta_{kp},h}(0;\boldsymbol{\theta}_0),
\end{aligned}
$$

as $n\to\infty$ for $k=1,\ldots,m$, $p=1,\ldots,p_k$.

Let $\boldsymbol{\mathcal{W}}(\boldsymbol{x};\boldsymbol{\theta}_0)$ be the limit of $\mathrm{Cov}\left(H_1(\boldsymbol{x};\boldsymbol{\theta}_0), \left.\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)$, then

$$
\begin{aligned}
&\mathrm{Cov}\left(\sqrt{n}(H_1(\boldsymbol{x};\boldsymbol{\theta}_0)-H_2(\boldsymbol{x};\boldsymbol{\theta}_0)), \sqrt{n}(H_1(\boldsymbol{y};\boldsymbol{\theta}_0)-H_2(\boldsymbol{y};\boldsymbol{\theta}_0))\right)\\
\to&2\pi\left(\sum_{j=0}^{m}\rho_j\right)\left(\sum_{h=0}^{m}\rho_h f_{\gamma_{\boldsymbol{x}},\gamma_{\boldsymbol{y}},h}(0;\boldsymbol{\theta}_0)\right)-\boldsymbol{K}(\boldsymbol{y};\boldsymbol{\theta}_0)\boldsymbol{\mathcal{W}}(\boldsymbol{x};\boldsymbol{\theta}_0)-\boldsymbol{K}(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\mathcal{W}}(\boldsymbol{y};\boldsymbol{\theta}_0) \quad (2.9)\\
&+\boldsymbol{K}(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\mathcal{V}}(\boldsymbol{\theta}_0)\boldsymbol{K}^\top(\boldsymbol{y};\boldsymbol{\theta}_0),
\end{aligned}
$$

as $n\to\infty$. Thus, $\sqrt{n}(\widetilde{G}-G)$ converges weakly to a zero-mean Gaussian process with covariance function being the limit in (2.9). For a given $\boldsymbol{x}$,

$$
\sqrt{n}(\widetilde{G}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})-G(\boldsymbol{x}))\xrightarrow{d} N(0,\nu(\boldsymbol{x};\boldsymbol{\theta}_0)),
$$

as $n \to \infty$, where

$$\begin{aligned}\nu(\boldsymbol{x};\boldsymbol{\theta}_0) =& 2\pi \left(\sum_{j=0}^{m} \rho_j\right) \left(\sum_{h=0}^{m} \rho_h f_{\gamma_{\boldsymbol{x}},\gamma_{\boldsymbol{x}},h}(0;\boldsymbol{\theta}_0)\right) - 2\boldsymbol{K}(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\mathcal{W}}(\boldsymbol{x};\boldsymbol{\theta}_0) \\ &+ \boldsymbol{K}(\boldsymbol{x};\boldsymbol{\theta}_0)\boldsymbol{\mathcal{V}}(\boldsymbol{\theta}_0)\boldsymbol{K}^\top(\boldsymbol{x};\boldsymbol{\theta}_0).\end{aligned} \tag{2.10}$$

2.2.4 Consistent Estimators of Asymptotic Variances

In this section, we establish the consistency of the asymptotic variances of $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{G}(\boldsymbol{x}, \widetilde{\boldsymbol{\theta}})$. Conclusions made in Section 1.2.4 can be applied directly for $\widetilde{\boldsymbol{S}}(\widetilde{\boldsymbol{\theta}})$, $\widetilde{A}_k(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$ and $\widetilde{\boldsymbol{B}}_k(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$. Thus, our main goal is to show the consistency of the plug-in estimators of the cross spectral densities. The following theorem is introduced to help establishing consistency of these estimators.

Theorem 2.4. *Let $\{\boldsymbol{X}_t\}$ be a d-dimensional strictly stationary and strong mixing process with mixing coefficient $\alpha_{\boldsymbol{X}}(n) \leq c(n+1)^{-a}$ for some $c \geq 1$ and $a > 1$. $\forall\ \boldsymbol{\theta}$, $g_1(\cdot;\boldsymbol{\theta})$ and $g_2(\cdot;\boldsymbol{\theta})$ satisfy that $\exists\ g$ and D such that*

$$|g_1(\boldsymbol{X}_t;\boldsymbol{\theta})| \leq g(\boldsymbol{X}_t), \ \left\| \frac{\partial g_1(\boldsymbol{X}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \leq D(\boldsymbol{X}_t),$$
$$|g_2(\boldsymbol{X}_t;\boldsymbol{\theta})| \leq g(\boldsymbol{X}_t), \ \left\| \frac{\partial g_2(\boldsymbol{X}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \leq D(\boldsymbol{X}_t),$$

almost surely for all t, and

$$E[g(\boldsymbol{X}_1)]^{2m} < \infty, \quad E[D(\boldsymbol{X}_1)]^m < \infty,$$

for $m > \frac{2a}{a-1}$.

The cross spectral density of $g_1(\boldsymbol{X}_t;\boldsymbol{\theta})$ and $g_2(\boldsymbol{X}_t;\boldsymbol{\theta})$ evaluated at 0 is given by

$$f(0;\boldsymbol{\theta}) = \frac{1}{2\pi}\sum_{u=-\infty}^{\infty} R(u;\boldsymbol{\theta}),$$

where $R(u;\boldsymbol{\theta}) = \mathrm{Cov}\left(g_1(\boldsymbol{X}_t;\boldsymbol{\theta}), g_2(\boldsymbol{X}_{t+u};\boldsymbol{\theta})\right)$. The corresponding spectral density estimator is given by

$$\widehat{f}_n(0;\boldsymbol{\theta}) = \frac{1}{2\pi}\sum_{|u|<n} w\left(\frac{u}{M_n}\right)\widehat{R}(u;\boldsymbol{\theta}),$$

where

$$\widehat{R}(u;\boldsymbol{\theta}) = \begin{cases} \frac{1}{n-u}\sum_{t=1}^{n-u} g_1(\boldsymbol{X}_t;\boldsymbol{\theta})g_2(\boldsymbol{X}_{t+u};\boldsymbol{\theta}) - \left(\frac{1}{n}\sum_{t=1}^{n} g_1(\boldsymbol{X}_t;\boldsymbol{\theta})\right)\left(\frac{1}{n}\sum_{t=1}^{n} g_2(\boldsymbol{X}_t;\boldsymbol{\theta})\right),\ u \geq 0, \\ \frac{1}{n-|u|}\sum_{t=|u|+1}^{n} g_1(\boldsymbol{X}_t;\boldsymbol{\theta})g_2(\boldsymbol{X}_{t+u};\boldsymbol{\theta}) - \left(\frac{1}{n}\sum_{t=1}^{n} g_1(\boldsymbol{X}_t;\boldsymbol{\theta})\right)\left(\frac{1}{n}\sum_{t=1}^{n} g_2(\boldsymbol{X}_t;\boldsymbol{\theta})\right),\ u < 0, \end{cases}$$

$M_n \sim o(n^b)$ for some $0 < b \leq \frac{1}{3}$ and the rectangular window function

$$w(x) = \begin{cases} 1, & |x| \leq 1, \\ 0, & |x| > 1. \end{cases}$$

Suppose $\widehat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}$ such that

$$\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| = o_p\left(M_n^{-1}\right),$$

then

$$\widehat{f}_n(0;\widehat{\boldsymbol{\theta}}_n) \xrightarrow{p} f(0;\boldsymbol{\theta}),$$

as $n \to \infty$.

Proof. Decompose $\widehat{f}_n(0;\widehat{\boldsymbol{\theta}}_n) - f(0;\boldsymbol{\theta})$ into two parts,

$$\widehat{f}_n(0;\widehat{\boldsymbol{\theta}}_n) - f(0;\boldsymbol{\theta}) = \left(\widehat{f}_n(0;\widehat{\boldsymbol{\theta}}_n) - \widehat{f}_n(0;\boldsymbol{\theta})\right) + \left(\widehat{f}_n(0;\boldsymbol{\theta}) - f(0;\boldsymbol{\theta})\right).$$

We first show that $\widehat{f}_n(0;\boldsymbol{\theta}) - f(0;\boldsymbol{\theta}) \xrightarrow{p} 0$ as $n \to \infty$. Based on the expression of window function $w(x)$,

$$\widehat{f}_n(0;\boldsymbol{\theta}) - f(0;\boldsymbol{\theta}) = \frac{1}{2\pi}\sum_{|u|\le M_n}\left(\widehat{R}(u;\boldsymbol{\theta}) - R(u;\boldsymbol{\theta})\right) + \frac{1}{2\pi}\sum_{|u|>M_n} R(u;\boldsymbol{\theta}).$$

By (2.2) in [10],

$$\begin{aligned} |\mathrm{Cov}(g_1(X_t;\boldsymbol{\theta}), g_2(X_{t+u};\boldsymbol{\theta}))| \le & 12 \cdot 2^{-\frac{m-2}{m}} (E|g(\boldsymbol{X}_1)|^m)^{\frac{2}{m}} \alpha_{\boldsymbol{X}}^{\frac{m-2}{m}}(|u|) \\ \le & 12 \cdot 2^{-\frac{m-2}{m}} (E|g(\boldsymbol{X}_1)|^m)^{\frac{2}{m}} c^{\frac{m-2}{m}} (|u|+1)^{-\frac{a(m-2)}{m}}. \end{aligned}$$

Since $m > \frac{2a}{a-1}$, then $\sum_{u=-\infty}^{\infty}(|u|+1)^{-\frac{a(m-2)}{m}}$ is convergent. Thus,

$$\sum_{u=-\infty}^{\infty} |\mathrm{Cov}(g_1(\boldsymbol{X}_t;\boldsymbol{\theta}), g_2(\boldsymbol{X}_{t+u};\boldsymbol{\theta}))| < \infty,$$

and

$$\frac{1}{2\pi}\sum_{|u|>M_n} R(u;\boldsymbol{\theta}) \to 0, \tag{2.11}$$

as $n \to \infty$.

$$
\begin{aligned}
&\sum_{|u|\le M_n} \left(\widehat{R}(u;\boldsymbol{\theta}) - R(u;\boldsymbol{\theta})\right) \\
=&\sum_{u=0}^{\lfloor M_n \rfloor} \left(\frac{1}{n-u} \sum_{t=1}^{n-u} g_1(\boldsymbol{X}_t;\boldsymbol{\theta}) g_2(\boldsymbol{X}_{t+u};\boldsymbol{\theta}) - Eg_1(\boldsymbol{X}_1;\boldsymbol{\theta}) g_2(\boldsymbol{X}_{1+u};\boldsymbol{\theta}) \right) + \\
&\sum_{u=-\lfloor M_n \rfloor}^{-1} \left(\frac{1}{n-|u|} \sum_{t=|u|+1}^{n} g_1(\boldsymbol{X}_t;\boldsymbol{\theta}) g_2(\boldsymbol{X}_{t+u};\boldsymbol{\theta}) - Eg_1(\boldsymbol{X}_{1+|u|};\boldsymbol{\theta}) g_2(\boldsymbol{X}_1;\boldsymbol{\theta}) \right) - \\
&\sum_{|u|\le M_n} \left[\left(\frac{1}{n} \sum_{t=1}^{n} g_1(\boldsymbol{X}_t;\boldsymbol{\theta}) \right) \left(\frac{1}{n} \sum_{t=1}^{n} g_2(\boldsymbol{X}_t;\boldsymbol{\theta}) \right) - Eg_1(\boldsymbol{X}_1;\boldsymbol{\theta}) Eg_2(\boldsymbol{X}_1;\boldsymbol{\theta}) \right].
\end{aligned}
$$

By Theorem 2.2,

$$
\begin{aligned}
&\frac{1}{n} \sum_{t=1}^{n} g_1(\boldsymbol{X}_t;\boldsymbol{\theta}) - Eg_1(\boldsymbol{X}_1;\boldsymbol{\theta}) \sim O_p\left(n^{-\frac{1}{2}}\right), \\
&\frac{1}{n} \sum_{t=1}^{n} g_2(\boldsymbol{X}_t;\boldsymbol{\theta}) - Eg_2(\boldsymbol{X}_1;\boldsymbol{\theta}) \sim O_p\left(n^{-\frac{1}{2}}\right).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\sum_{|u|\le M_n} \left[\left(\frac{1}{n} \sum_{t=1}^{n} g_1(\boldsymbol{X}_t;\boldsymbol{\theta}) \right) \left(\frac{1}{n} \sum_{t=1}^{n} g_2(\boldsymbol{X}_t;\boldsymbol{\theta}) \right) - Eg_1(\boldsymbol{X}_1;\boldsymbol{\theta}) Eg_2(\boldsymbol{X}_1;\boldsymbol{\theta}) \right] \\
\sim& O_p(M_n n^{-\frac{1}{2}}) \sim o_p(n^{-\frac{1}{6}}).
\end{aligned} \tag{2.12}
$$

Denote $\boldsymbol{Y}_{t,u} = \left(\boldsymbol{X}_t^\top, \boldsymbol{X}_{t+u}^\top\right)^\top$, then $\{\boldsymbol{Y}_{t,u}\}$ is strictly stationary and strong mixing with mixing coefficient

$$
\alpha_{\boldsymbol{Y}_u}(s) = \begin{cases} \frac{1}{2}, & 0 \le s \le |u|, \\ \alpha_{\boldsymbol{X}}(s-|u|), & s > |u|. \end{cases}
$$

Define $h(\boldsymbol{Y}_{t,u};\boldsymbol{\theta}) = g_1(\boldsymbol{X}_t;\boldsymbol{\theta})g_2(\boldsymbol{X}_{t+u};\boldsymbol{\theta})$, then by (2.2) in [10], $\forall\ s, u$,

$$\begin{aligned}|\mathrm{Cov}(h(\boldsymbol{Y}_{t,u};\boldsymbol{\theta}), h(\boldsymbol{Y}_{t+s,u};\boldsymbol{\theta}))| \leq & 12 \cdot 2^{-\frac{m-2}{m}}(E|h(\boldsymbol{Y}_{t,u};\boldsymbol{\theta})|^m)^{\frac{1}{m}}(E|h(\boldsymbol{Y}_{t+s,u};\boldsymbol{\theta})|^m)^{\frac{1}{m}}\alpha_{\mathbf{Y}_u}^{\frac{m-2}{m}}(|s|)\\ \leq & 12 \cdot 2^{-\frac{m-2}{m}}(E|g(\boldsymbol{X}_1)|^{2m})^{\frac{2}{m}}\alpha_{\mathbf{Y}_u}^{\frac{m-2}{m}}(|s|).\end{aligned}$$

For $0 \leq u \leq M_n$,

$$\begin{aligned}&\mathrm{Var}\left(\frac{1}{n-u}\sum_{t=1}^{n-u}h(\boldsymbol{Y}_{t,u};\boldsymbol{\theta})\right)\\ =&\frac{1}{(n-u)^2}\left[(n-u)\mathrm{Var}(h(\boldsymbol{Y}_{t,u};\boldsymbol{\theta})) + 2\sum_{s=1}^{n-u-1}(n-u-s)\mathrm{Cov}(h(\boldsymbol{Y}_{t,u};\boldsymbol{\theta}), h(\boldsymbol{Y}_{t+s,u};\boldsymbol{\theta}))\right]\\ \leq&\frac{1}{(n-u)^2}\left[2(n-u)\sum_{s=0}^{n-u-1}|\mathrm{Cov}(h(\boldsymbol{Y}_{t,u};\boldsymbol{\theta}), h(\boldsymbol{Y}_{t+s,u};\boldsymbol{\theta}))|\right]\\ \leq&\frac{C}{n-u}\sum_{s=0}^{n-u-1}\alpha_{\mathbf{Y}_u}^{\frac{m-2}{m}}(s)\\ \leq&\frac{C}{n-M_n}\left(\sum_{s=0}^{u}\alpha_{\mathbf{Y}_u}^{\frac{m-2}{m}}(s) + \sum_{s=u+1}^{\infty}\alpha_{\mathbf{Y}_u}^{\frac{m-2}{m}}(s)\right)\\ \leq&\frac{C}{n-M_n}\left(\sum_{s=0}^{u}1 + \sum_{s=1}^{\infty}\alpha_{\boldsymbol{X}}^{\frac{m-2}{m}}(s)\right)\\ \leq&\frac{C}{n-M_n}\left(M_n + 1 + \sum_{s=1}^{\infty}\alpha_{\boldsymbol{X}}^{\frac{m-2}{m}}(s)\right) \sim O_p(M_n n^{-1}),\end{aligned}$$

where $C = 24\cdot 2^{-\frac{m-2}{m}}(E|g(\boldsymbol{X}_1)|^{2m})^{\frac{2}{m}}$. Take $c(n)$ such that $c(n) \to \infty$ as $n \to \infty$, by Chebyshev's Inequality, $\forall\ \varepsilon > 0$,

$$\begin{aligned}P\left(c(n)\left|\frac{1}{n-u}\sum_{t=1}^{n-u}h(\boldsymbol{Y}_{t,u};\boldsymbol{\theta}) - Eh(\boldsymbol{Y}_{1,u};\boldsymbol{\theta})\right| > \varepsilon\right) \leq & \frac{c^2(n)\mathrm{Var}\left(\frac{1}{n-u}\sum_{t=1}^{n-u}h(\boldsymbol{Y}_{t,u};\boldsymbol{\theta})\right)}{\varepsilon^2}\\ \sim & o(1),\end{aligned}$$

if $c(n) \sim o(M_n^{-\frac{1}{2}} n^{\frac{1}{2}})$. Hence,

$$\left| \frac{1}{n-u} \sum_{t=1}^{n-u} h(\boldsymbol{Y}_{t,u};\boldsymbol{\theta}) - Eh(\boldsymbol{Y}_{1,u};\boldsymbol{\theta}) \right| \sim O_p\left(M_n^{\frac{1}{2}} n^{-\frac{1}{2}} \right).$$

Similarly, for $-M_n \leq u < 0$,

$$\left| \frac{1}{n-|u|} \sum_{t=|u|+1}^{n} h(\boldsymbol{Y}_{t,u};\boldsymbol{\theta}) - Eh(\boldsymbol{Y}_{1,u};\boldsymbol{\theta}) \right| \sim O_p\left(M_n^{\frac{1}{2}} n^{-\frac{1}{2}} \right).$$

Therefore,

$$\begin{aligned} \sum_{u=0}^{\lfloor M_n \rfloor} \left(\frac{1}{n-u} \sum_{t=1}^{n-u} h(\boldsymbol{Y}_{t,u};\boldsymbol{\theta}) - Eh(\boldsymbol{Y}_{1,u};\boldsymbol{\theta}) \right) &\sim O_p\left(M_n^{\frac{3}{2}} n^{-\frac{1}{2}} \right) \sim o_p(1), \\ \sum_{u=-\lfloor M_n \rfloor}^{-1} \left(\frac{1}{n-|u|} \sum_{t=|u|+1}^{n} h(\boldsymbol{Y}_{t,u};\boldsymbol{\theta}) - Eg(\boldsymbol{Y}_{1,u};\boldsymbol{\theta}) \right) &\sim O_p\left(M_n^{\frac{3}{2}} n^{-\frac{1}{2}} \right) \sim o_p(1). \end{aligned} \tag{2.13}$$

Then by (2.11), (2.12) and (2.13), $\widehat{f}_n(0;\boldsymbol{\theta}) - f(0;\boldsymbol{\theta}) \xrightarrow{p} 0$ as $n \to \infty$.

It remains to show $\widehat{f}_n(0;\widehat{\boldsymbol{\theta}}_n) - \widehat{f}_n(0;\boldsymbol{\theta})$ as $n \to \infty$.

$$\begin{aligned} &\widehat{f}_n(0;\widehat{\boldsymbol{\theta}}_n) - \widehat{f}_n(0;\boldsymbol{\theta}) \\ =&\frac{1}{2\pi} \sum_{u=0}^{\lfloor M_n \rfloor} \left(\frac{1}{n-u} \sum_{t=1}^{n-u} \left(g_1\left(\boldsymbol{X}_t;\widehat{\boldsymbol{\theta}}_n\right) g_2\left(\boldsymbol{X}_{t+u};\widehat{\boldsymbol{\theta}}_n\right) - g_1(\boldsymbol{X}_t;\boldsymbol{\theta}) g_2(\boldsymbol{X}_{t+u};\boldsymbol{\theta}) \right) \right) \\ &+ \frac{1}{2\pi} \sum_{u=-\lfloor M_n \rfloor}^{-1} \left(\frac{1}{n-|u|} \sum_{t=|u|+1}^{n} \left(g_1\left(\boldsymbol{X}_t;\widehat{\boldsymbol{\theta}}_n\right) g_2\left(\boldsymbol{X}_{t+u};\widehat{\boldsymbol{\theta}}_n\right) - g_1(\boldsymbol{X}_t;\boldsymbol{\theta}) g_2(\boldsymbol{X}_{t+u};\boldsymbol{\theta}) \right) \right) \\ &- \frac{1}{2\pi} \sum_{|u| \leq M_n} \left[\left(\frac{1}{n} \sum_{t=1}^{n} g_1\left(\boldsymbol{X}_t;\widehat{\boldsymbol{\theta}}_n\right) \right) \left(\frac{1}{n} \sum_{t=1}^{n} g_2\left(\boldsymbol{X}_t;\widehat{\boldsymbol{\theta}}_n\right) \right) \right. \\ &\left. - \left(\frac{1}{n} \sum_{t=1}^{n} g_1(\boldsymbol{X}_t;\boldsymbol{\theta}) \right) \left(\frac{1}{n} \sum_{t=1}^{n} g_2(\boldsymbol{X}_t;\boldsymbol{\theta}) \right) \right]. \end{aligned}$$

By Mean Value Theorem, Cauchy-Schwartz Inequality and Triangle Inequality, $\forall\, t, u$, $\exists\, \widehat{\boldsymbol{\theta}}_n^{(t,u)} \in B(\boldsymbol{\theta}, \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n\|)$,

$$\begin{aligned}
& g_1(\boldsymbol{X}_t; \widehat{\boldsymbol{\theta}}_n) g_2(\boldsymbol{X}_{t+u}; \widehat{\boldsymbol{\theta}}_n) - g_1(\boldsymbol{X}_t; \boldsymbol{\theta}) g_2(\boldsymbol{X}_{t+u}; \boldsymbol{\theta}) \\
=& \left(\left. \frac{\partial g_1\left(\boldsymbol{X}_t; \boldsymbol{\theta}\right) g_2\left(\boldsymbol{X}_{t+u}; \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_n^{(t,u)}} \right)^{\top} \left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \right) \\
\leq& \left\| \left. \frac{\partial g_1\left(\boldsymbol{X}_t; \boldsymbol{\theta}\right) g_2\left(\boldsymbol{X}_{t+u}; \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_n^{(t,u)}} \right\| \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| \\
\leq& \left(\left\| \left. \frac{\partial g_1\left(\boldsymbol{X}_t; \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_n^{(t,u)}} g_2(\boldsymbol{X}_{t+u}; \widehat{\boldsymbol{\theta}}_n^{(t,u)}) \right\| + \left\| \left. \frac{\partial g_2\left(\boldsymbol{X}_{t+u}; \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_n^{(t,u)}} g_1(\boldsymbol{X}_t; \widehat{\boldsymbol{\theta}}_n^{(t,u)}) \right\| \right) \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| \\
\leq& (D(\boldsymbol{X}_t) g(\boldsymbol{X}_{t+u}) + D(\boldsymbol{X}_{t+u}) g(\boldsymbol{X}_t)) \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| \\
=& B(\mathbf{Y}_{t,u}) \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|,
\end{aligned}$$

where $B(\mathbf{Y}_{t,u}) = D(\boldsymbol{X}_t) g(\boldsymbol{X}_{t+u}) + D(\boldsymbol{X}_{t+u}) g(\boldsymbol{X}_t)$. By Lemma 2.1 and Theorem 2.1,

$$\frac{1}{n-u} \sum_{t=1}^{n-u} B(\mathbf{Y}_{t,u}) \overset{a.s.}{\to} EB(\mathbf{Y}_{1,u}),$$

as $n \to \infty$. Therefore,

$$\begin{aligned}
& \sum_{u=0}^{\lfloor M_n \rfloor} \left(\frac{1}{n-u} \sum_{t=1}^{n-u} \left(g_1\left(\boldsymbol{X}_t; \widehat{\boldsymbol{\theta}}_n\right) g_2\left(\boldsymbol{X}_{t+u}; \widehat{\boldsymbol{\theta}}_n\right) - g_1(\boldsymbol{X}_t; \boldsymbol{\theta}) g_2(\boldsymbol{X}_{t+u}; \boldsymbol{\theta}) \right) \right) \\
\leq& \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| \sum_{u=0}^{\lfloor M_n \rfloor} \left(\frac{1}{n-u} \sum_{t=1}^{n-u} B(\mathbf{Y}_{t,u}) \right) \qquad (2.14) \\
\sim& o_p(1).
\end{aligned}$$

Similarly, for $M_n \le u < 0$,

$$\sum_{u=-\lfloor M_n \rfloor}^{-1} \left(\frac{1}{n-|u|} \sum_{t=|u|+1}^{n} \left(g_1\left(\boldsymbol{X}_t; \widehat{\boldsymbol{\theta}}_n\right) g_2\left(\boldsymbol{X}_{t+u}; \widehat{\boldsymbol{\theta}}_n\right) - g_1(\boldsymbol{X}_t;\boldsymbol{\theta}) g_2(\boldsymbol{X}_{t+u};\boldsymbol{\theta}) \right) \right) \sim o_p(1). \tag{2.15}$$

By the similar derivation of (2.14),

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\left(g_1\left(\boldsymbol{X}_t; \widehat{\boldsymbol{\theta}}_n\right) - g_1(\boldsymbol{X}_t;\boldsymbol{\theta}) \right) \le & \frac{1}{n}\sum_{t=1}^{n} D(\boldsymbol{X}_t)\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| \sim o_p\left(M_n^{-1}\right), \\
\frac{1}{n}\sum_{t=1}^{n}\left(g_2\left(\boldsymbol{X}_t; \widehat{\boldsymbol{\theta}}_n\right) - g_2(\boldsymbol{X}_t;\boldsymbol{\theta}) \right) \le & \frac{1}{n}\sum_{t=1}^{n} D(\boldsymbol{X}_t)\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\| \sim o_p\left(M_n^{-1}\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
\sum_{|u|\le M_n} & \left[\left(\frac{1}{n}\sum_{t=1}^{n} g_1\left(\boldsymbol{X}_t; \widehat{\boldsymbol{\theta}}_n\right) \right) \left(\frac{1}{n}\sum_{t=1}^{n} g_2\left(\boldsymbol{X}_t; \widehat{\boldsymbol{\theta}}_n\right) \right) \right. \\
& \left. - \left(\frac{1}{n}\sum_{t=1}^{n} g_1(\boldsymbol{X}_t;\boldsymbol{\theta}) \right) \left(\frac{1}{n}\sum_{t=1}^{n} g_2(\boldsymbol{X}_t;\boldsymbol{\theta}) \right) \right] \sim o_p(M_n^{-1}).
\end{aligned} \tag{2.16}$$

By (2.14), (2.15) and (2.16), $\widehat{f}_n(0;\widehat{\boldsymbol{\theta}}_n) - \widehat{f}_n(0;\boldsymbol{\theta})$ as $n \to \infty$, which completes the proof.

□

2.2.4.1 Consistency of Plug-in Estimator $\widetilde{\boldsymbol{\mathcal{U}}}(\widetilde{\boldsymbol{\theta}})$

Based on the expression of the asymptotic variance $\boldsymbol{\mathcal{U}}(\boldsymbol{\theta}_0) = \boldsymbol{S}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{\mathcal{V}}(\boldsymbol{\theta}_0)\boldsymbol{S}^{-1}(\boldsymbol{\theta}_0)$, the consistency of $\widetilde{\boldsymbol{\mathcal{U}}}(\widetilde{\boldsymbol{\theta}})$ can be guaranteed by the consistency of $\widetilde{\boldsymbol{S}}(\widetilde{\boldsymbol{\theta}})$ and $\widetilde{\boldsymbol{\mathcal{V}}}(\widetilde{\boldsymbol{\theta}})$. The consistency of $\widetilde{\boldsymbol{S}}(\widetilde{\boldsymbol{\theta}})$ can be established based on the derivation in Section 1.2.4.1. Thus, it remains to who the consistency of $\widetilde{\boldsymbol{\mathcal{V}}}(\widetilde{\boldsymbol{\theta}})$. We demonstrate the proof for one entry of $\widetilde{\boldsymbol{\mathcal{V}}}(\widetilde{\boldsymbol{\theta}})$ and the rest can be proved in the same way. Consider the entry corresponding to β_{kp} and $\beta_{k'p'}$ where $k \neq k'$. The plug-in estimator is given by

$$\frac{2\pi \sum_{h=0}^{m} \rho_h \widetilde{f}_{\beta_{kp},\beta_{k'p'},h}(0;\widetilde{\boldsymbol{\theta}})}{\sum_{j=0}^{m} \rho_j},$$

where $\widetilde{f}$ is the estimator of the cross spectral density following the expression in Theorem 2.4. Here we show the consistency

$$\widetilde{f}_{\beta_{kp},\beta_{k'p'},h}(0;\widetilde{\boldsymbol{\theta}}) \xrightarrow{p} f_{\beta_{kp},\beta_{k'p'},h}(0;\boldsymbol{\theta}_0),$$

as $n \to \infty$ for $h \neq k, k'$. The case of $h = k, k'$ can be proved in the same way.

By the definition of $f_{\beta_{kp},\beta_{k'p'},h}(0;\boldsymbol{\theta})$, it is the cross spectral density of $\{\varphi_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})h_{kp}(\boldsymbol{X}_l^{(h)})\}$ and $\{\varphi_{k'}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})h_{k'p'}(\boldsymbol{X}_l^{(h)})\}$. Since

$$\varphi_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}) \leq 1, \quad \varphi_{k'}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}) \leq 1,$$

then

$$\begin{aligned} |\varphi_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})h_{kp}(\boldsymbol{X}_l^{(h)})| \leq & |h_{kp}(\boldsymbol{X}_l^{(h)})| + |h_{k'p'}(\boldsymbol{X}_l^{(h)})|, \\ |\varphi_{k'}(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})h_{k'p'}(\boldsymbol{X}_l^{(h)})| \leq & |h_{kp}(\boldsymbol{X}_l^{(h)})| + |h_{k'p'}(\boldsymbol{X}_l^{(h)})|, \end{aligned}$$

with $E|h_{kp}(\boldsymbol{X}_l^{(h)})|^{m_h} < \infty$ and $E|h_{k'p'}(\boldsymbol{X}_l^{(h)})|^{m_h} < \infty$ for $m_h > \frac{4a_h}{a_h-1}$ by Assumption 2.2. By

Triangle Inequality,

$$
\begin{aligned}
&\left\|\frac{\partial\varphi_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})h_{kp}(\boldsymbol{X}_l^{(h)})}{\partial\boldsymbol{\theta}}\right\| \\
\leq & |h_{kp}(\boldsymbol{X}_l^{(h)})| \left(\frac{\rho_k w_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}) \left\|\frac{\partial \log w_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|}{\sum_{j=0}^m \rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})} \right. \\
& \left. + \frac{\rho_k w_k(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}) \left[\sum_{j=0}^m \rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta}) \left\|\frac{\partial \log w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|\right]}{\left[\sum_{j=0}^m \rho_j w_j(\boldsymbol{X}_l^{(h)};\boldsymbol{\theta})\right]^2} \right) \\
\leq & |h_{kp}(\boldsymbol{X}_l^{(h)})| \left(1 + \sum_{k=1}^m \sum_{p=1}^{p_k} |h_{kp}(\boldsymbol{X}_l^{(h)})|\right),
\end{aligned}
$$

and

$$
E\left|h_{kp}(\boldsymbol{X}_l^{(h)})\left(1+\sum_{k=1}^m\sum_{p=1}^{p_k}\left|h_{k,p}(\boldsymbol{X}_l^{(h)})\right|\right)\right|^{\frac{m_h}{2}} < \infty
$$

for $m_h > \frac{4a_h}{a_h - 1}$, by Assumption 2.2. Similarly, this also holds for k' and p'. Thus, by Assumption 2.1, 2.2 and Theorem 2.4,

$$
\widetilde{f}_{\beta_{kp},\beta_{k'p'},h}(0;\widetilde{\boldsymbol{\theta}}) \xrightarrow{p} f_{\beta_{kp},\beta_{k'p'},h}(0;\boldsymbol{\theta}_0),
$$

as $n \to \infty$. As mentioned above, the case of $h = k, k'$ and the other entries can be proved by the same derivation. Thus, the consistency of $\widetilde{\boldsymbol{\mathcal{V}}}(\widetilde{\boldsymbol{\theta}})$ is established and hence $\widetilde{\boldsymbol{\mathcal{U}}}(\widetilde{\boldsymbol{\theta}})$ is consistent.

2.2.4.2 Consistency of Plug-in Estimator $\widetilde{\nu}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$

By (2.10), the expression of $\widetilde{\nu}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$ can be written as

$$\begin{aligned}\widetilde{\nu}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}}) =& 2\pi\left(\sum_{j=0}^{m}\rho_j\right)\left(\sum_{h=0}^{m}\rho_h\widetilde{f}_{\gamma_{\boldsymbol{x}},\gamma_{\boldsymbol{x}},h}(0;\widetilde{\boldsymbol{\theta}})\right) - 2\widetilde{\boldsymbol{K}}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{\mathcal{W}}}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}}) \\ &+ \widetilde{\boldsymbol{K}}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{\mathcal{V}}}(\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{K}}^{\top}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}}).\end{aligned}$$

The consistency of $\widetilde{\boldsymbol{\mathcal{V}}}(\widetilde{\boldsymbol{\theta}})$ is established in Section 2.2.4.1. The consistency of $\widetilde{\boldsymbol{\mathcal{W}}}(\widetilde{\boldsymbol{\theta}})$ $\widetilde{f}_{\gamma_{\boldsymbol{x}},\gamma_{\boldsymbol{x}},h}(0;\widetilde{\boldsymbol{\theta}})$ for $h = 0,\dots,m$, can be proved by the same derivation in Section 2.2.4.1. $\widetilde{\boldsymbol{K}}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$ consists of $\widetilde{\boldsymbol{S}}(\widetilde{\boldsymbol{\theta}})$, $\widetilde{A}_k(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$ and $\widetilde{\boldsymbol{B}}_k(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$ for $k = 1,\dots,m$. The consistency of $\widetilde{\boldsymbol{S}}(\widetilde{\boldsymbol{\theta}})$ can be proved using the derivation in Section 1.2.4.1 while the consistency of $\widetilde{A}_k(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$ and $\widetilde{\boldsymbol{B}}_k(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$ can be shown by the derivation in Section 1.2.4.2. Thus, $\widetilde{\nu}(\boldsymbol{x};\widetilde{\boldsymbol{\theta}})$ is a consistent estimator of $\nu(\boldsymbol{x};\boldsymbol{\theta}_0)$.

2.3 Simulation

In the section, simulation results are presented to verify the asymptotic properties of $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{G}$. Moreover, comparison is also made between DRM and DRMWD to show the improvement when dependence is taken into consideration. The processes we use are Autoregressive Moving Average (ARMA) sequences since their strong mixing coefficients have a exponential decay by [49], which satisfies Assumption 2.1. Additionally, through out this section, all cross spectral densities are estimated based on the rectangular window with bandwidth $M_n = \lfloor n^{\frac{1}{3.5}} \rfloor$.

2.3.1 Confidence Interval for $\boldsymbol{\theta}$

We generate 95% CIs for each parameter of $\boldsymbol{\theta}$ based on DRM and DRMWD, and compare these two models by the coverage rates of the CIs. Note that we only examine the parameters of $\boldsymbol{\beta}$ since the constant $\boldsymbol{\alpha}$ depends on $\boldsymbol{\beta}$ based on (1.3) and it does not provide information regarding the significance of tilt functions. All sequences generated have length 5000, and the coverage rates are calculated by 1000 iterations.

2.3.1.1 Normal vs. Normal

Suppose g_0 and g_1 are densities such that $g_0 \sim N(\mu_0, \sigma_0^2)$ and $g_1 \sim N(\mu_1, \sigma_1^2)$, then

$$\begin{aligned}
\frac{g_1(x)}{g_0(x)} &= \exp(\alpha + \beta_1 x + \beta_2 x^2) \\
\alpha &= \log \frac{\sigma_0}{\sigma_1} + \frac{\mu_0^2}{2\sigma_0^2} - \frac{\mu_1^2}{2\sigma_1^2} \\
\beta_1 &= \frac{\mu_1}{\sigma_1^2} - \frac{\mu_0}{\sigma_0^2} \\
\beta_2 &= \frac{1}{2\sigma_0^2} - \frac{1}{2\sigma_1^2}.
\end{aligned}$$

Case 1. Let

$$\begin{aligned}
X_t^{(0)} &= 0.5 X_{t-1}^{(0)} + \varepsilon_t^{(0)} \\
X_t^{(1)} &= 0.3 + 0.7 X_{t-1}^{(1)} + \varepsilon_t^{(1)}
\end{aligned}$$

where $\varepsilon_t^{(0)}$'s are IID $N(0, 0.75)$ noise, and $\varepsilon_t^{(1)}$'s are IID $N(0,1)$ noise. Then $X_t^{(0)} \sim N(0,1)$ and $X_t^{(1)} \sim N(1, \frac{100}{51})$. Thus, $\beta_1 = \frac{51}{100}$ and $\beta_2 = \frac{49}{200}$.

Case 2. Let

$$\begin{aligned} X_t^{(0)} =& 0.5X_{t-1}^{(0)} + \varepsilon_t^{(0)} \\ X_t^{(1)} =& 0.3 + 0.7X_{t-1}^{(1)} + \varepsilon_t^{(1)} - 0.1\varepsilon_{t-1}^{(1)} \end{aligned}$$

where $\varepsilon_t^{(0)}$'s are IID $N(0, \frac{3}{4})$ noise, and $\varepsilon_t^{(1)}$'s are IID $N(0,1)$ noise. Then $X_t^{(0)} \sim N(0,1)$ and $X_t^{(1)} \sim N(1, \frac{87}{51})$. Thus, $\beta_1 = \frac{51}{87}$ and $\beta_2 = \frac{18}{87}$.

From Table 2.1, it is observed that the coverage rates of CIs generated by DRMWD are higher and closer to 0.95.

Table 2.1: Coverage rates of CIs for β_1 and β_2 in Case 1 and 2.

Case	Model	Coverage rate β_1	Coverage rate β_2
1	DRM	0.712	0.846
	DRMWD	0.924	0.945
2	DRM	0.729	0.835
	DRMWD	0.908	0.942

2.3.1.2 Gamma vs. Gamma

Suppose g_0 and g_1 are densities such that $g_0 \sim Gamma(a_0, b_0)$ and $g_1 \sim Gamma(a_1, b_1)$ where b_0 and b_1 are rate parameters, then

$$\begin{aligned} \frac{g_1(x)}{g_0(x)} =& \exp(\alpha + \beta_1 x + \beta_2 \log(x)) \\ \alpha =& \log \frac{b_1^{a_1}\Gamma(a_0)}{b_0^{a_0}\Gamma(a_1)} \\ \beta_1 =& b_0 - b_1 \\ \beta_2 =& a_1 - a_0. \end{aligned}$$

Case 3-6. Let

$$X_t^{(0)} = 0.5X_{t-1}^{(0)} + \varepsilon_t^{(0)}$$

$$X_t^{(1)} = 0.5X_{t-1}^{(1)} + \varepsilon_t^{(1)}$$

where $\varepsilon_t^{(0)}$'s and $\varepsilon_t^{(1)}$'s are generated based on [64] and [67] to ensure that

Case 3: $X_t^{(0)} \sim Gamma(2, 1)$ and $X_t^{(1)} \sim Gamma(2, 2)$. $\beta_1 = -1$ and $\beta_2 = 0$.

Case 4: $X_t^{(0)} \sim Gamma(2, 1)$ and $X_t^{(1)} \sim Gamma(3, 2)$. $\beta_1 = -1$ and $\beta_2 = 1$.

Case 5: $X_t^{(0)} \sim Gamma(7, 1)$ and $X_t^{(1)} \sim Gamma(8, 2)$. $\beta_1 = -1$ and $\beta_2 = 1$.

Case 6: $X_t^{(0)} \sim Gamma(2, \frac{1}{4})$ and $X_t^{(1)} \sim Gamma(3, \frac{1}{2})$. $\beta_1 = -\frac{1}{4}$ and $\beta_2 = 1$.

From Table 2.2, we see that the coverage rates in Case 3 where the tile function is $h(x) = x$, resemble the coverage rates in Case 1 and 2. However, in Case 4, the coverage rates of DRMWD are closed to 1 indicating that the CIs are too conservative. We postulate that this is caused by including $\log(x)$ in the tilt function. For $Gamma(a, b)$, more data are closer to 0 when a is small and b is large. Then the values of $\log(X)$ are large, which makes the spectral density estimates in $\widetilde{\boldsymbol{\mathcal{V}}}(\widetilde{\boldsymbol{\theta}})$ overestimate the variances. Recall that $\widetilde{\boldsymbol{\mathcal{V}}}(\widetilde{\boldsymbol{\theta}})$ depends on the estimators of spectral densities at 0 frequency, these estimators are less stable as mentioned by [48]. Thus, further investigation shall be done for finding alternative spectral density estimators. From Case 5 and Case 6, we observe that the overestimation issue is alleviated when a becomes larger and b becomes smaller.

Table 2.2: Coverage rates of CIs for β_1 and β_2 in Case 3-6. A hyphen "-" indicates that the corresponding parameter is not in the model.

Case	Model	Coverage rate β_1	Coverage rate β_2
3	DRM	0.747	-
	DRMWD	0.930	-
4	DRM	0.844	0.865
	DRMWD	0.995	0.997
5	DRM	0.892	0.878
	DRMWD	0.937	0.939
6	DRM	0.865	0.844
	DRMWD	0.977	0.982

2.3.1.3 Gamma vs. Lognormal

Suppose g_0 and g_1 are densities such that $g_0 \sim Gamma(a, b)$ and $g_1 \sim LN(\mu, \sigma^2)$ where b is the rate parameter, then

$$\begin{aligned}
\frac{g_1(x)}{g_0(x)} &= \exp(\alpha + \beta_1 x + \beta_2 \log(x) + \beta_3 (\log(x))^2) \\
\alpha &= \log \frac{\Gamma(a)}{b^a \sigma \sqrt{2\pi}} - \frac{\mu}{2\sigma^2} \\
\beta_1 &= b \\
\beta_2 &= \frac{\mu}{\sigma^2} - a \\
\beta_3 &= -\frac{1}{2\sigma^2}.
\end{aligned}$$

Case 7-8. Let

$$\begin{aligned}
X_t^{(0)} &= 0.5 X_{t-1}^{(0)} + \varepsilon_t^{(0)} \\
\log(X_t^{(1)}) &= 0.8 \log(X_{t-1}^{(1)}) + \varepsilon_t^{(1)}
\end{aligned}$$

where $\varepsilon_t^{(0)}$'s are generated by the method in Section 2.3.1.2 to ensure that $X_t^{(0)} \sim Gamma(3, 1)$ in Case 7, and $X_t^{(0)} \sim Gamma(3, \frac{1}{2})$ in Case 8. In both cases, $\varepsilon_t^{(1)}$'s are IID $N(0, 0.36)$ noise so that $X_t^{(1)} \sim LN(0,1)$. $\beta_1 = 1$ in Case 7 while $\beta_1 = \frac{1}{2}$ in Case 8. $\beta_2 = -3$ and $\beta_3 = -\frac{1}{2}$ in both cases.

As shown in Table 2.3, the overestimation is severer in Case 7 since the tilt function contains $(\log(x))^2$ but it is mitigated in Case 8 when b becomes smaller.

Table 2.3: Coverage rates of CIs for β_1 β_2 and β_3 in Case 7 and 8.

Case	Model	Coverage rate β_1	Coverage rate β_2	Coverage rate β_3
7	DRM	0.876	0.830	0.867
	DRMWD	1.000	1.000	1.000
8	DRM	0.869	0.794	0.863
	DRMWD	0.988	0.966	0.975

2.3.1.4 Halfnormal vs. Exponential

Suppose g_0 and g_1 are densities such that $g_0 \sim HN(\sigma)$ and $g_1 \sim Exp(\lambda)$ where λ is the rate parameter, then

$$\begin{aligned}\frac{g_1(x)}{g_0(x)} &= \exp(\alpha + \beta_1 x + \beta_2 x^2)\\ \alpha &= \log\left(\lambda\sigma\sqrt{\frac{\pi}{2}}\right)\\ \beta_1 &= -\lambda\\ \beta_2 &= \frac{1}{2\sigma^2}.\end{aligned}$$

Case 9-10. Let

$$\begin{aligned} Z_t^{(0)} =& 0.5 Z_{t-1}^{(0)} + \varepsilon_t^{(0)} \\ X_t^{(0)} =& \left| Z_t^{(0)} \right| \\ X_t^{(1)} =& 0.5 X_{t-1}^{(1)} + \varepsilon_t^{(1)} \end{aligned}$$

where $\varepsilon_t^{(0)}$'s are IID $N(0, \frac{3}{4})$ in both cases so that $X_t^{(0)} \sim HN(1)$. $\varepsilon_t^{(1)}$'s are generated by the method in Section 2.3.1.2 to ensure that $X_t^{(1)} \sim Exp(1)$ in Case 9 and $X_t^{(1)} \sim Exp(2)$ in Case 10. $\beta_1 = -1$ in Case 9 while $\beta_1 = -2$ in Case 10. $\beta_2 = \frac{1}{2}$ in both cases.

Based on Table 2.4, the coverage rates of DRMWD are also closer to 0.95 than that of DRM.

Table 2.4: Coverage rates of CIs for β_1 and β_2 in Case 9 and 10.

Case	Model	Coverage rate β_1	Coverage rate β_2
9	DRM	0.877	0.903
	DRMWD	0.985	0.984
10	DRM	0.892	0.925
	DRMWD	0.952	0.967

2.3.2 Confidence Interval for G

In this section, 95% CIs for univariate and bivariate threshold probabilities of the reference distribution are generated based on DRM and DRMWD, and a comparison is made based on their coverage rates. All sequences generated have length 5000, and the coverage rates are calculated by 1000 iterations.

2.3.2.1 Univariate Cases

Case 1. Let

$$
\begin{aligned}
X_t^{(0)} =& 2 + 0.5X_{t-1}^{(0)} + \varepsilon_t^{(0)} \\
X_t^{(1)} =& 0.6 + 0.8X_{t-1}^{(1)} + \varepsilon_t^{(1)}
\end{aligned}
$$

where $\varepsilon_t^{(0)}$'s are IID $N(0, 0.75)$ noise, and $\varepsilon_t^{(1)}$'s are IID $N(0, 0.72)$ noise. Then $X_t^{(0)} \sim N(4, 1)$ and $X_t^{(1)} \sim N(3, 2)$.

Case 2. Let

$$
\begin{aligned}
X_t^{(0)} =& 0.5X_{t-1}^{(0)} + \varepsilon_t^{(0)} \\
X_t^{(1)} =& 0.5X_{t-1}^{(1)} + \varepsilon_t^{(1)}
\end{aligned}
$$

where $\varepsilon_t^{(0)}$'s and $\varepsilon_t^{(1)}$'s are generated by the method in Section 2.3.1.2 to ensure that $X_t^{(0)} \sim Gamma(8, 2)$ and $X_t^{(1)} \sim Gamma(6, 1)$.

Case 3. Let

$$
\begin{aligned}
\log(X_t^{(0)}) =& 0.2 + 0.8\log(X_{t-1}^{(0)}) + \varepsilon_t^{(0)} \\
X_t^{(1)} =& 0.5X_{t-1}^{(1)} + \varepsilon_t^{(1)}
\end{aligned}
$$

where $\varepsilon_t^{(0)}$'s are IID $N(0, 0.36)$ noise so that $X_t^{(0)} \sim LN(1, 1)$, and $\varepsilon_t^{(1)}$'s are generated by the method in Section 2.3.1.2 to ensure $X_t^{(1)} \sim Gamma(5, 1)$.

Case 4. Let

$$\begin{aligned} Z_t^{(0)} &= 0.8 Z_{t-1}^{(0)} + \varepsilon_t^{(0)} \\ X_t^{(0)} &= \left| Z_t^{(0)} \right| \\ X_t^{(1)} &= 0.5 X_{t-1}^{(1)} + \varepsilon_t^{(1)} \end{aligned}$$

where $\varepsilon_t^{(0)}$'s are IID $N(0, 9)$ noise so that $X_t^{(0)} \sim HN(5)$. $\varepsilon_t^{(1)}$'s are generated by the method in Section 2.3.1.2 to ensure that $X_t^{(1)} \sim Exp(2)$.

The coverage rates of CIs for reference threshold probability $G(T)$ with $T = 3, 4, 5, 6, 7$, are given in Table 2.5. In most cases, the coverage rates are improved by DRMWD but there are few cases where the coverage rates of DRMWD is less than that of DRM. Moreover, the coverage rates of DRMWD are more sensitive to the thresholds compared to that of DRM. This is because both $\widetilde{\mathcal{W}}(x; \widetilde{\boldsymbol{\theta}})$ and $\widetilde{\mathcal{V}}(\widetilde{\boldsymbol{\theta}})$ are estimated based on cross spectral density estimators, which are less stable as discussed in Section 2.3.1.2.

Table 2.5: Coverage rates of CIs for threshold probability $G(T)$ with $T = 3, 4, 5, 6, 7$.

Case	Model	$T = 3$	$T = 4$	$T = 5$	$T = 6$	$T = 7$
1	DRM	0.798	0.774	0.806	0.860	0.885
	DRMWD	0.999	0.825	0.911	0.911	0.975
2	DRM	0.789	0.791	0.811	0.849	0.860
	DRMWD	0.952	0.953	0.956	0.957	0.950
3	DRM	0.508	0.539	0.543	0.555	0.567
	DRMWD	0.958	0.969	0.968	0.963	0.964
4	DRM	0.765	0.730	0.732	0.730	0.720
	DRMWD	0.694	0.763	0.839	0.897	0.944

2.3.2.2 Bivariate Cases

In time series analysis, it is of interest to obtain the joint distribution of neighboring observations that reveals the association between past and future. Here we present a comparison of the coverage rates of 95% CIs for the bivariate reference threshold probability $P(X_{t-1}^{(0)} \in A, X_t^{(0)} \in B)$.

Case 5. Let

$$\begin{aligned} X_t^{(0)} =& 0.5X_{t-1}^{(0)} + \varepsilon_t^{(0)} \\ X_t^{(1)} =& 0.8X_{t-1}^{(1)} + \varepsilon_t^{(1)} \end{aligned}$$

where $\varepsilon_t^{(0)}$'s are IID $N(0, 0.75)$ noise and $\varepsilon_t^{(1)}$'s are IID $N(0, 0.36)$ noise. Thus,

$$\begin{bmatrix} X_{t-1}^{(0)} \\ X_t^{(0)} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right), \quad \begin{bmatrix} X_{t-1}^{(1)} \\ X_t^{(1)} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}\right),$$

and the tilt function used is $\boldsymbol{h}(\boldsymbol{x}) = (x_1^2, x_1x_2, x_2^2)$.

Case 6. Let

$$\begin{aligned} X_t^{(0)} =& 0.8X_{t-1}^{(0)} + \varepsilon_t^{(0)} - 0.2\varepsilon_{t-1}^{(0)} \\ X_t^{(1)} =& 0.8X_{t-1}^{(1)} + \varepsilon_t^{(1)} \end{aligned}$$

where $\varepsilon_t^{(0)}$'s are IID $N(0, 0.5)$ noise and $\varepsilon_t^{(1)}$'s are IID $N(0, 0.36)$ noise. Thus,

$$\begin{bmatrix} X_{t-1}^{(0)} \\ X_t^{(0)} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix}\right), \quad \begin{bmatrix} X_{t-1}^{(1)} \\ X_t^{(1)} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}\right),$$

and the tilt function used is $\boldsymbol{h}(\boldsymbol{x}) = (x_1^2, x_1x_2, x_2^2)$.

From Table 2.6, in most cases, the coverage rates of DRMWD are closer to 0.95 and more stable compared to that of DRM.

Table 2.6: Coverage rates of CIs for the reference threshold probabilities $P_1 = P(X_{t-1}^{(0)} \leq 1, X_t^{(0)} > 1)$, $P_2 = P(X_{t-1}^{(0)} \leq 0, X_t^{(0)} \leq 1)$, $P_3 = P(X_{t-1}^{(0)} > 1, X_t^{(0)} \leq -1)$, $P_4 = P(X_{t-1}^{(0)} > 2, X_t^{(0)} > 1)$ and $P_5 = P(X_{t-1}^{(0)} > 0, X_t^{(0)} \leq 0)$.

Case	Model	P_1	P_2	P_3	P_4	P_5
5	DRM	0.963	0.709	0.938	0.831	0.994
	DRMWD	0.944	0.947	0.940	0.870	0.985
6	DRM	0.936	0.576	0.898	0.716	0.965
	DRMWD	0.941	0.947	0.897	0.815	0.984

2.3.3 Testing for Equality of Distributions

DRM can be used to test the equality of distributions based on the asymptotic normality of $\widetilde{\boldsymbol{\theta}}$ (see [21], [32]). Thus, if we rewrite (1.13) as

$$\sqrt{n}\left(\begin{bmatrix} \widetilde{\boldsymbol{\alpha}} \\ \widetilde{\boldsymbol{\beta}} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\alpha}_0 \\ \boldsymbol{\beta}_0 \end{bmatrix}\right) \xrightarrow{d} N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11}(\boldsymbol{\theta}_0) & \boldsymbol{\Sigma}_{12}(\boldsymbol{\theta}_0) \\ \boldsymbol{\Sigma}_{12}^\top(\boldsymbol{\theta}_0) & \boldsymbol{\Sigma}_{22}(\boldsymbol{\theta}_0) \end{bmatrix}\right),$$

then

$$n\widetilde{\boldsymbol{\beta}}^\top \widetilde{\boldsymbol{\Sigma}}_{22}^{-1}(\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{\beta}} \xrightarrow{d} \chi^2_{\sum_{k=1}^m p_k}$$

as $n \to \infty$ under $\boldsymbol{\beta}_0 = \mathbf{0}$. Similarly, for DRMWD, we have

$$n\widetilde{\boldsymbol{\beta}}^\top \widetilde{\boldsymbol{\mathcal{U}}}_{22}^{-1}(\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{\beta}} \xrightarrow{d} \chi^2_{\sum_{k=1}^m p_k}$$

as $n \to \infty$ under $\boldsymbol{\beta}_0 = \mathbf{0}$ based on (2.5). Hence, the equivalence of densities

$$H_0: \ g_0 = \cdots = g_m \quad \Leftrightarrow \quad H_0: \ \boldsymbol{\beta}_0 = \mathbf{0}$$

is rejected if

$$n\widetilde{\boldsymbol{\beta}}^\top \widetilde{\boldsymbol{\Sigma}}_{22}^{-1}(\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{\beta}} > \chi^2_{\alpha, \sum_{k=1}^m p_k}$$

for DRM, and

$$n\widetilde{\boldsymbol{\beta}}^\top \widetilde{\boldsymbol{\mathcal{U}}}_{22}^{-1}(\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{\beta}} > \chi^2_{\alpha, \sum_{k=1}^m p_k}$$

for DRMWD, at the significance level of α.

We use four cases to check the rejection rates of the two tests in both univariate case and bivariate case. Additionally, for univariate case, we also compare the two tests with Kolmogorov-Smirnov (K-S) test, a commonly used nonparametric test for equality of distributions. We use "ks.test()" in R package "stats" to perform K-S test. All tests are done at the significance level of 0.05, and all rejection rates are calculated by 1000 iterations. The sizes of samples are chosen to be $N = 100, 200, 500, 1000, 2000, 5000$. The tilt function used for univariate cases is $\boldsymbol{h}(x) = (x, x^2)^\top$, and for bivariate cases is $\boldsymbol{h}(\boldsymbol{x}) = (x_1, x_2, x_1^2, x_1x_2, x_2^2)^\top$.

Case 1. Let

$$
\begin{aligned}
X_t^{(0)} =& 0.5X_{t-1}^{(0)} + \varepsilon_t^{(0)} \\
X_t^{(1)} =& 0.5X_{t-1}^{(1)} + \varepsilon_t^{(1)}
\end{aligned}
$$

where $\varepsilon_t^{(0)}$'s and $\varepsilon_t^{(1)}$'s are IID $N(0, 0.75)$ noise. Thus, the distributions are identical for $X_t^{(0)}$ and $X_t^{(1)}$ in the univariate case, and for $(X_{t-1}^{(0)}, X_t^{(0)})^\top$ and $(X_{t-1}^{(1)}, X_t^{(1)})^\top$ in the bivariate case.

Case 2. Let

$$
\begin{aligned}
X_t^{(0)} =& 0.8X_{t-1}^{(0)} + \varepsilon_t^{(0)} - 0.2\varepsilon_{t-1}^{(0)} \\
X_t^{(1)} =& 0.8X_{t-1}^{(1)} + \varepsilon_t^{(1)} - 0.2\varepsilon_{t-1}^{(1)}
\end{aligned}
$$

where $\varepsilon_t^{(0)}$'s and $\varepsilon_t^{(1)}$'s are IID $N(0, 0.5)$ noise. This is a more complicated model compared to Case 1, while the distributions are still identical for $X_t^{(0)}$ and $X_t^{(1)}$ in the univariate case, and for $(X_{t-1}^{(0)}, X_t^{(0)})^\top$ and $(X_{t-1}^{(1)}, X_t^{(1)})^\top$ in the bivariate case.

Case 3. Let

$$
\begin{aligned}
X_t^{(0)} =& 0.5X_{t-1}^{(0)} + \varepsilon_t^{(0)} \\
X_t^{(1)} =& 0.8X_{t-1}^{(1)} + \varepsilon_t^{(1)}
\end{aligned}
$$

where $\varepsilon_t^{(0)}$'s are IID $N(0, 0.75)$ noise and $\varepsilon_t^{(1)}$'s are IID $N(0, 0.36)$ noise. The univariate distributions are still identical for $X_t^{(0)}$ and $X_t^{(1)}$ while distributions of in $(X_{t-1}^{(0)}, X_t^{(0)})^\top$ and $(X_{t-1}^{(1)}, X_t^{(1)})^\top$ are different.

Case 4. Let

$$X_t^{(0)} = 0.5X_{t-1}^{(0)} + \varepsilon_t^{(0)}$$

$$X_t^{(1)} = 0.2 + 0.8X_{t-1}^{(1)} + \varepsilon_t^{(1)}$$

where $\varepsilon_t^{(0)}$'s are IID $N(0, 0.75)$ noise and $\varepsilon_t^{(1)}$'s are IID $N(0, 0.72)$ noise. The univariate distributions are different since $X_t^{(0)} \sim N(0,1)$ while $X_t^{(1)} \sim N(1,2)$. The distributions of $(X_{t-1}^{(0)}, X_t^{(0)})^\top$ and $(X_{t-1}^{(1)}, X_t^{(1)})^\top$ are also different.

Based on Table 2.7, the rejection rates of DRMWD are closer to 0.05 when H_0 is true compared to that of DRM and K-S test. Moreover, the rejection rates of DRMWD under H_0 get closer to 0.05 as N increases though the approaching speed is rather slow. Additionally, all rejection rates approach 1 as N increases when H_0 is false. Thus, a substantial improvement has been made by DRMWD for the test for equality of distributions when data are weakly dependent.

2.4 Analysis of Air Monitoring Data

Nitrogen dioxide (NO_2) is an air pollutant stemming from the burning of fuel. Research works have been shown that the exposure to NO_2 is related to respiratory and cardiovascular diseases (e.g. [3], [6]). The outdoor NO_2 concentration data can be found on EPA website (`https://aqs.epa.gov/aqsweb/airdata/download_files.html`), and we will demonstrate the use of DRMWD by the analysis of daily NO_2 concentration data. The description of sites and monitors can be found on `https://aqs.epa.gov/aqsweb/airdata/download_files.html`. We select eight sites to perform the test for equality of distribution for two time periods, 2010-2012

Table 2.7: Rejection rates of tests for equality of distributions based on DRM, DRMWD and K-S test. For Case 1, Case 2 and the univariate case in Case 3, the rejection rate represents the estimated Type-I error. For the bivariate case in Case 3 and Case 4, the rejection rate represents the estimated power.

Case	Dimension	Test	$N = 100$	$N = 200$	$N = 500$	$N = 1000$	$N = 2000$	$N = 5000$
1	Univariate	DRM	0.250	0.260	0.251	0.258	0.267	0.284
		DRMWD	0.113	0.097	0.063	0.054	0.058	0.039
		K-S test	0.189	0.207	0.235	0.220	0.241	0.226
	Bivariate	DRM	0.458	0.468	0.471	0.471	0.461	0.490
		DRMWD	0.268	0.198	0.176	0.139	0.148	0.132
2	Univariate	DRM	0.535	0.587	0.542	0.575	0.584	0.582
		DRMWD	0.232	0.153	0.100	0.083	0.078	0.071
		K-S test	0.449	0.492	0.480	0.493	0.499	0.492
	Bivariate	DRM	0.675	0.718	0.671	0.706	0.713	0.678
		DRMWD	0.369	0.269	0.202	0.170	0.163	0.136
3	Univariate	DRM	0.470	0.489	0.498	0.497	0.514	0.507
		DRMWD	0.168	0.119	0.099	0.073	0.078	0.076
		K-S test	0.376	0.380	0.433	0.417	0.435	0.430
	Bivariate	DRM	0.998	1.000	1.000	1.000	1.000	1.000
		DRMWD	0.935	0.968	0.994	0.997	1.000	1.000
4	Univariate	DRM	0.973	1.000	1.000	1.000	1.000	1.000
		DRMWD	0.886	0.994	1.000	1.000	1.000	1.000
		K-S test	0.947	0.998	1.000	1.000	1.000	1.000
	Bivariate	DRM	0.994	1.000	1.000	1.000	1.000	1.000
		DRMWD	0.902	0.965	0.988	0.993	0.998	1.000

and 2020-2022. The Information of selected sites are given in Table 2.8.

Table 2.8: The information of selected sites.

Site	State name (number)	County name (number)	Site name (number)
1	Arizona (4)	Maricopa (13)	Central Phoenix (3002)
2	California (6)	Contra Costa (13)	Bethel Island (1002)
3	Georgia (13)	DeKalb (89)	South DeKalb (2)
4	Maryland (24)	Baltimore (5)	Essex (3001)
5	Missouri (29)	Jackson (95)	Troost (34)
6	North Carolina (37)	Mecklenburg (119)	Garinger High School (41)
7	Texas (48)	Bexar (29)	Calaveras Lake (59)
8	Wyoming (56)	Sweetwater (37)	Wamsutter (200)

For each site, we use daily records from 2020 to 2022 as the reference $Z_t^{(0)}$ and that from 2010 to 2012 as the alternative $Z_t^{(1)}$. Then we take log differences to obtain $X_t^{(0)}$ and $X_t^{(1)}$ such that

$$
\begin{aligned}
X_t^{(0)} =& \log(Z_t^{(0)}) - \log(Z_{t-1}^{(0)}), \\
X_t^{(1)} =& \log(Z_t^{(1}) - \log(Z_{t-1}^{(1)}).
\end{aligned}
$$

Then we employ the test the equality of univariate distributions of $X_t^{(0)}$ and $X_t^{(1)}$ based on the test in Section 2.3.3. Furthermore, we also test the equality of bivariate distribution for

$$
\begin{aligned}
\mathbf{Y}_t^{(0)} =& (X_{t-1}^{(0)}, X_t^{(0)})^\top, \\
\mathbf{Y}_t^{(1)} =& (X_{t-1}^{(1)}, X_t^{(1)})^\top,
\end{aligned}
$$

which detects the structural differences between the two time series regrading current time and time lag 1. All tests are at the significance level of 0.05, and cross spectral densities are estimated based on the rectangular window with bandwidth $M_n = \lfloor n^{\frac{1}{3.5}} \rfloor$. The result is shown in Table 2.9. For Site 1, 3 and 7, the null hypothesis is rejected in both tests so that the time series of the two periods are significantly different in term of univariate and bivariate distributions. For Site 2, the null hypotheses are not rejected indicating that the time series of the two periods do not have a significant difference regarding univariate and bivariate distributions. For Site 3, 6 and 8, the difference of the two time series are not observed in univariate distribution while it is detected in bivariate distribution. For Site 4, the result seems to be a bit problematic since there is a significant difference in the univariate case while the difference is insignificant in the bivariate case. We postulate that this is due to the fact that the correlation is taken into account and a different tilt function is used in the bivariate case.

Table 2.9: The result of the test for equality of distributions of the two time periods. Each entry indicates whether the corresponding null hypothesis is rejected or not at the significance level of 0.05.

Site	1	2	3	4	5	6	7	8
Univariate	Rejected	Not rejected	Not rejected	Rejected	Rejected	Not rejected	Rejected	Not rejected
Bivariate	Rejected	Not rejected	Rejected	Not rejected	Rejected	Rejected	Rejected	Rejected

In addition to testing the structural differences of the time series, we can also fuse them to obtain CIs for threshold probabilities. We demonstrate this with the bivariate case of Site 1. Again, we consider the bivariate sequence from 2020 to 2022 as the reference and estimate the bivariate CDF by fusing it with the bivariate sequence from 2010 to 2012. In Table 2.10, we can see the estimated reference threshold probabilities and the corresponding 95% CIs.

Table 2.10: Estimates of selected reference threshold probabilities and the corresponding 95% CIs.

Threshold probability	$\widetilde{P}$	95% CI
$P(X_{t-1}^{(0)} \leq 1, X_t^{(0)} > 1)$	0.0249	$(0.0168, 0.0329)$
$P(X_{t-1}^{(0)} \leq 0, X_t^{(0)} \leq 1)$	0.4755	$(0.4608, 0.4902)$
$P(X_{t-1}^{(0)} > 1, X_t^{(0)} \leq 2)$	0.0237	$(0.0162, 0.0311)$
$P(X_{t-1}^{(0)} > -1, X_t^{(0)} > 1)$	0.0164	$(0.0102, 0.0225)$
$P(X_{t-1}^{(0)} > 0, X_t^{(0)} \leq 0)$	0.2525	$(0.2426, 0.2624)$

Remark: The missing values are removed when we perform the analysis so that the strictly stationary condition does not hold. However, the univariate observations and bivariate pairs are still identically distributed so that the density ratio structure still holds. Moreover, the strong mixing coefficient of the sequence with missing values removed is bounded by that of the sequence without missing values. Another issue is that the cross spectral density estimates are also affected by the missing values. But since the proportion of missing values is small in this case, the issue is omitted here. Further investigation can be made to formulate a cross spectral density estimator that is resistant to such issue.

Chapter 3: Extension to Residual Coherence

3.1 Introduction

A multiple regression model with m covariates can be written as

$$Y_i = \beta_0 + \sum_{j=1}^{m} \beta_j X_{ji} + \varepsilon_i.$$

The parameters $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_m)^\top$ are estimated by the Ordinary Least Squares (OLS) method. When the response and covariates are time series, the regression model can be considered as

$$Y(t) = \beta_0 + \sum_{j=1}^{m} \beta_j X_j(t) + \varepsilon(t),$$

which resembles the model (11.3.1) in [27]. This model is also a special case of the Generalized Linear Model (GLM) for time series introduced by [33], and the parameters are estimated based on partial likelihood instead of OLS. A more complicated model is proposed by [27] such that

$$Y(t) = L_0 + \sum_{j=1}^{m} L_j(X_j(t)) + \varepsilon(t) \tag{3.1}$$

where L_0 is a constant and $L_1, \ldots, L_m$ are linear filters. In this case, the parameters have infinite dimension so that it is less viable to analyze the model in time domain. A frequency domain analysis of the model is provided by [27], in which multiple coherency spectrum and residual spectrum are introduced to measure the proportion of output predicted by the inputs and the noise, respectively.

A similar model is proposed by [44] to analyze a nonlinear system based on orthogonal projections of the output process. If $\{Y(t)\}$ is a quadratic system of $\{X(t)\}$ in general with $EY(t) = EX(t) = 0$, then for sufficiently large m, it can be approximated by

$$Y(t) = L(X(t)) + \sum_{j=1}^{m} \left(A_j(X(t)) + B_j(\widetilde{X}_j(t)) \right), \tag{3.2}$$

where A_j, B_j for $j = 1, \ldots, m$, are also linear filters, and $\{\widetilde{X}_j(t)\}$ is a lag process with lag j such that

$$\widetilde{X}_j(t) = X(t)X(t+j) - R_{XX}(j).$$

Moreover, $\{A_j(X(t)) + B_j(\widetilde{X}_j(t))\}$ is orthogonal to $\{L(X(t))\}$ for all j. (3.2) is indeed (3.1) with noise removed, first input being $\{X(t)\}$ and the rest inputs being lag processes. A follow up work [42] simplifies (3.2) to the case of one lag process with noise added to the system such that

$$Y(t) = L(X(t)) + \left(A(X(t)) + B(\widetilde{X}_j(t)) \right) + \varepsilon(t).$$

The concept lagged coherence is proposed based on this model to measure the effect of the lag process in the model. This concept is closely related to the squared partial coherency spectrum introduced by [27]. In [43] and [31], another criterion, maximum residual coherence, is proposed to measure the lag effect. Based on the orthogonal projection ideal in [44], [77] considered an orthogonal projection of $Y(t)$ in (3.1) with $L_0 = 0$ such that

$$
\begin{aligned}
Y(t) =& G_1(t) + \cdots + G_m(t) + \varepsilon(t),\\
G_1(t) =& L_{11}(X_1(t)),\\
G_2(t) =& L_{21}(X_1(t)) + L_{22}(X_2(t)),\\
&\vdots\\
G_m(t) =& L_{m1}(X_1(t)) + \cdots + L_{mm}(X_m(t)),
\end{aligned}
\tag{3.3}
$$

where $\{G_j(t)\}$'s are mutually orthogonal. This orthogonal decomposition has been extensively examined by [24]. Based on this orthogonal decomposition and [15], [24] developed statistical tests for the significance of input effects and applied them to the detection of the structural differences of brain functional connectivity.

In this chapter, we extend the notion of residual coherence based on the orthogonal processes in (3.3) to measure and test the effect of input processes following [77] and [24]. In Section 3.2, the discrete time version of (3.1) is analyzed based on the orthogonal decomposition in (3.3). Statistical tests for input effect are formulated based on the spectral representation of these orthogonal processes. In Section 3.3, we extend the notion of residual coherence in [43] and [31] to measure the input effect, and illustrate that nonlinear association between time series can be detected based on the tests developed in Section 3.2. In Section 3.4, we use simulation to show that the tests developed in Section 3.2 are satisfactory regarding the simulated Type-I error rates and powers. In Section 3.5, these tests are applied to blood oxygen level dependent (BOLD) sequences recorded from brain subregions to analyze the associations between different subregions.

3.2 A Multi-Input Single-Output Model

Without the loss of generality, we consider $L_0 = 0$ and all processes are zero-mean processes in (3.1). Then the model can be written as

$$Y(t) = \sum_{j=1}^{m} L_j(X_j(t)) + \varepsilon(t).$$

The discrete time version of this model can be expressed as

$$Y(t) = \sum_{j=1}^{m} \sum_{k_j=-\infty}^{\infty} b_j(k_j) X_j(t - k_j) + \varepsilon(t), \tag{3.4}$$

and we shall examine this model based on the orthogonal decomposition in (3.3).

3.2.1 Orthogonal Projection

The projection idea in (3.3) is indeed an analogy of the projection in regression. Consider a regression model without intercept

$$\boldsymbol{Y} = \sum_{j=1}^{m} \beta_j \boldsymbol{X}_j + \boldsymbol{\varepsilon}. \tag{3.5}$$

The OLS estimator $\widehat{\boldsymbol{Y}}$ is a summation of mutually orthogonal projections of $\boldsymbol{Y}$ such that

$$\widehat{\boldsymbol{Y}} = \boldsymbol{P}_1 \boldsymbol{Y} + \sum_{j=1}^{m-1} \left(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j\right) \boldsymbol{Y},$$

where $\boldsymbol{P}_1\boldsymbol{Y}, (\boldsymbol{P}_2-\boldsymbol{P}_1)\boldsymbol{Y}, \ldots, (\boldsymbol{P}_m-\boldsymbol{P}_{m-1})\boldsymbol{Y}$ are mutually orthogonal processes. $\boldsymbol{P}_1$ projects $\boldsymbol{Y}$ onto $\mathrm{Span}(\boldsymbol{X}_1)$, and $\boldsymbol{P}_j - \boldsymbol{P}_{j-1}$ projects $\boldsymbol{Y}$ onto $\mathrm{Span}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_j) \cap (\mathrm{Span}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{j-1}))^{\perp}$ for $j = 2, \ldots, m$.

Analogically, in (3.3), $G_1(t)$ can be considered as the projection of $Y(t)$ onto $\overline{\mathrm{Span}}(\{X_1(t)\})$, and $G_j(t)$ can be considered as the projection of $Y(t)$ onto $\overline{\mathrm{Span}}(\{X_1(t)\}, \ldots, \{X_j(t)\}) \cap (\overline{\mathrm{Span}}(\{X_1(t)\}, \ldots, \{X_{j-1}(t)\}))^{\perp}$ for $j = 2, \ldots, m$. Throughout this chapter, the norm used for the closure of span is L_2 norm. More assumptions are entailed to estimate and make statistical inferences on the orthogonal processes $\{G_j(t)\}$'s, and we will discuss these technical issues in the next section. For now, we shall discuss the tests formulated based on these orthogonal projections.

In the regression problem, one may test the effect of $\boldsymbol{X}_j$ based on the nested models

$$H_0 : \boldsymbol{Y} = \sum_{l=1}^{j-1} \beta_l \boldsymbol{X}_l + \boldsymbol{\varepsilon},$$
$$H_1 : \boldsymbol{Y} = \sum_{l=1}^{j-1} \beta_l \boldsymbol{X}_l + \beta_j \boldsymbol{X}_j + \boldsymbol{\varepsilon}.$$

If the variance of $\boldsymbol{Y}$ is known, a χ^2-test can be formulated based on the projection $(\boldsymbol{P}_j - \boldsymbol{P}_{j-1})\boldsymbol{Y}$. Moreover, if the goal is to examine the effect of $\boldsymbol{X}_j, \ldots, \boldsymbol{X}_m$, then we can consider the nested models

$$H_0 : \boldsymbol{Y} = \sum_{l=1}^{j-1} \beta_l \boldsymbol{X}_l + \boldsymbol{\varepsilon},$$
$$H_1 : \boldsymbol{Y} = \sum_{l=1}^{j-1} \beta_l \boldsymbol{X}_l + \sum_{h=j}^{m} \beta_h \boldsymbol{X}_h + \boldsymbol{\varepsilon}.$$

Again, a χ^2-test can be formulated based on the projection $(\boldsymbol{P}_m - \boldsymbol{P}_{j-1})\boldsymbol{Y}$.

Analogically, the effect of $\{X_j(t)\}$ can be tested based on

$$H_0 : Y(t) = \sum_{l=1}^{j-1} \sum_{k_l=-\infty}^{\infty} b_l(k_l) X_l(t-k_l) + \varepsilon(t),$$
$$H_1 : Y(t) = \sum_{l=1}^{j-1} \sum_{k_l=-\infty}^{\infty} b_l(k_l) X_l(t-k_l) + \sum_{k_j=-\infty}^{\infty} b_j(k_j) X_j(t-k_j) + \varepsilon(t).$$

The test should be formulated based on the estimated $G_j(t)$. Similarly, in order to test the effect $\{X_j(t)\}, \{X_{j+1}(t)\}, \ldots, \{X_m(t)\}$, based on

$$H_0 : Y(t) = \sum_{l=1}^{j-1} \sum_{k_l=-\infty}^{\infty} b_l(k_l) X_l(t-k_l) + \varepsilon(t),$$
$$H_1 : Y(t) = \sum_{l=1}^{j-1} \sum_{k_l=-\infty}^{\infty} b_l(k_l) X_l(t-k_l) + \sum_{h=j}^{m} \sum_{k_h=-\infty}^{\infty} b_h(k_h) X_h(t-k_h) + \varepsilon(t),$$

we can formulate a test based on the estimated $G_j(t), G_{j+1}(t), \ldots, G_m(t)$.

3.2.2 Spectral Representation of Orthogonal Processes

In this section, we rigorously examine the orthogonal processes $\{G_j(t)\}$'s. Let $\boldsymbol{F}_m(\lambda) = [f_{i,j}(\lambda)]_{i,j=1,\ldots,m}$ be the spectral matrix of the input processes $\{X_1(t)\}, \ldots, \{X_m(t)\}$, and $\boldsymbol{F}_{m|Y}(\lambda)$ be the spectral matrix of $\{X_1(t)\}, \ldots, \{X_m(t)\}, \{Y(t)\}$, such that

$$\boldsymbol{F}_{m|Y}(\lambda) = \begin{bmatrix} \boldsymbol{F}_m(\lambda) & \boldsymbol{f}_{m|Y}(\lambda) \\ \boldsymbol{f}_{m|Y}^H(\lambda) & f_{Y,Y}(\lambda) \end{bmatrix},$$

and $\boldsymbol{f}_{m|Y}(\lambda) = (f_{1,Y}(\lambda), \ldots, f_{m,Y}(\lambda))^\top$.

Assumption 3.1. *In (3.4),*

(1) $\{X_1(t)\}, \ldots, \{X_m(t)\}$ *are zero-mean jointly second order stationary processes with absolutely summable covariance functions.*

(2) $\sum_{k_j=-\infty}^{\infty} |b_j(k_j)| < \infty$ *for* $j = 1, \ldots, m$.

(3) $\{\varepsilon(t)\}$ *is a sequence of zero-mean IID noise that is independent of* $\{X_1(t)\}, \ldots, \{X_m(t)\}$. $E|\varepsilon(t)|^2 < \infty$.

(4) The spectral matrix $\boldsymbol{F}_{m|Y}(\lambda)$ *is non-singular and all elements are twice continuously differentiable* λ*-a.e.*

Under Assumption 3.1, $\{X_1(t)\}, \ldots, \{X_m(t)\}, \{Y(t)\}$ are joint second order stationary and $Y(t)$ admits the spectral representation

$$Y(t) = \sum_{j=1}^{m} \int_{-\pi}^{\pi} e^{it\lambda} B_j(\lambda) dZ_{X_j}(\lambda) + \int_{-\pi}^{\pi} e^{it\lambda} dZ_{\varepsilon}(\lambda),$$

where $B_j(\lambda) = \sum_{k_j=-\infty}^{\infty} e^{-ik_j\lambda} b_j(k_j)$ for $j = 1, \ldots, m$. Define

$$G_j(t) = \sum_{l=1}^{j} \int_{-\pi}^{\pi} e^{it\lambda} A_{j,l}(\lambda) dZ_{X_l}(\lambda)$$

for $j = 1, \ldots, m$, where $A_{j,l}(\lambda) = \sum_{h_{j,l}=-\infty}^{\infty} e^{-ih_{j,l}\lambda} a_{j,l}(h_{j,l})$.

Theorem 3.1. *Let*

$$\begin{aligned} \boldsymbol{f}_j(\lambda) =& - \left(f_{1,j}(\lambda), \ldots, f_{j-1,j}(\lambda)\right)^\top, \\ \boldsymbol{A}_j(\lambda) =& (A_{j,1}(\lambda), \ldots, A_{j,j-1}(\lambda))^\top, \end{aligned}$$

for $j = 2, \ldots, m$. *Let* $\boldsymbol{F}_{j|l}(\lambda)$ *be the matrix* $\boldsymbol{F}_j(\lambda)$ *with* l*th column replaced by* $\boldsymbol{f}_{j+1}(\lambda)$ *for*

$l=1,\dots,j$, and $\boldsymbol{F}_{j|j+1}(\lambda)=\boldsymbol{F}_j(\lambda)$. Suppose it holds for λ-a.e. that

$$\begin{aligned} A_{1,1}(\lambda) =& \frac{f_{Y,1}(\lambda)}{f_{1,1}(\lambda)}, \\ A_{j,j}(\lambda) =& \frac{\sum_{l=1}^{j} \det\left(\boldsymbol{F}_{j-1|l}(\lambda)\right) f_{Y,l}(\lambda)}{\det\left(\boldsymbol{F}_j(\lambda)\right)}, \\ \boldsymbol{A}_j(\lambda) =& \overline{\boldsymbol{F}_{j-1}^{-1}(\lambda)\boldsymbol{f}_j(\lambda)} A_{j,j}(\lambda), \end{aligned}$$

for $j=2,\dots,m$, then under Assumption 3.1, $G_1(t)$ is the projection of $Y(t)$ onto $\overline{\text{Span}}(\{X_1(t)\})$, and $G_j(t)$ is that onto $\overline{\text{Span}}(\{X_1(t)\},\dots,\{X_j(t)\}) \cap (\overline{\text{Span}}(\{X_1(t)\},\dots,\{X_{j-1}(t)\}))^{\perp}$ for $j=2,\dots,m$.

Proof. Under Assumption 3.1, for $j=1,\dots,m$ and $l=1,\dots,j$, $A_{j,l}(\lambda)$ is twice continuously differentiable λ-a.e. so that $\sum_{h_{j,l}=-\infty}^{\infty}|a_{j,l}(h_{j,l})|<\infty$ by Theorem 4.4 in [29]. Thus, $\sum_{h_{j,l}=-\infty}^{\infty}|a_{j,l}(h_{j,l})|^2<\infty$ and hence $G_j(t)\in\overline{\text{Span}}(\{X_1(t)\},\dots,\{X_j(t)\})$.

For $j\geq 2$ and $k=1,\dots,j$, $\forall\ h$,

$$EG_j(t+h)\overline{X_k(t)} = \int_{-\pi}^{\pi} e^{ih\lambda}\sum_{l=1}^{j} A_{j,l}(\lambda) f_{l,k}(\lambda) d\lambda.$$

$$\begin{aligned} \sum_{l=1}^{j} A_{j,l}(\lambda)\begin{bmatrix} f_{l,1}(\lambda) \\ \vdots \\ f_{l,j-1}(\lambda)\end{bmatrix} =& \sum_{l=1}^{j-1} A_{j,l}(\lambda)\begin{bmatrix} \overline{f_{1,l}(\lambda)} \\ \vdots \\ \overline{f_{j-1,l}(\lambda)}\end{bmatrix} + A_{j,j}(\lambda)\begin{bmatrix} f_{j,1}(\lambda) \\ \vdots \\ f_{j,j-1}(\lambda)\end{bmatrix} \\ =& \overline{\boldsymbol{F}_{j-1}(\lambda)}\boldsymbol{A}_j(\lambda) - A_{j,j}(\lambda)\overline{\boldsymbol{f}_j(\lambda)} \\ =& \boldsymbol{0}. \end{aligned}$$

Therefore, $\forall\ h$, $EG_j(t+h)\overline{X_k(t)}=0$ for $k=1,\dots,j-1$, which indicates that $G_j(t)\in$

$(\overline{\mathrm{Span}}(\{X_1(t)\},\ldots,\{X_{j-1}(t)\}))^{\perp}$.

It remains to show $B_l(\lambda)=\sum_{j=l}^{m}A_{j,l}(\lambda)$, λ-a.e. for $l=1,\ldots,m$. Since

$$G_m(t)\in(\overline{\mathrm{Span}}(\{X_1(t)\},\ldots,\{X_{m-1}(t)\}))^{\perp},$$

then $\forall\ h$,

$$EY(t+h)\overline{G_m(t)} = E\left(\int_{-\pi}^{\pi}e^{i(t+h)\lambda}B_m(\lambda)dZ_{X_m}(\lambda)\right)\overline{G_m(t)}$$
$$\int_{-\pi}^{\pi}e^{ih\lambda}\sum_{l=1}^{m}\overline{A_{m,l}(\lambda)}f_{Y,l}(\lambda)d\lambda=\int_{-\pi}^{\pi}e^{ih\lambda}B_m(\lambda)\sum_{l=1}^{m}\overline{A_{m,l}(\lambda)}f_{m,l}(\lambda)d\lambda,$$

so that

$$B_m(\lambda)=\frac{\sum_{l=1}^{m}\overline{A_{m,l}(\lambda)}f_{Y,l}(\lambda)}{\sum_{l=1}^{m}\overline{A_{m,l}(\lambda)}f_{m,l}(\lambda)},\quad \lambda\text{−a.e.}$$

Since $\boldsymbol{A}_j(\lambda)=\overline{\boldsymbol{F}_{j-1}^{-1}(\lambda)\boldsymbol{f}_j(\lambda)}A_{j,j}(\lambda)$, λ-a.e., by Cramer's rule,

$$\overline{A_{j,l}(\lambda)}=\frac{\det\left(\boldsymbol{F}_{j-1|l}(\lambda)\right)}{\det\left(\boldsymbol{F}_{j-1}(\lambda)\right)}\overline{A_{j,j}(\lambda)},\quad \lambda\text{−a.e.}$$

and hence

$$\begin{aligned}B_m(\lambda)=&\frac{\sum_{l=1}^{m}\det\left(\boldsymbol{F}_{m-1|l}(\lambda)\right)f_{Y,l}(\lambda)}{\sum_{l=1}^{m}\det\left(\boldsymbol{F}_{m-1|l}(\lambda)\right)f_{m,l}(\lambda)}\\ =&\frac{\sum_{l=1}^{m}\det\left(\boldsymbol{F}_{m-1|l}(\lambda)\right)f_{Y,l}(\lambda)}{\det\left(\boldsymbol{F}_{m}(\lambda)\right)}\\ =&A_{m,m}(\lambda),\quad \lambda\text{−a.e.}\end{aligned}$$

Note that the second equality is obtained by the fact that

$$\boldsymbol{F}_j(\lambda) = \begin{bmatrix} \boldsymbol{F}_{j-1}(\lambda) & -\boldsymbol{f}_j(\lambda) \\ -\boldsymbol{f}_j^H(\lambda) & f_{j,j}(\lambda) \end{bmatrix},$$

$$\begin{aligned} \det(\boldsymbol{F}_j(\lambda)) &= \det(\boldsymbol{F}_{j-1}(\lambda)) \det(f_{j,j}(\lambda) - \boldsymbol{f}_j^H(\lambda)\boldsymbol{F}_{j-1}^{-1}(\lambda)\boldsymbol{f}_j(\lambda)) \\ &= \det(\boldsymbol{F}_{j-1}(\lambda)) \left(f_{j,j}(\lambda) - \boldsymbol{f}_j^H(\lambda) \frac{\overline{\boldsymbol{A}_j(\lambda)}}{\overline{A_{j,j}(\lambda)}} \right) \\ &= \det(\boldsymbol{F}_{j-1}(\lambda)) \frac{\sum_{l=1}^{j} \det(\boldsymbol{F}_{j-1|l}(\lambda)) f_{j,l}(\lambda)}{\det(\boldsymbol{F}_{j-1}(\lambda))} \\ &= \sum_{l=1}^{j} \det(\boldsymbol{F}_{j-1|l}(\lambda)) f_{j,l}(\lambda), \end{aligned}$$

Based on the similar derivation, use

$$E(Y(t+h) - G_m(t+h))\overline{G_{m-1}(t)} = E\left(\int_{-\pi}^{\pi} e^{i(t+h)\lambda}(B_{m-1}(\lambda) - A_{m,m-1}(\lambda))\right)\overline{G_{m-1}(t)}$$

to obtain

$$B_{m-1}(\lambda) - A_{m,m-1}(\lambda) = A_{m-1,m-1}(\lambda), \quad \lambda\text{−a.e.}$$

Repeat this procedure by taking the expectation

$$E\left(Y(t+h) - \sum_{k=0}^{j} G_{m-k}(t+h)\right)\overline{G_{m-j-1}(t)}$$

for $j = 1, \ldots, m-2$, we obtain $B_l(\lambda) = \sum_{j=l}^{m} A_{j,l}(\lambda)$, λ-a.e. for $l = 1, \ldots, m$. $\square$

Remark: The existence and uniqueness of these projections are guaranteed by Theorem 2.3.1 in [5] because $\overline{\text{Span}}(\{X_1(t)\})$ and $\overline{\text{Span}}(\{X_1(t)\}, \ldots, \{X_j(t)\}) \cap (\overline{\text{Span}}(\{X_1(t)\}, \ldots, \{X_{j-1}(t)\}))^{\perp}$

for $j = 2, \ldots, m$, are closed subspaces of the L_2 space generated by $\{X_1(t)\}, \ldots, \{X_m(t)\}$. It seems to be more natural to first assume the projection $G_j(t) = \sum_{l=1}^{j} \sum_{h_{j,l}=-\infty}^{\infty} a_{j,l}(h_{j,l}) X_l(t - h_{j,l})$ and then find their spectral representation. However, here we prove this backwards, that is, we start with a spectral representation and then prove that it is a projection. That is because the closure of the span is taken under L_2 norm so that it only ensures the square summability $\sum_{h_{j,l}=-\infty}^{\infty} |a_{j,l}(h_{j,l})|^2 < \infty$, which is not sufficient for $G_j(t)$ to admit a spectral representation.

By Theorem 3.1, the spectral representations of the orthogonal projections $G_j(t)$'s are obtained. Subsequently,

$$f_{G_1G_1}(\lambda) = \frac{|f_{Y,1}(\lambda)|^2}{f_{1,1}(\lambda)},$$

and the spectral density of $G_j(t)$ for $j = 2, \ldots, m$, can be obtained by

$$\begin{aligned}
EG_j(t+h)\overline{G_j(t)} =& EG_j(t+h)\left(\int_{-\pi}^{\pi} e^{-it\lambda}\overline{A_{j,j}(\lambda)}d\overline{Z_{X_j}(\lambda)}\right) \\
\int_{-\pi}^{\pi} e^{ih\lambda} f_{G_jG_j}(\lambda)d\lambda =& \int_{-\pi}^{\pi} e^{ih\lambda}\overline{A_{j,j}(\lambda)}\sum_{l=1}^{j} A_{j,l}(\lambda) f_{l,j}(\lambda)d\lambda \\
\int_{-\pi}^{\pi} e^{ih\lambda} f_{G_jG_j}(\lambda)d\lambda =& \int_{-\pi}^{\pi} e^{ih\lambda}|A_{j,j}(\lambda)|^2 \frac{\sum_{l=1}^{j}\overline{\det(\boldsymbol{F}_{j-1|l}(\lambda)) f_{j,l}(\lambda)}}{\det(\boldsymbol{F}_{j-1}(\lambda))}d\lambda \\
\int_{-\pi}^{\pi} e^{ih\lambda} f_{G_jG_j}(\lambda)d\lambda =& \int_{-\pi}^{\pi} e^{ih\lambda}|A_{j,j}(\lambda)|^2 \frac{\det(\boldsymbol{F}_j(\lambda))}{\det(\boldsymbol{F}_{j-1}(\lambda))}d\lambda,
\end{aligned}$$

$$f_{G_jG_j}(\lambda) = \frac{\left|\sum_{l=1}^{j}\det(\boldsymbol{F}_{j-1|l}(\lambda)) f_{Y,l}(\lambda)\right|^2}{\det(\boldsymbol{F}_{j-1}(\lambda))\det(\boldsymbol{F}_j(\lambda))} \quad \lambda\text{−a.e.}$$

Let $\boldsymbol{Z}_{m+1} = [z_{i,j}]_{i,j=1,\ldots,m+1}$, $\boldsymbol{Z}_{j|j+1} = \boldsymbol{Z}_j$, and $\boldsymbol{Z}_{j|l}$ be the matrix $\boldsymbol{Z}_j$ with lth column replaced by $-(z_{1,j+1}, \ldots, z_{j,j+1})^\top$ for $l = 1, \ldots, j$, $j = 1, \ldots, m$. Denote

$$\Psi_{1|m}(\boldsymbol{Z}_{m+1}) = z_{m+1,1}, \quad \Psi_{j|m}(\boldsymbol{Z}_{m+1}) = \sum_{l=1}^{j}\det(\boldsymbol{Z}_{j-1|l}) z_{m+1,l},$$

for $j = 2, \ldots, m$. Thus,

$$f_{G_1G_1}(\lambda) = \frac{|\Psi_{1|m}(\boldsymbol{F}_{m|Y}(\lambda))|^2}{\det(\boldsymbol{F}_1(\lambda))}, \quad f_{G_jG_j}(\lambda) = \frac{\left|\Psi_{j|m}(\boldsymbol{F}_{m|Y}(\lambda))\right|^2}{\det(\boldsymbol{F}_{j-1}(\lambda))\det(\boldsymbol{F}_j(\lambda))}, \quad \lambda\text{−a.e.} \tag{3.6}$$

for $j = 2, \ldots, m$. Note that

$$\begin{aligned}
\Psi_{m|m}(\boldsymbol{F}_{m|Y}(\lambda)) &= \sum_{l=1}^{m} \det(\boldsymbol{F}_{m-1|l}(\lambda)) f_{Y,l}(\lambda) \\
&= (\det(\boldsymbol{F}_{m-1}(\lambda))) \left(f_{Y,m}(\lambda) + \sum_{l=1}^{m-1} \frac{\det(\boldsymbol{F}_{m-1|l}(\lambda))}{\det(\boldsymbol{F}_{m-1}(\lambda))} f_{Y,l}(\lambda) \right) \\
&= (\det(\boldsymbol{F}_{m-1}(\lambda))) \left(f_{Y,m}(\lambda) + \boldsymbol{f}^H_{m-1|Y}(\lambda) \frac{\overline{\boldsymbol{A}_m(\lambda)}}{\overline{A_{m,m}(\lambda)}} \right) \\
&= \det(\boldsymbol{F}_{m-1}(\lambda))(f_{Y,m}(\lambda) + \boldsymbol{f}^H_{m-1|Y}(\lambda)\boldsymbol{F}^{-1}_{m-1}(\lambda)\boldsymbol{f}_m(\lambda)).
\end{aligned} \tag{3.7}$$

3.2.3 Testing for Input Effect

Recall from Section 3.2.1 that we may investigate the input effect of $\{X_j(t)\}$ ($j \leq m$) by testing

$$\begin{aligned}
H_0 : Y(t) &= \sum_{l=1}^{j-1} \sum_{k_l=-\infty}^{\infty} b_l(k_l) X_l(t - k_l) + \varepsilon(t), \\
H_1 : Y(t) &= \sum_{l=1}^{j-1} \sum_{k_l=-\infty}^{\infty} b_l(k_l) X_l(t - k_l) + \sum_{k_j=-\infty}^{\infty} b_j(k_j) X_j(t - k_j) + \varepsilon(t).
\end{aligned}$$

By Theorem 3.1, this is equivalent to testing whether $G_j(t) = 0$. Thus, based on (3.6), we can formulate the hypotheses as

$$H_0 : \int_{-\pi}^{\pi} |\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda))|^2 d\lambda = 0, \quad H_1 : \int_{-\pi}^{\pi} |\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda))|^2 d\lambda > 0. \tag{3.8}$$

In order to construct a test statistic, stronger assumptions are made for (3.4) based on [15] and [24].

Assumption 3.2. *In (3.4),*

(1) $\{X_1(t)\}, \ldots, \{X_m(t)\}$ *are zero-mean jointly stationary processes for all orders. Moreover,* $\forall\, j \in \mathbb{Z}^+$, $\forall\, u_0, \ldots, u_j \in \{1, \ldots, m\}$

$$\sum_{t_1,\ldots,t_j=-\infty}^{\infty} \left(1 + \sum_{i=1}^{j} t_i^2\right) \left|\mathrm{cum}\left(X_{u_0}(0), X_{u_1}(t_1), \ldots, X_{u_j}(t_j)\right)\right| < \infty.$$

(2) $\sum_{k_j=-\infty}^{\infty} k_j^2 |b_j(k_j)| < \infty$ *for* $j = 1, \ldots, m$.

(3) $\{\varepsilon(t)\}$ *is a sequence of zero-mean IID noise that is independent of* $\{X_1(t)\}, \ldots, \{X_m(t)\}$. *All moments of* $\varepsilon(t)$ *are finite.*

(4) The spectral matrix $\boldsymbol{F}_{m|Y}(\lambda)$ *is non-singular* λ*-a.e.*

A sufficient condition for Assumption 3.2 (1) can be referred to Lemma 2.1 in [24]. The assumption of the cumulants can be replaced by strong mixing condition. If $\boldsymbol{X}(t) = (X_1(t), \ldots, X_m(t))^\top$ is strong mixing and the mixing coefficient has a exponential decay, then $\forall\, j \in \mathbb{Z}^+$, $\forall\, u_0, \ldots, u_j \in \{1, \ldots, m\}$,

$$\sum_{t_1,\ldots,t_j=-\infty}^{\infty} \left(1 + \sum_{i=1}^{j} t_i^d\right) \left|\mathrm{cum}\left(X_{u_0}(0), X_{u_1}(t_1), \ldots, X_{u_j}(t_j)\right)\right| < \infty$$

for all $d \in \mathbb{N}$. By [49], ARMA series satisfy this mixing condition so that our analysis can be applied to a wide range of time series. According to [24], Assumption 3.2 (1) and (2) imply that (1) holds for $u_0, \ldots, u_j \in \{0, 1, \ldots, m\}$ with $Y(t) = X_0(t)$. Additionally, Assumption 3.2 (1) ensures that the elements of $\boldsymbol{F}_{m|Y}(\lambda)$ are twice continuously differentiable λ-a.e.

Next, we define the estimator of cross spectral density, which largely follows [24]. The cross spectral density estimator of $\{U(t)\}$ and $\{V(t)\}$ is given by

$$
\begin{aligned}
\widehat{f}_{UV}(\lambda) =& \frac{1}{2\pi} \sum_{|h|<n} w\left(\frac{h}{M_n}\right) \widehat{R}_{UV}(h) e^{-ih\lambda}, \\
\widehat{R}_{UV}(h) =& \frac{1}{n-h} \sum_{t=1}^{n-h} (U(t+h) - \overline{U})(V(t) - \overline{V}), \\
\overline{U} =& \frac{1}{n} \sum_{t=1}^{n} U(t), \\
\overline{V} =& \frac{1}{n} \sum_{t=1}^{n} V(t).
\end{aligned}
$$

The assumptions on the window function w and bandwidth M_n refer to Assumption 3.1 in [24].

Assumption 3.3. *The window function w and bandwidth M_n satisfies*

(1) $nM_n^{-\frac{9}{2}} \to 0$ *and* $nM_n^{-3} \to \infty$ *as* $n \to \infty$.

(2) $C_{w,2} = \int_{-\infty}^{\infty} w^2(x)dx < \infty$, $C_{w,4} = \int_{-\infty}^{\infty} w^4(x)dx < \infty$ *and* $\sum_{|h|<n} w\left(\frac{h}{M_n}\right) \sim O(M_n)$.

(3) $W(\lambda) = \frac{1}{2\pi} \sum_{|h|<n} w\left(\frac{h}{M_n}\right) e^{-ih\lambda}$ *is bounded, symmetric, nonnegative and Lipschitz continuous satisfying* $\int_{-\infty}^{\infty} W(\lambda)d\lambda = 1$, $\int_{-\infty}^{\infty} \lambda^2 W(\lambda)d\lambda < \infty$ *and* $\limsup_{\lambda\to\infty} \lambda^2 W(\lambda) = 0$.

According to [24], under Assumption 3.2 and 3.3, the statistic

$$
Z_{n,j|j} = \frac{T_{n,j|j} - \mu_{n,j|j}}{\sigma_{n,j|j}} \xrightarrow{d} N(0,1),
$$

as $n \to \infty$ under H_0 in (3.8), where

$$
\begin{aligned}
T_{n,j|j} =& n M_n^{-\frac{1}{2}} \int_{-\pi}^{\pi} |\Psi_{j|j}(\widehat{\boldsymbol{F}}_{j|Y}(\lambda))|^2 d\lambda, \\
\mu_{n,j|j} =& M_n^{\frac{1}{2}} C_{w,2} \int_{-\pi}^{\pi} \operatorname{tr}\left(D^{\top}_{\Psi_{j|j}}(\widehat{\boldsymbol{F}}_{j|Y}(\lambda)) \widehat{\boldsymbol{F}}_{j|Y}(\lambda) \overline{D_{\Psi_{j|j}}(\widehat{\boldsymbol{F}}_{j|Y}(\lambda))} \widehat{\boldsymbol{F}}_{j|Y}(\lambda) \right) d\lambda, \\
\sigma^2_{n,j|j} =& 4\pi C_{w,4} \int_{-\pi}^{\pi} \left(\left| \operatorname{tr}\left(D^{\top}_{\Psi_{j|j}}(\widehat{\boldsymbol{F}}_{j|Y}(\lambda)) \widehat{\boldsymbol{F}}_{j|Y}(\lambda) \overline{D_{\Psi_{j|j}}(\widehat{\boldsymbol{F}}_{j|Y}(\lambda))} \widehat{\boldsymbol{F}}_{j|Y}(\lambda) \right) \right|^2 \right. \\
& \left. + \left| \operatorname{tr}\left(D^{\top}_{\Psi_{j|j}}(\widehat{\boldsymbol{F}}_{j|Y}(\lambda)) \widehat{\boldsymbol{F}}_{j|Y}(\lambda) D^{\top}_{\Psi_{j|j}}(\widehat{\boldsymbol{F}}_{j|Y}(\lambda)) \widehat{\boldsymbol{F}}_{j|Y}(\lambda) \right) \right|^2 \right) d\lambda, \\
D_{\Psi_{j|j}}(\widehat{\boldsymbol{F}}_{j|Y}(\lambda)) =& \left. \frac{\partial \Psi_{j|j}(\boldsymbol{Z}_{j+1})}{\partial \boldsymbol{Z}_{j+1}} \right|_{\boldsymbol{Z}_{j+1} = \widehat{\boldsymbol{F}}_{j|Y}(\lambda)},
\end{aligned}
$$

and we reject H_0 if $Z_{n,j|m} > z_\alpha$ at the significance level of α.

Theorem 3.2. *Under H_0 in (3.8),*

$$
\begin{aligned}
\operatorname{tr}\left(D^{\top}_{\Psi_{1|1}}(\boldsymbol{F}_{1|Y}(\lambda)) \boldsymbol{F}_{1|Y}(\lambda) \overline{D_{\Psi_{1|1}}(\boldsymbol{F}_{1|Y}(\lambda))} \boldsymbol{F}_{1|Y}(\lambda) \right) =& f_{1,1}(\lambda) f_{Y,Y}(\lambda), \\
\operatorname{tr}\left(D^{\top}_{\Psi_{1|1}}(\boldsymbol{F}_{1|Y}(\lambda)) \boldsymbol{F}_{1|Y}(\lambda) D^{\top}_{\Psi_{1|1}}(\boldsymbol{F}_{1|Y}(\lambda)) \boldsymbol{F}_{1|Y}(\lambda) \right) =& 0,
\end{aligned}
$$

and

$$
\begin{aligned}
& \operatorname{tr}\left(D^{\top}_{\Psi_{j|j}}(\boldsymbol{F}_{j|Y}(\lambda)) \boldsymbol{F}_{j|Y}(\lambda) \overline{D_{\Psi_{j|j}}(\boldsymbol{F}_{j|Y}(\lambda))} \boldsymbol{F}_{j|Y}(\lambda) \right) \\
=& \det\left(\boldsymbol{F}_{j-1}(\lambda)\right) \det\left(\boldsymbol{F}_{j}(\lambda)\right) (f_{Y,Y}(\lambda) - \boldsymbol{f}^H_{j-1|Y}(\lambda) \boldsymbol{F}^{-1}_{j-1}(\lambda) \boldsymbol{f}_{j-1|Y}(\lambda)), \\
& \operatorname{tr}\left(D^{\top}_{\Psi_{j|j}}(\boldsymbol{F}_{j|Y}(\lambda)) \boldsymbol{F}_{j|Y}(\lambda) D^{\top}_{\Psi_{j|j}}(\boldsymbol{F}_{j|Y}(\lambda)) \boldsymbol{F}_{j|Y}(\lambda) \right) \\
=& 0,
\end{aligned}
$$

for $j \geq 2$.

Proof. When $j = 1$,

$$D^{\top}_{\Psi_{1|1}}(\boldsymbol{F}_{1|Y}(\lambda))\boldsymbol{F}_{1|Y}(\lambda) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} f_{1,1}(\lambda) & f_{1,Y}(\lambda) \\ f_{Y,1}(\lambda) & f_{Y,Y}(\lambda) \end{bmatrix} = \begin{bmatrix} f_{Y,1}(\lambda) & f_{Y,Y}(\lambda) \\ 0 & 0 \end{bmatrix},$$

$$\overline{D_{\Psi_{1|1}}(\boldsymbol{F}_{1|Y}(\lambda))}\boldsymbol{F}_{1|Y}(\lambda) = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} f_{1,1}(\lambda) & f_{1,Y}(\lambda) \\ f_{Y,1}(\lambda) & f_{Y,Y}(\lambda) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ f_{1,1}(\lambda) & f_{1,Y}(\lambda) \end{bmatrix},$$

so that

$$\begin{aligned}
\operatorname{tr}\left(D^{\top}_{\Psi_{1|1}}(\boldsymbol{F}_{1|Y}(\lambda))\boldsymbol{F}_{1|Y}(\lambda)\overline{D_{\Psi_{1|1}}(\boldsymbol{F}_{1|Y}(\lambda))}\boldsymbol{F}_{1|Y}(\lambda)\right) &= f_{1,1}(\lambda)f_{Y,Y}(\lambda), \\
\operatorname{tr}\left(D^{\top}_{\Psi_{1|1}}(\boldsymbol{F}_{1|Y}(\lambda))\boldsymbol{F}_{1|Y}(\lambda)D^{\top}_{\Psi_{1|1}}(\boldsymbol{F}_{1|Y}(\lambda))\boldsymbol{F}_{1|Y}(\lambda)\right) &= |f_{Y,1}(\lambda)|^2 = |\Psi_{1|1}(\boldsymbol{F}_{1|Y}(\lambda))|^2 = 0,
\end{aligned}$$

under H_0. When $j \geq 2$,

$$\boldsymbol{F}_{j|Y}(\lambda) = \begin{bmatrix} \boldsymbol{F}_{j-1}(\lambda) & -\boldsymbol{f}_j(\lambda) & \boldsymbol{f}_{j-1|Y}(\lambda) \\ -\boldsymbol{f}_j^H(\lambda) & f_{j,j}(\lambda) & f_{j,Y}(\lambda) \\ \boldsymbol{f}^H_{j-1|Y}(\lambda) & f_{Y,j}(\lambda) & f_{Y,Y}(\lambda) \end{bmatrix},$$

$$D^{\top}_{\Psi_{j|j}}(\boldsymbol{F}_{j|Y}(\lambda)) = \begin{bmatrix} \boldsymbol{D}^{\top}_{j-1}(\lambda) & \boldsymbol{0} & (\det(\boldsymbol{F}_{j-1}(\lambda)))\boldsymbol{F}^{-1}_{j-1}(\lambda)\boldsymbol{f}_j(\lambda) \\ -(\det(\boldsymbol{F}_{j-1}(\lambda)))\boldsymbol{f}^H_{j-1|Y}(\lambda)\boldsymbol{F}^{-1}_{j-1}(\lambda) & 0 & \det(\boldsymbol{F}_{j-1}(\lambda)) \\ \boldsymbol{0} & 0 & 0 \end{bmatrix},$$

$$\overline{D_{\Psi_{j|j}}(\boldsymbol{F}_{j|Y}(\lambda))} = \begin{bmatrix} \overline{\boldsymbol{D}_{j-1}(\lambda)} & -(\det(\boldsymbol{F}_{j-1}(\lambda)))\boldsymbol{F}^{-1}_{j-1}(\lambda)\boldsymbol{f}_{j-1|Y}(\lambda) & \boldsymbol{0} \\ \boldsymbol{0} & 0 & 0 \\ (\det(\boldsymbol{F}_{j-1}(\lambda)))\boldsymbol{f}^H_j(\lambda)\boldsymbol{F}^{-1}_{j-1}(\lambda) & \det(\boldsymbol{F}_{j-1}(\lambda)) & 0 \end{bmatrix},$$

where

$$\begin{aligned}\boldsymbol{D}_{j-1}^{\top}(\lambda) =&(\det(\boldsymbol{F}_{j-1}(\lambda)))(f_{Y,j}(\lambda)+\boldsymbol{f}_{j-1|Y}^{H}(\lambda)\boldsymbol{F}_{j-1}^{-1}(\lambda)\boldsymbol{f}_{j}(\lambda))\boldsymbol{F}_{j-1}^{-1}(\lambda)\\ &-(\det(\boldsymbol{F}_{j-1}(\lambda)))\boldsymbol{F}_{j-1}^{-1}(\lambda)\boldsymbol{f}_{j}(\lambda)\boldsymbol{f}_{j-1|Y}^{H}(\lambda)\boldsymbol{F}_{j-1}^{-1}(\lambda),\\ \overline{\boldsymbol{D}_{j-1}(\lambda)} =&(\det(\boldsymbol{F}_{j-1}(\lambda)))\overline{(f_{Y,j}(\lambda)+\boldsymbol{f}_{j-1|Y}^{H}(\lambda)\boldsymbol{F}_{j-1}^{-1}(\lambda)\boldsymbol{f}_{j}(\lambda))}\boldsymbol{F}_{j-1}^{-1}(\lambda)\\ &-(\det(\boldsymbol{F}_{j-1}(\lambda)))\boldsymbol{F}_{j-1}^{-1}(\lambda)\boldsymbol{f}_{j-1|Y}(\lambda)\boldsymbol{f}_{j}^{H}(\lambda)\boldsymbol{F}_{j-1}^{-1}(\lambda).\end{aligned}$$

Then,

$$D_{\Psi_{j|j}}^{\top}(\boldsymbol{F}_{j|Y}(\lambda))\boldsymbol{F}_{j|Y}(\lambda)=\begin{bmatrix}\boldsymbol{A}_1(\lambda) & \boldsymbol{A}_2(\lambda)\\ \boldsymbol{A}_3(\lambda) & \boldsymbol{A}_4(\lambda)\end{bmatrix},$$

where

$$\begin{aligned}\boldsymbol{A}_1(\lambda) =&\boldsymbol{D}_{j-1}^{\top}(\lambda)\boldsymbol{F}_{j-1}(\lambda)+\begin{bmatrix}\boldsymbol{0} & (\det(\boldsymbol{F}_{j-1}(\lambda)))\boldsymbol{F}_{j-1}^{-1}(\lambda)\boldsymbol{f}_{j}(\lambda)\end{bmatrix}\begin{bmatrix}-\boldsymbol{f}_{j}^{H}(\lambda)\\ \boldsymbol{f}_{j-1|Y}^{H}(\lambda)\end{bmatrix}\\ =&(\det\boldsymbol{F}_{j-1}(\lambda))(f_{Y,j}(\lambda)+\boldsymbol{f}_{j-1|Y}^{H}(\lambda)\boldsymbol{F}_{j-1}^{-1}(\lambda)\boldsymbol{f}_{j}(\lambda))\boldsymbol{I}_{j-1}\\ =&\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda))\boldsymbol{I}_{j-1},\\ \boldsymbol{A}_2(\lambda) =&\boldsymbol{D}_{j-1}^{\top}(\lambda)\begin{bmatrix}-\boldsymbol{f}_{j}(\lambda) & \boldsymbol{f}_{j-1|Y}(\lambda)\end{bmatrix}\\ &+\begin{bmatrix}\boldsymbol{0} & (\det(\boldsymbol{F}_{j-1}(\lambda)))\boldsymbol{F}_{j-1}^{-1}(\lambda)\boldsymbol{f}_{j}(\lambda)\end{bmatrix}\begin{bmatrix}f_{j,j}(\lambda) & f_{j,Y}(\lambda)\\ f_{Y,j}(\lambda) & f_{Y,Y}(\lambda)\end{bmatrix},\end{aligned}$$

$$\boldsymbol{A}_3(\lambda) = \begin{bmatrix} -(\det(\boldsymbol{F}_{j-1}(\lambda)))\boldsymbol{f}^H_{j-1|Y}(\lambda)\boldsymbol{F}^{-1}_{j-1}(\lambda) \\ \boldsymbol{0} \end{bmatrix} \boldsymbol{F}_{j-1}(\lambda) + \begin{bmatrix} 0 & \det(\boldsymbol{F}_{j-1}(\lambda)) \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -\boldsymbol{f}^H_j \\ \boldsymbol{f}^H_{j-1|Y} \end{bmatrix}$$

$$=\boldsymbol{0},$$

$$\boldsymbol{A}_4(\lambda) = \begin{bmatrix} -(\det(\boldsymbol{F}_{j-1}(\lambda)))\boldsymbol{f}^H_{j-1|Y}(\lambda)\boldsymbol{F}^{-1}_{j-1}(\lambda) \\ \boldsymbol{0} \end{bmatrix} \begin{bmatrix} -\boldsymbol{f}_j(\lambda) & \boldsymbol{f}_{j-1|Y}(\lambda) \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & \det(\boldsymbol{F}_{j-1}(\lambda)) \\ 0 & 0 \end{bmatrix} \begin{bmatrix} f_{j,j}(\lambda) & f_{j,Y}(\lambda) \\ f_{Y,j}(\lambda) & f_{Y,Y}(\lambda) \end{bmatrix}$$

$$= \begin{bmatrix} \Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda)) & (\det(\boldsymbol{F}_{j-1}(\lambda)))(f_{Y,Y}(\lambda) - \boldsymbol{f}^H_{j-1|Y}(\lambda)\boldsymbol{F}^{-1}_{j-1}(\lambda)\boldsymbol{f}_{j-1|Y}(\lambda)) \\ 0 & 0 \end{bmatrix}.$$

$$\overline{D_{\Psi_{j|j}}(\boldsymbol{F}_{j|Y}(\lambda))}\boldsymbol{F}_{j|Y}(\lambda) = \begin{bmatrix} \boldsymbol{B}_1(\lambda) & \boldsymbol{B}_2(\lambda) \\ \boldsymbol{B}_3(\lambda) & \boldsymbol{B}_4(\lambda) \end{bmatrix},$$

where

$$\boldsymbol{B}_1(\lambda) = \overline{\boldsymbol{D}_{j-1}(\lambda)}\boldsymbol{F}_{j-1}(\lambda) + \begin{bmatrix} -(\det(\boldsymbol{F}_{j-1}(\lambda)))\boldsymbol{F}^{-1}_{j-1}(\lambda)\boldsymbol{f}_{j-1|Y}(\lambda) & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} -\boldsymbol{f}^H_j(\lambda) \\ \boldsymbol{f}^H_{j-1|Y}(\lambda) \end{bmatrix}$$

$$=\overline{\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda))}\boldsymbol{I}_{j-1},$$

$$\boldsymbol{B}_2(\lambda) = \overline{\boldsymbol{D}_{j-1}(\lambda)} \begin{bmatrix} -\boldsymbol{f}_j(\lambda) & \boldsymbol{f}_{j-1|Y}(\lambda) \end{bmatrix}$$

$$+ \begin{bmatrix} -(\det(\boldsymbol{F}_{j-1}(\lambda)))\boldsymbol{F}^{-1}_{j-1}(\lambda)\boldsymbol{f}_{j-1|Y}(\lambda) & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} f_{j,j}(\lambda) & f_{j,Y}(\lambda) \\ f_{Y,j}(\lambda) & f_{Y,Y}(\lambda) \end{bmatrix},$$

$$
\begin{aligned}
\boldsymbol{B}_3(\lambda) &= \begin{bmatrix} \mathbf{0} \\ (\det(\boldsymbol{F}_{j-1}(\lambda)))\boldsymbol{f}_j^H(\lambda)\boldsymbol{F}_{j-1}^{-1}(\lambda) \end{bmatrix} \boldsymbol{F}_{j-1}(\lambda) + \begin{bmatrix} 0 & 0 \\ \det(\boldsymbol{F}_{j-1}(\lambda)) & 0 \end{bmatrix} \begin{bmatrix} -\boldsymbol{f}_j^H(\lambda) \\ \boldsymbol{f}_{j-1|Y}^H(\lambda) \end{bmatrix} \\
&= \mathbf{0}, \\
\boldsymbol{B}_4(\lambda) &= \begin{bmatrix} \mathbf{0} \\ (\det(\boldsymbol{F}_{j-1}(\lambda)))\boldsymbol{f}_j^H(\lambda)\boldsymbol{F}_{j-1}^{-1}(\lambda) \end{bmatrix} \begin{bmatrix} -\boldsymbol{f}_j(\lambda) & \boldsymbol{f}_{j-1|Y}(\lambda) \end{bmatrix} \\
&\quad + \begin{bmatrix} 0 & 0 \\ \det(\boldsymbol{F}_{j-1}(\lambda)) & 0 \end{bmatrix} \begin{bmatrix} f_{j,j}(\lambda) & f_{j,Y}(\lambda) \\ f_{Y,j}(\lambda) & f_{Y,Y}(\lambda) \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 \\ \det(\boldsymbol{F}_j(\lambda)) & \overline{\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda))} \end{bmatrix}.
\end{aligned}
$$

Thus, under H_0,

$$
\begin{aligned}
&\operatorname{tr}\left(D^\top_{\Psi_{j|j}}(\boldsymbol{F}_{j|Y}(\lambda))\boldsymbol{F}_{j|Y}(\lambda)\overline{D_{\Psi_{j|j}}(\boldsymbol{F}_{j|Y}(\lambda))}\boldsymbol{F}_{j|Y}(\lambda)\right) \\
=&\operatorname{tr}(\boldsymbol{A}_1(\lambda)\boldsymbol{B}_1(\lambda)) + \operatorname{tr}(\boldsymbol{A}_4(\lambda)\boldsymbol{B}_4(\lambda)) \\
=&(j-1)|\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda))|^2 + \det(\boldsymbol{F}_{j-1}(\lambda))\det(\boldsymbol{F}_j(\lambda))(f_{Y,Y}(\lambda) - \boldsymbol{f}_{j-1|Y}^H(\lambda)\boldsymbol{F}_{j-1}^{-1}(\lambda)\boldsymbol{f}_{j-1|Y}(\lambda)) \\
=&\det(\boldsymbol{F}_{j-1}(\lambda))\det(\boldsymbol{F}_j(\lambda))(f_{Y,Y}(\lambda) - \boldsymbol{f}_{j-1|Y}^H(\lambda)\boldsymbol{F}_{j-1}^{-1}(\lambda)\boldsymbol{f}_{j-1|Y}(\lambda)), \\
&\operatorname{tr}\left(D^\top_{\Psi_{j|j}}(\boldsymbol{F}_{j|Y}(\lambda))\boldsymbol{F}_{j|Y}(\lambda)D^\top_{\Psi_{j|j}}(\boldsymbol{F}_{j|Y}(\lambda))\boldsymbol{F}_{j|Y}(\lambda)\right) \\
=&\operatorname{tr}(\boldsymbol{A}_1(\lambda)\boldsymbol{A}_1(\lambda)) + \operatorname{tr}(\boldsymbol{A}_4(\lambda)\boldsymbol{A}_4(\lambda)) \\
=&j\Psi^2_{j|j}(\boldsymbol{F}_{j|Y}(\lambda)) \\
=&0.
\end{aligned}
$$

□

By Theorem 3.2, $\mu_{n,j|j}$ and $\sigma^2_{n,j|j}$ can be simplified to

$$\begin{aligned}\mu_{n,j|j} =& M_n^{\frac{1}{2}} C_{w,2} \int_{-\pi}^{\pi} \psi_{j|j}(\widehat{\boldsymbol{F}}_{j|Y}(\lambda)) d\lambda,\\ \sigma^2_{n,j|j} =& 4\pi C_{w,4} \int_{-\pi}^{\pi} |\psi_{j|j}(\widehat{\boldsymbol{F}}_{j|Y}(\lambda))|^2 d\lambda,\end{aligned}$$

where

$$\psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda)) = \det(\boldsymbol{F}_{j-1}(\lambda)) \det(\boldsymbol{F}_j(\lambda))(f_{Y,Y}(\lambda) - \boldsymbol{f}^H_{j-1|Y}(\lambda)\boldsymbol{F}^{-1}_{j-1}(\lambda)\boldsymbol{f}_{j-1|Y}(\lambda)).$$

Now, if we want to test the effect of multiple inputs $\{X_j(t)\}, \{X_{j+1}(t)\}, \ldots, \{X_m(t)\}$, according to Section 3.2.1, we consider the nested models

$$\begin{aligned}H_0 : Y(t) &= \sum_{l=1}^{j-1} \sum_{k_l=-\infty}^{\infty} b_l(k_l) X_l(t-k_l) + \varepsilon(t),\\ H_1 : Y(t) &= \sum_{l=1}^{j-1} \sum_{k_l=-\infty}^{\infty} b_l(k_l) X_l(t-k_l) + \sum_{h=j}^{m} \sum_{k_h=-\infty}^{\infty} b_h(k_h) X_h(t-k_h) + \varepsilon(t).\end{aligned}$$

This is equivalent to testing

$$H_0 : \int_{-\pi}^{\pi} \left\| \boldsymbol{\Psi}_{j:m|m}(\boldsymbol{F}_{m|Y}(\lambda)) \right\|^2 d\lambda = 0, \quad H_1 : \int_{-\pi}^{\pi} \left\| \boldsymbol{\Psi}_{j:m|m}(\boldsymbol{F}_{m|Y}(\lambda)) \right\|^2 d\lambda > 0, \qquad (3.9)$$

where $\boldsymbol{\Psi}_{j:m|m}(\boldsymbol{F}_{m|Y}(\lambda)) = (\Psi_{j|m}(\boldsymbol{F}_{m|Y}(\lambda)), \ldots, \Psi_{m|m}(\boldsymbol{F}_{m|Y}(\lambda)))^\top$. However, the formula of $\Psi_{j|m}(\boldsymbol{F}_{m|Y}(\lambda))$ becomes more complicated as j increases. Thus, we design the test in an alternative way.

Let us first reexamine (3.8). This test is designed to test whether the process $\{X_j(t)\}$ should be included in the model given $\{X_1(t)\}, \ldots, \{X_{j-1}(t)\}$, are in the model. Since we have m input processes in total, $\{X_j(t)\}, \{X_{j+1}(t)\}, \ldots, \{X_m(t)\}$ are all possible candidates for jth input. Let $v = m - j + 1$ and denote $\{X_j(t)\}, \ldots, \{X_m(t)\}$ as $\{X(t; u_1)\}, \ldots, \{X(t; u_v)\}$, respectively. Then the test (3.8) can be rewritten as

$$H_0: \int_{-\pi}^{\pi} |\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda; u))|^2 d\lambda = 0, \quad H_1: \int_{-\pi}^{\pi} |\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda; u))|^2 d\lambda > 0, \tag{3.10}$$

where u indicates the process we want to test when there are more than one candidate. If we want to test $\{X_j(t)\}$, then $u = u_1$. The corresponding test statistic can be rewritten as

$$Z_{n,j}(u) = \frac{T_{n,j}(u) - \mu_{n,j}(u)}{\sigma_{n,j}(u)},$$

where

$$\begin{aligned} T_{n,j}(u) &= n M_n^{-\frac{1}{2}} \int_{-\pi}^{\pi} |\Psi_{j|j}(\widehat{\boldsymbol{F}}_{j|Y}(\lambda; u))|^2 d\lambda, \\ \mu_{n,j}(u) &= M_n^{\frac{1}{2}} C_{w,2} \int_{-\pi}^{\pi} \psi_{j|j}(\widehat{\boldsymbol{F}}_{j|Y}(\lambda; u)) d\lambda, \\ \sigma_{n,j}^2(u) &= 4\pi C_{w,4} \int_{-\pi}^{\pi} |\psi_{j|j}(\widehat{\boldsymbol{F}}_{j|Y}(\lambda; u))|^2 d\lambda. \end{aligned}$$

Thus, if we want to test whether $\{X_j(t)\}, \ldots, \{X_m(t)\}$ should be included in the model, we can formulate the hypotheses as

$$H_0: \int_{-\pi}^{\pi} \left\| \boldsymbol{\Phi}_j(\boldsymbol{F}_{m|Y}(\lambda; \boldsymbol{u})) \right\|^2 d\lambda = 0, \quad H_1: \int_{-\pi}^{\pi} \left\| \boldsymbol{\Phi}_j(\boldsymbol{F}_{m|Y}(\lambda; \boldsymbol{u})) \right\|^2 d\lambda > 0, \tag{3.11}$$

where

$$\boldsymbol{\Phi}_j(\boldsymbol{F}_{m|Y}(\lambda;\boldsymbol{u})) =(\phi_{j,1}(\boldsymbol{F}_{m|Y}(\lambda;\boldsymbol{u})),\ldots,\phi_{j,v}(\boldsymbol{F}_{m|Y}(\lambda;\boldsymbol{u})))^\top,$$
$$\phi_{j,w}(\boldsymbol{F}_{m|Y}(\lambda;\boldsymbol{u})) =\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda;u_w)),\quad w=1,\ldots,v.$$

According to [15] and [24], under Assumption 3.2 and 3.3,

$$\ddot{Z}_{n,j}(\boldsymbol{u}) = \frac{\ddot{T}_{n,j}(\boldsymbol{u})-\ddot{\mu}_{n,j}(\boldsymbol{u})}{\ddot{\sigma}_{n,j}(\boldsymbol{u})} \xrightarrow{d} N(0,1),$$

as $n\to\infty$ where

$$\ddot{T}_{n,j}(\boldsymbol{u}) =nM_n^{-\frac{1}{2}}\int_{-\pi}^{\pi}\left\|\boldsymbol{\Phi}_j(\widehat{\boldsymbol{F}}_{m|Y}(\lambda;\boldsymbol{u}))\right\|^2 d\lambda,$$
$$\ddot{\mu}_{n,j}(\boldsymbol{u}) =M_n^{\frac{1}{2}}C_{w,2}\int_{-\pi}^{\pi}\operatorname{tr}\left[\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Phi}_j}(\lambda;\boldsymbol{u})\left(\widehat{\boldsymbol{F}}_{m|Y}(\lambda;\boldsymbol{u})\otimes\widehat{\boldsymbol{F}}_{m|Y}(\lambda;\boldsymbol{u})\right)\right]d\lambda,$$
$$\ddot{\sigma}^2_{n,j}(\boldsymbol{u}) =4\pi C_{w,4}\int_{-\pi}^{\pi}\operatorname{tr}\left[\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Phi}_j}(\lambda;\boldsymbol{u})\left(\widehat{\boldsymbol{F}}_{m|Y}(\lambda;\boldsymbol{u})\otimes\widehat{\boldsymbol{F}}_{m|Y}(\lambda;\boldsymbol{u})\right)\left(\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Phi}_j}(\lambda;\boldsymbol{u})+\widehat{\boldsymbol{\Gamma}}^H_{\boldsymbol{\Phi}_j}(\lambda;\boldsymbol{u})\right)\right.$$
$$\left.\times\left(\widehat{\boldsymbol{F}}_{m|Y}(\lambda;\boldsymbol{u})\otimes\widehat{\boldsymbol{F}}_{m|Y}(\lambda;\boldsymbol{u})\right)\right]d\lambda,$$
$$\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Phi}_j}(\lambda;\boldsymbol{u}) = \sum_{w=1}^{v}\operatorname{vec}\left(\overline{\frac{\partial\phi_{j,w}(\boldsymbol{Z}_{m+1})}{\partial\boldsymbol{Z}_{m+1}}}\right)\operatorname{vec}\left(\frac{\partial\phi_{j,w}(\boldsymbol{Z}_{m+1})}{\partial\boldsymbol{Z}_{m+1}}\right)^\top\Bigg|_{\boldsymbol{Z}_{m+1}=\widehat{\boldsymbol{F}}_{m|Y}(\lambda;\boldsymbol{u})}.$$

Based on the new notation $\{X(t;u_1)\},\ldots,\{X(t;u_v)\}$ for $\{X_j(t)\},\ldots,\{X_m(t)\}$, $\boldsymbol{\Psi}_{j:m|m}(\boldsymbol{F}_{m|Y}(\lambda))$ in (3.9) can be rewritten as

$$\boldsymbol{\Psi}_{j:m|m}(\boldsymbol{F}_{m|Y}(\lambda;u_1,\ldots,u_v)) = (\Psi_{j|m}(\boldsymbol{F}_{m|Y}(\lambda;u_1,\ldots,u_v)),\ldots,\Psi_{m|m}(\boldsymbol{F}_{m|Y}(\lambda;u_1,\ldots,u_v)))^\top,$$

and by the definition of $\Psi_{j|m}(\boldsymbol{F}_{m|Y}(\lambda))$, we have

$$\begin{aligned}\Psi_{j|m}(\boldsymbol{F}_{m|Y}(\lambda;u_1,\dots,u_v)) =& \Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda;u_1)),\\ \Psi_{j+1|m}(\boldsymbol{F}_{m|Y}(\lambda;u_1,\dots,u_v)) =& \Psi_{j+1|j+1}(\boldsymbol{F}_{j+1|Y}(\lambda;u_1,u_2)),\\ &\vdots\\ \Psi_{m|m}(\boldsymbol{F}_{m|Y}(\lambda;u_1,\dots,u_v)) =& \Psi_{m|m}(\boldsymbol{F}_{m|Y}(\lambda;u_1,\dots,u_v)).\end{aligned}$$

Then we use the following theorem to show that (3.9) and (3.11) are indeed equivalent.

Theorem 3.3. *For* $v = m - j + 1$,

$$\boldsymbol{\Psi}_{j:m|m}(\boldsymbol{F}_{m|Y}(\lambda;u_1,\dots,u_v)) = \boldsymbol{0} \;\Leftrightarrow\; \boldsymbol{\Phi}_j(\boldsymbol{F}_{m|Y}(\lambda;u_1,\dots,u_v)) = \boldsymbol{0},\ \lambda\text{−a.e.}$$

Proof. It is sufficient to show that

$$\begin{aligned}&\Psi_{j+w-1|j+w-1}(\boldsymbol{F}_{j+w-1|Y}(\lambda;u_1,\dots,u_w)) = 0,\ w = 1,\dots,v\\ \Leftrightarrow\ &\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda;u_w)) = 0, \quad w = 1,\dots,v,\ \lambda\text{−a.e.}\end{aligned} \tag{3.12}$$

We first prove this with $w = 2$, that is,

$$\begin{aligned}&\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda;u_1)) = 0,\ \ \Psi_{j+1|j+1}(\boldsymbol{F}_{j+1|Y}(\lambda;u_1,u_2)) = 0\\ \Leftrightarrow\ &\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda;u_1)) = 0,\ \ \Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda;u_2)) = 0,\ \lambda\text{−a.e.}\end{aligned} \tag{3.13}$$

Notice that the $f_{j+1,j+1}(\lambda;u_2)$, $f_{Y,j+1}(\lambda;u_2)$, $\boldsymbol{f}_{j+1}(\lambda;u_2)$ in $\boldsymbol{F}_{j+1|Y}(\lambda;u_1,u_2)$ are identical to $f_{j,j}(\lambda;u_2)$, $f_{Y,j}(\lambda;u_2)$, $\boldsymbol{f}_j(\lambda;u_2)$ in $\boldsymbol{F}_{j|Y}(\lambda;u_2)$, respectively.

Thus, for $j = 1$, $f_{Y,2}(\lambda; u_2) = f_{Y,1}(\lambda; u_2)$ so that it is straightforward that

$$
\begin{aligned}
& f_{Y,1}(\lambda; u_1) = 0, \;\; f_{1,1}(\lambda; u_1) f_{Y,2}(\lambda; u_2) - f_{Y,1}(\lambda; u_1) f_{1,2}(\lambda; u_1, u_2) = 0 \\
\Leftrightarrow\ & f_{Y,1}(\lambda; u_1) = 0, \;\; f_{Y,1}(\lambda; u_2) = 0, \; \lambda\text{−a.e.}
\end{aligned}
$$

For $j \geq 2$, the expression of $\boldsymbol{F}_{j|Y}(\lambda; u_1)$ and $\boldsymbol{F}_{j|Y}(\lambda; u_2)$ are given by

$$
\boldsymbol{F}_{j|Y}(\lambda; u_1) = \begin{bmatrix} \boldsymbol{F}_{j-1}(\lambda) & -\boldsymbol{f}_j(\lambda; u_1) & \boldsymbol{f}_{j-1|Y}(\lambda) \\ -\boldsymbol{f}_j^H(\lambda; u_1) & f_{j,j}(\lambda; u_1) & f_{j,Y}(\lambda; u_1) \\ \boldsymbol{f}_{j-1|Y}^H(\lambda) & f_{Y,j}(\lambda; u_1) & f_{Y,Y}(\lambda) \end{bmatrix},
$$

and

$$
\boldsymbol{F}_{j|Y}(\lambda; u_2) = \begin{bmatrix} \boldsymbol{F}_{j-1}(\lambda) & -\boldsymbol{f}_j(\lambda; u_2) & \boldsymbol{f}_{j-1|Y}(\lambda) \\ -\boldsymbol{f}_j^H(\lambda; u_2) & f_{j,j}(\lambda; u_2) & f_{j,Y}(\lambda; u_2) \\ \boldsymbol{f}_{j-1|Y}^H(\lambda) & f_{Y,j}(\lambda; u_2) & f_{Y,Y}(\lambda) \end{bmatrix},
$$

so that $\boldsymbol{F}_{j+1|Y}(\lambda; u_1, u_2)$ can be expressed by

$$
\boldsymbol{F}_{j+1|Y}(\lambda; u_1, u_2) = \begin{bmatrix} \boldsymbol{F}_{j-1}(\lambda) & -\boldsymbol{f}_j(\lambda; u_1) & -\boldsymbol{f}_j(\lambda; u_2) & \boldsymbol{f}_{j-1|Y}(\lambda) \\ -\boldsymbol{f}_j^H(\lambda; u_1) & f_{j,j}(\lambda; u_1) & f_{j,j+1}(\lambda; u_1, u_2) & f_{j,Y}(\lambda; u_1) \\ -\boldsymbol{f}_j^H(\lambda; u_2) & f_{j+1,j}(\lambda; u_1, u_2) & f_{j,j}(\lambda; u_2) & f_{j,Y}(\lambda; u_2) \\ \boldsymbol{f}_{j-1|Y}^H(\lambda) & f_{Y,j}(\lambda; u_1) & f_{Y,j}(\lambda; u_2) & f_{Y,Y}(\lambda) \end{bmatrix}.
$$

By (3.7),

$$\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda;u_1))$$
$$=\det\left(\boldsymbol{F}_{j-1}(\lambda)\right)(f_{Y,j}(\lambda;u_1)+\boldsymbol{f}^H_{j-1|Y}(\lambda)\boldsymbol{F}^{-1}_{j-1}(\lambda)\boldsymbol{f}_j(\lambda;u_1)),$$
$$\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda;u_2))$$
$$=\det\left(\boldsymbol{F}_{j-1}(\lambda)\right)(f_{Y,j}(\lambda;u_2)+\boldsymbol{f}^H_{j-1|Y}(\lambda)\boldsymbol{F}^{-1}_{j-1}(\lambda)\boldsymbol{f}_j(\lambda;u_2)),$$
$$\Psi_{j+1|j+1}(\boldsymbol{F}_{j+1|Y}(\lambda;u_1,u_2))$$
$$=\det\begin{bmatrix}\boldsymbol{F}_{j-1}(\lambda) & -\boldsymbol{f}_j(\lambda;u_1)\\ -\boldsymbol{f}^H_j(\lambda;u_1) & f_{j,j}(\lambda;u_1)\end{bmatrix}\times\Bigg(f_{Y,j}(\lambda;u_2)$$
$$+\begin{bmatrix}\boldsymbol{f}^H_{j-1|Y}(\lambda) & f_{Y,j}(\lambda;u_1)\end{bmatrix}\begin{bmatrix}\boldsymbol{F}_{j-1}(\lambda) & -\boldsymbol{f}_j(\lambda;u_1)\\ -\boldsymbol{f}^H_j(\lambda;u_1) & f_{j,j}(\lambda;u_1)\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{f}_j(\lambda;u_2)\\ -f_{j,j+1}(\lambda;u_1,u_2)\end{bmatrix}\Bigg).$$

Since the spectral matrix is nonsingular by Assumption 3.2 (4) λ-a.e., then

$$\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda;u_1))=0\ \Leftrightarrow f_{Y,j}(\lambda;u_1)+\boldsymbol{f}^H_{j-1|Y}(\lambda)\boldsymbol{F}^{-1}_{j-1}(\lambda)\boldsymbol{f}_j(\lambda;u_1)=0,$$
$$\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda;u_2))=0\ \Leftrightarrow f_{Y,j}(\lambda;u_2)+\boldsymbol{f}^H_{j-1|Y}(\lambda)\boldsymbol{F}^{-1}_{j-1}(\lambda)\boldsymbol{f}_j(\lambda;u_2)=0,$$
$$\Psi_{j+1|j+1}(\boldsymbol{F}_{j+1|Y}(\lambda;u_1,u_2))=0\ \Leftrightarrow f_{Y,j}(\lambda;u_2)$$
$$+\begin{bmatrix}\boldsymbol{f}^H_{j-1|Y}(\lambda) & f_{Y,j}(\lambda;u_1)\end{bmatrix}\begin{bmatrix}\boldsymbol{F}_{j-1}(\lambda) & -\boldsymbol{f}_j(\lambda;u_1)\\ -\boldsymbol{f}^H_j(\lambda;u_1) & f_{j,j}(\lambda;u_1)\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{f}_j(\lambda;u_2)\\ -f_{j,j+1}(\lambda;u_1,u_2)\end{bmatrix}=0,$$

for λ-a.e.

$$\begin{bmatrix} \boldsymbol{F}_{j-1}(\lambda) & -\boldsymbol{f}_j(\lambda;u_1) \\ -\boldsymbol{f}_j^H(\lambda;u_1) & f_{j,j}(\lambda;u_1) \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \boldsymbol{F}_{j-1}^{-1}(\lambda) + \frac{\boldsymbol{F}_{j-1}^{-1}(\lambda)\boldsymbol{f}_j(\lambda;u_1)\boldsymbol{f}_j^H(\lambda;u_1)\boldsymbol{F}_{j-1}^{-1}(\lambda)}{\det(\boldsymbol{F}_j(\lambda;u_1))} & \frac{\boldsymbol{F}_{j-1}^{-1}(\lambda)\boldsymbol{f}_j(\lambda;u_1)}{\det(\boldsymbol{F}_j(\lambda;u_1))} \\ \frac{\boldsymbol{f}_j^H(\lambda;u_1)\boldsymbol{F}_{j-1}^{-1}(\lambda)}{\det(\boldsymbol{F}_j(\lambda;u_1))} & \frac{1}{\det(\boldsymbol{F}_j(\lambda;u_1))} \end{bmatrix},$$

so that

$$f_{Y,j}(\lambda;u_2) + \begin{bmatrix} \boldsymbol{f}_{j-1|Y}^H(\lambda) & f_{Y,j}(\lambda;u_1) \end{bmatrix} \begin{bmatrix} \boldsymbol{F}_{j-1}(\lambda) & -\boldsymbol{f}_j(\lambda;u_1) \\ -\boldsymbol{f}_j^H(\lambda;u_1) & f_{j,j}(\lambda;u_1) \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{f}_j(\lambda;u_2) \\ -f_{j,j+1}(\lambda;u_1,u_2) \end{bmatrix}$$

$$= f_{Y,j}(\lambda;u_2) + \boldsymbol{f}_{j-1|Y}^H(\lambda)\boldsymbol{F}_{j-1}(\lambda)\boldsymbol{f}_j(\lambda;u_2) + \frac{1}{\det(\boldsymbol{F}_j(\lambda;u_1))}$$
$$\times \left(f_{Y,j}(\lambda;u_1) + \boldsymbol{f}_{j-1|Y}^H(\lambda)\boldsymbol{F}_{j-1}(\lambda)\boldsymbol{f}_j(\lambda;u_1)\right)\left(\boldsymbol{f}_j^H(\lambda;u_1)\boldsymbol{F}_{j-1}(\lambda)\boldsymbol{f}_j(\lambda;u_2) - f_{j,j+1}(\lambda;u_1,u_2)\right).$$

Thus, given $f_{Y,j}(\lambda;u_1) + \boldsymbol{f}_{j-1|Y}^H(\lambda)\boldsymbol{F}_{j-1}(\lambda)\boldsymbol{f}_j(\lambda;u_1) = 0$,

$$f_{Y,j}(\lambda;u_2) + \boldsymbol{f}_{j-1|Y}^H(\lambda)\boldsymbol{F}_{j-1}(\lambda)\boldsymbol{f}_j(\lambda;u_2) = 0 \iff f_{Y,j}(\lambda;u_2)$$
$$+ \begin{bmatrix} \boldsymbol{f}_{j-1|Y}^H(\lambda) & f_{Y,j}(\lambda;u_1) \end{bmatrix} \begin{bmatrix} \boldsymbol{F}_{j-1}(\lambda) & -\boldsymbol{f}_j(\lambda;u_1) \\ -\boldsymbol{f}_j^H(\lambda;u_1) & f_{j,j}(\lambda;u_1) \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{f}_j(\lambda;u_2) \\ -f_{j,j+1}(\lambda;u_1,u_2) \end{bmatrix} = 0,$$

and thus (3.13) is proved.

Since j, u_1 and u_2 are arbitrary, then by (3.13),

$$
\begin{aligned}
&\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda; u_1)) = 0,\\
&\Psi_{j+1|j+1}(\boldsymbol{F}_{j+1|Y}(\lambda; u_1, u_2)) = 0,\\
&\Psi_{j+2|j+2}(\boldsymbol{F}_{j+2|Y}(\lambda; u_1, u_2, u_3)) = 0\\
\Leftrightarrow\ &\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda; u_1)) = 0,\\
&\Psi_{j+1|j+1}(\boldsymbol{F}_{j+1|Y}(\lambda; u_1, u_2)) = 0,\\
&\Psi_{j+1|j+1}(\boldsymbol{F}_{j+1|Y}(\lambda; u_1, u_3)) = 0\\
\Leftrightarrow\ &\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda; u_1)) = 0,\\
&\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda; u_2)) = 0,\\
&\Psi_{j|j}(\boldsymbol{F}_{j|Y}(\lambda; u_3)) = 0,\ \lambda\text{−a.e.}
\end{aligned}
$$

Repeat the above argument, (3.12) is proved, which completes the proof. □

The equivalence of (3.9) and (3.11) immediately follows Theorem 3.3.

3.3 Nonlinearity and Residual Coherence

When the input processes in (3.4) are zero-mean polynomial processes of a single process $\{X(t)\}$, the tests (3.10) and (3.11) can be used to test the nonlinearity between $\{X(t)\}$ and $\{Y(t)\}$. A typical example is the lagged process model proposed by [42]. The model can be considered as

$$
Y(t) = \sum_{k_1=-\infty}^{\infty} b_1(k_1)X_1(t-k_1) + \sum_{k_2=-\infty}^{\infty} b_2(k_2; u)X_2(t-k_2; u) + \varepsilon(t),
$$

where $X_1(t) = X(t)$ and $X_2(t;u) = \widetilde{X}_u(t) = X(t)X(t+u) - R_{XX}(u)$. $\widetilde{X}_u(t)$ in [42] is referred as the lagged process with lag u. Thus, (3.10) and (3.11) can be used to test whether a single lagged process or a set of lagged processes should be included in the model given that the linear process $\{X(t)\}$ is included as an input. A related concept, lagged coherence, is introduced by [42] to measure the effect of the lagged process. The lagged coherence corresponding to $X_2(t;u)$ is given by

$$S_2(\lambda;u) = \frac{f_{G_1G_1}(\lambda) + f_{G_2G_2}(\lambda;u)}{f_{Y,Y}(\lambda)},$$

and it is noted by [42] that

$$S_1(\lambda) = \frac{f_{G_1G_1}(\lambda)}{f_{Y,Y}(\lambda)}$$

is indeed the squared linear coherence between $\{X(t)\}$ and $\{Y(t)\}$. This is closely related to the squared partial coherency spectrum in [27], which also measures the effect of $\{X_2(t;u)\}$ given $\{X_1(t)\}$ is included in the model. The squared partial coherency spectrum of $\{X(t)\}$ and $\{Y(t)\}$ and $\{Y(t)\}$ given $\{X_1(t)\}$ has a connection with the lagged coherence such that

$$\kappa^2_{2Y|1}(\lambda) = \frac{f_{1,1}(\lambda)f_{Y,Y}(\lambda)}{f_{1,1}(\lambda)f_{Y,Y}(\lambda) - |f_{1,Y}(\lambda)|^2}(S_2(\lambda;u) - S_1(\lambda)).$$

Since $S_1(\lambda)$ does not depend on u so that $S_2(\lambda;u) - S_1(\lambda)$ can also be used to measure the effect of $\{X_2(t;u)\}$, and $\kappa^2_{2Y|1}$ can be regarded as a weighted version of $S_2(\lambda;u) - S_1(\lambda)$. However, these measures are functions of λ, so that they could be maximized by different u's for different frequencies. In order to compare the effect of different lagged processes, [43] and [31] introduced the residual coherence

$$RS(u) = \sup_{\lambda}(S_2(\lambda; u) - S_1(\lambda)) = \sup_{\lambda} \frac{f_{G_2G_2}(\lambda; u)}{f_{Y,Y}(\lambda)}.$$

Moreover, $\{X_2(t; u)\}$ is not necessarily a lagged process and it can be considered as a candidate for the second input in general as discussed in Section 3.2.3. Thus, the residual coherence can be used to measure the effect of the second input in general. Based on its formula, it is effortless to extend it to a measure of jth input. But let us first examine another criterion, integrated spectrum, proposed by [77]. The integrated spectrum measuring the effect of jth input is defined as

$$IS_j(u) = \int_{-\pi}^{\pi} f_{G_jG_j}(\lambda; u)d\lambda.$$

The integration is used to make the criterion independent of λ. Thus, we define the Generalized Residual Coherence (GRC) for jth input as

$$GRC_j(u; p, q, k(\cdot)) = \left\|k(\lambda) f_{G_jG_j}(\lambda; u)\right\|_p^q,$$

where $k(\cdot)$ is a positive and bounded function, $p = 1, 2, \ldots$, and $q > 0$. For example,

$$\begin{aligned} GRC_2(u; \infty, 1, f_{Y,Y}^{-1}(\cdot)) =& RS(u), \\ GRC_j(u; 1, 1, 1) =& IS_j(u). \end{aligned}$$

To simplify the notation, we denote $GRC_j(u) = GRC_j(u; 1, 1, 1)$, that is, we let $GRC_j(u; 1, 1, 1)$ be the default criterion. Thus, (3.10) and (3.11) can also be expressed as

$$H_0 : \ GRC_j(u) = 0, \quad H_1 : \ GRC_j(u) > 0, \tag{3.14}$$

and

$$\begin{aligned} &H_0 : \ GRC_j(u_w) = 0, \ w = 1, \ldots, v, \\ &H_1 : \ \text{At least one } GRC_j(u_w) > 0, \ w = 1, \ldots, v, \end{aligned} \tag{3.15}$$

respectively.

3.4 Simulation

In this section, we use 11 simulated cases to verify that the tests (3.10) and (3.11) have satisfactory Type-I error rate and power. All tests are performed at the significance level of 0.05, and all simulated Type-I error rates and powers are calculated based on 1000 iterations. All cross spectral densities are estimated by the Parzen window

$$w(x) = \begin{cases} 1 - 6|x|^2 + 6|x|^3, & |x| < \frac{1}{2}, \\ 2(1 - |x|)^3, & \frac{1}{2} \le |x| \le 1, \\ 0, & \text{otherwise}, \end{cases}$$

with bandwidth $M_n = \lfloor n^{\frac{1}{3.5}} \rfloor$, which satisfies Assumption 3.3. First, we generate an AR(1) process $Z(t)$ such that

$$Z(t) = 0.5Z(t-1) + e(t),$$

where $e(t)$'s are IID $N(0, 0.75)$ noise. Next, denote the candidate processes that we use to test as

$$\begin{aligned} X(t;0) =& Z(t), \\ X(t;1) =& Z(t)Z(t-1) - 0.5, \\ X(t;2) =& Z(t)Z(t-2) - 0.25. \end{aligned}$$

Then, the following three models are constructed for the simulation.

Model 1: $Y(t) = \varepsilon(t)$.

Model 2: $Y(t) = X(t;0) + \varepsilon(t)$.

Model 3: $Y(t) = X(t;0) + 0.8X(t;1) + \varepsilon(t)$.

$\{\varepsilon(t)\}$ is a sequence of IID $N(0, 1)$ noise. The description of simulated cases is given in Table 3.1.

Finally, we obtain the rejection rate of each case (Type-I error rate when H_0 is true and power when H_0 is false) with length of time series $N = 100, 200, 500, 1000, 2000$. As shown in Table 3.2, the simulated Type-error rates are closed to 0.05, and simulated powers are closed to 1 so that the tests we developed are valid.

Table 3.1: The models and tests used for the 11 simulated cases.

Case	Model	H_0 (True or False)		Description of the test
1	Model 1	$GRC_1(0) = 0$	(True)	Test $\{X(t;0)\}$ as the first input.
2		$GRC_1(0) = GRC_1(1) = 0$	(True)	Test whether the first input can be one of $\{X(t;0)\}$ and $\{X(t;1)\}$.
3	Model 2	$GRC_1(0) = 0$	(False)	Test $X(t;0)$ as the first input.
4		$GRC_1(0) = GRC_1(1) = 0$	(False)	Test whether the first input can be one of $\{X(t;0)\}$ and $\{X(t;1)\}$.
5		$GRC_2(1) = 0$	(True)	Test $\{X(t;1)\}$ as the second input given that $\{X(t;0)\}$ is the first input.
6		$GRC_2(1) = GRC_2(2) = 0$	(True)	Test whether the second input can be one of $\{X(t;1)\}$ and $\{X(t;2)\}$ given that $\{X(t;0)\}$ is the first input.
7	Model 3	$GRC_1(0) = 0$	(False)	Test $\{X(t;0)\}$ as the first input.
8		$GRC_1(0) = GRC_1(1) = 0$	(False)	Test whether the first input can be one of $\{X(t;0)\}$ and $\{X(t;1)\}$.
9		$GRC_2(1) = 0$	(False)	Test $\{X(t;1)\}$ as the second input given that $\{X(t;0)\}$ is the first input.
10		$GRC_2(1) = GRC_2(2) = 0$	(False)	Test whether the second input can be one of $\{X(t;1)\}$ and $\{X(t;2)\}$ given that $\{X(t;0)\}$ is the first input.
11		$GRC_3(2) = 0$	(True)	Test $\{X(t;2)\}$ as the third input given $\{X(t;0)\}$ and $X(t;1)$ are the first two inputs.

3.5 Analysis of Brain Functional Connectivity Data

In this section, we apply tests (3.10) and (3.11) to detecting associations between BOLD sequences extracted from 246 brain subregions. These subregions are defined based on the brainnetome atlas introduced by [17]. The data are obtained from 110 people including 61 healthy individuals and 49 schizophrenia patients. For each individual, 246 BOLD sequences that correspond to the 246 subregions are recorded with length 150. Further description of the data can be referred to [9].

For each individual, we can use (3.10) and (3.11) to test whether the BOLD sequences of two subregions have a significant association. Such association can be linear, quadratic or higher degree polynomials. One may use the test result from all individuals for further analysis. To

Table 3.2: Rejection rates of 11 simulated cases. For Case 1, 2, 5, 6, 11, the rejection rate represents the estimated Type-I error. For Case 3, 4, 7, 8, 9, 10, the rejection rate represents the estimated power.

Case	$N = 100$	$N = 200$	$N = 500$	$N = 1000$	$N = 2000$	$N = 5000$
1	0.070	0.077	0.081	0.079	0.080	0.068
2	0.034	0.037	0.039	0.035	0.036	0.026
3	1.000	1.000	1.000	1.000	1.000	1.000
4	1.000	1.000	1.000	1.000	1.000	1.000
5	0.072	0.080	0.064	0.089	0.084	0.068
6	0.033	0.038	0.029	0.049	0.046	0.035
7	0.995	1.000	1.000	1.000	1.000	1.000
8	1.000	1.000	1.000	1.000	1.000	1.000
9	1.000	1.000	1.000	1.000	1.000	1.000
10	0.999	1.000	1.000	1.000	1.000	1.000
11	0.057	0.042	0.043	0.061	0.051	0.054

avoid digression, we shall focus on demonstrating the use of the tests. Thus, we apply the tests to the BOLD sequences averaged over individuals instead of using them for each individual. But we should bear in mind that more information can be obtained if these tests are used for analyzing the BOLD sequences from each individual.

3.5.1 Data Processing

First, for each subregion, we average BOLD sequences in healthy control group and schizophrenia patient group such that

$$Z_{H,i}(t) = \frac{1}{61}\sum_{j=1}^{61} Z_{H,i,j}(t),$$
$$Z_{P,i}(t) = \frac{1}{49}\sum_{j=1}^{49} Z_{P,i,j}(t),$$

where $\{Z_{H,i,j}(t)\}$ represents the BOLD sequence of subregion i obtained from jth healthy individual, while $\{Z_{P,i,j}(t)\}$ represents that obtained from jth patient. Thus, for each subregion, we have one BOLD sequence for the healthy control group and one for schizophrenia patient group.

Moreover, to make the sequences sufficiently stationary, we take the first order difference and remove the first two observations for all $\{Z_{H,i}(t)\}$ and $\{Z_{P,i}(t)\}$. Furthermore, all sequences are centered so that the processed sequences have mean 0. The processed sequences are denoted as $\{X_{H,i}(t)\}$ and $\{X_{P,i}(t)\}$, $i = 1, \ldots, 246$.

3.5.2 Testing for Quadratic Association

For healthy control group, we have $\{X_{H,i}(t)\}$ for $i = 1, \ldots, 246$, representing the 246 subregions. For any two subregions i and j, the linear association can be tested based on

$$\begin{aligned} H_0 : \ Y(t) =& \varepsilon(t), \\ H_1 : \ Y(t) =& \sum_{k_1=-\infty}^{\infty} b_1(k_1)X_1(t-k_1) + \varepsilon(t), \end{aligned} \tag{3.16}$$

where $Y(t) = X_{H,i}(t)$ and $X_1(t) = X_{H,j}(t)$ when $\{X_{H,i}(t)\}$ is considered as the output. We can exchange $\{X_{H,i}(t)\}$ and $\{X_{H,j}(t)\}$ and perform the test with $\{X_{H,j}(t)\}$ being the output. This test corresponds to (3.10).

To detect possible quadratic association, we apply our tests to lagged processes. If the linear association is significant, we consider testing

$$\begin{aligned} H_0 : \ Y(t) =& \sum_{k_1=-\infty}^{\infty} b_1(k_1)X_1(t-k_1) + \varepsilon(t), \\ H_1 : \ Y(t) =& \sum_{k_1=-\infty}^{\infty} b_1(k_1)X_1(t-k_1) + \sum_{k_2=-\infty}^{\infty} b_2(k_2;u)X_2(t-k_2;u) + \varepsilon(t), \end{aligned} \tag{3.17}$$

where $Y(t) = X_{H,i}(t)$, $X_1(t) = X_{H,j}(t)$ and $X_2(t;u) = X_{H,j}(t)X_{H,j}(t+u) - R_{X_{H,j}X_{H,j}}(u)$ when $\{X_{H,i}(t)\}$ is considered as the output. If the linear association is insignificant, we consider testing

$$
\begin{aligned}
H_0 &: \ Y(t) = \varepsilon(t), \\
H_1 &: \ Y(t) = \sum_{k_2=-\infty}^{\infty} b_2(k_2; u) X_2(t - k_2; u) + \varepsilon(t).
\end{aligned}
\tag{3.18}
$$

Again, we can exchange $\{X_{H,i}(t)\}$ and $\{X_{H,j}(t)\}$ to obtain test result with $\{X_{H,j}(t)\}$ being the output. The tests (3.17) and (3.18) correspond to (3.10) when we test for a single u and (3.11) when we test for multiple u's. Note that (3.17) can also be considered as testing $GRC_2(u) = 0$ given $\{X_1(t)\}$ is the first input, and (3.18) can be considered as testing $GRC_1(u) = 0$.

Similarly, the above analysis can be repeated for schizophrenia patient group. Thus, we can use colored blocks to show the test result from the two groups. In Figure 3.1, the blue blocks indicate that the association is significant when $\{X_{H,2}(t)\}$ is the output of $\{X_{H,1}(t)\}$, and $\{X_{H,3}(t)\}$ is the output of $\{X_{H,2}(t)\}$. In Figure 3.2, the red blocks indicate that the association is significant when $\{X_{P,2}(t)\}$ is the output of $\{X_{P,1}(t)\}$, and $\{X_{P,1}(t)\}$ is the output of $\{X_{P,3}(t)\}$. Thus, Figure 3.3 is obtained by merging Figure 3.1 and 3.2. The four colors, white (W), blue (B), red (R), and purple (P), represent the four different test results of the two subregions such that

W: the association is insignificant in both groups.

B: the association is significant in healthy control group but insignificant in schizophrenia patient group.

R: the association is significant in schizophrenia patient group but insignificant in healthy control group.

P: the association is significant in both groups.

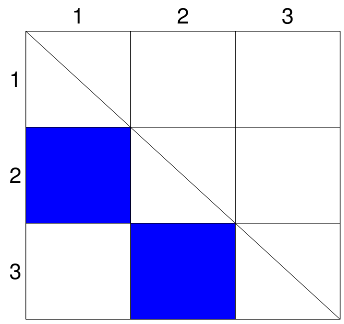

Figure 3.1: Test result of $\{X_{H,i}(t)\}$, $i = 1, 2, 3$. Blue indicates significance.

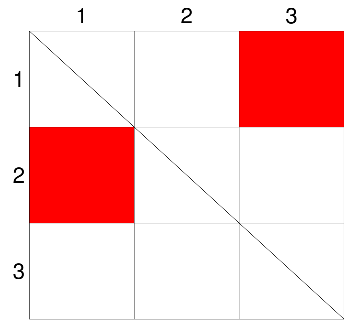

Figure 3.2: Test result of $\{X_{P,i}(t)\}$, $i = 1, 2, 3$. Red indicates significance.

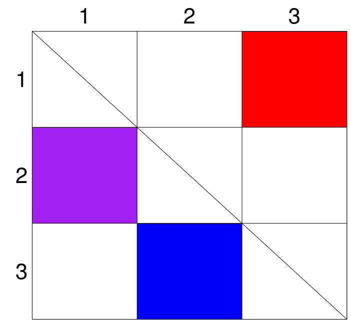

Figure 3.3: Colored blocks obtained by adding up the result of Figure 3.1 and 3.2.

Then, we apply the above tests discussed to the subregions of subcortical nuclei. These subregions are labeled by 211-246 in [17]. An odd number indicates that the corresponding subregion is in left hemisphere while an even number indicates otherwise. We relabel these subregions, letting 1-18 be the subregions in left hemisphere (211, 213,...,245) and 19-36 be the subregions in right hemisphere (212, 214,..., 246). All tests are performed at the significance level of 0.05 and all cross spectral densities are estimated by the Parzen window with bandwidth $M_n = \lfloor n^{\frac{1}{3.5}} \rfloor$.

Figure 3.4 is obtained by testing the linear association between the subregions, and Figure 3.5 is obtained by testing the lagged processes with lags $u = 0, \ldots, 4$. In Figure 3.4 and 3.5, the four blocks represent the test results from left and right hemisphere such that

Upper left: both input and output are from left hemisphere.

Upper right: output is from left hemisphere and input is from right hemisphere.

Lower left: input is from left hemisphere and output is from right hemisphere.

Lower right: both input and output are from right hemisphere.

From Figure 3.4, we can readily observe the discrepancy between the healthy control group and schizophrenia patient group in terms of the linear associations between the subregions. It seems that more subregions are linearly connected in the healthy control group since the blue area is larger than the red area. Comparing Figure 3.4 and 3.5, the test for lagged processes renders us more information about the differences between the two groups. In Figure 3.4, the blocks of row 27-36 and column 11-14 are mostly purple, indicating that we cannot observe any differences between the two groups regarding the linear association. Nonetheless, there are several red blocks in this area observed from Figure 3.5, which signifies the differences between the two groups in terms of the quadratic association.

3.5.3 A Forward Selection Method

A forward selection of input processes is developed based on test (3.10), or equivalently (3.14). Suppose we have candidate input processes $\{X(t;u)\}$ for $u = 1, \ldots, v$. Then we can start with the null model

$$Y(t) = \varepsilon(t),$$

and test

$$H_0: \; GRC_1(u) = 0,$$

for each $u \in \{1, \ldots, v\}$. Compute $GRC_1(u)$ for u such that H_0 is rejected, and find $u = u_1$ that maximizes $GRC_1(u)$. Then, test

$$H_0: \; GRC_2(u) = 0,$$

for each $u \in \{1, \ldots, v\} \backslash \{u_1\}$ given the first input is $\{X(t; u_1)\}$. Repeat this procedure until

$$H_0: \; GRC_j(u) = 0,$$

is not rejected for every $u \in \{1, \ldots, v\} \backslash \{u_1, \ldots, u_{j-1}\}$. Alternatively, one may test multiple u's all at once. But here we prefer to test each one of them without affected by the rest of inputs. Also, the multiplicity issue can be neglected since this is a selection method and we do not compute CIs or make statements about confidence levels.

We demonstrate this forward selection method based on the BOLD sequences (1-18) in left hemisphere. Take the BOLD sequence in subregion 1 as the output, and the rest (2-18) as inputs. For the healthy control group, Figure 3.6, 3.7 and 3.8 shows that the three inputs are subregion 3, 2 and 5 in the order of entering the model. For the schizophrenia patient group, the selection stopped at $j = 3$ with the first input being subregion 2 and the second input being subregion 3 as shown in Figure 3.9 and 3.10.

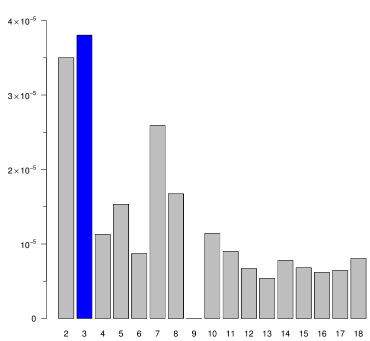

Figure 3.6: GRC_1 of subregion 2-18. The vacancy indicates the corresponding subregion is insignificant or has already been included as an input.

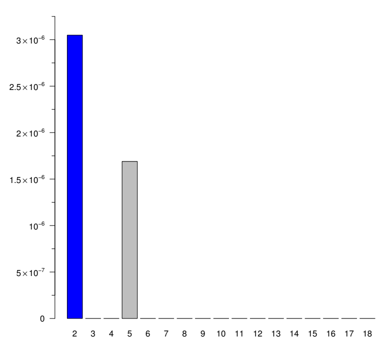

Figure 3.7: GRC_2 of subregion 2-18. The vacancy indicates the corresponding subregion is insignificant or has already been included as an input.

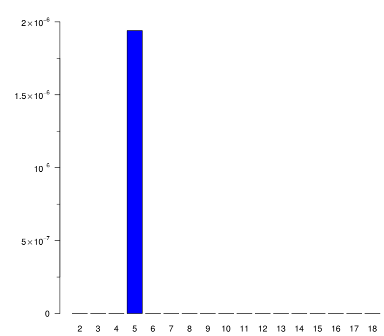

Figure 3.8: GRC_3 of subregion 2-18. The vacancy indicates the corresponding subregion is insignificant or has already been included as an input.

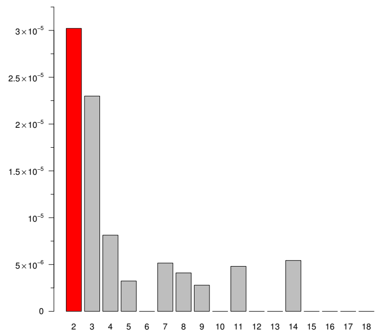

Figure 3.9: GRC_1 of subregion 2-18. The vacancy indicates the corresponding subregion is insignificant or has already been included as an input.

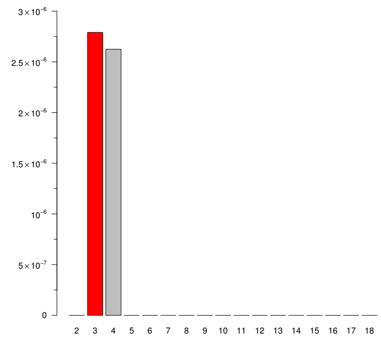

Figure 3.10: GRC_2 of subregion 2-18. The vacancy indicates the corresponding subregion is insignificant or has already been included as an input.

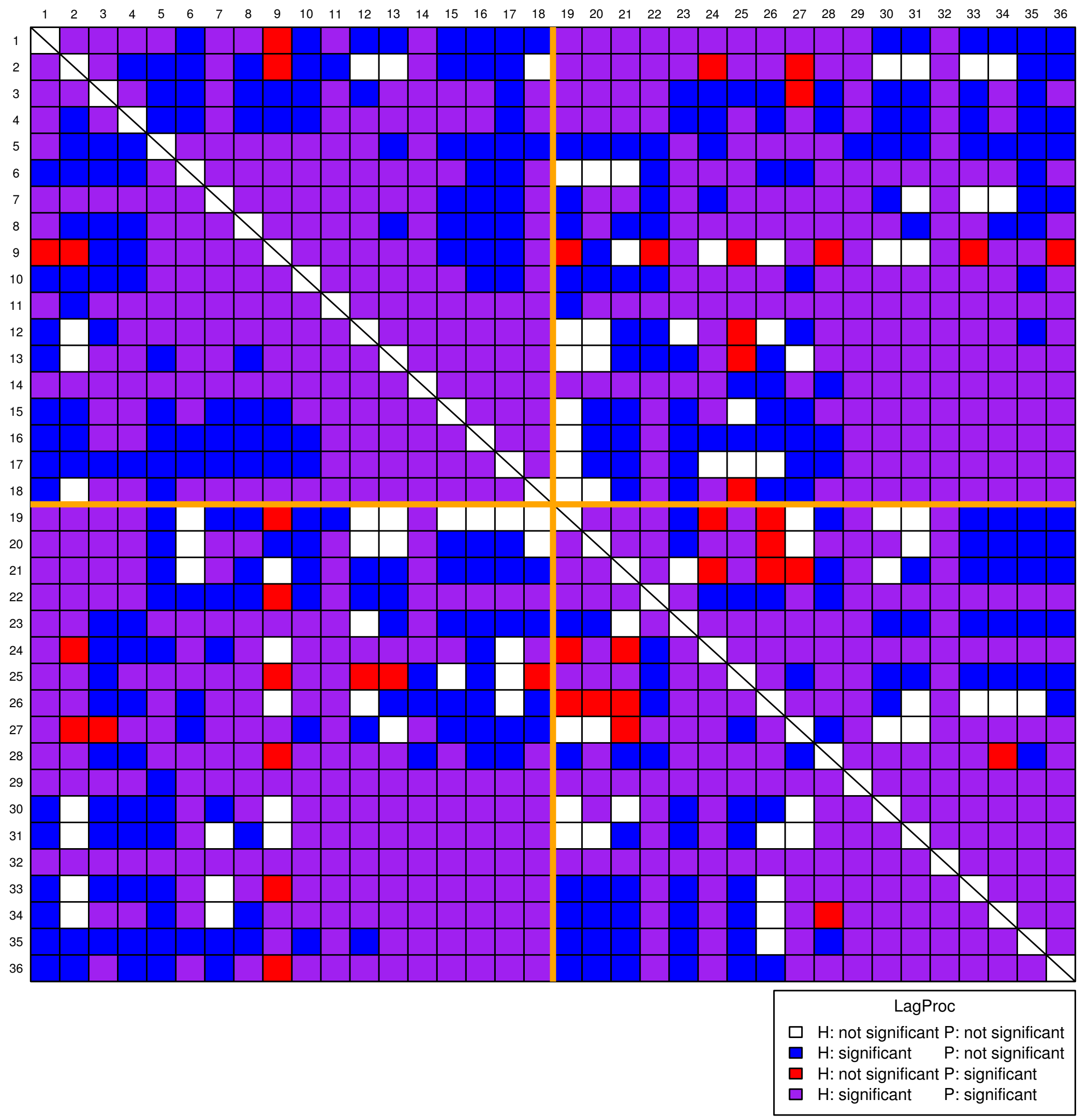

Figure 3.4: Colored blocks based on testing for linear association.

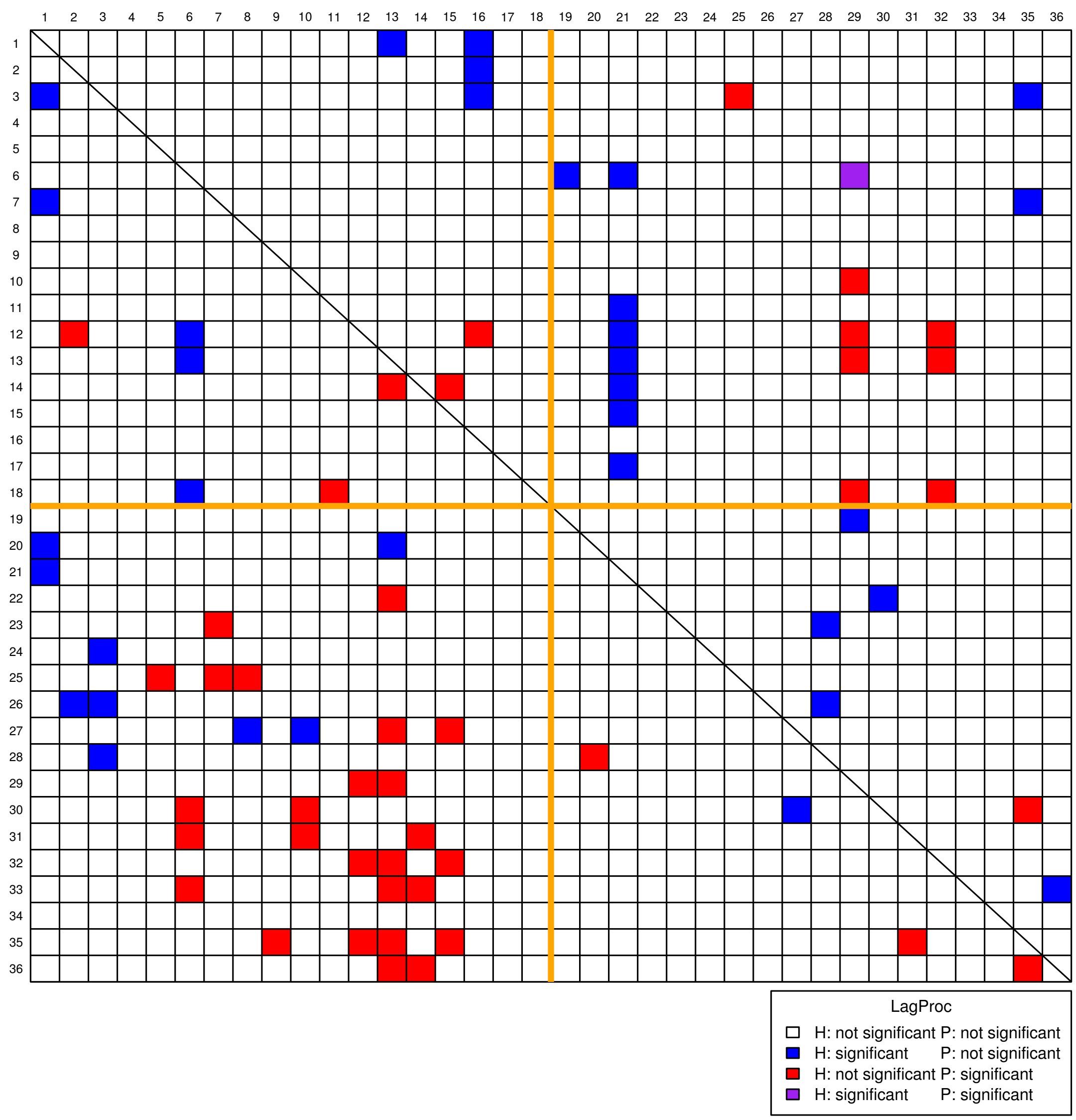

Figure 3.5: Colored blocks based on testing for lagged processes with lags $u = 0, \ldots, 4$.

Bibliography

[1] EPA Assessment of Risks from Radon in Homes: EPA 402-R-03-003. https://www.epa.gov/sites/production/files/2015-05/documents/402-r-03-003.pdf. Accessed: 2023-11-13.

[2] Radon in the Home. https://www.dep.pa.gov/Business/RadiationProtection/RadonDivision/Pages/Radon-in-the-home.aspx. Accessed: 2023-11-13.

[3] T. Bourdrel, M.-A. Bind, Y. Béjot, O. Morel, and J.-F. Argacha. Cardiovascular effects of air pollution. *Archives of cardiovascular diseases*, 110(11):634–642, 2017.

[4] R. C. Bradley. Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. *Probability Surveys*, 2:107 – 144, 2005.

[5] P. J. Brockwell and R. A. Davis. *Time series: theory and methods*. Springer science & business media, 1991.

[6] A. J. Chauhan and S. L. Johnston. Air pollution and infection in respiratory illness. *British medical bulletin*, 68(1):95–112, 2003.

[7] J. Chen and Y. Liu. Quantile and quantile-function estimations under density ratio model. *The Annals of Statistics*, 41(3):1669–1692, 2013.

[8] S. X. Chen, W. Härdle, and M. Li. An empirical likelihood goodness-of-fit test for time series. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(3):663–678, 2003.

[9] A. J. Culbreth, Q. Wu, S. Chen, B. M. Adhikari, L. E. Hong, J. M. Gold, and J. A. Waltz. Temporal-thalamic and cingulo-opercular connectivity in people with schizophrenia. *NeuroImage: Clinical*, 29:102531, 2021.

[10] Y. A. Davydov. Convergence of distributions generated by stationary stochastic processes. *Theory of Probability & Its Applications*, 13(4):691–696, 1968.

[11] M. De Carvalho and A. C. Davison. Spectral density ratio models for multivariate extremes. *Journal of the American Statistical Association*, 109(506):764–776, 2014.

[12] V. De Oliveira and B. Kedem. Bayesian analysis of a density ratio model. *Canadian Journal of Statistics*, 45(3):274–289, 2017.

[13] J. Dedecker, P. Doukhan, G. Lang, L. R. José Rafael, S. Louhichi, and C. Prieur. *Weak Dependence: With Examples and Applications*. Lecture Notes in Statistics. Springer New York, 2007.

[14] B. Efron and R. Tibshirani. Using specially designed exponential families for density estimation. *The Annals of Statistics*, 24(6):2431–2461, 1996.

[15] M. Eichler. Testing nonparametric and semiparametric hypotheses in vector stationary processes. *Journal of Multivariate Analysis*, 99(5):968–1009, 2008.

[16] A. El Ghouch, I. Van Keilegom, and I. W. McKeague. Empirical likelihood confidence intervals for dependent duration data. *Econometric Theory*, 27(1):178–198, 2011.

[17] L. Fan, H. Li, J. Zhuo, Y. Zhang, J. Wang, L. Chen, Z. Yang, C. Chu, S. Xie, A. R. Laird, P. T. Fox, S. B. Eickhoff, C. Yu, and T. Jiang. The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cerebral cortex*, 26(8):3508–3526, 2016.

[18] K. Fokianos. Merging information for semiparametric density estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(4):941–958, 2004.

[19] K. Fokianos. Density ratio model selection. *Journal of Statistical Computation and Simulation*, 77(9):805–819, 2007.

[20] K. Fokianos and I. Kaimi. On the effect of misspecifying the density ratio model. *Annals of the Institute of Statistical Mathematics*, 58(3):475–497, 2006.

[21] K. Fokianos, B. Kedem, J. Qin, and D. A. Short. A semiparametric approach to the one-way layout. *Technometrics*, 43(1):56–65, 2001.

[22] P. B. Gilbert, S. R. Lele, and Y. Vardi. Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika*, 86(1):27–43, 1999.

[23] R. D. Gill, Y. Vardi, and J. A. Wellner. Large sample theory of empirical distributions in biased sampling models. *The Annals of Statistics*, 16(3):1069–1112, 1988.

[24] Y. Goto, X. Zhang, B. Kedem, and S. Chen. Residual spectrum applied in brain functional connectivity. *arXiv preprint arXiv:2305.19461*, 2023.

[25] H. Guo. Generalized Volatility Model And Calculating VaR Using A New Semiparametric Model, 2005. Ph.D. Dissertation, University of Maryland, College Park.

[26] I. A. Ibragimov and Y. V. Linnik. *Independent and stationary sequences of random variables*. Walters-Noordhoff, Groningen, 1971.

[27] G. M. Jenkins and D. G. Watts. *Spectral analysis and its applications*. Holden-Day, San Francisco., 1968.

[28] S. Karlin and H. E. Taylor. *A First Course in Stochastic Processes*. Academic Press New York, second edition, 1975.

[29] Y. Katznelson. *An introduction to harmonic analysis*. Cambridge University Press, 2004.

[30] M. Katzoff, W. Zhou, D. Khan, G. Lu, and B. Kedem. Out-of-sample fusion in risk prediction. *Journal of Statistical Theory and Practice*, 8:444–459, 2014.

[31] B. Kedem. Coherence consideration in binary time series analysis. *Handbook of Discrete-Valued Time Series*, 311, 2016.

[32] B. Kedem, V. De Oliveira, and M. Sverchkov. *Statistical data fusion*. World Scientific, 2017.

[33] B. Kedem and K. Fokianos. *Regression models for time series analysis*. John Wiley & Sons, 2005.

[34] B. Kedem and R. E. Gagnon. Semiparametric distribution forecasting. *Journal of statistical planning and inference*, 140(12):3734–3741, 2010.

[35] B. Kedem, G. Lu, R. Wei, and P. D. Williams. Forecasting mortality rates via density ratio modeling. *Canadian Journal of Statistics*, 36(2):193–206, 2008.

[36] B. Kedem, L. Pan, P. Smith, and C. Wang. Repeated out of sample fusion in the estimation of small tail probabilities. *arXiv preprint arXiv:1803.10766*, 2018.

[37] B. Kedem, L. Pan, and W. Zhou. Out of sample fusion: A Monte Carlo method. In *2014 International Conference on Computational Science and Computational Intelligence*, volume 1, pages 364–367. IEEE, 2014.

[38] B. Kedem, L. Pan, W. Zhou, and C. A. Coelho. Interval estimation of small tail probabilities–applications in food safety. *Statistics in Medicine*, 35(18):3229–3240, 2016.

[39] B. Kedem and S. Pyne. Estimation of tail probabilities by repeated augmented reality. *Journal of statistical theory and practice*, 15:1–16, 2021.

[40] B. Kedem, R. M. Stauffer, X. Zhang, and S. Pyne. On the probabilities of environmental extremes. *International Journal of Statistics in Medical Research*, 10:72–84, 2021.

[41] B. Kedem and S. Wen. Semi-parametric cluster detection. *Journal of statistical theory and practice*, 1(1):49–72, 2007.

[42] B. Kedem-Kimelfeld. Estimating the lags of lag processes. *Journal of the American Statistical Association*, 70(351a):603–605, 1975.

[43] D. Khan, M. Katzoff, and B. Kedem. Coherence structure and its application in mortality forecasting. *Journal of Statistical Theory and Practice*, 8:578–590, 2014.

[44] B. Kimelfeld. Estimating the kernels of nonlinear orthogonal polynomial functionals. *The Annals of Statistics*, pages 353–358, 1974.

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

[46] A. N. Kolmogorov and Y. A. Rozanov. On strong mixing conditions for stationary gaussian processes. *Theory of Probability & Its Applications*, 5(2):204–208, 1960.

[47] G. Lu. Asymptotic Theory for Multiple-Sample Semiparametric Density Ratio Models and its Application to Mortality Forecasting, 2007. Ph.D. Dissertation, University of Maryland, College Park.

[48] T. McElroy and D. N. Politis. Estimating the spectral density at frequencies near zero. *Journal of the American Statistical Association*, 0(0):1–29, 2022.

[49] A. Mokkadem. Mixing properties of ARMA processes. *Stochastic processes and their applications*, 29(2):309–315, 1988.

[50] D. J. Nordman and S. N. Lahiri. A review of empirical likelihood methods for time series. *Journal of Statistical Planning and Inference*, 155:1–18, 2014.

[51] L. Pan. Semiparametric Methods in the Estimation of Tail Probabilities and Extreme Quantiles, 2016. Ph.D. Dissertation, University of Maryland, College Park.

[52] R. L. Prentice and R. Pyke. Logistic disease incidence models and case-control studies. *Biometrika*, 66(3):403–411, 1979.

[53] J. Qin and J. Lawless. Empirical likelihood and general estimating equations. *The Annals of Statistics*, pages 300–325, 1994.

[54] J. Qin and B. Zhang. A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*, 84(3):609–618, 1997.

[55] E. Rio. *Asymptotic theory of weakly dependent random processes*, volume 80. Springer, 2017.

[56] V. K. Rohatgi and A. M. E. Saleh. *An introduction to probability and statistics*. John Wiley & Sons, 2015.

[57] M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proceedings of the national Academy of Sciences*, 42(1):43–47, 1956.

[58] A. W. Van Der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer, 1996.

[59] Y. Vardi. Nonparametric estimation in the presence of length bias. *The Annals of Statistics*, 10(2):616–620, 1982.

[60] Y. Vardi. Empirical distributions in selection bias models. *The Annals of Statistics*, pages 178–203, 1985.

[61] V. A. Volkonskii and Y. A. Rozanov. Some limit theorems for random functions. I. *Theory of Probability & Its Applications*, 4(2):178–197, 1959.

[62] A. Voulgaraki. Semiparametric Regression and Mortality Rate Prediction, 2011. Ph.D. Dissertation, University of Maryland, College Park.

[63] A. Voulgaraki, B. Kedem, and B. I. Graubard. Semiparametric regression in testicular germ cell data. *The Annals of Applied Statistics*, 6(3):1185–1208, 2012.

[64] S. G. Walker. A note on the innovation distribution of a gamma distributed autoregressive process. *Scandinavian journal of statistics*, 27(3):575–576, 2000.

[65] C. Wang. Data Fusion based on the Density Ratio Model, 2018. Ph.D. Dissertation, University of Maryland, College Park.

[66] S. Wen. Semiparametric Cluster Detection, 2007. Ph.D. Dissertation, University of Maryland, College Park.

[67] R. L. Wolpert. Lecture notes on stationary gamma processes. *arXiv preprint arXiv:2106.00087*, 2021.

[68] L. Yu. Two Goodness-of-Fit Tests for the Density Ratio Model, 2017. Ph.D. Dissertation, University of Maryland, College Park.

[69] A. G. Zhang. Semiparametric Inferences under a Density Ratio Model, 2022. Ph.D. Dissertation, The University of British Columbia.

[70] A. G. Zhang and J. Chen. Density ratio model with data-adaptive basis function. *Journal of Multivariate Analysis*, 191:105043, 2022.

[71] A. G. Zhang, G. Zhu, and J. Chen. Empirical likelihood ratio test on quantiles under a density ratio model. *Electronic Journal of Statistics*, 15(2):6191–6227, 2021.

[72] B. Zhang. A chi-squared goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*, 86(3):531–539, 1999.

[73] B. Zhang. A goodness of fit test for multiplicative-intercept risk models based on case-control data. *Statistica Sinica*, pages 839–865, 2000.

[74] B. Zhang. M-estimation under a two-sample semiparametric model. *Scandinavian journal of statistics*, 27(2):263–280, 2000.

[75] B. Zhang. Quantile estimation under a two-sample semi-parametric model. *Bernoulli*, 6(3):491–511, 2000.

[76] B. Zhang. Assessing goodness-of-fit of generalized logit models based on case-control data. *Journal of multivariate analysis*, 82(1):17–38, 2002.

[77] X. Zhang and B. Kedem. Extended residual coherence with a financial application. *Statistics in Transition new series*, 22(2):1–14, 2021.

[78] X. Zhang, S. Pyne, and B. Kedem. Estimation of residential radon concentration in Pennsylvania counties by data fusion. *Applied Stochastic Models in Business and Industry*, 36(6):1094–1110, 2020.

[79] X. Zhang, S. Pyne, and B. Kedem. Model selection in radon data fusion. *Statistics in Transition New Series*, 21(4):159–165, 2020.

[80] X. Zhang, S. Pyne, and B. Kedem. Multivariate Tail Probabilities: Predicting Regional Pertussis Cases in Washington State. *Entropy*, 23(6):675, 2021.

[81] W. Zhou. Out-of-Sample Fusion, 2013. Ph.D. Dissertation, University of Maryland, College Park.