Spring 2023

ON SPECTRAL DENSITY-BASED GOODNESS-OF-FIT TESTS FOR TIME SERIES MODELS

Biao Zhang
University of New Hampshire, Durham

ON SPECTRAL DENSITY-BASED GOODNESS-OF-FIT TESTS FOR TIME SERIES MODELS

BY

BIAO ZHANG

B.S., Southwest University, CHINA, 2012
M.S., The George Washington University, USA, 2014

DISSERTATION

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy of
in
Statistics

May, 2023

This dissertation has been examined and approved in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics by:

Dissertation Director, Dr. Linyuan Li, Professor of Statistics
Department of Mathematics and Statistics
University of New Hampshire

Dr. Qi Zhang, Associate Professor of Statistics
Department of Mathematics and Statistics
University of New Hampshire

Dr. Rita Hibschweiler, Professor of Mathematics
Department of Mathematics and Statistics
University of New Hampshire

Dr. Karen Graham, Professor of Mathematics
Department of Mathematics and Statistics
University of New Hampshire

Dr. Junhao Shen, Professor of Mathematics
Department of Mathematics and Statistics
University of New Hampshire

On May, 2023

Original approval signatures are on file with the University of New Hampshire Graduate School.

DEDICATION

To my parents.

ACKNOWLEDGEMENTS

This dissertation would not have been accomplished without the help and support that I gained at the University of New Hampshire. Especially during the hard time in pandemic, it provided me with a quiet and peaceful environment to put all my concentration and focus on the research.

I would like to give special thanks to my advisor Dr. Linyuan Li, for all of his inspiring guidance, who helped me through my rigorous doctoral research. I would also like to thank my committee members: Dr. Junhao Shen, Dr. Karen Graham, Dr. Qi Zhang and Dr. Rita Hibschweiler. Especially for Dr. Rita Hibschweiler, even in my hardest time, she always encouraged me and believed in what I would finally achieve.

I received generous financial and academic support from the University of New Hampshire during my graduate study. I appreciate the teaching assistant offer by the Department of Mathematics and Statistics at University of New Hampshire. I am also thankful to the summer teaching fellowship granted by the Graduate School in 2021 which made my summer research possible.

In the end, I am truly grateful for my parents. Without their supports, I cannot be where I am right now. This paper is for my beloved family.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

ON SPECTRAL DENSITY-BASED GOODNESS-OF-FIT TESTS FOR TIME SERIES MODELS

by

Biao Zhang

University of New Hampshire, May, 2023

Advisor: Dr. Linyuan Li

The goodness-of-fit tests for time series models have been discussed for years. Most of the current goodness-of-fit tests require the calculation of residuals from the fitted model. Such calculations can be tedious, and in some cases, the tests are even unreliable. Furthermore, most of their results require assumptions that rule out long memory processes, whose autocorrelation decays at a hyperbolic rate as the lag increases and has become popular in time series modeling. In this thesis, a new goodness-of-fit test for time series data is proposed. The test statistic is based on the ratio between the periodogram and the parametric spectral density or its estimator under the null. The asymptotic distribution of the proposed test statistic is derived and its power properties are discussed. Unlike most current goodness-of-fit tests, the asymptotic distribution of our test statistic allows the null hypothesis to be either a short- or long-range dependence model. As our test is in the frequency domain, it is easy to compute, and it does not require the calculation of residuals from the fitted model. The finite sample performance of the test is investigated through simulation experiments.

INTRODUCTION

In applications, overall tests of fit are frequently based on the examination of the correlation behavior of model residuals, such as the Portmanteau type tests which are based on the examination of the M, squared residual autocorrelations. A particular (and popular) example in this context is the test of fit proposed by Box & Pierce (1970).

Box & Pierce (1970) give detailed discussions on the relationship of autocorrelation function obtained by errors and their estimates (residuals) respectively. They mainly discuss about the general autoregressive-integrated moving average (ARIMA) model. As motivated by the fact that the residuals from a correct fit should resemble the true errors of the process, Box & Pierce (1970) set their designed test as the autocorrelation function of residuals to check adequacy of fit. Based on the assumption of independent normal "white noise", it is well known that the autocorrelation function of true errors for moderate or large sample size n possess a multivariate normal distribution. As Box & Pierce (1970) reveal the autocorrelation function of residuals can be represented as a linear transformation of the autocorrelation function of errors, the asymptotic distribution of their designed test is easily determined. Although in their Monte Carlo simulation, the empirical mean of the null distribution is significantly different from its true mean, they argue that it is caused by a trivial approximation made in their designed test, which can be corrected by suitable modifications. Moreover, since the autocovarinace matrix of the sample autocorrelation departs the most from the common variance of $1/n$ for white noise autocorrelation, indicating ignoring their difference can be a serious underestimation of significance and a failure to detect lack of fit, they warn that extra cares need to be taken in their use in diagnostic checking.

Later the adequacy of the designed test in Box & Pierce (1970) was questioned by Prothero & Wallis (1976), Davies et al. (1977) further. Then a modified test is given by Ljung & Box (1978). In consideration of the asymptotic distribution, they claim that the reduction in the location bias brought by the modified test results in a markedly improved approximation, which should be adequate for most practical purposes. Unlike the noise variables ε_t previously are normally distributed, the modified test is insensitive to the departures from normality of the ε_t's. In their simulation studies, for both exponential and uniform distribution, the results agree closely with those obtained under the normality assumption. However, for its test statistics which has better approximation on asymptotic distribution than before, it is only supported by simulation, no formal justification is available in the literature.

Then Hong (1996) proposes some consistent tests, which are useful when no prior information about the true alternative is available. He compares a kernel-based normalized spectral density estimator to the null normalized spectral density, using a quadratic norm, the Hellinger metric, and the Kullback-Leibler information criterion respectively. The null limit distribution of those tests are all $N(0, 1)$ and they have the following breakthroughs in comparison with the previous conventional tests. The residuals $\hat{\varepsilon}_t$ (i.e. the use of residuals $\hat{\varepsilon}_t$ in place of ε_t) have no impact on the limiting distribution. This conclusion is in sharp contrast to those of the Box & Pierce (1970) and Ljung & Box (1978) tests. And the Box & Pierce (1970) test only applies applies to autoregressive moving average (ARMA) model of finite orders, it is derived under the normality assumption. Although the later Ljung & Box (1978) test is robust to nonnormality and this statement is supported by the simulation, the formal proof is not provided. The derived test in Hong (1996) remains invariant when regressors include both lagged dependent variables and exogenous variables, and it also keeps invariant with nonnormality. And Hong (1996)'s approach also provides an interpretation for the Box & Pierce (1970) test, which can be viewed as a test based on a quadratic norm with the use of a truncated periodogram. Moreover, unlike the fact that Box & Pierce (1970) puts equal weight for all sample autocorrelations, the choice of kernels and bandwidth adds more flexibility on the test, therefore we should expect the Hong (1996) test to have better power. In its simulation, the choice of bandwidth p_n has significantly effects on size and power, a faster p_n gives better size, while a slower p_n gives better power. Hong (1996) also considers the fractionally differenced ARIMA (0, 0.35, 0) process, whose autocorrelation decays at hyperbolic rate as the lag increases. But he does not discuss the long memory process in detail. He just shows that the tests have better power than other conventional tests in the long memory case in simulation.

It is clear to see that the tests we've discussed so far have the following limits and restrictions. They are all time-domain methods for detecting model misspecification where calculation of residuals can be tedious. And residuals are not uniquely defined when the model does not have a finite-order autoregressive representation. Moreover, they rule out the long memory process, which arises from aggregation of time series data and has become popular in time series modeling.

A test statistic that circumvents the computation of residuals from the fitted model is proposed by Milhoj (1981). The new test statistic $M_n^d = \{\sum_{j=1}^{N-1} V_j\}^{-2} \sum_{j=1}^{N-1} V_j^2$ is defined in the frequency domain, where $V_j = I(\lambda_j)/f(\lambda_j)$, and $I(\lambda) = (2\pi N)^{-1} | \sum_{t=1}^{N} x_t e^{-i\lambda t}|^2$ is the periodogram of the observations, $\lambda_j = 2\pi j/n$ is the jth Fourier frequency. The test statistic M_n^d is easily calculated since the periodogram can be found by fast Fourier transform. Its asymptotic distribution is normal distribution with mean $1/\pi$ and variance $2\pi^{-2}N^{-1}$, which does not depend on any of the parameters involved; the distribution only depends on N, the number of observations. When the parameter θ in $f(\cdot)$ is estimated, unlike the Ljung & Box (1978) test, the asymptotic distribution of Milhoj (1981) remains unchanged. However, in simulation, the Ljung & Box (1978) test is more sensitive than the Milhoj (1981) test, especially for a large number of observations. This is explained by the fact that the Ljung & Box (1978) test uses m residual autocorrelations, where m can be determined by the specific case, which adds more flexibility on the test, whereas the Milhoj (1981) test uses all residual autocorrelations. In other words, in the case when the residual autocorrelations die out quickly, the extra autocorrelations mainly add more variances to M_n^d.

Following Milhoj (1981), Beran (1992) extends Milhoj's results to long memory time series models that have unbounded spectral densities at the origin. Instead of restricting attention to a fixed set of correlations, Beran (1992) considers all estimable correlations together. This seems particularly appropriate for long memory process, since for these models single correlations can be small whereas their decay to 0 is very slow. To verify the statement, Beran (1992) applies the method to the yearly minimum water levels of the Nile river. This famous data set was one of the major motivations for Mandelbrot to introduce fractional Gaussian noise into statistics. But the data itself is not complicated, a simple fractional Gaussian noise model with only one parameter is able to fit the data well.

It is important to note that Beran (1992) obtains his results by claiming that M_n^d is asymptotically equivalent to its integral version $M_n = \{\int_0^{2\pi} V(\lambda)\, d\lambda\}^{-2} \int_0^{2\pi} V^2(\lambda)\, d\lambda$. Unfortunately, Deo & Chen (2000) show that even in the case of Gaussian white noise, M_n^d and M_n do not have the same asymptotic distribution and that the variance of the asymptotic distribution of M_n is two-thirds that of the variance of the asymptotic distribution of M_n^d. This would seem to suggest that test based on M_n should have greater power than those based on M_n^d and should be preferred. This is verified in simulations. However, in the simulation, both tests lack power in detecting the autoregressive fractionally integrated moving average (ARFIMA) process when the sample size is 100 and almost biased in this case. And it also shows, the finite sample variance of both M_n^d and M_n tend to underestimate (be smaller than) their respective asymptotic values. Therefore both tests are undersized at both the 5% and the 10% level of significance.

Then another frequency-domain test is proposed by Chen & Deo (2004), in which they extend the Hong (1996) test to the long memory process. The null distribution for short memory and long memory, for the true parameter θ_0 and the estimated parameter $\hat{\theta}$ are all asymptotic normal. In simulations, the test performance of rejection rates under H_0 is better than the respective tests in Hong (1996) and Milhoj (1981). Meanwhile both Chen & Deo (2004) and Hong (1996)'s tests have significantly higher power than the Milhoj (1981) test. Although the Chen & Deo (2004) test shows competitive results in simulations, when they establish the asymptotic distribution of their test for long memory process, the proof is based on the following assumption,

(iii) There exists a constant C with

$$|f(\boldsymbol{\theta_1}, \lambda) - f(\boldsymbol{\theta_2}, \lambda)| \leq C\|\boldsymbol{\theta_1} - \boldsymbol{\theta_2}\| f(\boldsymbol{\theta_2}, \lambda)$$

uniformly for all λ and all $\boldsymbol{\theta_1} = (\boldsymbol{\beta_1}, d_1)^\top$ and $\boldsymbol{\theta_2} = (\boldsymbol{\beta_2}, d_2)^\top$ such that $d_1 < d_2$.

This assumption originated in Dahlhaus (1989). When he shows efficiency and consistency of the exact maximum likelihood estimator for the parameters of a long range dependent Gaussian process, he set the above assumption for all long memory processes. However, Dahlhaus (2006) later removes this assumption from his original proof. Because this assumption is not valid, it rules out certain long range dependent processes.[See Dahlhaus (2006)]. From this perspective, the proof in Chen & Deo (2004) is not valid either. In their proof, they use this assumption multiple times. They may argue their test is still powerful based on its good performance in the simulation. However, to support the statement, their proofs are necessary to be corrected.

This thesis deals with another general goodness-of-test procedures, i.e. procedures which can be applied when no a priori information is available about what departures from the null should be anticipated; the asymptotic distribution of our test statistic does not rely on the condition (iii) given above, indicating it can be applied to the general long-memory case. The test proposed in this paper is based on the property according to which, if the model is correct, then for each non-zero frequency the asymptotic expected value of the ratio between the sample spectral density (periodogram) and the spectral density of the model equals one. The idea is to estimate the expected value of this ratio non-parametrically and then to compare the estimate obtained with its expected value under the assumption that the parametric model is correct. To evaluate the discrepancies between these two quantities globally an integrated squared deviation measure is used. Moreover, the distribution of the test statistic has two different expressions for the simple null hypothesis and the composite hypothesis respectively. They are given detailed and comprehensive discussions in this paper.

The organization of the dissertation is as follows. In Chapter 1, we recall briefly the elements of Fourier expansion. The asymptotic distribution of the coefficients of cosine series expansion are derived for testing a simple and a composite hypothesis, which are asymptotically normal distribution. Unlike the covariance matrix of the asymptotic distribution under a simple hypothesis, which is a simple identity matrix, indicating those coefficients are independent when sample size is large enough, under a composite hypothesis, the asymptotic distribution contains a specific covariance matrix which has to be calculated. Thus extra details are given in this section. Then based on those asymptotic results for those empirical coefficients, new test statistics are proposed. Chapter 2 presents all simulation studies under the null hypothesis and for several alternative hypotheses. We demonstrate that our proposed tests are very comparable to the current tests, especially to the most recent Chen & Deo (2004) test. We conclude in Chapter 3 with some remarks, while all technical proofs are provided in Chapter 4.

CHAPTER I

PRELIMINARIES AND THE TESTING PROBLEM

1.1 Model Diagnostics and The Testing Problem

In practice, building a time series model (ARIMA, ARFIMA, etc.), involves a list of steps. Analysts always attempt to visualize data or do some brief exploratory data analysis at the beginning. For time series data, plots help to determine whether the time series is stationary and the related transformation is necessary. When the stationarity is satisfied, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) can be used to identify the dependence orders of the model. Then the estimation is an important step. In this paper, we mainly consider the following two major methods for estimation, the Gaussian-Maximum-Likelihood estimation (MLE, as a time-domain method, it has be widely used in time series analysis), and the Quasi-Maximum-Likelihood estimation (QML, as a frequency-domain method, owe to the fast Fourier transform (FFT) algorithm, its computation efficiency has been significantly improved from $O(n^3)$ to $O(n \log n)$ in contrast to MLE). Details about this part will be given in later sections. In the end, there are several tests of randomness provided for model diagnostics. As discussed in the preceding chapter, time-domain tests in Box & Pierce (1970), Ljung & Box (1978) and Hong (1996) have their own limits and restrictions. This paper focuses on designing a new frequency-domain test to compete with the current ones.

In this paper we assume that time series X_t, $t = 1, 2, \cdots, n$ is a realization of a linear process

$$X_t = \sum_{j=0}^{\infty} \psi_j \zeta_{t-j}, \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty, \tag{1.1}$$

where $\{\zeta_t\}$ is a sequence of white noise with mean zero and variance σ^2. The class of the above linear process $\{X_t\}$ is very large. From Giraitis et al. (2012, p.38), every stationary process $\{X_t\}$ with mean zero, whose spectral density $f(\cdot)$ satisfies $\int_{-\pi}^{\pi} \log f(u) du > -\infty$, can be represented as a linear process above.

Known as the fundamental frequency domain tool, the spectral density can be written as

$$f(u) = \frac{\sigma^2}{2\pi} \left| \sum \psi_j e^{\mathtt{i} ju} \right|^2,$$

Denote a finite dimensional parametric class of spectral densities by $\mathcal{F} = \{f(\cdot, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$, where Θ is a parameter space for $\boldsymbol{\theta}$. To present the essential ideas of the test statistic, we consider first the basic case of testing a simple hypothesis

$$\begin{aligned} H_0 : \ f(u) &= f(u, \boldsymbol{\theta}_0), \qquad \text{for all } u \in (0, \pi), \\ H_1 : \ f(u) &\neq f(u, \boldsymbol{\theta}_0), \qquad \text{for some } u \in (0, \pi), \end{aligned} \tag{1.2}$$

where $f(\cdot)$ is the spectral density for observed process $\{X_t\}$ and $f(\cdot, \boldsymbol{\theta}_0) \in \mathcal{F}$ with that $\boldsymbol{\theta}_0$ is a fixed value and assumed to be known.

1.2 The Construction and Consistency of The Test Statistic

According to discussions in the preceding sections, let $h(u) = f(u)/f(u, \boldsymbol{\theta}_0)$. Then the preceding hypotheses (1.2) become

$$H_0 : h(u) = 1, \text{ for all } u \in (0, \pi), \qquad H_1 : h(u) \neq 1, \text{ for some } u \in (0, \pi). \tag{1.3}$$

For any periodic even function $h(\cdot) \in L^2([0,\ \pi])$, using basis $\{1/\sqrt{\pi}, \sqrt{2/\pi}\cos(ku), k = 1, 2, \cdots\}$, we have its following Fourier expansion

$$h(u) = \frac{1}{\sqrt{\pi}}\beta_0 + \sqrt{\frac{2}{\pi}}\sum_{k=1}^{\infty}\beta_k \cos(ku),$$

where $\beta_0 = \int_0^\pi h(u)/\sqrt{\pi}du$ and $\beta_k = (\sqrt{2}/\sqrt{\pi})\int_0^\pi h(u)\cos(ku)du$.

Under null hypotheses H_0 in (1.3) (or in (1.2)), one has $\beta_0 = \sqrt{\pi}$ and $\beta_k = 0$ for all $k > 1$, i.e., all the true coefficients except β_0 vanish. Under alternative hypothesis H_1, β_0 may not be $\sqrt{\pi}$ or there exist some k, such that $\beta_k \neq 0$. Therefore, the original null hypothesis H_0 in (1.3) is equivalent to the following hypothesis with the above coefficients:

$$\begin{aligned} H_0 :&\ \ \beta_0 = \sqrt{\pi},\ \beta_k = 0 \text{ for all } k \geq 1, \\ H_1 :&\ \ \beta_0 \neq \sqrt{\pi},\ \textit{or}\ \beta_k \neq 0 \text{ for some } k \geq 1. \end{aligned} \tag{1.4}$$

For all N observations, let $n = 2[N/2]$. When N is odd, we drop one data for notation simplicity. This is suggested by Moulines & Soulier (1999), they set $n_m = 2m[N/2m]$ for a fixed integer m, and claim this is asymptotically irrelevant since $n_m/n \to 1$.

Regarding the hypothesis testing problem (1.2), it is well known that the basic tool to make inference on the spectral density $f(u)$ of $\{X_t\}$ is its periodogram

$$I_X(u) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} X_t e^{-\mathrm{i}tu}\right|^2.$$

Also, denote $u_j = 2\pi j/n$, the Fourier frequencies, $j = 1, 2, \cdots, \nu, \nu = \nu_n = n/2 - 1$. We propose estimators for the above coefficients as

$$\tilde{\beta}_0 = \frac{2\sqrt{\pi}}{n} \sum_{j=1}^{\nu} \frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)}, \ \tilde{\beta}_k = \frac{2\sqrt{2\pi}}{n} \sum_{j=1}^{\nu} \cos(ku_j) \frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)}, \ k = 1, 2, \cdots, \nu/2. \tag{1.5}$$

In order to derive properties of the above empirical coefficients and control the bound on the covariance of periodogram ordinates, we need to impose some conditions on those coefficients ψ_j and innovations ζ_t in (1.1). In specific, we require the following assumptions.

A1: $\sum_j |j|^{1/2+\tau} |\psi_j| < \infty$ for an arbitrary small positive $0 < \tau < 1/2$.

Note above assumption **A1** is slightly stronger than $\sum_j |j|^{1/2} |\psi_j| < \infty$, which is usually assumed for short memory linear process.

A2: Innovations $\{\zeta_t, t \in \mathbb{Z}\}$ are *i.i.d.* random variables with $\mathrm{E}[\zeta_t] = 0$, $\mathrm{E}[\zeta_t^2] = \sigma_0^2$ and $u_4 = \mathrm{E}[\zeta_t^4] < \infty$.

We give some brief additional introductions on the notations of fourth order cumulant function and its spectral density, since it is mentioned frequently in later sections. Let assume

$$\sum_{h_1, h_2, h_3 \in \mathbb{Z}} \left| K(h_1, h_2, h_3) \right| < \infty,$$

where $K(h_1, h_2, h_3) = \mathrm{Cum}(X_0, X_{h_1}, X_{h_2}, X_{h_3})$ for $h_1, h_2, h_3 \in \mathbb{Z}$ is the fourth order cumulant function of the process X. Under the above condition, we may define the fourth order cumulant spectral density function as

$$\kappa(\lambda_1, \lambda_2, \lambda_3) = (2\pi)^{-3} \sum_{h_1 \in \mathbb{Z}} \sum_{h_2 \in \mathbb{Z}} \sum_{h_3 \in \mathbb{Z}} K(h_1, h_2, h_3) e^{-\mathrm{i}(\lambda_1 h_1 + \lambda_2 h_2 + \lambda_3 h_3)}.$$

Now, we are ready to present the following limit distributions for the above empirical coefficients in (1.5).

Theorem 1.2.1 Under the null hypothesis H_0 in (1.2), consider the linear process $\{X_t\}$ in (1.1) under conditions **A1**, **A2**, we have, as $n \to \infty$,

$$\begin{aligned}
&\sqrt{n}(\tilde{\beta}_0 - \sqrt{\pi}) \longrightarrow_d N(0, 2\pi q_0),\\
&\sqrt{n}\tilde{\beta}_k \longrightarrow_d N(0, 2\pi), \text{ for all } k = 1, 2, \cdots, m,\\
&\text{Cov}(\sqrt{n}\tilde{\beta}_{k_1}, \ \sqrt{n}\tilde{\beta}_{k_2}) \longrightarrow 0, \text{ for all } k_1 \neq k_2, k_1, k_2 = 1, 2, \cdots, m,\\
&\text{Cov}(\sqrt{n}(\tilde{\beta}_0 - \sqrt{\pi}), \sqrt{n}\tilde{\beta}_k) \longrightarrow 0, \text{ for all } k = 1, 2, \cdots, m,
\end{aligned}$$

where $q_0 = 1 + \text{Cum}_4(e_0)/2$ and m is a fixed integer less than $\nu/2$. Denote $\{e_t, t = 1, \cdots, n\}$ are the standardized innovations, $e_t = \zeta_t/\sigma_0$ and "$\longrightarrow_d$" stands for the convergence in distribution.

Remark 1.2.1 If innovations ζ_t follow the normal distribution, one has $\text{Cum}_4(e_0) = \text{E}[e_0^4] - 3 = u_4 - 3 = 0$. Thus one has $q_0 = 1$.

Since above empirical coefficients $\tilde{\beta}_0$ and $\tilde{\beta}_k$ are consistent estimator of the β_0 and β_k's, it is intuitive to propose the following test statistic $X_m^2(\boldsymbol{\theta}_0)$ for the hypothesis testing problem (1.2):

$$X_m^2(\boldsymbol{\theta}_0) = \frac{n}{2\pi q_0}(\tilde{\beta}_0 - \sqrt{\pi})^2 + \frac{n}{2\pi}\sum_{k=1}^{m}\tilde{\beta}_k^2.$$

From Theorem 1.2.1 and applying the continuous mapping theorem, we obtain the following results.

Theorem 1.2.2 Under the null hypothesis H_0 in (1.2) and consider the linear process $\{X_t\}$ in (1.1) under conditions **A1**, **A2**, we have, for any fixed m such that $1 \leq m < \nu/2$,

$$X_m^2(\boldsymbol{\theta}_0) \longrightarrow_d \chi^2(m+1), \text{ as } n \to \infty.$$

1.3 The Consistency of Whittle Likelihood Estimates

In applications, the null hypothesis of interest is the composite hypothesis that the process has spectral density $f(\cdot, \boldsymbol{\theta})$ for some unknown $\boldsymbol{\theta}$ in the parameter space Θ. Under this null, the empirical coefficients become

$$\hat{\beta}_0 = \frac{2\sqrt{\pi}}{n} \sum_{j=1}^{\nu} \frac{I_X(u_j)}{f(u_j, \hat{\boldsymbol{\theta}})}, \;\; \hat{\beta}_k = \frac{2\sqrt{2\pi}}{n} \sum_{j=1}^{\nu} \cos(k u_j) \frac{I_X(u_j)}{f(u_j, \hat{\boldsymbol{\theta}})}, \;\; k = 1, 2, \cdots, m, \tag{1.6}$$

and $\hat{\boldsymbol{\theta}}$ is some estimator of $\boldsymbol{\theta}$ based on the sample $X_1, \cdots, X_n$, m is a fixed integer less than $\nu/2$.

There are two main estimation methods which have been widely used in applications. One is a frequency-domain method, the Whittle-Likelihood(or Quasi-Maximum-Likelihood) estimation. Another is a time-domain method, the Maximum-Likelihood estimation. Both estimation methods bring changes on the asymptotic null distributions of $\hat{\beta}_k$, they are quite different with that of $\tilde{\beta}_k$, $k = 0, 1, \cdots, m$, in the preceding Theorem 1.2.1.

We first discuss the Whittle likelihood estimator. Hoaoya (1974) firstly proposed to minimize the following Whittle likelihood function

$$\mathcal{L}_n^W(\boldsymbol{\theta}) = \int_{-\pi}^{\pi} \left\{ \log f(u, \boldsymbol{\theta}) + \frac{I_X(u)}{f(u, \boldsymbol{\theta})} \right\} du, \tag{1.7}$$

with respect to $\boldsymbol{\theta}$ in order to find an estimator of $\boldsymbol{\theta}$. Let $\hat{\boldsymbol{\theta}}_W$ be a value that minimizes (1.7). Hoaoya (1974) also derived the asymptotic distribution of $\hat{\boldsymbol{\theta}}_W$ under appropriate regularity conditions. The above likelihood function is also known as an application of the Kullback-Leibler distance function, details can be found in Dahlhaus & Wefelmeyer (1996). Later Hosoya & Masanobu (1982) give a natural extension of (1.7) to multivariate processes, and developed the asymptotic theory of the related $\hat{\boldsymbol{\theta}}_W$.

For better understanding in later sections, we give more detailed introductions on the part related to the Kullback-Leibler distance function. We consider the following function,

$$D(\boldsymbol{\theta}, f) = \int_{-\pi}^{\pi} \left\{ \log f(u, \boldsymbol{\theta}) + \frac{f(u)}{f(u, \boldsymbol{\theta})} \right\} du, \tag{1.8}$$

where $f(\cdot)$ is defined as in (1.2), the spectral density of linear process $\{X_t\}$ in (1.1). And we fit $f(u, \boldsymbol{\theta}) \in \mathcal{F}$ to $f(u)$ by the criterion $D(\boldsymbol{\theta}, f)$ in (1.8), where $\mathcal{F} = \{f(\cdot, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$, it is a family of parametric spectral density models, which may not contain the true model. Let $\bar{\boldsymbol{\theta}} \in \Theta$ be value defined by

$$D(\bar{\boldsymbol{\theta}}, f) = \min_{\theta \in \Theta} D(\boldsymbol{\theta}, f).$$

To estimate $\bar{\boldsymbol{\theta}}$ we use $\hat{\boldsymbol{\theta}}_W$, by the requirement

$$D(\hat{\boldsymbol{\theta}}_W, f) = \min_{\theta \in \Theta} D(\boldsymbol{\theta}, I_X).$$

This is where the likelihood function (1.7) comes from. Note that if $\bar{\boldsymbol{\theta}}$ is the unique minimizer, under H_0 in (1.2) we have $f(u) = f(u, \boldsymbol{\theta}_0) = f(u, \bar{\boldsymbol{\theta}})$. The above context is widely discussed in Taniguchi & Kakizawa (2000, Section 3.1), Hosoya & Masanobu (1982), McElroy & Holan (2009) and Dahlhaus & Wefelmeyer (1996).

The remaining discussions in this section focus on the empirical coefficients in (1.6) through using Whittle likelihood estimator $\hat{\boldsymbol{\theta}}_W$ under the null composite hypothesis. We first represent the related asymptotic theory of $\hat{\beta}_k$, $k = 1, 2, \cdots, m$.

To show the theorem, we need to add some additional assumptions.

B1: Let Θ be a compact and convex finite-dimensional parameter space. Let the spectral density of the process be $\{f(\cdot, \boldsymbol{\theta}_0)\}$, where $\boldsymbol{\theta}_0$ is the true parameter vector that lies in the interior of Θ.

B2: $\bar{\boldsymbol{\theta}}$ denotes the minimizer of (1.8), suppose this pesudo-true value exists uniquely and lies in the interior of Θ. And $M_f = \int_{-\pi}^{\pi} \left[\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \frac{f_{\theta_0}(u)}{f_\theta(u)} + \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log f_{\boldsymbol{\theta}}(u) \right]_{\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}} du$, is the Hessian of Kullback-Leibler information divergence between $f_{\boldsymbol{\theta}}$ and f is a nonsingular matrix.

B3: The linear process $\{X_t\}$ satisfies those Hosoya-and-Taniguchi conditions describes in Taniguchi & Kakizawa (2000, p.55), **HT1-HT3** and **HT5**.

B4: The spectral density $f(u, \boldsymbol{\theta})$ satisfies the following conditions for $(\theta, u) \in \Theta \times [-\pi, \pi]$:

(i) $f(u, \boldsymbol{\theta})$ and $f^{-1}(u, \boldsymbol{\theta})$ are continuous at all $(u, \boldsymbol{\theta})$.

(ii) $\frac{\partial}{\partial \theta_\omega} f^{-1}(u, \boldsymbol{\theta})$ and $\frac{\partial^2}{\partial \theta_\omega \partial \theta_v} f^{-1}(u, \boldsymbol{\theta})$ are continuous and finite at all $(u, \boldsymbol{\theta})$.

Those assumptions are general for the short-memory model. In fact, **B4** is from Chen & Deo (2004). And **B1-B3** are from Theorem 3.1.2 in Taniguchi & Kakizawa (2000), they are used to show the consistency and efficiency of $\hat{\boldsymbol{\theta}}_W$. Following the theorem, the result $\left\|\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0\right\| = O_p(n^{-1/2})$ should be intuitive.

Unlike Chen & Deo (2004), about estimator $\hat{\boldsymbol{\theta}}$, they only give a general assumption $\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right\| = O_p(n^{-1/2})$. Here since we restrict our attention on a specific estimator, Whittle likelihood estimator $\hat{\boldsymbol{\theta}}_W$, **B1-B3** gives more detailed prerequisites for the linear process $\{X_t\}$ and its spectral density $f(\cdot, \boldsymbol{\theta})$.

The next theorem states the asymptotic distribution of the estimated empirical coefficient vector $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_m)$ when $\{X_t\}$ is a short memory process.

Theorem 1.3.1 Under the null hypothesis H_0 in (1.2), consider the linear process $\{X_t\}$ in (1.1) and its spectral density $f(\cdot, \boldsymbol{\theta})$ under conditions **A1**, **A2**, **B1-B4** with $\hat{\boldsymbol{\theta}}_W$ the Whittle likelihood estimator, we have

$$\sqrt{n}\hat{\boldsymbol{\beta}} \longrightarrow_d N(0, \Sigma(\boldsymbol{\theta}_0)),$$

with $n \to \infty$,where $\Sigma(\boldsymbol{\theta}_0)$ with the following entries

$$
\begin{aligned}
\Sigma_{k,l}(\boldsymbol{\theta}_0) &= 2\pi q_0 \mathrm{I}(k=l) - C_k^T(\boldsymbol{\theta}_0)\left\{4\pi\Gamma(\boldsymbol{\theta}_0)^{-1}\right\}C_l(\boldsymbol{\theta}_0),\\
C_k(\boldsymbol{\theta}_0) &= \frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi}\cos(ku)\boldsymbol{\nabla}_{\boldsymbol{\theta}}\log f(u,\boldsymbol{\theta}_0)du,\\
\Gamma(\boldsymbol{\theta}_0) &= \int_{-\pi}^{\pi}\boldsymbol{\nabla}_{\boldsymbol{\theta}}\log f(u,\boldsymbol{\theta}_0)\boldsymbol{\nabla}_{\boldsymbol{\theta}}^T\log f(u,\boldsymbol{\theta}_0)du.
\end{aligned}
$$

Note $k, l = 1, 2, \cdots, m$.

Remark 1.3.1 The matrix inside $\{\cdot\}$ in the covariance matrix $\Sigma(\boldsymbol{\theta}_0)$ is brought by $\sqrt{n}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0)$. Under the preceding simple null hypothesis, we do not have such worries, therefore the corresponding covariance matrix just end up with $2\pi q_0 \boldsymbol{I}_m$, where $\boldsymbol{I}_p$ is a $p\times p$ identity matrix.

In practice, $\Sigma(\boldsymbol{\theta}_0)$ is not provided. We can form the estimate $\Sigma(\hat{\boldsymbol{\theta}}_W)$, which is shown below to be consistent for $\Sigma(\boldsymbol{\theta}_0)$ under the null hypothesis.

Note if $\{X_t\}$ is a non-Gaussian linear process, $q_0 = 1 + \mathrm{Cum}_4(e_0)/2$. We need to do additional work to estimate $\eta_4 = \mathrm{Cum}_4(e_0)$, the related estimator is provided by Fragkeskou & Paparoditis (2016). To ensure the consistency of the estimator of η_4, we need to update our conditions **A1** and **A2** into **A0**,

A0: $\sum_j j^2\psi_j^2 < \infty$, $\phi_0 = 1$. Innovations $\{\zeta_t, t \in \mathbb{Z}\}$ are $i.i.d.$ random variables with $\mathrm{E}[\zeta_t] = 0$, $\mathrm{E}[\zeta_t^2] = \sigma_0^2$ and $\mathrm{E}[\zeta_t^8] < \infty$.

This condition is **Assumption 1** in Fragkeskou & Paparoditis (2016). Thus we have

Proposition 1.3.1 Under the null hypothesis H_0 in (1.2), consider the linear process $\{X_t\}$ in (1.1) and its spectral density $f(\cdot, \boldsymbol{\theta})$ under conditions **A0**, **B1-B4**, we have

$$\Sigma(\hat{\boldsymbol{\theta}}_W) \longrightarrow_p \Sigma(\boldsymbol{\theta}_0).$$

with $n \to \infty$, in the sense that each matrix entry converges in probability.

Lastly to complete our new test statistic, we need to find what changes brought by $\hat{\boldsymbol{\theta}}_W$ into the estimated $\hat{\beta}_0$. This can be easily explained in the following computational way.

In practice, it is normally not possible to solve the estimation equation $\nabla \mathcal{L}_n^W(\boldsymbol{\theta}) = 0$, where $\mathcal{L}_n^W(\boldsymbol{\theta})$ is Whittle likelihood function in (1.7). Instead one would determine the estimate, e.g., by a Newton iteration where integral in $\mathcal{L}_n^W(\boldsymbol{\theta})$ is replaced by a sum over the Fourier frequencies, i.e., one would minimize

$$\mathcal{L}_n^{\dagger}(\boldsymbol{\theta}) = \sum_{j=0}^{\nu} \left\{ \log f(u_j, \theta) + \frac{I_X(u_j)}{f(u_j, \theta)} \right\}. \tag{1.9}$$

It is well known that the resulting estimate $\hat{\boldsymbol{\theta}}_W^{\dagger}$ has the same asymptotic behaviour as $\hat{\boldsymbol{\theta}}_W$. Now in consideration of the estimator $\hat{\boldsymbol{\theta}}_W^{\dagger}$, we can see what changes they bring into the estimated $\hat{\beta}_0$.

Firstly consider the spectral density in the following form $\sigma^2 g(u, \boldsymbol{\phi})$. Such assumption has been made by Hannan (1973), when he gives general results of the consistency and efficiency of the maximum likelihood estimator (MLE) based on a Gaussian Likelihood function for a general autoregressive moving-average (ARMA) time series, which will be discussed in detail in the next section when we consider our estimated empirical coefficients by using MLE $\hat{\boldsymbol{\theta}}_M$. If the spectral density can be expressed in this form, then (1.9) can be written as

$$\begin{aligned} \mathcal{L}_n^{\dagger}(\boldsymbol{\phi}, \sigma^2) &= \sum_{j=0}^{\nu} \left\{ \log \sigma^2 g(u_j, \phi) + \frac{I_X(u_j)}{\sigma^2 g(u_j, \theta)} \right\} \\ &= \frac{n}{2} \log \sigma^2 + \sum_{j=0}^{\nu} \log g(u_j, \boldsymbol{\phi}) + \frac{1}{\sigma^2} \frac{I_X(u_j)}{g(u_j, \boldsymbol{\phi})}. \end{aligned}$$

Suppose $\hat{\beta}_0$ has been put under the same assumptions as those in the Theorem 1.3.1. Given $g(u, \boldsymbol{\phi})$ is sufficiently smooth with respect to $\boldsymbol{\phi}$, and that $(\hat{\boldsymbol{\phi}}^{\dagger}, \hat{\sigma}^{\dagger 2})$ satisfies

$$\frac{\partial}{\partial \sigma^2} \mathcal{L}_n^{\dagger}(\boldsymbol{\phi}, \sigma^2)|_{(\boldsymbol{\phi}, \sigma^2) = (\hat{\boldsymbol{\phi}}^{\dagger}, \hat{\sigma}^{\dagger 2})} = 0$$

We have,

$$\frac{n}{2}\frac{1}{\hat{\sigma}^{\dagger 2}} - \frac{1}{(\hat{\sigma}^{\dagger 2})^2}\sum_{j=0}^{\nu}\frac{I_X(u_j)}{g(u_j,\hat{\boldsymbol{\phi}}^{\dagger})} = 0$$

$$\hat{\sigma}^{\dagger 2} = \frac{1}{n/2}\sum_{j=0}^{\nu}\frac{I_X(u_j)}{g(u_j,\hat{\boldsymbol{\phi}}^{\dagger})}.$$

Then $\hat{\beta}_0$ can be written as

$$\begin{aligned}
\hat{\beta}_0 &= \frac{2\sqrt{\pi}}{n}\sum_{j=1}^{\nu}\frac{I_X(u_j)}{\hat{\sigma}^{\dagger 2}g(u_j,\hat{\boldsymbol{\phi}}^{\dagger})} \\
&= \sqrt{\pi}\frac{\frac{1}{n/2}\sum_{j=1}^{\nu}I_X(u_j)/g(u_j,\hat{\boldsymbol{\phi}}^{\dagger})}{\frac{1}{n/2}\sum_{j=0}^{\nu}I_X(u_j)/g(u_j,\hat{\boldsymbol{\phi}}^{\dagger})} \\
&= \sqrt{\pi}\left(1 - \frac{\frac{1}{n/2}I_X(u_0)/g(u_0,\hat{\boldsymbol{\phi}}^{\dagger})}{\frac{1}{n/2}\sum_{j=1}^{\nu}I_X(u_j)/g(u_j,\hat{\boldsymbol{\phi}}^{\dagger})}\right) \\
&= \sqrt{\pi} + O_p(n^{-1}).
\end{aligned} \tag{1.10}$$

The last equality is ensured by the fact that

$$\frac{2\sqrt{\pi}}{n}\sum_{j=1}^{\nu}\frac{I_X(u_j)}{\sigma_0^2 g(u_j,\hat{\boldsymbol{\phi}}^{\dagger})} \longrightarrow_p \sqrt{\pi}, \quad \frac{I_X(u_0)}{g(u_0,\hat{\boldsymbol{\phi}}^{\dagger})} = O_p(1).$$

The above results can be represented following the proof arguments in Theorem 1.3.1.

Note since the spectral density is assumed to be in the form $\sigma^2 g(u,\boldsymbol{\phi})$, all the preceding conditions **B1-B4** on the spectral density $f(\cdot,\boldsymbol{\theta})$, the parameter $\boldsymbol{\theta}$ and its space Θ are now can be updated to $g(\cdot,\boldsymbol{\phi})$, $\boldsymbol{\phi}$ and the corresponding space Φ.

Summarizing all the results we've obtained so far, it is intuitive to build the following test statistic and its corresponding asymptotic theory with using $\hat{\boldsymbol{\phi}}_W$.

Theorem 1.3.2 Let the spectral density of the linear process $\{X_t\}$ in (1.1) in the form $\sigma^2 g(u,\boldsymbol{\phi})$. Under the null hypothesis H_0 in (1.2), and assumptions **A0**, **B1-B4**, note $\boldsymbol{\beta}_m =$

$(\beta_1, \cdots, \beta_m)$ is a m-vector, for any fixed m such that $1 \leq m < \nu/2$, we have

$$X_m^2(\hat{\boldsymbol{\phi}}_W) = n(\hat{\beta}_0 - \sqrt{\pi})^2 + n\hat{\boldsymbol{\beta}}_m^T \Sigma(\hat{\boldsymbol{\phi}}_W)^{-1} \hat{\boldsymbol{\beta}}_m \longrightarrow_d \chi^2(m), \text{ as } n \to \infty.$$

1.4 The Consistency of Maximum Likelihood Estimates

In this section, we mainly discuss the maximum likelihood estimates. We can formally construct a Gaussian likelihood for $\{X_t\}$,

$$\mathcal{L}_n(\boldsymbol{\theta}) = (2\pi)^{-\frac{1}{2}n} |\Sigma_n|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} X_n^T \Sigma_n^{-1} X_n\right\}, \tag{1.11}$$

where Σ_n is the covariance matrix of $\{X_t, t = 1, \cdots, n\}$. Then the exact maximum likelihood estimates $\hat{\boldsymbol{\theta}}_M$ can be obtained by minimizing (1.11). It is well known that (1.7) in the preceding section is an approximation of $-4\pi n^{-1} \log \mathcal{L}_n(\boldsymbol{\theta})$. The fact that the Whittle likelihood estimator is called as quasi-Gaussian maximum likelihood estimator, stems from "approximated Gaussian likelihoods"(Taniguchi & Kakizawa (2000, p.53)). Before we show Theorem 1.3.1 remain consistent with $\hat{\boldsymbol{\theta}}_M$ the maximum likelihood estimate, we need to add two more assumptions on our $\{X_t\}$ and its spectral density $f(\cdot, \boldsymbol{\theta})$. We shall assume that $\{X_t\}$ is Gaussian. Moreover, we have an additional assumption on the spectral density model.

B5: $f_{\boldsymbol{\theta}}$ and its derivative are uniformly bounded and bounded away from 0.

Note the Gaussian assumption implies all Hosoya-and-Taniguchi conditions in **B3**, therefore we have the following same result as in the preceding section.

Theorem 1.4.1 Under the null hypothesis H_0 in (1.2), consider the Gaussian linear process $\{X_t\}$ in (1.1) and its spectral density $f(\cdot, \boldsymbol{\theta})$ under conditions **A1**, **A2**, **B1**, **B2**, **B4**, **B5** with $\hat{\boldsymbol{\theta}}_M$ the maximum likelihood estimator, we have

$$\sqrt{n}\hat{\boldsymbol{\beta}} \longrightarrow_d N(0, \Sigma(\boldsymbol{\theta}_0)),$$

as $n \to \infty$, where $\Sigma(\boldsymbol{\theta}_0)$ is the same as we described in Theorem 1.3.1.

It is easy to show $\Sigma(\hat{\boldsymbol{\theta}}_M)$ is a consistent estimator of $\Sigma(\boldsymbol{\theta}_0)$ through the smooth condition of $f_{\boldsymbol{\theta}}$. However, the asymptotic theory of $\hat{\beta}_0$ with $\hat{\boldsymbol{\theta}}_M$ is not that intuitive in comparison with that with $\hat{\boldsymbol{\theta}}_W$. Its proof is motivated by the connections between frequency and time domain method, thus we give more detailed discussions on this part in the rest of section.

The following work is mainly based on Hannan (1973). Hannan (1973) established the asymptotic theory for $\hat{\boldsymbol{\theta}}_M$ for a general ARMA time series. It is one of the most influential results in the classical time series. The result is simple and elegant. The imposed conditions are minimal. They are conditions B and (12) in Hannan (1973). They are necessary for the Theorem 3 in Hannan (1973), which gives us a result that $\left\|\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}_0\right\| = O_p(n^{-1/2})$. Some of those conditions also play an important role in the proof of our Theorem 1.4.2, thus we give a brief introduction here.

Firstly, Hannan (1973) assumes the spectral density is in the form

$$f(u) = f_{\boldsymbol{\phi},\sigma^2}(u) = \sigma^2 g(u, \boldsymbol{\phi}). \tag{1.12}$$

Then he assumes $c(h, \boldsymbol{\phi}_0)$, the hth Fourier coefficient of $g(u, \boldsymbol{\phi}_0)$, satisfies the condition (12), $c(h, \boldsymbol{\phi}_0) = O(|h|^{-2-\alpha})$, $\alpha > 0$. Thus we have

$$\sum |h||c(h, \boldsymbol{\phi}_0)| = L < \infty, \tag{1.13}$$

he also has the smoothness condition on $g(u, \boldsymbol{\phi})$, here we consider a stronger one for this part.

C0: Suppose Φ is a finite dimensional parameter space. The $g(u, \boldsymbol{\phi})$ satisfies the following conditions for $(\boldsymbol{\phi}, u) \in \Phi \times [-\pi, \pi]$:

(i) $g(u, \boldsymbol{\phi})$ and $g^{-1}(u, \boldsymbol{\phi})$ are continuous at all $(u, \boldsymbol{\phi})$.

(ii) $\frac{\partial}{\partial \theta_\omega} g^{-1}(u, \boldsymbol{\phi})$ and $\frac{\partial^2}{\partial \theta_\omega \partial \theta_v} g^{-1}(u, \boldsymbol{\phi})$ are continuous and finite at all $(u, \boldsymbol{\phi})$.

Since he puts the spectral density in the form (1.12), the Gaussian likelihood $\mathcal{L}_n(\boldsymbol{\theta})$

(1.11) can be written as

$$(2\pi\sigma^2)^{-\frac{1}{2}n}|G_n|^{-\frac{1}{2}}\exp\left\{-X_n^T G_n^{-1} X_n/2\sigma^2\right\}. \tag{1.14}$$

About the n rowed square matrix G_n, its structure just like the autocovraince matrix Σ_n, but its entries are consisted of $c(h,\boldsymbol{\phi}_0) = \gamma(h,\boldsymbol{\theta}_0)/\sigma_0^2$, where $\gamma(h,\boldsymbol{\theta}_0)$ is the hth Fourier coefficient of $f(\cdot,\boldsymbol{\theta}_0)$, an autocovariance function of a stationary time series $\{X_t\}$. It is intuitive to have the following estimator of σ^2 with the likelihood (1.14),

$$\sigma_n^2(\boldsymbol{\phi}) = n^{-1}X_n^T G_n^{-1}(\boldsymbol{\phi})X_n. \tag{1.15}$$

And $\hat{\boldsymbol{\theta}}_M$ is obtained by minimizing this quadratic form (1.15).

Now we can state the result of $\hat{\beta}_0$ under the above assumptions.

Theorem 1.4.2 Under the null hypothesis H_0 in (1.2), and all the conditions required for Theorem 3 in Hannan (1973) (the smoothness condition on $g(\cdot,\boldsymbol{\phi})$ has been updated by our **C0**), with $\hat{\boldsymbol{\phi}}_M$ the maximum likelihood estimator, we have

$$\hat{\beta}_0 = \sqrt{\pi} + \sqrt{\pi}\frac{\sigma_n^2(\boldsymbol{\phi}_0) - \sigma_n^2(\hat{\boldsymbol{\phi}}_M)}{\sigma_n^2(\hat{\boldsymbol{\phi}}_M)} + O_p(n^{-1}).$$

The above result is intriguing. From Brockwell & Davis (1991, Section 8.9), the second term is used to build the confidence intervals for the $p+q$ parameters of Gaussian causal invertible ARMA process.

Consider $\{X_t\}$ follows a stationary process defined by

$$X_t - b_1X_{t-1} - \cdots - b_pX_{t-p} = \zeta_t + a_1\zeta_{t-1} + \cdots + a_q\zeta_{t-q}, \tag{1.16}$$

where $\zeta_t \sim$ i.i.d $N(0,\sigma^2)$.

As Yao & Brockwell (2006) claim, for the below two operations,

$$b(z) = 1 - b_1 z - \cdots - b_p z^p \text{ and } a(z) = 1 + a_1 z + \cdots + a_q z^q,$$

if $b(z)a(z) \neq 0$ for all $|z| = 1$, there are 2^{p+q} stationary ARMA(p,q) models (with different $\boldsymbol{b}$ and $\boldsymbol{a}$) sharing the same autocorrelation function (ACF). To avoid the ambiguity, it is common practice to assume that $b(z) \neq 0$ and $a(z) \neq 0$ for all $|z| \leqslant 1$. This assumption guarantees that the parameters $\boldsymbol{b}$ and $\boldsymbol{a}$ are identifiable in terms of the ACF, which is necessary condition in the context of Gaussian MLE since the likelihood function depends the parameter $(\boldsymbol{b}, \boldsymbol{a})$ through the ACF only.

From Brockwell & Davis (1991, (8.9.7)), now we have

$$\frac{n-p-q}{p+q} \cdot \frac{\sigma_n^2(\boldsymbol{\phi}_0) - \sigma_n^2(\hat{\boldsymbol{\phi}}_M)}{\sigma_n^2(\hat{\boldsymbol{\phi}}_M)} \sim \mathrm{F}(p+q, n-p-q).$$

If we apply the Slutsky's theorem, as $n \to \infty$, it is straightforward to see $F(p+q, n-p-q) \to_d \chi^2_{p+q}/(p+q)$, indicating

$$\frac{\sigma_n^2(\boldsymbol{\phi}_0) - \sigma_n^2(\hat{\boldsymbol{\phi}}_M)}{\sigma_n^2(\hat{\boldsymbol{\phi}}_M)} = O_p(n^{-1}).$$

This gives us an asymptotic result of $\hat{\beta}_0$ with $\hat{\boldsymbol{\phi}}_M$, which is consistent with the preceding result (1.10) with $\hat{\boldsymbol{\phi}}_W$. Now we are ready to state our test statistic designed for $\hat{\boldsymbol{\phi}}_M$.

Theorem 1.4.3 Let $\{X_t\}$ be the stationary process defined by (1.16), and suppose the true value $\boldsymbol{\phi}_0 = (\boldsymbol{b}_0, \boldsymbol{a}_0)$ and its parameter space Φ are under the condition **B1**, we have

$$X_m^2(\hat{\boldsymbol{\phi}}_M) = n(\hat{\beta}_0 - \sqrt{\pi})^2 + n\hat{\boldsymbol{\beta}}_m^T \Sigma(\hat{\boldsymbol{\phi}}_M)^{-1} \hat{\boldsymbol{\beta}}_m \longrightarrow_d \chi^2(m), \text{ as } n \to \infty.$$

1.5 Long Memory Processes

We extend all theorems we discussed so far to long memory processes in this section. To establish the asymptotic theory of our test statistic when the process is a long memory process, we restrict the process $\{X_t\}$ to be Gaussian.

Under a simple hypothesis, Theorem 1.2.1 remains consistent.

Theorem 1.5.1 Under the null hypothesis H_0 in (1.2), consider the Gaussian linear process $\{X_t\}$ in (1.1) under condition **A2**, which has a spectral density $f(u) = u^{-2d} g^*(u)$, $d \in (0, 0.5)$ and $g^*(\cdot)$ is an even differentiable function on $[-\pi, \pi]$. Also let the spectral density satisfy $\inf_u f(u) > 0$. We have, as $n \to \infty$, for a fixed m such that $1 \leqslant m < \nu/2$,

$$
\begin{aligned}
&\sqrt{n}(\tilde{\beta}_0 - \sqrt{\pi}) \longrightarrow_d N(0, 2\pi), \\
&\sqrt{n}\tilde{\beta}_k \longrightarrow_d N(0, 2\pi), \text{ for all } k = 1, 2, \cdots, m, \\
&\mathrm{Cov}(\sqrt{n}\tilde{\beta}_{k_1}, \ \sqrt{n}\tilde{\beta}_{k_2}) \longrightarrow 0, \text{ for all } k_1 \neq k_2, k_1, k_2 = 1, 2, \cdots, m, \\
&\mathrm{Cov}(\sqrt{n}(\tilde{\beta}_0 - \sqrt{\pi}), \sqrt{n}\tilde{\beta}_k) \longrightarrow 0, \text{ for all } k = 1, 2, \cdots, m.
\end{aligned}
$$

Under a composite hypothesis, we mainly consider the Whittle likelihood estimator (1.7). To establish the corresponding asymptotic theories, we need the following assumptions on $\boldsymbol{\theta}$ and the spectral density $f(\cdot, \boldsymbol{\theta})$.

D0: Let $\boldsymbol{\Theta}_0$ be a compact subset of $\boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is a finite dimensional parameter space in $\mathbb{R}^s$ for some positive integer s. Let the spectral density for the process $\{X_t\}$ be $f(u, \boldsymbol{\theta}_0) = f^*(u, d_0) g^*(u, \boldsymbol{\varphi}_0)$, where f^* and g^* are even functions on $[-\pi, \pi]$, $f^*(u, d) \sim a_d u^{-2d}$ as $u \to 0$ for some $a_d > 0$, $g^*(u, \boldsymbol{\varphi})$ is differentiable on $[-\pi, \pi]$, and $\boldsymbol{\theta}_0 = (\boldsymbol{\varphi}_0, d_0)$ is the true parameter vector that lies in the interior of $\boldsymbol{\Theta}_0$. Furthermore, assume that the sth component of $\boldsymbol{\Theta}_0$ is contained in the segment $[\delta_1, 0.5 - \delta_1]$ for some $0 < \delta_1 < 0.25$.

D1: Assumptions **A1-A6** in Dahlhaus (1989). Note the original parameter space $\boldsymbol{\Theta}$ in Dahlhaus (1989) has been updated into $\boldsymbol{\Theta}_0$ following discussions in **D0**.

Assumption **D0** is from Chen & Deo (2004). And assumption **D1** is from Dahlhaus (1989), they are necessary to represent the consistency and efficiency of $\hat{\boldsymbol{\theta}}_W$ for long memory processes. Unlike Chen & Deo (2004), the asymptotic distribution of our test statistic does not rely on the following assumption for the spectral density $f(\cdot, \boldsymbol{\theta})$.

There exists a constant C with

$$|f(\boldsymbol{\theta_1}, \lambda) - f(\boldsymbol{\theta_2}, \lambda)| \leq C \|\boldsymbol{\theta_1} - \boldsymbol{\theta_2}\| f(\boldsymbol{\theta_2}, \lambda)$$

uniformly for all λ and all $\boldsymbol{\theta_1} = (\boldsymbol{\beta_1}, d_1)^\top$ and $\boldsymbol{\theta_2} = (\boldsymbol{\beta_2}, d_2)^\top$ such that $d_1 < d_2$.

This is from assumption **A9** in Dahlhaus (1989), Chen & Deo (2004) use it multiple times in some critical steps of their proof to establish the asymptotic distribution of their test statistic $T_n(\hat{\boldsymbol{\theta}})$ in a long memory process. However, Dahlhaus (2006) later removes this assumption and correct his original proof, since this condition rules out long range dependent processes.

Now we are ready to state the asymptotic theory of $\hat{\beta}_k$, $k = 1, 2, \cdots, m$ with $\hat{\boldsymbol{\theta}}_W$.

Theorem 1.5.2 Under the null hypothesis H_0 in (1.2), consider the Gaussian linear process $\{X_t\}$ in (1.1) satisfying the same assumptions as those of Theorem 1.5.1, let the parameter vector $\boldsymbol{\theta}$ satisfy assumption **D0** and the spectral density of $\{X_t\}$ satisfy assumption **D1**, consider the Whittle likelihood estimator $\hat{\boldsymbol{\theta}}_W$, we have

$$\sqrt{n}\hat{\boldsymbol{\beta}} \longrightarrow_d N(0, \Sigma(\boldsymbol{\theta}_0)),$$

as $n \to \infty$, where $\Sigma(\boldsymbol{\theta}_0)$ is the same as we described in Theorem 1.3.1.

Based on the smoothness assumptions of $f(\cdot, \boldsymbol{\theta})$, it is straightforward to show that $\Sigma(\hat{\boldsymbol{\theta}}_W)$ is a consistent estimator of $\Sigma(\boldsymbol{\theta}_0)$ as in Proposition 1.3.1.

To explain the asymptotic result of $\hat{\beta}_0$, we can again consider the discrete analog of the likelihood function (1.7),

$$\mathcal{L}_n^{\dagger}(\boldsymbol{\theta}) = \sum_{j=1}^{\nu} \left\{ \log f(u_j, \theta) + \frac{I_X(u_j)}{f(u_j, \theta)} \right\}, \tag{1.17}$$

see also Dahlhaus (1989) and Shumway & Stoffer (2017, p.247). Then following the same discussions as we did in the Section 1.3, we have $\hat{\beta}_0 - \sqrt{\pi} = o_p(n^{-1/2})$. In the end, our test statistic for long memory processes remains the same as that for short memory processes.

Theorem 1.5.3 Under the null hypothesis H_0 in (1.2), consider the spectral density of the Gaussian linear process $\{X_t\}$ in (1.1) in the form $\sigma^2 g(u, \boldsymbol{\phi})$ satisfying the same assumptions as those of Theorem 1.5.2 (parameter space $\boldsymbol{\Theta}$ can be updated to $\boldsymbol{\Phi}$). Note $\boldsymbol{\beta}_m = (\beta_1, \cdots, \beta_m)$ is a m-vector, for any fixed m such that $1 \leqslant m < \nu/2$. We have

$$X_m^2(\hat{\boldsymbol{\phi}}_W) = n(\hat{\beta}_0 - \sqrt{\pi})^2 + n\hat{\boldsymbol{\beta}}_m^T \Sigma(\hat{\boldsymbol{\phi}}_W)^{-1} \hat{\boldsymbol{\beta}}_m \longrightarrow_d \chi^2(m), \text{ as } n \to \infty.$$

CHAPTER II

SIMULATION RESULTS

In the previous chapter we introduced the new test statistic X_m^2 for both short memory and long memory processes. Since Chen and Deo showed their test T_n (2004, *Econometric Theory* 20, 382-416) has power comparable to that Hong's test (1996, *Econometrica* 64, 837-864), which is a generalization of the Box and Pierce (1970, *Journal of the American Statistical Association* 65, 1509-1526) test statistic, and superior to that of another frequency domain test by Milhoj (1981, *Biometrika* 68, 177-187). We mainly compare out test X_m^2 with Chen and Deo's test T_n, which is introduced in section 2.1. More precisely, the finite sample performance of those two test statistics in terms of their empirical levels and powers are investigated. In section 2.2, it is about the level study, we examined the empirical frequencies of rejection of the null hypothesis when it is in fact true. In section 2.3, it is about the power study, we compute the empirical frequencies of rejection of the null hypothesis under several alternatives. The common $\alpha = 5\%$ and $\alpha = 10\%$ significance level have been adopted and two sample size $n = 128$ and 512 are considered. All computations were done using scripts written in R 4.2.2.

2.1 Current Test Statistic T_n

We firstly introduce the test statistic of Hong (1996). Consider AR(p) process, $X_t = b_0 + b_1 X_{t-1} + \cdots + b_p X_{t-p} + \zeta_t$, where ζ_t are zero mean white noises. Let ξ_t be the residuals from the fitted model, then the test statistic of Hong (1996) is

$$H_n = \sum_{j=1}^{n-1} k^2(j/p_n)\hat{\rho}^2_{\xi,j},$$

where $k(\cdot)$ is a suitable kernel function such that $k(0) = 1$, $\hat{\rho}^2_{\xi,j}$ are the sample autocorrelations of the residuals. By Parseval's identity, H_n can be written as

$$\begin{aligned} H_n &= \frac{1}{2}\left(\sum_{j=-(n-1)}^{n-1} k^2(j/p_n)\hat{\rho}^2_{\xi,j} - 1\right) \\ &= \frac{1}{2}\left\{\left(\int_0^{2\pi} \hat{f}_\xi(u)du\right)^{-2}\left(2\pi\int_0^{2\pi} \hat{f}^2_\xi(u)du\right) - 1\right\}, \end{aligned} \tag{2.1}$$

where

$$\hat{f}_\xi(u) = \int_0^{2\pi} W(u-\omega)I_{n,\xi}(\omega)d\omega, \qquad W(u) = \frac{1}{2\pi}\sum_{|h|<n} k(h/p_n)e^{-\mathrm{i}hu},$$

$I_{n,\xi}$ is the mean corrected periodogram of the residuals.

Chen & Deo (2004) adapted (2.1) in the following way, they use the discrete version of $\hat{f}_\xi$ with $I_{n,\xi}$ is replaced by I_X/f_X, then update the integral of $\hat{f}_\xi$ into its discrete version. That is,

$$T_n = \left\{\frac{2\pi}{n}\sum_{l=0}^{n-1} \hat{f}_\xi(u_l)\right\}^{-2}\left\{\frac{2\pi}{n}\sum_{l=0}^{n-1} \hat{f}^2_\xi(u_l)\right\},$$

where

$$\hat{f}_\xi(u) = \frac{2\pi}{n}\sum_{j=1}^{n-1} \frac{W(u-u_j)I_X(u_j)}{f_X(u_j)}.$$

Thus they whiten the process in the frequency domain instead of in the time domain. Under the null hypothesis, T_n converges to normal distribution for suitable kernel functions $k(\cdot)$ and bandwidths p_n.

In their simulation study, they used the following three kernels.

(i) Bartlett $k(z) = 1 - |z|, |z| \leqslant 1,$

$= 0$, otherwise.

(ii) Tukey $k(z) = \frac{1}{2}(\cos(z\pi) + 1), |z| \leqslant 1,$

$= 0$, otherwise.

(iii) Quadratic spectral(QS) $k(z) = \frac{25}{12\pi^2 z^2}\left(\frac{\sin(6\pi z/5)}{6\pi z/5} - \cos(6\pi z/5)\right), z \in [-1, 1] \setminus \{0\},$

$= 1, z = 0.$

And they used three bandwidths, $p_n = [3n^{0.2}], [3n^{0.3}]$ and $[3n^{0.4}]$, where $[x]$ denotes the integer closest to the real number x. The bandwidths deliver $p_n = 8$, 13 and 21 for $n = 128$; and 11, 20, 37 for $n = 512$.

2.2 Level Study

Given the modern computing resources, we propose to use Monte Carlo methods to find the critical values and the rejection regions for a given finite sample size n under H_0. Monte Carlo methods can also be used to calculate the empirical powers of the test statistics under a given alternative H_1.

We compute the empirical levels using the asymptotic critical values (noted as ACV). We illustrate the steps for Monte Carlo computation of the empirical levels using the asymptotic critical values as follows:

1. For a specific test statistic, find the ACV which is the 95-th (90-th) quantile of the limiting distribution of the test statistic under the null hypothesis.
2. Generate a random sample $\{X_t\}_{t=1}^n$ under the null hypothesis.
3. Compute the test statistic under the null hypothesis based on the random sample $\{X_t\}_{t=1}^n$ generated in step 2.
4. Repeat step 2 and 3 for $N = 5000$ times to derive 5000 test statistics under the null hypothesis.
5. Compute the level which is the percentage of the 5000 test statistics that are larger than the ACV.

For our test statistic X_m^2, through all the proofs we showed before, for the finite small sample size, especially $n = 128$, it is recommended to use discrete version instead of integral version to calculate the asymptotic variance. That is,

$$
\begin{aligned}
\Sigma_{k,l}^{\dagger}(\boldsymbol{\theta}_0) &= 2\pi \mathrm{I}(k = l) - [C_k^{\dagger}(\boldsymbol{\theta}_0)]^T \left\{4\pi\Gamma^{\dagger}(\boldsymbol{\theta}_0)^{-1}\right\} C_l^{\dagger}(\boldsymbol{\theta}_0), \\
C_k^{\dagger}(\boldsymbol{\theta}_0) &= \frac{1}{\sqrt{2\pi}} \frac{2\pi}{n} \sum_j \cos(k u_j) \boldsymbol{\nabla}_{\boldsymbol{\theta}} \log f(u_j, \boldsymbol{\theta}_0), \\
\Gamma^{\dagger}(\boldsymbol{\theta}_0) &= \frac{2\pi}{n} \sum_j \boldsymbol{\nabla}_{\boldsymbol{\theta}} \log f(u_j, \boldsymbol{\theta}_0) \boldsymbol{\nabla}_{\boldsymbol{\theta}}^T \log f(u_j, \boldsymbol{\theta}_0).
\end{aligned}
$$

Table 1: Rejection rates in percentage under an AR(1) model

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	3.32	5.64	3.78	6.14	4.70	7.50	3.34	5.48	3.74	6.38	4.72	7.94
	TUK	3.18	5.38	3.90	6.08	4.76	7.54	3.36	5.56	3.90	6.74	4.96	8.14
	QS	3.86	5.98	4.16	7.12	6.22	9.82	3.62	6.08	4.30	7.40	5.58	9.34
m		1		2		3		1		2		3	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
X_m^2		6.04	11.1	5.28	10.2	4.70	9.72	5.18	10.6	5.42	10.5	4.78	9.96

Note: Model $X_t - 0.8X_{t-1} = \zeta_t$, $\zeta_t \sim N(0,1)$.

Table 2: Rejection rates in percentage under an ARFIMA(d) model

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	2.22	4.10	3.22	5.44	4.82	7.56	2.76	4.64	4.00	6.42	4.76	8.30
	TUK	2.10	4.18	3.46	5.74	4.74	7.52	2.92	4.76	4.32	6.74	5.14	8.72
	QS	2.90	5.16	4.18	6.78	6.44	9.88	3.54	5.46	4.74	7.62	5.62	9.66
m		1		2		3		1		2		3	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
X_m^2		5.02	10.1	4.92	9.68	4.76	8.48	5.04	10.2	4.62	9.58	5.04	9.80

Note: Model $X_t = (1-B)^{-0.4}\zeta_t$, $\zeta_t \sim N(0,1)$.

Table 3: Rejection rates in percentage under an AR(1) model with innovations from t distribution

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	3.12	5.18	3.58	6.34	4.64	7.84	3.30	5.12	3.36	6.36	4.74	7.68
	TUK	3.22	5.06	3.72	6.40	4.62	7.76	3.34	5.22	3.36	6.62	4.78	7.90
	QS	3.52	5.80	4.32	7.20	6.30	9.62	3.30	5.48	4.04	6.94	5.40	8.58
m		1		2		3		1		2		3	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
X_m^2		5.38	11.3	4.90	10.2	4.34	9.12	5.32	10.6	4.78	9.56	5.16	9.50

Note: Model $X_t - 0.8X_{t-1} = \zeta_t$, $\zeta_t \sim t_9$.

Table 4: Rejection rates in percentage under an ARFIMA(d) model with innovations from t distribution

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	2.28	3.78	3.06	5.06	4.28	6.80	2.82	4.84	3.66	6.32	4.78	8.12
	TUK	2.30	3.76	3.36	5.36	4.52	6.76	3.04	5.14	3.94	6.42	5.08	8.50
	QS	3.00	4.68	3.98	6.34	5.80	8.80	3.44	5.86	4.36	7.14	5.60	9.14
m		1		2		3		1		2		3	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
X_m^2		5.08	10.0	5.38	9.86	5.14	9.36	4.56	9.28	4.48	9.30	4.64	9.20

Note: Model $X_t = (1 - B)^{-0.4}\zeta_t$, $\zeta_t \sim t_9$.

And we choose $m = 1$, 2 and 3 to compete with the test statistic of Chen & Deo (2004) in different bandwidths p_n. We use the Whittle likelihood for estimation.

In Table 1 and Table 2, we report the size of the two tests under the composite null hypothesis of an AR(1) and an ARFIMA(d). As Chen & Deo (2004) mentioned, their test statistic has undersized issue at both 5% and 10% levels. The amount by which they are undersized decreases as the bandwidth p_n increases. It is impressed to see our test statistic does not have such issues for both $n = 128$ and $n = 512$ cases. And Chen & Deo (2004) use Table 3 and 4 to argue that their test statistic can be used in non-Gaussian cases. It is clear to see in those two tables, our test statistic also outperforms their results.

The four time series models in those four tables are the same as the models Chen & Deo (2004) use in their level study, when they compare test statistic with Hong (1996)'s test and Milhoj (1981)'s test (see Appendix A). In other words, our test statistic outperforms all other three statistics when we consider the preciseness of rejection rates under the null hypothesis.

2.3 Power Study

About the power study, Chen & Deo (2004) only calculate asymptotic critical values (ACV). We add empirical critical values (ECV) in our study, since they allow us to be able to compare the powers of both test statistics on an equal basis.

To compute the empirical powers based on the asymptotic (empirical) critical values, the following steps have been implemented:

1. Generate a random sample $\{X_t\}_{t=1}^n$ under an alternative hypothesis.
2. Compute the test statistic based on the random sample $\{X_t\}_{t=1}^n$ generated in step 1. All parameters are estimated by the Whittle likelihood.
3. Repeat steps 1 and 2 for $N = 5000$ times to derive 5000 test statistics.
4. Compute the empirical power which is the percentage of the 5000 test statistics that are greater than the asymptotic (empirical) critical values.

Table 5 reports the powers against the AR(2) process, the powers of our test statistics are uniformly higher than the powers of test statistic T_n, no matter in a small sample size $n = 128$ or a large sample size $n = 512$, no matter when we consider ACV or ECV. Since T_n has undersized issue reported in Table 1, especially for small bandwidth p_n, their powers increases from ACV to ECV. Our powers remain consistent for both two types of critical values.

Table 6 is to detect short memory alternatives. The differences in powers between two test statistics increase dramatically when the sample size increases from $n = 128$ to $n = 512$. X_m^2's powers are highly significantly greater than powers of T_n when we consider a large sample size.

Table 7 is to detect long memory alternatives. Both test statistics show the same powers in almost all cases. Note the highest powers in ACV and ECV, are all reached by our test statistic X_1^2 by setting $m = 1$.

Table 5: Rejection rates in percentage under an AR(2) alternative fitting model AR(1)

Asymptotic Critical Values (ACV)													
n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	23.54	29.02	24.42	30.60	24.24	31.22	79.14	84.18	74.96	81.34	68.10	75.76
	TUK	23.12	29.00	24.22	30.62	23.34	30.70	79.08	84.04	74.08	80.36	65.36	73.34
	QS	24.04	30.26	24.06	30.90	24.62	31.52	77.46	83.08	70.74	77.24	61.22	69.84
m		1		2		3		1		2		3	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
X_m^2		47.10	59.98	35.66	48.50	27.56	40.84	94.10	96.68	89.20	93.86	85.02	91.14
Empirical Critical Values (ECV)													
n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	27.38	38.40	26.08	36.28	22.38	33.32	84.24	89.38	78.04	85.58	68.80	78.90
	TUK	27.58	38.80	26.46	36.22	21.80	32.58	84.08	89.36	76.52	85.02	65.48	75.96
	QS	27.16	37.78	23.16	34.36	19.38	30.80	81.22	87.56	72.26	81.34	59.62	70.90
m		1		2		3		1		2		3	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
X_m^2		43.70	56.58	35.76	47.34	30.40	42.60	94.14	96.72	89.46	93.72	85.48	91.10

Note: Model $X_t - 0.8X_{t-1} + 0.15X_{t-2} = \zeta_t$, $\zeta_t \sim N(0,1)$.
Empirical levels are obtained by $X_t - 0.8X_{t-1} = \zeta_t$, $\zeta_t \sim N(0,1)$.

Table 6: Rejection rates in percentage under an ARMA(1,1) alternative fitting model ARFIMA(1,d,0)

Asymptotic Critical Values (ACV)													
n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	3.92	6.08	4.84	7.64	6.16	9.74	22.16	29.84	21.18	28.62	19.84	27.72
	TUK	4.14	6.36	5.04	7.90	6.20	9.72	22.20	30.14	20.18	27.42	18.80	26.60
	QS	4.64	7.34	5.74	9.36	7.84	11.66	21.34	28.88	19.56	27.06	18.34	26.14
m		1		2		3		1		2		3	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
X_m^2		11.66	21.88	9.08	16.82	6.66	13.64	57.24	73.36	44.30	60.44	36.66	51.70

Empirical Critical Values (ECV)													
n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	17.58	26.70	14.60	23.46	12.50	21.18	47.72	60.66	36.52	49.60	28.22	41.10
	TUK	17.90	27.46	14.10	23.36	11.60	20.76	47.36	60.74	34.18	47.60	26.08	38.04
	QS	15.32	24.92	12.94	21.34	11.06	19.14	40.76	54.02	29.56	42.70	22.52	35.42
m		1		2		3		1		2		3	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
X_m^2		20.60	38.98	12.68	23.40	8.48	17.78	70.44	84.16	55.60	72.74	41.84	60.30

Note: Model $X_t = 0.8X_{t-1} + \zeta_t + 0.2\zeta_{t-1}$, $\zeta_t \sim N(0,1)$.
Empirical levels are obtained by $X_t - 0.1X_{t-1} = (1-B)^{-0.4}\zeta_t$, $\zeta_t \sim N(0,1)$.

Table 7: Rejection rates in percentage under an ARFIMA(1,d,0) alternative fitting model AR(1)

Asymptotic Critical Values (ACV)													
n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	9.36	13.20	10.56	14.88	11.50	16.54	28.58	35.80	28.26	35.54	26.60	34.22
	TUK	9.48	13.12	10.74	14.96	11.28	16.44	29.10	36.32	28.22	35.64	25.66	33.54
	QS	10.28	14.48	11.24	16.08	12.62	18.04	29.14	36.46	26.78	34.82	24.62	32.28
m		1		2		3		1		2		3	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
X_m^2		17.28	27.70	12.86	21.46	10.40	18.02	33.96	45.44	27.12	38.58	25.04	35.98
Empirical Critical Values (ECV)													
n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	12.16	20.12	11.50	19.02	10.16	18.12	33.06	45.10	29.70	41.98	25.70	37.80
	TUK	12.10	20.34	12.02	19.62	10.02	18.06	33.46	45.56	29.20	41.30	24.52	36.56
	QS	12.14	20.38	10.98	18.78	8.84	17.00	30.98	44.66	26.70	39.26	22.78	33.86
m		1		2		3		1		2		3	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
X_m^2		16.28	25.98	12.48	21.44	10.90	18.62	34.54	46.12	27.86	38.98	25.14	36.34

Note: Model $X_t - 0.6X_{t-1} = (1-B)^{-0.3}\zeta_t$, $\zeta_t \sim N(0,1)$.
Empirical levels are obtained by $X_t - 0.8X_{t-1} = \zeta_t$, $\zeta_t \sim N(0,1)$.

Table 8: Rejection rates in percentage under an ARFIMA(1,d,0) alternative fitting model ARFIMA(d)

Asymptotic Critical Values (ACV)													
n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	5.32	7.88	6.68	9.38	7.34	10.98	11.42	16.22	11.58	16.90	11.56	17.48
	TUK	5.20	7.66	6.44	9.48	7.60	11.18	11.40	16.34	11.72	17.12	11.58	17.22
	QS	6.14	8.94	7.08	10.70	8.54	12.60	11.76	16.96	11.80	16.48	11.58	17.26
m		1		2		3		1		2		3	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
X_m^2		17.90	26.30	16.88	23.78	16.78	22.74	28.14	38.96	21.78	31.70	19.58	28.84
Empirical Critical Values (ECV)													
n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	10.10	17.92	9.24	16.26	8.62	14.68	17.08	27.24	15.02	23.68	12.16	20.66
	TUK	10.38	17.66	9.26	16.32	8.40	14.28	16.78	26.92	14.42	23.30	11.82	20.60
	QS	9.64	16.64	8.60	15.30	7.58	13.68	16.06	24.88	13.88	22.52	10.60	18.54
m		1		2		3		1		2		3	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
X_m^2		18.58	27.20	18.00	25.44	18.08	24.50	30.06	40.24	23.08	33.24	21.32	29.56

Note: Model $X_t - 0.1X_{t-1} = (1-B)^{-0.4}\zeta_t$, $\zeta_t \sim N(0,1)$.
Empirical levels are obtained by $X_t = (1-B)^{-0.4}\zeta_t$, $\zeta_t \sim N(0,1)$.

In Table 8, the stochastic process ARFIMA(1,d,0) is simulated. X_m^2 dominates the power performance again as in preceding short memory processes.

As Chen & Deo (2004) use the same time series models as those in Table 5, 6 and 8 in their power study to compare with test statistics from Hong (1996) and Milhoj (1981) (see details in Appendix A), it is reasonable to claim that our test has powers comparable to current ones, in some cases, it even outperforms them.

CHAPTER III

CONCLUSION

This paper builds a new goodness-of-fit test for stochastic processes. It has intriguing results in both theories and simulations. In theories, the test statistic is asymptotically normally distributed. And its assumptions are general, they cover general short memory process and Gaussian long memory process. In simulations, through our empirical level and power studies, it outperforms the most current test by utilizing Whittle estimation. It is tempting to extend those results to maximum likelihood estimation. Through our proofs, they can be achieved if we only consider the Gaussian short memory process. About the long memory process, more works are necessary to be done in this field. This gap is mainly caused by an dichotomy between time and frequency domain estimation. Some paper starts to build their connections, see also Rao & Yang (2021), but it is only restricted to a short memory process.

Moreover, all asymptotic theories of our test statistic we build so far are under the null hypothesis when the model is correctly specified. Formulating power properties of the test statistic under alternative hypothesis when the model is misspecified is intriguing. We believe it can be done in a short memory process, but it would still be much more complex in a long memory process.

An additional question of interest is the choice of m in our test statistic X_m^2. In our simulation study, since the empirical levels are very close to true levels, we may choose the proper m by finding the one which is able to provide the highest power. As $X_m^2(\hat{\boldsymbol{\phi}}_W)$ is asymptotically Chi-squared distributed with the degree of freedom m. The most intuitive way might be

$$\hat{m} = \underset{m}{\operatorname{argmax}}\{X_m^2(\hat{\boldsymbol{\phi}}_W) - 2m\},$$

where m is determined by AIC or BIC. But we believe further and more comprehensive study is required for this part.

CHAPTER IV

PROOF OF THEOREMS

4.1 Proof of Theorem 1.2.1

Lemma I Let $u_j = \frac{2\pi j}{n}, j = 1, 2, \cdots, \nu$ and $\nu = \frac{n}{2} - 1$. Then we have, as $n \to \infty$, for a fixed m,

$$\frac{1}{\nu}\sum_{j=1}^{\nu}\cos\big(2ku_j\big) \to 0,\ \text{uniformly for all integers } 1 \le k \le m < \nu/2,$$

$$\frac{1}{\nu}\sum_{j=1}^{\nu}\cos\big(ku_j\big) \to 0,\ \text{uniformly for all integers } 1 \le k \le m < \nu.$$

Proof:

Since the above proofs are alike, we only consider the proof for the second one.

We know the following result,

$$\sum_{j=1}^{\nu}\cos(\lambda j) = \frac{\sin[(\nu + \frac{1}{2})\lambda]}{2\sin(\lambda/2)} - \frac{1}{2}, \qquad \text{where } \lambda = 2k\pi/n.$$

Then

$$\begin{aligned}
\frac{1}{\sqrt{\nu}}\sum_{j=1}^{\nu}\cos\,(ku_j) &= \frac{1}{\sqrt{\nu}}\sum_{j=1}^{\nu}\cos\,(\frac{2\pi k}{n}j) \\
&= \frac{1}{\sqrt{\nu}}\left[\frac{\sin[(\nu+\frac{1}{2})\frac{2\pi k}{n}]}{2\sin\frac{\pi k}{n}} - \frac{1}{2}\right] \\
&= \frac{1}{\sqrt{\nu}}\left[\frac{\sin\frac{2\pi k\nu}{n}\cos\frac{\pi k}{n}}{2\sin\frac{\pi k}{n}} + \frac{\cos\frac{2\pi k\nu}{n}\sin\frac{\pi k}{n}}{2\sin\frac{\pi k}{n}} - \frac{1}{2}\right] \\
&= \frac{1}{2\sqrt{\nu}}\frac{\sin\frac{2\pi k\nu}{n}\cos\frac{\pi k}{n}}{\sin\frac{\pi k}{n}} + \frac{1}{2\sqrt{\nu}}\cos\frac{2\pi k\nu}{n} - \frac{1}{2\sqrt{\nu}}.
\end{aligned}$$

Since $\nu = \frac{n}{2} - 1 \to \infty$, as $n \to \infty$. We directly have

$$\frac{1}{2\sqrt{\nu}} \cos\frac{2\pi k\nu}{n} \xrightarrow{n\to\infty} 0, \qquad \frac{1}{2\sqrt{\nu}} \xrightarrow{n\to\infty} 0.$$

Now we only need to show

$$\frac{1}{\sqrt{\nu}} \frac{\sin\left(2\pi k\nu/n\right)}{\sin(\pi k/n)} \xrightarrow{n\to\infty} 0.$$

Since $1 \le k \le \nu = \frac{n}{2} - 1$, it indicates $0 \le \frac{k\pi}{n} \le \frac{n\pi/2}{n} = \frac{\pi}{2}$. As we know, for $x \in (0, \frac{\pi}{2})$, $\frac{2}{\pi} \le \frac{\sin x}{x} \le 1$. Thus this term can be bounded above in the following way,

$$\begin{aligned}
\frac{1}{\sqrt{\nu}} \frac{\sin\left(2\pi k\nu/n\right)}{\sin(\pi k/n)} &= \frac{1}{\sqrt{\nu}} \frac{\sin\left(2\pi k\nu/n\right)}{\sin\left(\frac{\pi k}{n}\right)/\frac{\pi k}{n}} \frac{1}{\frac{\pi k}{n}} \\
&\le \frac{\pi}{2\sqrt{\nu}} \frac{\sin\left(2\pi k\nu/n\right)}{\pi k/n} \\
&= \frac{n}{2\sqrt{\nu}} \frac{\sin\left(2\pi k\nu/n\right)}{k} \\
&= \frac{\sqrt{n}}{2\sqrt{\frac{n}{2} - 1}} \frac{\sqrt{n}\sin\left(2\pi k\nu/n\right)}{k}.
\end{aligned}$$

Meanwhile it can also be bounded below as the following,

$$\begin{aligned}
\frac{1}{\sqrt{\nu}} \frac{\sin\left(2\pi k\nu/n\right)}{\sin\left(\frac{\pi k}{n}\right)/\frac{\pi k}{n}} \frac{1}{\frac{\pi k}{n}} &\ge \frac{1}{\sqrt{\nu}} \frac{\sin\left(2\pi k\nu/n\right)}{\pi k/n} \\
&= \frac{n}{\pi\sqrt{\frac{n}{2} - 1}} \frac{\sin 2\pi k\nu/n}{k} \\
&= \frac{\sqrt{n}}{\pi\sqrt{\frac{n}{2} - 1}} \frac{\sqrt{n}\sin\left(2\pi k\nu/n\right)}{k}.
\end{aligned}$$

Therefore, to show the final result, it is equivalent to show

$$\frac{\sqrt{n}\sin\left(2\pi k\nu/n\right)}{k} \xrightarrow{n\to\infty} 0.$$

Then we apply the mean value theorem,

$$\begin{aligned}\frac{\sqrt{n}}{k}\left|\sin\frac{2\pi k\nu}{n}\right| &= \frac{\sqrt{n}}{k}\left|\sin\frac{2\pi k\nu}{n} - \sin k\pi\right| \\ &= \frac{\sqrt{n}}{k}\left|\cos\{k\pi[1+\eta(\frac{2\nu}{n}-1)]\}\right|\left|k\pi(\frac{2\nu}{n}-1)\right| \\ &= \pi\sqrt{n}\left|\frac{2\nu}{n}-1\right|\left|\cos\{k\pi[1+\eta(\frac{2\nu}{n}-1)]\}\right|, \qquad \text{where } \eta\in(0,1).\end{aligned}$$

Note

$$\sqrt{n}\left|\frac{2\nu}{n}-1\right| = \sqrt{n}\left|2\frac{\frac{n}{2}-1}{n}-1\right| = \sqrt{n}\left|-\frac{2}{n}\right| = \frac{2}{\sqrt{n}} \xrightarrow{n\to\infty} 0.$$

It concludes the proof. □

Lemma II Let $u_j = \frac{2\pi j}{n}, u_k = \frac{2\pi k}{n}, j,k = 1,2,\cdots,\nu$, and $\nu = \frac{n}{2}-1$. For the linear process $\{X_t\}$ in (1.1) under conditions **A1**, **A2**, we have,

$$\mathrm{Cov}(I_X(u_j), I_X(u_k)) = n^{-1}\eta_4 f(u_j,\boldsymbol{\theta}_0)f(u_k,\boldsymbol{\theta}_0) + o(n^{-1}),$$

for all u_j and u_k such that $u_j \neq u_k$. Above $o(\cdot)$ term is uniform in u_j and u_k.

Proof:

From (4.2) in Krogstad (1982, p.198), we have

$$\begin{aligned}\mathrm{Cov}(I_X(u_j), I_X(u_k)) &= \frac{1}{n^2}\sum\sum\sum\sum\nolimits_{p,q,l,m=0}^{n-1} K(q-p,l-p,m-p)e^{-\mathtt{i}(p-q)u_j-\mathtt{i}(l-m)u_k} \\ &+ \frac{1}{n^2}\sum\sum\sum\sum\nolimits_{p,q,l,m=0}^{n-1}[r(p-l)r(q-m)+r(p-m)r(q-l)] \\ &\times e^{-\mathtt{i}(p-q)u_j-\mathtt{i}(l-m)u_k} \\ &= Q_n(u_j,u_k) + C_n(u_j,u_k),\end{aligned}$$

where $K(\cdot,\cdot,\cdot)$ is the fourth order cumulant function for the process. From (4.7) in Krogstad (1982, p.199), $Q_n(u_j,u_k) = n^{-1}\kappa(u_j,u_k,-u_k)(1+o(1))$, where $\kappa(\cdot,\cdot,\cdot)$ is the fourth order cumulant spectral density defined as before, and $o(\cdot)$ term is uniform in both u_j and u_k.

From Rosenblatt (2012, p.130), $Q_n(u_j, u_k)$ can be also written as

$$n^{-1}\eta_4 f(u_j, \boldsymbol{\theta}_0) f(u_k, \boldsymbol{\theta}_0)(1 + o(1)),$$

where $\eta_4 = (\mathrm{E}[\zeta_0^4] - 3\sigma_0^4)/\sigma_0^4$ is the rescaled fourth-order cumulant of ζ_0. The same claim can be found in Fragkeskou & Paparoditis (2016, p.241).

About the second term, we have

$$\begin{aligned} C_n(u_j, u_k) = &\left(\frac{1}{n}\sum_{p=0}^{n-1}\sum_{l=0}^{n-1} r(p-l)e^{-\mathrm{i}pu_j}e^{-\mathrm{i}lu_k}\right)\left(\frac{1}{n}\sum_{q=0}^{n-1}\sum_{m=0}^{n-1} r(q-m)e^{\mathrm{i}qu_j}e^{\mathrm{i}mu_k}\right) \\ &+ \left(\frac{1}{n}\sum_{p=0}^{n-1}\sum_{m=0}^{n-1} r(p-m)e^{-\mathrm{i}pu_j}e^{\mathrm{i}mu_k}\right)\left(\frac{1}{n}\sum_{q=0}^{n-1}\sum_{l=0}^{n-1} r(q-l)e^{\mathrm{i}qu_j}e^{-\mathrm{i}lu_k}\right). \end{aligned}$$

we can show the result

$$\frac{1}{n}\sum_{p=0}^{n-1}\sum_{l=0}^{n-1} r(p-l)e^{-\mathrm{i}pu_j}e^{-\mathrm{i}lu_k} = O(n^{-\frac{1}{2}-\tau}), \tag{4.1}$$

where $\tau \in (0, 1/2)$ from **A1**. Then the rest of three factors can be proved in the same manner. The development of the rest proof closely matched Theorem 10.3.2 in Brockwell & Davis (1991, p.349). Note

$$\begin{aligned} \frac{1}{n}\sum_{p=0}^{n-1}\sum_{l=0}^{n-1} r(p-l)e^{-\mathrm{i}pu_j}e^{-\mathrm{i}lu_k} &= \frac{1}{n}\sum_{p=0}^{n-1}\sum_{s=p-n+1}^{p} r(s)e^{-\mathrm{i}pu_j}e^{-\mathrm{i}(p-s)u_k} \\ &= \frac{1}{n}\sum_{s=-(n-1)}^{n-1}\sum_{0\vee s}^{(n-1)\wedge(n-1+s)} r(s)e^{-\mathrm{i}p(u_j+u_k)}e^{\mathrm{i}su_k} \\ &= \frac{1}{n}\left(\sum_{s=0}^{n-1} r(s)e^{\mathrm{i}su_k}\sum_{p=s}^{n-1} e^{-\mathrm{i}p(u_j+u_k)} + \sum_{s=-(n-1)}^{-1} r(s)e^{\mathrm{i}su_k}\right. \\ &\quad \left.\times \sum_{p=0}^{n-1+s} e^{-\mathrm{i}p(u_j+u_k)}\right), \end{aligned}$$

since $u_j + u_k \neq 0$ or 2π, we have for $0 \leqslant s \leqslant n-1$,

$$\begin{aligned}\left|\sum_{p=s}^{n-1} e^{-\mathtt{i}p(u_j+u_k)}\right| &= \left|\sum_{p=1}^{n} e^{-\mathtt{i}p(u_j+u_k)} - \sum_{p=1}^{s-1} e^{-\mathtt{i}p(u_j+u_k)} - e^{-\mathtt{i}n(u_j+u_k)}\right| \\ &= \left|0 - \sum_{p=1}^{s-1} e^{-\mathtt{i}p(u_j+u_k)} - 1\right| \\ &\leqslant s.\end{aligned}$$

Similarly,

$$\left|\sum_{p=0}^{n-1+s} e^{-\mathtt{i}p(u_j+u_k)}\right| \leqslant |s|, \quad -n+1 \leqslant s \leqslant -1.$$

Then the left side of (4.1) is bounded above by

$$\begin{aligned}\frac{1}{n}\sum_{|s|<n} |s|\,|r(s)| &\leqslant \frac{\sigma_0^2}{n}\sum_{|s|<n}\sum_{j=-\infty}^{\infty} |s|\,|\psi_j\psi_{j+s}| \\ &\leqslant \frac{\sigma_0^2}{n^{\frac{1}{2}+\tau}}\sum_{|s|<n}\sum_{j=-\infty}^{\infty} |s|^{\frac{1}{2}+\tau}\,|\psi_j\psi_{j+s}| \\ &\leqslant \frac{\sigma_0^2}{n^{\frac{1}{2}+\tau}}\left(\sum_{s=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} |s+j|^{\frac{1}{2}+\tau}\,|\psi_j\psi_{j+s}| + \sum_{s=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} |j|^{\frac{1}{2}+\tau}\,|\psi_j\psi_{j+s}|\right) \\ &= \frac{2\sigma_0^2}{n^{\frac{1}{2}+\tau}}\left(\sum_{s=-\infty}^{\infty} |s|^{\frac{1}{2}+\tau}|\psi_s|\right)\left(\sum_{j=-\infty}^{\infty} |\psi_j|\right) \\ &= O(n^{-\frac{1}{2}-\tau}).\end{aligned}$$

It showed the result (4.1). In other words, $C_n(u_j, u_k)$ is bounded by $O(n^{-1-2\tau}) = o(n^{-1})$, which concludes the proof. □

The following powerful result can be applied to both short- and long-memory models. To develop the lemma, we consider the following two general assumptions

E1: The spectral density f_x of the process X_t, $t \in \mathbb{Z}$, satisfies

$$f_x(u) = |u|^{-2d} g(u), \quad |u| \leqslant \pi,$$

for some $d \in [0, 1/2)$, where $g(u)$ is an even differentiable function on $[-\pi, \pi]$.

E2: Consider the transfer function $A_X(u) = \sum_{j=0}^{\infty} e^{-\mathrm{i}ju}\psi_j$, $u \in [-\pi, \pi]$ of the linear process X_t in (1.1) is differentiable in $(0, \pi)$ and its derivative $\dot{A}_X$ has the property

$$|\dot{A}_X(u)| \leqslant C|u|^{-1-d}, \quad |u| \leqslant \pi.$$

Lemma III Suppose the linear process $\{X_t, t \in \mathbb{Z}\}$ of (1.1) satisfies **A2**, **E1** and **E2**. Consider $S_{n,X} = \sum_{j=1}^{\nu} b_{n,j} \frac{I_{X,j}}{f_{X,j}}$, where $\{b_{n,j}, j = 1, \cdots, \nu\}$ is an array of real numbers. Let

$$S_{n,\zeta} = \sum_{j=1}^{\nu} b_{n,j} \frac{I_{\zeta,j}}{f_{\zeta,j}} \equiv \sum_{j=1}^{\nu} b_{n,j} \frac{2\pi}{\sigma_0^2} I_{\zeta,j},$$

$$R_n = S_{n,X} - S_{n,\zeta}, \quad b_n = \max_{j=1,\cdots,\nu} |b_{n,j}|, \quad B_n^2 = \sum_{j=1}^{\nu} b_{n,j}^2,$$

$$q_n^2 = B_n^2 + \mathrm{Cum}_4(e_0) n^{-1} \left(\sum_{j=1}^{\nu} b_{n,j} \right)^2,$$

where $\{e_t, t = 1, \cdots, n\}$ are the standardized innovations, $e_t = \zeta_t / \sigma_0$, and assume

$$\frac{\max_{j=1,\cdots,\nu} |b_{n,j}|}{(\sum_{j=1}^{\nu} b_{n,j}^2)^{1/2}} = \frac{b_n}{B_n} \longrightarrow 0, \tag{4.2}$$

then we have,

$$q_n^{-1} \left(S_{n,X} - \sum_{j=1}^{\nu} b_{n,j} \right) \longrightarrow_d N(0, 1).$$

Proof:

The following proof is adapted from the proof of Theorem 6.2.1 in Giraitis et al. (2012, p.136). Note $I_{\zeta,j} = \frac{1}{2\pi}\left|\sum_{t=1}^{n}\zeta_t e^{-\mathrm{i}tu_j}\right|^2$, we have

$$\begin{aligned} S_{n,\zeta} &= \frac{1}{n}\sum_{t,s=1}^{n}\sum_{j=1}^{\nu} e^{\mathrm{i}(t-s)u_j} b_{n,j}\frac{\zeta_s}{\sigma_0}\frac{\zeta_t}{\sigma_0} \\ &= \sum_{t,s=1}^{n} c_n(t-s)e_s e_t \\ &= S_{n,e} \end{aligned}$$

where $c_n(t) = \frac{1}{n}\sum_{j=1}^{\nu} b_{n,j}\cos(tu_j)$, $t = 1, 2, \cdots$. Then it is obvious to see

$$\mathrm{E}[S_{n,\zeta}] = \mathrm{E}[S_{n,e}] = \sum_{t=1}^{n} c_n(0)E[e_0^2] = \sum_{j=1}^{\nu} b_{n,j}, \tag{4.3}$$

and

$$\begin{aligned} \mathrm{Var}[S_{n,\zeta}] &= \mathrm{Var}[S_{n,e}] \\ &= \mathrm{Cov}\left(\sum_{t,s=1}^{n} c_n(t-s)e_t e_s, \sum_{u,v=1}^{n} c_n(u-v)e_u e_v\right) \\ &= \sum_{t,s,u,v=1}^{n} c_n(t-s)c_n(u-v)\mathrm{Cov}(e_t e_s, e_u e_v) \\ &= 2\sum_{t,s=1;t\neq s}^{n} c_n^2(t-s) + \mathrm{Var}[e_0^2]\sum_{t=1}^{n} c_n^2(0) \\ &= 2\left\|C_n\right\|^2 + \mathrm{Cum}_4(e_0)n^{-1}(\sum_{j=1}^{\nu} b_{n,j})^2, \end{aligned}$$

The matrix $C_n = \{c_n(t-s)\}_{t,s=1,\cdots,n}$ is a symmetric matrix with real entries. Let $\|C_n\| = \left(\sum_{t,s=1}^{n} c_{n,ts}^2\right)^{1/2}$ and $\mathrm{Cum}_4(e_0) = \mathrm{E}[e_0^4] - 3 = \mathrm{Var}[e_0^2] - 2$. Thus the $\mathrm{Var}[S_{n,\zeta}]$ can be bounded below by

$$\mathrm{Var}[S_{n,\zeta}] \geqslant \min(2, \mathrm{Var}[e_0^2])\left\|C_n\right\|^2.$$

And from (6.2.29) in Giraitis et al. (2012, p.144), $\|Cn\|^2 = 2^{-1}B_n^2$, then the above inequality can be further written as

$$\min(1, \mathrm{Var}[e_0^2]/2)B_n^2 \leqslant \mathrm{Var}[S_{n,\zeta}] = \mathrm{Var}[S_{n,e}] = q_n^2. \tag{4.4}$$

Meanwhile following (6.2.21) in Giraitis et al. (2012, p.142), under conditions **E1** and **E2**, it can be shown

$$\mathrm{E}[R_n^2] = o(B_n^2), \quad \text{if } b_n = o(B_n). \tag{4.5}$$

Based on the result (4.3), we have

$$\begin{aligned} q_n^{-1}(S_{n,X} - \sum_{j=1}^{\nu} b_{n,j}) &= q_n^{-1}(S_{n,\zeta} - \mathrm{E}[S_{n,\zeta}] + R_n) \\ &= q_n^{-1}(S_{n,e} - \mathrm{E}[S_{n,e}]) + q_n^{-1} R_n. \end{aligned}$$

Then by (4.2), (4.4) and (4.5), we have

$$\mathrm{E}|R_n| \leqslant \left(\mathrm{E}[R_n^2]\right)^{1/2} = o(B_n) = o(q_n).$$

And

$$q_n^{-1}(S_{n,X} - \sum_{j=1}^{\nu} b_{n,j}) = \left(\mathrm{Var}[S_{n,e}]\right)^{-1/2} (S_{n,e} - \mathrm{E}[S_{n,e}]) + o_p(1).$$

Following (6.2.30) and Theorem 6.2.2 in Giraitis et al. (2012, p.143-144), the first term converges in the standard normal distribution,

$$\left(\mathrm{Var}[S_{n,e}]\right)^{-1/2} (S_{n,e} - \mathrm{E}[S_{n,e}]) \longrightarrow_d N(0,1).$$

It concludes the proof. □

Proof of Theorem 1.2.1: In the view of (1.5), we write

$$\sqrt{n}\tilde{\beta}_0 = \frac{\sqrt{2\nu}}{\sqrt{n}} \sum_{j=1}^{\nu} \frac{\sqrt{2\pi}}{\sqrt{\nu}} \frac{I_x(u_j)}{f(u_j, \boldsymbol{\theta}_0)}.$$

From $\nu = n/2 - 1$, we have $\sqrt{2\nu}/\sqrt{n} \to 1$. From Slutsky's Theorem, we only need to verify the limit distribution of the summation term. In order to do that, we apply the asymptotic normality result from **Lemma III** on the limit distribution of $S_{n,X} = \sum_{j=1}^{\nu} b_{n,j} \frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)}$. Thus we need to verify the corresponding assumptions. In our case, we have

$$b_{n,j} = \frac{\sqrt{2\pi}}{\sqrt{\nu}},\ b_n = \max_j |b_{n,j}| = \frac{\sqrt{2\pi}}{\sqrt{\nu}},\ B_n = \left(\sum_{j=1}^{\nu} b_{n,j}^2\right)^{1/2} = (2\pi)^{1/2},$$
$$q_n^2 = B_n^2 + \mathrm{Cum}_4(e_0)n^{-1}\left(\sum_{j=1}^{\nu} b_{n,j}\right)^2 = 2\pi\left(1 + \mathrm{Cum}_4(e_0)\frac{\nu}{n}\right).$$

From $b_n/B_n = \nu^{-1/2} \to 0$, the condition (4.2) is satisfied. As $\sum_{j=1}^{\nu} b_{n,j} = \sqrt{2\pi\nu}$, applying **Lemma III**, we have

$$q_n^{-1}\left(\sqrt{n}\tilde{\beta}_0 - \sqrt{2\pi\nu}\right) \longrightarrow_d N(0,1).$$

It is straightforward to see $q_n^2 \to 2\pi(1+\mathrm{Cum}_4(e_0)/2) = 2\pi q_0$, and $\sqrt{n\pi}-\sqrt{2\pi\nu} \to 0$. Following the Slutsky's theorem, we showed the first result in Theorem 1.2.1,

$$\sqrt{n}(\tilde{\beta}_0 - \sqrt{\pi}) \longrightarrow_d N(0, 2\pi q_0).$$

The second result can be showed in the same manner, note

$$\sqrt{n}\tilde{\beta}_k = \frac{\sqrt{2\nu}}{\sqrt{n}} \sum_{j=1}^{\nu} \frac{2\sqrt{\pi}}{\sqrt{\nu}} \cos(ku_j) \frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)}.$$

We only need to consider the limit distribution for the summation term as before. In this case, we have

$$b_{n,j} = \frac{2\sqrt{\pi}}{\sqrt{\nu}} \cos(ku_j), \quad b_n = \max_j |b_{n,j}| \leqslant \frac{2\sqrt{\pi}}{\sqrt{\nu}},$$

and $B_n^2 = \sum_{j=1}^{\nu} b_{n,j}^2 = 2\pi + 2\pi\nu^{-1}\sum_{j=1}^{\nu}\cos(2ku_j)$. From the preceding **Lemma I**, we have $B_n^2 \to 2\pi$. Thus we have $b_n/B_n \to 0$, the condition (4.2) in **Lemma III** is satisfied. We have

$$q_n^{-1}\left(\sqrt{n}\tilde{\beta}_k - \sum_{j=1}^{\nu} b_{n,j}\right) \longrightarrow_d N(0,1),$$

where $q_n^2 = B_n^2 + \mathrm{Cum}_4(e_0)n^{-1}\left(\sum_{j=1}^{\nu} b_{n,j}\right)^2$. From **Lemma I**, $\sum_{j=1}^{\nu} b_{n,j} \to 0$, we have $q_n^2 \to 2\pi$ and $\sqrt{n}\tilde{\beta}_k \longrightarrow_d N(0, 2\pi)$, which concludes the proof of the second result. We next show that those empirical coefficients are asymptotically uncorrelated.

$$
\begin{aligned}
\mathrm{Cov}(\sqrt{n}\tilde{\beta}_{k_1}, \sqrt{n}\tilde{\beta}_{k_2}) &= \frac{8\pi}{n}\sum_{j_1=1}^{\nu}\sum_{j_2=1}^{\nu}\cos(k_1 u_{j_1})\cos(k_2 u_{j_2})\mathrm{Cov}\left(\frac{I_X(u_{j_1})}{f(u_{j_1},\boldsymbol{\theta}_0)}, \frac{I_X(u_{j_2})}{f(u_{j_2},\boldsymbol{\theta}_0)}\right) \\
&= \frac{8\pi}{n}\sum_{j=1}^{\nu}\cos(k_1 u_j)\cos(k_2 u_j)\mathrm{Var}\left[\frac{I_X(u_j)}{f(u_j,\boldsymbol{\theta}_0)}\right] \\
&+ \sum\sum_{j_1 \neq j_2}\cos(k_1 u_{j_1})\cos(k_2 u_{j_2})\mathrm{Cov}\left(\frac{I_X(u_{j_1})}{f(u_{j_1},\boldsymbol{\theta}_0)}, \frac{I_X(u_{j_2})}{f(u_{j_2},\boldsymbol{\theta}_0)}\right) \\
&:= I_{1n} + I_{2n},
\end{aligned}
$$

Firstly consider I_{1n}. From Theorem 10.3.2 in Brockwell & Davis (1991, p.347), we know $\mathrm{Var}[I_X(u_j)] = f^2(u_j,\boldsymbol{\theta}_0) + O(n^{-1/2})$, then we have

$$
I_{1n} = \frac{8\pi}{n}\sum_{j=1}^{\nu}\cos(k_1 u_j)\cos(k_2 u_j) + R_{1n}, \tag{4.6}
$$

where $|R_{1n}| \leqslant Cn^{-1}\nu n^{-1/2} \leqslant Cn^{-1/2} \to 0$. Note $\cos(k_1 u_j)\cos(k_2 u_j) = \{\cos[(k_1+k_2)u_j] - \cos[(k_1-k_2)u_j]\}/2$, since $k_1 \neq k_2$ and $1 \leqslant k_1 + k_2 < \nu$, following the proof arguments of **Lemma I**, it is obvious to see the first term in I_{1n} also converges to 0 as $n \to \infty$, leading the result that $I_{1n} \to 0$.

Then consider I_{2n}, through **Lemma II**, we have

$$
I_{2n} = \frac{8\pi\eta_4}{n^2}\sum\sum_{j_1 \neq j_2}\cos(k_1 u_j)\cos(k_2 u_j) + R_{2n},
$$

where $|R_{2n}| = o(n^{-1}\nu^2 n^{-1})$, indicating $|R_{2n}| \to 0$. The first term in I_{1n} can be written as

$$
\frac{8\pi\eta_4}{n^2}\sum_{j_1=1}^{\nu}\sum_{j_2=1}^{\nu}\cos(k_1 u_{j_1})\cos(k_2 u_{j_2}) - \frac{8\pi\eta_4}{n^2}\sum_{j=1}^{\nu}\cos(k_1 u_j)\cos(k_2 u_j).
$$

It is straightforward to see the above first term goes to 0 following **Lemma I**. The second term is bounded above by $n^{-2}\nu$, which goes to 0 as well. It concludes the proof of the third result in the theorem. The last result can be done in the same fashion, since most part of its proof are repetitive, we skip them here. □

4.2 Proof of Theorem 1.3.1

Lemma IV Consider the linear process $\{X_t\}$ of (1.1) under conditions **A1** and **A2**, let the spectral density $f(\cdot)$ of the process satisfy the condition **B4**, and suppose $\boldsymbol{\theta}$ is a p-vector, then

$$\frac{1}{n}\sum_{j} \boldsymbol{\nabla}_{\boldsymbol{\theta}} \log f(u_j, \boldsymbol{\theta}_0) = O(n^{-1})\mathbf{1_p},$$

where the sum is over the Fourier frequencies $(-\pi, \pi) \setminus \{0\}$, $\mathbf{1_p}$ is a p-vector with all entries 1.

Proof:

It is equivalent to show

$$\frac{1}{n}\sum_{j} \frac{\partial}{\partial\theta_\omega} \log f(u_j, \boldsymbol{\theta}_0) = O(n^{-1}), \quad \omega = 1, 2, \cdots, p. \tag{4.7}$$

Note from Kolmogrov's Formula, Brockwell & Davis (1991, p.191), the one-step mean square prediction error of the stationary process $\{X_t\}$ is

$$\begin{aligned}
\sigma^2(f_{\boldsymbol{\theta}}) &= 2\pi \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(u, \boldsymbol{\theta}) du\right\} \\
\log\frac{\sigma^2(f_{\boldsymbol{\theta}})}{2\pi} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(u, \boldsymbol{\theta}) du \\
\frac{\partial}{\partial\theta_\omega}\log\frac{\sigma^2(f_{\boldsymbol{\theta}})}{2\pi} &= \frac{1}{2\pi}\frac{\partial}{\partial\theta_\omega}\int_{-\pi}^{\pi} \log f(u, \boldsymbol{\theta}) du \\
\frac{2\pi}{\sigma^2(f_{\boldsymbol{\theta}})}\frac{\partial}{\partial\theta_\omega}\sigma^2(f_{\boldsymbol{\theta}}) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_\omega}\log f(u, \boldsymbol{\theta}) du.
\end{aligned}$$

Since we have $\frac{\partial}{\partial\theta_\omega}\sigma^2(f_{\boldsymbol{\theta}})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0$, then

$$\int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_\omega}\log f(u, \boldsymbol{\theta}_0) du = 0. \tag{4.8}$$

This is also noted in Lemma 6 in Chen & Deo (2004). Note $f(u, \boldsymbol{\theta})$ is an even function for

$u \in [-\pi, \pi]$, the above result can also be written

$$\int_0^{\pi} \frac{\partial}{\partial \theta_\omega} \log f(u, \boldsymbol{\theta}_0) du = 0.$$

Now if we set $h(u) = \frac{\partial}{\partial \theta_\omega} \log f(u, \boldsymbol{\theta}_0)$, then we have

$$\begin{aligned}
\frac{2\pi}{n}\sum_{j=1}^{\nu} h(u_j) &= \frac{2\pi}{n}\sum_{j=1}^{\nu} h(u_j) - \int_0^{\pi} h(u)du \\
&= \frac{2\pi}{n}\sum_{j=1}^{\nu} h(u_j) - \sum_{j=1}^{\nu}\int_{u_{j-1}}^{u_j} h(u)du - \int_{u_\nu}^{u_{\frac{n}{2}}} h(u)du, \quad \text{where } u_j = \frac{2\pi j}{n}, \\
&= \sum_{j=1}^{\nu}\int_{u_{j-1}}^{u_j} h(u_j)du - \sum_{j=1}^{\nu}\int_{u_{j-1}}^{u_j} h(u)du - \int_{u_\nu}^{u_{\frac{n}{2}}} h(u)du \\
&= \sum_{j=1}^{\nu}\int_{u_{j-1}}^{u_j} [h(u_j) - h(u)]du - \int_{u_\nu}^{u_{\frac{n}{2}}} h(u)du \\
&= \sum_{j=1}^{\nu} h'(u_{\tilde{j}})\int_{u_{j-1}}^{u_j} [u_j - u]du - \int_{u_\nu}^{u_{\frac{n}{2}}} h(u)du, \quad \text{where } u_{j-1} \leqslant u_{\tilde{j}} \leqslant u_j, \\
&= \sum_{j=1}^{\nu} h'(u_{\tilde{j}})\int_{u_{j-1}}^{u_j} u_j du - \sum_{j=1}^{\nu} h'(u_{\tilde{j}})\int_{u_{j-1}}^{u_j} u du - \int_{u_\nu}^{u_{\frac{n}{2}}} h(u)du \\
&= \frac{2\pi}{n}\sum_{j=1}^{\nu} h'(u_{\tilde{j}})u_j - \sum_{j=1}^{\nu} h'(u_{\tilde{j}})\frac{1}{2}(u_j + u_{j-1})(u_j - u_{j-1}) - \int_{u_\nu}^{u_{\frac{n}{2}}} h(u)du \\
&= \frac{2\pi}{n}\sum_{j=1}^{\nu} h'(u_{\tilde{j}})u_j - \sum_{j=1}^{\nu} h'(u_{\tilde{j}})\frac{1}{2}(2u_j - \frac{2\pi}{n})\frac{2\pi}{n} - \int_{u_\nu}^{u_{\frac{n}{2}}} h(u)du \\
&= \frac{1}{2}\left(\frac{2\pi}{n}\right)\left\{\frac{2\pi}{n}\sum_{j=1}^{\nu} h'(u_{\tilde{j}})\right\} - \int_{u_\nu}^{u_{\frac{n}{2}}} h(u)du \\
&= O(n^{-1}).
\end{aligned}$$

□

Proof of Theorem 1.3.1: We use the Cramér-Wold device: take scalars $\alpha_1, \cdots, \alpha_m$ and consider $\sum_k \alpha_k \sqrt{n}\hat{\beta}_k$. We firstly study $\sqrt{n}\hat{\beta}_k$, that is

$$
\begin{aligned}
\sqrt{n}\hat{\beta}_k &= \sqrt{n}\frac{2\sqrt{2\pi}}{n}\sum_{j=1}^{\nu}\cos(ku_j)\frac{I_X(u_j)}{f(u_j,\hat{\boldsymbol{\theta}}_W)} \\
&= \sqrt{n}\frac{\sqrt{2\pi}}{n}\sum_{j}\cos(ku_j)\frac{I_X(u_j)}{f(u_j,\hat{\boldsymbol{\theta}}_W)} \\
&= \sqrt{n}\frac{\sqrt{2\pi}}{n}\sum_{j}\cos(ku_j)\frac{I_X(u_j)}{f(u_j,\boldsymbol{\theta}_0)} + \sqrt{n}\frac{\sqrt{2\pi}}{n}\sum_{j}\cos(ku_j)I_X(u_j)(f^{-1}(u_j,\hat{\boldsymbol{\theta}}_W) \\
&- f^{-1}(u_j,\boldsymbol{\theta}_0)) \\
&= \sqrt{n}\tilde{\beta}_k + \sqrt{n}\frac{\sqrt{2\pi}}{n}\sum_{j}\cos(ku_j)(I_X(u_j)-f(u_j,\boldsymbol{\theta}_0))(f^{-1}(u_j,\hat{\boldsymbol{\theta}}_W)-f^{-1}(u_j,\boldsymbol{\theta}_0)) \\
&+ \sqrt{n}\frac{\sqrt{2\pi}}{n}\sum_{j}\cos(ku_j)f(u_j,\boldsymbol{\theta}_0))(f^{-1}(u_j,\hat{\boldsymbol{\theta}}_W)-f^{-1}(u_j,\boldsymbol{\theta}_0)).
\end{aligned}
$$

According the smoothness conditions of $f_{\boldsymbol{\theta}}$ in **B4**, we have by a Taylor series expansion,

$$
f^{-1}(u,\hat{\boldsymbol{\theta}}_W) - f^{-1}(u,\boldsymbol{\theta}_0) = \boldsymbol{\nabla}_{\boldsymbol{\theta}}^T f^{-1}(u,\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_W-\boldsymbol{\theta}_0) + \frac{1}{2}(\hat{\boldsymbol{\theta}}_W-\boldsymbol{\theta}_0)^T \mathrm{H}_{\boldsymbol{\theta}} f^{-1}(u,\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_W-\boldsymbol{\theta}_0),
$$

where $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + \lambda(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0)$ for some $\lambda \in (0,1)$, and $\mathrm{H}_{\boldsymbol{\theta}} f^{-1}(u,\tilde{\boldsymbol{\theta}}) \longrightarrow_p \mathrm{H}_{\boldsymbol{\theta}} f^{-1}(u,\boldsymbol{\theta}_0)$. From Theorem 3.1.2 in Hosoya & Masanobu (1982), we also have $\left\|\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0\right\| = O_p(n^{-1/2})$, then the above Taylor expansion can also be written

$$
f^{-1}(u,\hat{\boldsymbol{\theta}}_W) - f^{-1}(u,\boldsymbol{\theta}_0) = \boldsymbol{\nabla}_{\boldsymbol{\theta}}^T f^{-1}(u,\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_W-\boldsymbol{\theta}_0) + O_p(n^{-1}). \tag{4.9}
$$

Then $\sqrt{n}\hat{\beta}_k$ can be further written

$$\begin{aligned}
\sqrt{n}\hat{\beta}_k &= \sqrt{n}\tilde{\beta}_k - \frac{1}{\sqrt{2\pi}}\frac{2\pi}{n}\sum_j \cos(ku_j)\boldsymbol{\nabla}_{\boldsymbol{\theta}}^T \log f(u_j,\boldsymbol{\theta}_0) \times \sqrt{n}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0) \\
&- \frac{1}{\sqrt{2\pi}}\frac{2\pi}{n}\sum_j \cos(ku_j)\boldsymbol{\nabla}_{\boldsymbol{\theta}}^T f(u_j,\boldsymbol{\theta}_0)(I_X(u_j) - f(u_j,\boldsymbol{\theta}_0)) \times \sqrt{n}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0) \\
&+ O_p(n^{-1})\sqrt{\frac{2\pi}{n}}\sum_j \cos(ku_j)(I_X(u_j) - f(u_j,\boldsymbol{\theta}_0)) \\
&+ O_p(n^{-1})\sqrt{\frac{2\pi}{n}}\sum_j \cos(ku_j) f(u_j,\boldsymbol{\theta}_0). \qquad (4.10)
\end{aligned}$$

We have the following three claims,

$$\frac{1}{\sqrt{2\pi}}\frac{2\pi}{n}\sum_j \cos(ku_j)\boldsymbol{\nabla}_{\boldsymbol{\theta}}^T f(u_j,\boldsymbol{\theta}_0)(I_X(u_j) - f(u_j,\boldsymbol{\theta}_0)) \times \sqrt{n}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0) = o_p(1), \qquad (4.11)$$

$$\sqrt{\frac{2\pi}{n}}\sum_j \cos(ku_j)(I_X(u_j) - f(u_j,\boldsymbol{\theta}_0)) = o_p(n), \qquad (4.12)$$

$$\sqrt{\frac{2\pi}{n}}\sum_j \cos(ku_j) f(u_j,\boldsymbol{\theta}_0) = o_p(n). \qquad (4.13)$$

We only present the proof of (4.11), the rest of two claims can be showed in the same fashion. (4.11) is true if

$$\sum_{j=1}^{\nu} \cos(ku_j)\frac{\partial}{\partial\theta_\omega}\log f(u_j,\boldsymbol{\theta}_0)\left(\frac{I_X(u_j)}{f(u_j,\boldsymbol{\theta}_0)} - 1\right) = o_p(n).$$

It is thus enough to show

$$\begin{aligned}
&\sum_{j=1}^{\nu}\sum_{h=1}^{\nu} \cos(ku_j)\cos(ku_h)\frac{\partial}{\partial\theta_\omega}\log f(u_j,\boldsymbol{\theta}_0)\frac{\partial}{\partial\theta_\omega}\log f(u_h,\boldsymbol{\theta}_0)\mathrm{E}\left[\left(\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}} - 1\right)\left(\frac{I_{X,h}}{f_{\boldsymbol{\theta}_0,h}} - 1\right)\right] \\
&= o(n^2).
\end{aligned}$$

If $j \neq h$, then LHS of the above result can be written as

$$\begin{aligned} \text{LHS} = &\sum_{j \neq h}\sum \cos(ku_j)\cos(ku_h)\frac{\partial}{\partial\theta_\omega}\log f(u_j, \boldsymbol{\theta}_0)\frac{\partial}{\partial\theta_\omega}\log f(u_h, \boldsymbol{\theta}_0) \\ &\times \left\{ \mathrm{E}\left[\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}\frac{I_{X,h}}{f_{\boldsymbol{\theta}_0,h}}\right] - \mathrm{E}\left[\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}\right] - \mathrm{E}\left[\frac{I_{X,h}}{f_{\boldsymbol{\theta}_0,h}}\right] + 1 \right\}. \end{aligned} \tag{4.14}$$

According to Brockwell & Davis (1991, Theorem 10.3.2), or our **Lemma I**, we have

$$\mathrm{Cov}\left(I_X(u_j), I_X(u_h)\right) = O(n^{-1}).$$

Under (**HT5**), the 5th condition from Hosoya & Masanobu (1982), $f(\cdot) \in \mathrm{Lip}(\alpha)$, $\alpha > 1/2$. From Krogstad (1982), it indicates $\|\mathrm{E}[I_X(u)] - f(u)\|_\infty \leqslant Cn^{-\alpha}$, $\alpha < 1$, or $\|\mathrm{E}[I_X(u)] - f(u)\|_\infty \leqslant C\log(n)/n$, $\alpha = 1$. Here we consider $\alpha \in (\frac{1}{2}, 1)$, that is $\mathrm{E}[I_X(u_j)] = f(u_j, \boldsymbol{\theta}_0) + O(n^{-\alpha})$, where $O(n^{-\alpha})$ is bounded uniformly in j by $Cn^{-\alpha}$. Then we have

$$\mathrm{E}\left[\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}\frac{I_{X,h}}{f_{\boldsymbol{\theta}_0,h}}\right] = \mathrm{Cov}\left(\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}, \frac{I_{X,h}}{f_{\boldsymbol{\theta}_0,h}}\right) + \mathrm{E}\left[\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}\right]\mathrm{E}\left[\frac{I_{X,h}}{f_{\boldsymbol{\theta}_0,h}}\right] = 1 + O(n^{-\alpha}), \tag{4.15}$$

therefore,

$$\mathrm{E}\left[\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}\frac{I_{X,h}}{f_{\boldsymbol{\theta}_0,h}}\right] - \mathrm{E}\left[\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}\right] - \mathrm{E}\left[\frac{I_{X,h}}{f_{\boldsymbol{\theta}_0,h}}\right] + 1 = O(n^{-\alpha}).$$

Then (4.14) ends up with

$$\text{LHS} = O(n^{-\alpha})\sum_{j \neq h}\sum \cos(ku_j)\cos(ku_h)\frac{\partial}{\partial\theta_\omega}\log f(u_j, \boldsymbol{\theta}_0)\frac{\partial}{\partial\theta_\omega}\log f(u_h, \boldsymbol{\theta}_0).$$

It is straightforward to see, this term is bounded by $O(n^{2-\alpha}) = o(n^2)$. For $\alpha = 1$, (4.15) can be shown as $1 + O(\log(n)/n)$, then the rest of the proof should be the same.

If $j = h$, LHS can be written as

$$\text{LHS} = \sum_{j=1}^{\nu} \cos(ku_j)^2 \left[\frac{\partial}{\partial\theta_\omega}\log f(u_j, \boldsymbol{\theta}_0)\right]^2 \left\{ \mathrm{E}\left[\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}\right]^2 + 1 - 2\mathrm{E}\left[\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}\right] \right\}. \tag{4.16}$$

As we know, from Brockwell & Davis (1991, Theorem 10.3.2),

$$\mathrm{Var}\left[I_X(u_j)\right] = f^2(u_j, \boldsymbol{\theta}_0) + O(n^{-1/2}).$$

Then following the same proof argument we did before, consider $f(\cdot) \in \mathrm{Lip}(\alpha)$, $\alpha \in (\frac{1}{2}, 1)$, we have

$$\mathrm{E}\left[\frac{I_{X,j}^2}{f_{\boldsymbol{\theta}_0,j}^2}\right] = \mathrm{Var}\left[\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}\right] + \left\{\mathrm{E}\left[\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}\right]\right\}^2 = 2 + O(n^{-1/2}), \tag{4.17}$$

therefore,

$$\mathrm{E}\left[\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}\right]^2 + 1 - 2\mathrm{E}\left[\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}\right] = 1 + O(n^{-1/2}).$$

In other words, (4.16) is

$$\mathrm{LHS} = \sum_{j=1}^{\nu} \cos(ku_j)^2 \left[\frac{\partial}{\partial\theta_\omega} \log f(u_j, \boldsymbol{\theta}_0)\right]^2 \left\{1 + O(n^{-1/2})\right\} = O(n) = o(n^2).$$

This concludes the proof of (4.11).

Now $\sqrt{n}\hat{\beta}_k$ can be written as

$$\sqrt{n}\hat{\beta}_k = \sqrt{n}\tilde{\beta}_k - \frac{1}{\sqrt{2\pi}}\frac{2\pi}{n}\sum_j \cos(ku_j)\boldsymbol{\nabla}_{\boldsymbol{\theta}}^T \log f(u_j, \boldsymbol{\theta}_0) \times \sqrt{n}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0) + o_p(1).$$

Moreover, about the second term, from the arguments in the proof of Theorem 3.1.2 in Taniguchi & Kakizawa (2000), we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0) = \Gamma(\boldsymbol{\theta}_0)^{-1}\sqrt{n}\frac{2\pi}{n}\sum_j \boldsymbol{\nabla}_{\boldsymbol{\theta}} \log f(u_j, \boldsymbol{\theta}_0)\left(\frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)} - 1\right) + o_p(1). \tag{4.18}$$

In other words,

$$\begin{aligned}\sqrt{n}\hat{\beta}_k &= \sqrt{n}\tilde{\beta}_k + \left[C_k^\dagger(\boldsymbol{\theta}_0)\right]^T \left\{-\Gamma(\boldsymbol{\theta}_0)^{-1}\sqrt{n}\frac{2\pi}{n}\sum_j \boldsymbol{\nabla}_{\boldsymbol{\theta}} \log f(u_j, \boldsymbol{\theta}_0)\left(\frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)} - 1\right)\right\} + o_p(1) \\ &:= A_{1k} + A_{2k} + o_p(1).\end{aligned}$$

Note the sum $C_k^{\dagger}(\boldsymbol{\theta}_0)$ is a discrete analog of the integral $C_k(\boldsymbol{\theta}_0)$.

About A_{1k}, A_{2k}, we have the following claims, as $n \to \infty$,

$$A_{1k} \longrightarrow_d N(0, 2\pi q_0), \tag{4.19}$$

$$A_{2k} \longrightarrow_d N(0, C_k^T(\boldsymbol{\theta}_0)\left\{4\pi\Gamma(\boldsymbol{\theta}_0)^{-1}\right\}C_k(\boldsymbol{\theta}_0)), \tag{4.20}$$

$$\mathrm{Cov}(A_{1k_1}, A_{1k_2}) \longrightarrow 0, \quad k_1 \neq k_2, \tag{4.21}$$

$$\mathrm{Cov}(A_{2k_1}, A_{2k_2}) \longrightarrow C_{k_1}^T(\boldsymbol{\theta}_0)\left\{4\pi\Gamma(\boldsymbol{\theta}_0)^{-1}\right\}C_{k_2}(\boldsymbol{\theta}_0), \tag{4.22}$$

$$\mathrm{Cov}(A_{1k_1}, A_{2k_2}) \longrightarrow -C_{k_1}^T(\boldsymbol{\theta}_0)\left\{4\pi\Gamma(\boldsymbol{\theta}_0)^{-1}\right\}C_{k_2}(\boldsymbol{\theta}_0). \tag{4.23}$$

Note (4.19) and (4.21) are from our Theorem 1.2.1.

Consider (4.20) and (4.22) since they are related, for better explanations, we firstly calculate (4.22).

$$\begin{aligned}
\mathrm{Cov}(A_{2k_1}, A_{2k_2}) &= \left\{\frac{1}{\sqrt{2\pi}}\frac{2\pi}{n}\sum_j \cos(k_1 u_j)\boldsymbol{\nabla}_{\boldsymbol{\theta}}^T \log f(u_j, \boldsymbol{\theta}_0)\right\}\left\{-2\pi\Gamma(\boldsymbol{\theta}_0)^{-1}\right\} \\
&\times \left\{\frac{1}{n}\sum_h \boldsymbol{\nabla}_{\boldsymbol{\theta}} \log f(u_h, \boldsymbol{\theta}_0)\sqrt{\frac{2\pi}{n}}\sum_l \cos(k_2 u_l)\boldsymbol{\nabla}_{\boldsymbol{\theta}}^T \log f(u_l, \boldsymbol{\theta}_0)\right\}\left\{-2\pi\Gamma(\boldsymbol{\theta}_0)^{-1}\right\} \\
&\times \left\{\frac{1}{\sqrt{n}}\sum_m \boldsymbol{\nabla}_{\boldsymbol{\theta}} \log f(u_m, \boldsymbol{\theta}_0)\mathrm{Cov}\left(\frac{I_X(u_h)}{f(u_h, \boldsymbol{\theta}_0)}, \frac{I_X(u_m)}{f(u_m, \boldsymbol{\theta}_0)}\right)\right\} \\
&= \left[C_{k_1}^{\dagger}(\boldsymbol{\theta}_0)\right]^T\left\{-2\pi\Gamma(\boldsymbol{\theta}_0)^{-1}\right\}\left\{\frac{1}{n}\sum_h\sum_m \boldsymbol{\nabla}_{\boldsymbol{\theta}} \log f(u_h, \boldsymbol{\theta}_0)\boldsymbol{\nabla}_{\boldsymbol{\theta}}^T \log f(u_m, \boldsymbol{\theta}_0)\right. \\
&\times \left.\mathrm{Cov}\left(\frac{I_X(u_h)}{f(u_h, \boldsymbol{\theta}_0)}, \frac{I_X(u_m)}{f(u_m, \boldsymbol{\theta}_0)}\right)\right\}\left\{-2\pi\Gamma(\boldsymbol{\theta}_0)^{-1}\right\}C_{k_2}^{\dagger}(\boldsymbol{\theta}_0).
\end{aligned}$$

To find the final result, the critical step is to show

$$\begin{aligned}
&\frac{1}{n}\sum_h\sum_m \frac{\partial}{\partial\theta_\omega}\log f(u_h, \boldsymbol{\theta}_0)\frac{\partial}{\partial\theta_\upsilon}\log f(u_m, \boldsymbol{\theta}_0)\mathrm{Cov}\left(\frac{I_X(u_h)}{f(u_h, \boldsymbol{\theta}_0)}, \frac{I_X(u_m)}{f(u_m, \boldsymbol{\theta}_0)}\right) \\
&= 2\cdot\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta_\omega}\log f(u, \boldsymbol{\theta}_0)\frac{\partial}{\partial\theta_\upsilon}\log f(u, \boldsymbol{\theta}_0)du + o(1).
\end{aligned} \tag{4.24}$$

where $o(1)$ is infinitely small as $n \to \infty$.

The proof idea is almost the same as the proof of Cov $\left(\sqrt{n}\tilde{\beta}_{k_1}, \sqrt{n}\tilde{\beta}_{k_2}\right)$, about the LHS of (4.24),

$$\begin{aligned}
2\pi \cdot \text{LHS} &= 4 \cdot \frac{2\pi}{n} \sum_{h=1}^{\nu} \sum_{m=1}^{\nu} \frac{\partial}{\partial \theta_\omega} \log f(u_h, \boldsymbol{\theta}_0) \frac{\partial}{\partial \theta_\upsilon} \log f(u_m, \boldsymbol{\theta}_0) \text{Cov}\left(\frac{I_X(u_h)}{f(u_h, \boldsymbol{\theta}_0)}, \frac{I_X(u_m)}{f(u_m, \boldsymbol{\theta}_0)}\right) \\
&= 4 \cdot \frac{2\pi}{n} \sum_{h=1}^{\nu} \frac{\partial}{\partial \theta_\omega} \log f(u_h, \boldsymbol{\theta}_0) \frac{\partial}{\partial \theta_\upsilon} \log f(u_h, \boldsymbol{\theta}_0) \text{Var}\left[\frac{I_X(u_h)}{f(u_h, \boldsymbol{\theta}_0)}\right] \\
&+ 4 \cdot \frac{2\pi}{n} \sum_{h \neq m}\sum \frac{\partial}{\partial \theta_\omega} \log f(u_h, \boldsymbol{\theta}_0) \frac{\partial}{\partial \theta_\upsilon} \log f(u_m, \boldsymbol{\theta}_0) \text{Cov}\left(\frac{I_X(u_h)}{f(u_h, \boldsymbol{\theta}_0)}, \frac{I_X(u_m)}{f(u_m, \boldsymbol{\theta}_0)}\right) \\
&:= J_{1n} + J_{2n}.
\end{aligned}$$

Consider J_{1n}, utilizing the result $\text{Var}\left(I_X(u_h)\right) = f^2(u_h, \boldsymbol{\theta}_0) + O(n^{-1/2})$ in Brockwell & Davis (1991, Theorem 10.3.2), we have

$$J_{1n} = 4 \cdot \frac{2\pi}{n} \sum_{h=1}^{\nu} \frac{\partial}{\partial \theta_\omega} \log f(u_h, \boldsymbol{\theta}_0) \frac{\partial}{\partial \theta_\upsilon} \log f(u_h, \boldsymbol{\theta}_0) + R_{1n},$$

where $|R_{1n}| \leq C n^{-1/2} \to 0$, as $n \to \infty$. Following the proof arguments of **Lemma IV**, it is obvious to see $\frac{2\pi}{n} \sum_h \frac{\partial}{\partial \theta_\omega} \log f(u_h, \boldsymbol{\theta}_0) \frac{\partial}{\partial \theta_\upsilon} \log f(u_h, \boldsymbol{\theta}_0) = \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_\omega} \log f(u, \boldsymbol{\theta}_0) \frac{\partial}{\partial \theta_\upsilon} \log f(u, \boldsymbol{\theta}_0) du + O(n^{-1})$. Therefore we showed $J_{1n} \to 2 \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_\omega} \log f(u, \boldsymbol{\theta}_0) \frac{\partial}{\partial \theta_\upsilon} \log f(u, \boldsymbol{\theta}_0) du$, as $n \to \infty$.

Consider J_{2n}, from **Lemma II**, we have

$$\begin{aligned}
J_{2n} &= 4 \cdot \frac{2\pi \eta^4}{n^2} \sum_{h=1}^{\nu} \sum_{m=1}^{\nu} \frac{\partial}{\partial \theta_\omega} \log f(u_h, \boldsymbol{\theta}_0) \frac{\partial}{\partial \theta_\upsilon} \log f(u_m, \boldsymbol{\theta}_0) \\
&- 4 \cdot \frac{2\pi \eta^4}{n^2} \sum_{h=1}^{\nu} \frac{\partial}{\partial \theta_\omega} \log f(u_h, \boldsymbol{\theta}_0) \frac{\partial}{\partial \theta_\upsilon} \log f(u_h, \boldsymbol{\theta}_0) \\
&+ o\left(\frac{2\pi}{n^2} \sum_{h \neq m}\sum \frac{\partial}{\partial \theta_\omega} \log f(u_h, \boldsymbol{\theta}_0) \frac{\partial}{\partial \theta_\upsilon} \log f(u_m, \boldsymbol{\theta}_0) \text{Cov}\left(\frac{I_X(u_h)}{f(u_h, \boldsymbol{\theta}_0)}, \frac{I_X(u_m)}{f(u_m, \boldsymbol{\theta}_0)}\right)\right)
\end{aligned} \tag{4.25}$$

Following the proof arguments in **Lemma IV**, it is straightforward to see the first item in (4.25) is $O(n^{-2})$, then the rest of terms can be showed in the same fashion. In the end, we have $J_{2n} = O(n^{-1})$. It converges to 0 as $n \to 0$. This concludes the proof of (4.24) and

(4.22).

About (4.20), the general result is given in Taniguchi & Kakizawa (2000, Theorem 3.1.2, (iii)). Since we discuss under the condition **A2**, the more specific result is given in Hosoya & Masanobu (1982, Proposition 3.1). Note the redundant part in its asymptotic variance in the Proposition 3.1 is the integral analog of the first term in our I_{2n} (4.25). As we mentioned before, under H_0, we have $f(u) = f(u, \boldsymbol{\theta}_0) = f(u, \bar{\boldsymbol{\theta}})$, then following our discussions on the Kolmogrov's formula in **Lemma IV**, it is straightforward to see the redundant part equals 0, which ends up with the same asymptotic variance as we calculate in (4.22). This is also discussed in Hosoya & Masanobu (1982, Remark 3.1).

Then the proof of (4.23) is almost the same as the preceding work in (4.22). Since it's repeated, we choose to skip this part.

After we obtain the results (4.19)-(4.22), we figured out what is inside of $\sqrt{n}\hat{\beta}_k$. Recall $\sum_k \alpha_k \sqrt{n}\hat{\beta}_k$, through the continuous mapping theorem, we have, as $n \to \infty$,

$$\sum_k \alpha_k \sqrt{n}\hat{\beta}_k = \sum_k \alpha_k (A_{1k} + A_{2k}) + o_p(1) \longrightarrow_d N(0, V_{\boldsymbol{\theta}_0}(\boldsymbol{\alpha})),$$

where

$$V_{\boldsymbol{\theta}_0}(\boldsymbol{\alpha}) = \sum_{k_1}\sum_{k_2} \alpha_{k_1}\alpha_{k_2} \left\{\mathrm{Cov}_\infty(A_{1k_1}, A_{1k_2}) + \mathrm{Cov}_\infty(A_{2k_1}, A_{2k_2}) + 2\mathrm{Cov}_\infty(A_{1k_1}, A_{2k_2})\right\}$$

$\mathrm{Cov}_\infty(\cdot,\cdot)$ denotes the asymptotic covariance values, they are represented in (4.19)-(4.23).

Through the Cramér-Wold device, we have

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_m \end{pmatrix} \longrightarrow_d N(0, \Sigma(\boldsymbol{\theta}_0)),$$

where $\Sigma(\boldsymbol{\theta}_0)$ with entries

$$\Sigma_{k,l}(\boldsymbol{\theta}_0) = \frac{1}{2}(V_{\boldsymbol{\theta}_0}(e_k + e_l) - V_{\boldsymbol{\theta}_0}(e_k) - V_{\boldsymbol{\theta}_0}(e_l)),$$

with e_k denotes the kth unit vector.

For off-diagonal terms, when $k \neq l$, we have

$$\begin{aligned}\Sigma_{k,l}(\boldsymbol{\theta}_0) &= \mathrm{Cov}_\infty(A_{1k}, A_{1l}) + \mathrm{Cov}_\infty(A_{2k}, A_{2l}) + 2\mathrm{Cov}_\infty(A_{1k}, A_{2l}) \\ &= -C_k^T(\boldsymbol{\theta}_0)\left\{4\pi\Gamma(\boldsymbol{\theta}_0)^{-1}\right\}C_l(\boldsymbol{\theta}_0).\end{aligned}$$

For diagonal terms, we have

$$\begin{aligned}\Sigma_{k,k}(\boldsymbol{\theta}_0) &= \mathrm{Var}_\infty[A_{1k}] + \mathrm{Var}_\infty[A_{2k}] + 2\mathrm{Cov}_\infty(A_{1k}, A_{2k}) \\ &= 2\pi q_0 - C_k^T(\boldsymbol{\theta}_0)\left\{4\pi\Gamma(\boldsymbol{\theta}_0)^{-1}\right\}C_k(\boldsymbol{\theta}_0).\end{aligned}$$

This concludes the proof. □

Proof of Proposition 1.3.1: The matrix $\Gamma(\boldsymbol{\theta})$ is a continuous function of θ by our assumptions on $f_{\boldsymbol{\theta}}$, likewise each entry of the inverse matrix is continuous in θ. $C_k(\boldsymbol{\theta})$ is also continuous.

As we have $\hat{\boldsymbol{\theta}}_W \to_p \boldsymbol{\theta}_0$, then the result follows immediately.

Note if it is a Gaussian time series, then $q_0 = 1$, there's no additional work to estimate η_4. Thus the condition **A0** can be relived to **A1** and **A2**. □

4.3 Proof of Theorem 1.4.1 & 1.4.2

Proof of Theorem 1.4.1: Through the preceding proof of Theorem 1.3.1, we only need to show (4.18) remains consistent with $\hat{\boldsymbol{\theta}}_M$, then the rest of work should be the same. In fact, the difference between $\hat{\boldsymbol{\theta}}_M$ and $\hat{\boldsymbol{\theta}}_W$ is $o_p(n^{-\frac{1}{2}})$, this is ensured by Theorem 3.2 and Theorem 3.3 from Dahlhaus & Wefelmeyer (1996), then the result follows immediately. $\square$

Lemma V Suppose n is large enough, under assumptions of Theorem 1.4.2, we have

$$n^{-1}X_n^T G_n^{-1}(\boldsymbol{\phi}_0)X_n - \frac{2\pi}{n}\sum_{j=0}^{n-1}\frac{I_X(u_j)}{g(u_j,\boldsymbol{\phi}_0)} = O_p(n^{-1}).$$

Proof:

We note

$$G_n(\boldsymbol{\phi}_0) = [c(s-t,\boldsymbol{\phi}_0)]_{s,t} = \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} g(u,\boldsymbol{\phi}_0)e^{\mathrm{i}(s-t)u}du\right]_{s,t}, \quad \text{where } s,t = 1,\cdots,n.$$

Consider its discrete analog as the following

$$G_n^{\dagger}(\boldsymbol{\phi}_0) = [c^{\dagger}(s-t,\boldsymbol{\phi}_0)]_{s,t} = \left[\frac{1}{2\pi}\cdot\frac{2\pi}{n}\sum_{j=0}^{n-1} g(u_j,\boldsymbol{\phi}_0)e^{\mathrm{i}(s-t)u_j}\right]_{s,t}.$$

We recognize $G_n^{\dagger}(\boldsymbol{\phi}_0)$ can be written as $U_n^* A_n(\boldsymbol{\phi}_0)U_n$, where $A_n(\boldsymbol{\phi}_0) = \mathrm{diag}\{g(u_0,\boldsymbol{\phi}_0),\cdots,$ $g(u_{n-1},\boldsymbol{\phi}_0)\}$ and $U_n = (\mu_0,\cdots,\mu_{n-1})$, where the row vector $\mu_j = \left\{\frac{1}{\sqrt{n}}e^{\mathrm{i}u_j t}\right\}_{t=0}^{n-1}$. It is obvious to see that U_n is an unitary matrix. That is, $U_n U_n^* = U_n^* U_n = I$, $|U_n| = 1$.

Then we have

$$X_n^T[G_n^{\dagger}(\boldsymbol{\theta}_0)]^{-1}X_n = X_n^T U_n^* D_n^{-1}(\boldsymbol{\phi}_0)U_n X_n = (X_n^T U_n^* D_n^{-1/2}(\boldsymbol{\phi}_0))(X_n^T U_n^* D_n^{-1/2}(\boldsymbol{\phi}_0))^*.$$

The jth entry of $X_n^T U_n^*$ is $X_n^T \bar{\mu}_j^T = \frac{1}{n}\sum_{t=0}^{n-1} X_t e^{-\mathrm{i}u_j t} = \sqrt{n}J_j$, thus we have the jth entry of

$X_n^T U_n^* D_n^{-1/2}(\boldsymbol{\phi}_0) = \sqrt{n} J_j / \sqrt{g(u_j, \boldsymbol{\phi}_0)}$, then

$$X_n^T [G_n^{\dagger}(\boldsymbol{\phi}_0)]^{-1} X_n = \sum_{j=0}^{n-1} \frac{n|J_j|^2}{g(u_j, \boldsymbol{\phi}_0)} = 2\pi \sum_{j=0}^{n} \frac{\frac{1}{2\pi n} \left|\sum_{t=0}^{n-1} X_t e^{-\mathtt{i} u_j t}\right|^2}{g(u_j, \boldsymbol{\phi}_0)}$$
$$= 2\pi \sum_{j=0}^{n-1} \frac{I_X(u_j)}{g(u_j, \boldsymbol{\phi}_0)}.$$

Therefore, to show the final result, it is equivalent to show

$$n^{-1} \left| X_n^T G_n^{-1}(\boldsymbol{\phi}_0) X_n - X_n^T [G_n^{\dagger}(\boldsymbol{\phi}_0)]^{-1} X_n \right| = O_p(n^{-1}).$$

As we have

$$\begin{aligned}\left| X_n^T G_n^{-1}(\boldsymbol{\phi}_0) X_n - X_n^T [G_n^{\dagger}(\boldsymbol{\phi}_0)]^{-1} X_n \right| &= \left| X_n^T G_n^{-1}(\boldsymbol{\phi}_0)[G_n(\boldsymbol{\phi}_0) - G_n^{\dagger}(\boldsymbol{\phi}_0)][G_n^{\dagger}(\boldsymbol{\phi}_0)]^{-1} X_n \right| \\ &= \left| X_n^T G_n^{-1}(\boldsymbol{\phi}_0) U_n^* U_n [G_n(\boldsymbol{\phi}_0) - G_n^{\dagger}(\boldsymbol{\phi}_0)] U_n^* U_n \right. \\ &\left. \times [G_n^{\dagger}(\boldsymbol{\phi}_0)]^{-1} X_n \right| \\ &= \left| X_n^T G_n^{-1}(\boldsymbol{\phi}_0) U_n^* [U_n G_n(\boldsymbol{\phi}_0) U_n^* - A_n(\boldsymbol{\phi}_0)] U_n \right. \\ &\left. \times [G_n^{\dagger}(\boldsymbol{\phi}_0)]^{-1} X_n \right|,\end{aligned}$$

About $U_n G_n(\boldsymbol{\phi}_0) U_n^* - A_n(\boldsymbol{\phi}_0)$, follow Fuller (1995, Corollary 4.2.1), and its proof argument, we have, with each entry in $U_n G_n(\boldsymbol{\phi}_0) U_n^* - A_n(\boldsymbol{\phi}_0)$,

$$\left| [a_{ij}(\boldsymbol{\phi}_0)]_{i,j} \right| = \left| [U_n G_n(\boldsymbol{\phi}_0) U_n^* - A_n(\boldsymbol{\phi}_0)]_{i,j} \right| < \frac{2L}{n},$$

where $i, j = 0, \cdots, n-1$, and $L = \sum |h||c(h, \boldsymbol{\phi}_0)| < \infty$ (1.13), which is discussed in the Section 1.4. Then the rest proof can simply follow the related proof arguments in Yao & Brockwell (2006, Lemma 3). That is, for $\boldsymbol{x} = (x_0, \cdots, x_{n-1}), \boldsymbol{y} = (y_0, \cdots, y_{n-1}) \in \mathbb{C}^n$, $\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = 1$, we have

$$\begin{aligned}\left|\boldsymbol{x}\left[U_nG_n(\boldsymbol{\phi}_0)U_n^* - A_n(\boldsymbol{\phi}_0)\right]\boldsymbol{y}\right| &= \left|\sum_{i,j=0}^{n-1} x_i a_{ij}(\boldsymbol{\phi}_0)y_j\right| \\ &\leqslant \sum_{i,j=0}^{n-1} |x_i||a_{ij}(\boldsymbol{\phi}_0)||y_j| \\ &< \frac{2L}{n}\sum_{i,j=0}^{n-1}|x_i||y_j| \\ &\leqslant \frac{2L}{n}\|\boldsymbol{x}\|\|\boldsymbol{y}\| \\ &= \frac{2L}{n}.\end{aligned}$$

Hence

$$\begin{aligned}&\left|X_n^T G_n^{-1}(\boldsymbol{\phi}_0)U_n^*[U_nG_n(\boldsymbol{\phi}_0)U_n^* - A_n(\boldsymbol{\phi}_0)]U_n[G_n^\dagger(\boldsymbol{\phi}_0)]^{-1}X_n\right| \\ &< \frac{2L}{n}\left\{\left|X_n^T G_n^{-2}(\boldsymbol{\phi}_0)X_n\right|\left|X_n^T[G_n^\dagger(\boldsymbol{\phi}_0)]^{-2}X_n\right|\right\}^{1/2},\end{aligned}$$

set $K^{-1} := \inf_{u\in[-\pi,\pi]} g(u,\boldsymbol{\phi}_0)$, through the construction of $G_n^\dagger(\boldsymbol{\phi}_0)$, we can see its eigenvalues are bounded below by K^{-1}. From Brockwell & Davis (1991, Proposition 4.5.3), all eigenvalues of $G_n(\boldsymbol{\phi}_0)$ are also bounded below by K^{-1}, thus their inverses are bounded above by K. Then we have

$$\left\{\left|X_n^T G_n^{-2}(\boldsymbol{\phi}_0)X_n\right|\left|X_n^T[G_n^\dagger(\boldsymbol{\phi}_0)]^{-2}X_n\right|\right\}^{1/2} \leqslant K^2 X_n^T X_n.$$

In other words,

$$\mathrm{E}\left|X_n^T\left[G_n^{-1}(\boldsymbol{\phi}_0) - [G_n^\dagger(\boldsymbol{\phi}_0)]^{-1}\right]X_n\right| < \frac{2LK^2}{n}\mathrm{E}\left[X_n^T X_n\right] = 2LK^2\gamma(0).$$

It concludes the proof. □

Proof of Theorem 1.4.2: Note

$$\begin{aligned}\hat{\beta}_0 &= \frac{2\sqrt{\pi}}{n}\sum_{j=1}^{\nu}\frac{I_X(u_j)/g(u_j,\hat{\boldsymbol{\phi}}_M)}{\sigma_n^2(\hat{\boldsymbol{\phi}}_M)/(2\pi)} \\ &= \sqrt{\pi}\frac{\frac{2\pi}{n}\sum_j I_X(u_j)/g(u_j,\hat{\boldsymbol{\phi}}_M)}{\sigma_n^2(\hat{\boldsymbol{\phi}}_M)} \\ &= \frac{\sqrt{\pi}}{\sigma_n^2(\hat{\boldsymbol{\phi}}_M)}\left\{\frac{2\pi}{n}\sum_j\frac{I_X(u_j)}{g(u_j,\hat{\boldsymbol{\phi}}_M)} - \frac{2\pi}{n}\sum_j\frac{I_X(u_j)}{g(u_j,\boldsymbol{\phi}_0)}\right\} + \frac{\sqrt{\pi}}{\sigma_n^2(\hat{\boldsymbol{\phi}}_M)}\left\{\frac{2\pi}{n}\sum_j\frac{I_X(u_j)}{g(u_j,\boldsymbol{\phi}_0)}\right. \\ &\left. -\sigma_n^2(\boldsymbol{\phi}_0)\right\} + \sqrt{\pi}\frac{\sigma_n^2(\boldsymbol{\phi}_0)-\sigma_n^2(\hat{\boldsymbol{\phi}}_M)}{\sigma_n^2(\hat{\boldsymbol{\phi}}_M)} + \sqrt{\pi}.\end{aligned}$$

Lemma V shows the second term is $O_p(n^{-1})$. We only need to show the first term also ends up with $O_p(n^{-1})$, that is

$$\frac{2\pi}{n}\sum_j I_X(u_j)[g^{-1}(u_j,\hat{\boldsymbol{\phi}}_M) - g^{-1}(u_j,\boldsymbol{\phi}_0)] = O_p(n^{-1}) \tag{4.26}$$

It is the same as we did before. According the smoothness conditions of $g_{\boldsymbol{\phi}}$ in **C0**, we have by a Taylor series expansion,

$$g^{-1}(u,\hat{\boldsymbol{\phi}}_M) - g^{-1}(u,\boldsymbol{\phi}_0) = \boldsymbol{\nabla}_{\boldsymbol{\phi}}^T g^{-1}(u,\boldsymbol{\phi}_0)(\hat{\boldsymbol{\phi}}_M-\boldsymbol{\phi}_0) + \frac{1}{2}(\hat{\boldsymbol{\phi}}_M-\boldsymbol{\phi}_0)^T \mathrm{H}_{\boldsymbol{\phi}} g^{-1}(u,\tilde{\boldsymbol{\phi}})(\hat{\boldsymbol{\phi}}_M-\boldsymbol{\phi}_0),$$

where $\tilde{\boldsymbol{\phi}} = \boldsymbol{\phi}_0 + \lambda(\hat{\boldsymbol{\phi}}_M - \boldsymbol{\phi}_0)$ for some $\lambda \in (0,1)$, and $\mathrm{H}_{\boldsymbol{\phi}} g^{-1}(u,\tilde{\boldsymbol{\phi}}) \longrightarrow_p \mathrm{H}_{\boldsymbol{\phi}} g^{-1}(u,\boldsymbol{\phi}_0)$. From Theorem 3 in Hannan (1973), we also have $\left\|\hat{\boldsymbol{\phi}}_M - \boldsymbol{\phi}_0\right\| = O_p(n^{-1/2})$, then the above Taylor expansion can also be written

$$g^{-1}(u,\hat{\boldsymbol{\phi}}_M) - g^{-1}(u,\boldsymbol{\phi}_0) = \boldsymbol{\nabla}_{\boldsymbol{\phi}}^T g^{-1}(u,\boldsymbol{\phi}_0)(\hat{\boldsymbol{\phi}}_M-\boldsymbol{\phi}_0) + O_p(n^{-1}).$$

To show (4.26), it is equivalent to show the following, for each ω,

$$\frac{2\pi}{n}\sum_j I_X(u_j)\frac{\partial}{\partial\phi_\omega}g^{-1}(u_j,\boldsymbol{\phi}_0)(\hat{\phi}_{M_\omega}-\phi_{0_\omega}) = O_p(n^{-1}) \tag{4.27}$$

$$O_p(n^{-1})\frac{2\pi}{n}\sum_j I_X(u_j) = O_p(n^{-1}) \tag{4.28}$$

About (4.27), it is equivalent to show

$$\sum_j I_X(u_j)\frac{\partial}{\partial\phi_\omega}g^{-1}(u_j,\boldsymbol{\phi}_0) = O_p(\sqrt{n}).$$

From the assumption, since the spectral density is in the form (1.12), the above result is equivalent to

$$\sum_j \frac{I_X(u_j)}{f(u_j,\boldsymbol{\phi}_0)}\frac{\partial}{\partial\phi_\omega}\log g(u_j,\boldsymbol{\phi}_0) = O_p(\sqrt{n}).$$

And this is true if

$$\sum_j\sum_h \frac{\partial}{\partial\phi_\omega}\log g(u_j,\boldsymbol{\phi}_0)\frac{\partial}{\partial\phi_\omega}\log g(u_h,\boldsymbol{\phi}_0)\mathrm{E}\left[\frac{I_X(u_j)}{f(u_j,\boldsymbol{\phi}_0)}\frac{I_X(u_h)}{f(u_h,\boldsymbol{\phi}_0)}\right] = O(n). \tag{4.29}$$

Recall (1.13), it indicates $\sum|h||\gamma(h)| < \infty$, then from Krogstad (1982), we have $\|\mathrm{E}I_X(u) - f(u,\boldsymbol{\theta}_0)\|_\infty < Cn^{-1}$.

If $j \neq h$, follow the same argument we have put on (4.15), it can be shown that

$$\mathrm{E}\left[\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}\frac{I_{X,h}}{f_{\boldsymbol{\theta}_0,h}}\right] = 1 + O(n^{-1}).$$

Now the LHS of (4.29) can be written as

$$\sum\sum_{j\neq h}\frac{\partial}{\partial\phi_\omega}\log g(u_j,\boldsymbol{\phi}_0)\frac{\partial}{\partial\phi_\omega}\log g(u_h,\boldsymbol{\phi}_0) + O\left(n^{-1}\sum\sum_{j\neq h}\frac{\partial}{\partial\phi_\omega}\log g(u_j,\boldsymbol{\phi}_0)\frac{\partial}{\partial\phi_\omega}\log g(u_h,\boldsymbol{\phi}_0)\right).$$

Then we only need to consider the first term, which can be written as

$$\sum_j \frac{\partial}{\partial \phi_\omega} \log g(u_j, \boldsymbol{\phi}_0) \sum_h \frac{\partial}{\partial \phi_\omega} \log g(u_h, \boldsymbol{\phi}_0) - \sum_j \left[\frac{\partial}{\partial \phi_\omega} \log g(u_j, \boldsymbol{\phi}_0) \right]^2 . \tag{4.30}$$

Since we are under assumptions of Hannan (1973), according to his result (3), and our proof arguments in **Lemma IV**, it can be easily shown that

$$\frac{1}{n} \sum_j \frac{\partial}{\partial \phi_\omega} \log g(u_j, \boldsymbol{\phi}_0) = O(n^{-1})$$

Then (4.30) ends up with $O(n)$, which concludes the proof of (4.29) in case when $j \neq h$.

If $j = h$, (4.29) can be written as

$$\sum_j \left(\frac{\partial}{\partial \phi_\omega} \log g(u_j, \boldsymbol{\phi}_0) \right)^2 \mathrm{E} \left[\frac{I_X(u_j)}{f(u_j, \boldsymbol{\phi}_0)} \right]^2 = O(n).$$

About $\mathrm{E}\left[\frac{I_X(u_j)}{f(u_j, \boldsymbol{\phi}_0)} \right]^2$, it ends up with the same as (4.17). Then LHS is

$$\sum_j \{2 + O(n^{-1/2})\} \left(\frac{\partial}{\partial \phi_\omega} \log g(u_j, \boldsymbol{\phi}_0) \right)^2$$

The result comes immediately after we note $\sum_j \left(\frac{\partial}{\partial \phi_\omega} \log g(u_j, \boldsymbol{\phi}_0) \right)^2 = O(n)$.

(4.28) can be showed in the same fashion. Then we obtain the result (4.26), it concludes the proof. $\square$

4.4 Proof of Theorem 1.5.1 & 1.5.2

Lemma VI Let $(X,Y) \sim N(0,0,\sigma_1^2,\sigma_2^2,\sigma_{12})$ with $\mathrm{Var}[X] = \sigma_1^2$, $\mathrm{Var}[Y] = \sigma_2^2$, $\mathrm{Cov}(X,Y) = \sigma_{12}$, then

$$\mathrm{Cov}(X^2,Y^2) = 2\sigma_{12}^2.$$

Proof:

Let $U = X/\sigma_1$ and $V = Y/\sigma_2$, then $(U,V) \sim N(0,0,1,1,\rho)$, where $\rho = \sigma_{12}/(\sigma_1\sigma_2)$. For such a standardized normal distribution, we have

$$f(u,v) = \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{u^2-2\rho uv+v^2}{2(1-\rho^2)}\right], \qquad \text{where } u,v \in \mathbb{R}.$$

Set $W_1 = U - \rho V$, $W_2 = \sqrt{1-\rho^2}V$. As $W_1, W_2 \in \mathbb{R}$, we also have $U = W_1 + \frac{\rho}{\sqrt{1-\rho^2}}W_2$, $V = \frac{W_2}{\sqrt{1-\rho^2}}$. Moreover, since

$$|J| = \begin{vmatrix} 1 & \frac{\rho}{\sqrt{1-\rho^2}} \\ 0 & \frac{1}{\sqrt{1-\rho^2}} \end{vmatrix} = \frac{1}{\sqrt{1-\rho^2}},$$

we have

$$\begin{aligned} f(w_1,w_2) &= \frac{1}{2\pi(1-\rho^2)}\exp\left[-\frac{w_1^2+w_2^2}{2(1-\rho^2)}\right] \\ &= \frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left[-\frac{w_1^2}{2(1-\rho^2)}\right] \times \frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left[-\frac{w_2^2}{2(1-\rho^2)}\right]. \end{aligned}$$

It is obvious to see W_1 and W_2 are independent, and $W_1, W_2 \sim N(0, 1-\rho^2)$. Now consider

$$\begin{aligned} \mathrm{Cov}(U^2,V^2) &= \mathrm{Cov}(W_1^2 + 2\frac{\rho}{\sqrt{1-\rho^2}}W_1W_2 + \frac{\rho^2}{1-\rho^2}W_2^2, \frac{1}{1-\rho^2}W_2^2) \\ &= 0 + 0 + \frac{\rho^2}{(1-\rho^2)^2}\mathrm{Cov}(W_2^2,W_2^2), \end{aligned}$$

we only need to figure out $\mathrm{E}[W_2^4] - (\mathrm{E}[W_2^2])^2$. It should not be difficult to see $\mathrm{E}[W_2^4] = 3(1-\rho^2)^2$ and $\mathrm{E}[W_2^2] = 1-\rho^2$. Therefore $\mathrm{Cov}(U^2, V^2) = \frac{\rho^2}{(1-\rho^2)^2} \cdot 2(1-\rho^2)^2 = 2\rho^2$, which concludes the proof. □

Lemma VII Let $u_j = \frac{2\pi j}{n}$, $u_k = \frac{2\pi k}{n}$, $j, k = 1, \cdots, \nu$. Then for the Gaussian linear process $\{X_t\}$ under the same conditions as those of Theorem 1.5.1, we have,

$$\mathrm{Cov}\left(\frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)}, \frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)}\right) = O\left(\left(\frac{k}{j}\right)^{2d_0} \frac{\log^2 k}{k^2}\right),$$

uniformly for $1 \leqslant j \leqslant k \leqslant \nu$.

Proof:

The proof of this Lemma is mainly based on Moulines & Soulier (1999). Note

$$\frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)} = \left(\frac{\frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n X_t \cos(u_j t)}{f^{1/2}(u_j, \boldsymbol{\theta}_0)}\right)^2 + \left(\frac{\frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n X_t \sin(u_j t)}{f^{1/2}(u_j, \boldsymbol{\theta}_0)}\right)^2 := V_{c,j}^2 + V_{s,j}^2,$$

then we have

$$\begin{aligned}\mathrm{Cov}\left(\frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)}, \frac{I_X(u_k)}{f(u_k, \boldsymbol{\theta}_0)}\right) &= \mathrm{Cov}(V_{c,j}^2 + V_{s,j}^2, V_{c,k}^2 + V_{s,k}^2) \\ &= \mathrm{Cov}(V_{c,j}^2, V_{c,k}^2) + \mathrm{Cov}(V_{c,j}^2, V_{s,k}^2) + \mathrm{Cov}(V_{s,j}^2, V_{c,k}^2) \\ &+ \mathrm{Cov}(V_{s,j}^2, V_{s,k}^2)\end{aligned}$$

Now $\mathrm{Cov}\left(\frac{I_X(u_j)}{f(u_j,\boldsymbol{\theta}_0)}, \frac{I_X(u_k)}{f(u_k,\boldsymbol{\theta}_0)}\right)$ can be directly evaluated by those four terms $\mathrm{Cov}(V_j^2, V_k^2)$. We use the first term $\mathrm{Cov}(V_{c,j}^2, V_{c,k}^2)$ as an example, since the rest of three terms can be showed in the same fashion.

Denote

$$\omega_n(u) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n X_t e^{\mathrm{i}ut},$$

then

$$\begin{aligned}
&\mathrm{E}\left[\frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n}X_t\cos(u_j t)\frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n}X_t\cos(u_k t)\right]\\
&=\frac{1}{4}\mathrm{E}\left[\omega_n(u_j)\omega_n(u_k)+\omega_n(-u_j)\omega_n(u_k)+\omega_n(u_j)\omega_n(-u_k)+\omega_n(-u_j)\omega_n(-u_k)\right].
\end{aligned}$$

Therefore we mainly consider $\mathrm{E}[\omega_n(\pm u_j)\omega_n(\pm u_k)]$. Here we only give the proof of $\mathrm{E}[\omega_n(u_j)\omega_n(u_k)]$, others can be done in the same way.

From (D.1) in Moulines & Soulier (1999), we have

$$\begin{aligned}
&\mathrm{E}[\omega_n(u_j)\omega_n(u_k)]\\
&=\frac{1}{2\pi n}\int_0^{\pi}\left[f\left(\frac{u_k-u_j}{2}-t\right)+f\left(\frac{u_k-u_j}{2}+t\right)\right]H_n(\frac{u_k+u_j}{2}-t)H_n(\frac{u_k+u_j}{2}+t)dt,
\end{aligned}$$

where $f(\cdot)$ is the spectral density of $\{X_t\}$, and

$$H_n(x)=\sum_{t=1}^{n}e^{\mathtt{i}tx}=\exp\left(\mathtt{i}(n+1)\frac{x}{2}\right)\frac{\sin(nx/2)}{\sin(x/2)}.$$

Now set $g(t,\omega)=f(\omega-t)+f(\omega+t)$, $\Delta_n(t,z)=H_n(z-t)H_n(z+t)$, note $g(\omega,t)$ satisfies condition (D.6) and (D.7) in Moulines & Soulier (1999), and it is not difficult to verify that

$$\begin{aligned}
&\int_0^{\pi}g\left(\frac{u_k-u_j}{2},\frac{u_k+u_j}{2}\right)\Delta_n\left(t,\frac{u_k+u_j}{2}\right)dt\\
&=[f(u_j)+f(u_k)]\int_0^{\pi}H_n\left(\frac{u_k+u_j}{2}-t\right)H_n\left(\frac{u_k+u_j}{2}+t\right)dt\\
&=\pi[f(u_j)+f(u_k)]H_n(u_k+u_j)\\
&=0.
\end{aligned}$$

Utilizing Lemma 5 in Moulines & Soulier (1999), we have

$$\begin{aligned}
|\mathrm{E}[\omega_n(u_j)\omega_n(u_k)]| &= \frac{1}{2\pi n}\left|\int_0^{\pi} g\left(t, \frac{u_k - u_j}{2}\right)\Delta_n\left(t, \frac{u_k + u_j}{2}\right)dt\right| \\
&= \frac{1}{2\pi n}\left|\int_0^{\pi}\left[g\left(t, \frac{u_k - u_j}{2}\right) - g\left(\frac{u_k - u_j}{2}, \frac{u_k + u_j}{2}\right)\right]\Delta_n\left(t, \frac{u_k + u_j}{2}\right)dt\right| \\
&\leqslant \frac{1}{2\pi n}\int_0^{\pi}\left|\left[g\left(\frac{u_k - u_j}{2}, t\right) - g\left(\frac{u_k - u_j}{2}, \frac{u_k + u_j}{2}\right)\right]\Delta_n\left(\frac{u_k + u_j}{2}, t\right)\right| dt \\
&\leqslant Cn^{2d_0}\log(k)[k^{-2d_0} + j^{-2d_0}]/k \\
&\leqslant Cn^{2d_0} j^{-2d_0}\log(k)/k \qquad (4.31)
\end{aligned}$$

Through the definition of $f(\cdot)$ in Theorem 1.5.1, it is straightforward to see $f(u) = \Theta(u^{-2d_0})$. With the boundary we obtained in (4.31), we have $\mathrm{Cov}(V_{c,j}, V_{c,k}) = O(j^{-d_0}k^{d_0-1} \cdot\log(k))$ for $1 \leqslant j \leqslant k \leqslant \nu$.

As $(V_{c,j}, V_{c,k})$ follows a bivariate normal distribution, the boundary of $\mathrm{Cov}(V_{c,j}^2, V_{c,k}^2)$ comes immediately after utilizing **Lemma VI**. □

Proof of Theorem 1.5.1: The proof of the first two results in Theorem 1.5.1 is the same as that of Theorem 1.2.1, since the **Lemma III** holds for both short and long memory processes. Therefore, we only need to consider the covariance result in Theorem 1.5.1.

$$\begin{aligned}
\mathrm{Cov}(\sqrt{n}\tilde{\beta}_{k_1}, \sqrt{n}\tilde{\beta}_{k_2}) &= \frac{8\pi}{n}\sum_{j=1}^{\nu}\cos(k_1 u_j)\cos(k_2 u_j)\mathrm{Var}\left[\frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)}\right] \\
&\quad + \frac{8\pi}{n}\sum\sum_{j_1 \neq j_2}\cos(k_1 u_{j_1})\cos(k_2 u_{j_2})\mathrm{Cov}\left(\frac{I_X(u_{j_1})}{f(u_{j_1}, \boldsymbol{\theta}_0)}, \frac{I_X(u_{j_2})}{f(u_{j_2}, \boldsymbol{\theta}_0)}\right) \\
&:= I_{1n} + I_{2n}
\end{aligned}$$

Consider the first term I_{1n}, following the Bartlett approximation for a standardized periodogram $I_X(u_j)/f(u_j, \boldsymbol{\theta}_0)$ of a linear process in Giraitis et al. (2012, Theorem 5.3.2), we have

$$\frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)} = \frac{I_\zeta(u_j)}{f_\zeta(u_j, \boldsymbol{\theta}_0)} + r_{n,j}, \qquad (4.32)$$

where $r_{n,j}$ is a remainder term, details can be found in Theorem 5.3.2. Then

$$\mathrm{Var}\left[\frac{I_X(u_j)}{f(u_j,\boldsymbol{\theta}_0)}\right] = \mathrm{Var}\left[\frac{I_\zeta(u_j)}{f_\zeta(u_j,\boldsymbol{\theta}_0)}\right] + \mathrm{Var}[r_{n,j}] + 2\mathrm{Cov}\left(\frac{I_\zeta(u_j)}{f_\zeta(u_j,\boldsymbol{\theta}_0)}, r_{n,j}\right), \tag{4.33}$$

I_{1n} thus can be written as $I_{1n} = I_{11n} + I_{12n} + I_{13n}$ in terms of (4.33). From proof arguments in Giraitis et al. (2012, (5.3.25)),

$$\mathrm{Var}\left[\frac{I_\zeta(u_j)}{f_\zeta(u_j,\boldsymbol{\theta}_0)}\right] = 1 + O(n^{-1}), \tag{4.34}$$

Then

$$I_{11n} = \frac{8\pi}{n}\sum_{j=1}^{\nu}\cos(k_1u_j)\cos(k_2u_j)\left[1 + O(n^{-1})\right].$$

Following the same proof arguments in (4.6), it can be shown $I_{11n} \to 0$ as $n \to \infty$.

About I_{12n}, from Giraitis et al. (2012, Theorem 5.3.2 (iii)),

$$\mathrm{E}|r_{n,j}|^2 \leqslant C\frac{\log j}{j}, \qquad \mathrm{E}|r_{n,j}| \leqslant C\left(\frac{\log j}{j}\right)^{1/2}.$$

Thus we have $\mathrm{Var}[r_{n,j}] \leqslant C\log(j)/j$. Then

$$|I_{12n}| \leqslant C\sum_{j=1}^{\nu}\frac{\log j}{j} \leqslant C\frac{\log^2(n)}{n} \to 0, \qquad \text{as } n \to \infty.$$

And when $n \to \infty$, $I_{13n} \to 0$ follows immediately by utilizing Cauchy-Schwartz inequality. We showed that $I_{1n} \to 0$.

As to I_{2n}, we have

$$\begin{aligned}|I_{2n}| &\leqslant \frac{C}{n}\sum_{j_1=1}^{\nu}\sum_{j_2>j_1}^{\nu}\left|\mathrm{Cov}\left(\frac{I_X(u_{j_1})}{f(u_{j_1},\boldsymbol{\theta}_0)},\frac{I_X(u_{j_2})}{f(u_{j_2},\boldsymbol{\theta}_0)}\right)\right| \\ &\leqslant \frac{C}{n}\sum_{j_1=1}^{\nu}\frac{1}{j_1^{2d_0}}\sum_{j_2>j_1}^{\nu}\frac{\log^2 j_2}{j_2^{2-2d_0}} \\ &\leqslant \frac{C\log^2 n}{n}\sum_{j_1=1}^{\nu}\frac{1}{j_1^{2d_0}} \\ &\leqslant C\frac{\log^2 n}{n^{2d_0}},\end{aligned}$$

The 2nd inequality is ensured by **Lemma VII**. We thus have $I_{2n} \to 0$. It concludes the proof of the third result. The last result can be proved in the same way. □

Proof of Theorem 1.5.2: The outline of the proof is the same as that of Theorem 1.3.1. However, because we are now under conditions specific to a long memory process, there are some extra details we have to take care of. Therefore we choose to go through all of them.

Firstly, utilizing Taylor series expansion on $f^{-1}(u,\hat{\boldsymbol{\theta}}_W) - f^{-1}(u,\boldsymbol{\theta}_0)$. Following Theorem 2.1 in Dahlhaus (1989), we have, $\left\|\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0\right\| = O_p(n^{-1/2})$. In addition to the assumption that $\frac{\partial^2}{\partial\theta_\omega\partial\theta_\upsilon}f^{-1}(u,\boldsymbol{\theta})$ is continuous on all $(u,\boldsymbol{\theta})$, the Taylor expansion can be written in the same way as in (4.9),

$$f^{-1}(u,\hat{\boldsymbol{\theta}}_W) - f^{-1}(u,\boldsymbol{\theta}_0) = \boldsymbol{\nabla}_{\boldsymbol{\theta}}^T f^{-1}(u,\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0) + O_p(n^{-1}).$$

Then $\sqrt{n}\hat{\beta}_k$ can be further written

$$\begin{aligned}\sqrt{n}\hat{\beta}_k &= \sqrt{n}\tilde{\beta}_k - \frac{1}{\sqrt{2\pi}}\frac{2\pi}{n}\sum_j \cos(ku_j)\nabla_{\boldsymbol{\theta}}^T \log f(u_j,\boldsymbol{\theta}_0) \times \sqrt{n}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0)\\ &- \frac{1}{\sqrt{2\pi}}\frac{2\pi}{n}\sum_j \cos(ku_j)\nabla_{\boldsymbol{\theta}}^T f(u_j,\boldsymbol{\theta}_0)(I_X(u_j) - f(u_j,\boldsymbol{\theta}_0)) \times \sqrt{n}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0)\\ &+ O_p(n^{-1})\sqrt{\frac{2\pi}{n}}\sum_j \cos(ku_j)(I_X(u_j) - f(u_j,\boldsymbol{\theta}_0))\\ &+ O_p(n^{-1})\sqrt{\frac{2\pi}{n}}\sum_j \cos(ku_j) f(u_j,\boldsymbol{\theta}_0).\end{aligned}$$

It is the same as what we've done in a short memory process (4.10). Then we have the same claims as before,

$$\frac{1}{\sqrt{2\pi}}\frac{2\pi}{n}\sum_j \cos(ku_j)\nabla_{\boldsymbol{\theta}}^T f(u_j,\boldsymbol{\theta}_0)(I_X(u_j) - f(u_j,\boldsymbol{\theta}_0)) \times \sqrt{n}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0) = o_p(1), \tag{4.35}$$

$$\sqrt{\frac{2\pi}{n}}\sum_j \cos(ku_j)(I_X(u_j) - f(u_j,\boldsymbol{\theta}_0)) = o_p(n), \tag{4.36}$$

$$\sqrt{\frac{2\pi}{n}}\sum_j \cos(ku_j) f(u_j,\boldsymbol{\theta}_0) = o_p(n). \tag{4.37}$$

We firstly consider (4.35), it is true if

$$\sum_{j=1}^{\nu} \cos(ku_j)\frac{\partial}{\partial\theta_\omega}\log f(u_j,\boldsymbol{\theta}_0)\left(\frac{I_X(u_j)}{f(u_j,\boldsymbol{\theta}_0)} - 1\right) = o_p(n). \tag{4.38}$$

Before to show this result, we have the following claim. under the assumption of Theorem 1.5.1,

$$\mathrm{E}\left[\left(\frac{I_X(u_j)}{f(u_j,\boldsymbol{\theta}_0)} - 1\right)\left(\frac{I_X(u_h)}{f(u_h,\boldsymbol{\theta}_0)} - 1\right)\right] = O(j^{-1}h^{-1}\log(j)\log(h)) \tag{4.39}$$

uniformly for $\log^2 n \leq j < h \leq \nu$, $\nu = \frac{n}{2} - 1$. Also

$$\max_{1\leq j\leq n} \mathrm{E}\left[\frac{I_X(u_j)}{f(u_j,\boldsymbol{\theta}_0)} - 1\right]^2 < \infty. \tag{4.40}$$

The proof of (4.39) and (4.40) can be obtained by following the proof arguments of Lemma 4 and Lemma 5 in Chen & Deo (2004). We do not include their proof because it is tedious but does not have any technical hurdles.

(4.38) is true if

$$\mathrm{E}\left[\sum_{j,h=1}^{\nu}\cos(ku_j)\cos(ku_h)\frac{\partial}{\partial\theta_\omega}\log f(u_j,\boldsymbol{\theta}_0)\frac{\partial}{\partial\theta_\omega}\log f(u_h,\boldsymbol{\theta}_0)\left(\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}-1\right)\left(\frac{I_{X,h}}{f_{\boldsymbol{\theta}_0,h}}-1\right)\right] \tag{4.41}$$
$$= o(n^2).$$

If $j = h$, LHS of (4.41) is

$$\sum_{j=1}^{\nu}\cos^2(ku_j)\left[\frac{\partial}{\partial\theta_\omega}\log f(u_j,\boldsymbol{\theta}_0)\right]^2\mathrm{E}\left[\frac{I_X(u_j)}{f(u_j,\boldsymbol{\theta}_0)}-1\right]^2.$$

Recall $\frac{\partial}{\partial\theta_\omega}\log f(u,\boldsymbol{\theta}_0) = O(|u|^{-\delta})$, for any $\delta > 0$. This is from Assumptions **A2-A6** in Dahlhaus (1989), more details can be found in Fox & Taqqu (1986) and Feller (1971, p.277). From those assumptions, in fact we have

$$\frac{\partial}{\partial\theta_\omega}\log f(u,\boldsymbol{\theta}_0) = -f(u,\boldsymbol{\theta}_0)\cdot\frac{\partial}{\partial\theta_\omega}f^{-1}(u,\boldsymbol{\theta}_0) = O(|u|^{-\delta}). \tag{4.42}$$

Utilizing (4.40), we have

$$\begin{aligned}
&\sum_{j=1}^{\nu}\cos^2(ku_j)\left[\frac{\partial}{\partial\theta_\omega}\log f(u_j,\boldsymbol{\theta}_0)\right]^2\mathrm{E}\left[\frac{I_X(u_j)}{f(u_j,\boldsymbol{\theta}_0)}-1\right]^2\\
&\leqslant \max_{1\leqslant j\leqslant n}\mathrm{E}\left[\frac{I_X(u_j)}{f(u_j,\boldsymbol{\theta}_0)}-1\right]^2\sum_{j=1}^{\nu}\left(\frac{2\pi j}{n}\right)^{-\delta}\\
&\leqslant Cn^{\delta}\sum_{j=1}^{\nu}\frac{1}{j^{\delta}} = O(n) = o(n^2).
\end{aligned}$$

If $j \neq h$, LHS of (4.41) can be written as

$$\left(\sum_{j=1}^{\log^2 n}\sum_{h\neq j} + 2\sum_{j=\log^2 n+1}^{\nu}\sum_{h=1}^{\log^2 n} + \sum_{j=\log^2 n+1}^{\nu}\sum_{h\neq j}\right)\cos(ku_j)\cos(ku_h)\frac{\partial}{\partial\theta_\omega}\log f(u_j,\boldsymbol{\theta}_0)\frac{\partial}{\partial\theta_\omega}\log f(u_h,\boldsymbol{\theta}_0)$$
$$\times \mathrm{E}\left[\left(\frac{I_X(u_j)}{f(u_j,\boldsymbol{\theta}_0)}-1\right)\left(\frac{I_X(u_h)}{f(u_h,\boldsymbol{\theta}_0)}-1\right)\right]$$
$$:= K_{1n} + 2K_{2n} + K_{3n}.$$

About K_{1n}, utilizing (4.40) and Cauchy-Schwartz inequality, it can be bounded in the following way,

$$\left|\sum_{j=1}^{\log^2 n}\sum_{h\neq j}\cos(ku_j)\cos(ku_h)\frac{\partial}{\partial\theta_\omega}\log f(u_j,\boldsymbol{\theta}_0)\frac{\partial}{\partial\theta_\omega}\log f(u_h,\boldsymbol{\theta}_0)\mathrm{E}\left[\left(\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}-1\right)\left(\frac{I_{X,h}}{f_{\boldsymbol{\theta}_0,h}}-1\right)\right]\right|$$
$$\leqslant \sum_{j=1}^{\log^2 n}\sum_{h\neq j}\left|\frac{\partial}{\partial\theta_\omega}\log f(u_j,\boldsymbol{\theta}_0)\right|\left|\frac{\partial}{\partial\theta_\omega}\log f(u_h,\boldsymbol{\theta}_0)\right|\sqrt{\mathrm{E}\left[\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}-1\right]^2}\sqrt{\mathrm{E}\left[\frac{I_{X,h}}{f_{\boldsymbol{\theta}_0,h}}-1\right]^2}$$
$$\leqslant C\sum_{j=1}^{\log^2 n}\sum_{h\neq j}\left|\frac{2\pi j}{n}\right|^{-\delta}\left|\frac{2\pi h}{n}\right|^{-\delta}$$
$$\leqslant Cn^{2\delta}\left[(\log^2 n)^{1-\delta}\right]^2 = o(n^2).$$

About K_{2n}, it can be bounded in the same way,

$$\sum_{j=1}^{\log^2 n}\sum_{h=\log^2 n+1}^{\nu}\left|\frac{2\pi j}{n}\right|^{-\delta}\left|\frac{2\pi h}{n}\right|^{-\delta} \leqslant Cn^{2\delta}(\log^2 n)^{1-\delta}n^{1-\delta} = o(n^2).$$

Last but not the least, we also have $K_{3n} = o(n^2)$ by utilizing (4.39),

$$\left|\sum_{j=\log^2 n+1}^{\nu}\sum_{h>j}\cos(ku_j)\cos(ku_h)\frac{\partial}{\partial\theta_\omega}\log f(u_j,\boldsymbol{\theta}_0)\frac{\partial}{\partial\theta_\omega}\log f(u_h,\boldsymbol{\theta}_0)\mathrm{E}\left[\left(\frac{I_{X,j}}{f_{\boldsymbol{\theta}_0,j}}-1\right)\left(\frac{I_{X,h}}{f_{\boldsymbol{\theta}_0,h}}-1\right)\right]\right|$$
$$\leqslant \sum_{j=\log^2 n+1}^{\nu}\sum_{h>j}\left|\frac{2\pi j}{n}\right|^{-\delta}\left|\frac{2\pi h}{n}\right|^{-\delta}\frac{\log(j)}{j}\frac{\log(h)}{h}$$
$$\leqslant Cn^{2\delta}\sum_{j=\log^2 n+1}^{\nu}\frac{\log(j)}{j^{1+\delta}}\sum_{h>j}\frac{\log(h)}{h^{1+\delta}} = o(n^2).$$

This concludes the proof of (4.41), indicating (4.35).

Then we consider (4.36), it is true if

$$\sum_{j=1}^{\nu} \cos(ku_j)(I_X(u_j) - f(u_j, \boldsymbol{\theta}_0)) = o_p(n^{3/2}).$$

And this is true if we can show

$$\mathrm{E}\left|\sum_{j=1}^{\nu} \cos(ku_j) f(u_j, \boldsymbol{\theta}_0)\left(\frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)} - 1\right)\right| = o(n^{3/2}). \tag{4.43}$$

From the assumption **A2** in Dahlhaus (1989), we have

$$f(u_j, \boldsymbol{\theta}_0) = O(|u_j|^{-2d_0-\delta}).$$

And from Theorem 5.3.2 and its proof in Giraitis et al. (2012, p.128), we have

$$\begin{aligned}
&\mathrm{E}\left|\frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)} - 1\right| \leqslant \mathrm{E}\left|\frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)} - \frac{I_\zeta(u_j)}{f_\zeta(u_j, \boldsymbol{\theta}_0)}\right| + \mathrm{E}\left|\frac{I_\zeta(u_j)}{f_\zeta(u_j, \boldsymbol{\theta}_0)} - 1\right|, \\
&\mathrm{E}\left|\frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)} - \frac{I_\zeta(u_j)}{f_\zeta(u_j, \boldsymbol{\theta}_0)}\right| \leqslant C\left(\frac{\log j}{j}\right)^{1/2}, \\
&\mathrm{E}\left|\frac{I_\zeta(u_j)}{f_\zeta(u_j, \boldsymbol{\theta}_0)} - 1\right| = \mathrm{E}\left|\frac{I_\zeta(u_j)}{f_\zeta(u_j, \boldsymbol{\theta}_0)} - \mathrm{E}\left[\frac{I_\zeta(u_j)}{f_\zeta(u_j, \boldsymbol{\theta}_0)}\right]\right| \\
&\leqslant \sqrt{\mathrm{E}\left[\frac{I_\zeta(u_j)}{f_\zeta(u_j, \boldsymbol{\theta}_0)} - \mathrm{E}\left[\frac{I_\zeta(u_j)}{f_\zeta(u_j, \boldsymbol{\theta}_0)}\right]\right]^2} \\
&= \sqrt{1 + O(n^{-1})}.
\end{aligned}$$

Therefore, LHS of (4.43) can be bounded as the following,

$$\begin{aligned}
\sum_{j=1}^{\nu} f(u_j, \boldsymbol{\theta}_0)\mathrm{E}\left|\frac{I_X(u_j)}{f(u_j, \boldsymbol{\theta}_0)} - 1\right| &\leqslant C^* \sum_{j=1}^{\nu}\left(\frac{2\pi j}{n}\right)^{-2d_0-\delta}\left\{C\left(\frac{\log j}{j}\right)^{1/2} + \sqrt{1 + O(n^{-1})}\right\} \\
&\leqslant Cn^{2d_0+\delta}\left\{\sum_{j=1}^{\nu}\frac{1}{j^{2d_0+\delta}} \cdot \frac{\log^{1/2} j}{j^{1/2}} + \sum_{j=1}^{\nu}\frac{1}{j^{2d_0+\delta}}\right\} \\
&= O(n\log^{1/2} n) = o(n^{3/2}).
\end{aligned}$$

Note the above proof also covers the result of (4.37). Therefore we showed both (4.36) and (4.37).

Now $\sqrt{n}\hat{\beta}_k$ can be written as

$$\sqrt{n}\hat{\beta}_k = \sqrt{n}\tilde{\beta}_k - \frac{1}{\sqrt{2\pi}}\frac{2\pi}{n}\sum_j \cos(ku_j)\nabla^T_{\boldsymbol{\theta}}\log f(u_j,\boldsymbol{\theta}_0)\times\sqrt{n}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0) + o_p(1).$$

Consider $\sqrt{n}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0)$. Following the proof arguments in Theorem 2.1 in Dahlhaus (1989), about the Whittle likelihood function (1.7), we have

$$\nabla\mathcal{L}^W_n(\hat{\boldsymbol{\theta}}_W) - \nabla\mathcal{L}^W_n(\boldsymbol{\theta}_0) = \nabla^2\mathcal{L}^W_n(\hat{\boldsymbol{\theta}}_W)(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0).$$

From smoothness conditions **D0** and **D1**, we have

$$\begin{aligned} \nabla^2\mathcal{L}^W_n(\hat{\boldsymbol{\theta}}_W) - \nabla^2\mathcal{L}^W_n(\boldsymbol{\theta}_0) &= o_p(1), \\ \nabla^2\mathcal{L}^W_n(\boldsymbol{\theta}_0) &\to_p \Gamma(\boldsymbol{\theta}_0), \\ \sqrt{n}\mathcal{L}^W_n(\hat{\boldsymbol{\theta}}_W) &= o_p(1), \end{aligned}$$

where

$$\Gamma(\boldsymbol{\theta}_0) = \int_{-\pi}^{\pi}\nabla_{\boldsymbol{\theta}}\log f(u,\boldsymbol{\theta}_0)\nabla^T_{\boldsymbol{\theta}}\log f(u,\boldsymbol{\theta}_0)du.$$

This indicates the following result

$$o_p(1) - \nabla\mathcal{L}^W_n(\boldsymbol{\theta}_0) = (\Gamma(\boldsymbol{\theta}_0) + o_p(1))\sqrt{n}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0).$$

Now consider $D(\boldsymbol{\theta}, f_{\boldsymbol{\theta}_0})$ in (1.8). Under the current conditions, it is obvious to see $\nabla D(\boldsymbol{\theta}, f_{\boldsymbol{\theta}_0})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0$. Therefore the above equality can be written as

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0) = -\Gamma^{-1}(\boldsymbol{\theta}_0)\sqrt{n}\{\nabla\mathcal{L}^W_n(\boldsymbol{\theta}_0) - \nabla D(\boldsymbol{\theta}, f_{\boldsymbol{\theta}_0})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\} + o_p(1). \qquad (4.44)$$

As Dahlhaus (1989) claims, we have

$$\frac{\partial}{\partial \theta_\omega}\int_{-\pi}^{\pi} \log f(u,\boldsymbol{\theta}) + \frac{I_X(u)}{f(u,\boldsymbol{\theta})}du\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} - \frac{\partial}{\partial \theta_\omega}\frac{2\pi}{n}\sum_j \left\{\log f(u_j,\boldsymbol{\theta}) + \frac{I_X(u_j)}{f(u_j,\boldsymbol{\theta})}\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = o_p(n^{-1/2}).$$

Moreover, according to those smoothness conditions, it is straightforward to see

$$\frac{\partial}{\partial \theta_\omega}\int_{-\pi}^{\pi} \log f(u,\boldsymbol{\theta}) + \frac{f(u,\boldsymbol{\theta}_0)}{f(u,\boldsymbol{\theta})}du\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} - \frac{\partial}{\partial \theta_\omega}\frac{2\pi}{n}\sum_j \left\{\log f(u_j,\boldsymbol{\theta}) + \frac{f(u,\boldsymbol{\theta}_0)}{f(u_j,\boldsymbol{\theta})}\right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = o(n^{-1/2}).$$

Then (4.44) can be further written as

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_0) = -\Gamma(\boldsymbol{\theta}_0)^{-1}\sqrt{n}\frac{2\pi}{n}\sum_j \boldsymbol{\nabla}_{\boldsymbol{\theta}} f^{-1}(u_j,\boldsymbol{\theta}_0)\left(I_X(u_j) - f(u_j,\boldsymbol{\theta}_0)\right) + o_p(1).$$

This is the same as what we have in the proof of Theorem 1.3.1 (4.18). In other words, for a long memory process, our $\hat{\beta}_k$ can also be written as

$$\begin{aligned}\sqrt{n}\hat{\beta}_k &= \sqrt{n}\tilde{\beta}_k + \left[C_k^{\dagger}(\boldsymbol{\theta}_0)\right]^T\left\{-\Gamma(\boldsymbol{\theta}_0)^{-1}\sqrt{n}\frac{2\pi}{n}\sum_j \boldsymbol{\nabla}_{\boldsymbol{\theta}}\log f(u_j,\boldsymbol{\theta}_0)\left(\frac{I_X(u_j)}{f(u_j,\boldsymbol{\theta}_0)} - 1\right)\right\} + o_p(1)\\ &:= A_{1k} + A_{2k} + o_p(1).\end{aligned}$$

Note the sum $C_k^{\dagger}(\boldsymbol{\theta}_0)$ is a discrete analog of the integral $C_k(\boldsymbol{\theta}_0)$.

About A_{1k}, A_{2k}, we have the same claims as before, as $n \to \infty$,

$$A_{1k} \longrightarrow_d N(0, 2\pi), \tag{4.45}$$

$$A_{2k} \longrightarrow_d N(0, C_k^T(\boldsymbol{\theta}_0)\left\{4\pi\Gamma(\boldsymbol{\theta}_0)^{-1}\right\}C_k(\boldsymbol{\theta}_0)), \tag{4.46}$$

$$\mathrm{Cov}(A_{1k_1}, A_{1k_2}) \longrightarrow 0, \quad k_1 \neq k_2, \tag{4.47}$$

$$\mathrm{Cov}(A_{2k_1}, A_{2k_2}) \longrightarrow C_{k_1}^T(\boldsymbol{\theta}_0)\left\{4\pi\Gamma(\boldsymbol{\theta}_0)^{-1}\right\}C_{k_2}(\boldsymbol{\theta}_0), \tag{4.48}$$

$$\mathrm{Cov}(A_{1k_1}, A_{2k_2}) \longrightarrow -C_{k_1}^T(\boldsymbol{\theta}_0)\left\{4\pi\Gamma(\boldsymbol{\theta}_0)^{-1}\right\}C_{k_2}(\boldsymbol{\theta}_0). \tag{4.49}$$

Note (4.45) and (4.47) are from Theorem 1.5.1. And (4.46) can be obtained from Theorem

2.1 in Dahlhaus (1989). We only need to calculate the rest of two asymptotic covariance results.

This time we mainly consider (4.49),

$$\mathrm{Cov}(A_{1k_1}, A_{2k_2}) = \left\{ \frac{1}{\sqrt{2\pi}} \frac{2\pi}{n} \sum_j \cos(k_1 u_j) \boldsymbol{\nabla}_{\boldsymbol{\theta}}^T \log f(u_j, \boldsymbol{\theta}_0) \right\} \left\{ -2\pi \Gamma(\boldsymbol{\theta}_0)^{-1} \right\} \left\{ \frac{1}{\sqrt{2\pi}} \frac{2\pi}{n} \right.$$
$$\left. \times \sum_h \sum_l \cos(k_2 u_h) \boldsymbol{\nabla}_{\boldsymbol{\theta}} \log f(u_l, \boldsymbol{\theta}_0) \mathrm{Cov}\left(\frac{I_X(u_h)}{f(u_h, \boldsymbol{\theta}_0)}, \frac{I_X(u_l)}{f(u_l, \boldsymbol{\theta}_0)} \right) \right\}.$$

The critical step is required to show the following result,

$$\begin{aligned} &\frac{2\pi}{n} \sum_h \sum_l \cos(k u_h) \frac{\partial}{\partial \theta_\omega} \log f(u_l, \boldsymbol{\theta}_0) \mathrm{Cov}\left(\frac{I_X(u_h)}{f(u_h, \boldsymbol{\theta}_0)}, \frac{I_X(u_l)}{f(u_l, \boldsymbol{\theta}_0)} \right) \\ &= 2 \int_{-\pi}^{\pi} \cos(k u) \frac{\partial}{\partial \theta_\omega} \log f(u, \boldsymbol{\theta}_0) du + o(1). \end{aligned} \tag{4.50}$$

The proof idea is almost the same as what we had before,

$$\begin{aligned} \mathrm{LHS} &= 4 \cdot \frac{2\pi}{n} \sum_{h=1}^{\nu} \sum_{l=1}^{\nu} \cos(k u_h) \frac{\partial}{\partial \theta_\omega} \log f(u_l, \boldsymbol{\theta}_0) \mathrm{Cov}\left(\frac{I_X(u_h)}{f(u_h, \boldsymbol{\theta}_0)}, \frac{I_X(u_l)}{f(u_l, \boldsymbol{\theta}_0)} \right) \\ &= 4 \cdot \frac{2\pi}{n} \sum_{h=1}^{\nu} \cos(k u_h) \frac{\partial}{\partial \theta_\omega} \log f(u_h, \boldsymbol{\theta}_0) \mathrm{Var}\left[\frac{I_X(u_h)}{f(u_h, \boldsymbol{\theta}_0)} \right] \\ &+ 4 \cdot \frac{2\pi}{n} \sum_{h \neq l} \sum \cos(k u_h) \frac{\partial}{\partial \theta_\omega} \log f(u_l, \boldsymbol{\theta}_0) \mathrm{Cov}\left(\frac{I_X(u_h)}{f(u_h, \boldsymbol{\theta}_0)}, \frac{I_X(u_l)}{f(u_l, \boldsymbol{\theta}_0)} \right) \\ &:= L_{1n} + L_{2n}. \end{aligned}$$

We firstly consider L_{1n}. Utilizing Bartlett approximation (4.32) again, then $\mathrm{Var}\left[\frac{I_X(u_h)}{f(u_h, \boldsymbol{\theta}_0)} \right]$ can be written as (4.33). Therefore, L_{1n} can also be written as $L_{1n} = L_{11n} + L_{12n} + L_{13n}$. L_{11n} is the term containing $\mathrm{Var}\left[\frac{I_\zeta(u_h)}{f_\zeta(u_h, \boldsymbol{\theta}_0)} \right]$. From Giraitis et al. (2012, 5.3.25), we have (4.34), thus L_{11n} ends up with

$$L_{11n} = 4 \cdot \frac{2\pi}{n} \sum_{h=1}^{\nu} \cos(k u_h) \frac{\partial}{\partial \theta_\omega} \log f(u_h, \boldsymbol{\theta}_0) [1 + O(n^{-1})].$$

Following **Lemma IV**, it can be easily shown that

$$\frac{2\pi}{n}\sum_{h}\cos(ku_h)\frac{\partial}{\partial\theta_\omega}\log f(u_h,\boldsymbol{\theta}_0) = \int_{-\pi}^{\pi}\cos(ku)\frac{\partial}{\partial\theta_\omega}\log f(u,\boldsymbol{\theta}_0)du + O(n^{-1+\delta})$$

Thus L_{11n} is

$$L_{11n} = 2\int_{-\pi}^{\pi}\cos(ku)\frac{\partial}{\partial\theta_\omega}\log f(u,\boldsymbol{\theta}_0)du + O(n^{-1+\delta}).$$

About L_{12n}, it is the term containing $\mathrm{Var}[r_{n,h}]$, which can be bounded as $\mathrm{Var}[r_{n,h}] \leqslant C\log(h)/h$. Recall $\frac{\partial}{\partial\theta_\omega}\log f(u_h,\boldsymbol{\theta}_0)$ can be bounded as (4.42), we thus have

$$L_{12n} = 4\cdot\frac{2\pi}{n}\sum_{h=1}^{\nu}\cos(ku_h)\frac{\partial}{\partial\theta_\omega}\log f(u_h,\boldsymbol{\theta}_0)\mathrm{Var}[r_{n,h}] = O\left(\frac{1}{n}\sum_{h=1}^{\nu}\left(\frac{n}{h}\right)^{\delta}\frac{\log h}{h}\right) = O\left(\frac{\log(n)}{n^{1-\delta}}\right),$$

which shows $L_{12n} \to 0$, as $n \to \infty$. Consider L_{13n} containing $\mathrm{Cov}\left(\frac{I_{\zeta,h}}{f_{\boldsymbol{\theta}_0,h}}, r_{n,h}\right)$, which is bounded by $\sqrt{\log(h)/h}$ through utilizing Cauchy-Schwartz inequality, Therefore

$$\begin{aligned} L_{13n} &= 4\cdot\frac{2\pi}{n}\sum_{h=1}^{\nu}\cos(ku_h)\frac{\partial}{\partial\theta_\omega}\log f(u_h,\boldsymbol{\theta}_0)\mathrm{Cov}\left(\frac{I_{\zeta,h}}{f_{\boldsymbol{\theta}_0,h}}, r_{n,h}\right) = O\left(\frac{1}{n}\sum_{h=1}^{\nu}\left(\frac{n}{h}\right)^{\delta}\sqrt{\frac{\log h}{h}}\right) \\ &= O\left(\sqrt{\frac{\log(n)}{n}}\right). \end{aligned}$$

We have $L_{13n} \to 0$, as $n \to \infty$.

Then we consider L_{2n}, utilizing **Lemma VII**, we have

$$\begin{aligned} |L_{2n}| &\leqslant \frac{C}{n}\sum_{h=1}^{\nu}\sum_{l>h}^{\nu}\left|\frac{\partial}{\partial\theta_\omega}\log f(u_l,\boldsymbol{\theta}_0)\right|\left|\mathrm{Cov}\left(\frac{I_X(u_h)}{f(u_h,\boldsymbol{\theta}_0)}, \frac{I_X(u_l)}{f(u_l,\boldsymbol{\theta}_0)}\right)\right| \\ &= O\left(\frac{1}{n}\sum_{h=1}^{\nu}\sum_{l>h}^{\nu}\left(\frac{n}{l}\right)^{\delta}\left(\frac{l}{h}\right)^{2d_0}\left(\frac{\log l}{l}\right)^{2}\right) \\ &= O\left(\frac{1}{n^{1-\delta}}\sum_{h=1}^{\nu}\frac{1}{h^{2d_0}}\sum_{l=1}^{\nu}\frac{\log^2 l}{l^{2+\delta-2d_0}}\right) \\ &= O\left(\frac{\log^2 n}{n^{2d_0-\delta}}\right), \end{aligned}$$

thus $L_{2n} \to 0$, as $n \to \infty$. This concludes the proof of (4.50), indicating the asymptotic covariance result (4.49).

About (4.48), it can be showed in the same way. Since we did it multiple times in the preceding proofs, we choose to skip details.

Then through the continuous mapping theorem, we ends up with the same expressions as in a short memory process, as $n \to \infty$,

$$\sum_k \alpha_k \sqrt{n} \hat{\beta}_k = \sum_k \alpha_k (A_{1k} + A_{2k}) + o_p(1) \longrightarrow_d N(0, V_{\boldsymbol{\theta}_0}(\boldsymbol{\alpha})),$$

where

$$V_{\boldsymbol{\theta}_0}(\boldsymbol{\alpha}) = \sum_{k_1} \sum_{k_2} \alpha_{k_1} \alpha_{k_2} \left\{ \text{Cov}_\infty(A_{1k_1}, A_{1k_2}) + \text{Cov}_\infty(A_{2k_1}, A_{2k_2}) + 2\text{Cov}_\infty(A_{1k_1}, A_{2k_2}) \right\}$$

$\text{Cov}_\infty(\cdot, \cdot)$ denotes the asymptotic covariance values, they are represented in (4.45)-(4.49).

The final result comes immediately after utilizing Cramér-Wold device, and the formula

$$\Sigma_{k,l}(\boldsymbol{\theta}_0) = \frac{1}{2} (V_{\boldsymbol{\theta}_0}(e_k + e_l) - V_{\boldsymbol{\theta}_0}(e_k) - V_{\boldsymbol{\theta}_0}(e_l)),$$

with e_k denotes the kth unit vector. □

LIST OF REFERENCES

Beran, J. (1992), 'A goodness-of-fit test for time series with long range dependence', *Journal of the Royal Statistical Society. Series B (Methodological)* **54**(3), 749–760.

Box, G. E. P. & Pierce, D. A. (1970), 'Distribution of residual autocorrelations in autoregressive-integrated moving average time series models', *Journal of the American Statistical Association* **65**(332), 1509–1526.

Brockwell, P. & Davis, R. (1991), *Time Series: Theory and Methods*, Springer Series in Statistics, Springer.

Chen, W. W. & Deo, R. S. (2004), 'A generalized portmanteau goodness-of-fit test for time series models', *Econometric Theory* **20**(2), 382–416.

Dahlhaus, R. (1989), 'Efficient Parameter Estimation for Self-Similar Processes', *The Annals of Statistics* **17**(4), 1749 – 1766.

Dahlhaus, R. (2006), 'Correction: Efficient parameter estimation for self-similar processes', *The Annals of Statistics* **34**(2), 1045–1047.

Dahlhaus, R. & Wefelmeyer, W. (1996), 'Asymptotically optimal estimation in misspecified time series models', *The Annals of Statistics* **24**(3), 952 – 974.

Davies, N., Triggs, C. M. & Newbold, P. (1977), 'Significance levels of the box-pierce portmanteau statistic in finite samples', *Biometrika* **64**(3), 517–522.

Deo, R. S. & Chen, W. W. (2000), 'On the integral of the squared periodogram', *Stochastic Processes and their Applications* **85**(1), 159–176.

Feller, W. (1971), *An introduction to probability theory and its applications. Vol. II.*, Second edition, John Wiley & Sons Inc., New York.

Fox, R. & Taqqu, M. S. (1986), 'Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series', *The Annals of Statistics* **14**(2), 517–532.

Fragkeskou, M. & Paparoditis, E. (2016), 'Inference for the fourth-order innovation cumulant in linear time series', *Journal of Time Series Analysis* **37**(2), 240–266.

Fuller, W. (1995), *Introduction to Statistical Time Series*, Wiley Series in Probability and Statistics, Wiley.

Giraitis, L., Koul, H. L. & Surgailis, D. (2012), *Large Sample Inference for Long Memory Processes*, Imperial College Press.

Hannan, E. J. (1973), 'The asymptotic theory of linear time-series models', *Journal of Applied Probability* **10**(1), 130–145.

Hoaoya, Y. (1974), Estimation problems on stationary time-series models, PhD thesis, Yale University.

Hong, Y. (1996), 'Consistent testing for serial correlation of unknown form', *Econometrica* **64**, 837–864.

Hosoya, Y. & Masanobu, T. (1982), 'A central limit theorem for stationary processes and the parameter estimation of linear processes', *The Annals of Statistics* **10**(1), 132–153.

Krogstad, H. E. (1982), 'On the covariance of the periodogram', *Journal of Time Series Analysis* **3**(3), 195–207.

Ljung, G. M. & Box, G. E. P. (1978), 'On a measure of lack of fit in time series models', *Biometrika* **65**(2), 297–303.

McElroy, T. & Holan, S. (2009), 'A local spectral approach for assessing time series model misspecification', *Journal of Multivariate Analysis* **100**(4), 604–621.

Milhoj, A. (1981), 'A test of fit in time series models', *Biometrika* **68**(1), 177–187.

Moulines, E. & Soulier, P. (1999), 'Broadband log-periodogram regression of time series with long-range dependence', *The Annals of Statistics* **27**(4), 1415 – 1439.

Prothero, D. L. & Wallis, K. F. (1976), 'Modelling macroeconomic time series', *Journal of the Royal Statistical Society. Series A (General)* **139**(4), 468–500.

Rao, S. S. & Yang, J. (2021), 'Reconciling the Gaussian and Whittle likelihood with an application to estimation in the frequency domain', *The Annals of Statistics* **49**(5), 2774 – 2802.

Rosenblatt, M. (2012), *Stationary Sequences and Random Fields*, Birkhäuser Boston.

Shumway, R. & Stoffer, D. (2017), *Time Series Analysis and Its Applications: With R Examples*, Springer Texts in Statistics, Springer International Publishing.

Taniguchi, M. & Kakizawa, Y. (2000), *Asymptotic Theory of Statistical Inference for Time Series*, Springer Series in Statistics, Springer New York.

Yao, Q. & Brockwell, P. J. (2006), 'Gaussian maximum likelihood estimation for arma models. i. time series', *Journal of Time Series Analysis* **27**(6), 857–875.

APPENDIX A

LIST OF SUPPLEMENTARY TABLES

Table A1: Rejection rates in percentage under an AR(1) model

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	3.08	5.02	4.04	6.12	4.90	7.80	3.82	5.82	4.32	6.98	5.06	8.02
	TUK	3.04	4.96	4.04	6.12	5.04	7.74	3.98	5.82	4.56	7.10	5.16	8.40
	QS	3.64	5.64	4.52	6.90	6.30	9.68	4.06	6.52	4.74	7.64	5.58	9.26
H_n	BAR	3.30	5.08	3.82	5.82	4.26	6.76	3.62	5.72	4.20	6.54	4.76	7.34
	TUK	3.16	4.90	3.78	5.92	4.46	6.96	3.76	5.78	4.26	6.84	4.88	7.48
	QS	3.52	5.52	4.22	6.44	4.82	7.40	4.02	6.20	4.36	7.12	5.08	8.36
M_n		4.34 at 5%			7.12 at 10%			5.14 at 5%			8.88 at 10%		

Note: Model $X_t - 0.8X_{t-1} = \zeta_t$, $\zeta_t \sim N(0,1)$.

Table A2: Rejection rates in percentage under an ARFIMA(d) model

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	2.62	4.08	3.52	5.28	4.90	7.42	2.64	4.60	3.74	6.00	4.80	8.20
	TUK	2.52	4.00	3.46	5.58	4.96	7.50	2.92	4.78	3.86	6.14	5.10	8.42
	QS	3.22	4.98	4.34	6.78	6.62	9.60	3.30	5.74	4.40	7.06	5.58	9.08
H_n	BAR	2.28	3.76	3.02	4.86	4.54	5.88	2.56	4.42	3.42	5.86	4.22	7.00
	TUK	2.20	3.52	3.20	5.10	3.90	5.88	3.12	5.32	4.14	6.52	4.70	7.86
	QS	2.82	4.46	3.66	5.36	4.10	7.04	2.72	4.54	3.70	5.98	4.44	7.44
M_n		4.70 at 5%			7.58 at 10%			4.50 at 5%			8.18 at 10%		

Note: Model $X_t = (1 - B)^{-0.4}\zeta_t$, $\zeta_t \sim N(0, 1)$.

Table A3: Rejection rates in percentage under an AR(1) model with innovations from t distribution

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	2.90	4.66	3.28	5.42	3.98	6.64	3.42	5.16	3.88	6.08	4.76	8.12
	TUK	2.92	4.50	3.36	5.40	3.98	6.84	3.52	5.28	4.00	6.14	4.90	8.02
	QS	3.20	5.14	3.44	6.08	5.36	8.52	3.60	5.76	4.32	6.90	5.74	8.96
H_n	BAR	3.10	4.76	3.32	5.20	3.34	5.88	3.22	4.92	3.66	5.60	4.48	7.08
	TUK	3.08	4.90	3.24	5.26	3.52	5.94	3.28	5.20	3.68	5.82	4.56	7.32
	QS	3.20	5.14	3.32	5.54	4.00	6.38	3.38	5.42	3.84	6.34	4.86	8.26
M_n		3.80 at 5%			6.26 at 10%			4.60 at 5%			8.48 at 10%		

Note: Model $X_t - 0.8X_{t-1} = \zeta_t$, $\zeta_t \sim t_9$.

Table A4: Rejection rates in percentage under an ARFIMA(d) model with innovations from t distribution

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	2.16	3.50	3.02	5.00	4.28	6.54	2.66	4.10	3.44	5.64	4.32	7.24
	TUK	2.08	3.56	3.08	5.00	4.46	6.78	2.86	4.30	3.64	6.04	4.46	7.52
	QS	2.66	4.40	3.92	6.06	5.58	8.84	3.12	5.12	3.90	6.64	5.10	8.66
H_n	BAR	1.86	3.30	2.64	4.74	3.48	5.60	2.70	4.18	3.24	5.32	4.00	6.78
	TUK	1.96	3.28	2.66	4.90	3.66	5.88	2.86	4.42	3.46	5.86	4.10	7.18
	QS	2.30	4.24	3.52	5.44	4.08	6.58	3.20	5.16	3.72	6.36	4.66	7.60
M_n		3.94 at 5%			7.04 at 10%			4.92 at 5%			8.56 at 10%		

Note: Model $X_t = (1 - B)^{-0.4}\zeta_t$, $\zeta_t \sim t_9$.

Table A5: Rejection rates in percentage under an AR(2) alternative fitting model AR(1)

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	22.48	28.60	22.92	29.04	22.88	29.62	80.18	84.96	76.02	81.64	68.54	75.74
	TUK	21.94	28.16	22.80	28.58	22.40	28.76	79.96	82.18	74.76	81.06	65.66	73.26
	QS	22.44	28.64	22.80	28.96	22.74	29.74	78.20	83.42	70.56	77.68	61.04	69.96
H_n	BAR	23.58	30.22	23.42	29.66	22.42	28.70	80.62	85.46	75.84	81.86	68.24	75.26
	TUK	23.18	29.36	23.22	29.42	21.98	28.22	80.32	85.24	74.90	80.90	65.02	72.66
	QS	23.28	29.90	22.76	28.54	21.12	27.40	78.34	83.96	70.46	77.14	59.60	68.88
M_n		8.84 at 5%			13.78 at 10%			17.78 at 5%			25.96 at 10%		

Note: Model $X_t - 0.8X_{t-1} + 0.15X_{t-2} = \zeta_t$, $\zeta_t \sim N(0,1)$.

Table A6: Rejection rates in percentage under an ARMA(1,1) alternative fitting model ARFIMA(1,d,0)

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	9.50	13.38	8.44	12.58	8.96	13.04	31.84	40.74	28.36	36.56	25.10	34.24
	TUK	7.24	11.28	8.04	12.06	8.80	12.80	31.34	40.54	26.94	35.12	23.36	32.48
	QS	8.74	12.26	8.04	12.02	10.04	14.74	29.20	37.80	25.00	33.62	22.78	31.50
H_n	BAR	12.68	17.04	11.28	15.52	8.92	13.20	33.02	42.28	28.94	37.36	24.82	33.70
	TUK	7.98	12.12	8.20	12.32	8.66	12.96	32.66	41.88	27.70	35.72	23.12	31.88
	QS	11.48	15.70	10.18	14.40	9.06	13.68	30.32	38.70	25.30	33.72	21.78	30.10
M_n		5.42 at 5%			8.76 at 10%			6.44 at 5%			10.38 at 10%		

Note: Model $X_t = 0.8X_{t-1} + \zeta_t + 0.2\zeta_{t-1}$, $\zeta_t \sim N(0,1)$.

Table A7: Rejection rates in percentage under an ARFIMA(d) alternative fitting model ARMA(1,1)

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	6.46	9.38	7.50	11.10	8.90	13.30	37.28	44.78	38.54	46.16	37.02	45.06
	TUK	6.54	9.10	7.50	11.32	8.68	13.14	38.06	45.20	39.14	46.74	36.20	44.42
	QS	7.20	10.22	8.24	12.56	10.44	15.34	39.70	46.54	37.90	45.72	34.36	43.00
H_n	BAR	5.26	7.54	6.22	8.90	6.84	10.34	36.14	43.38	37.28	44.80	35.00	42.88
	TUK	5.32	7.48	6.46	9.12	6.92	10.54	37.16	44.08	37.90	45.46	34.22	42.36
	QS	6.04	8.64	6.68	10.18	7.32	10.78	38.42	45.56	36.72	44.56	32.26	40.10
M_n		5.34 at 5%			8.92 at 10%			11.56 at 5%			17.96 at 10%		

Note: Model $X_t = (1 - B)^{-0.4}\zeta_t$, $\zeta_t \sim N(0, 1)$.

Table A8: Rejection rates in percentage under an ARFIMA(1,d,0) alternative fitting model ARFIMA(d)

n		128						512					
p_n		8		13		21		11		20		37	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T_n	BAR	8.52	12.48	8.76	12.68	9.68	14.16	16.92	22.42	14.94	21.14	13.32	19.42
	TUK	8.16	12.10	8.10	12.14	9.10	13.60	16.26	21.78	14.50	20.50	12.80	18.32
	QS	8.24	11.74	8.82	12.86	10.88	15.54	15.76	21.34	13.16	19.22	12.62	18.24
H_n	BAR	7.54	10.84	7.54	11.42	7.98	11.56	16.22	21.78	14.22	20.14	12.32	17.74
	TUK	7.36	10.68	7.26	11.06	7.60	11.36	15.28	20.38	12.52	18.14	10.88	16.32
	QS	7.32	10.70	7.32	11.34	8.12	11.53	15.84	20.98	13.88	19.68	11.60	16.82
M_n		6.14 at 5%			9.92 at 10%			6.82 at 5%			11.40 at 10%		

Note: Model $X_t - 0.1X_{t-1} = (1-B)^{-0.4}\zeta_t$, $\zeta_t \sim N(0,1)$.