THE UNIVERSITY OF CHICAGO

ESTIMATION AND INFERENCE FOR HIGH DIMENSIONAL TIME SERIES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS

BY
DANNA ZHANG

CHICAGO, ILLINOIS
AUGUST 2017

To my parents, Wenzhao Zhang and Youping Qian

"You are not a beautiful and unique snowflake. You are the same decaying organic matter as everyone else, and we are all part of the same compost pile." – Tyler Durden, *Fight Club*.

TABLE OF CONTENTS

LIST OF FIGURES vi

ACKNOWLEDGMENTS vii

ABSTRACT viii

1 INTRODUCTION 1

2 HIGH DIMENSIONAL STATIONARY PROCESSES 5
- 2.1 Introduction 5
- 2.2 High-dimensional Stationary Processes 7
- 2.3 Gaussian Approximations for Sample Mean Vectors 9
- 2.4 Simultaneous Inference of Covariances 13
- 2.5 A Uniform Test for Distributions of Time Series 15
- 2.6 Estimation of Long-run Covariance Matrices 22
  - 2.6.1 Computing Approximated Cutoff Values 24
- 2.7 Simulation Study 26
- 2.8 Inequalities for High-dimensional Time Series with Finite Polynomial Moments 30
- 2.9 Auxiliary Results 34
  - 2.9.1 Inequalities with Finite Polynomial Moments 34
  - 2.9.2 Inequalities with Finite Exponential Moments 36
  - 2.9.3 Inequalities for Sums of High-dimensional Random Vectors 37
- 2.10 Deferred Proofs 38
  - 2.10.1 An error bound of the Gaussian approximation 39
  - 2.10.2 Proofs of Theorem 2.3.2 and Theorem 2.3.3 42
  - 2.10.3 Proofs of Results in Section 2.6 43
  - 2.10.4 Proofs of Results in Section 2.10.1 49

3 HIGH DIMENSIONAL LOCALLY STATIONARY PROCESSES 53
- 3.1 Introduction 53
- 3.2 High Dimensional Locally Stationary Processes 54
- 3.3 Estimation of Autocovariance Matrix Functions 58
- 3.4 Spectral Density and Coherence Matrix 63
  - 3.4.1 Estimation of Long-run Covariance Matrix Functions 67
- 3.5 Graphical Model and Inverse Spectral Density Matrix 73
- 3.6 Hanson–Wright-Type Inequalities 76
- 3.7 Deferred Proofs 82
  - 3.7.1 Proofs of Corollaries 3.3.2, 3.4.2 and 3.4.5 82
  - 3.7.2 Some Lemmas used in this chapter 87
  - 3.7.3 Proofs of Theorem 3.6.2 and Theorem 3.6.3 90

REFERENCES 106

LIST OF FIGURES

2.1 QQ-plots for Gaussian approximation with heavier tails 28
2.2 QQ-plots for Gaussian approximation with lighter tails 29

ACKNOWLEDGMENTS

I would like to thank my advisors, Wei Biao Wu and Peter McCullagh, for their careful supervision, thoughtful encouragement, great patience and strong support over the last five years. They have always been readily available with their time and my conversation with them has always been inspiring. Han Yu, a poet of the Tang dynasty, said: “The scholar should disseminate ideas, impart knowledge and dispel confusion.” My advisors exactly fit the bill. They taught me to challenge and refine myself at a higher level and with a broader perspective. I would not be where I am now without them and I can never thank them enough for all that they have done for me.

I would like to thank my committee member Per Mykland for his deep insight, strong support and helpful feedback on my work. I thank Stephen Stigler, Steve Lalley, Michael Stein, Dan Nicolae, Mei Wang, Jian Ding, Tracy Ke and Chao Gao for contributing to my experience here and helping me better prepare for the next stage in my academic career. I also thank the faculty, staff and my fellow students who I have met in Chicago, especially Mengyu Xu, Si Tang, Yuancheng Zhu, Miaoyan Wang, Luyi Yang and Mengyin Lu. I am grateful to Tianxiao Pang at Zhejiang University for his willingness to continue working with me from half a world away.

I would like to thank my best friend, Yuefeng Han. Thanks for dreaming with me, making me strong and pushing me to be my best possible self. His companion in my darkest days is always cherished. He is like my brother and is always there for me.

I would like to thank my fiance, Zhipeng Lou. My life forever changed after meeting him. It is him who shows me that love is something to be done; something not just to be said, but also to be shown. I look forward to the future that awaits us.

Most of all, I would like to thank my parents, Wenzhao Zhang and Youping Qian, for all of their love and support. They have played and continue to play a role in my life like none other. They are my role models for always being passionate about work, family and life.

ABSTRACT

There is a well-developed asymptotic theory for sample means and sample second-order statistics of low dimensional stationary processes. However, many important problems on their asymptotic behaviors are still unanswered for time series which can be high-dimensional, nonstationary and non-Gaussian.

This thesis concerns the estimation and inference of high-dimensional time series under the framework of functional dependence measure. We first consider the problem of approximating sums of high dimensional stationary time series by Gaussian vectors. We also consider an estimator for long-run covariance matrices and study its convergence properties. Our results allow constructing simultaneous confidence intervals for mean vectors of high-dimensional time series with asymptotically correct coverage probabilities. As an application, we can do simultaneous inferences for covariance matrices of high-dimensional stationary time series. We also propose a Kolmogorov-Smirnov type statistic for testing distributions of high-dimensional time series.

This thesis also presents a systematic asymptotic theory for the estimates of time-varying second-order statistics for a general class of high-dimensional nonstationary processes. In particular, we investigate the estimation of time-varying autocovariance matrix functions, spectral density matrices and coherence matrices for high-dimensional locally stationary processes. Besides, we use the constrained ℓ_1 minimization approach to estimate the inverse of the spectral density matrix which can be used to identify the graphical structure for high-dimensional locally stationary processes. We derive the convergence rates of the estimates which depend on the sample size, the dimension, the moment condition and the dependence of the underlying processes.

CHAPTER 1

INTRODUCTION

During the past several decades, there has been a well-developed theory for low dimensional stationary processes. However, the assumption of low dimensionality and stationarity may not be adequate for some practical applications. High dimensional and/or nonstationary time series analysis has gained credibility recently in finance, signal processing, neuroscience, meteorology, seismology and many other areas.

As an important class of non-stationary processes, locally stationary processes have attracted considerable attention in the past years. Different approaches for modelling locally stationary processes have been developed. For example, Dahlhaus [1997, 2000a] adopted a time-varying spectral representation, which was first studied in detail by Priestley [1965, 1982, 1988a]. Mallat et al. [1998] considered processes whose covariance operators are time-varying convolutions. Another method of modelling is to approximate non-stationary processes by piecewise stationary processes; see Adak [1998] and Ombao et al. [2005]. Other notable work includes Nason et al. [2000], Moulines et al. [2005] and more recently Zhou [2010] and Vogt [2012]; see Dahlhaus [2012] for a comprehensive overview.

Parametric locally stationary processes with time-varying coefficients have been intensively studied; see, for example, time-varying AR models (Subba Rao [1970], Dahlhaus [1997], Moulines et al. [2005]), ARMA models (Grenier [1983], Dahlhaus and Polonik [2009]), ARCH and GARCH models (Dahlhaus and Subba Rao [2006, 2007], Hafner and Linton [2010], Fryzlewicz and Subba Rao [2011]). Despite the advantage of ease of interpretation and simplicity of prediction using parametric models, one may draw erroneous conclusions of the model suffering from misspecification. In view of this, we consider nonparametric locally stationary processes. Let $(X_{t,n})_{t=1}^{n}$ be the observed sequence generated from the model

$$X_{t,n} = G(t/n, \mathcal{F}^t) = (X_{t1,n}, \dots, X_{tp,n})^\top, \tag{1.0.1}$$

where $\mathcal{F}^t = (\ldots, \varepsilon_{t-1}, \varepsilon_t)$ and ε_t, $t \in \mathbb{Z}$ are i.i.d. random elements, $^\top$ denotes matrix transpose, $G(\cdot,\cdot) = (g_1(\cdot,\cdot), \ldots, g_p(\cdot,\cdot))^\top$ is an $\mathbb{R}^p$-valued measurable function such that $X_t(u) = G(u, \mathcal{F}^t)$ is a well-defined random vector and the uniform stochastic Lipschitz continuity holds: there exists $\mathcal{K} > 0$ for which

$$\max_{1\le j\le p} \|g_j(u, \mathcal{F}^t) - g_j(v, \mathcal{F}^t)\| \le \mathcal{K}|u - v|, \text{ for all } u, v \in [0,1], \tag{1.0.2}$$

where $\|X\| = (\mathbb{E}X^2)^{1/2}$ for a random variable X. In the scalar case with $p = 1$, Zhou and Wu [2009] considered the estimation of quantile curves under this framework and Zhou [2010] performed nonparametric specification tests of quantile curves. Wu and Zhou [2011] obtained an Gaussian approximation result on partial sums of the process when p is fixed. If $G(u,\cdot)$ does not depend on u, then (1.0.1) becomes $X_t = G(\mathcal{F}^t)$, which defines a large class of high-dimensional stationary processes. Under this framework, Chen et al. [2013] quantified the convergence rates in covariance and precision matrix estimation and extended it to the locally stationary case. Wu and Wu [2016] studied properties of estimates of the regression parameters in high-dimensional linear models with dependent covariates and errors.

We aim to perform simultaneous inference for mean vectors of high-dimensional stationary processes in Chapter 2. We first consider the problem of approximating sums of high dimensional stationary time series by Gaussian vectors. To perform statistical inference based on the Gaussian approximation result, one needs to estimate the long-run covariance matrix. The latter problem has been extensively studied in the scalar and the low-dimensional case; see Newey and West [1987], Politis et al. [1999], Bühlmann [2002], Lahiri [2003], Alexopoulos and Goldsman [2004], among others. We study the batched-mean estimate of long-run covariance matrices and derive a large deviation result about quadratic forms of stationary processes. The latter tail probability inequalities allow dependent and/or non-sub-Gaussian processes under mild conditions, which are expected to be useful in other high-dimensional inference problems for dependent vectors. Our results allow constructing simultaneous confidence intervals for mean vectors of high-dimensional time series with asymptotically correct coverage probabilities. As an application, we can do simultaneous inferences for covariance matrices of high-dimensional stationary time series. We also propose a Kolmogorov-Smirnov type statistic for testing distributions of high-dimensional time series.

The primary goal of Chapter 3 is to estimate second-order characteristics of a general class of locally stationary processes which can be possibly high-dimensional and non-Gaussian, and lay a theoretical foundation for estimation consistency. We first concern the estimation of time-varying autocovariance matrix functions and study the nonparametric estimation of time-varying spectral density and coherence matrices. In particular, we also introduce the overlapped batched mean estimate of long-run covariance matrix functions and use the constrained ℓ_1 minimization approach to estimate the inverse of the spectral density matrix which can be used to identify the graphical structure for high-dimensional locally stationary processes. We derive the convergence rates of the estimates which depend on the sample size, the dimension, the moment condition and the dependence of the underlying processes. We also provide Hanson–Wright-type inequalities for tail probabilities for non-stationary processes with finite polynomial moments, which could be quite useful in the estimation of second-order statistics.

We now introduce some notation. For a random variable X and $q \geq 1$, we write $X \in \mathcal{L}^q$ if $\|X\|_q := (\mathbb{E}|X|^q)^{1/q} < \infty$. Denote $\|X\| = \|X\|_2$ and the operator $\mathbb{E}_0$ with $\mathbb{E}_0(X) := X - \mathbb{E}X$. Define the projection operator $\mathcal{P}^t \cdot = \mathbb{E}(\cdot|\mathcal{F}^t) - \mathbb{E}(\cdot|\mathcal{F}^{t-1})$ where $\mathcal{F}^t = (\ldots, \varepsilon_{t-1}, \varepsilon_t)$. For a vector $v = (v_1, \ldots, v_p)^\top$ and $q \geq 1$, we define $|v|_q = (\sum_{j=1}^p |v_j|^q)^{1/q}$ and $|v|_\infty = \max_j |v_j|$. For a matrix $A = (a_{ij})_{i,j=1}^p$, define the elementwise ℓ_∞ norm $|A|_\infty = \max_{i,j} |a_{ij}|$ and the matrix ℓ_1 norm $|A|_{L_1} = \max_j \sum_i |a_{ij}|$. Write the $p \times p$ identity matrix as I_p. For an interval $\mathcal{I} \subset \mathbb{R}$, denote by $\mathcal{C}^i\mathcal{I}$, $i \in \mathbb{N}$, be the collection of functions that have i-th order continuous derivatives on $\mathcal{I}$. For two real numbers, set $x \vee y = \max(x, y)$ and $x \wedge y = \min(x, y)$. For two sequences of positive numbers (a_n) and (b_n), we write $a_n \asymp b_n$ (resp., $a_n \lesssim b_n$ or $a_n \ll b_n$) if there exists some constant $C > 0$ such that $C^{-1} \leq a_n/b_n \leq C$ (resp., $a_n/b_n \leq C$ or

$a_n/b_n \to 0$) for all large n. We use $C, C_1, C_2, \cdots$ to denote positive constants whose values may differ from place to place. A constant with a symbolic subscript is used to emphasize the dependence of the value on the subscript.

CHAPTER 2

HIGH DIMENSIONAL STATIONARY PROCESSES

2.1 Introduction

In this chapter, we shall consider simultaneous inference for mean vectors of high-dimensional stationary processes, so that one can perform family-wise multiple testing or construct simultaneous confidence intervals, an important problem in the analysis of spatial-temporal processes. To fix the idea, let (X_i) be a stationary process in $\mathbb{R}^p$ with mean $\mu = (\mu_1, \ldots, \mu_p)^\top$ and finite second moment in the sense that $\mathbb{E}(X_i^\top X_i) < \infty$. In the scalar case in which $p = 1$ or when p is fixed, under suitable weak dependence conditions, we can have the central limit theorem (CLT):

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \Rightarrow N(0, \Sigma), \text{ where } \Sigma = \sum_{k=-\infty}^{\infty} \mathbb{E}((X_0 - \mu)(X_k - \mu)^\top).$$

See, for example, Rosenblatt [1956], Ibragimov and Linnik [1971], Wu [2005], Dedecker et al. [2007], Bradley [2007] among others. In the high dimension case in which p can also diverge to infinity, Portnoy [1986] showed that the central limit theorem can fail for i.i.d. random vectors if $\sqrt{n} = o(p)$. We shall consider an alternative form: Gaussian approximation for the largest entry of the sample mean vector $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. For a vector $v = (v_1, \ldots, v_p)^\top$, let $|v|_\infty = \max_{j \le p} |v_j|$. Specifically, our primary goal is to establish the Gaussian Approximation (GA) in $\mathbb{R}^p$

$$\sup_{u \ge 0} |\mathbb{P}(\sqrt{n}|\bar{X}_n - \mu|_\infty \ge u) - \mathbb{P}(|Z|_\infty \ge u)| \to 0, \tag{2.1.1}$$

where both $n, p \to \infty$. Here, the Gaussian vector $Z = (Z_1, \ldots, Z_p)^\top \sim N(0, \Sigma)$. Chernozhukov et al. [2013a] studied the Gaussian approximation for independent random vectors. There has been limited research on high-dimensional inference under dependence. The associated statistical inference becomes considerably more challenging since the autocovariances with all lags should be considered. Zhang and Cheng [2014] extended the Gaussian approximation in Chernozhukov et al. [2013a] to very weakly dependent random vectors which satisfy a uniform geometric moment contraction condition. The latter condition is also adopted in Chen et al. [2015] for self-normalized sums. Chernozhukov et al. [2013b] did a similar extension to strong mixing random vectors. Here, we shall establish (2.1.1) for a wide class of high-dimensional stationary process under suitable conditions on the magnitudes of p, n and the mild dependence conditions on the process (X_i).

In Section 2.2, we shall introduce the framework of high-dimensional time series in detail and some concepts about functional dependence measures that are useful for establishing an asymptotic theory. The main result for Gaussian approximation of the normalized mean vector and the choice of the normalization matrix is presented in Section 2.3. Depending on the moment and the dependence conditions, both high dimension and ultra high dimension cases are discussed. In Section 2.4, we apply our Gaussian approximation result to simultaneous inference of entries of sample covariance matrices of high-dimensional time series. In Section 2.5, we shall develop a Kolmogorov–Smirnov-type statistic for testing distributions of high-dimensional time series.

To perform statistical inference based on (2.1.1), one needs to estimate the long-run covariance matrix Σ. The latter problem has been extensively studied in the scalar and the low-dimensional case; see Newey and West [1987], Politis et al. [1999], Bühlmann [2002], Lahiri [2003], Alexopoulos and Goldsman [2004], among others. In Section 2.6, we study the batched-mean estimate of long-run covariance matrices and derive a large deviation result about quadratic forms of stationary processes. The latter tail probability inequalities allow dependent and/or non-sub-Gaussian processes under mild conditions, which are expected to be useful in other high-dimensional inference problems for dependent vectors. The consistency of the batched-mean estimate ensures the validity of the quantile estimates of $\mathcal{L}^\infty$ norms of sample means; see Section 2.6.1.

We provide in Section 2.7 a simulation study. Section 2.8 includes some sharp inequalities for tail probabilities for high dimensional dependent processes in the polynomial tail case. The readers are referred to Section 2.9 for the tail probability inequalities in the one-dimensional case under finite polynomial moment and exponential moment conditions, respectively. Part of the proofs are relegated to Section 2.10.

2.2 High-dimensional Stationary Processes

Let $\varepsilon_i, i \in \mathbb{Z}$, be i.i.d. random elements and $\mathcal{F}^i = (\ldots, \varepsilon_{i-1}, \varepsilon_i)$; let (X_i) be a stationary process taking values in $\mathbb{R}^p$ that assumes the form

$$X_i = (X_{i1}, X_{i2}, \ldots, X_{ip})^\top = G(\mathcal{F}^i), \tag{2.2.1}$$

where $G(\cdot) = (g_1(\cdot), \ldots, g_p(\cdot))^\top$ is an $\mathbb{R}^p$-valued measurable function such that X_i is well-defined. The model (2.2.1) is a special case of (1.0.1) where $G(u, \cdot)$ does not depend on u. In the scalar case with $p = 1$, (2.2.1) allows a very general class of stationary processes (cf. Wiener [1958], Rosenblatt [1971], Priestley [1988b], Tong [1990], Wu [2005], Tsay [2005], Wu [2011]). It includes linear processes as well as a large class of nonlinear time series models. For example, if $\varepsilon_i, i \in \mathbb{Z}$, are i.i.d. d-dimensional random vectors with mean 0 and $\mathbb{E}(\varepsilon_i^\top \varepsilon_i) < \infty$, and $A_i, i \geq 0$, are $p \times d$ coefficient matrices with real entries such that $\sum_{i=0}^{\infty} \operatorname{tr}(A_i^\top A_i) < \infty$, where $\operatorname{tr}(\cdot)$ denotes the trace of a matrix. Then by Kolmogorov's three-series theorem, the linear process

$$X_i = \sum_{l=0}^{\infty} A_l \varepsilon_{i-l} \tag{2.2.2}$$

exists, and it is of form (2.2.1) with a linear functional G. In particular, the vector AR(1) process $X_i = AX_{i-1} + \varepsilon_i$ has form (2.2.2) with $A_l = A^l$ if $\max_{j \leq p} |\lambda_j(A)| < 1$, where A is a coefficient matrix and $\lambda_1(A), \ldots, \lambda_p(A)$ are eigenvalues of A. Within this framework,

(ε_i) can be viewed as independent inputs of a physical system and all the dependencies among the outputs (X_i) result from the underlying data-generating mechanism $G(\cdot)$. The function $g_j(\cdot)$, $1 \le j \le p$, is the jth coordinate projection of $G(\cdot)$. Unless otherwise specified, assume throughout this chapter that $\mathbb{E}X_i = 0$ and $\max_{j\le p} \|X_{ij}\|_q < \infty$ for some $q \ge 2$. Let $\Gamma(l) = (\gamma_{jk}(l))_{j,k=1}^p = \mathbb{E}(X_i X_{i+l}^\top)$ be the autocovariance matrix and recall the long-run covariance matrix

$$\Sigma = (\sigma_{jk})_{j,k=1}^p = \sum_{l=-\infty}^{\infty} \Gamma(l) \tag{2.2.3}$$

if it exists. Note that $\sigma_{jj} = \sum_{l=-\infty}^{\infty} \gamma_{jj}(l)$, $1 \le j \le p$, is the long-run variance of the component process $X_{\cdot j} = (X_{ij})_{i\in\mathbb{Z}}$. For the latter process, following Wu [2005] we define the functional dependence measure:

$$\delta_{i,q,j} = \|X_{ij} - X_{ij,\{0\}}\|_q = \|X_{ij} - g_j(\mathcal{F}^{i,\{0\}})\|_q, \tag{2.2.4}$$

where $\mathcal{F}^{i,\{k\}} = (\ldots, \varepsilon_{k-1}, \varepsilon'_k, \varepsilon_{k+1}, \ldots, \varepsilon_i)$ is a coupled version of $\mathcal{F}^i$ with ε_k in $\mathcal{F}^i$ replaced by ε'_k, and $\varepsilon_i, \varepsilon'_l$, $i, l \in \mathbb{Z}$, are i.i.d. random elements. Note that $\mathcal{F}^{i,\{k\}} = \mathcal{F}^i$ if $k > i$. To account for the dependence in the process $X_{\cdot j}$, we define the dependence adjusted norm

$$\|X_{\cdot j}\|_{q,\alpha} = \sup_{m\ge 0} (m+1)^\alpha \Delta_{m,q,j},\ \alpha \ge 0,\ \text{where } \Delta_{m,q,j} = \sum_{i=m}^{\infty} \delta_{i,q,j}. \tag{2.2.5}$$

Due to the dependence, it may happen that $\max_{j\le p} \|X_{ij}\|_q < \infty$ while $\|X_{\cdot j}\|_{q,\alpha} = \infty$. Elementary calculations show that, if $X_{ij}, i \in \mathbb{Z}$, are i.i.d., then $\|X_{ij}\|_q \le \|X_{\cdot j}\|_{q,\alpha} \le 2\|X_{ij}\|_q$, suggesting that the dependence adjusted norm is equivalent to the classical $\mathcal{L}^q$ norm.

To account for high-dimensionality, we define

$$\Psi_{q,\alpha} = \max_{1\le j\le p} \|X_{\cdot j}\|_{q,\alpha} \text{ and } \Upsilon_{q,\alpha} = \left(\sum_{j=1}^{p} \|X_{\cdot j}\|_{q,\alpha}^q\right)^{1/q},$$

which can be interpreted as the uniform and the overall dependence adjusted norms of $(X_i)_{i\in\mathbb{Z}}$, respectively. The form (2.2.1) and its associated dependence measures provide a convenient framework for studying high-dimensional time series. Zhang and Cheng [2014] considered the special case which imposes the stronger geometric moment contraction condition $\max_{1\le j\le p} \Delta_{m,q,j} \le C\rho^m$ with $\rho \in (0,1)$ and some constant C. This assumption can be fairly restrictive. In this thesis $\Psi_{q,\alpha}$ can be unbounded in p. Additionally, we define the $\mathcal{L}^\infty$ functional dependence measure and its corresponding dependence adjusted norm for the p-dimensional stationary process (X_i)

$$\omega_{i,q} = \||X_i - X_{i,\{0\}}|_\infty\|_q;$$

$$\||X_\cdot|_\infty\|_{q,\alpha} = \sup_{m\ge 0} (m+1)^\alpha \Omega_{m,q}, \alpha \ge 0, \text{ where } \Omega_{m,q} = \sum_{i=m}^{\infty} \omega_{i,q}.$$

Clearly, we have $\Psi_{q,\alpha} \le \||X_\cdot|_\infty\|_{q,\alpha} \le \Upsilon_{q,\alpha}$. Throughout this chapter, we assume $p = p_n \to \infty$ as $n \to \infty$.

2.3 Gaussian Approximations for Sample Mean Vectors

In this section, we shall present main results on Gaussian approximations. Theorem 2.3.2 concerns the finite polynomial moment case with both weaker and stronger temporal dependence. If the underlying process has finite dependence adjusted sub-exponential norms, Theorem 2.3.3 asserts that an ultra-high dimension p can be allowed. Theorem 2.10.4 in Section 2.10.1 provides a convergence rate of the Gaussian approximation.

Recall (2.2.3) for the long-run covariance matrix Σ. Let $\Sigma_0 = \mathrm{diag}(\Sigma)$ be the diagonal matrix of Σ, and $D_0 = \mathrm{diag}(\sigma_{11}^{1/2}, \ldots, \sigma_{pp}^{1/2}) = \Sigma_0^{1/2}$. Assume $\mu = 0$. We consider the following normalized version of (2.1.1):

$$\rho_n := \sup_{u \geq 0} |\mathbb{P}(\sqrt{n}|D_0^{-1}\bar{X}_n|_\infty \geq u) - \mathbb{P}(|D_0^{-1}Z|_\infty \geq u)| \to 0, \tag{2.3.1}$$

Assumption 2.3.1. *There exists a constant* $c > 0$ *such that* $\min_{1 \leq j \leq p} \sigma_{jj} \geq c$.

To state Theorem 2.3.2, we need to define the following quantities:

$$\begin{aligned}
\Theta_{q,\alpha} &= \Upsilon_{q,\alpha} \wedge (\||X_\cdot|_\infty\|_{q,\alpha}(\log p)^{3/2}),\ L_1 = (\Psi_{2,\alpha}\Psi_{2,0}(\log p)^2)^{1/\alpha}, \\
W_1 &= (\Psi_{3,0}^6 + \Psi_{4,0}^4)(\log(pn))^7,\ W_2 = \Psi_{2,\alpha}^2(\log(pn))^4, \\
W_3 &= (n^{-\alpha}(\log(pn))^{3/2}\Theta_{q,\alpha})^{1/(1/2-\alpha-1/q)}, \\
N_1 &= (n/\log p)^{q/2}/\Theta_{q,\alpha}^q,\ N_2 = n(\log p)^{-2}\Psi_{2,\alpha}^{-2}, \\
N_3 &= (n^{1/2}(\log p)^{-1/2}\Theta_{q,\alpha}^{-1})^{1/(1/2-\alpha)}.
\end{aligned}$$

Theorem 2.3.2. *Let Assumption 2.3.1 be satisfied. (i) Assume that* $\Theta_{q,\alpha} < \infty$ *holds with some* $q \geq 4$ *and* $\alpha > 1/2 - 1/q$ *(the weaker dependence case),*

$$\Theta_{q,\alpha} n^{1/q-1/2}(\log(pn))^{3/2} \to 0 \tag{2.3.2}$$

and

$$L_1 \max(W_1, W_2) = o(1) \min(N_1, N_2). \tag{2.3.3}$$

Then the Gaussian approximation (2.3.1) holds. (ii) Assume $0 < \alpha < 1/2 - 1/q$ *(the stronger dependence case). Then (2.3.1) holds if* $\Theta_{q,\alpha}(\log p)^{1/2} = o(n^\alpha)$ *and*

$$L_1 \max(W_1, W_2, W_3) = o(1) \min(N_2, N_3). \tag{2.3.4}$$

Remark 1. A careful check of the proof of Theorem 2.3.2 indicates that if it is further assumed that $\max_{1\le j\le p}\sigma_{jj}$ is bounded from above, the Gaussian approximation is also valid for the nonnormalized maximum, that is, for both cases of Theorem 2.3.2,

$$\sup_{u\ge 0}|\mathbb{P}(\sqrt{n}|\bar{X}_n|_\infty \ge u) - \mathbb{P}(|Z|_\infty \ge u)| \to 0. \tag{2.3.5}$$

Remark 2. *(Optimality of our result on the allowed dimension p)* Assume $\alpha > 1/2 - 1/q$. In the special case with $\Psi_{q,\alpha} \asymp 1$ and $\Theta_{q,\alpha} \asymp p^{1/q}$, (2.3.2) becomes

$$p(\log(pn))^{3q/2} = o(n^{q/2-1}), \tag{2.3.6}$$

which by elementary manipulations implies (2.3.3), and hence the GA (2.3.1). It turns out that condition (2.3.6), or equivalently $p(\log p)^{3q/2} = o(n^{q/2-1})$, is optimal up to a multiplicative logarithmic term. Consider the special case in which X_{ij}, $i,j\in\mathbb{Z}$, are i.i.d. symmetric random variables with $\mathbb{E}(X_{ij}^2) = 1$ and the tail probability $\mathbb{P}(X_{ij}\ge u) = u^{-q}\ell(u)$, $u\ge u_0$, where $\ell(u) = (\log u)^{-2}$. By Theorem 1.9 of Nagaev [1979], we have the expansion: for a sequence $y_n \ge \sqrt{n}$, as $n\to\infty$,

$$\frac{\mathbb{P}(X_{11}+\ldots+X_{n1}\ge y_n)}{n y_n^{-q}\ell(y_n) + 1 - \Phi(y_n/\sqrt{n})} \to 1. \tag{2.3.7}$$

Let $M_n = X_{11}+\ldots+X_{n1}$, $Z = (Z_1,\ldots,Z_p)^\top \sim N(0,\mathrm{I}_p)$ and assume

$$n^{q/2-1} = o(p(\log n)^{-2}(\log p)^{-q/2}). \tag{2.3.8}$$

Then the GA (2.3.1) *does not hold.* To see this, let $u = (2\log p)^{1/2}$. Then $p\mathbb{P}(|Z_1|\ge u)\to 0$, and, by (2.3.7) and (2.3.8), $p\mathbb{P}(M_n \ge \sqrt{n}u)\to\infty$. Hence, $\mathbb{P}^p(|M_n|\le\sqrt{n}u)\to 0$ and $\mathbb{P}^p(|Z_1|\le u)\to 1$, implying that

$$\begin{aligned}\rho_n &\ge |\mathbb{P}(\sqrt{n}|\bar{X}_n|_\infty \le u) - \mathbb{P}(|Z|_\infty\le u)| \\ &= |\mathbb{P}^p(|M_n|\le\sqrt{n}u) - \mathbb{P}^p(|Z_1|\le u)|\end{aligned}$$

$$= \quad |[1 - 2\mathbb{P}(M_n \geq \sqrt{n}u)]^p - \mathbb{P}^p(|Z_1| \leq u)| \to 1.$$

Note that (2.3.8) is equivalent to $n^{q/2-1} = o(p(\log p)^{-2-q/2})$, suggesting that (2.3.6) is optimal up to a logarithmic term. $\square$

Now suppose there exist $0 \leq \kappa_1 \leq \kappa_2$ such that $\Psi_{q,\alpha} \asymp p^{\kappa_1}$ and $\Theta_{q,\alpha} \asymp p^{\kappa_2}$, and $p^\tau \asymp n$. Elementary but tedious calculations show that, in the weaker dependence case $\alpha > 1/2 - 1/q$, if

$$\tau > \max\left\{\frac{\kappa_2}{1/2 - 1/q}, \frac{2\kappa_1}{\alpha} + 8\kappa_1, \frac{2}{q}\left(\frac{2\kappa_1}{\alpha} + 8\kappa_1\right) + 2\kappa_2\right\}, \tag{2.3.9}$$

then conditions in (i) of Theorem 2.3.2 are satisfied, while for the stronger dependence case with $0 < \alpha < 1/2 - 1/q$, a larger sample size n is required:

$$\tau > \max\left\{\frac{\kappa_2}{\alpha}, \frac{2\kappa_1}{\alpha} + 8\kappa_1, (1 - 2\alpha)\left(\frac{2\kappa_1}{\alpha} + 8\kappa_1\right) + 2\kappa_2\right\}. \tag{2.3.10}$$

The lower bounds in (2.3.9) and (2.3.10) are both nondecreasing of κ_1, κ_2 and nonincreasing in q, α.

Next we consider the sub-exponential case in which X_{ij} satisfies a stronger moment condition than the existence of finite qth moment. Assume that X_{ij} has finite moment with any order. For $\nu \geq 0$ and $\alpha \geq 0$, define the dependence adjusted sub-exponential norm

$$\|X_{\cdot j}\|_{\psi_\nu,\alpha} = \sup_{q \geq 2} \frac{\|X_{\cdot j}\|_{q,\alpha}}{q^\nu} \text{ and } \Phi_{\psi_\nu,\alpha} = \max_{j \leq p} \|X_{\cdot j}\|_{\psi_\nu,\alpha}$$

By this definition, if X_{ij}, $i \in \mathbb{Z}$ are i.i.d., $\|X_{\cdot j}\|_{\psi_\nu,\alpha}$ is equivalent to the sub-Gaussian norm ($\nu = 1$) or sub-exponential norm ($\nu = 1/2$), due to the equivalence of $\|X_{\cdot j}\|_{q,\alpha}$ and $\|X_{ij}\|_q$. The parameter ν measures how fast $\|X_{\cdot j}\|_{q,\alpha}$ increases with q.

To state Theorem 2.3.3, we let $\beta = 2/(1+2\nu)$ and define

$$\begin{aligned} L_2 &= ((\log p)^{1/\beta+1/2}\Phi_{\psi_\nu,\alpha})^{1/\alpha}, \ N_4 = n(\log p)^{-1-2/\beta}\Phi_{\psi_\nu,0}^{-2}, \\ W_4 &= (\log(pn))^{3+2/\beta}\Phi_{\psi_\nu,0}^{2} + (\log(pn))^4. \end{aligned}$$

Theorem 2.3.3. *Let Assumption 2.3.1 be satisfied. Assume that* $\Phi_{\psi_\nu,\alpha} < \infty$ *for some* $\nu \geq 0$, $\alpha > 0$ *and*

$$\max(L_1, L_2)\max(W_1, W_4) = o(N_4), \quad L_1^\alpha \max(W_1, W_4) = o(n). \tag{2.3.11}$$

Then the Gaussian approximation (2.3.1) holds.

If $\Phi_{\psi_\nu,\alpha} \asymp 1$, then the ultra high-dimensional case with $\log p = o(n^c)$ with some $c > 0$ is allowed, where specifically we can let

$$c = \begin{cases} 1/(8+2/\alpha+2/\beta), & 2/3 \leq \beta \leq 2 \\ 1/[7+(1/\beta+1/2)(1/\alpha+2)], & 1/2 \leq \beta < 2/3 \\ 1/[3+2/\beta+(1/\beta+1/2)(1/\alpha+2)], & 0 < \beta < 1/2 \end{cases}. \tag{2.3.12}$$

2.4 Simultaneous Inference of Covariances

Let $X_1, \ldots, X_n$ be i.i.d. p-dimensional vectors with mean 0 and covariance matrix $\Gamma_0 = (\gamma_{jk})_{j,k=1}^p = \mathbb{E}(X_i X_i^\top)$. We estimate Γ_0 by the sample covariance matrix $\hat{\Gamma}_0 = (\hat{\gamma}_{jk})_{j,k=1}^p = n^{-1}\sum_{i=1}^n X_i X_i^\top$. To perform simultaneous inference on $\gamma_{jk}, 1 \leq j, k \leq p$, one needs to derive the asymptotic distribution of the maximum deviation $\max_{j,k\leq p}|\hat{\gamma}_{jk} - \gamma_{jk}|$ or the normalized version $\max_{j,k\leq p}|\hat{\gamma}_{jk} - \gamma_{jk}|/\tau_{jk}$; cf. equation (2) in Xiao and Wu [2013]. The former is also referred to as the mutual coherence of the data matrix in the compressed sensing literature (see, e.g., Donoho et al. [2006]). Jiang [2004] established the Gumbel convergence of the maximum deviation under some polynomial moment condition and under the setup that all entries of X_i are also independent. See Li and Rosalsky [2006], Zhou [2007], Liu et al. [2008] and Li et al. [2010] for some refined results. Cai and Jiang [2011] showed that $\max_{|j-k|>s_n} |\hat{\gamma}_{jk} - \gamma_{jk}|$ also converges to the Gumbel distribution if $(X_{ij})_{1\le j\le p}$ is Gaussian and s_n-dependent for each i. Xiao and Wu [2013] considered the extension to the non-Gaussian case and allowed a general dependence structure among entries of X_i. However, the latter two paper both require that the vectors $X_1, \ldots, X_n$ are i.i.d. The problem of further extension to temporally dependent X_i is open. In analyzing fMRI functional connectivity in brain networks in the format of multivariate time series, researchers use the maximum correlation between time series to identify edges that connect the corresponding nodes in a network (cf. Hipp et al. [2012], Deco et al. [2013], Hutchison et al. [2013], Larson-Prior et al. [2013], among many others). Such applications suggest that an asymptotic theory for maximum deviations of sample covariances is needed.

Our Theorems 2.3.2 and 2.3.3 can be applied to the above problem of further extension to temporally dependent processes. Let (X_i) be a mean zero p-dimensional stationary process of form (2.2.1). To apply Theorems 2.3.2 and 2.3.3, one needs to deal with the key issue of computing the functional dependence measure of the p^2-dimensional vector $\mathcal{X}_i = \mathrm{vec}(X_i X_i^\top - \mathbb{E}(X_i X_i^\top))$. Interestingly, our framework allows a natural and elegant treatment. Let $a = (j,k)$, $j,k \le p$ and $\mathcal{X}_{ia} = X_{ij}X_{ik} - \gamma_a$, where $\gamma_a = \mathbb{E}(X_{ij}X_{ik})$. By Hölder's inequality, the functional dependence of the component process $(\mathcal{X}_{ia})_i$:

$$\begin{aligned} \varphi_{i,q/2,a} &:= \|X_{ij}X_{ik} - \mathbb{E}(X_{ij}X_{ik}) - X_{ij,\{0\}}X_{ik,\{0\}} + \mathbb{E}(X_{ij,\{0\}}X_{ik,\{0\}})\|_{q/2} \\ &\le 2\|X_{ij}X_{ik} - X_{ij,\{0\}}X_{ik,\{0\}}\|_{q/2} \\ &\le 2\|X_{ij}(X_{ik} - X_{ik,\{0\}})\|_{q/2} + 2\|(X_{ij} - X_{ij,\{0\}})X_{ik,\{0\}}\|_{q/2} \\ &\le 2\|X_{ij}\|_q \delta_{i,q,k} + 2\|X_{ik}\|_q \delta_{i,q,j}. \end{aligned} \tag{2.4.1}$$

Hence, we can have an upper bound of the dependence adjusted norm of $(\mathcal{X}_{ia})$

$$\begin{aligned} \|\mathcal{X}_{\cdot a}\|_{q/2,\alpha} &:= \sup_{m\ge 0} (m+1)^\alpha \sum_{i=m}^{\infty} \varphi_{i,q/2,j,k} \\ &\le 2\|X_{\cdot j}\|_{q,0}\|X_{\cdot k}\|_{q,\alpha} + 2\|X_{\cdot k}\|_{q,0}\|X_{\cdot j}\|_{q,\alpha}. \end{aligned} \tag{2.4.2}$$

Consequently, the uniform and the overall dependence adjusted norms of $\mathcal{X}_i$ are

$$\begin{aligned}
&\max_a \|\mathcal{X}_{\cdot a}\|_{q/2,\alpha} \leq 4\Psi_{q,0}\Psi_{q,\alpha}, \\
&\left(\sum_a \|\mathcal{X}_{\cdot a}\|_{q/2,\alpha}^{q/2}\right)^{2/q} \leq 4\left(\sum_{j=1}^p \|X_{\cdot j}\|_{q,0}^{q/2}\right)^{2/q}\left(\sum_{j=1}^p \|X_{\cdot j}\|_{q,\alpha}^{q/2}\right)^{2/q}.
\end{aligned} \tag{2.4.3}$$

Similarly, the $\mathcal{L}^\infty$ dependence adjusted norm for the process $(\mathcal{X}_i)$ can be calculated by

$$\||\mathcal{X}_\cdot|_\infty\|_{q/2,\alpha} \leq 4\||X_\cdot|_\infty\|_{q,0}\||X_\cdot|_\infty\|_{q,\alpha}. \tag{2.4.4}$$

With (2.4.1)-(2.4.4), conditions in Theorems 2.3.2 and 2.3.3 can be formulated accordingly, and under those conditions we can have the following Gaussian approximation:

$$\sup_{u\geq 0} |\mathbb{P}(\sqrt{n}\max_a |\hat{\gamma}_a - \gamma_a|/\tau_a \geq u) - \mathbb{P}(\max_a |Z_a/\tau_a| \geq u)| \to 0, \tag{2.4.5}$$

where $Z = (Z_a)_a \sim N(0, \Sigma_{\mathcal{X}})$, $\Sigma_{\mathcal{X}}$ is the $p^2 \times p^2$ long-run covariance matrix of $(\mathcal{X}_i)_i$ and $(\tau_a^2)_a$ is the diagonal matrix of $\Sigma_{\mathcal{X}}$.

2.5 A Uniform Test for Distributions of Time Series

In this section, we shall apply the Gaussian approximation result Theorem 2.3.2 and test distributions of time series. For the process (X_i) defined in (2.2.1), let $F_j(u) = \mathbb{P}(X_{ij} \leq u)$, $u \in \mathbb{R}$, be the cumulative distribution function (c.d.f.) of X_{ij}, $1 \leq j \leq p$; let $F_{j,0}(\cdot)$ be the reference c.d.f. We are interested in testing the null hypothesis:

$$H_0: F_j(\cdot) = F_{j,0}(\cdot) \text{ for all } j = 1, \ldots, p. \tag{2.5.1}$$

In the classical Kolmogorov–Smirnov test with $p = 1$ and i.i.d. data X_{i1}, $i \in \mathbb{Z}$, one uses a test statistic that involves the supremum distance between the empirical and the reference c.d.f.s. Here, we shall apply a smoothing procedure and consider testing an equivalent form of (2.5.1). In particular, we let $h(u) = H'(u)$ be a probability density function (p.d.f.) such that $h(u) > 0$ for all $u \in \mathbb{R}$, $\sup_u h(u) < \infty$ and let

$$H_j(u) = \int_{\mathbb{R}} F_j(v)h(u-v)dv \text{ and } H_{j,0}(u) = \int_{\mathbb{R}} F_{j,0}(v)h(u-v)dv. \tag{2.5.2}$$

For example, we can let $h(\cdot)$ be the standard Gaussian p.d.f. In this case, $H_j(\cdot)$ is the c.d.f. of $X_{ij} + \eta$, where $\eta \sim N(0,1)$ is independent of X_{ij}. Here, we shall consider testing the following equivalent form of (2.5.1):

$$H_0 : H_j(\cdot) = H_{j,0}(\cdot) \text{ for all } j = 1, \ldots, p, \tag{2.5.3}$$

by using the goodness-of-fit test statistic of the form $\sup_{u \in \mathcal{I}} |\hat{H}_j(u) - H_{j,0}(u)|$, where $\mathcal{I} \subset \mathbb{R}$ is an interval and $\hat{H}_j(u)$ is an unbiased estimate of $H_j(u)$:

$$\hat{H}_j(u) = \frac{1}{n}\sum_{i=1}^{n} H(u - X_{ij}). \tag{2.5.4}$$

Similar smoothing ideas appeared in the literature. Researchers applied kernel smoothing to overcome the shortcoming of discontinuity of empirical distribution functions; see, for example, Yamato [1973], Azzalini [1981], Reiss [1981], Falk [1985], Cheng and Peng [2002], Wang et al. [2013], among others.

Here, we shall develop a Gaussian approximation theory for

$$\Delta_n := \max_{1 \le j \le p} \sup_{u \in \mathcal{I}} \sqrt{n}|\hat{H}_j(u) - H_j(u)|. \tag{2.5.5}$$

To this end, we shall carry out a detailed calculation for the functional dependence measures defined in Section 2.2 of $H(u - X_{ij})$. For presentational clarity here, we only consider marginal distributions and linear processes (X_i) defined in (2.2.2). We remark that our approach also applies to testing for joint distributions and for nonlinear processes.

Assumption 2.5.1. *The process* (X_i) *is of form (2.2.2) with* $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{id})^\top$, *where* ε_{ij} *are i.i.d. with mean* 0 *and* $\|\varepsilon_{ij}\|_\gamma < \infty$, $\gamma > 2$; *and coefficient matrices* $A_i = (a_{i,jk})_{j \le p, k \le d}$ *satisfy* $\sum_{i=0}^{\infty} \operatorname{tr}(A_i^\top A_i) < \infty$.

For $j, k = 1, \ldots, p$ and $u, v \in \mathbb{R}$, define the long-run covariance function

$$\sigma_{j,k}(u, v) = \sum_{l=-\infty}^{\infty} \operatorname{Cov}(H(u - X_{0j}), H(v - X_{lk})). \tag{2.5.6}$$

Let $\{Z_j(u), j = 1, \ldots, p; u \in \mathbb{R}\}$ be a mean 0 Gaussian process such that its covariance function is given by (2.5.6).

Assumption 2.5.2. *There exists a constant* $c > 0$ *and a closed finite interval* $\mathcal{I} \subset \mathbb{R}$ *such that* $\min_{1 \le j \le p} \min_{u \in \mathcal{I}} \sigma_{j,j}(u, u) \ge c$.

Theorem 2.5.3. *Let Assumptions 2.5.1 and 2.5.2 be satisfied, and suppose there exists a constant* $C_1 > 0$ *such that for all* $m \ge 0$,

$$\sum_{i=m}^{\infty} \left(\sum_{k=1}^{d} \max_j |a_{i,jk}|^2 \right)^{\min(\gamma/q, 1)/2} \le C_1 (1 \vee m)^{-\alpha} \tag{2.5.7}$$

holds for some $q \ge 4$ *and* $\alpha > 0$. *Let* $\iota = \min(\gamma/q, 1)/2$. *There exists some constant* $\kappa > 0$ *depending on* α *and* ι *such that if* p *satisfies*

$$\log p = o(n^\kappa), \tag{2.5.8}$$

we have

$$\sup_{u \ge 0} \left| \mathbb{P}(\sqrt{n}\Delta_n \ge u) - \mathbb{P}\left(\max_{1 \le j \le d} \sup_{x \in \mathcal{I}} |Z_j(x)| \ge u \right) \right| \to 0. \tag{2.5.9}$$

Remark 3. *A careful check of the proof of Theorem 2.5.3 indicates that, for the index* κ *in (2.5.8), we can let* $\kappa = \kappa_1 = [(2\iota + 2)/\alpha + 8\iota + 11]^{-1}$ *if* $\alpha > 1/2 - 1/q$, *and* $\kappa =$

$\min(\kappa_1, \alpha/(3+\iota))$ *if* $0 < \alpha < 1/2 - 1/q$.

For i.i.d. random vectors, Kosorok and Ma [2007] considered uniform convergence of empirical distribution functions. Theorem 2.5.3 might be the first result in the literature concerning weak convergence of empirical processes in the high-dimensional setting under dependence.

Proof of Theorem 2.5.3. We shall divide the proof into 5 steps: discretization of the empirical process; representation of the covariance function; continuity of the approximating Gaussian process; computation of the functional dependence measures; and application of Theorem 2.3.2.

Step 1: discretization of the empirical process. Without loss of generality let $\mathcal{I} = [0, 1]$. Let $\mathcal{L} = n^2$ and $u_\ell = \ell/\mathcal{L}$, $\ell = 1, \ldots, \mathcal{L}$. For $\mathcal{V} = \{(j, \ell) : 1 \le j \le p, 1 \le \ell \le \mathcal{L}\}$, define the $(p\mathcal{L})$-dimensional vector $\mathcal{M}_i = (\mathcal{M}_{iv})_{v \in \mathcal{V}}$ with $\mathcal{M}_{iv} = H(u_\ell - X_{ij}) - \mathbb{E}H(u_\ell - X_{ij})$ for $v = (j, \ell) \in \mathcal{V}$. Let $\bar{\mathcal{M}}_n = n^{-1}\sum_{i=1}^n \mathcal{M}_i$. Since $H(\cdot)$ is increasing and $h_0 = \sup_u h(u) < \infty$, we have by the triangle inequality that

$$|\Delta_n - \sqrt{n}|\bar{\mathcal{M}}_n|_\infty| \le \frac{h_0\sqrt{n}}{\mathcal{L}} = \frac{h_0}{n\sqrt{n}}. \tag{2.5.10}$$

Step 2: representation of the covariance function. Define the projection operator $\mathcal{P}^i \cdot = \mathbb{E}(\cdot|\mathcal{F}^i) - \mathbb{E}(\cdot|\mathcal{F}^{i-1})$ and

$$D_j(u) = \sum_{l=0}^{\infty} \mathcal{P}^0 H(u - X_{lj}), \; j = 1, \ldots, p. \tag{2.5.11}$$

Recall (2.5.6) for $\sigma_{j,k}(u, v)$. By the orthogonal decomposition,

$$H(u - X_{0j}) - \mathbb{E}H(u - X_{0j}) = \sum_{m=-\infty}^{\infty} \mathcal{P}^m H(u - X_{0j})$$

and the stationarity of (X_i), we have the representation

$$\sigma_{j,k}(u,v) = \sum_{l=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \mathbb{E}[\mathcal{P}^m H(u - X_{0j}) \mathcal{P}^m H(v - X_{lk})] = \mathbb{E}[D_j(u) D_k(v)]. \quad (2.5.12)$$

Since $\mathcal{P}^0 H(u - X_{lj}) = \mathbb{E}[H(u - X_{lj}) - H(u - X_{lj,\{0\}}) | \mathcal{F}^0]$, by the first inequality in (2.5.21) and Jensen's inequality, we have

$$\|\mathcal{P}^0 H(u - X_{lj})\| \le \|H(u - X_{lj}) - H(u - X_{lj,\{0\}})\| \le 2h_0 b_l \|\varepsilon_{ij}\|, \quad (2.5.13)$$

where $b_i = (\sum_{k=1}^{d} \max_j |a_{i,jk}|^2)^{1/2}$. By (2.5.7), $\#\{i : b_i \ge 1\} \le C_1$. If $b_i < 1$, then $b_i \le b_i^{\min(1,\gamma/q)}$. Hence, $\sum_{i=0}^{\infty} b_i \le 2C_1$ and

$$(\sigma_{jj}(u,u))^{1/2} = \|D_j(u)\| \le \sum_{l=0}^{\infty} \|\mathcal{P}^0 H(u - X_{lj})\| \le 4C_1 h_0 \|\varepsilon_{ij}\|. \quad (2.5.14)$$

Step 3: continuity of the approximating Gaussian process. Let $\zeta = |u - v| \le 1$. Then $|H(u - X_{lj}) - H(v - X_{lj})| \le h_0 \zeta$. By (2.5.11) and (2.5.13),

$$\|D_j(v) - D_j(u)\| \le \sum_{l=0}^{\infty} \min(4h_0 b_l \|\varepsilon_{ij}\|,\, h_0 \zeta), \quad (2.5.15)$$

By (2.5.7), since $2\iota \le 1$ and $\zeta \le 1$, we have $\sum_{i=m}^{\infty} \min(b_i, \zeta) \le C_1 m^{-\alpha}$ for all $m \ge 1$. Let $J = \lceil \zeta^{-1/(1+\alpha)} \rceil$. Then

$$\begin{aligned} \sum_{i=0}^{J} \min(b_i, \zeta) + \sum_{i=J+1}^{\infty} \min(b_i, \zeta) &\le (J+1)\zeta + C_1 J^{-\alpha} \\ &\le (C_1 + 3)\zeta^{\alpha/(1+\alpha)}. \end{aligned} \quad (2.5.16)$$

Hence, by (2.5.12) and (2.5.15), for $C_2 = h_0(4\|\varepsilon_{ij}\| + 1)(C_1 + 3)$ we obtain

$$\|Z_j(u) - Z_j(v)\|^2 = \sigma_{j,j}(u,u) + \sigma_{j,j}(v,v) - 2\sigma_{j,j}(u,v)$$

$$= \|D_j(v) - D_j(u)\|^2 \leq C_2^2 |u-v|^{2\alpha/(1+\alpha)} \tag{2.5.17}$$

when $|u-v| \leq 1$. Let $0 < t \leq 1$ and $\lambda = \alpha/(1+\alpha)$. By (2.5.17) and the Fernique inequality (cf. Section 4.1.3 of Fernique [1975]), there exists constants $c_1, c_2, c_3 > 0$ only depending λ such that for all $w \geq c_2 C_2 t^{\lambda}$,

$$\mathbb{P}[\sup_{0 \leq y \leq t} |Z_j(v+y) - Z_j(v)| \geq w] \leq c_1[1 - \Phi(c_3 w/(C_2 t^{\lambda})], \tag{2.5.18}$$

where $\Phi(\cdot)$ is the standard normal c.d.f. For $u \in \mathcal{I} = [0,1]$, write $\lfloor u \rfloor_{\mathcal{L}} = \mathcal{L}^{-1} \lfloor \mathcal{L} u \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function. As u changes from 0 to 1, $\lfloor u \rfloor_{\mathcal{L}}$ take values $u_0, u_1, \ldots, u_{\mathcal{L}}$. Let

$$w = C_3 \mathcal{L}^{-\lambda} (\log(pn))^{1/2}, \tag{2.5.19}$$

where C_3 is a sufficiently large constant. Then by (2.5.18), we have

$$\mathbb{P}[\sup_{u \in \mathcal{I}, 1 \leq j \leq p} |Z_j(u) - Z_j(\lfloor u \rfloor_{\mathcal{L}})| \geq w] \leq p\mathcal{L} c_1[1 - \Phi(c_3 w \mathcal{L}^{\lambda}/C_2)] \leq \frac{C_4}{pn}. \tag{2.5.20}$$

Step 4: computation of the functional dependence measures. We shall first bound the functional dependence measures of the vector process $(\mathcal{M}_i)_i$ which is induced by $H(u - X_{ij})$. Let $\varepsilon_{ij}, \varepsilon_{i'j'}, i, i', j, j' \in \mathbb{Z}$, be i.i.d. random variables and $\varepsilon'_i = (\varepsilon'_{i1}, \ldots, \varepsilon'_{id})^{\top}$. Note that $X_{ij} - X_{ij,\{0\}} = a_{i,j\cdot}(\varepsilon_0 - \varepsilon'_0)$, where $a_{i,j\cdot}$ is the j-th row of the $A_i = (a_{i,jk})_{j \leq p, k \leq d}$. Then

$$\begin{aligned} \sup_u |H(u - X_{ij}) - H(u - X_{ij,\{0\}})| &\leq \min(1, h_0 |X_{ij} - X_{ij,\{0\}}|) \\ &= \min(1, h_0 |a_{i,j\cdot}(\varepsilon_0 - \varepsilon'_0)|) \\ &\leq (h_0 |a_{i,j\cdot}(\varepsilon_0 - \varepsilon'_0)|)^{\min(\gamma/q, 1)}. \end{aligned} \tag{2.5.21}$$

Recall $b_i = (\sum_{k=1}^{d} \max_j |a_{i,jk}|^2)^{1/2}$. By Lemma C.5, we have

$$\left\| \max_j |a_{i,j\cdot}(\varepsilon_0 - \varepsilon_0')| \right\|_{\min(\gamma,q)} \leq C_5 b_i \sqrt{\log p}, \tag{2.5.22}$$

where the constant C_5 depends on γ, q and $\|\varepsilon_{ij}\|_\gamma$. Hence,

$$\begin{aligned} \| \sup_{j,u} |H(u - X_{ij}) - H(u - X_{ij,\{0\}})| \|_q &\leq \left[\mathbb{E} \max_j \left(h_0 |a_{i,j\cdot}(\varepsilon_0 - \varepsilon_0')| \right)^{\min(\gamma,q)} \right]^{1/q} \\ &\leq C_6 (\log p)^\iota b_i^{2\iota}, \end{aligned}$$

which by (2.5.7) implies

$$\begin{aligned} \| |\mathcal{M}_\cdot|_\infty \|_{q,\alpha} &:= \sup_{m \geq 0} (m+1)^\alpha \sum_{i=m}^{\infty} \| \max_{j,\ell} |H(x_\ell - X_{ij}) - H(x_\ell - X_{ij,\{0\}})| \|_q \\ &\leq C_7 (\log p)^\iota. \end{aligned} \tag{2.5.23}$$

Then we can obtain the upper bounds of the dependence adjusted norms by

$$\Theta_{q,\alpha} \leq (\log(p\mathcal{L}))^{3/2} \| |\mathcal{M}_\cdot|_\infty \|_{q,\alpha}, \quad \Psi_{2,\alpha} \leq \Psi_{q,\alpha} \leq \| |\mathcal{M}_\cdot|_\infty \|_{q,\alpha}. \tag{2.5.24}$$

Step 5: application of Theorem 2.3.2. By Theorem 2.3.2 (cf (2.3.5) in Remark 1, which is applicable here in view of (2.5.14) and Assumption 2.5.2), we have

$$\sup_{u \geq 0} |\mathbb{P}(\sqrt{n} |\bar{\mathcal{M}}_n|_\infty \geq u) - \mathbb{P}(\max_{j \leq p} \max_{\ell \leq \mathcal{L}} |Z_j(u_\ell)| \geq u)| \to 0, \tag{2.5.25}$$

if the conditions of Theorem 2.3.2 are satisfied. Specifically, we have

$$L_1 = O([(\log p)^{2\iota} (\log pn)^2]^{1/\alpha}),$$

$$\max(W_1, W_2) = O((\log p)^{6\iota} (\log pn)^7),$$

as well as

$$n(\log pn)^{-4}(\log p)^{-2\iota} = O(\min(N_1, N_2)).$$

For $\alpha > 1/2 - 1/q$, there exists some κ depending on α and ι such that if $\log p = o(n^{\kappa})$, (2.3.2) and (2.3.3) hold. The other case with $0 < \alpha < 1/2 - 1/q$ can be dealt with similarly. Since

$$\left(\frac{\sqrt{n}}{\mathcal{L}} + \frac{1}{pn}\right)\sqrt{\log(p\mathcal{L})} \to 0, \tag{2.5.26}$$

by the triangle inequality and Theorem 3 of Chernozhukov et al. [2014], (2.5.9) follows in view of (2.5.10), (2.5.20), (2.5.25) and (2.5.26).

□

2.6 Estimation of Long-run Covariance Matrices

Given the realization $X_1, \ldots, X_n$, to apply the Gaussian approximation (2.3.1), we need to estimate the long-run covariance matrix Σ. Note that $\Sigma/(2\pi)$ is the value of the spectral density matrix of (X_i) at zero frequency. In the one or low-dimensional case, there is a large literature concerning spectral density estimation; see, for example, Anderson [1971], Priestley [1981], Rosenblatt [1985], Newey and West [1987], Liu and Wu [2010] among others. Assume $\mathbb{E}X_i = 0$. We then consider the batched mean estimate:

$$\hat{\Sigma} = \frac{1}{Mw}\sum_{b=1}^{w} Y_b Y_b^{\top} = \frac{1}{Mw}\sum_{b=1}^{w}(\sum_{i\in L_b} X_i)(\sum_{i\in L_b} X_i)^{\top}. \tag{2.6.1}$$

where the window $L_b = \{1+(b-1)M, \ldots, bM\}, b = 1, \ldots, w$, the window size $|L_b| = M \to \infty$ and the number of blocks $w = \lfloor n/M \rfloor$. Theorems 2.6.1 and 2.6.2 concern the convergence of the above estimate for processes with finite polynomial and finite sub-exponential dependence adjusted norms, respectively. The convergence rate depends in a subtle way on the temporal dependence characterized by α [cf. (2.2.5)], the uniform and the overall dependence adjusted norms $\Psi_{q,\alpha}$ and $\Upsilon_{q,\alpha}$, respectively, the same size n and the dimension p.

Theorem 2.6.1. *Assume* $\Psi_{q,\alpha} < \infty$ *with* $q > 4$, $\alpha > 0$, *and* $M = O(n^{\varsigma})$ *for some* $0 < \varsigma < 1$. *Let* $F_\alpha = wM$ *(resp.,* $wM^{q/2-\alpha q/2}$ *or* $w^{q/4-\alpha q/2}M^{q/2-\alpha q/2}$*) for* $\alpha > 1 - 2/q$ *(resp.,* $1/2 - 2/q < \alpha < 1 - 2/q$ *or* $\alpha < 1/2 - 2/q$*). Then for* $x \geq \sqrt{w}M\Psi_{q,\alpha}^2$, *we have*

$$\mathbb{P}(n|diag(\hat{\Sigma}) - \mathbb{E}diag(\hat{\Sigma})|_\infty \geq x) \lesssim \frac{F_\alpha \Upsilon_{q,\alpha}^q}{x^{q/2}} + p\exp\left(-\frac{C_{q,\alpha}x^2}{wM^2\Psi_{4,\alpha}^4}\right), \qquad (2.6.2)$$

$$\mathbb{P}(n|\hat{\Sigma} - \mathbb{E}\hat{\Sigma}|_\infty \geq x) \lesssim \frac{pF_\alpha \Upsilon_{q,\alpha}^q}{x^{q/2}} + p^2\exp\left(-\frac{C_{q,\alpha}x^2}{wM^2\Psi_{4,\alpha}^4}\right) \qquad (2.6.3)$$

for all large n, *where the constants in* $\lesssim$ *only depend on* ς, α *and* q.

Under stronger moment conditions, we can have an exponential inequality.

Theorem 2.6.2. *Assume* $\Phi_{\psi_\nu,0} < \infty$ *for some* $\nu \geq 0$. *Then for all* $x > 0$, *we have*

$$\mathbb{P}(n|diag(\hat{\Sigma}) - \mathbb{E}diag(\hat{\Sigma})|_\infty \geq x) \lesssim p\exp\left(-\frac{x^\gamma}{4e\gamma(\sqrt{w}M\Phi_{\psi_\nu,0}^2)^\gamma}\right), \qquad (2.6.4)$$

$$\mathbb{P}(n|\hat{\Sigma} - \mathbb{E}\hat{\Sigma}|_\infty \geq x) \lesssim p^2\exp\left(-\frac{x^\gamma}{4e\gamma(\sqrt{w}M\Phi_{\psi_\nu,0}^2)^\gamma}\right), \qquad (2.6.5)$$

where $\gamma = 1/(1+2\nu)$ *and the constants in* $\lesssim$ *only depend on* ν.

Remark 4. *An alternative estimate of* Σ, *which also works with unknown mean* $\mathbb{E}X_i$, *is*

$$\tilde{\Sigma} = \frac{1}{wM}\sum_{b=1}^{w}\Big(\sum_{i\in L_b} X_i - M\bar{X}\Big)\Big(\sum_{i\in L_b} X_i - M\bar{X}\Big)^\top, \qquad (2.6.6)$$

where $\bar{X} = (wM)^{-1}\sum_{i=1}^{wM} X_i$, $w = \lfloor n/M \rfloor$. *Then* $|\tilde{\Sigma} - \hat{\Sigma}|_\infty = M|\bar{X}|_\infty^2$. *Applying Lemma C.2 to* $\sum_{i=1}^{wM} X_{ij}$, *one can conclude that Theorems 2.6.1 and 2.6.2 still hold for* $\tilde{\Sigma}$ *with* $\mathbb{E}\hat{\Sigma}$ *therein replaced by* $\Sigma_M := \sum_{i=-M}^{M}(1 - |i|/M)\Gamma_i$ *(which equals to* $\mathbb{E}\hat{\Sigma}$ *if* $\mathbb{E}X_i = 0$*).*

Corollary 2.6.3. *(i) Under conditions in Theorem 2.6.1, we have* $|\tilde{\Sigma} - \Sigma|_\infty = O_{\mathbb{P}}(R_n)$, *where*

$$R_n = n^{-1}\max\{p^{2/q}F_\alpha^{2/q}\Upsilon_{q,\alpha}^2,\ \sqrt{w}M\Psi_{4,\alpha}^2\sqrt{\log p},\ \sqrt{w}M\Psi_{q,\alpha}^2\} + \Psi_{2,0}\Psi_{2,\alpha}v(M), \quad (2.6.7)$$

with $v(M) = 1/M$ *if* $\alpha > 1$, $v(M) = (\log M)/M$ *if* $\alpha = 1$ *and* $v(M) = 1/M^{\alpha}$ *if* $0 < \alpha < 1$.
(ii) Under conditions in Theorem 2.6.2, we have $|\tilde{\Sigma} - \Sigma|_\infty = O_{\mathbb{P}}(R_n^*)$ *with*

$$R_n^* = n^{-1}\sqrt{w} M \Phi^2_{\psi_\nu,0}(\log p)^{1/\gamma} + \Psi_{2,0}\Psi_{2,\alpha} v(M). \tag{2.6.8}$$

The above corollary easily follows from Theorems 2.6.1 and 2.6.2 since the bias $|\Sigma_M - \Sigma|_\infty \lesssim \Psi_{2,0}\Psi_{2,\alpha} v(M)$; see the proof of Lemma 2.10.3.

2.6.1 Computing Approximated Cutoff Values

To apply the Gaussian approximation (2.3.1) for hypothesis testing or construction of simultaneous confidence intervals, we need to compute χ_θ, the θth quantile of $|D_0^{-1}Z|_\infty$, $0 < \theta < 1$. The latter can be computed by simulation if the long-run covariance matrix Σ is known. When it is unknown, we shall use the estimate $\tilde{\Sigma}$ in (2.6.6). Let $\tilde{D}_0 = [\mathrm{diag}(\tilde{\Sigma})]^{1/2}$. We estimate χ_θ by $\tilde{\chi}_\theta$, the conditional θ-quantile of $|\tilde{D}_0^{-1}\tilde{\Sigma}^{1/2}\eta|_\infty$ given $(X_i)_{i=1}^n$, where $\eta \sim N(0, \mathrm{Id}_p)$ is independent of $(X_i)_{i=1}^n$. Note that $\tilde{\chi}_\theta$ can be computed by extensive simulations. This is a Gaussian multiplier resampling method using estimated long-run covariance matrices. Given the level $\alpha \in (0,1)$, we can reject the null hypothesis $H_0 : \mu = \mu_0$ at level α if $\sqrt{n}|\tilde{D}_0^{-1}(\bar{X}_n - \mu_0)|_\infty > \tilde{\chi}_{1-\alpha}$. The $(1-\alpha)$th simultaneous confidence intervals for $\mu = (\mu_1, \ldots, \mu_p)^\top$ can be constructed as $\hat{\mu}_j \pm \tilde{\chi}_{1-\alpha}\tilde{\sigma}_{jj}^{1/2}/\sqrt{n}$, $1 \le j \le p$. Corollary 2.6.4 concerns validity of this approach.

Corollary 2.6.4. *(i) Let conditions of Theorem 2.3.2 and Theorem 2.6.1 be satisfied. Further assume* $R_n \log^2 p \to 0$ *with* R_n *given by (2.6.7). Then*

$$\sup_{\theta \in (0,1)} |\mathbb{P}(\sqrt{n}|\tilde{D}_0^{-1}\bar{X}_n|_\infty \ge \tilde{\chi}_{1-\theta}) - \theta| \to 0. \tag{2.6.9}$$

(ii) Under conditions of Theorem 2.3.3 and Theorem 2.6.2, if $R_n^* \log^2 p \to 0$ *with* R_n^* *given by (2.6.8), we have (2.6.9).*

Proof of Corollary 2.6.4. (i) Recall (2.3.1) for ρ_n. Let $\Lambda_n = \sqrt{n}|(\tilde{D}_0^{-1} - D_0^{-1})\bar{X}_n|_\infty$. By the triangle inequality and Theorem 3 of Chernozhukov et al. [2014], for $w > 0$, we have

$$\begin{aligned}\tilde{\rho}_n &:= \sup_{u \in \mathbb{R}} \Big|\mathbb{P}(\sqrt{n}|\tilde{D}_0^{-1}\bar{X}_n|_\infty \geq u) - \mathbb{P}(|D_0^{-1}Z|_\infty \geq u)\Big| \\ &\leq \rho_n + \sup_{u \in \mathbb{R}} \mathbb{P}(||D_0^{-1}Z|_\infty - u| \leq w) + \mathbb{P}(\Lambda_n \geq w) \\ &\lesssim \rho_n + w\sqrt{\log p} + \mathbb{P}(\Lambda_n \geq w).\end{aligned}$$

Let $V_n = \max_{1 \leq j \leq p} |(\sigma_{jj}/\tilde{\sigma}_{jj})^{1/2} - 1|$ and $L_n = \max_{1 \leq j \leq p} |\sigma_{jj} - \tilde{\sigma}_{jj}|$. Then $\Lambda_n \leq V_n\sqrt{n}|D_0^{-1}\bar{X}_n|_\infty$. Let c be the constant in Assumption 2.3.1. On the event $\mathcal{A}_0 = \{L_n \leq x\}$ for $x \leq c/2$, we have $V_n \leq 2L_n/c$. Hence,

$$\begin{aligned}\mathbb{P}(\Lambda_n \geq w) &\leq \mathbb{P}\left(V_n \geq 2x/c\right) + \mathbb{P}(\sqrt{n}|D_0^{-1}\bar{X}_n|_\infty \geq cy/2) \\ &\leq \mathbb{P}\left(L_n \geq x\right) + \rho_n + \mathbb{P}(|D_0^{-1}Z|_\infty \geq cy/2),\end{aligned}$$

where $w = xy$, $0 < x < c/2$, $y > 0$. It follows that

$$\tilde{\rho}_n \lesssim \rho_n + xy\sqrt{\log p} + \mathbb{P}(L_n \geq x) + \mathbb{P}(|D_0^{-1}Z|_\infty \geq cy/2).$$

We let $y = C\sqrt{\log p}$, where $C > 0$ is a sufficiently large constant. Note that the marginal variances of $D_0^{-1}Z$ are 1. Let

$$r_n = \frac{1}{n}\max\{F_\alpha^{2/q}\Upsilon_{q,\alpha}^2,\ \sqrt{w}M\Psi_{4,\alpha}^2\sqrt{\log p},\ \sqrt{w}M\Psi_{q,\alpha}^2\} + \Psi_{2,0}\Psi_{2,\alpha}v(M).$$

Let $x = r_n\sqrt{\log p}$. Since $R_n \log^2 p \to 0$ and $r_n \leq R_n$, by Corollary 2.6.3, we have $\mathbb{P}(\mathcal{A}_0) \to 1$. Theorem 2.3.2 ensures $\rho_n \to 0$. Hence, $\tilde{\rho}_n \to 0$.

Let $T_n = |\tilde{\Sigma} - \Sigma|_\infty$ and $W_n = \max_{1 \leq j \leq p} |\tilde{\sigma}_{jj}/\sigma_{jj} - 1|$. By the elementary inequality

$|1-\sqrt{ab}| \leq |1-a| + (1-a)^2 + |1-b| + (1-b)^2$, we have

$$\begin{aligned} |\tilde{D}_0^{-1}\tilde{\Sigma}\tilde{D}_0^{-1} - D_0^{-1}\Sigma D_0^{-1}|_\infty &\leq \max_{1\leq j,k\leq p}\left(\left|\frac{\tilde{\sigma}_{jk}-\sigma_{jk}}{\sqrt{\sigma_{jj}\sigma_{kk}}}\right| + \left|1-\frac{\sqrt{\tilde{\sigma}_{jj}\tilde{\sigma}_{kk}}}{\sqrt{\sigma_{jj}\sigma_{kk}}}\right|\right) \\ &\leq \frac{T_n}{c} + 2W_n + 2W_n^2 \leq \frac{3T_n}{c} + \frac{2T_n^2}{c^2}. \end{aligned} \tag{2.6.10}$$

Let event $\mathcal{A} = \{T_n \leq z_n\}$ where $z_n = R_n^{1/2}/\log p$. Since $R_n \log^2 p \to 0$, we have $z_n/R_n \to \infty$. By Corollary 2.6.3, $\mathbb{P}(\mathcal{A}) \to 1$. Since $z_n \to 0$, by (2.6.10) and following the arguments of Theorem 3.1 in Chernozhukov et al. [2013a], we have

$$\sup_{\theta\in(0,1)} |\mathbb{P}(\sqrt{n}|\tilde{D}_0^{-1}\bar{X}_n|_\infty \geq \tilde{\chi}_{1-\theta}) - \theta| \lesssim \tilde{\rho}_n + \pi\left(\frac{3z_n}{c} + \frac{2z_n^2}{c^2}\right) + \mathbb{P}(T_n \geq z_n),$$

where $\pi(z) = z^{1/3}(1 \vee \log(p/z))^{2/3}$. Since $R_n \log^2 p \to 0$, (2.6.9) follows.

(ii) The proof is similar to (i), and thus is omitted. $\square$

2.7 Simulation Study

In this section we shall carry out a simulation and study how the dependence, moment condition and sample size affect the accuracy of the Gaussian approximation. We consider the following linear process with heteroscedastic errors: let $\varepsilon_{ij}, i, j \in \mathbb{Z}$, be i.i.d. random variables distributed as $t(d)/\sqrt{d/(d-2)}$, where $t(d)$ is Student's t with degrees of freedom d; let $\eta_{ij} = \varepsilon_{ij}(0.8\varepsilon_{(i-1)j}^2 + 0.2)^{1/2}$ and $\eta_i = (\eta_{i1}, \ldots, \eta_{ip})^\top$. Let

$$X_i = \sum_{k=0}^{\infty} A_k \eta_{i-k}, \tag{2.7.1}$$

where the coefficient matrices $A_k = (k+1)^{-a-1}M_k$ in which $a > 0$ and M_k are realizations from i.i.d. Ginibre matrices, namely all entries of M_k are i.i.d. $N(0,1)$. After those M_k are generated, we keep their values throughout the simulation. The parameter a controls the strength of dependency for the process. We consider the following numerical setups: a is taken to be 0.1 (stronger dependence) and 1.0 (weaker dependence); the degree of freedom d is taken to be 4 and 8. The empirical distributions of 1000 realizations are performed as an approximation of the theoretical distributions. In our simulation we truncate the sum in linear process (2.7.1) to $\sum_{k=0}^{1000}$. The normalization matrix $D_0 = [\mathrm{diag}(\Sigma)]^{1/2}$ where $\Sigma = (\sum_{k=0}^{1000} A_k)(\sum_{k=0}^{1000} A_k)^\top$. We take $n = 50, 100, 200$ and $p = 100$. The Gaussian vector Z is distributed as $N(0, \Sigma)$. For each case, we report the QQ-plot that compares the empirical distributions of $\sqrt{n}|D_0^{-1}\bar{X}_n|_\infty$ and $|D_0^{-1}Z|_\infty$.

Figures 2.1 and 2.2 indicate that, as expected from our theoretical results, the Gaussian approximation becomes better as the dependence is weaker, the tail of the process is lighter or the sample size is larger. A similar claim can be made for the 95% quantiles. The 95% quantiles for $|D_0^{-1}Z|_\infty$ is 3.476 and 3.399, for $a = 0.1$ and $a = 1.0$ respectively.

$d = 4$	$a = 0.1$	$a = 1.0$
$n = 50$	3.593	3.495
$n = 100$	3.513	3.476
$n = 200$	3.509	3.416
$n = \infty$	3.476	3.399
$d = 8$	$a = 0.1$	$a = 1.0$
$n = 50$	3.509	3.425
$n = 100$	3.489	3.413
$n = 200$	3.484	3.395
$n = \infty$	3.476	3.399

Table 1: 95% Quantiles for $\sqrt{n}|D_0^{-1}\bar{X}_n|_\infty$ vs $|D_0^{-1}Z|_\infty$.

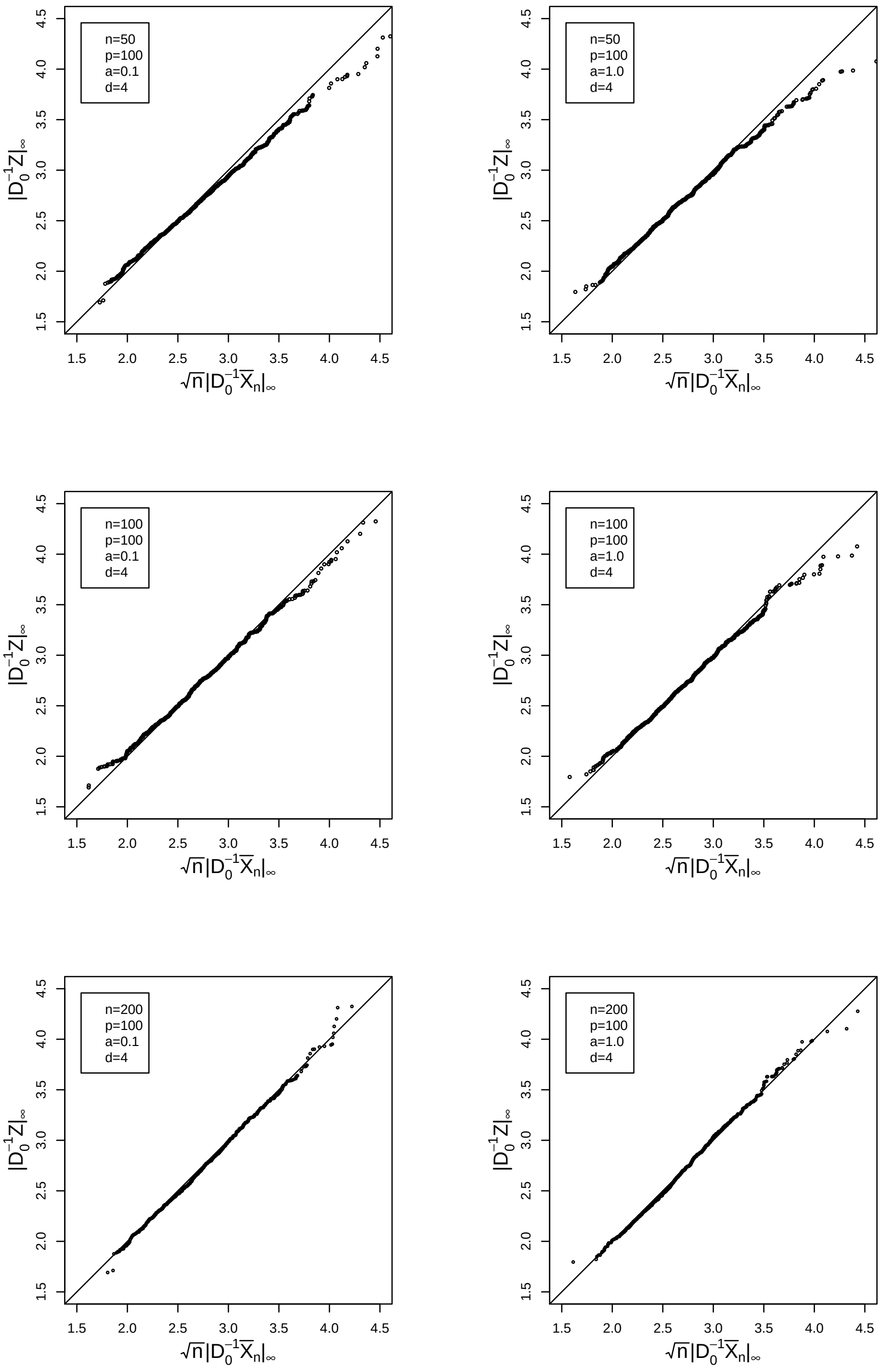

Figure 2.1: QQ-plots for $\sqrt{n}|D_0^{-1}\bar{X}_n|_\infty$ vs $|D_0^{-1}Z|_\infty$ with heavier tail innovations $d = 4$

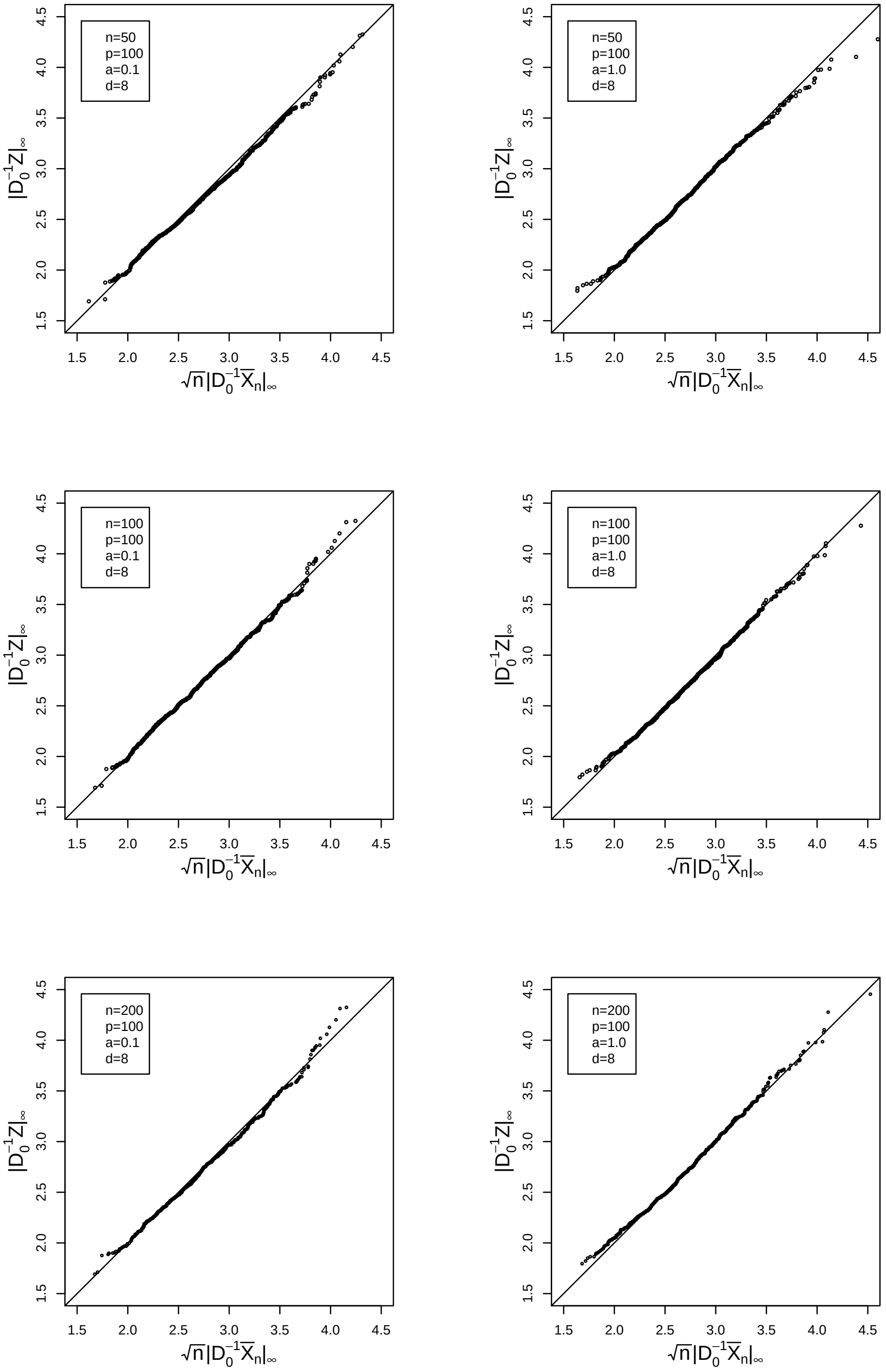

Figure 2.2: QQ-plots for $\sqrt{n}|D_0^{-1}\bar{X}_n|_\infty$ vs $|D_0^{-1}Z|_\infty$ with lighter tail innovations $d = 8$

2.8 Inequalities for High-dimensional Time Series with Finite Polynomial Moments

Tail probability inequalities play an important role in simultaneous inference. In this section, we shall derive powerful tail probability inequalities for high-dimensional stationary vectors; cf. Theorems 2.8.1 and 2.8.2. They are of independent interest. The proofs require Theorem 4.1 of Pinelis [1994], a deep Rosenthal–Burkholder-type bound on moments of Banach-spaced martingales, and Lemma C.6, a Fuk–Nagaev-type inequality for the sum of independent random vectors. We refer the readers to Appendix C for tail probability inequalities in the one-dimensional case under finite polynomial or exponential moment conditions.

Let X_i be a mean zero p-dimensional stationary process and $T_n = \sum_{i=1}^n X_i$, $T_{n,m} = \sum_{i=1}^n X_{i,m}$ where $X_{i,m} = \mathbb{E}(X_i|\varepsilon_{i-m}, \ldots, \varepsilon_i)$. We are interested in bounding the tail probabilities of $\mathbb{P}(|T_n - T_{n,m}|_\infty \geq x)$ and $\mathbb{P}(|T_n|_\infty \geq x)$ for large x. Write $\ell = \ell(p) = 1 \vee \log p$.

Theorem 2.8.1. *Assume* $\||X_\cdot|_\infty\|_{q,\alpha} < \infty$, *where* $q > 2$ *and* $\alpha \geq 0$, *and* $\Psi_{2,\alpha} < \infty$*:*
(i) If $\alpha > 1/2 - 1/q$, *for* $x \gtrsim \sqrt{n\ell}\Psi_{2,\alpha}m^{-\alpha} + n^{1/q}\ell^{3/2}\||X_\cdot|_\infty\|_{q,\alpha}m^{1/2-1/q-\alpha}$,

$$\mathbb{P}(|T_n - T_{n,m}|_\infty \geq x) \lesssim \frac{n\ell^{q/2}\||X_\cdot|_\infty\|_{q,\alpha}^q}{m^{\alpha q+1-q/2}x^q} + \exp\left(-\frac{C_{q,\alpha}x^2m^{2\alpha}}{n\Psi_{2,\alpha}^2}\right)$$

holds for all $1 \leq m \leq n$, *where the constant in* $\lesssim$ *only depends on* q *and* α. *(ii) If* $0 < \alpha < 1/2 - 1/q$, *then for* $x \gtrsim \sqrt{n\ell}\Psi_{2,\alpha}m^{-\alpha} + n^{1/2-\alpha}\ell^{3/2}\||X_\cdot|_\infty\|_{q,\alpha}$,

$$\mathbb{P}(|T_n - T_{n,m}|_\infty \geq x) \lesssim \frac{n^{q/2-\alpha q}\ell^{q/2}\||X_\cdot|_\infty\|_{q,\alpha}^q}{x^q} + \exp\left(-\frac{C_{q,\alpha}x^2m^{2\alpha}}{n\Psi_{2,\alpha}^2}\right).$$

Proof of Theorem 2.8.1. Let $s = \ell = 1 \vee \log p$. Then $\mathbb{P}(|T_n - T_{n,m}|_\infty \geq x)$ is equivalent to $\mathbb{P}(|T_n - T_{n,m}|_s \geq x)$, since for any vector $v = (v_1, \ldots, v_p)^\top$, $|v|_\infty \leq |v|_s \leq p^{1/s}|v|_\infty$. Let $L = \lfloor(\log n - \log m)/(\log 2)\rfloor$, $\varpi_l = 2^l$ if $1 \leq l < L$, $\varpi_L = \lfloor n/m \rfloor$ and $\tau_l = m\varpi_l$ for

$1 \le l < L$, $\tau_0 = m$, $\tau_L = n$. Define $M_{n,l} = T_{n,\tau_l} - T_{n,\tau_{l-1}}$ for $1 \le l \le L$ and write

$$T_n - T_{n,m} = T_n - T_{n,n} + \sum_{l=1}^{L} M_{n,l}. \tag{2.8.1}$$

Notice that $T_n - T_{n,n} = \sum_{j=n}^{\infty} T_{n,j+1} - T_{n,j}$. By Lemma C.5,

$$\| |T_n - T_{n,n}|_s \|_q \le \sum_{j=n}^{\infty} \| |T_{n,j+1} - T_{n,j}|_s \|_q \le \sum_{j=n}^{\infty} C_q (ns)^{1/2} \omega_{j+1,q},$$

where the constant C_q only depends on q. By Markov's inequality, we have

$$\mathbb{P}(|T_n - T_{n,n}|_s \ge x) \le \frac{\| |T_n - T_{n,n}|_s \|_q^q}{x^q} \le \frac{C_q (ns)^{q/2} \Omega_{n+1,q}^q}{x^q}. \tag{2.8.2}$$

For each $1 \le l \le L$, define

$$\begin{aligned} Y_{i,l} &= \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \left(X_{k,\tau_l} - X_{k,\tau_{l-1}} \right), \quad \text{for } 1 \le i \le \lfloor n/\tau_l \rfloor; \\ R_{n,l}^e &= \sum_{i \text{ is even}} Y_{i,l} \text{ and } R_{n,l}^o = \sum_{i \text{ is odd}} Y_{i,l}. \end{aligned}$$

Let $c = q/2 - 1 - \alpha q$; let $\lambda_l = l^{-2}/(\pi^2/3)$ if $1 \le l \le L/2$ and $\lambda_l = (L+1-l)^{-2}/(\pi^2/3)$ if $L/2 < l \le L$. Since $Y_{i,l}$ and $Y_{i',l}$ are independent for $|i - i'| > 1$, by Lemma C.6, for any $x > 0$,

$$\mathbb{P}(|R_{n,l}^e|_s - 2\mathbb{E}|R_{n,l}^e|_s \ge \lambda_l x) \le \frac{C_q \sum_{i \text{ is even}} \mathbb{E}|Y_{i,l}|_s^q}{(\lambda_l x)^q} + \exp\left(-\frac{(\lambda_l x)^2}{3 \sum_{i \text{ is even}} |\sigma_{Y_i,l}|_s^2} \right),$$

where $\sigma_{Y_i,l} = (\|Y_{i1,l}\|_2, \ldots, \|Y_{ip,l}\|_2)^\top$. By Lemma C.5,

$$\||Y_{i,l}|_s\|_q \le C_q(\tau_l s)^{1/2}\tilde{\omega}_{l,q}, \text{ where } \tilde{\omega}_{l,q} = \sum_{k=\tau_{l-1}+1}^{\tau_l} \omega_{k,q} \le \frac{\||X_\cdot|_\infty\|_{q,\alpha}}{\tau_{l-1}^\alpha}.$$

For $1 \le j \le p$, by Theorem 3.2 of Burkholder [1973],

$$\|Y_{ij,l}\|_2 \le \sqrt{\tau_l}\tilde{\delta}_{l,2,j}, \text{ where } \tilde{\delta}_{l,2,j} = \sum_{k=\tau_{l-1}+1}^{\tau_l} \delta_{k,2,j} \le \frac{\|X_{\cdot j}\|_{2,\alpha}}{\tau_{l-1}^\alpha},$$

which implies $|\sigma_{Y_i,l}|_s \lesssim \tau^{1/2}\tau_{l-1}^{-\alpha}\Psi_{2,\alpha}$. So, we obtain

$$\mathbb{P}(|R_{n,l}^e|_s - 2\mathbb{E}|R_{n,l}^e|_s \ge \lambda_l x) \le \frac{C_1 n s^{q/2}}{x^q} \cdot \frac{\tau_l^{q/2-1}\tilde{\omega}_{l,q}^q}{\lambda_l^q} + \exp\left(-\frac{C_2\,(\lambda_l x)^2\,\tau_{l-1}^{2\alpha}}{n\Psi_{2,\alpha}^2}\right). \tag{2.8.3}$$

By Lemma 8 in Chernozhukov et al. [2014],

$$\mathbb{E}|R_{n,l}^e|_s \lesssim \sqrt{ns}\tau_{l-1}^{-\alpha}\Psi_{2,\alpha} + n^{1/q}s^{3/2}\tau_l^{1/2-1/q}\tilde{\omega}_{l,q} \lesssim \frac{\sqrt{ns}\Psi_{2,\alpha}}{(m\varpi_l)^\alpha} + \frac{n^{1/q}s^{3/2}\||X_\cdot|_\infty\|_{q,\alpha}}{(m\varpi_l)^{-c/q}}.$$

Notice that $\lambda_l^{-1}(m\varpi_l)^{c/q} \lesssim n^{c/q}$ for $c > 0$ and $\min_{l\ge0}\lambda_l\varpi_l^{-c/q} > 0$ for $c < 0$, and $\min_{l\ge0}\lambda_l\varpi_l^\alpha > 0$. Hence, $\mathbb{E}|R_{n,l}^e|_s \lesssim \lambda_l x$ always holds and (2.8.3) implies

$$\mathbb{P}(|R_{n,l}^e|_s \ge \lambda_l x) \le \frac{C_1 n s^{q/2}}{x^q} \cdot \frac{\tau_l^{q/2-1}\tilde{\omega}_{l,q}^q}{\lambda_l^q} + \exp\left(-\frac{C_2\,(\lambda_l x)^2\,\tau_{l-1}^{2\alpha}}{n\Psi_{2,\alpha}^2}\right). \tag{2.8.4}$$

A similar inequality holds for $R_{n,l}^o$. Let

$$A = \sum_{l=1}^L \frac{\varpi_l^c}{\lambda_l^q} \text{ and } B = \sum_{l=1}^L \exp\left(-\frac{C_5 x^2\lambda_l^2\varpi_l^{2\alpha}}{nm^{-2\alpha}\Psi_{2,\alpha}^2}\right).$$

Since $\sum_{l=1}^{L} \lambda_l \le 1$ and $|M_{n,l}|_s \le |R^e_{n,l}|_s + |R^o_{n,l}|_s$, by (2.8.4),

$$\begin{aligned}
\mathbb{P}(|\sum_{l=1}^{L} M_{n,l}|_s \ge 2x) &\le \sum_{l=1}^{L} \mathbb{P}\left(|M_{n,l}|_s \ge 2\lambda_l x\right) \\
&\le \sum_{l=1}^{L} [\mathbb{P}(|R^e_{n,l}|_s \ge \lambda_l x) + \mathbb{P}(|R^o_{n,l}|_s \ge \lambda_l x)] \\
&\le \frac{C_3 n m^c s^{q/2} \||X_\cdot|_\infty\|^q_{q,\alpha}}{x^q} A + C_4 B. \qquad (2.8.5)
\end{aligned}$$

Let $\nu := \min_{l \ge 1} \lambda_l^2 \varpi_l^{2\alpha} > 0$. By the definition of ϖ_l and λ_l and by elementary calculations, there exists a constant $C_6 > 1$ such that for all $t \ge 1$,

$$\sum_{l=1}^{L} \exp(-C_5 t \lambda_l^2 \varpi_l^{2\alpha}) \le C_6 \exp(-C_5 t \nu), \qquad (2.8.6)$$

If $c > 0$, it can be obtained that $A \le C_7 \varpi_L^c \le C_7 n^c / m^c$. If $c < 0$, then $A \le C_8$. Hence, combining (2.8.1), (2.8.2), (2.8.5), (2.8.6), Theorem 2.8.1 follows. □

Theorem 2.8.2. *Assume $\||X_\cdot|_\infty\|_{q,\alpha} < \infty$, where $q > 2$ and $\alpha \ge 0$, and $\Psi_{2,\alpha} < \infty$: (i) If $\alpha > 1/2 - 1/q$, then for $x \gtrsim \sqrt{n\ell}\Psi_{2,\alpha} + n^{1/q}\ell^{3/2}\||X_\cdot|_\infty\|_{q,\alpha}$,*

$$\mathbb{P}(|T_n|_\infty \ge x) \le \frac{C_{q,\alpha} n \ell^{q/2} \||X_\cdot|_\infty\|^q_{q,\alpha}}{x^q} + C_{q,\alpha} \exp\left(-\frac{C_{q,\alpha} x^2}{n \Psi^2_{2,\alpha}}\right). \qquad (2.8.7)$$

(ii) If $0 < \alpha < 1/2 - 1/q$, then for $x \gtrsim \sqrt{n\ell}\Psi_{2,\alpha} + n^{1/2-\alpha}\ell^{3/2}\||X_\cdot|_\infty\|_{q,\alpha}$,

$$\mathbb{P}(|T_n|_\infty \ge x) \le \frac{C_{q,\alpha} n^{q/2 - \alpha q} \ell^{q/2} \||X_\cdot|_\infty\|^q_{q,\alpha}}{x^q} + C_{q,\alpha} \exp\left(-\frac{C_{q,\alpha} x^2}{n \Psi^2_{2,\alpha}}\right). \qquad (2.8.8)$$

Proof of Theorem 2.8.2. The proof is similar to that of Theorem 2.8.1, and thus is omitted.

□

2.9 Auxiliary Results

Here we shall provide some Nagaev-type tail probability inequalities for one dimensional time series. Let $\varepsilon_i, \varepsilon'_k, i, k, \in \mathbb{Z}$, be i.i.d. random elements. We consider the one-dimensional stationary process $(e_i)_{i=-\infty}^{\infty}$ of the form

$$e_i = g(\ldots, \varepsilon_{i-1}, \varepsilon_i), \tag{2.9.1}$$

where g is a measurable function such that e_i is well-defined. Recall $\mathcal{F}^i = (\ldots, \varepsilon_{i-1}, \varepsilon_i)$ and the projection operator $\mathcal{P}^i \cdot = \mathbb{E}(\cdot|\mathcal{F}^i) - \mathbb{E}(\cdot|\mathcal{F}^{i-1})$. The projection $(\mathcal{P}^i \cdot)_{i \in \mathbb{Z}}$ induces martingale differences with respect to $(\mathcal{F}^i)$. We define respectively the functional and predictive dependence measures

$$\delta_{i,q} = \|e_i - g(\mathcal{F}^{i,\{0\}})\|_q, \ \theta_{i,q} = \|\mathcal{P}^0 e_i\|_q. \tag{2.9.2}$$

where $\mathcal{F}^{i,\{0\}} = (\ldots, \varepsilon_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_i)$. Let $\delta_{i,q} = 0$ if $i < 0$; let $\Delta_{m,q} = \sum_{i=m}^{\infty} \delta_{i,q}$, $m \geq 0$, be the tail dependence measures, and the dependence adjusted norm

$$\|e.\|_{q,\alpha} := \sup_{m \geq 0} (m+1)^{\alpha} \Delta_{m,q}, \text{ for } \alpha \geq 0. \tag{2.9.3}$$

Here $\delta_{i,q}$ measures the dependence of e_i on ε_0 and $\Delta_{m,q}$ measures the cumulative impact of ε_0 on $(e_i)_{i \geq m}$. The predictive dependence measure provides an evaluation to the effect on the prediction of e_i when part of the previous inputs is concealed, and it satisfies $\theta_{i,q} \leq \delta_{i,q}$ in view of Jensen's inequality.

2.9.1 Inequalities with Finite Polynomial Moments

For $m \geq 0$, the m-dependence approximation of e_i is denoted by $e_{i,m}$ where

$$e_{i,m} = \mathbb{E}(e_i | \varepsilon_{i-m}, \varepsilon_{i-m+1}, \ldots, \varepsilon_i).$$

Let $S_n = \sum_{i=1}^n e_i$, $S_{n,m} = \sum_{i=1}^n e_{i,m}$. With the dependence adjusted norm (2.9.3), we are able to provide tail probability inequalities for error bounds when approximating (e_i) by the m-dependent process $(e_{i,m})$. In lemmas below the constant $C_{q,\alpha}$ only depends on q and α and its values may change from line to line.

Lemma 2.9.1. *Assume $\|e_\cdot\|_{q,\alpha} < \infty$, where $q > 2$ and $\alpha > 0$. (i) If $\alpha > 1/2 - 1/q$, then*

$$\mathbb{P}(|S_n - S_{n,m}| \geq x) \leq \frac{C_{q,\alpha} n m^{q/2-1-\alpha q} \|e_\cdot\|_{q,\alpha}^q}{x^q} + C_{q,\alpha} \exp\left(-\frac{C_{q,\alpha} x^2 m^{2\alpha}}{n \|e_\cdot\|_{2,\alpha}^2}\right)$$

holds for all $x > 0$ and $1 \leq m \leq n$. (ii) If $0 < \alpha < 1/2 - 1/q$, we have

$$\mathbb{P}(|S_n - S_{n,m}| \geq x) \leq \frac{C_{q,\alpha} n^{q/2-\alpha q} \|e_\cdot\|_{q,\alpha}^q}{x^q} + C_{q,\alpha} \exp\left(-\frac{C_{q,\alpha} x^2 m^{2\alpha}}{n \|e_\cdot\|_{2,\alpha}^2}\right).$$

Proof of Lemma 2.9.1. It is a special case of Theorem 2.8.1 for $p = 1$. □

Lemma 2.9.2 (cf. Theorem 2 of Wu and Wu [2016]). *Assume that $\|e_\cdot\|_{q,\alpha} < \infty$, where $q > 2$ and $\alpha > 0$. (i) If $\alpha > 1/2 - 1/q$, then there exists some constant $C_{q,\alpha}$ depending on q and α only such that, for $x > 0$,*

$$\mathbb{P}(|S_n| \geq x) \leq \frac{C_{q,\alpha} n \|e_\cdot\|_{q,\alpha}^q}{x^q} + C_{q,\alpha} \exp\left(-\frac{C_{q,\alpha} x^2}{n \|e_\cdot\|_{2,\alpha}^2}\right). \tag{2.9.4}$$

(ii) If $0 < \alpha < 1/2 - 1/q$, we have the following inequality,

$$\mathbb{P}(|S_n| \geq x) \leq \frac{C_{q,\alpha} n^{q/2-\alpha q} \|e_\cdot\|_{q,\alpha}^q}{x^q} + C_{q,\alpha} \exp\left(-\frac{C_{q,\alpha} x^2}{n \|e_\cdot\|_{2,\alpha}^2}\right). \tag{2.9.5}$$

Remark 5. *By Markov's inequality and Lemma 1 of Liu and Wu [2010], one obtains*

$$\mathbb{P}(|S_n - S_{n,m}| \geq x) \leq \frac{\|S_n - S_{n,m}\|_q^q}{x^q} \leq C_q \frac{n^{q/2} m^{-\alpha q} \|e_\cdot\|_{q,\alpha}^q}{x^q}. \tag{2.9.6}$$

In comparison, the polynomial tail bounds in (2.9.4) and (2.9.5) are sharper.

2.9.2 Inequalities with Finite Exponential Moments

If e_i satisfies stronger moment condition than the existence of finite q-th moment, we can have an exponential inequality. We shall assume $\|e.\|_{q,\alpha} < \infty$ for all $q > 0$ and some $\alpha \geq 0$ and we further assume for some $\nu \geq 0$, the dependence adjusted sub-exponential norm

$$\|e.\|_{\psi_\nu,\alpha} := \sup_{q\geq 2} q^{-\nu}\|e.\|_{q,\alpha} < \infty. \tag{2.9.7}$$

Lemma 2.9.3. *Assume (2.9.7). Let* $J_n = (S_n - S_{n,m})/\sqrt{n}$ *and* $\beta = 2/(1+2\nu)$. *Then*

$$h(t) := \sup_{n\in\mathbb{N}} \mathbb{E}[\exp(tJ_n^\beta)] \leq 1 + C_\beta(1-t/t_0)^{-1/2}t/t_0$$

holds for $0 \leq t < t_0$ *with* $t_0 = m^{\alpha\beta}/(e\beta\|e.\|_{\psi_\nu,\alpha}^\beta)$. *Consequently, letting* $t = t_0/2$, *for* $x > 0$,

$$\mathbb{P}(|J_n| \geq x) \leq \exp(-tx^\beta)h(t) \leq C_\beta \exp\left(-\frac{x^\beta m^{\alpha\beta}}{2e\beta\|e.\|_{\psi_\nu,\alpha}^\beta}\right). \tag{2.9.8}$$

Lemma 2.9.4 (cf. Theorem 3 of Wu and Wu [2016])**.** *Assume (2.9.7) holds for* $\alpha = 0$. *Let* $\beta = 2/(1+2\nu)$. *Then for* $x > 0$,

$$\mathbb{P}(|S_n/\sqrt{n}| \geq x) \leq C_\beta \exp\left(-\frac{x^\beta}{2e\beta\|e.\|_{\psi_\nu,0}^\beta}\right). \tag{2.9.9}$$

Proof of Lemma 2.9.3. Let $Q_{n,l} = \sum_{i=1}^n \mathcal{P}^{i-l}X_i$, $l \geq 0$. Then $Q_{n,l}$ is a martingale. By Theorem 2.1 of Rio [2009], we have

$$\|Q_{n,l}\|_q^2 \leq (q-1)\sum_{i=1}^n \|\mathcal{P}^{i-l}X_i\|_q^2 = (q-1)n(\theta_{l,q})^2.$$

By $\theta_{l,q} \leq \delta_{l,q}$, we have $\|J_n\|_q \leq (q-1)^{1/2}\Delta_{m+1,q}$ in view of $\sqrt{n}J_n = \sum_{l=m+1}^\infty Q_{n,l}$. Write the negative binomial expansion $(1-s)^{-1/2} = 1 + \sum_{k=1}^\infty a_k s^k$ with $a_k = (2k)!/(2^{2k}(k!)^2)$

for $|s| < 1$. By Stirling's formula, we have $a_k \sim (k\pi)^{-1/2}$ as $k \to \infty$. Hence, there exists absolute constants $C_1, C_2 > 0$ such that for all $k \geq 1$,

$$C_1 (k/e)^k a_k^{-1} \leq k! \leq C_2 (k/e)^k a_k^{-1}. \tag{2.9.10}$$

Under condition (2.9.7), if $k\beta \geq 2$, then $\|e.\|_{\beta k, \alpha} \leq \|e.\|_{\psi_\nu, \alpha} (\beta k)^\nu$ and hence

$$\frac{t^k \|J_n^\beta\|_k^k}{k!} \leq \frac{t^k (\beta k - 1)^{\beta k/2} \Delta_{m+1,\beta k}^{\beta k}}{C_1 (k/e)^k a_k^{-1}} \leq \frac{a_k t^k (\beta k - 1)^{\beta k/2}}{C_1 t_0^k (\beta k)^{\beta k/2}} \leq \frac{a_k t^k}{C_1 \sqrt{e} t_0^k}.$$

If $k\beta < 2$, then $\|J_n\|_{\beta k} \leq \|J_n\|_2 \leq 2^\nu m^{-\alpha} \|e.\|_{\psi_\nu, \alpha}$. In $e^y = \sum_{k=0}^\infty y^k / k!$, let $y = t J_n^\beta$, then

$$\begin{aligned} h(t) &\leq 1 + \sum_{1 \leq k < 2/\beta} \frac{t^k (2^\nu m^{-\alpha} \|e.\|_{\psi_\nu, \alpha})^{\beta k}}{k!} + \sum_{k \geq 2/\beta} \frac{a_k t^k}{C_1 \sqrt{e} t_0^k} \\ &\leq 1 + C_\beta \sum_{k=1}^\infty a_k \frac{t^k}{t_0^k} \leq 1 + C_\beta \frac{t/t_0}{(1 - t/t_0)^{1/2}}, \end{aligned}$$

where $C_\beta > 0$ only depends on β. So (2.9.8) follows by Markov's inequality. □

2.9.3 Inequalities for Sums of High-dimensional Random Vectors

In this section we shall present two useful inequalities for sums of high-dimensional random vectors. Lemma 2.9.5 provides a Rosenthal-Burkholder type bound on moments of Banach-spaced martingales and follows from Theorem 4.1 of Pinelis [1994]. Lemma 2.9.6 is a Fuk-Nagaev type inequality for the sum of independent random vectors. For a p-dimensional vector $v = (v_1, \ldots, v_p)$ recall the s-length $|v|_s = (\sum_{j=1}^p |v_j|^s)^{1/s}$, $s \geq 1$.

Lemma 2.9.5. *Let D_i, $1 \leq i \leq n$, be p-dimensional martingale difference vectors with* *respect to the σ-field $\mathcal{G}_i$. Let $s > 1$ and $q \geq 2$. Then*

$$\||D_1 + \ldots + D_n|_s\|_q \leq c \left\{ q\|\sup_i |D_i|_s\|_q + \sqrt{q(s-1)} \left\| \left[\sum_{i=1}^{n} \mathbb{E}(|D_i|_s^2|\mathcal{G}_{i-1}) \right]^{1/2} \right\|_q \right\},$$

where c is an absolute constant.

Lemma 2.9.6. *Assume $s > 1$. Let $X_1, \ldots, X_n$ be p-dimensional independent random vectors with mean zero such that for some $q > 2$, $\||X_i|_s\|_q < \infty$, $1 \leq i \leq n$. Let $T_n = \sum_{i=1}^n X_i$ and $\sigma_i = (\|X_{i1}\|_2, \ldots, \|X_{ip}\|_2)^\top$. Then for any $y > 0$,*

$$\mathbb{P}\left(|T_n|_s \geq 2\mathbb{E}|T_n|_s + y\right) \leq C_q y^{-q} \sum_{i=1}^{n} \mathbb{E}|X_i|_s^q + \exp\left(-\frac{y^2}{3\sum_{i=1}^n |\sigma_i|_s^2}\right), \tag{2.9.11}$$

where C_q is a positive constant only depending on q.

Proof of Lemma 2.9.6. For $s > 1$, we apply Theorem 3.1 of Einmahl and Li [2008] with the Banach space $(\mathbb{R}^p, |\cdot|_s)$ and $\eta = \delta = 1$. The unit ball of the dual of $(\mathbb{R}^p, |\cdot|_s)$ is the set of linear functions $\{u = (u_1, \ldots, u_p)^\top \mapsto \lambda^\top u : \lambda \in \mathbb{R}^p, |\lambda|_a \leq 1\}$ where $1/a + 1/s = 1$. By Minkowski's and Hölder's inequalities, we have

$$\|\lambda^\top X_i\|_2 \leq \sum_{j=1}^{p} |\lambda_j| \cdot \|X_{ij}\|_2 \leq |\lambda|_a |\sigma_i|_s.$$

Hence, the Λ_n therein is bounded by $\sum_{i=1}^n |\sigma_i|_s^2$. □

2.10 Deferred Proofs

In this section, we shall provide deferred proofs of the results in this chapter.

2.10.1 An error bound of the Gaussian approximation

We shall apply the m-dependence approximation approach. For $m \geq 0$, define

$$X_{i,m} = (X_{i1,m}, \ldots, X_{ip,m})^\top = \mathbb{E}(X_i | \varepsilon_{i-m}, \varepsilon_{i-m+1}, \ldots, \varepsilon_i). \tag{2.10.1}$$

Write $T_X = \sum_{i=1}^n X_i$ and $T_{X,m} = \sum_{i=1}^n X_{i,m}$. For simplicity, suppose $n = (M+m)w$, where $M \gg m$ and $M, m, w \to \infty$ (to be determined) as $n \to \infty$. We apply the block technique and split the interval $[1, n]$ into alternating large blocks $L_b = [(b-1)(M+m)+1, bM+(b-1)m]$ and small blocks $S_b = [bM + (b-1)m + 1, b(M+m)]$, $1 \leq b \leq w$. Let

$$Y_b = \sum_{i \in L_b} X_i,\ Y_{b,m} = \sum_{i \in L_b} X_{i,m},\ T_Y = \sum_{b=1}^w Y_b,\ T_{Y,m} = \sum_{b=1}^w Y_{b,m}.$$

Let Z_b, $1 \leq b \leq w$, be i.i.d. $N(0, MB)$ and $Z_{b,m}$ be i.i.d. $N(0, M\tilde{B})$, where the covariance matrices B and $\tilde{B}$ are respectively given by

$$B = (b_{ij})_{i,j=1}^p = \mathrm{Cov}(Y_b/\sqrt{M}) \text{ and } \tilde{B} = (\tilde{b}_{ij})_{i,j=1}^p = \mathrm{Cov}(Y_{b,m}/\sqrt{M}). \tag{2.10.2}$$

Write $T_{Z,m} = \sum_{b=1}^w Z_{b,m}$ and let $Z \sim N(0, \Sigma)$.

Lemma 2.10.1. *(i) Assume $\Theta_{q,\alpha} < \infty$ for some $q > 2$ and $\alpha > 0$. Then there exists some constant $C_{q,\alpha}$ such that for $y > 0$*

$$\mathbb{P}(|T_X - T_{Y,m}|_\infty \geq y) \lesssim f_1^*(y) + f_2^*(y) =: f^*(y) \tag{2.10.3}$$

where the constant in $\lesssim$ only depends on q and α,

$$f_1^*(y) = \begin{cases} y^{-q} n m^{q/2-1-\alpha q} \Theta_{q,\alpha}^q + p \exp\left(-\dfrac{C_{q,\alpha} y^2 m^{2\alpha}}{n \Psi_{2,\alpha}^2}\right), & \alpha > 1/2 - 1/q \\ y^{-q} n^{q/2-\alpha q} \Theta_{q,\alpha}^q + p \exp\left(-\dfrac{C_{q,\alpha} y^2 m^{2\alpha}}{n \Psi_{2,\alpha}^2}\right), & \alpha < 1/2 - 1/q \end{cases} \tag{2.10.4}$$

and

$$f_2^*(y) = \begin{cases} y^{-q} wm\Theta_{q,\alpha}^q + p\exp\left(-\frac{C_{q,\alpha}y^2}{mw\Psi_{2,\alpha}^2}\right), & \alpha > 1/2 - 1/q \\ y^{-q}(wm)^{q/2-\alpha q}\Theta_{q,\alpha}^q + p\exp\left(-\frac{C_{q,\alpha}y^2}{wm\Psi_{2,\alpha}^2}\right), & \alpha < 1/2 - 1/q \end{cases}. \tag{2.10.5}$$

(ii) Assume $\Phi_{\psi_\nu,\alpha} < \infty$ for some $\nu \geq 0$ and $\alpha > 0$. Let $\beta = 2/(1+2\nu)$. Then there exists a constant $C_\beta > 0$ such that for $y > 0$,

$$\mathbb{P}(|T_X - T_{Y,m}|_\infty \geq y) \lesssim f_1^\diamond(y) + f_2^\diamond(y) =: f^\diamond(y), \tag{2.10.6}$$

where the constant in $\lesssim$ only depends on β and α,

$$f_1^\diamond(y) = p\exp\left\{-C_\beta\left(\frac{ym^\alpha}{\sqrt{n}\Phi_{\psi_\nu,\alpha}}\right)^\beta\right\},\ f_2^\diamond(y) = p\exp\left\{-C_\beta\left(\frac{y}{\sqrt{mw}\Phi_{\psi_\nu,0}}\right)^\beta\right\}.$$

Lemma 2.10.2. *Let $D = (d_{ij})_{i,j=1}^p$ be a diagonal matrix. Assume that there exist constants $c > 0, c_2 > c_1 > 0$ such that $c < \min_{1\leq j\leq p} d_{jj}$ and $c_1 \leq \tilde{b}_{jj}/d_{jj} \leq c_2$ for all $1 \leq j \leq p$. Assume $\Psi_{q,0} < \infty$ for some $q \geq 4$. Then for all $\lambda \in (0,1)$,*

$$\begin{aligned} & \sup_{t\in\mathbb{R}} \left|\mathbb{P}(|D^{-1/2}T_{Y,m}/\sqrt{n}|_\infty \leq t) - \mathbb{P}(|D^{-1/2}T_{Z,m}/\sqrt{n}|_\infty \leq t)\right| \\ \lesssim\ & w^{-1/8}(\Psi_{3,0}^{3/4} \vee \Psi_{4,0}^{1/2})(\log(pw/\lambda))^{7/8} + w^{-1/2}(\log(pw/\lambda))^{3/2}u_m(\lambda) + \lambda \\ =:\ & h(\lambda, u_m(\lambda)), \end{aligned}$$

where the constant in $\lesssim$ depends on c, c_1, c_2 and q and α for (i), and β for (ii) below, and $u_m(\lambda) \leq u_m^*(\lambda)$ in (i), and $u_m(\lambda) \leq u_m^\diamond(\lambda)$ in (ii).

(i) Assume $\Theta_{q,\alpha} < \infty$ for some $q \geq 4$ and $\alpha > 0$, then

$$u_m^*(\lambda) = \begin{cases} \max\{\Theta_{q,\alpha}(\lambda^{-1}w)^{1/q}M^{1/q-1/2}, \Psi_{2,\alpha}\sqrt{\log(pw/\lambda)}\}, & \alpha > 1/2 - 1/q \\ \max\{\Theta_{q,\alpha}(\lambda^{-1}w)^{1/q}M^{-\alpha}, \Psi_{2,\alpha}\sqrt{\log(pw/\lambda)}\}, & \alpha < 1/2 - 1/q. \end{cases} \tag{2.10.7}$$

(ii) Assume $\Phi_{\psi_\nu,0} < \infty$ *for some* $\nu \geq 0$*. Then*

$$u_m^\diamond(\lambda) = \max\{\Phi_{\psi_\nu,0}(\log(pw/\lambda))^{1/\beta}, \sqrt{\log(pw/\lambda)}\}. \tag{2.10.8}$$

Lemma 2.10.3. *Assume* $\Psi_{2,\alpha} < \infty$ *for some* $\alpha > 0$*. Let* $D = (d_{ij})_{i,j=1}^p$ *be a diagonal matrix such that there exist some constants* $0 < C_1 < C_2$ *such that* $C_1 \leq \sigma_{jj}/d_{jj} \leq C_2$ *for all* $1 \leq j \leq p$*. Then we have*

$$\begin{aligned} & \sup_{t\in\mathbb{R}} \left| \mathbb{P}(|D^{-1/2} T_{Z,m}/\sqrt{n}|_\infty \leq t) - \mathbb{P}(|D^{-1/2} Z|_\infty \leq t) \right| \\ \lesssim \; & \pi(\max_{1\leq j\leq p} d_{jj}^{-1} \Psi_{2,\alpha}\Psi_{2,0}(m^{-\alpha} + v(M)) + wm/n), \end{aligned}$$

where $\pi(x) = x^{1/3}(1 \vee \log(p/x))^{2/3}$ *for* $x > 0$ *and* $v(M)$ *is the same as defined in Corollary 2.6.3.*

Theorem 2.10.4. *Let* $\Sigma_0 = \operatorname{diag}(\Sigma)$ *and* $D_0 = \Sigma_0^{1/2}$*. Let Assumption 2.3.1 be satisfied. (i) Assume* $\Theta_{q,\alpha} < \infty$*, where* $q \geq 4$ *and* $\alpha > 0$*. Let* $\chi(m, M) = \Psi_{2,\alpha}\Psi_{2,0}(m^{-\alpha} + v(M)) + wm/n$*, where* $v(M)$ *is given in Corollary 2.6.3. Recall (2.3.1) for* ρ_n*. Then for every* $\lambda \in (0, 1)$ *and* $\eta > 0$*,*

$$\rho_n \lesssim f^*(\sqrt{n}\eta) + \eta\sqrt{\log p} + h(\lambda, u_m^*(\lambda)) + \pi(\chi(m, M)). \tag{2.10.9}$$

(ii) Assume $\Phi_{\psi_\nu,\alpha} < \infty$*, where* $\nu \geq 0$ *and* $\alpha > 0$*. Then for every* $\lambda \in (0, 1)$ *and* $\eta > 0$*,*

$$\rho_n \lesssim f^\diamond(\sqrt{n}\eta) + \eta\sqrt{\log p} + h(\lambda, u_m^\diamond(\lambda)) + \pi(\chi(m, M)). \tag{2.10.10}$$

Proof of Theorem 2.10.4. (i) By Lemma 2.10.2 (i) and Lemma 2.10.3, we have for every

$\lambda \in (0,1)$,

$$\begin{aligned} &\sup_{t\in\mathbb{R}} \Big|\mathbb{P}(|D_0^{-1}T_{Y,m}/\sqrt{n}|_\infty \le t) - \mathbb{P}(|D_0^{-1}Z|_\infty \le t)\Big| \\ \lesssim\ & h(\lambda, u_m^*(\lambda)) + \pi(\Psi_{2,\alpha}\Psi_{2,0}(m^{-\alpha} + v(M)) + wm/n). \end{aligned} \tag{2.10.11}$$

Observe that the Gaussian vector $D_0^{-1}Z$ has marginal variance 1. By Theorem 3 of Chernozhukov et al. [2014], for every $\eta > 0$,

$$\sup_{t\in\mathbb{R}} \mathbb{P}(\Big||D_0^{-1}Z|_\infty - t\Big| \le \eta) \lesssim \eta\sqrt{\log p}. \tag{2.10.12}$$

By the triangle inequality, for every $\eta > 0$, we have

$$\begin{aligned} &\sup_{t\in\mathbb{R}} \Big|\mathbb{P}(|D_0^{-1}T_X/\sqrt{n}|_\infty > t) - \mathbb{P}(|D_0^{-1}T_{Y,m}/\sqrt{n}|_\infty > t)\Big| \\ \le\ & \mathbb{P}(|D_0^{-1}(T_X - T_{Y,m})/\sqrt{n}|_\infty > \eta) + \sup_{t\in\mathbb{R}} \mathbb{P}(\Big||D_0^{-1}T_{Y,m}/\sqrt{n}|_\infty - t\Big| \le \eta), \end{aligned}$$

which implies Theorem 2.10.4 (i) in view of Lemma 2.10.1 (i), (2.10.11) and (2.10.12).

(ii) Inequality (2.10.10) can be obtained by replacing f^* and u_m^* with $f^\diamond$ and $u_m^\diamond$ in the above proof. □

2.10.2 Proofs of Theorem 2.3.2 and Theorem 2.3.3

Proof. Recall (2.10.3) for $f^*(\cdot)$. By Theorem 2.10.4, for $\alpha > 1/2 - 1/q$, to have (2.3.1), we need

$$\pi(\Psi_{2,\alpha}\Psi_{2,0}(m^{-\alpha} + v(M)) + wm/n) \to 0 \tag{2.10.13}$$

and for some $\eta > 0$ and $\lambda \in (0,1)$,

$$f^*(\sqrt{n}\eta) + \eta\sqrt{\log p} \to 0, \tag{2.10.14}$$

$$h(\lambda, u_m^*(\lambda)) \to 0. \tag{2.10.15}$$

First, (2.10.13) requires $m \gg L_1$, $wm \ll n(\log p)^{-2}$, $w \ll n(\log p)^{-2}(\Psi_{2,\alpha}\Psi_{2,0})^{-1}$ if $\alpha > 1$ and $w \ll n/L_1$ if $0 < \alpha < 1$. Moreover, (2.10.14) requires $m \gg \max(L_0, (\Psi_{2,\alpha} \log p)^{1/\alpha})$ with $L_0 = (n^{1/q-1/2}(\log p)^{1/2}\Theta_{q,\alpha})^{1/(\alpha-1/2+1/q)}$ and $wm \ll \min(N_1, N_2)$. And (2.10.15) needs (2.3.2) and $w \gg \max(W_1, W_2)$. We also need $M \asymp n/w \gg m$. Notice that $(\Psi_{2,\alpha} \log p)^{1/\alpha} \lesssim L_1$, $N_2 \lesssim n(\log p)^{-2}$, $N_2 \leq n(\log p)^{-2}(\Psi_{2,\alpha}\Psi_{2,0})^{-1}$ and under (2.3.2), $L_0 \to 0$. If

$$L_1 \max(W_1, W_2) = o(1) \min(n, N_1, N_2), \tag{2.10.16}$$

then we can always choose m and w such that (2.3.1) holds. Observe that $N_2 \lesssim n$, then (2.10.16) is reduced to (2.3.3).

For $0 < \alpha < 1/2 - 1/q$, the function f^* in (2.10.14) is replaced by $f^\diamond$ [cf. (2.10.6)], which implies $\Theta_{q,\alpha}(\log p)^{1/2} = o(n^\alpha)$, $m \gg (\Psi_{2,\alpha} \log p)^{1/\alpha}$ and $wm \ll \min(N_2, N_3)$. And u_m^* in (2.10.15) is replaced by $u_m^\diamond$, implying $w \gg \max(W_1, W_2, W_3)$. By the similar argument, if (2.3.4) is further assumed, then (2.3.1) also holds for the case $0 < \alpha < 1/2 - 1/q$.

The proof of Theorem 2.3.3 is similar to that of Theorem 2.3.2, and thus is omitted. $\square$

Remark 6. *In the proof of Theorem 2.3.2, we exclude the case* $\alpha = 1$ *when* $\alpha > 1/2 - 1/q$. *If* $\alpha = 1$, *we need to impose the additional assumption*

$$\max(W_1, W_2) = o(n/(L_1 \log n)) \tag{2.10.17}$$

to ensure (2.10.13). The above condition is very mild since (2.3.3) implies $\max(W_1, W_2) = o(n/L_1)$. *If* $\log n \lesssim (\log p)^2\Psi_{2,\alpha}^2$, *which trivially holds in the high-dimensional case* $p \asymp n^\kappa$ *with some* $\kappa > 0$, *we have* $N_2 = O(n/\log n)$, *and hence (2.3.3) implies (2.10.17). Similarly, in Theorem 2.3.3 we shall further assume* $\max(W_1, W_4) = o(n/(L_1 \log n))$ *if* $\alpha = 1$.

2.10.3 Proofs of Results in Section 2.6

For a random variable X, we define the operator $\mathbb{E}_0$ as $\mathbb{E}_0(X) := X - \mathbb{E}X$. For $\mathcal{F}^i = (\ldots, \varepsilon_{i-1}, \varepsilon_i)$, define the projection operator $\mathcal{P}^i \cdot = \mathbb{E}(\cdot|\mathcal{F}^i) - \mathbb{E}(\cdot|\mathcal{F}^{i-1})$.

Proof of Theorem 2.6.1. Fix $1 \leq j, k \leq p$; let

$$T = \sum_{b=1}^{w} Y_{bj} Y_{bk},$$

where $Y_{bj} = \sum_{i \in L_b} X_{ij}$. For $\tau \geq 0$, define $X_{ij,\tau} = \mathbb{E}(X_{ij} | \varepsilon_{i-\tau}, \ldots, \varepsilon_i)$, $Y_{bj,\tau} = \sum_{i \in L_b} X_{ij,\tau}$ and $T_\tau = \sum_{b=1}^{w} Y_{bj,\tau} Y_{bk,\tau}$. We shall first prove for $\alpha > 1/2 - 1/q$ and $\alpha < 1/2 - 1/q$ respectively,

$$\mathbb{P}(|\mathbb{E}_0(T - T_M)| \geq x) \lesssim \begin{cases} x^{-q/2} w M^{q/2 - \alpha q/2} \xi_{q,\alpha}^{q/2} + E_{q,\alpha}(x), \\ x^{-q/2} w^{q/4 - \alpha q/2} M^{q/2 - \alpha q/2} \xi_{q,\alpha}^{q/2} + E_{q,\alpha}(x), \end{cases} \tag{2.10.18}$$

where the constants in $\lesssim$ only depend on ς, α and q, and

$$\begin{aligned} \xi_{q,\alpha} &= \|X_{\cdot j}\|_{q,0} \|X_{\cdot k}\|_{q,\alpha} + \|X_{\cdot k}\|_{q,0} \|X_{\cdot j}\|_{q,\alpha}, \\ E_{q,\alpha}(x) &= \exp\{-C_{q,\alpha} (w M^{2-2\alpha} \xi_{4,\alpha}^2)^{-1} x^2\}. \end{aligned}$$

Following the argument in the proof of Theorem 2.8.1, let $L = \lfloor (\log w)/(\log 2) \rfloor$, $\varpi_l = 2^l$, $1 \leq l < L$, $\varpi_L = w$ and $\tau_l = M \varpi_l$ for $1 \leq l \leq L$. Let $\varpi_0 = 1$ and $\tau_0 = M$. Write

$$T - T_M = T - T_{Mw} + \sum_{l=1}^{L} V_{w,l}, \text{ where } V_{w,l} = T_{\tau_l} - T_{\tau_{l-1}}. \tag{2.10.19}$$

By the argument in Lemma 9 of Xiao and Wu [2012], we have

$$\begin{aligned} \|\mathbb{E}_0(T - T_{Mw})\|_{q/2} &\leq C_q M \sqrt{w} (\Delta_{0,q,j} \Delta_{Mw+1,q,k} + \Delta_{Mw+1,q,j} \Delta_{0,q,k}) \\ &\leq C_q M \sqrt{w} (Mw)^{-\alpha} \xi_{q,\alpha} \end{aligned} \tag{2.10.20}$$

for some constant $C_q > 0$. By Markov's inequality, for $x > 0$,

$$\mathbb{P}(|\mathbb{E}_0(T - T_{Mw})| \geq x) \leq \frac{C_q M^{q/2-\alpha q/2} w^{q/4-\alpha q/2} \xi_{q,\alpha}^{q/2}}{x^{q/2}}. \tag{2.10.21}$$

By the same argument for proving (2.10.20), we have

$$\|\mathbb{E}_0(V_{w,l})\|_{q/2} \leq C_q M \sqrt{w} \tau_l^{-\alpha} \xi_{q,\alpha}.$$

Let $c = q/4 - 1 - \alpha q/2$, $\lambda_l = 3l^{-2}\pi^{-2}$ if $1 \leq l \leq L/2$ and $\lambda_l = 3(L+1-l)^{-2}\pi^{-2}$ if $L/2 < l \leq L$. Let $I = \sum_{l=1}^{L} \varpi_l^c / \lambda_l^{q/2}$. Elementary calculations show that

$$I \leq C_5 \text{ for } c < 0 \text{ and } I \leq C_6 \varpi_L^c = C_6 w^c \text{ for } c > 0. \tag{2.10.22}$$

Notice that $\min_{l \geq 1} \lambda_l^2 \varpi_l^{2\alpha} > 0$ and $\xi_{4,\alpha} \leq 2\Psi_{4,\alpha}^2$. For $x \geq \sqrt{w} M \Psi_{q,\alpha}^2$, we have

$$II := \sum_{l=1}^{L} E_{q,\alpha}(\lambda_l \varpi_l^{\alpha} x) \lesssim E_{q,\alpha}(x). \tag{2.10.23}$$

By Corollary 1.8 of Nagaev [1979], it follows that

$$\mathbb{P}(|\mathbb{E}_0(V_{w,l})| \geq \lambda_l x) \leq \frac{C_1 w (M \varpi_l^{1/2} \tau_l^{-\alpha})^{q/2} \xi_{q,\alpha}^{q/2}}{\varpi_l (\lambda_l x)^{q/2}} + \exp\left(-\frac{C_2 (\lambda_l x)^2 \tau_l^{2\alpha}}{w M^2 \xi_{4,\alpha}^2}\right).$$

Since $\sum_{l=1}^{L} \lambda_l < 1$, by the above inequality,

$$\mathbb{P}(|\sum_{l=1}^{L} \mathbb{E}_0(V_{w,l})| \geq x) \leq \frac{C_3 w M^{q/2-\alpha q/2} \xi_{q,\alpha}^{q/2}}{x^{q/2}} I + C_4 II. \tag{2.10.24}$$

Putting (2.10.19), (2.10.21), (2.10.22), (2.10.23) and (2.10.24) together, we have (2.10.18).

Now it suffices to consider $\mathbb{P}(|\mathbb{E}_0(T_M)| \geq x)$. Observe $(Y_{bj,M} Y_{bk,M})_{b \text{ is odd}}$ are independent and so are $(Y_{bj,M}Y_{bk,M})_{b \text{ is even}}$. By Corollary 1.7 of Nagaev [1979], for any $J > 1$,

$$\begin{aligned}\mathbb{P}(|\mathbb{E}_0(T_M)| \geq x) &\leq \sum_{b=1}^{w} \mathbb{P}\left(|\mathbb{E}_0(Y_{bj,M}Y_{bk,M})| \geq \frac{x}{2J}\right) \\ &\quad + 2\left(\frac{\sum_{b=1}^{w} \|\mathbb{E}_0(Y_{bj,M}Y_{bk,M})\|_{q/2}^{q/2}}{Jx^{q/2}}\right)^J \\ &\quad + 4\exp\left\{-\frac{C_q x^2}{\sum_{b=1}^{w} \|\mathbb{E}_0(Y_{bj,M}Y_{bk,M})\|_2^2}\right\}.\end{aligned}$$

Note that $\|Y_{bj,M}\|_q \leq C_q\sqrt{M}\|X_{\cdot j}\|_{q,0}$. Hence for $1 \leq b \leq w$, $1 \leq j,k \leq p$ and $q \geq 4$,

$$\begin{aligned}\|\mathbb{E}_0(Y_{bj,M}Y_{bk,M})\|_{q/2} &\leq 2\|Y_{bj,M}Y_{bk,M}\|_{q/2} \\ &\leq 2\|Y_{bj,M}\|_q\|Y_{bk,M}\|_q \leq C_q M\|X_{\cdot j}\|_{q,0}\|X_{\cdot k}\|_{q,0}.\end{aligned}$$

Since

$$\mathbb{E}|Y_{bj,M}Y_{bk,M}| \leq \|Y_{bj,M}\|_2\|Y_{bk,M}\|_2 \leq M\|X_{\cdot j}\|_{2,0}\|X_{\cdot k}\|_{2,0} \leq \frac{x}{\sqrt{w}},$$

we have

$$\begin{aligned}\mathbb{P}(|\mathbb{E}_0(T_M)| \geq x) &\leq \sum_{b=1}^{w} \mathbb{P}(|Y_{bj,M}Y_{bk,M}| \geq x/(4J)) \\ &\quad + 4\exp\left(-\frac{C_q x^2}{wM^2\Psi_{4,0}^4}\right) \\ &\quad + 2\left(\frac{wM^{q/2}\|X_{\cdot j}\|_{q,0}^{q/2}\|X_{\cdot k}\|_{q,0}^{q/2}}{Jx^{q/2}}\right)^J.\end{aligned}$$

Recall that $M = O(n^{\varsigma})$ with $0 < \varsigma < 1$. Let $J = 1 + (2q-2)(q-4)^{-1}(1-\varsigma)^{-1}$. Since $x \geq \sqrt{w}M\|X_{\cdot j}\|_{q,0}\|X_{\cdot k}\|_{q,0}$, elementary calculations show that for sufficiently large n the second term in the above expression is no greater than $C_J wM\|X_{\cdot j}\|_{q,0}^{q/2}\|X_{\cdot k}\|_{q,0}^{q/2}/x^{q/2}$. As for the first term, we have

$$\mathbb{P}(|Y_{bj,M}Y_{bk,M}| \geq x/(4J)) \leq \mathbb{P}(|Y_{bj,M}| \geq \sqrt{x/(4J)}) + \mathbb{P}(|Y_{bk,M}| \geq \sqrt{x/(4J)}).$$

By Lemma 2.9.2, for $\alpha > 1/2 - 1/q$ and $\alpha < 1/2 - 1/q$, respectively, we have

$$\mathbb{P}(|Y_{bj,M}| \geq \sqrt{x}) \leq \begin{cases} C_{q,\alpha} x^{-q/2} M \|X_{\cdot j}\|_{q,\alpha}^q + C_{q,\alpha} \exp\left(-\frac{C_{q,\alpha} x}{M\|X_{\cdot j}\|_{2,\alpha}^2}\right), \\ C_{q,\alpha} x^{-q/2} M^{q/2-\alpha q} \|X_{\cdot j}\|_{q,\alpha}^q + C_{q,\alpha} \exp\left(-\frac{C_{q,\alpha} x}{M\|X_{\cdot j}\|_{2,\alpha}^2}\right). \end{cases}$$

A similar inequality holds for $\mathbb{P}(|Y_{bk,M}| \geq \sqrt{x})$. Let $\phi_{q,\alpha} = \|X_{\cdot j}\|_{q,\alpha}^q + \|X_{\cdot k}\|_{q,\alpha}^q$. Hence, it follows that for $\alpha > 1/2 - 1/q$ and $\alpha < 1/2 - 1/q$ respectively,

$$\mathbb{P}(|\mathbb{E}_0(T_M)| \geq x) \leq \begin{cases} C_{q,\alpha} x^{-q/2} wM \phi_{q,\alpha} + C_{q,\alpha} \exp\left(-\frac{C_{q,\alpha} x^2}{wM^2 \Psi_{4,\alpha}^4}\right), \\ C_{q,\alpha} x^{-q/2} wM^{q/2-\alpha q} \phi_{q,\alpha} + C_{q,\alpha} \exp\left(-\frac{C_{q,\alpha} x^2}{wM^2 \Psi_{4,\alpha}^4}\right). \end{cases} \tag{2.10.25}$$

Combining (2.10.18) and (2.10.25), and noticing that $\xi_{q,\alpha}^{q/2} \leq C_q \phi_{q,\alpha}$, it follows that

$$\mathbb{P}(|\mathbb{E}_0(T)| \geq x) \leq C_{q,\alpha} x^{-q/2} F_\alpha \phi_{q,\alpha} + C_{q,\alpha} \exp\left(-\frac{C_{q,\alpha} x^2}{wM^2 \Psi_{4,\alpha}^4}\right),$$

which implies (2.6.2) and (2.6.3) by the Bonferroni inequality by summing over j and k. □

Proof of Theorem 2.6.2. Let $T = \sum_{b=1}^w Y_{bj} Y_{bk}$. By Theorem 2.1 of Rio [2009], we have

$$\begin{aligned} \|\mathbb{E}_0 T\|_{q/2}^2 &\leq (q/2-1) \sum_{l=-\infty}^{wM} \|\mathcal{P}^l T\|_{q/2}^2 \\ &\leq (q/2-1) \sum_{l=-\infty}^{wM} \left(\sum_{b=1}^{w} \|\mathcal{P}^l Y_{bj} Y_{bk}\|_{q/2}\right)^2. \end{aligned} \tag{2.10.26}$$

By Theorem 3 in Wu [2011], $\|Y_{bj}\|_q \le (q-1)^{1/2}\sqrt{M}\|X_{\cdot j}\|_{q,0}$. Note that

$$\begin{aligned}\|\mathcal{P}^l Y_{bj}Y_{bk}\|_{q/2} &\le \|Y_{bj}Y_{bk} - Y_{bj,\{l\}}Y_{bk,\{l\}}\|_{q/2} \\ &\le \|Y_{bj}\|_q\|Y_{bk} - Y_{bk,\{l\}}\|_q + \|Y_{bj} - Y_{bj,\{l\}}\|_q\|Y_{bk,\{l\}}\|_q.\end{aligned}$$

Since $\|Y_{bk} - Y_{bk,\{l\}}\|_q \le \sum_{h=1+(b-1)M}^{bM} \delta_{h-l,q,k}$, we have by (2.10.26) that

$$\begin{aligned}\|\mathbb{E}_0 T\|_{q/2}^2 &\le (q/2-1)\sum_{l=-\infty}^{wM} \|\mathcal{P}^l T\|_{q/2}^2 \\ &\le (q-2)(q-1)wM^2\|X_{\cdot j}\|_{q,0}^2\|X_{\cdot k}\|_{q,0}^2.\end{aligned}$$

Let $R_{jk} = \mathbb{E}_0 T/(\sqrt{w}M)$. Similarly as the argument for proving Lemma 2.9.3, if $\gamma h \ge 2$, it follows that

$$\|R_{jk}\|_{\gamma h} \le (2\gamma h - 1)(2\gamma h)^{2\nu}\|X_{\cdot j}\|_{\psi_\nu,0}\|X_{\cdot k}\|_{\psi_\nu,0}.$$

Let $\tau_0 = (2e\gamma\|X_{\cdot j}\|_{\psi_\nu,0}^\gamma\|X_{\cdot k}\|_{\psi_\nu,0}^\gamma)^{-1}$. Notice that $-2\nu = 1 - 1/\gamma$. Then

$$\begin{aligned}\frac{t^h\|R_{jk}^\gamma\|_h^h}{h!} &\le \frac{t^h(2\gamma h-1)^{\gamma h}(2\gamma h)^{2\nu\gamma h}\|X_{\cdot j}\|_{\psi_\nu,0}^{\gamma h}\|X_{\cdot k}\|_{\psi_\nu,0}^{\gamma h}}{C_1(h/e)^h a_h^{-1}} \\ &\le \frac{a_h t^h(2\gamma h-1)^{\gamma h}}{C_1\tau_0^h(2\gamma h)^{\gamma h}} \le \frac{a_h t^h}{C_1\sqrt{e}\tau_0^h}.\end{aligned}$$

If $\gamma h < 2$, then $\|R_{jk}\|_{\gamma h} \le \|R_{jk}\|_2 \le 4^{2\nu}\sqrt{6}\|X_{\cdot j}\|_{\psi_\nu,0}\|X_{\cdot k}\|_{\psi_\nu,0}$. So we have

$$\begin{aligned}\mathbb{E}[\exp(tR_{jk}^\gamma)] &\le 1 + \sum_{1\le h<2/\gamma}\frac{t^h(4^{2\nu}\sqrt{6}\|X_{\cdot j}\|_{\psi_\nu,0}\|X_{\cdot k}\|_{\psi_\nu,0})^{\gamma h}}{h!} + \sum_{h\ge 2/\gamma}\frac{a_h t^h}{C_1\sqrt{e}\tau_0^h} \\ &\le 1 + C_\gamma\sum_{h=1}^\infty a_h\frac{t^h}{\tau_0^h} \le 1 + C_\gamma\frac{t/\tau_0}{(1-t/\tau_0)^{1/2}}.\end{aligned}$$

By choosing $t = \tau_0/2$, and applying the Markov inequality and the Bonferroni inequality, (2.6.4) and (2.6.5) are obtained. □

2.10.4 Proofs of Results in Section 2.10.1

Proof of Lemma 2.10.1. Let $P_1 = \mathbb{P}(|T_X - T_{X,m}|_\infty \geq y/2)$ and $P_2 = \mathbb{P}(|T_{X,m} - T_{Y,m}|_\infty \geq y/2)$. Lemma 2.9.1 and Theorem 2.8.1 imply that $P_1 \leq f_1^*(y)$. Write $T_{X,m} - T_{Y,m} = \sum_{b=1}^{w} \sum_{i \in S_b} X_{i,m}$. By Lemma 2.9.2 and Theorem 2.8.2, we also have $P_2 \leq f_2^*(y)$. Hence both cases with $\alpha > 1/2 - 1/q$ and $\alpha < 1/2 - 1/q$ of Lemma 2.10.1(i) follow in view of $\mathbb{P}(|T_X - T_{Y,m}|_\infty \geq y) \leq P_1 + P_2$.

The exponential moment case (ii) similarly follows from $P_1 \leq f_1^\diamond(y)$ and $P_2 \leq f_2^\diamond(y)$. $\square$

Proof of Lemma 2.10.2. Define $R_l = \max_{1 \leq j \leq p} \|M^{-1/2} Y_{bj,m}\|_l$, $1 < l \leq q$. Since $X_{ij,m} = \sum_{k=0}^{m} \mathcal{P}^{i-k} X_{ij}$, by Theorem 3.2 of Burkholder [1973],

$$\|\sum_{i=1}^{M} \mathcal{P}^{i-k} X_{ij}\|_l^2 \leq C_l \sum_{i=1}^{M} \|\mathcal{P}^{i-k} X_{ij}\|_l^2 \leq C_l M (\delta_{k,l,j})^2,$$

then we have

$$\|\sum_{i=1}^{M} X_{ij,m}\|_l \leq C_l \sum_{k=0}^{m} \|\sum_{i=1}^{M} \mathcal{P}^{i-k} X_{ij}\|_l \leq C_l M^{1/2} \Delta_{0,l,j}, \qquad (2.10.27)$$

which implies $R_l \leq C_l \Psi_{l,0}$. For $0 < \lambda < 1$ and the diagonal matrix $D = (d_{ij})_{i,j=1}^{p}$, define $u_{Y,m}(\lambda)$ as the infimum over all numbers $u > 0$ such that

$$\mathbb{P}(|M^{-1/2} d_{jj}^{-1/2} Y_{bj,m}| \leq u, 1 \leq b \leq w, 1 \leq j \leq p) \geq 1 - \lambda.$$

Also define $u_{Z,m}(\lambda)$ by the corresponding quantity for the analogue Gaussian case, namely with $Y_{b,m}$ replaced by $Z_{b,m}$ in the above definition. Let $u_m(\lambda) := u_{Y,m}(\lambda) \vee u_{Z,m}(\lambda)$. By Theorem 2.2 of Chernozhukov et al. [2013a], for all $\lambda \in (0,1)$,

$$\begin{aligned} & \sup_{t \in \mathbb{R}} \Big| \mathbb{P}(|D^{-1/2} T_{Y,m}/\sqrt{n}|_\infty \leq t) - \mathbb{P}(|D^{-1/2} T_{Z,m}/\sqrt{n}|_\infty \leq t) \Big| \\ \lesssim\ & w^{-1/8} (R_3^{3/4} \vee R_4^{1/2}) (\log(pw/\lambda))^{7/8} + w^{-1/2} (\log(pw/\lambda))^{3/2} u_m(\lambda) + \lambda, \end{aligned}$$

Now we shall find a bound on the function $u_m(\lambda)$. (i) By Lemma 2.9.2 and Theorem 2.8.2, we have

$$\begin{aligned}
&\mathbb{P}(|M^{-1/2}d_{jj}^{-1/2}Y_{bj,m}| > u \text{ for some } b, j) \le \mathbb{P}(|M^{-1/2}Y_{b,m}|_\infty > c^{1/2}u) \\
&\le \begin{cases} C_{q,\alpha}u^{-q}wM^{1-q/2}\Theta_{q,\alpha}^q + C_{q,\alpha}pw\exp\left(-\frac{C_{q,\alpha}u^2}{\Psi_{2,\alpha}^2}\right), & \alpha > 1/2 - 1/q \\ C_{q,\alpha}u^{-q}wM^{-\alpha q}\Theta_{q,\alpha}^q + C_{q,\alpha}pw\exp\left(-\frac{C_{q,\alpha}u^2}{\Psi_{2,\alpha}^2}\right), & \alpha < 1/2 - 1/q \end{cases},
\end{aligned}$$

which implies (2.10.7). For $u_{Z,m}(\lambda)$, since $M^{-1/2}Z_{bj,m} \sim N(0, \tilde{b}_{jj})$, we have

$$\mathbb{E}(\exp\{M^{-1}Z_{bj,m}^2/(4\tilde{b}_{jj})\}) \le C.$$

Hence

$$\begin{aligned}
\mathbb{P}(|M^{-1/2}d_{jj}^{-1/2}Z_{bj,m}| > u \text{ for some } b, j) &\le \sum_{b=1}^{w}\sum_{j=1}^{p}\mathbb{P}(|M^{-1/2}Z_{bj,m}| > d_{jj}^{1/2}u) \\
&\le Cpw\exp(-d_{jj}u^2/(4\tilde{b}_{jj})). \qquad (2.10.28)
\end{aligned}$$

With the assumption $c_1 \le \tilde{b}_{jj}/d_{jj} \le c_2$, $u_{Z,m}(\lambda) \le C\sqrt{\log(pw/\lambda)}$.

(ii) By Bonferroni inequality and Lemma 2.9.4,

$$\mathbb{P}(|M^{-1/2}d_{jj}^{-1/2}Y_{bj,m}| > u \text{ for some } b, j) \le C_\beta pw\exp\left\{-C_\beta\frac{u^\beta}{\Phi_{\psi_\nu,0}^\beta}\right\}, \qquad (2.10.29)$$

where $\beta = 2/(1 + 2\nu)$ and C_β is a constant that depends on β only. Combining (2.10.28) and (2.10.29), (2.10.8) follows. □

Proof of Lemma 2.10.3. By the definition of $T_{Z,m}$ and Z and (2.10.2),

$$\begin{aligned}
\Sigma^{Z,m} &:= \mathrm{Cov}(D^{-1/2}T_{Z,m}/\sqrt{n}) = \frac{Mw}{n}D^{-1/2}\tilde{B}D^{-1/2}, \\
\Sigma^{Z} &:= \mathrm{Cov}(D^{-1/2}Z) = D^{-1/2}\Sigma D^{-1/2}.
\end{aligned}$$

Let $S_{Mj} = \sum_{i=1}^{M} X_{ij}$ and $S_{Mj,m} = \sum_{i=1}^{M} X_{ij,m}$. By Theorem 3 in Wu [2011], $\|S_{Mj}\|_2 \leq M^{1/2}\Delta_{0,2,j}$, $\|S_{Mj,m}\|_2 \leq M^{1/2}\Delta_{0,2,j}$ and $\|S_{Mj} - S_{Mj,m}\|_2 \leq M^{1/2}\Delta_{m+1,2,j}$. Note $b_{jk} = M^{-1}\mathbb{E}(S_{Mj}S_{Mk})$ and $\tilde{b}_{jk} = M^{-1}\mathbb{E}(S_{Mj,m}S_{Mk,m})$. Then

$$\begin{aligned} |b_{jk} - \tilde{b}_{jk}| &= \frac{1}{M}|\mathbb{E}(S_{Mj}S_{Mk} - S_{Mj,m}S_{Mk,m})| \\ &\leq \frac{1}{M}\left(\|S_{Mj}\|_2\|S_{Mk} - S_{Mk,m}\|_2 + \|S_{Mk,m}\|_2\|S_{Mj} - S_{Mj,m}\|_2\right) \\ &\leq 2\Psi_{2,\alpha}\Psi_{2,0}m^{-\alpha}. \end{aligned}$$

Recall that $\sigma_{jk} = \sum_{l=-\infty}^{\infty} \gamma_{jk}(l)$ and

$$b_{jk} = M^{-1}\mathbb{E}(S_{Mj}S_{Mk}) = M^{-1}\sum_{l=-M}^{M}(M - |l|)\gamma_{jk}(l).$$

It follows that

$$\sigma_{jk} - b_{jk} = \sum_{|l|>M}\gamma_{jk}(l) + M^{-1}\sum_{l=-M}^{M}|l|\gamma_{jk}(l).$$

By $X_{ij} = \sum_{h=0}^{\infty}\mathcal{P}^{i-h}X_{ij}$, we have

$$\begin{aligned} |\gamma_{jk}(l)| &= |\sum_{h=0}^{\infty}\mathbb{E}[(\mathcal{P}^{-h}X_{0j})(\mathcal{P}^{-h}X_{lk})]| \\ &\leq \sum_{h=0}^{\infty}|\mathbb{E}[(\mathcal{P}^{-h}X_{0j})(\mathcal{P}^{-h}X_{lk})]| \leq \sum_{h=0}^{\infty}\delta_{h,2,j}\delta_{h+l,2,k}. \end{aligned}$$

Hence, it can be obtained that

$$\left|\sum_{|l|>M}\gamma_{jk}(l)\right| \leq 2\sum_{l=M+1}^{\infty}\sum_{h=0}^{\infty}\delta_{h,2,j}\delta_{h+l,2,k} \leq 2\Delta_{0,2,j}\Delta_{M+1,2,k},$$

and

$$\left|\frac{1}{M}\sum_{l=-M}^{M}|l|\gamma_{jk}(l)\right| \leq \frac{2}{M}\sum_{l=1}^{M}\sum_{\iota=k}^{M}\sum_{h=0}^{\infty}\delta_{h,2,j}\delta_{h+\iota,2,k} \leq \frac{2}{M}\Delta_{0,2,j}\sum_{l=1}^{M}\Delta_{l,2,k}.$$

Since $\Delta_{0,2,j} \leq \Psi_{2,0}$ and $\Delta_{m,2,j} \leq \Psi_{2,\alpha} m^{-\alpha}$, we have

$$\max_{1 \leq j,k \leq p} |b_{jk} - \sigma_{jk}| \leq \Psi_{2,\alpha} \Psi_{2,0} v(M).$$

Let $V = D^{-1/2} \Sigma D^{-1/2}$. Hence,

$$\begin{aligned} |\Sigma^{Z,m} - \Sigma^{Z}|_\infty &\leq \max_{1 \leq j \leq p} d_{jj}^{-1} (|\tilde{B} - B|_\infty + |B - \Sigma|_\infty) + \left(1 - \frac{Mw}{n}\right) |V|_\infty \\ &\leq \max_{1 \leq j \leq p} d_{jj}^{-1} \Psi_{2,\alpha} \Psi_{2,0} (m^{-\alpha} + v(M)) + C_2 wm/n. \end{aligned}$$

By Theorem 2 of Chernozhukov et al. [2014], the result follows. $\square$

CHAPTER 3

HIGH DIMENSIONAL LOCALLY STATIONARY PROCESSES

3.1 Introduction

It is still an open problem on whether an asymptotic theory for the estimate of second-order characteristics can be developed for high-dimensional and/or nonstationary processes via a general data-generating mechanism. Estimating second-order characteristics is of fundamental importance in many aspects of statistics. During the past decades, various cases of second-order statistic estimation have been studied for dependent and non-stationary processes. For example, in finance, Jacquier et al. [2004] concerned multivariate stochastic volatility models parameterized by time-varying covariance matrices with fat tails and correlated errors. In environmental science, Wikle and Hooten [2010] proposed nonlinear spatio-temporal dynamic models to accommodate quadratic interactions between processes which are critical for many geophysical (Kondrashov et al. [2005], Majda et al. [2005]) and ecological (Hooten and Wikle [2008]) processes. In electroencephalographic (EEG) studies, Prado et al. [2001] considered dynamic regression models with time-varying lag-lead structure to analyse multichannel EEG recordings of scalp electrical potential activity, and Park et al. [2014] developed multivariate locally stationary wavelet processes to capture the time-evolving scale-specific cross-dependence between components of the non-stationary signals. In essence, researchers face a numbers of challenges in solving these real-world problems: (i) nonlinear dynamics of data generating systems, (ii) temporally dependent and non-stationary observations, (iii) non-Gaussian distributions and/or (iv) high-dimensional data.

Motivated by these real-world applications, the primary goal of this chapter is to estimate second-order characteristics of a general class of locally stationary processes which can be possibly high-dimensional and non-Gaussian, and lay a theoretical foundation for estimation consistency. In Section 3.2, we shall introduce the framework of high-dimensional locally stationary processes and some concepts about functional dependence measures that are useful for establishing an asymptotic theory. Section 3.3 concerns the estimation of time-varying autocovariance matrix functions. Section 3.4 introduces the nonparametric estimation of time-varying spectral density and coherence matrices. In particular, we study the overlapped batched mean estimate of long-run covariance matrix functions in Section 3.4.1. In Section 3.5, we use the constrained ℓ_1 minimization approach to estimate the inverse of the spectral density matrix which can be used to identify the graphical structure for high-dimensional locally stationary processes. We provide in Section 3.6 Hanson–Wright-type inequalities for tail probabilities for non-stationary processes with finite polynomial moments. We relegate some proofs to Section 3.7. Throughout this chapter, we use r, s, t to denote time indexes and use i, j to denote dimension indexes.

3.2 High Dimensional Locally Stationary Processes

Consider the p-dimensional process $(X_{t,n})$ generated from the model (1.0.1). For convenience of notation, we shall abbreviate $X_{t,n}$ as X_t. The stochastic continuity condition (1.0.2) indicates that $X_{tj}(u) = g_j(u, \mathcal{F}^t)$ changes smoothly with respect to u. One has local stationarity in the sense that for $1 \le j \le p$, the subsequence $(X_{tj})_{t=s}^{s+r-1}$ of length r can be approximated by the stationary process $X_{tj}^* = g_j(s/n, \mathcal{F}^t)$, $t = s, \ldots, s+r-1$ in view of

$$\|X_{tj} - g_j(s/n, \mathcal{F}^t)\| \le \mathcal{K}r/n$$

if $\mathcal{K}r/n = o(1)$. In the stationary case where $G(\cdot, \cdot)$ does not depend on u, one can let $\mathcal{K} = 0$ in (1.0.2). With the condition (1.0.2), the form (1.0.1) provides a convenient framework for studying high-dimensional locally stationary processes and covers a large range of non-stationary time series models. In the scalar case in which $p = 1$, Wiener [1958] studied the stationary processes that can be coded by using i.i.d. random variables ε_t via a possibly nonlinear function G; see also Rosenblatt [1971], Priestley [1988a], Tong [1990], Wu [2005], Tsay [2005]) for the huge class of processes of this form. The representation $X_t = G(\mathcal{F}^t)$

also includes recursive model of the form $X_t = G(X_{t-1}, \varepsilon_t)$, which includes Markov chain models and nonlinear autoregressive models such as threshold autoregressive models, autoregressive models with conditional heteroscedasticity and exponential autoregressive models. By allowing the data-generating function G to change flexibly over time u, it extends a large number of existing stationary processes into their non-stationary counterparts in a natural way.

To develop an asymptotic theory for the time-varying second-order characteristic estimate, we need to introduce appropriate dependence measures. Assume that

$$\max_{1 \le j \le p} \sup_{u \in [0,1]} \|g_j(u, \mathcal{F}^0)\|_q < \infty \text{ for some } q \ge 1$$

Let ε'_t, $t \in \mathbb{Z}$ be an i.i.d. copy of ε_t, $t \in \mathbb{Z}$. For $t \ge 0$ and $1 \le j \le p$, we define the functional dependence measure

$$\delta_{t,q,j} = \sup_{u \in [0,1]} \|g_j(u, \mathcal{F}^t) - g_j(u, \mathcal{F}^{t,\{0\}})\|_q,$$

where $\mathcal{F}^{t,\{0\}} = (\ldots, \varepsilon_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_{t-1}, \varepsilon_t)$ is a coupled version of $\mathcal{F}^t$ with ε_0 in $\mathcal{F}^t$ replaced by ε'_0. Note that $\mathcal{F}^{t,\{0\}} = \mathcal{F}^t$ if $t < 0$. Hence, $\delta_{t,q,j} = 0$ for $t < 0$. In Chapter 2, (2.2.4) introduced a functional dependence measure for stationary processes in which the data-generating mechanism g_j does not vary with time u. In this chapter, the quantity $\delta_{t,q,j}$ measures the dependence of $g_j(u, \mathcal{F}^t)$ on the single input ε_0 over $u \in [0,1]$, which can be viewed as the uniform dependence measure with lag t for locally stationary processes.

Equipped with the above dependence measures, we define in the following the dependence adjusted norm

$$\|X_{\cdot j}\|_{q,\alpha} = \sup_{m \ge 0} (m+1)^{\alpha} \Delta_{m,q,j}, \ \alpha \ge 0, \ \text{where } \Delta_{m,q,j} = \sum_{t=m}^{\infty} \delta_{t,q,j}, \tag{3.2.1}$$

and the associated overall and uniform dependence adjusted norms

$$\Theta_{q,\alpha} = \left(\sum_{j=1}^{p} \|X_{\cdot j}\|_{q,\alpha}^{q/2} \right)^{2/q}, \quad \Phi_{q,\alpha} = \max_{1 \le j \le p} \|X_{\cdot j}\|_{q,\alpha}. \tag{3.2.2}$$

The quantities $\Theta_{q,\alpha}$ and $\Phi_{q,\alpha}$ are used to account for high dimensionality and they may be unbounded functions in terms of the dimension p.

Before stating some main results, we shall provide two examples of high-dimensional locally stationary time series for which one can bound $\Theta_{q,\alpha}$ and $\Phi_{q,\alpha}$, a key step in applying these theorems.

Example 3.2.1 (High-dimensional locally stationary linear processes). *Let ε_{tj}, $t, j \in \mathbb{Z}$ be i.i.d. random variables with mean 0, variance 1 and finite q-th moment for some $q > 2$. Let $A_m(u) = (a_{m,ij}(u))_{i,j=1}^{p}$ be $p \times p$ matrices with real entries such that $a_{m,ij}(u) \in \mathcal{C}^1[0,1]$, $m \ge 0$, $1 \le i, j \le p$, and $\sup_{u \in [0,1]} \sum_{m=0}^{\infty} tr[A_m(u)A_m(u)^\top] < \infty$. Write $\varepsilon_t = (\varepsilon_{t1}, \ldots, \varepsilon_{tp})^\top$ and define the p-dimensional non-stationary linear processes*

$$X_t(u) = \sum_{m=0}^{\infty} A_m(u)\varepsilon_{t-m}. \tag{3.2.3}$$

By Kolmogorov's three series theorem, the process (3.2.3) is well-defined for all $u \in [0,1]$, and the assumptions on $A_m(u)$ ensure the local stationarity. Let $A_{m,j}(u)$ be the j-th row of $A_m(u)$. By Rosenthal's inequality (Rosenthal [1970]), the functional dependence measure

$$\delta_{t,q,j} = \sup_{u \in [0,1]} \|A_{t,j}(u)\varepsilon_0\|_q \le (q-1)^{1/2} \sup_{u \in [0,1]} |A_{t,j}(u)|_2 \|\varepsilon_{00}\|_q. \tag{3.2.4}$$

If there exist $w > 1$ and $K > 0$ such that $\sup_{u \in [0,1]} |A_{t,j}(u)|_2 \le K(t+1)^{-w}$ holds for all $t \ge 0$ and $1 \le j \le p$, then with $\alpha = w - 1$, we have

$$\Phi_{q,\alpha} \le C_1 K \|\varepsilon_{00}\|_q \text{ and } \Theta_{q,\alpha} \le C_2 K p^{2/q} \|\varepsilon_{00}\|_q,$$

where the constants C_1 and C_2 both only depend on q and α.

Example 3.2.2 (Time-varying threshold vector autoregressive (TVTVAR) models)**.** *Tong [1990] studied one-dimensional threshold autoregressive models. Here we shall consider high-dimensional time-varying threshold vector AR models. Let ε_{tj}, $t, j \in \mathbb{Z}$ be i.i.d random variables with mean 0, variance 1 and finite q-th moment for some $q > 2$. Write $\varepsilon_t = (\varepsilon_{t1}, \ldots, \varepsilon_{tp})^\top$. For a vector $v = (v_1, \ldots, v_p)^\top$, define $v^+ = (v_1^+, \ldots, v_p^+)^\top$ where $v_j^+ = \max\{v_j, 0\}$. Consider the model*

$$X_t(u) = A(u)[X_{t-1}(u)]^+ + B(u)[-X_{t-1}(u)]^+ + \varepsilon_t =: G_{\varepsilon_t}(u, X_{t-1}(u)), \tag{3.2.5}$$

where $A(u) = (a_{jk}(u))_{j,k=1}^p$ and $B(u) = (b_{jk}(u))_{j,k=1}^p$ are transition matrices satisfying $a_{ij}(u) \in \mathcal{C}^1[0,1]$, $b_{jk}(u) \in \mathcal{C}^1[0,1]$, $1 \leq j, k \leq p$. Let $G_{\varepsilon_t}(\cdot,\cdot) = (G^{(1)}_{\varepsilon_t}(\cdot,\cdot), \ldots, G^{(p)}_{\varepsilon_t}(\cdot,\cdot))^\top$. We abbreviate (3.2.5) as $X_t(u) = G_{\varepsilon_t}(u, X_{t-1}(u))$. Let $\rho_j(u) = \sum_{k=1}^p (|a_{jk}(u)| \vee |b_{jk}(u)|)$ and $\rho_j = \sup_{u\in[0,1]} \rho_j(u)$. For any $u \in [0,1]$, note that

$$|G^{(j)}_{\varepsilon_t}(u, \mathbf{x}) - G^{(j)}_{\varepsilon_t}(u, \mathbf{y})| \leq \rho_j(u)|\mathbf{x} - \mathbf{y}|_\infty. \tag{3.2.6}$$

Hence $|G_{\varepsilon_t}(u, \mathbf{x}) - G_{\varepsilon_t}(u, \mathbf{y})|_\infty \leq \max_j \rho_j(u)|\mathbf{x} - \mathbf{y}|_\infty$. Assume that

$$\rho := \sup_{u\in[0,1]} \max_{1\leq j\leq p} \rho_j(u) < 1.$$

Let $K_0 = \sup_{u\in[0,1]} |G_{\varepsilon_t}(u, \mathbf{0})|_\infty = \max_{1\leq j\leq p} |\varepsilon_{0j}|$. By the arguments for Theorem 2 in Wu and Shao [2004], for any $u \in [0,1]$, (3.2.5) admits a unique stationary solution and iterations of (3.2.5) lead to

$$X_t(u) = H(u, \mathcal{F}^t) = (h_1(u, \mathcal{F}^t), \ldots, h_p(u, \mathcal{F}^t))^\top$$

with $\mathcal{F}^t = (\ldots, \varepsilon_{t-1}, \varepsilon_t)$*. It also implies the* ℓ^∞ *functional dependence measure*

$$\sup_{u\in[0,1]} \| |H(u, \mathcal{F}^t) - H(u, \mathcal{F}^{t,\{0\}})|_\infty \|_q \le C_q \rho^t \|K_0\|_q/(1-\rho), \tag{3.2.7}$$

where the constant C_q *only depends on* q*. Hence by (3.2.5) and (3.2.7), the functional dependence measure for the* j*-th component process satisfies*

$$\delta_{t,q,j} := \sup_{u\in[0,1]} \|h_j(u, \mathcal{F}^t) - h_j(u, \mathcal{F}^{t,\{0\}})\|_q \le C_q \rho_j \rho^t \|K_0\|_q/(1-\rho).$$

Since $\rho < 1$*, the corresponding dependence adjusted norm* $\|X_{\cdot j}\|_{q,\alpha} < \infty$ *for all* $\alpha \ge 0$*, and the quantities* $\Theta_{q,\alpha}$ *and* $\Phi_{q,\alpha}$ *can be computed accordingly.*

3.3 Estimation of Autocovariance Matrix Functions

Autocovariances play an important role in almost every aspect of time series analysis. In the non-stationary setting, our goal is to estimate the time-varying autocovariance matrices

$$\Gamma_l(u) = \mathbb{E} X_0(u) X_l(u)^\top, \text{ where } X_t(u) = G(u, \mathcal{F}^t). \tag{3.3.1}$$

For a fixed $u \in [0, 1]$, we can use observations X_t with t close to $\lfloor nu \rfloor$ to have the estimator $\hat{\Gamma}_l(u)$. Specifically, let b_n be a bandwidth sequence satisfying

$$b_n \to 0, \text{ and } nb_n \to \infty.$$

Let $T_1(u) = \lfloor nu \rfloor - \lfloor nb_n \rfloor + 1$, $T_2(u) = \lfloor nu \rfloor + \lfloor nb_n \rfloor$ and $M = 2\lfloor nb_n \rfloor$. For $u \in [b_n, 1 - b_n]$ and $0 \le l < M$, a natural estimator of $\Gamma_l(u)$ from the sample X_t, $t = T_1(u), \ldots, T_2(u)$ is given by

$$\hat{\Gamma}_l(u) = \frac{1}{M} \sum_{r=l+T_1(u)}^{T_2(u)} X_{r-l} X_r^\top, \tag{3.3.2}$$

and $\hat{\Gamma}_l(u) = \hat{\Gamma}_{-l}^{\top}(u)$ for $l < 0$. In the special case of stationary processes, we estimate the autocovariance matrices using the entire sample $X_1, \ldots, X_n$ by letting $b_n = 1$ and $M = n$.

We concern the maximum deviation over the range $0 \le l < M$, i.e.,

$$\psi_n^\star := \max_{0 \le l < M} \sup_{u \in [b_n, 1-b_n]} |\hat{\Gamma}_l(u) - \Gamma_l(u)|_\infty. \tag{3.3.3}$$

For univariate stationary processes with $p = 1$, the uniform convergence of autocovariances is closely related to the estimation of orders of ARMA processes or linear systems in general. The pioneer works in this direction were given by E. J. Hannan and his collaborators; see, for example, Hannan [1974] and An et al. [1982]. Readers can find a summary of these works and references therein in Section 5.3 of Hannan and Deistler [1988]. Giurcanu and Spokoiny [2004] obtained an upper bound of $\max_{0 \le l < M} |\hat{\Gamma}_l - \Gamma_l|$ for Gaussian stationary processes and also extended to the locally stationary case (cf. Propositions 3.2 and 3.4 therein). More recently Xiao and Wu [2014] considered the maximum deviation for sample autocovariances of univariate stationary processes in the polynomial tail case.

Since the process X_i can be possibly nonlinear, non-stationary, non-Gaussian and high-dimensional, it can be quite involved to derive an upper bound for $\psi_n^\star$. Theorem 3.3.1 below provides a non-asymptotic bound for the stochastic part

$$\psi_n := \max_{0 \le l < M} \sup_{u \in [b_n, 1-b_n]} |\hat{\Gamma}_l(u) - \mathbb{E}\hat{\Gamma}_l(u)|_\infty.$$

In our setting, with the framework of functional dependence measures, it turns out that we can have a close form of the upper bound in the form of (3.3.4). The convergence rate depends in a subtle way on the temporal dependence characterized by α [cf. (3.2.1)], the overall and the uniform dependence adjusted norms $\Theta_{q,\alpha}$ and $\Phi_{q,\alpha}$, the same size n and the dimension p.

Theorem 3.3.1. *Assume that* $\mathbb{E}(X_t) = 0$ *and* $\Theta_{q,\alpha} < \infty$ *for some* $q > 4$ *and* $\alpha > 0$. *Let* b_n *be a bandwidth sequence satisfying* $b_n \to 0$ *and* $nb_n \to \infty$. *Define* $M = 2\lfloor nb_n \rfloor$. *Assume*

$M = O(n^\eta)$ for some $0 < \eta < 1$. Then for any $0 < \beta < 1$ and $x \geq M^{1/2}\Phi_{q,\alpha}^2$, we have

$$\mathbb{P}(M\psi_n \geq x) \lesssim \frac{nH_M\Theta_{q,\alpha}^q}{x^{q/2}} + nMp^2 \exp\left(-\frac{C_{q,\alpha}x^2}{M\Phi_{4,\alpha}^4}\right), \tag{3.3.4}$$

where $H_M = M(\log M)^{q+1}$ for $\alpha > 1/2 - 2/q$ and $H_M = M^{q/4-\alpha\beta q/2}$ for $\alpha \leq 1/2 - 2/q$, and the constant in $\lesssim$ depends on q, α and β.

Corollary 3.3.2. *Let $\psi_n^\star$ be the maximum deviation defined in (3.3.3). Under the condition (1.0.2) and the assumptions of Theorem 3.3.1, we have*

$$\psi_n^\star = O_{\mathbb{P}}\left(\frac{(nH_M)^{2/q}}{M}\Theta_{q,\alpha}^2 + \sqrt{\frac{\log(pn)}{M}}\Phi_{q,\alpha}^2 + \mathcal{U}_{M,n} + M^{-\min\{\alpha,1\}}\Phi_{2,0}\Phi_{2,\alpha}\right), \tag{3.3.5}$$

where H_M is the same as defined in Theorem 3.3.1 and

$$\mathcal{U}_{M,n} = \min\left\{\frac{\mathcal{K}M}{n}\Phi_{2,0}, \left(\frac{\mathcal{K}M}{n}\right)^{\frac{\alpha}{1+\alpha}} \frac{\Phi_{2,0}\Phi_{2,\alpha}^{1/(1+\alpha)}}{M^{1/2}}\right\}.$$

Remark 7. *If the function $G(u, \cdot)$ in (1.0.1) does not depend on u, i.e., $X_t = G(\mathcal{F}^t)$ is a stationary process, the autocovariance matrices are estimated based on the observations X_t, $1 \leq t \leq n$, with $b_n = 1$ and $M = n$ in (3.3.2). The convergence rate can be derived similarly without extra technical difficulties. For the special case of univariate stationary processes, i.e., for $p = 1$, the proof of (3.3.4) implies $\psi_n = O_{\mathbb{P}}(\sqrt{\log n/n})$ under the conditions $\|X_{\cdot 1}\|_{q,\alpha} < \infty$ and $\alpha q > 2$, which is consistent with the result in Theorem 2 of Xiao and Wu [2014].*

Next we provide the proof of Theorem 3.3.1. The proof of Corollary 3.3.2 is deferred to Section 3.7.

Proof of Theorem 3.3.1. For $1 \le i, j \le p$, $l \ge 0$ and $b_n \le u \le 1 - b_n$, let

$$\hat{\Gamma}_{l,ij}(u) = \frac{1}{M} \sum_{r=l+T_1(u)}^{T_2(u)} X_{(r-l)i} X_{rj}.$$

Define $u_t = t/n$ for $t = \lfloor nb_n \rfloor, \ldots, \lfloor n(1-b_n) \rfloor$. Let

$$T = \lceil (\lfloor n(1-b_n) \rfloor - \lfloor nb_n \rfloor + 1)/M \rceil,$$

and for $k = 1, \ldots, T$, let

$$\mathcal{D}_k = \{\lfloor nb_n \rfloor + (k-1)M, \ldots, (\lfloor nb_n \rfloor + kM - 1) \wedge \lfloor n(1-b_n) \rfloor\}$$

and $t_k = \min \mathcal{D}_k = \lfloor nb_n \rfloor + (k-1)M$. Since

$$\begin{aligned} \max_{t \in \mathcal{D}_k} M |\mathbb{E}_0 \hat{\Gamma}_{l,ij}(u_t)| &\le \max_{t \in \mathcal{D}_k} \left| \mathbb{E}_0 \sum\nolimits_{r=l+T_1(u_{t_k})}^{l+T_1(u_t)} X_{(r-l)i} X_{rj} \right| \\ &\quad + \max_{t \in \mathcal{D}_k} \left| \mathbb{E}_0 \sum\nolimits_{r=l+T_1(u_{t_k})}^{T_2(u_t)} X_{(r-l)i} X_{rj} \right| \\ &\le 2 \max_{l \le t \le 2M-1} \left| \mathbb{E}_0 \sum\nolimits_{r=l+T_1(u_{t_k})}^{T_1(u_{t_k})+t} X_{(r-l)i} X_{rj} \right| =: 2U_{kl,ij}, \quad (3.3.6) \end{aligned}$$

by the Bonferroni inequality,

$$\begin{aligned} \mathbb{P}(M\psi_n \ge x) &\le \sum_{k=1}^{T} \sum_{i,j=1}^{p} \sum_{l=0}^{M-1} \mathbb{P}\left(\max_{t \in \mathcal{D}_k} |M \mathbb{E}_0 \hat{\Gamma}_{l,ij}(u_t)| \ge x \right) \\ &\le \sum_{k=1}^{T} \sum_{i,j=1}^{p} \sum_{l=0}^{M-1} \mathbb{P}\left(U_{kl,ij} \ge x/2 \right). \qquad (3.3.7) \end{aligned}$$

First consider the case where $k = 1$. Since $T_1(u_{t_1}) = 1$, we have

$$U_{1l,ij} = \max_{l+1 \le t \le 2M} \left| \mathbb{E}_0 \sum\nolimits_{r=l+1}^{t} X_{(r-l)i} X_{rj} \right|.$$

For $0 \le l < M$, let $Y_r(l) = \mathbb{E}_0 X_{(r-l)i} X_{rj}$ and $Z_t(l)$ be the partial sum given by $Z_t(l) = \sum_{r=l+1}^{t} Y_r(l)$. Let $d = \lceil \log_2(2M) \rceil$. For any positive integer $h \le 2^d$, write its dyadic expansion as $h = 2^{s_1} + \cdots + 2^{s_L}$, where $0 \le s_L < \cdots < s_1 \le d$. Define $h(\ell) = 2^{s_1} + \cdots + 2^{s_\ell}$, for $\ell = 1, 2, \ldots, L$, and $h(0) = 0$. We then have

$$\begin{aligned} |Z_h(l)| &\le \sum_{\ell=1}^{L} |Z_{h(\ell)}(l) - Z_{h(\ell-1)}(l)| \\ &\le \sum_{\ell=1}^{L} \max_{1 \le v \le 2^{d-s_\ell}} |Z_{2^{s_\ell} v}(l) - Z_{2^{s_\ell}(v-1)}(l)| \\ &\le \sum_{s=0}^{d} \max_{1 \le v \le 2^{d-s}} |Z_{2^s v}(l) - Z_{2^s(v-1)}(l)|. \end{aligned} \tag{3.3.8}$$

Let $(\lambda_s)_{s=0}^d$ be a positive sequence such that $\sum_{s=0}^d \lambda_s \le 1$; specifically, let $\lambda_s = (s+1)^{-2}/(\pi^2/3)$ if $0 \le s \le d/2$ and $\lambda_s = (d+1-s)^{-2}/(\pi^2/3)$ if $d/2 < s \le d$. By (3.3.8) and the Bonferroni inequality,

$$\begin{aligned} \mathbb{P}\left(U_{1l,ij} \ge x\right) &\le \mathbb{P}\left(\max_{1 \le t \le 2M} |Z_t(l)| \ge x\right) \\ &\le \sum_{s=0}^{d} \mathbb{P}\left(\max_{1 \le v \le 2^{d-s}} |Z_{2^s v}(l) - Z_{2^s(v-1)}(l)| \ge \lambda_s x\right) \\ &\le \sum_{s=0}^{d} 2^{d-s} \max_{1 \le v \le 2^{d-s}} \mathbb{P}\left(|Z_{2^s v}(l) - Z_{2^s(v-1)}(l)| \ge \lambda_s x\right). \end{aligned} \tag{3.3.9}$$

Since $\min_{s \ge 0} \lambda_s^2 2^s > 0$, by Theorem 3.6.3, given v, $0 \le l < M$ and $0 < \beta < 1$, for $x \ge M^{1/2} \|X_{\cdot i}\|_{q,\alpha} \|X_{\cdot j}\|_{q,\alpha}$,

$$\begin{aligned} &\mathbb{P}(|Z_{2^s v}(l) - Z_{2^s(v-1)}(l)| \ge \lambda_s x) \\ \lesssim\ & \frac{[2^s + 2^{s(q/4 - \alpha\beta q/2)}] \|X_{\cdot i}\|_{q,\alpha}^{q/2} \|X_{\cdot j}\|_{q,\alpha}^{q/2}}{(\lambda_s x)^{q/2}} + \exp\left(-\frac{C_{q,\alpha} \lambda_s^2 x^2}{2^s \Phi_{4,\alpha}^4}\right). \end{aligned} \tag{3.3.10}$$

We can deal with $U_{kl,ij}$ for $k > 1$ similarly and obtain the same upper bound for the tail probability. Let

$$\mathrm{I} = \sum_{s=0}^{d} \frac{2^{d-s}[2^s + 2^{s(q/4-\alpha\beta q/2)}]}{\lambda_s^{q/2}}, \quad \mathrm{II} = \sum_{s=0}^{d} \exp\left(-\frac{C_{q,\alpha}\lambda_s^2 x^2}{2^s \Phi_{4,\alpha}^4}\right).$$

Notice that $T \asymp n/M$. By (3.3.7), (3.3.9) and (3.3.10), we have for $x \geq M^{1/2}\Phi_{q,\alpha}^2$,

$$\mathbb{P}(M\psi_n \geq x) \leq C_{q,\alpha} n \frac{\Theta_{q,\alpha}^q}{x^{q/2}} \cdot \mathrm{I} + C_{q,\alpha} n M p^2 \cdot \mathrm{II}. \tag{3.3.11}$$

By the choice of λ_s and d, when $\alpha > 1/2 - 2/q$, we can choose β such that $(1/2-2/q)/\alpha \leq \beta < 1$ and hence $\mathrm{I} = O(M(\log M)^{q+1})$. If $\alpha \leq 1/2 - 2/q$, elementary calculation shows $\mathrm{I} = O(M^{q/4-\alpha\beta q/2})$. Notice that $\min_{s\geq 0} \lambda_s^2 2^s > 0$. Then for $x \geq \sqrt{M}\Phi_{4,\alpha}^2$, we have

$$\mathrm{II} \leq \sum_{s=0}^{d} \exp\left(-\frac{C_{q,\alpha}\lambda_s^2 2^s x^2}{M\Phi_{4,\alpha}^4}\right) \lesssim \exp\left(-\frac{C_{q,\alpha} x^2}{M\Phi_{4,\alpha}^4}\right).$$

By (3.3.11) and the above bounds of I and II, (3.3.4) follows. □

3.4 Spectral Density and Coherence Matrix

Spectral analysis is a fundamental tool to have an insight into the cyclical behavior of time series. The spectrum was traditionally considered as an adequate description of the frequency domain characteristics of stationary processes. Estimation of spectral density has been extensively studied in the univariate stationary case; see for example Anderson [1971], Priestley [1981], Rosenblatt [1985], Newey and West [1987] among many others. Coherence, also known as the time series analogue in the frequency domain of the standard correlation coefficient, measures the linear relationship between a pair of time series as a function of frequency; see, for example, Brillinger [1975] and Brockwell and Davis [1991]. Since non-stationary data with a time-varying structural change is increasingly common in diverse fields, time-varying spectrum and coherence has been a popular tool to reveal the dynamics of the underlying mechanism. For example, in EEG data analysis, it has been widely used to measure brain functional connectivity; see Liu et al. [2010], Simpson et al. [2013], Lindquist et al. [2014] among others.

Various models and methods to obtain the time-varying spectra and coherence for non-stationary processes have been investigated in the literature. Priestley and Tong [1973] concerned the cross-spectrum and coherence between oscillatory processes stemming from a time-varying spectral representation, which was later investigated by Dahlhaus [2000a] allowing for rigorous asymptotic considerations. Ombao et al. [2001] proposed an method based on the smooth localized complex exponentials to select the span which can be used to obtain the smoothed estimates of the time-varying spectra and coherence. Sanderson et al. [2010] and Park et al. [2014] considered the problem of estimating time-evolving cross-dependence in a collection of locally stationary wavelet processes. Ombao and Bellegem [2008] developed a coherence estimation procedure using time-localized linear filtering. Many of the previous results require restrictive structural condition on the underlying processes such as linearity or Gaussianity.

More formally, under the framework (1.0.1), the spectral density matrix evolving over time u is defined by

$$F(u,\theta) = \frac{1}{2\pi}\sum_{l\in\mathbb{Z}} \Gamma_l(u)\exp(-\iota l\theta), \text{ where } \iota = \sqrt{-1}.$$

and the time-varying coherence matrix is defined by

$$\mathcal{C}(u,\theta) = \text{diag}[F(u,\theta)]^{-1/2}F(u,\theta)\text{diag}[F(u,\theta)]^{-1/2}.$$

To estimate the time-varying spectral density matrix consistently, it is common to use the idea of smoothing. In particular, we consider the lag window estimate

$$\hat{F}(u,\theta) = \frac{1}{2\pi} \sum_{l=-m}^{m} K(l/m)\hat{\Gamma}_l(u)\exp(-\iota l\theta). \tag{3.4.1}$$

where $\hat{\Gamma}_l(u)$ is the estimate of the autocovariance matrix function with lag l defined in (3.3.2), m is the bandwidth satisfying the natural conditions $m := m_M \to \infty$ and $m/M \to 0$, and $K(\cdot)$ is a symmetric kernel function satisfying

$$K(0) = 0,\ |K(x)| \le 1,\ \text{and } K(x) = 0 \text{ for } |x| > 1.$$

We estimate the coherence matrix $\mathcal{C}(u,\theta)$ by

$$\hat{\mathcal{C}}(u,\theta) = \text{diag}[\hat{F}(u,\theta)]^{-1/2}\hat{F}(u,\theta)\text{diag}[\hat{F}(u,\theta)]^{-1/2}, \tag{3.4.2}$$

where $\hat{F}(u,\theta)$ is the estimate of the spectral density matrix given by (3.4.1).

Theorem 3.4.1. *Assume that* $\mathbb{E}(X_t) = 0$ *and* $\Theta_{q,\alpha} < \infty$ *for some* $q > 4$ *and* $\alpha > 0$*. Let* b_n *be a bandwidth sequence satisfying* $b_n \to 0$ *and* $nb_n \to \infty$*. Define* $M = 2\lfloor nb_n \rfloor$*. Let* $M = O(n^\eta)$ *and* $m = O(M^\beta)$ *for some* $0 < \eta, \beta < 1$*. Let*

$$\varphi_n = \sup_{u\in[b_n,1-b_n]} \max_{\theta} |\hat{F}(u,\theta) - \mathbb{E}\hat{F}(u,\theta)|_\infty.$$

Then for any $x \ge \sqrt{Mm}\Phi_{q,\alpha}^2$*, we have*

$$\mathbb{P}(M\varphi_n \ge x) \lesssim \frac{nmR_{M,m}\Theta_{q,\alpha}^q}{x^{q/2}} + np^2 \exp\left(-\frac{C_{q,\alpha}x^2}{Mm\Phi_{4,\alpha}^4}\right), \tag{3.4.3}$$

where $R_{M,m} = m$ *for* $\alpha > 1 - 2/q$*,* $R_{M,m} = m + m^{q/2-1-\alpha q/2}(\log M)^{q+1}$ *for* $1/2 - 2/q < \alpha < 1 - 2/q$ *and* $R_{M,m} = M^{q/4-1-\alpha q/2}m^{q/4}$ *for* $\alpha < 1/2 - 2/q$*, and the constant in* $\lesssim$ *depends on* q, α *and* β*.*

Corollary 3.4.2. *Under the condition (1.0.2) and the assumptions of Theorem 3.4.1, we have*

$$\begin{aligned}\varphi_n^\star &:= \sup_{u\in[b_n,1-b_n]} \max_\theta |\hat{F}(u,\theta) - F(u,\theta)|_\infty \\ &= O_{\mathbb{P}}\left(\frac{(nmR_{M,m})^{2/q}}{M}\Theta_{q,\alpha}^2 + \sqrt{\frac{m\log(pn)}{M}}\Phi_{q,\alpha}^2 + \mathcal{V}_{m,M,n} + \mathcal{W}_m\right), \qquad (3.4.4)\end{aligned}$$

with

$$\begin{aligned}\mathcal{V}_{m,M,n} &= \sqrt{m}\,\mathcal{U}_{M,n} + \Big(m^{-\alpha} + \frac{r(m)}{M}\Big)\Phi_{2,0}\Phi_{2,\alpha}, \\ \mathcal{W}_m &= \sup_u \sum_{l=1}^m (1 - K(l/m))|\Gamma_l(u)|_\infty,\end{aligned}$$

where $\mathcal{U}_{M,n}$ *is defined in Corollary 3.3.2,* $R_{M,m}$ *is defined in Theorem 3.4.1, and* $r(m) = 1$ *if* $\alpha > 1$, $r(m) = \log m$ *if* $\alpha = 1$ *and* $r(m) = m^{-\alpha+1}$ *if* $\alpha < 1$.

The term $\mathcal{W}_m$ depends on the kernel function. Its order of magnitude is determined by the smoothness of $K(\cdot)$ at zero. In particular, this term vanishes if $K(\cdot)$ is the rectangular kernel. If $1 - K(x) = O(|x|^\nu)$ at $x = 0$ for some $a > 0$ and $\sup_u |\Gamma_l(u)|_\infty = O(l^{-b})$ for some $b > 1$, then

$$\mathcal{W}_m = O(m^{-a} + m^{1-b}).$$

Corollary 3.4.3. *Let*

$$c_0 = \inf_u \min_\theta \min_{1\le j\le p} F_{jj}(u,\theta).$$

Under the assumptions of Theorem 3.4.1, we have

$$\rho_n := \sup_{u\in[b_n,1-b_n]} \max_\theta |\hat{\mathcal{C}}(u,\theta) - \mathcal{C}(u,\theta)|_\infty \le \frac{3\varphi_n^\star}{c_0} + \frac{2{\varphi_n^\star}^2}{c_0^2}. \qquad (3.4.5)$$

where $\varphi_n^\star = \sup_{u\in[b_n,1-b_n]} \max_\theta |\hat{F}(u,\theta) - F(u,\theta)|_\infty$.

3.4.1 Estimation of Long-run Covariance Matrix Functions

The estimation of long-run covariance matrix Σ is an important problem in statistical inference for stationary processes. It has been extensively studied in the scalar and the low-dimensional case; see Newey and West [1987], Politis et al. [1999], Bühlmann [2002], Lahiri [2003], Alexopoulos and Goldsman [2004]. Also, $\Sigma/(2\pi)$ is the value of the spectral density matrix of (X_i) at zero frequency. For locally stationary processes, the long-run covariance matrix is time-varying, which is given by

$$\Sigma(u) = \sum_{l=-\infty}^{\infty} \Gamma_l(u), \text{ where } \Gamma_l(u) = \mathbb{E}X_0(u)X_l(u)^\top.$$

We investigate the overlapped batched mean estimate

$$\hat{\Sigma}(u) = \frac{1}{(M-m+1)m} \sum_{r=T_1(u)}^{T_2(u)-m+1} \left(\sum_{t=r}^{r+m-1} X_t \right) \left(\sum_{t=r}^{r+m-1} X_t \right)^\top, \tag{3.4.6}$$

where $M = o(n)$ is the largest size of the segment $\{t : |t/n-u| \le b_n\}$, which is also the largest lag we concern for the estimate of autocovariance matrix functions in Section 3.3, $m = o(M)$ is the block size. Theorem 3.4.4 below concerns the convergence rate of the estimate (3.4.6), which also applies to the nonoverlapped batched mean estimate; see Remark 9.

Theorem 3.4.4. *Let the assumptions of Theorem 3.4.1 be satisfied. Define*

$$\phi_n = \sup_{u\in[b_n,1-b_n]} |\hat{\Sigma}(u) - \mathbb{E}\hat{\Sigma}(u)|_\infty. \tag{3.4.7}$$

Then for any $x \ge \sqrt{Mm}\Phi_{q,\alpha}^2$, we have

$$\mathbb{P}(M\phi_n \ge x) \lesssim \frac{nR_{M,m}\Theta_{q,\alpha}^q}{x^{q/2}} + \frac{np^2}{m} \exp\left(-\frac{C_{q,\alpha}x^2}{Mm\Phi_{4,\alpha}^4} \right), \tag{3.4.8}$$

where $R_{M,m}$ is defined in Theorem 3.4.1 and the constant in $\lesssim$ depends on q, α and β.

Corollary 3.4.5. *Under the condition (1.0.2) and the assumptions of Theorem 3.4.4, we have*

$$\begin{aligned} \phi_n^\star &:= \sup_{u \in [b_n, 1-b_n]} |\hat{\Sigma}(u) - \Sigma(u)|_\infty \\ &= O_{\mathbb{P}}\left(\frac{(nR_{M,m})^{2/q}}{M} \mathcal{R}_{M,m} \Theta_{q,\alpha}^2 + \sqrt{\frac{m \log(pn)}{M}} \Phi_{q,\alpha}^2 + \mathcal{Y}_{m,M,n} \right), \end{aligned} \tag{3.4.9}$$

with

$$\mathcal{Y}_{m,M,n} = \sqrt{m}\, \mathcal{U}_{M,n} + \frac{r(m)}{m} \Phi_{2,0} \Phi_{2,\alpha},$$

where $R_{M,m}$ is defined in Theorem 3.4.1, $\mathcal{U}_{M,n}$ is defined in Corollary 3.3.2 and $r(m)$ is defined in Corollary 3.4.2.

Remark 8. *Consider the lag window estimate (3.4.1) of spectral density matrices with the rectangular kernel $K(x) = 1$ for $|x| \le 1$. Comparing $\mathcal{V}_{m,M,n}$ in Corollary 3.4.2 and $\mathcal{Y}_{m,M,n}$ in Corollary 3.4.5, we can find the overlapped batched mean estimate leads to a slightly larger bias, but it has the additional advantage that it is always positive definite.*

Remark 9. *Following Section 2.6, an alternative but less efficient estimate of $\Sigma(u)$, which is also called non-overlapped batched mean estimate, is*

$$\tilde{\Sigma}(u) = \frac{1}{wm} \sum_{r=1}^{w} \Big(\textstyle\sum_{t \in \mathcal{B}_r(u)} X_t \Big) \Big(\textstyle\sum_{t \in \mathcal{B}_r(u)} X_t \Big)^\top, \tag{3.4.10}$$

where the window $\mathcal{B}_r(u) = \{T_1(u) + (r-1)m, \ldots, T_1(u) + rm - 1\}$ for $r = 1, \ldots, w$, the window size $|\mathcal{B}_r(u)| = m \to \infty$ and the number of windows $w = \lfloor M/m \rfloor$. A careful check of the proof of Theorem 3.4.4 indicates that the convergence rate of the nonoverlapped batched mean estimate has the same upper bound as shown in (3.4.8).

Next we provide the proofs of Theorem 3.4.1 and Corollary 3.4.3. The proof of Theorem 3.4.4 is similar to that of Theorem 3.4.1 and hence is omitted. Corollary 3.4.2 and Corollary 3.4.5 are proved in the Appendix.

Proofs of Theorem 3.4.1. Let

$$a_l(\theta) = K(l/m)\exp(-\iota l\theta), \text{ for } l \in \mathbb{Z}. \tag{3.4.11}$$

Then $|a_l(\theta)| \le 1$ if $|l| \le m$ and $a_l(\theta) = 0$ if $|l| > m$. For $1 \le i, j \le p$, we write

$$2\pi M\hat{F}_{ij}(u,\theta) = M\sum\nolimits_{l=-m}^{m} a_l(\theta)\hat{\Gamma}_{l,ij}(u) =: Q_{1,ij}(u,\theta) + Q_{2,ij}(u,\theta), \tag{3.4.12}$$

where

$$\begin{aligned} Q_{1,ij}(u,\theta) &= \sum\nolimits_{l=0}^{m}\sum\nolimits_{r=l+T_1(u)}^{T_2(u)} a_l(\theta)X_{(r-l)i}X_{rj}, \\ Q_{2,ij}(u,\theta) &= \sum\nolimits_{l=1}^{m}\sum\nolimits_{r=l+T_1(u)}^{T_2(u)} a_{-l}(\theta)X_{(r-l)j}X_{ri}. \end{aligned}$$

Let $\theta_h = \pi h/2m$ for $0 \le h \le 4m$. Since $\hat{F}_{ij}(u,\theta)$ is a trigonometric polynomial of order m for fixed u, by Lemma 3.7.1 with $\delta = 1$,

$$\max_{\theta} |\mathbb{E}_0\hat{F}(u,\theta)|_\infty \le 2 \max_{0\le h\le 4m} |\mathbb{E}_0\hat{F}(u,\theta_h)|_\infty. \tag{3.4.13}$$

By (3.4.12), (3.4.13) and the Bonferroni inequality,

$$\mathbb{P}(2\pi M\varphi_n \ge x) \le (4m+1)\sum_{i,j=1}^{p}\sum_{w=1}^{2}\max_h \mathbb{P}\left(\sup_{u\in[b_n,1-b_n]} |\mathbb{E}_0 Q_{w,ij}(u,\theta_h)| \ge x/4\right).$$

Let $u_t = t/n$ for $t = \lfloor nb_n\rfloor, \ldots, \lfloor n(1-b_n)\rfloor$ and let

$$T = \lceil(\lfloor n(1-b_n)\rfloor - \lfloor nb_n\rfloor + 1)/M\rceil.$$

for $k = 1, \ldots, T$, define

$$\mathcal{D}_k = \{\lfloor nb_n\rfloor + (k-1)M, \ldots, (\lfloor nb_n\rfloor + kM - 1) \wedge \lfloor n(1-b_n)\rfloor\},$$

$t_k = \min \mathcal{D}_k = \lfloor nb_n \rfloor + (k-1)M$. Given $1 \le k \le p$, θ_h and $1 \le i, j \le p$, let

$$W_{k,h,ij} = \max_{0 \le t \le 2M-1} \left| \sum\nolimits_{s=T_1(u_{t_k})}^{T_1(u_{t_k})+t} \sum\nolimits_{r=(s-m)\vee T_1(u_{t_k})}^{s} \mathbb{E}_0 a_{s-r}(\theta_h) X_{ri} X_{sj} \right|.$$

By similar arguments as (3.3.7), we have

$$\mathbb{P}\left(\sup_{u \in [b_n, 1-b_n]} |\mathbb{E}_0 Q_{1,ij}(u, \theta_h)| \ge x \right) \le \sum_{k=1}^{T} \sum_{i,j=1}^{p} \mathbb{P}\left(W_{k,h,ij} \ge x/2 \right). \tag{3.4.14}$$

First consider the case where $k = 1$. Then $T_1(u_{t_1}) = 1$. Let S_r be the partial sum given by

$$S_t = \sum\nolimits_{s=1}^{t} \sum\nolimits_{r=(s-m)\vee 1}^{s} \mathbb{E}_0 a_{s-r}(\theta_h) X_{ri} X_{sj}.$$

Let $b = \lfloor 2M/m \rfloor$, $d^\circ = \lceil \log_2(2M/m) \rceil$ and $d^* = \lceil \log_2(m-1) \rceil$. Let $(\lambda_s^\circ)_{s=0}^{d^\circ}$ and $(\lambda_s^*)_{s=0}^{d^*}$ be positive sequences such that $\sum_{s=0}^{d^\circ} \lambda_s^\circ \le 1$ and $\sum_{s=0}^{d^*} \lambda_s^* \le 1$; specifically, let $\lambda_s^\circ = (s+1)^{-2}/(\pi^2/3)$ if $0 \le s \le d^\circ/2$, $\lambda_s^\circ = (d^\circ + 1 - s)^{-2}/(\pi^2/3)$ if $d^\circ/2 < s \le d^\circ$, $\lambda_s^* = (s+1)^{-2}/(\pi^2/3)$ if $0 \le s \le d^*/2$, $\lambda_s^* = (d^* + 1 - s)^{-2}/(\pi^2/3)$ if $d^*/2 < s \le d^*$. We can write

$$\mathbb{P}(W_{1,h,ij} \ge x) \le \mathbb{P}\left(\max_{0 \le \ell \le b} \max_{0 \le s \le m-1} |S_{\ell m+s}| \ge x \right) =: \mathcal{S}_1 + \mathcal{S}_2, \tag{3.4.15}$$

where

$$\begin{aligned} \mathcal{S}_1 &= \mathbb{P}\left(\max_{0 \le \ell \le b} |S_{\ell m}| \ge x/2 \right), \\ \mathcal{S}_2 &= \sum_{\ell=0}^{b} \mathbb{P}\left(\max_{1 \le s \le m-1} |S_{\ell m+s} - S_{\ell m}| \ge x/2 \right). \end{aligned}$$

Using a similar argument as (3.3.9), we have

$$\mathcal{S}_1 \le \sum_{s=0}^{d^\circ} 2^{d^\circ - s} \max_{1 \le r \le 2^{d^\circ - s}} \mathbb{P}\left(|S_{2^s mr} - S_{2^s m(r-1)}| \ge \lambda_s^\circ x/2 \right). \tag{3.4.16}$$

By Theorem 3.6.1 and Theorem 3.6.2, we have for $x \geq \sqrt{Mm}\|X_{\cdot i}\|_{q,\alpha}\|X_{\cdot j}\|_{q,\alpha}$,

$$\mathbb{P}\left(|S_{2^s mr} - S_{2^s m(r-1)}| \geq \lambda_s^\circ x\right) \lesssim \frac{\varpi_s \|X_{\cdot i}\|_{q,\alpha}^{q/2}\|X_{\cdot j}\|_{q,\alpha}^{q/2}}{(\lambda_s^\circ x)^{q/2}} + \exp\left(-\frac{C_{q,\alpha}(\lambda_s^\circ x)^2}{2^s m^2 \Phi_{4,\alpha}^4}\right), \quad (3.4.17)$$

where $\varpi_s = 2^s m$ (resp. $2^s m^{q/2-\alpha q/2}$ or $2^{s(q/4-\alpha q/2)} m^{q/2-\alpha q/2}$) for $\alpha > 1 - 2/q$ (resp. $1/2 - 2/q < \alpha < 1 - 2/q$ or $\alpha < 1/2 - 2/q$), and the constant in $\lesssim$ only depends on q, α and β.

Let $Y_{sj} = \sum_{t=-\infty}^{s} a_{s-t}(\theta_h) X_{tj}$. Since $a_l(\theta_h) = 0$ for $|l| > m$, we can write

$$S_{\ell m+t} - S_{\ell m} = \sum_{r=\ell m}^{\ell m+t} \mathbb{E}_0 X_{ri} Y_{rj} := Z_{\ell,t}.$$

Using the argument when proving (3.3.9) once again, we obtain

$$\mathcal{S}_2 \leq \sum_{\ell=0}^{b}\sum_{s=0}^{d^*} 2^{d^*-s} \max_{1\leq t\leq 2^{d^*-s}} \mathbb{P}\left(|Z_{\ell,2^s r} - Z_{\ell,2^s(r-1)}| \geq \lambda_s^* x/2\right).$$

Theorem 3.6.1 and Theorem 3.6.2 also imply for $x \geq \sqrt{Mm}\|X_{\cdot i}\|_{q,\alpha}\|X_{\cdot j}\|_{q,\alpha}$,

$$\mathbb{P}(|Z_{\ell,2^s r} - Z_{\ell,2^s(r-1)}| \geq \lambda_s^* x) \lesssim \frac{\tau_s \|X_{\cdot i}\|_{q,\alpha}^{q/2}\|X_{\cdot j}\|_{q,\alpha}^{q/2}}{(\lambda_s^* x)^{q/2}} + \exp\left(-\frac{C_{q,\alpha}(\lambda_s^* x)^2}{2^s m \Phi_{4,\alpha}^4}\right), \quad (3.4.18)$$

where $\tau_s = m$ (resp. $m2^{s(q/2-\alpha q/2-1)}$ or $m^{q/4-\alpha q/2}2^{sq/4}$) for $\alpha > 1-2/q$ (resp. $1/2-2/q < \alpha < 1-2/q$ or $\alpha < 1/2-2/q$), and the constant in $\lesssim$ only depends on q and α.

We can deal with $W_{k,h,ij}$ for $k > 1$ and $\mathbb{E}_0 Q_{2,ij}(u,\theta_h)$ similarly. Let

$$\begin{aligned}
\mathrm{I} &= \sum_{s=0}^{d^\circ} \frac{2^{d^\circ - s}\varpi_s}{(\lambda_s^\circ)^{q/2}} + (b+1)\sum_{s=0}^{d^*} \frac{2^{d^*-s}\tau_s}{(\lambda_s^*)^{q/2}}, \\
\mathrm{II} &= \sum_{s=0}^{d^\circ} \exp\left(-\frac{C_{q,\alpha}(\lambda_s^\circ x)^2}{2^s m^2 \Phi_{4,\alpha}^4}\right) + m^2 \sum_{s=0}^{d^*} \exp\left(-\frac{C_{q,\alpha}(\lambda_s^* x)^2}{2^s m \Phi_{4,\alpha}^4}\right).
\end{aligned}$$

With (3.4.14), (3.4.15), (3.4.18) and (3.4.17), we have

$$\mathbb{P}(2\pi M\varphi_n \geq x) \lesssim \frac{nm}{M}\frac{\Theta_{q,\alpha}^q}{x^{q/2}} \cdot \mathrm{I} + np^2 \cdot \mathrm{II}. \tag{3.4.19}$$

By the choices of λ_s° and λ_s^* and the definitions of τ_s, ϖ_s, d° and d^*, elementary calculations show that for all cases,

$$\mathrm{I} = O(MR_{M,m}). \tag{3.4.20}$$

Notice that $\min_{s\geq 0}(\lambda_s^\circ)^2 2^s > 0$ and $\min_{s\geq 0}(\lambda_s^*)^2 2^s > 0$. Then for $x \geq \sqrt{Mm}\Phi_{4,\alpha}^2$, we have

$$\begin{aligned} \mathrm{II} &\lesssim \sum_{s=0}^{d^\circ} \exp\left(-\frac{C_{q,\alpha}(\lambda_s^\circ)^2 2^s x^2}{Mm\Phi_{4,\alpha}^4}\right) + m^2 \sum_{s=0}^{d^*} \exp\left(-\frac{C_{q,\alpha}(\lambda_s^*)^2 2^s x^2}{m^2\Phi_{4,\alpha}^4}\right) \\ &\lesssim \exp\left(-\frac{C_{q,\alpha}x^2}{Mm\Phi_{4,\alpha}^4}\right). \end{aligned} \tag{3.4.21}$$

By (3.4.19), (3.4.20) and (3.4.21), (3.4.8) follows. □

Proof of Corollary 3.4.3. Let

$$\chi_n = \sup_{u\in[b_n,1-b_n]} \max_\theta \max_{1\leq j\leq p} \left|\frac{\hat{F}_{jj}(u,\theta)}{F_{jj}(u,\theta)} - 1\right|.$$

Recall the definition of $\varphi_n^\star$. Then $\chi_n \leq \varphi_n^\star/c_0$. By the elementary inequality $|1-\sqrt{ab}| \leq |1-a| + (1-a)^2 + |1-b| + (1-b)^2$, we have

$$\begin{aligned} \rho_n &\leq \sup_{u\in[b_n,1-b_n]} \max_\theta \max_{1\leq j,k\leq p} \left|\frac{\hat{F}_{jk}(u,\theta) - F_{jk}(u,\theta)}{\sqrt{F_{jj}(u,\theta)F_{kk}(u,\theta)}}\right| \\ &\quad + \sup_{u\in[b_n,1-b_n]} \max_\theta \max_{1\leq j,k\leq p} \left|1 - \frac{\sqrt{\hat{F}_{jj}(u,\theta)\hat{F}_{kk}(u,\theta)}}{\sqrt{F_{jj}(u,\theta)F_{kk}(u,\theta)}}\right| \\ &\leq \frac{\varphi_n^\star}{c_0} + 2\chi_n + 2\chi_n^2 \leq \frac{3\varphi_n^\star}{c_0} + \frac{2{\varphi_n^\star}^2}{c_0^2}. \end{aligned}$$

□

3.5 Graphical Model and Inverse Spectral Density Matrix

The concept of graphical model for multivariate data has been extended to multivariate time series (e.g. Brillinger [1996], Dahlhaus [2000b], Timmer et al. [2000], Eichler [2012] among others). Each vertex of the graphical model represents a component of a process and each edge indicates the partial correlation of the two corresponding components given others. Hence, for stationary Gaussian processes, this induced graph is a conditional independence graph in the frequency domain, the properties of which has been investigated largely (cf. Dahlhaus [2000b], Fried and Didelez [2003], Bach and Jordan [2004], etc.). For non-Gaussian processes, it is termed partial correlation graph in Dahlhaus [2000b] using partial spectral coherence as a measure for the dependence between two marginal time series after removing the linear effects of some other components. Partial spectral coherence has been widely used in many real-world applications; see for example Gather et al. [2002], Salvador et al. [2005], Eichler [2007], Medkour et al. [2009] (to list only a few).

Theorem 2.4 in Dahlhaus [2000b] indicates that if the spectral density matrix if of full rank, the partial coherence can be obtained as the negative value of the rescaled inverse of the spectral density. Hence, a natural way to identify the partial correlation graph is to invert and rescale the estimate of the spectral density. However, in the high-dimensional case where the dimension p can be even much larger than the sample size T, since the estimated spectral density may not be invertable, classical methods under the low dimensional setting are no longer applicable. It is a challenge to find a new approach to estimate the inverse of the spectral density matrix. In this section, we consider the more general case in which the process can be locally stationary and hence the inverse spectral density matrix varies with time.

For $0 \leq u \leq 1$ and θ, the inverse of the spectral density matrix is $\Omega^0(u,\theta) = F(u,\theta)^{-1}$. We estimate the spectral density matrix by the lag window estimate [cf. (3.4.1)]. For simplicity, we consider the rectangular kernel, i.e., $K(x) = 1$ for $|x| \leq 1$ and $K(x) = 0$

otherwise. Then we use the constrained ℓ_1 minimization approach to estimate $\Omega^0(u,\theta)$. Let

$$\hat{\Omega}(u,\theta) = \arg\min |\Omega(u,\theta)|_1 \text{ subject to } |\hat{F}(u,\theta)\Omega(u,\theta) - \mathrm{I}_p|_\infty \le \lambda, \tag{3.5.1}$$

where $\lambda > 0$ is a tuning parameter. The constrained ℓ_1 minimization approach has been adopted in many applications; see Candes and Tao [2007], Bickel et al. [2009], Cai et al. [2011] among many others. The optimization program (3.5.1) can be decomposed into p parallel vector minimization problems. Let e_i be a standard unit vector in $\mathbb{R}^p$ with 1 in the i-th coordinate and 0 in all other coordinates. For $1 \le i \le p$, let $\hat{w}_i(u,\theta)$ be the solution of the following convex optimization problem:

$$\min |w|_1 \text{ subject to } |\hat{F}(u,\theta)w - e_i|_\infty \le \lambda, \tag{3.5.2}$$

where w is a vector in $\mathbb{R}^p$. By similar argument as Lemma 1 of Cai et al. [2011], we can show that solving the optimization problem (3.5.1) is equivalent to solving the p optimization problems (3.5.2), i.e.,

$$\hat{\Omega}(u,\theta) = (\hat{w}_1(u,\theta), \ldots, \hat{w}_p(u,\theta)). \tag{3.5.3}$$

We estimate $\Omega^0(u,\theta)$ by

$$\tilde{\Omega}(u,\theta) = \frac{\hat{\Omega}(u,\theta) + \hat{\Omega}^\dagger(u,\theta)}{2}, \tag{3.5.4}$$

where † is the conjugate transpose of a matrix.

Theorem 3.5.1. *Let the assumptions of Theorem 3.4.1 be satisfied. Define*

$$\kappa_0 = \sup_{0\le u\le 1} \max_\theta |\Omega^0(u,\theta)|_{L_1}$$

and

$$\varrho_n := \sup_{u\in[b_n,1-b_n]} \max_\theta |\tilde{\Omega}(u,\theta) - \Omega^0(u,\theta)|_\infty. \tag{3.5.5}$$

Let

$$\lambda = C\kappa_0\Big(\frac{(nmR_{M,m})^{2/q}}{M}\Theta^2_{q,\alpha} + \sqrt{\frac{m\log(pn)}{M}}\Phi^2_{q,\alpha} + \mathcal{V}_{m,M,n}\Big),$$

where $R_{M,m}$ is defined in Theorem 3.4.1, $\mathcal{V}_{m,M,n}$ is defined in Corollary 3.4.2 and C is a sufficiently large constant. Then for any $x \geq \lambda$, we have

$$\mathbb{P}(\varrho_n \geq 4x\kappa_0) \lesssim \frac{nmR_{M,m}\kappa_0^{q/2}\Theta^q_{q,\alpha}}{(Mx)^{q/2}} + np^2\exp\left(-\frac{C_{q,\alpha}Mx^2}{m\kappa_0^2\Phi^4_{4,\alpha}}\right). \qquad (3.5.6)$$

Proof of Theorem 3.5.1. For $x \geq \lambda$, let

$$E_x = \left\{\sup_{u\in[b_n,1-b_n]}\max_\theta|\Omega^0(u,\theta)|_{L_1}|\hat{F}(u,\theta) - F(u,\theta)|_\infty \leq x\right\}.$$

Notice that $|\hat{\Omega}(u,\theta)|_{L_1} \leq |\Omega^0(u,\theta)|_{L_1}$ uniformly over $u \in [b_n, 1-b_n]$ and θ. On the event E_x, by the triangle inequality, we have

$$\begin{aligned}\Lambda &:= \sup_{u\in[b_n,1-b_n]}\max_\theta|\hat{F}(u,\theta)(\hat{\Omega}(u,\theta) - \Omega^0(u,\theta))|_\infty \\ &\leq \sup_{u\in[b_n,1-b_n]}\max_\theta(|\hat{F}(u,\theta)\hat{\Omega}(u,\theta) - \mathrm{I}_p|_\infty + |\hat{F}(u,\theta)\Omega^0(u,\theta) - \mathrm{I}_p|_\infty) \\ &\leq \lambda + x, \qquad (3.5.7)\end{aligned}$$

which further implies

$$\begin{aligned}&\sup_{u\in[b_n,1-b_n]}\max_\theta|F(u,\theta)(\hat{\Omega}(u,\theta) - \Omega^0(u,\theta))|_\infty \\ \leq\ &\sup_{u\in[b_n,1-b_n]}\max_\theta|(\hat{F}(u,\theta) - F(u,\theta))(\hat{\Omega}(u,\theta) - \Omega^0(u,\theta))|_\infty + \Lambda \\ \leq\ &2\sup_{u\in[b_n,1-b_n]}\max_\theta|\Omega^0(u,\theta)|_{L_1}|\hat{F}(u,\theta) - F(u,\theta)|_\infty + \Lambda \leq 4x.\end{aligned}$$

Since

$$\varrho_T \leq \sup_{u\in[b_n,1-b_n]}\max_\theta|\hat{\Omega}(u,\theta) - \Omega^0(u,\theta)|_\infty$$

$$\leq \sup_{u\in[b_n,1-b_n]} \max_{\theta} |\Omega^0(u,\theta)|_{L_1} |F(u,\theta)(\hat{\Omega}(u,\theta)-\Omega^0(u,\theta))|_\infty \leq 4x\kappa_0,$$

and $\mathbb{P}(E_x) \to 1$ for $x \geq \lambda$, we then have

$$\mathbb{P}(\varrho_n \geq 4x\kappa_0) \leq \mathbb{P}(E_x^c) \leq \mathbb{P}\Big(\sup_{u\in[b_n,1-b_n]} \max_{\theta} |\hat{F}(u,\theta) - F(u,\theta)|_\infty \geq x/\kappa_0 \Big).$$

By Theorem 3.4.1 and Corollary 3.4.2, (3.5.6) follows. □

3.6 Hanson–Wright-Type Inequalities

In this section, we shall provide Hanson–Wright-type tail probability inequalities for locally stationary processes. The celebrated Hanson–Wright inequality provided a concentration result for quadratic forms of sub-Gaussian i.i.d. random variables; see Hanson and Wright [1971], Wright [1973] and Rudelson and Vershynin [2013]. There has been a large literature concerning large/moderate deviations for quadratic forms of Gaussian processes; see, for example, Bercu et al. [1997], Bryc and Dembo [1997], Zani [2002], Kakizawa [2007] among others. Xiao and Wu [2012] provided sharp tail probability upper bounds for quadratic forms of stationary processes with finite polynomial moments. We aim to relax the i.i.d., Gaussian/sub-Gaussian or stationary assumptions which were imposed in previous works, and establish tail probability inequalities for quadratic forms of locally stationary processes.

Let $\varepsilon_t, \varepsilon'_s,\ t, s \in \mathbb{Z}$ be i.i.d. random elements. In particular, we consider the one-dimensional locally stationary process $(X_t)_{t=1}^n$ of the form

$$X_t = g(t/n, \mathcal{F}^t), \tag{3.6.1}$$

where $\mathcal{F}^t = (\ldots, \varepsilon_{t-1}, \varepsilon_t)$, g is a measurable function such that $X_t(u) = g(u, \mathcal{F}^t)$ is well-defined. Let $\mathcal{F}^{t,\{0\}} = (\ldots, \varepsilon_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_{t-1}, \varepsilon_t)$. Assume that $\sup_{u\in[0,1]} \|g(u, \mathcal{F}_0)\|_q < \infty$

for some $q \geq 1$. For $t \geq 0$, we define the uniform functional dependence measure

$$\delta_{t,q} = \sup_{u \in [0,1]} \|g(u, \mathcal{F}^t) - g(u, \mathcal{F}^{t,\{0\}})\|_q \tag{3.6.2}$$

and the dependence adjusted norm

$$\|X_\cdot\|_{q,\alpha} = \sup_{m \geq 0} (m+1)^\alpha \Delta_{m,q}, \alpha \geq 0, \text{ where } \Delta_{m,q} = \sum_{t=m}^{\infty} \delta_{t,q}.$$

For $m \geq 0$, let $X_{t,m}$ be the m-dependence approximation of X_t given by

$$X_{t,m} = \mathbb{E}(X_t | \varepsilon_{t-m}, \ldots, \varepsilon_{t-1}, \varepsilon_t).$$

The quadratic form of the process (X_t) is written as

$$Q_T = \sum_{1 \leq s \leq t \leq T} a_{s,t} X_s X_t,$$

where the coefficients under our setting satisfy $a_{s,t} = a_{t-s}$ $(t \geq s)$, which only depends on the distance $t - s$. Moreover, we assume that $\sup_{s,t} |a_{s,t}| \leq 1$ and $a_{s,t} = 0$ if $t - s > B$, where $B \to \infty$ and $B = o(T)$. Theorem 3.6.1 provides a tail probability inequality for the quadratic form Q_T. For a random variable X, recall the operator $\mathbb{E}_0$ with $\mathbb{E}_0 X = X - \mathbb{E}X$ and the projection operator $\mathcal{P}^t \cdot = \mathbb{E}(\cdot|\mathcal{F}^t) - \mathbb{E}(\cdot|\mathcal{F}^{t-1})$.

Theorem 3.6.1. *Assume $\mathbb{E}(X_t) = 0$ and $X_t \in \mathcal{L}^q$ for some $q > 4$. Further assume $\|X_\cdot\|_{q,\alpha} < \infty$ for some $\alpha > 0$. Let $B \to \infty$ and $B = O(T^\eta)$ for some $0 < \eta < 1$. Then for $x \geq \sqrt{TB}\|X_\cdot\|_{q,\alpha}^2$, there exist positive constants $C_{q,\alpha,\eta}$ and $C_{q,\alpha}$ such that*

$$\mathbb{P}(|\mathbb{E}_0 Q_T| \geq x) \leq C_{q,\alpha,\eta} x^{-q/2} \|X_\cdot\|_{q,\alpha}^q F_{T,B} + C_{q,\alpha} \exp\left(-\frac{C_{q,\alpha} x^2}{\|X_\cdot\|_{4,\alpha}^4 TB} \right),$$

where $F_{T,B} = T$ (resp. $TB^{q/2-\alpha q/2-1}$ or $T^{q/4-\alpha q/2}B^{q/4}$) if $\alpha > 1-2/q$ (resp. $1/2-2/q <$

$\alpha < 1 - 2/q$ or $\alpha < 1/2 - 2/q$).

Proof of Theorem 3.6.1. Let $n = \lceil T/B \rceil$. For $k = 1, 2, \ldots, n$, define

$$V_k = \sum_{t=(k-1)B+1}^{(kB)\wedge T} \sum_{1 \le s \le t} a_{s,t} X_s X_t.$$

For $j \in \mathbb{Z}$, we define the innovation sets

$$\eta_j = \left(\varepsilon_{(j-1)B+1}, \varepsilon_{(j-1)B+2}, \ldots, \varepsilon_{jB}\right).$$

Let $L = \lfloor \log n / \log 2 \rfloor$, $\tau_l = 2^l$ for $1 \le l \le L-1$ and $\tau_L = n$. Define

$$\begin{aligned} V_{k,\tau} &= \mathbb{E}(V_k | \eta_{k-\tau}, \eta_{k-\tau+1}, \ldots, \eta_k), \text{ for } \tau \ge 0, \\ M_{n,l} &= \sum_{k=1}^{n} (V_{k,\tau_l} - V_{k,\tau_{l-1}}), \text{ for } 2 \le l \le L. \end{aligned}$$

We can write Q_T as

$$Q_T = \sum_{k=1}^{n} (V_k - V_{k,n}) + \sum_{l=2}^{L} M_{n,l} + \sum_{k=1}^{n} V_{k,2}. \tag{3.6.3}$$

Notice that $V_k - V_{k,n} = \sum_{j=n+1}^{\infty} (V_{k,j} - V_{k,j-1})$. By Lemma 3.7.4 and Theorem 3.2 of Burkholder [1973],

$$\begin{aligned} \left\| \sum_{k=1}^{n} (V_k - V_{k,n}) \right\|_{q/2} &\le \sum_{j=n+1}^{\infty} \left\| \sum_{k=1}^{n} (V_{k,j} - V_{k,j-1}) \right\|_{q/2} \\ &\le \sum_{j=n+1}^{\infty} C_q \sqrt{n} B \Delta_{0,q} \sum_{h=(j-2)B+1}^{(j+1)B} \delta_{h,q} \\ &\le C_q \sqrt{n} B \Delta_{0,q} \Delta_{(n-1)B+1,q}. \end{aligned} \tag{3.6.4}$$

By Markov's inequality, we have

$$\begin{aligned}\mathbb{P}\left(\left|\sum\nolimits_{k=1}^{n}(V_k - V_{k,n})\right| \geq x\right) &\leq \frac{C_q n^{q/4} B^{q/2} \Delta_{0,q}^{q/2} \Delta_{(n-1)B+1,q}^{q/2}}{x^{q/2}} \\ &\leq C_{q,\alpha} x^{-q/2} T^{q/4-\alpha q/2} B^{q/4} \|X.\|_{q,\alpha}^{q}.\end{aligned}$$

For each $2 \leq l \leq L$, define

$$\begin{aligned}Y_{i,l} &= \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \left(V_{k,\tau_l} - V_{k,\tau_{l-1}}\right), \quad \text{for } 1 \leq i \leq \lceil n/\tau_l \rceil; \\ R_{n,l}^{e} &= \sum_{i \text{ is even}} Y_{i,l} \text{ and } R_{n,l}^{o} = \sum_{i \text{ is odd}} Y_{i,l}.\end{aligned}$$

Let $\lambda_2, \ldots, \lambda_L$ be a positive sequence such that $\sum_{l=2}^{L} \lambda_l \leq 1$. In particular, we take $\lambda_l = (l-1)^{-2}/(\pi^2/3)$ if $2 \leq l \leq L/2$ and $\lambda_l = (L+1-l)^{-2}/(\pi^2/3)$ if $L/2 < l \leq L$. Since $Y_{i,l}$ and $Y_{i',l}$ are independent for $|i - i'| > 1$, by Corollary 1.8 of Nagaev [1979], for any $x > 0$,

$$\mathbb{P}(\left|R_{n,l}^{e}\right| \geq \lambda_l x) \leq \frac{C_q \sum_{i \text{ is even}} \mathbb{E}\left|Y_{i,l}\right|^{q/2}}{(\lambda_l x)^{q/2}} + 2\exp\left(-\frac{C_q (\lambda_l x)^2}{\sum_{i \text{ is even}} \mathbb{E}\left|Y_{i,l}\right|^2}\right).$$

Similarly as (3.6.4), we have

$$\|Y_{i,l}\|_{q/2} \leq C_q \tau_l^{1/2} B \Delta_{0,q} \phi_{l,q},$$

where $\phi_{l,q} = \sum_{h=(\tau_{l-1}-1)B+1}^{(\tau_l+1)B} \delta_{h,q} \leq \Delta_{(\tau_{l-1}-1)B+1,q}$. A similar inequality holds for $R_{n,l}^{o}$. Preceding arguments indicate

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{l=2}^{L} M_{n,l}\right| \geq 2x\right) &\leq \sum_{l=2}^{L} \mathbb{P}\left(\left|M_{n,l}\right| \geq 2\lambda_l x\right) \\ &\leq \sum_{l=2}^{L} \mathbb{P}\left(\left|R_{n,l}^{e}\right| \geq \lambda_l x\right) + \mathbb{P}\left(\left|R_{n,l}^{o}\right| \geq \lambda_l x\right)\end{aligned}$$

$$\leq \frac{C_{q,\alpha}\|X_\cdot\|_{q,\alpha}^q TB^{q/2-\alpha q/2-1}}{x^{q/2}} \cdot \mathrm{I} + \mathrm{II}. \tag{3.6.5}$$

where

$$\mathrm{I} = \sum_{l=2}^{L} \frac{\tau_l^{\varpi}}{\lambda_l^{q/2}}, \varpi = q/4 - 1 - \alpha q/2, \quad \mathrm{II} = 4\sum_{l=2}^{L} \exp\left(-\frac{C_{q,\alpha}x^2\lambda_l^2\tau_l^{2\alpha}}{\|X_\cdot\|_{4,\alpha}^4 TB}\right).$$

By the definitions of τ_l's and λ_l's, elementary calculations show $\mathrm{I} = O(1)$ if $\varpi < 0$, and $\mathrm{I} = O(n^{\varpi})$ if $\varpi > 0$. Also, since $\min_{l\geq 1}\lambda_l^2\tau_l^{2\alpha} > 0$, for $x \geq \sqrt{TB}\|X_\cdot\|_{q,\alpha}^2$,

$$\mathrm{II} \leq C_{q,\alpha}\exp\left(-\frac{C_{q,\alpha}x^2}{\|X_\cdot\|_{4,\alpha}^4 TB}\right).$$

Therefore, it follows that

$$\mathbb{P}\left(\left|\sum_{l=2}^{L} M_{n,l}\right| \geq x\right) \leq \frac{C_{q,\alpha}\|X_\cdot\|_{q,\alpha}^q F_{T,B}}{x^{q/2}} + C_{q,\alpha}\exp\left(-\frac{C_{q,\alpha}x^2}{\|X_\cdot\|_{4,\alpha}^4 TB}\right).$$

Now it remains to deal with the term $\sum_{k=1}^n \mathbb{E}_0 V_{k,2}$. By the definition of $V_{k,2}$ and the independence of η_k, $k \in \mathbb{Z}$, $V_{k,2}$ and $V_{k',2}$ are independent for $|k-k'| > 2$. By Corollary 1.7 of Nagaev [1979] and Lemma 3.7.2 (iv), we have for any $M > 1$ and $x \geq \sqrt{TB}\|X_\cdot\|_{q,\alpha}^2$,

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{k=1}^{n}\mathbb{E}_0 V_{k,2}\right| \geq x\right) &\leq \sum_{k=1}^{n}\mathbb{P}\left(\left|\mathbb{E}_0 V_{k,2}\right| \geq \frac{x}{3M}\right) + 6\exp\left(\frac{-C_q x^2}{\|X_\cdot\|_{4,\alpha}^4 TB}\right) \\ &\quad + \left(\frac{C_q\|X_\cdot\|_{q,\alpha}^q TB^{q/2-1}}{x^{q/2}}\right)^M.\end{aligned}$$

Recall that $B = O(T^\eta)$ with $0 < \eta < 1$. Let $M = 1 + (2q-2)\eta(q-4)^{-1}(1-\eta)^{-1}$. Since $x \geq \sqrt{TB}\|X_\cdot\|_{q,\alpha}^2$, elementary calculations show that for sufficiently large T, the last term in the above expression is no greater than $C_q T\|X_\cdot\|_{q,\alpha}^q/x^{q/2}$. Let $\eta_k^* = (\eta_{k-2}, \eta_{k-1}, \eta_k)$. Notice that $\mathbb{E}_0 V_{k,2} = \mathbb{E}(\mathbb{E}_0 V_k | \eta_k^*)$. For $2 < q^* < q/2$, we have

$$\begin{aligned}
\mathbb{P}\left(\left|\mathbb{E}_0 V_{k,2}\right| \geq x\right) &\leq \mathbb{P}\left(\left|\mathbb{E}(\mathbb{E}_0 V_k\ \mathbf{1}\{|\mathbb{E}_0 V_k| \geq x/2\}|\eta_k^*)\right| \geq x/3\right) \\
&\leq C_q x^{-q^*} \mathbb{E}\left|\mathbb{E}(\mathbb{E}_0 V_k \mathbf{1}\{|V_k - \mathbb{E}(V_k)| \geq x/2\}|\eta_k^*)\right|^{q^*} \\
&\leq \frac{C_q}{x^{q^*}} \mathbb{E}\left(|\mathbb{E}_0 V_k|^{q^*}\ \mathbf{1}\{|\mathbb{E}_0 V_k| \geq x/2\}\right) \\
&\leq \frac{C_q}{x^{q^*}} \int_{x/2}^{\infty} z^{q^*-1} \mathbb{P}\left(|\mathbb{E}_0 V_k| \geq z\right) dz \leq \frac{C_{q,\alpha} \|X.\|_{q,\alpha}^q \zeta_B}{x^{q/2}}. \qquad (3.6.6)
\end{aligned}$$

where $\zeta_B = B$ (resp. $B^{q/2-\alpha q/2}$) for $\alpha > 1 - 2/q$ (resp. $\alpha < 1 - 2/q$), and the last step in (3.6.6) is an immediate consequence by plugging in the bound of $\mathbb{P}(|\mathbb{E}_0 V_k| \geq z)$ applying Theorem 3.6.2.

Putting all of the pieces together, the proof is complete. □

In Theorem 3.6.2, we consider the tail probability inequality of a general quadratic form

$$Q_B := \sum\nolimits_{1 \leq s \leq t \leq B} a_{s,t} X_s X_t$$

with the only restriction $\sup_{s,t} |a_{s,t}| \leq 1$ on the coefficients. Theorem 3.6.3 concerns the quadratic form

$$L_T(l) := \sum_{l+1 \leq t \leq T} a_t X_{t-l} X_t$$

with $\sup_t |a_t| \leq 1$, for $0 \leq l < B$, and $B = o(T)$. The proofs of Theorem 3.6.2 and Theorem 3.6.3 will be given in the Appendix.

Theorem 3.6.2. *Assume $\mathbb{E}(X_t) = 0$ and $X_t \in \mathcal{L}^q$ for some $q > 4$. Further assume $\|X.\|_{q,\alpha} < \infty$ for some $\alpha > 0$. Let $r = q/2$. If $x > 0$ satisfies $B^{1+\kappa}\|X.\|_{2r,\alpha}^2 = o(x)$ for some $\kappa > 0$, then there exists some positive constant $C_{\kappa,r,\alpha}$ depending on κ, r and α such that*

$$\mathbb{P}(|\mathbb{E}_0 Q_B| \geq x) \leq C_{\kappa,r,\alpha} x^{-r} \|X.\|_{2r,\alpha}^{2r} \zeta_B, \qquad (3.6.7)$$

where $\zeta_B = B$ for $\alpha > 1 - 1/r$ and $\zeta_B = B^{r-\alpha r}$ if $\alpha < 1 - 1/r$.

Theorem 3.6.3. *Assume* $\mathbb{E}(X_t) = 0$ *and* $X_t \in \mathcal{L}^q$ *for some* $q > 4$. *Further assume* $\|X_\cdot\|_{q,\alpha} < \infty$ *for some* $\alpha > 0$. *For* $0 < \beta < 1$, *let* $B = \lfloor T^\beta \rfloor$. *(i) If* $l > 3B$, *for* $x \geq T^{1/2}\|X_\cdot\|_{q,\alpha}^2$, *we have*

$$\mathbb{P}(|\mathbb{E}_0 L_T(l)| \geq x) \lesssim \frac{\|X_\cdot\|_{q,\alpha}^q (T + T^{q/4-\alpha\beta q/2})}{x^{q/2}},$$

where the constant in $\lesssim$ *only depends on* q, α *and* β. *(ii) If* $0 \leq l \leq 3B$, *the inequality is*

$$\mathbb{P}(|\mathbb{E}_0 L_T(l)| \geq x) \lesssim \frac{\|X_\cdot\|_{q,\alpha}^q (D_T + TB^{q/4-\alpha\beta q/2-1})}{x^{q/2}} + \exp\left(-\frac{C_{q,\alpha}x^2}{\|X_\cdot\|_{4,\alpha}^4 T}\right),$$

where $D_T = T$ *(resp.* $T(\log T)^{1+q}$ *or* $T^{q/4-\alpha q/2}$*) for* $\alpha > 1/2 - 2/q$ *(resp.* $\alpha = 1/2 - 2/q$ *or* $\alpha < 1/2 - 2/q$*).*

3.7 Deferred Proofs

In this section, we shall provide the proofs of Corollary 3.3.2, Corollary 3.4.2 and Corollary 3.4.5, some lemmas that are useful in proofs of this chapter, and the proofs of Theorem 3.6.2 and Theorem 3.6.3.

3.7.1 Proofs of Corollaries 3.3.2, 3.4.2 and 3.4.5

Proof of Corollary 3.3.2. By Theorem 3.3.1, it follows that

$$\psi_n = O_{\mathbb{P}}\left(\frac{(nH_M)^{2/q}}{M}\Theta_{q,\alpha}^2 + \sqrt{\frac{\log(pn)}{M}}\Phi_{q,\alpha}^2\right). \tag{3.7.1}$$

It remains to deal with the bias $\max_{0 \leq l < M} \sup_{u \in [b_n, 1-b_n]} |\mathbb{E}\hat{\Gamma}_l(u) - \Gamma_l(u)|_\infty$. For any $u \in [b_n, 1-b_n]$ and $0 \leq l < M$, by the triangle inequality,

$$|\mathbb{E}\hat{\Gamma}_l(u) - \Gamma_l(u)|_\infty \quad \leq \quad \left|\mathbb{E}\hat{\Gamma}_l(u) - \frac{M-l}{M}\Gamma_l(u)\right|_\infty + \frac{l}{M}|\Gamma_l(u)|_\infty$$

$$=: \quad \mathrm{I} + \mathrm{II}. \tag{3.7.2}$$

Notice that $\max_{1\le j\le p}\sup_{u\in[0,1]}\|X_{rj}(u)\|_q \le \Phi_{q,0}$. Under the condition (1.0.2), by the triangle inequality, we have

$$\begin{aligned}
\mathrm{I} &= \frac{1}{M}\Big|\sum_{r=l+T_1(u)}^{T_2(u)} \mathbb{E}(X_{r-l}X_r^\top - X_{r-l}(u)X_r(u)^\top)\Big|_\infty \\
&\le \frac{1}{M}\Big|\sum_{r=l+T_1(u)}^{T_2(u)} \mathbb{E}(X_{r-l}X_r^\top - X_{r-l}(u)X_r^\top)\Big|_\infty \\
&\quad + \frac{1}{M}\Big|\sum_{r=l+T_1(u)}^{T_2(u)} \mathbb{E}(X_{r-l}(u)X_r^\top - X_{r-l}(u)X_r(u)^\top)\Big|_\infty \\
&=: \mathrm{I}_1 + \mathrm{I}_2.
\end{aligned} \tag{3.7.3}$$

Under the condition (1.0.2), we have

$$\mathrm{I}_1 \le \frac{M-l}{M}\mathcal{K}b_n \max_{1\le j\le p}\|X_{rj}\|_2.$$

Similarly, it also applies to I_2. Hence,

$$\mathrm{I} = O\Big(\frac{\mathcal{K}M}{n}\Phi_{2,0}\Big). \tag{3.7.4}$$

Notice that $X_{rj} - X_{rj}(u) = \sum_{t=0}^{\infty}\mathcal{P}_{r-t}(X_{rj} - X_{rj}(u))$. Under the condition (1.0.2), for $r \in [T_1(u), T_2(u)]$,

$$\|\mathcal{P}_{r-t}(X_{rj} - X_{rj}(u))\|_2 \le \min\{2\delta_{t,2,j}, \mathcal{K}M/n\}.$$

We then have

$$\mathrm{I}_1 \le \frac{1}{M}\Phi_{2,0}\sum_{t=0}^{\infty}\sqrt{M-l}\cdot\min\{2\delta_{t,2,j}, \mathcal{K}M/n\}$$

$$\leq \quad C_\alpha M^{-1/2}\Big(\frac{\mathcal{K}M}{n}\Big)^{\frac{\alpha}{1+\alpha}}\Phi_{2,0}\Phi_{2,\alpha}^{1/(1+\alpha)}.$$

Since I_2 has the same bound, we also have

$$\mathrm{I} = O\Big(M^{-1/2}\Big(\frac{\mathcal{K}M}{n}\Big)^{\frac{\alpha}{1+\alpha}}\Phi_{2,0}\Phi_{2,\alpha}^{1/(1+\alpha)}\Big). \tag{3.7.5}$$

By $X_{rj}(u) = \sum_{t=0}^{\infty}\mathcal{P}_{r-t}X_{rj}(u)$, $\Delta_{0,2,j} \leq \Phi_{2,0}$ and $\Delta_{m,2,j} \leq \Phi_{2,\alpha}m^{-\alpha}$,

$$\begin{aligned}
\mathrm{II} &= \frac{l}{M}\max_{1\leq i,j\leq p}\Big|\sum_{t=0}^{\infty}\mathbb{E}[(\mathcal{P}_{r-t}X_{(r-l)i}(u))(\mathcal{P}_{r-t}X_{rj}(u))]\Big| \\
&\leq \frac{l}{M}\max_{1\leq i,j\leq p}\sum_{t=0}^{\infty}\delta_{t,2,i}\delta_{t+l,2,j} \leq \frac{l}{M}\Delta_{0,2,i}\Delta_{l,2,j} \\
&\leq \frac{l^{-\alpha+1}}{M}\Phi_{2,0}\Phi_{2,\alpha} \leq M^{-\min\{\alpha,1\}}\Phi_{2,0}\Phi_{2,\alpha}.
\end{aligned} \tag{3.7.6}$$

By (3.7.2), (3.7.4), (3.7.5) and (3.7.6), we have

$$|\mathbb{E}\hat{\Gamma}_l(u) - \Gamma_l(u)|_\infty \leq \mathcal{U}_{M,n} + M^{-\min\{\alpha,1\}}\Phi_{2,0}\Phi_{2,\alpha} \tag{3.7.7}$$

uniformly over $u \in [b_n, 1-b_n]$ and $0 \leq l < M$, which implies (3.3.5) immediately. $\square$

Proof of Corollary 3.4.2. By Theorem 3.4.1, it suffices to show

$$|\mathbb{E}\hat{F}(u,\theta) - F(u,\theta)| = O(\mathcal{V}_{m,M,n} + \mathcal{W}_m).$$

Notice that

$$\begin{aligned}
&2\pi|\mathbb{E}\hat{F}(u,\theta) - F(u,\theta)| \\
\leq\ & \sum_{|l|>m}|\Gamma_l(u)|_\infty + \Big|\sum_{l=-m}^{m}[K(l/m)\mathbb{E}\hat{\Gamma}_l(u) - \Gamma_l(u)]\exp(-\iota l\theta)\Big|_\infty \\
=:\ & \mathrm{I} + \mathrm{II}.
\end{aligned} \tag{3.7.8}$$

Denote $\Gamma_l(u) = (\gamma_{l,ij}(u))$. By similar arguments as (3.7.6), we have

$$|\gamma_{l,ij}(u)| \leq \sum_{t=0}^{\infty} \delta_{t,2,i}\delta_{t+l,2,j}. \tag{3.7.9}$$

Hence it can be obtained that

$$\Big| \sum_{|l|>m} \gamma_{l,ij}(u) \Big| \leq \sum_{|l|>m} |\gamma_{l,ij}(u)| \leq 2\Delta_{0,2,i}\Delta_{m+1,2,j}. \tag{3.7.10}$$

Since $\Delta_{0,2,j} \leq \Phi_{2,0}$ and $\Delta_{m,2,j} \leq \Phi_{2,\alpha} m^{-\alpha}$, we have

$$\mathrm{I} \leq 2\Delta_{0,2,i}\Delta_{m+1,2,j} \leq 2m^{-\alpha}\Phi_{2,0}\Phi_{2,\alpha}. \tag{3.7.11}$$

Recall the definition of $a_l(\theta)$ in (3.4.11). By the triangle inequality,

$$\begin{aligned}
\mathrm{II} \ &\leq \ \frac{2}{M}\Big| \sum_{l=0}^{m} \sum_{r=l+T_1(u)}^{T_2(u)} a_l(\theta)\mathbb{E}(X_{r-l}X_r^\top - X_{r-l}(u)X_r(u)^\top) \Big|_\infty \\
&\quad + \frac{2}{M}\Big| \sum_{l=1}^{m} |l| a_l(\theta)\Gamma_l(u) \Big|_\infty + 2\sup_u \sum_{l=1}^{m} (1 - K(l/m))|\Gamma_l(u)|_\infty \\
&=: \ \mathrm{II}_1 + \mathrm{II}_2 + 2\mathcal{W}_m.
\end{aligned} \tag{3.7.12}$$

Since $X_{rj}(u) = \sum_{t=0}^{\infty} \mathcal{P}_{r-t} X_{rj}(u)$ and $|a_l(\theta)| \leq 1$, by Lemma 3.7.2 (iii),

$$\| \sum_{t=1}^{m} a_l(\theta) X_{tj}(u) \|_2 \leq \sqrt{m}\Delta_{0,2,j} \leq \sqrt{m}\Phi_{2,0}.$$

Similarly, we have $\| \sum_{t=1}^{m} a_l(\theta) X_{tj} \|_2 \leq \sqrt{m}\Phi_{2,0}$. Under the condition (1.0.2), by the triangle inequality and Holder's inequality, we have

$$\mathrm{II}_1 \ \leq \ \frac{2}{M}\Big| \sum_{r=T_1(u)}^{T_2(u)} \sum_{s=r}^{r+m} a_{s-r}(\theta)\mathbb{E}[(X_r - X_r(u))X_s^\top] \Big|_\infty$$

$$+\frac{2}{M}\Big|\sum_{s=T_1(u)}^{T_2(u)}\sum_{r=(s-m)\vee 1}^{s} a_{s-r}(\theta)\mathbb{E}[(X_s - X_s(u))X_r(u)^\top]\Big|_\infty$$

By the similar argument as (3.7.4) and (3.7.5), we obtain

$$\mathrm{II}_1 = O\left(\sqrt{m}\,\mathcal{U}_{M,n}\right). \tag{3.7.13}$$

By (3.7.9), we have

$$\mathrm{II}_2 \le \frac{2}{M}\sum_{l=1}^{m}\sum_{s=l}^{m}\sum_{t=0}^{\infty}\delta_{t,2,i}\delta_{t+s,2,j} \le \frac{2}{M}\Delta_{0,2,i}\sum_{l=1}^{m}\Delta_{l,2,j}.$$

Since $\Delta_{0,2,j} \le \Phi_{2,0}$ and $\Delta_{m,2,j} \le \Phi_{2,\alpha}m^{-\alpha}$, it follows that

$$\mathrm{II}_2 \le \frac{2r(m)}{M}\Phi_{2,0}\Phi_{2,\alpha}. \tag{3.7.14}$$

The result follows in view of (3.7.8), (3.7.11), (3.7.12), (3.7.13) and (3.7.14). □

Proof of Corollary 3.4.5. The proof is similar to that of Corollary 3.4.2 with $\theta = 0$. By Theorem 3.4.4, we have

$$\sup_{u\in[b_n,1-b_n]}|\hat{\Sigma}(u) - \mathbb{E}\hat{\Sigma}(u)|_\infty = O_{\mathbb{P}}\left(\frac{(nR_{M,m})^{2/q}}{M}\Theta_{q,\alpha}^2 + \sqrt{\frac{m\log(pn)}{M}}\Phi_{q,\alpha}^2\right).$$

Now we consider the bias $\sup_u|\mathbb{E}\hat{\Sigma}(u) - \Sigma(u)|_\infty$. For $u \in [b_n, 1-b_n]$, let

$$B(u) = \frac{1}{(M-m+1)m}\sum_{r=T_1(u)}^{T_2(u)-m+1}\left(\sum_{t=r}^{r+m-1}X_t(u)\right)\left(\sum_{t=r}^{r+m-1}X_t(u)\right)^\top.$$

Similarly as (3.7.13), we have

$$|\mathbb{E}\hat{\Sigma}(u) - \mathbb{E}B(u)|_\infty = O\left(\sqrt{m}\,\mathcal{U}_{M,n}\right).$$

Notice that

$$\mathbb{E}B(u) = \frac{1}{m}\mathbb{E}\left(\sum_{t=1}^{m} X_t(u)\right)\left(\sum_{t=1}^{m} X_t(u)\right)^{\top} = \frac{1}{m}\sum_{l=-m}^{m}(m-|l|)\Gamma_l(u),$$

and recall that $\Sigma(u) = \sum_{l=-\infty}^{\infty}\Gamma_l(u)$. It follow that

$$\Sigma(u) - \mathbb{E}B(u) = \sum_{|l|>m}\Gamma_l(u) + m^{-1}\sum_{l=-m}^{m}|l|\Gamma_l(u).$$

By (3.7.11), we have

$$\sum_{|l|>m}|\Gamma_l(u)|_{\infty} = O(m^{-\alpha}\Phi_{2,0}\Phi_{2,\alpha}). \tag{3.7.15}$$

And (3.7.14) implies

$$m^{-1}\Big|\sum_{l=-m}^{m}|l|\Gamma_l(u)\Big|_{\infty} \le \frac{r(m)}{m}\Phi_{2,0}\Phi_{2,\alpha}. \tag{3.7.16}$$

Putting all of the pieces together, the proof is complete. □

3.7.2 Some Lemmas used in this chapter

Lemma 3.7.1. *Let* $S(\theta) = \frac{1}{2}a_0 + \sum_{k=1}^{n}[a_k cos(k\theta) + b_k sin(k\theta)]$ *be a trigonometric polynomial of order* n*. For any* $\theta^* \in \mathbb{R}$*,* $\delta > 0$ *and* $l \ge 2(1+\delta)n$*, let* $\theta_h = \theta^* + 2\pi h/l$ *for* $0 \le h \le l$*, then*

$$\max_{\theta}|S(\theta)| \le (1+\delta^{-1})\max_{0\le h\le l}|S(\theta_h)|.$$

Lemma 3.7.1 is adapted from Theorem 7.28 in Ch. X, Zygmund [2002]. It shows that for a trigonometric polynomial, we can bound its maximum by the maximum over a fine grid.

Lemma 3.7.2, Lemma 3.7.3 and Lemma 3.7.4 below concern the one-dimensional process (X_t) of the form (3.6.1). The uniform dependence measure $\delta_{\cdot,q}$ for the sequence (X_t) is defined by (3.6.2).

Lemma 3.7.2. *Assume* $\mathbb{E}(X_t) = 0$ *and* $X_t \in \mathcal{L}^q$ *for some* $q \ge 2$*. Let* C_q *be some positive* *constant that depends on q only. Then*

$$\begin{aligned}
&(i) \quad \|\mathcal{P}^s X_t\|_q \le \delta_{t-s,q},\\
&(ii) \quad |\mathbb{E}(X_s X_t)| \le \sum\nolimits_{r=0}^{\infty} \delta_{r,2}\delta_{r+t-s,2},\\
&(iii) \quad \Big\| \sum\nolimits_{t=1}^{T} c_t X_t \Big\|_q \le C_q \Delta_{0,q} \mathcal{D}_1,\\
&(iv) \quad \Big\| \sum\nolimits_{s,t=1}^{T} c_{s,t} \mathbb{E}_0(X_s X_t) \Big\|_{q/2} \le C_q \Delta_{0,q}^2 \mathcal{D}_2 \sqrt{T}, \ \textit{for } q \ge 4,
\end{aligned}$$

where $\mathcal{D}_1^2 = \sum_{t=1}^{T} c_t^2$ *and* $\mathcal{D}_2^2 = \max\left\{ \max_{1\le t\le T} \sum_{s=1}^{T} c_{s,t}^2, \max_{1\le s\le T} \sum_{t=1}^{T} c_{s,t}^2 \right\}$.

Lemma 3.7.3. *Assume* $\mathbb{E}(X_t) = 0$, $X_t \in \mathcal{L}^q$ *for some* $q \ge 4$ *and* $\Delta_{0,q} < \infty$. *Let* $Q_{T,m} = \sum_{1\le s\le t\le T} a_{s,t} X_{s,m} X_{t,m}$ *for* $m \ge 0$. *Then we have*

$$\|\mathbb{E}_0 Q_T - \mathbb{E}_0 Q_{T,m}\|_{q/2} \le C_q \sqrt{TB} \Delta_{0,q} \Delta_{\lfloor m/2 \rfloor,q}.$$

Lemma 3.7.2 extends some results of Proposition 1 in Xiao and Wu [2014] to the non-stationary case. A detailed proof for these conclusions can also be found in that paper. Equipped with the uniform dependence measure (3.6.2), we can obtain Lemma 3.7.2 for the non-stationary case by a similar proof without extra technical difficulties and the proof is hence omitted.

Lemma 3.7.3 provides a moment inequality for the difference between the quadratic form and its m-dependence approximation for the non-stationary processes. It follows by similar arguments when proving Lemma 9 in Xiao and Wu [2012]. Both of the lemmas can be very useful in the proof.

Lemma 3.7.4. *Assume* $\mathbb{E}(X_t) = 0$, $X_t \in \mathcal{L}^q$ *for some* $q > 4$ *and* $\Delta_{0,q} < \infty$. *Let* $\eta_j = \left(\varepsilon_{(j-1)B+1}, \varepsilon_{(j-1)B+2}, \dots, \varepsilon_{jB} \right)$. *For* $1 \le k \le \lceil T/B \rceil$ *and* $\tau \ge 0$, *define*

$$V_k = \sum_{t=(k-1)B+1}^{(kB)\wedge T} \sum_{1\le s\le t} a_{s,t} X_s X_t,$$

and $V_{k,\tau} = \mathbb{E}(V_k|\eta_{k-\tau}, \eta_{k-\tau+1} \ldots, \eta_k)$. *Then for* $j \geq 1$, *there exists some constant* $C_q > 0$ *only depending on* q *such that*

$$\left\|V_{k,j} - V_{k,j-1}\right\|_{q/2} \leq C_q B \Delta_{0,q} \sum_{h=(j-2)B+1}^{(j+1)B} \delta_{h,q}. \tag{3.7.17}$$

Proof of Lemma 3.7.4. Let $\mathcal{F}_i^j = (\varepsilon_i, \varepsilon_{i+1}, \ldots, \varepsilon_j)$. For a random variable X which is $\mathcal{F}_j$-measurable, define $\mathcal{P}_i X = \mathbb{E}(X|\mathcal{F}_i^j) - \mathbb{E}(X|\mathcal{F}_{i+1}^j)$. For $j \geq 1$, we can write

$$\begin{aligned} V_{k,j} - V_{k,j-1} &= \mathbb{E}(V_k|\eta_{k-j}, \eta_{k-j+1}, \ldots, \eta_k) - \mathbb{E}(V_k|\eta_{k-j+1}, \eta_{k-j+2}, \ldots, \eta_k) \\ &= \sum_{i=1}^{B} \mathbb{E}(V_k|\mathcal{F}_{(k-j-1)B+i}^{kB}) - \mathbb{E}(V_k|\mathcal{F}_{(k-j-1)B+i+1}^{kB}) \\ &= \sum_{i=1}^{B} \mathcal{P}_{(k-j-1)B+i} V_k. \end{aligned}$$

By Jensen's inequality and the triangle inequality, we have

$$\|\mathcal{P}_{(k-j-1)B+i} V_k\|_{q/2} \leq \mathrm{I} + \mathrm{II}, \tag{3.7.18}$$

where

$$\begin{aligned} \mathrm{I} &= \left\| \sum_{t=(k-1)B+1}^{(kB)\wedge T} (X_t - X_{t,\{(k-j-1)B+i\}}) \sum_{s=(t-B)\vee 1}^{t} a_{s,t} X_s \right\|_{q/2}, \\ \mathrm{II} &= \left\| \sum_{s=[(k-2)B+1]\vee 1}^{(kB)\wedge T} (X_s - X_{s,\{(k-j-1)B+i\}}) \sum_{t=s}^{(s+B)\wedge T} a_{s,t} X_{t,\{(k-j-1)B+i\}} \right\|_{q/2}. \end{aligned}$$

We apply Hölder's inequality, Lemma 3.7.2 (i) and (iii) to have

$$\begin{aligned}
\mathrm{I} &\le \sum_{t=(k-1)B+1}^{(kB)\wedge T} \|X_t - X_{t,\{(k-j-1)B+i\}}\|_q \left\| \sum_{s=(t-B)\vee 1}^{t} a_{s,t} X_s \right\|_q \\
&\le C_q \sqrt{B} \Delta_{0,q} \sum_{t=(k-1)B+1}^{kB} \delta_{t-(k-j-1)B-i,q}. \qquad (3.7.19)
\end{aligned}$$

Similarly, we have

$$\mathrm{II} \le C_q \sqrt{B} \Delta_{0,q} \sum_{s=(k-2)B+1}^{kB} \delta_{s-(k-j-1)B-i,q}. \qquad (3.7.20)$$

Combining (3.7.18), (3.7.19) and (3.7.20), it follows that

$$\|\mathcal{P}_{(k-j-1)B+i} V_k\|_{q/2} \le C_q \sqrt{B} \Delta_{0,q} \sum_{h=(j-1)B-i+1}^{(j+1)B-i} \delta_{h,q}. \qquad (3.7.21)$$

Note that $\mathcal{P}_{(k-j-1)B+i} V_k$ are backward martingale differences w.r.t. $(\mathcal{F}_{(k-j-1)B+i}^{kB})_{1 \le i \le B}$. By Theorem 3.2 in Burkholder [1973], we have

$$\begin{aligned}
\|V_{k,j} - V_{k,j-1}\|_{q/2}^2 &\le C_q \sum_{i=1}^{B} \|\mathcal{P}_{(k-j-1)B+i} V_k\|_{q/2}^2 \\
&\le C_q B \Delta_{0,q}^2 \sum_{i=1}^{B} \left(\sum_{h=(j-1)B-i+1}^{(j+1)B-i} \delta_{h,q} \right)^2 \\
&\le C_q B^2 \Delta_{0,q}^2 \left(\sum_{h=(j-2)B+1}^{(j+1)B} \delta_{h,q} \right)^2,
\end{aligned}$$

which implies (3.7.17) immediately. □

3.7.3 Proofs of Theorem 3.6.2 and Theorem 3.6.3

Proof of Theorem 3.6.2. First we make the convention that if a term X_t in the summation has the subscript $t \notin [1, T]$, then that term should be replaced by zero. And we set $a_{s,t} = 0$

if $s > t$. In the proof, we consider the process normalized by the dependence adjusted norm $\|X.\|_{2r,\alpha}$. Then it suffices to show for x satisfying $B^{1+\kappa} = o(x)$,

$$\mathbb{P}(|\mathbb{E}_0 Q_B| \geq x) \leq C_{\kappa,r,\alpha} x^{-r} \zeta_B. \tag{3.7.22}$$

Step 1: Decomposition of $\mathbb{P}(|\mathbb{E}_0 Q_B| \geq x)$.

Let $\tau = \lfloor \log B / \log 4 \rfloor$. Let $B = m_0 > m_1 > \ldots > m_\tau \geq 1$ be a decreasing sequence with $m_i = \lfloor B/4^i \rfloor$ for $1 \leq i \leq \tau$ and $\lambda_1, \ldots, \lambda_\tau$ be a positive sequence such that $\sum_{i=1}^{\tau} \lambda_i \leq 1$; specifically, $\lambda_i = i^{-2}/(\pi^2/3)$ if $1 \leq i \leq \tau/2$ and $\lambda_i = (\tau + 1 - i)^{-2}/(\pi^2/3)$ if $\tau/2 < l \leq \tau$. For each m_i, define the m_i-dependence approximation of Q_B as $Q_{B,m_i} = \sum_{1 \leq s \leq t \leq B} a_{s,t} X_{s,m_i} X_{t,m_i}$. Then

$$\begin{aligned} \mathbb{P}(|\mathbb{E}_0 Q_B| \geq 2x) \quad \leq \quad & \mathbb{P}(|\mathbb{E}_0 Q_{B,m_\tau}| \geq x) + \mathbb{P}(|\mathbb{E}_0 (Q_B - Q_{B,m_1})| \geq \lambda_1 x) \\ & + \sum_{i=2}^{\tau} \mathbb{P}(|\mathbb{E}_0 (Q_{B,m_{i-1}} - Q_{B,m_i})| \geq \lambda_\tau x). \end{aligned} \tag{3.7.23}$$

By Lemma 3.7.3, we have

$$\mathbb{P}(|\mathbb{E}_0 (Q_B - Q_{B,m_1})| \geq \lambda_1 x) \leq \frac{C_r (B m_1^{-\alpha})^r}{(\lambda_1 x)^r} =: \mathrm{I}_1. \tag{3.7.24}$$

For $2 \leq i \leq \tau$, write

$$\mathbb{P}(|\mathbb{E}_0 (Q_{B,m_{i-1}} - Q_{B,m_i})| \geq \lambda_i x) \leq \mathrm{I}_i + \mathrm{II}_i. \tag{3.7.25}$$

where

$$\begin{aligned} \mathrm{I}_i &= \mathbb{P}\Big(|\mathbb{E}_0 \sum_{0 \leq t-s \leq 3m_{i-1}} (a_{s,t} X_{s,m_{i-1}} X_{t,m_{i-1}} - a_{s,t} X_{s,m_i} X_{t,m_i})| \geq \lambda_i x/2\Big), \\ \mathrm{II}_i &= \mathbb{P}\Big(|\mathbb{E}_0 \sum_{t-s > 3m_{i-1}} (a_{s,t} X_{s,m_{i-1}} X_{t,m_{i-1}} - a_{s,t} X_{s,m_i} X_{t,m_i})| \geq \lambda_i x/2\Big). \end{aligned}$$

Step 2: Bounding I_i, $1 \le i \le \tau$.

Let $b^* = \lceil B/(4m_{i-1}) \rceil$. For $1 \le b \le b^*$, define $\mathcal{B}_b^* = [4(b-1)m_{i-1}+1, (4bm_{i-1}) \wedge B]$ and

$$W_b = \sum_{t \in \mathcal{B}_b^*} \sum_{t-3m_{i-1} \le s \le t} \left(a_{s,t} X_{s,m_{i-1}} X_{t,m_{i-1}} - a_{s,t} X_{s,m_i} X_{t,m_i} \right).$$

Observe that for $|b - b'| > 1$, W_b and $W_{b'}$ are independent. By Corollary 1.6 of Nagaev [1979], Lemma 3.7.2 (iv) and Lemma 3.7.3, we have for any $M > 1$ and for x satisfying $B^{1+\kappa} = o(x)$, there exist constants $C_{\kappa,r,M}$ and $C_{\kappa,M}$ such that

$$\begin{aligned} \mathrm{I}_i &\le C_{\kappa,r,M} x^{-M} + \sum_{b=1}^{b^*} \mathbb{P}(|\mathbb{E}_0(W_b)| \ge \lambda_i x/(2C_{\kappa,M})) \\ &\le C_{\kappa,r,M} x^{-M} + C_{\kappa,r,M} \frac{B m_{i-1}^{r-1} m_i^{-\alpha r}}{(\lambda_i x)^r}. \end{aligned}$$

We can choose a sufficiently large M such that the first term in the above expression is no greater than the second term. To sum up, it becomes

$$\sum_{i=1}^{\tau} \mathrm{I}_i \le C_r x^{-r} B \sum_{i=1}^{\tau} \frac{m_{i-1}^{r-1} m_i^{-\alpha r}}{\lambda_i^r}.$$

Let $\omega = \sum_{i=1}^{\tau} m_{i-1}^{r-1} m_i^{-\alpha r}/\lambda_i^r$. By the definitions of m_i's and λ_i's, elementary calculations show $\omega = O(1)$ if $\alpha > 1 - 1/r$, and $\omega = O(B^{r-1-\alpha r})$ if $\alpha < 1 - 1/r$. Therefore, it follows that

$$\sum_{i=1}^{\tau} \mathrm{I}_i \le C_r x^{-r} \zeta_B. \tag{3.7.26}$$

Step 3: Bounding II_i, $2 \le i \le \tau$.

For non-negative integers n, m, N with $n < N$, define three functions as follows

$$\begin{aligned} U(n,m,x) &= \sup_{\{a_{s,t}\}} \mathbb{P}\Big(|\mathbb{E}_0 \sum_{t-s>n} a_{s,t} X_{s,m} X_{t,m}| \geq x\Big), \\ V(n,m,x) &= \sup_{\{a_{s,t}\}} \mathbb{P}\Big(|\mathbb{E}_0 \sum_{t-s\leq n} a_{s,t} X_{s,m} X_{t,m}| \geq x\Big), \\ W(n,N,m,x) &= \sup_{\{a_{s,t}\}} \mathbb{P}\Big(|\mathbb{E}_0 \sum_{n<t-s\leq N} a_{s,t} X_{s,m} X_{t,m}| \geq x\Big). \end{aligned}$$

Hence, we have

$$\begin{aligned} \mathrm{II}_i &\leq U(3m_{i-1}, m_{i-1}, \lambda_2 x/6) + U(3m_i, m_i, \lambda_2 x/6) \\ &\quad + W(3m_i, 3m_{i-1}, m_i, \lambda_2 x/6). \end{aligned} \tag{3.7.27}$$

Step 3.1: Bounding $U(3m_i, m_i, x)$, $1 \leq i \leq \tau$.

Define $Y^*_{t,m_i} = \sum_{s=1}^{t-3m_i-1} a_{s,t} X_{s,m_i}$. Let $b^\diamond = \lceil B/m_i \rceil$. For $1 \leq b \leq b^\diamond$, define $\mathcal{B}^\diamond_b = [(b-1)m_i + 1, (bm_i) \wedge B]$ and define

$$R_{b,m_i} = \sum_{t \in \mathcal{B}^\diamond_b} X_{t,m_i} Y^*_{t,m_i}.$$

Then we have $\sum_{t-s>3m_i} a_{s,t} X_{s,m_i} X_{t,m_i} = \sum_{b=1}^{b^\diamond} R_{b,m_i}$. Let σ_b be the σ-fields generated by $\{\varepsilon_{I_b}, \varepsilon_{I_b - 1}, \ldots\}$ where $I_b = \max \mathcal{B}^\diamond_b$. It can be seen that $(R_{b,m_i})_{b \text{ is odd}}$ is a martingale difference sequence with respect to $(\sigma_b)_{b \text{ is odd}}$, and so is $(R_{b,m_i})_{b \text{ is even}}$ with respect to $(\sigma_b)_{b \text{ is even}}$. By Lemma 1 of Haeusler [1984], for any $M > 1$, there exists some constant $C_{\kappa,M}$ such that

$$\mathbb{P}(|\sum_{t-s>3m_i} a_{s,t} X_{s,m_i} X_{t,m_i}| \geq x) \leq C_{\kappa,M} x^{-M} + \mathrm{III}_B + \mathrm{IV}_B, \tag{3.7.28}$$

where

$$\begin{aligned}
\mathrm{III}_B &= 4\mathbb{P}\left(\sum_{b=1}^{b^\diamond} \mathbb{E}(R_{b,m_i}^2 | \sigma_{b-2}) \geq \frac{x^2}{(\log x)^{3/2}}\right), \\
\mathrm{IV}_B &= \sum_{b=1}^{b^\diamond} \mathbb{P}\left(|R_{b,m_i}| \geq \frac{x}{\log x}\right).
\end{aligned}$$

Denote $\gamma_{s-t,m_i} = \mathbb{E}(X_{s,m_i} X_{t,m_i})$. We then have

$$\begin{aligned}
\sum_{b=1}^{b^\diamond} \mathbb{E}(R_{b,m_i}^2 | \sigma_{b-2}) &\leq \sum_{b=1}^{b^\diamond} \sum_{s,t \in \mathcal{B}_b^\diamond} \gamma_{s-t,m_i} Y_{s,m_i}^* Y_{t,m_i}^* \\
&\leq \sum_{1 \leq s \leq t \leq B} a_{s,t}^\diamond X_{s,m_i} X_{t,m_i}.
\end{aligned}$$

By Lemma 3.7.2 (ii), we have $\sum_{k \in \mathbb{Z}} \left|\gamma_{k,m_i}\right| \leq \|X_\cdot\|_{2,\alpha}^2 / \|X_\cdot\|_{q,\alpha}^2 \leq 1$, which implies $\left|a_{s,t}^\diamond\right| \leq B$ and $\mathbb{E} \sum_{1 \leq s \leq t \leq B} a_{s,t}^\diamond X_{s,m_i} X_{t,m_i} \leq B^2$. Hence it follows that

$$\begin{aligned}
\mathrm{III}_B &\leq 4\mathbb{P}\left(\sum_{1 \leq s \leq t \leq B} a_{s,t}^\diamond X_{s,m_i} X_{t,m_i} \geq \frac{x^2}{(\log x)^{3/2}}\right) \\
&\lesssim U\left(3m_i, m_i, \frac{x^2}{B(\log x)^2}\right) + V\left(3m_i, m_i, \frac{x^2}{B(\log x)^2}\right). \qquad (3.7.29)
\end{aligned}$$

Now we deal with the term V_B. Conditioned on $Y_{t,m_i}^* = y_{t,m_i}$, by Theorem 2 in Wu and Wu [2016],

$$\mathbb{P}(|\sum_{t \in \mathcal{B}_b^\diamond} X_{t,m_i} y_{t,m_i}| \geq x) \lesssim x^{-2r} \xi_{m_i} \sum_{t \in \mathcal{B}_b^\diamond} |y_{t,m_i}|^{2r} + \exp\left(-\frac{C_{r,\alpha} x^2}{\sum_{t \in \mathcal{B}_b^\diamond} |y_{t,m_i}|^2}\right),$$

where $\xi_{m_i} = 1$ (resp. $(\log m_i)^{1+4r}$ or $m_i^{r-1-2\alpha r}$) if $\alpha > 1/2 - 1/(2r)$ (resp. $\alpha = 1/2 - 1/(2r)$

or $\alpha < 1/2 - 1/(2r)$). By the independence of X_{t,m_i} and Y^*_{t,m_i}, we have

$$\begin{aligned}\mathbb{P}(|R_{b,m_i}| \geq x) &\lesssim x^{-2r}\xi_{m_i}\sum_{t\in\mathcal{B}_b^\diamond}\mathbb{E}|Y^*_{t,m_i}|^{2r} + \mathbb{E}\exp\left(-\frac{C_{r,\alpha}x^2}{\sum_{t\in\mathcal{B}_b^\diamond}|Y^*_{t,m_i}|^2}\right)\\ &=: x^{-2r}\xi_{m_i}D_1 + D_2. \end{aligned} \tag{3.7.30}$$

Lemma 3.7.2 (iii) implies $D_1 \leq C_r m_i B^r$.

Also, for any $M > 1$, there exists a constant $C_{\kappa,r,\alpha,M}$ such that

$$\begin{aligned} D_2 &= \mathbb{E}\exp\left(-\frac{C_{r,\alpha}x^2}{\sum_{t\in\mathcal{B}_b^\diamond}|Y^*_{t,m_i}|^2}\right)\mathbf{1}\left\{\sum_{t\in\mathcal{B}_b^\diamond}|Y^*_{t,m_i}|^2 > \frac{x^2}{(\log x)^{3/2}}\right\}\\ &\quad+\mathbb{E}\exp\left\{-\frac{C_{r,\alpha}x^2}{\sum_{t\in\mathcal{B}_b^\diamond}|Y^*_{t,m_i}|^2}\right\}\mathbf{1}\left\{\sum_{t\in\mathcal{B}_b^\diamond}|Y^*_{t,m_i}|^2 \leq \frac{x^2}{(\log x)^{3/2}}\right\}\\ &\leq \mathbb{P}\left(\sum_{t\in\mathcal{B}_b^\diamond}|Y^*_{t,m_i}|^2 > \frac{x^2}{(\log x)^{3/2}}\right) + C_{\kappa,r,\alpha,M}x^{-M}, \end{aligned}$$

the first term of which can be further formulated by a similar argument as III_B. Then we have

$$\begin{aligned}\mathbb{P}(|R_{b,m_i}| \geq x) &\leq C_r\frac{m_iB^r\xi_{m_i}}{x^{2r}} + C_{\kappa,r,\alpha,M}x^{-M}\\ &\quad+C_{r,\alpha}\mathbb{P}\left(\sum_{1\leq s\leq t\leq B}a'_{s,t}X_{s,m_i}X_{t,m_i} \geq \frac{x^2}{m_i(\log x)^2}\right), \end{aligned}\tag{3.7.31}$$

with $\sup_{s,t}|a'_{s,t}| \leq 1$. Hence

$$\mathrm{IV}_B \lesssim \frac{B^{r+1}\xi_B}{(x/\log x)^{2r}} + \frac{B}{m_i}\mathbb{P}\left(\sum_{1\leq s\leq t\leq B}a'_{s,t}X_{s,m_i}X_{t,m_i} \geq \frac{x^2}{m_i(\log x)^6}\right) + x^{-M}, \tag{3.7.32}$$

where $\xi_B = 1$ (resp. $(\log B)^{1+4r}$ or $B^{r-1-2\alpha r}$) if $\alpha > 1/2 - 1/(2r)$ (resp. $\alpha = 1/2 - 1/(2r)$ or $\alpha < 1/2 - 1/(2r)$), and the constant in $\lesssim$ only depends on κ, r, α and M. Combining

(3.7.28), (3.7.29) and (3.7.32), we have

$$\begin{aligned} U(3m_i, m_i, x) &\lesssim U\left(3m_i, m_i, \frac{x^2}{B(\log x)^2}\right) + V\left(3m_i, m_i, \frac{x^2}{B(\log x)^2}\right) \\ &\quad + \frac{B}{m_i}\mathbb{P}\left(\sum_{1\le s\le t\le B} a'_{s,t}X_{s,m_i}X_{t,m_i} \ge \frac{x^2}{m_i(\log x)^6}\right) \\ &\quad + \frac{B^{r+1}\xi_B}{(x/\log x)^{2r}} + x^{-M}. \end{aligned} \tag{3.7.33}$$

By applying (3.7.33) for J times such that $(x/B)^{-2^{J+1}r} = O(x^{-(M+1)})$, we have

$$\begin{aligned} U(3m_i, m_i, x) &\lesssim V\left(3m_i, m_i, \frac{x^2}{B(\log x)^2}\right) + \frac{B^{r+1}\xi_B}{(x/\log x)^{2r}} + x^{-M} \\ &\quad + \frac{B}{m_i}\mathbb{P}\left(\sum_{1\le s\le t\le B} a'_{s,t}X_{s,m_i}X_{t,m_i} \ge \frac{x^2}{m_i(\log x)^6}\right). \end{aligned} \tag{3.7.34}$$

We further split the last term in (3.7.34) into

$$\frac{B}{m_i}V\left(3m_i, m_i, \frac{x^2}{m_i(\log x)^6}\right) + \frac{B}{m_i}U\left(3m_i, m_i, \frac{x^2}{m_i(\log x)^6}\right).$$

Applying the new recursion for J times as well, and it becomes

$$\begin{aligned} U(3m_i, m_i, x) &\lesssim \sum_{j=1}^{J}\left(\frac{B}{m_i}\right)^{j-1} V\left(3m_i, m_i, x_{j,B}\right) \\ &\quad + \sum_{j=1}^{J}\left(\frac{B}{m_i}\right)^{j} V\left(3m_i, m_i, x_{j,m_i}\right) \\ &\quad + \frac{B^{r+1}\xi_B}{(x/\log x)^{2r}} + x^{-M}, \end{aligned} \tag{3.7.35}$$

where

$$
\begin{aligned}
x_{1,B} &= \frac{x^2}{B(\log x)^2}, x_{j,B} = \frac{x_{j-1,m_i}^2}{B(\log x_{j-1,m_i})^2}, \quad j \geq 2, \\
x_{1,m_i} &= \frac{x^2}{m_i(\log x)^6}, x_{j,m_i} = \frac{x_{j-1,m_i}^2}{m_i(\log x_{j-1,m_i})^6}, \quad j \geq 2,
\end{aligned}
$$

and the constant in $\lesssim$ only depends on κ, r, α, M.

Step 3.2: Bounding $V(3m_i, m_i, x)$, $1 \leq i \leq \tau$.

Define $y = x/(2\tau)$ with $\tau = \lfloor \log B / \log 4 \rfloor$. Then y satisfies $B^{1+\kappa/2} = o(y)$. Let $\{a_{s,t}^\circ\}$ be the set of coefficients such that

$$\mathbb{P}(|\mathbb{E}_0 \sum_{t-s \leq 3m_i} a_{s,t}^\circ X_{s,m_i} X_{t,m_i}| \geq x) = V(3m_i, m_i, x).$$

Split $V(3m_i, m_i, x)$ into three parts.

$$V(3m_i, m_i, x) \leq \mathrm{V}_B + \mathrm{VI}_B + V(3m_{i+1}, m_{i+1}, x - 2y), \tag{3.7.36}$$

where

$$
\begin{aligned}
\mathrm{V}_B &= \mathbb{P}(|\mathbb{E}_0 \sum_{t-s \leq 3m_i} (a_{s,t}^\circ X_{s,m_i} X_{t,m_i} - a_{s,t}^\circ X_{s,m_{i+1}} X_{t,m_{i+1}})| \geq y), \\
\mathrm{VI}_B &= \mathbb{P}(|\mathbb{E}_0 \sum_{3m_{i+1} < t-s \leq 3m_i} a_{s,t}^\circ X_{s,m_{i+1}} X_{t,m_{i+1}}| \geq y).
\end{aligned}
$$

Using similar arguments when dealing with I_i, we have

$$\mathrm{V}_B \leq \frac{C_r B m_i^{r-1} m_{i+1}^{-\alpha r}}{y^r}. \tag{3.7.37}$$

To deal with VI_B, let $b^\circ = \lceil B/4m_i \rceil$. For $1 \leq b \leq b^\circ$, define $\mathcal{B}_b^\circ = [4(b-1)m_i + 1, (4bm_i) \wedge B]$ and $R_{b,m_i}^\circ = \sum_{t \in \mathcal{B}_b^\circ, 3m_{i+1} < t-s \leq 3m_i} a_{s,t}^\circ X_{s,m_{i+1}} X_{t,m_{i+1}}$. By Corollary 1.6 of Nagaev [1979], for any $M > 1$,

$$\mathrm{VI}_B \lesssim \sum_{b=1}^{b^\circ} \mathbb{P}(|\mathbb{E}_0 R^\circ_{b,m_i}| \geq y/2C_{\kappa,M}) + y^{-M}. \tag{3.7.38}$$

Observe that when $t-s > 3m_2$, X_{s,m_2} and X_{t,m_2} are independent. By the similar argument as (3.7.31), we have from (3.7.38)

$$\begin{aligned} W(3m_{i+1}, 3m_i, m_{i+1}, y) &\lesssim \frac{B}{m_i} W\left(3m_{i+1}, 3m_i, m_{i+1}, y_{1,m_i}\right) + y^{-M} \\ &\quad + \frac{B}{m_i} V\left(3m_{i+1}, m_{i+1}, y_{1,m_i}\right) + \frac{B^{r+1}\xi_B}{y^{2r}}, \end{aligned} \tag{3.7.39}$$

where $y_{1,m_i} = y^2/[m_i(\log y)^2]$. By applying (3.7.39) recursively for v times such that $(y/B)^{-2^{v+1}r} = O(y^{-(M+1)})$,

$$\begin{aligned} W(3m_{i+1}, 3m_i, m_{i+1}, y) &\lesssim \sum_{j=1}^{v} \left(\frac{B}{m_i}\right)^j V\left(3m_{i+1}, m_{i+1}, y_{j,m_i}\right) \\ &\quad + \frac{B^{r+1}\xi_B}{y^{2r}} + y^{-M}, \end{aligned} \tag{3.7.40}$$

where $y_{j,m_i} = y^2_{j-1,m_i}/[m_i(\log y^2_{j-1,m_i})^2]$ for $j \geq 2$. Hence (3.7.36) becomes

$$\begin{aligned} V(3m_i, m_i, x) &\lesssim \sum_{j=1}^{v} \left(\frac{B}{m_i}\right)^j V\left(3m_{i+1}, m_{i+1}, y_{j,m_i}\right) + \frac{B^{r+1}\xi_B}{y^{2r}} \\ &\quad + \frac{B m_i^{r-1} m_{i+1}^{-\alpha r}}{y^r} + y^{-M} + V(3m_{i+1}, m_{i+1}, x-2y). \end{aligned} \tag{3.7.41}$$

By applying (3.7.41) for at most $\tau - i$ times such that $1 \leq m_\tau \leq 2$, we have

$$\begin{aligned} V(3m_i, m_i, x) &\lesssim \sum_{k=1}^{\tau-1} \sum_{j=1}^{v} \left(\frac{B}{m_k}\right)^j V\left(3m_{k+1}, m_{k+1}, y_{j,m_k}\right) + \frac{\tau B^{r+1}\xi_B}{y^{2r}} \\ &\quad + \frac{B}{y^r} \sum_{k=1}^{\tau-1} m_k^{r-1} m_{k+1}^{-\alpha r} + \tau y^{-M} + V(3m_\tau, m_\tau, 2y). \end{aligned} \tag{3.7.42}$$

By the block technique and Lemma 3.7.2 (3), it can be easily obtained that

$$V(3m_\tau, m_\tau, 2y) \le \frac{C_r B}{y^r}. \tag{3.7.43}$$

Notice that $\sum_{k=1}^{\tau-1} m_k^{r-1} m_{k+1}^{-\alpha r} = O(\zeta_B/B)$. We then have

$$\begin{aligned} V\left(3m_i, m_i, x\right) &\lesssim \sum_{k=1}^{\tau-1}\sum_{j=1}^{v}\left(\frac{B}{m_k}\right)^j V\left(3m_{k+1}, m_{k+1}, y_{j,m_k}\right) \\ &\quad + \frac{\zeta_B}{y^r} + \frac{\tau B^{r+1}\xi_B}{y^{2r}} + \tau y^{-M}. \end{aligned} \tag{3.7.44}$$

For each $1 \le k \le \tau - 1$, $V(3m_{k+1}, m_{k+1}, y_{j,m_k})$ can be further dealt with by applying (3.7.44) recursively for at most v times. Consequently we have

$$\begin{aligned} V(3m_i, m_i, x) &\lesssim \frac{\tau\zeta_B}{y^r} + \frac{\tau^2 B^{r+1}\xi_B}{y^{2r}} + \tau^2 y^{-M} \\ &\lesssim \frac{(\log B)^{r+1}\zeta_B}{x^r} + \frac{(\log B)^{2r+2}B^{r+1}\xi_B}{x^{2r}} + x^{-M}. \end{aligned} \tag{3.7.45}$$

Step 3.3: Bounding $W(3m_i, 3m_{i-1}, m_i, x)$ and II_i, $2 \le i \le \tau$.

Plugging (3.7.45) into (3.7.35) and (3.7.40), we obtain a bound of $U(3m_i, m_i, x)$ and $W(3m_{i+1}, 3m_i, m_{i+1}, x)$, $1 \le i \le \tau$ respectively,

$$\begin{aligned} U(3m_i, m_i, x) &\lesssim \frac{B^r\zeta_B(\log x)^{7r+1}}{x^{2r}} + \frac{B^{3r+1}\xi_B(\log x)^{14r+2}}{x^{4r}} + x^{-M}, \\ W(3m_{i+1}, 3m_i, m_{i+1}, x) &\lesssim \frac{B^r\zeta_B(\log x)^{3r+1}}{x^{2r}} + \frac{B^{3r+1}\xi_B(\log x)^{6r+2}}{x^{4r}} + x^{-M}. \end{aligned}$$

With the above two bounds, (3.7.27) implies

$$\mathrm{II}_i \lesssim \frac{(\log x)^{7r+1}B^r\zeta_B}{(\lambda_i x)^{2r}} + \frac{(\log x)^{14r+2}B^{3r+1}\xi_B}{(\lambda_i x)^{4r}} + x^{-M}. \tag{3.7.46}$$

Since $B^{1+\kappa} = o(x)$, we can choose a sufficiently large M such that the last term of (3.7.46) is no greater than the first term. By the choice for λ_i's, it holds that for x satisfying $B^{1+\kappa} = o(x)$,

$$\sum_{i=2}^{\tau} \mathrm{II}_i \leq C_{\kappa,r,\alpha} x^{-r} \zeta_B. \tag{3.7.47}$$

Step 4: Bounding $\mathbb{P}\left(\left|\mathbb{E}_0 Q_{B,m_\tau}\right| \geq x\right)$ with $1 \leq m_\tau \leq 2$.

Notice that

$$\mathbb{P}(|\mathbb{E}_0 Q_{B,m_\tau}| \geq x) \leq V(3m_\tau, m_\tau, x/2) + U(3m_\tau, m_\tau, x/2).$$

By (3.7.43) and the bound of $U(3m_\tau, m_\tau, x)$ in Step 2.3, it follows that

$$\mathbb{P}\left(\left|\mathbb{E}_0 Q_{B,m_\tau}\right| \geq x\right) \leq C_{\kappa,r,\alpha} x^{-r} \zeta_B. \tag{3.7.48}$$

Combining (3.7.23), (3.7.24), (3.7.25), (3.7.26), (3.7.47) and (3.7.48), (3.7.22) follows, which implies (3.6.7) immediately. □

Proof of Theorem 3.6.3. Similarly as the proof of Theorem 3.6.2, we first normalized the process by $\|X.\|_{q,\alpha}$. For $0 < \beta < 1$, let $B = B_1 = \lfloor T^\beta \rfloor$. It remains to show for $x \geq T^{1/2}$, (i) if $l \geq 3B$,

$$\mathbb{P}(|\mathbb{E}_0 L_T(l)| \geq x) \lesssim x^{-q/2}(T + T^{q/4-\alpha\beta q/2}) \tag{3.7.49}$$

and (ii) if $0 \leq l \leq 3B$,

$$\mathbb{P}(|\mathbb{E}_0 L_T(l)| \geq x) \lesssim \frac{D_T + TB^{q/4-\alpha\beta q/2-1}}{x^{q/2}} + \exp\left(-\frac{C_{q,\alpha} x^2 \|X.\|_{q,\alpha}^4}{\|X.\|_{4,\alpha}^4 T}\right). \tag{3.7.50}$$

Case 1: $l > 3B$.

Define the B-approximation of $L_T(l)$ as $L_{T,B}(l) = \sum_{l+1 \leq t \leq T} a_t X_{t-l,B} X_{t,B}$. We can follow the similar proof of Lemma 3.7.3 to have

$$\|\mathbb{E}_0 L_T(l) - \mathbb{E}_0 L_{T,B}(l)\|_{q/2} \leq C_{q,\alpha}(T-l)^{1/2} B^{-\alpha},$$

which implies

$$\mathbb{P}(|\mathbb{E}_0 L_T(l) - \mathbb{E}_0 L_{T,B}(l)| \geq x) \leq C_{q,\alpha} x^{-q/2} (T-l)^{q/4} B^{-\alpha q/2}. \tag{3.7.51}$$

Let $b^\star = \lceil (T-l)/B \rceil$. For $1 \leq b \leq b^\star$, define $\mathcal{B}_b^\star = [l + (b-1)B + 1, (l + bB) \wedge T]$ and

$$L_{b,B}^\star = \sum_{t \in \mathcal{B}_b^\star} a_t X_{t-l,B} X_{t,B}.$$

Let $\sigma_b^\star$ be the σ-fields generated by $\{\varepsilon_{I_b^\star}, \varepsilon_{I_b^\star - 1}, \ldots\}$ where $I_b^\star = \max \mathcal{B}_b^\star$. We can apply similar arguments as Step 3.1 (cf. (3.7.28)) when proving Theorem 3.6.2 to obtan for any $M \geq 1$,

$$\mathbb{P}(|L_{T,B}(l)| \geq x) \leq C_M x^{-M} + L_T^{(1)} + L_T^{(2)}, \tag{3.7.52}$$

where

$$\begin{aligned} L_T^{(1)} &= 4\mathbb{P}\left(\sum_{b=1}^{b^\star} \mathbb{E}(L_{b,B}^{\star 2} | \sigma_{b-2}^\star) \geq \frac{x^2}{(\log x)^{3/2}} \right), \\ L_T^{(2)} &= \sum_{b=1}^{b^\star} \mathbb{P}\left(|L_{b,B}^\star| \geq \frac{x}{\log x} \right). \end{aligned}$$

Denote $\gamma_{s-t,B} = \mathbb{E}(X_{s,B} X_{t,B})$. We then have

$$\sum_{b=1}^{b^\star} \mathbb{E}(L_{b,B}^{\star 2} | \sigma_{b-2}^\star) \leq \sum_{b=1}^{b^\star} \sum_{s,t \in \mathcal{B}_b^\star} a_s a_t \gamma_{s-t,B} X_{s-l,B} X_{t-l,B}.$$

Notice that $\sup_{s,t} |a_s a_t \gamma_{s-t,B}| \leq 1$. By Theorem 3.6.1, for $x \geq T^{1/2}$, we have

$$\begin{aligned} L_T^{(1)} &\lesssim \frac{(\log x)^{3q/4} F_{T,B}}{x^q} + \exp\left(-\frac{C_{q,\alpha} x^4 \|X_\cdot\|_{q,\alpha}^4}{(\log x)^3 \|X_\cdot\|_{4,\alpha}^4 TB} \right) \\ &\lesssim \frac{(\log x)^{3q/4} F_{T,B}}{x^q}, \end{aligned} \tag{3.7.53}$$

where the constant in $\lesssim$ only depends on q, α and β, and $F_{T,B}$ has been defined in Theorem 3.6.1. Let

$$D_1^\star = \sum_{t\in\mathcal{B}_b^\star} \mathbb{E}|X_{t-l,B}|^q \quad \text{and} \quad D_2^\star = \mathbb{E}\exp\left(-\frac{C_{q,\alpha}(x/\log x)^2}{\sum_{t\in\mathcal{B}_b^\star}|X_{t-l,B}|^2}\right).$$

By the independence of $X_{t,B}$ and $X_{t-l,B}$, similarly as (3.7.30), we have

$$\mathbb{P}\left(|L_{b,B}^\star| \geq \frac{x}{\log x}\right) \lesssim \frac{\xi_B D_1^\star}{(x/\log x)^q} + D_2^\star,$$

where $\xi_B = 1$ (resp. $(\log B)^{1+2q}$ or $B^{q/2-1-\alpha q}$) if $\alpha > 1/2-1/q$ (resp. $\alpha = 1/2-1/q$ or $\alpha < 1/2-1/q$). It is easy to have $D_1^\star \leq B$. Notice that $\sum_{t\in\mathcal{B}_b^\star}\mathbb{E}|X_{t-l,B}|^2 \leq B\|X_\cdot\|_{2,\alpha}^2/\|X_\cdot\|_{q,\alpha}^2$, which is negligible compared to $x^2/(\log x)^{7/2}$ when $x \geq T^{1/2}$. By Theorem 2 in Wu and Wu [2016] for the process $|X_{t-l,B}|^2$, for any $M \geq 1$,

$$\begin{aligned}
D_2^\star &\leq \left(\sum_{t\in\mathcal{B}_b^\star}|X_{t-l,B}|^2 > \frac{x^2}{(\log x)^{7/2}}\right) + C_{q,\alpha,M}x^{-M}\\
&\leq \left(\sum_{t\in\mathcal{B}_b^\star}|X_{t-l,B}|^2 - \mathbb{E}|X_{t-l,B}|^2 > \frac{x^2}{2(\log x)^{7/2}}\right) + C_{q,\alpha,M}x^{-M}\\
&\lesssim \frac{(\log x)^{7q/4}\xi_B^\star}{x^q} + \exp\left(-\frac{C_{q,\alpha}x^4\|X_\cdot\|_{q,\alpha}^4}{(\log x)^7\|X_\cdot\|_{4,\alpha}^4 B}\right) + x^{-M},
\end{aligned}$$

where $\xi_B^\star = B$ (resp. $B(\log B)^{1+q}$ or $B^{q/4-\alpha q/2}$) if $\alpha > 1/2-2/q$ (resp. $\alpha = 1/2-2/q$ or $\alpha < 1/2-2/q$). Summing over b, for $x \geq T^{1/2}$, we obtain

$$L_T^{(2)} \lesssim \frac{(\log x)^q T\xi_B}{x^q} + \frac{(\log x)^{7q/4}T\xi_B^\star}{Bx^q} + TB^{-1}x^{-M}, \tag{3.7.54}$$

where the constant in $\lesssim$ only depends on q, α, β and M. Combining (3.7.52), (3.7.53), (3.7.54) and choosing a sufficiently large M, we have for $x \geq T^{1/2}$,

$$\mathbb{P}(|L_{T,B}(l)| \geq x) \leq C_{q,\alpha,\beta} x^{-q} (\log x)^{2q} F_{T,B} \leq C_{q,\alpha} x^{-q/2} D_T. \tag{3.7.55}$$

Hence, for all cases of α, (3.7.49) follows by (3.7.51) and (3.7.55).

Case 2: $3e < l \leq 3B$.

Let $B_0 = T$, $B_j = \lfloor B_{j-1}^{\beta} \rfloor$ and $n^{(j)} = \lceil (T-l)/(4B_j) \rceil$ for $j \geq 1$. Let

$$J = \lceil -\log(\log T)/\log \beta \rceil.$$

Then $B_J \leq \lfloor T^{\beta^J} \rfloor \leq e$. For $3e < l \leq 3B$, we can find a J_0 ($1 \leq J_0 \leq J-1$) such that $3B_{J_0+1} < l \leq 3B_{J_0}$. For $k = 1, 2, \ldots, n^{(J_0)}$, define

$$L_k^{(J_0)} = \sum_{t=l+4(k-1)B_{J_0}+1}^{(l+4kB_{J_0})\wedge T} a_t X_{t-l} X_t.$$

For $h \in \mathbb{Z}$, we define the innovation sets

$$\eta_h^{(J_0)} = (\varepsilon_{4(h-1)B_{J_0}+1}, \varepsilon_{4(h-1)B_{J_0}+2}, \ldots, \varepsilon_{4hB_{J_0}}).$$

and

$$L_{k,\tau}^{(J_0)} = \mathbb{E}\left(L_k^{(J_0)} | \eta_{k-\tau}^{(J_0)}, \eta_{k-\tau+1}^{(J_0)}, \ldots, \eta_k^{(J_0)} \right), \text{ for } \tau \geq 0$$

We can apply the similar arguments when proving Theorem 3.6.1 to have for $x \geq T^{1/2}$,

$$\begin{aligned} \mathbb{P}(|\mathbb{E}_0 L_T(l)| \geq x) \ &\lesssim \ \frac{W_{T,B_{J_0}}}{x^{q/2}} + \exp\left(-\frac{C_{q,\alpha} x^2 \|X.\|_{q,\alpha}^4}{\|X.\|_{4,\alpha}^4 T} \right) \\ &\quad + \sum_{k=1}^{n^{(J_0)}} \mathbb{P}(|\mathbb{E}_0 L_{k,0}^{(J_0)}| \geq x/C_{q,\alpha,\beta}), \end{aligned} \tag{3.7.56}$$

where $W_{T,B_{J_0}} = TB_{J_0}{}^{q/4-\alpha q/2-1}$ (resp. $T(\log T)^{1+q}$ or $T^{q/4-\alpha q/2}$) for $\alpha > 1/2-2/q$ (resp. $\alpha = 1/2-2/q$ or $\alpha < 1/2-2/q$). To deal with $\mathbb{P}(|\mathbb{E}_0 L_{k,0}^{(J_0)}| \geq x)$, similarly as (3.7.51), we take B_{J_0+1}-approximation of $L_{k,0}^{(J_0)}$ and bound the tail probability of the difference by

$$\mathbb{P}\left(\left|\mathbb{E}_0 L_{k,0}^{(J_0)} - \mathbb{E}_0 L_{k,0,B_{J_0+1}}^{(J_0)}\right| \geq x\right) \leq C_{q,\alpha} x^{-q/2} B_{J_0}^{q/4} B_{J_0+1}^{-\alpha q/2}. \tag{3.7.57}$$

Since $l > 3B_{J_0+1}$, by (3.7.55), for $x \geq T^{1/2}$, we have

$$\mathbb{P}\left(\left|\mathbb{E}_0 L_{k,0,B_{J_0+1}}^{(J_0)}\right| \geq x\right) \leq C_{q,\alpha,\beta} x^{-q} (\log x)^{2q} F_{B_{J_0},B_{J_0+1}}. \tag{3.7.58}$$

Since $x \geq T^{1/2}$, (3.7.50) follows in view of (3.7.56), (3.7.57) and (3.7.58).

Case 3: $0 \leq l \leq 3e$.

This case is easier. Let $n^\star = \lceil (T-l)/4e \rceil$. For $k = 1, 2, \ldots, n^\star$, define

$$L_k^\star = \sum_{t=l+4e(k-1)+1}^{(l+4ek)\wedge T} a_t X_{t-l} X_t$$

and

$$L_{k,0}^\star = \mathbb{E}(L_k^\star | \varepsilon_{4e(k-1)+1}, \varepsilon_{4e(k-1)+2}, \ldots, \varepsilon_{4ek}).$$

Following the proof of Theorem 3.6.1, (3.7.56) then becomes

$$\begin{aligned} \mathbb{P}(|\mathbb{E}_0 L_T(l)| \geq x) \quad &\lesssim \quad \frac{D_T}{x^{q/2}} + \exp\left(-\frac{C_{q,\alpha} x^2 \|X_\cdot\|_{q,\alpha}^4}{\|X_\cdot\|_{4,\alpha}^4 T}\right) \\ &\quad + \sum_{k=1}^{n^\star} \mathbb{P}(|\mathbb{E}_0 L_{k,0}^\star| \geq x / C_{q,\alpha,\beta}). \end{aligned} \tag{3.7.59}$$

Notice that $L_{k,0}^\star$ takes the sum over a block with size $4e$. By Lemma 3.7.2 (iv) and the Markov inequality,

$$\mathbb{P}(|\mathbb{E}_0 L_{k,0}^\star| \geq x) \leq C_{q,\alpha} x^{-q/2}.$$

So (3.7.50) satisfies in view of (3.7.59). □

REFERENCES

S. Adak. Time-dependent spectral analysis of nonstationary time series. *Journal of the American Statistical Association*, 93(444):1488–1501, 1998.

C. Alexopoulos and D. Goldsman. To batch or not to batch? *ACM Trans. Model. Comput. Simul.*, 14(1):76–114, 2004.

H.Z. An, Z.G. Chen, and E.J. Hannan. Autocorrelation, autoregression and autoregressive approximation. *The Annals of Statistics*, pages 926–936, 1982.

T.W. Anderson. *The Statistical Analysis of Time Series.* Wiley, 1971.

A. Azzalini. A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, 68(1):326–328, 1981.

F.R. Bach and M.I. Jordan. Learning graphical models for stationary time series. *IEEE transactions on signal processing*, 52(8):2189–2199, 2004.

B. Bercu, F. Gamboa, and A. Rouault. Large deviations for quadratic forms of stationary gaussian processes. *Stochastic Processes and their Applications*, 71(1):75 – 90, 1997.

P.J. Bickel, Y. Ritov, and A.B. Tsybakov. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, pages 1705–1732, 2009.

R.C. Bradley. *Introduction to Strong Mixing Conditions.* Kendrick Press, 2007.

D.R. Brillinger. *Time Series: Data Analysis and Theory.* Holt, Rinehart, and Winston, 1975.

D.R. Brillinger. Remarks concerning graphical models for time series and point processes. *Revista de Econometria*, 16(1):23, 1996.

P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods.* Springer, 1991.

W. Bryc and A. Dembo. Large deviations for quadratic functionals of gaussian processes. *Journal of Theoretical Probability*, 10(2):307–332, 1997.

P. Bühlmann. Bootstraps for time series. *Statistical Science*, 17(1):52–72, 05 2002.

D.L. Burkholder. Distribution function inequalities for martingales. *The Annals of Probability*, 1(1):19–42, 02 1973.

T. Cai and T. Jiang. Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *The Annals of Statistics*, 39(3):1496–1525, 2011.

T. Cai, W. Liu, and X. Luo. A constrained l1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011.

E. Candes and T. Tao. The dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, pages 2313–2351, 2007.

X. Chen, M. Xu, and W.B. Wu. Covariance and precision matrix estimation for high-dimensional time series. *The Annals of Statistics*, 41(6):2994–3021, 2013.

X. Chen, Q.M. Shao, W.B. Wu, and Lihu Xu. Self-normalized Cramér type moderate deviations under dependence. *arXiv preprint arXiv:1409.3642*, 2015.

M.Y. Cheng and L. Peng. Regression modeling for nonparametric estimation of distribution and quantile functions. *Statistica Sinica*, pages 1043–1060, 2002.

V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819, 12 2013a.

V. Chernozhukov, D. Chetverikov, and K. Kato. Testing many moment inequalities. *arXiv preprint arXiv:1312.7614*, 2013b.

V. Chernozhukov, D. Chetverikov, and K. Kato. Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields*, 162(1-2): 47–70, 2014.

R. Dahlhaus. Fitting time series models to nonstationary processes. *The Annals of Statistics*, 25(1):1–37, 02 1997.

R. Dahlhaus. A likelihood approximation for locally stationary processes. *The Annals of Statistics*, 28(6):1762–1794, 12 2000a.

R. Dahlhaus. Graphical interaction models for multivariate time series1. *Metrika*, 51(2): 157–172, 2000b.

R. Dahlhaus. Locally stationary processes. *Handbook of Statistics*, 30:351–412, 2012.

R. Dahlhaus and W. Polonik. Empirical spectral processes for locally stationary time series. *Bernoulli*, 15(1):1–39, 02 2009.

R. Dahlhaus and S. Subba Rao. Statistical inference for time-varying arch processes. *The Annals of Statistics*, 34(3):1075–1114, 2006.

R. Dahlhaus and S. Subba Rao. A recursive online algorithm for the estimation of time-varying arch parameters. *Bernoulli*, 13(2):389–422, 2007.

G. Deco, V.K. Jirsa, and A.R. McIntosh. Resting brains never rest: computational insights into potential cognitive architectures. *Trends in neurosciences*, 36(5):268–274, 2013.

J. Dedecker, P. Doukhan, G. Lang, L. Rafael, S. Louhichi, and C. Prieur. *Weak Dependence: With Examples and Applications*. Springer, 2007.

D.L. Donoho, M. Elad, and V.N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1): 6–18, 2006.

M. Eichler. Granger causality and path diagrams for multivariate time series. *Journal of Econometrics*, 137(2):334–353, 2007.

M. Eichler. Graphical modelling of multivariate time series. *Probability Theory and Related Fields*, 153(1-2):233–268, 2012.

U. Einmahl and D. Li. Characterization of lil behavior in banach space. *Transactions of the American Mathematical Society*, 360(12):6677–6693, 2008.

M. Falk. Asymptotic normality of the kernel quantile estimator. *The Annals of Statistics*, 13(1):428–433, 1985.

X. Fernique. Regularité des trajectoires des fonctions aléatoires gaussiennes. In *École d'Été de Probabilités de Saint-Flour, IV-1974*, pages 1–96. Lecture Notes in Math., Vol. 480. Springer, Berlin, 1975.

R. Fried and V. Didelez. Decomposability and selection of graphical models for multivariate time series. *Biometrika*, 90(2):251–267, 2003.

P. Fryzlewicz and S. Subba Rao. Mixing properties of arch and time-varying arch processes. *Bernoulli*, 17(1):320–346, 2011.

U. Gather, M. Imhoff, and R. Fried. Graphical models for multivariate time series from intensive care monitoring. *Statistics in medicine*, 21(18):2685–2701, 2002.

M. Giurcanu and V. Spokoiny. Confidence estimation of the covariance function of stationary and locally stationary processes. *Statistics and Decisions*, 22:283–300, 2004.

Y. Grenier. Time-dependent arma modeling of nonstationary signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(4):899–911, 1983.

Erich Haeusler. An exact rate of convergence in the functional central limit theorem for special martingale difference arrays. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 65(4):523–534, 1984.

C.M. Hafner and O. Linton. Efficient estimation of a multivariate multiplicative volatility model. *Journal of Econometrics*, 159(1):55–73, 2010.

E.J. Hannan. The uniform convergence of autocovariances. *The Annals of Statistics*, 2(4): 803–806, 1974.

E.J. Hannan and M. Deistler. *The Statistical Theory of Linear Systems*. John Wiley & Sons Inc., New York., 1988.

D.L. Hanson and F.T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Mathematical Statistics*, 42(3):1079–1083, 1971.

J.F. Hipp, D.J. Hawellek, M. Corbetta, M. Siegel, and A.K. Engel. Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nature neuroscience*, 15(6):884–890, 2012.

M.B. Hooten and C.K. Wikle. A hierarchical bayesian non-linear spatio-temporal model for the spread of invasive species with application to the eurasian collared-dove. *Environmental and Ecological Statistics*, 15(1):59–70, 2008.

R.M. Hutchison, T. Womelsdorf, E.A. Allen, P.A. Bandettini, V.D. Calhoun, M. Corbetta, Della Penna S., J.H. Duyn, G.H. Glover, J. Gonzalez-Castillo, D.A. Handwerker, S. Keilholz, V. Kiviniemi, D.A. Leopold, F. de Pasquale, O. Sporns, M. Walter, and C. Chang. Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, 80: 360–378, 2013.

I. A. Ibragimov and Yu. V. Linnik. *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing, Groningen, 1971.

E. Jacquier, N.G. Polson, and P.E. Rossi. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122(1):185–212, 2004.

T. Jiang. The asymptotic distributions of the largest entries of sample correlation matrices. *The Annals of Applied Probability*, 14(2):865–880, 05 2004.

Y. Kakizawa. Moderate deviations for quadratic forms in gaussian stationary processes. *Journal of Multivariate Analysis*, 98(5):992–1017, 2007.

D. Kondrashov, S. Kravtsov, A.W. Robertson, and M. Ghil. A hierarchy of data-based enso models. *Journal of Climate*, 18(21):4425–4444, 2005.

M.R. Kosorok and S. Ma. Marginal asymptotics for the large p, small n paradigm: with applications to microarray data. *The Annals of Statistics*, 35(4):1456–1486, 2007.

S.N. Lahiri. *Resampling Methods for Dependent Data*. Springer, 2003.

L.J. Larson-Prior, R. Oostenveld, S. Della Penna, G. Michalareas, F. Prior, A. Babajani-Feremi, J.-M. Schoffelen, L. Marzetti, F. de Pasquale, F. Di Pompeo, J. Stout, M. Woolrich, Q. Luo, R. Bucholz, P. Fries, V. Pizzella, G.L. Romani, M. Corbetta, and A.Z. Snyder. Adding dynamics to the human connectome project with {MEG}. *NeuroImage*, 80:190 – 201, 2013.

D. Li and A. Rosalsky. Some strong limit theorems for the largest entries of sample correlation matrices. *The Annals of Applied Probability*, 16(1):423–447, 2006.

D. Li, W. Liu, and A. Rosalsky. Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix. *Probability theory and related fields*, 148(1-2):5–35, 2010.

M.A. Lindquist, Y. Xu, M.B. Nebel, and B.S. Caffo. Evaluating dynamic bivariate correlations in resting-state fmri: A comparison study and a new approach. *Neuroimage*, 101: 531–546, 2014.

C. Liu, W. Gaetz, and H. Zhu. Estimation of time-varying coherence and its application in understanding brain functional connectivity. *EURASIP Journal on Advances in Signal Processing*, 2010(1), 2010.

W. Liu and W.B. Wu. Asymptotics of spectral density estimates. *Econometric Theory*, 26 (4):1218–1245, 2010.

W. Liu, Z. Lin, and Q.M. Shao. The asymptotic distribution and berrycesseen bound of a new test for independence in high dimension with an application to stochastic optimization. *The Annals of Applied Probability*, 18(6):2337–2366, 12 2008.

A.J. Majda, R.V. Abramov, and M.J. Grote. *Information theory and stochastics for multiscale nonlinear systems*, volume 25. American Mathematical Society, 2005.

S. Mallat, G. Papanicolaou, and Z. Zhang. Adaptive covariance estimation of locally stationary processes. *The Annals of Statistics*, 26(1):1–47, 02 1998.

T. Medkour, A.T. Walden, and A. Burgess. Graphical modelling for brain connectivity via partial coherence. *Journal of neuroscience methods*, 180(2):374–383, 2009.

E. Moulines, P. Priouret, and F. Roueff. On recursive estimation for time varying autoregressive processes. *The Annals of statistics*, 33(6):2610–2654, 2005.

S.V. Nagaev. Large deviations of sums of independent random variables. *The Annals of Probability*, pages 745–789, 1979.

G.P. Nason, R. von Sachs, and G. Kroisandt. Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 62(2):271–292, 2000.

W. Newey and K. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987.

H.C. Ombao and V.S. Bellegem. Evolutionary coherence of nonstationary signals. *IEEE Transactions on Signal Processing*, 56(6):2259–2266, 2008.

H.C. Ombao, J.A. Raz, R. von Sachs, and B.A. Malow. Automatic statistical analysis of bivariate nonstationary time series. *Journal of the American Statistical Association*, 96 (454):543–560, 2001.

H.C. Ombao, R. von Sachs, and W. Guo. Slex analysis of multivariate nonstationary time series. *Journal of the American Statistical Association*, 100(470):519–531, 2005.

T. Park, I.A. Eckley, and H.C. Ombao. Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes. *IEEE Transactions on Signal Processing*, 62(20):5240–5250, 2014.

I. Pinelis. Optimum bounds for the distributions of martingales in banach spaces. *The Annals of Probability*, pages 1679–1706, 1994.

D.N. Politis, J.P. Romano, and M. Wolf. *Subsampling.* Springer, 1999.

S. Portnoy. On the central limit theorem in R^p when $p \to \infty$. *Probability Theory and Related Fields*, 73(4):571–583, 1986.

R. Prado, M. West, and A.D. Krystal. Multichannel electroencephalographic analyses via dynamic regression models with time-varying lag–lead structure. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(1):95–109, 2001.

M. B. Priestley. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 27(2):204–237, 1965.

M.B. Priestley. *Spectral Analysis and Time Series.* Academic Press, 1981.

M.B. Priestley. *Spectral Analysis and Time Series.* Academic Press, London, 1982.

M.B. Priestley. *Nonlinear and Nonstationary Time Series Analysis.* Academic Press, London, 1988a.

M.B. Priestley. *Non-linear and Non-stationary Time Series Analysis.* Academic Press, 1988b.

M.B. Priestley and H. Tong. On the analysis of bivariate non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 35(2):153–166, 1973.

R.D. Reiss. Nonparametric estimation of smooth distribution functions. *Scandinavian Journal of Statistics*, pages 116–119, 1981.

E. Rio. Moment inequalities for sums of dependent random variables under projective conditions. *Journal of Theoretical Probability*, 22(1):146–163, 2009.

M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1):43, 1956.

M. Rosenblatt. *Markov Processes: Structure and Asymptotic Behavior.* Springer, 1971.

M. Rosenblatt. *Stationary Sequences and Random Fields.* Springer, 1985.

H.P. Rosenthal. On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables. *Israel Journal of Mathematics*, 8(3):273–303, 1970.

M. Rudelson and R. Vershynin. Hanson-wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18(82):1–9, 2013.

R. Salvador, J. Suckling, C. Schwarzbauer, and Ed Bullmore. Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 360(1457):937–946, 2005.

J. Sanderson, P. Fryzlewicz, and M.W. Jones. Estimating linear dependence between nonstationary time series using the locally stationary wavelet model. *Biometrika*, 97(2):435–446, 2010.

S.L. Simpson, F.D. Bowman, and P.J. Laurienti. Analyzing complex functional brain networks: fusing statistics and network science to understand the brain. *Statistics surveys*, 7:1, 2013.

T. Subba Rao. The fitting of non-stationary time-series models with time-dependent parameters. *Journal of the Royal Statistical Society. Series B (Methodological)*, 32(2):312–322, 1970.

J. Timmer, M. Lauk, S. Häußler, V. Radt, B. Köster, B. Hellwig, B. Guschlbauer, C.H. Lücking, M. Eichler, and G. Deuschl. Cross-spectral analysis of tremor time series. *international Journal of Bifurcation and chaos*, 10(11):2595–2610, 2000.

H. Tong. *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press, 1990.

R.S. Tsay. *Analysis of Financial Time Series*, volume 543. John Wiley & Sons, 2005.

M. Vogt. Nonparametric regression for locally stationary time series. *The Annals of Statistics*, 40(5):2601–2633, 10 2012.

J. Wang, F. Cheng, and L. Yang. Smooth simultaneous confidence bands for cumulative distribution functions. *Journal of Nonparametric Statistics*, 25(2):395–407, 2013.

N. Wiener. *Nonlinear Problems in Random Theory*. Wiley, New York, 1958.

C.K. Wikle and M.B. Hooten. A general science-based framework for dynamical spatio-temporal models. *Test*, 19(3):417–451, 2010.

F.T. Wright. A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *The Annals of Probability*, 1 (6):1068–1070, 1973.

W. B. Wu and X. Shao. Limit theorems for iterated random functions. *Journal of Applied Probability*, pages 425–436, 2004.

W.B. Wu. Nonlinear system theory: another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America*, 102(40):pp. 14150–14154, 2005.

W.B. Wu. Asymptotic theory for stationary processes. *Statistics and Its Interface*, 0:1–20, 2011.

W.B. Wu and Y.N. Wu. Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electronic Journal of Statistics*, 10(1):352–379, 2016.

Wei Biao Wu and Zhou Zhou. Gaussian approximations for non-stationary multiple time series. *Statistica Sinica*, pages 1397–1413, 2011.

H. Xiao and W.B. Wu. Covariance matrix estimation for stationary time series. *The Annals of Statistics*, 40(1):466–493, 2012.

H. Xiao and W.B. Wu. Asymptotic theory for maximum deviations of sample covariance matrix estimates. *Stochastic Processes and their Applications*, 123(7):2899 – 2920, 2013.

H. Xiao and W.B. Wu. Portmanteau test and simultaneous inference for serial covariances. *Statistica Sinica*, pages 577–599, 2014.

H. Yamato. Uniform convergence of an estimator of a distribution function. *Bulletin of Mathematical Statistics*, 15:69–78, 1973.

M. Zani. Large deviations for quadratic forms of locally stationary processes. *Journal of multivariate analysis*, 81(2):205–228, 2002.

X. Zhang and G. Cheng. Bootstrapping high dimensional time series. *arXiv preprint arXiv:1406.1037*, 2014.

W. Zhou. Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Transactions of the American Mathematical Society*, 359(11):5345–5363, 2007.

Z. Zhou. Nonparametric inference of quantile curves for nonstationary time series. *The Annals of Statistics*, 38(4):2187–2217, 08 2010.

Z. Zhou and W. B. Wu. Local linear quantile estimation for nonstationary time series. *The Annals of Statistics*, 37(5B):2696–2729, 10 2009.

A. Zygmund. *Trigonometric Series. Vol I, II.* Third Ed. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2002.