Danksagung

Ich möchte mich zuallererst bei Giang für die unermüdliche Unterstützung und die großen Schwierigkeiten, die sie in Kauf genommen hat, bedanken. Dann möchte ich mich bei meiner Familie, insbesondere bei meinen Eltern, bedanken, die mir immer ein Refugium geboten hat. Besonderer Dank gilt auch den Kollegen vom Institut für Statistik, in dem ich eine sehr schöne Zeit verbringen durfte. Die vielen Diskussionen, vor allem diejenigen nicht mathematischer Natur, waren immer eine willkommene Abwechslung und Bereicherung. In diesem Sinne möchte ich mich auch bei den Kollegen vom Institut für Numerik bedanken.

Ganz speziell möchte ich mich bei meinem Betreuer, Prof. Dr. István Berkes, für seine große Unterstützung und die zahllosen Gespräche bedanken, die es mir ermöglicht haben, einen Einblick in ein spannendes Gebiet der Mathematik zu erhalten. Von seinem großen Wissen und der Begeisterung, dieses mitzuteilen, habe ich enorm profitiert.

Abstract

The rapid development in many sciences, such as economics, biology, sociology, geology, physics and others, had a great impact on statistics, and vice versa. In particular, the progress in computer science and hardware technology led to a substantial increase in the amount of data being collected and analyzed. As the size of the data grows, so does often the dimensionality, and tools such as principal component analysis become inevitable for analyzing and modeling. A fundamental tool in this theory is the multivariate central limit theorem. It allows to deal with important issues such as parameter estimation and model diagnosis. For example, suppose that we have a sample of n random variables $\mathbf{X}_{(n)} = (X_1, \ldots, X_n)$ and a collection of estimators for some parameters

$$\mathbf{g}_{(d)} = \big(g_1(X_1, \ldots, X_n), ..., g_d(X_1, \ldots, X_n)\big)^T,$$

based on the sample $\mathbf{X}_{(n)}$. If the multivariate central limit theorem holds, one can use T-tests and F-tests to decide upon the redundance of various parameters and the model quality. We may summarize these procedures in the confidence ellipsoids

$$\big\{\mathbf{\Theta}_d \mid \big(\mathbf{g}_{(d)} - \mathbf{\Theta}_d\big)^T \widehat{\Gamma}^{-1} \big(\mathbf{g}_{(d)} - \mathbf{\Theta}_d\big) \leq n^{-1}\chi^2_{1-\alpha}(d)\big\}, \tag{1}$$

where $\widehat{\Gamma}$ is an estimator of the covariance matrix, and $\chi^2_{1-\alpha}(d)$ the quantile of the χ^2 distribution with d degrees of freedom, corresponding to the confidence level $1-\alpha$. Unfortunately, as the dimension d grows, the above ellipsoids become less and less informative, as everything is 'summed up'. It is therefore more convenient to use the maximum function instead of adding all the elements, i.e. to consider

$$\mathcal{V}_d = \sqrt{n} \max_{1 \leq h \leq d} \widehat{\gamma}_{h,h}^{-1} \big| g_h(X_1, ..., X_n) - \theta_h \big|,$$

where $\widehat{\gamma}_{h,h}^2$ denotes an estimator of the diagonal elements $\gamma_{h,h}^2$ of the covariance matrix Γ. Suppose that we have

$$a_d^{-1}\big(\mathcal{V}_d - b_d\big) \xrightarrow{w} G \qquad \text{as } d \to \infty \tag{2}$$

for some sequences a_d, b_d, where G is an extreme value distribution. Then we can formulate the simultaneous confidence band

$$\big\{\mathbf{\Theta}_d \mid \sqrt{n} \max_{1 \leq h \leq d} \widehat{\gamma}_{h,h}^{-1} \big| g_h(X_1, ..., X_n) - \theta_h \big| \leq a_d G_{1-\alpha} + b_d\big\}, \tag{3}$$

where $G_{1-\alpha}$ corresponds to the quantile of an extreme value distribution with confidence level $1-\alpha$. It turns out that in many cases the actual growth rate

of $a_d G_{1-\alpha} + b_d$ is of the magnitude $\mathcal{O}\left(\sqrt{\log d}\right)$, which is much smaller than $\sqrt{d}$ for large d, reflecting the growth rate of the ellipsoids in (1). Moreover, using the maximum function allows for immediate inference for subsets or single components, whereas the ellipsoids in (1) are rather inappropriate for such purpose. However, establishing weak convergence to an extreme value distribution of functionals of the above type is a very delicate problem, as the difficulties in the proofs of Erdős-Darling type limit theorems clearly indicate. The purpose of this thesis is to develop techniques allowing to establish (2) for a variety of different settings, which include covariance estimators, Yule-Walker estimators in case of autoregressive processes, and also change-point analysis in increasing dimension. The statistical gains from our program are considerable, e.g. an improved efficiency of a number of statistical procedures used previously in the literature. Our methods lead also to some new theoretical facts, among others limit theorems for random vectors with increasing dimensions and an analogue of the Berry-Esseen inequality in the context of the Cramér-Wold device, used repeatedly in our thesis.

Contents

Chapter 1

Why Extreme Inference?

1.1 Introduction

Data analysis grew out from mathematical statistics and by now it has become a fundamental tool in applied sciences. In the 1960's, John Tukey was one of the first mathematicians recognizing the great importance of its development as an independent discipline. It uses methods from various mathematical fields ranging from number theory to functional analysis, but its interplay with other disciplines such as economics, biology, chemistry, geology, physics and other sciences, has also had a tremendous impact on its development. In particular, the progress in computer science and hardware technology have completely changed the scenery. Over the last few decades, data management and data processing have become crucial factors in science and technology and substantial progress has taken place in data gathering and data processing mechanisms. Recent examples include biotech data, economical data, financial data, and imagery.

Inevitably, many of the data sets encountered in practice are multidimensional, leading to profound mathematical difficulties. Working with high dimensional data has both its advantages and disadvantages, or as Donoho [40] puts it, "the blessings and curses of high dimensionality". A cartesian grid of spacing 1/10 on the unit cube in 10 dimensions has 10^{10} points; if the cube in 20 dimensions were considered, we would have 10^{20} points. Thus optimizing a function over a continuous product domain of a few dozen variables by performing a search on some discrete space defined by a crude discretization, we will be faced with the problem of making tens of trillions of evaluations of the function. This precludes, under almost any computational scheme foreseeable today, the use of exhaustive enumeration strategies, requiring completely new techniques, such as principal component analysis or dynamic programming, developed by Richard Bellman. In order to give some more insight on this subject, I will present a brief overview of some classical examples and possible solutions offered by the literature.

1.1.1 The 'curses' of high dimension

Non-parametric estimation

Suppose we have a data set of d-dimensional variables whose first coordinate depends on the others through a model of the form

$$X_{i,1} = f(X_{i,2}, \dots, X_{i,d}) + \epsilon_i. \tag{1.1.1}$$

Suppose that f is of unknown form, i.e. we are not using a specific model for f such as, for example, a linear model. Instead, we assume merely that f is a Lipschitz function and the noise variables $\{\epsilon_i\}_{i\in\mathbb{N}}$ are i.i.d. Gaussian r.v.'s with mean 0 and variance 1. How does the accuracy of estimation depend on n, the number of observations in our data set? Denote with $\mathcal{F}$ the class of Lipschitz functions on $[0,1]^d$. A standard calculation in minimax decision theory [30] shows that for any estimator $\widehat{f}$ we have

$$\sup_{f\in\mathcal{F}} \left\|\widehat{f}(x) - f(x)\right\|_2 \geq Cn^{-1/(2+d)}, \tag{1.1.2}$$

where $\|.\|_2$ denotes the L^2 norm. This lower bound is nonasymptotic. Thus getting an estimate for f with accuracy 0.01 for large d requires trillions of samples. The very slow rate of convergence in high dimensions is the ugly consequence of dimensionality.

Model selection

Suppose that, contrary to the previous example, we model f as a linear function, i.e. we consider the linear regression problem, where there is a dependent variable $X_{i,1}$ which we want to model as a linear function of $X_{i,2}, ..., X_{i,d}$ as in (1.1.1). Standard techniques provide us with estimators $\widehat{\alpha} = \left(\widehat{\alpha}_1, ..., \widehat{\alpha}_d\right)^T$ for the regression coefficients. To investigate whether the model fits or not, one can use the F-test. If d is very large, we might have overparameterized the model, and having many irrelevant variables can easily lead to wrong conclusions and poor performance. Hence we are highly interested in reducing the dimension of the parameter vector $\widehat{\boldsymbol{\alpha}}$. For this reason, statisticians have, for a long time, considered model selection by searching among subsets of the possible explanatory variables, trying to find a few variables among the many which will adequately explain the dependent variable. The history of this approach goes back to the early 1960's when computers began to be used for data analysis and automatic variable selection became a possibility. A natural approach is via hypothesis testing; for instance, we can use the T-test to accept or reject the hypothesis $\alpha_i = 0$ for some $1 \leq i \leq d$. If d is large, we need to test simultaneously the hypothesis $\alpha_{i_1} = \alpha_{i_2} = ... = \alpha_{i_m} = 0$, for some $m \in \mathbb{N}$. By combining the T-tests, a popular approach is to consider the statistic $\big(\widehat{\alpha}_{i_1}, ..., \widehat{\alpha}_{i_m}\big)\widehat{\Sigma}\big(\widehat{\alpha}_{i_1}, ..., \widehat{\alpha}_{i_m}\big)^T$. Let $\boldsymbol{\alpha} = \big(\alpha_{i_1}, ..., \alpha_{i_m}\big)^T$ be the true parameter vector. Inference may then be based on the ellipsoids

$$\Big\{\boldsymbol{\alpha} \in \mathbb{R}^m \mid \big(\widehat{\alpha}_{i_1} - \alpha_{i_1}, ..., \widehat{\alpha}_{i_m} - \alpha_{i_m}\big)\widehat{\Gamma}^{-1}\big(\widehat{\alpha}_{i_1} - \alpha_{i_1}, ..., \widehat{\alpha}_{i_m} - \alpha_{i_m}\big)^T \leq m^{-1}\chi_q^2(m)\Big\}, \tag{1.1.3}$$

where $\widehat{\Gamma}$ denotes an estimator of the covariance matrix Γ and $\chi_q^2(m)$ is the q-quantile of the chi-square distribution with m degrees of freedom. This approach seems reasonable, but it has some drawbacks. Since one essentially sums up all the errors, it is difficult to provide inference for single elements. This is especially the case if m is large. Hence it is desirable to have a procedure which is essentially invariant of m and which allows for simple inference for the single elements.

Before investigating which, if any, parameters are redundant, it has become common practice to first establish an upper bound for d. This approach goes back to the 1960's, path breaking contributions are due to [1] and [86, 87]. The idea is to optimize over the penalized form

$$\min\left\{RSS\,(Model) + \lambda(d)\,Model\,Complexity\right\}, \tag{1.1.4}$$

where RSS denotes the sum of squares of some residuals, and the model complexity is the number of variables $\boldsymbol{\alpha} = \big(\alpha_{i_1}, ..., \alpha_{i_d}\big)^T$ used in forming the model. Early formulations used $\lambda(d) = 2\sigma^2$, where σ^2 is the assumed variance of the noise $\big\{\epsilon_i\big\}_{i\in\mathbb{N}}$ in (1.1.1). The overall idea is to impose a cost on large complex models. More recently, one sees proposals of the form $\lambda(d) = 2\sigma^2 \log d$ or $\lambda(d) = 2\sigma^2 \log\log d$. With these logarithmic penalties, one takes into account in an appropriate way the true effects of searching for variables to be included among many variables. This approach stems from information theoretic arguments, and a variety of results indicate that this form of logarithmic penalty is quite satisfactory; for a survey see [31] and [82]. That is, with this logarithmic penalty, one can mine one's data to one's taste, while controlling the risk of finding spurious structure. The logarithm increases slowly with d, a faster increase would indicate that automatic variable selection is in general hopeless: one loses too much by searching for the right variables. Having decided upon a specific choice of d, one can then use the ellipsoid in (1.1.3) to single out redundant parameters. This procedure is commonly referred to as *subset-modelling*, and is an active field of research.

Covariance Estimation

An important concept in probability theory and statistics is dependence. A lot of different dependence concepts and models have been established over the past years, such as Markov chains, martingales, mixing, etc. More recently, emphasis has been put on various projective dependence measures, see for instance [41, 129]. A fundamental concept in multivariate analysis is covariance, and more recently, also copulas. Suppose that we have data $\{X_{i,j}\}_{1\leq i\leq n, 1\leq j\leq d}$ where the vectors $X^{(i)} = (X_{i,j} \mid 1 \leq j \leq d)$ are assumed to be samples from a Gaussian distribution with mean zero and covariance Γ. We are interested in knowing whether $\Gamma = I$ as compared to $\Gamma \neq I$, where I denotes the identity. In the spirit of principal component analysis, and depending on the alternative hypothesis, it is natural to rephrase the question as $\lambda_1 = 1$ versus $\lambda_1 > 1$, where λ_1 is the top eigenvalue of the covariance matrix. It then becomes natural to develop a test based on l_1, the top eigenvalue of the empirical covariance matrix $\widehat{\Gamma}$. Thus, finding the asymptotic null distribution of l_1 is very important. Exact formulas go back to Anderson (1962), but are not very useful in practice; already for moderate d, and also if n is proportional to d, one cannot apply them since they are extremely complicated. We are interested, instead, in a simple approach assuming that d is large. Suppose we are in a setting of many observations and many variables. What is the behavior of the top eigenvalue of $\widehat{\Gamma}$? Consider the case where $d/n \to \beta$, i.e. we have large dimension and large sample size. This problem has been studied for decades; see [32] for references. Classical results in random matrix theory, e.g. the Wigner semicircle law, concern infinite matrices and give information about the bulk spectrum of $\widehat{\Gamma}$; but unfortunately they do not accurately predict the top eigenvalue. Tracy and Widom introduced substantial new ideas in the study of the top eigenvalues of certain ensembles of infinite random matrices, and finally Johnstone [72], extending and adapting this work to a statistical setting, has been able to obtain asymptotic results for the top eigenvalue of $\widehat{\Gamma}$ in the so-called null case. This, however, requires knowledge of the single parameters $\mu_j = \mathbb{E}\big(X_j^{(i)}\big)$ and $\sigma_i^2 = \mathrm{Var}\big(X_j^{(i)}\big)$, in particular one needs that $\mu_1 = ... = \mu_d$, and $\sigma_1^2 = ... = \sigma_d^2$. There seems to be evidence that even if these conditions are violated, l_1 will asymptotically follow the Tracy-Widom law. There is another promising approach due to T. Jiang [71]. Instead of considering the maximum eigenvalue l_1, he considers the maximum entry of the empirical correlation matrix $\widehat{\Sigma}$. Denote this entry with $\max\big|\widehat{\Sigma}\big|$. Then, based on methods developed by Dembo et al. [36, 37], he shows that, given appropriate moment assumptions,

$$n\big(\max\big|\widehat{\Sigma}\big|\big)^2 - 4\log n + \log\log n \xrightarrow{w} G, \tag{1.1.5}$$

where G is an extreme value distribution. Empirical studies show that the results, though derived asymptotically, are useful for small d as well. Hence one obtains, from a high-dimensional analysis, useful results in moderate dimensions.

A related problem is the following. Suppose that we are given a time series $\{X_k\}_{k\in\mathbb{N}}$, possibly non-Gaussian, and we want to test $\mathbb{E}(X_h X_0) = \phi_h = 0$ for $h \in \mathcal{I}$, where $\mathcal{I}$ is some index set with $|\mathcal{I}| = d$. Such a test is important in the inference of stochastic processes. For example, after a model is fitted to some observed data, one would like to inspect the residuals and perform model diagnostics. If the residuals do not behave like a white noise sequence, one may need to find a better model which can capture more structural information from the data. Various tests for white noise have been proposed in the literature: they include Fisher's test, generalized likelihood ratio test, and the Neyman test, see Section 7.4 in [49]. Another general test is the Ljung-Box test, proposed in [83], which has the form

$$\mathcal{Q}_{LB} = n(n+2)\sum_{k=1}^{d}\frac{\widehat{\rho}_k^2}{n-k}, \quad \text{where } \widehat{\rho}_k^2 = \frac{\widehat{\phi}_k}{\widehat{\phi}_0}, \tag{1.1.6}$$

and behaves as $\chi^2(d)$ if the sample size n is large enough. Recently, Wu [131] proposed the test statistic

$$\mathcal{Q}_W = \max_{1\le h\le d}\left|\frac{\widehat{\phi}_k}{\widehat{\phi}_0}\right|, \tag{1.1.7}$$

and established an upper tail bound for the asymptotic distribution of $\mathcal{Q}_W$ (suitably normalized) under general dependence conditions, if $d = d_n$ is of the magnitude $\mathcal{O}\left(n^{1/2}(\log n)^{-1/2}\right)$. He then used his result to demonstrate that a test based on (1.1.7) has far more power than the Ljung-Box test based on (1.1.6). Moreover, his test is practically invariant to the dimension $d = d_n$, which is another attractive feature. However, showing that $\mathcal{Q}_W$ (suitably normalized) converges weakly to an extreme value distribution is a much harder problem than it was in the previous case of the correlation matrix $\widehat{\Sigma}$. Due to the more complicated dependence structure, arguments based on [36, 37] cannot be used, and another approach needs to be developed.

1.1.2 The 'blessings' of high dimension and the main question

In the previous examples we have seen that using the maximum function as a statistic yields very good results. A possible explanation is the *concentration of* *measure* phenomenon, a terminology introduced by V. Milman for a remarkable fact about probabilities on product spaces in high dimensions. Suppose we have a Lipschitz function f on the d-dimensional sphere. Place a uniform measure P on the sphere, and let X be a random variable with law P. Then

$$P\big(|f(X) - \mathbb{E}(f(X))|_d > t\big) \leq C_1 \exp(-C_2\, t^2), \tag{1.1.8}$$

where $|.|_d$ denotes the usual Euclidian norm, and C_1, C_2 are constants independent of f and of the dimension. In short, a Lipschitz function is nearly constant for large d. But even more importantly: the tails behave at worst like a scalar Gaussian random variable with controlled mean and variance. Variants of this phenomenon are known for many high-dimensional situations; e.g. discrete hypercubes $\mathbb{Z}_2^d$ with the Hamming distance. The roots of these phenomena are quite old: they go back to the isoperimetric problem of classical geometry. Milman credits Paul Lévy with the first modern general recognition of the phenomenon. There is a vast literature on this problem and we find that using the maximal function fits well into this framework. Suppose we take the maximum of d i.i.d. Gaussian random variables $X_1, ..., X_d$. As the maximum is a Lipschitz functional, we know from the concentration of measure principle that the distribution of the maximum behaves not worse than a standard normal distribution in the tails. Using other arguments, one can show that

$$\limsup_{d\to\infty}\ (2\log d)^{1/2}\frac{\max(X_1, ..., X_d) - (2\log d)^{1/2}}{\log\log d} = \frac{1}{2}, \tag{1.1.9}$$

hence the chance that the maximum exceeds $\sqrt{2\log d} + t$ decays very rapidly in t. These properties indicate that the statistic $\max(X_1, ..., X_d)$ is a promising candidate when one is dealing with higher dimensions. To highlight the advantages, let us, for comparison, consider the well-known $\chi^2(d)$ based test statistics, and the corresponding ellipsoids. Suppose that we have a sample of n random variables $\mathbf{X}_{(n)} = (X_1, \ldots, X_n)$, and we have a collection of estimators

$$\mathbf{g}_{(d)} = \big(g_1(X_1, \ldots, X_n), ..., g_d(X_1, ..., X_n)\big)^T, \tag{1.1.10}$$

based on the sample $\mathbf{X}_{(n)}$. Suppose that a multivariate CLT is valid, i.e.

$$\sqrt{n}\,\big(\mathbf{g}_{(d)} - \mathbb{E}\big(\mathbf{g}_{(d)}\big)\big) \xrightarrow{w} \mathcal{N}\big(0, \Gamma\big), \tag{1.1.11}$$

where Γ is some covariance matrix. Let $\mathbb{E}\big(\mathbf{g}_{(d)}\big)^T = \mathbf{\Theta}_d = \big(\theta_1, ..., \theta_d\big)$. Then, in spirit of the $\chi^2(d)$ based tests mentioned earlier, we can formulate the confidence ellipsoids

$$\big\{\mathbf{\Theta}_d \mid \big(\mathbf{g}_{(d)} - \mathbf{\Theta}_d\big)^T\,\widehat{\Gamma}^{-1}\,\big(\mathbf{g}_{(d)} - \mathbf{\Theta}_d\big) \leq n^{-1}\chi^2_{1-\alpha}(d)\big\}, \tag{1.1.12}$$

where $\widehat{\Gamma}$ is an estimator of the covariance matrix, and $\chi^2_{1-\alpha}(d)$ the quantile of the χ^2 distribution with d degrees of freedom, corresponding to the confidence level $1-\alpha$. In contrast, consider now the statistic

$$\mathcal{V}_d = \sqrt{n} \max_{1 \leq h \leq d} \widehat{\gamma}_{h,h}^{-1} \big| g_h(X_1, ..., X_n) - \theta_h \big|, \tag{1.1.13}$$

where $\widehat{\gamma}^2_{h,h}$ denotes an estimator of the diagonal elements $\gamma^2_{h,h}$ of the covariance matrix Γ. Suppose that we have

$$a_d^{-1} \big(\mathcal{V}_d - b_d\big) \xrightarrow{w} G, \tag{1.1.14}$$

for some sequences a_d, b_d, where G is an extreme value distribution. Then we can formulate the simultaneous confidence band

$$\big\{ \mathbf{\Theta}_d \mid \sqrt{n} \max_{1 \leq h \leq d} \widehat{\gamma}_{h,h}^{-1} \big| g_h(X_1, ..., X_n) - \theta_h \big| \leq a_d G_{1-\alpha} + b_d \big\}, \tag{1.1.15}$$

where $G_{1-\alpha}$ corresponds to the quantile of an extreme value distribution with confidence level $1-\alpha$. Since we assumed that the (multivariate) CLT is valid, one can expect (and indeed this is the case), that $a_d G_{1-\alpha} + b_d$ is of the magnitude $\mathcal{O}\big(\sqrt{2 \log d}\big)$. Compared to the ellipsoid given in (1.1.12), the above confidence band now has the following advantages:

(i) it is essentially invariant to the dimension d if d is large enough,

(ii) yields a much smaller confidence region since $2 \log d \ll d$,

(iii) only requires to estimate the variances $\gamma^2_{h,h}$,

(iv) provides immediate inference for the single estimators $g_h(X_1, ..., X_n)$.

It is therefore natural to pose the following question:

When does (1.1.14)*, i.e. weak convergence to an extreme value distribution hold?*

The answer to this question depends on the relation between the growth rate of the dimension $d = d_n$ and the sample size n. Note that in the extreme case of independence we can choose $d = d_n = n$. It is, however, impossible to reach this rate in general. On the other hand, it is not hard to show that if the multivariate CLT holds for all fixed d, then there exists a sequence $d_n \to \infty$ such that (1.1.14) is valid. However, a logarithmic growth rate of d would be undesirable, since it provides no practical use. Hence, we should reformulate the above question to

> *Given a dependence condition, what is the highest growth rate of $d = d_n$ such that* (1.1.14) *holds?*

The aim of this thesis is to provide partial answers to this question in several different settings. The thesis is structured as follows. In Chapter 2, we outline some technical difficulties, and which tools one may use to circumvent them. Motivated by covariance estimators, a very general approach is then presented in Chapter 3, which allows for a growth rate of essentially $d_n = \mathcal{o}(\log n)$, if the sample size is n. In chapter 4, we show that by strengthening some of the structural assumptions, one may obtain a growth rate of $\mathcal{O}\left(n^{1/6}(\log n)^{-\alpha/3}\right)$, $\alpha > 3$. Chapter 5 addresses the problem of establishing simultaneous confidence bands for the Yule-Walker estimators in case of autoregressive processes, which also leads to new estimators for the possible order of such processes. Finally, chapter 6 deals with change-point analysis if the dimension increases, and the problem of detecting global and local changes.

Chapter 2

Tools

As outlined in the previous chapter, our ultimate goal is to establish weak convergence of extremal statistics such that the growth rate of the dimension is as large as possible. There is a wide literature on extreme values of dependent random variables which gives precise conditions when and how fast the maximum of a dependent sequence converges weakly to an extreme value distribution. In particular, the well known conditions $D(u_n)$ and $D^*(u_n)$ of Leadbetter [79] (see also [80, 84]) provide a satisfactory solution. However, verifying $D(u_n)$ and $D^*(u_n)$ in the situations investigated in our dissertation (such as extremal statistics involving covariances) is a very difficult technical problem. If u_n is an increasing sequence tending to infinity and we set $Y_{h,n} = \sqrt{n}\big(g_h(X_1, ..., X_n) - \theta_h\big)$, $1 \leq h \leq d_n$, where $g_h(X_1, ..., X_n)$ is as in (1.1.10), condition $D(u_n)$ can be formulated as

Condition $D(u_n)$**.** *There exists a sequence* $\alpha_{n,l}$*, with* $\lim_{n\to\infty} \lim_{l\to\infty} \alpha_{n,l} = 0$ *such that for any* $1 \leq i \leq j \leq d_n$ *with* $j - i \geq l$ *we have*

$$\left| P\big(\max_{1\leq h\leq d_n} Y_{h,n} \leq u_n\big) - P\big(\max_{1\leq h\leq i} Y_{h,n} \leq u_n\big) P\big(\max_{j\leq h\leq d_n} Y_{h,n} \leq u_n\big) \right| \leq \alpha_{n,l}.$$

Typically, the sequence u_n is chosen to satisfy the condition

$$P\big(Y_{h,n} > u_n\big) = \frac{z}{d_n} + o\left(d_n^{-1}\right), \tag{2.0.1}$$

since then one can show that

$$\exp(-z) \leq \liminf_{n\to\infty} P\big(\max_{1\leq h\leq d_n} Y_{h,n} \leq u_n\big).$$

We will discuss this point in detail later. Now let us take a closer look at condition $D(u_n)$. If we assume that $\left\{Y_{h,n}\right\}_{1\leq h\leq d_n}$ is an array of Gaussian random variables, then it is possible (cf. [17, 38]) to verify condition $D(u_n)$ if the array satisfies one of the conditions

(a) $\sum_{h=1}^{\infty} r_h^2 < \infty$,

(b) For some $\beta > 0$ we have $r_h(\log h)^{2+\beta} \to 0$,

where $r_h := \limsup_{n\to\infty} \sup_{|i-j|\geq h} \big|\mathbb{E}\big(Y_{i,n}Y_{j,n}\big)\big|$, and we require in addition $r_1 < 1$. Unfortunately, however, the array $\big\{Y_{h,n}\big\}_{1\leq h\leq d_n}$ is not Gaussian in our case. If the multivariate CLT holds, we are close to a Gaussian distribution, but we need to quantify this closeness by either a distributional or a.s. remainder term. Hence the following approach seems reasonable:

(1) Show that $\big\{Y_{h,n}\big\}_{1\leq h\leq d_n}$ is 'close enough' to a Gaussian vector,

(2) use this to verify that the conditions $D(u_n)$ and $D^*(u_n)$ are valid.

Note that point (2) actually amounts to verifying the conditions (a), (b) given above, hence the major task is to deal with (1). The issue of Gaussian approximation has been studied for decades in the literature, and is still a very active field of research. Depending on the actual need, one may essentially distinguish between *strong invariance principles*, and *Berry-Esséen* type results; a small overview is given in the next sections.

2.1 Berry-Esséen type results

The first results concerning the remainder term in the CLT were obtained by A.M. Lyapunov in his works of 1900-1901. His investigations inspired many scientists to begin the analysis of approximation errors in limit theorems. A famous contribution is the Berry-Esséen theorem proved in the early 1940's. Since then, many authors have studied this problem in a variety of different situations. Roughly speaking, one can classify them according to the following three categories:

(1) the probability metric used,

(2) the dependence condition of the involved random variables,

(3) the space of the random variables.

In this section, we will restrict ourselves to the space $\mathbb{R}^d$; for results in more general spaces, see for instance [77, 107, 110, 139] and the references there. The classical Berry-Esséen theorem says that, given an i.i.d. sequence of zero mean random variables $X_1, ..., X_n$, we have for some $C > 0$

$$\sup_{x\in\mathbb{R}} \big| P\big(n^{-1/2}\sigma^{-1}(X_1 + ... + X_n\big) \le x\big) - \Phi(x)\big| \le \frac{C\,\mathbb{E}|X_1|^3}{\sqrt{n}}, \tag{2.1.1}$$

where $\sigma^2 = \mathbb{E}\big(X_1^2\big)$, and $\Phi(x)$ denotes the standard normal distribution function. The function that measures the discrepancy between the two probability measures above, is in fact, a metric. Let P, Q be two probability measures on $\mathbb{R}^d$, and denote with P_n the measure induced by the normalized sum $S_n = n^{-1/2}\sum_{k=1}^n X_k$. Then the metric $d_{\mathcal{U}}(P, Q)$ defined by

$$d_{\mathcal{U}}(P, Q) = \sup_{x\in\mathbb{R}^d} \big| P(x) - Q(x) \big|, \tag{2.1.2}$$

is often referred to as the *Kolmogorov* or *uniform* metric, and we may rewrite (2.1.1) as

$$d_{\mathcal{U}}(P_n, \Phi) \le C\frac{C\,\mathbb{E}|X_1|^3}{\sqrt{n}}. \tag{2.1.3}$$

Several other metrics have also been proposed, e.g., the *mean metric*

$$\zeta_1(P, Q) = \int |P(x) - Q(x)|dx, \tag{2.1.4}$$

or the distance of total variation

$$d_{\mathcal{V}}(P, Q) = \sup\big\{|P(A) - Q(A)| \,:\, A \in \mathfrak{B}\big\}, \tag{2.1.5}$$

where $\mathfrak{B}$ denotes the class of Borel sets. The concept of considering a metric on a particular type of sets has proved to be very useful; frequently used classes are $\mathfrak{C}$, the set of all convex sets, and $\mathfrak{S}$, the set of all spheres. One of the most convenient metrics in general probability theory is the Lévy-Prokhorov metric defined as

$$\pi(P, Q, \mathfrak{B}) = \inf\big\{\epsilon \,:\, Q(A) \le P(A^\epsilon) + \epsilon,\, P(A) \le Q(A^\epsilon) + \epsilon \text{ for all } A \in \mathfrak{B}\big\}, \tag{2.1.6}$$

where $A^\epsilon = \{x \mid d(x, A) < \epsilon\}$ and $d(.,.)$ it is some metric on $\mathbb{R}^d$. This metric has remarkable properties, and in some sense, is the weakest possible (cf. [110, Section 2.1]).

2.1.1 I.I.D. random variables

In the one dimensional case, it is known that the Berry-Esséen estimate in (2.1.1) is optimal, namely there exist measures P_n such that

$$\lim_{n\to\infty} \sqrt{n}\, d_{\mathcal{U}}(P_n, \Phi) = C > 0.$$

Ibragimov [67] and Ibragimov and Linnik [68] showed that the finiteness of the moment $\mathbb{E}|X_1|^{2+\delta}$, $0 < \delta < 1$, guarantees the decrease rate $n^{-\delta/2}$ for $d_{\mathcal{U}}(P_n, \Phi)$ as n tends to infinity, and this rate is unimprovable. For $\delta \geq 1$, we have $d_{\mathcal{U}}(P_n, \Phi) = \mathcal{O}\left(n^{-1/2}\right)$, and it is already mentioned above that this estimate is sharp in the general case. The convergence rate in the CLT has been studied for practically all known metrics; for most of them the analogue of the Berry-Esséen estimate holds, i.e. the convergence rate is $\mathcal{O}\left(n^{-1/2}\right)$. A notable exception is the total variation metric, see e.g. [110].

In theory, the distribution of the sum $S_n = n^{-1/2}\sum_{k=1}^{n} X_k$ can be expressed by convolutions, but evaluating these directly is generally impossible. A natural tool to handle them is Fourier analysis, and in one dimension the use of characteristic functions is a very effective method. However, in higher dimensions characteristic functions no longer lead to optimal results, for a discussion see [110, Section 3.6]. Thus finding a different approach is an important question, which was exhaustively answered by Zolotarev [139], who introduced the s-metric ζ_s, which is an example of the so-called *ideal metrics*. Ideal metrics have the properties of *semiadditivity* and *homogeneity of order s* (cf. [110]), which makes deductions of convergence rate estimates very simple. Let $s > 0$. Then we can represent s as $s = m + \alpha$, where $[s] = m$ denotes the integer part, and $0 \leq \alpha < 1$. Let $\mathfrak{F}_s$ be the class of all real-valued functions f such that the m-th derivative exists, is bounded and satisfies

$$\left|f^{(m)}(x) - f^{(m)}(y)\right| \leq \left|x - y\right|^{\alpha}. \tag{2.1.7}$$

The metric ζ_s for two probability measures P, Q is then defined as

$$\zeta_s(P, Q) = \sup\left\{\left|\int f(x)(P-Q)(dx)\right| : f \in \mathfrak{F}_s\right\}.$$

The s-metrics are not very strong metrics, they are, however, stronger than the Lévy-Prokhorov metric; in fact, the following estimates (cf. [110, Theoerem 6.4.2]) are valid.

Theorem 2.1.1. *Let P, Q be probability measures on $\mathbb{R}^d$. Then the inequalities*

$$\pi(P, Q, \mathfrak{C}) \leq c\, d^{1/8} \zeta_s(P, Q)^{1/4},$$
$$\pi(P, Q, \mathfrak{B}) \leq c\, d^{1/4} \zeta_s(P, Q)^{1/4},$$

are valid, where c is an absolute constant.

These inequalities, together with the properties of the s-metric, provide an essential tool, and will in fact give the largest growth rate of the dimension d_n in (1.1.14).

2.1.2 Dependent random variables

In case of dependent random variables, the techniques presented in the previous section are no longer directly applicable. However, by various blocking and truncation arguments, one may construct an approximation with independent random variables, and then use the results from the previous section. There exist, however, completely different methods, which rely heavily on the assumed dependence condition. The most popular dependence structure is the class of martingales. The problem of embedding a martingale $\{M_t\}_{t\geq 0}$ into Brownian motion goes back to Skorohod, and has been considered by many authors. For more details, we refer to the next section. This embedding procedure is a very powerful tool, and leads to the theorem below, proved by Brown and Heyde in [27] (see also [58]). To formulate the result, let $\{X_n, \mathcal{F}_n, n = 0, 1, 2...\}$, be a martingale with $X_0 = 0$ a.s., $X_n = \sum_{i=1}^n Y_i$, $n \geq 1$, and $\mathcal{F}_n$ the σ-field generated by $X_0, X_1, ..., X_n$. Put

$$\sigma_n^2 = \mathbb{E}\big(Y_n^2 \mid \mathcal{F}_{n-1}\big), \quad s_n^2 = \sum_{k=1} \mathbb{E}\big(\sigma_k^2\big), \tag{2.1.8}$$

and suppose that there is a constant $0 < \delta \leq 1$ such that $\mathbb{E}\big|Y_n\big|^{2+2\delta} < \infty$, $n = 1, 2, \ldots$.

Theorem 2.1.2. *There exist positive constants K_1, K_2, depending only on δ, such that*

$$\begin{aligned} &\sup_{x\in\mathbb{R}}\big|P\big(X_n \leq s_n x\big) - \Phi\big(x\big)\big| \\ &\leq K_1 \left(s_n^{-2-2\delta} \sum_{k=1}^n \mathbb{E}\big|Y_k\big|^{2\delta} + \mathbb{E}\big|\big(\sum_{k=1}^n \sigma_k^2\big) - \sigma_n^2\big|^{1+\delta} \right)^{1/(3+2\delta)} && (2.1.9) \\ &\leq K_2 \left(s_n^{-2-2\delta} \sum_{k=1}^n \mathbb{E}\big|Y_k\big|^{2\delta} + \mathbb{E}\big|\big(\sum_{k=1}^n Y_k^2\big) - \sigma_n^2\big|^{1+\delta} \right)^{1/(3+2\delta)}, && (2.1.10) \end{aligned}$$

where $\Phi(x)$ denotes the distribution function of the standard normal distribution.

This result has been generalized and refined over the years, see for instance [28, 70] and the references there. Many classes of sequences of r.v.'s can be approximated by martingales, leading to more general results. Another powerful method is *Stein's method*, originally developed to deal with the CLT. In connection with Mallivian calculus, this enables one to obtain Berry-Esseen type results and applies even for noncentral limit theorems when the underlying processes have long memory (cf. [23, 94, 95]).

2.2 Strong invariance principles

In many applications, a normal approximation presented as in the previous section is not sufficient. Instead of quantifying the difference of the measures via various metrics, one is rather interested in constructing normal approximations on the same probability space. The method and the quality of such approximations depend heavily on the dependence structure of the underlying random variables $X_1, X_2, \ldots$. The most frequently studied cases are where the X_k are either independent, or martingale differences. All other cases can be reduced, in some way or other, to one of the latter problems.

2.2.1 Martingales

Many important stochastic processes, when suitably scaled, can be embedded into a Brownian motion $\{W_t\}_{t\geq 0}$. Such an embedding was first introduced by Skorohod [117], and was later extended to martingales and semimartingales, (cf. [43, 44, 90, 118]). For the question which probability measures can be embedded into a Brownian motion, we refer to [91] and in particular to [96], which gives a complete survey on this topic.

Theorem 2.2.1. *Let $\{S_n = \sum_{k=1}^n X_k,\, \mathcal{F}_n, n \geq 1\}$ be a zero mean, square integrable martingale. Then there exists a probability space supporting a standard Brownian motion $\{W_t\}_{t\geq 0}$ and a sequence of nonnegative variables $\tau_1, \tau_2 \ldots$ with the following properties. If $T_n = \sum_{h=1}^n \tau_h$, $S'_n = W(T_n)$, $X'_1 = S'_1$, $X'_n = S'_n - S'_{n-1}$, for $n \geq 2$, and $\mathcal{G}_n$ is the σ-field generated by $S'_1, \ldots, S'_n$ and W_t for $0 \leq t \leq T_n$, then*

(i) $\{S_n, n \geq 1\} \stackrel{d}{=} \{S'_n, n \geq 1\}$,

(ii) T_n *is* $\mathcal{G}_n$*-measurable,*

(iii) For each real number $r \geq 1$,

$$\mathbb{E}\big(\tau_n^r \mid \mathcal{G}_{n-1}\big) \leq C_r \mathbb{E}\big(|X'_n|^{2r} \mid \mathcal{G}_{n-1}\big) = C_r \mathbb{E}\big(|X'_n|^{2r} \mid X'_{n-1}, ..., X'_1\big) \quad a.s.$$

where $C_r = 2(8/\pi^2)^{r-1}\Gamma(r+1)$, *and*

(v) $\mathbb{E}\big(\tau_n \mid \mathcal{G}_{n-1}\big) = \mathbb{E}\big({X'_n}^2 \mid \mathcal{G}_{n-1}\big)$.

The above theorem is an essential tool in deriving both CLT's, strong invariance principles and the LIL for various processes that can be approximated by martingales. For example, consider a stationary, ergodic sequence $\{X_k\}_{k\in\mathbb{N}}$. The usual steps are

(i) Decompose $S_n = \sum_{k=1}^n X_k$ into $M_n + A_n$, where M_n is a martingale with respect to some filtration $\mathcal{F}_n$.

(ii) Show that $\max_{1\le k\le n} |A_k| = o\left(n^{1/2}\right)$ a.s.

(iii) Prove that the increments $\Delta M_k = M_k - M_{k-1}$ satisfy

$$\sum_{k=1}^{n} (\Delta M_k)^2 = n\sigma^2 + o(n) \quad \text{a.s.} \quad \text{for some } \sigma^2 \ge 0.$$

(iv) Use results about the increments of a Brownian motion (cf.) to deduce that

$$S_n - \sigma W_n = o\left(n^{1/2}\right) \quad \text{a.s.} \tag{2.2.1}$$

Much of the research in this area has naturally been devoted to the question under what conditions this approach is possible, and which approximation rates can be obtained. Note that point (iii) implies that the best possible approximation rate in (2.2.1) is $\mathcal{O}\left(n^{1/4}\right)$, since we cannot obtain a better convergence rate than $\mathcal{O}\left(n^{1/2}\right)$ in (iii). Very recent and optimal results which address this question are due Wu [130], Wu and Woodroofe and . One of the main advantages of the embedding method lies in its generality, but there are also some drawbacks. One of those is the convergence rate, which is not sufficient in many cases, another one is that, in general, this method breaks down in higher dimensions. It is possible to embed a d-dimensional process in a d-dimensional Brownian motion, see e.g. the excellent survey of Obloj [96] and the references therein, but the stopping times are no longer as tractable as in Theorem 2.2.1.

2.2.2 I.I.D. approximations

Let $X_1, X_2, ...$ be independent and identically distributed random variables with zero mean and unit variance, and define $S_n = \sum_{k=1}^n X_k$. By the strong law of large numbers we have $n^{-1}S_n \to 0$ almost surely as $n \to \infty$. Rates of convergence in this result are provided by the CLT and the LIL. For a long time, the Skorohod embedding scheme described in the previous section provided the best known Wiener approximation for the partial sums S_n and the question was posed whether this constitutes the best possible approximation or not. This question was answered in the negative by Csörgő and Révész in [30], who introduced a completely different method to approximate partial sums of i.i.d. random variables and used this to show that under some (rather restrictive) conditions there is a version of S_n and a Brownian motion $\{W_t\}_{t\geq 0}$ such that almost surely

$$\left|S_n - W_n\right| = o\left(n^{1/6+\epsilon}\right), \quad \epsilon > 0.$$

Their key idea, namely constructing an approximating sequence via the quantile transformation, was substantially refined in the celebrated Hungarian construction of Komlós, Major and Tusnády [74, 75, 76], who proved the following.

Theorem 2.2.2. *Let $X, X_1, X_2, \ldots$ be i.i.d. random variables with mean zero, variance 1 and satisfying $\mathbb{E}\big(\exp(tX)\big) < \infty$ for some $|t| \leq t_0$. Then the sequence $(X_k)_{k\geq 1}$ can be redefined on a suitable probability space together with a sequence $(Y_k)_{k\geq 1}$ of i.i.d. standard Gaussian random variables such that putting $S_n = \sum_{k=1}^n X_k$, $T_n = \sum_{k=1}^n Y_k$ we have for all $x > 0$ and every n*

$$P\left(\max_{k\leq n}\left|S_k - T_k\right| > C\log n + x\right) < K\exp(-\lambda x),$$

where C, K, λ depend only on F, and λ can be taken as large as desired by choosing C large enough. Consequently, $|S_n - T_n| = \mathcal{O}\left(\log n\right)$ a.s.

If only the first $p > 2$ moments of X exist, the analogous result to the above is

Theorem 2.2.3. *Suppose that $\mathbb{E}\left|X\right|^p < \infty$, $p > 2$. Then for an appropriate construction we have*

$$\left|S_n - T_n\right| = o\left(n^{1/p}\right) \qquad a.s.$$

As it was also shown in [76], the above convergence rates are optimal. Sahanenko (1982) extended the Komlós-Major-Tusnády theorems for independent, not necessarily identically distributed random variables. A different, but very general approach was proposed by Berkes and Philipp [16], which provides approximation rates for possibly dependent random variables in $\mathbb{R}^d$ or more generally, in Hilbert spaces. This approach has been extended by various authors to a variety of different dependence conditions, see [31] for a general overview on these results. A novelty of this approach was that it allows approximation for sequences where $X_n \in \mathbb{R}^{d_n}$ with $d_n \to \infty$. Unfortunately, approximation rates like in Theorems 2.2.2 and 2.2.3 could not be obtained with this approach. It was therefore natural to ask if it is possible to extend the results of the Hungarian construction presented earlier to higher dimensions, i.e. for $d \geq 2$. Affirmative answers were given by Einmahl [46], Einmahl and Mason [47], and Zaitsev [137, 138]. Recently, Berthet and Mason [18] (see also Einmahl and Mason [47]) pointed out that the Strassen–Dudley theorem (see Theorem 11.6.2 in [45]) in combination with a special case of Theorem 1.1 and Example 1.2 of Zaitsev [136] yields the following coupling inequality.

Lemma 2.2.4. *Let $X_1, ..., X_n$ be independent, mean zero random vectors in $\mathbb{R}^d$, $d \geq 1$, such that for some $B > 0$, $|X_i| \leq B$, $i = 1, ..., n$, where $|.|$ denotes the usual Euclidian norm in $\mathbb{R}^d$. If the probability space is rich enough, then for each $\delta > 0$, one can define independent normally distributed mean zero random vectors $\xi_1, \ldots, \xi_n$ with ξ_i and X_i having the same covariance matrix for $i = 1, \ldots, n$, such that for universal constants $C_1 > 0$ and $C_2 > 0$,*

$$P\left\{ \Big|\sum_{i=1}^{N} (X_i - \xi_i)\Big| > \delta \right\} \leq C_1 d^2 \exp\left(-\frac{C_2 \delta}{B d^2} \right).$$

This coupling inequality will be an essential tool in proving some of the results presented in Chapters , and . Note that the dimension d is contained explicitly in the bound, hence we will be able to give precise conditions on the growth rate of $d = d_n$. A drawback of Lemma 2.2.4 is that it is only valid for a sequence of bounded random variables, thus various truncation and blocking arguments are needed to make this result applicable in more general cases.

Chapter 3

Extremes of Covariances and a General Device

3.1 Introduction

Let $\{X_k\}_{k\in\mathbb{Z}}$ be a stationary process with mean zero and finite variances and let $\phi_h = \mathbb{E}\big(X_k X_{k+h}\big)$, $k, h \in \mathbb{Z}$ be the covariance function. A wide range of statistical techniques, including regression analysis, principal component analysis, cluster analysis, linear and quadratic discriminant analysis, require the estimation of covariances. A natural estimate is the sample covariance $\widehat{\phi}_{n,h} = \frac{1}{n}\sum_{i=h+1}^{n} X_i X_{i-h}$. Depending on the magnitude of h, a different normalization, such as $(n-h)^{-1}$ is often convenient. Studying the asymptotic properties of $\widehat{\phi}_{n,h}$ is very important for applications, and has been extensively discussed in the literature, see for instance [4, 25, 58, 60, 61, 124, 131] and the references therein. We formulate here some important contributions in more detail. Let $X_k = \sum_{i=0}^{\infty} \alpha_i \epsilon_{k-i}$, $k \in \mathbb{Z}$ be a linear process, where $\{\epsilon_k\}_{k\in\mathbb{Z}}$ is an IID sequence of real-valued random variables with $\mathbb{E}\big(\epsilon_1\big) = 0$, $\mathbb{E}\big(\epsilon_1^2\big) < \infty$, and $\alpha_i \in \mathbb{R}$. In Brockwell and Davis [25, Section 7.2] it is shown that for bounded lags $0 \le h \le d$, a multidimensional CLT holds under the short memory condition $\sum_{i=0}^{\infty} |\alpha_i| < \infty$. More precisely, for any $d \in \mathbb{N}$ we have

$$n^{1/2}\Big\{\big(\widehat{\phi}_{n,0}, \widehat{\phi}_{n,1}, ..., \widehat{\phi}_{n,d}\big)^{\mathbf{T}} - \big(\phi_0, \phi_1, ..., \phi_d\big)^{\mathbf{T}}\Big\} \xrightarrow{d} \big\{\xi_h\big\}_{0\le h\le d}, \tag{3.1.1}$$

where $\{\xi_h\}_{h\in\mathbb{N}}$ is a mean zero Gaussian process, whose covariance structure can be explicitly expressed in terms of the coefficients $\{\alpha_i\}_{i\in\mathbb{N}}$ by Bartlett's formula, see for instance [4, 25, 60]. The case of unbounded lags h_n was discussed in Keenan [73] for stationary processes satisfying a strong mixing assumption, and it was proved that under $h_n \to \infty$ and $h_n = \mathcal{o}\big(\log n\big)$ we have

$$n^{1/2}\left\{\big(\widehat{\phi}_{n,h_n}, \widehat{\phi}_{n,h_n+1}, ..., \widehat{\phi}_{n,h_n+d}\big)^{\mathbf{T}} - \big(\phi_{h_n}, \phi_{h_n+1}, ..., \phi_{h_n+d}\big)^{\mathbf{T}}\right\} \xrightarrow{d} \big\{G_h\big\}_{0 \leq h \leq d}, \tag{3.1.2}$$

where $\big\{G_h\big\}_{h\in\mathbb{N}}$ is the stationary Gaussian linear process defined by

$$G_h = \sum_{i\in\mathbb{Z}} \phi_i \eta_{h-i},$$

where $\{\eta_i, i \in \mathbb{Z}\}$ is an IID sequence of mean zero Gaussian random variables, see for instance Wu [131]. Note that the covariance structure of the Gaussian limit processes in (3.1.1) and (3.1.2) are different. It seems that so far this linear process $\big\{G_h\big\}_{h\in\mathbb{N}}$ has not been explicitly studied in the literature. However, it is worth mentioning that in the case of a long memory process $\big\{X_k\big\}_{k\in\mathbb{Z}}$, a variety of different limit processes can be obtained, see for instance [66, 132]. The condition $h_n = \mathcal{o}\big(\log n\big)$ was later weakened for linear processes by Harris et al. [64], and, quite recently, by Wu [131], whose results we will discuss now in more detail. For bounded lags, Wu established a similar result as in (3.1.1) under a very general dependence condition that includes linear processes, but also many nonlinear processes. In case of unbounded lags, he showed that the condition $h_n = \mathcal{o}\big(\log n\big)$ can be relaxed to $h_n = \mathcal{o}\big(n\big)$. In case of $d = 1$, this can again be weakened to $h_n \to \infty$. In this context, Wu also raised the issue of simultaneous confidence bands, and proposed to study the asymptotic behavior, as $n \to \infty$, of the object

$$\max_{0\leq h\leq d_n} \big|\widehat{\phi}_{n,h} - \phi_h\big|. \tag{3.1.3}$$

He noted that obtaining the asymptotic distribution of (3.1.3) can be used to construct confidence intervals for $\{\phi_k,\ k \geq 0\}$, which in turn can be used to test the hypothesis of white noises $\phi_1 = \phi_2 = ... = 0$. Wu established an asymptotic upper distributional bound, and conjectured ([131, Conjecture 1]) that

Conjecture 3.1.1.

$$P\left(a_n^{-1}\big(\max_{0\leq h\leq d_n} \sigma^{-1/2} n^{1/2}\big|\widehat{\phi}_{n,h} - \phi_h\big| - b_n\big) \leq z\right) \to \exp(-2e^{-z}), \tag{3.1.4}$$

provided $d_n \to \infty$, $d_n = \mathcal{o}\left(n^{1/2}(\log n)^{-2}\right)$, where

$$a_n = (2\log d_n)^{-1/2}, \quad b_n = (2\log d_n)^{1/2} - 1/2(2\log d_n)^{-1/2}(\log\log d_n + \log 4\pi),$$

and $\sigma = \sum_{k\in\mathbb{Z}} \phi_k^2$.

The aim of this chapter is to verify this conjecture for linear processes under the condition $d_n = \mathcal{O}(\log n / \log\log n)$. Due to (3.1.1) one can expect that the limiting behavior of (3.1.3) is the same as if one replaces $n^{1/2}(\widehat{\phi}_{n,h} - \phi_h)$ by its limiting distribution $\{\xi_h\}$ and one can try to make this heuristics precise by an almost sure invariance principle. Many strong invariance techniques rely on martingale approximation and Skorohod embedding (see e.g. Hall and Heyde [58]), but this method already breaks down for dimension two. Other approaches are based on approximations with IID sequences, see for instance [16, 75, 109], in which case strong approximation procedures with increasing dimension are possible. However, applying these results in this particular situation (which requires some careful truncation arguments, see e.g. [14]) is not easy and leads to more restrictive conditions than martingale approximation. We will circumvent the difficulty of increasing dimension by using a general estimate for the rate of convergence of the Cramér-Wold device, which puts us back to dimension one.

This chapter is structured as follows. In Section 3.2 the main results are presented. This includes a partial verification of Conjecture (3.1.1), but also a Berry-Esséen type result (Theorem 3.2.4) for the sample covariances $\left\{\widehat{\phi}_{n,h}\right\}_{1\leq h\leq d_n}$. In Section 3.3, the main tool for the proofs is presented, which can be described as a quantitative Cramér-Wold device, and is formulated in Theorem 3.3.1. This result essentially allows to replace the normalized sample covariances $\left\{n^{1/2}\left(\widehat{\phi}_{n,h} - \phi_h\right)\right\}_{1\leq h\leq d_n}$ in (3.1.4) with their limiting processes $\{\xi_h\}_{0\leq h\leq d_n}$. Section 3.4 is devoted to establishing a Gaussian approximation result that is very useful for verifying the assumptions made in Theorem 3.3.1. Finally, the remaining proofs are given in Section 3.5.

3.2 Main results

Before discussing bounds for the magnitude of the dimension d_n such that (3.1.4) holds, we address the question under what conditions such a sequence $d_n \to \infty$ exists. To this end, we introduce the quantity

$$\zeta_{n,h} := \sqrt{n\mathrm{Var}(\widehat{\phi}_{n,h})^{-1}}\left(\widehat{\phi}_{n,h} - \phi_h\right),\ 1 \leq h \leq d.$$

Theorem 3.2.1. *Let $\{X_k\}_{k\in\mathbb{Z}}$ be a stationary process such that* (3.1.1) *is valid for every $d \geq 1$. Put $\phi_{i,j} = Cov(\xi_i, \xi_j)$, $r_n := \sup_{|i-j|\geq n} |\phi_{i,j}|$, and let $r_1 < 1$. Assume that one of the following two conditions is satisfied:*

(a) $\sum_{n=1}^{\infty} r_n^2 < \infty$,

(b) For some $\beta > 0$, $r_n(\log n)^{2+\beta} \to 0$.

Then there exists a sequence $d_n \to \infty$ such that

$$P\left(a_n^{-1}\big(\max_{0\le h\le d_n}\big|\zeta_{n,h}\big| - b_n\big) \le z\right) \to \exp(-e^{-z}),$$

where $a_n = (2\log d_n)^{-1/2}$ and $b_n = (2\log d_n)^{1/2} - (8\log d_n)^{-1/2}(\log\log d_n + 4\pi - 4)$.

Note that the difference in the limiting extreme distributions in Conjecture 3.1.1 and the above theorem are due to the different centering sequences b_n. As already mentioned, very general conditions for (3.1.1) are given e.g. in [131]. In order to establish a bound for the growth rate of d_n, we need to estimate the convergence rate in (3.1.1). This issue is dealt with in detail in Section 3.3, and allows us to formulate our main result. For simplicity, we focus on linear, short memory processes $\big\{X_k\big\}_{k\in\mathbb{Z}}$, $X_k = \sum_{i=0}^{\infty}\alpha_i\epsilon_{k-i}$ such that $\big\{\epsilon_k\big\}_{k\in\mathbb{Z}}$ is an IID sequence of real-valued random variables and $\alpha_i \in \mathbb{R}$.

Theorem 3.2.2. *Let $X_k = \sum_{i=0}^{\infty}\alpha_i\epsilon_{k-i}$, $k \in \mathbb{Z}$ be a linear process such that*

(1) $\sum_{i=0}^{\infty}\sqrt{\sum_{j=i}^{\infty}\alpha_j^2} < \infty$, $\sum_{i=0}^{\infty}\alpha_i^2 > \sup_{h\ge 1}\sum_{i=0}^{\infty}|\alpha_i\alpha_{i+h}|$.

(2) $\mathbb{E}\big(\epsilon_k\big) = 0$, $\mathbb{E}\big(\epsilon_k^2\big) = 1$, $\mathbb{E}\big(\epsilon_k^8\big) < \infty$, $k \in \mathbb{Z}$.

(3) The density function of X_1 exists and is continuous.

Then we have

$$P\left(a_n^{-1}\big(\max_{0\le h\le d_n}\big|\zeta_{n,h}\big| - b_n\big) \le z\right) \to \exp(-e^{-z}),$$

where $d_n = \lambda\log n/\log\log n$, $\lambda > 0$ sufficiently small, $a_n = (2\log d_n)^{-1/2}$ and $b_n = (2\log d_n)^{1/2} - (8\log d_n)^{-1/2}(\log\log d_n + 4\pi - 4)$.

Remark 3.2.3. Note that condition (1) is valid if $\alpha_n \ll n^{-\gamma}$, $\gamma > 3/2$.

Theorem 3.2.2 can be deduced by the following more general result, which can be viewed as a Berry-Esséen type result for increasing dimension.

Theorem 3.2.4. *Assume that the conditions of Theorem 3.2.2 hold. If $d_n = \lambda\log n/\log\log n$ with $\lambda > 0$ sufficiently small, then*

$$\lim_{n\to\infty}\sup_{x_1,\ldots,x_{d_n}\in\mathbb{R}}\Big|P\big(\zeta_{n,1}\le x_1,\ldots,\zeta_{n,d_n}\le x_{d_n}\big) - P\big(\xi_1\le x_1,\ldots,\xi_{d_n}\le x_{d_n}\big)\Big| \to 0,$$

where $\big\{\xi_k\big\}_{k\in\mathbb{N}}$ is the normalized Gaussian process appearing in (3.1.1).

Remark 3.2.5. We did not try for maximum generality in the above theorems; the assumption of linearity and in particular the dependence condition can be relaxed by using ideas from [130] and [131]. This allows for instance to consider martingale differences $\{\epsilon_k\}_{k\in\mathbb{Z}}$ instead of an IID sequence. We will see in the next chapter, that one can obtain a significantly larger growth rate for d_n if one only allows for an IID sequence $\{\epsilon_k\}_{k\in\mathbb{Z}}$.

It is interesting that the same result (same growth rate for the dimension d_n) was obtained in [73], if the process $\{X_k\}_{k\in\mathbb{Z}}$ is strongly mixing. This, however, is a fairly strong assumption, and is generally not true for linear processes, see for instance Andrews [5].

As already mentioned, the proofs of the Theorems 3.2.2 and 3.2.4 are based on a general estimate for the rate of convergence of the Cramér-Wold device. Loosely speaking, this quantitative Cramér-Wold device essentially tells us that the difference of the distribution functions of two random vectors $\mathbf{X} = (X_1, \ldots, X_d)$ and $\mathbf{Z} = (Z_1, \ldots, Z_d)$ is small, if the difference of the distribution functions of the linear combinations $s_1X_1 + ... + s_dX_d$ and $s_1Z_1 + ... + s_dZ_d$, is small. For more details, we refer to the next section.

3.3 A quantitative Cramér-Wold device

Given vectors $\mathbf{s}, \mathbf{t} \in \mathbb{R}^d$, we denote the usual scalar product with $\mathbf{s}^{\mathbf{T}}\mathbf{t}$, and put $|\mathbf{s}| = \sqrt{\mathbf{s}^{\mathbf{T}}\mathbf{s}}$. For a function $f \in L^2$ or $f \in L^1$, we write $\mathcal{F}(f)$ and φ_f simultaneously for the Fourier transform of f, and, given a random variable X, we write $\varphi_X(s) = \mathbb{E}\big(\exp(\mathrm{i}sX)\big)$.

Theorem 3.3.1. *Let $\mathbf{X} = (X_1, \ldots, X_d)$ and $\mathbf{Z} = (Z_1, \ldots, Z_d)$ be d-dimensional random vectors with mean $\mathbf{0}$ such that $\mathbf{X}$ has a continuous density and $\mathbf{Z}$ is Gaussian. Assume $\mathbb{E}(X_iX_j) = \phi_{i,j}$, $\phi_{i,i} = 1$, $1 \leq i, j \leq d$, and put $R = \sum_{i,j=1}^d |\phi_{i,j}|$. Assume also that*

$$\sup_x \big|P\big(X_j \leq x\big) - P\big(Z_j \leq x\big)\big| \leq \mathcal{X}, \quad 1 \leq j \leq d, \tag{3.3.1}$$

and

$$\sup_{\{x:\,|x|\geq\theta\}} \big|P\big(\mathbf{s}^{\mathbf{T}}\mathbf{X} \leq x\big) - P\big(\mathbf{s}^{\mathbf{T}}\mathbf{Z} \leq x\big)\big| \leq \mathcal{X}, \quad \theta, \mathcal{X} > 0, \tag{3.3.2}$$

for all $\mathbf{s} = (s_1, ..., s_d)$ with $\max_{1\leq i\leq d} |s_i| \leq M$. Then for any $a \geq 0$ we have

$$\big|P\big(|X_1| \leq a, ..., |X_d| \leq a\big) - P\big(|Z_1| \leq a, ..., |Z_d| \leq a\big)\big| \leq \epsilon = \min\{\epsilon_1, \epsilon_2\},$$

where

$$\epsilon_1 = C(a+d)^d \left(M\mathcal{X} + \theta + d\,M^{-1/4}\right) + C\exp(-a^2/2) + RM^{-1/2}\exp(-M(\sqrt{2}R)^{-1}),$$
$$\epsilon_2 = \mathcal{X}d + Cd\exp(-a^2/2), \tag{3.3.3}$$

with an absolute constant C.

Remark 3.3.2. The condition that the density function of $\mathbf{X}$ exists can be weakened, but it simplifies the calculations substantially and since Theorem 3.3.1 suffices for the purposes of the present paper, we will keep this condition. Also, Theorem 3.3.1 can be modified for sets of the form

$$\left\{b_1 \le X_1 \le a_1, ..., b_d \le X_d \le a_d\right\},$$

which will be apparent from the proof.

Remark 3.3.3. Condition (3.3.1) is the Kolmogorov distance of the two random variables X_j, Z_j, and it is contained in condition (3.3.2) if $\theta = 0$ and $M \ge 1$. Also, note that the assumption of Gaussianity of the vector $\mathbf{Z}$ is not a necessity, and is only reflected via the tail estimate given in (3.3.4). Thus, Theorem 3.3.1 can be adapted to non-Gaussian cases.

The classical Cramér-Wold device states that a sequence of real-valued d-dimensional vectors $\mathbf{X}_n = (X_{1,n}, \ldots, X_{d,n})$ (here d is fixed) converges in distribution to the vector $\mathbf{Z} = (Z_1, \ldots Z_d)$, if and only if all the linear combinations $s_1X_{1,n} + ... + s_dX_{d,n}$ converge in distribution to $s_1Z_1 + ... + s_dZ_d$, a fact that can readily be proved via characteristic functions. Note that this reduces the problem of establishing weak convergence in $\mathbb{R}^d$ to establishing weak convergence in $\mathbb{R}$ and vice versa. The new feature of Theorem 3.3.1 is that it gives an explicit upper bound for the approximation error in terms of the dimension d. This allows us to treat cases where $d = d_n \to \infty$, which is exactly the case encountered in Theorems 3.2.2 and 3.2.4. However, the idea of reducing the dimension from d to one is also the main idea in the proof of Theorem 3.3.1.

The proof requires some preliminary results. Denote with $F_X(x)$ the distribution function of a random variable X, and with $\Phi(x)$ the standard normal distribution function. We will use repeatedly the tail estimate

$$1 - \Phi(x) \le (2\pi)^{-1/2}x^{-1}\exp(-x^2/2) \qquad (x > 0) \tag{3.3.4}$$

and the fact that

$$\begin{aligned} |\varphi_X(1) - \varphi_Y(1)| &\le \mathbb{E}(|X|) + \mathbb{E}(|Y|) + 2\left(\mathbb{E}(X^2) + \mathbb{E}(Y^2)\right) \\ &\le 2\left(\mathbb{E}(X^2) + \sqrt{\mathbb{E}(X^2)} + \mathbb{E}(Y^2) + \sqrt{\mathbb{E}(Y^2)}\right) \end{aligned} \tag{3.3.5}$$

for any square integrable random variables X, Y. We will further use the following well known connection between Fourier transform and convolution.

Lemma 3.3.4. *Let* $\{h_j(x)\}_{1\le j\le d}$ *and* $f(x_1, ..., x_d)$ *be a collection of real valued, integrable functions such that their Fourier transform is integrable. Put*

$$T(y_1, ..., y_d) = \int_{R^d} f(x_1, ..., x_d) \prod_{j=1}^{h} h_j(y_j - x_j)\, d\,x_1 ... d\,x_d$$

Then we have

$$\mathcal{F}(T)(s_1, ..., s_d) = \mathcal{F}(f)(s_1, ..., s_d) \prod_{j=1}^{d} \mathcal{F}(h_j)(s_j),$$

for any $s_1, ..., s_d$*, and in particular*

$$\sup_{y_1,...,y_d} \big|T(y_1, ..., y_d)\big| \le (2\pi)^{-d/2} \big\|\mathcal{F}(T)(s_1, ..., s_d)\big\|_1 \quad a.s.$$

where $\|\cdot\|_1$ *denotes* $L^1(-\infty, \infty)$ *norm.*

The next lemma is a continuity result.

Lemma 3.3.5. *Under the conditions of Theorem 3.3.1 we have*

$$\begin{aligned}&\Big|P\big(|X_1| \le a_1, ..., |X_d| \le a_d\big) - P\big(|X_1| \le a_1 + \delta, ..., |X_d| \le a_d + \delta\big)\Big| \\ &\le C\left(d\,\mathcal{X} + \delta \sum_{j=1}^{d} \exp(-a_j^2/2)\right),\end{aligned}$$

for $a_i > 0$*,* $1 \le i \le d$ *and* $\delta \ge 0$*, with an absolute constant* $C > 0$*.*

Proof. First observe that

$$\begin{aligned}&\Big|P\big(|X_1| \le a_1, ..., |X_d| \le a_d\big) - P\big(|X_1| \le a_1 + \delta, ..., |X_d| \le a_d + \delta\big)\Big| \\ &\le \sum_{j=1}^{d}\Big(P\big(-a_j - \delta \le X_j \le -a_j\big) + P\big(a_j \le X_j \le a_j + \delta\big)\Big). \qquad (3.3.6)\end{aligned}$$

By assumption (3.3.1) of Theorem 3.3.1 and by properties of the Gaussian distribution function, this is smaller than

$$2d\,\mathcal{X} + C\sum_{j=1}^{d}\big(\Phi(a_j + \delta) - \Phi(a_j)\big) \le C\left(d\,\mathcal{X} + \delta \sum_{j=1}^{d} \exp(-a_j^2/2)\right).$$

□

Proof of Theorem 3.3.1. The idea of the proof is to approximate the probabilities

$$P\big(-a \le X_1 \le a, ..., -a \le X_d \le a\big) \quad \text{and} \quad P\big(-a \le Z_1 \le a, ..., -a \le Z_d \le a\big)$$

using mollifiers (smooth truncation functions), and then estimate the various approximation errors with the help of Fourier transforms. We will repeatedly use the fact that $\varphi_{\mathbf{s}^T\mathbf{X}}(1) = \varphi_{\mathbf{X}}(s_1, ..., s_d)$. Throughout this proof, C denotes absolute constants that may vary from one formula to another.

Let us first establish the bound ϵ_2, whose derivation is rather straightforward. We have that

$$\begin{aligned}
&\big|P\big(-a \le X_1 \le a, ..., -a \le X_d \le a\big) - P\big(-a \le Z_1 \le a, ..., -a \le Z_d \le a\big)\big| \\
&= \big|1 - P\big(\max_{1\le j\le d} |X_j| \ge a\big) - 1 + P\big(\max_{1\le j\le d} |Z_j| \ge a\big)\big| \\
&\le d\left(\max_{1\le j\le d} P\big(|X_j| \ge a\big) + \max_{1\le j\le d} P\big(|Z_j| \ge a\big)\right).
\end{aligned}$$

By the assumptions and relation (3.3.4), we conclude that

$$d\left(\max_{1\le j\le d} P\big(|X_j| \le a\big) + \max_{1\le j\le d} P\big(|Z_j| \le a\big)\right) \le C\big(d\,\mathcal{X} + d\exp(-a^2/2)\big),$$

which yields the claim. Unfortunately, establishing the bound ϵ_1 is more involved. To this end, let $B_M := \{x \in \mathbb{R}^1 : |x| \le M\}$ and denote $\mathbf{1}_{B_M}(x)$ $(x \in \mathbb{R}^1)$ and $\mathbf{1}_{B_M}(\mathbf{x})$ $(\mathbf{x} \in \mathbb{R}^d)$ the indicator function of B_M and the indicator function of $B_M \times \cdots \times B_M$, respectively. For any $a > 0$, put $I_{\mathbf{a}}(x) := \mathbf{1}_{[-a,a]}(x)$. Fix $a > 0$, let $h = h(x_1, ..., x_d)$ be a real-valued function, and define

$$H_h(y_1, ..., y_d) := \int_{\mathbb{R}^d} I_{\mathbf{a}}(y_1 - x_1) I_{\mathbf{a}}(y_2 - x_2) \dots I_{\mathbf{a}}(y_d - x_d) h(x_1, ..., x_d)\, dx_1 ... d\, x_d$$

Let $f = f_{X_1,...,X_d}(x_1, ..., x_d)$ and $g = g_{Z_1,...,Z_d}(z_1, ..., z_d)$ be the density functions of $\mathbf{X}$ and $\mathbf{Z}$. Clearly

$$P\big(-a \le X_1 \le a, ..., -a \le X_d \le a\big) = \int_{[-a,a]^d} f_{X_1,...,X_d}(x_1, ..., x_d)\, dx_1 ... dx_d = H_f(0, ..., 0),$$

and the same is valid for g. Thus

$$\begin{aligned}
&\big|P\big(-a \le X_1 \le a, ..., -a \le X_d \le a\big) - P\big(-a \le Z_1 \le a, ..., -a \le Z_d \le a\big)\big| \\
&= \big|H_f(0, ..., 0) - H_g(0, ..., 0)\big| = \big|H_{f-g}(0, ..., 0)\big|.
\end{aligned}$$

For some $b > 0$ we introduce the quantity

$$J_{f-g}(z_1, ..., z_d) := \int_{\mathbb{R}^d} \prod_{j=1}^{d} (2b)^{-1} I_{\mathbf{b}}(z_j - y_j) H_{f-g}(y_1, ..., y_d) d\, y_1 ... d\, y_d.$$

Then we have the following bound for $\big|H_{f-g}(0,...,0)\big|$:

$$\big|H_{f-g}(0,...,0)\big| = \big|\int_{\mathbb{R}^d}\prod_{j=1}^{d}(2b)^{-1}I_{\mathbf{b}}(-y_j)H_{f-g}(0,...,0)dy_1,...,dy_d\big|$$

$$\leq \big|\int_{\mathbb{R}^d}\prod_{j=1}^{d}(2b)^{-1}I_{\mathbf{b}}(-y_j)H_{f-g}(y_1,...,y_d)dy_1,...,dy_d\big| + \sup_{|y_j|\leq b, 1\leq j\leq d}\big|H_{f-g}(y_1,...,y_d) - H_{f-g}(0,...,0)\big|$$

$$\leq \big|J_{f-g}(0,...,0)\big| + \sup_{|y_j|\leq b, 1\leq j\leq d}\big|H_{f-g}(y_1,...,y_d) - H_{f-g}(0,...,0)\big|.$$

By Lemma 3.3.5, we have

$$\sup_{|y_j|\leq b, 1\leq j\leq d}\big|H_{f-g}(y_1,...,y_d) - H_{f-g}(0,...,0)\big| \leq C\left(d\,\mathcal{X} + d\,b\exp(-a^2/2)\right).$$

Choosing $b^{-1} = d$, we obtain

$$\sup_{|y_j|\leq b, 1\leq j\leq d}\big|H_{f-g}(y_1,...,y_d) - H_{f-g}(0,...,0)\big| \leq C\left(d\,\mathcal{X} + \exp(-a^2/2)\right). \quad (3.3.7)$$

Hence we need to study $\big|J_{f-g}(0,...,0)\big|$ if $b^{-1} = d$. By Lemma 3.3.4, we have that

$$\sup_{z_1,...,z_d}\big|J_{f-g}(z_1,...,z_d)\big| \leq (2\pi)^{-d/2}\big\|\mathcal{F}\big(J_{f-g}\big)\big\|_1$$

$$= (2\pi)^{-d/2}\big\|\mathcal{F}(H_{f-g})\prod_{j=1}^{d}\mathcal{F}\big(\frac{1}{2b}I_{\mathbf{b}}\big)\big\|_1 = (2\pi)^{-d/2}\big\|\mathcal{F}(f-g)\prod_{j=1}^{d}\mathcal{F}(I_{\mathbf{a}})(s_j)\prod_{j=1}^{d}\mathcal{F}\big(\frac{1}{2b}I_{\mathbf{b}}\big)\big\|_1$$

$$= (2\pi)^{-d/2}\big\|\mathcal{F}(f-g)\prod_{j=1}^{d}\left(\mathcal{F}\big(\frac{1}{2b}I_{\mathbf{b}}\big)\mathcal{F}(I_{\mathbf{a}})(s_j)\right)\big\|_1 := \Omega_d. \quad (3.3.8)$$

Notice that

$$\begin{aligned}\big|\mathcal{F}(f-g)(\mathbf{s_d})\big| &\leq \big|\mathcal{F}(f-g)(\mathbf{s_d})\mathbf{1}_{B_{\sqrt{M}}}(\mathbf{s_d})\big| + \big|\mathcal{F}(f-g)(\mathbf{s_d})\mathbf{1}_{\mathbb{R}^d\setminus B_{\sqrt{M}}}(\mathbf{s_d})\big| \\ &= \big|\varphi_{\mathbf{s_d^T}\cdot\mathbf{X}}(1) - \varphi_{\mathbf{s_d^T}\cdot\mathbf{Z}}(1)\big|\mathbf{1}_{B_{\sqrt{M}}}(\mathbf{s_d}) + \big|\mathcal{F}(f-g)(\mathbf{s_d})\big|\mathbf{1}_{\mathbb{R}^d\setminus B_{\sqrt{M}}}(\mathbf{s_d}) \\ &:= A(\mathbf{s_d}) + B(\mathbf{s_d}).\end{aligned}$$

To treat $A(\mathbf{s_d})$, observe that

$$\begin{aligned}\big|\varphi_{\mathbf{s_d}^{\mathbf{T}}\cdot X}(1) - \varphi_{\mathbf{s_d^T}\cdot\mathbf{Z}}(1)\big| &\leq \big|\mathbb{E}\big(\exp(\mathrm{i}\mathbf{s_d^T}\cdot\mathbf{X})\mathbf{1}_{B_M}(\mathbf{s_d^T}\cdot\mathbf{X})\big) - \mathbb{E}\big(\exp(\mathrm{i}\mathbf{s_d^T}\cdot\mathbf{Z})\mathbf{1}_{B_M}(\mathbf{s_d^T}\cdot\mathbf{Z})\big)\big| \\ &+ \big|\mathbb{E}\big(\exp(\mathrm{i}\mathbf{s_d^T}\cdot\mathbf{X})\mathbf{1}_{\mathbb{R}\setminus B_M}(\mathbf{s_d^T}\cdot\mathbf{X})\big)\big| + \big|\mathbb{E}\big(\exp(\mathrm{i}\mathbf{s_d^T}\cdot\mathbf{Z})\mathbf{1}_{\mathbb{R}\setminus B_M}(\mathbf{s_d^T}\cdot\mathbf{Z})\big)\big| \\ &\leq \big|\mathbb{E}\big(\exp(\mathrm{i}\mathbf{s_d^T}\cdot\mathbf{X})\mathbf{1}_{B_M}(\mathbf{s_d^T}\cdot\mathbf{X})\big) - \mathbb{E}\big(\exp(\mathrm{i}\mathbf{s_d^T}\cdot\mathbf{Z})\mathbf{1}_{B_M}(\mathbf{s_d^T}\cdot\mathbf{Z})\big)\big| \\ &+ P\big(\mathbf{s_d^T}\cdot\mathbf{X}\in\mathbb{R}\setminus B_M\big) + P\big(\mathbf{s_d^T}\cdot\mathbf{Z}\in\mathbb{R}\setminus B_M\big) \\ &:= C(\mathbf{s_d}) + D(\mathbf{s_d}) + E(\mathbf{s_d}).\end{aligned}$$

Let $B_{\theta,M} := \{x \mid \theta \le |x| \le M\}$. Then by Lemma 3.3.5, we have that

$$\begin{aligned}
C(\mathbf{s_d}) =& \big|\mathbb{E}\big(\exp(\mathrm{i}\mathbf{s_d^T}\cdot\mathbf{X})\mathbf{1}_{B_M}(\mathbf{s_d}^{\mathbf{T}}\cdot X)\big) - \mathbb{E}\big(\exp(\mathrm{i}\mathbf{s_d^T}\cdot\mathbf{Z})\mathbf{1}_{B_M}(\mathbf{s_d^T}\cdot\mathbf{Z})\big)\big| \\
&\le \big|\mathbb{E}\big(\exp(\mathrm{i}\mathbf{s_d^T}\cdot\mathbf{X})\mathbf{1}_{B_\theta}(\mathbf{s_d^T}\cdot\mathbf{X})\big) - \mathbb{E}\big(\exp(\mathrm{i}\mathbf{s_d^T}\cdot\mathbf{Z})\mathbf{1}_{B_\theta}(\mathbf{s_d^T}\cdot\mathbf{Z})\big)\big| \\
&+ \big|\mathbb{E}\big(\exp(\mathrm{i}\mathbf{s_d^T}\cdot\mathbf{X})\mathbf{1}_{B_{\theta,M}}(\mathbf{s_d^T}\cdot\mathbf{X})\big) - \mathbb{E}\big(\exp(\mathrm{i}\mathbf{s_d^T}\cdot\mathbf{Z})\mathbf{1}_{B_{\theta,M}}(\mathbf{s_d^T}\cdot\mathbf{Z})\big)\big| \\
&\le 8\theta + \big|\mathbb{E}\big(\exp(\mathrm{i}\mathbf{s_d^T}\cdot\mathbf{X})\mathbf{1}_{B_{\theta,M}}(\mathbf{s_d^T}\cdot\mathbf{X})\big) - \mathbb{E}\big(\exp(\mathrm{i}\mathbf{s_d^T}\cdot\mathbf{Z})\mathbf{1}_{B_{\theta,M}}(\mathbf{s_d^T}\cdot\mathbf{Z})\big)\big| \\
&:= 8\theta + F(\mathbf{s_d}).
\end{aligned}$$

Put

$$\mu_{\mathbf{s_d}}(u) := P\big(0 \le \mathbf{s_d^T}\cdot\mathbf{X} \le u\big) - P\big(0 \le \mathbf{s_d^T}\cdot\mathbf{Z} \le u\big) = F_{\mathbf{s_d^T}\cdot\mathbf{X}}(u) - F_{\mathbf{s_d^T}\cdot\mathbf{Z}}(u).$$

By the conditions of Theorem 3.3.1, this gives us the following bound for $F(\mathbf{s_d})$

$$\begin{aligned}
F(\mathbf{s_d}) &= \left|\int_{B_{M,\theta_n}} \big(\cos u + \mathrm{i}\sin u\big) d\big(F_{\mathbf{s_d}^{\mathbf{T}}\cdot X}(u) - F_{\mathbf{s_d}^{\mathbf{T}}\cdot Z}(u)\big)\right| \\
&\le \sum_{j=1}^{4M}\left(\left|\int_{j\pi/2}^{(j+1)\pi/2}\sin x\mu_{\mathbf{s_d}}(dx)\right| + \left|\int_{j\pi/2}^{(j+1)\pi/2}\cos x\mu_{\mathbf{s_d}}(dx)\right|\right) \\
&+ \left|\int_{\theta}^{\pi/2}\sin x\mu_{\mathbf{s_d}}(dx)\right| + \left|\int_{\theta}^{\pi/2}\cos x\mu_{\mathbf{s_d}}(dx)\right|.
\end{aligned}$$

For a continuous random variable $\theta \le U \le \pi/2$, an application of integration by parts gives us

$$\begin{aligned}
\int_{\theta}^{\pi/2}\cos x dF_U(x) &= -\cos\theta F_U(\theta) + \cos\pi/2 F_U(\pi/2) - \int_{\theta}^{\pi/2} F_U(x) d\cos x \\
&= -\cos\theta F_U(\theta) + \int_{\theta}^{\pi/2} F_U(x)\sin x dx.
\end{aligned}$$

Since $|\cos x|$ and $|\sin x|$ are bounded by one, we obtain from the conditions of Theorem 3.3.1 and integration by parts

$$\begin{aligned}
\left|\int_{\theta}^{\pi/2}\cos x\mu_{\mathbf{s_d}}(dx)\right| &\le C\big|P\big(\theta \le \mathbf{s_d^T}\cdot\mathbf{X} \le \pi/2\big) - P\big(\theta \le \mathbf{s_d^T}\cdot\mathbf{Z} \le \pi/2\big)\big| \\
&\le C\mathcal{X}.
\end{aligned}$$

Similarly, one obtains the same bound for $\int_{j\pi/2}^{(j+1)\pi/2}\cos x\mu_{\mathbf{s_d}}(dx)$ and $\int_{j\pi/2}^{(j+1)\pi/2}\sin x\mu_{\mathbf{s_d}}(dx)$ for $1 \le j \le M$, hence we obtain

$$F(\mathbf{s_d}) \le CM\,\mathcal{X}.$$

In order to treat $D(\mathbf{s_d})$ and $E(\mathbf{s_d})$, notice that the assumptions of Theorem 3.3.1 give us

$$\big|E(\mathbf{s_d}) - D(\mathbf{s_d})\big| \leq C\mathcal{X}.$$

On the other hand, using Lemma 3.3.4, we get

$$\begin{aligned}\mathbf{1}_{B_{\sqrt{M}}}(\mathbf{s_d})E(\mathbf{s_d}) &\leq \sup_{\mathbf{s_d}\in B_{\sqrt{M}}} \sqrt{\frac{2\sigma^2_{\mathbf{s_d}}}{\pi}}M^{-1}\exp(-M^2(2\sigma^2_{\mathbf{s_d}})^{-1}) \\ &\leq \sqrt{\frac{2R_d^2 M}{\pi M^2}}\exp(-M(\sqrt{2}R_d)^{-1}) \\ &\leq \sqrt{\frac{2R_d^2}{\pi M}}\exp(-M(\sqrt{2}R_d)^{-1}) := G,\end{aligned}$$

where $\sigma^2_{\mathbf{s_d}} = \sum_{i,j=1}^{d} s_j s_i \rho_{i,j} \leq R_d \sup_{1\leq i\leq d}|s_i|^2$. Hence we obtain

$$\mathbf{1}_{B_{\sqrt{M}}}(\mathbf{s_d})\big|E(\mathbf{s_d}) + D(\mathbf{s_d})\big| \leq 2G + C\mathcal{X},$$

and thus

$$A(\mathbf{s_d}) \leq C\left(M\ \mathcal{X} + \theta + G + \mathcal{X}\right)\mathbf{1}_{B_{\sqrt{M}}}(\mathbf{s_d}).$$

Continuing in equation 3.3.8, we obtain

$$\begin{aligned}\Omega_d &= (2\pi)^{-d/2}\big\|\mathcal{F}(f-g)\prod_{j=1}^{d}\Big(\mathcal{F}(\frac{1}{2b}I_{\mathbf{b}})\mathcal{F}(I_{\mathbf{a}})(s_j)\Big)\big\|_1 \\ &\leq (2\pi)^{-d/2}\big\|\big(A_n(\mathbf{s_d}) + B_n(\mathbf{s_d})\big)\prod_{j=1}^{d}\Big(\mathcal{F}(\frac{1}{2b}I_{\mathbf{b}})\mathcal{F}(I_{\mathbf{a}})(s_j)\Big)\big\|_1 \\ &\leq (2\pi)^{-d/2}\big\|\mathbf{1}_{B_{\sqrt{M}}}(\mathbf{s_d})\prod_{j=1}^{d}\Big(\mathcal{F}(\frac{1}{2b}I_{\mathbf{b}})\mathcal{F}(I_{\mathbf{a}})(s_j)\Big)\big\|_1 C\left(M\ \mathcal{X} + \theta + G\right) \\ &+ (2\pi)^{-d/2}\big\|\prod_{j=1}^{d}\Big(\mathcal{F}(\frac{1}{2b}I_{\mathbf{b}})\mathcal{F}(I_{\mathbf{a}})(s_j)\Big)\mathbf{1}_{\mathbb{R}^d\setminus B_{\sqrt{M}}}(\mathbf{s_d})B_n(\mathbf{s_d})\big\|_1.\end{aligned}$$

Since $\mathcal{F}(I_{\mathbf{a}})(s) = (2\sin as)s^{-1}$, we obtain

$$\begin{aligned}&(2\pi)^{-d/2}\Big\|\prod_{j=1}^{d}\mathbf{1}_{B_{\sqrt{M}}}(\mathbf{s_d})\Big(\mathcal{F}(\frac{1}{2b}I_{\mathbf{b}})\mathcal{F}(I_{\mathbf{a}})(s_j)\Big)\Big\|_1 \leq \prod_{j=1}^{d}\big\|\frac{\sin bs_j \sin as_j}{bs_j^2}\big\|_1 \\ &\leq C\big(a + b^{-1}\big)^d,\end{aligned}$$

and similarly, since $|B_n(s)| \leq 2$,

$$\begin{aligned}
&(2\pi)^{-d/2}\Big\|\prod_{j=1}^{d}\mathbf{1}_{\mathbb{R}^d\setminus B_{\sqrt{M}}}(\mathbf{s_d})B_n(\mathbf{s_d})\left(\mathcal{F}\big(\frac{1}{2b}I_{\mathbf{b}}\big)\mathcal{F}(I_{\mathbf{a}})(s_j)\right)\Big\|_1\\
&\leq Cd\big\|(bs^2)^{-1}\mathbf{1}_{\mathbb{R}\setminus B_{\sqrt{M}}}(s)\big\|_1\prod_{j=2}^{d}\Big\|\frac{\sin bs_j \sin as_j}{bs_j^2}\Big\|_1\\
&\leq Cd(b^{-1}+a)^d M^{-1/2}.
\end{aligned}$$

Piecing everything together, we obtain the bound

$$\begin{aligned}
\Omega_{n,d} &\leq C(b^{-1}+a)^d\left(M\,\mathcal{X}+\theta+G+\mathcal{X}+d\,M^{-1/2}\right)\\
&\leq C(a+d)^d\left(M\,\mathcal{X}+\theta+G+\mathcal{X}+d\,M^{-1/2}\right),
\end{aligned}$$

which completes the proof. □

3.4 Normal approximation

Let

$$S_{n,h} = n^{-1/2}\sum_{k=0}^{n-h}\big(X_k X_{k+h}-\phi_h\big) \quad (1\leq h\leq d_n), \qquad S_n=\sum_{h=0}^{d_n} s_h S_{n,h} \qquad (3.4.1)$$

with some sequence $d_n \to \infty$ of positive integers and real coefficients s_h. In this section we prove, under suitable assumptions, a normal approximation for the r.v.'s $S_{n,h}, 1 \leq h \leq d_n$ and S_n. Using the quantitative Cramér-Wold device obtained in the previous section, it will then follow that the distribution of $\mathbf{X}_n = (S_{n,1},\ldots,S_{n,d_n})$ is close to the distribution of a Gaussian vector $\mathbf{Z}_n = (Z_{n,1},\ldots,Z_{n,d_n})$, with the same covariance structure. In Section 3.5 we will show that the distribution of $\mathbf{Z}_n$ converges to the finite dimensional distributions of a Gaussian process $\big\{\xi_k\big\}_{k\in\mathbb{Z}}$ and using the extremal theory of Gaussian processes, Theorem 3.2.2 will follow.

Theorem 3.4.1. *Let $X_k = \sum_{i=0}^{\infty}\alpha_i\epsilon_{k-i}$, $k \in \mathbb{Z}$ be a linear process such that the ϵ_i are i.i.d. random variables and*

(i) $\sum_{i=0}^{\infty}\sqrt{\sum_{j=i}^{\infty}\alpha_j^2} < \infty$,

(ii) $\mathbb{E}\big(\epsilon_1\big) = 0$, $\mathbb{E}\big(\epsilon_1^2\big) = 1$, $\mathbb{E}\big(\epsilon_1^8\big) < \infty$.

Let $d_n \to \infty, M_n \to \infty, \theta_n \to 0$ be positive sequences satisfying $d_n^4 M_n^4 = \mathcal{O}(n^r)$, $0 < r < 1/2$ and assume also $\max_{0 \leq h \leq d_n} |s_h| \leq M_n$. Then there exist a Gaussian random variable Z_n with $Var(Z_n) = Var(S_n)$ and a Gaussian vector $\{Z_{n,h}, 1 \leq h \leq d_n\}$ with the same covariance structure as $\{S_{n,h}, 1 \leq h \leq d_n\}$ such that

$$\sup_{\{x:|x|\geq\theta_n\}} \big|P\big(S_n \leq x\big) - P\big(Z_n \leq x\big)\big| \leq C\left(n^{-1/10+r} + n^{-1/2}\theta_n^{-2}\right),$$
$$\sup_x \big|P\big(S_{n,h} \leq x\big) - P\big(Z_{n,h} \leq x\big)\big| \leq C\left(n^{-1/10+r}\right), \quad 0 \leq h \leq d_n,$$

Proof. The proof is based on martingale approximation and an estimate of the speed of convergence in the martingale CLT by Heyde and Brown [27]. Denote with $\mathcal{G}_i = \sigma(\epsilon_k,\ k \leq i)$ the σ-algebra generated by the innovations $\big\{\epsilon_k\big\}_{k\leq i}$. For the proof, it is convenient to introduce the following projection operator

$$\mathcal{P}_i X = \mathbb{E}\big(X\big|\mathcal{G}_i\big) - \mathbb{E}\big(X\big|\mathcal{G}_{i-1}\big).$$

For $j < k$, put

$$\begin{aligned} X_k &:= X_k^{(k\leq j)} + X_k^{(k>j)} = X_k^{(k<j)} + X_k^{(k=j)} + X_k^{(k>j)} \\ &= \sum_{i=0}^{\infty} \alpha_{k-j+i}\epsilon_{j-i} + \sum_{i=0}^{k-j-1} \alpha_i \epsilon_{k-i} \\ &= \sum_{i=0}^{\infty} \alpha_{k-j+1+i}\epsilon_{j-i-1} + \alpha_{k-j}\epsilon_j + \sum_{i=0}^{k-j-1} \alpha_i \epsilon_{k-i}. \end{aligned}$$

Note that for $h \geq 0$,

$$\mathcal{P}_i X_k X_{k+h} = X_k^{(k=i)} X_{k+h}^{(k+h=i)} - \mathbb{E}\big(X_k^{(k=i)} X_{k+h}^{(k+h=i)}\big) + X_k^{(k=i)} X_{k+h}^{(k+h<i)} + X_k^{(k<i)} X_{k+h}^{(k+h=i)}. \tag{3.4.2}$$

For fixed n, we introduce the martingale

$$M_l := \sum_{k=1}^{\infty} \sum_{h=1}^{d} s_h \big(\mathbb{E}\big(X_k X_{k+h}\big|\mathcal{G}_l\big) - \mathbb{E}\big(X_k X_{k+h}\big|\mathcal{G}_0\big)\big),\ l \geq 0,$$

and the process

$$R_l := \sum_{k=l+1}^{\infty} \sum_{h=1}^{d} s_h \big(\mathbb{E}\big(X_k X_{k+h} - \phi_h\big|\mathcal{G}_l\big)\big),\ l \geq 0.$$

Thus, we obtain the decomposition

$$S_n = M_n + R_0 - R_{n+d}.$$

Note that by stationarity, for $k \geq 0$ it holds that

$$\big\|R_k\big\|_2^2 = \big\|R_0\big\|_2^2,$$

and using (3.4.2) in connection with Assumption $(i),(ii)$, we obtain

$$\big\|R_k\big\|_2^2 = \big\|R_0\big\|_2^2 \leq Cd^2 \max_{1\leq h\leq d} |s_h|^2 = \mathcal{O}\big(n^{r/2}\big). \tag{3.4.3}$$

For the martingale differences $M_l - M_{l-1}$, we have that

$$\Delta M_l = M_l - M_{l-1} = \sum_{i=l-d}^{\infty} \mathcal{P}_l\big(\sum_{h=1}^{d} s_h X_i X_{i+h}\big),$$

in particular, proceeding as in the case of the process $\big\{R_l\big\}_{l\in\mathbb{N}}$, we have that

$$\big\|\Delta M_l\big\|_4^4 \leq Cd^4 \max_{1\leq h\leq d} |s_h|^4. \tag{3.4.4}$$

In addition, we put

$$\sigma_{n,d} := \sum_{l=1}^{n} \mathbb{E}\big(\Delta M_l^2\big). \tag{3.4.5}$$

For computational reasons, we now introduce the martingales N_l^n, defined as

$$N_l^n := \sum_{k=1}^{n} \mathbb{E}\big(\Delta M_k^2 - \mathbb{E}(\Delta M_k^2)\big|\mathcal{G}_l\big),$$

and the corresponding martingale differences

$$\Delta N_l^n = N_l^n - N_{l-1}^n = \sum_{i=l}^{n} \mathcal{P}_l\big(\Delta M_i^2\big).$$

For a discrete time martingale $\{M_k\}_{k\in\mathbb{N}}$, denote with

$$[M,M]_k := \sum_{l=0}^{k} \Delta M_l^2$$

the square bracket of a martingale. Using the L^2 orthogonality of the martingale differences, we then have that

$$\begin{aligned}\big\|[M,M]_n - \mathbb{E}\big([M,M]_n\big)\big\|_2^2 &= \big\|N_n^n\big\|_2^2 = \big\|\sum_{k=0}^{n} \Delta N_k^n\big\|_2^2 = \sum_{k=0}^{n}\big\|\Delta N_k^n\big\|_2^2 \\ &= \sum_{k=0}^{n}\big\|\sum_{i=0}^{n} \mathcal{P}_k\big(\Delta M_i^2\big)\big\|_2^2 = \sum_{k=0}^{n}\sum_{i=0}^{n}\big\|\mathcal{P}_k\big(\Delta M_i^2\big)\big\|_2^2.\end{aligned}$$

Proceeding as in the case of the process $\{R_l\}_{l\in\mathbb{N}}$, one readily computes that

$$\sum_{i=0}^{\infty}\big\|\Delta M_i^2\big\|_2^2 \leq Cd^4 \max_{1\leq h\leq d}|s_h|^4. \tag{3.4.6}$$

It now follows from [27, Theorem], that

$$\sup_{x\in\mathbb{R}}\big|P\big(\sigma_{n,d}^{-1/2}M_n \leq x\big) - \Phi(x)\big| \leq C\big(\sigma_{n,d}^{-1}d^4 \max_{1\leq h\leq d}|s_h|^4\big)^{1/5}. \tag{3.4.7}$$

We will now show that we can essentially replace M_n with $S_{n,d}$ in (3.4.7). To this end, let γ_n be a positive, monotone decreasing sequence. For a random variable U, we define the following sets:

$$\begin{aligned}
A_U &:= \{\omega||U(\omega)| \leq \gamma_n\},\\
A_U^+ &:= A \cap \{\omega|U(\omega) \geq 0\},\\
A_U^- &:= A \cap \{\omega|U(\omega) < 0\}.
\end{aligned}$$

For a random variable V, we now have that

$$\begin{aligned}
&\big|P\big(V+U\leq x\big) - P\big(V\leq x\big)\big| \leq P\big(\{x-U\leq V\leq x\}\cap A_U\big) + P\big(A_U^c\big)\\
&\leq P\big(\{x-U\leq V\leq x\}\cap A_U^+\big) + P\big(\{x\leq V\leq x-U\}\cap A_U^-\big) + P\big(A_U^c\big)\\
&\leq \big|P\big(\{x-\gamma_n\leq V\leq x\}\cap A_U^+\big) + P\big(\{x\leq V\leq x+\gamma_n\}\cap A_U^-\big) + P\big(A_U^c\big)\\
&\leq P\big(x-\gamma_n\leq V\leq x\big) + P\big(x\leq V\leq \gamma_n + x\big) + P\big(A_U^c\big)\\
&\leq 2P\big(x-\gamma_n\leq V\leq x+\gamma_n\big) + P\big(A_U^c\big).
\end{aligned}$$

The Markov inequality yields

$$P\big(A^c\big) \leq \frac{\mathbb{E}\big(U^2\big)}{\gamma_n^2},$$

hence, substituting $U = \sigma_{n,d}^{-1/2}(R_0 - R_{n+d})$, $V = \sigma_{n,d}^{-1/2}M_n$, we obtain

$$\begin{aligned}
&\big|P\big(\sigma_{n,d}^{-1/2}S_n \leq x\big) - P\big(\sigma_{n,d}^{-1/2}M_n \leq x\big)\big|\\
&\leq 2P\big(x-\gamma_n \leq \sigma_{n,d}^{-1/2}M_n \leq x+\gamma_n\big) + \frac{C}{\gamma_n^2\,\sigma_{n,d}}.
\end{aligned} \tag{3.4.8}$$

According to (3.4.7), the above is smaller than

$$\begin{aligned}
&C\big(\sigma_{n,d}^{-1}d^4 \max_{1\leq h\leq d}|s_h|^4\big)^{1/5} + 2\sup_{\substack{x\in\mathbb{R},\\ |y|\leq\gamma_n}}\big|\Phi(x-y) - \Phi(x+y)\big| + \frac{C}{\gamma_n^2\,\sigma_{n,d}}\\
&\leq C\big(\sigma_{n,d}^{-1}d^4 \max_{1\leq h\leq d}|s_h|^4\big)^{1/5} + 4\gamma_n + \frac{C}{\gamma_n^2\,\sigma_{n,d}}.
\end{aligned}$$

By equating the last two terms, we obtain $\gamma_n = \sigma_{n,d}^{-1/3}$, hence

$$\sup_{x\in\mathbb{R}}\big|P\big(\sigma_{n,d}^{-1/2}S_n \le x\big) - \Phi(x)\big| \le C\big(\sigma_{n,d}^{-1}d^4 \max_{1\le h\le d}|s_h|^4\big)^{1/5}. \tag{3.4.9}$$

We will now consider the two cases $\mathrm{Var}(S_n) \le n^{2/3}$ and $\mathrm{Var}(S_n) > n^{2/3}$. In the first case, note that

$$\begin{aligned}\sup_{\{x||x|\ge\theta_n\}}\big|P\big(n^{-1/2}S_n \le x\big) - P\big(Z_n \le x\big)\big| &= \sup_{\{x|x\ge\theta_n\}}\big|P\big(n^{-1/2}S_n > x\big) - P\big(Z_n > x\big)\big| \\ &\quad + \sup_{\{x|-x\ge\theta_n\}}\big|P\big(n^{-1/2}S_n \le x\big) - P\big(Z_n \le x\big)\big| \\ &\le 4\frac{\mathrm{Var}(S_n)n^{-1}}{\theta_n^2} \le 4n^{-1/3}\theta_n^{-2}.\end{aligned}$$

In order to treat the second case, by the Cauchy-Schwarz inequality, we have

$$\big|\mathrm{Var}(S_n) - \sigma_{n,d}\big| \le \mathrm{Var}\big(R_0 - R_{n+d}\big) + 2\sqrt{\mathrm{Var}\big(R_0 - R_{n+d}\big)\mathrm{Var}(S_n)},$$

which implies that $\sigma_{n,d} \ge Cn^{2/3}$. Put

$$\Delta_{n,d} := \sigma_{n,d}^{-1}\left(\mathrm{Var}\big(R_0 - R_{n+d}\big) + 2\sqrt{\mathrm{Var}\big(R_0 - R_{n+d}\big)\mathrm{Var}(S_n)}\right), \tag{3.4.10}$$

and note that $\Delta_{n,d} = \mathcal{O}\left(n^{-1/3+r/4}\right)$. In addition, one readily verifies

$$\big|\mathrm{Var}(S_n)^{1/2}\sigma_{n,d}^{-1/2} - 1\big| \le \Delta_{n,d}. \tag{3.4.11}$$

This gives us the following bound

$$\begin{aligned}&\sup_{\{x||x|\ge\theta_n\}}\big|P\big(n^{-1/2}S_n \le x\big) - P\big(Z_n \le x\big)\big| \le C\big(\sigma_{n,d}^{-1}d^4 \max_{1\le h\le d}|s_h|^4\big)^{1/5} \\ &+ \sup_{\{x||x|\ge\theta_n\}}\big|\Phi\big(xn^{1/2}\sigma_{n,d}^{-1/2}\big) - \Phi\big(xn^{1/2}\mathrm{Var}(S_n)^{-1/2}\big)\big|.\end{aligned}$$

Due to (3.4.11), we have

$$\begin{aligned}&\sup_{\{x||x|\ge\theta_n\}}\big|\Phi\big(xn^{1/2}\sigma_{n,d}^{-1/2}\big) - \Phi\big(xn^{1/2}\mathrm{Var}(S_n)^{-1/2}\big)\big| \\ &\le \sup_{\{x||x|\ge\theta_n\}}\sup_{\{y||y|\le\Delta_{n,d}\}}\big|\Phi\big(xn^{1/2}\mathrm{Var}(S_n)^{-1/2}\big) - \Phi\big((x+y)n^{1/2}\mathrm{Var}(S_n)^{-1/2}\big)\big|,\end{aligned}$$

and since $\Delta_{n,d}n^{1/2}\mathrm{Var}(S_n)^{-1/2} = \mathcal{O}(n^{-1/6+r/4})$, this is further smaller than

$$\big|\Phi(0) - \Phi(Cn^{-1/6})\big| \le Cn^{-1/6+r/4}.$$

Piecing everything together, we obtain

$$\sup_{\{x||x|\geq\theta_n\}}\left|P\left(n^{-1/2}S_n\leq x\right)-P\left(Z_n\leq x\right)\right|\leq C\left(n^{-1/10+r}+n^{-1/6+r/4}+n^{-1/2}\theta_n^{-2}\right)$$
$$\leq C\left(n^{-1/10+r}+n^{-1/3}\theta_n^{-2}\right).$$

It is clear that from the previous computations, we also have that

$$\sup_x\left|P\left(n^{-1/2}S_{n,h}\leq x\right)-P\left(Z_{n,h}\leq x\right)\right|\leq C\left(n^{-1/10+r}\right),\quad 0\leq h\leq d,$$

which completes the proof. □

3.5 Proof of the theorems

For the proof of the Theorems 3.2.1, Theorem 3.2.2 and Theorem 3.2.4, we require some additional results (Lemma 3.5.1, Lemma 3.5.2, Lemma 5.7.1 and Corollary 3.5.3), whose proof will be given at the end of this section.

Lemma 3.5.1. *Let Z_n, Z be mean zero Gaussian random variables such that*

$$\left|Var(Z_n)-Var(Z)\right|\leq n^{-q},\ q>0.$$

Then for any $\theta_n>0$

$$\sup_{\{x||x|\geq\theta_n\}}\left|P\left(Z_n\leq x\right)-P\left(Z\leq x\right)\right|\leq C\left(n^{-q/2}\theta_n^{-2}\right).$$

Let $S_{n,h}$ be defined by (3.4.1), put $\rho_{n,i,j}=\mathrm{Cov}(S_{n,i},S_{n,j})$, $\rho_{i,j}=\lim_n\mathrm{Cov}(S_{n,i},S_{n,j})$, and $r_n:=\sup_{|i-j|\geq n}|\rho_{i,j}|$, provided the limit exists.

Lemma 3.5.2. *Let $\lambda>0$, $d_n=\lambda\log n/(\log\log n)$, let $\{X_k\}_{k\in\mathbb{Z}}$ be a linear process satisfying the conditions of Theorem 3.4.1. Then the limit $\rho_{k,l}$ exists for all $k,l\in\mathbb{N}$ and*

a) $\max_{0\leq k,l\leq d_n}|\rho_{n,k,l}-\rho_{k,l}|=\mathcal{O}\left(n^{-2/3}\right)$,

b) $\min_{0\leq k\leq d_n}|\rho_{n,k,k}|\geq c>0$,

c) $r_n\to 0$,

d) $\sum_{1\leq k,l\leq n}|\rho_{k,l}|=\mathcal{O}\left(n\right)$.

Corollary 3.5.3. *Let $\{X_k\}_{k\in\mathbb{Z}}$ be a linear process satisfying the conditions of Theorem 3.4.1. Assume in addition that*

- $\max_{0\leq h\leq d_n} |s_h| \leq M_n$,
- $d_n^2 M_n^2 = \mathcal{O}\big(n^r\big)$, $0 < r < 1/6$.

Then we have

$$\sup_{\{x||x|\geq\theta_n\}} \big|P\big(n^{-1/2}S_n \leq x\big) - P\big(Z_{d_n} \leq x\big)\big| \leq C\left(n^{-1/10+r} + n^{-1/4}\theta_n^{-2}\right),$$

where S_n is as in Proposition 3.4.1, and Z_{d_n} is a mean zero Gaussian random variable with Var$(Z_{d_n}) = \sum_{i,j=0}^{d_n} s_i s_j \rho_{i,j}$.

The following result is a key ingredient, and is due to C. Deo and can be found in [38, Theorem 1].

Lemma 3.5.4. *Let $\big\{\xi_i\big\}_{i\in\mathbb{N}}$ be Gaussian process, where $\mathbb{E}\big(\xi_i\big) = 0$, $\mathbb{E}\big(\xi_i^2\big) = 1$ for all $i \in \mathbb{N}$. Put $\phi_{i,j} = Cov(\xi_i, \xi_j)$, $r_n := \sup_{|i-j|\geq n} |\phi_{i,j}|$, and let $r_1 < 1$. Assume that one of the following two conditions is satisfied:*

(a) $\sum_{n=1}^{\infty} r_n^2 < \infty$,

(b) For some $\beta > 0$, $r_n(\log n)^{2+\beta} \to 0$.

Then it holds

$$P\left(a_n^{-1}\big(\max_{1\leq h\leq n}\big|\xi_h\big| - b_n\big) \leq z\right) \to \exp(-e^{-z}),$$

where $a_n = (2\log n)^{-1/2}$ and $b_n = (2\log n)^{1/2} - (8\log n)^{-1/2}(\log\log n + 4\pi - 4)$.

Proof of Theorem 3.2.2. In this proof, C always denotes a generic, positive constant, that may vary from one formulae to another. We will only consider the case of $z \geq 0$, the other case $z < 0$ follows in the same manner. Put $\sigma_{n,h} = \mathrm{Var}(S_{n,h})$, $\sigma_h = \mathrm{Var}(\xi_h)$, and

$$\begin{aligned}
\Psi_n(z) &:= P\left(a_n^{-1}\big(\max_{0\leq h\leq d}\sigma_{n,h}^{-1/2}\big|S_{n,h}\big| - b_n\big) \leq z\right),\\
\Phi_n(z) &:= P\left(a_n^{-1}\big(\max_{0\leq h\leq d}\sigma_{h}^{-1/2}\big|\xi_h\big| - b_n\big) \leq z\right),\\
\Psi_n(A(n)) &:= P\big(\sigma_{n,1}^{-1/2}\big|S_{n,1}\big| \leq A(n), ...\sigma_{n,d}^{-1/2}\big|S_{n,d_n}\big| \leq A(n)\big),\\
\Phi_n(A(n)) &:= P\big(\sigma_{1}^{-1/2}\big|\xi_1\big| \leq A(n), ...\sigma_{d}^{-1/2}\big|\xi_d\big| \leq A(n)\big),
\end{aligned}$$

where $\big\{\xi_h\big\}_{h\in\mathbb{N}}$ is a mean zero Gaussian process with $\mathrm{Cov}(\xi_i, \xi_j) = \rho_{i,j}$, $i, j \geq 0$. Put $A(n) = za_n + b_n$, $\theta_n = n^{-1/24}$, $p = \min\{1/10 - r, 1/48\}$ and $M(n) = n^{p/4}$, and note that $d_n^4 M(n)^4 = n^{3p/2} \leq n^{1/8}$. Then by Corollary 3.5.3 and Theorem 3.3.1 we have

$$\big|\Psi_n(z) - \Phi_n(z)\big| = \big|\Psi_n(A(n)) - \Phi_n(A(n))\big| = \mathcal{o}(1), \tag{3.5.1}$$

hence it suffices tho show that

$$\Phi_n(z) \to \exp(-e^{-z}).$$

Per assumption, we have $\sum_{i=0}^{\infty} \alpha_i^2 > \sup_{h\geq 1} \sum_{i=0}^{\infty} |\alpha_i \alpha_{i+h}|$, hence we deduce from the proof of Lemma 3.5.2 that

$$\sup_{k,l:|k-l|=1} \big|(\sigma_k \sigma_l)^{-1} \rho_{k,l}\big| < 1,$$

thus the claim follows from Lemma 3.5.2 and Lemma 3.5.4. □

Proof of Theorem 3.2.1. Put $\sigma_{n,h} = \operatorname{Var}(S_{n,h})$, $\sigma_h = \operatorname{Var}(\xi_h)$, and denote with $\mathcal{L}(.,.)$ the Lévy distance, i.e.

$$\mathcal{L}(X,Y) = \inf\{\epsilon > 0 : F(x) \leq G(x+\epsilon) + \epsilon \text{ and } G(x) \leq F(x+\epsilon) + \epsilon, \text{ for all } x\},$$

where F and G are distribution functions of the r.v. X and Y. By assumption, we have

$$n^{-1/2}\Big\{\big(\widehat{\phi}_0, \widehat{\phi}_1, ..., \widehat{\phi}_d\big)^{\mathbf{T}} - \big(\phi_0, \phi_1, ..., \phi_d\big)^{\mathbf{T}}\Big\} \xrightarrow{d} \big\{\xi_h\big\}_{0\leq h\leq d}$$

for any finite $d \in \mathbb{N}$, where $\big\{\xi_h\big\}_{h\in\mathbb{N}}$ is a mean zero Gaussian process. On the other hand, it follows from Lemma 5.7.1 that

$$P\left(a_n^{-1}\big(\max_{0\leq h\leq n} \sigma_h^{-1/2}\big|\xi_h\big| - b_n\big) \leq z\right) \to \exp(-e^{-z}),$$

where $a_n = (2\log n)^{-1/2}$ and $b_n = (2\log n)^{1/2} - (8\log n)^{-1/2}(\log\log n + 4\pi - 4)$. Let V be a r.v. with cdf $F(z) = \exp(-e^{-z})$, and put

$$E_{X,d_n} = a_{d_n}^{-1}\big(\max_{0\leq h\leq d_n} \sigma_{d_n,h}^{-1/2}\big|S_{n_{d_n},h}\big| - b_{d_n}\big),$$
$$E_{\xi,d_n} = a_{d_n}^{-1}\big(\max_{0\leq h\leq d_n} \sigma_h^{-1/2}\big|\xi_h\big| - b_{d_n}\big).$$

Then for any $\epsilon > 0$, we can chose a d_n and a corresponding n_{d_n} such that

$$\mathcal{L}\left(V, E_{X,d_n}\right) \leq \mathcal{L}\left(V, E_{\xi,d_n}\right) + \mathcal{L}\left(E_{X,d_n}, E_{\xi,d_n}\right) \leq \epsilon,$$

hence the claim follows. □

Proof of Theorem 3.2.4. Let $A_d^{(n)} = \big\{1 \leq j \leq d \;\mid |x_j| \geq \log n\big\}$, $1 \leq j \leq d$ and denote with $|A_d^{(n)}|$ the cardinality of the set $A_d^{(n)}$. Note that the sets $A_d^{(n)}$ may substantially vary for each d. Let $A_d^{(n),c}$ be the complement of $A_d^{(n)}$ with respect to the total set $\{1, ..., d\}$. We then have

$$\begin{aligned}&\big|P\big(\bigcap_{j\in A_d^{(n),c}} \{\zeta_{n,j} \leq x_j\} \cap \bigcap_{j\in A_d^{(n)}} \{\zeta_{n,j} \leq x_j\}\big) - P\big(\bigcap_{j\in A_d^{(n),c}} \{\zeta_{n,j} \leq x_j\}\big)\big| \\ &\leq \sum_{j\in A_d^{(n)}} P\big(\zeta_{n,j} > |x_j|\big) \leq \mathcal{X}\, d + C\, d \exp\big(-(\log n)^2/2\big),\end{aligned}$$

and the same bound is valid for the vector $\big(\xi_1, ..., \xi_d\big)^T$. Since $d_n = o\,(\log n)$, the above bound converges to zero as n increases, hence it suffices to establish the claim for $\max_{1\leq j\leq d} |x_j| \leq \log n$. To this end, note that under the assumptions of Theorem 3.3.1 we have the more general conclusion

$$\begin{aligned}&\big|P\big(-b < \zeta_{n,1} \leq x_1, \ldots, -b < \zeta_{n,d} \leq x_d\big) - P\big(-b < \xi_1 \leq x_1, \ldots, -b < \xi_d \leq x_d\big)\big| \\ &\leq C(x + b + d\log d)^d \left(M\mathcal{X} + \theta + d\, M^{-1/4}\right) + RM^{-1/2} \exp(-M(\sqrt{2}R)^{-1}) \\ &+ Cd^{-1} \sum_{j=1}^{d} \exp(-x_j^2/2) := \epsilon,\end{aligned} \tag{3.5.2}$$

where $x = \max_j\{x_j\}$. In addition, observe that

$$\begin{aligned}&\big|P\big(\zeta_{n,1} \leq x_1, ..., \zeta_{n,d} \leq x_d\big) - P\big(-b < \zeta_{n,1} \leq x_1, ..., -b < \zeta_{n,1} \leq x_d\big)\big| \\ &\leq d \max_{1\leq j\leq d} P\big(|\zeta_{n,j}| \geq b\big) \leq d\mathcal{X} + CdP\big(|\xi_j| \geq b\big) \\ &\leq C\, d\left(\mathcal{X} + \exp(-b^2/2)\right).\end{aligned}$$

We choose now $d = d_n = \lambda \log n / \log\log n$, $\lambda > 0$ and $b = b_n = d_n \log d_n$. Suppose now that

$$\limsup_{d\to\infty} d^{-1} \sum_{j=1}^{d} \exp(-x_j^2/2) = 0. \tag{3.5.3}$$

Then by virtue of Theorem 3.4.1 and arguing as in the proof of Theorem 3.2.2, one easily verifies that $\lim_{n\to\infty} \epsilon_n = 0$ and $\lim_{n\to\infty} d_n\, \mathcal{X}_n + d_n \exp(-b_n^2/2) = 0$ for appropriately increasing sequences $\mathcal{X}_n$, θ_n^{-1}, M_n and sufficiently small $\lambda > 0$, where ϵ_n is the quantity in formula (3.5.2) with variable parameters $d_n, \mathcal{X}_n, M_n, , R_n, \theta_n$.

This leaves us to consider the case were (3.5.3) is violated. This implies that we must have

$$\limsup_{d\to\infty} d^{-1}\sum_{j=1}^{d}\exp(-x_j^2/2) \geq \delta > 0. \tag{3.5.4}$$

Define the sets $B_d^{(n)} = \{1 \leq j \leq d \ \mid \exp(-x_j^2/2) \geq \delta/2\}$, and note that these sets may substantially differ for each d. Due to (3.5.4), we have that $|B_d^{(n)}| \to \infty$ as n (and hence also d) increases, where $|B_d^{(n)}|$ denotes the cardinality of the set. We thus obtain

$$\begin{aligned}
&\big|P\big(\zeta_{n,1} \leq x_1, ..., \zeta_{n,d} \leq x_d\big) - P\big(\xi_1 \leq x_1, ..., \xi_d \leq x_d\big)\big| \\
&\leq P\big(\bigcap_{j\in B_d^{(n)}} \{\zeta_{n,j} \leq x_j\}\big) + P\big(\bigcap_{j\in B_d^{(n)}} \{\xi_j \leq x_j\}\big) \\
&\leq P\big(\bigcap_{j\in B_d^{(n)}} \{\zeta_{n,j} \leq \sqrt{-2\ln\delta/2}\}\big) + P\big(\bigcap_{j\in B_d^{(n)}} \{\xi_j \leq \sqrt{-2\ln\delta/2}\}\big).
\end{aligned}$$

Since $|B_d^{(n)}| \to \infty$ as n (and hence also d) increases, [39, Theorem] implies that $P\big(\bigcap_{j\in B_d^{(n)}}\{\xi_j \leq \sqrt{-2\ln\delta/2}\}\big) \to 0$. On the other hand, it follows from the proof of Theorem 3.2.1 that there exists a sequence of subsets $B_d^{(n),*} \subset B_d^{(n)}$, such that $\lim_n |B_d^{(n),*}| \to \infty$ and

$$\lim_n \big|P\big(\bigcap_{j\in B_d^{(n),*}} \{\zeta_{n,j} \leq \sqrt{-2\ln\delta/2}\}\big) - P\big(\bigcap_{j\in B_d^{(n),*}} \{\xi_j \leq \sqrt{-2\ln\delta/2}\}\big)\big| \to 0.$$

Since $P\big(\bigcap_{j\in B_d^{(n)}}\{\zeta_{n,j} \leq \sqrt{-2\ln\delta/2}\}\big) \leq P\big(\bigcap_{j\in B_d^{(n),*}}\{\zeta_{n,j} \leq \sqrt{-2\ln\delta/2}\}\big)$, we conclude that

$$\lim_n \big|P\big(\zeta_{n,1} \leq x_1, ..., \zeta_{n,d} \leq x_d\big) - P\big(\xi_1 \leq x_1, ..., \xi_d \leq x_d\big)\big| = 0,$$

if (3.5.3) is violated. Piecing everything together, the claim follows. This completes the proof of Theorem 3.2.4. □

Proof of Lemma 3.5.1. Put $\sigma_n = \mathrm{Var}(Z_n)$, $\sigma = \mathrm{Var}(Z)$, $r_n := \big|\sigma_n - \sigma\big|$, $L := r_n/(\sigma\sigma_n)$, and

$$\phi_\sigma(x) = (\sqrt{2\pi\sigma})^{-1/2}\exp(-x^2/(2\sigma)).$$

Then

$$
\begin{aligned}
&\Big|\int_0^\infty \Big((\sqrt{2\pi\sigma})^{-1/2}\exp(-x^2/(2\sigma)) - (\sqrt{2\pi\sigma_n})^{-1/2}\exp(-x^2/(2\sigma_n))\Big)\,dx\Big| \\
&\leq \int_0^\infty \varphi_\sigma(x)\sqrt{1+\frac{r_n}{\sigma_n}}\big|\exp(x^2 r_n/(\sigma\sigma_n) - 1\big|dx + \Big|\sqrt{1+\frac{r_n}{\sigma_n}} - 1\Big| \\
&\leq \big|r_n\sigma_n^{-1}\big| + \sqrt{1+\frac{r_n}{\sigma_n}}\left(\int_0^{L^{-1}} \varphi_\sigma(x)x^2 L dx + 2\int_{L^{-1}}^\infty \varphi_\sigma(x)\exp(x^2 L)dx\right) \\
&\leq \big|r_n\sigma_n^{-1}\big| + \sqrt{1+\frac{r_n}{\sigma_n}}\left(\sigma L + (1-2\sigma\sigma_n^{-1}L)\int_{L^{-1}}^\infty \varphi_\sigma(x)dx\right) := A. \qquad (3.5.5)
\end{aligned}
$$

If we have that $\sigma_n \geq 2\sqrt{r_n} = 2n^{-q}$, $q > 0$, an application of Lemma 3.3.4 gives the following upper bound for A.

$$
A \leq 4\left(\sqrt{r_n} + \int_{L^{-1}}^\infty \varphi_\sigma(x)dx\right) \leq 5\sqrt{r_n} = 5n^{-q/2}.
$$

On the other hand, if $\sigma_n < 2\sqrt{r_n}$, we obtain as in the proof of Theorem 3.4.1

$$
\sup_{\{x||x|\geq\theta_n\}} \big|P\big(Z_n \leq x\big) - P\big(Z \leq x\big)\big| \leq C\left(\sqrt{r_n}\theta_n^{-2}\right),
$$

which completes the proof. □

Proof of Lemma 3.5.2. Put $\mathbb{E}\big(\epsilon^4\big) = \eta$ and let $\Lambda_{n,h} = n^{-1/2}\sum_{j=0}^n\big(X_j X_{j+h} - \phi_h\big)$. Since $d_n = \mathcal{O}(\log n)$, we have that

$$
\begin{aligned}
\mathbb{E}\big(S_{n,k}, S_{n,l}\big) &= \mathbb{E}\big(\Lambda_{n,k}, \Lambda_{n,l}\big) + \mathcal{O}\left(n^{-1}(\log n)^2\right) \\
&= \mathbb{E}\big(\Lambda_{n,k}, \Lambda_{n,l}\big) + \mathcal{O}\left(n^{-2/3}\right),
\end{aligned}
$$

and it thus suffices to consider $\mathbb{E}\big(\Lambda_{n,k}, \Lambda_{n,l}\big)$. Due to [25, section 7.2], it holds that

$$
\mathbb{E}\big(\Lambda_{n,k}, \Lambda_{n,l}\big) = \sum_{|m|<n} \frac{n-|m|}{n} T_m, \qquad (3.5.6)
$$

where

$$
T_m = \phi_m\phi_{m+k-l} + \phi_{m+k}\phi_{m-l} + (\eta-3)\sum_i \alpha_i\alpha_{i+k}\alpha_{i+m}\alpha_{i+m+l}.
$$

In particular, it holds that

$$
\rho_{k,l} = \lim_n \rho_{n,k,l} = (\eta-3)\phi_k\phi_l + \sum_{m=-\infty}^{\infty}\big(\phi_m\phi_{m+k-l} + \phi_{m+k}\phi_{m-l}\big). \qquad (3.5.7)
$$

We will now show a). Since

$$\int_1^\infty \frac{1}{\big(x(x+m)^{3/2}\big)}dx = \mathcal{O}\left((m\sqrt{m})^{-1}\right), \tag{3.5.8}$$

we have that

$$\phi_m = \sum_{i=0}^{\infty} \alpha_i \alpha_{i+m} = \mathcal{O}\left((m\sqrt{m})^{-1}\right). \tag{3.5.9}$$

We thus obtain for $0 \leq k,l \leq d_n$

$$\sum_{|m|>n^{1/3}} \left(\phi_m \phi_{m+k-l} + \phi_{m+k}\phi_{m-l}\right) = \mathcal{O}\left(n^{-2/3}\right).$$

Using this, we obtain for $0 \leq k,l \leq d_n$ the decomposition

$$\rho_{k,l} = \sum_{|m|=0}^{\infty} T_m = \sum_{|m|\leq n^{1/3}} T_m + \mathcal{O}\left(n^{-2/3}\right), \tag{3.5.10}$$

which yields

$$\max_{0\leq k,l\leq d_n} |\rho_{n,k,l} - \rho_{k,l}| = \mathcal{O}\left(n^{-2/3}\right),$$

and thus a). b) follows from (6.5.16), while c) follows from (6.5.24) and (3.5.8). Finally, d) follows from (6.5.24) and (3.5.8). □

Proof of Corollary 3.5.3. By [131, Theorem 1], we have for any fixed $d \in \mathbb{N}$, that

$$\big\{S_{n,h}\big\}_{0\leq h\leq d} \xrightarrow{d} \big\{\xi_h\big\}_{0\leq h\leq d},$$

where $\big\{\xi_h\big\}_{0\leq h\leq d}$ is a Gaussian process with $\mathrm{Cov}(\xi_i,\xi_j) = \rho_{i,j}$, $0 \leq i,j \leq d$. Put $S_{n,h,s} = n^{-1/2}\sum_{k=0}^n s_h\big(X_k X_{k+h} - \phi_h\big)$, and $\xi_{h,s} = s_h \xi_h$, where $\max_{0\leq h\leq d}|s_h| \leq M$. Then Proposition 3.5.2 implies

$$\Big|\mathrm{Var}\big(\sum_{h=0}^{d} S_{n,h,s}\big) - \mathrm{Var}\big(\sum_{h=0}^{d} \xi_{h,s}\big)\Big| \leq n^{-2/3} d^2 M^2 \leq n^{-1/2},$$

and the claim follows from Lemma 3.5.1 and Theorem 3.4.1. □

Chapter 4

Obtaining a larger Growth Rate

4.1 Introduction

Let us reconsider the problem discussed in the previous chapter, i.e; we want to study the asymptotic behavior, as $n \to \infty$, of the object

$$\max_{0 \leq h \leq d_n} \left|\widehat{\phi}_{n,h} - \phi_h\right|. \tag{4.1.1}$$

We essentially established a growth rate of $d_n = \scriptstyle\mathcal{O}\displaystyle(\log n)$ using the quantiative Cramér-Wold device discussed in Section 3.3. The prove was based on martingale approximations and characteristic functions. The advantage of this approach was that it allows for very general processes. On the other hand, as for instance Senatov [110] pointed out, using characteristic functions as a tool in higher dimensions cannot lead to optimal results. Moreover, this is also the case when using martingale approximations for obtaining Gaussian approximations (cf. Chapter 2). Narrowing down the class of potential processes allows to use different tools which yield much larger growth rates for d_n, more precisely, we will establish analogues of Theorem 3.2.2 with the growth rate $d_n = \mathcal{O}\left(n^{1/6}(\log n)^{-\alpha/3}\right)$, for some $\alpha > 3$. To this end, note that in case of linear processes, one may write

$$\widehat{\phi}_{n,h} = \frac{1}{n}\sum_{k=1}^{n} g_h(\epsilon_{k+h}, \epsilon_{k+h-1}, \ldots), \tag{4.1.2}$$

where the functions g_h are defined via

$$g_h(\epsilon_{k+h}, \epsilon_{k+h-1}, \ldots) := X_{k+h}X_k = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \alpha_i \alpha_j \epsilon_{k+h-i}\epsilon_{k-j}. \tag{4.1.3}$$

Instead of considering the particular case of the covariance estimators $\left\{\widehat{\phi}_{n,h}\right\}_{1 \leq h \leq d_n}$, we will generally discuss functions $g_h(\epsilon_{k+h}, \epsilon_{k+h-1}, \ldots)$ under a mild dependence condition, that is easy to verify.

4.2 Main results

Let $\{\epsilon_k\}_{k\in\mathbb{Z}}$ be a sequence of zero mean IID random variables. In the sequel, we will consider the array of zero mean random variables $X_{k,h} = g_h(\epsilon_{k+h}, \epsilon_{k+h-1}, ...)$, $k \in \mathbb{Z}$, $1 \leq h \leq d_n$, where g_h are measurable functions such that $X_{k,h}$ are proper random variables. For convenience, we will also write $g_h(\xi_{k+h})$, with $\xi_k = (\epsilon_k, \epsilon_{k-1}, ...)$. The class of processes that fits into this framework is large, and contains a variety of linear and nonlinear processes including ARCH, GARCH and related processes, see for instance [52, 104, 120, 121]. A very nice feature of the representation given above is that it allows to give simple, yet very efficient and general dependence conditions. Following Wu [133], let $\{\epsilon'_k\}_{k\in\mathbb{Z}}$ be an independent copy of $\{\epsilon_k\}_{k\in\mathbb{Z}}$ on the same probability space, and define the 'filters' $\xi_{k,}^{(m,')}$, $\xi_{k,h}^{(m,*)}$ as $\xi_{k,h}^{(m,')} = (\epsilon_{k+h}, \epsilon_{k+h-1}, ..., \epsilon'_{k-m}, \epsilon_{k-m-1}, ...)$ and $\xi_{k,h}^{(m,*)} = (\epsilon_{k+h}, \epsilon_{k+h-1}, ..., \epsilon_{k-m}, \epsilon'_{k-m-1}, ...)$. We put $\xi'_{k,h} = \xi_{k,h}^{(0,')} = (\epsilon_{k+h}, \epsilon_{k+h-1}, ..., \epsilon'_0, \epsilon_{-1}, ...)$ and $\xi^*_{k,h} = \xi_{k,h}^{(0,*)} = (\epsilon_{k+h}, \epsilon_{k+h-1}, ..., \epsilon_0, \epsilon'_{-1}, ...)$. In analogy, we put $X_{k,h}^{(m,')} = g_h\big(\xi_{k,h}^{(m,')}\big)$ and $X_{k,h}^{(m,*)} = g_h\big(\xi_{k,h}^{(m,*)}\big)$, in particular we have $X'_{k,h} = X_{k,h}^{(0,')}$ and $X^*_{k,h} = X_{k,h}^{(0,*)}$.

As a dependence measure, one may now consider the quantities $\big\|X_{k,h} - X'_{k,h}\big\|_p$ or $\big\|X_{k,h} - X^*_{k,h}\big\|_p$, $p \geq 1$, where $\|.\|_p^p = \mathbb{E}(|.|^p)$. For example, if we define the linear processes $X_{k,h} = \sum_{i=0}^{\infty} \alpha_{i,h}\epsilon_{k-i}$, the condition

$$\sum_{k=0}^{\infty} \big\|X_{k,h} - X'_{k,h}\big\|_2 < \infty \tag{4.2.1}$$

is valid if $\sum_{i=0}^{\infty} |\alpha_{i,h}| < \infty$, provided that $\mathbb{E}\big(\epsilon_0^2\big) < \infty$. Dependence conditions of the type of (4.2.1) are often quite general and easy to verify in many cases, see for instance [15, 34, 42, 130] and the references there. For fixed h, we will always express the dependence condition in terms of $\big\|X_{k,h} - X'_{k,h}\big\|_p$ in the sequel (cf. Assumption 4.2.1).

For $1 \leq h \leq d_n$, we denote the partial sums with

$$S_{n,h} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_{k,h},$$

and define the d_n-dimensional vector

$$\mathbf{S}_n = \big(S_{n,1}, S_{n,2}, ..., S_{n,d_n}\big)^T.$$

In order to state the main results, we require the following additional notation. We formally define the limit covariance

$$\varphi_{h,l} = \limsup_{n\to\infty} \left|\mathbb{E}\big(S_{n,h}^{(m_n)}, S_{n,l}^{(m_n)}\big)\right|, \quad h,l \in \mathbb{N}, \tag{4.2.2}$$

and for $k \geq 0$ the total limit covariance

$$r_k = \sup_{|h-l|\geq k} |\varphi_{h,l}|. \tag{4.2.3}$$

The main results are derived under the following dependence and regularity conditions.

Assumption 4.2.1. *Assume that for some $p \geq 4$ it holds that*

$$(i)\ \max_{1\leq h\leq d_n} \left\|X_{k,h} - X_{k,h}'\right\|_p = \mathcal{O}\left(k^{-\beta}\right), \quad \beta > 3/2,$$

$$(ii)\ \liminf_{n\to\infty} \min_{1\leq h\leq d_n} Var\big(S_{n,h}\big) > 0.$$

Note that (i) implies in particular that $\max_{1\leq h\leq d_n}\left\|X_{k,h}\right\|_p < \infty$, this follows for instance from Lemma 4.3.3. We also mention that instead of (i), one may also use the condition

$$\max_{1\leq h\leq d_n} \left\|X_{k,h} - X_{k,h}^{(h,')}\right\|_p = \mathcal{O}\left(k^{-\beta}\right), \quad \beta > 3/2. \tag{4.2.4}$$

We can now give the main result.

Theorem 4.2.2. *Assume that $r_1 < 1$ and that $r_n(\log n)^{2+\gamma} \to 0$, for some $\gamma > 0$. Then, if in addition Assumption 4.2.1 is valid, we have for $z \in \mathbb{R}$*

$$\lim_{n\to\infty} P\left(a_n^{-1}\big(\max_{1\leq h\leq d_n} \left|S_{n,h}\, Var\big(S_{n,h}\big)^{-1/2}\right| - b_n\big) \leq z\right) = \exp(-e^{-z}),$$

where $a_n = (2\log d_n)^{-1/2}$ and $b_n = (2\log d_n)^{1/2} - (8\log d_n)^{-1/2}(\log\log d_n + 4\pi - 4)$, and $d_n = \mathcal{O}\left(n^{1/6}(\log n)^{-\alpha/3}\right)$, for some $\alpha > 4$.

The conditions needed to prove the above theorem can essentially be divided into two classes. Assumption 4.2.1 is needed to allow for a suitable Gaussian approximation for the random vector $\mathbf{S}_n$, whereas condition (a) or (b) is required to establish weak convergence to an extreme-value type distribution. On the whole, the above conditions are quite general and, as pointed out earlier, include many weakly dependent processes.

In general, the variance $\mathrm{Var}\big(S_{n,h}\big)$ is not known in practice and needs to be estimated. One may hope that the above Theorems are still valid if one replaces $\mathrm{Var}\big(S_{n,h}\big)$ with the corresponding estimates $\widehat{\mathrm{Var}}\big(S_{n,h}\big)$, and indeed this is the case if the following mild condition is imposed on potential variance estimators $\widehat{\psi}_h$.

Assumption 4.2.3.

$$\limsup_{n\to\infty} P\big(\max_{1\le h\le d_n}\big|\widehat{Var}\big(S_{n,h}\big) - Var\big(S_{n,h}\big)\big| > (\log n)^{-\alpha}\big) = 0,\ \alpha > 1.$$

We then have that

Theorem 4.2.4. *Assume that the assumptions of Theorem 4.2.2 are satisfied. If in addition Assumption 4.2.3 is valid, we have*

$$\lim_{n\to\infty} P\left(a_n^{-1}\big(\max_{1\le h\le d_n}\big|S_{n,h}\widehat{Var}\big(S_{n,h}\big)^{-1/2}\big| - b_n\big) \le z\right) = \exp(-e^{-z}),$$

where $a_n = (2\log d_n)^{-1/2}$ *and* $b_n = (2\log d_n)^{1/2} - (8\log d_n)^{-1/2}(\log\log d_n + 4\pi - 4)$, *and* $d_n = \mathcal{O}\left(n^{1/6}(\log n)^{-\alpha/3}\right)$, *for some* $\alpha > 3$.

The literature (cf. [4, 25, 60]) provides many potential candidates to estimate the long run variance $\sigma_h^2 = \lim_{n\to\infty}\operatorname{Var}\big(S_{n,h}\big)$. A popular estimator is Bartlett's estimator, or more general, estimators of the form

$$\widehat{\sigma}_h^2 = \sum_{|j|\le r}\omega(k/r)\widehat{\gamma}_{j,h} \tag{4.2.5}$$

with weight function $\omega(x)$, where $\gamma_{j,h} = \mathbb{E}\big(X_{0,h}X_{j,h}\big)$ and $\widehat{\gamma}_{j,h} = n^{-1}\sum_{k=1}^{n-j}X_{k,h}X_{k+j,h}$. Considering the triangular weight function $\omega(x) = 1-|x|$ for $|x|\le 1$ and $\omega(x) = 0$ for $|x| > 1$, one recovers the Bartlett estimator in (4.2.5). One may also use the plain estimate

$$\widehat{\sigma}_h^2 = \widehat{\gamma}_{0,h} + 2\sum_{i=1}^{l_n}\widehat{\gamma}_{i,h}, \tag{4.2.6}$$

see for instance [112, 113]. In particular, Wu [131, Proposition 1] provides the following result, which we have reformulated for our setting.

Proposition 4.2.5. *Let* $l_n \in \mathbb{N}$, $l_n \to \infty$ *as* n *increases with* $l_n = \mathcal{O}\left(\sqrt{n}\big(d_n(\log n)^{\alpha}\big)^{-1}\right)$, *where* $\alpha > 1$. *If Assumption 4.2.1 holds, then*

$$\limsup_{n\to\infty} P\big(\max_{1\le h\le d_n}\big|\widehat{Var}\big(S_{n,h}\big) - Var\big(S_{n,h}\big)\big| > (\log n)^{-\alpha}\big) = 0,\ \alpha > 1.$$

Consequently, Theorem 4.2.4 is valid if one uses the variance estimator given in (4.2.6).

Let us briefly reconsider the case of the covariance estimators $\widehat{\phi}_{n,h}$ in case of linear processes $\{L_k\}_{k\in\mathbb{Z}}$, more precisely, let $L_k = \sum_{i=0}^{\infty}\alpha_i\epsilon_{k-i}$ be a linear process, where $\{\epsilon_k\}_{k\in\mathbb{Z}}$ is a mean zero IID sequence. We are interested in establishing simultaneous confidence bands for the covariances $\phi_h = \mathbb{E}\big(L_h L_0\big)$, which, as mentioned in Chapters 1 and 3, is an important issue. To this end, let

$$S_{n,h} = \frac{1}{\sqrt{n}}\sum_{k=0}^{n-h}\big(L_k L_{k+h} - \phi_h\big), \quad 0 \leq h \leq d_n,$$

where $\phi_h = \sum_{i=0}^{\infty}\alpha_i\alpha_{i+h}$.

Then the following variant of Theorem 4.2.2 holds.

Theorem 4.2.6. *Assume that* $0 < \|\epsilon_0\|_8 < \infty$*, and that*

- $|\alpha_i| > 0$ *for at least one* $i \in \mathbb{N}_0$*,*
- $|\alpha_i| = \mathcal{O}\left(i^{-\beta}\right)$*, with* $\beta > 3/2$*.*

Then for $z \in \mathbb{R}$ *it holds that*

$$\lim_{n\to\infty} P\left(a_n^{-1}\big(\max_{1\leq h\leq d_n}\big|S_{n,h}\,Var\big(S_{n,h}\big)^{-1/2}\big| - b_n\big) \leq z\right) = \exp(-e^{-z}),$$

where $a_n = (2\log d_n)^{-1/2}$ *and* $b_n = (2\log d_n)^{1/2} - (8\log d_n)^{-1/2}(\log\log d_n + 4\pi - 4)$*, and* $d_n = \mathcal{O}\left(n^{1/6}(\log n)^{-\alpha/3}\right)$*, for some* $\alpha > 3$*.*

Naturally, an analogue version of Theorem 4.2.4 is valid, and Proposition 4.2.5 is also valid under the conditions given in Theorem 4.2.6.

4.3 Proofs

The proof of Theorem 4.2.2 is developed in a series of Lemmas. To this end, we formally introduce the following notation. Let $\{U_k\}_{k\in\mathbb{Z}}$ be a stationary process, adapted to some filtration $\mathcal{F}_k$. We define the projection operator $\mathcal{P}_k U_i$ as

$$\mathcal{P}_k U_i = \mathbb{E}\big(U_i \mid \mathcal{F}_k\big) - \mathbb{E}\big(U_i \mid \mathcal{F}_{k-1}\big),\ k, i \in \mathbb{Z}.$$

Let us consider the partial sums $\Lambda_n = \sum_{i=1}^{n} U_i$. Many of the following results are based on martingale approximations for Λ_n. Various different approximating martingale sequences have been proposed in the literature, see for instance [70, 100, 101, 128, 130] and the references there. In our setting, the following approximating sequence of martingales, introduced by Gordin [55], is appropriate. Define the martingale $\{M_k\}_{k\in\mathbb{Z}}$ and the remainder process $\{R_k\}_{k\in\mathbb{Z}}$ as

$$M_k = \sum_{i=1}^{\infty} \mathbb{E}\big(U_i \mid \mathcal{F}_k\big), \quad R_k = \sum_{i=k+1}^{\infty} \mathbb{E}\big(U_i \mid \mathcal{F}_k\big).$$

Note that both processes are stationary (if the above series converge), this follows for instance from Lemma 3.84 in [70]. We can now decompose Λ_n as

$$\Lambda_n = M_n - M_1 + M_1 - R_n, \tag{4.3.1}$$

where we note that $M_1 = R_0$. We will frequently use the above decomposition for varying underlying processes $\{U_k\}_{k\in\mathbb{Z}}$. We will, however, abuse the notation by always writing M_n, R_n for the corresponding martingale and remainder process, regardless of the specific process $\{U_k\}_{k\in\mathbb{Z}}$. In addition, we define the martingale differences

$$D_k = M_{k+1} - M_k. \tag{4.3.2}$$

Another essential tool will be approximations with m_n-dependent random variables. To this end, we introduce the following notation. Let $\{\epsilon_k\}_{k\in\mathbb{Z}}$ be a sequence of zero mean IID random variables. We define the following two σ-algebras

$$\mathcal{F}_k = \sigma\big(\epsilon_j,\, j \le k\big), \quad \mathcal{F}_{k-m}^{k+m} = \sigma\big(\epsilon_j,\, k-m \le j \le k+m\big), \tag{4.3.3}$$

and for $1 \le h \le d_n$ the random variables

$$\begin{aligned} Y_{k,h}^{(\le m)} &= \mathbb{E}\big(X_{k,h} \mid \mathcal{F}_{k-m}^{k+m}\big),\\ Y_{k,h}^{(>m)} &= X_{k,h} - Y_{k,h}^{(\le m)} = X_{k,h} - \mathbb{E}\big(X_{k,h} \mid \mathcal{F}_{k-m}^{k+m}\big). \end{aligned} \tag{4.3.4}$$

In addition to the conditional approximations defined above, we denote the corresponding partial sums as

$$S_{n,h}^{(m)} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Y_{k,h}^{(\le m)}, \quad S_{n,h}^{(>m)} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Y_{k,h}^{(>m)}, \quad 1 \le h \le d_n,$$

and the random vector

$$\mathbf{S}_n^{(m)} = \big(S_{n,1}^{(m)}, S_{n,2}^{(m)}, \ldots, S_{n,d_n}^{(m)}\big)^T, \quad \mathbf{S}_n^{(>m)} = \big(S_{n,1}^{(>m)}, S_{n,2}^{(>m)}, \ldots, S_{n,d_n}^{(>m)}\big)^T.$$

When dealing with extreme value-type distributions, the following notation also turns out to be quite convenient. For $z \in \mathbb{R}$ we put

$$u_{d_n}(z) = a_n z + b_n,$$

where a_n, b_n are as in Theorem 4.2.2, and we usually write u_{d_n} instead of $u_{d_n}(z)$ if the dependence on z is not important. Note in particular the following asymptotic expansion

$$u_{d_n}^2(z) = u_{d_n}^2 = 2 \log d_n - \log \log d_n + \mathcal{O}(1), \tag{4.3.5}$$

and, for a standard Gaussian random variable Z the tail estimate

$$P\big(|Z| \geq u_{d_n}(z)\big) = \frac{z}{d_n} + o\left(d_n^{-1}\right), \tag{4.3.6}$$

which we will extensively use in the sequel. Throughout the proofs, we will always assume that the sequences $m = m_n$ and d_n satisfy

- $d_n = \mathcal{O}\left(n^{1/6}(\log n)^{-\alpha/3}\right)$, for some $\alpha > 3$.
- $2d_n \leq m_n = \mathcal{O}\left(n^{1/6}(\log n)^{-\alpha/3}\right)$, where $\alpha > 3$ is as above.

The proof of Theorem 4.2.2 essentially consists of three steps. In the first step we collect some preliminary results concerning the magnitude of $S_{n,h}^{(m_n)}, S_{n,h}^{(>m_n)}$, and related quantities. The second step consists of various truncation arguments, which essentially allows to consider m_n-dependent sequences $\{X_{k,h}\}_{k \in \mathbb{N}}$. Finally, the third steps presents an appropriate Gaussian approximation, which allows to apply results from the literature to deduce the result.

4.3.1 Step one - preliminary results

Lemma 4.3.1. *Assume that Assumption 4.2.1 is valid. Then*

$$\max_{1 \leq h \leq d_n} \big\| S_{n,h}^{(>m_n)} \big\|_2 = \mathcal{O}\left(m_n^{3/2-\beta} + n^{-1/2}\right).$$

Remark 4.3.2. Note that the assumption $d_n \geq n^{\delta}$ for some $\delta > 0$ implies that we have in particular

$$\max_{1 \leq h \leq d_n} \big\| S_{n,h}^{(>m_n)} \big\|_2 = o\left((\log n)^{-q}\right), \tag{4.3.7}$$

for any $q > 0$, since $d_n < m_n$.

Proof of Lemma 4.3.1. First note that the Cauchy-Schwarz inequality implies

$$\big\|X_{k,h} - X'_{k,h}\big\|_2 \leq \big\|X_{k,h} - X'_{k,h}\big\|_4 = \mathcal{O}\left(k^{-\beta}\right). \tag{4.3.8}$$

Using that $S_{n,h}^{(>m_n)} = M_n - M_1 - (R_n - R_0)$, we obtain

$$\big\|S_{n,h}^{(>m_n)}\big\|_2 \leq 2\big\|R_0\big\|_2 + \big\|M_n\big\|_2.$$

Using the orthogonality of the martingale increments D_k we have

$$\big\|M_n\big\|_2 = \big\|\sum_{k=0}^{\infty} \mathcal{P}_0(Y_{k,h}^{(>m_n)})\big\|_2 \leq \sum_{k=0}^{\infty}\big\|\mathcal{P}_0(Y_{k,h}^{(>m_n)})\big\|_2.$$

Since $\mathbb{E}\big\{\mathbb{E}\big(X_{k,h} \mid \mathcal{F}_{k-m}^{k+m}\big) \mid \mathcal{F}_0\big\} = 0$ for $k > m$, we obtain that

$$\big\|\mathcal{P}_0(Y_{k,h}^{(>m_n)})\big\|_2 \leq 2\min\big\{\big\|Y_{k,h}^{(>m_n)}\big\|_2, \big\|X_{k,h} - X'_{k,h}\big\|_2\big\}. \tag{4.3.9}$$

Let $\mathcal{F}'_k = \sigma(\epsilon'_k, \epsilon'_{k-1}, ...)$. Then for any $p \geq 1$ we have by the triangular and Jensen's inequality

$$\begin{aligned}\big\|Y_{k,h}^{(>m_n)}\big\|_p &\leq \big\|X_{k,h} - X_{k,h}^{(m_n,*)}\big\|_p + \big\|\mathbb{E}\big(X_{k,h} - X_{k,h}^{(m_n,*)} \mid \sigma\big(\mathcal{F}_{k-m_n}^{k+m_n} \cup \mathcal{F}'_{k-m_n-1}\big)\big)\big\|_p \\ &\leq 2\big\|X_{k,h} - X_{k,h}^{(m_n,*)}\big\|_p,\end{aligned}$$

where we also used the fact that

$$X_{k,h}^{(m_n,*)} - \mathbb{E}\big(X_{k,h}^{(m_n,*)} \mid \mathcal{F}_{k-m_n}^{k+m_n}\big) = \mathbb{E}\big(X_{k,h}^{(m_n,*)} - X_{k,h} \mid \sigma\big(\mathcal{F}_{k-m_n}^{k+m_n} \cup \mathcal{F}'_{k-m_n-1}\big)\big).$$

Hence we obtain from the above that

$$\begin{aligned}\sum_{k=1}^{\infty}\big\|\mathcal{P}_0(Y_{k,h}^{(>m_n)})\big\|_2 &\leq \sum_{k=1}^{m_n} 2\big\|X_{k,h} - X_{k,h}^{(m_n,*)}\big\|_2 + \sum_{k=m_n+1}^{\infty}\big\|X_{k,h} - X'_{k,h}\big\|_2 \\ &= 2m_n\big\|X_{1,h} - X_{1,h}^{(m_n,*)}\big\|_2 + \sum_{k=m_n+1}^{\infty}\big\|X_{k,h} - X'_{k,h}\big\|_2.\end{aligned} \tag{4.3.10}$$

By [129, Theorem 1 (iii)] and Assumption 4.2.1 (i) we have for $p \geq 2$

$$\big\|X_{1,h} - X_{1,h}^{(m_n,*)}\big\|_p^2 \leq C\sum_{i=-\infty}^{0}\big\|X_{m_n+h-i,h} - X'_{m_n-i,h}\big\|_p^2 = \mathcal{O}\left(m_n^{1-2\beta}\right), \tag{4.3.11}$$

and we thus obtain that

$$\big\|M_n\big\|_2 \leq \sum_{k=1}^{\infty}\big\|\mathcal{P}_0(Y_{k,h}^{(>m_n)})\big\|_2 = \mathcal{O}\left(m_n^{3/2-\beta}\right). \tag{4.3.12}$$

On the other hand, using (4.3.11) and similar arguments as above, we have

$$\sqrt{n}\big\|R_0\big\|_2 \leq 2\sum_{k=0}^{\infty}\big\|X_{k,h} - X_{k,h}^*\big\|_2 = \mathcal{O}\left(\sum_{k=1}^{\infty} k^{1/2-\beta}\right) = \mathcal{O}(1).$$

Piecing everything together, the claim follows. □

Lemma 4.3.3. *Assume that Assumption 4.2.1 is valid. Then*

$$\limsup_{n\to\infty} \max_{1\leq h\leq d_n} \left\| \frac{1}{\sqrt{L}} \sum_{k=1}^{L} Y_{k,h}^{(\leq m_n)} \right\|_4 < \infty,$$

for any $1 \leq L \leq n$.

Proof of Lemma 4.3.3. Using the martingale decomposition $S_{n,h}^{(\leq m_n)} = M_n - M_1 - (R_n - R_1)$, we obtain as in Lemma 4.3.1

$$\big\|S_{L,h}^{(\leq m_n)}\big\|_4 \leq 2\big\|R_1\big\|_4 + \big\|M_L\big\|_4 = \mathcal{O}\left(L^{-1/2}\right) + \big\|M_L\big\|_4.$$

Let $D_k = M_{k+1} - M_k$ be the martingale differences. Then an application of [35, Proposition 4] yields

$$\big\|M_L\big\|_4^2 \leq \frac{8}{L}\sum_{k=0}^{L}\big\|D_k^2\big\|_2 = \mathcal{O}\left(1\right). \tag{4.3.13}$$

□

For $x \geq 1$, let r_x be the solution of the equation

$$x = (1 + r_x)^9 \exp(x^2/2).$$

As $x \to \infty$, one has the expansion $r_x^2 = 2\log x - (18 + \mathcal{o}(1))\log(1 + \sqrt{2\log x})$, see [57]. The following result is a reformulation of Theorem 3 in [131], which we have adapted for our cause.

Lemma 4.3.4. *Assume that Assumption 4.2.1 is valid. Let* $b_n = d_n/n + n^{-2/3}$*. Then*

$$\max_{1\leq h\leq d_n}\left| P\left(|\sum_{k=1}^{n} X_{k,h}| > \sigma_h\sqrt{n}r_x \right) - 2\big(\Phi(r_x) - 1\big)\right| \leq C\left(b_n x\right)^{1/5}\left(\Phi(r_x) - 1\right),$$

uniformly over $x \in [1, b_n^{-1}]$*, where* C *is independent of* h, n*.*

4.3.2 Step two - truncation

We will frequently use the following result.

Lemma 4.3.5. *Let* $\mathbf{Z}_n^{(1)} = \big(Z_1^{(1)}, ..., Z_n^{(1)}\big)^T$ *be a zero mean* n*-dimensional Gaussian random vector, such that*

$$Var\big(Z_i^{(1)}\big) = 1, \ 1 \le i \le d_n.$$

Suppose that $\mathbf{Z}_n^{(2)} = \big(Z_1^{(2)}, ..., Z_n^{(2)}\big)^T$ *is another* n*-dimensional random vector such that*

$$\lim_{n\to\infty} P\big(\max|\mathbf{Z}_n^{(2)}| > (\log n)^{-\delta}\big) = 0, \tag{4.3.14}$$

for some $\delta > 1/2$*. Then*

$$\big|P\big(\max|\mathbf{Z}_n^{(1)} + \mathbf{Z}_n^{(2)}| \le u_n\big) - P\big(\max|\mathbf{Z}_n^{(1)}| \le u_n\big)\big| \to 0, \tag{4.3.15}$$

as n *tends to infinity.*

Proof of Lemma 4.3.5. We have

$$\begin{aligned}&\big|P\big(\max|\mathbf{Z}_n^{(1)} + \mathbf{Z}_n^{(2)}| \le u_n\big) - P\big(\{\max|\mathbf{Z}_n^{(1)} + \mathbf{Z}_n^{(2)}| \le u_n\} \cap \{\max|\mathbf{Z}_n^{(2)}| \le (\log n)^{-\delta}\}\big)\big|\\ &\le P\big(\max|\mathbf{Z}_n^{(2)}| > (\log n)^{-\delta}\big),\end{aligned}$$

which tends to zero as n increases due to condition (4.3.14). Moreover, it holds that

$$\begin{aligned}&\big|P\big(\{\max|\mathbf{Z}_n^{(1)} + \mathbf{Z}_n^{(2)}| \le u_n\} \cap \{\max|\mathbf{Z}_n^{(2)}| \le (\log n)^{-\delta}\}\big)\big|\\ &\le P\big(u_n - (\log n)^{-\delta} \le \max|\mathbf{Z}_n^{(1)}| \le u_n\big) + P\big(u_n \le \max|\mathbf{Z}_n^{(1)}| \le u_n + (\log n)^{-\delta}\big)\\ &\le n\left(\max_{1\le i\le n} P\big(u_n - (\log n)^{-\delta}|Z_i^{(1)}| \le u_n\big) + \max_{1\le i\le n} P\big(u_n \le |Z_i^{(1)}| \le u_n + (\log n)^{-\delta}\big)\right).\end{aligned}$$

Using the tail estimate $1 - \Phi(x) \le (2\pi)^{-1/2}x^{-1}\exp(-x^2/2)$ for $x > 0$, the above is smaller than

$$C\,n\,\big((\log n)^{-\delta}u_n^{-1}\exp\big(-u_n^2/2 + u_n(\log n)^{-\delta}\big) + (\log n)^{-\delta}u_n^{-1}\exp\big(-u_n^2/2 - u_n(\log n)^{-\delta}\big)\big).$$

Note that $u_n(\log n)^{-\delta} \to 0$ as n increases, thus by relation (4.3.5) the above is bounded by $C\,(\log n)^{-\delta}$. Hence the claim follows. □

Lemma 4.3.6. *Assume that*

$$\lim_{n\to\infty} P\big(\max_{1\le h\le d_n}\big|S_{n,h}\,Var\big(S_{n,h}^{(m)}\big)^{-1/2}\big| \le u_{d_n}\big) = \exp(-e(-z)), \tag{4.3.16}$$

and that Assumption 4.2.1 is valid. Then

$$\lim_{n\to\infty} P\big(\max_{1\le h\le d_n}\big|S_{n,h}\,Var\big(S_{n,h}\big)^{-1/2}\big| \le u_{d_n}\big) = \exp(-e(-z))$$

Proof of Lemma 4.3.6. Put

$$\begin{aligned}
I_{n,h}^{(m)} &:= \frac{S_{n,h}}{\sqrt{\mathrm{Var}\big(S_{n,h}\big)}} - \frac{S_{n,h}}{\sqrt{\mathrm{Var}\big(S_{n,h}^{(m)}\big)}} \\
&= \frac{S_{n,h}}{\sqrt{\mathrm{Var}\big(S_{n,h}^{(m)}\big)}} \frac{\mathrm{Var}\big(S_{n,h}^{(m)}\big) - \mathrm{Var}\big(S_{n,h}\big)}{\sqrt{\mathrm{Var}\big(S_{n,h}\big)}\big(\sqrt{\mathrm{Var}\big(S_{n,h}^{(m)}\big)} + \sqrt{\mathrm{Var}\big(S_{n,h}\big)}\big)} \\
&:= \frac{S_{n,h}}{\sqrt{\mathrm{Var}\big(S_{n,h}^{(m)}\big)}} II_{n,h}^{(m)},
\end{aligned}$$

and note that relation (4.3.7) implies that

$$\limsup_{n\to\infty} \max_{1\le h\le d_n} \big|II_{n,h}^{(m)}\big|(\log n)^{1/2+\delta} = 0,\ \delta > 1/2. \tag{4.3.17}$$

We thus obtain that for sufficiently large n

$$\begin{aligned}
P\big(\max_{1\le h\le d_n}\big|I_{n,h}^{(m)}\big| > (\log n)^{-\delta}\big) &\le P\big(\max_{1\le h\le d_n}\big|S_{n,h}\mathrm{Var}\big(S_{n,h}^{(m)}\big)^{-1/2}\big| > (\log n)^{-\delta}\big|II_{n,h}^{(m)}\big|^{-1}\big) \\
&\le P\big(\max_{1\le h\le d_n}\big|S_{n,h}\mathrm{Var}\big(S_{n,h}^{(m)}\big)^{-1/2}\big| > \sqrt{3\log n}\big) \\
&= o(1).
\end{aligned}$$

Then, arguing as in the proof of Lemma 4.3.5 yields

$$\begin{aligned}
&\big|P\big(\max_{1\le h\le d_n}\big|S_{n,h}\mathrm{Var}\big(S_{n,h}\big)^{-1/2}\big| \le u_n\big) - P\big(\max_{1\le h\le d_n}\big|S_{n,h}\mathrm{Var}\big(S_{n,h}^{(m)}\big)^{-1/2}\big| \le u_n\big)\big| \\
&\le P\big(u_n - (\log n)^{-\delta} \le \max_{1\le h\le d_n}\big|S_{n,h}\mathrm{Var}\big(S_{n,h}^{(m)}\big)^{-1/2}\big| \le u_n\big) \\
&+ P\big(u_n \le \max_{1\le h\le d_n}\big|S_{n,h}\mathrm{Var}\big(S_{n,h}^{(m)}\big)^{-1/2}\big| \le u_n + (\log n)^{-\delta}\big) \\
&= o(1),
\end{aligned}$$

hence the claim follows. □

Lemma 4.3.7. *Assume that*

$$\lim_{n\to\infty} P\big(\max_{1\le h\le d_n}\big|S_{n,h}^{(m_n)}\, Var\big(S_{n,h}^{(m)}\big)^{-1/2}\big| \le u_{d_n}\big) = \exp(-e(-z)), \tag{4.3.18}$$

and that Assumption 4.2.1 is valid. Then

$$\lim_{n\to\infty} P\big(\max_{1\le h\le d_n}\big|S_{n,h}\, Var\big(S_{n,h}^{(m)}\big)^{-1/2}\big| \le u_{d_n}\big) = \exp(-e(-z)).$$

Proof of Lemma 4.3.7. Let $\delta > 1/2$. Then, proceeding as in the proof of Lemma 4.3.5, we obtain that

$$\begin{aligned}
&\big|P\big(\max_{1\le h\le d_n}\big|S_{n,h}^{(m_n)}\mathrm{Var}\big(S_{n,h}^{(m_n)}\big)^{-1/2}\big|\le u_n\big)-P\big(\max_{1\le h\le d_n}\big|S_{n,h}\mathrm{Var}\big(S_{n,h}^{(m_n)}\big)^{-1/2}\big|\le u_n\big)\big|\\
&\le P\big(u_{d_n}-(\log n)^{-\delta}\le \max_{1\le h\le d_n}\big|S_{n,h}^{(m_n)}\mathrm{Var}\big(S_{n,h}^{(m_n)}\big)^{-1/2}\big|\le u_{d_n}\big)\\
&+P\big(u_{d_n}\le \max_{1\le h\le d_n}\big|S_{n,h}^{(m_n)}\mathrm{Var}\big(S_{n,h}^{(m_n)}\big)^{-1/2}\big|\le u_{d_n}+(\log n)^{-\delta}\big)\\
&+P\big(\max_{1\le h\le d_n}\big|S_{n,h}^{(>m_n)}\mathrm{Var}\big(S_{n,h}^{(m_n)}\big)^{-1/2}\big|\ge (\log n)^{-\delta}\big).
\end{aligned}$$

Due to Lemma 4.3.3 and relation (4.3.7) we have for sufficiently large n that

$$\begin{aligned}
P\big(\max_{1\le h\le d_n}\big|S_{n,h}^{(>m_n)}\mathrm{Var}\big(S_{n,h}^{(m_n)}\big)^{-1/2}\big|\ge (\log n)^{-\delta}\big)&\\
\le d_n \max_{1\le h\le d_n} P\big(|S_{n,h}^{(>m_n)}|\mathrm{Var}\big(S_{n,h}^{(>m_n)}\big)^{-1/2}>\sqrt{\log n}\big).&
\end{aligned}$$

Setting $b_n = n^{-2/3}$, $x = b_n^{-1}$, an application of Lemma 4.3.4 yields that the above is bounded by $Cd_n\big(\Phi(r_x)-1\big)$, for some $C > 0$, where $r_x^2 \ge 1/2\log n$ for sufficiently large n. Hence, using the well known tail estimate $1-\Phi(x)\le (2\pi)^{-1/2}x^{-1}\exp(-x^2/2)$ for $x > 0$, we thus obtain

$$P\big(\max_{1\le h\le d_n}\big|S_{n,h}^{(>m_n)}\mathrm{Var}\big(S_{n,h}^{(m_n)}\big)^{-1/2}\big|\ge (\log n)^{-\delta}\big)=\mathcal{O}\left(d_n n^{-1/2}\right)=o(1). \quad (4.3.19)$$

On the other hand, for any $z' < z < z''$, we have for sufficiently large n the inequalities

$$\begin{aligned}
P\big(u_{d_n}(z)-(\log n)^{-\delta}&\le \max_{1\le h\le d_n}\big|S_{n,h}^{(m_n)}\mathrm{Var}\big(S_{n,h}^{(m_n)}\big)^{-1/2}\big|\le u_{d_n}(z)\big)\\
&\le P\big(u_{d_n}(z')\le \max_{1\le h\le d_n}\big|S_{n,h}^{(m_n)}\mathrm{Var}\big(S_{n,h}^{(m_n)}\big)^{-1/2}\big|\le u_{d_n}(z)\big),\\
P\big(u_{d_n}(z)&\le \max_{1\le h\le d_n}\big|S_{n,h}^{(m_n)}\mathrm{Var}\big(S_{n,h}^{(m_n)}\big)^{-1/2}\big|\le u_{d_n}(z)+(\log n)^{-\delta}\big)\\
&\le P\big(u_{d_n}(z)\le \max_{1\le h\le d_n}\big|S_{n,h}^{(m_n)}\mathrm{Var}\big(S_{n,h}^{(m_n)}\big)^{-1/2}\big|\le u_{d_n}(z'')\big).
\end{aligned}$$

In addition, condition (4.3.18) yields that

$$\begin{aligned}
&P\big(u_{d_n}(z')\le \max_{1\le h\le d_n}\big|S_{n,h}^{(m_n)}\mathrm{Var}\big(S_{n,h}^{(m_n)}\big)^{-1/2}\big|\le u_{d_n}(z)\big)\\
&+P\big(u_{d_n}(z)\le \max_{1\le h\le d_n}\big|S_{n,h}^{(m_n)}\mathrm{Var}\big(S_{n,h}^{(m_n)}\big)^{-1/2}\big|\le u_{d_n}(z'')\big)\to \exp(-e(-z''))-\exp(-e(-z')),
\end{aligned}$$

as n tends to infinity. This implies that we can chose sequences $z_n' \uparrow z$, $z_n'' \downarrow z$, such that

$$\lim_{n\to\infty} P\big(u_{d_n}(z_n') \le \max_{1\le h\le d_n} \big|S_{n,h}^{(m_n)} \mathrm{Var}\big(S_{n,h}^{(m_n)}\big)^{-1/2}\big| \le u_{d_n}(z)\big)$$
$$= \exp(-e(-z)) - \exp(-e(-z)) = 0,$$
$$\lim_{n\to\infty} P\big(u_{d_n}(z) \le \max_{1\le h\le d_n} \big|S_{n,h}^{(m_n)} \mathrm{Var}\big(S_{n,h}^{(m_n)}\big)^{-1/2}\big| \le u_{d_n}(z_n'')\big)$$
$$= \exp(-e(-z)) - \exp(-e(-z)) = 0,$$

which completes the proof. □

We will now approximate $\{X_{k,h}\}_{k\in\mathbb{N}}$ with an m_n-dependent sequence. To this end, let $d_n^* = d_n + d_n(\log n)^\alpha$. We divide the set of integers $\{1, 2, ...\}$ into consecutive blocks $\mathcal{H}_1, \mathcal{I}_1, \mathcal{H}_2, \mathcal{I}_2...$. The blocks are defined by recursion, more precisely, define

$$\mathcal{H}_j := \big\{k \in \mathbb{N} \mid 1 + (j-1)d_n^* \le k \le jd_n^*\big\},$$

and

$$\mathcal{I}_j = \big\{k \in \mathbb{N} \mid jd_n^* - 2m_n - 1 \le k \le jd_n^*\big\}.$$

Note that unlike to many other authors, we do not use a dyadic (or triadic) scheme, since it turns out that the Gaussian approximation used in Section 4.3.3 works better with the blocks defined above.

For $1 \le j \le \lceil n/d_n^* \rceil$ we define the random variables

$$\eta_{j,h}^{(m,1)} = \sum_{k\in\mathcal{H}_j} Y_{k,h}^{(\le m)}, \quad \eta_{j,h}^{(m,2)} := \sum_{k\in\mathcal{I}_j} Y_{k,h}^{(\le m)}, \tag{4.3.20}$$

and the random vectors

$$\boldsymbol{\eta}_j^{(m,i)} = \big(\eta_{j,1}^{(m,i)}, \eta_{j,2}^{(m,i)}, ..., \eta_{j,d_n}^{(m,i)}\big)^T, \quad i \in \{1, 2\}. \tag{4.3.21}$$

In this spirit, we also define the partial sums as

$$S_{n,h}^{(m,i)} = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lceil n/d_n^* \rceil} \eta_{j,h}^{(m,i)}, \quad i \in \{1, 2\},$$

and the random vectors

$$\mathbf{S}_n^{(m,i)} = \big(S_{n,1}^{(m,i)}, S_{n,2}^{(m,i)}, ..., S_{n,d_n}^{(m,i)}\big)^T, \quad i \in \{1, 2\}.$$

Note that we have the representation

$$\mathbf{S}_n^{(m)} = \mathbf{S}_n^{(m,1)} + \mathbf{S}_n^{(m,2)} = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lceil n/d_n^* \rceil} \big(\boldsymbol{\eta}_j^{(m,1)} + \boldsymbol{\eta}_j^{(m,2)}\big).$$

In addition, we assume throughout the remainder of this paper that $\Lambda_{n,h}^{(m,i)}$, $i \in \{1,2\}$ are Gaussian copies of $S_{n,h}^{(m,i)}$,$i \in \{1,2\}$, in other words, Gaussian processes with the same covariance structure as $S_{n,h}^{(m,i)}$,$i \in \{1,2\}$.

Lemma 4.3.8. *Let $d_n^* = d_n + d_n(\log n)^\alpha$, $\alpha > 3$, and assume that*

$$\lim_{n\to\infty} P\big(\max_{1\le h\le d_n} \big|S_{n,h}^{(m_n)}\, Var\big(S_{n,h}^{(m_n,1)}\big)^{-1/2}\big| \le u_{d_n}\big) = \exp(-e(-z)), \tag{4.3.22}$$

and that Assumption 4.2.1 is valid. Then

$$\lim_{n\to\infty} P\big(\max_{1\le h\le d_n} \big|S_{n,h}^{(m_n)}\, Var\big(S_{n,h}^{(m_n)}\big)^{-1/2}\big| \le u_{d_n}\big) = \exp(-e(-z)).$$

Proof of Lemma 4.3.8. By the cauchy-Schwarz inequality, we have

$$\big|\mathrm{Var}\big(S_{n,h}^{(m_n,1)}\big) - \mathrm{Var}\big(S_{n,h}^{(m_n)}\big)\big| \le \mathrm{Var}\big(S_{n,h}^{(m_n,2)}\big) + 2\sqrt{\mathrm{Var}\big(S_{n,h}^{(m_n,2)}\big)\mathrm{Var}\big(S_{n,h}^{(m_n,1)}\big)},$$

hence

$$\max_{1\le h\le d_n} \big|\mathrm{Var}\big(S_{n,h}^{(m_n,1)}\big) - \mathrm{Var}\big(S_{n,h}^{(m_n)}\big)\big| = \mathcal{O}\left(\max_{1\le h\le d_n} \sqrt{\mathrm{Var}\big(S_{n,h}^{(m_n,2)}\big)}\right). \tag{4.3.23}$$

Per construction, we have

$$\mathrm{Var}\big(S_{n,h}^{(m_n,2)}\big) = \frac{1}{n} \sum_{j=1}^{\lceil n/d_n^* \rceil} \mathrm{Var}\big(\eta_{j,2}^{(m_n,2)}\big) \le C(d_n^*)^{-1}\mathrm{Var}\big(\eta_{1,2}^{(m_n,2)}\big),$$

and due to Assumption 4.2.1 (ii) the right hand side is bounded by

$$C(d_n^*)^{-1} d_n = \mathcal{O}\left((\log n)^{-\alpha/2}\right). \tag{4.3.24}$$

One can now proceed exactly as in the proof of Lemma 4.3.6. □

4.3.3 Step three - normal approximation

To simplify the notation a little, we will assume from now on that

$$\mathrm{Var}\big(S_{n,h}^{(m_n,1)}\big) = 1, \quad \max_{1 \le h \le d_n} \mathrm{Var}\big(S_{n,h}^{(m_n,2)}\big) = \mathcal{O}\left((\log n)^{-\alpha}\right), \ \alpha > 3.$$

Note that this does not pose any additional restraint due to the previous results. In the sequel, we will be dealing with d-dimensional spheres. To this end, we define the open sphere with center x as

$$\mathbf{B}(\mathbf{x}, r, d) = \{\mathbf{y} \in \mathbb{R}^d \mid |\mathbf{x} - \mathbf{y}| < r\},$$

and we denote with $\mathfrak{S}$ the system of all such spheres. In this spirit, we define the Lévy-Prokhorov metric as

$$\pi\big(U, V; \mathfrak{S}\big) = \inf\big\{\epsilon : \ V(A) \le U(A^{\epsilon}) + \epsilon, U(A) \le V(A^{\epsilon}) + \epsilon \text{ for all } A \in \mathfrak{S}\big\},$$

where U, V are probability distributions, and $A^{\epsilon} = \{x : \ d(x, A) < \epsilon\}$ is the ϵ-neighborhood of a set A, and $d(x, A)$ is the distance between x and the set A. For a random variable X, we will write $P_X(A)$ for the probability $P(X \in A)$, and $\Phi_X(A)$ if the distribution is Gaussian with the same covariance structure as X.

An important tool for estimating convergence rates for the Lévy-Prokhorov metric are the ζ_s-metrics (cf. [110, 139]), which are examples of the so-called ideal metrics. Ideal metrics have the properties of *semiadditivity* and *homogeneity of order s* (cf. [110]), which make deductions of convergence rate estimates very simple. Let $s > 0$. Then we can represent s as $s = m + \alpha$, where $[s] = m$ denotes the integer part, and $0 \le \alpha < 1$. Let $\mathfrak{F}_s$ be the class of all real-valued functions f, such that the m-th derivative exists, is bounded and satisfies

$$\big|f^{(m)}(x) - f^{(m)}(y)\big| \le \big|x - y\big|^{\alpha}. \tag{4.3.25}$$

The metric ζ_s for two probability measures P, Q is then defined as

$$\zeta_s(P, Q) = \sup\left\{\left|\int f(x)(P - Q)(dx)\right| : f \in \mathfrak{F}_s\right\}.$$

Based on the ζ_s-metrics, we have the following estimate for the the Lévy-Prokhorov metric.

Lemma 4.3.9. *We have*

$$\pi\big(P_{\mathbf{S}_n^{(m_n,1)}}, \Phi_{\mathbf{S}_n^{(m_n,1)}}, \mathfrak{S}\big)^{1+s} \le c(s)\big(n^{s/2-1} d_n^*\big)^{-1} \zeta_s\big(P_{\boldsymbol{\eta}_1^{(m_n,1)}}, \Phi_{\boldsymbol{\eta}_1^{(m_n,1)}}\big),$$
$$\pi\big(P_{\mathbf{S}_n^{(m_n,2)}}, \Phi_{\mathbf{S}_n^{(m_n,2)}}, \mathfrak{S}\big)^{1+s} \le c(s)\big(n^{s/2-1} d_n^*\big)^{-1} \zeta_s\big(P_{\boldsymbol{\eta}_1^{(m_n,2)}}, \Phi_{\boldsymbol{\eta}_1^{(m_n,2)}}\big).$$

Proof of Lemma 4.3.9. By inequality [110, 6.1.3] (see also Dudley), we have

$$\pi\big(P_{\mathbf{S}_n^{(m_n,1)}}, \Phi_{\mathbf{S}_n^{(m_n,1)}}, \mathfrak{S}\big) \le c(s)\zeta_s\big(P_{\mathbf{S}_n^{(m_n,1)}}, \Phi_{\mathbf{S}_n^{(m_n,1)}}\big).$$

Using the semi-addivity of the ζ_s metric, we obtain

$$\begin{aligned}
\zeta_s\big(P_{\mathbf{S}_n^{(m_n,1)}}, \Phi_{\mathbf{S}_n^{(m_n,1)}}\big) &= \zeta_s\big(P_{n^{-1/2}(\boldsymbol{\eta}_1^{(m_n,1)}+\ldots+\boldsymbol{\eta}_{\lceil n/d_n^*\rceil}^{(m_n,1)})}, \Phi_{n^{-1/2}(\boldsymbol{\eta}_1^{(m_n,1)}+\ldots+\boldsymbol{\eta}_{\lceil n/d_n^*\rceil}^{(m_n,1)})}\big) \\
&\le \zeta_s\big(P_{n^{-1/2}\boldsymbol{\eta}_1^{(m_n,1)}}, \Phi_{n^{-1/2}\boldsymbol{\eta}_1^{(m_n,1)}}\big) \\
&+ \zeta_s\big(P_{n^{-1/2}(\boldsymbol{\eta}_2^{(m_n,1)}+\ldots+\boldsymbol{\eta}_{\lceil n/d_n^*\rceil}^{(m_n,1)})}, \Phi_{n^{-1/2}(\boldsymbol{\eta}_2^{(m_n,1)}+\ldots+\boldsymbol{\eta}_{\lceil n/d_n^*\rceil}^{(m_n,1)})}\big) \\
&\le \sum_{j=1}^{\lceil n/d_n^*\rceil} \zeta_s\big(P_{n^{-1/2}\boldsymbol{\eta}_j^{(m_n,1)}}, \Phi_{n^{-1/2}\boldsymbol{\eta}_1^{(m_n,1)}}\big).
\end{aligned}$$

In addition, the homogeneity of the ζ_s metric implies that

$$\begin{aligned}
\sum_{j=1}^{\lceil n/d_n^*\rceil} \zeta_s\big(P_{n^{-1/2}\boldsymbol{\eta}_j^{(m_n,1)}}, \Phi_{n^{-1/2}\boldsymbol{\eta}_j^{(m_n,1)}}\big) &\le \lceil n/d_n^*\rceil \zeta_s\big(P_{n^{-1/2}\boldsymbol{\eta}_1^{(m_n,1)}}, \Phi_{n^{-1/2}\boldsymbol{\eta}_1^{(m_n,1)}}\big) \\
&\le 2(n^{s/2-1}d_n)^{-1}\zeta_s\big(P_{\boldsymbol{\eta}_1^{(m_n,1)}}, \Phi_{\boldsymbol{\eta}_1^{(m_n,1)}}\big),
\end{aligned}$$

which completes the proof. □

Lemma 4.3.10. *If Assumption 4.2.1 is valid, then*

$$\zeta_3\big(P_{\boldsymbol{\eta}_1^{(m_n,1)}}, \Phi_{\boldsymbol{\eta}_1^{(m_n,1)}}\big) = \mathcal{O}\left(d_n^{5/2}(d_n^*)^{3/2}\right), \tag{4.3.26}$$

$$\zeta_3\big(P_{\boldsymbol{\eta}_1^{(m_n,2)}}, \Phi_{\boldsymbol{\eta}_1^{(m_n,2)}}\big) = \mathcal{O}\left(d_n^{5/2}(d_n^*)^{3/2}\right) \tag{4.3.27}$$

The proof goes along the lines of [110], see also Zolotarev [139] for the original argument.

Proof of Lemma 4.3.10. Fix any function $f \in \mathfrak{F}_3$. Using Taylor's formula with the integral representation for the remainder term, we have.

$$f(\mathbf{x}) = f(\mathbf{0}) + \sum_{i=1}^{d_n} \frac{\partial f(\mathbf{0})}{\partial x_i} x_i + (2!)^{-1} \sum_{i_1,i_2=1}^{d_n} x_{i_1}x_{i_2} \int_0^1 \left(\frac{\partial^2 f(\lambda \mathbf{x})}{\partial x_{i_1}\partial x_{i_2}}(1-\lambda)\right) d\lambda.$$

Using this, we obtain for the expectation given below the estimate

$$\begin{aligned}
&\int f(\mathbf{x})(P_{\boldsymbol{\eta}_1^{(m_n,1)}} - \Phi_{\boldsymbol{\eta}_1^{(m_n,1)}})(d\mathbf{x}) \\
&= (2!)^{-1} \int \left(\sum_{i_1,i_2=1}^{d_n} x_{i_1}x_{i_2} \int_0^1 \left(\frac{\partial^2 f(\lambda \mathbf{x})}{\partial x_{i_1}\partial x_{i_2}}(1-\lambda)\right) d\lambda\right) (P_{\boldsymbol{\eta}_1^{(m_n,1)}} - \Phi_{\boldsymbol{\eta}_1^{(m_n,1)}})(d\mathbf{x}).
\end{aligned}$$

Moreover, using that

$$\begin{aligned}\left|\int_0^1 \left(\frac{\partial^2 f(\lambda\mathbf{x})}{\partial x_{i_1}\partial x_{i_2}} - \frac{\partial^2 f(\lambda\mathbf{0})}{\partial x_{i_1}\partial x_{i_2}}\right)(1-\lambda)d\lambda\right| &\leq \int_0^1 \left(\left|\lambda\mathbf{x}\right|(1-\lambda)\right)d\lambda \\ &\leq \left|\mathbf{x}\right| \int_0^1 \lambda(1-\lambda)d\lambda \\ &\leq C\left|\mathbf{x}\right|,\end{aligned}$$

the above estimate simplifies to

$$\begin{aligned}\left|\int f(\mathbf{x})(P_{\boldsymbol{\eta}_1^{(m_n,1)}} - \Phi_{\boldsymbol{\eta}_1^{(m_n,1)}})(d\mathbf{x})\right| &\leq C\sum_{i_1,i_2=1}^{d_n}\int \left(\left|\mathbf{x}\right|\left|x_{i_1}x_{i_2}\right|\right)\left|P_{\boldsymbol{\eta}_1^{(m_n,1)}} - \Phi_{\boldsymbol{\eta}_1^{(m_n,1)}}\right|(d\mathbf{x}) \\ &\leq C\sum_{i_1,i_2=1}^{d_n}\int \left(\left|\mathbf{x}\right|\left|x_{i_1}x_{i_2}\right|\right)\left(P_{\boldsymbol{\eta}_1^{(m_n,1)}} + \Phi_{\boldsymbol{\eta}_1^{(m_n,1)}}\right)(d\mathbf{x}).\end{aligned}$$

An application of the Cauchy-Schwarz inequality now yields

$$\int \left(\left|\mathbf{x}\right|\left|x_{i_1}x_{i_2}\right|\right)P_{\boldsymbol{\eta}_1^{(m_n,1)}} \leq \sqrt{\int \left|\mathbf{x}\right|^2 P_{\boldsymbol{\eta}_1^{(m_n,1)}}}\sqrt{\int (x_{i_1}x_{i_2})^2 P_{\boldsymbol{\eta}_1^{(m_n,1)}}}.$$

Lemma 4.3.3 now implies that

$$\int \left|\mathbf{x}\right|^2 P_{\boldsymbol{\eta}_1^{(m_n,1)}} = \sum_{i=1}^{d_n}\mathbb{E}\left(\left(\eta_{i,1}^{(m_n,1)}\right)^2\right) = \mathcal{O}\left(d_n d_n^*\right). \tag{4.3.28}$$

In addition, an application of the Cauchy-Schwarz inequality yields

$$\left(\int (x_{i_1}x_{i_2})^2 P_{\boldsymbol{\eta}_1^{(m_n,1)}}\right)^{1/2} \leq \left(\int x_{i_2}^4 P_{\boldsymbol{\eta}_1^{(m_n,1)}} \int x_{i_1}^4 P_{\boldsymbol{\eta}_1^{(m_n,1)}}\right)^{1/4} = \mathcal{O}\left(d_n^*\right),$$

where the last assertion follows from Lemma 4.3.3. Hence we obtain that

$$\sum_{i_1,i_2=1}^{d_n}\int (x_{i_1}x_{i_2})^2 P_{\boldsymbol{\eta}_1^{(m_n,1)}} = \mathcal{O}\left(d_n^2 d_n^*\right), \tag{4.3.29}$$

which together with (4.3.28) yields

$$\left|\int f(\mathbf{x})(P_{\boldsymbol{\eta}_1^{(m_n,1)}} - \Phi_{\boldsymbol{\eta}_1^{(m_n,1)}})(d\mathbf{x})\right| = \mathcal{O}\left(d_n^{5/2}(d_n^*)^{3/2}\right).$$

Since the function f was arbitrary, the claim follows. □

Lemma 4.3.11. *Let $v_n \geq 0$. If Assumption 4.2.1 holds, we have*

$$\left|P_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0}, v_n, d_n)\big) - \Phi_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0}, v_n, d_n)\big)\right| = o(1),$$
$$\left|P_{\mathbf{S}_n^{(m_n,2)}}\big(\mathbf{B}(\mathbf{0}, v_n, d_n)\big) - \Phi_{\mathbf{S}_n^{(m_n,2)}}\big(\mathbf{B}(\mathbf{0}, v_n, d_n)\big)\right| = o(1)$$

Proof of Lemma 4.3.11. It follows from Lemma 4.3.9 and 4.3.10 that

$$\pi\big(P_{\mathbf{S}_n^{(m_n,1)}}, \Phi_{\mathbf{S}_n^{(m_n,1)}}, \mathfrak{S}\big)^3 \leq c(3)\zeta_3\big(P_{\mathbf{S}_n^{(m_n,1)}}, \Phi_{\mathbf{S}_n^{(m_n,1)}}\big) = \mathcal{O}(\epsilon_n),$$

where

$$\epsilon_n = n^{-1/2} d_n^{5/2} (d_n^*)^{1/2} = (\log n)^{-\alpha/2}, \tag{4.3.30}$$

where we point out that $\alpha/6 > 1/2$. Hence we obtain the inequalities

$$\begin{aligned}
&P_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0}, v_n, d_n)\big) - \Phi_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0}, v_n, d_n)\big) \\
&\qquad \leq \Phi_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0}, v_n + \epsilon_n, d_n)\big) + \epsilon_n^{1/3} - \Phi_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0}, v_n, d_n)\big),
\end{aligned}$$

and

$$\begin{aligned}
&\Phi_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0}, v_n, d_n)\big) - P_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0}, v_n, d_n)\big) \\
&\qquad \leq P_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0}, v_n + \epsilon_n, d_n)\big) + \epsilon_n^{1/3} - P_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0}, v_n, d_n)\big).
\end{aligned}$$

Arguing as in the proof of Lemma 4.3.5, we obtain that

$$\begin{aligned}
&\Phi_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0}, v_n + \epsilon_n, d_n)\big) - \Phi_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0}, v_n, d_n)\big) \\
&\leq d_n \left(\max_{1 \leq i \leq d_n} P\big(v_n - \epsilon_n^{1/3} \leq |\Lambda_{n,i}^{(m_n,1)}| \leq v_n\big) + \max_{1 \leq i \leq d_n} P\big(v_n \leq |\Lambda_{n,i}^{(m_n,1)}| \leq v_n + \epsilon_n^{1/3}\big) \right) \\
&= o(1),
\end{aligned}$$

since $\epsilon_n^{1/3} = (\log n)^{-\alpha/6}$. Hence we conclude that

$$\Phi_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0}, v_n, d_n)\big) - \Phi_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0}, v_n, d_n)\big) = o(1). \tag{4.3.31}$$

In the same manner, we obtain

$$\begin{aligned}
&P_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0}, v_n, d_n)\big) - \Phi_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0}, v_n, d_n)\big) \\
&\leq d_n \left(\max_{1 \leq i \leq d_n} P\big(v_n - \epsilon_n^{1/3} \leq |S_{n,i}^{(m_n,1)}| \leq v_n\big) + \max_{1 \leq i \leq d_n} P\big(v_n \leq |S_{n,i}^{(m_n,1)}| \leq v_n + \epsilon_n^{1/3}\big) \right).
\end{aligned} \tag{4.3.32}$$

By the classical Berry-Esséen bound, we have that

$$\max_{1\leq i\leq d_n}\sup_{x\in\mathbb{R}}\left|P\big(S_{n,i}^{(m_n,1)}\leq x\big)-P\big(\Lambda_{n,i}^{(m_n,1)}\leq x\big)\right|\leq\sqrt{\frac{d_n^*}{n}}\max_{1\leq i\leq d_n}\mathbb{E}\left(\left((d_n^*)^{-1/2}\eta_{i,1}^{(m_n,1)}\right)^3\right)$$
$$=\mathcal{O}\left(\sqrt{\frac{d_n^*}{n}}\right).$$

Since $d_n(d_n^*)^{1/2}n^{-1/2}=\mathcal{o}(1)$, we can replace the random variables $S_{n,i}^{(m_n,1)}$ with their Gaussian versions $\Lambda_{n,i}^{(m_n,1)}$ in (4.3.32), and we already know from the above that the magnitude of the resulting expression is $\mathcal{o}(1)$. We thus conclude that

$$P_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0},\upsilon_n,d_n)\big)-P_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0},\upsilon_n,d_n)\big)=\mathcal{o}(1), \tag{4.3.33}$$

which together with (4.3.32) yields

$$\left|P_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0},\upsilon_n,d_n)\big)-\Phi_{\mathbf{S}_n^{(m_n,1)}}\big(\mathbf{B}(\mathbf{0},\upsilon_n,d_n)\big)\right|=\mathcal{o}(1). \tag{4.3.34}$$

□

Lemma 4.3.12. *Let $d_n^*=d_n+d_n(\log n)^\alpha$, $\alpha>3$, and assume that Assumption 4.2.1 is valid. Then*

$$\lim_{n\to\infty}\left|P\big(\max_{1\leq h\leq d_n}\left|S_{n,h}^{(m_n)}\right|\leq u_n\big)-P\big(\max_{1\leq h\leq d_n}\left|\Lambda_{n,h}^{(m_n)}\right|\leq u_n\big)\right|=0.$$

Proof of Lemma 4.3.12. Let Z be a standard Gaussian random variable, and $\delta>1/2$. We have

$$P\big(\max_{1\leq h\leq d_n}\left|\Lambda_{n,h}^{(m_n,2)}\right|>(\log n)^{-\delta}\big)\leq d_n\max_{1\leq h\leq d_n}P\big(\left|\Lambda_{n,h}^{(m_n,2)}\right|>(\log n)^{-\delta}\big)$$
$$=d_n\max_{1\leq h\leq d_n}P\big(\left|Z\right|^2>(\log n)^{-2\delta}\mathrm{Var}\big(\Lambda_{n,h}^{(m_n,2)}\big)\big)$$
$$\leq d_nP\big(\left|Z\right|>(\log n)^{-\delta+\alpha/2}\big)\big).$$

Since $\alpha>3$, we can chose $\delta>1/2$ such that $\alpha/2-\delta>1/2$. Then, arguing as in the proof of Lemma 4.3.5, we obtain

$$d_nP\big(\left|Z\right|>(\log n)^{-\delta+\alpha/2}\big)\big)=\mathcal{o}(1),$$

hence

$$P\big(\max_{1\leq h\leq d_n}\left|\Lambda_{n,h}^{(m_n,2)}\right|>(\log n)^{-\delta}\big)=\mathcal{o}(1),\quad \delta>1/2. \tag{4.3.35}$$

Similarly, using Lemma 4.3.11 and relation (4.3.35) we obtain

$$P\big(\max\big|\mathbf{S}_n^{(m_n,2)}\big| > (\log n)^{-\delta}\big) = \mathcal{o}(1), \quad \delta > 1/2. \tag{4.3.36}$$

Since $u_n^2 = \mathcal{O}(\log n)$, repeating the arguments of Lemma 4.3.14 together with the approximation given by Lemma 4.3.11 yields that

$$\big|P\big(\max_{1\leq h\leq d_n}\big|S_{n,h}^{(m_n,1)}\big| \leq u_n\big) - P\big(\max_{1\leq h\leq d_n}\big|S_{n,h}^{(m_n)}\big| \leq u_n\big)\big| = \mathcal{o}(1), \tag{4.3.37}$$

and the same is true for $\Lambda_{n,h}^{(m_n)}$. Hence we obtain that,

$$\begin{aligned}\big|P\big(\max_{1\leq h\leq d_n}\big|S_{n,h}^{(m_n)}\big| \leq u_n\big)-&P\big(\max_{1\leq h\leq d_n}\big|\Lambda_{n,h}^{(m_n)}\big| \leq u_n\big)\big|\\ &\leq \big|P\big(\max_{1\leq h\leq d_n}\big|S_{n,h}^{(m_n,1)}\big| \leq u_n\big) - P\big(\max_{1\leq h\leq d_n}\big|\Lambda_{n,h}^{(m_n,1)}\big| \leq u_n\big)\big| + \mathcal{o}(1).\end{aligned}$$

But by Lemma 4.3.11 we have

$$\big|P\big(\max_{1\leq h\leq d_n}\big|S_{n,h}^{(m_n,1)}\big| \leq u_n\big) - P\big(\max_{1\leq h\leq d_n}\big|\Lambda_{n,h}^{(m_n,1)}\big| \leq u_n\big)\big| = \mathcal{o}(1),$$

hence the claim follows. □

4.3.4 Proofs of Theorem 4.2.2 and 4.2.4

Proof of Theorem 4.2.2. Due to the results of Sections 4.3.1, 4.3.2 and 4.3.3, it suffices to establish that

$$P\big(\max_{1\leq h\leq d_n}\big|\Lambda_{n,h}^{(m_n)}\big| \leq u_n\big) \to \exp(-z). \tag{4.3.38}$$

Note however, that the truncation and blocking arguments have altered the covariance structure, i.e. the covariance matrices of $S_{n,h}$ and $\Lambda_{n,h}^{(m_n)}$, $1 \leq h \leq d_n$ are not the same. Fortunately, it will be sufficient that the covariance structures are close enough. To this end, note that the Cauchy-Schwarz inequality and Lemma 4.3.3 imply that for $1 \leq i, j \leq d_n$

$$\Big|\mathrm{Cov}\big(S_{n,i}, S_{n,j}\big) - \mathrm{Cov}\big(\Lambda_{n,i}^{(m_n)}, \Lambda_{n,j}^{(m_n)}\big)\Big| \leq C \max_{1\leq h\leq d_n}\Big(\big\|S_{n,h}^{(>m)}\big\|_2 + \big\|S_{n,h}^{(m,2)}\big\|_2\Big).$$

By Lemma 4.3.1 we have $\big\|S_{n,h}^{(m,2)}\big\|_2 = \mathcal{O}\big(m_n^{3/2-\beta} + n^{-1/2}\big)$, and Lemma 4.3.3 implies that $\big\|S_{n,h}^{(m,2)}\big\|_2 = \mathcal{O}\big(d_n^{1/2}(d_n^*)^{-1/2}\big) = \mathcal{O}\big((\log n)^{\alpha/2}\big)$. Hence we obtain that

$$\Big|\mathrm{Cov}\big(S_{n,i}, S_{n,j}\big) - \mathrm{Cov}\big(\Lambda_{n,i}^{(m_n)}, \Lambda_{n,j}^{(m_n)}\big)\Big| = \mathcal{O}\big((\log n)^{2+\gamma}\big),\ \gamma > 0. \tag{4.3.39}$$

Per assumption, we have that $r_1 < 1$, and in addition $r_n(\log n)^{2+\gamma} \to 0$ for some $\gamma > 0$, where $\varphi_{h,l} = \limsup_{n\to\infty}\big|\mathbb{E}\big(S_{n,h}, S_{n,l}\big)\big|$, $h, l \in \mathbb{N}$, and $r_k = \sup_{|h-l|\geq k}|\varphi_{h,l}|$. Hence, for large enough n, it follows from (4.3.39), that these conditions are also valid if we replace $S_{n,h}$ with $\Lambda_{n,h}^{(m_n)}$. Thus (4.3.38) follows from Theorem 1 in [38]. □

Proof of Theorem 4.2.4. In order to increase the readability, we introduce the following notation. Denote with $\sigma_h^2 = \mathrm{Var}\big(S_{n,h}\big)$, and $\widehat{\sigma}_h^2 = \widehat{\mathrm{Var}}\big(S_{n,h}\big)$. We need to show that the error

$$\max_{1\leq h\leq d_n}\big|\big(\sigma_h^{-1} - \widehat{\sigma}_h^{-1}\big)S_{n,h}\big| \tag{4.3.40}$$

is sufficiently small in probability, since then the claim follows from Theorem 4.3.6 and arguments used in the proof of Lemma 4.3.6. To this end, we have

$$\big|\big(\sigma_h^{-1} - \widehat{\sigma}_h^{-1}\big)S_{n,h}\big| = \left|\frac{S_{n,h}}{\sigma_h}\frac{\widehat{\sigma}_h^2 - \sigma_h^2}{\widehat{\sigma}_h\big(\widehat{\sigma}_h + \sigma_h\big)}\right| \leq \left|\frac{S_{n,h}}{\sigma_h}\frac{\widehat{\sigma}_h^2 - \sigma_h^2}{\widehat{\sigma}_h^2}\right|.$$

In addition, for $0 < \epsilon_n \leq 1$ we have

$$\begin{aligned} P\Big(\max_{1\leq h\leq d_n}\big(|\widehat{\sigma}_h^2 - \sigma_h^2|\widehat{\sigma}_h^{-2}\big) \geq \epsilon_n\Big) &\leq P\Big(\max_{1\leq h\leq d_n}\big(|\widehat{\sigma}_h^2 - \sigma_h^2|(1+\epsilon_n)\sigma_h^{-2}\big) \geq \epsilon_n\Big) \\ &\leq P\Big(\max_{1\leq h\leq d_n}\big(|\widehat{\sigma}_h^2 - \sigma_h^2|\sigma_h^{-2}\big) \geq \epsilon_n/2\Big), \end{aligned}$$

which due to Assumption 4.2.1 is bounded by $P\big(\max_{1\leq h\leq d_n}|\widehat{\sigma}_h^2 - \sigma_h^2| \geq C\,\epsilon\big)$, for some $C > 0$, which does not depend on n or h. Choosing $\epsilon_n = (\log n)^{-\alpha}$, $\alpha > 1$, we thus obtain from Assumption 4.2.3

$$P\Big(\max_{1\leq h\leq d_n}\big(|\widehat{\sigma}_h^2 - \sigma_h^2|\widehat{\sigma}_h^{-2}\big) \geq \epsilon_n\Big) = \mathcal{O}(1). \tag{4.3.41}$$

In addition, it follows from the above that for some $\delta > 1/2$

$$\begin{aligned} P\Big(\max_{1\leq h\leq d_n}\big|\big(\sigma_h^{-1} - \widehat{\sigma}_h^{-1}\big)S_{n,h}\big| > (\log n)^{-\delta}\Big) &\leq P\Big(\max_{1\leq h\leq d_n}\big|S_{n,h}\sigma_h^{-1}\big| \geq (\log n)^{-\delta+\alpha}\Big) \\ &+ P\Big(\max_{1\leq h\leq d_n}\big(|\widehat{\sigma}_h^2 - \sigma_h^2|\widehat{\sigma}_h^{-2}\big) \geq (\log n)^{-\alpha}\Big). \end{aligned}$$

Since $\alpha > 1$, we can choose a $\delta > 1/2$ such that $\alpha - \delta > 1/2$. Then, arguing as in the proof of Lemma 4.3.11 or using Theorem 4.2.2, one deduces that

$$P\Big(\max_{1\leq h\leq d_n}\big|S_{n,h}\sigma_h^{-1}\big| \geq (\log n)^{-\delta+\alpha}\Big) = \mathcal{O}(1). \tag{4.3.42}$$

Using (4.3.41), we thus obtain

$$P\Big(\max_{1\leq h\leq d_n}\big|\big(\sigma_h^{-1} - \widehat{\sigma}_h^{-1}\big)S_{n,h}\big| > (\log n)^{-\delta}\Big) = \mathcal{O}(1), \quad \delta > 1/2. \tag{4.3.43}$$

Due to Theorem 4.2.2, we can now use the same arguments as in the proof of Lemma 4.3.6, which yields the claim. □

4.3.5 Proof of Theorem 4.2.6

In order to proof Theorem 4.2.6, we need to validate Assumption 4.2.1. We will do so in the lemmas given below. To this end, let

$$X_{k,h} = L_{k+h-1}L_k - \phi_{h-1}. \tag{4.3.44}$$

In order to verify Assumption 4.2.1 (i), the following result is useful.

Lemma 4.3.13. *Assume that* $\|\epsilon\|_8 < \infty$ *and that* $\sum_{i=0}^{\infty} |\alpha_i| < \infty$*. Then*

$$\big\|X_{k,h} - X'_{k,h}\big\|_4 \le C\big(|\alpha_k| + |\alpha_{k+h}|\big),$$

where C *does not depend on* h, k*.*

For $m \ge 0$ and $k \ge l \ge 0$, let

$$T_m = \phi_m\phi_{m+k-l} + \phi_{m+k}\phi_{m-l} + (\eta - 3)\sum_{i=0}^{\infty} \alpha_i\alpha_{i+k}\alpha_{i+m}\alpha_{i+m+l}, \tag{4.3.45}$$

with the convention that $\alpha_i = 0$ for $i < 0$, and $\phi_m = \phi_{|m|}$, if $m < 0$. In addition, put $\rho_{i,j}^{(n)} = \mathbb{E}\big(S_{n,i}S_{n,j}\big)/\sqrt{\mathbb{E}\big(S_{n,i}^2\big)\mathbb{E}\big(S_{n,j}^2\big)}$. We then have the following two results.

Lemma 4.3.14. *Suppose that* $0 < \|\epsilon\|_4 < \infty$*. If*

- $|\alpha_i| > 0$ *for at least one* $i \in \mathbb{N}_0$*,*
- $|\alpha_i| = \mathcal{O}\left(i^{-\beta}\right)$*, with* $\beta > 3/2$*,*

then

$$\mathbb{E}\big(S_{n,k}S_{n,l}\big) = \sum_{m=-\infty}^{\infty} T_m + \mathcal{O}\left(n^{-1/3}d_n^{1/2}\right). \tag{4.3.46}$$

In particular, it holds that $\inf_{h\ge L}\varphi_{h,h} > 0$*, for some finite* $L \ge 0$*.*

Lemma 4.3.15. *Suppose that* $\|\epsilon\|_4 < \infty$*, and that*

(i) $\sum_{m=0}^{\infty} |\phi_m| < \infty$*,*

(ii) $\limsup_{h\to\infty}\sup_{m\ge L}\left|\frac{\phi_{m+h}}{\phi_m}\right| < 1$*, for some finite* $L > 0$*.*

Then we have

$$\limsup_{n\to\infty} \sup_{i,j \geq M_0 \colon 1 \leq |i-j|} \left|\rho_{i,j}^{(n)}\right| < 1,$$

for some finite $M_0 > 0$.

Remark 4.3.16. Note that since $\lim_{m\to\infty} |\phi_m| = 0$, one can actually assume that $L = 0$ in condition (ii) of Lemma 4.3.15.

We are now ready to proof Theorem 4.2.6.

Proof of Theorem 4.2.6. The assumption $|\alpha_i| = \mathcal{O}\left(i^{-\beta}\right)$ of Theorem 4.2.6 together with Lemma 4.3.13 implies that Assumption 4.2.1 (i) is valid, and (ii) follows from Lemma 4.3.14. Moreover, discarding the first M_0 elements of

$$\max_{1\leq h\leq d_n} \left|S_{n,h} \mathrm{Var}\big(S_{n,h}\big)^{-1/2}\right|$$

has no effect on the limit distribution, and thus it suffices to establish that $r_1 = \sup_{|h-l|\geq 1} |\varphi_{h,l}| < 1$ for $\min\{h,l\} \geq M_0$, where M_0 is finite. This, however, follows directly from Lemma 4.3.15. In addition, the estimate (4.3.46) together with the definition of T_m in (4.3.45) implies that $r_n (\log n)^3 \to 0$, which completes the proof. □

Proof of Lemma 4.3.13. We have that

$$\begin{aligned}
\left\|X_{k,h} - X'_{k,h}\right\|_4 &\leq \Big\|\epsilon'_0\alpha_k \sum_{\substack{i=0,\\ i\neq k+h}}^{\infty} \alpha_i\epsilon_{k+h-i} + \epsilon'_0\alpha_{k+h} \sum_{\substack{i=0,\\ i\neq k}}^{\infty} \alpha_i\epsilon_{k-i} + (\epsilon'_0)^2\alpha_k\alpha_{k+h}\Big\|_4 \\
&\quad + \Big\|\epsilon_0\alpha_k \sum_{\substack{i=0,\\ i\neq k+h}}^{\infty} \alpha_i\epsilon_{k+h-i} + \epsilon_0\alpha_{k+h} \sum_{\substack{i=0,\\ i\neq k}}^{\infty} \alpha_i\epsilon_{k-i} + (\epsilon_0)^2\alpha_k\alpha_{k+h}\Big\|_4 \\
&\leq C\big(|\alpha_k| + |\alpha_{k+h}|\big),
\end{aligned}$$

which completes the proof. □

Proof of Lemma 4.3.14. Put $R_{n,k} = n^{-1/2}\sum_{k=1}^{n}\big(L_kL_{k+h} - \phi_h\big)$.

An application of the Cauchy-Schwarz inequality yields

$$\mathbb{E}\big(S_{n,k}S_{n,l}\big) = \mathbb{E}\big(R_{n,k}R_{n,l}\big) + \mathcal{O}\left(n^{-1/2}d_n^{1/2}\right),$$

hence it suffices to consider $\mathbb{E}\big(R_{n,k}R_{n,l}\big)$. Due to [25, section 7.2], it holds that

$$\mathbb{E}\big(R_{n,k}, R_{n,l}\big) = \sum_{|m|<n} \frac{n-|m|}{n} T_m, \tag{4.3.47}$$

in particular, we have

$$\varpi_{k,l} = \lim_{n\to\infty} \mathbb{E}\big(R_{n,k}, R_{n,l}\big) = (\eta-3)\phi_k\phi_l + \sum_{m=-\infty}^{\infty} \big(\phi_m\phi_{m+k-l} + \phi_{m+k}\phi_{m-l}\big). \tag{4.3.48}$$

Moreover, we have the decomposition

$$\sum_{|m|\leq n} \frac{n-m}{n} T_m = \sum_{m=-\infty}^{\infty} T_m + \sum_{|m|>n} T_m + n^{-1} \sum_{|m|\leq n} mT_m.$$

Note that for $k \geq l$, we have

$$\begin{aligned}\sum_{|m|>K} \big|T_m\big| &\leq C\left\{\sum_{|m|>K} |\alpha_m| + \sum_{|m|>K} \big(|\phi_m| + |\phi_{m+k}|\big)\right\}\\ &\leq C\left\{\sum_{i>K}|\alpha_i| + \sum_{|m|>K}|\phi_m|\right\} := \theta_K,\end{aligned}$$

with $\lim_{K\to\infty}\theta_K = 0$. Thus we get the estimate

$$n^{-1}\sum_{|m|\leq n} m|T_m| \leq \lambda \sum_{|m|\leq \lambda n} |T_m| + \sum_{|m|>n\lambda} |T_m| = \mathcal{O}\left(\theta_{\lambda n} + \lambda\right).$$

Note that

$$\sum_{|m|>K}^{\infty} |\phi_m| \leq 2\sum_{m>K}^{\infty}\sum_{i=0}^{\infty} |\alpha_i\alpha_{i+m}| \leq C\sum_{m>K}^{\infty}|\alpha_i|, \tag{4.3.49}$$

which implies $\theta_K = o\left(K^{-1/2}\right)$. Hence choosing $\lambda = \lambda_n = n^{-1/3}$ gives

$$n^{-1}\sum_{|m|\leq n} m|T_m| = \mathcal{O}\left(n^{-1/3}\right),$$

which implies (4.3.46). Finally note that since $\lim_{m\to\infty}|\phi_m| = 0$, we have

$$\limsup_h \sup_{k\geq h}\Big|\sum_{m=-\infty}^{\infty} T_m - \sum_{m=-\infty}^{\infty}\phi_m^2\Big| = 0, \tag{4.3.50}$$

hence $\inf_{h\geq L}\varphi_{h,h} > 0$, for some finite $L \geq 0$. □

Proof of Lemma 4.3.15. Due to Lemma 4.3.14, we can assume that

$$\mathbb{E}\big(S_{n,k}, S_{n,l}\big) = (\eta - 3)\phi_k\phi_l + \sum_{m=-\infty}^{\infty} \big(\phi_m\phi_{m+k-l} + \phi_{m+k}\phi_{m-l}\big). \tag{4.3.51}$$

Now suppose first that $\min\{k, l\} \geq M_0$, for some $M_0 > 0$. Thus, since $\lim_{m\to\infty} |\phi_m| = 0$, we have from (4.3.51) that

$$\varpi_{k,l} = \sum_{m=-\infty}^{\infty} \phi_m\phi_{m+k-l} + \epsilon^{(M_0)}, \tag{4.3.52}$$

where $\epsilon^{(M_0)} \downarrow 0$ as M_0 increases. Condition (ii) and Remark 4.3.16 imply that for some $K \geq L$ large enough, we have

$$\sup_m \left|\frac{\phi_{m+h}}{\phi_m}\right| \leq \vartheta^{(K)} < 1, \;\; K \leq h. \tag{4.3.53}$$

Hence we obtain

$$\sqrt{\varpi_{k,k}\varpi_{l,l}} \geq \sum_{m=-\infty}^{\infty} \phi_m^2 - \mathcal{O}\left(\epsilon^{(M_0)}\right), \tag{4.3.54}$$

and

$$\sum_{m=-\infty}^{\infty} \phi_m^2 - \sum_{m=-\infty}^{\infty} \phi_m\phi_{m+h} \geq \left(1 - \vartheta^{(K)}\right) \sum_{m=-\infty}^{\infty} \phi_m^2.$$

Hence, for large enough but fixed M_0 and K, we deduce that that the Cauchy-Schwarz inequality is strict, i.e.

$$\big|\rho_{i,j}^{(n)}\big| < 1. \tag{4.3.55}$$

Now suppose that $h < K$. Then the Cauchy-Schwarz inequality (in l_2) implies that

$$\left|\sum_{m=-\infty}^{\infty} \phi_m\phi_{m+h}\right| \leq \sum_{m=-\infty}^{\infty} \phi_m^2,$$

and we have equality if and only if $v_1 = \lambda v_2$, $\lambda \in \mathbb{R}$ and $v_1 = (..., \phi_m, ...)^T$, $v_2 = (...\phi_{m+h}, ...)^T$. This implies that $\phi_m = \lambda\phi_{m+h}$, and consequently

$$\phi_0 = \lambda\phi_h = \lambda^2\phi_{2h} = ... = \lambda^n\phi_{nh}. \tag{4.3.56}$$

Since $|\phi_{nh}| \to 0$, we must have $|\lambda| > 1$. We thus conclude that

$$\left| \sum_{m=-\infty}^{\infty} \phi_m - \sum_{m=-\infty}^{\infty} \phi_m \phi_{m+h} \right| = |\lambda - 1| \sum_{m=-\infty}^{\infty} \phi_m^2 > 0.$$

Since K is finite, we deduce that

$$\min_{1 \leq h < K} \left| \sum_{m=-\infty}^{\infty} \phi_m - \sum_{m=-\infty}^{\infty} \phi_m \phi_{m+h} \right| = \epsilon^{(K)} > 0,$$

which together with (4.3.54) implies that for large enough (but finite) M_0, we have (4.3.55), which completes the proof. □

Chapter 5

Extremes of Yule-Walker Estimators

5.1 Introduction

Let $\{X_k\}_{k\in\mathbb{Z}}$ be a q-th order autoregressive process AR(q) with coefficient vector $\mathbf{\Theta}_q \in \mathbb{R}^q$. A considerable literature in the past years dealt with various aspects and problems on AR(q)-processes, see for instance [4, 25, 60, 85] and the references therein. More recently, people have moved on to more complicated models such as ARCH [22, 48], GARCH [21] and related models, which again have been extended in many different directions. However, in many applications, AR(q)-processes still form the backbone and are often used as first approximations for further analysis, in particular, many estimation and fitting procedures can be based on preliminary AR(q) approximations. This includes for instance ARMA, ARCH and GARCH models ([20, 56, 62]). Thus, AR(q) processes have moved from the spotlight to the backstage area, yet their significance remains unchallenged.

When fitting an AR(q) model, two important questions arise: how to choose the order q, and having done so, which estimators are to be used. Naturally, these two problems can hardly be separated and are often dealt with simultaneously, or at least so in preliminary estimates. An extensive literature has evolved around these two issues, pioneering contributions in this direction are due to Akaike [1, 2], Mallows [86, 87], Walker [123] and Yule [134], for more details we refer to [4, 24, 25, 60, 85]. In order to be able to describe some of the basic results, we recall that an AR(q) process $\{X_k\}_{k\in\mathbb{Z}}$ is defined through the recurrence relation

$$X_k = \theta_1 X_{k-1} + ... + \theta_q X_{k-q} + \epsilon_k, \tag{5.1.1}$$

where it is often assumed that $\{\epsilon_k\}_{k\in\mathbb{Z}}$ is an IID sequence. Let $\phi_h = \mathbb{E}\big(X_k X_{k+h}\big)$, $k, h \in \mathbb{Z}$ be the covariance function. A natural estimate for ϕ_h is the sample covariance $\widehat{\phi}_{n,h} = \frac{1}{n}\sum_{i=h+1}^{n} X_i X_{i-h}$. Depending on the magnitude of h, a different normalization, such as $(n-h)^{-1}$ is sometimes more convenient. Denote with $\mathbf{\Theta}_q = \big(\theta_1, ..., \theta_q\big)^T$ the parameter vector and put $\mathbf{\Phi}_q = \big(\phi_1, ..., \phi_q\big)^T$, and let $\mathbf{\Gamma}_q = \big(\phi_{|i-j|}\big)_{1 \leq i,j \leq q}$ be the $q \times q$ dimensional covariance matrix. Then it follows from (5.1.1) that

$$\mathbf{\Gamma}_q \mathbf{\Theta}_q = \mathbf{\Phi}_q. \tag{5.1.2}$$

Hence a natural idea is to replace the corresponding quantities by estimators $\widehat{\mathbf{\Phi}}_q = \big(\widehat{\phi}_{n,1}, ..., \widehat{\phi}_{n,q}\big)^T$, $\widehat{\mathbf{\Gamma}}_q = \big(\widehat{\phi}_{n,|i-j|}\big)_{1 \leq i,j \leq q}$, and thus define the estimator $\widehat{\mathbf{\Theta}}_q = \big(\widehat{\theta}_1, ..., \widehat{\theta}_q\big)^T$ via

$$\widehat{\mathbf{\Gamma}}_q^{-1} \widehat{\mathbf{\Phi}}_q = \widehat{\mathbf{\Theta}}_q \text{ and } \widehat{\sigma}^2(q) = \widehat{\phi}_0 - \widehat{\mathbf{\Theta}}_q^T \widehat{\mathbf{\Phi}}_q, \tag{5.1.3}$$

where $\sigma^2 = \mathbb{E}\big(\epsilon_0^2\big)$. These estimators are commonly referred to as the Yule-Walker estimators, and they have some remarkable properties. For example, if $\big\{X_k\big\}_{k \in \mathbb{Z}}$ is causal, then the fitted model

$$X_k = \widehat{\theta}_1 X_{k-1} + ... + \widehat{\theta}_p X_{k-q} + \epsilon_k$$

is still causal, see for instance [25] and [93]. Another interesting feature is that even though the Yule-Walker estimators are obtained via moment matching methods, their variance is asymptotically equivalent with those obtained via a maximum likelihood approach. More precisely, for $m \geq q$ it holds that

$$\sqrt{n}\big(\widehat{\mathbf{\Theta}}_m - \mathbf{\Theta}_m\big) \xrightarrow{d} \mathcal{N}\big(0, \sigma^2 \mathbf{\Gamma}_m^{-1}\big), \tag{5.1.4}$$

where $\mathbf{\Theta}_m = \big(\theta_1, ..., \theta_q, 0, ..., 0\big)^T$, see for instance [25]. These asymptotic results form the basis for earlier estimation methods of the order q ([105, 122, 126]), which focused on a fixed, finite number of possible orders and consist of multiple-testing-procedures, which in practice leads to the difficulty of having a required level. On the other hand, as it was pointed out by Shwarz [116], a direct likelihood approach fails, since it invariably chooses the highest possible dimension. Akaike [1] and Mallows [86, 87], developed a different approach, which is based on a 'generalized likelihood function', which has been further developed by various authors. Shibata [114] investigated the asymptotic distribution and showed that the estimator defined as in (5.1.5) is not consistent. This issue was successfully dealt with by Akaike [2] (BIC), Hannan and Quinn [63] (HQC), Parzen [99], Rissanen [108] and Shwarz [116] (SIC), who introduced consistent modifications (Parzen's CAT-criterion is conceptually different). For more recent advances and generalizations, see for instance Barron et al. [12], Foster and

George [50], Shao [111] and the detailed review on model selection given by Leeb and Pötscher [82]. Here and now, we will content ourselves with briefly discussing Akaike's approach and closely related criteria. Akaike's generalized likelihood function leads to the expression

$$\mathrm{AIC}(m) = n \log \widehat{\sigma}^2(m) + 2m, \tag{5.1.5}$$

where n is the sample size and $\widehat{\sigma}^2(m)$ is as in (5.1.3). The order q is then obtained by minimizing $\mathrm{AIC}(m)$, $m \in \{0, 1, ..., K\}$, for some predefined $K > 0$. Consistent modifications are obtained by inserting an increasing sequence C_n, and $\mathrm{AIC}(m)$ then becomes

$$\widetilde{\mathrm{AIC}}(m) = n \log \widehat{\sigma}^2(m) + 2C_n m. \tag{5.1.6}$$

Most modifications result in $C_n = \mathcal{O}(\log n)$, even though the arguments are sometimes quite different. A notable exception is the idea of Hannan and Quinn [63], who successfully employed the LIL to obtain $C_n = \mathcal{O}(\log \log n)$.

The aim of this chapter is to introduce a different approach, based on the quantity $\max_{1 \leq i \leq d_n} \big|\widehat{\theta}_i - \theta_i\big|$, where d_n is an increasing function in n. It is shown for instance that appropriately normalized, this expression converges weakly to a Gumbel-type distribution. On one hand, this allows to construct simultaneous confidence bands for the Yule-Walker-estimators $\widehat{\mathbf{\Theta}}_{d_n}$, but also permits us to construct a variety of different, consistent estimators for the order q of an autoregressive process. The asymptotic distribution of such a particular estimator is also derived. As a byproduct, it is shown that known consistent criteria such as BIC, SIC and HQC are also consistent if the parameter space is increasing, i.e. consistency even holds if $q \in \{0, ..., d_n\}$, where $d_n = \mathcal{O}\left(n^{\delta}\right)$. This partially gives answers to questions raised by Hannan and Quinn [63], Shibata [114], and extends results given by An et al. [3]. In addition, the general method seems to be very useful for model fitting for subset autoregressive processes (see for instance [89]), which is highlighted in Remark 5.2.7 and Section 5.3.

5.2 Main Results

We will frequently use the following notation. For a matrix $\mathbf{A} = \left(a_{i,j}\right)_{\left\{\substack{1 \leq i \leq r, \\ 1 \leq j \leq s}\right\}}$, $r, s \in \mathbb{N}$ we put

$$\max\big|\mathbf{A}\big| = \max_{1 \leq i \leq r, 1 \leq j \leq s} \big|a_{i,j}\big|. \tag{5.2.1}$$

In addition, we will use the abbreviation $\|.\|_p = \left(\mathbb{E}(|.|^p)\right)^{1/p}$. The main results are dealing with an array of AR(q) processes, more precisely, we consider the family of AR(d_n) processes $\{X_k^{(r)}\}_{k\in\mathbb{Z}}$, $1 \leq r \leq d_n$, where $d_n = \mathcal{O}(n^\delta)$ (more details are given later). Since we are always only dealing with a single member of this array, the index (r) is dropped for convenience, and we just consider an AR(d_n) process $\{X_k\}_{k\in\mathbb{Z}}$, keeping in mind that the parameters $\{\theta_i\}_{1\leq i\leq d_n}$ may depend on n. This implies that X_k satisfies the recurrence relation

$$X_k = \theta_1 X_{k-1} + ... + \theta_{d_n} X_{k-d_n} + \epsilon_k, \quad k \in \mathbb{Z}, \tag{5.2.2}$$

where $\{\epsilon_k\}_{k\in\mathbb{Z}}$ defines the usual innovations. We specifically point out that $\{\theta_i\}_{1\leq i\leq d_n}$ are not required to be all different from zero, only one parameter should be nonzero to exclude the trivial case. All of the results are derived under the following assumption regarding the AR(d_n) process $\{X_k\}_{k\in\mathbb{Z}}$.

Assumption 5.2.1.

- $\sum_{i=1}^{d_n}|\theta_i| \leq \vartheta < 1$, *where* ϑ *does not depend on* n,
- $\{\epsilon_k\}_{k\in\mathbb{Z}}$ *is a mean zero IID-sequence of random variables, such that* $\|\epsilon_k\|_p < \infty$ *for some* $p > 4$, $\|\epsilon_k\|_2^2 = \sigma^2$, $k \in \mathbb{Z}$.

We point out that Assumption 5.2.1 implies that the AR(d_n)-process $\{X_k\}_{k\in\mathbb{Z}}$ is causal, and admits the representation

$$X_k = \sum_{i=0}^{\infty} \psi_i \epsilon_{k-i}, \quad \text{with } |\psi_i| \leq \vartheta^i,$$

see also Lemma 5.6.1. Also, it is clear from the proofs that condition $\sum_{i=1}^{d_n}|\theta_i| \leq \vartheta < 1$ can be considerably relaxed and reformulated for instance in terms of the eigenvalues of the covariance matrix $\mathbf{\Gamma}_{d_n}$ and the spectral density function of the process $\{X_k\}_{\{k\in\mathbb{Z}\}}$ (see also [13]). However, the current formulation significantly simplifies the notation and some of the proofs.

In accordance to the previously established notation, we introduce the inverse and estimated inverse matrix

$$\mathbf{\Gamma}_{d_n}^{-1} = (\gamma_{i,j}^*)_{1\leq i,j\leq d_n}, \quad \widehat{\mathbf{\Gamma}}_{d_n}^{-1} = (\widehat{\gamma}_{i,j}^*)_{1\leq i,j\leq d_n}. \tag{5.2.3}$$

In addition, we will use the convention that $\theta_0 = \widehat{\theta}_0 = -1$. Put

$$r_{k,n} = \max_{k\leq h, i\leq d_n}\left|\sum_{r=0}^{i-1}\theta_r\theta_{r+h} - \sum_{r=d_n+1-i-h}^{d_n-h}\theta_r\theta_{r+h}\right|, \tag{5.2.4}$$

where a sum is taken to be zero if its upper limit is less than its lower limit. We can now formulate our main result.

Theorem 5.2.2. *Let $\{X_k\}_{k\in\mathbb{Z}}$ be an AR(d_n) process satisfying Assumption 5.2.1. Suppose that $d_n = \mathcal{O}\left(n^{\delta(p)}\right)$ and that $\sup_n r_{k,n}(\log k)^{2+\beta} \to 0$, for some $\beta > 0$. Then for $z \in \mathbb{R}$*

$$P\left(a_n^{-1}\big(\sqrt{n}\max_{1\le i\le d_n}\big|(\widehat{\gamma}^*_{i,i}\widehat{\sigma}^2(d_n))^{-1/2}(\widehat{\theta}_i-\theta_i)\big| - b_n\big) \le z\right) \to \exp(-e^{-z}),$$

where $a_n = (2\log d_n)^{-1/2}$ and $b_n = (2\log d_n)^{1/2} - (8\log d_n)^{-1/2}(\log\log d_n + 4\pi - 4)$.

Remark 5.2.3. The exact specification of $\delta(p)$ is not very simple and involves a maximization-minimization problem, which is given in Theorem 5.4.3. An upper bound is $\delta(p) < 1/10$, and thus for smaller n the restriction $n^{\delta(p)}$ is rather strong. However, numerical results (see Section 5.3) indicate that at least in the cases of $n \in \{125, 250, 500, 1000\}$ the true bound is bigger, good results were obtained for instance with $d_n \in \{2\log n, 4\log n, 6\log n\}$.

The proof of Theorem 5.2.2 essentially consists of two main steps. The first step deals with estimating the error $\max\big|\widehat{\mathbf{\Gamma}}_{d_n}^{-1} - \mathbf{\Gamma}_{d_n}^{-1}\big|$ (Theorem 5.4.2), the second step with approximating $\sqrt{n}(\mathbf{\Theta}_{d_n} - \mathbf{\Theta}_{d_n})$ by an appropriate Gaussian random vector (Theorem 6.6.4). The partial result of the Gaussian approximation also has a practical relevance, since it gives us an upper bound on how close the expression $\sqrt{n}\max_{1\le i\le d_n}\big|(\widehat{\gamma}^*_{i,i}\widehat{\sigma}^2(d_n))^{-1/2}(\widehat{\theta}_i - \theta_i)\big|$ is to a correspondingly transformed Gaussian distribution, see Remark 5.4.7 for more details.

The above result allows us to construct the simultaneous confidence bands

$$\big\{\mathbf{\Theta}_{d_n} \in \mathbb{R}^{d_n} \,\big|\, a_n^{-1}\big(\max_{1\le i\le d_n}\big|(\widehat{\gamma}^*_{i,i})^{-1/2}(\widehat{\theta}_i - \theta_i)\big| - b_n\big) \le n^{-1/2}\sqrt{\widehat{\sigma}^2(d_n)}V_{1-\alpha}\big\}, \tag{5.2.5}$$

where $V_{1-\alpha}$ denotes the $1-\alpha$ quantile of the Gumbel-type distribution given above. In the literature ([4, 25, 60]) one often finds the confidence ellipsoids

$$\big\{\mathbf{\Theta}_m \in \mathbb{R}^m \,\big|\, \big(\widehat{\mathbf{\Theta}}_m - \mathbf{\Theta}_m\big)\mathbf{\Gamma}_m\big(\widehat{\mathbf{\Theta}}_m - \mathbf{\Theta}_m\big)^T \le n^{-1}\widehat{\sigma}^2(d_n)\chi^2_{1-\alpha}(m)\big\}, \tag{5.2.6}$$

where $\chi^2_{1-\alpha}(m)$ denotes the $1-\alpha$ quantile of the chi-square distribution with m degrees of freedom. By the law of large numbers, we have $\chi^2_{1-\alpha}(d_n) \approx d_n$, and since $\big(a_nV_{1-\alpha} + b_n\big)^2 \approx \log d_n$ the confidence band given in (5.2.5) is asymptotically much smaller than the one in (5.2.6). In addition, (5.2.5) immediately gives bounds for the single elements $\big\{\big|\widehat{\theta}_i - \theta_i\big|\big\}_{1\le i\le d_n}$.

Theorem 5.2.2 can not only be used to construct simultaneous confidence bands for the Yule-Walker estimators $\widehat{\mathbf{\Theta}}_{d_n}$, but also provides a test for the degree of an an AR(q)-process. To be more precise, for an AR(q)-process $\{X_k\}_{k\in\mathbb{Z}}$ satisfying the assumptions of Theorem 5.2.2, we formulate the null hypothesis

$\mathcal{H}_0 : q \leq q_0$, and the alternative $\mathcal{H}_A : q > q_0$. It follows immediately from Theorem 5.2.2 that under $\mathcal{H}_0$ we have

$$P\left(a_n^{-1}\sqrt{n}\big(\max_{q_0+k\leq i\leq d_n}\big|(\widehat{\gamma}^*_{i,i}\widehat{\sigma}^2(d_n))^{-1/2}\widehat{\theta}_i\big| - b_n\big) \leq z\right) \to \exp(-e^{-z}),$$

for any fixed integer $k > 0$, since we are assuming that $\theta_i = 0$ for $i > q_0$. Conversely, it is not hard to verify (see the proof of Theorem 5.2.4 for details) that the quantity

$$a_n^{-1}\sqrt{n}\big(\max_{q_0+1\leq i\leq d_n}\big|(\widehat{\gamma}^*_{i,i}\widehat{\sigma}^2(d_n))^{-1/2}\widehat{\theta}_i\big| - b_n\big)$$

explodes under the alternative $\mathcal{H}_A : q > q_0$. This can be used to establish a lower bound for the order q or to test if the order was chosen sufficiently large. This is particularly useful if q is large compared with the sample size and the magnitude of $\mathbf{\Theta}_q$, in which case the AIC and related criteria sometimes heavily fail to get anywhere near the true order. More details on this subject and examples are given in Section 5.3. Generally speaking, such situations are often encountered in so-called *subset autoregressive* models, see Remark (5.2.7).

The above conclusions lead to the following family of estimators $\widehat{q}^{(1)}_{z_n}$ for q. Let z_n be a monotone sequence that tends to infinity as n increases. Then we define the estimator

$$\widehat{q}^{(1)}_{z_n} = \min\big\{q \in \mathbb{N}\big|\, a_n^{-1}\sqrt{n}\big(\max_{q+1\leq i\leq d_n}\big|(\widehat{\gamma}^*_{i,i}\widehat{\sigma}^2(d_n))^{-1/2}(\widehat{\theta}_i - \theta_i)\big| - b_n\big) \leq z_n\big\}. \tag{5.2.7}$$

Using the above ideas, it is not hard to show that the estimators $\widehat{q}^{(1)}_{z_n}$ are consistent if z_n does not grow too fast. In fact, under some more conditions imposed on the sequence z_n, we can even derive the asymptotic distribution of the estimators.

Theorem 5.2.4. *Let $\{X_k\}_{k\in\mathbb{Z}}$ be an AR(q)-process, and assume that the conditions of Theorem 5.2.2 are satisfied. Assume in addition that $z_n = o\left(\sqrt{n(\log n)^{-1}}\right)$, and $\kappa_n d_n \to 0$ as n increases, where κ_n is as in Theorem 5.4.3. Then if $z_n \to \infty$, the estimator $\widehat{q}^{(1)}_{z_n}$ in (5.2.7) is consistent. Moreover, the following expansion is valid.*

$$P\big(\widehat{q}^{(1)}_{z_n} = k + q\big) = \frac{e^{-z_n}}{d_n} + o\left(\frac{e^{-z_n}}{d_n} + d_n^{-z_n^2+1}\right) + \mathcal{O}\left(\kappa_n d_n\right)$$

for $k \in \mathbb{N}$, $k > 0$.

This immediately gives us the following corollary.

Corollary 5.2.5. *Assume that the conditions of Theorem 5.2.4 are satisfied. If* $z_n = \mathcal{O}\left(|\log \kappa_n| + |2\log d_n|\right)$*, then*

$$P\big(\widehat{q}^{(1)}_{z_n} = k+p\big) = \frac{e^{-z_n}}{d_n} + \mathcal{O}\left(\frac{e^{-z_n}}{d_n}\right)$$

for $k \in \mathbb{N}$*,* $k > 0$*.*

Remark 5.2.6. From the above corollary we obtain that in some sense, the estimators $\widehat{q}^{(1)}_{z_n}$ possess a discrete uniform asymptotic distribution, which yields the surprising conclusion

$$P\big(\widehat{q}^{(1)}_{z_n} = 1+q\big) \approx P\big(\widehat{q}^{(1)}_{z_n} = 1000+q\big).$$

However, this fact can be explained by the maximum function in the definition of $\widehat{q}^{(1)}_{z_n}$, more precisely, due to the weak dependence of the Yule-Walker estimators $\widehat{\mathbf{\Theta}}_{d_n}$. The maximum function essentially does not care at which index i the boundary z_n is exceeded, and this results in the uniform distribution. It turns out (see Section 5.3) that a modified version of the estimator $\widehat{q}^{(1)}_{z_n}$ is a very efficient preliminary estimator that establishes a decent lower bound.

An asymptotic uniform-type distribution clearly is not a desirable property for an estimator. However, similar to Akaike's method, we can introduce a penalty function and construct different yet also consistent estimators for the order q. To this end, for $x \in \mathbb{R}$ put $(x)^+ = \max(0,x)$ and let $\Upsilon_{n,i} = a_n^{-1}\sqrt{n}\big(\big|(\widehat{\gamma}^*_{i,i}\widehat{\sigma}^2(d_n))^{-1/2}\widehat{\theta}_i\big| - b_n\big)$. Then we introduce a new estimator $\widehat{q}^{(2)}_{z_n}$ as

$$\widehat{q}^{(2)}_{z_n} = \underset{q\in\mathbb{N}}{\operatorname{argmin}}\Big\{\max_{q+1\le i\le d_n}\big\{\big(\Upsilon_{n,i} - z_n\big)^+\big\} + \log(1+q)\Big\}.$$

More generally, let $\mathcal{F} = \big(f_d\big)_{d\in\mathbb{N}}$ be a collection of continuous functions such that

- f_d is a map from $\mathbb{R}^{d+2}$ to $\mathbb{R}$,
- $f_d(0,...,0,q,d) < f_d(0,...,0,q+1,d)$ for all $d,q \in \mathbb{N}$,
- if $a_n, d_n \to \infty$ as n increases, then $f_{d_n}(...,a_n,...,q,d_n) \to \infty$ as n increases, regardless of the values of the other coordinates.

Define

$$\widehat{q}^{(f)}_{z_n} = \underset{q\in\mathbb{N}}{\operatorname{argmin}} f_{d_n}\big(0,...,0,(\Upsilon_{n,q+1} - z_n)^+,...,(\Upsilon_{n,d_n} - z_n)^+, q, d_n\big). \tag{5.2.8}$$

Then arguing as in the proof of Theorem 5.2.4 it can be shown that this constitutes a consistent estimator for the true value q. For example, the following estimator

$$\widehat{q}_{z_n}^{(3)} = \operatorname*{argmin}_{q \in \mathbb{N}} \Big\{ \sum_{q+1 \leq i \leq d_n} \big(\Upsilon_{n,i} - z_n\big)^+ + q \Big\},$$

satisfies the conditions above and is consistent.

Remark 5.2.7. Note that instead of defining a specific order q, one can also consider a special lag configuration, for example, $\mathbf{\Theta}_q = \big(\theta_1, \theta_2, 0, ...0, \theta_{10}, \theta_{11}, ..., \theta_q\big)^T$. Such configurations are commonly referred to as *subset autoregressive models*, see for instance [26, 88, 89, 119, 127] and the references therein. The AIC(m) and especially related consistent criteria have problems dealing with such *subset autoregressive models*, which can be seen as follows. By Hannan [60, Chapter VI], we have for $m \in \mathbb{N}$

$$\widetilde{\mathrm{AIC}}(m)n^{-1} = \log\big(\widehat{\sigma}^2(m)\big) + 2n^{-1}C_n m = \log \widehat{\phi}_{n,0} + \sum_{j=1}^{m} \log(1 - \widehat{\theta}_j^2(m)) + 2n^{-1}C_n m. \tag{5.2.9}$$

For large enough n we can assume that $\widehat{\theta}_j^2(m) < 1$ and hence

$$\sum_{j=1}^{m} \log(1 - \widehat{\theta}_j^2(m)) \approx -\sum_{j=1}^{m} \widehat{\theta}_j^2(m).$$

This shows that in case of *subset autoregressive models*, the penalty function $2n^{-1}C_n m$ is too severe and should be replaced, at least in theory, by $2n^{-1}C_n \sum_{\left\{\substack{1 \leq i \leq m, \\ \theta_i \neq 0}\right\}} 1$, since this is impossible in practice. Of course the same problem arises if some of the $\big\{\theta_i\big\}_{1 \leq i \leq q}$ are close to zero. On the other hand, a maximum based estimator like $\widehat{q}_{z_n}^{(1)}$ does not face this problem. This is empirically confirmed in Section 5.3.

An important theoretical assumption for estimators related to AIC(m) is that the parameter space for q is finite, i.e; it is usually assumed in advance that $q \in \{1, .., K\}$, where K is 'chosen sufficiently large', but finite. In [63], K is allowed to increase with the sample size with unknown rate, which was specified later by An et al. [3]. Note however that for the estimators defined above we allow $K = K_n = d_n$. Before extending this result, we give precise definitions of BIC, HQC and SIC, as the literature does not seem to be very clear on this subject, in particular in the case of the BIC and SIC. In the sequel, the following definitions are used.

$$\begin{aligned}
\mathrm{BIC}(m) &= \log \widehat{\sigma}^2(m) + mn^{-1} \log n, \\
\mathrm{SIC}(m) &= \log \widehat{\sigma}^2(m) + m/2n^{-1} \log n, \\
\mathrm{HQC}(m) &= \log \widehat{\sigma}^2(m) + n^{-1} 2cm \log\log n, \quad c > 1.
\end{aligned} \tag{5.2.10}$$

In case of the BIC, HQC, An et al. [3] obtained the rates $K_n = \mathcal{O}(\log(n))$ (BIC) and $K_n = \mathcal{O}(\log\log(n))$ (HQC), in particular, it is always assumed that the true order q is fixed.

Using some of the results of Section 5.4 and 5.5, this result can be extended.

Theorem 5.2.8. *Let $\{X_k\}_{k\in\mathbb{Z}}$ be an AR(q) process satisfying Assumption 5.2.1. Assume in addition that C_n, K_n are positive sequences such that*

- $\lim_n C_n(2\log\log n)^{-1} > 1$, $C_n = \scriptstyle\mathcal{O}\displaystyle(n)$,
- $4\log K_n \leq C_n$, $K_n = \scriptstyle\mathcal{O}\displaystyle(\sqrt{n})$.

Then the estimators for the order q defined as

$$\widehat{q}_n^* = \underset{0\leq m\leq K_n}{\operatorname{argmin}}\big(\log\widehat{\sigma}^2(m) + n^{-1}C_n m\big),$$

are consistent. If $\{X_k\}_{k\in\mathbb{Z}}$ is an AR(d_n) process where $d_n = \mathcal{O}\left(n^{\delta(p)}\right)$ is as in Theorem 5.2.2, then the above defined $\widehat{q}_n^*$ is also consistent.

Theorem 5.2.8 thus implies the bounds $K_n \in \{n^{1/4}, n^{1/8}, (\log n)^{c/2}\}$ for BIC, SIC and HQC. In particular, it is not required that the true order q is fixed. On the other hand, the setting in An et al. [3] is more general, and they also show that the estimators are strongly consistent.

5.3 Simulation and numerical results

In this section we will perform a small simulation study to compare some of the previously mentioned estimators.

We will look at the performance in case of AR(6), AR(12) and AR(24) processes. The sample size n satisfies $n \in \{125, 250, 500, 1000\}$, as for the dimension d_n, we chose the functions $d_n \in \{2\log n, 4\log n, 6\log n\}$, and rounded up the values. This implies that the parameter space $q \in \{0, .., K\}$ satisfies $K \in \{10, 12, 13, 14\}$, $K \in \{20, 23, 25, 29\}$, $K \in \{29, 34, 38, 42\}$.

To introduce the estimators $\widehat{q}_{z_n}^{(4)}(d_n), \widehat{q}_{z_n}^{(5)}(d_n)$, we require some additional notation. For $1 \leq k \leq d_n$, define $\{\widehat{\gamma}_{i,i}^*(k)\}_{1\leq i\leq k}$ and $\{\widehat{\theta}_i(k)\}_{1\leq i\leq k}$ via the usual relation

$$\widehat{\Theta}_k = \widehat{\mathbf{\Gamma}}_k^{-1}\widehat{\Phi}_k. \tag{5.3.1}$$

The estimators are now defined as

$$\widehat{q}_{z_n}^{(4)}(k) = \min\big\{q\in\mathbb{N}\,\big|\, a_n^{-1}\sqrt{n}\big(\max_{q+1\leq i\leq k}\big|(\widehat{\gamma}_{i,i}^*(k)\widehat{\sigma}^2(k))^{-1/2}\widehat{\theta}_i(k)\big| - b_n\big) \leq z_n\big\},$$

$$\widehat{q}_{z_n}^{(5)}(d_n) = \max_{1\leq k\leq d_n}\widehat{q}_{z_n}^{(4)}(k).$$

Note that the definition of a_n, b_n remains unchanged. Usually, the 'maximum' version $\widehat{q}^{(5)}_{z_n}(d_n)$ outperforms its counterpart $\widehat{q}^{(4)}_{z_n}(d_n)$, at least in the examples given below. The values for z_n were chosen as $z_n \in \{x_n, y_n\}$, where x_n satisfies $a_n x_n + b_n = 2.71$ for $n \in \{125, 250\}$, $a_n x_n + b_n = 2.91$ for $n \in \{500, 1000\}$. Similarly, we have $a_n y_n + b_n = 3$ for $n \in \{125, 250\}$, $a_n y_n + b_n = 3.2$ for $n \in \{500, 1000\}$. This means that the estimators get less parsimonious when d_n increases. Of course an adaption to maintain the same confidence level is possible, but the general picture remains the same.

For the criteria AIC, BIC , HQC and SIC we use the definitions given in (5.1.5) and (5.2.10), in case of HQC we chose $c = 1$, since, as pointed out by Hannan and Quinn [63], "it would seem pedantic to chose values as $c = 1.01$". The following modifications are also considered.

$$\begin{aligned} &\mathrm{AIC}(m)^* = \max\big\{\mathrm{AIC}(m), \widehat{q}^{(5)}_{y_n}(d_n)\big\},\ \mathrm{BIC}(m)^* = \max\big\{\mathrm{BIC}(m), \widehat{q}^{(5)}_{y_n}(d_n)\big\},\\ &\mathrm{HQC}(m)^* = \max\big\{\mathrm{HQC}(m), \widehat{q}^{(5)}_{y_n}(d_n)\big\},\ \mathrm{SIC}(m)^* = \max\big\{\mathrm{SIC}(m), \widehat{q}^{(5)}_{y_n}(d_n)\big\}. \end{aligned} \tag{5.3.2}$$

All simulations were carried out using the program *R*. In order to get a sample of size n, a sample path of size $1000 + n$ was simulated and the first 1000 observations were discarded.

Generally speaking, unreported simulations show that in many cases the modified criteria $\mathrm{AIC}(m)^*, \mathrm{BIC}(m)^*$... perform nearly identical as the none-modified ones $\mathrm{AIC}(m), \mathrm{BIC}(m)$... . This is in particular the case when dealing with full parameter sets, i.e. $\theta_i \neq 0$, $1 \leq i \leq q$, and θ_q is sufficiently large. If this is the case, the estimators $\widehat{q}^{(5)}_{x_n}(d_n), \widehat{q}^{(5)}_{y_n}(d_n)$ performance is somewhere between the $\mathrm{BIC}(m)$ and $\mathrm{HQC}(m)$. On the other hand, if the model is not full and/or the order q is sufficiently large, then the differences can be quite striking. The aim of the following examples is to illustrate this behavior.

5.3.1 AR(6)

First note that the definitions of x_n, y_n result in

$$\begin{aligned} &P\big(\max|\boldsymbol{\xi}| \leq 2.71\big) \geq 0.92, \quad P\big(\max|\boldsymbol{\xi}| \leq 3\big) \geq 0.97,\ d_n \in \{10, 12\},\\ &P\big(\max|\boldsymbol{\xi}| \leq 2.91\big) \geq 0.95, \quad P\big(\max|\boldsymbol{\xi}| \leq 3.2\big) \geq 0.98,\ d_n \in \{13, 14\}, \end{aligned}$$

where $\boldsymbol{\xi} = \big(\xi_1, ..., \xi_{d_n}\big)^T$ is a d_n-dimensional mean zero Gaussian random vector where the covariance matrix is the identity.

The results shown in Tables 5.1, 5.2 hint at what is to be expected in case of full models, namely that the modifications $\mathrm{AIC}(m)^*, \mathrm{BIC}(m)^*$... perform nearly as

n	$\widehat{q}$	AIC	AIC*	BIC	BIC*	HQC	HQC*	SIC	SIC*	$\widehat{q}^{(5)}_{y_n}$	$\widehat{q}^{(5)}_{x_n}$
125	< 5	428	427	943	808	746	704	550	545	816	701
	5	65	65	10	30	32	40	58	58	28	41
	6	**344**	**341**	**45**	**143**	**191**	**214**	**295**	**294**	**137**	**196**
	7	66	65	1	5	23	24	54	53	5	14
	< 7	97	102	1	14	8	18	43	50	14	48
250	< 5	93	89	693	432	328	282	202	188	440	299
	5	24	23	14	32	32	32	33	31	42	38
	6	**646**	**632**	**287**	**481**	**586**	**595**	**649**	**634**	**467**	**543**
	7	96	95	5	8	37	35	74	73	4	9
	> 7	141	161	1	47	17	56	42	74	47	111

Table 5.1: Simulation of an AR(6) process with coefficients $\Theta_6 = (0.1, -0.3, 0.05, 0.2, -0.1, 0.2)^T$, $\epsilon \sim \mathcal{N}(0,1)$, 1000 repetitions, $d_n \in \{10, 12\}$.

good as the normal versions $\mathrm{AIC}(m), \mathrm{BIC}(m)...$. The estimators $\widehat{q}^{(5)}_{x_n}(d_n), \widehat{q}^{(5)}_{y_n}(d_n)$ perform also quite well.

Contrary to the previous results, Tables 5.5 and 5.6 show the difference of the modified estimators (and $\widehat{q}^{(5)}_{x_n}(d_n), \widehat{q}^{(5)}_{y_n}(d_n)$), if the model is very sparse. Except for the case $n = 1000$, the modifications are notable better.

5.3.2 AR(12)

The definitions of x_n, y_n result in

$$P\big(\max|\boldsymbol{\xi}| \le 2.71\big) \ge 0.85, \quad P\big(\max|\boldsymbol{\xi}| \le 3\big) \ge 0.94, \; d_n \in \{20, 23\},$$
$$P\big(\max|\boldsymbol{\xi}| \le 2.91\big) \ge 0.9, \quad P\big(\max|\boldsymbol{\xi}| \le 3.2\big) \ge 0.96, \; d_n \in \{25, 29\},$$

where $\boldsymbol{\xi} = \big(\xi_1, ..., \xi_{d_n}\big)^T$ is a d_n-dimensional mean zero Gaussian random vector where the covariance matrix is the identity.

The results are depicted in the Tables 5.5, 5.6 and 5.7, 5.8, and are quite similar as in the case of the AR(6)-processes. If the model is rather full, $\mathrm{AIC}(m)^*, \mathrm{BIC}(m)^*...$ perform nearly as good as the normal versions $\mathrm{AIC}(m), \mathrm{BIC}(m)...$, whereas in case of the sparse model, a significant difference can be observed.

5.3.3 AR(24)

In this case the definitions of x_n, y_n result in

$$P\big(\max|\boldsymbol{\xi}| \le 2.71\big) \ge 0.795, \quad P\big(\max|\boldsymbol{\xi}| \le 3\big) \ge 0.912, \; d_n \in \{29, 34\},$$
$$P\big(\max|\boldsymbol{\xi}| \le 2.91\big) \ge 0.86, \quad P\big(\max|\boldsymbol{\xi}| \le 3.2\big) \ge 0.94, \; d_n \in \{38, 42\},$$

n	$\widehat{q}$	AIC	AIC*	BIC	BIC*	HQC	HQC*	SIC	SIC*	$\widehat{q}_{y_n}^{(5)}$	$\widehat{q}_{x_n}^{(5)}$
500	< 5	1	1	177	75	29	25	15	15	86	52
	5	3	3	9	11	6	6	3	3	17	14
	6	**730**	**713**	**805**	**874**	**913**	**889**	**892**	**867**	**865**	**849**
	7	108	108	8	8	42	42	57	57	0	2
	< 7	158	175	1	32	10	38	33	58	32	83
1000	< 5	0	0	3	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0
	6	**724**	**709**	**990**	**951**	**952**	**917**	**934**	**901**	**955**	**885**
	7	103	101	7	9	36	34	47	44	5	7
	> 7	173	190	0	40	12	49	19	55	40	108

Table 5.2: Simulation of an AR(6) process with coefficients $\Theta_6 = (0.1, -0.3, 0.05, 0.2, -0.1, 0.2)^T$, $\epsilon \sim \mathcal{N}(0,1)$, 1000 repetitions, $d_n \in \{13, 14\}$.

n	$\widehat{q}$	AIC	AIC*	BIC	BIC*	HQC	HQC*	SIC	SIC*	$\widehat{q}_{y_n}^{(5)}$	$\widehat{q}_{x_n}^{(5)}$
125	< 5	719	699	998	854	944	842	839	787	854	747
	5	11	11	0	0	2	2	7	7	0	11
	6	**168**	**181**	**2**	**124**	**43**	**126**	**107**	**145**	**124**	**184**
	7	44	44	0	4	8	11	23	24	4	8
	< 7	58	65	0	18	3	19	24	37	18	50
250	< 5	290	276	960	437	723	424	550	396	438	321
	5	6	6	0	3	2	3	5	5	3	5
	6	**491**	**488**	**39**	**513**	**245**	**503**	**376**	**494**	**513**	**573**
	7	91	90	1	2	21	21	40	40	1	7
	> 7	122	140	0	45	9	49	29	65	45	94

Table 5.3: Simulation of an AR(6) process with coefficients $\Theta_6 = (0.1, 0, 0.05, 0, 0, 0.2)^T$, $\epsilon \sim \mathcal{N}(0,1)$, 1000 repetitions, $d_n \in \{10, 12\}$.

n	$\widehat{q}$	AIC	AIC*	BIC	BIC*	HQC	HQC*	SIC	SIC*	$\widehat{q}^{(5)}_{y_n}$	$\widehat{q}^{(5)}_{x_n}$
500	< 5	21	21	761	102	267	98	164	85	102	56
	5	0	0	1	0	0	0	0	0	0	1
	6	**663**	**655**	**234**	**871**	**675**	**822**	**736**	**796**	**874**	**863**
	7	125	124	4	3	50	49	69	68	0	10
	< 7	191	200	0	24	8	31	31	51	24	70
1000	< 5	0	0	168	1	3	1	1	1	1	0
	5	0	0	0	0	0	0	0	0	0	0
	6	**702**	**683**	**822**	**949**	**940**	**905**	**919**	**887**	**955**	**898**
	7	121	119	9	9	43	42	52	52	3	9
	> 7	177	198	1	41	14	52	28	60	41	93

Table 5.4: Simulation of an AR(6) process with coefficients $\Theta_6 = (0.1, 0, 0.05, 0, 0, 0.2)^T$, $\epsilon \sim \mathcal{N}(0,1)$, 1000 repetitions, $d_n \in \{13, 14\}$.

n	$\widehat{q}$	AIC	AIC*	BIC	BIC*	HQC	HQC*	SIC	SIC*	$\widehat{q}^{(5)}_{y_n}$	$\widehat{q}^{(5)}_{x_n}$
125	< 11	705	701	995	966	931	917	812	807	969	929
	11	79	79	2	3	22	22	54	54	1	2
	12	**141**	**141**	**3**	**23**	**40**	**47**	**97**	**98**	**22**	**47**
	13	48	48	0	4	6	9	30	30	4	11
	> 13	27	31	0	4	1	5	7	11	4	11
250	< 11	257	257	854	730	573	560	423	421	748	620
	11	39	39	9	10	31	31	39	39	3	11
	12	**495**	**493**	**135**	**247**	**349**	**356**	**442**	**441**	**237**	**313**
	13	115	115	2	4	40	40	65	65	3	13
	> 13	94	96	0	9	7	13	31	34	9	43

Table 5.5: Simulation of an AR(12) process with nonzero coefficients $\theta_1 = 0.1$, $\theta_3 = -0.4$, $\theta_5 = 0.5$,$\theta_7 = -0.1$, $\theta_8 = 0.05$,$\theta_{10} = -0.3$, $\theta_{12} = 0.2$ $\epsilon \sim \mathcal{N}(0,1)$, 1000 repetitions, $d_n \in \{20, 23\}$.

n	$\widehat{q}$	AIC	AIC*	BIC	BIC*	HQC	HQC*	SIC	SIC*	$\widehat{q}^{(5)}_{y_n}$	$\widehat{q}^{(5)}_{x_n}$
500	< 11	19	19	367	256	110	106	75	73	269	183
	11	4	4	4	4	6	6	6	6	2	2
	12	**684**	**680**	**618**	**705**	**808**	**793**	**808**	**797**	**702**	**758**
	13	129	128	10	12	63	62	78	76	4	8
	> 13	164	169	1	23	13	33	33	48	23	49
1000	< 11	0	0	11	2	0	0	0	0	2	1
	11	0	0	0	0	0	0	0	0	0	0
	12	**679**	**676**	**970**	**947**	**925**	**900**	**896**	**873**	**958**	**914**
	13	151	150	17	17	61	60	79	78	6	13
	> 13	170	174	2	34	14	40	25	49	34	72

Table 5.6: Simulation of an AR(12) process with nonzero coefficients $\theta_1 = 0.1$, $\theta_3 = -0.4$, $\theta_5 = 0.5$,$\theta_7 = -0.1$, $\theta_8 = 0.05$,$\theta_{10} = -0.3$, $\theta_{12} = 0.2$ $\epsilon \sim \mathcal{N}(0,1)$, 1000 repetitions, $d_n \in \{25, 28\}$.

n	$\widehat{q}$	AIC	AIC*	BIC	BIC*	HQC	HQC*	SIC	SIC*	$\widehat{q}^{(5)}_{y_n}$	$\widehat{q}^{(5)}_{x_n}$
125	< 10	884	853	1000	920	995	920	963	910	920	861
	11	3	3	0	0	0	0	1	1	0	3
	12	**68**	**94**	**0**	**71**	**5**	**71**	**25**	**70**	**71**	**114**
	13	11	13	0	3	0	3	4	7	3	5
	> 13	34	37	0	6	0	6	7	12	6	17
250	< 10	509	421	999	555	934	552	792	530	555	424
	11	3	3	0	3	0	2	2	3	3	4
	12	**340**	**419**	**1**	**421**	**59**	**419**	**170**	**416**	**421**	**514**
	13	67	68	0	2	4	6	18	19	2	5
	> 13	81	89	0	19	3	21	18	32	19	53

Table 5.7: Simulation of an AR(12) process with nonzero coefficients $\theta_1 = 0.1$, $\theta_3 = -0.4$, $\theta_{12} = 0.2$ $\epsilon \sim \mathcal{N}(0,1)$, 1000 repetitions, $d_n \in \{20, 23\}$.

n	$\widehat{q}$	AIC	AIC*	BIC	BIC*	HQC	HQC*	SIC	SIC*	$\widehat{q}^{(5)}_{y_n}$	$\widehat{q}^{(5)}_{x_n}$
500	< 11	77	58	983	125	613	125	402	115	125	78
	11	0	0	0	2	0	2	0	1	2	1
	12	**663**	**678**	**17**	**858**	**360**	**834**	**532**	**808**	**858**	**870**
	13	104	103	0	3	15	16	39	40	3	4
	> 13	156	161	0	12	12	23	27	36	12	47
1000	< 11	0	0	689	2	67	2	35	2	2	2
	11	0	0	0	0	0	0	0	0	0	0
	12	**706**	**701**	**307**	**971**	**880**	**926**	**893**	**907**	**972**	**936**
	13	124	123	2	2	39	38	54	53	1	3
	> 13	170	176	2	25	14	34	18	38	25	59

Table 5.8: Simulation of an AR(12) process with nonzero coefficients $\theta_1 = 0.1$, $\theta_3 = -0.4$, $\theta_{12} = 0.2$ $\epsilon \sim \mathcal{N}(0,1)$, 1000 repetitions, $d_n \in \{25, 28\}$.

where $\boldsymbol{\xi} = \left(\xi_1, ..., \xi_{d_n}\right)^T$ is a d_n-dimensional mean zero Gaussian random vector where the covariance matrix is the identity. The behavior shown in Tables 5.9, 5.10 and 5.11, 5.12, is as in the previous two cases. The difference in the sparse model is perhaps the most striking one.

5.4 Proofs and ramification

In this section, we will prove the Theorems 5.2.2, 5.2.4, 5.2.8, and also explicitly mention some auxiliary results which have interest in themselves. Throughout this section, we will assume that Assumption 5.2.1 is valid. For $d_n \leq m$ let $\boldsymbol{\Gamma}_m^{-1} = \left(\gamma^*_{i,j}\right)_{1\leq i,j\leq m}$ be the inverse of the covariance matrix $\boldsymbol{\Gamma}_m = \left(\gamma_{i,j}\right)_{1\leq i,j\leq m}$ associated to the AR(d_n)-process $\left\{X_k\right\}_{k\in\mathbb{Z}}$. Due to Galbraith and Galbraith [51], it holds that

$$\sigma^2\gamma^*_{i,j} = \sum_{r=0}^{\alpha}\theta_r\theta_{r+j-i} - \sum_{r=\beta}^{d_n+i-j}\theta_r\theta_{r+j-i}, \quad 1 \leq i \leq j \leq m, \tag{5.4.1}$$

where

$$\alpha = \min\left\{i-1, d_n+i-j, m-j\right\}, \quad \beta = \max\left\{i-1, m-j\right\},$$

and either of the sums is taken to be zero if its upper limit is less than its lower limit. The second sum is zero unless $m - d_n + 1 \leq i \leq j \leq d_n$ while both sums are zero if $j - i > d_n$. Note that this implies $\sigma^2(m)\gamma^*_{m,m} = 1$ for $m > d_n$. In addition, throughout this section and particularly in the proofs of the presented results, we use the notation $\widehat{\sigma}^2 = \widehat{\sigma}^2(d_n)$.

n	$\widehat{q}$	AIC	AIC*	BIC	BIC*	HQC	HQC*	SIC	SIC*	$\widehat{q}_{y_n}^{(5)}$	$\widehat{q}_{x_n}^{(5)}$
125	< 23	972	970	1000	996	1000	996	992	990	996	989
	23	12	12	0	1	0	1	5	5	1	2
	24	**3**	**3**	**0**	**1**	**0**	**1**	**1**	**1**	**1**	**6**
	25	10	10	0	0	0	0	2	2	0	1
	> 25	3	5	0	2	0	2	0	2	2	2
250	< 23	518	516	995	923	872	840	727	717	924	845
	23	120	120	2	13	48	50	77	78	12	25
	24	**185**	**186**	**3**	**57**	**67**	**90**	**135**	**138**	**57**	**98**
	25	89	89	0	1	7	8	38	38	1	10
	> 25	88	89	0	6	6	12	23	29	6	22

Table 5.9: Simulation of an AR(24) process with nonzero coefficients $\theta_1 = 0.6$, $\theta_2 = -0.1$,$\theta_4 = 0.05$, $\theta_7 = 0.15$, $\theta_8 = -0.27$, $\theta_{10} = 0.1$, $\theta_{12} = -0.2$, $\theta_{15} = -0.25$, $\theta_{18} = 0.05$, $\theta_{20} = 0.1$, $\theta_{21} = -0.3$,$\theta_{24} = 0.17$, $\epsilon \sim \mathcal{N}(0,1)$, 1000 repetitions, $d_n \in \{29, 34\}$.

n	$\widehat{q}$	AIC	AIC*	BIC	BIC*	HQC	HQC*	SIC	SIC*	$\widehat{q}_{y_n}^{(5)}$	$\widehat{q}_{x_n}^{(5)}$
500	< 23	63	62	716	545	302	288	210	205	589	430
	23	38	38	55	60	87	87	85	85	58	71
	24	**513**	**512**	**208**	**357**	**490**	**500**	**525**	**526**	**326**	**437**
	25	192	192	18	28	93	93	129	129	19	27
	> 25	194	196	3	10	28	32	51	55	8	35
1000	< 23	0	0	81	30	6	5	3	3	42	18
	23	0	0	34	31	8	7	6	6	48	35
	24	**562**	**552**	**835**	**857**	**796**	**775**	**761**	**741**	**868**	**842**
	25	197	195	48	45	140	137	160	156	7	24
	> 25	241	253	2	37	50	76	70	94	35	81

Table 5.10: Simulation of an AR(24) process with nonzero coefficients $\theta_1 = 0.6$, $\theta_2 = -0.1$,$\theta_4 = 0.05$, $\theta_7 = 0.15$, $\theta_8 = -0.27$, $\theta_{10} = 0.1$, $\theta_{12} = -0.2$, $\theta_{15} = -0.25$, $\theta_{18} = 0.05$, $\theta_{20} = 0.1$, $\theta_{21} = -0.3$,$\theta_{24} = 0.17$, $\epsilon \sim \mathcal{N}(0,1)$, 1000 repetitions, $d_n \in \{38, 42\}$.

n	$\widehat{q}$	AIC	AIC*	BIC	BIC*	HQC	HQC*	SIC	SIC*	$\widehat{q}^{(5)}_{y_n}$	$\widehat{q}^{(5)}_{x_n}$
125	< 23	1000	991	1000	991	1000	991	1000	991	991	969
	23	0	2	0	2	0	2	0	2	2	6
	24	**0**	**6**	**0**	**6**	**0**	**6**	**0**	**6**	**6**	**20**
	25	0	1	0	1	0	1	0	1	1	1
	> 25	0	0	0	0	0	0	0	0	0	4
250	< 23	857	768	1000	817	998	817	986	815	817	702
	23	1	15	0	27	0	26	0	25	27	39
	24	**99**	**166**	**0**	**142**	**2**	**143**	**13**	**145**	**142**	**225**
	25	20	22	0	3	0	3	0	3	3	5
	> 25	23	29	0	11	0	11	1	12	11	29

Table 5.11: Simulation of an AR(24) process with nonzero coefficients $\theta_1 = 0.6$, $\theta_2 = -0.1$,$\theta_4 = 0.05$, $\theta_{10} = 0.1$, $\theta_{12} = -0.2$, $\theta_{24} = 0.17$, $\epsilon \sim \mathcal{N}(0,1)$, 1000 repetitions, $d_n \in \{29, 34\}$.

n	$\widehat{q}$	AIC	AIC*	BIC	BIC*	HQC	HQC*	SIC	SIC*	$\widehat{q}^{(5)}_{y_n}$	$\widehat{q}^{(5)}_{x_n}$
500	< 23	351	270	1000	383	952	380	854	379	383	256
	23	2	8	0	51	0	48	0	41	51	61
	24	**451**	**522**	**0**	**547**	**45**	**550**	**130**	**545**	**547**	**637**
	25	74	73	0	0	3	3	13	13	0	2
	> 25	122	127	0	19	0	19	3	22	19	44
1000	< 23	10	6	986	15	440	15	280	15	15	3
	23	0	0	0	14	0	13	0	11	14	12
	24	**718**	**715**	**14**	**941**	**522**	**908**	**659**	**887**	**941**	**905**
	25	121	118	0	3	32	31	46	45	3	8
	> 25	151	161	0	27	6	33	15	42	27	72

Table 5.12: Simulation of an AR(24) process with nonzero coefficients $\theta_1 = 0.6$, $\theta_2 = -0.1$,$\theta_4 = 0.05$, $\theta_{10} = 0.1$, $\theta_{12} = -0.2$, $\theta_{24} = 0.17$, $\epsilon \sim \mathcal{N}(0,1)$, 1000 repetitions, $d_n \in \{38, 42\}$.

Note that we can rewrite the equation defining the AR(d_n)-process as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\Phi}_{\mathbf{d_n}} + \mathbf{Z}, \tag{5.4.2}$$

where $\mathbf{Y} = \big(X_1, ..., X_n\big)^T$, $\mathbf{Z} = \big(\epsilon_1, ..., \epsilon_n\big)^T$, and the $n \times d_n$ design matrix $\mathbf{X}$ is given as

$$\mathbf{X} = \begin{pmatrix} X_0 & X_{-1} & ... & X_{1-d_n} \\ X_1 & X_0 & ... & X_{2-d_n} \\ ... & ... & & \\ X_{n-1} & X_{n-2} & ... & X_{n-d_n} \end{pmatrix}.$$

The following two results are key ingredients.

Proposition 5.4.1. *Let $\big\{X_k\big\}_{k\in\mathbb{Z}}$ be an AR(d_n) process, such that Assumption 5.2.1 is valid. Let d_n be an increasing sequence in n, $\epsilon > 0$. Then*

$$P\bigg(\max\big|\boldsymbol{\Gamma}_{d_n}^{-1} - \widehat{\boldsymbol{\Gamma}}_{\mathbf{d_n}}^{-1}\big| > \epsilon\bigg) = \mathcal{O}\left(\epsilon^{-p} n^{-p/2} d_n^p + \frac{(d_n \log n)^p}{n^{p/2}}\right).$$

Theorem 5.4.2. *Let $\big\{X_k\big\}_{k\in\mathbb{Z}}$ be an AR(d_n) process, such that Assumption 5.2.1 is valid. Then we have*

$$P\big(\max\big|n^{1/2}(\widehat{\boldsymbol{\Theta}}_{d_n} - \boldsymbol{\Theta}_{d_n}) - n^{-1/2}\boldsymbol{\Gamma}^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Z}\big| \geq \epsilon\big) = \mathcal{O}\left(\epsilon^{-p} n^{-p/2} d_n^{2p} (\log n)^p + n^{-1}\right),$$

where d_n is chosen as $d_n = \mathcal{O}\left(n^{\delta}\right)$, and δ needs to satisfy the conditions given in Theorem 5.4.3 given below.

The proofs of Proposition 5.4.1 and Theorem 5.4.2 are given in Section 5.6.1. Based on the above results, we can now prove the following.

Theorem 5.4.3. *Under the same conditions as in Theorem 5.4.2, we have for $x > 0$ that*

$$\left|P\big(\max\big|n^{1/2}(\widehat{\boldsymbol{\Theta}}_{d_n} - \boldsymbol{\Theta}_{d_n})\big| \leq x\big) - P\bigg(\max\big|\xi_{\mathbf{d_n}}\big| \leq x\bigg)\right| = \mathcal{O}\left(\kappa_n\right),$$

where $\xi_{\mathbf{d_n}} = \big(\xi_{n,1}, ..., \xi_{n,d_n}\big)^T$ is a d_n-dimensional mean zero Gaussian random vector with covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{\xi}^{d_n}}$, such that $\max\big|n^{-1}\boldsymbol{\Gamma}_{\boldsymbol{\xi}^{d_n}} - \sigma^2\boldsymbol{\Gamma}_{d_n}\big| = \mathcal{O}\left(n^{-\delta/2}\right)$, where $d_n = \mathcal{O}\big(n^{\delta}\big)$, $\kappa_n = \mathcal{O}\left(n^{-1/8+\delta} + n^{1/\nu-1/2+\delta} + n^{-p(1/3-2\delta)}(\log n)^p\right)$ and δ must satisfy

$$\delta < \min\left\{\frac{8 + 3p - (4+p)\nu}{(8+5p)\nu}, \frac{(3\nu - 4)(4+5p)}{(52+25p)\nu - 16}\right\}.$$

Remark 5.4.4. For fixed $p > 4$, the largest possible values for δ can be obtained by computing

$$\max_{2\leq\nu}\left(\min\left\{\frac{8+3p-(4+p)\nu}{(8+5p)\nu}, \frac{(3\nu-4)(4+5p)}{(52+25p)\nu-16}, 1/2-1/\nu\right\}\right).$$

For example, for $p = 16$ we obtain $\delta <$, with $\nu =$.

Using the above results, we obtain the following corollaries.

Corollary 5.4.5. *Under the same conditions as in Theorem 5.4.2, we have*

$$P\big(\big|\widehat{\sigma}^2-\sigma^2\big|\geq\epsilon\big)=\mathcal{O}\left(n^{-1}+\epsilon^{-p}n^{-p/2}d_n(\log n)^{p/2}+\epsilon^{-p/2}n^{-p/2}d_n(d_n\log n)^{p/2}\right).$$

Corollary 5.4.6. *Theorem 5.4.3 remains valid if one replaces* $\max\big|n^{1/2}(\widehat{\mathbf{\Theta}}_{d_n}-\mathbf{\Theta}_{d_n})\big|$ *with the normalized version* $\max_{1\leq i\leq n}\big|\sqrt{n}(\widehat{\gamma}^*_{i,i}\widehat{\sigma}^2)^{-1/2}(\widehat{\theta}_i-\theta_i)\big|$.

Remark 5.4.7. The above theorem also has the following practical relevance: The rate of convergence to an extreme-value type distribution as given in Theorem 5.2.2 can be rather slow, see for instance [10, 97]. On the other hand, we get from the above corollary that

$$\left|P\big(\max_{1\leq i\leq n}\big|\sqrt{n}(\widehat{\gamma}^*_{i,i}\widehat{\sigma}^2)^{-1/2}(\widehat{\theta}_i-\theta_i)\big|\leq x\big)-P\left(\max\big|\xi_{\mathbf{d_n}}\big|\leq x\right)\right|=\mathcal{O}\left(\kappa_n\right),$$

where $\boldsymbol{\xi}_{d_n}=\big(\xi_{n,1},...,\xi_{n,d_n}\big)^T$ is a d_n-dimensional mean zero Gaussian random vector with covariance matrix $\mathbf{\Gamma}_{\boldsymbol{\xi}^{d_n}}$, that is close to the matrix

$$\mathbf{\Gamma}^*_{d_n}=\big(\gamma^*_{i,j}/\sqrt{\gamma^*_{i,i}\gamma^*_{j,j}}\big)_{1\leq,i,j\leq d_n}=\big\{\rho_{i,j}\big\}_{1\leq,i,j\leq d_n}.$$

Corresponding quantiles can be obtained for instance via a Monte Carlo technique. However, if d_n is sufficiently large, one has that

$$P\left(\max\big|\boldsymbol{\xi}_{d_n}\big|\leq x\right)\approx P\left(\max\big|\boldsymbol{\eta}_{d_n}\big|\leq x\right),$$

where $\boldsymbol{\eta}_{d_n}=\big(\eta_{n,1},...,\eta_{n,d_n}\big)^T$ is a sequence of IID mean zero Gaussian random variables with unit variance. A bound for the error can be given by using the techniques developed by Berman [17] and Deo [38], see also the proof of Theorem 5.2.4.

Throughout the proofs, the following inequality will be frequently used. For random variables $X_1,...,X_q$, and $\epsilon > 0$, the inequality between the geometric and arithmetic mean implies

$$P\big(\prod_{i=1}^q|X_i|\geq\epsilon\big)\leq\sum_{i=1}^q P\big(|X_i|\geq\epsilon^{1/q}\big). \tag{5.4.3}$$

Proof of Theorem 5.4.3. Let $\chi_{n,\epsilon} = \epsilon^{-p} n^{-p/2} d_n^{2p} (\log n)^p + n^{-1}$. Then Theorem 5.4.2 implies

$$\begin{aligned}
P\big(\max\big|n^{1/2}(\widehat{\mathbf{\Theta}}_{d_n} - \mathbf{\Theta}_{d_n})\big| \le x\big) &\le P\big(\max\big|n^{1/2}(\widehat{\mathbf{\Theta}}_{d_n} - \mathbf{\Theta}_{d_n}) - n^{-1/2}\mathbf{\Gamma}^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Z}\big| \ge 2\epsilon\big) \\
&\quad + P\big(\max\big|n^{-1/2}\mathbf{\Gamma}^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Z}\big| \le x + 2\epsilon\big) \\
&= \mathcal{O}\left(\chi_{n,\epsilon}\right) + P\big(\max\big|n^{-1/2}\mathbf{\Gamma}^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Z}\big| \le x + 2\epsilon\big).
\end{aligned}$$

As in Brockwell and Davis [25, Section 8.10], by setting $\mathbf{U}_k = \big(X_{k-1}, ..., X_{k-d_n}\big)^T \epsilon_k$, $k \in \mathbb{N}$, we have

$$n^{-1/2}\mathbf{X}^{\mathbf{T}}\mathbf{Z} = n^{-1/2} \sum_{k=1}^{n} \mathbf{U}_k.$$

Note in particular that $\mathbb{E}\big(\mathbf{U}_k\big) = 0$, and

$$\mathbb{E}\big(\mathbf{U}_k \mathbf{U}_{k+h}^T\big) = \begin{cases} \sigma^2 \Gamma_{d_n}, & \text{if } h = 0, \\ 0_{d_n \times d_n}, & \text{if } h \neq 0, \end{cases} \tag{5.4.4}$$

since $\epsilon_{\mathbf{k}}$ is independent of $\big\{X_{k-i}\big\}_{i \ge 1}$. Note that since $|\theta_i| < 1$, using the representation (5.4.1) for $\big\{\gamma_{i,j}^*\big\}_{1 \le i \le d_n}$, we obtain that

$$\sigma^2 \max\big|\mathbf{\Gamma}_{d_n}^{-1}\big| < 2.$$

From Corollary 5.5.3 we obtain that on a possible larger probability space there exists a d_n-dimensional Gaussian vector $\xi_{\mathbf{d_n}} = \big(\xi_{n,1}, ..., \xi_{n,d_n}\big)^T$ with the covariance matrix $\mathbf{\Gamma}_{\boldsymbol{\xi}^{d_n}}$, such that for $2 \le \nu$

$$P\left(\max\big|n^{-1/2}\mathbf{\Gamma}_{d_n}^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Z} - \boldsymbol{\xi}_{d_n}\big| > n^{1/\nu - 1/2}\right) = \mathcal{O}\left(n^{-1}\right). \tag{5.4.5}$$

Using this we obtain

$$P\big(\max\big|n^{1/2}(\widehat{\mathbf{\Theta}}_{d_n} - \mathbf{\Theta}_{d_n})\big| \le x\big) \le P\left(\max\big|\boldsymbol{\xi}_{d_n}\big| \le x + \epsilon + n^{1/\nu - 1/2}\right) + \mathcal{O}\left(\chi_{n,\epsilon}\right).$$

Observe that for $y > 0$

$$\begin{aligned}
P\left(\max\big|\boldsymbol{\xi}_{d_n}\big| \le x + y\right) - P\left(\max\big|\boldsymbol{\xi}_{d_n}\big| \le x\right) &\le \sum_{h=0}^{d_n} P\left(x \le \big|\xi_h\big| \le x + y\right) \\
&\le \sum_{h=0}^{d_n} P\left(\big|\xi_h\big| \le y\right) = \mathcal{O}\left(y d_n\right),
\end{aligned}$$

hence

$$P\big(\max\big|n^{1/2}(\widehat{\boldsymbol{\Theta}}_{d_n} - \boldsymbol{\Theta}_{d_n})\big| \leq x\big) \leq P\bigg(\max\big|\boldsymbol{\xi}_{d_n}\big| \leq x\bigg) + \mathcal{O}\left((\epsilon + n^{1/\nu-1/2})d_n + \chi_{n,\epsilon}\right). \tag{5.4.6}$$

Similarly, one establishes the lower bound

$$P\big(\max\big|n^{1/2}(\widehat{\boldsymbol{\Theta}}_{d_n} - \boldsymbol{\Theta}_{d_n})\big| \leq x\big) \geq P\bigg(\max\big|\boldsymbol{\xi}_{d_n}\big| \leq x\bigg) - \mathcal{O}\left((\epsilon + n^{1/\nu-1/2})d_n + \chi_{n,\epsilon}\right). \tag{5.4.7}$$

Setting $\epsilon = n^{-c}$, we obtain the inequalities

$$\delta < \frac{1}{2} - \frac{1}{\nu}, \quad \delta < \frac{1-2c}{4}, \quad \delta < c.$$

Hence we can choose $c = 1/8$, and thus obtain

$$(\epsilon + n^{1/\nu-1/2})d_n + \chi_{n,\epsilon} \leq n^{-1/8+\delta} + n^{1/\nu-1/2+\delta} + n^{-p(1/3-2\delta)}(\log n)^p,$$

which completes the proof. □

Proof of Corollary 5.4.5. Trivially, it holds that

$$\begin{aligned}\widehat{\sigma}^2 - \sigma^2 &= \widehat{\phi}_0 - \phi_0 + \widehat{\boldsymbol{\Theta}}_{d_n}^T \widehat{\boldsymbol{\Phi}}_{d_n} - \boldsymbol{\Theta}_{d_n}^T \boldsymbol{\Phi}_{d_n} \\ &= \widehat{\phi}_0 - \phi_0 + \big(\widehat{\boldsymbol{\Theta}}_{d_n}^T - \boldsymbol{\Theta}_{d_n}^T\big)\big(\widehat{\boldsymbol{\Phi}}_{d_n} - \boldsymbol{\Phi}_{d_n}\big) + \big(\widehat{\boldsymbol{\Theta}}_{d_n}^T - \boldsymbol{\Theta}_{d_n}^T\big)\boldsymbol{\Phi}_{d_n} + \big(\widehat{\boldsymbol{\Theta}}_{d_n}^T - \boldsymbol{\Theta}_{d_n}^T\big)\boldsymbol{\Phi}_{d_n}.\end{aligned}$$

Since $\sum_{i=1}^{d_n} |\theta_i| < 1$, Lemma 5.6.1, the Markov and Minikowski inequality yield

$$\begin{aligned}P\big(\big|\boldsymbol{\Theta}_{d_n}^T\big(\widehat{\boldsymbol{\Phi}}_{d_n} - \boldsymbol{\Phi}_{d_n}\big| \geq \epsilon\big) &\leq \bigg\|\sum_{i=1}^{d_n} |\theta_i||\widehat{\phi}_i - \phi_i|\bigg\|_p^p \epsilon^{-p} \\ &\leq \bigg(\sum_{i=1}^{d_n} |\theta_i|\big\|\widehat{\phi}_i - \phi_i\big\|_p\bigg)^p \epsilon^{-p} = \mathcal{O}\left(n^{-p/2}\epsilon^{-p}\right).\end{aligned}$$

Since $\sum_{h=0}^{\infty} |\phi_h| \leq C < \infty$ we have

$$P\big(\big|\big(\widehat{\boldsymbol{\Theta}}_{d_n}^T - \boldsymbol{\Theta}_{d_n}^T\big)\boldsymbol{\Phi}_{d_n}\big| \geq \epsilon\big) \leq P\big(C\max\big|\widehat{\boldsymbol{\Theta}}_{d_n} - \boldsymbol{\Theta}_{d_n}\big| \geq \epsilon\big).$$

An application of Theorem 5.4.2 yields

$$P\big(\max\big|(\widehat{\boldsymbol{\Theta}}_{d_n} - \boldsymbol{\Theta}_{d_n}) - n^{-1}\boldsymbol{\Gamma}^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Z}\big| \geq \epsilon\big) = \mathcal{O}\left(\epsilon^{-p}n^{-p}d_n^{2p}(\log n)^p + n^{-1}\right).$$

Using the notation of the proof of Theorem 6.6.4 we have

$$\mathbf{\Gamma}^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Z} = \sum_{i \le i \le d_n} \mathbf{U_i}.$$

By Lemma 5.6.1, we can present X_k as $X_k = \sum_{i=0}^{\infty} \psi_i \epsilon_{k-i}$. Set $X_k^{(n)} = \sum_{i=0}^{\lceil A \log n \rceil} \psi_i \epsilon_{k-i}$, $A > 0$, $Y_k^{(n)} = \sum_{i > \lceil A \log n \rceil}^{\infty} \psi_i \epsilon_{k-i}$ and define $\mathbf{U}_k^{(n)} = \big(U_k^{(1)}, ..., U_{k-d_n+1}^{(d_n)}\big)^T$, where $U_k^{(h)} = X_{k-h-1}^{(n)} \epsilon_k$, $0 \le h \le d_n - 1$, and similarly $\mathbf{V}_k^{(n)} = \big(V_k^{(1)}, ..., V_{k-d_n+1}^{(d_n)}\big)^T$, where $V_k^{(h)} = Y_{k-h-1}^{(n)} \epsilon_k$, $0 \le h \le d_n - 1$. This implies

$$\begin{aligned} P\big(\max\big|\mathbf{\Gamma}^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Z}\big| \ge n\epsilon\big) &\le P\big(\max\big|\sum_{1 \le i \le d_n} \mathbf{U_i}^{(n)}\big| \ge n\epsilon/2\big) + P\big(\max\big|\sum_{1 \le i \le d_n} \mathbf{V_i}^{(n)}\big| \ge n\epsilon/2\big) \\ &\le \sum_{h=1}^{d_n} P\big(\big|\sum_{1 \le i \le n} U_i^{(h)}\big| \ge n\epsilon/2\big) + \sum_{h=1}^{d_n} P\big(\big|\sum_{1 \le i \le n} V_i^{(h)}\big| \ge n\epsilon/2\big). \end{aligned}$$

By the Markov inequality and Lemma 6.6.10 we have

$$\begin{aligned} &\sum_{h=1}^{d_n} P\big(\big|\sum_{1 \le i \le n} U_i^{(h)}\big| \ge n\epsilon/2\big) \le (n\epsilon/2)^{-p} \sum_{h=1}^{d_n} \Big\|\sum_{l \le i \le n} U_i^{(h)}\Big\|_p^p \le C(n\epsilon)^{-p} d_n (n \log n)^{p/2} \\ &= C\epsilon^{-p} n^{-p/2} d_n (\log n)^{p/2}. \end{aligned}$$

On the other hand, for large enough $A > 0$ we have $\big\|V_i^{(h)}\big\|_p = \mathcal{O}\left(n^{-1}\right)$, and thus the Markov and Minikowski inequality imply

$$\begin{aligned} \sum_{h=1}^{d_n} P\big(\big|\sum_{1 \le i \le n} V_i^{(h)}\big| \ge n\epsilon/2\big) &\le C(n\epsilon/2)^{-p} \sum_{h=1}^{d_n} \Big\|\sum_{1 \le i \le n} V_i^{(h)}\Big\|_p^p \\ &\le C(n\epsilon/2)^{-p} \sum_{h=1}^{d_n} \left(\sum_{1 \le i \le n} \big\|V_i^{(h)}\big\|_p\right)^p \\ &= \mathcal{O}\left((n\epsilon)^{-p} d_n\right). \end{aligned}$$

Since $d_n = o\left(n^{-1/8}\right)$ this results in a total bound

$$P\big(\big|\big(\widehat{\mathbf{\Theta}}_{d_n}^T - \mathbf{\Theta}_{d_n}^T\big)\mathbf{\Phi}_{d_n}\big| \ge \epsilon\big) = \mathcal{O}\left(n^{-1} + \epsilon^{-p} n^{-p/2} d_n (\log n)^{p/2}\right). \qquad (5.4.8)$$

In addition, observe that

$$P\big(\big|\big(\widehat{\mathbf{\Theta}}_{d_n}^T - \mathbf{\Theta}_{d_n}^T\big)\big(\widehat{\mathbf{\Phi}}_{d_n} - \mathbf{\Phi}_{d_n}\big)\big| \ge \epsilon\big) \le P\big(d_n \max\big|\widehat{\mathbf{\Theta}}_{d_n}^T - \mathbf{\Theta}_{d_n}^T\big| \max\big|\widehat{\mathbf{\Phi}}_{d_n} - \mathbf{\Phi}_{d_n}\big| \ge \epsilon\big),$$

which by the inequality given in (5.4.3) is smaller than

$$P\big(\max\big|\widehat{\mathbf{\Theta}}_{d_n}^T - \mathbf{\Theta}_{d_n}^T\big| \geq \sqrt{\epsilon d_n^{-1}}\big) + P\big(\max\big|\widehat{\mathbf{\Phi}}_{d_n} - \mathbf{\Phi}_{d_n}\big| \geq \sqrt{\epsilon d_n^{-1}}\big).$$

By the bounds derived above we thus obtain

$$\begin{aligned} P\big(\big|\big(\widehat{\mathbf{\Theta}}_{d_n}^T - \mathbf{\Theta}_{d_n}^T\big)\big(\widehat{\mathbf{\Phi}}_{d_n} - \mathbf{\Phi}_{d_n}\big)\big| \geq \epsilon\big) &= \mathcal{O}\left(n^{-1} + \epsilon^{-p/2} n^{-p/2} d_n (d_n \log n)^{p/2} + \epsilon^{-p/2} d_n^{p/2} n^{-p/2}\right) \\ &= \mathcal{O}\left(n^{-1} + \epsilon^{-p/2} n^{-p/2} d_n (d_n \log n)^{p/2}\right). \end{aligned}$$

Piecing everything together, we obtain

$$\begin{aligned} P\big(\big|\widehat{\sigma}^2 - \sigma^2\big| \geq \epsilon\big) &= \mathcal{O}\left(n^{-p/2}\epsilon^{-p} + n^{-1} + \epsilon^{-p} n^{-p/2} d_n (\log n)^{p/2} + \epsilon^{-p/2} n^{-p/2} d_n (d_n \log n)^{p/2}\right) \\ &= \mathcal{O}\left(n^{-1} + \epsilon^{-p} n^{-p/2} d_n (\log n)^{p/2} + \epsilon^{-p/2} n^{-p/2} d_n (d_n \log n)^{p/2}\right), \end{aligned}$$

which completes the proof. □

Proof of Corollary 5.4.6. First note that using the representation (5.4.1) for $\{\gamma_{i,j}^*\}_{1 \leq i \leq d_n}$, we obtain from Assumption 5.2.5 that

$$\sigma^2 \min_{1 \leq i \leq d_n} \big|\gamma_{i,i}^*\big| \geq 1 - \sum_{i=1}^{d_n} |\theta_i| \geq \vartheta > 0. \tag{5.4.9}$$

Since $\sigma^2 > 0$, it suffices to show that the error difference

$$\sqrt{n}\big|(\widehat{\gamma}_{i,i}^*\widehat{\sigma}^2)^{-1/2}(\widehat{\theta}_i - \theta_i) - (\gamma_{i,i}^*\sigma^2)^{-1/2}(\widehat{\theta}_i - \theta_i)\big| \tag{5.4.10}$$

is sufficiently small in probability, in which case the claim follows from Theorem 5.4.3.

We will first derive bounds for the difference $\Delta_i = \big|(\widehat{\gamma}_{i,i}^*\widehat{\sigma}^2)^{-1/2}(\widehat{\theta}_i - \theta_i) - (\gamma_{i,i}^*\sigma^2)^{-1/2}(\widehat{\theta}_i - \theta_i)\big|$. It holds that

$$\begin{aligned} &P\big(\max_{1 \leq i \leq d_n} \big|(\widehat{\gamma}_{i,i}^*\widehat{\sigma}^2)^{-1/2}(\widehat{\theta}_i - \theta_i) - (\gamma_{i,i}^*\sigma^2)^{-1/2}(\widehat{\theta}_i - \theta_i)\big| \geq \epsilon\big) \\ &\leq P\big(\max_{1 \leq i \leq d_n} \big|\big((\widehat{\gamma}_{i,i}^*\widehat{\sigma}^2)^{1/2} - (\gamma_{i,i}^*\sigma^2)^{1/2}\big)(\widehat{\gamma}_{i,i}^*\widehat{\sigma}^2)^{-1/2}(\widehat{\theta}_i - \theta_i)(\gamma_{i,i}^*\sigma^2)^{-1/2}\big| \geq \epsilon\big). \end{aligned}$$

Due to (5.4.9), the expression $\big|(\widehat{\theta}_i - \theta_i)(\gamma_{i,i}^*\sigma^2)^{-1/2}\big|$ can be controlled by Theorem 6.6.4 (we give some details later), hence we need to study $\big|(\widehat{\gamma}_{i,i}^*\widehat{\sigma}^2)^{1/2} - (\gamma_{i,i}^*\sigma^2)^{1/2}\big|(\widehat{\gamma}_{i,i}^*\widehat{\sigma}^2)^{-1/2}$. Since

$$\begin{aligned} \left|\frac{(\widehat{\gamma}_{i,i}^*\widehat{\sigma}^2)^{1/2} - (\gamma_{i,i}^*\sigma^2)^{1/2}}{(\widehat{\gamma}_{i,i}^*\widehat{\sigma}^2)^{1/2}}\right| &= \left|\frac{\widehat{\gamma}_{i,i}^*\widehat{\sigma}^2 - \gamma_{i,i}^*\sigma^2}{(\widehat{\gamma}_{i,i}^*\widehat{\sigma}^2)^{1/2} + (\gamma_{i,i}^*\sigma^2)^{1/2}} \frac{1}{(\widehat{\gamma}_{i,i}^*\widehat{\sigma}^2)^{1/2}}\right| \\ &\leq \left|\frac{\big(\widehat{\gamma}_{i,i}^*\widehat{\sigma}^2 - \gamma_{i,i}^*\sigma^2\big)}{\widehat{\gamma}_{i,i}^*\widehat{\sigma}^2}\right|, \end{aligned}$$

it suffices to treat $\big(\widehat{\gamma}^*_{i,i}\widehat{\sigma}^2 - \gamma^*_{i,i}\sigma^2\big)\big(\widehat{\gamma}^*_{i,i}\widehat{\sigma}^2\big)^{-1}$.

For $0 < \epsilon \leq 1$ we have

$$\begin{aligned}\big\{\big|\sigma^2\gamma^*_{i,i} - \widehat{\sigma}^2\widehat{\gamma}^*_{i,i}\big| \geq \epsilon\widehat{\gamma}^*_{i,i}\widehat{\sigma}^2\big\} &\subseteq \big\{\big|\sigma^2\gamma^*_{i,i} - \widehat{\sigma}^2\widehat{\gamma}^*_{i,i}\big|(1+\epsilon) \geq \epsilon\gamma^*_{i,i}\sigma^2\big\}\\ &\subseteq \big\{\big|\sigma^2\gamma^*_{i,i} - \widehat{\sigma}^2\widehat{\gamma}^*_{i,i}\big| \geq \epsilon\gamma^*_{i,i}\sigma^2/2\big\}\\ &\subseteq \big\{\big|\sigma^2\widehat{\gamma}^*_{i,i} - \widehat{\sigma}^2\widehat{\gamma}^*_{i,i}\big| + \big|\sigma^2\gamma^*_{i,i} - \sigma^2\widehat{\gamma}^*_{i,i}\big| \geq \epsilon\gamma^*_{i,i}\sigma^2/2\big\}\end{aligned}$$

Since $\sigma^2, \gamma^*_{i,i} \geq C > 0$, we have from Proposition 5.4.1 that

$$P\big(\max_{1\leq i\leq d_n}\big|\sigma^2\gamma^*_{i,i} - \sigma^2\widehat{\gamma}^*_{i,i}\big| \geq \epsilon\min_{1\leq i\leq d_n}\gamma^*_{i,i}\big) = \mathcal{O}\left(\epsilon^{-p}n^{-p/2}d_n^p + \frac{(d_n\log n)^p}{n^{p/2}}\right).$$

In order to treat $\big|\sigma^2\widehat{\gamma}^*_{i,i} - \widehat{\sigma}^2\widehat{\gamma}^*_{i,i}\big|$, note that

$$\big|\sigma^2\widehat{\gamma}^*_{i,i} - \widehat{\gamma}^*_{i,i}\widehat{\sigma}^2\big| \leq \big|\sigma^2 - \widehat{\sigma}^2\big|\big|\gamma^*_{i,i} - \widehat{\gamma}^*_{i,i}\big| + \gamma^*_{i,i}\big|\sigma^2 - \widehat{\sigma}^2\big|.$$

Using the inequality given in (5.4.3) we obtain from Proposition 5.4.1 and Corollary 5.4.5

$$P\big(\max_{1\leq i\leq d_n}\big|\sigma^2\widehat{\gamma}^*_{i,i} - \widehat{\gamma}^*_{i,i}\widehat{\sigma}^2\big| \geq \epsilon\big) = \mathcal{O}\left(\zeta_{n,\epsilon}\right), \tag{5.4.11}$$

where $\zeta_{n,\epsilon} = n^{-1} + \epsilon^{-p}n^{-p/2}d_n(\log n)^{p/2} + \epsilon^{-p/2}n^{-p/2}d_n(d_n\log n)^{p/2}$, and consequently for $1 \leq i \leq d_n$

$$P\big(\max_{1\leq i\leq d_n}\big|\sigma^2\gamma^*_{i,i} - \widehat{\sigma}^2\widehat{\gamma}^*_{i,i}\big|(\widehat{\gamma}^*_{i,i})^{-1} \geq \epsilon\widehat{\sigma}^2\big) = \mathcal{O}\left(\epsilon^{-p}n^{-p/2}d_n^p + \chi_{n,\epsilon}\right),$$

if $0 < \epsilon \leq 1$. This implies that for $\epsilon^* > 0$ we have

$$\begin{aligned}&P\big(\sqrt{n}\max_{1\leq i\leq d_n}\Delta_i \geq \epsilon\big)\\ &\leq P\big(\max_{1\leq i\leq d_n}\big|\big((\widehat{\theta}_i - \theta_i)(\gamma^*_{i,i}\sigma^2)^{-1/2}\big| \geq \epsilon(\epsilon^*)^{-1}\big) + P\big(\big|(\widehat{\gamma}^*_{i,i}\widehat{\sigma}^2)^{1/2} - (\gamma^*_{i,i}\sigma^2)^{1/2}\big| \geq \epsilon^*\widehat{\gamma}^*_{i,i}\widehat{\sigma}^2\big)\\ &= P\big(\sqrt{n}\max_{1\leq i\leq d_n}\big|(\widehat{\theta}_i - \theta_i)(\gamma^*_{i,i}\sigma^2)^{-1/2}\big| \geq \epsilon(\epsilon^*)^{-1}\big) + \mathcal{O}\left((\epsilon^*)^{-p}n^{-p/2}d_n^p + \zeta_{n,\epsilon^*}\right).\end{aligned}$$

Since $\sigma^2, \gamma^*_{i,i} \geq C > 0$, and application of Theorem 5.4.3 (remains valid) yields

$$\Big|P\big(\sqrt{n}\max_{1\leq i\leq d_n}\big|(\widehat{\theta}_i - \theta_i)(\gamma^*_{i,i}\sigma^2)^{-1/2}\big| \geq \epsilon(\epsilon^*)^{-1}\big) - P\big(\max_{1\leq i\leq d_n}\big|\xi_i(\gamma^*_{i,i}\sigma^2)^{-1/2}\big| \geq \epsilon(\epsilon^*)^{-1}\big)\Big| = \mathcal{O}(\kappa_n)$$

if $\epsilon(\epsilon^*)^{-1} \geq 1$, where $\xi_{\mathbf{d_n}} = \big(\xi_{n,1}, ..., \xi_{n,d_n}\big)^T$ is a d_n-dimensional Gaussian vector. Choosing $\epsilon^* = \epsilon(\log n)^{-1}$ we obtain

$$P\big(\max_{1\leq i\leq d_n}\big|\xi_i(\gamma^*_{i,i}\sigma^2)^{-1/2}\big| \geq \epsilon(\epsilon^*)^{-1}\big) \leq \sum_{i=1}^{d_n}P\big(\big|\xi_i(\gamma^*_{i,i}\sigma^2)^{-1/2}\big| \geq \log n\big) = \mathcal{O}\left(n^{-1}\right).$$

Piecing everything together, we get

$$P\big(\sqrt{n}\max_{1\le i\le d_n}\Delta_i \ge \epsilon\big) = \mathcal{O}\left(\epsilon^{-p}n^{-p/2}(d_n\log n)^p + \zeta_{n,\epsilon(\log n)^{-1}} + \kappa_n\right). \tag{5.4.12}$$

This finally entitles us to establish a decent approximation of $P\big(\max_{1\le i\le d_n}\big|(\widehat{\gamma}^*_{i,i}\widehat{\sigma}^2)^{-1/2}(\widehat{\theta}_i - \theta_i)\big| \le x\big)$. From the above we have

$$\Big|P\big(\sqrt{n}\max_{1\le i\le d_n}\big|(\widehat{\gamma}^*_{i,i}\widehat{\sigma}^2)^{-1/2}(\widehat{\theta}_i-\theta_i)\big|\le x\big) - P\big(\sqrt{n}\max_{1\le i\le d_n}\big|(\gamma^*_{i,i}\sigma^2)^{-1/2}(\widehat{\theta}_i-\theta_i)\big|\le x+\epsilon\big)\Big|$$
$$\le P\big(\sqrt{n}\max_{1\le i\le d_n}\Delta_i\ge\epsilon\big),$$

and as before an application of Theorem 5.4.3 gives us

$$\Big|P\big(\sqrt{n}\max_{1\le i\le d_n}\big|(\gamma^*_{i,i}\sigma^2)^{-1/2}(\widehat{\theta}_i-\theta_i)\big|\le x+\epsilon\big) - P\big(\sqrt{n}\max_{1\le i\le d_n}\big|(\gamma^*_{i,i}\sigma^2)^{-1/2}(\widehat{\theta}_i-\theta_i)\big|\le x\big)\Big|$$
$$= P\big(x\le\max_{1\le i\le d_n}\big|\xi_i(\gamma^*_{i,i}\sigma^2)^{-1/2}\big|\le x+\epsilon\big)+\mathcal{O}(\kappa_n)$$
$$\le\sum_{i=1}^{d_n}P\big(x\le\big|\xi_i(\gamma^*_{i,i}\sigma^2)^{-1/2}\big|\le x+\epsilon\big)+\mathcal{O}(\kappa_n)=\mathcal{O}\left(\kappa_n+d_n\epsilon\right).$$

This yields a total approximation error of the magnitude

$$P\big(\sqrt{n}\max_{1\le i\le d_n}\Delta_i\ge\epsilon\big)+\mathcal{O}\left(\kappa_n+d_n\epsilon\right)$$
$$=\mathcal{O}\left(\epsilon^{-p}n^{-p/2}(d_n\log n)^p+\zeta_{n,\epsilon(\log n)^{-1}}+\kappa_n+d_n\epsilon\right). \tag{5.4.13}$$

By comparing this error bound with the one given in the proof of Theorem 5.4.3, one finds that the latter is the dominating one, hence the conditions given in Theorem 5.4.3 remain unchanged. □

We are now ready to prove Theorems 5.2.2 and 5.2.4.

Proof of Theorem 5.2.2. By Corollary 5.4.6, it suffices to consider the absolute value of a sequence of mean zero Gaussian random variables $\boldsymbol{\xi}_{d_n} = \big(\xi_{n,1},\dots,\xi_{n,d_n}\big)^T$, with covariance matrix $\boldsymbol{\Gamma}^*_{\boldsymbol{\xi}_{d_n}}$, that satisfies $\max\big|\boldsymbol{\Gamma}^*_{\boldsymbol{\xi}_{d_n}} - \boldsymbol{\Gamma}^*_{d_n}\big| = \mathcal{O}\left(n^{-\delta/2}\right)$, where $\boldsymbol{\Gamma}^*_{d_n} = \big(\gamma^*_{i,j}/\sqrt{\gamma^*_{i,i}\gamma^*_{j,j}}\big)_{1\le,i,j\le d_n} = \big\{\rho_{i,j}\big\}_{1\le,i,j\le d_n}$. Results for the maximum of the absolute value of Gaussian sequences is for example treated in [38] (see also [17] for earlier contributions), and is stated in Section 5.7 as Lemma 5.7.1. It thus remains to verify the conditions of Lemma 5.7.1. To this end, let us first assume that $\boldsymbol{\Gamma}^*_{\boldsymbol{\xi}_{d_n}} = \boldsymbol{\Gamma}^*_{d_n}$.

Equation (5.4.9) in the proof of Corollary 5.4.6 implies that

$$\sigma^2 \min_{1 \leq i \leq d_n} \left|\gamma_{i,i}^*\right| \geq \vartheta > 0.$$

In addition, the Cauchy-Schwarz inequality yields $\rho_{i,j} \leq 1$ which gives us

$$\sup_{h,k} \left|\mathrm{Cov}\big(\xi_k, \xi_{k+h}\big)\right| \leq 1.$$

Suppose now that for some subsequence n' and corresponding sequences k_n', h_n', with $h_n' \geq 1$ we actually have

$$\lim_{n'} \left|\mathrm{Cov}\big(\xi_{k_n'}, \xi_{k_n'+h_n'}\big)\right| = |\phi_{h_n'}| = 1. \tag{5.4.14}$$

If we consider the corresponding subsequence of the 2×2 submatrices

$$\mathbf{A}_{n'} = \big((\phi_0, \phi_{h_n'})^t, (\phi_{h_n'}, \phi_0)^t\big),$$

then it follows that the smaller eigenvalue $\lambda_{\mathbf{A}_{n'},2}$ converges to zero. Hence we obtain from Cauchy's interlacing ([106]) theorem that the smallest eigenvalue λ_n of $\mathbf{\Gamma}_{d_n}$ tends to zero, which however contradicts Proposition 5.6.2. Hence we must have

$$\limsup_{n \to \infty} \sup_{h,k,\, h \geq 1} \left|\mathrm{Cov}\big(\xi_k, \xi_{k+h}\big)\right| < 1. \tag{5.4.15}$$

The remaining conditions of Lemma 5.7.1 are now explicitly assumed in the conditions of Theorem 5.2.2, hence the claim follows if $\mathbf{\Gamma}^*_{\boldsymbol{\xi}_{d_n}} = \mathbf{\Gamma}^*_{d_n}$. However, since $(\log n)^{2+\beta} = o\big(n^{-\delta/2}\big)$, it follows that the conditions are also verified in case of $\max\left|\mathbf{\Gamma}^*_{\boldsymbol{\xi}_{d_n}} - \mathbf{\Gamma}^*_{d_n}\right| = \mathcal{O}\left(n^{-\delta/2}\right)$, for large enough n. □

Proof of Theorem 5.2.4. Let $q_0 = q$ be the true order of the AR(q)-process $\{X_k\}_{k \in \mathbb{Z}}$, put

$$\overline{\theta}_{i,n} = a_n^{-1}\sqrt{n}\big(\left|(\widehat{\gamma}_{i,i}^*\widehat{\sigma}^2)^{-1/2}(\widehat{\theta}_i - \theta_i)\right| - b_n\big),$$

and assume first that $k \in \mathbb{N}$, $k > 0$. Note that $\theta_i = 0$ for $i > q$. Then we have that

$$\begin{aligned} P\big(\widehat{q}_{z_n} = k + q\big) &= P\big(\{\overline{\theta}_{q+k,n} > z_n\} \cap \{\max_{k+q+1 \leq i \leq d_n} \overline{\theta}_{i,n} \leq z_n\}\big) \\ &= P\big(\max_{k+q \leq i \leq d_n} \overline{\theta}_{i,n} \leq z_n\big) - P\big(\max_{k+q+1 \leq i \leq d_n} \overline{\theta}_{i,n} \leq z_n\big). \end{aligned}$$

As in the proof of Theorem 5.2.2, we can approximate the sequence $\{\overline{\theta}_{i,n}\}_{1\leq i\leq d_n}$ by a suitably transformed corresponding sequence of mean zero Gaussian random variables $\boldsymbol{\xi}_{d_n} = (\xi_{n,1}, ..., \xi_{n,d_n})^T$ with covariance matrix $\boldsymbol{\Gamma}^*_{\boldsymbol{\xi}_{d_n}}$. Note that this approximation gives us an error term of the order $\mathcal{O}(\kappa_n)$, where κ_n is as in Theorem 5.4.3. Let $\boldsymbol{\eta}_{d_n} = (\eta_{n,1}, ..., \eta_{n,d_n})^T$ be another sequence of IID mean zero Gaussian random variables with unit variance. Following Deo [38], we obtain from $\max\left|\boldsymbol{\Gamma}^*_{\boldsymbol{\xi}_{d_n}} - \boldsymbol{\Gamma}^*_{d_n}\right| = \mathcal{O}(n^{-\delta/2})$ and the Assumptions regarding the covariance structure of $\{\overline{\theta}_{i,n}\}_{1\leq i\leq d_n}$ that for fixed $l \in \mathbb{N}$

$$\begin{aligned} &\left|P\big(\max_{q+l\leq i\leq d_n} a_n^{-1}(|\xi_{n,i}| - b_n) \leq z_n\big) - P\big(\max_{q+l\leq i\leq d_n} a_n^{-1}(|\eta_{n,i}| - b_n) \leq z_n\big)\right| \\ &\leq C \sum_{1\leq i<j\leq d_n} |\rho_{i,j}|\big(d_n^{-\frac{2z_n^2}{1+|\rho_{i,j}|}}\big). \end{aligned}$$

Imitating the technique in Berman [17], we obtain that the above quantity is of the magnitude $o\left(d_n^{(-z_n^2+1)/2}\right)$. We thus obtain that

$$\begin{aligned} &P\big(\widehat{q}_{z_n} = k+q\big) \\ &\qquad = P\big(\max_{q+k+1\leq i\leq d_n} a_n^{-1}(|\eta_{n,i}| - b_n) \leq z_n\big) - P\big(\max_{q+k\leq i\leq d_n} a_n^{-1}(|\eta_{n,i}| - b_n) \leq z_n\big) \\ &\qquad + \mathcal{O}(\kappa_n) + o\left(d_n^{(-z_n^2+1)/2}\right) \\ &\qquad = P\big(a_n^{-1}(|\eta_{n,1}| - b_n) \leq z_n\big)^{d_n-k-q}\left(1 - P\big(a_n^{-1}(|\eta_{n,1}| - b_n) \leq z_n\big)\right) \\ &\qquad + \mathcal{O}(\kappa_n) + o\left(d_n^{(-z_n^2+1)/2}\right). \end{aligned}$$

From the definition of a_n, b_n, and since $z_n \to \infty$, we obtain that (Deo [38])

$$\lim_n P\big(a_n^{-1}(|\eta_{n,1}| - b_n) \leq z_n\big)^{d_n-k-q} \to 1, \tag{5.4.16}$$

$$P\big(a_n^{-1}(|\eta_{n,1}| - b_n) > z_n\big) = \frac{e^{-z_n}}{d_n} + o\left(\frac{e^{-z_n}}{d_n}\right). \tag{5.4.17}$$

This yields

$$P\big(\widehat{q}_{z_n} = k+q\big) = \frac{e^{-z_n}}{d_n} + o\left(\frac{e^{-z_n}}{d_n} + d_n^{(-z_n^2+1)/2}\right) + \mathcal{O}(\kappa_n), \tag{5.4.18}$$

and in particular

$$P\big(\widehat{q}_{z_n} > q\big) = \sum_{k=1}^{d_n} P\big(\widehat{q}_{z_n} = k+q\big) = e^{-z_n} + o\left(e^{-z_n} + d_n^{-z_n^2+2}\right) + \mathcal{O}(\kappa_n d_n), \tag{5.4.19}$$

and per assumption the right hand side goes to zero as n increases. We now consider the case $P\big(\widehat{q}_{z_n} < q\big)$. To this end, let $k \in \mathbb{N}$, $k > 0$. Then we have

$$\begin{aligned} P\big(\widehat{q}_{z_n} = q-k\big) &\le P\big(\overline{\theta}_{q-k,n} \le z_n\big) \\ &= P\Big(a_n^{-1}\big(\big|\xi_{n,q-k} + \sqrt{n}\theta_{q-k}\big| - b_n\big) \le z_n\Big) + \mathcal{O}(\kappa_n). \end{aligned}$$

Since $|\theta_{q-k}| > 0$ and $z_n = o\left(\sqrt{n(\log n)^{-1}}\right)$, one readily verifies by known properties of the Gaussian cdf that $P\Big(a_n^{-1}\big(\big|\xi_{n,q-k} + \sqrt{n}\theta_{q-k}\big| - b_n\big) \le z_n\Big) = \mathcal{O}(\kappa_n)$, and hence

$$P\big(\widehat{q}_{z_n} = q-k\big) = \mathcal{O}(r_n), \tag{5.4.20}$$

and in particular

$$P\big(\widehat{q}_{z_n} < q\big) = \mathcal{O}(\kappa_n q) \to 0, \tag{5.4.21}$$

as n increases. This together with (5.4.19) establishes consistency. □

Proof of Theorem 5.2.8. Let us first assume that the the true order q_0 of the autoregressive process is fixed, i.e. $q_0 = q$ for some finite q.

The proof then consists of two parts. It is first shown that $P\big(\widehat{q}_n^* < q\big) \to 0$, whereas in the second part the claim $P\big(\widehat{q}_n^* > q\big) \to 0$ is established.

By Hannan [60, Chapter VI], it holds that for $k \in \mathbb{N}$

$$\log\big(\widehat{\sigma}^2(k)\big) = \log\widehat{\phi}_{n,0} + \sum_{j=1}^{k}\log(1 - \widehat{\theta}_j^2(k)). \tag{5.4.22}$$

Then, arguing as in Hannan and Quinn [63], we have due to $C_n = o(n)$ that for large enough n

$$f_n(k) = \log\big(\widehat{\sigma}^2(k)\big) + n^{-1}C_n k$$

is a decreasing function in k for $0 \le k < q$, and strictly decreasing for $q-1 \le k \le q$ (since $\theta_q^2 > 0$) with probability approaching one. This implies that eventually $\widehat{q}_n^* \ge q$, hence it suffices to establish that the probability of overestimating the order goes to zero as n increases, i.e.

$$\lim_n P\left((\min\big\{q \le k \le K_n \;\big|\; \log\big(\widehat{\sigma}^2(k)\big) + n^{-1}C_n k\big\} \ge q+1\right) = 0. \tag{5.4.23}$$

Using relation (5.4.22), the above can be rewritten as

$$P\left(\min\big\{q \le k \le K_n \;\big|\; \log\big(\widehat{\sigma}^2(k)\big) + n^{-1}C_n k\big\} \ge q+1\right) \tag{5.4.24}$$

$$= P\left(\min\Big\{q \le k \le K_n \;\Big|\; \log\widehat{\phi}_{n,0} + \sum_{j=1}^{k}\log(1-\widehat{\theta}_j^2(k)) + n^{-1}C_n k\Big\} \ge q+1\right). \tag{5.4.25}$$

Using the same arguments as in [3], it follows that it suffices to establish

$$\lim_n P\left(\max_{1\leq k\leq K_n-q}\left(\sum_{j=1+q}^{k+q} -\log(1-\widehat{\theta}_j^2(k))\right)\geq n^{-1}C_n k\right)=0. \tag{5.4.26}$$

However, it holds that

$$\begin{aligned} &P\left(\max_{1\leq k\leq K_n-q}\left(\sum_{j=1+q}^{k+q} -\log(1-\widehat{\theta}_j^2(k))\right)\geq n^{-1}C_n k\right)\\ &\leq \sum_{k=1}^{K_n-q} P\left(\sum_{j=1+q}^{k+q} -\log(1-\widehat{\theta}_j^2(k))\geq n^{-1}C_n k\right)\\ &\leq \sum_{k=1}^{K_n-q}\sum_{j=1+q}^{k+q} P\left(-\log(1-\widehat{\theta}_j^2(k))\geq n^{-1}C_n\right), \end{aligned} \tag{5.4.27}$$

hence it suffices to treat

$$P\left(-\log(1-\widehat{\theta}_j^2(k))\geq n^{-1}C_n\right)\leq P\left(|\widehat{\theta}_j(k)|\geq n^{-1/2}\sqrt{C_n}\right),$$

for large enough n and $q<j\leq K_n$. Since it is no longer of any relevance, we drop the index k for the sake of simplicity. Since $\theta_j=0$ for $q<j\leq K_n$, following the proof of Theorem 5.4.2 (which amounts to setting $d_n=1$ in Theorem 5.4.2), we obtain that

$$P\Big(\Big|n^{1/2}\widehat{\theta}_j - n^{-1/2}(\gamma_{j,j}^*)^{-1}\sum_{i=1}^n U_i^{(j)}\Big|\geq\epsilon\Big)=\mathcal{O}\left(\epsilon^{-p}n^{-p/2}(\log n)^p+n^{-1}\right).$$

In addition, we have from Corollary 5.5.3 that on a possible larger probability space

$$P\left(\left|\left(n^{-1/2}(\gamma_{j,j}^*)^{-1}\sum_{i=1}^n U_i^{(j)}\right)-\xi_j\right|\geq n^{1/\nu-1/2}\right)=\mathcal{O}\left(n^{-1}\right),$$

where $\nu\geq 2$ and ξ_j is a Gaussian random variable with positive variance. Setting $\epsilon=2n^{1/\nu-1/2}$ we thus obtain

$$P\Big(\Big|n^{1/2}\widehat{\theta}_j-\xi_j\Big|\geq\epsilon\Big)=\mathcal{O}\left(\epsilon^{-p}n^{-p/2}(\log n)^p+n^{-1}\right).$$

By known properties of the Gaussian cdf, we have

$$P\left(|\xi_j|\geq\sqrt{C_n}\right)=\mathcal{O}\left(C_n^{-1/2}\exp(-C_n/2)\right),$$

thus, using the bounds derived above, we obtain

$$P\left(|\widehat{\theta}_j(k)| \geq n^{-1/2}\sqrt{C_n} - \epsilon\right) = \mathcal{O}\left(C_n^{-1/2}\exp(-C_n/2) + \epsilon^{-p}n^{-p/2}(\log n)^p + n^{-1}\right),$$

where $\epsilon = 2n^{1/\nu-1/2}$. Having (5.4.27) in mind, we thus require that

$$K_n^2\left(C_n^{-1/2}\exp(-C_n/2) + \epsilon^{-p}n^{-p/2}(\log n)^p + n^{-1}\right) \to 0$$

as n increases. Since $p > 4$, this amounts to

$$K_n = \scriptstyle{\mathcal{O}}\left(n^{1/2}\right), \quad 4\log K_n \leq C_n,$$

which completes the proof if the order of the process is finite.

Let us now consider the case of unbounded order, i.e. $q_0 = d_n$. The above arguments are still valid to establish $\lim_n P\big(\widehat{q}_n^* > d_n\big) = 0$, however, showing that $\lim_n P\big(\widehat{q}_n^* < d_n\big) = 0$ turns out to be a little more involved. To establish this claim, note that by Theorem 5.4.3 we have

$$\max\big|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\big| \leq n^{-1/4}$$

with probability approaching one. Thus we obtain

$$\max\big|\widehat{\mathbf{\Theta}}^2 - \mathbf{\Theta}^2\big| \leq \max\big|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\left(\big|\max\big|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\big| + 2\max\big|\mathbf{\Theta}\big|\right) \leq 3n^{-1/4} \quad (5.4.28)$$

with probability approaching one, since $\max\big|\mathbf{\Theta}^2\big| \leq \vartheta < 1$. Let $x_i \in \{\theta_i - 3n^{-1/4}, \theta_i + 3n^{-1/4}\}$. Then we deduce from the above, that for large enough n, $x_i \in \{-3n^{-1/4}, \vartheta^*\}$ where $|\vartheta^*| < 1$. This in turn implies that for large enough n

$$\begin{aligned}\sum_{j=0}^{k}\big|\log(1-\widehat{\theta}_j^2(k)) - \log(1-\theta_j^2(k))\big| &\leq \sum_{j=0}^{k}\big|\widehat{\theta}_j^2(k) - \theta_j^2(k)\big|\frac{1}{|1-x_i|} \\ &\leq 3kn^{-1/4}\frac{1}{1-\vartheta^*},\end{aligned}$$

with probability approaching one. Since $d_n n^{-1/4} = \scriptstyle{\mathcal{O}}(1)$, relation (5.4.22) thus yields that

$$\log\big(\widehat{\sigma}^2(k)\big) - \log\widehat{\phi}_{n,0} = \sum_{j=1}^{k}\log(1-\widehat{\theta}_j^2(k)) = \sum_{j=1}^{k}\log(1-\theta_j^2(k)) + \mathcal{O}\big(kn^{-1/4}\frac{1}{1-\vartheta^*}\big)$$

is a monotone decreasing function in $0 \leq k < d_n$, strictly decreasing in $k = d_n$, for large enough n with probability approaching one. This yields that $\lim_n P\big(\widehat{q}_n^* < q\big) = 0$. □

5.5 Gaussian approximation

In this section we obtain under suitable assumptions, a normal approximation for the quantity $n^{-1/2}\mathbf{X}^{\mathbf{T}}\mathbf{Z}$, where we used the notation introduce in Section 5.4.

As in the proof of Theorem 5.4.3, let $\mathbf{U}_k = \big(X_{k-1}, ..., X_{k-d_n}\big)^T \epsilon_k$, $k \in \mathbb{N}$. We have

$$n^{-1/2}\mathbf{X}^{\mathbf{T}}\mathbf{Z} = n^{-1/2}\sum_{k=1}^{n}\mathbf{U}_k,$$

with $\mathbb{E}\big(\mathbf{U}_k\big) = 0$ and

$$\mathbb{E}\big(\mathbf{U}_k\mathbf{U}_{k+h}^T\big) = \begin{cases} \sigma^2\Gamma_{d_n}, & \text{if } h = 0, \\ 0_{d_n\times d_n}, & \text{if } h \neq 0, \end{cases} \tag{5.5.1}$$

since $\epsilon_{\mathbf{k}}$ is independent of $\big\{X_{k-i}\big\}_{i\geq 1}$.

The main Theorem is formulated below.

Theorem 5.5.1. *Let $\nu \geq 2$. Then under Assumption 5.2.1, on a possible larger probability space, there exists a d_n-dimensional Gaussian vector $\boldsymbol{\xi}^{d_n} = \big(\xi^{(1)}, ..., \xi^{(d_n)}\big)^T$ with covariance matrix $\mathbf{\Gamma}_{\boldsymbol{\xi}^{d_n}}$, such that* $\max\big|n^{-1}\mathbf{\Gamma}_{\boldsymbol{\xi}^{d_n}} - \sigma^2\mathbf{\Gamma}_{d_n}\big| = \mathcal{O}\left(n^{-\delta/2}\right)$*, and*

$$P\left(\max\left|\left(\sum_{k=1}^{n}\mathbf{U}_k\right) - \boldsymbol{\xi}^{d_n}\right| \geq n^{1/\nu}\right) = \mathcal{O}\left(n^{-1}\right),$$

where $d_n = n^\delta$ and δ is chosen such that

$$\delta < \min\left\{\frac{8+3p-(4+p)\nu}{(8+5p)\nu}, \frac{(3\nu-4)(4+5p)}{(52+25p)\nu-16}\right\}.$$

Remark 5.5.2. Note that for $\nu = 2$ and $p \geq 4$ this results in $0 < \delta < 1/14$.

Corollary 5.5.3. *Let $\mathbf{\Gamma}_{d_n}^{-1} = \big(\gamma_{i,j}^*\big)_{1\leq i,j\leq d_n}$ be the inverse of the covariance matrix $\mathbf{\Gamma}_{d_n}$. Then under the same conditions as in Theorem 6.6.4, we have that*

$$P\left(\max\left|\mathbf{\Gamma}_{d_n}^{-1}\left\{\left(\sum_{k=1}^{n}\mathbf{U}_k\right) - \boldsymbol{\xi}^{d_n}\right\}\right| \geq n^{1/\nu}\right) = \mathcal{O}\left(n^{-1}\right).$$

Proof. By relation (5.4.1) and condition $\sum_{i=1}^{d_n}\big|\theta_i\big| = \vartheta < 1$ we have that for fixed i

$$\sum_{j=1}^{d_n}|\gamma_{i,j}^*| \leq \sum_{j=1}^{d_n}\left(\sum_{r=0}^{\alpha}\big|\theta_r\theta_{r+j-i}\big| + \sum_{r=\beta}^{d_n}\big|\theta_r\theta_{r+j-i}\big|\right) \leq 2\vartheta^2.$$

Hence the claim readily follows from

$$\max\left|\mathbf{\Gamma}_{d_n}^{-1}\left\{\left(\sum_{k=1}^{n}\mathbf{U}_k\right)-\boldsymbol{\xi}^{d_n}\right\}\right|\leq 2\max\left|\left(\sum_{k=1}^{n}\mathbf{U}_k\right)-\boldsymbol{\xi}^{d_n}\right|.$$

□

The proof of Theorem 6.6.4 follows [14, Theorem 4.1] in broad brushes, with some essential changes in the details. To this end, we require some preliminary results. For an n-dimensional vector $\mathbf{x}=(x_1,...,x_n)$, we denote with $|\mathbf{x}|_n=\left(\sum_{i=1}^n (x_i)^2\right)^{1/2}$ the usual euclidian norm. The following coupling inequality is due to Berthet and Mason [19].

Lemma 5.5.4 (Coupling inequality). *Let $X_1,...,X_N$ be independent, mean zero random vectors in $\mathbb{R}^n$, $n\geq 1$, such that for some $B>0$, $|X_i|_n\leq B$, $i=1,...,N$. If the probability space is rich enough, then for each $\delta>0$, one can define independent normally distributed mean zero random vectors $\xi_1,...,\xi_N$ with ξ_i and X_i having the same variance/covariance matrix for $i=1,...,N$, such that for universal constants $C_1>0$ and $C_2>0$,*

$$P\left\{\big|\sum_{i=1}^{N}(X_i-\xi_i)\big|_n>\delta\right\}\leq C_1 n^2\exp\left(-\frac{C_2\delta}{Bn^2}\right).$$

The proof of Theorem 6.6.4 is based on a blocking argument, which in turn requires carefully truncated random variables. By Lemma 5.6.1, we can present X_k as $X_k=\sum_{i=0}^{\infty}\psi_i\epsilon_{k-i}$. Set $X_k^{(n)}=\sum_{i=0}^{\lceil A\log n\rceil}\psi_i\epsilon_{k-i}$, $A>0$, and define $\mathbf{X}^{(n)}$ and $\mathbf{U}_k^{(n)}$ in an analogue manner, and $\mathbf{V}_k^{(n)}$ such that $\mathbf{U}_k=\mathbf{U}_k^{(n)}+\mathbf{V}_k^{(n)}$. Then $\big\{\mathbf{U}_k^{(n)}\big\}_{0\leq k\leq n}$ are $2\lceil A\log n\rceil$-dependent sequences, with $\mathbb{E}\big(\mathbf{U}_k^{(n)}\big)=\mathbb{E}\big(\mathbf{V}_k^{(n)}\big)=0$ and

$$\mathbb{E}\big(\mathbf{U}_k^{(n)}\mathbf{U}_{k+h}^{(n),T}\big)=\begin{cases}\sigma^2\Gamma_{d_n}^{(n)}, & \text{if } h=0,\\ 0_{d_n\times d_n}, & \text{if } h\neq 0,\end{cases}\tag{5.5.2}$$

Denote with $\mathbf{U}_k^{(n)}=\big(U_k^{(1)},...,U_k^{(d_n)}\big)^t$, $\mathbf{V}_k^{(n)}=\big(V_k^{(1)},...,V_k^{(d_n)}\big)^t$ the single components of $\mathbf{U}_k^{(n)}$ and $\mathbf{V}_k^{(n)}$.

Lemma 5.5.5. *There is an absolute constant C such that*

$$\mathbb{E}\Big|\sum_{l\leq i\leq k}U_i^{(h)}\Big|^p\leq C\big((k-l+1)\lceil A\log n\rceil\big)^{p/2}.$$

Remark 5.5.6. An analogue result is valid for $\mathbf{V}_k^{(n)}$.

Proof of Lemma 6.6.10. Put $K = \lceil A \log n \rceil$, and denote with $\|.\|_p = (\mathbb{E}|.|^p)^{1/p}$. Then per construction, we can rewrite

$$\sum_{l \le i \le k} U_i^{(h)} = S_1 + ... + S_K,$$

where S_i is a sum of independent random variables with at most $(k - l + 1)/K$ terms. Minikowski's inequality gives us

$$\|S_1 + ... + S_K\|_p \le \|S_1\|_p + ... + \|S_K\|_p.$$

By Rosenthal's inequality, we have

$$\mathbb{E}\big|S_i\big|^p \le C\big((k - l + 1)/K\big)^{p/2} = C\left((k - l + 1)/K\right)^{p/2},$$

hence

$$\Big\|\sum_{l \le i \le k} U_i^{(h)}\Big\|_p^p \le C\big((k - l + 1)K\big)^{p/2}.$$

□

Proof of Theorem 6.6.4. The proof is based on two truncation and a blocking argument. The first truncation consist of approximating $\sum_{k=1}^n \mathbf{U}_k$ with $\sum_{k=1}^n \mathbf{U}_k^{(n)}$. Due to the Markov inequality and Lemma 5.6.1, we can always find an A large enough such that

$$P\bigg(\max\Big|\sum_{j=1}^n \big(\mathbf{U_j} - \mathbf{U_j}^{(n)}\big)\Big| \ge n^{1/\nu}\bigg) = \mathcal{O}\left(n^{-1}\right),$$

and we can move on to the second truncation step, which is a little more involved. For a random variable ϵ, let $\mathbf{I}_B^{\epsilon} = \mathbf{1}(\epsilon)_{\{|\epsilon| \le B\}}$ for $B > 0$, and similarly, $\mathbf{I}_{B^c}^{\epsilon} = \mathbf{1}(\epsilon)_{\{|\epsilon| > B\}}$. Consider the quantity

$$I_B = \epsilon_k \sum_{i=0}^{\infty} \psi_i \epsilon_{k-i-1} - \epsilon_k \mathbf{I}_B^{\epsilon_k} \sum_{i=0}^{\infty} \psi_i \epsilon_{k-i-1} \mathbf{I}_B^{\epsilon_{k-i-1}}.$$

Applying the Minikowski, Cauchy-Schwarz and Markov inequality we obtain for $p \ge 1$

$$\begin{aligned}\big\|I_B\big\|_2 &\le \sum_{i=0}^{\infty} |\psi_i| \big\|\epsilon_k \mathbf{I}_{B^c}^{\epsilon_k} \epsilon_{k-i-1}\big\|_2 + \sum_{i=0}^{\infty} |\psi_i| \big\|\epsilon_k \epsilon_{k-i-1} \mathbf{I}_{B^c}^{\epsilon_{k-i-1}}\big\|_2 \\ &\le C\big\|\epsilon_1\big\|_4^2 \big(\mathbb{E}(\mathbf{I}_{B^c}^{\epsilon_1})\big)^{1/4} \le C\big\|\epsilon_1\big\|_4^2 \big\|\epsilon_1\big\|_p^{p/4} B^{-p/4}. \qquad (5.5.3)\end{aligned}$$

One readily verifies that this bound is still valid if we replace $\epsilon_{k-i}\mathbf{I}_B^{\epsilon_{k-i}}$ by the centered version $\epsilon_{k-i}\mathbf{I}_B^{\epsilon_{k-i}} - \mathbb{E}\big(\epsilon_{k-i}\mathbf{I}_B^{\epsilon_{k-i}}\big)$. If we us the truncation introduced above, we can construct the truncated version $\mathbf{U}_k^{(n,B)} = \big(U_k^{(1,B)}, ..., U_k^{(d_n,B)}\big)^t$ of $\mathbf{U}_k^{(n)}$. Using the Markov and Cauchy-Schwarz inequality together with Lemma 6.6.10 we have

$$\begin{aligned} P\big(\max\big|\sum_{k=1}^n \mathbf{U}_k^{(n)} - \mathbf{U}_k^{(n,B)}\big| \geq x\big) &\leq C d_n x^{-1} \max_{1\leq h\leq d_n} \big\|\sum_{k=1}^n U_k^{(h)} - U_k^{(h,B)}\big\|_2 \\ &\leq C d_n x^{-1} (n\log n)^{1/2} \big\|\epsilon_{k-i}\mathbf{I}_B^{\epsilon_{k-i}} - \mathbb{E}\big(\epsilon_{k-i}\mathbf{I}_B^{\epsilon_{k-i}}\big)\big\|_2, \end{aligned}$$

and using (5.5.3) the above is of the magnitude

$$P\big(\max\big|\sum_{k=1}^n \mathbf{U}_k^{(n)} - \mathbf{U}_k^{(n,B)}\big| \geq x\big) = \mathcal{O}\left(d_n x^{-1} (n\log n)^{1/2} B^{-p/4}\right). \tag{5.5.4}$$

We will discuss the this bound later. We will now construct a Gaussian approximation for the random vectors $\mathbf{U}_k^{(n,B)}$. To this end, let $\nu, \beta, \delta, b, p > 0$ be numbers such that

$$\frac{1}{\nu} - \frac{\beta}{1+\beta} - 2b - 3\delta > 0, \quad 2/\nu - 1/(1+\beta) - 2\delta > 0, \tag{5.5.5}$$

$$\frac{\beta}{\beta+1} > \delta, \quad -pb/4 + 1/2 - 1/\nu + \delta < 0. \tag{5.5.6}$$

We will now construct an approximation for the truncated random vector $\mathbf{U}_k^{(n,B)}$. To this end, we first divide the set of integers $\{1, 2, ...\}$ into consecutive blocks $H_1, J_1, H_2, J_2 ....$ The blocks are defined by recursion. Fix $\beta > 0$. If the largest element of J_{i-1} is k_{i-1}, then $H_i = \{k_{i-1}+1, ..., k_{i-1}+i^\beta\}$ and $J_i = \{k_{i-1} + i^\beta + 1, ..., k_i\}$, where $k_i = \min\{l : l - 2\lceil d_n\rceil \geq k_{i-1} + i^\beta\}$. Let $|.|$ denote the cardinality of a set. It follows from the definition of H_i, J_i that $|H_i| = i^\beta$ and $|J_i| = \mathcal{O}(d_n)$. Note that the total number of blocks is approximately $m = n^{1/(1+\beta)}$ if $\beta(1+\beta)^{-1} > \delta$. For $1 \leq h \leq d_n$, let

$$\xi_k^{(h)} = \sum_{i\in H_k} \big(U_i^{(h,B)}\big) \quad \text{and} \quad \eta_k^{(h)} = \sum_{i\in J_k} U_i^{(h,B)},$$

and define the vectors

$$\boldsymbol{\xi}_k = \big(\xi_k^{(1)}, \xi_k^{(2)}, ..., \xi_k^{(d_n)}\big)^T \quad \text{and} \quad \boldsymbol{\eta}_k = \big(\eta_k^{(1)}, \eta_k^{(2)}, ..., \eta_k^{(d_n)}\big)^T.$$

Note that per construction, we have that $\big\{\boldsymbol{\xi}_k\big\}_{k\in\mathbb{N}}$ is a sequence of independent random vectors with $|\boldsymbol{\xi}_k|_{d_n} \leq d_n m^\beta B^2$. By Lemma 6.6.9, we can define a sequence of independent normal random vectors $\boldsymbol{\xi}_k^* = \left(\xi_k^{(1,*)}, \xi_k^{(2,*)}, ..., \xi_k^{(d_n,*)}\right)^T$, such that for $x > 0$

$$\begin{aligned} P\left(\max_{0 \le h \le d_n} \left| \sum_{j=1}^m \left(\xi_j^{(h)} - \xi_j^{(h),*}\right) \right| \ge x \right) &\le \sum_{h=1}^{d_n} P\left(\left| \sum_{j=1}^m \left(\xi_j^{(h)} - \xi_j^{(h),*}\right) \right| \ge x \right) \\ &\le \sum_{h=0}^{d_n} P\left(\left| \sum_{j=1}^m \left(\xi_{\mathbf{j}} - \xi_{\mathbf{j}}^*\right) \right|_{d_n} \ge x \right) \\ &\le C_1 d_n^3 \exp\left(-\frac{C_2 x}{2 d_n^3 m^\beta B^2} \right). \end{aligned}$$

Hence, if $1/\nu - \beta/(1+\beta) - 2b - 3\delta > 0$, we obtain

$$P\left(\max_{0 \le h \le d_n} \left| \sum_{j=1}^m \left(\xi_j^{(h)} - \xi_j^{(h),*}\right) \right| \ge n^{1/\nu} \right) = \mathcal{O}\left(n^{-1}\right). \tag{5.5.7}$$

Similar arguments show that under the same conditions as above, there exists a sequence of independent normal random vectors $\boldsymbol{\eta}_k^* = \left(\eta_k^{(1,*)}, \eta_k^{(2,*)}, ..., \eta_k^{(d_n,*)}\right)^T$, such that

$$P\left(\max_{0 \le h \le d_n} \left| \sum_{j=1}^m \left(\eta_j^{(h)} - \eta_j^{(h),*}\right) \right| \ge n^{1/\nu} \right) = \mathcal{O}\left(n^{-1}\right).$$

By Lemma 6.6.10, we have that $\mathrm{Var}\left(\eta_j^{(h),*}\right) \le C d_n^2$ for all $j \le m$, $h \le d_n$. Hence if $2/\nu - 1/(1+\beta) - \delta > 0$, by known properties of the tails of a normal cdf, we obtain that

$$P\left(\max_{0 \le h \le d_n} \left| \sum_{j=1}^m \eta_j^{(h),*} \right| \ge n^{1/\nu} \right) \le \sum_{h=0}^{d_n} P\left(\left| \sum_{j=1}^m \eta_j^{(h),*} \right| \ge n^{1/\nu} \right) = \mathcal{O}\left(n^{-1}\right). \tag{5.5.8}$$

This yields

$$P\left(\max_{0 \le h \le d_n} \left| \sum_{j=1}^m \left(\eta_j^{(h)} + \xi_j^{(h),} - \xi_j^{(h),*}\right) \right| \ge n^{1/\nu} \right) = \mathcal{O}\left(n^{-1}\right). \tag{5.5.9}$$

Let $\boldsymbol{\eta}_k^{**} = \left(\eta_k^{(1,**)}, \eta_k^{(2,**)}, ..., \eta_k^{(d_n,**)}\right)^T$ be a copy of $\boldsymbol{\eta}_k^*$ such that $\boldsymbol{\eta}_i^{**}$ and $\boldsymbol{\xi}_j^*$ are independent for $i \ne j$. By enlarging the probability space if necessary, we can then construct $\boldsymbol{\eta}_k^{**}$ in such a manner that

$$\mathrm{Var}\left(\boldsymbol{\xi}_k^* + \boldsymbol{\eta}_k^{**}\right) = (|H_k| + |J_k|)\mathrm{Var}\left(\mathbf{U}_k^{(n,B)}\right). \tag{5.5.10}$$

Piecing everything together, we obtain that

$$P\Big(\max_{0\le h\le d_n}\Big|\sum_{j=1}^{m}\big(\xi_j^{(h)}+\eta_j^{(h)}-\xi_j^{(h,*)}-\eta_j^{(h,**)}\big)\Big|\ge n^{1/\nu}\Big)=\mathcal{O}\left(n^{-1}\right).$$

Finally, we obtain from the above

$$P\Big(\max\Big|\sum_{j=1}^{n}\mathbf{U}_k-\sum_{j=1}^{m}\big(\xi_{\mathbf{j}}{}^{*}-\eta_{\mathbf{j}}{}^{**}\big)\Big|\ge 2n^{1/\nu}\Big)\le P\Big(\max\Big|\sum_{j=1}^{n}\big(\mathbf{U_j}-\mathbf{U_j}^{(n,B)}\big)\Big|\ge n^{1/\nu}\Big)+\mathcal{O}\left(n^{-1}\right).$$

Due to (5.5.4), we have that

$$P\Big(\max\Big|\sum_{j=1}^{n}\big(\mathbf{U_j}-\mathbf{U_j}^{(n,B)}\big)\Big|\ge 2n^{1/\nu}\Big)\le P\Big(\max\Big|\sum_{j=1}^{n}\big(\mathbf{U_j}-\mathbf{U_j}^{(n)}\big)\Big|\ge n^{1/\nu}\Big)$$
$$+\,C\,d_n n^{1/2-1/\nu}(\log n)B^{-p/4}.$$

Now choose $B=n^b$. Then we obtain from the inequalities in (5.5.5), that δ is largest for $\beta=\frac{-4+3\nu-2b\nu}{4+2\nu+4b\nu}$, which results in $\delta<\min\{1/5\,(-1-2b+3/\nu)\,,\beta(1+\beta)^{-1}\}$. On the other hand, we obtain from the above that $-pb/4+1/2-1/\nu+\delta<0$. Hence, choosing $b=(6\nu-8)(8\nu+5\nu p)^{-1}$ yields

$$\delta<\min\left\{\frac{8+3p-(4+p)\nu}{(8+5p)\nu},\frac{(3\nu-4)(4+5p)}{(52+25p)\nu-16}\right\},$$

which gives the desired bound. It thus remains to deal with the covariances. Note that the truncation and blocking arguments have altered the covariance structure. To quantify the error which stems from the first truncation, note that by the Cauchy-Schwarz inequality we have

$$\max\Big|\mathrm{Cov}\big(\sum_{k=1}^{n}\mathbf{U_k},\sum_{k=1}^{n}\mathbf{U_k}\big)-\mathrm{Cov}\big(\sum_{k=1}^{n}\mathbf{U_k}^{(n)},\sum_{k=1}^{n}\mathbf{U_k}^{(n)}\big)\Big|$$
$$\le\max_{1\le i,j\le d_n}\left(\sqrt{\mathrm{Var}\big(\sum_{k=1}^{n}U_k^{(j)}\big)\mathrm{Var}\big(\sum_{k=1}^{n}V_k^{(j)}\big)}+\mathrm{Var}\big(\sum_{k=1}^{n}V_k^{(j)}\big)\right),$$

which by Lemma 6.6.10 is bounded by

$$C\,n\log n\max_{1\le j\le d_n}\sqrt{V_1^{(j)}}=\mathcal{O}\left(n^{-1}\right),$$

where the last inequality is valid if we choose A large enough. Similar arguments, using the results from (5.5.4), one obtains that the error in the covariance structure which results from bounding the random vectors $\mathbf{U_k}^{(n)}$ is of the order $\mathcal{O}\left(n^{3/2}(\log n)^{1/2}B^{-p/8}\right)$. Finally, in a similar manner, one obtains that the error which results from the blocking argument is of the size $\mathcal{O}\left(n^{\frac{2+\beta}{2(1+\beta)}}d_n^{1/2}\right)$. Piecing these bounds together, we obtain that

$$\max\left|n^{-1}\mathbf{\Gamma}_{\boldsymbol{\xi}^{d_n}} - \sigma^2\mathbf{\Gamma}_{d_n}\right| = \mathcal{O}\left(n^{\frac{-\beta}{2(1+\beta)}}d_n^{1/2}(n\log n)^{1/2}n^{-pb/8}\right).$$

Using the relations (5.5.5), we deduce that

$$\max\left|n^{-1}\mathbf{\Gamma}_{\boldsymbol{\xi}^{d_n}} - \sigma^2\mathbf{\Gamma}_{d_n}\right| = \mathcal{O}\left(n^{1/\nu-\delta/2}\right) = \mathcal{O}\left(n^{-\delta/2}\right). \tag{5.5.11}$$

□

5.6 Proofs of the auxiliary results

5.6.1 Proofs of Section 5.4

We will first show Proposition 5.4.1. To this end, we require some auxiliary results.

Lemma 5.6.1. *Let $\{X_k\}_{k\in\mathbb{Z}}$ be an AR(q) process such that Assumption 5.2.1 is satisfied. Then*

(i) $X_k = \sum_{i=0}^{\infty}\psi_i\epsilon_{k-i}$, *with* $|\psi_i| \leq \vartheta^i$,

(ii) $\left|Cov\big(X_k, X_{k+h}\big)\right| = \mathcal{O}\left(\rho^h\right)$, $|\rho| < 1$,

(iii) $\sqrt{n}\left\|\widehat{\phi}_{n,h} - \phi_h\right\|_p = \mathcal{O}(1)$, $p \geq 1$.

Proof. Property (i) simply follows by iterating the recurrence relation of an AR(q) process, and (ii) follows readily from (i). In order to show (ii), let $\mathcal{F}_j = \sigma\big(\epsilon_i,\ i \leq j\big)$ and $S_j^{(h)} = \sum_{k=1}^{j}\big(X_kX_{k+h} - \mathbb{E}(X_kX_{k+h})\big)$. Note that by (i) we have

- $\left\|X_kX_{k+h} - \mathbb{E}(X_kX_{k+h})\right\|_p < \infty$, and
- $\delta_{\infty,p} = \sum_{j=1}^{\infty}j^{-3/2}\left\|\mathbb{E}\big(S_j^{(h)}\big|\mathcal{F}_0\big)\right\|_p < \infty$.

Then by Theorem 1 in [102] we have

$$\sqrt{n}\left\|\widehat{\phi}_{n,h} - \phi_h\right\|_p \leq C_p\left(\left\|X_kX_{k+h} - \mathbb{E}(X_1X_{1+h})\right\|_p + 80\delta_{\infty,p}\right) = \mathcal{O}\big(1\big),$$

hence the claim follows. □

The next two Propositions help us to establish bounds for the eigenvalues of the covariance matrix $\mathbf{\Gamma}_{d_n}$.

Proposition 5.6.2. *Let $\{X_k\}_{k\in\mathbb{Z}}$ be an AR(q) process, such that Assumption 5.2.1 is satisfied. Denote with $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ the eigenvalues of the covariance matrix Γ_n of $\big(X_1, ..., X_n\big)^T$. Then we have that*

$$\frac{\sigma^2}{(1+\vartheta)^2} \leq \lambda_1 \leq \lambda_n \leq \frac{\sigma^2}{\vartheta^2}.$$

In order to proof Proposition 5.6.2, we require the following auxiliary result.

Proposition 5.6.3. *Let $\{X_k\}_{k\in\mathbb{Z}}$ be an AR(q) process, such that Assumption 5.2.1 is satisfied. Then we have that*

$$\sup_\lambda f_X(\lambda) \leq \frac{\sigma^2}{2\pi\vartheta^2}, \quad \inf_\lambda f_X(\lambda) \geq \frac{\sigma^2}{2\pi(1+\vartheta)^2},$$

where $\sigma^2 = \mathbb{E}\big(\epsilon^2\big)$.

Proof of Proposition 5.6.3. By Lemma 5.6.1 we have that $\{X_k\}_{k\in\mathbb{Z}}$ is causal, and can be represented as $X_k = \sum_{i=0}^{\infty} \psi_i \epsilon_{k-i}$. Since the innovations $\{\epsilon_k\}_{k\in\mathbb{Z}}$ have a spectral density function $f_\epsilon(\lambda) = \sigma^2/(2\pi)$, we obtain from [25, Theorem 4.4.1] that the density function $f_X(\lambda)$ exists, and by [25, Theorem 4.4.2] $f_X(\lambda)$ is given as

$$f_X(\lambda) = \frac{\sigma^2}{2\pi\big|\underline{\theta}\big(e^{-\mathrm{i}\lambda}\big)\big|^2},$$

where $\underline{\theta}(s) = 1 - \sum_{j=1}^{q} \theta_j s^j$. Since $\theta\big(e^{-\mathrm{i}\lambda}\big) = 1 - \sum_{j=1}^{q} \theta_i e^{-\mathrm{i}\lambda j}$, it holds that

$$\big|1 - \sum_{j=1}^{q} |\theta_j|\big|^2 \leq \big|\theta\big(e^{-\mathrm{i}\lambda}\big)\big|^2 \leq \big|1 + \sum_{j=1}^{q} |\theta_j|\big)\big|^2,$$

hence the claim follows. □

Proposition 5.6.2 can now readily be deduced.

Proof of Proposition 5.6.2. By [25, Proposition 4.5.3], we have that

$$2\pi \inf_\lambda f_X(\lambda) \leq \lambda_1 \leq \lambda_n \leq 2\pi \sup_\lambda f_X(\lambda),$$

hence the claim follows from Proposition 5.6.3. □

We are now in the position to show Proposition 5.4.1. To this end, we will introduce the following matrix norm. For a matrix $\mathbf{A} = \big(a_{i,j}\big)_{1\le i,j\le d_n}$ we define

$$\big\|\mathbf{A}\big\| = \sup\big|\mathbf{A}\mathbf{x}\big|_{d_n},\ \big|\mathbf{x}\big|_{d_n} \le 1,\ \mathbf{x} \in \mathbb{R}^{d_n},$$

where $|.|_{d_n}$ denotes the usual d_n-dimensional euclidian norm.

Proof of Proposition 5.4.1. We first remark that Proposition 5.6.2 implies that

$$\big\|\mathbf{\Gamma}_{d_n}\big\| \le \frac{\sigma^2}{\vartheta^2}, \quad \big\|\mathbf{\Gamma}_{d_n}^{-1}\big\| \le \frac{(1+\vartheta)^2}{\sigma^2}.$$

We introduce the following abbreviations. Put

$$E = \big\|\mathbf{\Gamma}_{d_n}^{-1}\big\|,\ F = \big\|\widehat{\mathbf{\Gamma}}_{d_n}^{-1} - \mathbf{\Gamma}_{d_n}^{-1}\big\|, G = \big\|\widehat{\mathbf{\Gamma}}_{d_n} - \mathbf{\Gamma}_{d_n}\big\|.$$

One readily verifies that

$$G = \big\|\widehat{\mathbf{\Gamma}}_{d_n} - \mathbf{\Gamma}_{d_n}\big\| \le \sum_{i,j\le d_n} \big(\widehat{\phi}_{n,|i-j|} - \phi_{|i-j|}\big)^2. \tag{5.6.1}$$

By Lemma 5.6.1 we have $\sqrt{n}\big\|\widehat{\phi}_{n,|i-j|} - \phi_{|i-j|}\big\|_p \le C_p$ for some finite constant C_p. Thus the Markov inequality in connection with Minikowski's inequality implies

$$P\big(\big\|\widehat{\mathbf{\Gamma}}_{d_n} - \mathbf{\Gamma}_{d_n}\big\| \ge (\log n)^{-1}\big) = \mathcal{O}\left(\frac{(d_n \log n)^p}{n^{p/2}}\right). \tag{5.6.2}$$

It follows from Lemma 3 in [13] that

$$F \le (E+F)G\,E,$$

and in particular if $E\,G < 1$

$$F \le E^2 G/(1-EG).$$

Hence we have for sufficiently large n

$$P\big(F \ge \epsilon\big) \le P\big(G \ge (\log n)^{-1}\big) + P\big(G \ge E^2/2\epsilon\big).$$

Since $E < C$, where C does not depend on n, we obtain from (5.6.1), (5.6.2) and the Markov inequality

$$P\big(F \ge \epsilon\big) = \mathcal{O}\left(\epsilon^{-p} n^{-p/2} d_n^p + \frac{(d_n \log n)^p}{n^{p/2}}\right).$$

Since

$$\max\big|\widehat{\mathbf{\Gamma}}_{d_n}^{-1} - \mathbf{\Gamma}_{d_n}^{-1}\big| \le \big\|\widehat{\mathbf{\Gamma}}_{d_n}^{-1} - \mathbf{\Gamma}_{d_n}^{-1}\big\|,$$

the claim follows. □

We now introduce the estimator $\widetilde{\boldsymbol{\Theta}} = \big(\widetilde{\theta}_1, ..., \widetilde{\theta}_d\big)^T$. Recall that we can write

$$\mathbf{Y} = X\mathbf{\Phi_d} + \mathbf{Z}, \tag{5.6.3}$$

where $\mathbf{Y} = \big(X_1, ..., X_n\big)^T$, $\mathbf{Z} = \big(\epsilon_1, ..., \epsilon_n\big)^T$. Define the $n \times d_n$ design matrix $\mathbf{X}$ as

$$\mathbf{X} = \begin{pmatrix} X_0 & X_{-1} & ... & X_{1-d_n} \\ X_1 & X_0 & ... & X_{2-d_n} \\ ... & ... & & \\ X_{n-1} & X_{n-2} & ... & X_{n-d_n} \end{pmatrix},$$

and the estimator $\widetilde{\boldsymbol{\Theta}}$ as

$$\widetilde{\boldsymbol{\Theta}} = \big(\mathbf{X}^T\mathbf{X}\big)^{-1}\mathbf{X}^T\mathbf{Y}. \tag{5.6.4}$$

Proposition 5.6.4. *Let $\big\{X_k\big\}_{k\in\mathbb{Z}}$ be an AR(d_n) process, such that Assumption 5.2.1 is satisfied. Then for $\epsilon > 0$*

$$P\big(\max\big|\sqrt{n}\big(\widehat{\boldsymbol{\Theta}} - \widetilde{\boldsymbol{\Theta}}\big)\big| \geq \epsilon\big) = \mathcal{O}\left(\epsilon^{-p}n^{-p/2}d_n^{2p} + \frac{(d_n \log n)^p}{n^{p/2}} + n^{-1}\right),$$

where d_n is as in Theorem 6.6.4.

Proof. Following the proof of [25, Theorem 8.10.1], we have the following decomposition.

$$\sqrt{n}\big(\widehat{\boldsymbol{\Theta}} - \widetilde{\boldsymbol{\Theta}}\big) = \sqrt{n}\widehat{\mathbf{\Gamma}}_{\mathbf{d_n}}^{-1}\big(\widehat{\mathbf{\Phi}}_{\mathbf{d_n}} - n^{-1}\mathbf{X^TY}\big) + n^{1/2}\big(\widehat{\mathbf{\Gamma}}_{\mathbf{d_n}}^{-1} - n(\mathbf{X^TX})^{-1}\big)n^{-1}\mathbf{X^TY}.$$

For the i-th component of $\sqrt{n}\big(\widehat{\mathbf{\Phi}}_{\mathbf{d_n}} - n^{-1}\mathbf{X^TY}\big)$ we have

$$n^{-1/2}\sum_{k=1-i}^{0} X_k X_{k+i} + \sqrt{n}\overline{X}_n\big((1 - n^{-1}i)\overline{X}_n - n^{-1}\sum_{k=1}^{n-i}(X_k + X_{k+i})\big).$$

Using the Minikowski and the Cauchy-Schwarz in equality we get

$$\begin{aligned}
&\big\|n^{-1/2}\sum_{k=1-i}^{0} X_k X_{k+i} + \sqrt{n}\overline{X}_n\big((1 - n^{-1}i)\overline{X}_n - n^{-1}\sum_{k=1}^{n-i}(X_k + X_{k+i})\big)\big\|_{p/2} \\
&\leq \sqrt{\frac{|1-i|}{n}}\big\||1-i|^{-1/2}\sum_{k=1-i}^{0}(X_k X_{k+i} - \phi_i)\big\|_{p/2} + n^{-1/2}\sum_{k=1-i}^{0}|\phi_i| \\
&+ \big\|\sqrt{n}X_n\big\|_p\left(\big\|\overline{X}_n\big\|_p + n^{-1/2}\big\|n^{-1/2}\sum_{k=1}^{n-i}(X_k + X_{k+i})\big\|_p\right) \\
&:= A_n.
\end{aligned}$$

Since $0 \leq i \leq d_n$, we obtain from Lemma 5.6.1 that $A_n = \mathcal{O}\left(n^{-1/2} d_n^{1/2}\right)$, and hence by the Markov inequality

$$P\big(\max\big|\sqrt{n}\big(\widehat{\mathbf{\Phi}}_{\mathbf{d_n}} - n^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}\big)\big| \geq \epsilon\big) = \mathcal{O}\left(\epsilon^{-p} n^{-p/2} d_n^{p/2}\right). \tag{5.6.5}$$

From the proof of Corollary 5.5.3, we have that $\sum_{j=1}^{d_n} |\gamma_{i,j}^*| \leq 2\vartheta^2$. Then we obtain from Proposition 5.4.1

$$\begin{aligned} P\big(\max\big|\sqrt{n}\widehat{\mathbf{\Gamma}}_{\mathbf{d_n}}^{-1}\big(\widehat{\mathbf{\Phi}}_{\mathbf{d_n}} - n^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}\big)\big| \geq \epsilon\big) &\leq P\big(\max\big|\widehat{\mathbf{\Gamma}}_{\mathbf{d_n}}^{-1} - \mathbf{\Gamma}_{\mathbf{d_n}}{}^{-1}\big| \geq d_n^{-1}\big) \\ &\quad + P\big(\max\big|\sqrt{n}\big(\widehat{\mathbf{\Phi}}_{\mathbf{d_n}} - n^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}\big)\big| \geq 2\epsilon\big) \\ &= \mathcal{O}\left(\epsilon^{-p} n^{-p/2} d_n^p + \frac{(d_n \log n)^p}{n^{p/2}}\right). \end{aligned} \tag{5.6.6}$$

We will now treat the second part. It holds that

$$\sqrt{n}\big(\widehat{\mathbf{\Gamma}}_{\mathbf{d_n}}^{-1} - n(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\big) = \widehat{\mathbf{\Gamma}}_{\mathbf{d_n}}^{-1}\sqrt{n}\big(n^{-1}(\mathbf{X}^{\mathbf{T}}\mathbf{X}) - \widehat{\mathbf{\Gamma}}_{\mathbf{d_n}}\big)n(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}.$$

Note that by (5.6.5), we obtain from the Markov inequality that

$$P\big(\max\big|n^{-1/2}\mathbf{X}^{\mathbf{T}}\mathbf{X} - n^{1/2}\widehat{\mathbf{\Gamma}}_{\mathbf{d_n}}\big| \geq \epsilon\big) = \mathcal{O}\left(\epsilon^{-p} n^{-p/2} d_n^{p/2}\right). \tag{5.6.7}$$

One readily verifies that Proposition 5.4.1 is also valid if one replaces $\widehat{\mathbf{\Gamma}}_{\mathbf{d_n}}^{-1}$ with $n(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}$, in fact, it is evident from the definition that $n(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}$ is actually the better estimator, hence it holds that for $\epsilon > 0$

$$P\big(\max\big|n(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1} - \mathbf{\Gamma}_{\mathbf{d_n}}{}^{-1}\big| \geq \epsilon\big) = \mathcal{O}\left(\epsilon^{-p} n^{-p/2} d_n^p + \frac{(d_n \log n)^p}{n^{p/2}}\right). \tag{5.6.8}$$

Recalling that $\sum_{j=1}^{d_n} |\gamma_{i,j}^*| \leq 2\vartheta^2$, we obtain from the inequality given in (5.4.3) that by appropriately adding and subtracting $\mathbf{\Gamma}_{d_n}^{-1}$ and using Proposition 5.4.1, (5.6.7) and (5.6.8), that

$$P\big(\max\big|\sqrt{n}\big(\widehat{\mathbf{\Gamma}}_{\mathbf{d_n}}^{-1} - n(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\big)\big| \geq \epsilon\big) = \mathcal{O}\left(\epsilon^{-p} n^{-p/2} d_n^p + \frac{(d_n \log n)^p}{n^{p/2}}\right). \tag{5.6.9}$$

Moreover, it holds that

$$n^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y} = \big(n^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{X} - \mathbf{\Gamma}_{d_n}\big)\mathbf{\Phi}_{d_n} + \mathbf{\Gamma}_{d_n}\mathbf{\Phi}_{d_n} + n^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Z}.$$

By Lemma 5.6.1 and Assumption 5.2.1 we have that $\left|\mathbf{\Gamma}_{d_n}\mathbf{\Phi}_{d_n}\right| \leq C$, where C does not depend on n. In addition, we deduce from Theorem 6.6.4 that on a possible larger probability space, there exists a sequence $\{\xi_{n,h}\}_{0\le h\le d_n}$ of mean zero Gaussian random variables with strictly positive variance such that

$$P\left(\max_{0\le h\le d_n}\left|\sqrt{n}(\widehat{\phi}_{n,h}-\phi_h)-\xi_{n,h}\right|>1\right)=\mathcal{O}\left(n^{-1}\right). \tag{5.6.10}$$

By known properties of the Gaussian distribution function, we obtain that

$$P\left(n^{-1/2}\max_{0\le h\le d_n}\left|\xi_{n,h}\right|\ge 1\right)\le\sum_{h=0}^{d_n}P\left(\left|\xi_{n,h}\right|\ge n^{1/2}\right)=\mathcal{O}\left(d_ne^{-n/2}\right)=\mathcal{O}\left(n^{-1}\right). \tag{5.6.11}$$

From the proof of Proposition 5.4.1, an application of the Markov inequality yields

$$P\big(\max\left|n^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{X}-\mathbf{\Gamma}_{d_n}\right|\ge 1\big)=\mathcal{O}\left(n^{-p/2}d_n^p\right),$$

hence for large enough $C>0$ we obtain

$$P\big(\max\left|n^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}\right|\ge C\big)=\mathcal{O}\left(n^{-p/2}d_n^p\right), \tag{5.6.12}$$

which gives us

$$\begin{aligned}
&P\big(\max\left|n^{-1/2}\big(\widehat{\mathbf{\Gamma}}_{\mathbf{d_n}}^{-1}-n(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\big)n^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}\right|\ge\epsilon\big)\le P\big(\max\left|n^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}\right|\ge C\big)\\
&+P\big(\max\left|\sqrt{n}\big(\widehat{\mathbf{\Gamma}}_{\mathbf{d_n}}^{-1}-n(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\big)\right|\ge Cd_n^{-1}\epsilon\big)\\
&=\mathcal{O}\left(\epsilon^{-p}n^{-p/2}d_n^{2p}+\frac{(d_n\log n)^p}{n^{p/2}}+n^{-p/2}d_n^p\right).
\end{aligned}$$

Thus, we finally obtain from the previous calculations and (5.6.6) that

$$P\big(\max\left|\sqrt{n}(\widehat{\mathbf{\Theta}}-\widetilde{\mathbf{\Theta}})\right|\ge\epsilon\big)=\mathcal{O}\left(\epsilon^{-p}n^{-p/2}d_n^{2p}+\frac{(d_n\log n)^p}{n^{p/2}}+n^{-1}\right). \tag{5.6.13}$$

□

We are now in the position to proof Theorem 5.4.2.

Proof of Theorem 5.4.2. We have that

$$\begin{aligned}
&P\big(\max\left|n^{1/2}(\widehat{\mathbf{\Theta}}-\mathbf{\Theta})-n^{-1/2}\mathbf{\Gamma}^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Z}\right|\ge 2\epsilon\big)\\
&\le P\big(\max\left|n^{1/2}(\widehat{\mathbf{\Theta}}-\widetilde{\mathbf{\Theta}})\right|\ge\epsilon\big)+P\big(\max\left|n^{1/2}(\widetilde{\mathbf{\Theta}}-\mathbf{\Theta})-n^{-1/2}\mathbf{\Gamma}^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Z}\right|\ge\epsilon\big).
\end{aligned}$$

From Proposition 5.6.4, we have

$$P\big(\max\big|n^{1/2}(\widehat{\boldsymbol{\Theta}} - \widetilde{\boldsymbol{\Theta}})\big| \geq \epsilon\big) = \mathcal{O}\left(\epsilon^{-p} n^{-p/2} d_n^{2p} + \frac{(d_n \log n)^p}{n^{p/2}}\right). \tag{5.6.14}$$

The proof of Proposition 5.6.4 gives us

$$n^{1/2}(\widetilde{\boldsymbol{\Theta}} - \boldsymbol{\Theta}) - n^{-1/2}\boldsymbol{\Gamma}^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Z} = \big(n(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1} - \boldsymbol{\Gamma}^{-1}\big)\, n^{-1/2}\mathbf{X}^{\mathbf{T}}\mathbf{Z}, \tag{5.6.15}$$

and

$$P\big(\max\big|n(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1} - \boldsymbol{\Gamma}_{\mathbf{d_n}}{}^{-1}\big| \geq \epsilon\big) = \mathcal{O}\left(\epsilon^{-p} n^{-p/2} d_n^{p} + \frac{(d_n \log n)^p}{n^{p/2}}\right). \tag{5.6.16}$$

Due to Theorem 6.6.4, on a possible larger probability space, there exists a d_n-dimensional Gaussian vector $\boldsymbol{\xi}_{d_n} = \big(\xi_{n,1}, ..., \xi_{n,d_n}\big)^T$ with strictly positive variance such that

$$P\left(\max\big|n^{-1/2}\mathbf{X}^{\mathbf{T}}\mathbf{Z} - \boldsymbol{\xi}_{d_n}\big| \geq 1\right) = \mathcal{O}\left(n^{-1}\right). \tag{5.6.17}$$

Proceeding as in the proof of Proposition 5.6.4 we have

$$P\left(n^{-1/2} \max_{0 \leq h \leq d_n} \big|\xi_{n,h}\big| \geq \log n\right) = \mathcal{O}\left(n^{-1}\right),$$

hence

$$P\big(\max\big|n^{-1/2}\mathbf{X}^{\mathbf{T}}\mathbf{Z}\big| \geq \log n\big) = \mathcal{O}\left(n^{-1}\right), \tag{5.6.18}$$

and we conclude

$$\begin{aligned} &P\big(\max\big|\big(n(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1} - \boldsymbol{\Gamma}^{-1}\big)n^{-1/2}\mathbf{X}^{\mathbf{T}}\mathbf{Z}\big| \geq \epsilon d_n \log n\big) \\ &\leq P\big(\max\big|n^{-1/2}\mathbf{X}^{\mathbf{T}}\mathbf{Z}\big| \geq \log n\big) + P\big(\max\big|n(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1} - \boldsymbol{\Gamma}_{\mathbf{d_n}}{}^{-1}\big| \geq \epsilon\big) \\ &= \mathcal{O}\left(\epsilon^{-p} n^{-p/2} d_n^{p} + \frac{(d_n \log n)^p}{n^{p/2}} + n^{-1}\right), \end{aligned}$$

yielding

$$P\big(\max\big|n^{1/2}(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}) - n^{-1/2}\boldsymbol{\Gamma}^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Z}\big| \geq \epsilon\big) = \mathcal{O}\left(\epsilon^{-p} n^{-p/2} d_n^{2p} (\log n)^p + n^{-1}\right).$$

□

5.7 Additional results

The following result is due to C.Deo [38, Theorem 1] and describes the asymptotic behavior of the absolute value of a Gaussian, possible nonstationary sequence $\{\xi_i\}_{i\in\mathbb{N}}$.

Lemma 5.7.1. *Let $\{\xi_i\}_{i\in\mathbb{N}}$ be Gaussian process, where $\mathbb{E}(\xi_i) = 0$, $\mathbb{E}(\xi_i^2) = 1$ for all $i \in \mathbb{N}$. Put $\phi_{i,j} = Cov(\xi_i, \xi_j)$, $r_n := \sup_{|i-j|\geq n} |\phi_{i,j}|$, and let $r_1 < 1$. Assume that one of the following two conditions is satisfied:*

(a) $\sum_{n=1}^{\infty} r_n^2 < \infty$,

(b) For some $\beta > 0$, $r_n(\log n)^{2+\beta} \to 0$.

Then it holds

$$P\left(a_n^{-1}\big(\max_{1\leq h\leq n}|\xi_h| - b_n\big) \leq z\right) \to \exp(-e^{-z}),$$

where $a_n = (2\log n)^{-1/2}$ *and* $b_n = (2\log n)^{1/2} - (8\log n)^{-1/2}(\log\log n + 4\pi - 4)$.

Chapter 6

Change-Point Analysis with increasing Dimension - Global and Local Changes

6.1 Introduction

Let $X_1, X_2, ..., X_n$ denote some collected observations. Structural stability is a very important topic in statistics and econometrics, excellent surveys can be found in Banerjee and Urga [11] and Perron [103], for deeper mathematical insights we refer to Csörgő and Horváth [29, 32]. Many authors studied testing for the stability of the mean $\mu_i = \mathbb{E}(X_i)$, $1 \leq i \leq n$ in case of independent and dependent observations, whereas others considered tests for a change in variance or some other parameters, see for instance [6, 7, 9, 14, 54, 65, 69] and the references therein. A very popular method to detect possible changes are so called CUSUM statistics, which are based on the CUSUM process defined by

$$S_n(t) \equiv \begin{cases} n^{-1/2} \sum_{i=1}^{[(n+1)t]} \big(X_i - \overline{X}_n\big), & \text{if } 0 \leq t < 1, \\ 0, & \text{if } t = 1, \end{cases} \tag{6.1.1}$$

where $\overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. Usually, the point where the statistic reaches its maximum is considered as the change point, if the test statistic exceeds a certain critical value. Naturally, these quantiles arise from the asymptotic distribution of the CUSUM process $S_n(t)$. If a functional limit theorem holds for the process $M_n(t) = n^{-1/2} \sum_{i=1}^{[nt]} X_i$, i.e.

$$M_n(t) \xrightarrow{\mathbb{D}[0,1]} \sigma W_t, \ 0 \leq t \leq 1,$$

where W_t is a Brownian motion and $\mathbb{D}[0,1]$ stands for the space of c?dl?g functions on $[0,1]$, then it follows for instance that

$$\sup_{0\le t\le 1}\sigma^{-1}\big|S_n(t)\big| \xrightarrow{w} \sup_{0\le t\le 1}\big|B_t\big|, \tag{6.1.2}$$

where B_t denotes a Brownian Bridge, and $\xrightarrow{w}$ stands for weak convergence. It is well established in the literature (cf. [29, 78]) that a weight function $w(t)$ will increase the power of testing procedures against certain alternatives. In particular, the weight function $\big(t(1-t)\big)^{-1/2}$ has received considerable attention, we refer to [29, 32] and the references there for more details on the subject. A theoretical drawback of weight functions is that usually, a functional limit theorem is no longer sufficient to determine the asymptotic distribution, more refined methods need to be used, such as strong or almost sure approximations, often also called strong invariance principles, for details see [16, 74, 76, 109, 130, 135] and the references there. Based on these methods, under appropriate assumptions, one obtains that

$$\sup_{0\le t\le 1}\sigma^{-1}\frac{\big|M_n(t)\big|}{w(t)} \xrightarrow{w} \sup_{0\le t\le 1}\frac{\big|B_t\big|}{w(t)}. \tag{6.1.3}$$

When testing for a change point, it is usually assumed that the basic model is known, and a certain parameter space is fixed. However, when first fitting a model, these facts are not known a priori, nor is the dimension of a possible parameter space. For example, suppose that, given a mean zero time series $\big\{X_k\big\}_{k\in\mathbb{N}}$, we would like to fit an AR(p) model. Various procedures are known (cf. [4, 25, 60]), and, in some way or other, they all depend on the covariances $\phi_h = \mathbb{E}\big(X_h X_0\big)$, hence it is crucial to know whether or not the covariance structure can assumed to be stable, and, if not, where a change point is located. As a rule of thumb, one may say that fitting an AR(p) model requires estimating the first $p+1$ covariances $\phi_0, ..., \phi_p$. Unfortunately, we do not know anything yet about the possible order p. The situation is quite similar when considering MA(q) processes, or, more generally speaking, ARMA(p,q) processes, or linear or nonlinear regression.

Summarizing, we see that we would require a procedure that, in some sense, is invariant with respect to the number of parameters of the underlying model. If we stick to the example of the covariances, a possible approach is to proceed as in Aue et al. [7] or Lee et al. [81]. Given a meanzero, stationary sequence $\big\{X_k\big\}_{k\in\mathbb{N}}$, they considered the usual covariance estimators $\widehat{\phi}_{h,n} = \frac{1}{n}\sum_{k=1}^{n-h} X_k X_{k+h}$, and defined the CUSUM statistic as

$$\Lambda_n(t) = \left(\frac{[nt]}{\sqrt{n}}\big(\widehat{\phi}_{0,[nt]} - \widehat{\phi}_{0,n}\big), ..., \frac{[nt]}{\sqrt{n}}\big(\widehat{\phi}_{p,[nt]} - \widehat{\phi}_{p,n}\big)\right)^T, \quad 0\le t\le 1.$$

It was then established that the ellipsoid $\Lambda_n(t)^T\widehat{\Gamma}_n^{-1}\Lambda_n(t)$ satisfies

$$\Lambda_n(t)^T\widehat{\Gamma}_n^{-1}\Lambda_n(t) \xrightarrow{\mathbb{D}[0,1]} \sum_{i=0}^{p}\big(W_{t,i}\big)^2, \quad \text{for fixed } p \tag{6.1.4}$$

where $\{W_{t,i}\}_{t\geq 0}$ is a sequence of independent Brownian motions, and $\widehat{\Gamma}_n$ is a consistent estimator of the covariance. According to Lee et al; one may then use

$$\Lambda_n = \max_{p\leq k\leq n} \Lambda_n(k/n)^T\widehat{\Gamma}_n^{-1}\Lambda_n(k/n)$$

as test statistic, whose quantiles are determined via (6.1.4). This approach is reasonable, in particular, since a change in a single parameter will often result in a *global change*, i.e. most, if not all of the covariances get altered, and summing up all the errors measures the total discrepancy. In view of the forgoing remarks, it is therefore desirable to extend this result to the case where one allows $p = p_n$ to increase in n. In addition, using a weighted modification $\Lambda_n(t)^\nu = (t(1-t)^{1/\nu}\Lambda_n(t)$, $1/2 \leq \nu$ will increase the power (cf. [32]). It is, however, possible that only a few or even a single covariance changes, see for instance Example 6.3.8. In this case, we only have a local change, and measuring this with $\Lambda_n(t)$ may not be the best thing to do. Similarly, suppose that we have a d-dimensional time series $\{X_{k,h}\}_{k\in\mathbb{Z},\,1\leq h\leq d}$, where d is fairly large, and we are interested in testing simultaneously for the stability of the mean $\mu_h = \mathbb{E}\big(X_{k,h}\big)$ via the estimators $\widehat{\mu}_{h,n} = \frac{1}{n}\sum_{k=1}^{n-h} X_{k,h}$. If we consider the previous approach, based on the ellipsoids in (6.1.4), then as before we will detect a global change if most of the μ_h change. However, if only a single or a few components have undergone a change in mean, we may not detect it since the ellipsoids in (6.1.4) measure the discrepancy of the whole set. Hence, in order to detect local changes, we propose to consider the (possible weighted) maximum, i.e

$$\Upsilon_{n,d} = \max_{0\leq h\leq d}\left(\widehat{\psi}_h^{-1}\sup_{l\leq t\leq 1-l} w(t)^{-1}\left|\frac{[nt]}{\sqrt{n}}\big(\widehat{\mu}_{h,[nt]} - \widehat{\mu}_{h,n}\big)\right|\right), \tag{6.1.5}$$

for some $0 < l < 1/2$, where $w(t)$ is some weight function and $\widehat{\psi}_h^2$ is a variance estimator. One can expect that for some increasing sequence d_n, and suitably normalized with sequences a_n, b_n, it holds that

$$a_n^{-1}\big(\Upsilon_{n,d_n} - b_n\big) \xrightarrow{w} G,$$

where G denotes an extreme value distribution. The following two question arise.

(i) What growth rate for d_n is possible?

(ii) What can we say about the magnitude of a_n, b_n?

Answering these question is important, and indeed we will show that (see Section 5.2)

(i) d_n may be chosen as $d_n = \mathcal{O}\left(n^{\delta}\right)$ for some $\delta > 0$,

(ii) for any fixed $x \geq 0$, we have $(a_n x + b_n)^2 = \mathcal{O}\left(\log d_n\right) = \mathcal{O}\left(\log n\right)$.

The statistic $\Upsilon_{n,d}$ now has the property that it is practical invariant of $d = d_n$, if d_n is sufficiently large. Moreover, we are no longer required to estimate the complete covariance matrix Γ_n, we only need the elements of the diagonal.

The aim of this chapter is twofold. On one hand, we will generalize the approach based on the ellipsoids (6.1.4) to a weighted version which allows for an increase in dimension with an explicit growth rate. On the other hand, we will provide answers to the questions (i) and (ii) in a general setting, which includes, among others, the statistic $\Upsilon_{n,d}$. The chapter is structured as follows. In Section 6.2 the main results are presented, alongside some comments and remarks. Based on two general key results, the proofs are presented in Section 6.5. In Sections 6.6 and 6.7, the above mentioned key results are shown, which may have interest in themselves.

6.2 Main results

Let $\left\{X_{k,h}\right\}_{k,h\geq 1}$ be a collection of random variables such that for each h_0, $\left\{X_{k,h_0}\right\}_{k\geq 1}$ is a zero mean stationary sequence. Given a sequence $\left\{\epsilon_k\right\}_{k\in\mathbb{Z}}$ of independent and identically distributed random variables, we define the following two σ-algebras.

$$\mathcal{F}_k = \sigma\big(\epsilon_j,\, j \leq k\big), \quad \mathcal{F}_{k-m}^{k+m} = \sigma\big(\epsilon_j,\, k-m \leq j \leq k+m\big). \tag{6.2.1}$$

We will always assume that $\left\{X_{k,h}\right\}_{k,h\geq 1}$ is adapted to $\mathcal{F}_{k+h}$, more specifically, we assume that $X_{k,h}$ is $\mathcal{F}_{k+h}$ measurable for each $k, h \geq 1$. Hence we implicitly assume that $X_{k,h}$ can be written as as function $g_{h,k} = g_h(\epsilon_{k+h}, \epsilon_{k+h-1}, ...)$. For convenience, we will write $g_h(\xi_{k+h})$, with $\xi_k = (\epsilon_k, \epsilon_{k-1}, ...)$. The class of processes that fits into this framework is large, and contains a variety of linear and nonlinear processes including ARCH, GARCH and related processes, see for instance [52, 104, 120, 121]. A very nice feature of the representation given above is that it allows to give simple, yet very efficient and general dependence conditions. Following Wu [133], let $\left\{\epsilon_k'\right\}_{k\in\mathbb{Z}}$ be an independent copy of $\left\{\epsilon_k\right\}_{k\in\mathbb{Z}}$ on the same probability space, and define the 'filters' $\xi_{k,}^{(m,')}$, $\xi_{k,h}^{(m,*)}$ as $\xi_{k,h}^{(m,')} = (\epsilon_{k+h}, \epsilon_{k+h-1}, ..., \epsilon_{k-m}', \epsilon_{k-m-1}, ...)$ and $\xi_{k,h}^{(m,*)} = (\epsilon_{k+h}, \epsilon_{k+h-1}, ..., \epsilon_{k-m}, \epsilon_{k-m-1}', ...)$.

We put $\xi'_{k,h} = \xi^{(0,')}_{k,h} = (\epsilon_{k+h}, \epsilon_{k+h-1}, ..., \epsilon'_0, \epsilon_{-1}, ...)$ and $\xi^*_{k,h} = \xi^{(0,*)}_{k,h} = (\epsilon_{k+h}, \epsilon_{k+h-1}, ..., \epsilon_0, \epsilon'_{-1}, ...)$. In analogy, we put $X^{(m,')}_{k,h} = g_h\big(\xi^{(m,')}_{k,h}\big)$ and $X^{(m,*)}_{k,h} = g_h\big(\xi^{(m,*)}_{k,h}\big)$, in particular we have $X'_{k,h} = X^{(0,')}_{k,h}$ and $X^*_{k,h} = X^{(0,*)}_{k,h}$.

As a dependence measure, one may now consider the quantities $\big\|X_{k,h} - X'_{k,h}\big\|_p$ or $\big\|X_{k,h} - X^*_{k,h}\big\|_p$, $p \geq 1$, where $\|.\|_p^p = \mathbb{E}(|.|^p)$. For example, if we define the linear processes $X_{k,h} = \sum_{i=0}^{\infty} \alpha_{i,h}\epsilon_{k-i}$, the condition

$$\sum_{k=0}^{\infty} \big\|X_{k,h} - X'_{k,h}\big\|_2 < \infty \tag{6.2.2}$$

is valid if $\sum_{i=0}^{\infty} |\alpha_{i,h}| < \infty$, provided that $\mathbb{E}\big(\epsilon_0^2\big) < \infty$. Dependence conditions of the type of (6.2.2) are often quite general and easy to verify in many cases, see for instance [15, 34, 42, 130] and the references there.

Another feature of the above representation is that it allows to quantify approximations with m-dependent variables. To this end, let

$$Y^{(\leq m)}_{k,h} = \mathbb{E}\big(X_{k,h} \mid \mathcal{F}^{k+m}_{k-m}\big), \quad Y^{(>m)}_{k,h} = X_{k,h} - Y^{(\leq m)}_{k,h} = X_{k,h} - \mathbb{E}\big(X_{k,h} \mid \mathcal{F}^{k+m}_{k-m}\big). \tag{6.2.3}$$

Then one can show (cf. Proposition 6.2.4), that

$$\big\|Y^{(>m)}_{k,h}\big\|_p \leq C \sum_{i=0}^{\infty} \big\|X_{m_n+h+i,h} - X'_{m_n+i,h}\big\|_p^2.$$

Remark 6.2.1. Note that we have not defined a sample space for the sequence $\big\{\epsilon_k\big\}_{k\in\mathbb{Z}}$. One may both consider cases where $\big\{\epsilon_k\big\}_{k\in\mathbb{Z}} \in \mathbb{R}$ or $\mathbb{R}^{\infty}$.

We also introduce the following notation. Put

$$S^{(n,l)}_h = \sum_{k=1}^{l} X_{k,h}, \quad M^{(n)}_{t,h} = n^{-1/2}\left(\sum_{k=1}^{[nt]} X_{k,h} - t\sum_{k=1}^{n} X_{k,h}\right), \tag{6.2.4}$$

and, for $0 < l < 1$, the weighted version of $M^{(n)}_{t,h}$

$$Z^{(n,l)}_h = n^{-1/2} \sup_{l\leq t\leq 1-l} \frac{\big|\sum_{k=1}^{[nt]} X_{k,h} - t\sum_{k=1}^{n} X_{k,h}\big|}{\sqrt{t(1-t)}}. \tag{6.2.5}$$

We use the abbreviation $S^{(n)}_h = S^{(n,n)}_h$, and we denote the corresponding random vectors with $\mathbf{S}^{(n)} = \big(S^{(n)}_1, S^{(n)}_2, ..., S^{(n)}_{d_n}\big)^t$, $\mathbf{M}^{(n)}_t = \big(M^{(n)}_{t,1}, M^{(n)}_{t,2}, ..., M^{(n)}_{t,d_n}\big)^t$, and

$\mathbf{Z}_l^{(n)} = \big(Z_1^{(n,l)}, Z_2^{(n,l)}, ..., Z_{d_n}^{(n,l)}\big)^t$. Put $\phi_{i,j} = \mathbb{E}\big(X_{k,i}, X_{k,j}\big)$, $k \geq 1$, where we point out that this is well defined since we can write $X_{k,h} = g_h(\xi_{k+h})$, and thus $\phi_{i,j}$ does not depend on k. We formally define the variance as

$$\psi_h = \lim_n n^{-1}\mathbb{E}\big(S_h^{(n)} S_h^{(n)}\big), \tag{6.2.6}$$

and for $1 \leq i \leq j \leq d_n$ the sample correlation

$$\rho_{i,j}^{(n)} = \mathbb{E}\big(S_i^{(n)} S_j^{(n)}\big) \Big(\mathrm{Var}\big(S_i^{(n)}\big)\mathrm{Var}\big(S_j^{(n)}\big)\Big)^{-1/2}. \tag{6.2.7}$$

We will also frequently use the following notation. For a matrix $\mathbf{A} = \big(a_{i,j}\big)_{\left\{\substack{1\leq i\leq r,\\ 1\leq j\leq s}\right\}}$, $r, s \in \mathbb{N}$ we put

$$\max\big|\mathbf{A}\big| = \max_{1\leq i\leq r, 1\leq j\leq s}\big|a_{i,j}\big|. \tag{6.2.8}$$

To state our main results, we need some dependence assumptions that are given below.

Assumption 6.2.2. *For $m = m_n = \mathcal{O}\left(n^\theta\right)$, $0 < \theta < 1$, $d = d_n = \mathcal{O}\left(n^\delta\right)$, $\delta > 0$ we suppose that*

$$(i)\ \sup_h \|X_{1,h}\|_p < \infty, \quad \textit{for some}\, p > 8, \quad \mathbb{E}\big(X_{1,h}\big) = 0, \ \textit{for all}\ 1 \leq h \leq d_n,$$

$$(ii)\ \sup_h \sum_{j=0}^{\infty} j|\phi_{j,h}| < \infty,$$

$$(iii)\ \limsup_{n\to\infty} \max_{1\leq h\leq d_n} \max_{1\leq l\leq n} \left\|\sum_{j=1}^{l} Y_{j,h}^{(>m_n)}\right\|_p = \mathcal{O}\left(1\right), \ p > 8,$$

$$(iv)\ \limsup_{n\to\infty} \sup_{i,j:1\leq|i-j|} \big|\rho_{i,j}^{(n)}\big| < 1,$$

$$(v)\ \limsup_{n\to\infty} (\log n)^2 \left(\sum_{r=\sqrt{\log n}}^{d_n} \sup_{r\leq|i-j|} \big|\rho_{i,j}^{(n)}\big|\right) = 0.$$

Remark 6.2.3. Note that Assumption 6.2.2 (ii) implies that

$$\psi_h^2 = \lim_n n^{-1}\mathrm{Var}\big(\sum_{1\leq k\leq n} X_{k,h}\big) < \infty, \tag{6.2.9}$$

and in particular $\sup_h \psi_h^2 < \infty$.

The conditions in Assumption 6.2.2 can be divided into the classes C1:(i)-(iii) and C2:(iv),(v). Conditions C1 are necessary to construct appropriate approximating Gaussian processes for $S_h^{(n)}$ and $M_{t,h}^{(n)}$, whereas conditions C2 are required to establish weak convergence to an extreme value distribution G, and reflect the well-known conditions for Gaussian sequences, see for instance [17], [38] and [79]. Note that condition (v) may be considerably weakened (cf. Theorem 6.7.5), this, however, would lead to a much less tractable condition, that is hard to verify directly. In view of the previous discussion on dependence measures, it is desirable to provide easy conditions for (ii) and (iii) in terms of $\|X_{k,h} - X'_{k,h}\|_p$, $k, h \geq 1$.

Proposition 6.2.4. *Suppose that* $\max_{1\leq h\leq d_n}\|X_{k,h} - X'_{k,h}\|_p = \mathcal{O}\left(k^{-\beta}\right)$, *with* $\theta \geq \frac{2}{2\beta-1}$, $p > 8$, *where* $\beta > 5/3$. *Then Assumption 6.2.2 (ii), (iii) are valid.*

As noted by Aue et al. [8], one may weaken the moment assumptions by strengthening the dependence conditions. This is accomplished by considering the transformation $U_{k,h} = |X_{k,h}|^{\delta}$. Indeed one then obtains for $\rho \in (0,1]$ that $\|U_{k,h} - U'_{k,h}\|_p \leq \|X_{k,h} - X'_{k,h}\|^{\rho}_{p\rho}$, and consequently

$$\sum_{k=1}^{\infty}\|U_{k,h} - U'_{k,h}\|_p \leq \sum_{k=1}^{\infty}\|X_{k,h} - X'_{k,h}\|^{\rho}_{p\rho}.$$

6.3 Global changes

In this section, we present approximation results that can be used to detect global changes. To this end, we introduce the random vector

$$\mathbf{M}_t^{(n)} = \left(M_{t,1}^{(n)}, M_{t,2}^{(n)}, ..., M_{t,d_n}^{(n)}\right)^t,$$

where $B_{t,h} = W_{t,h} - tW_{1,h}$, $1 \leq h \leq d_n$ is a sequence of Brownian Bridges, where the Brownian motions $\left\{W_{t,h}\right\}_{0\leq t\leq 1,\ 1\leq h\leq d_n}$ have the covariance matrix $\mathbf{\Gamma}_{\mathbf{W}^{(n)}}$. Denote with $\mathbf{\Gamma}_{\mathbf{S}^{(n)}}$ the covariance matrix of the vector $n^{-1}\mathbf{S}^{(n)}$, defined in (6.2.4).

Theorem 6.3.1. *Suppose that Assumption 6.2.2 (i) - (iii) holds, and let* $\mathbf{\Gamma}_n$ *be a sequence of regular matrices such that* $\lim_{L\to\infty}\lim_{n\to\infty}P\big(\max|\mathbf{\Gamma}_n^{-1}| \geq L\big) = 0$. *Then on a possible larger probability space, we have that*

$$\left|\sup_{\lambda_n/n\leq t\leq 1-\lambda_n/n}\left|w(t)^{-1}\big(\mathbf{M}_t^{(n)}\big)^t\mathbf{\Gamma}_n^{-1}\mathbf{M}_t^{(n)}\right| - \sup_{\lambda_n/n\leq t\leq 1-\lambda_n/n}\left|w(t)^{-1}\mathbf{B}_t^t\mathbf{\Gamma}_n^{-1}\mathbf{B}_t\right|\right| = o_p(1).$$

The dimension $d_n = \mathcal{O}\big(n^{\delta}\big)$ *must satisfy the relation*

$$\theta \leq \delta < \min\left\{\frac{4(p-2\nu)}{(-4+3p)\nu}, \frac{2+p-2(1+\theta)\nu}{(2+4\theta+p(4+\theta))\nu}\right\}, \tag{6.3.1}$$

where we require $p > 4\nu$ and $d_n^2 \lambda_n^{-1/2+1/\nu} = \mathcal{O}\left(n^{-\kappa}\right)$, for some $\kappa > 0$. Moreover, it holds that $\max\left|\mathbf{\Gamma}_{\mathbf{S}^{(n)}} - \mathbf{\Gamma}_{\mathbf{W}^{(n)}}\right| = \mathcal{O}\big(n^{-\gamma}\big)$, for some $\gamma > 0$. Alternatively, if one sets $d_n = \mathcal{O}\big((\log n)^{\delta}\big)$, for arbitrary $\delta > 0$, then we require

$$\nu < \min\left\{\frac{4+2p}{4+4\theta+p\theta+4\theta^2+p\theta^2}, \frac{2}{1-2\theta}, p/4\right\}, \tag{6.3.2}$$

and $d_n^2 \lambda_n^{-1/2+1/\nu} = \mathcal{O}\left((\log n)^{-\kappa}\right)$, for some $\kappa > 0$.

Remark 6.3.2. Conditions (i) and (iii) of Assumption 6.2.2 can in fact be weakened to $p > 4$. This, however, leads to a less tractable bound for δ.

Remark 6.3.3. Instead of considering the statistic $\Lambda_t^{(n,1)} = \sup_{\lambda_n/n \le t \le 1-\lambda_n/n}\left|w(t)^{-1}\Omega_t^{d_n}\right|$, where $\Omega_t^{d_n} = \left|w(t)^{-1}\big(\mathbf{M}_t^{(n)}\big)^t \mathbf{\Gamma}_n^{-1}\mathbf{M}_t^{(n)}\right|$, one may also use $\Lambda_t^{(n,2)} = n^{-1}\sum_{k=1}^n \Omega_{k/n}^{d_n}$. In fact, Theorems 6.3.1, 6.3.5 and 6.3.6 also apply to $\Lambda_t^{(n,2)}$, and we will therefore not mention it any further.

It would be desirable to strengthen the above result such that we actually have equality in the covariance structures, i.e. $\mathbf{\Gamma}_{\mathbf{S}^{(n)}} = \mathbf{\Gamma}_{\mathbf{W}^{(n)}}$, since then we can replace $\mathbf{B}_t^t \mathbf{\Gamma}_n^{-1}\mathbf{B}_t$ with $\sum_{h=1}^{d_n}\big(B_{t,h}^{(*)}\big)^2$, where $\big\{B_{t,h}^{(*)}\big\}_{0 \le t \le 1,\, 1 \le h \le d_n}$ are independent Brownian Bridges. Unfortunately, this requires more knowledge about the matrix $\mathbf{\Gamma}_{\mathbf{S}^{(n)}}$, and even then, this seems to be rather difficult to establish in general. Before we discuss this in more detail, we will briefly touch on the weight function $w(t)$ in Theorem 6.3.1.

It is well established in the literature, that a weight function increases the power of a testing procedure against an alternative. The specific choice $w(t) = \sqrt{t(1-t)}$ is particularly interesting, since it standardizes the Brownian Bridge $B_t = W_t - tW_1$, for details on this subject, we refer to [29, 32]. We will, however, briefly discuss the choice of λ_n in Theorem 6.3.1. It is reported in [32] that the sequence $\lambda_n = (\log n)^{3/2}$ yields good results in practice. Evaluating the conditions in Theorem 6.3.1 yields that one may choose the dimension d_n such that $d_n = \mathcal{O}\left((\log n)^{\delta}\right)$, with $\delta < 3/4 - 3/(2\nu)$ where ν satisfies (6.3.2).

Instead of considering the weight function $w(t) = \sqrt{t(1-t)}$, one may also work with functions $v(t)$ satisfying the following conditions.

- $v(t)$ is a function on $(0,1)$ increasing in a neighborhood of 0, and decreasing in a neighborhood of 1,
- $\inf_{c \le t \le 1-c} v(t) > 0$ for all $0 < c < 1/2$,
- the function $I(v,c) = \int_0^1 \frac{1}{t(t-1)}\exp(-\frac{cv^2(t)}{t(t-1)})dt$ is finite for some $0 < c < 1/2$.

It is then possible (cf. [14, 78]) to establish an analogue version of Theorem 6.4.1, where $w(t)$ is replaced with $v(t)$, satisfying the conditions above. In particular, one may also choose $v(t) = 1$, i.e. no weight function at all.

We will now continue to discuss the issue as to when and how we have $\mathbf{\Gamma}_{\mathbf{S}^{(n)}} = \mathbf{\Gamma}_{\mathbf{W}^{(n)}}$. As pointed out earlier, we need additional structural assumptions concerning the sequence of matrices $\mathbf{\Gamma}_{\mathbf{S}^{(n)}}$. One possibility is to formulate conditions in terms of the eigenvalues $\lambda_1, ..., \lambda_n$, which are given below together with some additional assumptions regarding the process $\{X_{k,h}\}_{k,h\geq 1}$.

Assumption 6.3.4. *Assume that for* $p > 8$

(i) $\sup_h \|X_{1,h}\|_p < \infty, \quad \mathbb{E}\big(X_{1,h}\big) = 0,$

(ii) $\max_{1\leq h\leq d_n}\big\|X_{k,h} - X'_{k,h}\big\|_p = \mathcal{O}\big(k^{\beta}\big)$, *where* $\beta \geq (4+\sqrt{82})(2\sqrt{81}-16)^{-1} \approx 6.185...$,

(iii) $P\big(\max\big|\widehat{\mathbf{\Gamma}}_{\mathbf{S}^{(n)}} - \mathbf{\Gamma}_{\mathbf{S}^{(n)}}\big| \geq n^{-\gamma}\big) = o\big(1\big)$, $\gamma > 0$,

(iv) The eigenvalues of the matrix $\mathbf{\Gamma}_{\mathbf{S}^{(n)}}$ *satisfy* $1/M_n \leq \lambda_1 \leq ... \leq \lambda_{d_n} \leq M_n$, *where* $M_n = \mathcal{O}\big(d_n\big)$.

Under the above assumptions, the following result is valid.

Theorem 6.3.5. *Assume that Assumption 6.3.4 holds. Then on a possible larger probability space, we have that*

$$\left|\sup_{\lambda_n/n\leq t\leq 1-\lambda_n/n}\big|w(t)^{-1}\big(\mathbf{M}_t^{(n)}\big)^t\widehat{\mathbf{\Gamma}}^{-1}_{\mathbf{S}^{(n)}}\mathbf{M}_t^{(n)}\big| - \sup_{\lambda_n/n\leq t\leq 1-\lambda_n/n}\big|w(t)^{-1}\sum_{h=1}^{d_n}\big(B_{t,h}^{(*)}\big)^2\big|\right| = o_p(1),$$

where $\{B_{t,h}^{(*)}\}_{0\leq t\leq 1,\,1\leq h\leq d_n}$ *are independent Brownian Bridges,* $\lambda_n = (\log n)^{\lambda}$, $\lambda > 0$ *and the dimension satisfies* $d_n = \mathcal{O}\big((\log n)^{\lambda/6} \wedge (\log n)(\log\log n)^{-\delta}\big)$, $\delta > 1$.

Now suppose that we want to test for changes in the mean of $\mathbf{S}^{(n)}$. Denote with $\boldsymbol{\mu} = \big(\mathbb{E}\big(X_{k,1}\big), ...\mathbb{E}\big(X_{k,d_n}\big)\big)^t$ the vector of the means, and recall that per assumption $\mathbb{E}\big(X_{k,h}\big) = 0$ for $k, h \geq 1$. We may thus formulate our null hypothesis as

$$\mathcal{H}_0 : 0 = \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = ... = \boldsymbol{\mu}_n,$$

and the alternative

$$\mathcal{H}_A : 0 = \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = ... = \boldsymbol{\mu}_{k^*} \neq \boldsymbol{\mu}_{k^*+1} = ... = \boldsymbol{\mu}_n.$$

Theorem 6.3.5 provides us with a parameter free (except for the dimension d_n, which is known however) asymptotic expression. For $l > 0$, the exact distribution of $\Omega_l^{(d)} = \sup_{l \leq t \leq 1-l} w(t)^{-1} \sum_{h=1}^{d} \big(B_{t,h}^{(*)}\big)^2$ is not known, however, there exist various asymptotic results (cf. [29]), and it holds that for instance that

$$P\left(\sqrt{\Omega_l^{(d)}} \geq x\right) = \frac{x^d \exp(-x^2/2)}{2^{d/2}\Gamma(d/2)} \left(\log \frac{(1-l)^2}{l^2} - \frac{d}{x^2} \log \frac{(1-l)^2}{l^2} + \frac{4}{x^4} + \mathcal{O}\big(x^{-4}\big)\right),$$

as x tends to infinity. Naturally, one may also use simulated quantiles to provide inference. Next, we turn our attention to the behavior of the statistic $\Omega_{\lambda_n/n}^{(d_n)}$ if the alternative $\mathcal{H}_A$ holds. As is common practice in the literature, we assume that the time of change $k^* = \lceil \tau n \rceil$, $\tau \in (0,1)$ depends on n. Let $\mathbf{S}^{(n)} = \mathbf{S}^{(\leq \lceil \tau n \rceil)} + \mathbf{S}^{(> \lceil \tau n \rceil)}$, where $\mathbf{S}^{(\leq \lceil \tau n \rceil)}$ denotes the pre-change vector, and $\mathbf{S}^{(> \lceil \tau n \rceil)}$ the post-change vector, and define $\mathbf{M}_t^{(\leq \lceil \tau n \rceil)}$, $\mathbf{M}_t^{(> \lceil \tau n \rceil)}$ in an analogue manner.

If we assume that the mean has changed, and that Assumption 6.3.4 is valid for both $\big\{X_{k,h}\big\}_{1 \leq h, 1 \leq k \leq k^*}$ and $\big\{X_{k,h} - \mathbb{E}\big(X_{k,h}\big)\big\}_{1 \leq h, k^* < k}$, then Theorem 6.3.5 remains valid for both sequences. In particular, we obtain that

Theorem 6.3.6. *Assume that Assumption 6.3.4 is valid for both* $\big\{X_{k,h}\big\}_{1 \leq h, 1 \leq k \leq k^*}$ *and* $\big\{X_{k,h} - \mathbb{E}\big(X_{k,h}\big)\big\}_{1 \leq h, k^* < k}$*, and let* $\chi_n = o\big(n^2 d_n^{-2}\big)$*. Then*

$$\liminf_{n \to \infty} \chi_n^{-1} \sup_{\lambda_n/n \leq t \leq 1-\lambda_n/n} \left| w(t)^{-1} \big(\mathbf{M}_t^{(n)}\big)^t \widehat{\mathbf{\Gamma}}_{\mathbf{S}^{(n)}}^{-1} \mathbf{M}_t^{(n)} \right| = \infty,$$

in probability.

Consider now the following examples.

Example 6.3.7. *Let* $\big\{L_k\big\}_{k \in \mathbb{Z}}$ *be an AR(q) process where* q *may depend on* n*, and the innovations* $\big\{\epsilon_k\big\}_{k \in \mathbb{Z}}$ *are a zero mean IID sequence, with* $\big\|\epsilon_k\big\|_p < \infty$*,* $p > 16$*. For simplicity, suppose that the parameters* $\boldsymbol{\zeta}_q = \big(1, \zeta_1, \zeta_2, ..., \zeta_q\big)$ *satisfy* $\sum_{j=1}^q |\zeta_j| \leq \vartheta < 1$*. Hence the process* $\big\{X_k\big\}_{k \in \mathbb{Z}}$ *can be represented as* $L_k = \sum_{i=0}^{\infty} \varphi_i \epsilon_{k-i}$*, where* $|\varphi_n| = \mathcal{O}\big(\rho^n\big)$*,* $0 < \rho < 1$*. We can test for stability in the parameter* $\boldsymbol{\zeta}_q$ *by testing the stability of the covariances* $\phi_h = \mathbb{E}\big(L_0 L_h\big)$*. Note that a change in a parameter will most likely result in a change in most, if not all covariances (one may construct counterexamples though). Naturally, this reflects a global change, and thus using* $\Omega_{\lambda_n/n}^{(d_n)}$ *as test statistic seems reasonable. Put* $S_h^{(n)} = (n-h)\widehat{\phi}_h = \sum_{k=1}^{n-h} L_k L_{k+h}$*. Let* d_n *be as in Theorem 6.3.5. Then, using the causal representation of* L_k *given above together with Lemma 6.5.8, it is not hard to verify that conditions (i)-(iii) of Assumption 6.3.4 are valid. Unfortunately, validating condition (iv) seems to be impossible in general, however, numerical examples indicate that this seems to be valid in most cases.*

Example 6.3.8. *Contrary to the previous example, let* $\{L_k\}_{k\in\mathbb{Z}}$ *be an MA(q) process, where q may depend on n. As before, we want to test for stability in the parameters by testing the stability of the covariances. Suppose that at time* k^*+1, $\{L_k\}_{k\in\mathbb{Z}}$ *becomes an MA(q + 1) process. If we use the first* d_n *covariance estimators* $S_h^{(n)} = (n-h)\widehat{\phi}_h = \sum_{k=1}^{n-h} L_k L_{k+h}$ *to form* $\Omega_{\lambda_n/n}^{(d_n)}$, *and if* $q \ll d_n$, *then the statistic* $\Omega_{\lambda_n/n}^{(d_n)}$ *will often fail to detect a change, since the change is no longer global. The reason for this is that* $\phi_h = 0$ *for* $h > q$, *before the change, and* $\phi_h = 0$ *for* $h > q+1$ *after the change, and hence the vast majority of the covariances remains unaltered.*

Example 6.3.9. *Let* $X_{k,h} = \sum_{i=0}^{\infty} \alpha_i L_{k-i,h}$, *where* $\{L_{k,h}\}_{k\in\mathbb{Z}}$ *is an IID sequence for every fixed h, and* $\{L_{k,h}\}_{h\in\mathbb{Z}}$ *is an AR(*d_n*) process with parameter* $\boldsymbol{\zeta} = \big(1, \zeta_1, ..., \zeta_{d_n}\big)$ *for every fixed k, i.e.* $L_{k,h} = \zeta_1 L_{k,h-1} + ... + \zeta_{d_n} L_{k,h-d_n} + \epsilon_{k,h}$, *where* $\{\epsilon_{k,h}\}_{h\in\mathbb{Z}}$ *is a zero mean white noise sequence, i.e. it holds that* $\mathbb{E}\big(\epsilon_{k,i}\epsilon_{k,j}\big) = 0$ *for* $i \neq j$. *Note that this results in* $n^{-1} Cov\big(S_i^{(n)}, S_j^{(n)}\big) = \mathbb{E}\big(L_{0,i}L_{0,j}\big)\sum_{r=0}^{\infty}\alpha_r^2$, *where* $S_h^{(n)} = \sum_{k=1}^{n} X_{k,h}$. *Suppose that* $\sum_{j=1}^{q} |\zeta_j| \leq \vartheta < 1$. *Proposition 6.5.6 now implies that the eigenvalues of the covariance matrix* $\mathbf{\Gamma}_{\mathbf{S}}^{(n)}$ *are bounded from below and above, i.e; Assumption 6.3.4 (iv) is valid. Assume in addition* $|\alpha_k| = \mathcal{O}\big(k^{-\beta}\big)$, *where* β *is as in Assumption 6.3.4 (ii), and that* $\big\|\epsilon_{k,h}\big\|_p < \infty$ *for* $p > 8$, $k \in \mathbb{Z}$, $1 \leq h \leq d_n$. *Then Assumption 6.3.4 is valid, and we may use the statistic* $\Omega_{\lambda_n/n}^{(d_n)}$ *to test for global changes in the mean* $\boldsymbol{\mu} = \big(\mathbb{E}\big(X_{k,1}\big), ...\mathbb{E}\big(X_{k,d_n}\big)\big)^t$. *Note however, that this test has low power if* d_n *is rather large, and we only have local changes in the vector* $\boldsymbol{\mu}$, *i.e; we only have changes in* $q \ll d_n$ *components of* $\boldsymbol{\mu}$.

6.4 Local changes

In this section, we present approximation results that can be used to detect local changes. Under the hypothesis of Assumption 6.2.2, we can formulate the following two results.

Theorem 6.4.1. *Suppose that Assumption 6.2.2 (i) - (v) holds, and that* $\inf_h \psi_h^2 > 0$. *Then*

$$\lim_{n\to\infty} P\left(\max_{1\leq h\leq d_n} \psi_h^{-1} Z_h^{(n,l)} \leq u_n(z)\right) = \exp(-z),$$

where $z = \exp(-x)$, *and* $\theta_l = \big(2\log(1-l) - 2\log l\big)$ *and* $u_n = u_n(z) = a_n x + b_n$, *with* $a_n = (2\log n)^{-1/2}$ *and* $b_n = \sqrt{2\log n} + (2\log n)^{-1/2}\left(\frac{1}{2}\log\log n + \log\theta_l - \frac{1}{2}\log\pi\right)$.

For $\nu > 2$, the dimension $d_n = \mathcal{O}(n^\delta)$ must satisfy the relation

$$\theta \leq \delta < \min\left\{\frac{1}{\nu} - \frac{2}{p}, \frac{2+p-2(1+\theta)\nu}{(4+4\theta+p(7+\theta))\nu}\right\}, \tag{6.4.1}$$

where we require $p > 4\nu$. Alternatively, if one sets $d_n = \mathcal{O}\big((\log n)^\delta\big)$, for arbitrary $\delta > 0$, then we require

$$\nu < \min\left\{\frac{4+2p}{4+4\theta+p\theta+4\theta^2+p\theta^2}, \frac{2}{1-2\theta}, p/4\right\}. \tag{6.4.2}$$

Theorem 6.4.2. *Suppose that Assumption 6.2.2 (i) - (v) holds, and that $\inf_h \psi_h^2 > 0$. Then*

$$\lim_{n\to\infty} P\left(\max_{1\leq h\leq d_n} \psi_h^{-1} \sup_{0\leq t\leq t} \big|M_{t,h}^{(n)}\big| \leq v_n(z)\right) = \exp(-z),$$

where $v_n = v_n(z) = e_n x + f_n$, with $e_n = 1/4\big(\log(2n)/2\big)^{-1/2}$, $f_n = \sqrt{1/2\log(2n)}$, and $z = \exp(-x)$. The parameters d_n, θ and p must satisfy the same conditions as in Theorem 6.4.1.

Remark 6.4.3. In both Theorems, conditions (i) and (iii) of Assumption 6.2.2 can in fact be weakened to $p > 4$. This, however, leads to a less tractable bound for δ.

Note that in contrast to Theorem 6.4.1, Theorem 6.4.2 does not include any weight function, and in particular not the parameter l. For more details on other possible weight functions we refer to the previous section or [29, 32].

A very important issue in practice is the rate of convergence. It is well known that the actual rate of convergence of extremes to an extreme value distribution can be very slow, and depends on the underlying distribution (cf. [33, 59, 97, 98]). On the other hand, the approximation error resulting from Gaussian approximations can be remarkably small, in particular in the non weighted case (cf. Theorem 6.6.4). This suggests that using simulated quantiles is more appropriate. On the first glance, this would require to estimate the covariances $\rho_{i,j}^{(n)}$, $1 \leq i,j \leq d_n$. However, the proof of Theorem 6.7.5 (cf. [79, Theorem 2.1]) shows that one can approximate $\max_{1\leq h\leq d_n} \psi_h^{-1} Z_h^{(n,l)}$ (resp. $\max_{1\leq h\leq d_n} \psi_h^{-1} \sup_{0\leq t\leq t}\big|M_{t,h}^{(n)}\big|$) with a sequence of independent, weighted Brownian Bridges $\max_{1\leq h\leq d_n} \sup_{l\leq t\leq 1-l} |B_{t,h}(t(1-t))^{-1/2}|$ (resp. $\max_{1\leq h\leq d_n} \sup_{l\leq t\leq 1-l} |B_{t,h}|$), hence estimating the covariance matrix is not necessary if d_n is large enough.

In general, the variance ψ_h is not known in practice and needs to be estimated.

One may hope that the above Theorems are still valid if one replaces ψ_h with the corresponding estimates $\widehat{\psi}_h$, and indeed this is the case if the following mild condition is imposed on potential variance estimators $\widehat{\psi}_h$.

Assumption 6.4.4. *For some $\alpha > 1$, the estimators $\widehat{\psi}_h$ satisfy*

$$P\big(\max_{1\leq h\leq d_n} |\widehat{\psi}_h^2 - \psi_h^2| \geq (\log n)^{\alpha}\big) = o(1). \tag{6.4.3}$$

We then have the corresponding analogue of the above results.

Theorem 6.4.5. *Assume that the conditions of Theorem 6.4.1 hold, and that Assumption 6.4.4 holds. Then*

$$\lim_{n\to\infty} P\left(\max_{1\leq h\leq d_n} \widehat{\psi}_h^{-1} Z_h^{(n,l)} \leq u_n(z)\right) = \exp(-z),$$

where $u_n(z)$ and d_n are as in Theorem 6.4.1.

Remark 6.4.6. An analogue result is valid for $\max_{1\leq h\leq d_n} \psi_h^{-1} \sup_{0\leq t\leq t}\big|M_{t,h}^{(n)}\big|$.

The literature (cf. [4, 25, 60]) provides many potential candidates to estimate the long run variance ψ_h^2. A popular estimator is Bartlett's estimator, or more general, estimators of the form

$$\widehat{\psi}_h^2 = \sum_{|j|\leq r} \omega(k/r)\widehat{\gamma}_{j,h} \tag{6.4.4}$$

with weight function $\omega(x)$, where $\gamma_{j,h} = \mathbb{E}\big(Y_{0,h}Y_{j,h}\big)$ and $\widehat{\gamma}_{j,h} = n^{-1}\sum_{k=1}^{n-j} Y_{k,h}Y_{k+j,h}$. Considering the triangular weight function $\omega(x) = 1-|x|$ for $|x| \leq 1$ and $\omega(x) = 0$ for $|x| > 1$, one recovers the Bartlett estimator in (6.4.4). One may also use the plain estimate

$$\widehat{\psi}_h^2 = \widehat{\gamma}_{0,h} + 2\sum_{i=1}^{l_n} \widehat{\gamma}_{i,h}, \tag{6.4.5}$$

see for instance [112, 113]. In particular, Wu [131, Proposition 1] provides the following result, which we have reformulated for our setting.

Proposition 6.4.7. *Let $l_n \in \mathbb{N}$, $l_n \to \infty$ as n increases with $l_n = \mathcal{O}\left(\sqrt{n}\big(d_n(\log n)^{\alpha}\big)^{-1}\right)$, where $\alpha > 1$. If Assumption 6.2.2 holds, then*

$$\limsup_{n\to\infty} P\big(\max_{1\leq h\leq d_n} \big|\widehat{\psi}_h^2 - \psi_h^2\big| > (\log n)^{-\alpha}\big) = 0, \ \alpha > 1$$

where $\widehat{\psi}_h^2$ is as in (6.4.5).

Consequently, Theorem 6.4.5 is valid if one uses the variance estimator given in (6.4.5). As in Section 6.3, using the therein established notation, we may now formulate the null hypothesis and the alternative as

$$\mathcal{H}_0 : 0 = \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = ... = \boldsymbol{\mu}_n,$$

and

$$\mathcal{H}_A : 0 = \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = ... = \boldsymbol{\mu}_{k^*} \neq \boldsymbol{\mu}_{k^*+1} = ... = \boldsymbol{\mu}_n.$$

Using $\Upsilon_{n,d_n} = \max_{1 \leq h \leq d_n} \widehat{\psi}_h^{-1} Z_h^{(n,l)}$ as a test statistic, Theorem 6.4.5 provides us with a parameter free asymptotic distribution. In addition, it is a relatively easy task to show that this test is consistent, i.e. Υ_{n,d_n} will explode under the alternative $\mathcal{H}_A$. To verify this, suppose that $\boldsymbol{\mu}_{k^*}$ changes at least in element $\mu_{k^*,h}$, and that $n^{-1/2}\widehat{\psi}_h^{-1} S_h^{(n,[nt])} \xrightarrow{\mathbb{D}[0,1]} W_t$. Then

$$\Upsilon_{n,d_n} \geq \widehat{\psi}_h^{-1} Z_h^{(n,l)} = \mathcal{O}_P\big(1\big) + \mathcal{O}\big(\sqrt{n}\big),$$

which proves consistency.

So far, nothing was said about the dependence conditions C2 made in Assumption 6.2.2. For an illustrative purpose, we consider the following very important example. Let $L_k = \sum_{i=0}^{\infty} \alpha_i \epsilon_{k-i}$ be a linear process, where $\big\{\epsilon_k\big\}_{k \in \mathbb{Z}}$ is a mean zero IID sequence. We are interested in the stability of the covariances $\phi_h = \mathbb{E}\big(L_h L_0\big)$, which, as previously mentioned, is an important issue. To this end, we consider $Y_{k,h} = L_{k+h-1} L_k - \phi_{h-1}$.

Theorem 6.4.8. *Assume that Assumption 6.4.4 holds, and that*

(i) $\big\|\epsilon_1\big\|_{12} < \infty$,

(ii) $\sum_{i=0}^{\infty} i^{\lambda} |\alpha_i| < \infty$, *where* λ *satisfies the condition given below,*

(iii) $\limsup_{h \to \infty} \sup_{m \geq L} \frac{\phi_{m+h}}{\phi_m} < 1$, *for some finite* $L > 0$.

Then

$$\lim_{n \to \infty} P\left(\max_{1 \leq h \leq d_n} \widehat{\psi}_h^{-1} Z_h^{(n,l)} \leq u_n(z)\right) = \exp(-z),$$

where $u_n(z)$ *is as in Theorem 6.4.1, and*

$$\delta < \frac{1}{5}\left(\sqrt{154} - 12\right) \approx 0.08193, \ \textit{and}\ \lambda > \frac{2\sqrt{154} - 19}{4(\sqrt{154} - 48)} \approx 3.5512.$$

Remark 6.4.9. An analogue result is valid for $\max_{1\leq h\leq d_n}\psi_h^{-1}\sup_{0\leq t\leq t}\big|M_{t,h}^{(n)}\big|$.

To estimate the variance, one may use the estimator proposed in Proposition 6.4.7. However, if the assumptions of Theorem 6.4.8 hold, lengthy and tedious calculations show that by [4, Theorem 9.3.4] (see also [53]), we have in case of the Bartlett weights that

$$\limsup_{n\to\infty}\max_{1\leq h\leq d_n}\frac{n}{r}\mathrm{Var}\left(\widehat{\psi}_h^2\right)<\infty,$$

provided that $r/n\to 0$. Hence if we choose $r=r_n=\mathcal{O}\left(nd_n^{-1}(\log n)^{-2\alpha}\right)$, $\alpha>1$, the Markov inequality implies

$$P\big(\max_{1\leq h\leq d_n}|\widehat{\psi}_h^2-\psi_h^2|\geq(\log n)^{\alpha}\big)\leq d_n(\log n)^{2\alpha}\max_{1\leq h\leq d_n}\mathrm{Var}\left(\widehat{\psi}_h^2\right)=\mathcal{O}(1),$$

thus Assumption 6.4.4 is valid. Let us briefly reconsider examples 6.3.7, 6.3.8 and 6.3.9 of the previous section. The conditions stated in example 6.3.7 satisfy those of Theorem 6.4.8 if we assume in addition that $\limsup_{h\to\infty}\sup_{m\geq L}\frac{\phi_{m+h}}{\phi_m}<1$, for some finite $L>0$, an this is also the case in example 6.3.8. In case of Example 6.3.9, we can apply Theorem 6.4.5, but we need to verify Assumption 6.2.2. Condition (i) is trivially valid, conditions (ii)-(iii) follow from Proposition 6.2.4, whereas condition (v) holds since $\sup_{i,j:k\leq|i-j|}\rho_{i,j}^{(n)}$ decays exponentially fast in k. Verifying condition (iv) however is a little more involved. First, note that the Cauchy-Schwarz inequality yields $\varphi_{|i-j|}^{(n)}:=\rho_{i,j}^{(n)}\leq 1$. Suppose now that for some subsequence n' and corresponding sequences k_n', h_n', with $h_n'\geq 1$ we actually have

$$\lim_{n'}\big|\mathrm{Corr}\big(S_{k_n'}^{(n')},S_{k_n'+h_n'}^{(n')}\big)\big|=|\varphi_{|h_n'|}^{(n')}|=1. \tag{6.4.6}$$

If we consider the corresponding subsequence of the 2×2 submatrices

$$\mathbf{A}_{n'}=\big((\varphi_0^{(n)},\varphi_{|h_n'|}^{(n)})^t,(\varphi_{|h_n'|}^{(n)},\varphi_0^{(n)})^t\big),$$

then it follows that the smaller eigenvalue $\lambda_{\mathbf{A}_{n'},2}$ converges to zero. Hence we obtain from Cauchy's interlacing theorem ([106]) that the smallest eigenvalue λ_n of $\mathbf{\Gamma}_{\mathbf{S^{(n)}}}$ tends to zero, which however contradicts Proposition 6.5.6. Hence we must have

$$\limsup_{n\to\infty}\sup_{h,k,\,h\geq 1}\big|\mathrm{Corr}\big(S_k^{(n)},S_{k+h}^{(n)}\big)\big|<1. \tag{6.4.7}$$

6.5 Proofs and ramifications

Throughout the proofs, C denotes a generic constant that may vary from one formula to another. The proofs are essentially based on the following two theorems, whose proof is given in Sections 5.5 and 6.7. We point out that the proof of Theorem 6.4.2 is almost identical with the one of Theorem 6.4.1, this is also true for all the related results. The sole difference is that one has to apply Theorem 6.7.6 instead of 6.5.2. We therefore omit all proofs involving the non-weighted quantity $\sup_{0\leq t\leq t}\big|M_{t,h}^{(n)}\big|$.

Theorem 6.5.1. *Assume that the assumptions of Theorem 6.6.4 hold, and let $\lambda_n = o(n)$ be a monotone increasing sequence. Then, for each n, we can define a d_n-dimensional Brownian Bridge $\big\{\mathbf{B}_t^{(n)}\big\}_{t\geq 0} = \big\{B_{t,h}^{(n)}\big\}_{\substack{0\leq t\leq 1,\\ 0\leq h\leq d_n}}$ such that*

$$\max_{0\leq h\leq d_n}\sup_{\lambda_n\leq t\, n\leq n-\lambda_n}\frac{\big|M_{t,h}^{(n)}-\psi_h B_{t,h}^{(n)}\big|}{\big(t(1-t)\big)^{1/2}} = \mathcal{O}_P\big(\lambda_n^{-1/2+1/\nu}\big),$$

for $\nu \geq 2$. The dimension $d_n = \mathcal{O}\big(n^{\delta}\big)$ must satisfy the relation

$$\theta \leq \delta < \min\left\{\frac{4(p-2\nu)}{(-4+3p)\nu}, \frac{2+p-2(1+\theta)\nu}{(2+4\theta+p(4+\theta))\nu}\right\},$$

where we require $p > 4\nu$. In addition, we have that $\max\big|\mathbf{\Gamma}_{\mathbf{M}}^{(n)} - \mathbf{\Gamma}_{\mathbf{B}}^{(n)}\big| = \mathcal{O}\big(n^{-\gamma}\big)$, for some $\gamma > 0$. Alternatively, if one sets $d_n = \mathcal{O}\big((\log n)^{\delta}\big)$, for arbitrary $\delta > 0$, then we require

$$\nu < \min\left\{\frac{4+2p}{4+4\theta+p\theta+4\theta^2+p\theta^2}, \frac{2}{1-2\theta}, p/4\right\}.$$

Let $\big\{W_t\big\}_{t\geq 0}$ be a Brownian motion, and denote with

$$B_h^{(l)} = \sup_{l\leq t\leq 1-l}\left|\frac{W_{t,h}-tW_{1,h}}{\sqrt{t(1-t)}}\right| \tag{6.5.1}$$

the weighted Brownian Bridge. Then Theorem 6.5.1 essentially allows us to replace $\max_{1\leq h\leq d_n}\psi_h^{-1}Z_h^{(n,l)}$ with $\max_{1\leq h\leq d_n}B_h^{(l)}$, hence it suffices to show that $\max_{1\leq h\leq d_n}B_h^{(l)}$, appropriately normalized, converges in distribution to the desired extreme value distribution. This step is accomplished in the Theorem given below.

Theorem 6.5.2. *Suppose that*

(i) $\limsup_{n\to\infty}(\log n)^2\Big(\sum_{r=\sqrt{\log n}}^{d_n}\sup_{r\leq|i-j|}\big|\rho_{i,j}\big|\Big) = 0,$

(ii) $\max_{1\le h\le n}\sup_{i,j:2\le|i-j|}\left|\rho_{i,j}\right| < 1.$

Then

$$\lim_{n\to\infty} P\left(\max_{1\le h\le n} B_h^{(l)} \le u_n(z)\right) = \exp(-z),$$

where $z = \exp(-x)$*, and* $\theta_l = \big(2\log(1-l) - 2\log l\big)$ *and* $u_n = u_n(z) = a_n x + b_n$*, with* $a_n = (2\log n)^{-1/2}$ *and* $b_n = \sqrt{2\log n} + (2\log n)^{-1/2}\left(\frac{1}{2}\log\log n + \log\theta_l - \frac{1}{2}\log\pi\right)$.

Before proving the results of Section 6.3 and 6.4, we will show the validity of Proposition 6.2.4.

Proof of Proposition 6.2.4. Let $\mathcal{F}'_k = \sigma(\epsilon'_k, \epsilon'_{k-1}, \ldots)$. Then for any $p \ge 1$ we have by Jensen's and the triangular inequality

$$\begin{aligned}\big\|Y_{k,h}^{(>m_n)}\big\|_p &\le \big\|X_{k,h} - X_{k,h}^{(m_n,*)}\big\|_p + \big\|\mathbb{E}\big(X_{k,h} - X_{k,h}^{(m_n,*)} \mid \sigma\big(\mathcal{F}_{k-m_n}^{k+m_n} \cup \mathcal{F}'_{k-m_n-1}\big)\big)\big\|_p \\ &\le 2\big\|X_{k,h} - X_{k,h}^{(m_n,*)}\big\|_p,\end{aligned}$$

where we also used the fact that

$$X_{k,h}^{(m_n,*)} - \mathbb{E}\big(X_{k,h}^{(m_n,*)} \mid \mathcal{F}_{k-m_n}^{k+m_n}\big) = \mathbb{E}\big(X_{k,h}^{(m_n,*)} - X_{k,h} \mid \sigma\big(\mathcal{F}_{k-m_n}^{k+m_n} \cup \mathcal{F}'_{k-m_n-1}\big)\big).$$

By [129, Theorem 1 (iii)] and we have for $p \ge 2$

$$\big\|X_{1,h} - X_{1,h}^{(m_n,*)}\big\|_p^2 \le C\sum_{i=-\infty}^{0}\big\|X_{m_n+h-i,h} - X'_{m_n-i,h}\big\|_p^2 = \mathcal{O}\left(m_n^{1-2\beta}\right), \qquad (6.5.2)$$

which leads to

$$\big\|Y_{k,h}^{(>m_n)}\big\|_p = \mathcal{O}\left(m_n^{1/2-\beta}\right). \qquad (6.5.3)$$

Consequently, using the triangular inequality and the above, we obtain that

$$\max_{1\le l\le n}\left\|\sum_{j=1}^{l} Y_{j,h}^{(>m_n)}\right\|_p \le \sum_{k=1}^{n}\big\|Y_{k,h}^{(>m_n)}\big\|_p = \mathcal{O}\left(n\, m_n^{1/2-\beta}\right) = \mathcal{O}(1),$$

which proves Assumption 6.2.2 (iii). In order to show (ii), note that the Cauchy-Schwarz inequality implies

$$\begin{aligned}\big|\mathbb{E}\big(X_{k,h}X_{0,h}\big)\big| = \big|\mathbb{E}\big(X_{0,h}\mathbb{E}\big(X_{k,h} \mid \mathcal{F}_h\big)\big| &\le \big\|X_{0,h}\big\|_2\big\|\mathbb{E}\big(X_{k,h} \mid \mathcal{F}_h\big)\big\|_2 \\ &\le \big\|X_{0,h}\big\|_2\big\|X_{k,h} - X_{k,h}^{(k,*)}\big\|_2,\end{aligned}$$

and it follows from (6.5.2) that

$$\sum_{j=0}^{\infty} j|\phi_j| \le C\sum_{j=0}^{\infty}\big\|X_{j,h} - X_{j,h}^{(j,*)}\big\|_2 = \mathcal{O}\left(\sum_{j=1}^{\infty} j^{3/2-\beta}\right) = \mathcal{O}(1).$$

□

$$B_h^{(l)} = \sup_{l \le t \le 1-l} \left| \frac{W_{t,h} - tW_{1,h}}{\sqrt{t(1-t)}} \right| \tag{6.5.4}$$

6.5.1 Proofs of Section 6.3

We will first give the proof of Theorem 6.3.1.

Proof of Theorem 6.3.1. Define the random vector

$$\mathbf{Z}_n = \left(Z_1^{(n,\lambda_n/n)}, Z_2^{(n,\lambda_n/n)}, ..., Z_{d_n}^{(n,\lambda_n/n)}\right)^t,$$
$$\mathbf{B}_n = \left(B_1^{(\lambda_n/n)}, B_2^{(\lambda_n/n)}, ..., B_{d_n}^{(\lambda_n/n)}\right)^t,$$

and for a function $f(t)$, we denote with $|f|_n^* := \sup_{\lambda_n/n \le t \le 1-\lambda/n} w(t)^{-1}\big|f(t)\big|$. Let us first assume that $\max|\mathbf{\Gamma}_n^{-1}| \le L$, for some constant L. We have that

$$\big||\mathbf{M}_t^{(n)}\mathbf{\Gamma}_n^{-1}\mathbf{M}_t^{(n)}|_n^* - |\mathbf{B}_t\mathbf{\Gamma}_n^{-1}\mathbf{B}_t|_n^*\big| \le \big||\mathbf{M}_t^{(n)} - \mathbf{B}_t)^t\mathbf{\Gamma}_n^{-1}\mathbf{B}_t|_n^*\big| + \big||\mathbf{B}_t^t\mathbf{\Gamma}_n^{-1}(\mathbf{M}_t^{(n)} - \mathbf{B}_t)|_n^*\big| + \big||(\mathbf{M}_t^{(n)} - \mathbf{B}_t)^t\mathbf{\Gamma}_n^{-1}(\mathbf{M}_t^{(n)} - \mathbf{B}_t)|_n^*\big|.$$

Using that $\max|\mathbf{\Gamma}_n^{-1}| \le L$, this is further smaller than

$$\big||\mathbf{M}_t^{(n)}\mathbf{\Gamma}_n^{-1}\mathbf{M}_t^{(n)}|_n^* - |\mathbf{B}_t\mathbf{\Gamma}_n^{-1}\mathbf{B}_t|_n^*\big| \le 2Cd_n^2\left(\max\big|\mathbf{Z}_n - \mathbf{B}_n\big|\right)\left(\max\big|\mathbf{B}_n\big|\right) + d_n^2\left(\max\big|\mathbf{Z}_n - \mathbf{B}_n\big|\right)^2.$$

By Theorem 6.6.7, we have that that $\max\big|\mathbf{Z}_n - \mathbf{B}_n\big| = \mathcal{O}_p\left(\lambda_n^{-1/2+1/\nu}\right)$. Let $K > 0$. Since we have $d_n^2\lambda_n^{-1/2+1/\nu} = \mathcal{O}\left(n^{-\kappa}\right)$, for some $\kappa > 0$, we obtain the bound

$$P\left(\big|\mathbf{Z}_n^t\mathbf{\Gamma}_n^{-1}\mathbf{Z}_n - \mathbf{B}_n^t\mathbf{\Gamma}_n^{-1}\mathbf{Z}_n\big| \ge K\right) \le o(1) + 2P\left(\max\big|\mathbf{B}_n\big| \ge (C\,d_n\lambda_n)^{-1}\right)$$
$$\le Cd_nP\left(B_1^{(\lambda_n/n)} \ge n^\eta\right),$$

for some $\eta > 0$. Lemma 6.7.1 implies that $P\left(B_1^{(\lambda_n/n)} \ge n^\eta\right) = \mathcal{O}\left(d_n^{-2}\right)$ which yields

$$d_nP\left(B_1^{(\lambda_n/n)} = o(1).\right.$$

Hence we can choose a sequence K_n that tends to zero as n increases, such that

$$P\left(\big||\mathbf{M}_t^{(n)}\mathbf{\Gamma}_n^{-1}\mathbf{M}_t^{(n)}|_n^* - |\mathbf{B}_t\mathbf{\Gamma}_n^{-1}\mathbf{B}_t|_n^*\big| \ge K\right) = o(1). \tag{6.5.5}$$

Moreover, we can choose a sequence L_n with $\lim_n L_n = \infty$ and $\max|\mathbf{\Gamma}_n^{-1}| \le L_n$, such that (6.5.5) remains valid, hence the claim follows. □

The proof of Theorem 6.3.5 is more involved, and will be developed in a series of Lemmas. The difficulty mainly consists in controlling the error of $\max\big|\widehat{\mathbf{\Gamma}}_{\mathbf{S}^{(n)}}^{-1} - \mathbf{\Gamma}_{\mathbf{W}^{(n)}}^{-1}\big|$, which, however, is unpleasant enough. To this end, we require the following Lemma which may be folklore, however, we could not find a reference for it.

Lemma 6.5.3. *Let* $\mathbf{A} = \big(a_{i,j}\big)_{1\leq i,j\leq d}$ *and* $\mathbf{B} = \big(b_{i,j}\big)_{1\leq i,j\leq d}$ *be two regular* $d \times d$ *dimensional matrices, such that*

- $\max\big|\mathbf{A}\big| \leq 1$, $\max\big|\mathbf{B}\big| \leq 1$
- $\max\big|\mathbf{A} - \mathbf{B}\big| \leq \epsilon$, *for some* $\epsilon > 0$.

Denote with $\mathbf{A}^{-1} = \big(a^*_{i,j}\big)_{1\leq i,j\leq d}$, $\mathbf{B}^{-1} = \big(b^*_{i,j}\big)_{1\leq i,j\leq d}$ *the inverse matrices of* $\mathbf{A}, \mathbf{B}$*. Then it holds that*

$$\max\big|\mathbf{A}^{-1} - \mathbf{B}^{-1}\big| \leq \frac{d!d\epsilon\big(\big|\det(\mathbf{A})\big| + d!d\epsilon\big) + \max|\mathbf{A}^{-1}|d!d\epsilon}{\big|\det(\mathbf{A})^2 - |\det(\mathbf{A})d!d\epsilon|\big|}\left(1 - \frac{d!d\epsilon}{\big|\det(\mathbf{A})^2 - |\det(\mathbf{A})d!d\epsilon|\big|}\right)^{-1}.$$

Lemma 6.5.4. *Let* $\mathbf{A}$ *be a* d*-dimensional regular matrix, such that the eigenvalues satisfy* $0 < 1/M \leq \lambda_1 \leq ... \leq \lambda_n \leq M < \infty$*. Then it holds that* $\max\big|\mathbf{A}^{-1}\big| \leq \sqrt{d}M$*.*

Proof of Lemma 6.5.4. It holds that

$$\max\big|\mathbf{A}^{-1}\big| \leq \sqrt{\sum_{1\leq i,j\leq d}\big|a^*_{i,j}\big|^2} = \sqrt{\sum_{i=1}^{d}\lambda_i^{-2}} \leq \sqrt{d}M,$$

where $\mathbf{A}^{-1} = \big(a^*_{i,j}\big)_{1\leq i,j\leq d}$. □

Put

$$\mathbf{\Gamma_{d_n}} = \big(\phi_{|i-j|}\big)_{1\leq i,j\leq d_n}, \quad \mathbf{\Gamma}_{d_n}^{-1} = \big(\gamma^*_{i,j}\big)_{1\leq i,j\leq d_n}. \tag{6.5.6}$$

Let $\mathbf{A} = \big(a_{i,j}\big)_{1\leq i,j\leq d_n}$ be a $d_n \times d_n$ regular matrix. Based on these two results, we can now control the distance $\max\big|\mathbf{A} - \mathbf{\Gamma}_{d_n}\big|$ as follows.

Corollary 6.5.5. *Let* $d = d_n = \log n(\log\log n)^{-\delta}$, $\delta > 1$, *and* $0 < \gamma < \gamma^+$*. If*

$$\max\big|\mathbf{A} - \mathbf{\Gamma}_{d_n}\big| \leq n^{-\gamma^+},$$

then

$$\big|\mathbf{A}^{-1} - \mathbf{\Gamma}_{d_n}^{-1}\big| = \mathcal{O}\left(n^{-\gamma}\right).$$

Proof. This follows directly by evaluating the bound in Lemma 6.5.3, using $\det \mathbf{A} = \prod_{i=1}^{n} \lambda_i$ and the relation

$$\log d_n! = d_n \log d_n - d_n + \mathcal{O}\left(\log d_n\right).$$

□

We are now ready to prove Theorem 6.3.5.

Proof of Theorem 6.3.5. First note that by Proposition 6.2.4, the assumptions of Theorem 6.3.1 are validated. Next, we notice that

$$\begin{aligned} \big||\mathbf{M}_t^{(n)}\mathbf{\Gamma}_{\mathbf{S}^{(n)}}^{-1}\mathbf{M}_t^{(n)}|_n^* - |\mathbf{M}_t^{(n)}\widehat{\mathbf{\Gamma}}_{\mathbf{S}^{(n)}}^{-1}\mathbf{M}_t^{(n)}|_n^*\big| &\leq \big||\mathbf{M}_t^{(n)}(\mathbf{\Gamma}_{\mathbf{S}^{(n)}}^{-1} - \widehat{\mathbf{\Gamma}}_{\mathbf{S}^{(n)}}^{-1})\mathbf{M}_t^{(n)}|_n^* \\ &\leq d_n^2 \max\big|\mathbf{\Gamma}_{\mathbf{S}^{(n)}}^{-1} - \widehat{\mathbf{\Gamma}}_{\mathbf{S}^{(n)}}^{-1}\big| \max\big|\mathbf{Z}_n\big|^2. \end{aligned} \tag{6.5.7}$$

Let $\mathbf{A}$ be a matrix such that $\max\big|\mathbf{\Gamma}_{\mathbf{S}^{(n)}} - \mathbf{A}\big| = \mathcal{O}\big(n^{-\gamma}\big)$. Then Assumption implies that

$$P\left(\max\big|\widehat{\mathbf{\Gamma}}_{\mathbf{S}^{(n)}} - \mathbf{A}\big| \geq n^{-\gamma}\right) = \mathcal{o}(1). \tag{6.5.8}$$

Using Corollary 6.5.5 we thus obtain

$$P\left(\max\big|\widehat{\mathbf{\Gamma}}_{\mathbf{S}^{(n)}}^{-1} - \mathbf{\Gamma}_{\mathbf{W}^{(n)}}^{-1}\big| \geq\geq n^{-\gamma^-}\right) = \mathcal{o}\big(1\big), \tag{6.5.9}$$

where $0 < \gamma^- < \gamma$.

Hence we obtain from Theorem 6.6.4 and (6.5.7) that for $K > 0$

$$\begin{aligned} P\Big(\big||\mathbf{M}_t^{(n)}\widehat{\mathbf{\Gamma}}_{\mathbf{S}^{(n)}}^{-1}\mathbf{M}_t^{(n)}|_n^* - |\mathbf{M}_t^{(n)}\mathbf{\Gamma}_{\mathbf{W}^{(n)}}^{-1}\mathbf{M}_t^{(n)}|_n^*\big|^{1/2} > K\Big) &\leq \mathcal{o}(1) + P\big(\max\big|\mathbf{Z}_n\big| \geq K\, d_n n^{-1/4}\big) \\ &\leq \mathcal{o}(1) + P\big(\max\big|\mathbf{B}_n\big| \geq K\, d_n n^{-\gamma^-/2}\big). \end{aligned}$$

Arguing as in the proof of Theorem 6.3.1, one obtains that

$$P\big(\max\big|\mathbf{B}_n\big| \geq K\, d_n n^{-\gamma^-/2}\big) = \mathcal{o}(1),$$

hence it remains to evaluate the bounds provided by Theorem 6.3.1. The condition $\beta \geq (4+\sqrt{82})(2\sqrt{81}-16)^{-1}$ implies that we may choose $\theta = 1/24(4\sqrt{82}-32)$, which allows us to choose $\nu > 3$. This in turn implies that $d_n = \mathcal{O}\big((\log n)^{\lambda/6} \wedge (\log n)(\log\log n)^{-\delta}\big)$, $\delta > 1$, which completes the proof. □

Proof of Theorem 6.3.6. As in the proof of Theorem 6.3.5 one derives that

$$\widehat{\mathbf{\Gamma}}^{-1}_{\mathbf{S}^{(n)}} = \tau \mathbf{\Gamma}^{-1}_{\mathbf{s}^{(n,1)}} + (1-\tau) + \tau \mathbf{\Gamma}^{-1}_{\mathbf{s}^{(n,2)}} + \mathcal{O}_p\big(1\big),$$

where the matrices $\mathbf{\Gamma}_{\mathbf{s}^{(n,1)}}, \mathbf{\Gamma}_{\mathbf{s}^{(n,2)}}$ denote the pre respectively post-change covariance matrices. Moreover, Assumption 6.3.4 implies that

$$\begin{aligned}\boldsymbol{\mu}^t_{k^*+1}\big(\tau \mathbf{\Gamma}^{-1}_{\mathbf{s}^{(n,1)}} + (1-\tau) + \tau \mathbf{\Gamma}^{-1}_{\mathbf{s}^{(n,2)}}\big)\boldsymbol{\mu}_{k^*+1} &= \tau \sum_{i=1}^{d_n} x_i^2 \lambda_i^{(1)} + (1-\tau) \sum_{i=1}^{d_n} y_i^2 \lambda_i^{(2)} \\ &\geq d_n^{-1}\big(\tau \sum_{i=1}^{d_n} x_i^2 + (1-\tau) \sum_{i=1}^{d_n} y_i^2\big),\end{aligned}$$

for some $x_i, y_i \in \mathbb{R}$, $1 \leq i \leq d_n$ where at least one x_j and one y_k are non-zero, where $1 \leq j, k \leq d_n$. Using Theorem 6.3.5, one thus obtains for some $C > 0$

$$\sup_{\lambda_n/n \leq t \leq 1-\lambda_n/n} \big|w(t)^{-1}\big(\mathbf{M}_t^{(n)}\big)^t \widehat{\mathbf{\Gamma}}^{-1}_{\mathbf{S}^{(n)}} \mathbf{M}_t^{(n)}\big| \geq \mathcal{O}_P\big(n^2 d_n^{-2}\big) + C n^2 d_n^{-2},$$

hence the claim follows. □

Proof of Lemma 6.5.3. Let

$$\mathrm{adj}\big(\mathbf{A}\big) = \big(\widetilde{\mathbf{A}}\big)^T,$$

be the complementary matrix, where $\widetilde{\mathbf{A}} = \big(\widetilde{a}_{i,j}\big)_{1 \leq i,j \leq d}$ and $\widetilde{a}_{i,j}$ denotes the cofactors of the matrix $\mathbf{A}$. The cofactors satisfy

$$\widetilde{a}_{i,j} = (-1)^{i+j} M_{i,j},$$

where $\big(M_{i,j}\big)_{1 \leq i,j \leq d}$ stands for the minors of the matrix $\mathbf{A}$, and are computed via the subdeterminants. It then holds that

$$\mathbf{A}^{-1} = \mathrm{adj}\big(\mathbf{A}\big) / \det(\mathbf{A}). \tag{6.5.10}$$

This formula can be found in any textbook about linear algebra. We also recall the following property of the determinant $\det(\mathbf{A})$ of the matrix $\mathbf{A}$.

$$\det(\mathbf{A}) = \sum_{\sigma \in \mathcal{S}_d} \mathrm{sign}(\sigma) \prod_{i=1}^{d} a_{i,\sigma(i)},$$

where $\mathcal{S}_d$ denotes the set of all permutations σ, and $\text{sign}(\sigma) = 1$ (-1) if σ is even (uneven). For regular matrixes $\mathbf{A}, \mathbf{B}$, denote with $\mathbf{A}^{-1} = \big(a^*_{i,j}\big)_{1\leq i,j\leq d}$, $\mathbf{B}^{-1} = \big(b^*_{i,j}\big)_{1\leq i,j\leq d}$ their inverses. We obtain from (6.5.10) that

$$\big|a^*_{i,j} - b^*_{i,j}\big| \leq \left|\frac{\widetilde{a}^*_{j,i}}{\det(\mathbf{A})} - \frac{\widetilde{b}^*_{j,i}}{\det(\mathbf{B})}\right|.$$

Since we have that $|\mathcal{S}_d| = d!$, we get

$$\big|\det(\mathbf{A}) - \det(\mathbf{B})\big| \leq d!d \max_{1\leq i,j\leq d}\big|a_{i,j} - b_{i,j}\big| \leq d!d\epsilon.$$

hence we conclude

$$\big|\widetilde{a}^*_{j,i} - \widetilde{b}^*_{j,i}\big| \leq d!d\epsilon,$$

and

$$\begin{aligned}\big|a^*_{i,j} - b^*_{i,j}\big| &\leq \frac{\big|\det(\mathbf{B})(a^*_{i,j} - b^*_{i,j})\big| + \big|b^*_{i,j}(\det(\mathbf{A}) - \det(\mathbf{B}))\big|}{\big|\det(\mathbf{A})^2 - |\det(\mathbf{A})d!d\epsilon|\big|} \\ &\leq \frac{d!d\epsilon\big(\big|\det(\mathbf{A})\big| + d!d\epsilon\big) + \big|b^*_{i,j}(\det(\mathbf{A}) - \det(\mathbf{B}))\big|}{\big|\det(\mathbf{A})^2 - |\det(\mathbf{A})d!d\epsilon|\big|} \\ &\leq \frac{d!d\epsilon\big(\big|\det(\mathbf{A})\big| + d!d\epsilon\big) + a^*_{i,j}d!d\epsilon + d!d\epsilon\big|a^*_{i,j} - b^*_{i,j}\big|}{\big|\det(\mathbf{A})^2 - |\det(\mathbf{A})d!d\epsilon|\big|}.\end{aligned}$$

Since $|a^*_{i,j}| \leq \max|\mathbf{A}^{-1}|$ we thus obtain

$$\big|a^*_{i,j} - b^*_{i,j}\big|\left(1 - \frac{d!d\epsilon}{\big|\det(\mathbf{A})^2 - |\det(\mathbf{A})d!d\epsilon|\big|}\right) \leq \frac{d!d\epsilon\big(\big|\det(\mathbf{A})\big| + d!d\epsilon\big) + \max|\mathbf{A}^{-1}|d!d\epsilon}{\big|\det(\mathbf{A})^2 - |\det(\mathbf{A})d!d\epsilon|\big|},$$

which completes the proof. □

Proposition 6.5.6. *Let $\big\{X_k\big\}_{k\in\mathbb{Z}}$ be an AR(q) process with parameter $\boldsymbol{\zeta} = \big(1, \zeta_1, ..., \zeta_{d_n}\big)$, such that $\sum_{j=1}^q |\zeta_i| \leq \vartheta < 1$. Denote with $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ the eigenvalues of the covariance matrix Γ_n of $\big(X_1, ..., X_n\big)^T$. Then we have that*

$$\frac{\sigma^2}{(1+\vartheta)^2} \leq \lambda_1 \leq \lambda_n \leq \frac{\sigma^2}{\vartheta^2}.$$

In order to proof Proposition 6.5.6, we require the following auxiliary result.

Proposition 6.5.7. *Let $\{X_k\}_{k\in\mathbb{Z}}$ be an AR(q) process with parameter $\boldsymbol{\zeta} = (1, \zeta_1, ..., \zeta_{d_n})$, such that $\sum_{j=1}^q |\zeta_i| \leq \vartheta < 1$. Then we have that*

$$\sup_\lambda f_X(\lambda) \leq \frac{\sigma^2}{2\pi\vartheta^2}, \quad \inf_\lambda f_X(\lambda) \geq \frac{\sigma^2}{2\pi(1+\vartheta)^2},$$

where $\sigma^2 = \mathbb{E}(\epsilon^2)$.

Proof of Proposition 6.5.7. It is easy to show that $\{X_k\}_{k\in\mathbb{Z}}$ can be represented as $X_k = \sum_{i=0}^\infty \psi_i \epsilon_{k-i}$. Since the innovations $\{\epsilon_k\}_{k\in\mathbb{Z}}$ have a spectral density function $f_\epsilon(\lambda) = \sigma^2/(2\pi)$, we obtain from [25, Theorem 4.4.1] that the density function $f_X(\lambda)$ exists, and by [25, Theorem 4.4.2] $f_X(\lambda)$ is given as

$$f_X(\lambda) = \frac{\sigma^2}{2\pi\big|\underline{\zeta}\big(e^{-\mathrm{i}\lambda}\big)\big|^2},$$

where $\underline{\zeta}(s) = 1 - \sum_{j=1}^q \zeta_j s^j$. Since $\underline{\zeta}\big(e^{-\mathrm{i}\lambda}\big) = 1 - \sum_{j=1}^q \zeta_i e^{-\mathrm{i}\lambda j}$, it holds that

$$\big|1 - \sum_{j=1}^q |\zeta_j|\big|^2 \leq \big|\zeta\big(e^{-\mathrm{i}\lambda}\big)\big|^2 \leq \big|1 + \sum_{j=1}^q |\zeta_j|\big)\big|^2,$$

hence the claim follows. □

Proposition 6.5.6 can now readily be deduced.

Proof of Proposition 6.5.6. By [25, Proposition 4.5.3], we have that

$$2\pi \inf_\lambda f_X(\lambda) \leq \lambda_1 \leq \lambda_n \leq 2\pi \sup_\lambda f_X(\lambda),$$

hence the claim follows from Proposition 6.5.7. □

6.5.2 Proofs of Section 6.4

We first give the proof of Theorem 6.4.1.

Proof of Theorem 6.4.1. By Theorem 6.5.1 we can approximate $\max_{1\leq h\leq d_n} \psi_h^{-1} Z_h^{(n,l)}$ with $\max_{1\leq h\leq d_n} B_h$, such that the error has magnitude $\mathcal{O}_P(n^{-\epsilon})$. Since $u_n(z)$ is defined up to an error of magnitude $\mathcal{o}\left((\log n)^{-1/2}\right)$, we thus obtain

$$P\big(\max_{1\leq h\leq d_n} \psi_h^{-1} Z_h^{(n,l)} \leq u_n(z)\big) \leq \mathcal{o}\big(1\big) + P\big(\max_{1\leq h\leq d_n} B_h \leq u_n(z')\big),$$

where $z' < z$. Similarly, one obtains a lower bound with $z'' > z$. The claim will follow from Theorem 6.5.2 if we can show that the conditions of class C2 are verified for $\max_{1\le h\le d_n} B_h$. Setting $\delta < \gamma$ in Theorem 6.6.4 and evaluating the set of inequalities (A)-(G) in Theorem 6.6.4, one obtains that $\max\big|\mathbf{\Gamma}_{\mathbf{M}}^{(n)} - \mathbf{\Gamma}_{\mathbf{B}}^{(n)}\big| = \mathcal{O}\big(n^{-\delta}\big)$ and the two conditions (6.4.1) and (6.4.2). This implies that conditions C2 are satisfied. We thus obtain for large enough n

$$\exp(-z'') \le P\big(\max_{1\le h\le d_n} \psi_h^{-1} Z_h^{(n,l)} \le u_n(z)\big) \le \exp(-z').$$

Choosing appropriate subsequences $z_n' \downarrow z$, $z_n'' \uparrow z$, we conclude that

$$\lim_{n\to\infty} P\big(\max_{1\le h\le d_n} \psi_h^{-1} Z_h^{(n,l)} \le u_n(z)\big) = \exp(z).$$

□

In order to proof Theorem 6.4.5, we need to show that the error

$$\max_{1\le h\le d_n} \big|\big(\psi_h^{-1} - \widehat{\psi}_h^{-1}\big) Z_h^{(n,l)}\big| \tag{6.5.11}$$

is sufficiently small in probability, since then the claim follows from Theorem 6.4.1.

Proof of Theorem 6.4.5. It holds that

$$\big|\big(\psi_h^{-1} - \widehat{\psi}_h^{-1}\big) Z_h^{(n,l)}\big| = \left|\frac{Z_h^{(n,l)}}{\psi_h} \frac{\widehat{\psi}_h^2 - \psi_h^2}{\widehat{\psi}_h\big(\widehat{\psi}_h + \psi_h\big)}\right| \le \left|\frac{Z_h^{(n,l)}}{\psi_h} \frac{\widehat{\psi}_h^2 - \psi_h^2}{\widehat{\psi}_h^2}\right|.$$

In addition, for $0 < \epsilon_n \le 1$ we have

$$\begin{aligned} P\big(\max_{1\le h\le d_n} \big(|\widehat{\psi}_h^2 - \psi_h^2|\widehat{\psi}_h^{-2}\big) \ge \epsilon_n\big) &\le P\big(\max_{1\le h\le d_n} \big(|\widehat{\psi}_h^2 - \psi_h^2|(1+\epsilon_n)\psi_h^{-2}\big) \ge \epsilon_n\big) \\ &\le P\big(\max_{1\le h\le d_n} \big(|\widehat{\psi}_h^2 - \psi_h^2|\psi_h^{-2}\big) \ge \epsilon_n/2\big), \end{aligned}$$

which due to Remark 6.2.3 is bounded by $P\big(\max_{1\le h\le d_n} |\widehat{\psi}_h^2 - \psi_h^2| \ge C\,\epsilon\big)$, for some $C > 0$, which does not depend on n or h. Choosing $\epsilon_n = (\log n)^{-\alpha}$, $\alpha > 1$, we thus obtain from Assumption 6.4.4

$$P\big(\max_{1\le h\le d_n} \big(|\widehat{\psi}_h^2 - \psi_h^2|\widehat{\psi}_h^{-2}\big) \ge \epsilon_n\big) = \mathcal{O}(1). \tag{6.5.12}$$

In addition, it follows from the above that for some $\delta > 1/2$

$$\begin{aligned} P\big(\max_{1\le h\le d_n} \big|\big(\psi_h^{-1} - \widehat{\psi}_h^{-1}\big) S_{n,h}\big| > (\log n)^{-\delta}\big) &\le P\big(\max_{1\le h\le d_n} \big|Z_h^{(n,l)}\psi_h^{-1}\big| \ge (\log n)^{-\delta+\alpha}\big) \\ &+ P\big(\max_{1\le h\le d_n} \big(|\widehat{\psi}_h^2 - \psi_h^2|\widehat{\psi}_h^{-2}\big) \ge (\log n)^{-\alpha}\big). \end{aligned}$$

Since $\alpha > 1$, we can choose a $\delta > 1/2$ such that $\alpha - \delta > 1/2$. Then one readily deduces from Theorem 6.4.1

$$P\big(\max_{1\leq h\leq d_n} \big|Z_h^{(n,l)}\psi_h^{-1}\big| \geq (\log n)^{-\delta+\alpha}\big) = \scriptstyle\mathcal{O}(1). \tag{6.5.13}$$

Using (6.5.12), we thus obtain

$$P\big(\max_{1\leq h\leq d_n} \big|\big(\psi_h^{-1} - \widehat{\psi}_h^{-1}\big) Z_h^{(n,l)}\big| > (\log n)^{-\delta}\big) = \scriptstyle\mathcal{O}(1), \quad \delta > 1/2, \tag{6.5.14}$$

and hence the claim follows from Theorem 6.4.1. □

In order to proof Theorem 6.4.8, we need to validate Assumption 6.2.2. We will do so in a series of lemmas. To this end, for $m \geq 0$ and $k \geq l \geq 0$, let

$$T_m = \phi_m\phi_{m+k-l} + \phi_{m+k}\phi_{m-l} + (\eta - 3)\sum_{i=0}^{\infty} \alpha_i\alpha_{i+k}\alpha_{i+m}\alpha_{i+m+l}, \tag{6.5.15}$$

with the convention that $\alpha_i = 0$ for $i < 0$, and $\phi_m = \phi_{|m|}$, if $m < 0$. We then have the following Lemmas.

Lemma 6.5.8. *Suppose that* $0 < \|\epsilon\|_4 < \infty$. *Then*

(i) if $\sum_{i\geq 0} |\alpha_i| < \infty$, *then* $\mathbb{E}\big(S_{n,k}S_{n,l}\big) = \sum_{m=-\infty}^{\infty} T_m + \scriptstyle\mathcal{O}(1)$.

(ii) if $\sum_{i\geq 0} i|\alpha_i| < \infty$, *then* $\mathbb{E}\big(S_{n,k}S_{n,l}\big) = \sum_{m=-\infty}^{\infty} T_m + \mathcal{O}\left(n^{-1/2}d_n^{1/2}\right)$.

In particular, it holds that $\inf_{h\geq L}\psi_h > 0$, *for some finite* $L \geq 0$.

Lemma 6.5.9. *Suppose that*

(i) $\big\|\epsilon\big\|_4 < \infty$,

(ii) $\sum_{i=0}^{\infty} i^3\log(1+i)^{\lambda}|\alpha_i| < \infty$, $\lambda > 0$.

Then

$$\limsup_{n\to\infty} (\log n)^2 \left(\sum_{r=\sqrt{\log n}}^{d_n} \sup_{r\leq|i-j|} \big|\rho_{i,j}^{(n)}\big|\right) = 0.$$

Lemma 6.5.10. *Suppose that* $\|\epsilon\|_4 < \infty$, *and that*

(i) $\sum_{m=0}^{\infty} |\phi_m| < \infty$,

(ii) $\limsup_{h\to\infty} \sup_{m\geq L} \left|\frac{\phi_{m+h}}{\phi_m}\right| < 1$, *for some finite* $L > 0$.

Then we have

$$\limsup_{n\to\infty} \sup_{i,j\geq M_0\colon 1\leq|i-j|} \left|\rho_{i,j}^{(n)}\right| < 1,$$

for some finite $M_0 > 0$.

Remark 6.5.11. Note that since $\lim_{m\to\infty} |\phi_m| = 0$, one can actually assume that $L = 0$ in condition (ii) of Lemma 6.5.10.

Lemma 6.5.12. *Let* $\left\|\epsilon\right\|_{12} < \infty$, *and assume that*

(i) $\sum_{i=0}^{\infty} i^{\lambda}|\alpha_i| < \infty$, *for some* $\lambda \geq 2$.

(ii) $m = m_n = n^{\theta} \geq 2d_n$, $\theta > 0$.

Then

$$\limsup_{n\to\infty} \max_{1\leq h\leq d_n} \max_{1\leq l\leq n} \left\| \sum_{j=1}^{l} Y_{j,h}^{(>m_n)} \right\|_6 = \mathcal{O}\left(n^{-(4\lambda-2)\theta+1}\right).$$

We are now ready to proof Theorem 6.4.8.

Proof of Theorem 6.4.8. The assumptions of Theorem 6.4.8 together with Lemma 6.5.12 and the bound given in (6.5.19) imply that Theorem 6.5.1 is applicable. Hence, for any finite $M_0 > 0$ we have

$$P\left(\max_{1\leq h\leq M_0} \psi_h^{-1} Z_h^{(n,l)} \leq u_n(z)\right) = \mathcal{o}(1).$$

Moreover, discarding the first M_0 elements has no effect on the limit distribution, and thus it suffices to establish Assumption 6.2.2 (iv) for $\min\{k, l\} \geq M_0$, where M_0 is finite. This, however, follows directly from Lemma 6.5.10. In addition, Assumption 6.2.2 (v) is validated by Lemma 6.5.9. Setting $\theta = \delta$, $p = 6$ and explicitly evaluating the bounds (A) - (E), (G) in the proof of Theorem 6.6.4 and taking (6.6.17) also into account (instead of (F)), leads to the inequalities

$$\delta < \frac{1}{5}\left(\sqrt{154} - 12\right) \approx 0.08193, \text{ and } \lambda > \frac{2\sqrt{154} - 19}{4(\sqrt{154} - 48)} \approx 3.5512.$$

In case of $d_n = \mathcal{O}\left((\log n)^{\gamma}\right)$, setting $\lambda \geq 2$ and $p = 6$, we obtain from Lemma 6.5.12 that $\theta \leq 1/6$. One readily verifies that the conditions of Theorem 6.4.1 are satisfied, which yields the claim. $\square$

Proof of Lemma 6.5.8. Put $R_{n,k} = n^{-1/2} \sum_{k=1}^{n} \big(L_k L_{k+h} - \phi_h\big)$.

An application of the Cauchy-Schwarz inequality yields

$$\mathbb{E}\big(S_{n,k} S_{n,l}\big) = \mathbb{E}\big(R_{n,k} R_{n,l}\big) + \mathcal{O}\left(n^{-1/2} d_n^{1/2}\right),$$

hence it suffices to consider $\mathbb{E}\big(R_{n,k} R_{n,l}\big)$. Due to [25, section 7.2], it holds that

$$\mathbb{E}\big(R_{n,k}, R_{n,l}\big) = \sum_{|m|<n} \frac{n - |m|}{n} T_m, \tag{6.5.16}$$

in particular, we have

$$\varpi_{k,l} = \lim_{n\to\infty} \mathbb{E}\big(R_{n,k}, R_{n,l}\big) = (\eta - 3)\phi_k \phi_l + \sum_{m=-\infty}^{\infty} \big(\phi_m \phi_{m+k-l} + \phi_{m+k}\phi_{m-l}\big). \tag{6.5.17}$$

Moreover, we have the decomposition

$$\sum_{|m|\leq n} \frac{n-m}{n} T_m = \sum_{m=-\infty}^{\infty} T_m + \sum_{|m|>n} T_m + n^{-1} \sum_{|m|\leq n} m T_m.$$

Note that for $k \geq l$, we have

$$\begin{aligned}\sum_{|m|>K} \big|T_m\big| &\leq C \left\{ \sum_{|m|>K} |\alpha_m| + \sum_{|m|>K} \big(|\phi_m| + |\phi_{m+k}|\big) \right\} \\ &\leq C \left\{ \sum_{i>K} |\alpha_i| + \sum_{|m|>K} |\phi_m| \right\} := \theta_K,\end{aligned}$$

with $\lim_{K\to\infty} \theta_K = 0$. Thus we get the estimate

$$n^{-1} \sum_{|m|\leq n} m|T_m| \leq \lambda \sum_{|m|\leq \lambda n} |T_m| + \sum_{|m|>n\lambda} |T_m| = \mathcal{O}\left(\theta_{\lambda n} + \lambda\right),$$

hence choosing an appropriate sequence $\lambda_n > 0$ with $\lim_{n\to\infty} \lambda_n = 0$, we conclude

$$\sum_{|m|\leq n} \frac{n-m}{n} T_m = \sum_{m=-\infty}^{\infty} T_m + o(1), \tag{6.5.18}$$

which gives (i). In order to establish (ii), note that

$$\sum_{j=0}^{\infty} j|\phi_j| \leq \sum_{j=0}^{\infty} j \sum_{i=0}^{\infty} |\alpha_i \alpha_{i+j}| \leq C \sum_{i=0}^{\infty} i|\alpha_i|, \tag{6.5.19}$$

which implies $\theta_K = \mathcal{O}(K^{-1})$. Hence choosing $\lambda = \lambda_n = n^{-1/2}$ gives

$$n^{-1} \sum_{|m| \le n} m|T_m| = \mathcal{O}\left(n^{-1/2}\right),$$

which implies (ii). Finally note that since $\lim_{m\to\infty} |\phi_m| = 0$, we have

$$\limsup_h \sup_{k \ge h} \Big| \sum_{m=-\infty}^{\infty} T_m - \sum_{m=-\infty}^{\infty} \phi_m^2 \Big| = 0, \tag{6.5.20}$$

hence $\inf_{h \ge L} \psi_h > 0$, for some finite $L \ge 0$. □

Proof of Lemma 6.5.9. Due to Lemma 6.5.8, we can assume that

$$\mathbb{E}\big(S_{n,k}, S_{n,l}\big) = (\eta - 3)\phi_k\phi_l + \sum_{m=-\infty}^{\infty} \big(\phi_m\phi_{m+k-l} + \phi_{m+k}\phi_{m-l}\big), \tag{6.5.21}$$

since $(\log n)^2 d_n^2 n^{-1/2} = o(1)$. Condition (ii) implies that $\alpha_i = o(i^{-3})$, and since $\int_1^\infty (y(y+m))^{-3}\, dy = \mathcal{O}(m^{-3})$, we conclude that $\phi_m = o(m^{-3})$. Using this, similar computations yield that

$$\mathbb{E}\big(S_{n,k}, S_{n,l}\big) = o\left(|k-l|^{-3}\right). \tag{6.5.22}$$

Since $\inf_{h \ge L} \psi_h > 0$, for some finite $L > 0$, we obtain that

$$(\log n)^2 \sum_{r=\sqrt{\log n}}^{d_n} \sup_{r \le |i-j|} \big|\rho_{i,j}^{(n)}\big| = o\left((\log n)^2)^2 \sum_{r=\sqrt{\log n}}^{\infty} r^{-3}\right) = o(1), \tag{6.5.23}$$

which completes the proof. □

Proof of Lemma 6.5.10. Due to Lemma 6.5.8, we can assume that

$$\mathbb{E}\big(S_{n,k}, S_{n,l}\big) = (\eta - 3)\phi_k\phi_l + \sum_{m=-\infty}^{\infty} \big(\phi_m\phi_{m+k-l} + \phi_{m+k}\phi_{m-l}\big). \tag{6.5.24}$$

Now suppose first that $\min\{k, l\} \ge M_0$, for some $M_0 > 0$. Thus, since $\lim_{m\to\infty} |\phi_m| = 0$, we have from (6.5.24) that

$$\varpi_{k,l} = \sum_{m=-\infty}^{\infty} \phi_m\phi_{m+k-l} + \epsilon^{(M_0)}, \tag{6.5.25}$$

where $\epsilon^{(M_0)} \downarrow 0$ as M_0 increases. Condition (ii) and Remark 6.5.11 imply that for some $K \geq L$ large enough, we have

$$\sup_m \left| \frac{\phi_{m+h}}{\phi_m} \right| \leq \vartheta^{(K)} < 1, \ K \leq h. \tag{6.5.26}$$

Hence we obtain

$$\sqrt{\varpi_{k,k}\varpi_{l,l}} \geq \sum_{m=-\infty}^{\infty} \phi_m^2 - \mathcal{O}\left(\epsilon^{(M_0)}\right), \tag{6.5.27}$$

and

$$\sum_{m=-\infty}^{\infty} \phi_m^2 - \sum_{m=-\infty}^{\infty} \phi_m \phi_{m+h} \geq (1 - \vartheta^{(K)}) \sum_{m=-\infty}^{\infty} \phi_m^2.$$

Hence, for large enough but fixed M_0 and K, we deduce that that the Cauchy-Schwarz inequality is strict, i.e.

$$\left|\rho_{i,j}^{(n)}\right| < 1. \tag{6.5.28}$$

Now suppose that $h < K$. Then the Cauchy-Schwarz inequality (in l_2) implies that

$$\left| \sum_{m=-\infty}^{\infty} \phi_m \phi_{m+h} \right| \leq \sum_{m=-\infty}^{\infty} \phi_m^2,$$

and we have equality if and only if $v_1 = \lambda v_2$, $\lambda \in \mathbb{R}$ and $v_1 = (..., \phi_m, ...)^T$, $v_2 = (...\phi_{m+h}, ...)^T$. This implies that $\phi_m = \lambda \phi_{m+h}$, and consequently

$$\phi_0 = \lambda \phi_h = \lambda^2 \phi_{2h} = ... = \lambda^n \phi_{nh}. \tag{6.5.29}$$

Since $|\phi_{nh}| \to 0$, we must have $|\lambda| > 1$. We thus conclude that

$$\left| \sum_{m=-\infty}^{\infty} \phi_m - \sum_{m=-\infty}^{\infty} \phi_m \phi_{m+h} \right| = |\lambda - 1| \sum_{m=-\infty}^{\infty} \phi_m^2 > 0.$$

Since K is finite, we deduce that

$$\min_{1 \leq h < K} \left| \sum_{m=-\infty}^{\infty} \phi_m - \sum_{m=-\infty}^{\infty} \phi_m \phi_{m+h} \right| = \epsilon^{(K)} > 0,$$

which together with (6.5.27) implies that for large enough (but finite) M_0, we have (6.5.28), which completes the proof. □

Proof of Lemma 6.5.12. Let

$$L_k = L_k^{(\leq m)} + L_k^{(>m)} = \sum_{i\leq m}^{\infty} \alpha_i \epsilon_{k-i} + \sum_{i>m}^{\infty} \alpha_i \epsilon_{k-i}.$$

Then

$$Y_{k,h}^{(>m)} = L_k^{(>m)} L_{k+h}^{(\leq m)} + L_k^{(\leq m)} L_{k+h}^{(>m)} + L_k^{(>m)} L_{k+h}^{(>m)} - \mathbb{E}\big(L_k^{(>m)} L_{k+h}^{(>m)}\big).$$

Lengthy and tedious calculations show that

$$\max_{1\leq l\leq n} \left\| \sum_{k=1}^{l} Y_{k,h}^{(>m)} \right\|_6^6 = \mathcal{O}\left(\left\| \sum_{k=1}^{n} Y_{k,h}^{(>m)} \right\|_2^6 \right).$$

The Minikowski and Cauchy-Schwarz inequality imply

$$\left\| \sum_{k=1}^{l} Y_{k,h}^{(>m)} \right\|_2 \leq \sum_{k=1}^{n} \big\| Y_{k,h}^{(>m)} \big\|_2 \leq C \sum_{k=1}^{n} \left(\big\| L_k^{(>m)} \big\|_4^2 + \big\| L_{k+h}^{(>m)} \big\|_4^2 \right).$$

Evaluating this bound using $m \geq 2d_n$ and $\limsup_{n\to\infty} \big|\alpha_n n^\lambda\big| < \infty$ yields

$$n^{-1} \left\| \sum_{k=1}^{l} Y_{k,h}^{(>m)} \right\|_2 = \mathcal{O}\left(\sqrt{ \sum_{i>m-h} \alpha_i^4 + \left(\sum_{i>m-h} \alpha_i^2 \right)^2 } \right) = \mathcal{O}\left(m^{-(4q-2)} \right) = \mathcal{O}\left(n^{-(4\lambda-2)\theta} \right).$$

Hence the claim follows. □

6.6 Gaussian approximation

Let $\{X_{k,h}\}_{k,h\geq 1}$ be a collection of random variables such that for each h_0, $\{X_{k,h_0}\}_{k\geq 1}$ is a zero mean stationary sequence.

Throughout the proofs, C denotes a generic constant that may vary from one formula to another. Recall the notation

$$Y_{k,h}^{(\leq m)} = \mathbb{E}\big(X_{k,h} \mid \mathcal{F}_{k-m}^{k+m}\big), \tag{6.6.1}$$

$$Y_{k,h}^{(>m)} = X_{k,h} - Y_{k,h}^{(\leq m)} = X_{k,h} - \mathbb{E}\big(X_{k,h} \mid \mathcal{F}_{k-m}^{k+m}\big). \tag{6.6.2}$$

The Gaussian approximation is obtained under the following Assumption.

Assumption 6.6.1. *For $m = m_n = n^\theta$, $0 < \theta < 1$, $d = d_n = n^\delta$, $0 < \delta$, $\psi_h > 0$ we suppose that*

$$(i)\ \limsup_{n\to\infty} \max_{1\le h\le d_n} \left\| Y_{k,h}^{(\le m)} \right\|_p < \infty,$$

$$(ii)\ \limsup_{n\to\infty} \max_{1\le h\le d_n} \max_{1\le l\le n} \left\| \sum_{j=1}^{l} Y_{j,h}^{(>m_n)} \right\|_p = \mathcal{O}(1),\ p > 8,$$

$$(iii)\ \limsup_{n\to\infty} \max_{1\le h\le d_n} \left| Var\left(\sum_{k=1}^{n} X_{k,h} \right) - \psi_h n \right| < \infty.$$

Remark 6.6.2. If the above assumptions hold for some $m = m_n = (\log n)^\lambda$, one can set $\theta = 0$ in all the conditions given below that involve θ.

Lemma 6.6.3. *Suppose that* $\sup_h \sum_{j=0}^{\infty} j|\phi_{j,h}| < \infty$. *Then Assumption 6.6.1 (iii) holds, and* $\psi_h = \phi_{0,h} + 2\sum_{j=1}^{\infty} \phi_{j,h}$.

Proof of Lemma 6.6.3. We have

$$\begin{aligned}\mathrm{Var}\big(\sum_{i=1}^{n} X_{i,h}\big) &= \sum_{1\le i,j} \phi_{|i-j|,h} = \sum_{i=1}^{k} \sum_{j=1-i}^{k-i} \phi_{|j|,h} = \psi_h + \mathcal{O}\left(\sum_{i=0}^{\infty} \sum_{j=i}^{\infty} |\phi_{j,h}| \right) \\ &= \psi_h + \mathcal{O}\left(\sup_h \sum_{j=0}^{\infty} j|\phi_{j,h}| < \infty \right) = \psi_h + \mathcal{O}(1).\end{aligned}$$

□

For a d_n-dimensional Brownian motion $\{\mathbf{W}_\mathbf{t}^{(\mathbf{n})}\}_{t\ge0} = \{W_{t,h}^{(n)}\}_{\substack{t\ge0,\\ 0\le h\le d_n}}$, we denote the covariance matrix with $\mathbf{\Gamma}_\mathbf{W}^{(n)}$, and similarly, we write $\mathbf{\Gamma}_\mathbf{S}^{(n)}$ for the covariance matrix of the vector $n^{-1/2}\mathbf{S}^{(n)}$.

The main Theorem is formulated below.

Theorem 6.6.4. *Suppose that Assumption 6.6.1 is valid. Then for each n and $\nu \ge 2$, on a possible larger probability space, there exists a d_n-dimensional Brownian motion $\{\mathbf{W}_\mathbf{t}^{(\mathbf{n})}\}_{t\ge0} = \{W_{t,h}^{(n)}\}_{\substack{t\ge0,\\ 0\le h\le d_n}}$ such that for some $q > 1$*

$$P\left(\max_{0\le h\le d_n} \max_{1\le i\le n} \left| \sum_{k=1}^{i} X_{k,h} - \psi_h W_{i,h}^{(n)} \right| \ge n^{1/\nu} \right) = \mathcal{O}\left(n^{-q}\right),$$

where $0 < \psi_h^2 = \lim_n n^{-1}\, Var\big(\sum_{1\le i\le n} X_{k,h}\big) < \infty$, *and*

$$\theta \le \delta < \min\left\{ \frac{4(p-2\nu)}{(-4+3p)\nu}, \frac{2+p-2(1+\theta)\nu}{(2+4\theta+p(4+\theta))\nu} \right\},$$

where we require $p > 4\nu$. In addition, we have that $\max\big|\mathbf{\Gamma}_{\mathbf{W}}^{(n)} - \mathbf{\Gamma}_{\mathbf{S}}^{(n)}\big| = \mathcal{O}\big(n^{-\gamma}\big)$, *for some $\gamma > 0$. Alternatively, if one sets $d_n = \mathcal{O}\big((\log n)^{\delta}\big)$, for arbitrary $\delta > 0$, then we require*

$$\nu < \min\left\{\frac{4+2p}{4+4\theta+p\theta+4\theta^2+p\theta^2}, \frac{2}{1-2\theta}, p/4\right\}.$$

Remark 6.6.5. Note that by setting $\theta = 0$ and letting $p \to \infty$, we obtain the upper bound $\delta < 1/9$. In addition, we point out that conditions (i),(ii) of Assumption 6.6.1 can be weakened too $p > 4$, which however leads to a less tractable bound for δ.

Based on this result, we can derive the following two Theorems.

Theorem 6.6.6. *Assume that the assumptions of Theorem 6.6.4 hold. Then, for each n, we can define two independent d_n-dimensional Brownian motions* $\big\{\mathbf{W}_{t,h}^{(1,n)}\big\}_{t\geq 0} = \big\{W_{t,h}^{(1,n)}\big\}_{\substack{t\geq 0,\\ 0\leq h\leq d_n}}$, $\big\{\mathbf{W}_{t,h}^{(2,n)}\big\}_{t\geq 0} = \big\{W_{t,h}^{(2,n)}\big\}_{\substack{t\geq 0,\\ 0\leq h\leq d_n}}$

$$P\left(\max_{0\leq h\leq d_n}\sup_{1\leq x\leq n/2}\left|\sum_{1\leq i\leq x} X_{i,h} - \psi_h W_{x,h}^{(n)}\right|/x^{1/\nu}\right) = \mathcal{O}(1),$$

and

$$P\left(\max_{0\leq h\leq d_n}\sup_{1\leq x\leq n/2}\left|\sum_{n-x\leq i\leq n} X_{i,h} - \psi_h W_{x,h}^{(n)}\right|/x^{1/\nu}\right) = \mathcal{O}(1),$$

with $\nu > 2$.

Theorem 6.6.7. *Assume that the assumptions of Theorem 6.6.4 hold, and let $\lambda_n = o(n)$ be a monotone increasing sequence. Then, for each n, we can define a d_n-dimensional Brownian Bridge* $\big\{\mathbf{B}_t^{(n)}\big\}_{t\geq 0} = \big\{B_{t,h}^{(n)}\big\}_{\substack{0\leq t\leq 1,\\ 0\leq h\leq d_n}}$ *such that*

$$\max_{0\leq h\leq d_n}\sup_{\lambda_n\leq t\,n\leq n-\lambda_n}\frac{\big|M_{t,h}^{(n)} - \psi_h B_{t,h}^{(n)}\big|}{\big(t(1-t)\big)^{1/2}} = \mathcal{O}_P\big(\lambda_n^{-1/2+1/\nu}\big),$$

for $\nu \geq 2$.

Remark 6.6.8. Note that in both Theorems, we still have the relation $\max\big|\mathbf{\Gamma}_{\mathbf{W}}^{(n)} - \mathbf{\Gamma}_{\mathbf{S}}^{(n)}\big| = \mathcal{O}\big(n^{-\gamma}\big)$, for some $\gamma > 0$, for the corresponding Brownian motion.

The proof of Theorem 6.6.4 follows [14, Theorem 4.1] in broad brushes, with some (essential) changes in the details. To this end, we require some preliminary results. The following coupling inequality is due do Berthet and Mason [19].

Lemma 6.6.9 (Coupling inequality)**.** *Let $X_1, ..., X_N$ be independent, mean zero random vectors in $\mathbb{R}^n$, $n \geq 1$, such that for some $B > 0$, $|X_i|_n \leq B$, $i = 1, ..., N$. If the probability space is rich enough, then for each $\delta > 0$, one can define independent normally distributed mean zero random vectors $\xi_1, ..., \xi_N$ with ξ_i and X_i having the same variance/covariance matrix for $i = 1, ..., N$, such that for universal constants $C_1 > 0$ and $C_2 > 0$,*

$$P\Big\{ \big|\sum_{i=1}^{N}(X_i - \xi_i)\big|_n > \delta \Big\} \leq C_1 n^2 \exp\left(-\frac{C_2\delta}{Bn^2}\right).$$

Lemma 6.6.10. *There is an absolute constant C such that*

$$\mathbb{E}\Big|\sum_{l \leq i \leq k} Y_{k,h}^{(\leq m)}\Big|^p \leq C\big((k-l+1)\lceil m+1 \rceil\big)^{p/2}.$$

Proof of Lemma 6.6.10. Put $K = 2\lceil m+1 \rceil$, and denote with $\|.\|_p = (\mathbb{E}|.|^p)^{1/p}$. Then per construction, we can rewrite

$$\sum_{l \leq i \leq k} Y_{k,h}^{(\leq m)} = R_1 + ... + R_K,$$

where R_i is a sum of independent random variables with at most $(k-l+1)/K$ terms. Minikowski's inequality gives us

$$\|R_1 + ... + R_K\|_p \leq \|R_1\|_p + ... + \|R_K\|_p.$$

By Rosenthal's inequality and Assumption 6.6.1 (i), we have

$$\mathbb{E}\big|R_i\big|^p \leq C\big((k-l+1)/K\big)^{p/2} = C\left((k-l+1)/K\right)^{p/2},$$

hence

$$\Big\|\sum_{l \leq i \leq k} Y_{k,h}^{(\leq m)}\Big\|_p^p \leq C\big((k-l+1)K\big)^{p/2}.$$

□

Proof of Theorem 6.6.4. The proof is based on a blocking and truncation argument, which requires us to have numbers $\beta, \delta, \kappa, \theta, p, q, \nu$ that satisfy the following conditions

(A) $\max\{\theta, \delta\} < \beta(\beta+1)^{-1}$,

(B) $\nu^{-1} - \beta(2+2\beta)^{-1} - \kappa - 3\delta > 0$,

(C) $\nu^{-1} - (1-\beta)(2+2\beta)^{-1} > 0$,

(D) $\nu^{-1} - (\beta/4 + 1/p)(1+\beta)^{-1} - \delta/4 - (1+\delta)/p > 0$,

(E) $1 < \nu^{-1} + p\kappa/2 - \delta - \theta\beta(1+\beta)^{-1}(p/4+1) - \theta$,

(F) $p > 4\nu$,

(G) $\beta(2(1+\beta))^{-1} = \gamma + \delta$.

which we will use as reference, and therefore they are not completely simplified. If we fix γ, θ, p, ν, and suppose that $\theta \leq \delta$, then using the above inequalities we obtain

$$\frac{\nu-2}{\nu+2} < \beta < \min\left\{\frac{-2(2+p-2\nu+2\gamma\nu+3\gamma p\nu-2\theta\nu)}{4+2p-6\nu+4\gamma\nu-4p\nu+6\gamma p\nu-8\theta\nu-p\theta\nu}, \frac{-16\nu+8\gamma\nu+8p+2\gamma\nu p}{12\nu-8\gamma\nu+8p-3\nu p+2\gamma\nu p}\right.$$

where $x \vee y = \min(x,y)$ if $x, y \geq 0$, and $x \vee y = y$ if $x < 0$. Using relation (G) one thus obtains a bound for δ. Note that if we just require $\gamma > 0$, then the above simplifies to

$$\frac{\nu-2}{\nu+2} < \beta < \min\left\{\frac{-2(2+p-2\nu-2\theta\nu)}{4+2p-6\nu-4p\nu-8\theta\nu-p\theta\nu}, \frac{-16\nu+8p}{12\nu+8p-3\nu p} \vee 0\right\}.$$

Alternatively, if we set $d_n = \mathcal{O}\big((\log n)^{\delta^*}\big)$, then we may set $\delta = 0$ in (A)-(G), and an evaluation amounts to

$$\nu < \min\left\{\frac{4+2p}{4+4\theta+p\theta+4\theta^2+p\theta^2}, \frac{2}{1-2\theta}, p/4\right\}.$$

We will now construct an approximation for the random variables $R_i^{(h)}$. To this end, we first divide the set of integers $\{1, 2, ...\}$ into consecutive blocks $\mathcal{H}_1, \mathcal{I}_1, \mathcal{H}_2, \mathcal{I}_2 ...$. The blocks are defined by recursion. Fix $\beta > 0$. If the largest element of $\mathcal{I}_{i-1}$ is k_{i-1}, then $\mathcal{H}_i = \{k_{i-1}+1, ..., k_{i-1}+i^{\beta}\}$ and $\mathcal{I}_i = \{k_{i-1}+i^{\beta}+1, ..., k_i\}$, where $k_i = \min\{l : l - (\Upsilon\, d_n) \vee m_n - 1 \geq k_{i-1} + i^{\beta}\}$, for some constant $\Upsilon > 0$, where $x \vee y = \max(x,y)$ for $x, y \in \mathbb{R}$. Let $|.|$ denote the cardinality of a set. It follows from the definition of $\mathcal{H}_i$, $\mathcal{I}_i$ that $|\mathcal{H}_i| = i^{\beta}$ and $|\mathcal{I}_i| \geq d_n + 1$. Note that the total number of blocks is approximately $c_n = n^{1/(1+\beta)}$, due to (A). For $1 \leq h \leq d_n$, let

$$U_{k,h}^{(m,1)} = \sum_{i \in \mathcal{H}_k} Y_{i,h}^{(\leq m)} \quad \text{and} \quad U_{k,h}^{(m,2)} = \sum_{i \in \mathcal{I}_k} Y_{k,h}^{(\leq m)},$$

$$V_{k,h}^{(m,1)} = \sum_{i \in \mathcal{H}_k} Y_{i,h}^{(>m)} \quad \text{and} \quad V_{k,h}^{(m,2)} = \sum_{i \in \mathcal{I}_k} Y_{k,h}^{(>m)},$$

and define the vectors

$$\mathbf{U}_k^{(m,i)} = \big(U_{k,1}^{(m,i)}, U_{k,2}^{(m,i)}, ..., U_{k,d_n}^{(m,i)}\big)^T,$$
$$\mathbf{V}_k^{(m,i)} = \big(V_{k,1}^{(m,i)}, V_{k,2}^{(m,i)}, ..., V_{k,d_n}^{(m,i)}\big)^T, \ i \in \{1,2\},$$

Throughout this proof, we will always assume that $m = m_n = n^\theta$. For a random variable X, let $\mathbf{I}_B^X = \mathbf{1}(X)_{\{|X|\le B\}}$ for $B > 0$, and similarly, $\mathbf{I}_{B^c}^X = 1 - \mathbf{I}_B^X = \mathbf{1}(X)_{\{|X|>B\}}$. In addition, we put $E_B^X = \mathbb{E}\big(X\mathbf{I}_B^X\big)$. Let

$$\xi_{k,h}^{(m)} = U_{k,h}^{(m,1)}\mathbf{I}_{B_n}^{U_{k,h}^{(m,1)}} - E_{B_n}^{U_{k,h}^{(m,1)}}, \quad \eta_{k,h}^{(m)} = U_{k,h}^{(m,2)}\mathbf{I}_{B_n}^{U_{k,h}^{(m,2)}} - E_{B_n}^{U_{k,h}^{(m,2)}},$$

and define the random vectors

$$\boldsymbol{\xi}_j^{(m)} = \big(\xi_{j,1}^{(m)}, \xi_{j,2}^{(m)}, ..., \xi_{j,d_n}^{(m)}\big)^T, \quad \boldsymbol{\eta}_j^{(m)} = \big(\eta_{j,1}^{(m)}, \eta_{j,2}^{(m)}, ..., \eta_{j,d_n}^{(m)}\big)^T.$$

As a first step, we will show that the truncation error is negligible, more precisely, we will show that

$$P\bigg(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\big|\sum_{j=1}^{i} U_{j,h}^{(m,1)} + U_{j,h}^{(m,2)} - \boldsymbol{\xi}_j^{(m)} - \boldsymbol{\eta}_j^{(m)}\big| \ge n^{1/\nu}\bigg) = \mathcal{O}\left(n^{-q}\right). \quad (6.6.3)$$

To this end, let $x > 0$. Then the Markov and Lévy's maximal inequality imply that

$$P\bigg(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\big|\sum_{j=1}^{i} U_{j,h}^{(m,1)} - \xi_{j,h}^{(m)}\big| \ge x\bigg) \le Cx^{-2}d_n \max_{1\le h\le d_n}\sum_{i=1}^{c_n}\big\|U_{i,h}^{(m,1)} - \xi_{i,h}^{(m)}\big\|_2^2.$$

Using the Cauchy-Schwarz inequality, we obtain

$$\max_{1\le h\le d_n}\big\|U_{i,h}^{(m,1)} - \xi_{i,h}^{(m)}\big\|_2^2 \le \big\|U_{i,h}^{(m,1)}\big\|_4^2\big\|\mathbf{I}_B^{U_{i,h}^{(m,1)}}\big\|_4^2 \le \big\|U_{i,h}^{(m,1)}\big\|_4^2\big\|U_{i,h}^{(m,1)}\big\|_p^{p/2}B^{-p/2},$$

which, by Lemma 6.6.10, is of the magnitude $\mathcal{O}\left((m\, i^\beta)^{p/4+1}\right)$. We thus obtain that

$$P\bigg(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\big|\sum_{j=1}^{i} U_{j,h}^{(m,1)} - \xi_{j,h}^{(m)}\big| \ge x\bigg) \le Cx^{-1}d_n\, m^{p/4+1}B^{-p/2}\sum_{i=1}^{c_n}(i^\beta)^{p/4+1}$$
$$= \mathcal{O}\left(x^{-1}d_n\, c_n^{\beta(p/4+4)+1}B^{-p/2}\right).$$

Setting $x = 2n^{1/\nu}$ and $B = B_n = n^\kappa$, we find that relation (E) establishes $\mathcal{O}\left(x^{-1}d_n\, c_n^{\beta(p/4+4)+1}B^{-p/2}\right) = \mathcal{O}\left(n^{-q}\right)$.

By the same argument, one also establishes that

$$P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\left|\sum_{j=1}^{i} U_{j,h}^{(m,2)}\right| \ge x\right) = \mathcal{O}\left(n^{-q}\right), \tag{6.6.4}$$

which together with the previous result gives us (6.6.3).

Note that per construction and relation (A), choosing the constant Υ big enough, we have that $\{\boldsymbol{\xi}_j^{(m)}\}_{j\in\mathbb{N}}$ and $\{\boldsymbol{\eta}_j^{(m)}\}_{j\in\mathbb{N}}$ are sequences of independent random vectors. In addition, we have the bound

$$\left|\boldsymbol{\xi}_j^{(m)}\right|_{d_n} \le d_n\, B_n \quad \left|\boldsymbol{\eta}_j^{(m)}\right|_{d_n} \le d_n\, B_n. \tag{6.6.5}$$

Hence, by Lemma 6.6.9, we can define a sequence of independent normal random vectors $\boldsymbol{\xi}_j^{(m,*)} = \left(\xi_{j,1}^{(m,*)}, \xi_{j,2}^{(m,*)}, ..., \xi_{j,d_n}^{(m,*)}\right)^T$, such that for $x > 0$

$$\begin{aligned} P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\left|\sum_{j=1}^{i}\left(\xi_{j,h}^{(m)} - \xi_{j,h}^{(m,*)}\right)\right| \ge x\right) &\le \sum_{h=1}^{d_n}\sum_{i=1}^{c_n} P\left(\left|\sum_{j=1}^{i}\left(\xi_{j,h}^{(m)} - \xi_{j,h}^{(m,*)}\right)\right| \ge x\right) \\ &= \sum_{h=1}^{d_n}\sum_{i=1}^{c_n} P\left(\left|\sum_{j=1}^{i}\left(\xi_{j,h}^{(m)} - \xi_{j,h}^{(m,*)}\right)\right|_2 \ge x\right) \\ &\le \sum_{h=1}^{d_n}\sum_{i=1}^{c_n} P\left(\left|\sum_{j=1}^{i}\left(\boldsymbol{\xi}_{j,h}^{(m)} - \boldsymbol{\xi}_{j,h}^{(m,*)}\right)\right|_{d_n} \ge x\right) \\ &\le C_1 \sum_{i=1}^{c_n} d_n^3 \exp\left(-\frac{C_2 x}{2 d_n^3 B_n}\right) \\ &\le C_1 c_n d_n^3 \exp\left(-\frac{C_2 x}{2 d_n^3 B_n}\right). \end{aligned}$$

Hence due to (B), we obtain

$$P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\left|\sum_{j=1}^{i}\left(\xi_{j,h}^{(m)} - \xi_{j,h}^{(m,*)}\right)\right| \ge n^{1/\nu}\right) = \mathcal{O}\left(n^{-q}\right), \tag{6.6.6}$$

for $q > 1$. Similar arguments show that under the same conditions as above, there exists a sequence of independent normal random vectors $\boldsymbol{\eta}_j^{(m,*)} = \left(\eta_{j,1}^{(m,*)}, \eta_{j,2}^{(m,*)}, ..., \eta_{j,d_n}^{(m,*)}\right)^T$, such that

$$P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\left|\sum_{j=1}^{i}\left(\eta_{j,h}^{(m)} - \eta_{j,h}^{(m,*)}\right)\right| \ge n^{1/\nu}\right) = \mathcal{O}\left(n^{-q}\right),$$

for $q > 1$. By Lévy's maximal inequality, we have

$$P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\left|\sum_{j=1}^{i}\eta_{j,h}^{(m,*)}\right|\ge n^{1/\nu}\right)\le 2\sum_{h=1}^{d_n}P\left(\left|\sum_{j=1}^{c_n}\eta_{j,h}^{(m,*)}\right|\ge n^{1/\nu}\right).$$

By Lemma 6.6.10, we have that $\operatorname{Var}\big(\eta_{j,h}^{(m,*)}\big)\le Cd_n^2$ for all $j\le c_n$, $h\le d_n$. Hence if (D) holds, by known properties of the tails of a normal cdf, we obtain that

$$P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\left|\sum_{j=1}^{i}\eta_{j,h}^{(m,*)}\right|\ge n^{1/\nu}\right)=\mathcal{O}\left(n^{-q}\right), \tag{6.6.7}$$

for $q > 1$. This yields

$$P\left(\max_{1\le h\le d_n}\max_{1\le k\le c_n}\left|\sum_{j=1}^{i}\big(\xi_{j,h}^{(m)}+\eta_{j,h}^{(m)}-\xi_{j,h}^{(m,*)}\big)\right|\ge n^{1/\nu}\right)=\mathcal{O}\left(n^{-q}\right), \tag{6.6.8}$$

for $q > 1$.

Let $\boldsymbol{\eta}_j^{(m,**)} = \big(\eta_{j,1}^{(m,**)},\eta_{j,2}^{(m,**)},...,\eta_{j,d_n}^{(m,**)}\big)^T$ be an independent copy of $\boldsymbol{\eta}_j^{(m,*)}$ such that $\boldsymbol{\eta}_j^{(m,*)}$ and $\boldsymbol{\xi}_j^{(m,**)}$ are independent. Proceeding as in the proof of [14, Theorem 4.1], by enlarging the probability space if necessary, there exists a d_n-dimensional Brownian motion $\big\{\mathbf{W}_t\big\}_{t\ge 0}=\big\{W_t^{(h)}\big\}_{\substack{t\ge 0,\\ 0\le h\le d_n}}$, such that

$$W_{k_i}^{(h)}=\sum_{1\le j\le i}d_j^{(h)}\big(\xi_{j,h}^{(m,*)}+\eta_{j,h}^{(m,**)}\big),$$

where $d_j^{(h)}$ is chosen such that $\big\|d_j^{(h)}(\xi_{j,h}^{(m,*)}+\eta_{j,h}^{(m,**)})\big\|_2^2=|H_j|+|J_j|$.

We will now establish that

$$d_j^{(h)}=1/\psi_h(1+\mathcal{O}(j^{-\beta/2})). \tag{6.6.9}$$

To this end, note that per construction

$$\big\|\xi_{j,h}^{(m,*)}+\eta_{j,h}^{(m,**)}\big\|_2^2=\big\|\xi_{j,h}^{(m,*)}\big\|_2^2+\big\|\eta_{j,h}^{(m,**)}\big\|_2^2=\big\|\xi_{j,h}^{(m)}\big\|_2^2+\big\|\eta_{j,h}^{(m)}\big\|_2^2.$$

In addition, Assumption 6.6.1 (ii) implies that

$$\begin{aligned}\big|\big\|U_{j,h}^{(m,1)}+V_{j,h}^{(m,1)}\big\|_2-\big\|\xi_{j,h}^{(m)}\big\|_2\big| &\le \big\|\xi_{j,h}^{(m)}-U_{j,h}^{(m,1)}\big\|_2+\big|\big\|V_{j,h}^{(m,1)}\big\|_2\\ &= \big\|\xi_{j,h}^{(m)}-U_{j,h}^{(m,1)}\big\|_2+\mathcal{O}(1).\end{aligned}$$

Moreover, it follows from computations performed when establishing (6.6.3) that $\|\xi_{j,h}^{(m)} - U_{j,h}^{(m,1)}\|_2 = \mathcal{O}(1)$, hence

$$\left|\|U_{j,h}^{(m,1)} + V_{j,h}^{(m,1)}\|_2 - \|\xi_{j,h}^{(m)}\|_2\right| = \mathcal{O}(1).$$

Similarly, one gets that

$$\left|\|U_{j,h}^{(m,2)} + V_{j,h}^{(m,2)}\|_2 - \|\eta_{j,h}^{(m)}\|_2\right| = \mathcal{O}(1).$$

Hence Assumption 6.6.1 (iii) implies that

$$\|(\xi_{j,h}^{(m,*)} + \eta_{j,h}^{(m,**)})\|_2^2 = \psi_H\big(|H_j| + |J_j|\big) + \mathcal{O}(1),$$

and thus (6.6.9) follows. Relation (6.6.9) implies that

$$\operatorname{Var}\left(\sum_{j=1}^{c_n}(1-\psi_h d_j^{(h)})(\xi_{j,h}^{(m,*)} + \eta_{j,h}^{(m,**)})\right) = \mathcal{O}\left(\sum_{j=1}^{c_n} j^{-\beta}\right) = \mathcal{O}\left(n^{(1-\beta)/(1+\beta)}\right).$$

Hence by Lévy's maximal inequality, it follows from (C) that

$$\begin{aligned} P\bigg(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\bigg|\sum_{j=1}^{i}\big(\xi_{j,h}^{(m,*)} + \eta_{j,h}^{(m,**)}\big) - \psi_h W_{k_i}^{(h)}\bigg| \ge n^{1/\nu}\bigg) \\ &\le 2\sum_{h=1}^{d_n} P\bigg(\bigg|\sum_{j=1}^{c_n}\big(\xi_{j,h}^{(m,*)} + \eta_{j,h}^{(m,**)}\big)(1-\psi_h d_j^{(h)})\bigg| \ge n^{1/\nu}\bigg) \\ &\le C d_n P\big(Z_n \ge n^{1/\nu}\big) = \mathcal{O}\left(n^{-q}\right) \end{aligned}$$

for $q > 1$, where Z_n is a mean zero Gaussian random variable with $\operatorname{Var}(Z_n) = \mathcal{O}\big(n^{(1-\beta)/(1+\beta)}\big)$. Next, it is shown that.

$$P\bigg(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\max_{k_i\le l\le k_{i+1}}\bigg|\sum_{j=k_i+1}^{l} Y_{j,h}^{(\le m)}\bigg| > n^{1/\nu}\bigg) = \mathcal{O}\left(n^{-q}\right), \tag{6.6.10}$$

and

$$P\bigg(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\sup_{k_i\le s\le k_{i+1}}\bigg|W_s^{(h)} - W_{k_i}^{(h)}\bigg| > n^{1/\nu}\bigg) = \mathcal{O}\left(n^{-q}\right). \tag{6.6.11}$$

To this end, note that by Lemma 6.6.10 and Móriz et al. [92][Theorem 3.1], it holds that

$$\mathbb{E}\left(\max_{k_i\le l\le k_{i+1}}\bigg|\sum_{j=k_i+1}^{l} Y_{j,h}^{(\le m)}\bigg|^{p/2}\right) \le C\big(k_{i+1}-k_i\big)^{p/4}\big(m+1\big)^{p/4} \tag{6.6.12}$$

$$= \mathcal{O}\left((i^\beta m)^{p/4}\right). \tag{6.6.13}$$

Using the Markov inequality, we thus obtain

$$
\begin{aligned}
P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\max_{k_i\le l\le k_{i+1}}\left|\sum_{j=k_i+1}^{l} Y_{j,h}^{(\le m)}\right| > n^{1/\nu}\right) &\le \sum_{h=1}^{d_n}\sum_{i=1}^{c_n} P\left(\max_{k_i\le l\le k_{i+1}}\left|\sum_{j=k_i+1}^{l} Y_{j,h}^{(\le m)}\right| > n^{1/\nu}\right)\\
&\le C n^{-p/\nu}\sum_{h=1}^{d_n}\sum_{i=1}^{c_n}\left(k_{i+1}-k_i\right)^{p/4}\left(d_n+1\right)^{p/4}\\
&\le C\, n^{-p/\nu}\, d_n^{(p+4)/4}\sum_{i=1}^{c_n} i^{\beta p/4} = \mathcal{O}\left(n^{-p/\nu}c_n^{(\beta p+4)/4}d_n^{(p+}\right.\\
&= \mathcal{O}\left(n^{-p/\nu+(\beta p+4)/(4+4\beta)+\delta(p+4)/4}\right),
\end{aligned}
$$

which proves (6.6.10) due to relation (D). The same argument also of applies to (6.6.11), by replacing the maximal inequality of Móriz et al. [92] by the corresponding results for the increments of the Wiener process in Csörgö and Horváth [32]. Piecing everything together, we obtain that

$$
P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\sup_{k_i\le s\le k_{i+1}}\left|\psi_h W_s^{(h)} - \sum_{j=1}^{i} U_{j,h}^{(m,1)} + U_{j,h}^{(m,2)}\right| > n^{1/\nu}\right) = \mathcal{O}\left(n^{-q}\right). \tag{6.6.14}
$$

Suppose now that

$$
P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\max_{k_i\le l\le k_{i+1}}\left|\sum_{j=0}^{l} Y_{j,h}^{(>m)}\right| > n^{1/\nu}\right) = \mathcal{O}\left(n^{-q}\right). \tag{6.6.15}
$$

This together with (6.6.10) yields

$$
P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\max_{k_i\le l\le k_{i+1}}\left|\sum_{r=0}^{l} X_{r,h} - \sum_{j=1}^{i} U_{j,h}^{(m,1)} + U_{j,h}^{(m,2)}\right| > n^{1/\nu}\right) = \mathcal{O}\left(n^{-q}\right), \tag{6.6.16}
$$

which together with (6.6.14) gives the desired approximation result. Hence we need to verify (6.6.15). To this end, note that Assumption 6.6.1 (ii) implies that

$$
P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\max_{k_i\le l\le k_{i+1}}\left|\sum_{j=0}^{l} Y_{j,h}^{(>m)}\right| > n^{1/\nu}\right) \le C d_n c_n n^{-p/\nu+1}\Lambda_{n,p}, \tag{6.6.17}
$$

where $\Lambda_{n,p} = \max_{1\le l\le n}\left\|\sum_{j=1}^{l} Y_{j,h}^{(>m)}\right\|_p^p$. We thus require $1/\nu - 2/p\delta - p^{-1}(\beta + 1)^{-1} - (q+1)/p > 0$. Using condition (A), this is true if $1/\nu - (2\beta+1)(p\beta +$

$p)^{-1} - (1+q)/p > 0$. Since by (F) $p > 4\nu$, we can choose a $q > 1$ such that $p > (3+q)\nu$, and it follows that

$$\frac{(2+q)\nu - p}{p-(3+q)\nu} < 0 < \beta, \tag{6.6.18}$$

hence this imposes no additional restriction, and (6.6.15) thus holds.

Regarding the covariance structure of $\{\mathbf{W}_t\}_{t\geq 0}$, note that the blocking and truncation argument has slightly changed the covariance structure. In order to quantify the error, note first that by stationarity, the covariance structure within the vectors $\boldsymbol{\xi}_j^{(m)}$ and $\boldsymbol{\eta}_j^{(m)}$ is the same for all j, and hence this is also true for the approximations $\boldsymbol{\xi}_j^{(m,*)}$ and $\boldsymbol{\eta}_j^{(m,**)}$. Put $I_{k,i} := \sum_{r\in\mathcal{H}_k} X_{r,i}$ and $II_{k,i} := \sum_{r\in\mathcal{I}_k} X_{r,i}$, and define $I_{k,i}^{(*)}$,$II_{k,i}^{(*)}$ in the same manner. Then the Cauchy-Schwarz inequality implies

$$\begin{aligned}
&\left|\mathbb{E}\left(\sum_{k=1}^{l}(I_{k,i}+II_{k,i})\sum_{k=1}^{l}(I_{k,j}+II_{k,j})\right) - \mathbb{E}\left(\sum_{k=1}^{l}(I_{k,i}^{(*)}+II_{k,i}^{(*)})\sum_{k=1}^{l}(I_{k,j}^{(*)}+II_{k,j}^{(*)})\right)\right| \\
&\leq \sqrt{\operatorname{Var}\big(\sum_{k=1}^{l} I_{k,i}\big)\operatorname{Var}\big(\sum_{k=1}^{l} II_{k,j}\big)} + \sqrt{\operatorname{Var}\big(\sum_{k=1}^{l} I_{k,j}\big)\operatorname{Var}\big(\sum_{k=1}^{l} II_{k,i}\big)} \\
&= \mathcal{O}\left(\sqrt{\sum_{k=1}^{l}|\mathcal{I}_k|\sum_{k=1}^{l}|\mathcal{H}_k|}\right) = \mathcal{O}\left(l^{\beta/2+1}\sqrt{m_n \vee d_n}\right) \\
&= \mathcal{O}\left(n^{\frac{2+\beta}{2(1+\beta)}}\sqrt{m_n\vee d_n}\right),
\end{aligned}$$

which gives us an upper bound for the error which stems from the blocking argument. Using the bound given in (6.6.3), a similar argument shows that the error which arises from the truncation of the random vectors $\mathbf{U}_{k,h}^{(m,i)}, \mathbf{V}_{k,h}^{(m,i)}$, $i \in \{1,2\}$ is of the magnitude $\mathcal{O}\big(n^{\frac{2+\beta}{2(1+\beta)}-1/4}\sqrt{m_n \vee d_n}\big)$. Finally, the error which comes from the conditioning argument is of the order $\mathcal{O}\big(n^{\frac{2+\beta}{2(1+\beta)}}\big)$, this follows again by a similar argument as before, using Assumption 6.6.1 (ii). Combining all bounds, we obtain that the total error is of the magnitude $\mathcal{O}\big(n^{\frac{2+\beta}{2(1+\beta)}}\sqrt{m_n\vee d_n}\big)$. Using relation (G), we thus obtain that $\max\big|\big| = \mathcal{O}\big(n^{-\gamma}\big)$.

which completes the proof. □

Proof of Theorem 6.6.6. Using Theorem 6.6.4, one can proceed exactly as in the proof of [14, Theorem 4.2]. □

Proof of Theorem 6.6.6. Using Theorem 6.6.4, one can proceed exactly as in the proofs of Theorem 4.3 and 4.4 in [14]. □

Proof of Theorem 6.5.1. This immediately follows from Lemma 6.6.3 and Theorem 6.6.7. The bound for δ simplifies since $\nu = 2$, $p > 8$ and thus $\delta < 3/4 < 1 - 2/p$, which is always valid due to Remark 6.6.5. □

6.7 Extremes of weighted extremes of Brownian Bridges

In this Section, we present some general results about the weak convergence of the maximum of extremes of (weighted) Brownian Bridges. The main results, together with some notation and preliminary remarks are given in Section 6.7.1, whereas Section 6.7.2 is solely devoted to the proofs.

6.7.1 Preliminary remarks and main results

Throughout the proofs, C denotes a generic constant that may vary from one formula to another.

For $1 \leq h \leq n$, let $\{W_{t,h}\}_{t\geq 0}$, $\{W_{t,h}^{(*)}\}_{t\geq 0}$ and $\{W_{t,h}^{(**)}\}_{t\geq 0}$ be independent Brownian motions, and $\{W_{t,h}^{\circ}\}_{t\geq 0}$ be an additional Brownian motion, which may depend on the latter three. Denote the covariance with $\rho_{i,j} = \mathbb{E}\big(W_{1,i}W_{1,j}\big)$.

For $1 \leq i \leq j \leq n$, consider the Hilbert spaces

$$\mathcal{S}_{i,j} = \big\{X \mid \langle X, X\rangle_H < \infty,\ X \in \sigma\big(W_{t,i}, W_{t,i+1}, ..., W_{t,j},\ 1 \leq t \leq 0\big)\big\},$$

equipped with the skalar product $\langle X, Y\rangle_H = \int_0^1 \mathbb{E}(X_t, Y_t)dt$. If $|\rho_{i,j}| > 0$ for some $1 \leq i < j$, then we can decompose $\mathcal{S}_{i,j}$ in distribution as

$$\mathcal{S}_{1,n} \stackrel{d}{=} \mathcal{S}_{1,i} \oplus \mathcal{S}_{i+1,j-1} \oplus \mathcal{S}_{j,n}^{(*)},$$

such that $\langle X, Y\rangle_H = 0$ for $X \in \mathcal{S}_{1,i}$, $Y \in \mathcal{S}_{j,n}^{(*)}$. In particular, $\mathcal{S}_{1,n}$ can be redefined such that any $W_t^{\circ} \in \mathcal{S}_{j,n}$ can be written as

$$W_{t,h}^{\circ} = W_{t,h}^{(*)}\sqrt{1 - \rho_{i,j,h}^2} + W_{t,h}\rho_{i,j,h}, \tag{6.7.1}$$

the numbers $\rho_{i,j,h}$ can for instance determined via the Gram-Schmidt procedure. Note that we have the following inequality

$$|\rho_{i,j,h}| \leq \sum_{r=1}^{i} |\rho_{r,h}| \leq \sum_{r=|h-i|}^{n} \sup_{r\leq|k-l|} |\rho_{k,l}|. \tag{6.7.2}$$

In addition, we require the following notation, where we drop the dependence on l for the sake of readability.

$$B_h = \sup_{l\le t\le 1-l}\left|\frac{W_{t,h}-tW_{1,h}}{\sqrt{t(1-t)}}\right|, \quad B_h^{(*)} = \sup_{l\le t\le 1-l}\left|\frac{W_{t,h}^{(*)}-tW_{1,h}^{(*)}}{\sqrt{t(1-t)}}\right|$$

$$B_h^{(*,i,j)} = \sup_{l\le t\le 1-l}\left|\frac{\sqrt{1-\rho_{i,j,h}^2}\big(W_{t,h}^{(*)}-tW_{1,h}^{(*)}\big)+\rho_{i,j,h}\big(W_{t,h}-tW_{1,h}\big)}{\sqrt{t(1-t)}}\right|, \ 1\le i\le j\le h\le n$$

$$B_h^{(**,i,j)} = \sup_{l\le t\le 1-l}\left|\frac{\sqrt{1-\rho_{i,j,h}^2}\big(W_{t,h}^{(*)}-tW_{1,h}^{(*)}\big)+\rho_{i,j,h}\big(W_{t,h}^{(**)}-tW_{1,h}^{(**)}\big)}{\sqrt{t(1-t)}}\right|, \ 1\le i\le j\le h\le n$$

We require the following tail estimate (cf. [78]).

Lemma 6.7.1. *Let $0 < l < 1$. Then we have*

$$P\left(B_h \ge x\right) = \frac{x\exp(-x^2/2)}{\sqrt{2\pi}}\left(\log\frac{(1-l)^2}{l^2}-\frac{1}{x^2}\log\frac{(1-l)^2}{l^2}+\frac{4}{x^4}+\mathcal{O}\big(x^{-4}\big)\right).$$

Corollary 6.7.2. *Let $\theta_l = \big(2\log(1-l)-2\log l\big)$, $z=\exp(-x)$. Then*

$$P\big(B_h\ge u_n\big) = \frac{z}{n}+\mathcal{o}\left(\frac{1}{n}\right),$$

where $u_n = u_n(z) = a_n x + b_n$, with $a_n = (2\log n)^{-1/2}$ and $b_n = \sqrt{2\log n} + (2\log n)^{-1/2}\left(\frac{1}{2}\log\log n+\log\theta_l-\frac{1}{2}\log\pi\right)$.

Remark 6.7.3. Note that u_n is similar to the well-known normalizing sequence in case of a Gaussian random variable ξ. The difference, apart form a constant, stems from the fact that $1-\Phi(u_n)\asymp\phi(u_n)/u_n$, whereas $P\left(B_h\ge u_n\right)\asymp\phi(u_n)u_n$, where $\Phi(x)$ denotes the cdf of ξ, with corresponding density function $\phi(x)$. Also note that

$$u_n^2 = 2\log n+\frac{1}{2}\log\log n+\mathcal{o}(\log\log n). \tag{6.7.3}$$

Proof of Corollary 6.7.2. By Lemma 6.7.1 we have

$$\frac{z}{n}+\mathcal{o}\left(\frac{1}{n}\right) = u_n\exp(-u_n^2/2)\frac{\theta_l}{\sqrt{2\pi}}.$$

Taking logarithms one obtains

$$u_n^2 = 2\left(x+\log n+\log u_n+\log\theta_l-\frac{1}{2}\log 2\pi\right)+\mathcal{o}(1), \tag{6.7.4}$$

which implies that $u_n/\sqrt{2\log n} \to 1$. Plugging this in equation (6.7.4) yields that

$$u_n = (2\log n)^{-1/2}\left(x + \frac{1}{2}\log\log n + \log\theta_l - \frac{1}{2}\log\pi\right) + \sqrt{2\log n} + o\left((\log n)^{-1/2}\right),$$

which yields the claim. □

In case of the non weighted Brownian Bridge, we have the following tail estimate (cf. [115]).

Lemma 6.7.4. *It holds that*

$$P\big(\sup_{0\le t\le 1}\big|W_t - tW_1\big| \ge x\big) = 2\sum_{k=1}^{\infty}(-1)^{k+1}\exp(-2k^2x^2).$$

Consequently, one obtains for $z = \exp(-x)$ that

$$P\big(\sup_{0\le t\le 1}\big|W_t - tW_1\big| \ge v_n(z)\big) = \frac{z}{n} + o\left(n^{-1}\right),$$

where $v_n = v_n(z) = e_n x + f_n$, with $e_n = 1/4\big(\log(2n)/2\big)^{-1/2}$, $f_n = \sqrt{1/2\log(2n)}$.

The proof of Theorem 6.5.2 is based on the more general Theorem 6.7.5, that is given below.

Theorem 6.7.5. *Suppose that $u_n = u_n(z)$ is as in Corollary 6.7.2, and that*

(i) $\max_{1\le h\le n}\sup_{i,j:u_n\le|i-j|}\big|\rho_{i,j,h}\big| = o(u_n^{-1})$,

(ii) $\sup_{i,j:u_N\le|i-j|}\big|\rho_{i,j}\big| = o(u_N^{-2})$,

(iii) $\sup_{i,j:1\le|i-j|}\big|\rho_{i,j}\big| < 1$.

Then

$$\lim_{n\to\infty} P\left(\max_{1\le h\le n} B_h \le u_n(z)\right) = \exp(-z).$$

Proof of Theorem 6.5.2. We need to show that the assumptions in Theorem 6.5.2 imply those needed in Theorem 6.7.5. Since $u_n^2 = \mathcal{O}(\log n)$, Assumption (i) of Theorem 6.5.2 implies (i), (ii) of Theorem 6.7.5 using the inequality given in (6.7.2). In addition, condition (ii) of Theorem 6.5.2 is identical with condition (iii) of Theorem 6.7.5. □

In an analogue manner, using Lemma 6.7.4 one obtains that

Theorem 6.7.6. *Suppose that $v_n = v_n(z)$ is as in Lemma 6.7.4, and that*

(i) $\max_{1\le h\le n} \sup_{i,j:v_n\le|i-j|} \big|\rho_{i,j,h}\big| = o(v_n^{-1})$,

(ii) $\sup_{i,j:v_N\le|i-j|} \big|\rho_{i,j}\big| = o(v_N^{-2})$,

(iii) $\sup_{i,j:1\le|i-j|} \big|\rho_{i,j}\big| < 1$.

Then

$$\lim_{n\to\infty} P\left(\max_{1\le h\le n} \sup_{0\le t\le 1} \big|W_t - tW_1\big| \le v_n(z)\right) = \exp(-z).$$

6.7.2 Proofs

Limit theorems for the maximum of a sequence of dependent random variables have a long tradition, early contributions are due to Berman [17], Loynes [84], and Watson [125]. Finally, in [79], Leadbetter provided the famous and quite general conditions $D(u_n)$ and $D^*(u_N)$, which imply weak convergence to an extreme value distribution (see also [84] for earlier formulations). We will slightly adapt these conditions to our cause.

Condition $D(u_n)$**.** *There exists a sequence $\alpha_n \downarrow 0$, such that for any $1 \le i \le j$ with $\sqrt{\log n} \le j - i$ we have*

$$\Big|P\big(\max_{1\le h\le i} B_h \le u_n, \max_{j\le h\le n} B_h^{(*,i,j)} \le u_n\big) - P\big(\max_{1\le h\le i} B_h \le u_n, \max_{j\le h\le n} B_h^{(**,i,j)} \le u_n\big)\Big| \le \alpha_n,$$

where u_n is as in Corollary 6.7.2.

Condition $D^*(u_n)$**.** *For any $k \in N$, it holds that*

$$\limsup_{n\to\infty} \sum_{1\le|i-j|\le n} P\big(B_i \ge u_{nk}, B_j \ge u_{nk}\big) = o\left(\frac{1}{k}\right), \qquad (6.7.5)$$

where u_n is as in Corollary 6.7.2.

We have the following Lemma, which reflects Condition $D(u_n)$.

Lemma 6.7.7. *Let L_n be a monotone increasing sequence such that*

$$\max_{1\le h\le n} \sup_{i,j:L_n\le|i-j|} \big|\rho_{i,j,h}\big| = o(u_n^{-1}). \qquad (6.7.6)$$

Then there exists a sequence α_n with $\lim_{n\to\infty} \alpha_n = 0$ such that for any $1 \le i \le j$ with $L_n \le j - i$ we have condition $D(u_n)$

$$\Big|P\big(\max_{1\le h\le i} B_h \le u_n, \max_{j\le h\le n} B_h^{(*,i,j)} \le u_n\big) - P\big(\max_{1\le h\le i} B_h \le u_n, \max_{j\le h\le n} B_h^{(**,i,j)} \le u_n\big)\Big| \le \alpha_n.$$

Proof of Lemma 6.7.7. We have that

$$\begin{aligned}
&\Big|P\big(\max_{1\le h\le i} B_h \le u_n, \max_{j\le h\le n} B_h^{(*,i,j)} \le u_n\big) - P\big(\max_{1\le h\le i} B_h \le u_n, \max_{j\le h\le n} B_h^{(**,i,j)} \le u_n\big)\Big| \\
&= \Big|\mathbb{E}\left(\mathbf{1}_{\{\max_{1\le h\le i} B_h\le u_n\}}\mathbf{1}_{\{\max_{j\le h\le n} B_h^{(*,i,j)}\le u_n\}} - \mathbf{1}_{\{\max_{1\le h\le i} B_h\le u_n\}}\mathbf{1}_{\{\max_{j\le h\le n} B_h^{(**,i,j)}\le u_n\}}\right)\Big| \\
&\le \mathbb{E}\big|\mathbf{1}_{\{\max_{j\le h\le n} B_h^{(*,i,j)}\le u_n\}} - \mathbf{1}_{\{\max_{j\le h\le n} B_h^{(**,i,j)}\le u_n\}}\big| \\
&= P\big(\max_{j\le h\le n} B_h^{(*,i,j)} \le u_n, \max_{j\le h\le n} B_h^{(**,i,j)} > u_n\big) + P\big(\max_{j\le h\le n} B_h^{(*,i,j)} > u_n, \max_{j\le h\le n} B_h^{(**,i,j)} \le u_n\big) \\
&= 2P\big(\max_{j\le h\le n} B_h^{(*,i,j)} > u_n, \max_{j\le h\le n} B_h^{(**,i,j)} \le u_n\big) \le 2\sum_{j=h}^{n} P\big(B_h^{(*,i,j)} > u_n, B_h^{(**,i,j)} \le u_n\big).
\end{aligned}$$

We can split up the probabilities on the right hand side as

$$\begin{aligned}
P\big(B_h^{(*,i,j)} > u_n, B_h^{(**,i,j)} \le u_n\big) \le{}& P\big(\sqrt{1-\rho_{i,j,h}^2}B_h^{(*)} > u_n - \epsilon_n, \sqrt{1-\rho_{i,j,h}^2}B_h^{(*)} \le u_n + \epsilon\big) \\
&+ 2P\big(B_h^{(**)} > |\rho_{i,j,h}^{-1}|\epsilon_n\big) := I_n + II_n,
\end{aligned}$$

where $\epsilon_n > 0$. Note that condition (6.7.6) implies that we can chose ϵ_n such that $\lim_{n\to\infty}\epsilon_n = 0$, and

$$\max_{1\le h\le n}\sup_{i,j:L_n\le|i-j|}\big|\rho_{i,j,h}\big|\epsilon_n^{-1} \le (2u_n)^{-1}. \tag{6.7.7}$$

Hence we have

$$P\big(B_h^{(**)} > |\rho_{i,j,h}^{-1}|\epsilon_n\big) \le P\big(B_h^{(**)} > 2u_n\big),$$

and Lemma 6.7.1 now yields that

$$I_n \le P\big(B_h > 2u_n\big) = o\left(n^{-1}\right). \tag{6.7.8}$$

We will now treat II_n. To this end, denote with $f(x)$ the continuous density function of the random variable $\sqrt{1-\rho_{i,j,h}^2}B_h^{(*)}$. Since $\epsilon_n \downarrow 0$, we have

$$\begin{aligned}
P\left(\sqrt{1-\rho_{i,j,h}^2}B_h^{(*)} > u_n - \epsilon_n, \sqrt{1-\rho_{i,j,h}^2}B_h^{(*)} \le u_n + \epsilon_n\right) &= \int_{u_n-\epsilon_n}^{u_n+\epsilon_n} f(x)dx \\
&\le C\epsilon_n f(u_n) \le C\epsilon_n\int_{u_n}^{\infty} f(x)dx = C\epsilon_n P\big(B_h^{(*)} \ge u_n(1-\rho_{i,j,h}^2)^{-1/2}\big) \\
&\le C\epsilon_n P\big(B_h^{(*)} \ge u_n\big) = \mathcal{O}\left(n^{-1}\epsilon_n\right).
\end{aligned}$$

Using (6.7.8) we thus obtain

$$P\big(B_h^{(*,i,j)} > u_n, B_h^{(**,i,j)} \le u_n\big) = \mathcal{o}\left(n^{-1}\right), \tag{6.7.9}$$

which yields

$$\sum_{j=h}^{n} P\big(B_h^{(*,i,j)} > u_n, B_h^{(**,i,j)} \le u_n\big) = \mathcal{o}(1). \tag{6.7.10}$$

Hence the claim follows. □

Lemma 6.7.8, given below, corresponds to Condition $D^*(u_n)$.

Lemma 6.7.8. *Let $N = nk$, for some $k \in \mathbb{N}$, and suppose that*

(i) $\sup_{i,j: u_N \le |i-j|} \big|\rho_{i,j}\big| = \mathcal{o}(u_N^{-2})$,

(ii) $\sup_{i,j: 1 \le |i-j|} \big|\rho_{i,j}\big| < 1$.

Then we have condition $D^*(u_N)$

$$\limsup_{n\to\infty} \sum_{1\le|i-j|\le n} P\big(B_i \ge u_{nk}, B_j \ge u_{nk}\big) = \mathcal{o}\left(\frac{1}{k}\right).$$

Proof of Lemma 6.7.8. We have the following decomposition

$$\begin{aligned}\sum_{1\le|i-j|\le n} P\big(B_i \ge u_{nk}, B_j \ge u_{nk}\big) = &\sum_{1\le|i-j|\le u_{nk}} P\big(B_i \ge u_{nk}, B_j \ge u_{nk}\big)\\ &+ \sum_{u_{nk}<|i-j|\le n} P\big(B_i \ge u_{nk}, B_j \ge u_{nk}\big).\end{aligned} \tag{6.7.11}$$

Note that the following inequality is valid.

$$\begin{aligned}P\big(B_i \ge u_{nk}, B_j \ge u_{nk}\big) &= P\big(B_i \ge u_{nk}, B_j^{(*,i,j)} \ge u_{nk}\big)\\ &\le P\big(B_i \ge u_{nk}, |\rho_{i,j,j}|B_i + \sqrt{1-\rho_{i,j,j}^2}B_j^{(*)} \ge u_{nk}\big)\end{aligned}$$

We will now treat the second sum on the right hand side in (6.7.11). To this end, let $u_{nk} \le |i-j|$ and $\epsilon_n > 0$. Then

$$\begin{aligned}&P\big(B_i \ge u_{nk}, |\rho_{i,j,j}|B_i + \sqrt{1-\rho_{i,j,j}^2}B_j^{(*)} \ge u_{nk}\big)\\ &\qquad\le P\big(B_i \ge u_{nk}, B_j^{(*)} \ge u_{nk} - \epsilon_n\big) + P\big(|\rho_{i,j,j}|B_i \ge \epsilon_n\big)\\ &\qquad= P\big(B_i \ge u_{nk}\big)P\big(B_j^{(*)} \ge u_{nk} - \epsilon_n\big) + P\big(|\rho_{i,j,j}|B_i \ge \epsilon_n\big).\end{aligned}$$

Arguing as in the proof of Lemma 6.7.7, we obtain from condition (i) that

$$P\big(|\rho_{i,j,j}|B_i \geq \epsilon_n\big) \leq P\big(B_i \geq 2u_{nk}\big) = \mathcal{O}\left((nk)^{-1-\theta}\right), \tag{6.7.12}$$

for some $\theta > 0$, such that $\epsilon_n = o\left(u_{nk}^{-1}\right)$. Moreover, we obtain - again in analogy to the proof of Lemma 6.7.6 - that

$$P\big(u_n \geq B_j^{(*)} \geq u_{nk} - \epsilon_n\big) = \mathcal{O}\left(\epsilon_n (nk)^{-1}\right) = o\left((u_{nk} nk)^{-1}\right). \tag{6.7.13}$$

We thus deduce that

$$P\big(B_i \geq u_{nk}, |\rho_{i,j,j}|B_i + \sqrt{1-\rho_{i,j,j}^2}B_j^{(*)} \geq u_{nk}\big) = \mathcal{O}\left((nk)^{-2}\right) + o\left((u_{nk} nk)^{-1}\right), \tag{6.7.14}$$

if $u_{nk} \leq |i-j|$. Since $u_n^2 = \mathcal{O}(\log n)$ by Remark 6.7.3, we obtain

$$\sum_{u_{nk} < |i-j| \leq n} P\big(B_i \geq u_{nk}, B_j \geq u_{nk}\big) = \mathcal{O}\left(k^{-1-\theta}\right), \tag{6.7.15}$$

for some $\theta > 0$. It remains to treat the first sum on the right hand side in (6.7.11). To this end, note that for some $\lambda > 1$

$$\begin{aligned} P\big(B_i \geq u_{nk}, B_j \geq u_{nk}\big) &\leq P\big(B_i \geq u_{nk}, |\rho_{i,j,j}|B_i + \sqrt{1-\rho_{i,j,j}^2}B_j^{(*)} \geq u_{nk}\big) \\ &= \int_{u_{nk}}^{\infty} P\left(B_j^{(*)} \geq \frac{u_{nk} - x|\rho_{i,j,j}|}{\sqrt{1-\rho_{i,j,j}^2}}\right) P_{B_i}\big(dx\big) \\ &\leq \int_{u_{nk}}^{\lambda u_{nk}} P\left(B_j^{(*)} \geq \frac{u_{nk} - x|\rho_{i,j,j}|}{\sqrt{1-\rho_{i,j,j}^2}}\right) P_{B_i}\big(dx\big) + P\big(B_i \geq \lambda u_{nk}\big) \\ &\leq P\left(B_j^{(*)} \geq u_{nk}\frac{1 - \lambda|\rho_{i,j,j}|}{\sqrt{1-\rho_{i,j,j}^2}}\right) + P\big(B_i \geq \lambda u_{nk}\big). \end{aligned}$$

Since $|\rho_{i,j,j}| < 1$ by condition (ii), we can chose $\lambda > 1$ such that $(1-\lambda|\rho_{i,j,j}|)(1-\rho_{i,j,j}^2)^{-1/2} > 0$. Hence, arguing as before, we obtain that

$$P\big(B_i \geq u_{nk}, B_j \geq u_{nk}\big) = \mathcal{O}\left((nk)^{-1-\theta}\right), \tag{6.7.16}$$

for some $\theta > 0$. Since $u_n^2 = \mathcal{O}(\log n)$ by Remark 6.7.3, we obtain

$$\sum_{1 \leq |i-j| \leq u_{nk}} P\big(B_i \geq u_{nk}, B_j \geq u_{nk}\big) = \mathcal{O}\left(k^{-1-\theta}\right), \tag{6.7.17}$$

which completes the proof. □

We will now proceed as in [79] to obtain the desired results. To this end, write $M(E) = \max\{B_h : h \in E\}$ for any set E of integers.

Lemma 6.7.9. *Suppose that $D(u_n)$ holds. Let N, r, k be fixed integers and $E_1, E_2, ...E_r$ subintervals of $(1, 2...d_n)$, such that E_i and E_j are separated by at least k when $i \neq j$. Then*

$$\left| P\left(\bigcap_{j=1}^{r}\big(M(E_j) \leq u_n\big)\right) - \prod_{j=1}^{r} P\left(M(E_j) \leq u_n\right)\right| \leq (r-1)\alpha_n$$

Proof of Lemma 6.7.9. One can proceed exactly as in the proof of Lemma 2.3 in [79]. □

Let k be a fixed integer, and write $N = nk$, $n = 1, 2....$ We divide the first $N = nk$ integers into $2k$ consecutive intervals as follows. Let m be a fixed integer and write $I_1 = (1, 2, ..., n-m)$, $I_1^* = (n-m+1, ..., n)$, $I_2 = (n+1, ..., 2n-m)$, $I_1^* = (2n-m+1, ..., 2n)$, and so on.

Lemma 6.7.10. *Assume that $D(u_n)$ holds. Then*

(i) $0 \leq P\left(\bigcap_{j=1}^{r}\big(M(I_j) \leq u_n\big)\right) - P\big(\max_{1\leq h\leq N} B_h \leq u_N\big) \leq \sum_{j=1}^{k} P\big(M(I_j) \leq u_N \leq M(I_j^*)$,

(ii) $P\left|\left(\bigcap_{j=1}^{r}\big(M(I_j) \leq u_n\big)\right) - \prod_{j=1}^{k}\big(M(I_j) \leq u_n\big)\right| \leq k\alpha_n$,

(iii) $\left|\prod_{j=1}^{k}\big(M(I_j) \leq u_n\big) - \prod_{j=1}^{k} P\big(\max_{1+(j-1)n\leq h\leq jn} B_h \leq u_n\big) \leq K \max_{1\leq j\leq k} P\big(M(I_j) \leq u_N \leq M(I_j^*)\big)\right.$,

for some constant K. Hence, by combining (i), (ii), (iii),

$$\begin{aligned}\Big|P\big(\max_{1\leq h\leq N} B_h \leq u_N\big) - \prod_{j=1}^{k} P\big(\max_{1+(j-1)n\leq h\leq jn} B_h \leq u_n\big)\\ \leq (k+K)\max_{1\leq j\leq k} P\big(M(I_j) \leq u_N \leq M(I_j^*)\big) + k\alpha_n.\end{aligned} \tag{6.7.18}$$

Proof of Lemma 6.7.10. One can proceed exactly as in the proof of Lemma 2.4 in [79]. □

Lemma 6.7.11. *Suppose that*

$$\min_{1\leq h\leq N} \inf_{i,j:1\leq|i-j|} \left|\rho_{i,j,h}\right| < 1.$$

Then

$$\max_{1\leq j\leq k} P\big(M(I_j) \leq u_N \leq M(I_j^*) = o(1). \tag{6.7.19}$$

Proof of Lemma 6.7.11. The claim follows by using the same arguments as in the proof of Lemma 6.7.7 and 6.7.8. □

It then follows from Lemma 6.7.10 that

$$\left| P\big(\max_{1\le h\le N} B_h \le u_N\big) - \prod_{j=1}^{k} P\big(\max_{1+(j-1)n\le h\le jn} B_h \le u_n\big)\right| = \mathcal{o}(1). \tag{6.7.20}$$

Proof of Theorem 6.7.5. First note that the assumptions cover the conditions needed for Lemmas 6.7.7 and 6.7.8. Now let $k > 0$ be an integer, and $N = nk$. Note that for $1 \le i \le k$

$$\begin{aligned} P\big(\max_{1+(i-1)n\le h\le in} B_h \le u_N\big) &= 1 - P\left(\bigcup_{j=1+(i-1)n}^{in} \{B_h > u_N\}\right) \\ &\ge 1 - \sum_{j=1+(i-1)n}^{in} P\big(B_j > u_N\big). \end{aligned}$$

Hence we deduce from Corollary 6.7.2 that

$$\liminf_{n\to\infty} P\big(\max_{1+(i-1)n\le h\le in} B_h \le u_N\big) \ge 1 - \frac{z}{k},$$

and thus (6.7.20) implies that

$$\left(1 - \frac{z}{k}\right)^k \le \liminf_{n\to\infty} P\big(\max_{1\le h\le N} B_h \le u_N\big). \tag{6.7.21}$$

Corresponding we also have by Lemma 6.7.8 and Corollary 6.7.2 that

$$\begin{aligned} P\big(\max_{1+(i-1)n\le h\le in} B_h \le u_N\big) &\le 1 - \sum_{j=1+(i-1)n}^{in} P\big(B_j > u_N\big) + \sum_{\substack{1\le|h-j|\le n,\\ 1+(i-1)n\le h,j\le in}} P\big(B_j > u_N, B_h > u_N\big) \\ &= 1 - \frac{z}{k} + \mathcal{o}\left(\frac{1}{k}\right). \end{aligned}$$

Hence we obtain

$$\limsup_{n\to\infty} P\big(\max_{1+(i-1)n\le h\le in} B_h \le u_N\big) \le 1 - \frac{z}{k} + \mathcal{o}\left(\frac{1}{k}\right), \tag{6.7.22}$$

which together with (6.7.21) yields

$$\begin{aligned} \left(1 - \frac{z}{k}\right)^k &\le \liminf_{n\to\infty} P\big(\max_{1\le h\le N} B_h \le u_N\big) \le \limsup_{n\to\infty} P\big(\max_{1+(i-1)n\le h\le in} B_h \le u_N\big) \\ &\le \left(1 - \frac{z}{k} + \mathcal{o}\left(\frac{1}{k}\right)\right)^k. \end{aligned} \tag{6.7.23}$$

Proceeding as in the proof of Theorem 3.1 in [79], one shows that one can replace $N = nk$ with n in (6.7.23). The result then clearly follows by letting $k \to \infty$. $\square$

Bibliography

[1] H. Akaike. Fitting autoregressive models for prediction. *Ann. Inst. Statist. Math.*, 21:243–247, 1969.

[2] H. Akaike. On entropy maximization principle. In *Applications of statistics (Proc. Sympos., Wright State Univ., Dayton, Ohio, 1976)*, pages 27–41. North-Holland, Amsterdam, 1977.

[3] H. Z. An, Z. G. Chen, and E. J. Hannan. Autocorrelation, autoregression and autoregressive approximation. *Ann. Statist.*, 10(3):926–936, 1982.

[4] T. W. Anderson. *The statistical analysis of time series*. John Wiley & Sons Inc., New York, 1971.

[5] D. W. K. Andrews. Non-strong mixing autoregressive processes. *Journal of Applied Probability*, 21(4):pp. 930–934, 1984.

[6] A. Aue, R. Gabrys, L. Horváth, and P. Kokoszka. Estimation of a change-point in the mean function of functional data. *J. Multivariate Anal.*, 100(10):2254–2269, 2009.

[7] A. Aue, S. Hörmann, L. Horváth, and M. Reimherr. Break detection in the covariance structure of multivariate time series models. *Ann. Statist.*, 37(6B):4046–4087, 2009.

[8] A. Aue, S. Hörmann, L. Horváth, and M. Reimherr. Break detection in the covariance structure of multivariate time series models. *Ann. Statist.*, 37(6B):4046–4087, 2009.

[9] A. Aue, L. Horváth, M. Hušková, and P. Kokoszka. Change-point monitoring in linear models. *Econom. J.*, 9(3):373–403, 2006.

[10] A. A. Balkema and L. De Haan. A convergence rate in extreme-value theory. *Journal of Applied Probability*, 27(3):pp. 577–585, 1990.

[11] A. Banerjee and G. Urga. Modelling structural breaks, long memory and stock market volatility: an overview. *Journal of Econometrics*, 129(1-2):1–34, 2005.

[12] A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413, 1999.

[13] K. N. Berk. Consistent autoregressive spectral estimates. *Ann. Statist.*, 2:489–502, 1974. Collection of articles dedicated to Jerzy Neyman on his 80th birthday.

[14] I. Berkes, E. Gombay, and L. Horváth. Testing for changes in the covariance structure of linear processes. *J. Statist. Plann. Inference*, 139(6):2044–2063, 2009.

[15] I. Berkes, S. Hörmann, and J. Schauer. Asymptotic results for the empirical process of stationary sequences. *Stochastic Process. Appl.*, 119(4):1298–1324, 2009.

[16] I. Berkes and W. Philipp. Approximation theorems for independent and weakly dependent random vectors. *Ann. Probab.*, 7(1):29–54, 1979.

[17] S. M. Berman. Limit theorems for the maximum term in stationary sequences. *Ann. Math. Statist.*, 35:502–516, 1964.

[18] P. Berthet and D. M. Mason. Revisiting two strong approximation results of Dudley and Philipp. In *High dimensional probability*, volume 51 of *IMS Lecture Notes Monogr. Ser.*, pages 155–172. Inst. Math. Statist., Beachwood, OH, 2006.

[19] P. Berthet and D. M. Mason. Revisiting two strong approximation results of Dudley and Philipp. In *High dimensional probability*, volume 51 of *IMS Lecture Notes Monogr. Ser.*, pages 155–172. Inst. Math. Statist., Beachwood, OH, 2006.

[20] R. J. Bhansali. Consistent recursive estimation of the order of an autoregressive moving average process. *International Statistical Review / Revue Internationale de Statistique*, 59(1):pp. 81–96, 1991.

[21] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 31(3):307–327, 1986.

[22] T. E. Bollerslev, F. Robert, and D. B. Nelson. Arch models. In *Handbook of econometrics, Vol. IV*, volume 2 of *Handbooks in Econom.*, pages 2959–3038. North-Holland, Amsterdam, 1994.

[23] S. Bourguin and C. A. Tudor. Berry-esseen bounds for long memory moving averages via stein's method and malliavin calculus. Technical report.

[24] G. E. P. Box, G. M. Jenkins, and G. Reinsel. *Time series analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc., Hoboken, NJ, fourth edition, 2008. Forecasting and control.

[25] P. J. Brockwel and R. A. Davis. *Time series: theory and methods*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1991.

[26] P. J. Brockwell, R. Dahlhaus, and A. Trindade. Modified Burg algorithms for multivariate subset autoregression. *Statist. Sinica*, 15(1):197–213, 2005.

[27] B. M. Brown and C. C. Heyde. On the departure from normality of a certain class of martingales. *Ann. Math. Statist.*, 41:2161–2165, 1970.

[28] B. Courbot. Rates of convergence in the functional clt for martingales. *Comptes Rendus de l'Acad?mie des Sciences - Series I - Mathematics*, 328(6):509 – 513, 1999.

[29] M. Csörgő and L. Horváth. *Weighted approximations in probability and statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester, 1993. With a foreword by David Kendall.

[30] M. Csörgő and P. Révész. A new method to prove Strassen type laws of invariance principle. I, II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31:255–259; ibid. 31 (1974/75), 261–269, 1974/75.

[31] S. CSÖRGÖ and P. Hall. The komlÓs-major-tusnÁdy approximations and their applications. *Australian Journal of Statistics*, 26(2):189–218, 1984.

[32] M. Csörgo and L. Horváth. *Limit Theorems in Change-Point Analysis*. 1997.

[33] L. de Haan and S. Resnick. Second-order regular variation and rates of convergence in extreme-value theory. *Ann. Probab.*, 24(1):97–124, 1996.

[34] J. Dedecker and P. Doukhan. A new covariance inequality and applications. *Stochastic Process. Appl.*, 106(1):63–80, 2003.

[35] J. Dedecker and P. Doukhan. A new covariance inequality and applications. *Stochastic Processes and their Applications*, 106(1):63 – 80, 2003.

[36] A. Dembo, S. Karlin, and O. Zeitouni. Critical phenomena for sequence matching with scoring. *Ann. Probab.*, 22(4):1993–2021, 1994.

[37] A. Dembo, S. Karlin, and O. Zeitouni. Limit distribution of maximal non-aligned two-sequence segmental score. *Ann. Probab.*, 22(4):2022–2039, 1994.

[38] C. M. Deo. Some limit theorems for maxima of absolute values of Gaussian sequences. *Sankhyā Ser. A*, 34:289–292, 1972.

[39] C. M. Deo. An iterated logarithm law for maxima of nonstationary gaussian processes. *Journal of Applied Probability*, 10(2):pp. 402–408, 1973.

[40] D. L. Donoho. High-dimensional data analysis: the curses and blessings of dimensionality. In *American Mathematical Society Conf. Math Challenges of the 21st Century*. 2000.

[41] P. Doukhan, P. Massart, and E. Rio. The functional central limit theorem for strongly mixing processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 30(1):63–82, 1994.

[42] P. Doukhan and O. Wintenberger. An invariance principle for weakly dependent stationary general models. *Probab. Math. Statist.*, 27(1):45–73, 2007.

[43] L. E. Dubins. On a theorem of Skorohod. *Ann. Math. Statist.*, 39:2094–2097, 1968.

[44] L. E. Dubins and G. Schwarz. A sharp inequality for sub-martingales and stopping-times. *Astérisque*, (157-158):129–145, 1988. Colloque Paul Lévy sur les Processus Stochastiques (Palaiseau, 1987).

[45] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[46] U. Einmahl. Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *J. Multivariate Anal.*, 28(1):20–68, 1989.

[47] U. Einmahl and D. Mason. Gaussian approximation of local empirical processes indexed by functions. *Probab. Theory Related Fields*, 107(3):283–311, 1997.

[48] R. Engle. Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *J. Bus. Econom. Statist.*, 20(3):339–350, 2002.

[49] J. Fan and Q. Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods*. (Springer Series in Statistics). Springer, August 2005.

[50] D. P. Foster and E. I. George. The risk inflation criterion for multiple regression. *Ann. Statist.*, 22(4):1947–1975, 1994.

[51] R. F. Galbraith and J. I. Galbraith. On the inverses of some patterned matrices arising in the theory of stationary time series. *Journal of Applied Probability*, 11(1):pp. 63–71, 1974.

[52] J. Gao. *Nonlinear time series*, volume 108 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, FL, 2007. Semiparametric and nonparametric methods.

[53] L. Giraitis, P. Kokoszka, R. Leipus, and G. Teyssière. Rescaled variance and related tests for long memory in volatility and levels. *J. Econometrics*, 112(2):265–294, 2003.

[54] E. Gombay and L. Horváth. Sequential tests and change detection in the covariance structure of weakly stationary time series. *Comm. Statist. Theory Methods*, 38(16-17):2872–2883, 2009.

[55] M. I. Gordin. The central limit theorem for stationary processes. *Dokl. Akad. Nauk SSSR*, 188:739–741, 1969.

[56] Christian Gouriéroux. *ARCH models and financial applications*. Springer Series in Statistics. Springer-Verlag, New York, 1997.

[57] I. G. Grama. On moderate deviations for martingales. *Ann. Probab.*, 25(1):152–183, 1997.

[58] P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980. Probability and Mathematical Statistics.

[59] W.J. Hall and J.A. Wellner. The rate of convergence in law of the maximum of an exponential sample. *Statistica Neerlandica*, 33(3):151–154, 1979.

[60] E. J. Hannan. *Multiple time series*. John Wiley and Sons, Inc., New York-London-Sydney, 1970.

[61] E. J. Hannan. The asymptotic distribution of serial covariances. *Ann. Statist.*, 4(2):396–399, 1976.

[62] E. J. Hannan. The estimation of the order of an ARMA process. *Ann. Statist.*, 8(5):1071–1081, 1980.

[63] E. J. Hannan and B. G. Quinn. The determination of the order of an autoregression. *J. Roy. Statist. Soc. Ser. B*, 41(2):190–195, 1979.

[64] D. Harris, B. McCabe, and S. Leybourne. Some limit theory for autocovariances whose order depends on sample size. *Econometric Theory*, 19(5):829–864, 2003.

[65] L. Horváth, M. Hušková, and P. Kokoszka. Testing the stability of the functional autoregressive process. *J. Multivariate Anal.*, 101(2):352–367, 2010.

[66] L. Horváth and P. Kokoszka. Sample autocovariances of long-memory time series. *Bernoulli*, 14(2):405–418, 2008.

[67] I. A. Ibragimov. On the accuracy of approximation by the normal distribution of distribution functions of sums of independent random variables. *Teor. Verojatnost. i Primenen*, 11:632–655, 1966.

[68] I. A. Ibragimov and Yu. V. Linnik. *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing, Groningen, 1971. With a supplementary chapter by I. A. Ibragimov and V. V. Petrov, Translation from the Russian edited by J. F. C. Kingman.

[69] C. Inclán and G.C. Tiao. Use of cumulative sums of squares for retrospective detection of changes of variance. *J. Amer. Statist. Assoc.*, 89(427):913–923, 1994.

[70] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.

[71] T. Jiang. The limiting distributions of eigenvalues of sample correlation matrices. *Sankhyā*, 66(1):35–48, 2004.

[72] I. M. Johnstone. On the distribution of the largest principal component. *ANN. STATIST*, 29:295–327, 2000.

[73] Daniel M. Keenan. A central limit theorem for $m(n)$ autocovariances. *J. Time Ser. Anal.*, 18(1):61–78, 1997.

[74] J. Komlós, P. Major, and G. Tusnády. Weak convergence and embedding. In *Limit theorems of probability theory (Colloq., Keszthely, 1974)*, pages 149–165. Colloq. Math. Soc. János Bolyai, Vol. 11. North-Holland, Amsterdam, 1975.

[75] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 34(1):33–58, 1976.

[76] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 34(1):33–58, 1976.

[77] J. Kuelbs and T. Kurtz. Berry-Esseen estimates in Hilbert space and an application to the law of the iterated logarithm. *Ann. Probability*, 2:387–407, 1974.

[78] M. Csörgö L. and Horváth. *Limit Theorems in Change-Point Analysis.* John Wiley & Sons, January 1997.

[79] M. R. Leadbetter. On extreme values in stationary sequences. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 28:289–303, 1973/74.

[80] M.R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes.* Springer Series in Statistics. Springer-Verlag, New York, 1983.

[81] S. Lee, J. Ha, O. Na, and S. Na. The cusum test for parameter change in time series models. *Scandinavian Journal of Statistics*, 30(4):pp. 781–796, 2003.

[82] H. Leeb and B. M. Pötscher. Model selection. *In, Handbook of Financial Time Series. New York: Springer.*, 2008.

[83] G. M. Ljung and G.E.P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, August 1978.

[84] R. M. Loynes. Extreme values in uniformly mixing stationary stochastic processes. *Ann. Math. Statist.*, 36:993–999, 1965.

[85] H. Lütkepohl. *New introduction to multiple time series analysis.* Springer-Verlag, Berlin, 2005.

[86] C. L. Mallows. Choosing variables in a linear regression: A graphical aid. *Presented at the Central Regional Meeting of the Institute of Mathematical Statistics, Manhattan, Kansas. 5*, 1964.

[87] C. L. Mallows. Choosing a subset regression. *Presented at the Joint Statistical Meeting, Los Angeles, Calif.*, 1966.

[88] J. McClave. Subset autoregression. *Technometrics*, 17:213–220, 1975.

[89] A. I. McLeod and Ying Zhang. Improved subset autoregression: With R package. *Journal of Statistical Software*, 28(2), 2008.

[90] I. Monroe. On embedding right continuous martingales in Brownian motion. *Ann. Math. Statist.*, 43:1293–1311, 1972.

[91] I. Monroe. Processes that can be embedded in Brownian motion. *Ann. Probability*, 6(1):42–56, 1978.

[92] F. A. Moricz, R. J. Serfling, and W. F. Stout. Moment and probability bounds with quasi-superadditive structure for the maximum partial sum. *The Annals of Probability*, 10(4):pp. 1032–1040, 1982.

[93] T. Nakatsuka. Regions of autocorrelation coefficients in ar(p) and ex(p) processes. *Annals of the Institute of Statistical Mathematics*, 30:315–319, 1978. 10.1007/BF02480221.

[94] I. Nourdin and G. Peccati. Stein's method on wiener chaos. *Probability Theory and Related Fields*, 145:75–118, 2009. 10.1007/s00440-008-0162-x.

[95] I. Nourdin, G. Peccati, and M. Podolskij. Quantitative breuer-major theorems. CREATES Research Papers 2010-22, School of Economics and Management, University of Aarhus, May 2010.

[96] J. Obłój. The Skorokhod embedding problem and its offspring. *Probab. Surv.*, 1:321–390 (electronic), 2004.

[97] E. Omey. On the rate of convergence in extreme value theory. In Vladimir Kalashnikov and Vladimir Zolotarev, editors, *Stability Problems for Stochastic Models*, volume 1412 of *Lecture Notes in Mathematics*, pages 270–279. Springer Berlin / Heidelberg, 1989. 10.1007/BFb0084179.

[98] E. Omey and S. Rachev. On the rate of convergence in extreme value theory. *Teor. Veroyatnost. i Primenen.*, 33(3):601–607, 1988.

[99] E. Parzen. Some recent advances in time series modeling. *IEEE Trans. Automatic Control*, AC-19:723–730, 1974. System identification and time-series analysis.

[100] M. Peligrad and S. Utev. A new maximal inequality and invariance principle for stationary sequences. *Ann. Probab.*, 33(2):798–815, 2005.

[101] M. Peligrad and S. Utev. Central limit theorem for stationary linear processes. *Ann. Probab.*, 34(4):1608–1622, 2006.

[102] M. Peligrad, S. Utev, and W. B. Wu. A maximal lp inequality for stationary sequences and its applications.

[103] P. Perron. Dealing with structural breaks. Boston University - Department of Economics - Working Papers Series WP2005-017, Boston University - Department of Economics, April 2005.

[104] M. B. Priestley. *Nonlinear and nonstationary time series analysis*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1988.

[105] M. H. Quenouille. A large-sample test for the goodness of fit of autoregressive schemes. *J. Roy. Statist. Soc. (N.S.)*, 110:123–129, 1947.

[106] M. Reed and B. Simon. *Methods of modern mathematical physics. IV. Analysis of operators*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1978.

[107] W. Rhee and M. Talagrand. Uniform bound in the central limit theorem for Banach space valued dependent random variables. *J. Multivariate Anal.*, 20(2):303–320, 1986.

[108] J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465 – 471, 1978.

[109] A. I. Sahanenko. Estimates of the rate of convergence in the invariance principle. *Dokl. Akad. Nauk SSSR*, 219:1076–1078, 1974.

[110] V.V. Senatov. *Normal approximation: new results, methods and problems*. Modern Probability and Statistics. VSP, Utrecht, 1998. Translated from the Russian manuscript by A. V. Kolchin.

[111] J. Shao. An asymptotic theory for linear model selection. *Statist. Sinica*, 7(2):221–264, 1997. With comments and a rejoinder by the author.

[112] X. Shao and W. B. Wu. Asymptotic spectral theory for nonlinear time series. *Ann. Statist.*, 35(4):1773–1801, 2007.

[113] X. Shao and W. B. Wu. A limit theorem for quadratic forms and its applications. *Econometric Theory*, 23(5):930–951, 2007.

[114] R. Shibata. Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, 63(1):117–126, 1976.

[115] G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[116] G. Shwarz. Estimating the dimension of a model. *Ann. Statist.*, 6(2):461–464, 1978.

[117] A. V. Skorokhod. *Studies in the theory of random processes*. Translated from the Russian by Scripta Technica, Inc. Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965.

[118] V. Strassen. Almost sure behavior of sums of independent random variables and martingales. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66)*, pages Vol. II: Contributions to Probability Theory, Part 1, pp. 315–343. Univ. California Press, Berkeley, Calif., 1967.

[119] H. Tong. Some comments on the canadian lynx data. *Journal of the Royal Statistical Society. Series A (General)*, 140(4):pp. 432–436, 1977.

[120] H. Tong. *Nonlinear time series*, volume 6 of *Oxford Statistical Science Series*. The Clarendon Press Oxford University Press, New York, 1990. A dynamical system approach, With an appendix by K. S. Chan, Oxford Science Publications.

[121] R. S. Tsay. *Analysis of financial time series*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2005.

[122] A. M. Walker. Some properties of the asymptotic power functions of goodness-of-fit tests for linear autoregressive schemes. *J. Roy. Statist. Soc. Ser. B.*, 14:117–134, 1952.

[123] G. WALKER. On periodicity in series of related terms. *Monthly Weather Review*, 59(7):277–278, 1931.

[124] M. Wan and W.B Wu. On linear processes with dependent innovations. *Stochastic Process. Appl.*, 115(6):939–958, 2005.

[125] G. S. Watson. Extreme values in samples from m-dependent stationary stochastic processes. *Ann. Math. Statistics*, 25:798–800, 1954.

[126] P. Whittle. Tests of fit in time series. *Biometrika*, 39:309–318, 1952.

[127] P. Whittle. On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika*, 50:129–134, 1963.

[128] M. Woodroofe and W. B. Wu. Martingale approximations for sums of stationary processes. *Ann. Probab.*, 32(2):1674–1690, 2004.

[129] W. B. Wu. Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA*, 102(40):14150–14154 (electronic), 2005.

[130] W. B. Wu. Strong invariance principles for dependent random variables. *Ann. Probab.*, 35(6):2294–2320, 2007.

[131] W. B. Wu. An asymptotic theory for sample covariances of Bernoulli shifts. *Stochastic Process. Appl.*, 119(2):453–467, 2009.

[132] W. B. Wu, Y. Huang, and W. Zheng. Covariances estimation for long-memory processes. *Adv. in Appl. Probab.*, 42(1):137–157, 2010.

[133] W.B. Wu. Nonlinear system theory : Another look at dependence. *Proceedings of the National Academy of Sciences USA.*, 102:14150–14154, 2005.

[134] U.G. Yule. On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers.

[135] A. Zaitsev. Estimates for the rate of strong approximation in the multidimensional invariance principle. *Journal of Mathematical Sciences*, 145:4856–4865, 2007. 10.1007/s10958-007-0319-7.

[136] A. Yu. Zaitsev. Estimates of the lévy-prokhorov distance in the multivariate central limit theorem for random variables with finite exponential moments. *I-III, Theory Probab. Appl*, 31:203—-220, 1986.

[137] A. Yu. Zaitsev. Multidimensional version of a result of sakhanenko in the invariance principle for vectors with finite exponential moments. *I-III, Theory Probab. Appl*, 45:624–641, 1998.

[138] A. Yu. Zaitsev. Multidimensional version of the results of komlos, major and tusnady for vectors with finite exponential moments. *ESAIM: PS*, 2:41–108, 1998.

[139] V. M. Zolotarev. Ideal metrics in the problem of approximating the distributions of sums of independent random variables. *Teor. Verojatnost. i Primenen.*, 22(3):449–465, 1977.