# A NEW SPECTRAL METHOD IN TIME SERIES ANALYSIS

A Dissertation

by

JUNHO YANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

| | |
|---|---|
| Chair of Committee, | Suhasini Subba Rao |
| Co-Chair of Committee, | Mikyoung Jun |
| Committee Members, | Moshen Pourahmadi |
| | Michael Anshelevich |
| Head of Department, | Branislav Vidakovic |

May 2021

Major Subject: Statistics

# ABSTRACT

In this dissertation, we propose a new spectral method that could be used to overcome two issues in time series analysis.

The first issue is the small sample problem. The periodogram is widely used to analyze second order stationary time series, since an expectation of the periodogram is approximately equal to the underlying spectral density of the time series. However, it is well known that the periodogram suffers from a finite sample bias. We show that the bias arises because of the finite boundary of observation in the discrete Fourier transforms (DFT), which is used in the construction of the periodogram. Moreover, we show that by using the best linear predictors of the time series outside the observed domain, we can obtain the "complete periodogram" that is an unbiased estimator of the spectral density. We propose a method for estimating the best linear predictors and prove, both theoretically and empirically, that the resulting estimated complete periodogram has a smaller bias than the regular periodogram. The estimated complete periodogram can be used to estimate parameters, which is expressed as a weighted sum of the spectral density.

The second issue is the discrepancy between time and frequency domain methods in parameter estimation. In time series analysis, there is a clear distinction between the two domain methods. We draw connections between two domain methods by deriving an exact and interpretable bound between the Gaussian and Whittle likelihood of a second order stationary time series. The derivation is based on obtaining the transformation, which is biorthogonal to the DFT of the time series. Such a transformation yields a new decomposition for the inverse of a Toeplitz matrix and enables the representation of the Gaussian likelihood within the frequency domain. Based on this result, we obtain an approximation for the difference between the Gaussian and Whittle likelihoods and define two new frequency domain quasi-likelihood criteria. We show that these new criteria are computationally fast and yield a better approximation of the spectral divergence criterion, as compared to both the Gaussian and Whittle likelihoods.

# ACKNOWLEDGMENTS

First and foremost, I would like to thank my Ph.D. advisors Suhasini Subba Rao and Mikyoung Jun. Suhasini introduced to me the beauty of the spectral methods in time series analysis. She was always willing to answer any questions and taught me many invaluable writing strategies. Working with Mikyoung taught me how to come up with a new idea. Her expertise in spatial statistics and its applications spiked my interest in many practical problems. Without their help and support, these works wouldn't have been possible.

I also would like to extend my gratitude to my dissertation committee, Mohsen Pourahmadi and Michael Anshelevich. Moshen was always available for discussion, and his pedagogies of teaching greatly influenced my teaching philosophy. Michael introduced a random matrix theory, which was sensational at that time and led me to great interest in this area. I am also grateful to Gregory Berkolaiko for his great support and advice.

Thanks to Sourav Das, R. Saravanan, and Courtney Schumacher for their amazing collaborative work during my Ph.D. study, to Alan Dabney and Darren Homrighausen for their help in teaching undergraduate courses, to Andrea Dawson for her assistance for graduate study, and to Danny Nam for his positive influence since we were in high school.

Thanks to my colleagues at Texas A&M. Changwoo, Hyunwoong, Junsouk, Myeongjong, Rakheon, Sangyoon, and Se Yoon for Korean gatherings. Biraj, Daniel, Huiling, Pallavi, Reza, Sandipan, Sean, Xiaomeng, and Yan for having good rapports since the first year of graduate study.

Finally, this dissertation is dedicated to my family, who always trust me and give unconditional support.

CONTRIBUTORS AND FUNDING SOURCES

## Contributors

This work was supported by a dissertation committee consisting of Professors Suhasini Subba Rao (advisor), Mikyoung Jun (co-advisor), and Moshen Pourahmadi of the Department of Statistics and Professor Michael Anshelevich of the Department of Mathematics.

The Ball bearing data analysis in Section 3.5.1 were conducted in part by Professor Sourav Das at James Cook University.

All other work conducted for the dissertation was completed by the student independently under the advisement of committee members.

## Funding Sources

Graduate study was supported by a teaching assistantship from Department of Statistics at Texas A&M University and a research assistantship is partially surpported from the National Science Foundation (NSF).

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

# 1. INTRODUCTION

## 1.1 A review of the spectral methods

The analysis of a time series in the frequency domain has a long history dating back to Schuster (1897, 1906). Schuster first defined the periodogram as a method of identifying periodicities in sunspot activity. Today, spectral analysis remains an active area of research with widespread applications in several disciplines from astronomical data to the analysis of EEG signals in the Neuroscience. Regardless of the discipline, the periodogram remains one of the most widely used tools in spectral analysis, as the periodogram is primarily a tool for detecting periodicities in a signal and various types of second order behavior in a time series.

Despite the popularity of the periodogram, it well known that it can have a severe finite sample bias (see Tukey (1967)). To be precise, we recall that $\{X_t\}_{t\in\mathbb{Z}}$ is a second order stationary time series (we will simply call it a stationary time series) if $\mathbb{E}[X_t] = \mu$ and the autocovariance function can be written as $c(r) = \mathrm{cov}(X_t, X_{t+r})$ for all $r$ and $t \in \mathbb{Z}$. Further, if $\sum_{r\in\mathbb{Z}} c(r)^2 < \infty$, then $f(\omega) = \sum_{r\in\mathbb{Z}} c(r)e^{ir\omega}$ is the corresponding (well-defined) spectral density function. To simplify the derivations, we assume $\{X_t\}$ is a demeaned time series, i.e., $\mu = 0$. The periodogram of an observed time series $\{X_t\}_{t=1}^n$ is defined as $I_n(\omega) = |J_n(\omega)|^2$, where $J_n(\omega)$ is the "regular" discrete Fourier transform (DFT), which is defined by

$$J_n(\omega) = n^{-1/2}\sum_{t=1}^{n} X_t e^{it\omega} \quad \text{with} \quad i = \sqrt{-1}.$$

It is well known that if $\sum_{r\in\mathbb{Z}} |rc(r)| < \infty$, then

$$\mathbb{E}[I_n(\omega)] = f_n(\omega) = f(\omega) + O(n^{-1}).$$

However, the seemingly small $O(n^{-1})$ error can be large in certain situation. A more detailed analysis shows $f_n(\omega)$ is the convolution between the true spectral density and the $n$th order Fejér

kernel $F_n(\lambda) = \frac{1}{n}\left(\frac{\sin(n\lambda/2)}{\sin(\lambda/2)}\right)^2$. This convolution smooths out the peaks in the spectral density function due to the "sidelobes" in the Fejér kernel. This effect is often called the leakage effect and it is greatest when the spectral density has a large peak and the sample size is small. Tukey (1967) showed that an effective method for reducing leakage is to taper the data and evaluate the periodogram of the tapered data. Brillinger (1981) and Dahlhaus (1983) showed that asymptotically the periodogram based on tapered time series shared many properties similar to the non-tapered periodogram. The number of points that are tapered will impact the bias, thus Hurvich (1988) proposed a method for selecting the amount of tapering. A theoretical justification for the reduced bias of the tapered periodogram is derived in Dahlhaus (1988), Lemma 5.4, where for the data tapers of degree $(k, \kappa) = (1, 0)$, he showed that the bias of the tapered periodogram (precise definition of the tapered periodogram is in Section 3.1) is $O(n^{-2})$. An imputation based approach to correct for the bias has recently been proposed in Lee and Zhu (2009) and Guinness (2019).

Many parameters in time series can be written in terms of the weighted average of the spectral density and we construct a statistic by replacing spectral density with periodogram. Therefore, the leakage effect in spectral analysis could be, in a subtle manner, a reason for the bias issue of the parameter estimations. Some non-trivial but related example is the Whittle's likelihood approximation. Whittle (1951, 1953) introduced the Whittle likelihood as an approximation of the Gaussian likelihood. To be more precise, suppose we fit a parametric second order stationary model with spectral density $f_\theta(\omega)$ and corresponding autocovariance function $\{c_{f_\theta}(r)\}_{r\in\mathbb{Z}}$ to the observed time series $\{X_t\}_{t=1}^n$. The (quasi) negative log-Gaussian likelihood (we simple call it the Gaussian likelihood) defined in the time domain is proportional to

$$\mathcal{L}_n(\theta; \underline{X}_n) = n^{-1}\left(\underline{X}_n' \Gamma_n(f_\theta)^{-1} \underline{X}_n + \log|\Gamma_n(f_\theta)|\right) \tag{1.1}$$

where $\Gamma_n(f_\theta)_{s,t} = c_{f_\theta}(s-t)$ is a Toeplitz matrix, $|A|$ denotes the determinant of the matrix $A$ and $\underline{X}_n' = (X_1, \ldots, X_n)$. In contrast, the Whittle likelihood is a "spectral divergence" between the

periodogram and the conjectured spectral density

$$K_n(\theta; \underline{X}_n) = n^{-1} \sum_{k=1}^{n} \left( \frac{|J_n(\omega_{k,n})|^2}{f_\theta(\omega_{k,n})} + \log f_\theta(\omega_{k,n}) \right) \quad \omega_{k,n} = \frac{2\pi k}{n}. \tag{1.2}$$

A decade later from Whittle, Walker (1964) derived the large sample properties of moving average models fitted using the Whittle likelihood. Subsequently, the Whittle likelihood has become a popular method for parameter estimation of various stationary time series (both long and short memory) and spatial models. The Whittle likelihood is computationally a very attractive method for estimation. Despite the considerable improvements in computation algorithms, interest in the Whittle likelihood has not abated. Several diverse applications of the Whittle likelihood can be found in Dahlhaus and Künsch (1987) (for spatial processes), Fox and Taqqu (1986), Robinson (1995) Hurvich and Chen (2000), Giraitis and Robinson (2001), Abadir et al. (2007), Shao and Wu (2007), Giraitis et al. (2012) (long memory time series and local Whittle methods), Choudhuri et al. (2004), Kirch et al. (2019) (Bayesian spectral methods), and Panaretos and Tavakoli (2013), van Delft and Eichler (2020) (functional time series), to name but few.

The Whittle likelihood can be interpreted as an "estimator" of the spectral divergence (precise definition is in (4.1)) which is a weighted average of the true spectral density. Therefore, despite its advantages, the Whittle likelihood can give rise to estimators with a substantial bias due to the leakage effect (see Priestley (1981) and Dahlhaus (1988)). Dahlhaus (1988) showed that the finite sample bias in the periodogram impacts the performance of the Whittle likelihood. Motivated by this discrepancy, Sykulski et al. (2019) proposed the debiased Whittle likelihood, which fits directly to the expectation of the periodogram rather than the limiting spectral density. Alternatively, Dahlhaus (1988) used the tapered Whittle likelihood to improve the bias. Empirical studies show that the tapered Whittle likelihood yields a smaller bias than the regular Whittle likelihood. As a theoretical justification, Dahlhaus (1988, 1990) used an alternative asymptotic framework to show that tapering yields a good approximation to the inverse of the Toeplitz matrix. It is worth mentioning that within the time domain, several authors, including Shaman (1975, 1976); Bhansali (1982)

and Coursol and Dacunha-Castelle (1982), have studied approximations to the inverse variance matrix. These results can be used to approximate the Gaussian likelihood.

However, as far as we are aware, there are no results which investigate the exact bias term of the spectral methods. Our main objective of this dissertation is to quantify the "loss" when using the periodogram as a primary tool to analysis the time series data. The benefits of such insight is not only of theoretical interest but also lead to the development of computationally simple frequency domain methods which are comparable with the Gaussian likelihood.

## 1.2 Contributions

Our contributions in this dissertation are threefold. The first contribution is that we obtain the linear transformation of the observed time series $\{X_t\}_{t=1}^n$ which is biorthogonal to the regular DFT, $\{J_n(\omega_{k,n})\}_{k=1}^n$.

A brief construction of such transformation is as follows. Assume that the spectral density of the underlying stationary time series is bounded and strictly positive. Under these conditions, for any $\tau \in \mathbb{Z}$ we can define the best linear predictor of $X_\tau$ given the observed time series $\{X_t\}_{t=1}^n$. We denote this predictor as $\widehat{X}_{\tau,n}$. Based on these predictors we define a new DFT

$$\widetilde{J}_n(\omega; f) = n^{-1/2} \sum_{\tau=-\infty}^{\infty} \widehat{X}_{\tau,n} e^{i\tau\omega}. \tag{1.3}$$

By its definition, it is obvious that $\widetilde{J}_n(\omega; f) \in \mathrm{sp}(\underline{X}_n)$ where $\mathrm{sp}(\underline{X}_n)$ is a span of $X_1, ..., X_n$ on the complex field. Moreover, using the property of the best linear predictors, in particular for $1 \leqslant t \leqslant n$ and $\tau \in \mathbb{Z}$ that $\mathrm{cov}(X_t, \widehat{X}_{\tau,n}) = c(t-\tau)$, we can show that $\{\widetilde{J}_n(\omega_{k,n}; f)\}_{k=1}^n$ is biorthogonal to the regular DFTs in the sense that

$$\mathrm{cov}(\widetilde{J}_n(\omega_{k_1,n}; f), J_n(\omega_{k_2,n})) = f(\omega_{k_1,n})\delta_{k_1,k_2} \qquad 1 \leqslant k_1, k_2 \leqslant n \tag{1.4}$$

where $\delta_{k_1,k_2} = 1$ when $k_1 = k_2$ and zero otherwise.

Using that $\widehat{X}_{\tau,n} = X_\tau$ for $1 \leqslant \tau \leqslant n$, (1.3) can be written as $\widetilde{J}_n(\cdot; f) = J_n(\cdot) + \widehat{J}_n(\cdot; f)$ where

$$\widehat{J}_n(\omega; f) = n^{-1/2} \sum_{\tau=-\infty}^{0} \widehat{X}_{\tau,n} e^{i\tau\omega} + n^{-1/2} \sum_{\tau=n+1}^{\infty} \widehat{X}_{\tau,n} e^{i\tau\omega}.$$

Therefore, the biorthgonal transform (to the regular DFT) is the regular DFT plus the Fourier transform of the best linear predictors of the time series outside the domain of observation. Since $\widehat{J}_n(\omega; f)$ is a DFT of all linear predictors, we call it the predictive DFT. Moreover, we call $\widetilde{J}_n(\cdot; f)$ the complete DFT as it "completes" the information not found in the regular DFT. Details of the complete and predictive DFT is described in Section 2.2.

The second contribution is that using the complete DFT, we provide an alternative approach, which yields a "periodogram" with a bias of order less than $O(n^{-1})$. The complete DFT defined as in (1.3) also satisfies

$$\operatorname{cov}(\widetilde{J}_n(\omega; f), J_n(\omega)) = f(\omega) \qquad 0 \leqslant \omega \leqslant \pi. \tag{1.5}$$

Based on (1.5), we define the unbiased complete periodogram $I_n(\omega; f) = \widetilde{J}_n(\omega; f)\overline{J_n(\omega)}$.

Unlike the regular periodogram, $I_n(\omega; f)$ depends on the (unknown) underlying spectral density and thus it needs to be estimated. For most time series models (an important exception for the autoregressive model of finite order), $I_n(\omega; f)$ does not have a simple analytic form. Instead in Section 2.3, we derive an approximation of $I_n(\omega; f)$, and propose a method for estimating the approximation. Both the approximation and estimation will induce errors in $I_n(\omega; f)$. However, we prove, under mild conditions, that the bias of the resulting estimator of $I_n(\omega; f)$ is less than $O(n^{-1})$. We show in the simulations (Section 3.4.1), that the resulting estimated complete periodogram outperforms than the classical periodogram and tends to better capture the peaks of the underlying spectral density.

The last contribution is that we derive an exact, interpretable, bound between the Gaussian and Whittle likelihood of a stationary time series. The key to the derivation is the complete and

predictive DFT. In Theorem 2.4.1, we show that the first term of the Gaussian likelihood is

$$n^{-1}\underline{X}_n'\Gamma_n(f_\theta)^{-1}\underline{X}_n = \frac{1}{n}\sum_{k=1}^{n}\frac{\widetilde{J}_n(\omega_{k,n};f_\theta)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} \tag{1.6}$$

thus the (first term of the) difference between the Gaussian and Whittle likelihood is

$$\mathcal{L}_n(\theta) - K_n(\theta) = \frac{1}{n}\sum_{k=1}^{n}\frac{\widehat{J}_n(\omega_{k,n};f_\theta)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})}.$$

Therefore, the difference between the Gaussian and Whittle likelihood is due to the omission of these linear predictors outside of the observed domain.

It is common to use the Cholesky decomposition to decompose the inverse of a Toeplitz matrix, $\Gamma_n(f_\theta)^{-1}$. However, an interesting aspect of (1.6) is that it provides an alternative decomposition of the inverse of a Toeplitz matrix using the complete DFT. In general processes, the complete DFT does not have a simple analytic expression. In Section 2.4.1, we obtain an approximation of the complete DFT and thus an approximation of the inverse Toeplitz matrix in terms of the infinite order causal and minimum phase autogregressive factorization of $f_\theta$. We prove in Theorem 2.4.3 that under mild conditions, element-wise $\ell_1$ norm of an approximation error converges to zero rapidly.

### 1.3 Organization

The rest of the dissertation is organized as follows. In Section 3, we discuss greater detail of the complete periodogram in Section 2.3. In Section 3.1, we propose a variant of the estimated complete periodogram, which tapers the regular DFT. In the simulations, it appears to improve on the non-tapered complete periodogram. In Section 3.2, we consider the integrated periodogram estimators, where the spectral density is replaced with the estimated complete periodogram (both tapered and non-tapered). Some examples can be found in Section 3.3. In Section 3.4, we illustrate the proposed variant of the complete periodogram method with simulations. Two real data analyses (ball bearing and sunspot data) are considered in Section 3.5. The various estimated

complete periodograms, proposed in this section, are available as an R package called `cspec` on CRAN(`https://cran.r-project.org`). Lastly, proof for the results in Sections 2.2, 2.3, and 3 can be found in Section 3.6.

In Section 4, we discuss greater detail of the frequency domain representation of the Gaussian likelihood in Section 2.4. In Section 4.1, we use an approximation for the difference in likelihoods, $\mathcal{L}_n(\theta) - K_n(\theta)$ in Section 2.4.1, to define two new spectral divergence criteria: The boundary corrected and hybrid Whittle. In Section 4.2, we describe a set of assumptions which is required to prove the sampling properties of the new likelihoods. In Section 4.3, we show consistency results of the new likelihood estimators and in Section 4.4, we calculate an asymptotic bias and variance of new likelihood estimators. In Section 4.5, we discuss of the implementation issues of the new estimators. In Section 4.6, we illustrate and compare the proposed frequency domain estimators through some simulations. We study the performance of the estimation scheme when the parametric model is both correctly specified (Section 4.6.1) and misspecified (Section 4.6.2). Also, empirical results when fitting lower order model (Section 4.6.3) and alternative estimating methods (Section 4.6.4) are presented. Finally, proof for the results in Sections 2.4 and 4 can be found in Section 4.7.

Some additional results and simulations can be found in Appendix. In Appendix A, Baxter-type inequalities for the finite predictors and their derivatives are introduced with proof. These inequalities play an important role to prove approximation results in the main part. In Appendix B, we derive an expression for the asymptotic bias of the Gaussian, Whittle likelihoods, and the new frequency domain likelihoods. Additional simulations for Section 4 are in Appendix C and some technical lemmas are derived in Appendix D.

# 2. MAIN RESULTS *

## 2.1 Notational conventions

In this section, we introduce most of the notation used in the dissertation. Let $\{X_t\}_{t\in\mathbb{Z}}$ be a second order stationary time series and we assume that $\mathbb{E}[X_t] = 0$ (as it makes the derivations cleaner). We use $\{c_f(r)\}_{r\in\mathbb{Z}}$ to denote an autocovariance function and $f(\omega) = \sum_{r\in\mathbb{Z}} c_f(r)e^{ir\omega}$ its corresponding spectral density. Sometimes, it will be necessary to make explicit the true underlying covariance (equivalently the spectral density) of the process. In this case, we use the notation $\mathrm{cov}_f(X_t, X_{t+r}) = \mathbb{E}_f[X_tX_{t+r}] = c_f(r)$. We define a Toeplitz matrix corresponds to the spectral density $f$, denotes $\Gamma_n(f)$, is an $n \times n$ matrix with entries $\Gamma_n(f)_{s,t} = c_f(s-t)$.

Let $g$ be a function on $[0, 2\pi]$. We define the $n \times n$ circulant matrix $C_n(g)$ with entries $(C_n(g))_{s,t} = n^{-1}\sum_{k=1}^n g(\omega_{k,n})e^{-i(s-t)\omega_{k,n}}$. Let $A^*$ denote the conjugate transpose of the matrix $A$. Then, the circulant matrix $C_n(g)$ can be written as a matrix form $C_n(g) = F_n^*\Delta_n(g)F_n$, where $\Delta_n(g) = \mathrm{diag}(g(\omega_{1,n}), \ldots, g(\omega_{n,n}))$ is a diagonal matrix and $F_n$ is the $n \times n$ DFT matrix with entries $(F_n)_{s,t} = n^{-1/2}e^{is\omega_{t,n}}$. We recall that the eigenvalues and the corresponding eigenvectors of any circulant matrix $C_n(g)$ are $\{g(\omega_{k,n})\}_{k=1}^n$ and $\{\underline{e}'_{k,n} = (e^{ik\omega_{1,n}}, \ldots, e^{ik\omega_{n,n}})\}_{k=1}^n$ respectively.

Next, we define the norms we will use. Suppose $A$ is a $n \times n$ square matrix, let $\|A\|_p = (\sum_{i,j=1}^n |a_{i,j}|^p)^{1/p}$ be an element-wise $p$-norm for $p \geqslant 1$, and $\|A\|_{spec}$ denote the spectral norm. Let $\|X\|_{\mathbb{E},p} = (\mathbb{E}|X|^p)^{1/p}$, where $X$ is a random variable. For the $2\pi$-periodic square integrable function $g$ with $g(\omega) = \sum_{r\in\mathbb{Z}} g_r e^{ir\omega}$, we use the sub-multiplicative norm $\|g\|_K = \sum_{r\in\mathbb{Z}}(2^K + |r|^K)|g_r|$. Note that if $\sum_{j=0}^{\lfloor K\rfloor+2}\sup_\omega |g^{(j)}(\omega)| < \infty$ then $\|g\|_K < \infty$, where $\lfloor K\rfloor$ is the largest integer smaller or equal to $K$ and $g^{(j)}(\cdot)$ denotes the $j$th derivative of $g$.

Suppose $f, g : [0, 2\pi] \to \mathbb{R}$ are bounded functions, that are strictly larger than zero and are symmetric about $\pi$. By using the classical factorization results in Szegö (1921) and Baxter (1962)

*Parts of this section have been modified with permission from [S. Das, S. Subba Rao, and J. Yang. Spectral methods for small sample time series: A complete periodogram approach. *Journal of Time Series Analysis (To appear)*, 2021, https://doi.org/10.1111/jtsa.12584.] and [S. Subba Rao and J. Yang. Reconciling the Gaussian and Whittle likelihood with an application to estimation in the frequency domain. *Annals of Statistics (To appear), arXiv:2001.06966*, 2021.]

we can write $f(\cdot) = \sigma_f^2|\psi_f(\cdot)|^2 = \sigma_f^2|\phi_f(\cdot)|^{-2}$, where $\phi_f(\omega) = 1 - \sum_{j=1}^{\infty}\phi_j(f)e^{-ij\omega}$ and $\psi_f(\omega) = 1 + \sum_{j=1}^{\infty}\psi_j(f)e^{-ij\omega}$, the terms $\sigma_g$, $\phi_g(\cdot)$, and $\psi_g(\cdot)$ are defined similarly. In Sections 2.4.1 and 4, we require the following notation

$$\begin{aligned} \rho_{n,K}(f) &= \sum_{r=n+1}^{\infty}|r^K\phi_r(f)|, \\ A_K(f,g) &= 2\sigma_g^{-2}\|\psi_f\|_0\|\phi_g\|_0^2\|\phi_f\|_K, \\ \text{and} \qquad C_{f,K} &= \frac{3-\varepsilon}{1-\varepsilon}\,\|\phi_f\|_K^2\,\|\psi_f\|_K^2 \end{aligned}$$

for some $0 < \varepsilon < 1$.

Lastly, we denote $\mathrm{Re}$ and $\mathrm{Im}$ as the real and imaginary part of a complex variable respectively.

## 2.2 The biorthogonal transform to the discrete Fourier transform

We recall that the DFT of the time series plays a fundamental role in the frequency domain methods of the second order stationary time series. With this in mind, our first goal in this section is to derive the transformation $\{Z_{k,n}\}_{k=1}^n \subset \mathrm{sp}(\underline{X}_n)$, which is biorthogonal to $\{J_n(\omega_{k,n})\}_{k=1}^n$. That is, we derive a transformation $\{Z_{k,n}\}_{k=1}^n$ which when coupled with $\{J_n(\omega_{k,n})\}_{k=1}^n$ satisfies the following condition

$$\mathrm{cov}_f\left(Z_{k_1,n}, J_n(\omega_{k_2,n})\right) = f(\omega_{k_1,n})\delta_{k_1,k_2}$$

where $\delta_{k_1,k_2} = 1$ if $k_1 = k_2$ (and zero otherwise). Since $\underline{Z}_n' = (Z_{1,n},\ldots,Z_{n,n}) \in \mathrm{sp}(\underline{X}_n)^n$, there exists an $n \times n$ complex matrix $U_n$, such that $\underline{Z}_n = U_n\underline{X}_n$. Since $(J_n(\omega_{k,1}),\ldots,J_n(\omega_{n,n}))' = F_n\underline{X}_n$, the biorthogonality of $U_n\underline{X}_n$ and $F_n\underline{X}_n$ gives $\mathrm{cov}_f\left(U_n\underline{X}_n, F_n\underline{X}_n\right) = \Delta_n(f)$.

To understand how $U_n\underline{X}_n$ is related to $F_n\underline{X}_n$ we rewrite $U_n = F_n + D_n(f)$. We show in the following theorem that $D_n(f)$ has a specific form with an intuitive interpretation. In order to develop these ideas, we use methods from linear prediction. In particular, we define the best linear predictor of $X_\tau$ for $\tau \in \mathbb{Z}$ given $\{X_t\}_{t=1}^n$ as

$$\widehat{X}_{\tau,n} = \sum_{t=1}^{n}\phi_{t,n}(\tau;f)X_t, \tag{2.1}$$

where $\{\phi_{t,n}(\tau;f)\}_{t=1}^{n}$ are the coefficients which minimize the $L_2$-distance

$$\mathbb{E}[X_\tau - \sum_{t=1}^{n}\phi_{t,n}(\tau;f)X_t]^2 = \frac{1}{2\pi}\int_0^{2\pi}\left|e^{i\tau\omega} - \sum_{t=1}^{n}\phi_{t,n}(\tau;f)e^{it\omega}\right|^2 f(\omega)d\omega.$$

We observe that for $1 \leqslant \tau \leqslant n$, $\phi_{t,n}(\tau;f) = \delta_{\tau,t}$. Furthermore, due to the stationarity, the finite predictor coefficients $\phi_{t,n}(\tau;f)$ are reflective i.e. the predictors of $X_m$ (for $m > n$) and $X_{n+1-m}$ share the same set of prediction coefficients (just reflected) such that

$$\phi_{t,n}(m;f) = \phi_{n+1-t,n}(n+1-m;f) \qquad \text{for } m > n.$$

Using a notation of the finite predictor coefficients, we obtain the following biorthogonal theorem.

**Theorem 2.2.1** (The biorthogonal transform)**.** *Let $\{X_t\}$ be a zero mean second order stationary time series with spectral density $f$. Suppose that $f$ bounded away from zero and whose autocovariance satisfies $\sum_{r\in\mathbb{Z}}|rc_f(r)| < \infty$. Let $\widehat{X}_{\tau,n}$ denote the best linear predictor of $X_\tau$ as defined in (2.1) and $\{\phi_{t,n}(\tau;f)\}$ the corresponding finite predictor coefficients. Then*

$$\mathrm{cov}_f\big((F_n + D_n(f))\underline{X}_n, F_n\underline{X}_n\big) = \Delta_n(f), \tag{2.2}$$

*where $D_n(f)$ has entries*

$$D_n(f)_{k,t} = n^{-1/2}\sum_{\tau\leqslant 0}\left(\phi_{t,n}(\tau;f)e^{i\tau\omega_{k,n}} + \phi_{n+1-t,n}(\tau;f)e^{-i(\tau-1)\omega_{k,n}}\right), \tag{2.3}$$

*for $1 \leqslant k,t \leqslant n$. And, entrywise $1 \leqslant k_1,k_2 \leqslant n$, we have*

$$\mathrm{cov}_f\left(\widetilde{J}_n(\omega_{k_1,n};f), J_n(\omega_{k_2,n})\right) = f(\omega_{k_1,n})\delta_{k_1,k_2} \tag{2.4}$$

*where* $\widetilde{J}_n(\omega; f) = (F_n + D_n(f))\underline{X}_n = J_n(\omega) + \widehat{J}_n(\omega; f)$ *and*

$$\widehat{J}_n(\omega; f) = D_n(f)\underline{X}_n = n^{-1/2} \sum_{\tau \leqslant 0} \widehat{X}_{\tau,n} e^{i\tau\omega} + n^{-1/2} \sum_{\tau > n} \widehat{X}_{\tau,n} e^{i\tau\omega}. \tag{2.5}$$

PROOF. See Section 3.6.1 (note that identity (2.4) can be directly verified using results on best linear predictors). □

What we observe is that the biorthogonal transformation $(F_n + D_n(f))\underline{X}_n$ extends the domain of observation by predicting outside the boundary. A visualization of the observations and the predictors that are involved in the construction of $\widetilde{J}_n(\omega; f)$ is given in Figure 2.1.

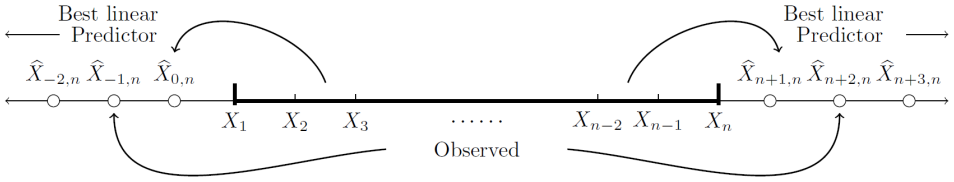



Figure 2.1: $\widetilde{J}_n(\omega; f)$ is the Fourier transform over both the observed time series and its predictors outside this domain.

It is quite surprising that only a small modification of the regular DFT leads to its biorthogonal transformation. Furthermore, we can show that the contribution of the additional DFT term is $\widehat{J}_n(\omega; f) = O_p(n^{-1/2})$. This is why the regular DFT satisfies the well known "near" orthogonal property

$$\mathrm{cov}_f(J_n(\omega_{k_1,n}), J_n(\omega_{k_2,n})) = f(\omega_{k_1})\delta_{k_1,k_2} + O(n^{-1}),$$

see Brillinger (1981) and Lahiri (2003). For future reference, we will use the following definitions.

**Definition 2.2.1.** *We refer to* $\widehat{J}_n(\omega; f)$ *in (2.5) as the predictive DFT (as it is the Fourier transform*

*of all the linear predictors). Nothing that basic algebra yields the expression*

$$\widehat{J}_n(\omega; f) = n^{-1/2} \sum_{t=1}^{n} X_t \sum_{\tau \leqslant 0} (\phi_{t,n}(\tau; f) e^{i\tau\omega} + e^{in\omega} \phi_{n+1-t,n}(\tau; f) e^{-i(\tau-1)\omega}). \tag{2.6}$$

*Note that when $\omega = \omega_{k,n}$, the term $e^{in\omega}$ in (2.6) vanishes. Further, we refer to $\widetilde{J}_n(\omega; f)$ as the complete DFT (as it contains the classical DFT of the time series together with the predictive DFT). Note that both $\widetilde{J}_n(\omega; f)$ and $\widehat{J}_n(\omega; f)$ are functions of $f$ since they involve the spectral density $f(\cdot)$, unlike the regular DFT which is model-free.*

**Remark 2.2.1.** *Biorthogonality of random variables is rarely used in statistics. An interesting exception is Kasahara et al. (2009). They apply the notion of biorthogonality to problems in prediction. In particular they consider the biorthogonal transform of $\underline{X}_n$, which is the random vector $\underline{\widetilde{X}}_n = \Gamma_n(f)^{-1} \underline{X}_n$ (since $\mathrm{cov}_f(\underline{\widetilde{X}}_n, \underline{X}_n) = I_n$). They obtain an expression for the entries of $\underline{\widetilde{X}}_n$ in terms of the Cholesky decomposition of $\Gamma_n(f)^{-1}$. However, there is an interesting duality between $\underline{\widetilde{X}}_n$ and $\underline{\widetilde{J}}_n = (\widetilde{J}_n(\omega_{1,n}; f), \ldots, \widetilde{J}_n(\omega_{n,n}; f))'$. In particular, applying identity (2.28) to the DFT of $\underline{\widetilde{X}}_n$ gives*

$$F_n \underline{\widetilde{X}}_n = F_n \Gamma_n(f)^{-1} \underline{X}_n = \Delta_n(f^{-1}) \underline{\widetilde{J}}_n.$$

*This shows that the DFT of the biorthogonal transform of $\underline{X}_n$ is the* standardized *complete DFT. Conversely, the inverse DFT of the standardized complete DFT gives the biorthogonal transform to the original time series, where the entries of $\underline{\widetilde{X}}_n$ are*

$$\widetilde{X}_{j,n} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{\widetilde{J}_n(\omega_{k,n}; f)}{f(\omega_{k,n})} e^{-ij\omega_{k,n}}.$$

**Remark 2.2.2** (Connection to the orthogonal increment process)**.** *Suppose that $Z(\omega)$ is the orthogonal increment process associated with the stationary time series $\{X_t\}$ and $f$ the corresponding spectral density. If $\{X_t\}$ is a Gaussian time series, then by using Theorem 4.9.1 in Brockwell and*

*Davis (2006) we can show that*

$$\widehat{X}_{\tau,n} = \mathbb{E}\left[X_\tau|\underline{X}_n\right] = \frac{1}{2\pi}\int_0^{2\pi} e^{-i\omega\tau}\mathbb{E}[Z(d\omega)|\underline{X}_n] = \frac{\sqrt{n}}{2\pi}\int_0^{2\pi} e^{-i\omega\tau}\widetilde{J}_n(\omega;f)d\omega.$$

*Based on the above, heuristically,* $\mathbb{E}[dZ(\omega)|\underline{X}_n] = \sqrt{n}\widetilde{J}_n(\omega;f)d\omega$ *and* $\sqrt{n}\widetilde{J}_n(\omega;f)$ *is the derivative of the orthogonal increment process conditioned on the observed time series. Under Assumption 2.3.1, below, it can be shown that* $\mathrm{var}[\widehat{J}_n(\omega;f)] = O(n^{-1})$*, whereas* $\mathrm{var}[J_n(\omega)] = f_n = O(1)$*. Based on this, since* $\widetilde{J}_n(\omega;f) = J_n(\omega) + \widehat{J}_n(\omega;f)$*, then* $\sqrt{n}\widetilde{J}_n(\omega;f) \approx \sqrt{n}J_n(\omega)$*. Thus the regular DFT,* $\sqrt{n}J_n(\omega)$*, can be viewed as an approximation of the derivative of the orthogonal increment process conditioned on the observed time series.*

### 2.2.1 The predictive DFT for the AR($p$) process

In this section, we derive an explicit form of the predictive DFT for the AR($p$) process. To begin with, we calculate the predictive DFT for the AR(1) process.

**Example 2.2.1** (The AR(1) process)**.** *Suppose that* $X_t$ *has an AR*(1) *representation* $X_t = \phi X_{t-1} + \varepsilon_t$ *(*$|\phi| < 1$*). Then the best linear predictors are simply a function of the observations at the two endpoints. That is for* $\tau \leqslant 0$*,* $\widehat{X}_{\tau,n} = \phi^{|\tau|+1}X_1$ *and for* $\tau > n$ $\widehat{X}_{\tau,n} = \phi^{\tau-n}X_n$*. Then the predictive DFT for the AR*(1) *model is*

$$\widehat{J}_n(\omega;f_\phi) = \frac{\phi}{\sqrt{n}}\left(\frac{1}{\phi(\omega)}X_1 + \frac{e^{i(n+1)\omega}}{\overline{\phi(\omega)}}X_n\right) \quad \textit{where} \quad \phi(\omega) = 1 - \phi e^{-i\omega}.$$

*An illustration is given in Figure 2.2.*

We now generalize Example 2.2.1 to the AR($p$) process. Suppose that $f_p(\omega) = \sigma^2|1 - \sum_{j=1}^p \phi_j e^{-ij\omega}|^{-2}$ is the spectral density of the time series $\{X_t\}_{t\in\mathbb{Z}}$ (it is a finite order autoregressive model AR($p$)) and where the characteristic polynomial associated with $\{\phi_j\}_{j=1}^p$ has roots lying outside the unit circle. Clearly, we can represent the time series $\{X_t\}_{t\in\mathbb{Z}}$ as

$$X_t = \sum_{j=1}^p \phi_j X_{t-j} + \varepsilon_t \quad t \in \mathbb{Z}$$


Figure 2.2: The past and future best linear predictors based on a AR(1) model.

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are uncorrelated random variables with $\mathbb{E}[\varepsilon_t] = 0$ and $\mathrm{var}[\varepsilon_t] = \sigma^2$. For finite order AR$(p)$ processes with autoregressive coefficients $\{\phi_j\}_{j=1}^p$, the best linear predictor of $X_0$ and $X_{n+1}$ given $\{X_t\}_{t=1}^n$ are $\widehat{X}_{0,n} = \sum_{j=1}^p \phi_j X_j$ and $\widehat{X}_{n+1,n} = \sum_{j=1}^p \phi_j X_{n+1-j}$ respectively. In general, we can recursively define the best linear predictors $\widehat{X}_{1-\tau,n}$ and $\widehat{X}_{n+\tau,n}$ to be

$$\widehat{X}_{1-\tau,n} = \sum_{j=1}^p \phi_j \widehat{X}_{1-\tau+j,n} \quad \text{and} \quad \widehat{X}_{n+\tau,n} = \sum_{j=1}^p \phi_j \widehat{X}_{n+\tau-j,n} \qquad \text{for } \tau \geqslant 1, \tag{2.7}$$

where $\widehat{X}_{t,n} = X_t$ for $1 \leqslant t \leqslant n$. Therefore, using (2.7) and similar to Example 2.2.1 for AR$(1)$ model, we can obtain the expression of the predictive DFT of an AR$(p)$ model in terms of the AR coefficients $\{\phi_j\}_{j=1}^p$.

**Theorem 2.2.2** (Predictive DFT for a finite order autoregressive models)**.** *Suppose that* $f_p(\omega) = \sigma^2|\phi_p(\omega)|^{-2}$ *where* $\phi_p(\omega) = 1 - \sum_{j=1}^p \phi_j e^{-ij\omega}$ *(the roots of the corresponding characteristic polynomial lie outside the unit circle) and* $p \leqslant n$*. Then, the predictive DFT has the analytic form*

$$\widehat{J}_n(\omega; f_p) = \frac{n^{-1/2}}{\phi_p(\omega)} \sum_{\ell=1}^p X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{-is\omega} + e^{in\omega} \frac{n^{-1/2}}{\overline{\phi_p(\omega)}} \sum_{\ell=1}^p X_{n+1-\ell} \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{i(s+1)\omega}. \tag{2.8}$$

PROOF. See Section 3.6.1. □

## 2.3 The complete periodogram

Using the notation of the complete and predictive DFT, we define the "complete" periodogram

$$I_n(\omega; f) = \widetilde{J}_n(\omega)\overline{J_n(\omega)} = |J_n(\omega)|^2 + \widehat{J}_n(\omega)\overline{J_n(\omega)} \qquad \omega \in [0, 2\pi]. \tag{2.9}$$

Then, using the similar technique in the proof of Theorem 2.2.1, it can be shown that $\mathbb{E}[I_n(\omega; f)] = f(\omega)$ for all $\omega \in [0, 2\pi]$. Therefore, the complete periodogram is an unbiased estimator of the spectral density for the entire frequency. However, the complete periodogram involves $\widehat{J}_n(\omega; f)$ which is a function of the unknown spectral density. Thus, the complete periodogram cannot be directly evaluated. In next sections, we obtain feasible approximations of the complete periodogram and the corresponding errors.

### 2.3.1 The AR($p$) model and an AR($\infty$) approximation

Recall from (2.6), the predictive DFT can be expressed in terms of the finite predictor coefficients which is, in general, an unwieldy function of the autocovariance function. Therefore, obtaining an approximation of the predictive DFT could be challenging. However, for certain spectral density functions, it is approachable. In Theorem 2.2.2, we show that when $f = f_p$ corresponds to the AR($p$) spectral density, $\widehat{J}_n(\omega; f_p)$ has a relatively simple analytic form in terms of the AR coefficients. This tells us that for finite order autoregressive models, estimation of the predictive DFT only requires us to estimate $p$ number of autoregressive parameters.

For general stationary time series, such simple expressions are not possible. But (2.8) provides a clue to obtaining a near approximation, based on the AR($\infty$) and MA($\infty$) representation that many stationary time series satisfy. It is well known that if the spectral density $f$ is strictly positive, then it has an AR($\infty$) and MA($\infty$) representation see Baxter (1962) (see also equation (2.3) in Kreiss et al. (2011))

$$\begin{aligned} X_t - \sum_{j=1}^{\infty} \phi_j(f) X_{t-j} &= \varepsilon_t \\ X_t &= \varepsilon_t + \sum_{j=1}^{\infty} \psi_j(f) \varepsilon_{t-j}, \end{aligned}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is an uncorrelated White noise process with $\mathbb{E}[\varepsilon_t^2] = \sigma^2$. Unlike finite order autoregressive models, $\widehat{J}_n(\omega; f)$ cannot be represented in terms of $\{\phi_j(f)\}_{j=1}^{\infty}$, since it only involves the sum of the best finite predictors (not infinite predictors). Instead, we define an approximation

based on (2.8), but using the AR($\infty$) and MA($\infty$) representation

$$\begin{aligned}\widehat{J}_{\infty,n}(\omega;f) &= \frac{n^{-1/2}}{\phi(\omega;f)}\sum_{\ell=1}^{n}X_\ell\sum_{s=0}^{\infty}\phi_{\ell+s}(f)e^{-is\omega}+e^{in\omega}\frac{n^{-1/2}}{\overline{\phi(\omega;f)}}\sum_{\ell=1}^{n}X_{n+1-\ell}\sum_{s=0}^{\infty}\phi_{\ell+s}(f)e^{i(s+1)\omega}\\ &= \frac{\psi(\omega;f)}{\sqrt{n}}\sum_{\ell=1}^{n}X_\ell\sum_{s=0}^{\infty}\phi_{\ell+s}(f)e^{-is\omega}+e^{in\omega}\frac{\overline{\psi(\omega;f)}}{\sqrt{n}}\sum_{\ell=1}^{n}X_{n+1-\ell}\sum_{s=0}^{\infty}\phi_{\ell+s}(f)e^{i(s+1)\omega}\end{aligned} \tag{2.10}$$

where $\phi(\omega;f) = 1-\sum_{j=1}^{\infty}\phi_j(f)e^{-ij\omega}$ and $\psi(\omega;f) = \sum_{j=0}^{\infty}\psi_j(f)e^{-ij\omega}$ (we set $\psi_0(f) = 1$ by convention). Though seemingly unwieldy, (2.10) has a simple interpretation. It corresponds to the Fourier transform of the best linear predictors of $X_\tau$ given the infinite future $\{X_t\}_{t=1}^{\infty}$ (if $\tau \leqslant 0$) and $X_\tau$ given in the infinite past $\{X_t\}_{t=-\infty}^{n}$ (if $\tau > n$), but are truncated to the observed terms $\{X_t\}_{t=1}^{n}$. Of course, this is not $\widehat{J}_n(\omega;f)$. However, we show that

$$I_{\infty,n}(\omega;f) = \Big(J_n(\omega)+\widehat{J}_{\infty,n}(\omega;f)\Big)\overline{J_n(\omega)} \tag{2.11}$$

is a close approximation of the complete periodogram, $I_n(\omega;f)$. To do so, we require the following assumptions.

The first set of assumptions is on the second order structure of the time series.

**Assumption 2.3.1.** *$\{X_t\}_{t\in\mathbb{Z}}$ is a second order stationary time series, where*

*(i) The spectral density $f$, is a bounded and strictly positive function.*

*(ii) For some $K > 1$, the autocovariance function is such that $\sum_{r\in\mathbb{Z}}|r^K c_f(r)| < \infty$.*

Assumption 2.3.1(ii) is related to the smoothness of the spectral density function. Assumption 2.3.1(ii) implies that $f$ is $s$-times differentiable, where the $s$th derivative is bounded for all $s < K$. Conversely, Assumption 2.3.1(ii) is satisfied for all $2\pi$-periodic functions which are $s$-times continuously differentiable for some $s > K+1$. We also mention that under Assumption 2.3.1, the corresponding AR($\infty$) and MA($\infty$) coefficients are such that $\sum_{j=1}^{\infty}|j^K\phi_j(f)| < \infty$ and $\sum_{j=1}^{\infty}|j^K\psi_j(f)| < \infty$ (see Lemma 2.1 in Kreiss et al. (2011)).

The next set of assumptions are on the higher order cumulants structure of the time series.

**Assumption 2.3.2.** *$\{X_t\}$ is an $2m$-order stationary time series such that $\mathbb{E}[|X_t|^{2m}] < \infty$ and $\mathrm{cum}(X_t, X_{t+s_1}, ..., X_{t+s_{h-1}}) = \mathrm{cum}(X_0, X_{s_1}, \ldots, X_{s_{h-1}}) = \kappa_h(s_1, ..., s_{h-1})$ for all $t, s_1, ..., s_{h-1} \in \mathbb{Z}$ with $h \leqslant 2m$. Further, the joint cumulant $\{\kappa_h(s_1, ..., s_{h-1})\}$ satisfies*

$$\sum_{s_1,...,s_{h-1}\in\mathbb{Z}} |\kappa_h(s_1, \ldots, s_{h-1})| < \infty \quad \textit{for} \quad 2 \leqslant h \leqslant 2m.$$

Before studying the approximation error when replacing $I_n(\omega; f)$ with $I_{\infty,n}(\omega; f)$ we first obtain some preliminary results on the complete periodogram $I_n(\omega; f)$. The following result concerns the order of contribution of the predictive DFT in the complete periodogram. Suppose Assumptions 2.3.1 (with $K \geqslant 1$) and 2.3.2 (for $m = 2$) hold. Let $\widehat{J}_n(\omega; f)$ be defined as in (2.5). Then

$$\mathbb{E}[\widehat{J}_n(\omega; f)\overline{J_n(\omega)}] = O(n^{-1}), \quad \mathrm{var}(\widehat{J}_n(\omega; f)\overline{J_n(\omega)}) = O(n^{-2}). \tag{2.12}$$

The details of the proof of the above can be found in Section 3.6.2.

Moreover, there are two main differences between the complete periodogram and the regular periodogram. The first is that the complete periodogram can be complex, however the imaginary part is mean zero and the variance is of order $O(n^{-1})$. Thus without loss of generality, we can focus on the real part of the complete periodogram $\widetilde{J}_n(\omega; f)\overline{J_n(\omega)}$, denotes $\mathrm{Re}\, \widetilde{J}_n(\omega; f)\overline{J_n(\omega)}$. Second, unlike the regular periodogram, $\mathrm{Re}\, \widetilde{J}_n(\omega; f)\overline{J_n(\omega)}$, can be negative. Therefore if positivity is desired it makes sense to threshold $\mathrm{Re}\, \widetilde{J}_n(\omega; f)\overline{J_n(\omega)}$ to be non-zero. Thresholding $\mathrm{Re}\, \widetilde{J}_n(\omega; f)\overline{J_n(\omega)}$ to be non-zero induces a small bias. But we observe from the simulations in Section 3.4 that the bias is small (see the middle column in Figures 3.1$-$3.3 where the average of the thresholded true complete periodogram for various models is given).

Lastly, we mention the variance of the complete periodogram. In the simulations, we observe that the variance of the complete periodogram tends to be larger than the regular periodogram, especially at frequencies where the spectral density peaks. To understand why, we focus on the

Gaussian time series. For the complete periodogram, it can be shown that

$$\mathrm{var}[I_n(\omega; f)] = \mathrm{var}[\widetilde{J}_n(\omega; f)] \cdot \mathrm{var}[J_n(\omega)] + O(n^{-2}).$$

By Cauchy-Schwarz inequality, we have for all $n$ that

$$\mathrm{var}[\widetilde{J}_n(\omega; f)] \cdot \mathrm{var}[J_n(\omega)] \geqslant |\mathrm{cov}[\widetilde{J}_n(\omega; f), J_n(\omega)]|^2 = f(\omega)^2.$$

Thus the variance of the complete periodogram is such that $\mathrm{var}[I_n(\omega; f)] \geqslant f(\omega)^2$. By contrast the variance of the regular periodogram is $\mathrm{var}[I_n(\omega)] \approx f_n(\omega)^2 < f(\omega)^2$. Nevertheless, despite an increase in variance of the periodogram, our empirical results suggest that this may be outweighed by a substantial reduction in the bias of the complete periodogram (see Figures 3.1−3.3 and Table 3.1).

### 2.3.2 The estimated complete periodogram and its approximation bound

Our aim is to estimate the predictive component in the complete periodogram; $\widehat{J}_n(\omega; f)\overline{J_n(\omega)}$. As a starting point, we use the Assumptions in Section 2.3.1 to bound the difference between $I_n(\omega; f)$ and $I_{\infty,n}(\omega; f)$.

**Theorem 2.3.1.** *Suppose Assumption 2.3.1 and 2.3.2 (for $m = 2$) hold. Let $I_n(\omega; f)$ and $I_{\infty,n}(\omega; f)$ is defined as in (2.9) and (2.11) respectively. Then*

$$I_{\infty,n}(\omega; f) = I_n(\omega; f) + \Delta_0(\omega), \tag{2.13}$$

*where* $\sup_\omega \mathbb{E}[\Delta_0(\omega)] = O\left(n^{-K}\right)$, $\sup_\omega \mathrm{var}[\Delta_0(\omega)] = O\left(n^{-2K}\right)$.

PROOF. See Section 3.6.2. □

A few comments on the above approximation are in order. Observe that the approximation error between the complete periodogram and its infinite approximation is of order $O(n^{-K})$. For AR($p$) processes (where $p \leqslant n$) this term would not be there. For AR($\infty$) representations with coefficients

that geometrically decay (e.g., an ARMA process), then $|I_{\infty,n}(\omega; f) - I_n(\omega; f)| = O_p(\rho^n)$, for some $0 \leqslant \rho < 1$. On the other hand, if the AR($\infty$) representation has an algebraic decaying coefficients, $\phi_j(f) \sim |j|^{-K-1-\delta}$ (for some $\delta > 0$), then $|I_{\infty,n}(\omega; f) - I_n(\omega; f)| = O_p(n^{-K})$. In summary, nothing that $I_n(\omega; f)$ is an unbiased estimator of $f$, if $K > 1$, then $I_{\infty,n}(\omega; f)$ has a smaller bias than the regular periodogram.

Now the aim is to estimate $\widehat{J}_{\infty,n}(\omega; f)$. There are various ways this can be done. In this dissertation, we approximate the underlying time series with an AR($p$) process and estimate the AR($p$) parameters. This approximation will incur two sources of errors. The first is approximating an AR($\infty$) process with a finite order AR($p$) model, the second is the estimation error when estimating the parameters in the AR($p$) model. In the following section, we obtain bounds for these errors.

**Remark 2.3.1** (Alternative estimation methods)**.** *If the underlying spectral density is highly complex with several peaks, fitting a finite order AR($p$) model may not be able to reduce the bias. An alternative method is to use the smooth periodogram to estimate the predictive DFT. That is to estimate the AR($\infty$) parameters and MA($\infty$) transfer function $\psi(\omega)$ in (2.10) using an estimate of the spectral density function. This can be done by first estimating the cepstral coefficients (Fourier coefficients of $\log f(\omega)$) using the method Wilson (1972). Then, by using the recursive algorithms obtained in Pourahmadi (1983, 1984, 2001) and Krampe et al. (2018) one can extract estimators of AR($\infty$) and MA($\infty$) parameters from the cepstral coefficients. It is possible that the probabilistic bounds for the estimates obtained in Krampe et al. (2018) can be used to obtain bounds for the resulting predictive DFT, but this remains an avenue for future research.*

Next, we return to the definition of the predictive DFT in (2.5), which is comprised of the best linear predictors outside the domain of observation. In time series, it is common to approximate the best linear predictors with the predictors based on a finite AR($p$) recursion (the so called plug-in estimators; see Bhansali (1996) and Kley et al. (2019)). This approximation corresponds to replacing $f$ in $\widehat{J}_n(\omega; f)$ with $f_p$, where $f_p$ is the spectral density corresponding to "best fitting" AR($p$) model based on $f$.

It is well known that the best fitting AR($p$) coefficients, given the covariances $\{c(r)\}$, are

$$\underline{\phi}_p = (\phi_{1,p}, ..., \phi_{p,p})' = R_p^{-1}\underline{r}_p, \tag{2.14}$$

where $R_p$ is the $p\times p$ Toeplitz variance matrix with $(R_p)_{(s,t)} = c_f(s-t)$ and $\underline{r}_p = (c_f(1), \ldots, c_f(p))'$. This leads to the AR($p$) spectral density approximation of $f$

$$f_p(\omega) = \sigma^2|\phi_p(\omega)|^{-2} = \sigma^2\big|1 - \sum_{j=1}^{p}\phi_{j,p}e^{-ij\omega}\big|^{-2}.$$

The coefficients $\{\phi_{j,p}\}_{j=1}^p$ are used to construct the plug-in prediction estimators for $X_\tau$ ($\tau \leqslant 0$ or $\tau > n$). This in turn gives the approximation of the predictive DFT $\widehat{J}_n(\omega; f_p)$ where the analytic form for $\widehat{J}_n(\omega; f_p)$ is given in (2.8), with the coefficients $\phi_j$ replaced with $\phi_{j,p}$.

Using $\widetilde{J}_n(\omega; f_p) = J_n(\omega) + \widehat{J}_n(\omega; f_p)$ we define the following approximation of the complete periodogram

$$I_n(\omega; f_p) = \widetilde{J}_n(\omega; f_p)\overline{J_n(\omega)}. \tag{2.15}$$

We now obtain a bound for the approximation error, where we replace $I_{\infty,n}(\omega; f)$ with $I_n(\omega; f_p)$.

**Theorem 2.3.2.** *Suppose Assumption 2.3.1 holds with $K > 1$. Let $I_{\infty,n}(\omega; f)$ and $I_n(\omega; f_p)$, be defined as in (2.11) and (2.15) respectively. Then we have*

$$I_n(\omega; f_p) = I_{\infty,n}(\omega; f) + \Delta_1(\omega), \tag{2.16}$$

*where* $\sup_\omega \mathbb{E}[\Delta_1(\omega)] = O\left((np^{K-1})^{-1}\right)$, $\sup_\omega \operatorname{var}[\Delta_1(\omega)] = O\left((np^{K-1})^{-2}\right)$.

PROOF. See Section 3.6.2. □

Applying Theorems 2.3.1 and 2.3.2, we observe that $I_n(\omega; f_p)$ has a smaller bias than the

regular periodogram

$$\mathbb{E}[I_n(\omega; f_p)] = f(\omega) + O\left(\frac{1}{np^{K-1}}\right).$$

In particular, the bias is substantially smaller than the usual $O(n^{-1})$ bias. Indeed, if the true underlying process has an AR($p^*$) representation where $p^* < p$, then the bias is zero.

However, in reality, the true spectral density and best fitting AR($p$) approximation $f$ and $f_p$ respectively are unknown, and they need to be estimated from the observed data.

To estimate the best fitting AR($p$) model, we replace the autocovariances with the sample autocovariances to yield the Yule-Walker estimator of the best fitting AR($p$) parameters

$$\underline{\widehat{\phi}}_p = (\widehat{\phi}_{1,p}, \ldots, \widehat{\phi}_{p,p})' = \widehat{R}_{p,n}^{-1}\, \underline{\widehat{r}}_{p,n}, \tag{2.17}$$

where $\widehat{R}_{p,n}$ is the $p \times p$ sample covariance matrix with $(\widehat{R}_{p,n})_{(s,t)} = \widehat{c}_n(s-t)$ and $\underline{\widehat{r}}_{p,n} = (\widehat{c}_n(1), \ldots, \widehat{c}_n(p))'$ where $\widehat{c}_n(k) = n^{-1} \sum_{t=1}^{n-|k|} X_t X_{t+k}$. We define the estimated AR($p$) spectral density

$$\widehat{f}_p(\omega) = |\widehat{\phi}_p(\omega)|^{-2} = \big|1 - \sum_{j=1}^{p} \widehat{\phi}_{j,p} e^{-ij\omega}\big|^{-2}.$$

Observe that we have ignored including an estimate of the innovation variance in $\widehat{f}_p(\omega)$ as it plays no role in the definition of $\widehat{J}_n(\omega; f_p)$. Using this we define the estimated complete DFT as $\widetilde{J}_n(\omega; \widehat{f}_p) = J_n(\omega) + \widehat{J}_n(\omega; \widehat{f}_p)$, where

$$\widehat{J}_n(\omega; \widehat{f}_p) = \frac{n^{-1/2}}{\widehat{\phi}_p(\omega)} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \widehat{\phi}_{\ell+s,p} e^{-is\omega} + e^{in\omega} \frac{n^{-1/2}}{\overline{\widehat{\phi}_p(\omega)}} \sum_{\ell=1}^{p} X_{n+1-\ell} \sum_{s=0}^{p-\ell} \widehat{\phi}_{\ell+s,p} e^{i(s+1)\omega} \tag{2.18}$$

and corresponding estimated complete periodogram based on $\widehat{f}_p$ is

$$I_n(\omega; \widehat{f}_p) = \widetilde{J}_n(\omega; \widehat{f}_p)\overline{J_n(\omega)}. \tag{2.19}$$

We now show that with the estimated AR($p$) parameters the resulting estimated complete periodogram has a smaller bias (in the sense of Bartlett (1953)) than the regular periodogram.

**Theorem 2.3.3.** *Suppose Assumptions 2.3.1(i) and 2.3.2 (where $m \geqslant 6$ and is multiple of two) hold. Let $I_n(\omega; f_p)$ and $I_n(\omega; \widehat{f}_p)$ be defined as in (2.15) and (2.19) respectively. Then we have the following decomposition*

$$I_n(\omega; \widehat{f}_p) = I_n(\omega; f_p) + \Delta_2(\omega) + R_n(\omega) \tag{2.20}$$

*where $\Delta_2(\omega)$ is the dominating error with*

$$\sup_{\omega} \mathbb{E}[\Delta_2(\omega)] = O\left(\frac{p^3}{n^2}\right), \quad \sup_{\omega} \operatorname{var}[\Delta_2(\omega)] = O\left(\frac{p^4}{n^2}\right)$$

*and $R_n(\omega)$ is such that $\sup_\omega |R_n(\omega)| = O_p\left((p^2/n)^{m/4}\right)$.*

PROOF. See Section 3.6.2. □

We now apply Theorems 2.3.1−2.3.3 to obtain a bound for the approximation error between the estimated complete periodogram $I_n(\omega; \widehat{f}_p)$ and the complete periodogram.

**Theorem 2.3.4.** *Suppose Assumptions 2.3.1 ($K > 1$) and 2.3.2 (where $m \geqslant 6$ and is a multiple of two) hold. Let $I_n(\omega; f) = \widetilde{J}_n(\omega; f)\overline{J_n(\omega)}$ and $I_n(\omega; \widehat{f}_p)$ be defined as in (2.19) respectively. Then we have*

$$I_n(\omega; \widehat{f}_p) = I_n(\omega; f) + \Delta(\omega) + O_p\left(\frac{p^{m/2}}{n^{m/4}}\right),$$

*where $\Delta(\omega) = \Delta_0(\omega) + \Delta_1(\omega) + \Delta_2(\omega)$ (with $\Delta_j(\cdot)$ as defined in Theorems 2.3.1−2.3.3),*
*$\sup_\omega \mathbb{E}[\Delta(\omega)] = O((np^{K-1})^{-1} + p^3/n^2)$ and $\sup_\omega \operatorname{var}[\Delta(\omega)] = O(p^4/n^2)$.*

PROOF. The result immediately follows from Theorems 2.3.1−2.3.3. □

To summarize, by predicting across the boundary using the estimated AR($p$) parameters heuristically we have reduced the "bias" of the periodogram. More precisely, if the probabilistic error in

Theorem 2.3.4 is such that $\frac{p^{m/2}}{n^{m/4}} << \frac{p^3}{n^2}$. Then the "bias" in the sense of Bartlett (1953) is

$$\mathbb{E}[I_{\underline{h},n}(\omega; \widehat{f}_p)] = f(\omega) + O\left(\frac{1}{np^{K-1}} + \frac{p^3}{n^2}\right).$$

Consequently, for $K > 1$, and $p$ chosen such that

$$p^3/n \to 0, \quad \text{as} \quad p, n \to \infty, \tag{2.21}$$

then the "bias" will be less than the $O(n^{-1})$ order. This can make a substantial difference when $n$ is small or the underlying spectral density has a large peak.

In practice the order $p$ of the best AR process needs to be selected. This is usually done using the AIC. In which case the above results need to be replaced with $\widehat{p}$, where $\widehat{p}$ is selected to minimize the AIC

$$\text{AIC}(p) = \log \widehat{\sigma}^2_{p,n} + \frac{2p}{n},$$

$\widehat{\sigma}^2_{p,n} = \frac{1}{n-K_n}\sum_{t=K_n}^{n}(X_t - \sum_{j=1}^{p}\widehat{a}_{j,p}X_{t-j})^2$, $K_n$ is such that $K_n^{2+\delta} \sim n$ for some $\delta > 0$ and the order $p$ is chosen such that $\widehat{p} = \arg\min_{1\leqslant k\leqslant K_n} \text{AIC}(k)$. To show that the selected $\widehat{p}$ satisfies (2.21), we use the conditions in Ing and Wei (2005) who assume that the underlying time series is a linear, stationary time series with an AR($\infty$) that satisfies Assumption K.1$-$K.4 in Ing and Wei (2005). Under Assumption 2.3.1, and applying Baxter's inequality, the AR($\infty$) coefficients satisfy

$$\sum_{j=1}^{\infty}|a_j - a_{j,p}|^2 \leqslant \big(\sum_{j=1}^{\infty}|a_j - a_{j,p}|\big)^2 \leqslant C\big(\sum_{j=p+1}^{\infty}|a_j|\big)^2 = O\left(p^{-2K}\right). \tag{2.22}$$

Under these conditions, Ing and Wei (2005) obtain a bound for $\widehat{p}$. In particular, if the underlying time series has an exponential decaying AR coefficients, then $\widehat{p} = O_p(\log n)$ (see Example 1 in Ing and Wei (2005)) on the other hand if the rate of decay is polynomial order satisfying (2.22), then $\widehat{p} = O_p(n^{1/(1+2K)})$ (see Example 2 in Ing and Wei (2005)). Thus, for for both these cases we have

$\widehat{p}^3/n \xrightarrow{\mathcal{P}} 0$ and $\widehat{p} \xrightarrow{\mathcal{P}} \infty$ as $n \to \infty$.

In summary, using the AIC as a method for selecting $p$, yields an estimated complete periodogram that has a lower bias than the regular periodogram.

### 2.4 The Gaussian likelihood in the frequency domain

In this section, we calculate the exact difference between the Gaussian and Whittle likelihood, and using this exact bound, we obtain the frequency domain representation of the Gaussian likelihood. To compare the Gaussian and Whittle likelihood, we rewrite the Whittle likelihood (defined as in (1.2)) in a matrix form. Using the circulant matrix notation, the Whittle likelihood $K_n(\theta; \underline{X}_n)$ can be written as

$$K_n(\theta; \underline{X}_n) = n^{-1}\big(\underline{X}_n' C_n(f_\theta^{-1})\underline{X}_n + \sum_{k=1}^{n} \log f_\theta(\omega_{k,n})\big). \tag{2.23}$$

Since the focus in this dissertation will be on the first terms in the Gaussian and Whittle likelihoods, we use $\mathcal{L}_n(\theta)$ and $K_n(\theta)$ to denote only these terms:

$$\mathcal{L}_n(\theta) = n^{-1}\underline{X}_n'\Gamma_n(f_\theta)^{-1}\underline{X}_n \quad \text{and} \quad K_n(\theta) = n^{-1}\underline{X}_n' C_n(f_\theta^{-1})\underline{X}_n. \tag{2.24}$$

Therefore, the difference of two likelihoods is

$$\mathcal{L}_n(\theta) - K_n(\theta) = n^{-1}\underline{X}_n'\left(\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1})\right)\underline{X}_n.$$

Since the inverse circulant matrix $C_n(f_\theta^{-1}) = F_n^* \Delta_n(f_\theta^{-1}) F_n$ has a relatively easy form, to obtain the exact bound, it is essential to obtain a "good" expression for the inverse of the variance matrix.

To motivate our approach, we first study the difference in the bias of the AR$(1)$ parameter estimator using both the Gaussian and Whittle likelihood. In Figure 2.3, we plot the bias in the estimator of $\phi$ in the AR$(1)$ model $X_t = \phi X_{t-1} + \varepsilon_t$ for different values of $\phi$ (based on sample size $n = 20$). We observe that the difference between the bias of the two estimators increases as $|\phi|$ approaches one. Further, the Gaussian likelihood clearly has a smaller bias than the Whittle

likelihood (which is more pronounced when $|\phi|$ is close to one).

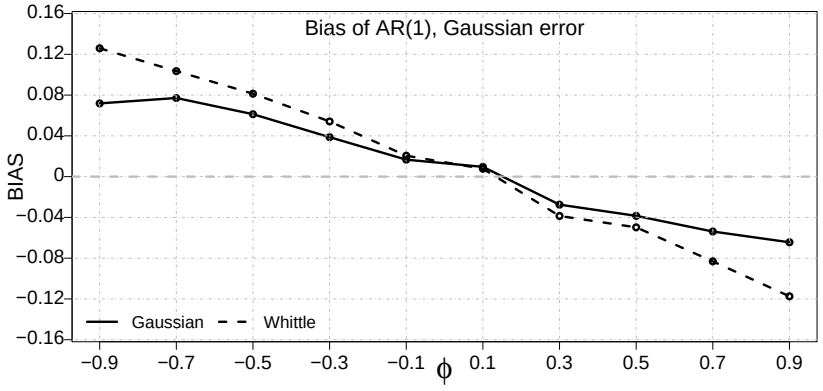



Figure 2.3: The model $X_t = \phi X_{t-1} + \varepsilon_t$ with independent standard normal errors is simulated. The bias of the estimator of $\phi$ based on sample size $n = 20$ over 1,000 replications.

Straightforward calculations give explicit forms for $\Gamma_n(f_\phi)^{-1}$ and $C_n(f_\phi^{-1})$

$$\Gamma_n(f_\phi)^{-1} = \begin{pmatrix} 1 & -\phi & 0 & 0 & \dots & 1 \\ -\phi & 1+\phi^2 & -\phi & 0 & \dots & 0 \\ 0 & -\phi & 1+\phi^2 & -\phi & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix},$$

$$C_n(f_\phi^{-1}) = \begin{pmatrix} 1+\phi^2 & -\phi & 0 & 0 & \dots & -\phi \\ -\phi & 1+\phi^2 & \phi & 0 & \dots & 0 \\ 0 & -\phi & 1+\phi^2 & -\phi & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\phi & 0 & 0 & 0 & \dots & 1+\phi^2 \end{pmatrix}. \tag{2.25}$$

Therefore, based on (2.25), the difference between the Gaussian and Whittle likelihoods for an

AR$(1)$ model is

$$\mathcal{L}_n(\phi) - K_n(\phi) = n^{-1}\left[2\phi X_1 X_n - \phi^2(X_1^2 + X_n^2)\right] \tag{2.26}$$

Thus we observe that the closer $|\phi|$ is to one, the larger the expected difference between the likelihoods. Using (2.26) and the Bartlett correction (see Bartlett (1953) and Cox and Snell (1968)), it is possible to obtain an asymptotic expression for the difference in the biases (see also Appendix B.2). Generalization of this result to higher order AR$(p)$ models may also be possible using the analytic expression for the inverse of the Toeplitz matrix corresponding to an AR$(p)$ model derived in Siddiqui (1958) and Galbraith and Galbraith (1974).

However, for more general models, such as the MA$(q)$ or ARMA$(p, q)$ models, using brute force calculations for deriving the difference $\mathcal{L}_n(\theta) - K_n(\theta)$ and its derivatives is extremely difficult. Furthermore, such results do not offer any insight on how the Gaussian and Whittle likelihood are related, nor what is "lost" when going from the Gaussian likelihood to the Whittle likelihood. Therefore, we use a different approach to obtain an inverse Toeplitz matrix representation. The key is on the matrix representation of the biorthogornal transform theorem (Theorem 2.2.1, (2.2)). The benefit of biorthogonality between $U_n = F_n + D_n(f)$ and $F_n$ is that it leads to the following simple identity on the inverse of the variance matrix.

**Lemma 2.4.1.** *Suppose that $U_n$ and $V_n$ are biorthogonal matrices with respect to the variance matrix* $\mathrm{var}(\underline{X}_n)$*, such that* $\mathrm{cov}(U_n\underline{X}_n, V_n\underline{X}_n) = \Delta_n$*, where $\Delta_n$ is an invertible diagonal matrix. Then, we have the representation*

$$\mathrm{var}(\underline{X}_n)^{-1} = V_n^*\Delta_n^{-1}U_n. \tag{2.27}$$

PROOF. We first note that $\mathrm{cov}(U_n\underline{X}_n, V_n\underline{X}_n) = U_n\mathrm{var}(\underline{X}_n)V_n^* = \Delta_n$. By taking determinant on both side, we have $|U_n||\mathrm{var}(\underline{X}_n)||V_n| = |D_n| \neq 0$. Therefore, $|U_n|, |V_n| \neq 0$, i.e., $U_n$ and $V_n$ are invertible. (2.27) follows immediately from inverting the identity $\mathrm{var}(\underline{X}_n) = U_n^{-1}\Delta_n(V_n^*)^{-1}$. □

Therefore, using Lemma 2.4.1 together with (2.2), we obtain representation of the inverse Toeplitz matrix.

**Corollary 2.4.1** (Inverse Toeplitz identity)**.** *Let $\Gamma_n(f)$ denote an $n \times n$ Toeplitz matrix generated by the spectral density $f$. Then equations (2.27) and (2.2) yield the following identity*

$$\Gamma_n(f)^{-1} = F_n^* \Delta_n(f^{-1})(F_n + D_n(f)), \tag{2.28}$$

*where $D_n(f)$ is defined in (2.3). Observe that two spectral density functions $f_1(\omega)$ and $f_2(\omega)$ with the same autocovariance up to lag $(n-1)$, $\{c(r)\}_{r=0}^{n-1}$, can give rise to two different representations*

$$\Gamma_n(f_1)^{-1} = F_n^* \Delta_n(f_1^{-1})(F_n + D_n(f_1)) = F_n^* \Delta_n(f_2^{-1})(F_n + D_n(f_2)) = \Gamma_n(f_2)^{-1}.$$

In the following theorem, we exploit the biorthogonality between the regular DFT and the complete DFT to yield an *exact* "frequency domain" representation for the Gaussian likelihood. We use the notation defined in Theorem 2.2.1.

**Theorem 2.4.1** (A frequency domain representation of the Gaussian likelihood)**.** *Suppose the spectral density $f_\theta$ is bounded away from zero, and the corresponding autocovariance is such that $\sum_r |r c_{f_\theta}(r)| < \infty$. Let $\widehat{J}_n(\omega_{k,n}; f_\theta)$ be the predictive DFT defined as in (2.5) but replacing $f$ with $f_\theta$ and $\widetilde{J}_n(\omega_{k,n}; f_\theta) = J_n(\omega_{k,n}) + \widehat{J}_n(\omega_{k,n}; f_\theta)$ be the complete DFT. Then, the Gaussian likelihood has the following representation*

$$\mathcal{L}_n(\theta) = \frac{1}{n} \underline{X}_n' \Gamma_n(f_\theta)^{-1} \underline{X}_n = \frac{1}{n} \sum_{k=1}^{n} \frac{\widetilde{J}_n(\omega_{k,n}; f_\theta) \overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})}. \tag{2.29}$$

*Further*

$$\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1}) = F_n^* \Delta_n(f_\theta^{-1}) D_n(f_\theta), \tag{2.30}$$

*where $D_n(f_\theta)$ is defined as in (2.3) but replacing $f$ with $f_\theta$. This yields the difference between the*

*Gaussian and Whittle likelihood*

$$\begin{aligned} \mathcal{L}_n(\theta) - K_n(\theta) &= n^{-1}\underline{X}_n' \left[\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1})\right] \underline{X}_n \\ &= \frac{1}{n}\sum_{k=1}^{n} \frac{\widehat{J}_n(\omega_{k,n}; f_\theta)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})}. \end{aligned} \tag{2.31}$$

PROOF. (2.30) follows immediately from Corollary 2.4.1. Next, we note that $F_n\underline{X}_n = \underline{J}_n$ and $(F_n + D_n(f_\theta))\underline{X}_n = \underline{\widetilde{J}}_n$, thus we immediately obtain equation (2.29), and since $\widetilde{J}_n(\omega_{k,n}; f_\theta) = J_n(\omega_{k,n}) + \widehat{J}_n(\omega_{k,n}; f_\theta)$, it proves (2.31). □

From the above theorem, we observe that the Gaussian likelihood is the Whittle likelihood plus an additional "correction"

$$\mathcal{L}_n(\theta) = \underbrace{\frac{1}{n}\sum_{k=1}^{n} \frac{|J_n(\omega_{k,n})|^2}{f_\theta(\omega_{k,n})}}_{=K_n(\theta)} + \frac{1}{n}\sum_{k=1}^{n} \frac{\widehat{J}_n(\omega_{k,n}; f_\theta)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})}.$$

To summarize, the Gaussian likelihood compensates for the well known boundary effect in the Whittle likelihood, by predicting outside the domain of observation. The Whittle likelihood estimator selects the spectral density $f_\theta$ which best fits the periodogram. On the other hand, since $\mathbb{E}_{f_\theta}[\widetilde{J}_n(\omega_{k,n}; f_\theta)\overline{J_n(\omega_{k,n})}] = f_\theta(\omega_{k,n})$, the Gaussian likelihood estimator selects the spectral density which best fits $\widetilde{J}_n(\omega_{k,n}; f_\theta)\overline{J_n(\omega_{k,n})}$ by simultaneously predicting and fitting. Therefore, the "larger" the level of "persistence" in the time series, the greater the predictive DFT $\widehat{J}_n(\omega_{k,n}; f_\theta)$, and subsequently the larger the approximation error between the two likelihoods. This fits with the insights of Dahlhaus (1988), who shows that the more peaked the spectral density the greater the leakage effect in the Whittle likelihood, leading to a large finite sample bias.

By using Theorem 2.4.1, we have

$$\mathcal{L}_n(\theta) - K_n(\theta) = \frac{1}{n}\sum_{k=1}^{n} \frac{\widehat{J}_n(\omega_{k,n}; f_\theta)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} = n^{-1}\underline{X}_n' F_n^* \Delta_n(f_\theta^{-1}) D_n(f_\theta)\underline{X}_n,$$

where the entries of $F_n^* \Delta_n(f_\theta^{-1}) D_n(f_\theta)$ are

$$
\begin{aligned}
&(F_n^* \Delta_n(f_\theta^{-1}) D_n(f_\theta))_{s,t} \\
&\quad = \sum_{\tau \leqslant 0} \left[ \phi_{t,n}(\tau; f_\theta) G_{1,n}(s, \tau; f_\theta) + \phi_{n+1-t,n}(\tau; f_\theta) G_{2,n}(s, \tau; f_\theta) \right] \qquad (2.32)
\end{aligned}
$$

with

$$
\begin{aligned}
G_{1,n}(s, \tau; f_\theta) &= \frac{1}{n} \sum_{k=1}^{n} \frac{1}{f_\theta(\omega_{k,n})} e^{i(\tau - s)\omega_{k,n}} = \sum_{a \in \mathbb{Z}} K_{f_\theta^{-1}}(\tau - s + an) \\
G_{2,n}(s, \tau; f_\theta) &= \frac{1}{n} \sum_{k=1}^{n} \frac{1}{f_\theta(\omega_{k,n})} e^{-i(\tau + s - 1)\omega_{k,n}} = \sum_{a \in \mathbb{Z}} K_{f_\theta^{-1}}(\tau + s - 1 + an)
\end{aligned}
$$

and $K_{f_\theta^{-1}}(r) = \int_0^{2\pi} f_\theta(\omega)^{-1} e^{ir\omega} d\omega$. We observe that for $1 << t << n$, $\phi_{t,n}(\tau; f_\theta)$ and $\phi_{n+1-t,n}(\tau; f_\theta)$ will be "small" as compared with $t$ close to one or $n$. The same is true for $G_{1,n}(s, \tau; f_\theta)$ and $G_{2,n}(s, \tau; f_\theta)$ when $1 << s << n$. Thus the entries of $F_n^* \Delta_n(f_\theta^{-1}) D_n(f_\theta)$ will be "small" far from the four corners of the matrix. In contrast, the entries of $F_n^* \Delta_n(f_\theta^{-1}) D_n(f_\theta)$ will be largest at the four corners at the matrix. This can be clearly seen for AR$(1)$ model in (2.25). Moreover, in the following theorem we generalize our observation to the case of AR$(p)$ models.

**Theorem 2.4.2** (Finite order autoregressive models)**.** *Suppose that* $f_\theta(\omega) = \sigma^2 |\phi_p(\omega)|^{-2}$ *where* $\phi_p(\omega) = 1 - \sum_{u=1}^{p} \phi_u e^{-iu\omega}$ *(the roots of the corresponding characteristic polynomial lie outside the unit circle) and* $p < n$*. Then, by Theorem 2.2.2, the predictive DFT has the analytic form*

$$
\widehat{J}_n(\omega_{k,n}; f_\theta) = \frac{n^{-1/2}}{\phi_p(\omega_{k,n})} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{-is\omega_{k,n}} + \frac{n^{-1/2}}{\overline{\phi_p(\omega_{k,n})}} \sum_{\ell=1}^{p} X_{n+1-\ell} \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{i(s+1)\omega_{k,n}}.
$$

*If $p \leqslant n/2$, then $D_n(f_\theta)$ is a rank $2p$ matrix where*

$$D_n(f_\theta) = n^{-1/2} \begin{pmatrix} \phi_{1,p}(\omega_{1,n}) & \dots & \phi_{p,p}(\omega_{1,n}) & 0 & \dots & 0 & e^{i\omega_{1,n}}\overline{\phi_{p,p}(\omega_{1,n})} & \dots & e^{i\omega_{1,n}}\overline{\phi_{1,p}(\omega_{1,n})} \\ \phi_{1,p}(\omega_{2,n}) & \dots & \phi_{p,p}(\omega_{2,n}) & 0 & \dots & 0 & e^{i\omega_{2,n}}\overline{\phi_{p,p}(\omega_{2,n})} & \dots & e^{i\omega_{2,n}}\overline{\phi_{1,p}(\omega_{2,n})} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \phi_{1,p}(\omega_{n,n}) & \dots & \phi_{p,p}(\omega_{n,n}) & 0 & \dots & 0 & e^{i\omega_{n,n}}\overline{\phi_{p,p}(\omega_{n,n})} & \dots & e^{i\omega_{n,n}}\overline{\phi_{1,p}(\omega_{n,n})} \end{pmatrix} \tag{2.33}$$

*and $\phi_{j,p}(\omega) = \phi_p(\omega)^{-1} \sum_{s=0}^{p-j} \phi_{j+s} e^{-is\omega}$. Note, if $n/2 < p \leqslant n$, then the entries of $D_n(f_\theta)$ will overlap. Let $\widetilde{\phi}_0 = 1$ and for $1 \leqslant s \leqslant p$, $\widetilde{\phi}_s = -\phi_s$ (zero otherwise), then if $1 \leqslant p \leqslant n/2$ we have*

$$\begin{aligned} \left(\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1})\right)_{s,t} &= (F_n^* \Delta_n(f_\theta^{-1}) D_n(f_\theta))_{s,t} \\ &= \begin{cases} \sigma^{-2} \sum_{\ell=0}^{p-t} \phi_{\ell+t} \widetilde{\phi}_{(\ell+s) \bmod n} & 1 \leqslant t \leqslant p \\ \sigma^{-2} \sum_{\ell=1}^{p-(n-t)} \phi_{\ell+(n-t)} \widetilde{\phi}_{(\ell-s) \bmod n} & n-p+1 \leqslant t \leqslant n \\ 0 & otherwise \end{cases}. \end{aligned} \tag{2.34}$$

PROOF. See Section 4.7.1. □

Theorem 2.4.2 shows that for AR($p$) models, the predictive DFT only involves the $p$ observations on each side of the observational boundary $X_1, \dots, X_p$ and $X_{n-p+1}, \dots, X_n$, where the coefficients in the prediction are a linear combination of the AR parameters (excluding the denominator $\phi_p(\omega)$). The well known result (see Siddiqui (1958) and Shaman (1975), equation (10)) that $F_n^* \Delta_n(f_\theta^{-1}) D_n(f_\theta)$ is non-zero only at the $(p \times p)$ submatrices located in the four corners of $F_n^* \Delta_n(f_\theta^{-1}) D_n(f_\theta)$ follows from equation (2.34).

By using (2.8) we obtain an analytic expression for the Gaussian likelihood of the AR($p$) model in terms of the autoregressive coefficients. In particular, the Gaussian likelihood (written in

the frequency domain) corresponding to the AR($p$) model $X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t$ is

$$\begin{aligned}\mathcal{L}_n(\phi) = & \frac{\sigma^{-2}}{n} \sum_{k=1}^{n} |J_n(\omega_{k,n})|^2 |\phi_p(\omega_{k,n})|^2 \\ & + \frac{\sigma^{-2}}{n} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s} \left( X_{(-s) \bmod n} - \sum_{j=1}^{p} \phi_j X_{(j-s) \bmod n} \right) \\ & + \frac{\sigma^{-2}}{n} \sum_{\ell=1}^{p} X_{n+1-\ell} \sum_{s=0}^{p-\ell} \phi_{\ell+s} \left( X_{(s+1) \bmod n} - \sum_{j=1}^{p} \phi_j X_{(s+1-j) \bmod n} \right), \end{aligned} \tag{2.35}$$

where $\phi = (\phi_1, ..., \phi_p)'$ and $\phi_p(\omega) = 1 - \sum_{j=1}^{p} \phi_j e^{-ij\omega}$. A proof of the above identity can be found in Section 4.7.1. Equation (2.35) offers a simple representation of the Gaussian likelihood in terms of a Whittle likelihood plus an additional term in terms of the AR($p$) coefficients.

### 2.4.1 Approximation of the Gaussian likelihood in the frequency domain

In this section, we obtain the approximation of $\Gamma_n(f)^{-1} - C_n(f^{-1}) = F_n^* \Delta_n(f^{-1}) D_n(f)$ for spectral density $f$. This is equivalant to obtain an approximation of $D_n(f)$.

In Theorem 2.2.1, we replace $\phi_{s,n}(\tau; f)$ in $D_n(f)$ with $\phi_s(\tau; f)$ which are the coefficients of the best linear predictor of $X_\tau$ (for $\tau \leqslant 0$) given infinite future of time series $\{X_t\}_{t=1}^{\infty}$ i.e. $\widehat{X}_\tau = \sum_{t=1}^{\infty} \phi_t(\tau; f) X_t$. This gives the approximation $D_{\infty,n}(f)$, where

$$(D_{\infty,n}(f))_{k,t} = n^{-1/2} \sum_{\tau \leqslant 0} \left( \phi_t(\tau; f) e^{i\tau\omega_{k,n}} + \phi_{n+1-t}(\tau; f) e^{-i(\tau-1)\omega_{k,n}} \right).$$

One advantage of this approximation is that the infinite prediction coefficients $\phi_s(\tau; f)$ (for $\tau \leqslant 0$) admits a simple convolution-like expression

$$\phi_j(\tau; f) = \sum_{j=0}^{|\tau|} \phi_{s+j}(f) \psi_{|\tau|-j}(f) \qquad \tau \leqslant 0, \tag{2.36}$$

where $\{\phi_j(f)\}_{j \geqslant 1}$ and $\{\psi_j(f)\}_{j \geqslant 1}$ are AR($\infty$) and MA($\infty$) coefficients of $\{X_t\}$ (with underlying spectral density $f$) respectively. By convention, we set $\phi_0(f) = \psi_0(f) = 1$.

Using (2.36), it can be shown that for $1 \leqslant k, t \leqslant n$,

$$(D_{\infty,n}(f))_{k,t} = n^{-1/2}\frac{\phi_t^{\infty}(\omega_{k,n};f)}{\phi(\omega_{k,n};f)} + n^{-1/2}e^{i\omega_{k,n}}\frac{\overline{\phi_{n+1-t}^{\infty}(\omega_{k,n};f)}}{\overline{\phi(\omega_{k,n};f)}}, \tag{2.37}$$

where $\phi_t^{\infty}(\omega;f) = \sum_{s=0}^{\infty}\phi_{t+s}(f)e^{-is\omega}$. The proof of the above identity can be found in Section 4.7.1. Using the above we can show that $(D_{\infty,n}(f)\underline{X}_n)_k = \widehat{J}_{\infty,n}(\omega_{k,n};f)$ where

$$\widehat{J}_{\infty,n}(\omega;f) = \frac{n^{-1/2}}{\phi(\omega;f)}\sum_{t=1}^{n}X_t\phi_t^{\infty}(\omega;f) + e^{i(n+1)\omega}\frac{n^{-1/2}}{\overline{\phi(\omega;f)}}\sum_{t=1}^{n}X_{n+1-t}\overline{\phi_t^{\infty}(\omega;f)}. \tag{2.38}$$

It is not surprising that the expression of $\widehat{J}_{\infty,n}(\omega;f)$ in (2.38) is identical to the first identity of (2.10). This is because both constructions are based on the infinite order AR representation of the stationary time series.

We show below that $\widehat{J}_{\infty,n}(\omega_{k,n};f)$ is an approximation of $\widehat{J}_n(\omega_{k,n};f)$.

**Theorem 2.4.3** (An AR($\infty$) approximation for general processes)**.** *Suppose $f$ satisfies Assumption 2.3.1 and $f_\theta$ is bounded away from zero and $\|f_\theta\|_0 < \infty$ (with $f_\theta(\omega) = \sigma_\theta^2|\phi_\theta(\omega)|^{-2}$). Let $D_n(f)$, $D_{\infty,n}(f)$ and $\widehat{J}_{\infty,n}(\omega_{k,n};f)$ be defined as in (2.3) and (2.37) and (2.38) respectively. Then we have*

$$\begin{aligned}&\underline{X}_n'F_n^*\Delta_n(f_\theta^{-1})\left(D_n(f) - D_{\infty,n}(f)\right)\underline{X}_n \\ &\quad= \sum_{k=1}^{n}\frac{\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})}\big(\widehat{J}_n(\omega_{k,n};f) - \widehat{J}_{\infty,n}(\omega_{k,n};f)\big)\end{aligned} \tag{2.39}$$

*and*

$$\left\|F_n^*\Delta_n(f_\theta^{-1})\left(D_n(f) - D_{\infty,n}(f)\right)\right\|_1 \leqslant \frac{C_{f,0}\rho_{n,K}(f)}{n^{K-1}}A_K(f,f_\theta). \tag{2.40}$$

*Further, if $\{X_t\}$ is a time series where $\sup_t\|X_t\|_{\mathbb{E},2q} = \|X\|_{\mathbb{E},2q} < \infty$ (for some $q > 1$), then*

$$\begin{aligned}&n^{-1}\left\|\underline{X}_n'F_n^*\Delta_n(f_\theta^{-1})\left(D_n(f) - D_{\infty,n}(f)\right)\underline{X}_n\right\|_{\mathbb{E},q} \\ &\quad\leqslant \frac{C_{f,0}\rho_{n,K}(f)}{n^K}A_K(f,f_\theta)\|X\|_{\mathbb{E},2q}^2.\end{aligned} \tag{2.41}$$

PROOF. See Section 4.7.1. □

We state the above theorem is the general case that the spectral density $f$ is used to construct the predictors $D_n(f)$. It does not necessarily have to be the same as $f_\theta$. This is to allow generalizations of the Whittle and Gaussian likelihoods, which we discuss in Section 4.1.

Applying the above theorem to the Gaussian likelihood gives an approximation which is analogous to (2.35)

$$\begin{aligned}\mathcal{L}_n(\theta) &= K_n(\theta) + \frac{1}{n}\sum_{k=1}^{n}\frac{\widehat{J}_{\infty,n}(\omega_{k,n};f_\theta)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} + O_p(n^{-K}) \\ &= K_n(\theta) + \frac{1}{n}\sum_{s,t=1}^{n} X_s X_t \frac{1}{n}\sum_{k=1}^{n} e^{-is\omega_{k,n}}\varphi_{t,n}(\omega_{k,n};f_\theta) + O_p(n^{-K}), \qquad (2.42)\end{aligned}$$

where $\varphi_{t,n}(\omega;f_\theta) = \sigma^{-2}\left[\overline{\phi(\omega;f_\theta)}\phi_t^\infty(\omega;f_\theta) + e^{i\omega}\phi(\omega;f_\theta)\overline{\phi_{n+1-t}^\infty(\omega;f_\theta)}\right]$. The above approximation shows that if the autocovariance function, corresponding to $f_\theta$ decays sufficiently fast (in the sense that $\sum_{r\in\mathbb{Z}}|r^K c_{f_\theta}(r)| < \infty$ for some $K > 1$). Then replacing the finite predictions with the predictors using the infinite past (or future) gives a close approximation of the Gaussian likelihood.

**Remark 2.4.1.** *Following from the above, the entrywise difference between the two matrices is approximately*

$$(\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1}))_{s,t} \approx (F_n^*\Delta_n(f_\theta^{-1})D_{\infty,n}(f_\theta))_{s,t} = \frac{1}{n}\sum_{k=1}^{n} e^{-is\omega_{k,n}}\varphi_{t,n}(\omega_{k,n};f_\theta),$$

*thus giving an analytic approximation to (2.32).*

We conclude this section by obtaining a bound between the Gaussian and Whittle likelihood.

**Theorem 2.4.4** (The difference in the likelihoods)**.** *Suppose $f_\theta$ satisfies Assumption 2.3.1. Let $D_n(f_\theta)$ and $D_{\infty,n}(f_\theta)$ be defined as in (2.3) and (2.37) respectively. Then we have*

$$\left\|F_n^*\Delta_n(f_\theta^{-1})D_{\infty,n}(f_\theta)\right\|_1 \leqslant A_1(f_\theta, f_\theta) \qquad (2.43)$$

*and*

$$\left\|\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1})\right\|_1 \leqslant \left(A_1(f_\theta, f_\theta) + \frac{C_{f_\theta,0}\rho_{n,K}(f_\theta)}{n^{K-1}} A_K(f_\theta, f_\theta)\right). \tag{2.44}$$

*Further, if* $\{X_t\}$ *is a time series where* $\sup_t \|X_t\|_{\mathbb{E},2q} = \|X\|_{\mathbb{E},2q} < \infty$ *(for some* $q > 1$*), then*

$$\|\mathcal{L}_n(\theta) - K_n(\theta)\|_{\mathbb{E},q} \leqslant n^{-1}\left(A_1(f_\theta, f_\theta) + \frac{C_{f_\theta,0}\rho_{n,K}(f_\theta)}{n^{K-1}} A_K(f_\theta, f_\theta)\right)\|X\|^2_{\mathbb{E},2q}. \tag{2.45}$$

PROOF. See Section 4.7.1. □

The above result shows that under the stated conditions

$$n^{-1}\left\|\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1})\right\|_1 = O(n^{-1}),$$

and the difference between the Whittle and Gaussian likelihoods is of order $O(n^{-1})$.

# 3. THE COMPLETE PERIODOGRAM *

In this section, we discuss greater detail of the complete periodogram in Section 2.3.

## 3.1 The tapered complete periodogram

We recall from Section 2.3 that the complete periodogram extends the "domain" of observation by predicting across the boundary for one of the DFTs, but keeping the other DFT the same. Our simulations suggest that a further improvement can be made by "softening" the boundary of the regular DFT by using a data taper. Unusually, unlike the classical data taper, we only taper the regular DFT, but keep the complete DFT as is. Precisely we define the tapered complete periodogram as

$$I_{\underline{h},n}(\omega; f) = \widetilde{J}_n(\omega; f)\overline{J_{\underline{h},n}(\omega)}, \qquad J_{\underline{h},n}(\omega) = n^{-1/2}\sum_{t=1}^{n} h_{t,n}X_t e^{it\omega}$$

and $\underline{h} = \{h_{t,n}\}_{t=1}^n$ are positive weights. Again by using that $\operatorname{cov}(\widehat{X}_{\tau,n}, X_t) = c(t-\tau)$ for $1 \leqslant t \leqslant n$ and $\tau \in \mathbb{Z}$ it is straightforward to show that

$$\mathbb{E}[I_{\underline{h},n}(\omega; f)] = \big(n^{-1}\sum_{t=1}^{n} h_{t,n}\big) \cdot f(\omega) \qquad \omega \in [0, 2\pi].$$

Thus to ensure that $I_{\underline{h},n}(\omega; f)$ is an unbiased estimator of $f$, we constrain the tapered weights to be such that $\sum_{t=1}^n h_{t,n} = n$. Unlike the regular tapered periodogram, for any choice of $\{h_{t,n}\}$ (under the constraint $\sum_{t=1}^n h_{t,n} = n$), $I_{\underline{h},n}(\omega; f)$ will be an unbiased estimator of (no smoothness assumptions on the taper is required). But it seems reasonable to use standard tapers when defining $\{h_{t,n}\}$. In particular, to let

$$h_{t,n} = c_n h_n(t/n)$$

*Parts of this section have been modified with permission from [S. Das, S. Subba Rao, and J. Yang. Spectral methods for small sample time series: A complete periodogram approach. *Journal of Time Series Analysis (To appear)*, 2021, https://doi.org/10.1111/jtsa.12584.]

where $c_n = n/H_{1,n}$ and

$$H_{q,n} = \sum_{t=1}^{n} h_n(t/n)^q, \qquad q \geqslant 1. \tag{3.1}$$

A commonly used taper is the Tukey (also called the cosine-bell) taper, where

$$h_n\left(\frac{t}{n}\right) = \begin{cases} \frac{1}{2}[1-\cos(\pi(t-\frac{1}{2})/d)] & 1 \leqslant t \leqslant d \\ 1 & d+1 \leqslant t \leqslant n-d \\ \frac{1}{2}[1-\cos(\pi(n-t+\frac{1}{2})/d)] & n-d+1 \leqslant t \leqslant n \end{cases}. \tag{3.2}$$

Since we do not observe the spectral density $f$, we use the estimated tapered complete periodogram

$$I_{\underline{h},n}(\omega; \widehat{f}_p) = \widetilde{J}_n(\omega; \widehat{f}_p)\overline{J_{\underline{h},n}(\omega)} \tag{3.3}$$

where $\widetilde{J}_n(\omega; \widehat{f}_p) = J_n(\omega) + \widehat{J}_n(\omega; \widehat{f}_p)$ where $\widehat{J}_n(\omega; \widehat{f}_p)$ is defined as in (2.18). In the theorem below we obtain that the asymptotic bias of the estimated tapered complete periodogram, this result is analogous to the non-tapered result in Theorem 2.3.4 (Noting that the tapered complete periodogram includes the non-tapered case where we set $h_{t,n} \equiv 1$).

**Theorem 3.1.1.** *Suppose Assumptions 2.3.1 ($K > 1$) and 2.3.2 (where $m \geqslant 6$ and is a multiple of two) hold. Let $I_{\underline{h},n}(\omega; \widehat{f}_p)$ be defined as in (3.3) where $\sum_{t=1}^{n} h_{t,n} = n$ and $\sup_{t,n} |h_{t,n}| < \infty$. Then we have*

$$I_{\underline{h},n}(\omega; \widehat{f}_p) = I_{\underline{h},n}(\omega; f) + \Delta_{\underline{h}}(\omega) + O_p\left(\frac{p^{m/2}}{n^{m/4}}\right),$$

*where* $\sup_\omega \mathbb{E}[\Delta_{\underline{h}}(\omega)] = O\left((np^{K-1})^{-1} + p^3/n^2\right)$ *and* $\sup_\omega \operatorname{var}[\Delta_{\underline{h}}(\omega)] = O\left(p^4/n^2\right)$.

PROOF. See Section 3.6.3. □

Comparing Theorem 3.1.1 with Theorem 2.3.4, if the taper satistifies $\sum_{t=1}^{n} h_{t,n} = n$ and $\sup_{t,n} |h_{t,n}| < \infty$, then the tapered complete periodogram has the same order of approximation

error with the regular complete periodogram. Therefore, if $p$ is chosen using AIC, then this yields an estimated tapered complete peridogram has a lower bias than the regular periodogram.

Theoretically, it is unclear using the tapered estimated complete improves on the non-tapered estimated complete periodogram. But in the simulations, we do observe an improvement in the bias of the estimator when using (3.2) with $d = n/10$ (this will require further research). In contrast, in Section 3.2 we show that the choice of data taper does have an impact on the variance of estimators based on the complete periodogram.

### 3.2 The integrated complete periodogram

We now apply the estimated (tapered) complete periodogram to estimating parameters in a time series. Many parameters in time series can be rewritten in terms of the integrated spectral mean

$$A(g) = \frac{1}{2\pi}\int_0^{2\pi} g(\omega)f(\omega)d\omega,$$

where $g(\cdot)$ is an integrable function that determines an underlying parameter, $A(g)$. Examples of useful functions $g$ are discussed in Section 3.3.

The above representation motivates the following estimator of $A(g)$, where we replace the spectral density function $f$ with the regular periodogram, to yield the following estimators

$$A_{I,n}(g) = \frac{1}{2\pi}\int_0^{2\pi} g(\omega)I_n(\omega)d\omega \quad \text{or} \quad A_{S,n}(g) = \frac{1}{n}\sum_{k=1}^{n} g(\omega_{k,n})I_n(\omega_{k,n}), \tag{3.4}$$

of $A(g)$ where $\omega_{k,n} = \frac{2\pi k}{n}$. See, for example, Milhøj (1981); Dahlhaus and Janas (1996); Bardet et al. (2008); Eichler (2008); Niebuhr and Kreiss (2014); Mikosch and Zhao (2015) and Subba Rao (2018). However, similar to the regular periodogram, the integrated regular periodogram has an $O(n^{-1})$ bias

$$\mathbb{E}[A_{x,n}(g)] = A(g) + O(n^{-1}) \quad x \in \{I, S\}$$

which can be severe for "peaky" spectral density functions and small sample sizes. The bias in the

case that an appropriate tapered periodogram is used instead of the regular periodogram will be considerably smaller and of order $O(n^{-2})$. Ideally, we could replace the periodogram in (3.4) with the complete periodogram $I_n(\omega; f)$ this would produce an unbiased estimator. Of course, this is infeasible, since $f$ is unknown. Thus motivated by the results in Section 2.3.2, to reduce the bias in $A_{x,n}(g)$ we propose replacing $I_n(\omega)$ with the estimated complete periodogram $I_n(\omega; \widehat{f}_p)$ or the tapered complete periodogram $I_{\underline{h},n}(\omega; \widehat{f}_p)$ to yield the estimated integrated complete periodogram

$$A_{I,n}(g; \widehat{f}_p) = \int_0^{2\pi} g(\omega) I_{\underline{h},n}(\omega; \widehat{f}_p) d\omega \quad \text{and} \quad A_{S,n}(g; \widehat{f}_p) = \frac{1}{n} \sum_{k=1}^{n} g(\omega_{k,n}) I_{\underline{h},n}(\omega_{k,n}; \widehat{f}_p) \tag{3.5}$$

of $A(g)$. Note that the above formulation allows for the non-tapered complete periodogram (by setting $h_{t,n} \equiv 1$ for $1 \leqslant t \leqslant n$).

In the following theorem, we show that the (estimated) integrated complete periodogram has a bias that has lower order than the integrated regular periodogram and is asymptotically "closer" to the ideal integrated complete periodogram $A_{x,n}(g; f)$ than the integrated regular periodogram.

**Theorem 3.2.1.** *Suppose the assumptions in Theorem 3.1.1 hold. Further, suppose that the functions $g$ and its derivative are continuous on the torus $[0, 2\pi]$. For $x \in \{I, S\}$, define $A_{x,n}(g; f)$ and $A_{x,n}(g; \widehat{f}_p)$ as in (3.4) and (3.5) respectively, where $\sum_{t=1}^{n} h_{t,n} = n$ and $\sup_{t,n} |h_{t,n}| < \infty$. Then*

$$A_{x,n}(g; \widehat{f}_p) = A_{x,n}(g; f) + \Delta(g) + O_p\left(\frac{p^{m/2}}{n^{m/4}}\right)$$

*where $\mathbb{E}[\Delta(g)] = O\left((np^{K-1})^{-1} + p^3/n^2\right)$ and $\operatorname{var}[\Delta(g)] = O\left((np^{K-1})^{-2} + p^6/n^3\right)$.*

PROOF. See Section 3.6.3. □

From the above theorem we observe that if $m \geqslant 6$, then the term $\Delta(g) = O_p((np^{K-1})^{-1} + p^3/n^{3/2})$ dominates the probablistic error. This gives

$$A_{x,n}(g; \widehat{f}_p) = A_{x,n}(g; f) + O_p\left(\frac{1}{np^{K-1}} + \frac{p^3}{n^{3/2}}\right) \qquad x \in \{I, S\}.$$

Further, the bias (in the sense of Bartlett (1953)) is

$$\mathbb{E}[A_{I,n}(g;\widehat{f}_p)] = A(g) + O\left(\frac{1}{np^{K-1}} + \frac{p^3}{n^2}\right).$$

since $\mathbb{E}[A_{I,n}(g;f)] = A(g)$.

### 3.2.1 Distributional properties of $A_{x,n}(g;\widehat{f}_p)$

In this section, we study the distributional properties of the (estimated) integrated tapered complete periodogram. To do so, we evaluate an expression for the asymptotic variance of $A_{x,n}(g;\widehat{f}_p)$. We show that asymptotically the variance is same as if the predictive part of the periodogram; $\widehat{J}_n(\omega;\widehat{f}_p)\overline{J_{\underline{h},n}(\omega)}$ were not included in the definition of $I_{\underline{h},n}(\omega;\widehat{f}_p)$. To do so, we require the condition

$$\frac{H_{1,n}}{H_{2,n}^{1/2}}\left(\frac{p^3}{n^{3/2}}\right) \to 0 \quad \text{as} \quad p,n \to \infty, \tag{3.6}$$

which ensures the predictive term is negligible as compared to the main term. Observe that, by using the Cauchy-Schwarz inequality, (3.6) holds for all tapers if $p^3/n \to 0$ as $p,n \to \infty$. Therefore, by the same argument at the end of Section 2.3.2, if the order $p$ is selected using the AIC, (3.6) holds for any taper.

**Corollary 3.2.1.** *Suppose the assumptions in Theorem 3.1.1 hold. Let the data taper $\{h_{t,n}\}$ be such that $h_{t,n} = c_n h_n(t/n)$ where $c_n = n/H_{1,n}$ and $h_n : [0,1] \to \mathbb{R}$ is a sequence of taper functions which satisfy the taper conditions in Section 5, Dahlhaus (1988). For $x \in \{I,S\}$, define $A_{x,n}(g;\widehat{f}_p)$ as in (3.5) and suppose $p,n$ satisfy (3.6). Then*

$$\frac{H_{1,n}^2}{H_{2,n}}\text{var}[A_{x,n}(g;\widehat{f}_p)] = (V_1 + V_2 + V_3) + o(1)$$

*where $H_{q,n}$ is defined in (3.1),*

$$\begin{aligned} V_1 &= \frac{1}{2\pi}\int_0^{2\pi} g(\omega)\overline{g(-\omega)}f(\omega)^2 d\omega, \quad V_2 = \frac{1}{2\pi}\int_0^{2\pi} |g(\omega)|^2 f(\omega)^2 d\omega \quad \textit{and} \\ V_3 &= \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} g(\omega_1)\overline{g(\omega_2)}f_4(\omega_1, -\omega_1, \omega_2)d\omega_1 d\omega_2, \end{aligned}$$

*where $f_4$ is the fourth order cumulant spectrum.*

PROOF. See Section 3.6.1. □

From the above, we observe that when tapering is used, the asymptotic variance of $A_{x,n}(g; \widehat{f}_p)$ is $O(H_{2,n}/H_{1,n}^2)$. If $h_n \equiv h$ for all $n$ for some $h : [0,1] \to \mathbb{R}$ with bounded variation, then above rate has the limit

$$\frac{nH_{2,n}}{H_{1,n}^2} \to \frac{\int_0^1 h(x)^2 dx}{\left(\int_0^1 h(x)dx\right)^2} \geqslant 1.$$

In general, to understand how it compares to the case where no tapering is used, we note that by the Cauchy-Schwarz inequality $H_{2,n}/H_{1,n}^2 \geqslant n^{-1}$, where we attain equality $H_{2,n}/H_{1,n}^2 = n^{-1}$ if and only if no tapering is used. Thus, typically the integrated tapered complete periodogram will be less efficient than the integrated (non-tapered) complete periodogram. However if $nH_{2,n}/H_{1,n}^2 \to 1$ as $n \to \infty$, then using the tapered complete periodogram in the estimator leads to an estimator that is asymptotically as efficient as the tapered complete periodogram (and regular periodogram).

**Remark 3.2.1** (Distributional properties of $A_{x,n}(g; \widehat{f}_p)$)**.** *By using Theorems 3.2.1 and Corollary 3.2.1 $A_{x,n}(g; \widehat{f}_p)$, $A_{x,n}(g; f)$ and $A_{x,\underline{h}}(g)$ (where $A_{x,\underline{h}}(g)$ is defined as in (3.4) but with $I_{\underline{h},n}(\omega)$ replacing $I_n(\omega)$) share the same asymptotic distributional properties. In particular, if (3.6) holds, then the asymptotic distributions $A_{x,n}(g; \widehat{f}_p)$ and $A_{x,\underline{h}}(g)$ are equivalent. Thus if asymptotic normality of $A_{x,\underline{h}}(g)$ can be shown, then $A_{x,n}(g; \widehat{f}_p)$ is also asymptotically normal with the same limiting variance (given in Corollary 3.2.1).*

### 3.3 Examples of the integrated complete periodogram

In this section, we apply the integrated complete periodogram to estimating various parameters.

#### 3.3.1 Autocovariance estimation

By Bochner's theorem, the autocovariance function at lag $r$, $c(r)$, can be represented as

$$c(r) = A\left(\cos(r\cdot)\right) = \frac{1}{2\pi}\int_0^{2\pi} \cos(r\omega) f(\omega) d\omega.$$

In order to estimate $\{c(r)\}$, we replace $f$ with the integrated complete periodogram to yield the estimator

$$\widehat{c}_n(r; \widehat{f}_p) = A_{I,n}(\cos(r\cdot); \widehat{f}_p) \ = \frac{1}{2\pi}\int_0^{2\pi} \cos(r\omega) I_{\underline{h},n}(\omega; \widehat{f}_p) d\omega.$$

$I_{\underline{h},n}(\omega; \widehat{f}_p)$ can be negative, in such situations, the sample autocovariance is not necessarily positive definite. To ensure a positive definiteness, we threshold the complete periodogram to be greater than a small cutoff value $\delta > 0$. This results in a sample autocovariance $\{\widehat{c}_{T,n}(r; \widehat{f}_p)\}$ which is guaranteed to be positive definite, where

$$\widehat{c}_{T,n}(r; \widehat{f}_p) = \frac{1}{2\pi}\int_0^{2\pi} \cos(r\omega) \max\{I_{\underline{h},n}(\omega; \widehat{f}_p), \delta\} d\omega.$$

This method is illustrated with simulations in Section 3.4.2.

#### 3.3.2 Spectral density estimation

Typically, to estimate the spectral density one "smooths" the periodogram using the spectral window function. The same method can be applied to the complete periodogram. Let $W$ be a non-negative symmetric function where $\int W(u)du = 2\pi$ and $\int W(u)^2 du < \infty$. Define $W_h(\cdot) = (1/h)W(\cdot/h)$, where $h$ is a bandwidth. A review of different spectral windows and their properties can be found in Priestley (1981) and Section 10.4 of Brockwell and Davis (2006) and references therein. For $\lambda \in [0, \pi]$, we choose $g(\omega) = g_\lambda(\omega) = W_h(\lambda - \omega)$. Then the (estimated) integrated

complete periodogram of the spectral density $f$ is

$$\widehat{f}_n(\lambda;\widehat{f}_p) = A_{I,n}(g_\lambda;\widehat{f}_p) = \frac{1}{2\pi}\int_0^{2\pi} W_h(\lambda-\omega) I_{\underline{h},n}(\omega;\widehat{f}_p)d\omega.$$

The method is illustrated with simulations in Section 3.4.3.

### 3.3.3 Whittle likelihood

Suppose that $\mathcal{F} = \{f_\theta(\cdot) : \theta \in \Theta\}$ for some compact $\Theta \in \mathbb{R}^d$ is a parametric family of spectral density functions. The parameter which minimizes the Whittle likelihood is used as an estimator of the spectral density. Replacing the periodogram with the complete periodogram we define a variant of the Whittle likelihood as

$$\begin{aligned} K_n(\theta) &= \frac{1}{2\pi}\int_0^{2\pi}\left(\frac{I_{\underline{h},n}(\omega;\widehat{f}_p)}{f_\theta(\omega)} + \log f_\theta(\omega)\right)d\omega \\ &= A_{I,n}(f_\theta^{-1};\widehat{f}_p) + \frac{1}{2\pi}\int_0^{2\pi}\log f_\theta(\omega)d\omega. \end{aligned}$$

In Section 2.4, we show that using the non-tapered DFT $A_{S,n}(f_\theta^{-1};f_\theta) = \underline{X}_n'\Gamma(f_\theta)^{-1}\underline{X}_n$ where $\underline{X}_n' = (X_1,\ldots,X_n)$ and $\Gamma(f_\theta)$ is the Toeplitz matrix corresponding to the spectral density $f_\theta$. $K_n(\theta)$ is a variant of the frequency domain quasi-likelihoods descirbe in Section 4.1. We mention that there aren't any general theoretical guarantees that the bias corresponding to estimators based on $K_n(\theta)$ is lower than the bias of the Whittle likelihood (though simulations suggest this is usually the case). Expression for the asymptotic bias of $K_n(\theta)$ are given in Appendix B and the method is illustrated with simulations in Section 4.6 (and Appendix C).

## 3.4 Simulations

To understand the utility of the proposed methods, we now present some simulations. For reasons of space, we focus on the Gaussian time series (noting that the methods also apply to non-Gaussian time series). In the simulations we use the following AR$(2)$ and ARMA$(3,2)$ models (we let $B$ denote the backshift operator)

**(M1)** $\phi(B)X_t = \varepsilon_t$ with $\phi(z) = (1 - \lambda e^{\frac{\pi}{2}i}z)(1 - \lambda e^{-\frac{\pi}{2}i}z)$ for $\lambda \in \{0.7, 0.9, 0.95\}$.

**(M2)** $\phi(B)X_t = \psi(B)\varepsilon_t$ with $\begin{cases} \phi(z) = (1 - 0.7z)(1 - 0.9e^{i}z)(1 - 0.9e^{-i}z) \\ \psi(z) = 1 + 0.5z + 0.5z^2 \end{cases}$.

where $\mathbb{E}[\varepsilon_t] = 0$ and $\mathrm{var}[\varepsilon_t] = 1$. We observe that the peak of the spectral density for the AR$(2)$ model (M1) becomes more pronounced as $\lambda$ approaches one (at frequency $\pi/2$). The ARMA$(3, 2)$ model (M2) has peaks at zero and $\pi/2$, further, it clearly does not have a finite order autoregressive representation.

We consider three different sample sizes: $n = 20$ (extremely small), 50 (small), and 300 (large) to understand how the proposed methods perform over different sample sizes. All simulations are conducted at over $B = 5,000$ replications.

Our focus will be on accessing the validity of our method in terms of bias, standard deviation, and mean squared error. We will compare (a) various periodograms; (b) the spectral density estimators based on smoothing the various periodograms; and the autocorrelation function based on the various periodograms. The periodograms we will consider are (i) the regular periodogram (ii) the tapered periodogram $I_{\underline{h},n}(\omega)$, where

$$I_{\underline{h},n}(\omega) = \left|H_{2,n}^{-1/2}\sum_{t=1}^{n} h_n\left(t/n\right) X_t e^{it\omega}\right|^2,$$

$H_{2,n}$ is defined in (3.1), (iii) the estimated complete periodogram (2.19) and (iv) the tapered complete periodogram (3.3). To understand the impact estimation has on the complete periodogram, for a model (M1) we also evaluate the complete periodogram using the *true* AR$(2)$ parameters, as this is an AR$(2)$ model the complete periodogram has an analytic form in terms of the AR parameters. This allows us to compare the infeasible complete periodogram $I_n(\omega; f)$ with the feasible estimated complete periodogram $I_n(\omega, \widehat{f}_p)$.

For the tapered periodogram and tapered complete periodogram, we use the Tukey taper defined in (3.2). Following Tukey's rule of thumb, we set the level of tapering to $10\%$ (which corresponds to $d = n/10$). When evaluating the estimated complete and tapered complete periodogram,

we select the order $p$ using the AIC, and we estimate the AR coefficients using the Yule-Walker estimator.

For both the complete and tapered complete periodogram, it is possible to have an estimator that is complex and/or the real part is negative. In the simulations, we found that a negative $\mathrm{Re}\, I_n(\omega_{k,n}; \widehat{f}_p)$ tends to happen more for the spectral densities with large peaks and the true spectral density is close to zero. To avoid such issues, for each frequency, we take the real part of the estimator and thresholding with a small positive value. In practice, we take the threshold value $\delta = 10^{-3}$. Thresholding induces a small bias in the estimator, but, at least in our models, the effect is negligible (see the middle column in Figures 3.1$-$3.3).

### 3.4.1 Comparing the different periodograms

In this section, we compare the bias and variance of the various periodograms for models (M1) and (M2).

Figures 3.1$-$3.3 give the average (left panels), bias (middle panels), and standard deviation (right panels) of the various periodograms for the different models and samples sizes. The dashed line in each panel is the true spectral density. It is well known that $\mathrm{var}[I_n(\omega)] \approx f(\omega)^2$ for $0 < \omega < \pi$ and $\mathrm{var}[I_n(\omega)] \approx 2f(\omega)^2$ for $\omega = 0, \pi$. Therefore, for a fair comparision in the standard deviation plot for the true spectral density we replace $\sqrt{2}f(0)$ and $\sqrt{2}f(\pi)$ with $f(0)$ and $f(\pi)$ respectively.

In Figures 3.1$-$3.3 (left and middle panels), we observe that in general, the various complete periodograms give a smaller bias than the regular periodogram and the tapered periodogram. This corroborates our theoretical findings that that complete periodogram smaller bias than the $O(n^{-1})$ rate. As expected, we observe that the true (based on the true AR parameters) complete periodogram (red) has a smaller bias than the estimated complete (orange) and tapered complete periodograms (green). Such an improvement is most pronounced near the peak of the spectral density and it is most clear when the sample size $n$ is small. For example, in Figure 3.1, when the sample size is extremely small ($n = 20$), the bias of the various complete periodograms reduce by more than a half the bias of the regular and tapered periodogram. As expected, the true complete

periodogram (red) for (M1) has very little bias even for the sample size $n = 20$. The slight bias that is observed is due to thresholding the true complete periodogram to be positive (which as we mentioned above induces a small, additional bias). We also observe that for the same sample size that the regular tapered periodogram (blue) gives a slight improvement in the bias over the regular periodogram (black), but it is not as noticeable as the improvements seen when using the complete periodograms. It is interesting to observe that even for model (M2), which does not have a finite autoregressive representation (thus the estimated complete periodogram incurs additional errors) also has a considerable improvement in bias.

As compared with the regular periodogram, the estimated complete periodogram incurs two additional sources of errors. In Section 2.3.1, we show that the variance of the true complete periodogram tends to be larger than the variance of the regular periodogram. Further in Theorem 2.3.3 we showed that using the estimated Yule-Walker estimators in the predictive DFT leads to an additional $O(p^4/n^2)$ variance in the estimated complete periodogram. This means for small sample sizes and large $p$ the variance can be quite large. We observe both these effects in the right panels in Figures 3.1$-$3.3. In particular, the standard deviation of the various complete periodograms tends to be greater than the asymptotic standard deviation $f(\omega)$ close to the peaks. On the other hand, the standard deviation of the regular periodogram tends to be smaller than $f(\omega)$.

In order to globally access bias/variance trade-off for the different periodograms, we evaluate their mean squared errors. We consider two widely used metrics (see, for example, Hurvich (1988)). The first is the integrated relative mean squared error

$$\text{IMSE} = \frac{1}{nB} \sum_{k=1}^{n} \sum_{j=1}^{B} \left( \frac{\widetilde{I}^{(j)}(\omega_{k,n})}{f(\omega_{k,n})} - 1 \right)^2 \tag{3.7}$$

where $\widetilde{I}^{(j)}(\cdot)$ is the $j$th replication of one of the periodograms. The second metric is the integrated relative bias

$$\text{IBIAS} = \frac{1}{n} \sum_{k=1}^{n} \left( \frac{B^{-1} \sum_{j=1}^{B} \widetilde{I}^{(j)}(\omega_{k,n})}{f(\omega_{k,n})} - 1 \right)^2 . \tag{3.8}$$

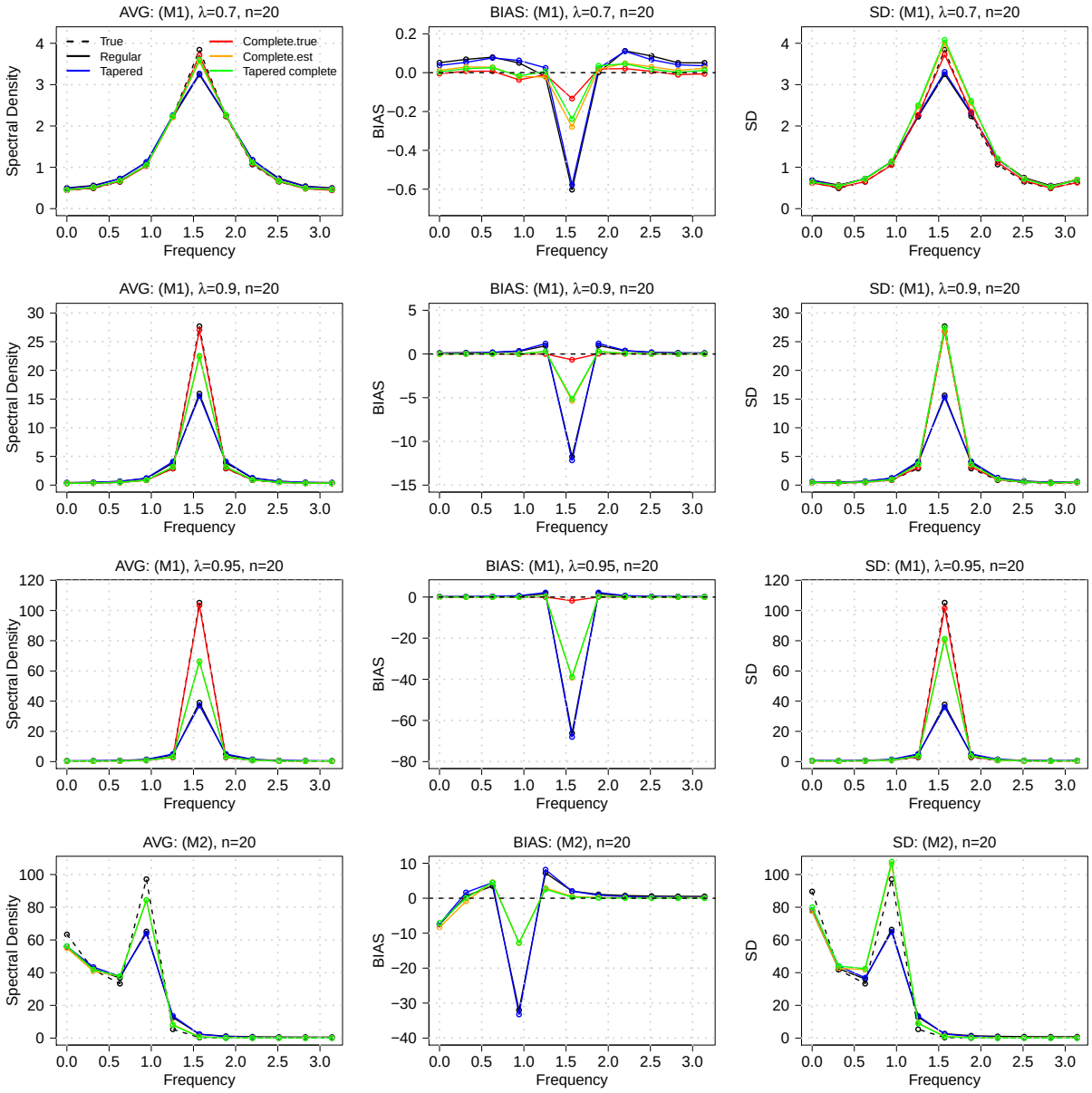



Figure 3.1: The average (left), bias (middle), and standard deviation (right) of the spectral density (black dashed) and the five different periodograms for Models (M1) and (M2). Length of the time series $n = 20$.

Table 3.1 summarizes the IMSE and IBIAS of each periodogram over the different models and sample sizes. In most cases, the tapered periodogram, true complete periodogram (when it can be evaluated) and the two estimated complete periodograms have a smaller IMSE and IBIAS than the regular periodogram. As expected, the IBIAS of the (true) complete periodogram is almost

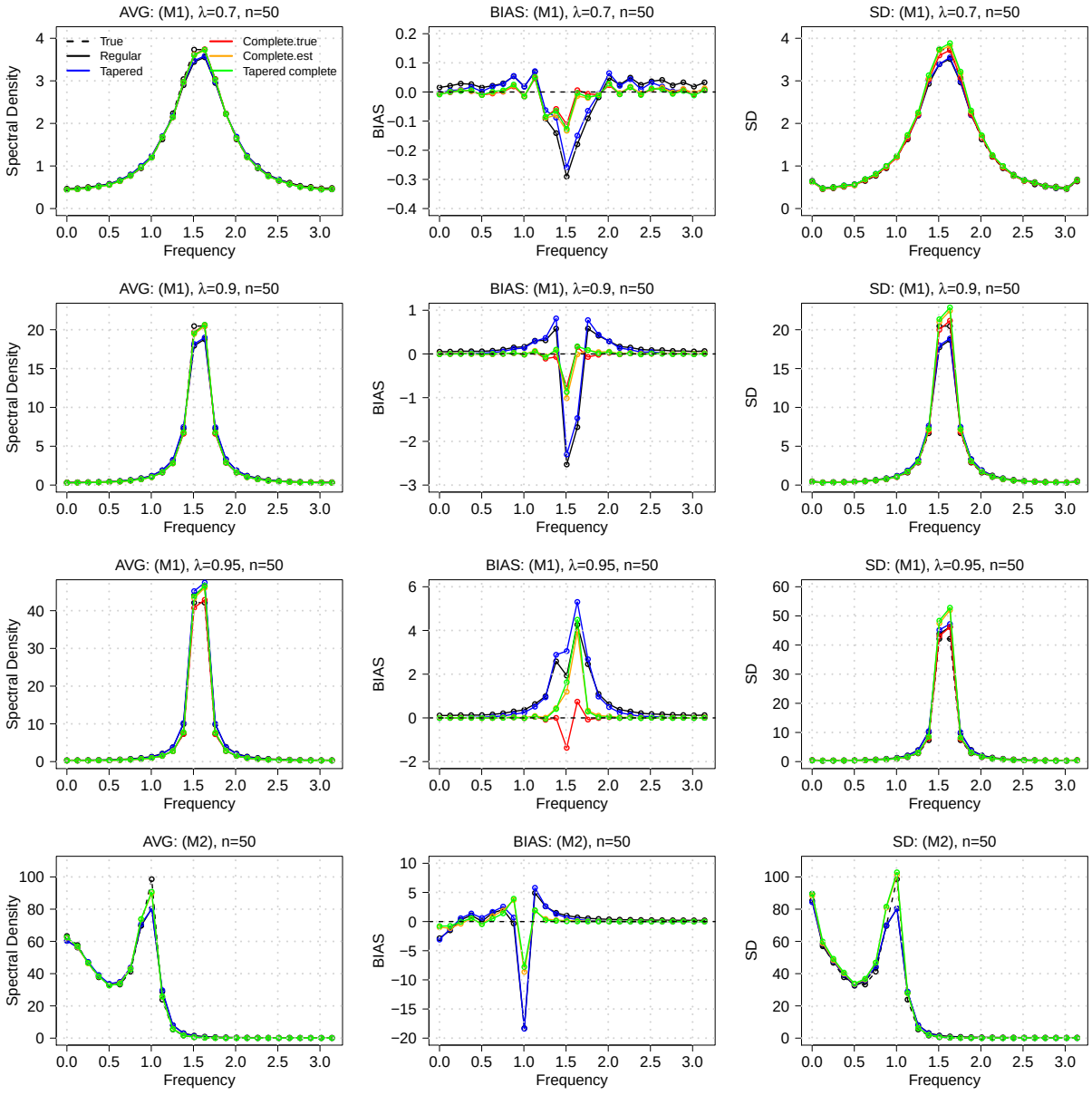



Figure 3.2: The average (left), bias (middle), and standard deviation (right) of the spectral density (black dashed) and the five different periodograms for Models (M1) and (M2). Length of the time series $n = 50$.

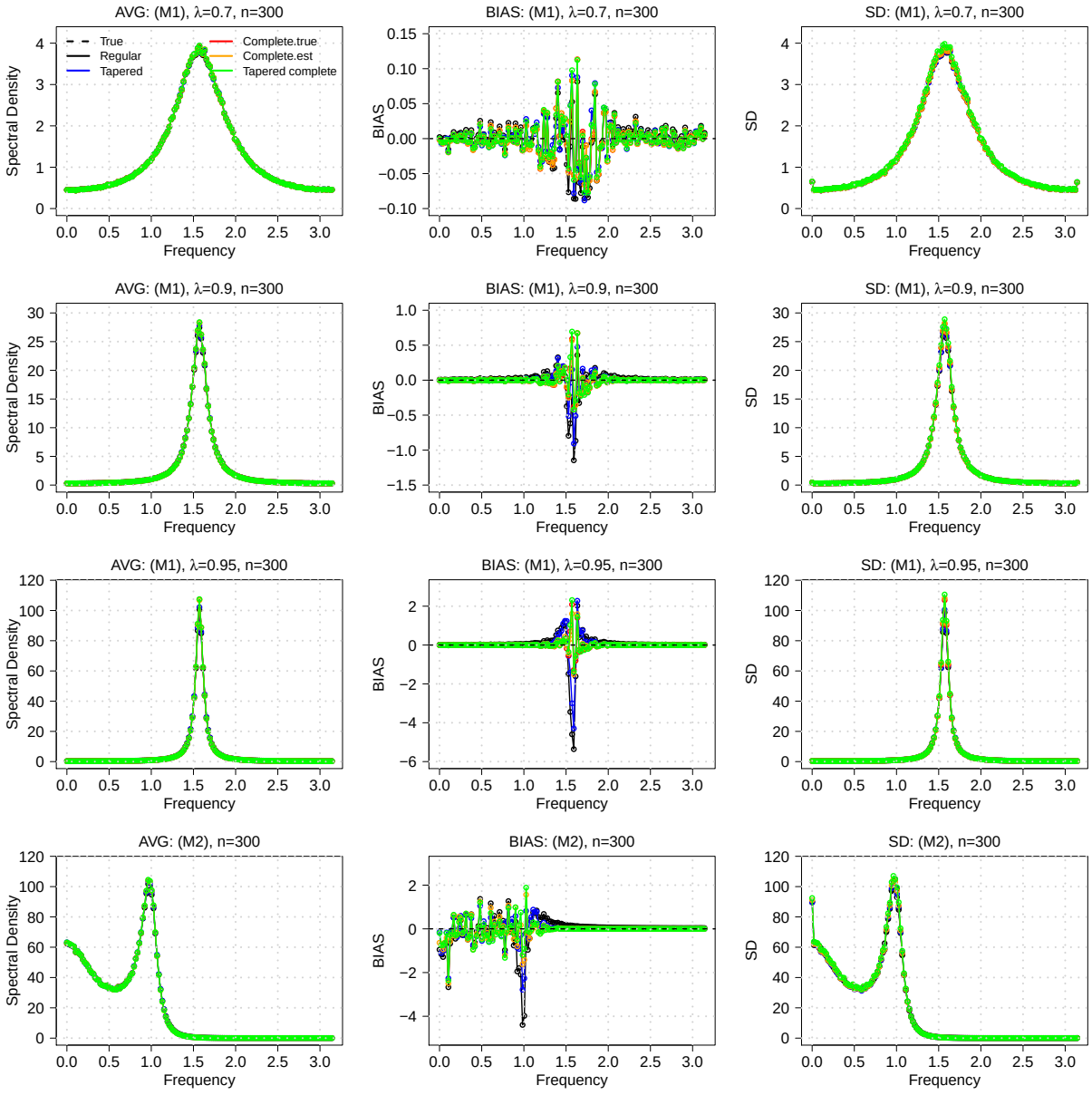



Figure 3.3: The average (left), bias (middle), and standard deviation (right) of the spectral density (black dashed) and the five different periodograms for Models (M1) and (M2). Length of the time series $n = 300$.

zero (rounded off to three decimal digits) for (M1). The estimated complete and tapered complete periodogram has significantly small IBIAS than the regular and tapered periodogram. But interestingly, when the spectral density is "more peaky" the estimated complete periodograms tend to have a smaller IMSE than the regular and tapered periodogram. Suggesting that for peaky spectral densities, the improvement in bias outweighs the increase in the variance. Comparing the tapered complete periodogram with the non-tapered complete periodogram we observe that the tapered complete periodogram tends to have a smaller IBIAS (and IMSE) than the non-tapered (estimated) complete periodogram.

The above results suggest that the proposed periodograms can considerably reduce the small sample bias without increasing the variance by too much.

| Model | $n$ | metric | Regular | Tapered | Complete(True) | Complete(Est) | Tapered complete |
|---|---|---|---|---|---|---|---|
| (M1), $\lambda = 0.7$ | 20 | IMSE | 1.284 | 1.262 | 1.127 | 1.323 | 1.325 |
| | | IBIAS | 0.011 | 0.009 | 0 | 0.002 | 0.001 |
| | 50 | IMSE | 1.101 | 1.069 | 1.055 | 1.098 | 1.117 |
| | | IBIAS | 0.002 | 0.001 | 0 | 0 | 0 |
| | 300 | IMSE | 1.014 | 1.006 | 1.007 | 1.009 | 1.046 |
| | | IBIAS | 0 | 0 | 0 | 0 | 0 |
| (M1), $\lambda = 0.9$ | 20 | IMSE | 2.184 | 2.155 | 1.226 | 1.466 | 1.447 |
| | | IBIAS | 0.152 | 0.159 | 0 | 0.009 | 0.007 |
| | 50 | IMSE | 1.434 | 1.217 | 1.112 | 1.166 | 1.145 |
| | | IBIAS | 0.029 | 0.011 | 0 | 0.001 | 0 |
| | 300 | IMSE | 1.059 | 1.010 | 1.017 | 1.020 | 1.047 |
| | | IBIAS | 0.001 | 0 | 0 | 0 | 0 |
| (M1), $\lambda = 0.95$ | 20 | IMSE | 3.120 | 4.102 | 1.298 | 1.527 | 1.560 |
| | | IBIAS | 0.368 | 0.664 | 0 | 0.022 | 0.018 |
| | 50 | IMSE | 2.238 | 1.486 | 1.211 | 1.295 | 1.200 |
| | | IBIAS | 0.151 | 0.045 | 0 | 0.002 | 0.001 |
| | 300 | IMSE | 1.133 | 1.017 | 1.033 | 1.037 | 1.049 |
| | | IBIAS | 0.004 | 0 | 0 | 0 | 0 |
| (M2) | 20 | IMSE | 457.717 | 136.830 | – | 26.998 | 4.836 |
| | | IBIAS | 157.749 | 58.717 | – | 4.660 | 0.421 |
| | 50 | IMSE | 81.822 | 3.368 | – | 3.853 | 1.357 |
| | | IBIAS | 26.701 | 0.692 | – | 0.288 | 0.002 |
| | 300 | IMSE | 4.376 | 1.015 | – | 1.274 | 1.049 |
| | | IBIAS | 0.787 | 0 | – | 0.003 | 0 |

Table 3.1: IMSE and IBIAS for the different periodograms and models.

### 3.4.2 The autocorrelation estimator

In this section, we estimate the autocorrelation function (ACF) using the integrated periodogram estimator in Section 3.2. Recall that we estimate the autocovariances using

$$\breve{c}_n(r) = \frac{1}{2\pi}\int_0^{2\pi} cos(r\omega)\widetilde{I}_n(\omega)d\omega \tag{3.9}$$

where $\widetilde{I}_n(\cdot)$ is one of the periodograms in Section 3.4. Based on $\breve{c}_n(r)$, the natural estimator of the ACF at lag $r$ is

$$\breve{\rho}_n(r) = \frac{\breve{c}_n(r)}{\breve{c}_n(0)}.$$

Note that if $\widetilde{I}_n(\cdot)$ is the regular periodogram, $\breve{c}_n(\cdot)$ and $\breve{\rho}_n(\cdot)$ become the classical sample autocovariances and sample ACFs respectively.

We generate the Gaussian time series from (M1) and (M2) in Section 3.4 and evaluate the ACF estimators at lag $r = 0, 1, ..., 10$. For the computational purpose, we approximate (3.9) using the Reimann sum over 500 uniform partitions on $[0, 2\pi]$.

Figures 3.4−3.6 show the average (left panels), bias (middle panels), and the mean squared error (MSE; right panels) of the ACF estimators at each lag for different models and sample sizes. Analogous to the results in Section 3.4.1, we observe that the complete and complete tapered periodogram significantly reduce the bias as compared to the regular (black) and tapered (blue) periodogram for all the models.

The MSE paints a complex picture. From the left panels in Figures 3.4−3.6 for (M1), we observe when the lag $r$ is odd, the true $\rho(r) = 0$. For these lags, all the ACF estimators are almost unbiased, and the variance dominates. This is why we observe the oscillation the MSE in (M1) over $r$. For (M2), the bias of all estimators are very small even for an extremely small sample size $n = 20$, and thus the variance dominates. For the small sample sizes ($n = 20$ and 50), MSE of the complete periodograms is larger then the classical methods (Regular and tapered). Whereas for the

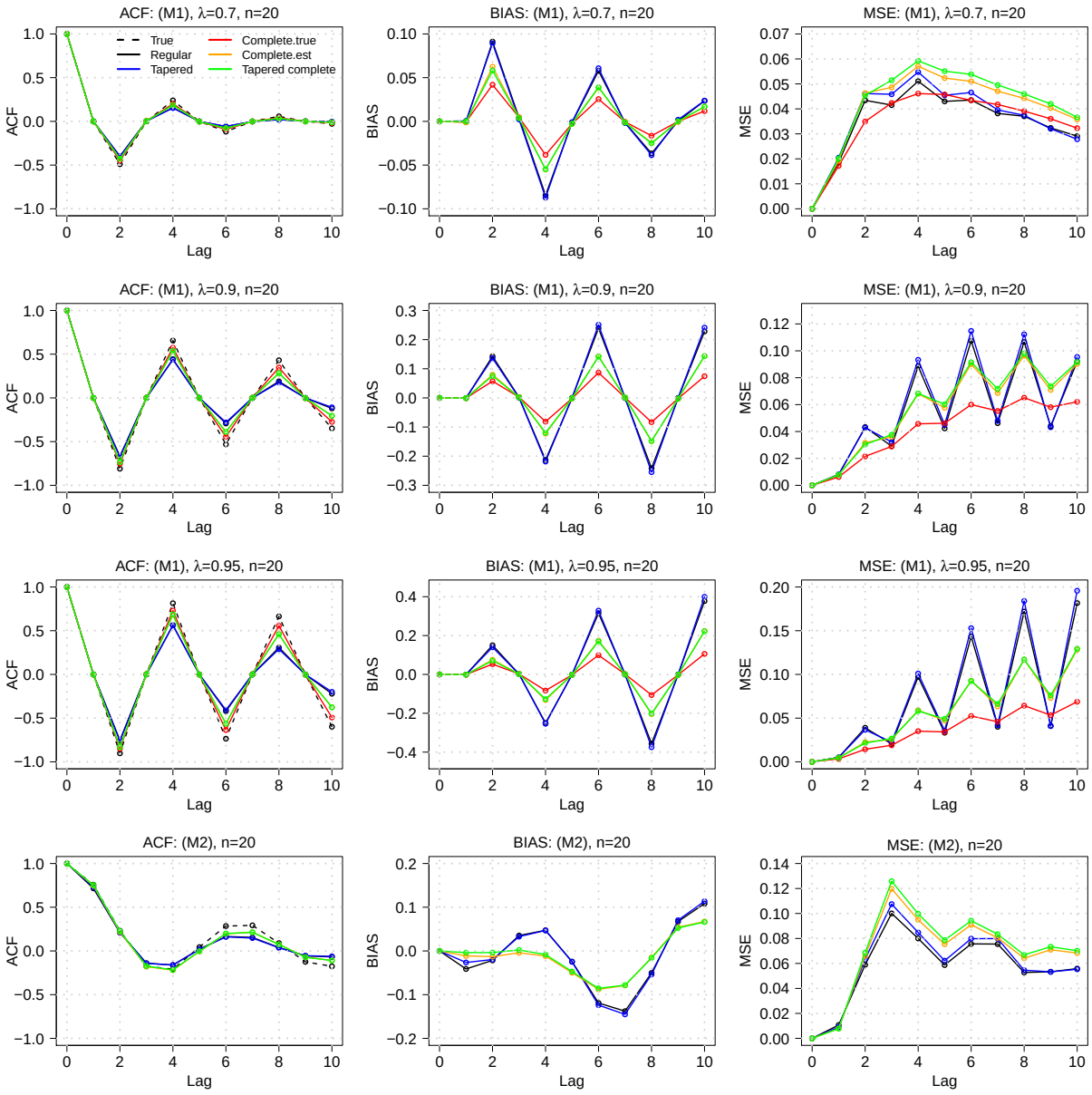



Figure 3.4: ACF: The average (left), bias (middle), and MSE (right) of the ACF estimators at lag $r = 0, ..., 10$. The length of the time series $n = 20$.

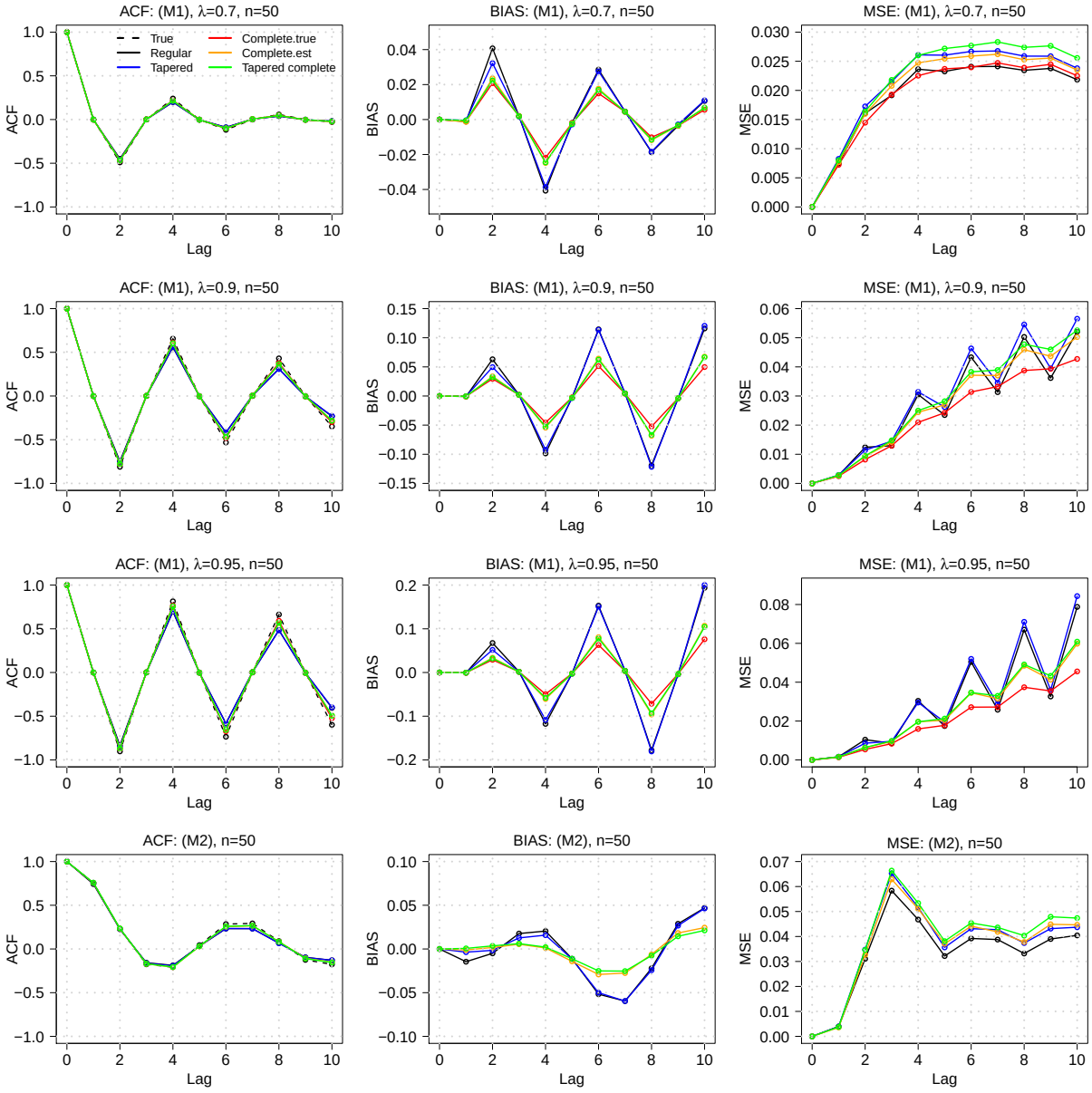



Figure 3.5: ACF: The average (left), bias (middle), and MSE (right) of the ACF estimators at lag $r = 0, ..., 10$. The length of the time series $n = 50$.

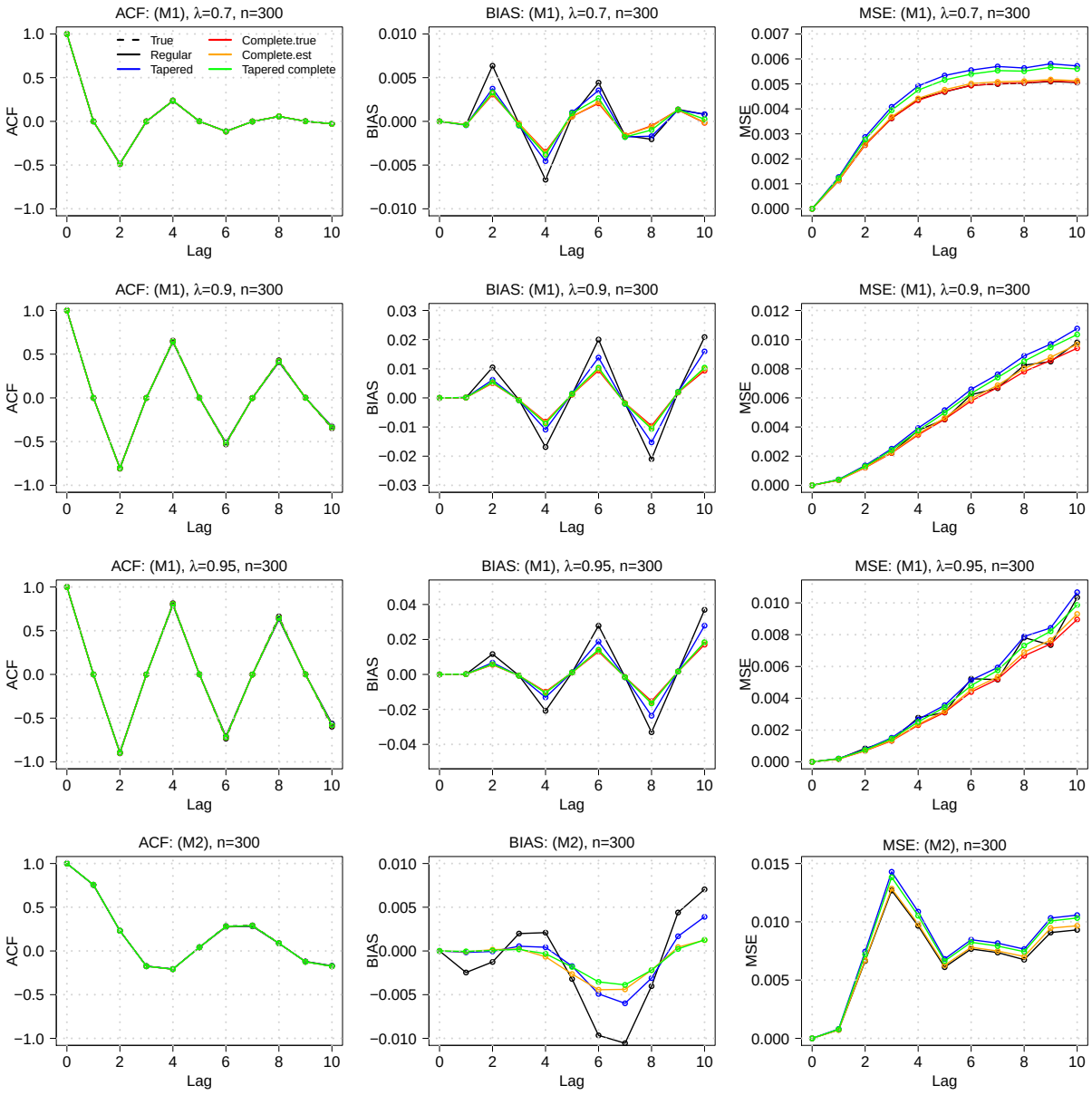



Figure 3.6: ACF: The average (left), bias (middle), and MSE (right) of the ACF estimators at lag $r = 0, ..., 10$. The length of the time series $n = 300$.

large sample size ($n$=300), it seems that the tapering increases the MSE.

To assess the overall performance of the ACF estimators, we evaluate the averaged mean squared error (MSE) and squared bias (BIAS)

$$\text{MSE} = \frac{1}{10B}\sum_{r=1}^{10}\sum_{j=1}^{B}(\breve{\rho}_n^{(j)}(r) - \rho(r))^2, \quad \text{BIAS} = \frac{1}{10}\sum_{r=1}^{10}\left(B^{-1}\sum_{j=1}^{B}\breve{\rho}_n^{(j)}(r) - \rho(r)\right)^2$$

where $\breve{\rho}^{(j)}$ is the $j$th replication of one of the ACF estimators. The results are summarized in Table 3.2. As described above, our method has a marked gain in the BIAS compared to the classical ACF estimators for all models. Moreover, the MSE is comparable, at least for our models, and even has a smaller MSE when the sample size is small and/or there is a strong dependent in the lags.

| Model | $n$ | metric | Regular | Tapered | Complete(True) | Complete(Est) | Tapered complete |
|---|---|---|---|---|---|---|---|
| (M1), $\lambda = 0.7$ | 20 | MSE | 0.038 | 0.040 | 0.038 | 0.044 | 0.046 |
| | | BIAS | 0.002 | 0.002 | 0 | 0.001 | 0.001 |
| | 50 | MSE | 0.021 | 0.023 | 0.021 | 0.022 | 0.024 |
| | | BIAS | 0 | 0 | 0 | 0 | 0 |
| | 300 | MSE | 0.004 | 0.005 | 0.004 | 0.004 | 0.004 |
| | | BIAS | 0 | 0 | 0 | 0 | 0 |
| (M1), $\lambda = 0.9$ | 20 | MSE | 0.061 | 0.064 | 0.045 | 0.062 | 0.063 |
| | | BIAS | 0.023 | 0.025 | 0.003 | 0.008 | 0.008 |
| | 50 | MSE | 0.030 | 0.032 | 0.025 | 0.029 | 0.030 |
| | | BIAS | 0.005 | 0.005 | 0.001 | 0.002 | 0.002 |
| | 300 | MSE | 0.005 | 0.006 | 0.005 | 0.005 | 0.005 |
| | | BIAS | 0 | 0 | 0 | 0 | 0 |
| (M1), $\lambda = 0.95$ | 20 | MSE | 0.077 | 0.082 | 0.039 | 0.063 | 0.064 |
| | | BIAS | 0.045 | 0.049 | 0.004 | 0.015 | 0.014 |
| | 50 | MSE | 0.032 | 0.034 | 0.022 | 0.027 | 0.028 |
| | | BIAS | 0.011 | 0.011 | 0.002 | 0.003 | 0.003 |
| | 300 | MSE | 0.005 | 0.005 | 0.004 | 0.004 | 0.004 |
| | | BIAS | 0 | 0 | 0 | 0 | 0 |
| (M2) | 20 | MSE | 0.062 | 0.065 | – | 0.074 | 0.077 |
| | | BIAS | 0.006 | 0.006 | – | 0.002 | 0.002 |
| | 50 | MSE | 0.036 | 0.040 | – | 0.040 | 0.042 |
| | | BIAS | 0.001 | 0.001 | – | 0 | 0 |
| | 300 | MSE | 0.008 | 0.009 | – | 0.008 | 0.008 |
| | | BIAS | 0 | 0 | – | 0 | 0 |

Table 3.2: MSE and BIAS of an ACF estimators.

### 3.4.3 Spectral density estimation

Finally, we estimate the spectral density function by smoothing the periodogram. We consider the smoothed periodogram of the form

$$\breve{f}(\omega_{k,n}) = \sum_{|j|\leqslant m} W(j)\widetilde{I}_n(\omega_{j+k,n})$$

where $\widetilde{I}_n(\cdot)$ is one of the candidate periodograms described in the previous section and $\{W(\cdot)\}$ are the positive symmetric weights satisfy the conditions (i) $\sum_{|j|\leqslant m} W(j) = 1$ and (ii) $\sum_{|j|\leqslant m} W^2(j) \to 0$. The bandwidth $m = m(n)$ satisfies the condition $m/n \to 0$ as $m, n \to \infty$. We use the following three spectral window functions:

- (The Daniell Window) $\widetilde{W}(j) = \frac{1}{2m+1}$, $|j| \leqslant m$.
- (The Bartlett Window) $\widetilde{W}(j) = 1 - \frac{|j|}{m}$, $|j| \leqslant m$.
- (The Hann Window) $\widetilde{W}(j) = \frac{1}{2}[1 - \cos(\frac{\pi(j+m)}{m})]$, $|j| \leqslant m$.

and normalize using $W(j) = \widetilde{W}(j)/\sum_{|j|\leqslant m} \widetilde{W}(j)$.

In this section, we only focus on estimating the spectral density of model (M2). We smooth the various periodogram using the three window functions described above. For each simulation, we calculate the IMSE and IBIAS (analogous to (3.7) and (3.8)). The bandwidth selection is also very important. One can extend the cross-validation developed for smoothing the regular periodogram (see Hurvich (1985), Beltrão and Bloomfield (1987) and Ombao et al. (2001)) to the complete periodogram and this may be an avenue of future research. In this dissertation, we simply use the bandwidth $m \approx n^{1/5}$ (in terms of order this corresponds to the optimal MSE).

The results are summarized in Table 3.3. We observe that smoothing with the tapered periodogram and the two different complete periodograms have a smaller IMSE and IBIAS as compared to the smooth regular periodogram. This is uniformly true for all the models, sample sizes, and window functions. When the sample size is small ($n = 20$ and $50$), the smooth complete and tapered complete periodogram has a uniformly smaller IMSE and IBIAS than the smooth tapered

periodogram for all window functions. For the large sample size ($n = 300$), smoothing with the tapered periodogram and tapered complete periodogram gave similar results, whereas smoothing using the complete periodogram gives a slightly worse bias and MSE.

| $n$ | $m$ | Window | Metric | Regular | Tapered | Complete | Tapered complete |
|---|---|---|---|---|---|---|---|
| 20 | No smoothing | | IMSE | 457.717 | 136.830 | 26.998 | 4.836 |
| | | | IBIAS | 157.749 | 58.717 | 4.660 | 0.421 |
| | 2 | Daniell | IMSE | 1775.789 | 1399.366 | 1008.590 | 943.855 |
| | | | IBIAS | 882.576 | 780.363 | 444.727 | 408.325 |
| | | Bartlett | IMSE | 538.477 | 203.217 | 43.347 | 17.489 |
| | | | IBIAS | 203.010 | 100.178 | 13.270 | 6.391 |
| | | Hann | IMSE | 538.477 | 203.217 | 43.347 | 17.489 |
| | | | IBIAS | 203.010 | 100.178 | 13.270 | 6.391 |
| 50 | No smoothing | | IMSE | 81.822 | 3.368 | 3.853 | 1.357 |
| | | | IBIAS | 26.701 | 0.692 | 0.288 | 0.002 |
| | 2 | Daniell | IMSE | 87.485 | 7.227 | 5.138 | 3.308 |
| | | | IBIAS | 33.327 | 3.947 | 1.954 | 1.346 |
| | | Bartlett | IMSE | 78.939 | 2.797 | 2.479 | 0.796 |
| | | | IBIAS | 27.883 | 1.106 | 0.425 | 0.074 |
| | | Hann | IMSE | 78.939 | 2.797 | 2.479 | 0.796 |
| | | | IBIAS | 27.883 | 1.106 | 0.425 | 0.074 |
| 300 | No smoothing | | IMSE | 4.376 | 1.015 | 1.274 | 1.049 |
| | | | IBIAS | 0.787 | 0 | 0.003 | 0 |
| | 3 | Daniell | IMSE | 2.514 | 0.176 | 0.210 | 0.173 |
| | | | IBIAS | 0.812 | 0.006 | 0.008 | 0.005 |
| | | Bartlett | IMSE | 2.685 | 0.257 | 0.312 | 0.256 |
| | | | IBIAS | 0.795 | 0.002 | 0.004 | 0.001 |
| | | Hann | IMSE | 2.717 | 0.272 | 0.330 | 0.272 |
| | | | IBIAS | 0.794 | 0.001 | 0.004 | 0.001 |

Table 3.3: IMSE and IBIAS of the smoothed periodogram for (M2).

It is intriguing to note that the smooth complete tapered periodogram gives one the smallest IBIAS and IMSE as compared with all the other methods. These results suggest that spectral smoothing using the tapered complete periodogram may be very useful for studying the spectral density of short time series. Such data sets can arise in many situations, which as the analyses of nonstationary time series, where the local periodograms are often used.

### 3.5 Data analysis

In this section, we present two data analysis using the (tapered) complete periodogram.

### 3.5.1 Analysis of ball bearing data

Vibration analysis, which is the tracking and predicting faults in engineering devices is an important problem in mechanical signal processing. Sensitive fault diagnostic tools can prevent significant financial and health risks for a business. A primary interest is to detect the frequency and amplitude of evolving faults in different component parts of a machine, see Randall and Antoni (2011) for further details.

The Bearing Data Center of the Case Western Reserve University (CWRU; `https://csegroups.case.edu/bearingdatacenter/pages/download-data-file`) maintains a repository of times series sampled from simulated experiments that were conducted to test the robustness of components of ball bearings. The aim of this study is not to detect when a fault has occurred (but this will be the ultimate aim), but to understand the "signature" of the fault. In order to classify (a) no fault, fault and the type of fault, our aim is to detect the features of different fault signals in ball bearings, where the damage occurs in (b) inner race, (b) outer race, and (d) ball spin. Please refer to Figure 3.7 for a schematic diagram of a typical ball bearing and locations where faults can occur. The ball bearing either with no fault or the three different faults described above were part of drive end of test rig motor. Vibration signals were sampled over the course of 10 seconds at 12,000 per second (12 kHz) using an accelerometer.

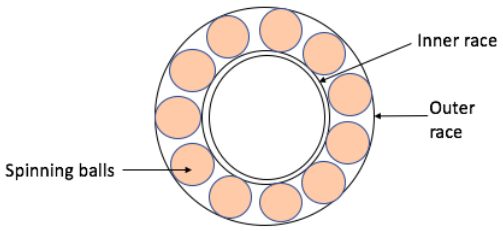



Figure 3.7: A schematic diagram of a ball bearing and the location of the three faults ((b) inner race, (c) outer race, and (d) ball spin).

A commonly used analytic tool in vibration analysis is the envelope spectrum. This is where a

smoothing filter is applied to the regular periodogram to extract the dominant frequencies. Using the envelope spectrum, Randall and Antoni (2011) and Smith and Randall (2015), have shown that a normal ball bearing has power distributed in the relatively lower frequency bandwidth of $60-150$ Hz ($0.05-0.1$, radian). Whereas, faults in the ball bearings lead to deviation from the usual spectral distribution with significant power in the $300-500$ Hz ($0.18-0.26$, radian) bandwidth, depending on the location of the fault. Note that the following are equally important in a vibration analysis, frequencies where the power is greatest but also the amplitude of the power at these frequencies.

The time series in the repository are extremely long, of the order $10^6$. But as the ultimate aim is to devise an online detection scheme based on shorter time series, we focus on shorter segments of the time series ($n = 609$, approximately $0.05$ seconds). A plot of the four different time series is given in Figure 3.8. In this study, we estimate the spectral density of the four time series signals by smoothing the different periodograms; regular, tapered, complete, and tapered complete periodogram. Our aim is to highlight the differences in the dominant frequencies in the spectral distribution of the normal ball bearing signal with three faulty signals. For the tapered and the tapered complete periodogram, we use the Tukey taper defined in (3.2) with $10\%$ tapering (which corresponds to $d = n/10$). For all the periodograms we smooth using the Bartlett window. For the time series (length $609$) we used $m = 16$ (where $m$ is defined in Section 3.4.3).

A plot of the estimated spectral densities is given in Figure 3.9. We observe that all the four spectral density estimators (based on the different periodograms) are very similar. Further, for the normal ball bearing the main power is in the frequency range $0.05 - 0.1(60 - 175$ Hz$)$. Interestingly, the spectral density estimator based on the tapered complete periodogram gives a larger amplitude at the principal frequency. Suggesting that the"normal signal" has greater power at that main frequency than is suggested by the other estimation methods. In contrast, for the faulty ball bearings, the power spectrum is very different from the normal signal. Most of the dominant frequencies are in the range $0.21 - 0.26(375 - 490$ Hz$)$. There appears to be differences between the power spectrum of the three different faults, but the difference is not as striking as the difference between no fault and fault. Whether the differences between the faults are statistically significant

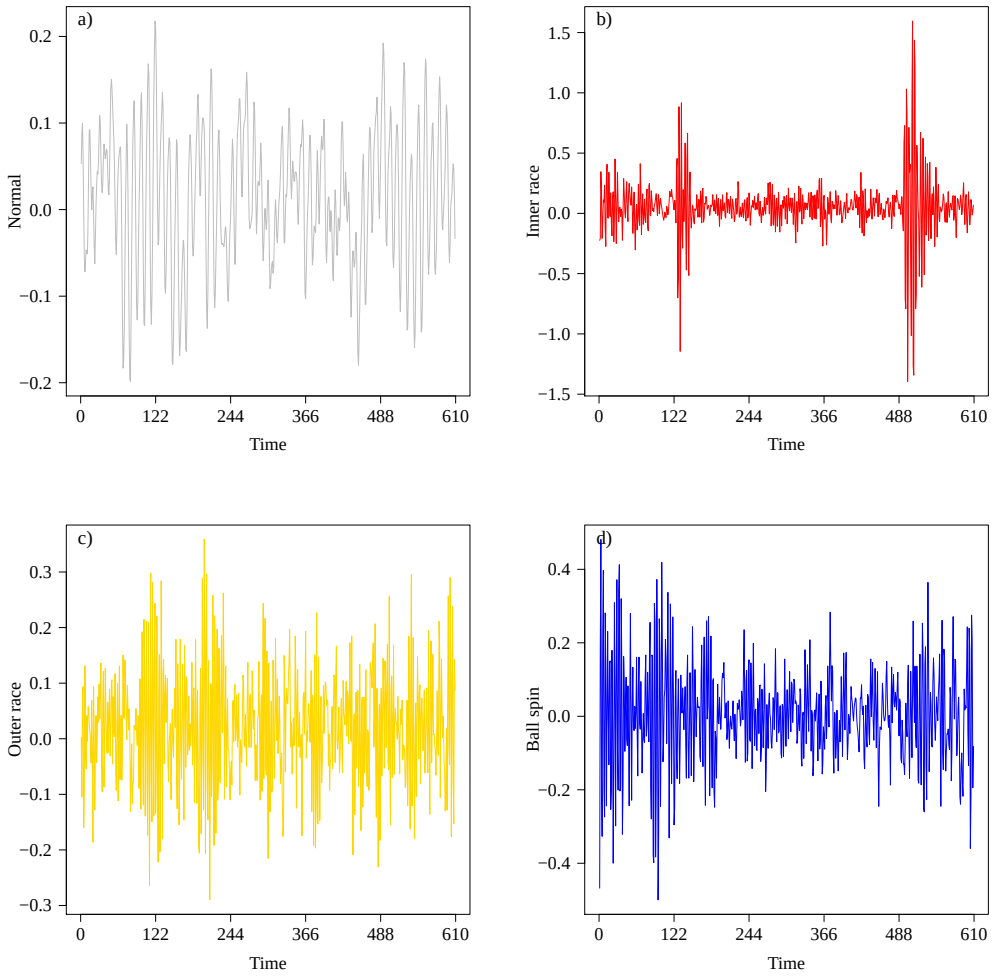



Figure 3.8: Panels in the figure show time series plots of signals recorded from a) Normal ball bearing b) Time series of bearing with fault in inner race, c) Time series of bearing with fault in outer race and, d) Time series of bearing with fault in ball spin. Each time series is of length 609 (0.05 seconds).

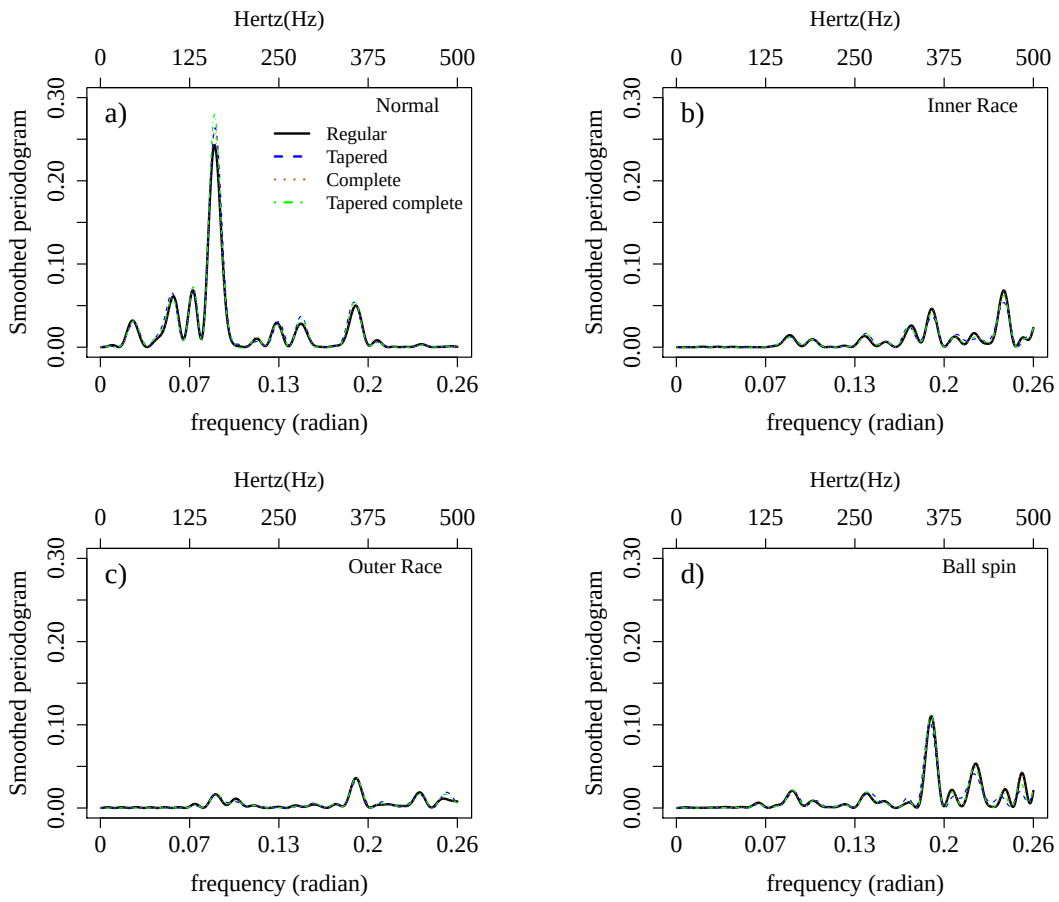



Figure 3.9: Plots show that smoothed periodograms of the four time series signals based on sample size $n = 609$. Top left: Normal, Top Right: Inner Race, Bottom Left: Outer Race and Bottom Right: Ball spin. The top axis shows frequencies in Hertz(Hz).

will be an avenue of future investigation. These observations corroborate the findings of the previous analysis of similar data, see for example Smith and Randall (2015). Despite the similarities in the different estimators the smooth tapered complete periodogram appears to better capture the dominant frequencies in the normal ball bearing. This is reassuring as one objective in vibration analysis is the estimation of power of the vibration at the dominant frequencies.

### 3.5.2 Analysis of sunspot data

We conclude by returning to the sunspot data which first motivated Schuster to define the periodogram 120 years ago.

Sunspots are visibly darker areas that are apparent on the surface of the Sun that are captured from satellite imagery or man-made orbiting telescopes. The darker appearance of these areas is due to their relatively cooler temperatures compared to other parts of the Sun that are attributed to the relatively stronger magnetic fields.

There is a rich history of analysis of the sunspot data and probably Schuster (1897, 1906) is the first one who analyzed this data in a frequency domain. Schuster developed the "periodogram" to study periodicities in sunspot activity. As mentioned in the introduction the Sunspot data has since served as a benchmark for developing several theories and methodologies and theories related to spectral analysis of time series. A broader account of these analyses can be found in Chapter 6−8 of Bloomfield (2004) and references therein.

In this section we implement the four comparator periodograms in Section 3.4 to estimate and corroborate the spectrum of the sunspot data. The dataset we have used is a subset of the data available at the World Data Center Sunspot Index and Long-term Solar Observations (WDC-SILSO), Royal Observatory of Belgium, Brussels (`http://sidc.be/silso/`). We use length $n$=3168 total monthly count of sunspots from Jan 1749 to Dec 2013. All periodograms are computed after removing the sample mean from the data. Figure 3.10 shows the time series plot (right), four different periodograms (middle) and smoothed periodograms (right). We smooth the periodogram using the Bartlett window function from Section 3.4.3 with the bandwidth $m = 5$ ($\approx n^{1/5}$).

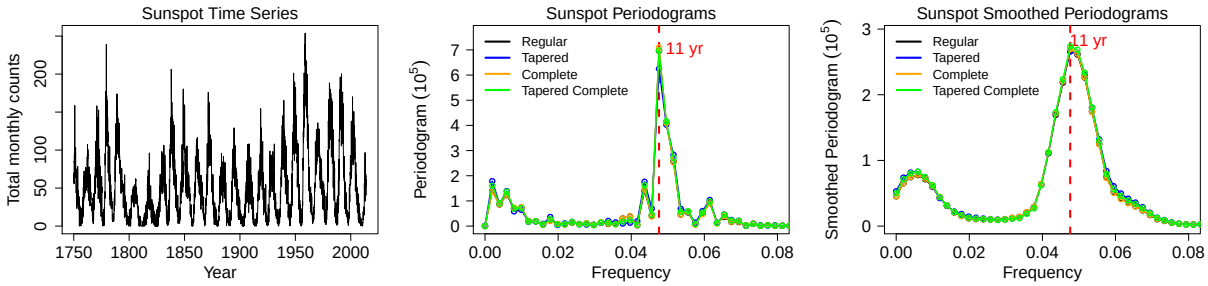



Figure 3.10: Right: Monthly Sunspot time series plot of length 3168 (264 years) starting from Jan 1749. Middle: Trajectories of the four different periodograms; regular, regular tapered, complete and tapered complete periodogram. left: Smoothed periodograms using Bartlett window.

From the middle panel of Figure 3.10, we observe that all the periodograms detect the peak corresponding to maximum sunspot activity at the 11-year cycle. The peak at the 11-year cycle (frequency $0.046$) for the complete periodogram (orange) is the largest, at about $7.1 \times 10^5$, the regular (black) and complete tapered(green) periodogram is slightly lower at about $6.98 \times 10^5$. Whereas, the tapered periodogram (blue) is the lowest at about $6.25 \times 10^5$. Looking at in the neighborhood of the main peak, we observe that there is very little difference between all the periodograms. This suggests that these "side peaks" in the neighborhood of $0.046$ are not an artifact of the periodogram but a feature of the data. Which further suggests that the sunspot data does not contain a fixed period but a quasi-dominant period in the frequencies range $0.042 - 0.058$ ($9.1 - 12.6$ years). The effect is clearer after smoothing the periodogram (right panel of Figure 3.10). Smoothing the complete and tapered complete periodogram yields a more dominant peak at $0.046$ ($11$ years), but the quasi-frequency band remains. Further, a secondary dominate frequency is seen in the very low frequency around $0.006$ ($88$ years) which is more pronounced when the smoothing is done using the (regular) tapered periodogram and tapered complete periodogam. In summary, due to the large sample size all the different periodograms exhibit very similar behaviour. However, even within the large sample setting (where theoretically all the periodograms are asymptotically equivalent) the complete periodograms appear to better capture the amplitude of the peak.

## 3.6 Proofs

In this section, we give a proof of Sections 2.2, 2.3, and 3.

### 3.6.1 Proof of Section 2.2

PROOF of Theorem 2.2.1 We recall that Theorem entails obtaining a transform $U_n\underline{X}_n$ where $\mathrm{cov}_f\left(U_n\underline{X}_n, F_n\underline{X}_n\right) = \Delta_n(f)$. Pre and post multiplying this covariance with $F_n^*$ and $F_n$ gives

$$F_n^*\mathrm{cov}_f\left(U_n\underline{X}_n, F_n\underline{X}_n\right)F_n = \mathrm{cov}_\theta\left(F_n^*U_n\underline{X}_n, \underline{X}_n\right) = F_n^*\Delta_n(f)F_n = C_n(f).$$

Thus our objective is to find the transform $\underline{Y}_n = F_n^*U_n\underline{X}_n$ such that $\mathrm{cov}_f(\underline{Y}_n, \underline{X}_n) = C_n(f)$. Then, the vector $F_n\underline{Y}_n = U_n\underline{X}_n$ will be biorthogonal to $F_n\underline{X}_n$, as required. We observe that the

entries of the circulant matrix $C_n(f)$ are

$$(C_n(f))_{u,v} = n^{-1} \sum_{k=1}^{n} f(\omega_{k,n}) \exp(-i(u-v)\omega_{k,n}) = \sum_{\ell \in \mathbb{Z}} c_f(u-v+\ell n),$$

where the second equality is due to the Poisson summation. The random vector $\underline{Y}_n = \{Y_{u,n}\}_{u=1}^n$ is such that $\mathrm{cov}_f(Y_{u,n}, X_v) = \sum_{\ell \in \mathbb{Z}} c_f(u-v+\ell n)$ and $Y_u \in \mathrm{sp}(\underline{X}_n)$. Since $\mathrm{cov}_f(X_{u+\ell n}, X_v) = c_f(u-v+\ell n)$, at least "formally" $\mathrm{cov}_f(\sum_{\ell \in \mathbb{Z}} X_{u+\ell n}, X_v) = \sum_{\ell \in \mathbb{Z}} c_f(u-v+\ell n)$. However, $\sum_{\ell \in \mathbb{Z}} X_{u+\ell n}$ is neither a well defined random variable nor does not it belong to $\mathrm{sp}(\underline{X}_n)$. We replace each element in the sum $\sum_{\ell \in \mathbb{Z}} X_{u+\ell n}$ with an element that belongs to $\mathrm{sp}(\underline{X}_n)$ and gives the same covariance. To do this we use the following well known result. Let $Z$ and $\underline{X}$ denote a random variable and vector respectively. Let $P_{\underline{X}}(Z)$ denote the projection of $Z$ onto $\mathrm{sp}(\underline{X})$, i.e., the best linear predictor of $Z$ given $\underline{X}$, then $\mathrm{cov}_f(Z, \underline{X}) = \mathrm{cov}_f(P_{\underline{X}}(Z), \underline{X})$. Let $\widehat{X}_{\tau,n}$ denote best linear predictor of $X_\tau$ given $\underline{X}_n = (X_1, \ldots, X_n)$ (as defined in (2.1)). $\widehat{X}_{\tau,n}$ retains the pertinent properties of $X_\tau$ in the sense that $\mathrm{cov}_f(\widehat{X}_{\tau,n}, X_t) = c_f(\tau - t)$ for all $\tau \in \mathbb{Z}$ and $1 \leqslant t \leqslant n$. Define

$$Y_{u,n} = \sum_{\ell \in \mathbb{Z}} \widehat{X}_{u+\ell n,n} = \sum_{s=1}^{n} \left( \sum_{\ell \in \mathbb{Z}} \phi_{s,n}(u+\ell n; f) \right) X_s \in \mathrm{sp}(\underline{X}_n),$$

where we note that $Y_{u,n}$ a well defined random variable, since by using Lemma A.1.1 it can be shown that $\sup_n \sum_{s=1}^n \sum_{\ell=-\infty}^{\infty} |\phi_{s,n}(u+\ell n; f)| < \infty$. Thus by definition of $Y_{u,n}$ the following holds

$$\mathrm{cov}_f\left(Y_{u,n}, X_v\right) = \sum_{\ell \in \mathbb{Z}} c_f(u-v+\ell n) = \left(C_n(f)\right)_{u,v}, \tag{3.10}$$

and $\underline{Y}_n = F_n^* U_n \underline{X}_n$, gives the desired transformation of the time series. Thus, based on this construction, $F_n \underline{Y}_n = U_n \underline{X}_n$ and $F_n \underline{X}_n$ are biorthogonal transforms, with entries $(F_n \underline{X}_n)_k =$

$J_n(\omega_{k,n})$ and

$$\begin{aligned}(U_n\underline{X}_n)_k = (F_n\underline{Y}_n)_k &= n^{-1/2}\sum_{\ell\in\mathbb{Z}}\sum_{u=1}^{n}\widehat{X}_{u+\ell n,n}e^{iu\omega_{k,n}} \\ &= n^{-1/2}\sum_{\tau\in\mathbb{Z}}\widehat{X}_{\tau,n}e^{i\tau\omega_{k,n}} \\ &= n^{-1/2}\sum_{t=1}^{n}X_t\sum_{\tau\in\mathbb{Z}}\phi_{t,n}(\tau;f)e^{i\tau\omega_{k,n}}. \end{aligned} \tag{3.11}$$

The entries of the matrix $U_n$ are $(U_n)_{k,t} = n^{-1/2}\sum_{\tau\in\mathbb{Z}}\phi_{t,n}(\tau;f)e^{i\tau\omega_{k,n}}$. To show that $U_n$ "embeds" the regular DFT, we observe that for $1 \leqslant \tau \leqslant n$, $\phi_{t,n}(\tau;f) = \delta_{\tau,t}$, furthermore, due to second order stationarity the coefficients $\phi_{t,n}(\tau;f)$ are reflective i.e. the predictors of $X_m$ (for $m > n$) and $X_{n+1-m}$ share the same set of prediction coefficients (just reflected) such that

$$\phi_{t,n}(m;f) = \phi_{n+1-t,n}(n+1-m;f) \qquad \text{for } m > n.$$

Using these two observations we can decompose $(U_n)_{k,t}$ as

$$\begin{aligned}(U_n)_{k,t} &= n^{-1/2}\left(e^{it\omega_{k,n}} + \sum_{\tau\leqslant 0}\phi_{t,n}(\tau;f)e^{i\tau\omega_{k,n}} + \sum_{\tau\geqslant n+1}\phi_{t,n}(\tau;f)e^{i\tau\omega_{k,n}}\right) \\ &= n^{-1/2}e^{it\omega_{k,n}} + n^{-1/2}\sum_{\tau\leqslant 0}\left(\phi_{t,n}(\tau;f)e^{i\tau\omega_{k,n}} + \phi_{n+1-t,n}(\tau;f)e^{-i(\tau-1-n)\omega_{k,n}}\right).\end{aligned}$$

It immediately follows from the above decomposition that $U_n = F_n + D_n(f)$ where $D_n(f)$ is defined in (2.3). Thus proving (2.2).

To prove (2.4), we first observe that (2.2) implies

$$\operatorname{cov}_f\left(((F_n + D_n(f))\underline{X}_n)_{k_1}, (F_n\underline{X}_n)_{k_2}\right) = f(\omega_{k_1,n})\delta_{k_1,k_2}.$$

It is clear that $(F_n \underline{X}_n)_k = J_n(\omega_{k,n})$ and from the representation of $F_n \underline{Y}_n$ given in (3.11) we have

$$\begin{aligned}(F_n \underline{Y}_n)_k &= n^{-1/2} \sum_{\tau=1}^{n} X_\tau e^{i\tau\omega_{k,n}} + n^{-1/2} \sum_{\tau \notin \{1,\ldots,n\}} \widehat{X}_{\tau,n} e^{i\tau\omega_{k,n}} \\ &= J_n(\omega_{k,n}) + \widehat{J}_n(\omega_{k,n}; f).\end{aligned}$$

This immediately proves (2.4). □

PROOF of Theorem 2.2.2 To prove Theorem 2.2.2 we study the predictive DFT for autoregressive processes. We start by obtaining an explicit expression for $\widehat{J}_n(\omega; f_\theta)$ where $f_\theta(\omega) = \sigma^2|1 - \sum_{u=1}^{p} \phi_u e^{-iu\omega}|^{-2}$ (the spectral density corresponding to an AR($p$) process). It is straightforward to show that predictive DFT predictor based on the AR(1) model is

$$\begin{aligned}\widehat{J}_n(\omega; f_\theta) &= n^{-1/2} \sum_{\tau=-\infty}^{0} \phi^{-\tau+1} X_1 e^{i\tau\omega} + n^{-1/2} \sum_{\tau=n+1}^{\infty} \phi^{\tau+1-n} X_n e^{i\tau\omega} \\ &= \frac{n^{-1/2}\phi}{\phi_1(\omega)} X_1 + \frac{n^{-1/2}\phi}{\overline{\phi_1(\omega)}} X_n e^{i(n+1)\omega},\end{aligned}$$

where $\phi_1(\omega) = 1 - \phi e^{-i\omega}$. In order to prove Theorem 2.2.2, which generalizes the above expression to AR($p$) processes, we partition $\widehat{J}_n(\omega; f_\theta)$ into the predictions involving the past and future terms

$$\widehat{J}_n(\omega; f_\theta) = \widehat{J}_{n,L}(\omega; f_\theta) + \widehat{J}_{n,R}(\omega; f_\theta)$$

where

$$\widehat{J}_{n,L}(\omega; f_\theta) = n^{-1/2} \sum_{\tau=-\infty}^{0} \widehat{X}_{\tau,n} e^{i\tau\omega} \quad \text{and} \quad \widehat{J}_{n,R}(\omega; f_\theta) = n^{-1/2} \sum_{\tau=n+1}^{\infty} \widehat{X}_{\tau,n} e^{i\tau\omega}.$$

We now obtain expressions for $\widehat{J}_{n,L}(\omega; f_\theta)$ and $\widehat{J}_{n,R}(\omega; f_\theta)$ separately, in the case the predictors are based on the AR($p$) parameters where $f_\theta(\omega) = \sigma^2|1 - \sum_{j=1}^{p} \phi_j e^{ij\omega}|^{-2}$ and the $\{\phi_j\}_{j=1}^{p}$ correspond to the causal AR($p$) representation. To do so, we define the $p$-dimension vector $\underline{\phi}' = (\phi_1, \ldots, \phi_p)$

and the matrix $A_p(\underline{\phi})$ as

$$A_p(\underline{\phi}) = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}. \tag{3.12}$$

Therefore, for $\tau \leqslant 0$, since $\widehat{X}_{\tau,n} = \left[A_p(\underline{\phi})^{|\tau|+1}\underline{X}_p\right]_{(1)}$, where $\underline{X}_p = (X_1, \dots, X_p)$, we can write

$$\widehat{J}_{n,L}(\omega; f_\theta) = n^{-1/2} \sum_{\tau=-\infty}^{0} \left[A_p(\underline{\phi})^{|\tau|+1}\underline{X}_p\right]_{(1)} e^{i\tau\omega}. \tag{3.13}$$

By using Lemma D.2.1 in Appendix, we have

$$[A_p(\underline{\phi})^{|\tau|+1}\underline{X}_p]_{(1)} = \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s}\psi_{|\tau|-s}.$$

Therefore, using (3.13) and the change of variables $\tau \leftarrow -\tau$

$$\begin{aligned} \widehat{J}_{n,L}(\omega; f_\theta) &= n^{-1/2} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s} \sum_{\tau=0}^{\infty} \psi_{\tau-s} e^{-i\tau\omega} \\ &= n^{-1/2} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{-is\omega} \sum_{\tau=0}^{\infty} \psi_{\tau-s} e^{-i(\tau-s)\omega} \\ &= n^{-1/2} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{-is\omega} \sum_{\tau=s}^{\infty} \psi_{\tau-s} e^{-i(\tau-s)\omega}. \end{aligned}$$

Let $\sum_{s=0}^{\infty} \psi_s e^{-is\omega} = \psi(\omega) = \phi_p(\omega)^{-1}$, and substitute this into the above to give

$$\widehat{J}_{n,L}(\omega; f_\theta) = \frac{n^{-1/2}}{\phi_p(\omega)} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{-is\omega}, \tag{3.14}$$

Thus we obtain an expression for the left hand side of the predictive DFT. Using the similar tech-

nique, it can be shown that the right hand side predictive DFT $\widehat{J}_{n,R}(\omega; f_\theta)$ has the representation

$$\widehat{J}_{n,R}(\omega; f_\theta) = e^{in\omega} \frac{n^{-1/2}}{\overline{\phi_p(\omega)}} \sum_{\ell=1}^{p} X_{n+1-\ell} \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{i(s+1)\omega}.$$

Thus proving equation (2.8). □

### 3.6.2 Proof of Section 2.3

PROOF of Theorem 2.3.1 We recall that

$$\begin{aligned} \widehat{J}_n(\omega; f) &= n^{-1/2} \sum_{\tau=-\infty}^{0} \widehat{X}_{\tau,n} e^{i\tau\omega} + n^{-1/2} \sum_{\tau=n+1}^{\infty} \widehat{X}_{\tau,n} e^{i\tau\omega} \\ &= n^{-1/2} \sum_{t=1}^{n} X_t \big( \sum_{\tau \leqslant 0} \big[ \phi_{t,n}(\tau; f) e^{i\tau\omega} + \phi_{n+1-t,n}(\tau; f) e^{-i(\tau-1-n)\omega} \big] \big) \end{aligned}$$

Using the above we write $\widehat{J}_n(\omega; f)$ as an inner product. Let

$$D_{t,n}(f) = n^{-1/2} \sum_{\tau \leqslant 0} \big[ \phi_{t,n}(\tau; f) e^{i\tau\omega} + \phi_{n+1-t,n}(\tau; f) e^{-i(\tau-1-n)\omega} \big].$$

Next, define the vectors

$$\underline{e}_n' = n^{-1/2}(e^{-i\omega}, ..., e^{-in\omega}) \quad \text{and} \quad \underline{D}_n(f)' = (D_{1,n}(f), ..., D_{n,n}(f)),$$

note that $\underline{e}_n$ and $\underline{D}_n(f)$ are both functions of $\omega$, but we have suppressed this dependence in our notation. Then, $J_n(\omega)$ and $\widehat{J}_n(\omega; f)$ can be represented as the inner products

$$J_n(\omega) = \underline{e}_n^* \underline{X}_n \text{ and } \widehat{J}_n(\omega; f) = \underline{X}_n' \underline{D}_n(f)$$

where $*$ denotes the Hermitian of a matrix. In the same vein we write $\widehat{J}_{\infty,n}(\omega;f)$ as an inner product. Let

$$\begin{aligned} D_t(f) &= n^{-1/2}\sum_{\tau\leqslant 0}\left[\phi_t(\tau;f)e^{i\tau\omega}+\phi_{n+1-t}(\tau;f)e^{-i(\tau-1-n)\omega}\right] \\ \underline{D}_{\infty,n}(f)' &= (D_1(f),...,D_n(f)), \end{aligned}$$

then we can write $\widehat{J}_{\infty,n}(\omega;f)=\underline{X}_n'\underline{D}_n(f)$. Therefore,

$$\begin{aligned} \left(\widehat{J}_{\infty,n}(\omega;f)-\widehat{J}_n(\omega;f)\right)\overline{J_n(\omega)} &= I_{\infty,n}(\omega;f)-I_n(\omega;f) \\ &= n^{-1/2}\sum_{s,t=1}^{n}X_tX_s(\underline{D}_{\infty,n}(f)-\underline{D}_n(f))_{(t)}e^{-is\omega} \\ &= \underline{X}_n'\left(\underline{D}_{\infty,n}(f)-\underline{D}_n(f)\right)\underline{e}_n'\underline{X}_n \\ &= \underline{X}_n'A_1(\omega)\underline{X}_n \end{aligned}$$

where $A_1(\omega)=\left(\underline{D}_{\infty,n}(f)-\underline{D}_n(f)\right)\underline{e}_n'$, an $(n\times n)$ matrix. For the remainder of this proof we drop the dependence of $A_1(\omega)$ on $\omega$. However, if we integrate over $\omega$ this dependence does become important. Using this notation, we have

$$\begin{aligned} \mathbb{E}\left[\left(\widehat{J}_{\infty,n}(\omega;f)-\widehat{J}_n(\omega;f)\right)\overline{J_n(\omega)}\right] &= \mathbb{E}[\underline{X}_n'A_1\underline{X}_n] \\ \mathrm{var}\left[\left(\widehat{J}_{\infty,n}(\omega;f)-\widehat{J}_n(\omega;f)\right)\overline{J_n(\omega)}\right] &= \mathrm{var}[\underline{X}_n'A_1\underline{X}_n]. \end{aligned}$$

By simple algebra

$$\begin{aligned} \mathbb{E}[\underline{X}_n'A_1\underline{X}_n] &= \mathrm{tr}(A_1R_n) \\ \mathrm{var}[\underline{X}_n'A_1\underline{X}_n] &= 2\mathrm{tr}(A_1R_nA_1R_n)+\sum_{s,t,u,v=1}^{n}(A_1)_{s,t}(A_1)_{u,v}\mathrm{cum}\left(X_s,X_t,X_u,X_v\right), \end{aligned} \quad (3.15)$$

where $R_n = \mathrm{var}[\underline{X}_n]$ (noting that $R_n$ is a Toeplitz matrix). To bound the expectation

$$\begin{aligned}
|\mathbb{E}[\underline{X}_n' A_1 \underline{X}_n]| = |\mathrm{tr}(A_1 R_n)| &\leqslant n^{-1/2} \sum_{s,t=1}^{n} |(\underline{D}_{\infty,n}(f) - \underline{D}_n(f))_{(t)} e^{-is\omega} (R_n)_{t,s}| \\
&= n^{-1/2} \sum_{s,t=1}^{n} |D_t(f) - D_{t,n}(f)||c(t-s)| \\
&\leqslant n^{-1/2} \sum_{t=1}^{n} |D_t(f) - D_{t,n}(f)| \left( \sum_{r\in\mathbb{Z}} |c(r)| \right). \qquad (3.16)
\end{aligned}$$

To bound the above, we observe that the sum over $t$ is

$$\begin{aligned}
&n^{-1/2} \sum_{t=1}^{n} |D_t(f) - D_{t,n}(f)| \\
&= n^{-1} \sum_{t=1}^{n} \sum_{\tau\leqslant 0} \left| (\phi_t(\tau; f) - \phi_{t,n}(\tau; f)) e^{i\tau\omega} + (\phi_{n+1-t}(\tau; f) - \phi_{n+1-t,n}(\tau; f)) e^{-i(\tau-1-n)\omega} \right| \\
&\leqslant n^{-1} \sum_{t=1}^{n} \sum_{\tau\leqslant 0} |\phi_t(\tau; f) - \phi_{t,n}(\tau; f)| + \sum_{t=1}^{n} \sum_{\tau\leqslant 0} |\phi_{n+1-t}(\tau; f) - \phi_{n+1-t,n}(\tau; f)| \\
&= 2n^{-1} \sum_{t=1}^{n} \sum_{\tau\leqslant 0} |\phi_t(\tau; f) - \phi_{t,n}(\tau; f)|.
\end{aligned}$$

To bound the above, we use the generalized Baxter's inequality in Lemma A.1.1. Using (A.1) with $K = 0$ the we have

$$\begin{aligned}
\sum_{t=1}^{n} \sum_{\tau\leqslant 0} |\phi_t(\tau; f) - \phi_{t,n}(\tau; f)| &\leqslant C_{f,0} \sum_{\tau\leqslant 0} \sum_{t=n+1}^{\infty} |\phi_t(\tau; f)| \\
&\leqslant C_{f,0} \sum_{\tau=0}^{\infty} \sum_{t=n+1}^{\infty} \sum_{j=0}^{\infty} |\phi_{t+j}||\psi_{\tau-j}| \leqslant C_{f,0} \sum_{\ell\in\mathbb{Z}} |\psi_\ell| \sum_{t=n+1}^{\infty} \sum_{j=0}^{\infty} |\phi_{t+j}| \\
&\leqslant C_{f,0} \sum_{\ell\in\mathbb{Z}} |\psi_\ell| \sum_{u=n+1}^{\infty} |u\phi_u| \leqslant \frac{C_{f,0}}{n^{K-1}} \sum_{\ell\in\mathbb{Z}} |\psi_\ell| \sum_{u=n+1}^{\infty} |u^K \phi_u|
\end{aligned}$$

To bound the above we use Assumption 2.3.1. By using Lemma 2.1 of Kreiss et al. (2011), under Assumption 2.3.1, we have $\sum_{u=1}^{\infty} |u^K \phi_u| \leqslant \infty$. Therefore,

$$\sum_{t=1}^{n} \sum_{\tau\leqslant 0} |\phi_t(\tau; f) - \phi_{t,n}(\tau; f)| = O(n^{-K+1}),$$

which gives

$$n^{-1/2}\sum_{t=1}^{n}|D_t(f)-D_{t,n}(f)| \leqslant 2n^{-1}\sum_{t=1}^{n}\sum_{\tau\leqslant 0}|\phi_t(\tau;f)-\phi_{t,n}(\tau;f)| = O(n^{-K}). \tag{3.17}$$

Substituting the above bound into (3.16) gives

$$\mathbb{E}[\underline{X}_n' A_1 \underline{X}_n] = \mathrm{tr}(A_1 R_n) \leqslant n^{-1/2}\sum_{t=1}^{n}|D_t(f)-D_{t,n}(f)|\left(\sum_{r\in\mathbb{Z}}|c(r)|\right) = O(n^{-K}). \tag{3.18}$$

Next we consider the variance. The first term in the variance (3.15) is bounded with

$$\begin{aligned}
|\mathrm{tr}(A_1R_nA_1R_n)| &\leqslant n^{-1}\sum_{s,t,u,v=1}^{n}|(D_s(f)-D_{s,n}(f))(D_t(f)-D_{t,n}(f))e^{-iu\omega}e^{-iv\omega}(R_n)_{s,u}(R_n)_{t,v}| \\
&= n^{-1}\sum_{s,t,u,v=1}^{n}|D_s(f)-D_{s,n}(f)||D_t(f)-D_{t,n}(f)||c(s-u)||c(t-v)| \\
&\leqslant \left(n^{-1/2}\sum_{t=1}^{n}|D_t(f)-D_{t,n}(f)|\right)^2\left(\sum_{r\in\mathbb{Z}}|c(r)|\right)^2 = O(n^{-2K}),
\end{aligned}$$

where the last line follows from (3.17). The second term in (3.15) is bounded by

$$\begin{aligned}
&\sum_{s,t,u,v=1}^{n}|(A_1)_{s,t}(A_1)_{u,v}\mathrm{cum}\left(X_s,X_t,X_u,X_v\right)| \\
&= n^{-1}\sum_{s,t,u,v=1}^{n}|D_t(f)-D_{t,n}(f)||D_v(f)-D_{v,n}(f)|\kappa_4\left(t-s,u-s,v-s\right)| \\
&\leqslant n^{-1}\sum_{t,v=1}^{n}|D_t(f)-D_{t,n}(f)||D_v(f)-D_{v,n}(f)|\sum_{s,u=1}^{n}|\kappa_4\left(t-s,u-s,v-s\right)| \\
&\leqslant n^{-1}\sum_{t,v=1}^{n}|D_t(f)-D_{t,n}(f)||D_v(f)-D_{v,n}(f)|\sum_{i,j,k\in\mathbb{Z}}|\kappa_4\left(i,j,k\right)| \\
&= \left(n^{-1/2}\sum_{t=1}^{n}|D_t(f)-D_{t,n}(f)|\right)^2\sum_{i,j,k\in\mathbb{Z}}|\kappa_4\left(i,j,k\right)| = O\left(n^{-2K}\right)
\end{aligned}$$

where the above follows from (3.17) and Assumption 2.3.2. Altogether this gives $\mathrm{var}[\underline{X}_n' A_1 \underline{X}_n] = O(n^{-2K})$. This proves theorem. □

PROOF of Theorem 2.3.2 To prove the theorem, we use the following observation. In the special case that $f = f_p$ corresponds to the AR($p$) model, the best finite linear predictor (given $p$ observations) and the best infinite predictor are the same in this case, $\underline{D}_n(f_p) = \underline{D}_{\infty,n}(f_p)$. Therefore, we have

$$\begin{aligned}\left(\widehat{J}_n(\omega; f_p) - \widehat{J}_{\infty,n}(\omega; f)\right)\overline{J_n(\omega)} &= n^{-1/2}\sum_{s,t=1}^{n} X_s X_t (\underline{D}_{\infty,n}(f) - \underline{D}_n(f_p))_{(t)} e^{-is\omega} \\ &= \underline{X}_n' \left(\underline{D}_{\infty,n}(f_p) - \underline{D}_{\infty,n}(f)\right)\underline{e}_n' \underline{X}_n \\ &= \underline{X}_n' A_2(\omega)\underline{X}_n \end{aligned} \tag{3.19}$$

where $A_2(\omega) = \left(\underline{D}_{\infty,n}(f_p) - \underline{D}_{\infty,n}(f)\right)\underline{e}_n'$. Again we drop the dependence of $A_2$ on $\omega$, but it will play a role in the proof of Theorem 3.2.1. To bound the mean and variance of $\underline{X}_n' A_2 \underline{X}_n$ we use similar expressions to (3.15). Thus by using the same method described above leads to our requiring bounds for

$$\begin{aligned}&|\mathbb{E}[\underline{X}_n' A_2 \underline{X}_n]| \leqslant n^{-1/2}\sum_{t=1}^{n} |D_t(f_p) - D_t(f)| \left(\sum_{r\in\mathbb{Z}} |c(r)|\right) \\ &|\mathrm{tr}(A_2 R_n A_2 R_n)| \leqslant \left(n^{-1/2}\sum_{t=1}^{n} |D_t(f_p) - D_t(f)|\right)^2 \left(\sum_{r\in\mathbb{Z}} |c(r)|\right)^2 \\ &\sum_{s,t,u,v=1}^{n} |(A_2)_{s,t}(A_2)_{u,v}\mathrm{cum}\,(X_s, X_t, X_u, X_v)\,| \\ &\qquad\qquad \leqslant \left(n^{-1/2}\sum_{t=1}^{n} |D_t(f_p) - D_t(f)|\right)^2 \sum_{i,j,k\in\mathbb{Z}} |\kappa_4\,(i,j,k)\,|. \end{aligned} \tag{3.20}$$

The above three bounds require a bound for $\sum_{t=1}^{n} |D_t(f_p) - D_t(f)|$. To obtain such a bound we use the Lemma D.2.1 (as $p \to \infty$) that

$$\phi_t(\tau; f) = \sum_{j=0}^{\infty} \phi_{t+j}\psi_{|\tau|-j} \qquad \phi_t(\tau; f_p) = \sum_{j=0}^{\infty} \phi_{t+j,p}\psi_{|\tau|-j,p}$$

where $\{\phi_s\}_{s=1}^{\infty}$, $\{\phi_{s,p}\}_{s=1}^{p}$, $\{\psi_j\}_{j=0}^{\infty}$ and $\{\psi_{j,p}\}_{j=0}^{\infty}$ are the AR($\infty$), AR($p$) and MA($\infty$) coefficients

corresponding to the spectral density $f$ and $f_p$ respectively. Taking differences gives

$$\begin{aligned} n^{-1/2}\sum_{t=1}^{n}|D_t(f)-D_t(f_p)| &\leqslant 2n^{-1}\sum_{t=1}^{n}\sum_{\tau\leqslant 0}|\phi_t(\tau;f)-\phi_t(\tau;f_p)| \\ &\leqslant 2n^{-1}\sum_{t=1}^{n}\sum_{\tau\leqslant 0}\sum_{j=0}^{\infty}\left|a_{t+j}\psi_{|\tau|-j}-\phi_{t+j,p}\psi_{|\tau|-j,p}\right| \\ &\leqslant 2n^{-1}\sum_{t=1}^{n}\sum_{\tau\leqslant 0}\sum_{j=0}^{\infty}|\phi_{t+j}-\phi_{t+j,p}|\,|\psi_{|\tau|-j}| \\ &\quad +2n^{-1}\sum_{t=1}^{n}\sum_{\tau\leqslant 0}\sum_{j=0}^{\infty}\left|\psi_{|\tau|-j}-\psi_{|\tau|-j,p}\right|\,|\phi_{t+j,p}| = I_1+I_2. \end{aligned}$$

We consider first term $I_1$. Reordering the summands gives

$$\begin{aligned} I_1 &= 2n^{-1}\sum_{t=1}^{n}\sum_{j=0}^{\infty}|\phi_{t+j}-\phi_{t+j,p}|\sum_{\tau\leqslant 0}|\psi_{|\tau|-j}| \\ &\leqslant 2n^{-1}\sum_{\ell=0}^{\infty}|\psi_\ell|\sum_{t=1}^{n}\sum_{j=0}^{\infty}|\phi_{t+j}-\phi_{t+j,p}| \qquad (\text{let } u=t+j) \\ &\leqslant 2n^{-1}\sum_{\ell=0}^{\infty}|\psi_\ell|\sum_{u=0}^{\infty}u\,|\phi_u-\phi_{u,p}|\,. \end{aligned}$$

By applying the Baxter's inequality to the above we have

$$I_1 \leqslant 2(1+C)n^{-1}\sum_{\ell=0}^{\infty}|\psi_\ell|\sum_{u=p+1}^{\infty}|u\phi_u| = O\left(\frac{1}{np^{K-1}}\right).$$

To bound $I_2$ we use a similar method

$$\begin{aligned} I_2 &= 2n^{-1}\sum_{\tau\geqslant 0}\sum_{j=0}^{\infty}|\psi_{\tau-j}-\psi_{\tau-j,p}|\sum_{t=1}^{n}|\phi_{t+j,p}| \\ &\leqslant 2n^{-1}\sum_{t=1}^{p}|\phi_{t,p}|\sum_{u=0}^{\infty}u\,|\psi_u-\psi_{u,p}|\,. \end{aligned}$$

By using the inequality on page 2126 of Kreiss et al. (2011), for a large enough $n$, we have

$\sum_{u=0}^{\infty} u|\psi_u - \psi_{u,p}| \leqslant C \sum_{u=p+1}^{\infty} |u\phi_u| = O(p^{-K+1})$. Substituting this into the above gives

$$I_2 \leqslant Cn^{-1} \sum_{t=1}^{p} |\phi_{t,p}| \sum_{u=p+1}^{\infty} |u\phi_u| = O\left(\frac{1}{np^{K-1}}\right),$$

where we note that $\sup_t \sum_{t=1}^{p} |\phi_{t,p}| = O(1)$. Altogether this gives

$$n^{-1/2} \sum_{t=1}^{n} |D_t(f_p) - D_t(f)| \quad = \quad O\left(\frac{1}{np^{K-1}}\right).$$

Substituting the above bound into (3.20) and using a similar proof to Theorem 2.3.1, we get desired results. □

PROOF of (2.12) We note that

$$\widehat{J}_n(\omega; f)\overline{J_n(\omega)} = \left(\widehat{J}_n(\omega; f) - \widehat{J}_{\infty,n}(\omega; f)\right) \overline{J_n(\omega)} + \widehat{J}_{\infty,n}(\omega; f)\overline{J_n(\omega)}.$$

The mean and variance of the first term on the right hand side of the above was evaluated in Theorem 2.3.1 and has a lower order. Now we focus on the second term. Using the same methods as those given in (3.16) we have

$$\begin{aligned}
\left|\mathbb{E}[\widehat{J}_{\infty,n}(\omega; f)\overline{J_n(\omega)}]\right| &\leqslant n^{-1/2} \sum_{s,t=1}^{n} |(\underline{D}_{\infty,n}(f))_{(t)} e^{-is\omega} (R_n)_{t,s}| \\
&= n^{-1/2} \sum_{s,t=1}^{n} |D_t(f)||c(t-s)| \\
&\leqslant n^{-1/2} \sum_{t=1}^{n} |D_t(f)| \left(\sum_{r\in\mathbb{Z}} |c(r)|\right) \\
&\leqslant 2n^{-1} \sum_{t=1}^{n} \sum_{\tau\leqslant 0} |\phi_t(\tau; f)| \left(\sum_{r\in\mathbb{Z}} |c(r)|\right) = O(n^{-1}).
\end{aligned}$$

Following a similar argument for the variance we have $\mathrm{var}[\widehat{J}_{\infty,n}(\omega; f)\overline{J_n(\omega)}] = O(n^{-2})$ and this proves the equation (2.12) □

PROOF of Theorem 2.3.3. Consider the expansion

$$I_n(\omega;\widehat{f}_p) - I_n(\omega,f_p) = \left[\widehat{J}_n(\omega;\widehat{f}_p) - \widehat{J}_n(\omega;f_p)\right]\overline{J_n(\omega)} = E_n(\omega).$$

The main idea of the proof is to decompose $E_n(\omega)$ into terms whose expectation (and variance) can be evaluated plus an additional error whose expectation cannot be evaluated (since it involves ratios of random variables), but whose probabilistic bound is less than the expectation. We will make a Taylor expansion of the estimated parameters about the true parameters. The order of the Taylor expansion used will be determined by the order of summability of the cumulants in Assumption 2.3.2. For a given even $m$, the order of the Taylor expansion will be $(m/2-1)$. The reason for this will be clear in the proof, but roughly speaking we need to evaluate the mean and variance of the terms in the Taylor expansion. The higher the order of the expansion we make, the higher the cumulant asssumptions we require. To simplify the proof, we prove the result in the specific case that Assumption 2.3.2 holds for $m = 8$ (summability of all cumulants up to the 16th order). This, we will show, corresponds to making a third order Taylor expansion of the sample autocovariance function about the true autocovariance function. Note that the third order expansion requires summability of the 16th-order cumulants.

We now make the above discussion precise. By using equation (2.8) and (2.18) we have

$$\begin{aligned}
\widehat{J}_n(\omega;f_p) &= \frac{n^{-1/2}}{\phi_p(\omega)}\sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell}\phi_{\ell+s}e^{-is\omega} + e^{in\omega}\frac{n^{-1/2}}{\overline{\phi_p(\omega)}}\sum_{\ell=1}^{p} X_{n+1-\ell}\sum_{s=0}^{p-\ell}\phi_{\ell+s}e^{i(s+1)\omega} \\
&= \frac{1}{\sqrt{n}}\left(\sum_{\ell=1}^{p} X_\ell \frac{a_{\ell,p}(\omega)}{1-a_{0,p}(\omega)} + e^{i(n+1)\omega}\sum_{\ell=1}^{p} X_{n+1-\ell}\frac{\overline{a_{\ell,p}(\omega)}}{1-\overline{a_{0,p}(\omega)}}\right)
\end{aligned}$$

and

$$\widehat{J}_n(\omega;\widehat{f}_p) = \frac{1}{\sqrt{n}}\left(\sum_{\ell=1}^{p} X_\ell \frac{\widehat{a}_{\ell,p}(\omega)}{1-\widehat{a}_{0,p}(\omega)} + e^{i(n+1)\omega}\sum_{\ell=1}^{p} X_{n+1-\ell}\frac{\overline{\widehat{a}_{\ell,p}(\omega)}}{1-\overline{\widehat{a}_{0,p}(\omega)}}\right),$$

where for $\ell \geqslant 0$

$$a_{\ell,p}(\omega) = \sum_{s=0}^{p-\ell}\phi_{\ell+s}e^{-is\omega} \quad (a_0 \equiv 0)$$

and $\widehat{a}_{\ell,p}(\omega)$ is defined similarly but with the estimated Yule-Walker coefficients. Therefore

$$\widehat{J}_n(\omega; f_p) - \widehat{J}_n(\omega; f_p) = E_n(\omega)$$

where

$$\begin{aligned}
E_n(\omega) &= \frac{1}{n}\sum_{t=1}^{n}\sum_{\ell=1}^{p} X_\ell X_t e^{it\omega}\left[\frac{\widehat{a}_{\ell,p}(\omega)}{1-\widehat{a}_{0,p}(\omega)} - \frac{a_{\ell,p}(\omega)}{1-a_{0,p}(\omega)}\right] \\
&\quad + e^{i(n+1)\omega}\frac{1}{n}\sum_{t=1}^{n}\sum_{\ell=1}^{p} X_{n+1-\ell} X_t e^{it\omega}\left[\frac{\overline{\widehat{a}_{\ell,p}(\omega)}}{1-\overline{\widehat{a}_{0,p}(\omega)}} - \frac{\overline{a_{\ell,p}(\omega)}}{1-\overline{a_{0,p}(\omega)}}\right] \\
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{\ell=1}^{p} X_\ell X_t e^{it\omega}\left[g_{\ell,p}(\omega, \underline{\widehat{c}}_{p,n}) - g_{\ell,p}(\omega, \underline{c}_p)\right] \\
&\quad + e^{i(n+1)\omega}\frac{1}{n}\sum_{t=1}^{n}\sum_{\ell=1}^{p} X_{n+1-\ell} X_t e^{it\omega}\left[\overline{g_{\ell,p}(\omega, \underline{\widehat{c}}_{p,n})} - \overline{g_{\ell,p}(\omega, \underline{c}_p)}\right] \\
&= E_{n,L}(\omega) + E_{n,R}(\omega),
\end{aligned}$$

where $\underline{c}_p' = (c(0), c(1), \ldots, c(p))$, $\underline{\widehat{c}}_{p,n}' = (\widehat{c}_n(0), \widehat{c}_n(1), \ldots, \widehat{c}_n(p))$,

$$g_{\ell,p}(\omega, \underline{c}_{p,n}) = \frac{a_{\ell,p}(\omega)}{1-a_{0,p}(\omega)} \quad \text{and} \quad g_{\ell,p}(\omega, \underline{\widehat{c}}_{p,n}) = \frac{\widehat{a}_{\ell,p}(\omega)}{1-\widehat{a}_{0,p}(\omega)}. \tag{3.21}$$

For the notational convenience, we denote by $\{c_k\}$ and $\{\widehat{c}_k\}$ the autocovariances and sample autocovariances of the time series respectively.

Let $(R_p)_{s,t} = c(s-t)$, $(\underline{r}_p)_k = c(k)$, $(\widehat{R}_p)_{s,t} = \widehat{c}_n(s-t)$ and $(\underline{\widehat{r}}_p)_k = \widehat{c}_n(k)$. Then since Since $\underline{a}_p = R_p^{-1}\underline{r}_p$ and $\underline{\widehat{a}}_p = \widehat{R}_{p,n}^{-1}\underline{\widehat{r}}_{p,n}$, an explicit expression for $g_{\ell,p}(\omega, \underline{c}_p)$ and $g_{\ell,p}(\omega, \underline{\widehat{c}}_{p,n})$ is

$$g_{\ell,p}(\omega, \underline{c}_p) = \frac{\underline{r}_p' R_p^{-1}\underline{e}_\ell(\omega)}{1-\underline{r}_p' R_p^{-1}\underline{e}_0(\omega)} \quad \text{and} \quad g_{\ell,p}(\omega, \underline{\widehat{c}}_{p,n}) = \frac{\underline{\widehat{r}}_{p,n}' \widehat{R}_{p,n}^{-1}\underline{e}_\ell(\omega)}{1-\underline{\widehat{r}}_{p,n}' \widehat{R}_{p,n}^{-1}\underline{e}_0(\omega)}, \tag{3.22}$$

where $\underline{e}_\ell(\omega)$ are $p$-dimension vectors, with

$$\underline{e}_\ell(\omega)' = (\underbrace{0, \ldots, 0}_{\ell-\text{zeros}}, e^{-i\omega}, \ldots, e^{-i(p-\ell)\omega}) \text{ for } 0 \leqslant \ell \leqslant p. \tag{3.23}$$

Since $E_{n,L}(\omega)$ and $E_{n,R}(\omega)$ are near identical expressions, we will only study $E_{n,L}(\omega)$, noting the same analysis and bounds also apply to $E_{n,R}(\omega)$. We observe that the random functions $\widehat{a}_{\ell,p}(\omega)$ form the main part of $E_{n,L}(\omega)$. $\widehat{a}_{\ell,p}(\omega)$ are rather complex and directly evaluating their mean and variance is extremely difficult if not impossible. However, on careful examination we observe that they are functions of the autocovariance function whose sampling properties are well known. For this reason, we make a third order Taylor expansion of $g_{\ell,p}(\omega, \widehat{\underline{c}}_{p,n})$ about $g_{\ell,p}(\omega, \underline{c}_p)$:

$$\begin{aligned} g_{\ell,p}(\omega, \widehat{\underline{c}}_{p,n}) - g_{\ell,p}(\omega, \underline{c}_p) &= \sum_{j=0}^{p} (\widehat{c}_j - c_j) \frac{\partial g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_j} + \frac{1}{2!} \sum_{j_1,j_2=0}^{p} (\widehat{c}_{j_1} - c_{j_1}) (\widehat{c}_{j_2} - c_{j_2}) \frac{\partial^2 g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_{j_1} \partial c_{j_2}} \\ &\quad + \frac{1}{3!} \sum_{j_1,j_2,j_3=0}^{p} (\widehat{c}_{j_1} - c_{j_1}) (\widehat{c}_{j_2} - c_{j_2}) (\widehat{c}_{j_3} - c_{j_3}) \frac{\partial^3 g_{\ell,p}(\omega, \widetilde{\underline{c}}_{p,n})}{\partial \widetilde{c}_{j_1} \partial \widetilde{c}_{j_2} \partial \widetilde{c}_{j_3}} \end{aligned}$$

where $\widetilde{\underline{c}}_{p,n}$ is a convex combination of $\underline{c}_p$ and $\widehat{\underline{c}}_{p,n}$. Such an expansion draws the sample autocovariance function out of the sum, allowing us to evaluate the mean and variance for the first and second term. Substituting the third order expansion into $E_{n,L}(\omega)$ gives the sum

$$E_{n,L}(\omega) = \underbrace{E_{11}(\omega) + E_{12}(\omega) + E_{21}(\omega) + E_{22}(\omega)}_{=\Delta_{2,L}(\omega)} + \underbrace{E_{31}(\omega) + E_{32}(\omega)}_{=R_L(\omega)},$$

where

$$\begin{aligned} E_{11}(\omega) &= \sum_{j=0}^{p} \sum_{\ell=1}^{p} \frac{1}{n} \sum_{t=1}^{n} (X_\ell X_t - \mathbb{E}[X_\ell X_t]) e^{it\omega} (\widehat{c}_j - c_j) \frac{\partial g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_j} \\ E_{12}(\omega) &= \sum_{j=0}^{p} \sum_{\ell=1}^{p} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}[X_\ell X_t] e^{it\omega} (\widehat{c}_j - c_j) \frac{\partial g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_j} \\ E_{21}(\omega) &= \frac{1}{2} \sum_{j_1,j_2=0}^{p} \sum_{\ell=1}^{p} \frac{1}{n} \sum_{t=1}^{n} (X_\ell X_t - \mathbb{E}[X_\ell X_t]) e^{it\omega} (\widehat{c}_{j_1} - c_{j_1}) (\widehat{c}_{j_2} - c_{j_2}) \frac{\partial^2 g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_{j_1} \partial c_{j_2}} \\ E_{22}(\omega) &= \frac{1}{2} \sum_{j_1,j_2=0}^{p} \sum_{\ell=1}^{p} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}[X_\ell X_t] e^{it\omega} (\widehat{c}_{j_1} - c_{j_1}) (\widehat{c}_{j_2} - c_{j_2}) \frac{\partial^2 g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_{j_1} \partial c_{j_2}} \end{aligned}$$

and

$$E_{31}(\omega) = \frac{1}{3!}\sum_{j_1,j_2=0}^{p}\sum_{\ell=1}^{p}\frac{1}{n}\sum_{t=1}^{n}\left(X_\ell X_t - \mathbb{E}[X_\ell X_t]\right)e^{it\omega}\left(\widehat{c}_{j_1}-c_{j_1}\right)\left(\widehat{c}_{j_2}-c_{j_2}\right)\left(\widehat{c}_{j_3}-c_{j_3}\right)\frac{\partial^3 g_{\ell,p}(\omega,\widetilde{\underline{c}}_{p,n})}{\partial\widetilde{c}_{j_1}\partial\widetilde{c}_{j_2}\partial\widetilde{c}_{j_3}}$$

$$E_{32}(\omega) = \frac{1}{3!}\sum_{j_1,j_2,j_3=0}^{p}\sum_{\ell=1}^{p}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}[X_\ell X_t]e^{it\omega}\left(\widehat{c}_{j_1,n}-c_{j_1}\right)\left(\widehat{c}_{j_2}-c_{j_2}\right)\left(\widehat{c}_{j_3}-c_{j_3}\right)\frac{\partial^3 g_{\ell,p}(\omega,\widetilde{\underline{c}}_{p})}{\partial\widetilde{c}_{j_1}\partial\widetilde{c}_{j_2}\partial\widetilde{c}_{j_3}}.$$

Our aim is to evaluate the expectation and variance of $E_{11}(\omega)$, $E_{12}(\omega), E_{21}(\omega)$ and $E_{22}(\omega)$. This will give the asymptotic bias of $I_n(\omega,\widehat{f}_p)$ in the sense of Bartlett (1953). Further we show that $E_{31}(\omega)$, $E_{32}(\omega)$ are both of lower order in probabilistic sense. To do so, we define some additional notations. Let

$$\breve{\mu}_\ell(\omega) = n^{-1}\sum_{t=1}^{n}(X_tX_\ell - \mathbb{E}[X_tX_\ell])e^{it\omega} \quad \text{and} \quad \breve{c}_j = \widehat{c}_{j,n} - \mathbb{E}[\widehat{c}_{j,n}].$$

For $I=\{i_1,...,i_r\}$ and $J=\{j_1,...,j_s\}$, define the joint cumulant of an order $(r+s)$

$$\operatorname{cum}\left(\breve{\mu}_I^{\otimes^r},\breve{c}_J^{\otimes^s}\right) = \operatorname{cum}\left(\breve{\mu}_{i_1}(\omega),\ldots,\breve{\mu}_{i_r}(\omega),\breve{c}_{j_1},\ldots,\breve{c}_{j_s}\right).$$

Note that in the proofs below we often suppress the notation $\omega$ in $\breve{\mu}_\ell(\omega)$ to make the notation less cumbersome. To further reduce notation define the "half" spectral density

$$f_{\ell,n}(\omega) = \sum_{t=1}^{n}\mathbb{E}[X_tX_\ell]e^{it\omega}.$$

We note that since $\mathbb{E}[X_tX_\ell] = c(t-\ell)$ and by assumption of absolute summability of the autocovariance function we have the bound

$$\sup_{\omega,\ell,n}|f_{\ell,n}(\omega)| \leqslant \sum_{r\in\mathbb{Z}}|c(r)| < \infty. \tag{3.24}$$

Using the notation above we can write $E_{11}(\omega)$, $E_{21}(\omega)$ and $E_{31}(\omega)$ as

$$\begin{aligned}
E_{11}(\omega) &= \sum_{j=0}^{p}\sum_{\ell=1}^{p} \breve{\mu}_\ell \left(\widehat{c}_j - c_j\right) \frac{\partial g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_j}, \\
E_{21}(\omega) &= \frac{1}{2}\sum_{j_1,j_2=0}^{p}\sum_{\ell=1}^{p} \breve{\mu}_\ell \left(\widehat{c}_{j_1} - c_{j_1}\right)\left(\widehat{c}_{j_2} - c_{j_2}\right) \frac{\partial^2 g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_{j_1}\partial c_{j_2}}, \\
E_{31}(\omega) &= \frac{1}{3!}\sum_{j_1,j_2,j_3=0}^{p}\sum_{\ell=1}^{p} \breve{\mu}_\ell \left(\widehat{c}_{j_1} - c_{j_1}\right)\left(\widehat{c}_{j_2} - c_{j_2}\right)\left(\widehat{c}_{j_3} - c_{j_3}\right) \frac{\partial^3 g_{\ell,p}(\omega, \widetilde{\underline{c}}_{p,n})}{\partial \widetilde{c}_{j_1}\partial \widetilde{c}_{j_2}\partial \widetilde{c}_{j_3}}
\end{aligned} \tag{3.25}$$

Using Lemma D.1.3, We summarize the pertinent bounds from the above. The first order expansion yields the bounds

$$\begin{aligned}
\mathbb{E}[E_{11}(\omega)] &= O\left(\frac{p^2}{n^2}\right), \quad \text{var}[E_{11}(\omega)] = O\left(\frac{p^4}{n^2}\right), \\
\mathbb{E}[E_{12}(\omega)] &= O\left(\frac{p^2}{n^2}\right), \quad \text{var}[E_{12}(\omega)] = O\left(\frac{p^4}{n^3}\right).
\end{aligned}$$

The second order expansion yields the bounds

$$\begin{aligned}
\mathbb{E}[E_{21}(\omega)] &= O\left(\frac{p^3}{n^2}\right), \quad \text{var}[E_{21}(\omega)] = O\left(\frac{p^6}{n^3}\right), \\
\mathbb{E}[E_{22}(\omega)] &= O\left(\frac{p^3}{n^2}\right), \quad \text{var}[E_{22}(\omega)] = O\left(\frac{p^6}{n^4}\right).
\end{aligned}$$

All together, the third order expansion yields the probabilistic bounds

$$E_{31}(\omega) = O_p\left(\frac{p^4}{n^2}\right) \qquad E_{32}(\omega) = O_p\left(\frac{p^4}{n^{5/2}}\right).$$

The above are bounds hold for the expansion of $E_{n,L}(\omega)$. A similar set of bounds also apply to $E_{n,R}(\omega)$. Then we have

$$\mathbb{E}[\Delta_2(\omega)] = O\left(\frac{p^3}{n^2}\right) \quad \text{var}[\Delta_2(\omega)] = O\left(\frac{p^4}{n^2}\right).$$

On the other hand

$$R_n(\omega) \;=\; O_p\left(\frac{p^4}{n^2}\right).$$

This proves the result for $m = 8$. The proof for $m = 6$ and all even $m > 8$ is similar, just the order of the Taylor expansion needs to be adjusted accordingly. □

### 3.6.3 Proof of Section 3

PROOF of Theorem 3.1.1 The proof is almost identical with the proof of Theorems 2.3.1−2.3.3, thus we only give a brief outline. As with Theorems 2.3.1−2.3.3 we can show that

$$\begin{aligned}\left(J_n(\omega) + \widehat{J}_{\infty,n}(\omega; f)\right)\overline{J_{\underline{h},n}(\omega)} &= I_{\underline{h},n}(\omega; f) + \Delta^{(0)}_{\underline{h},n}(\omega)\\ I_{\underline{h},n}(\omega; f_p) &= \left(J_n(\omega) + \widehat{J}_{\infty,n}(\omega; f)\right)\overline{J_{\underline{h},n}(\omega)} + \Delta^{(1)}_{\underline{h},n}(\omega)\\ I_{\underline{h},n}(\omega; \widehat{f}_p) &= I_{\underline{h},n}(\omega; f_p) + \Delta^{(2)}_{\underline{h},n}(\omega) + R_{\underline{h},n}(\omega).\end{aligned}$$

Since $\sup_t h_{t,n} \leqslant C$ for some constant, it is easy to verify that $|\Delta^{(i)}_{\underline{h},n}(\omega)| \leqslant C|\Delta_{i,n}(\omega)|$ for $i = 0, 1, 2$ and $|R_{\underline{h},n}(\omega)| \leqslant C|R_n(\omega)|$, where where $\Delta_{0,n}(\omega)$, $\Delta_{1,n}(\omega)$, $\Delta_{2,n}(\omega)$ and $R_n(\omega)$ are the error terms from Theorems 2.3.1−2.3.3. Thus by using the bounds in Theorems 2.3.1−2.3.3 we have proved the result. □

PROOF of Theorem 3.2.1 To simplify notation we focus on the case that the regular DFT is not tapered and consider the case that $A_{x,n}(g; f)$ is a sum (and not an integral). We will use the sequence of approximations in Theorems 2.3.1−2.3.3. We will obtain bounds between the "ideal" criterion $A_{S,n}(g; f)$ and the intermediate terms. Define the infinite predictor integrated sum as

$$A_{\infty,S,n}(g; f) = \frac{1}{n}\sum_{k=1}^{n} g(\omega_{k,n}) I_{\infty,n}(\omega_{k,n}; f).$$

We use the sequence of differences to prove the result:

$$\begin{aligned} A_{S,n}(g;\widehat{f}_p) - A_{S,n}(g;f) &= (A_{S,n}(g;\widehat{f}_p) - A_{S,n}(g;f_p)) + (A_{S,n}(g;f_p) - A_{\infty,S,n}(g;f)) \\ &\quad +(A_{\infty,S,n}(g;f) - A_{S,n}(g;f)). \end{aligned} \tag{3.26}$$

We start with the third term $A_{\infty,S,n}(g;f) - A_{S,n}(g;f)$

$$\begin{aligned} |A_{S,n}(g;f) - A_{\infty,S,n}(g;f)| &\leqslant \frac{1}{n}\sum_{k=1}^{n} |g(\omega_{k,n})| \left| \left( \widehat{J}_n(\omega_{k,n};f) - \widehat{J}_{\infty,n}(\omega_{k,n};f) \right) \overline{J_n(\omega_{k,n})} \right| \\ &= \sup_{\omega} \left| \left( \widehat{J}_n(\omega;f) - \widehat{J}_{\infty,n}(\omega;f) \right) \overline{J_n(\omega)} \right| \cdot \frac{1}{n}\sum_{k=1}^{n} |g(\omega_{k,n})| = R_0. \end{aligned}$$

Using Theorem 2.3.1, we have that $\mathbb{E}[R_0] = O(n^{-K})$ and $\mathrm{var}[R_0] = O(n^{-2K})$. Using a similar method we can show that the second term of above

$$|A_{\infty,S,n}(g;f_p) - A_{S,n}(g;f)| \leqslant \sup_{\omega} \left| \left( \widehat{J}_{\infty,n}(\omega;f) - \widehat{J}_n(\omega;f_p) \right) \overline{J_n(\omega)} \right| \cdot \frac{1}{n}\sum_{k=1}^{n} |g(\omega_{k,n})| = R_1$$

where $\mathbb{E}[R_1] = O(n^{-1}p^{-K+1})$ and $\mathrm{var}[R_1] = O(n^{-2}p^{-2K+2})$.

To bound the first term $A_{S,n}(g;\widehat{f}_p) - A_{S,n}(g;f_p)$ a little more care is required. We use the expansion and notation from the proof of Theorem 2.3.3;

$$A_{S,n}(g;\widehat{f}_p) - A_{S,n}(g;f_p) = U_L + U_R$$

where

$$U_L = \frac{1}{n}\sum_{k=1}^{n} g(\omega_{k,n}) E_{n,L}(\omega_{k,n}) \quad \text{and} \quad U_R = \frac{1}{n}\sum_{k=1}^{n} g(\omega_{k,n}) E_{n,R}(\omega_{k,n}).$$

We further decompose $U_L$ into

$$\begin{aligned} U_L &= \frac{1}{n}\sum_{k=1}^{n} g(\omega_{k,n})[E_{111}(\omega_{k,n}) + E_{112}(\omega_{k,n}) + E_{12}(\omega_{k,n}) + E_{21}(\omega_{k,n}) + E_{22}(\omega_{k,n}) \\ &\qquad + E_{31}(\omega_{k,n}) + E_{32}(\omega_{k,n})] = U_{1,n} + U_{2,n} + U_{3,n}, \end{aligned}$$

where

$$\begin{aligned} U_{1,n} &= \frac{1}{n}\sum_{k=1}^{n} g(\omega_{k,n}) E_{111}(\omega_{k,n}) \\ U_{2,n} &= \frac{1}{n}\sum_{k=1}^{n} g(\omega_{k,n}) \left[E_{112}(\omega_{k,n}) + E_{12}(\omega_{k,n}) + E_{21}(\omega_{k,n}) + E_{22}(\omega_{k,n})\right] \\ U_{3,n} &= \frac{1}{n}\sum_{k=1}^{n} g(\omega_{k,n}) \left[E_{31}(\omega_{k,n}) + E_{32}(\omega_{k,n})\right]. \end{aligned}$$

We note that a similar decomposition applies to the right hand decomposition, $U_R$. Thus the bounds we obtain for $U_L$ can also be applied to $U_R$. To bound $U_{i,n}$ for $i = 1, 2, 3$, we will treat the terms differently. Since

$$|U_{2,n}| \leqslant \sup_{\omega} \left(|E_{112}(\omega)| + |E_{12}(\omega)| + |E_{21}(\omega)| + |E_{22}(\omega)|\right) \cdot \frac{1}{n}\sum_{k=1}^{n} |g(\omega_{k,n})|,$$

we can use the bounds in the Lemma D.1.3 in Appendix to show that $\mathbb{E}[U_{2,n}] = O(p^3 n^{-2})$ and $\mathrm{var}[U_{2,n}] = O(p^6 n^{-3})$. Similarly we can show that $U_{3,n} = O_p(p^{m/2} n^{-m/4})$. However, directly applying the bounds for $E_{111}(\omega)$ to bound $U_{1,n}$ leads to a suboptimal bound for the variance (of order $p^4/n^2$). By applying a more subtle approach, we utilize the sum over $k$. By using Lemma D.1.3 in Appendix, we can show that $\mathbb{E}[U_{1,n}] = O(p^2 n^{-2})$. To obtain the variance we expand

$\mathrm{var}[U_{1,n}]$

$$\begin{aligned}
\mathrm{var}[U_{1,n}] &= \frac{1}{n^2}\sum_{k_1,k_2=1}^{n} g(\omega_{k_1,n})g(\omega_{k_2,n})\mathrm{cov}[E_{111}(\omega_{k_1,n}), E_{111}(\omega_{k_2,n})] \\
&= \frac{1}{n^2}\sum_{k_1,k_2=1}^{n} g(\omega_{k_1,n})g(\omega_{k_2,n}) \times \\
&\quad \sum_{\ell_1,\ell_2=1}^{p}\sum_{j_1,j_2=0}^{p} \mathrm{cov}\left(\breve{\mu}_{\ell_1}(\omega_{k_1,n})\breve{c}_{j_1}, \breve{\mu}_{\ell_2}(\omega_{k_2,n})\breve{c}_{j_2}\right)\right) \frac{\partial g_{\ell_1,p}(\omega_{k_1,n},\underline{c}_p)}{\partial c_{j_1}} \frac{\partial g_{\ell_2,p}(\omega_{k_2,n},\underline{c}_p)}{\partial c_{j_2}} \\
&= T_1 + T_2 + T_3
\end{aligned}$$

where

$$\begin{aligned}
T_1 &= \frac{1}{n^2}\sum_{k_1,k_2=1}^{n}\sum_{\ell_1,\ell_2=1}^{p}\sum_{j_1,j_2=0}^{p} h_{j_1,j_2}(\omega_{k_1,n},\omega_{k_2,n})\mathrm{cov}\left[\breve{\mu}_{\ell_1}(\omega_{k_1,n}), \breve{\mu}_{\ell_2}(\omega_{k_2,n})\right]\mathrm{cov}\left[\breve{c}_{j_1}, \breve{c}_{j_2}\right] \\
T_2 &= \frac{1}{n^2}\sum_{k_1,k_2=1}^{n}\sum_{\ell_1,\ell_2=1}^{p}\sum_{j_1,j_2=0}^{p} h_{j_1,j_2}(\omega_{k_1,n},\omega_{k_2,n})\mathrm{cov}\left[\breve{\mu}_{\ell_1}(\omega_{k_1,n}), \breve{c}_{j_2}\right]\mathrm{cov}\left[\breve{\mu}_{\ell_2}(\omega_{k_2,n}), \breve{c}_{j_1}\right] \\
T_3 &= \frac{1}{n^2}\sum_{k_1,k_2=1}^{n}\sum_{\ell_1,\ell_2=1}^{p}\sum_{j_1,j_2=0}^{p} h_{j_1,j_2}(\omega_{k_1,n},\omega_{k_2,n})\mathrm{cum}\left[\breve{\mu}_{\ell_1}(\omega_{k_1,n}), \breve{\mu}_{\ell_2}(\omega_{k_2,n}), \breve{c}_{j_1}, \breve{c}_{j_2}\right]
\end{aligned}$$

and $h_{j_1,j_2}(\omega_{k_1,n},\omega_{k_2,n}) = g(\omega_{k_1,n})g(\omega_{k_2,n}) \cdot \partial g_{\ell_1,p}(\omega_{k_1,n},\underline{c}_p)/\partial c_{j_1} \cdot \partial g_{\ell_2,p}(\omega_{k_2,n},\underline{c}_p)/\partial c_{j_2}$. Then, by Lemma D.1.2, we have

$$\sup_{0\leqslant j_1,j_2\leqslant p}\sup_{\omega_1,\omega_2}|h_{j_1,j_2}(\omega_1,\omega_2)| \leqslant C < \infty.$$

To bound above three terms, we first consider $T_2$. We directly apply Lemma D.1.1 and this gives $\mathrm{cov}\left[\breve{\mu}_{\ell_1}(\omega_{k_1,n}), \breve{c}_{j_2}\right] \cdot \mathrm{cov}\left[\breve{\mu}_{\ell_2}(\omega_{k_2,n}), \breve{c}_{j_1}\right] = O(n^{-4})$ and thus $T_2 = O(p^4 n^{-4})$.

To bound $T_1$, we expand $\mathrm{cov}\left[\breve{\mu}_{\ell_1}(\omega_{k_1,n}), \breve{\mu}_{\ell_2}(\omega_{k_2,n})\right]$

$$\begin{aligned}\mathrm{cov}\left[\breve{\mu}_{\ell_1}(\omega_{k_1,n}), \breve{\mu}_{\ell_2}(\omega_{k_2,n})\right] &= \frac{1}{n^2}\sum_{t_1,t_2=1}^{n}\Big(c(t_1-t_2)c(\ell_1-\ell_2)+c(t_1-\ell_2)c(t_2-\ell_1) \\ &\quad +\kappa_4(\ell_1-t_1,t_2-t_1,\ell_2-t_1)\Big)e^{it_1\omega_{k_1,n}-it_2\omega_{k_2,n}} \\ &= \frac{1}{n^2}\sum_{t_1,t_2=1}^{n} C_{\ell_1,\ell_2}(t_1,t_2)e^{it_1\omega_{k_1,n}-it_2\omega_{k_2,n}}.\end{aligned}$$

Substituting the above into $T_1$

$$T_1 = \frac{1}{n^2}\sum_{\ell_1,\ell_2=1}^{p}\sum_{j_1,j_2=0}^{p}\mathrm{cov}\left[\breve{c}_{j_1},\breve{c}_{j_2}\right]\sum_{t_1,t_2=1}^{n}C_{\ell_1,\ell_2}(t_1,t_2)\frac{1}{n^2}\sum_{k_1,k_2=1}^{n}h_{j_1,j_2}(\omega_{k_1,n},\omega_{k_2,n})e^{it_1\omega_{k_1,n}-it_2\omega_{k_2,n}}.$$

Since by assumption the function $g(\cdot)$ and its derivative are continuous on the torus $[0,2\pi]$ and $h_{j_1,j_2}(\cdot,\cdot)$ and its partial derivatives are continuous of $[0,2\pi]^2$, then by the Poisson summation formula

$$\frac{1}{n^2}\sum_{k_1,k_2=1}^{n}h_{j_1,j_2}(\omega_{k_1,n},\omega_{k_2,n})e^{it_1\omega_{k_1,n}-it_2\omega_{k_2,n}} = \sum_{s_1,s_2\in\mathbb{Z}}a^{(j_1,j_2)}(t_1+s_1n,-t_2+s_2n)$$

where $a^{(j_1,j_2)}(r_1,r_2)$ are the $(r_1,r_2)th$ Fourier coefficients of $h_{j_1,j_2}(\cdot,\cdot)$ and are absolutely summable. Substituting the above into $T_1$ and by Lemma D.1.2,

$$\begin{aligned}|T_1| &\leqslant \frac{1}{n^2}\sum_{\ell_1,\ell_2=1}^{p}\sum_{j_1,j_2=0}^{p}|\mathrm{cov}\left[\breve{c}_{j_1},\breve{c}_{j_2}\right]|\sum_{t_1,t_2=1}^{n}\sum_{s_1,s_2\in\mathbb{Z}}|C_{\ell_1,\ell_2}(t_1,t_2)|\cdot|a^{(j_1,j_2)}(t_1+s_1n,-t_2+s_2n)| \\ &\leqslant \frac{C}{n^3}\sum_{\ell_1,\ell_2=1}^{p}\sum_{j_1,j_2=0}^{p}\sum_{t_1,t_2=1}^{n}\sum_{s_1,s_2\in\mathbb{Z}}|a^{(j_1,j_2)}(t_1+s_1n,-t_2+s_2n)| \\ &= \frac{Cp^2}{n^3}\sum_{j_1,j_2=0}^{p}\sum_{r_1,r_2\in\mathbb{Z}}|a^{(j_1,j_2)}(r_1,r_2)| = O\left(\frac{p^4}{n^3}\right).\end{aligned}$$

Therefore, $T_1 = O(p^4n^{-3})$. Finally, we consider $T_3$. We use the expansions for $\mathrm{cum}[\breve{\mu}_{\ell_1}(\omega_{k_1,n}), \breve{\mu}_{\ell_2}(\omega_{k_2,n}), \breve{c}_{j_1}, \breve{c}_{j_2}]$ given in the proof of Lemma D.1.1 together with the same proof used to bound $T_1$. This once again gives the bound $T_3 = O(p^4n^{-3})$. Putting these bounds together gives

(i) $\mathbb{E}[U_{1,n}] = O(p^2n^{-2})$ and $\mathrm{var}[U_{1,n}] = O(p^4n^{-3})$.

(ii) $\mathbb{E}[U_{2,n}] = O(p^3n^{-2})$ and $\mathrm{var}[U_{2,n}] = O(p^6n^{-3})$

(iii) $U_{3,n} = O_p(p^{m/2}n^{-m/4})$.

The above covers $U_L$. The same set of bounds apply to $U_R$. Thus altogether we have that

$$A_{S,n}(g;\widehat{f}_p) - A_{S,n}(g;f_p) = U_L + U_R = R_2 + \mathcal{E},$$

where $R_2$ is the term whose mean and variance can be evaluated and is $\mathbb{E}[R_2] = O(p^2n^{-2})$ and $\mathrm{var}[R_2] = O(p^6n^{-3})$ and $\mathcal{E}$ is the term which has probabilistic bound $\mathcal{E} = O_p(p^{m/2}n^{-m/4})$. Finally, placing all the bounds into (3.26) we have

$$A_{S,n}(g;\widehat{f}_p) - A_{S,n}(g;f) = R_0 + R_1 + R_2 + \mathcal{E} = \Delta(g) + \mathcal{E},$$

where $\mathbb{E}[\Delta(g)] = O(n^{-1}p^{-K+1}+p^2n^{-2})$, $\mathrm{var}[\Delta(g)] = O(n^{-2}p^{-K-2}+p^6n^{-3})$ and $\mathcal{E} = O_p(p^{m/2}n^{-m/4})$ thus yielding the desired result. □

PROOF of Corollary 3.2.1. We prove the result for $A_{I,n}(g;\widehat{f}_p)$, noting that a similar result holds for $A_{S,n}(g;\widehat{f}_p)$. We recall

$$\begin{aligned} A_{I,n}(g;\widehat{f}_p) &= \frac{1}{2\pi}\int_0^{2\pi} g(\omega) I_{\underline{h},n}(\omega;\widehat{f}_p)d\omega \\ &= \frac{1}{2\pi}\int_0^{2\pi} g(\omega) J_n(\omega)\overline{J_{\underline{h},n}(\omega)}d\omega + \frac{1}{2\pi}\int_0^{2\pi} g(\omega)\left(\widehat{J}_n(\omega;\widehat{f}_p) - \widehat{J}_n(\omega;f)\right)\overline{J_{\underline{h},n}(\omega)}d\omega \\ &\quad + \frac{1}{2\pi}\int_0^{2\pi} g(\omega)\widehat{J}_n(\omega;f)\overline{J_{\underline{h},n}(\omega)}d\omega \end{aligned} \tag{3.27}$$

Using Theorem 3.2.1, we can bound the second term

$$\left|\frac{1}{2\pi}\int_0^{2\pi} g(\omega)\left(\widehat{J}_n(\omega;\widehat{f}_p) - \widehat{J}_n(\omega;f)\right)\overline{J_{\underline{h},n}(\omega)}d\omega\right| = O_p\left(\frac{p^{m/2}}{n^{m/4}} + \frac{1}{np^{K-1}} + \frac{p^3}{n^{3/2}}\right).$$

For the third term, we use similar technique to prove equation (2.12), we have $\widehat{J}_n(\omega; f)\overline{J_{\underline{h},n}(\omega)} = O_p(n^{-1})$. Therefore, integrability of $g$ gives that the third term in (3.27) is $O_p(n^{-1})$. Combining above results, for $m \geqslant 6$ where $m$ from Assumption 4.2.1

$$\begin{aligned}
&\frac{1}{2\pi}\int_0^{2\pi} g(\omega)\widehat{J}_n(\omega; \widehat{f}_p)\overline{J_{\underline{h},n}(\omega)}d\omega \\
&= \frac{1}{2\pi}\int_0^{2\pi} g(\omega)\left(\widehat{J}_n(\omega; \widehat{f}_p) - \widehat{J}_n(\omega; f)\right)\overline{J_{\underline{h},n}(\omega)}d\omega + \frac{1}{2\pi}\int_0^{2\pi} g(\omega)\widehat{J}_n(\omega; f)\overline{J_{\underline{h},n}(\omega)}d\omega \\
&= O_p\left(\frac{1}{n} + \frac{p^{m/2}}{n^{m/4}} + \frac{1}{np^{K-1}} + \frac{p^3}{n^{3/2}}\right) = O_p\left(\frac{1}{n} + \frac{p^3}{n^{3/2}}\right). \qquad (3.28)
\end{aligned}$$

Thus we focus on the first term of (3.27), which we define as

$$A_{\underline{h},n}(g) = \frac{1}{2\pi}\int_0^{2\pi} g(\omega)J_n(\omega)\overline{J_{\underline{h},n}(\omega)}d\omega.$$

From (3.28) if

$$\frac{H_{1,n}}{H_{2,n}^{1/2}}\left(\frac{1}{n} + \frac{p^3}{n^{3/2}}\right) \to 0$$

as $p, n \to \infty$, then $(H_{1,n}/H_{2,n}^{1/2})A_{\underline{h},n}(g)$ is the dominating term in $(H_{1,n}/H_{2,n}^{1/2})A_{I,n}(g; \widehat{f}_p)$. Moreover, by Cauchy-Schwarz inequality, we have $H_{1,n}/H_{2,n}^{1/2} \leqslant n^{1/2}$, thus we can omit the first term of the above condition and get condition (3.6).

Finally, by applying the techniques in Dahlhaus (1983) to $(H_{1,n}/H_{2,n}^{1/2})A_{\underline{h},n}(g)$ we can show that

$$\frac{H_{1,n}^2}{H_{2,n}}\text{var}[A_{\underline{h},n}(g)] = (V_1 + V_2 + V_3) + o(1).$$

Since $(H_{1,n}/H_{2,n}^{1/2})A_{I,n}(g; \widehat{f}_p) = (H_{1,n}/H_{2,n}^{1/2})A_{\underline{h},n}(g) + o_p(1)$, this proves the result. □

# 4. THE GAUSSIAN LIKELIHOOD IN THE FREQUENCY DOMAIN *

In this section, we discuss greater detail of the frequency domain representation of the Gaussian likelihood in Section 2.4.

## 4.1 New frequency domain quasi-likelihoods

In this section, we apply the approximations from Section 2.4.1 to define two new spectral divergence criteria.

To motivate the criteria, we recall from Theorem 2.4.1 that the Gaussian likelihood can be written as a contrast between $\widetilde{J}_n(\omega; f_\theta)\overline{J_n(\omega)}$ and $f_\theta(\omega)$. The resulting estimator is based on simultaneously predicting and fitting the spectral density. In the case that the model is correctly specified, in the sense there exists a $\theta \in \Theta$ where $f = f_\theta$ (and $f$ is the true spectral density). Then

$$\mathbb{E}_{f_\theta}[\widetilde{J}_n(\omega; f_\theta)\overline{J_n(\omega)}] = f_\theta(\omega)$$

and the Gaussian criterion has a clear interpretation. However, if the model is misspecified (which for real data is likely), $\mathbb{E}_f[\widetilde{J}_n(\omega; f_\theta)\overline{J_n(\omega)}]$ has no clear interpretation. Instead, to understand what the Gaussian likelihood is estimating, we use that $\mathbb{E}_f[\widehat{J}_n(\omega; f_\theta)\overline{J_n(\omega)}] = O(n^{-1})$, which leads to the approximation $\mathbb{E}_f[\widetilde{J}_n(\omega; f_\theta)\overline{J_n(\omega)}] = f(\omega) + O(n^{-1})$. From this, we observe that the expected negative log Gaussian likelihood is

$$n^{-1}\mathbb{E}_f[\underline{X}_n'\Gamma_n(f_\theta)^{-1}\underline{X}_n] + n^{-1}\log|\Gamma_n(f_\theta)| = I(f, f_\theta) + O(n^{-1}),$$

where

$$I_n(f; f_\theta) = \frac{1}{n}\sum_{k=1}^{n}\left(\frac{f(\omega_{k,n})}{f_\theta(\omega_{k,n})} + \log f_\theta(\omega_{k,n})\right). \tag{4.1}$$

Since $I_n(f; f_\theta)$ is the spectral divergence between the true spectral $f$ density and parametric spec-

*Parts of this section have been modified with permission from [S. Subba Rao and J. Yang. Reconciling the Gaussian and Whittle likelihood with an application to estimation in the frequency domain. *Annals of Statistics (To appear), arXiv:2001.06966*, 2021.]

tral density $f_\theta$, asymptotically the misspecified Gaussian likelihood estimator has a meaningful interpretation. However, there is still a finite sample bias in the Gaussian likelihood of order $O(n^{-1})$. This can have a knock-on effect, by increasing the finite sample bias in the resulting Gaussian likelihood estimator. To remedy this, in the following section, we obtain a frequency domain criterion which approximates the spectral divergence $I_n(f; f_\theta)$ to a greater degree of accuracy. This may lead to estimators which may give a more accurate fit of the underlying spectral density. We should emphasis at this point, that reducing the bias in the likelihood, does not necessarily translate to a provable reduction in the bias of the resulting estimators. It is worth noting that, strictly, the spectral divergence is defined as $n^{-1}\sum_{k=1}^{n}\left(\frac{f(\omega_{k,n})}{f_\theta(\omega_{k,n})} - \log\frac{f(\omega_{k,n})}{f_\theta(\omega_{k,n})} - 1\right)$. It is zero when $f_\theta = f$ and positive for other values of $f_\theta$. But since $-\log f - 1$ does not depend on $\theta$ we ignore this term.

### 4.1.1 The boundary corrected Whittle likelihood

In order to address some of the issues raised above, we recall from Theorem 2.2.1 that $\mathbb{E}_f[\widetilde{J}_n(\omega; f)\overline{J_n(\omega)}] = f(\omega)$. In other words, by predicting over the boundary using the (unobserved) spectral density which *generates* the data, the "complete periodogram" $\widetilde{J}_n(\omega; f)\overline{J_n(\omega)}$ is an inconsistent but *unbiased* of the true spectral density $f$. This motivates the (infeasible) boundary corrected Whittle likelihood

$$W_n(\theta) = \frac{1}{n}\sum_{k=1}^{n}\frac{\widetilde{J}_n(\omega_{k,n}; f)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} + \frac{1}{n}\sum_{k=1}^{n}\log f_\theta(\omega_{k,n}). \tag{4.2}$$

Thus, if $\{X_t\}$ is a second order stationary time series with spectral density $f$, then we have $\mathbb{E}_f[W_n(\theta)] = I_n(f; f_\theta)$.

Of course $f$ and thus $\widetilde{J}_n(\omega_{k,n}; f)$ are unknown. However, using steps of approximation in Section 2.3.2, we show that the predictive DFT (and thus the complete DFT) can be well approximated with relatively small error. The first step is to replacing $f$ in $\widetilde{J}_n(\omega_{k,n}; f)$ with the spectral density function corresponding to the best fitting AR($p$) process $\widetilde{J}_n(\omega_{k,n}; f_p)$, where an analytic form is given in (2.8). Since we have replaced $f$ with $f_p$, the "periodogram" $\widetilde{J}_n(\omega_{k,n}; f_p)\overline{J_n(\omega_{k,n})}$ does have a bias, but it is considerably smaller than the bias of the usual periodogram. In particular, it

follows from Theorems 2.3.1 and 2.3.2 that

$$\mathbb{E}_f[\widetilde{J}_n(\omega_{k,n};f_p)\overline{J_n(\omega_{k,n})}] = f(\omega_{k,n}) + O\left(\frac{1}{np^{K-1}}\right).$$

The above result leads to an approximation of the boundary corrected Whittle likelihood

$$W_{p,n}(\theta) = \frac{1}{n}\sum_{k=1}^{n}\frac{\widetilde{J}_n(\omega_{k,n};f_p)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} + \frac{1}{n}\sum_{k=1}^{n}\log f_\theta(\omega_{k,n}). \tag{4.3}$$

In the following lemma, we obtain a bound between the "ideal" boundary corrected Whittle likelihood $W_n(\theta)$ and $W_{p,n}(\theta)$.

**Lemma 4.1.1.** *Suppose $f$ satisfies Assumption 2.3.1, $f_\theta$ is bounded away from zero and $\|f_\theta\|_0 < \infty$. Let $\{a_j(p)\}$ denote the coefficients of the best fitting AR($p$) model corresponding to the spectral density $f$ and define $f_p(\omega) = |1 - \sum_{j=1}^{p} a_j(p)e^{-ij\omega}|^{-2}$. Suppose $1 \leqslant p < n$, then we have*

$$\begin{aligned}&\left\|F_n^*\Delta_n(f_\theta^{-1})\left(D_n(f) - D_n(f_p)\right)\right\|_1 \\ &\quad\leqslant \rho_{p,K}(f)A_K(f,f_\theta)\left(\frac{(C_{f,1}+1)}{p^{K-1}} + \frac{2(C_{f,1}+1)^2}{p^K}\|\psi_f\|_0\|\phi_f\|_1 + \frac{C_{f,0}}{n^{K-1}}\right).\end{aligned} \tag{4.4}$$

*Further, if $\{X_t\}$ is a time series where $\sup_t \|X_t\|_{\mathbb{E},2q} = \|X\|_{\mathbb{E},2q} < \infty$ (for some $q > 1$), then*

$$\begin{aligned}\|W_n(\theta) - W_{p,n}(\theta)\|_{\mathbb{E},q} \;\leqslant\; &\rho_{p,K}(f)A_K(f,f_\theta)\times \\ &\left(\frac{(C_{f,1}+1)}{np^{K-1}} + \frac{2(C_{f,1}+1)^2}{np^K}\|\psi_f\|_0\|\phi_f\|_1 + \frac{C_{f,0}}{n^K}\right)\|X\|_{\mathbb{E},2q}^2.\end{aligned} \tag{4.5}$$

PROOF. See Section 4.7.2. □

**Remark 4.1.1.** *We briefly discuss what the above bounds mean for different types of spectral densities $f$.*

*(i) Suppose $f$ is the spectral density of a finite order AR($p_0$). If $p \geqslant p_0$, then $\left\|F_n^*\Delta_n(f_\theta^{-1})\left(D_n(f) - D_n(f_p)\right)\right\|_1 = 0$ and $\|W_n(\theta) - W_{p,n}(\theta)\|_{\mathbb{E},q} = 0$. On the other hand, if*

*$p < p_0$ we replace the $p^K$ and $p^{K-1}$ terms in Lemma 4.1.1 with $\sum_{j=p+1}^{p_0} |\phi_j|$ and $\sum_{j=p+1}^{p_0} |j\phi_j|$ respectively, where $\{\phi_j\}_{j=1}^{p}$ are the AR($p$) coefficients corresponding to $f$.*

*(ii) If the autocovariances corresponding to $f$ decay geometrically fast to zero (for example an ARMA processes), then for some $0 \leqslant \rho < 1$ we have*

$$\|W_n(\theta) - W_{p,n}(\theta)\|_{\mathbb{E},q} = O\left(\frac{\rho^p}{n} + \rho^n\right). \tag{4.6}$$

*(iii) If the autocovariances corresponding to $f$ decay to zero at a polynomial rate with $\sum_r |r^K c(r)| < \infty$, then*

$$\|W_n(\theta) - W_{p,n}(\theta)\|_{\mathbb{E},q} = O\left(\frac{1}{np^{K-1}}\right). \tag{4.7}$$

*Roughly speaking, the faster the rate of decay of the autocovariance function, the "closer" $W_{p,n}(\theta)$ will be to $W_n(\theta)$ for a given $p$.*

It follows from the lemma above that if $1 \leqslant p < n$, $\mathbb{E}_f[W_{p,n}(\theta)] = I_n(f; f_\theta) + O((np^{K-1})^{-1})$ and

$$W_{p,n}(\theta) = W_n(\theta) + O_p\left(\frac{1}{np^{K-1}}\right).$$

Thus if $p \to \infty$ as $n \to \infty$, then $W_{p,n}(\theta)$ yields a better approximation to the "ideal" $W_n(\theta)$ than both the Whittle and the Gaussian likelihood.

Since $f$ is unknown, $f_p$ is also unknown. The second step is to estimate $f_p$ from the data. We use the Yule-Walker estimator to fit an AR($p$) process to the observed time series, where we select the order $p$ using the AIC. This leads a feasible estimator $\widetilde{J}_n(\omega; \widehat{f}_p) = J_n(\omega) + \widehat{J}_n(\omega; \widehat{f}_p)$, where $\widehat{J}_n(\omega; \widehat{f}_p)$ is defined as in (2.18).

This estimator allows us to replace $\widetilde{J}_n(\omega_{k,n}; f_p)$ in $W_{p,n}(\theta)$ with $\widetilde{J}_n(\omega_{k,n}; \widehat{f}_p)$ to give the "ob-

served" boundary corrected Whittle likelihood

$$\widehat{W}_{p,n}(\theta) = \frac{1}{n}\sum_{k=1}^{n} \frac{\widetilde{J}_n(\omega_{k,n};\widehat{f}_p)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} + \frac{1}{n}\sum_{k=1}^{n}\log f_\theta(\omega_{k,n}). \tag{4.8}$$

We use as an estimator of $\theta$, $\widehat{\theta}_n = \arg\min \widehat{W}_{p,n}(\theta)$. It is worth bearing in mind that

$$\mathrm{Im}\,\frac{\widetilde{J}_n(\omega_{k,n};\widehat{f}_p)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} = -\,\mathrm{Im}\,\frac{\widetilde{J}_n(\omega_{n-k,n};\widehat{f}_p)\overline{J_n(\omega_{n-k,n})}}{f_\theta(\omega_{n-k,n})}$$

thus $\widehat{W}_{p,n}(\theta)$ is real for all $\theta$. However, due to rounding errors it is prudent to use $\mathrm{Re}\,\widehat{W}_{p,n}(\theta)$ in the minimization algorithm. Sometimes $\mathrm{Re}\,\widetilde{J}_n(\omega_{k,n};\widehat{f}_p)\overline{J_n(\omega_{k,n})}$ can be negative, when this arises we threshold it to be positive (the method we use is given in Section 4.6).

In this dissertation, we focus on estimating $\widehat{J}_n(\omega_{k,n};f_p)$ using the Yule-Walker estimator. However, other estimators could be used. These may, in certain situations, give better results. For example, in the case that $f$ has a more peaked spectral density (corresponding to AR parameters close to the unit circle) it may be better to replace the Yule-Walker estimator with the tapered Yule-Walker estimator (as described in Dahlhaus (1988) and Zhang (1992)) or the Burg estimator. We show in Section 4.6.4, that using the tapered Yule-Walker estimator tends to give better results for peaked spectral density functions. Alternatively one could directly estimate $\widehat{J}_{\infty,n}(\omega_{k,n};f)$, where we use a non-parametric spectral density estimator of $f$. This is described in greater detail in Section 4.6.4 together with the results of some simulations.

### 4.1.2 The hybrid Whittle likelihood

The simulations in Section 4.6 suggest that the boundary corrected Whittle likelihood estimator (defined in (4.8)) yields an estimator with a smaller bias than the regular Whittle likelihood. However, the bias of the tapered Whittle likelihood (and often the Gaussian likelihood) is in some cases lower. The tapered Whittle likelihood (first proposed in Dahlhaus (1988)) gives a better resolution at the peaks in the spectral density. It also "softens" the observed domain of observation. With this in mind, we propose the hybrid Whittle likelihood which incorporates the notion of tapering.

Suppose $\underline{h} = \{h_{t,n}\}_{t=1}^n$ is a data taper, where the weights $\{h_{t,n}\}$ are non-negative and $\sum_{t=1}^n h_{t,n} = n$. Then, using results in Section 3.1, the tapered complete periodogram $I_{\underline{h},n}(\omega; f) = \widetilde{J}_n(\omega; f)\overline{J_{\underline{h},n}(\omega)}$ where $J_{\underline{h},n}(\omega) = n^{-1/2}\sum_{t=1}^n h_{t,n}X_t e^{it\omega}$ is an unbiased estimator of $f(\omega)$.

Based on the above result we define the infeasible hybrid Whittle likelihood

$$H_n(\theta) = \frac{1}{n}\sum_{k=1}^n \frac{\widetilde{J}_n(\omega_{k,n}; f)\overline{J_{\underline{h},n}(\omega_{k,n})}}{f_\theta(\omega_{k,n})} + \frac{1}{n}\sum_{k=1}^n \log f_\theta(\omega_{k,n}) \tag{4.9}$$

and $\mathbb{E}_f[H_n(\theta)] = I_n(f; f_\theta)$. Thus $H_n(\theta)$ is an unbiased estimator of $I_n(f; f_\theta)$. Clearly, it is not possible to estimate $\theta$ using the (unobserved) criterion $H_n(\theta)$. Instead we replace $\widetilde{J}_n(\omega_{k,n}; f)$ with its estimator $\widetilde{J}_n(\omega_{k,n}; \widehat{f}_p)$ and define

$$\widehat{H}_{p,n}(\theta) = \frac{1}{n}\sum_{k=1}^n \frac{\widetilde{J}_n(\omega_{k,n}; \widehat{f}_p)\overline{J_{n,\underline{h}_n}(\omega_{k,n})}}{f_\theta(\omega_{k,n})} + \frac{1}{n}\sum_{k=1}^n \log f_\theta(\omega_{k,n}). \tag{4.10}$$

We then use as an estimator of $\theta$, $\widehat{\theta}_n = \arg\min \widehat{H}_{p,n}(\theta)$. An illustration which visualises and compares the boundary corrected Whittle likelihood and hybrid Whittle likelihood is given in Figure 4.1.

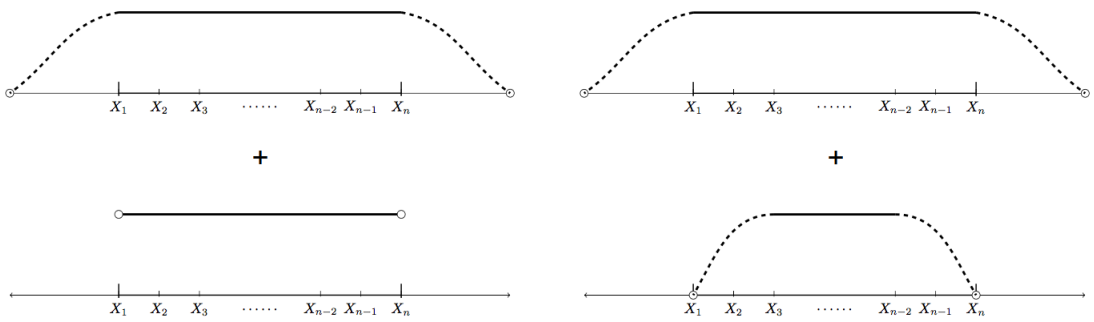



Figure 4.1: Left: The estimated complete DFT and the regular DFT which yields the boundary corrected Whittle likelihood. Right: The estimated complete DFT and the tapered DFT which forms the hybrid Whittle likelihood.

## 4.2 Assumptions

In this section, we make sets of assumptions that are used to study the sampling properties of the boundary corrected and hybrid Whittle likelihood. Our focus will be on the hybrid Whittle likelihood as it includes the boundary corrected likelihood as a special case, when $h_{t,n} \equiv 1$. In Section 3.2, we study the sampling properties of the estimated integrated complete periodogram, which is a weighted sum of $\widetilde{J}_n(\omega; \widehat{f}_p)\overline{J_{n,\underline{h}_n}(\omega)}$. Using these results and the results in Section B, we obtain the bias and variance of the boundary corrected and hybrid Whittle likelihood.

Suppose we fit the spectral density $f_\theta(\omega)$ (where $\theta$ is an unknown $d$-dimension parameter vector) to the stationary time series $\{X_t\}_{t=1}^n$ whose true spectral density is $f$. The best fitting spectral density is $f_{\theta_n}$, where $\theta_n = \arg\min I_n(f; f_\theta)$. Let $\widehat{\theta}_n = (\widehat{\theta}_{1,n}, \ldots, \widehat{\theta}_{d,n})$ be its estimator, where $\widehat{\theta}_n = \arg\min \widehat{H}_{p,n}(\theta)$.

To derive the sampling properties of $\widehat{\theta}_n$ we assume the data taper has the following form

$$h_{t,n} = c_n h_n(t/n), \tag{4.11}$$

where $h_n : [0,1] \to \mathbb{R}$ is a sequence of positive functions that satisfy the taper assumptions in Section 5, Dahlhaus (1988) and $c_n = n/H_{1,n}$ with $H_{q,n} = \sum_{t=1}^n h_n(t/n)^q$. We will assume $\sup_{t,n} h_{t,n} < \infty$, using this it is straightforward to show $H_{2,n}/H_{1,n}^2 = O(n^{-1})$. Under this condition, the hybrid Whittle is $n^{1/2}$–consistency and the equivalence result in Theorem 4.3.1 holds. This assumption is used in Dahlhaus (1983) and in practice one often assumes that a fixed percentage of the data is tapered. A relaxation of the condition $H_{2,n}/H_{1,n}^2 = O(n^{-1})$ will lead to a change of rate in Theorem 4.3.1.

**Assumption 4.2.1** (Assumptions on the parameter space). *(i) The parameter space $\Theta \subset \mathbb{R}^d$ is compact, $0 < \inf_{\theta\in\Theta} \inf_\omega f_\theta(\omega) \leqslant \sup_{\theta\in\Theta} \sup_\omega f_\theta(\omega) < \infty$ and $\theta_n$ lies in the interior of $\Theta$.*

*(ii) The one-step ahead prediction error $\sigma^2 = \exp((2\pi)^{-1}\int_0^{2\pi} \log f_\theta(\omega) d\omega)$ is not a function of the parameter $\theta$.*

*(iii) Let $\{\phi_j(f_\theta)\}$ and $\{\psi_j(f_\theta)\}$ denote the AR($\infty$) and MA($\infty$) coefficients corresponding to the spectral density $f_\theta$ respectively. Then for all $\theta \in \Theta$ and $0 \leqslant s \leqslant \kappa$ (for some $\kappa \geqslant 4$), we have*

$$(a) \sup_{\theta\in\Theta} \sum_{j=1}^{\infty} \|j^K \nabla_\theta^s \phi_j(f_\theta)\|_1 < \infty \qquad (b) \sup_{\theta\in\Theta} \sum_{j=1}^{\infty} \|j^K \nabla_\theta^s \psi_j(f_\theta)\|_1 < \infty,$$

*where $K > 3/2$, $\nabla_\theta^a g(f_\theta)$ is the $a$th order partial derivative of $g$ with respect to $\theta$, and $\|\nabla_\theta^a g(f_\theta)\|_1$ denotes the absolute sum of all the partial derivatives in $\nabla_\theta^a g(f_\theta)$.*

We use Assumption 4.2.1(ii, iii) to show that the $n^{-1}\sum_{k=1}^n \log f_\theta(\omega_{k,n})$ term in boundary corrected and hybrid Whittle likelihoods are negligible with respect the other bias terms. This allows us to simplify some of the bias expansions. Without Assumption 4.2.1(ii, iii-a) the asymptotic bias of the new-frequency domain likelihood estimators would contain some additional terms. Assumption 4.2.1(iii-b) is used to bound the $s$th derivative of the spectral density.

**Assumption 4.2.2** (Assumptions on the time series)**.** *(i) $\{X_t\}$ is a stationary time series. Let $\kappa_\ell(t_1, \ldots, t_{\ell-1})$ denote the joint cumulant* $\operatorname{cum}(X_0, X_{t_1}, \ldots, X_{t_{\ell-1}})$.

*Then for all $1 \leqslant j \leqslant \ell \leqslant 12$,*

$$\sum_{t_1,\ldots,t_{\ell-1}} |(1 + t_j)\kappa_\ell(t_1, \ldots, t_{\ell-1})| < \infty.$$

*(ii) The spectral density of $\{X_t\}$ is such that the spectral density $f$ is bounded away from zero and for some $K > 1$, the autocovariance function satisfies $\sum_{r\in\mathbb{Z}} |r^K c_f(r)| < \infty$.*

*(iii) $I(\theta_n)$ is invertible where*

$$I(\theta) = -\frac{1}{2\pi}\int_0^{2\pi} [\nabla_\theta^2 f_\theta(\omega)^{-1}] f(\omega) d\omega. \tag{4.12}$$

We require Assumption 4.2.2(i), when $\ell = 4$ and 6 to obtain a bound for the expectation of the terms in the bias expansions and $\ell = 12$ to show equivalence between the feasible estimator based on $\widehat{H}_{p,n}(\theta)$ and its infeasible counterparts $H_n(\theta)$. Under Assumption 4.2.2(i,ii), we show in

Theorem 3.2.1 that

$$\widehat{H}_{p,n}(\theta) = H_n(\theta) + O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).$$

Under Assumption 4.2.1(i,iii) the above error is uniform over the parameter space. If the model is an AR($p_0$) and $p_0 \leqslant p$, then the term $O((np^{K-1})^{-1})$ in the above disappears.

## 4.3 Rates of convergence of the new likelihood estimators

### 4.3.1 The criteria

To begin with, we state the assumptions required to obtain rates of convergence of the new criteria and asymptotic equivalence to the infeasible criteria. These results will be used to derive the asymptotic sampling properties of the new likelihood estimators, including their asymptotic bias (in a later section). To do this, we start by defining the criteria we will be considering.

We assume that $\{X_t\}$ is a stationary time series with spectral density $f$, where $f$ is bounded away from zero (and bounded above). We fit the model with spectral density $f_\theta$ to the observed time series. We do not necessarily assume that there exists a $\theta_0 \in \Theta$ where $f = f_{\theta_0}$. Since we allow the misspecified case, for a given $n$, it seems natural that the "ideal" best fitting parameter is

$$\theta_n = \arg\min_\theta I_n(f, f_\theta). \tag{4.13}$$

where $I_n(f, f_\theta)$ is defined in (4.1). Note that in the case the spectral density is correctly specified, then $\theta_n = \theta_0$ for all $n$ where $f = f_{\theta_0}$.

We now show that Assumption 4.2.1(ii,iii) allows us to ignore the $n^{-1}\sum_{k=1}^n \log f_\theta(\omega_{k,n})$ in the Whittle, boundary corrected Whittle and hybrid Whittle likelihoods. To show why this is true, we obtain the Fourier expansion of $\log f_\theta(\omega) = \sum_{r\in\mathbb{Z}} \alpha_r(f_\theta)e^{ir\omega}$, where $\alpha_0(f_\theta) = \log\sigma^2$, in terms of the corresponding MA($\infty$) coefficients. We use the well known Szegö's identity

$$\log f_\theta(\cdot) = \log\sigma^2|\psi(\cdot; f_\theta)|^2 = \log\sigma^2 + \log\psi(\cdot; f_\theta) + \log\overline{\psi(\cdot; f_\theta)}$$

where $\psi(\omega; f_\theta) = \sum_{j=0}^{\infty} \psi_j(f_\theta) e^{-ij\omega}$ with $\psi_0(f_\theta) = 1$ and the roots of the MA transfer function $\sum_{j=0}^{\infty} \psi_j(f_\theta) z^j$ lie outside the unit circle (minimum phased). Comparing
$\log f_\theta(\omega) = \sum_{r\in\mathbb{Z}} \alpha_r(f_\theta) e^{ir\omega}$ with the positive half of the above expansion gives

$$\log(\sum_{j=0}^{\infty} \psi_j(f_\theta) z^j) = \sum_{j=1}^{\infty} \alpha_j(f_\theta) z^j \quad \text{for } |z| < 1,$$

and since $\log f_\theta$ is real and symmetric about $\pi$, $\alpha_{-j}(f_\theta) = \alpha_j(f_\theta) \in \mathbb{R}$. This allows us to obtain coefficients $\{\alpha_j(f_\theta)\}$ in terms of the MA($\infty$) coefficients (it is interesting to note that Pourahmadi (2001) gives a recursion for $\alpha_j(f_\theta)$ in terms of the MA($\infty$) coefficient). The result is given in Lemma D.2.2, but we summarize it below. Under Assumption 4.2.1(iii) we have for $0 \leqslant s \leqslant \kappa$ (for some $\kappa \geqslant 4$)

$$\sum_{j=1}^{\infty} \|j^K \nabla_\theta^s \alpha_j(f_\theta)\|_1 < \infty.$$

Using this result, we bound $n^{-1} \sum_{k=1}^{n} \log f_\theta(\omega_{k,n})$. Applying the Poisson summation formula to this sum we have

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^{n} \log f_\theta(\omega_{k,n}) = \sum_{r\in\mathbb{Z}} \alpha_{rn}(f_\theta) &= \alpha_0(f_\theta) + \sum_{r\in\mathbb{Z}\setminus\{0\}} \alpha_{rn}(f_\theta) \\ &= \log(\sigma^2) + \sum_{r\in\mathbb{Z}\setminus\{0\}} \alpha_{rn}(f_\theta). \end{aligned} \tag{4.14}$$

The $s$th-order derivative ($s \geqslant 1$) with respect to $\theta$ (and using Assumption 4.2.1(ii) that $\sigma^2$ does not depend on $\theta$) we have

$$\frac{1}{n} \sum_{k=1}^{n} \nabla_\theta^s \log f_\theta(\omega_{k,n}) = \sum_{r\in\mathbb{Z}\setminus\{0\}} \nabla_\theta^s \alpha_{rn}(f_\theta).$$

By using Lemma D.2.2 for $0 \leqslant s \leqslant \kappa$ we have

$$\| \sum_{r\in\mathbb{Z}\setminus\{0\}} \nabla_\theta^s \alpha_{rn}(f_\theta)\|_1 \leqslant 2 \sum_{j\geqslant n} \|\nabla_\theta^s \alpha_j(f_\theta)\|_1 = O(n^{-K}). \tag{4.15}$$

Substituting the bound in (4.15) (for $s = 0$) into (4.14) gives

$$\left| \frac{1}{n} \sum_{k=1}^{n} \log f_\theta(\omega_{k,n}) - \log \sigma^2 \right| = O(n^{-K}).$$

Using (4.15) for $1 \leqslant s \leqslant \kappa$ we have

$$\left\| \frac{1}{n} \sum_{k=1}^{n} \nabla_\theta^s \log f_\theta(\omega_{k,n}) \right\|_1 = O(n^{-K}).$$

Therefore if $K > 1$, the log determinant term in the Whittle, boundary corrected, and hybrid Whittle likelihood is negligible as compared with $O(n^{-1})$ (which we show is the leading order in the bias).

However, for the Gaussian likelihood, the log determinant cannot be ignored. Specifically, by applying the strong Szegö's theorem (see e.g., Theorem 10.29 of Böttcher and Silbermann (2013)) to $\Gamma_n(f_\theta)$ we have

$$\frac{1}{n} \log |\Gamma_n(f_\theta)| = \log \sigma^2 + \frac{1}{n} E(\theta) + o(n^{-1})$$

where $E(\theta) = \sum_{k=1}^{\infty} \alpha_k(f_\theta)^2$. Therefore, unlike the other three quasi-likelihoods, the error in $\log |\Gamma_n(f_\theta)|$ is of order $O(n^{-1})$, which is of the same order as the bias. In Section B.2, we show that the inclusion and exclusion of $n^{-1} \log |\Gamma_n(f_\theta)|$ leads to Gaussian likelihood estimators with substantial differences in their bias. Further, there is no clear rule whether the inclusion of the $n^{-1} \log |\Gamma_n(f_\theta)|$ in the Gaussian likelihood improves the bias or makes it worse. In the case that $n^{-1} \log |\Gamma_n(f_\theta)|$ is included in the Gaussian likelihood, then the expression for the bias will include the derivatives of $E(\theta)$. Except for a few simple models (such as the AR$(1)$ model) the expression for the derivatives of $E(\theta)$ will be extremely unwieldy.

Based on the above, to make the derivations cleaner, we define all the quasi-likelihoods without

the log term and let

$$\begin{aligned}
\mathcal{L}_n(\theta) &= n^{-1}\underline{X}_n'\Gamma_n(f_\theta)^{-1}\underline{X}_n = \frac{1}{n}\sum_{k=1}^{n}\frac{\widetilde{J}_n(\omega_{k,n};f_\theta)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} \\
K_n(\theta) &= \frac{1}{n}\sum_{k=1}^{n}\frac{|J_n(\omega_{k,n})|^2}{f_\theta(\omega_{k,n})} \\
\widehat{W}_{p,n}(\theta) &= \frac{1}{n}\sum_{k=1}^{n}\frac{\widetilde{J}_n(\omega_{k,n};\widehat{f}_p)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} \\
\widehat{H}_{p,n}(\theta) &= \frac{1}{n}\sum_{k=1}^{n}\frac{\widetilde{J}_n(\omega_{k,n};\widehat{f}_p)\overline{J_{n,\underline{h}_n}(\omega_{k,n})}}{f_\theta(\omega_{k,n})}. \qquad (4.16)
\end{aligned}$$

In the case of the hybrid Whittle likelihood, we make the assumption the data taper $\{h_{t,n}\}$ is such that $h_{t,n} = c_n h_n(t/n)$ where $c_n = n/H_{1,n}$ and $h_n : [0,1] \to \mathbb{R}$ is a sequence of taper functions which satisfy the taper conditions in Section 5, Dahlhaus (1988).

We define the parameter estimators as

$$\begin{aligned}
\widehat{\theta}_n^{(G)} &= \arg\min \mathcal{L}_n(\theta), \quad \widehat{\theta}_n^{(K)} = \arg\min K_n(\theta), \\
\widehat{\theta}_n^{(W)} &= \arg\min \widehat{W}_{p,n}(\theta), \text{ and } \widehat{\theta}_n^{(H)} = \arg\min \widehat{H}_{p,n}(\theta) \qquad (4.17)
\end{aligned}$$

### 4.3.2 Asymptotic equivalence to the infeasible criteria

In this section we analyze the feasible estimators $\widehat{\theta}_n^{(W)}$ and $\widehat{\theta}_n^{(H)}$ and show it is asymptotic equivalence to the corresponding infeasible criteria which replace $\widehat{f}_p$ with $f$, the true spectral density.

Before that, we discuss the condition on data taper such that $H_{2,n}/H_{1,n}^2 = O(n^{-1})$. This has some benefits. The first is that the rates for the hybrid Whittle and the boundary corrected Whittle are the same. In particular, by using Theorems 2.3.4 and 3.2.1 (under Assumption 4.2.2) we have

$$[\widehat{J}_n(\omega;\widehat{f}_p) - \widehat{J}_n(\omega;f)]\overline{J_{n,\underline{h}_n}(\omega)} = O_p\left(\frac{p^2}{n} + \frac{p^3}{n^{3/2}}\right) \qquad (4.18)$$

and

$$\widehat{H}_{p,n}(\theta) = H_n(\theta) + O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right). \tag{4.19}$$

Using this, we show below that the Hybrid Whittle estimator has the classical $n^{1/2}$–rate. If we were to relax the rate on $H_{2,n}/H_{1,n}^2 = O(n^{-1})$, then the $n^{1/2}$–rate and the rates in (4.18) and (4.19) would change. This will make the proofs more technical. Thus for ease of notation and presentation we will assume that $H_{2,n}/H_{1,n}^2 = O(n^{-1})$.

We start by obtaining a "crude" bound for $\nabla_\theta^s \widehat{W}_{p,n}(\theta) - \nabla_\theta^s W_n(\theta)$.

**Lemma 4.3.1.** *Suppose that Assumptions 4.2.1(i,iii) and 4.2.2(i,ii) hold. Then for* $0 \leqslant s \leqslant \kappa$ *(for some* $\kappa \geqslant 4$*) we have*

$$\sup_{\theta\in\Theta} \left\| \nabla_\theta^s \widehat{W}_{p,n}(\theta) - \nabla_\theta^s W_n(\theta) \right\|_1 = O_p\left(\frac{p^2}{n}\right)$$

*and*

$$\sup_{\theta\in\Theta} \left\| \nabla_\theta^s \widehat{H}_{p,n}(\theta) - \nabla_\theta^s H_n(\theta) \right\|_1 = O_p\left(\frac{p^2}{n}\right).$$

PROOF. See Section 4.7.2. □

**Lemma 4.3.2.** *Suppose that Assumptions 4.2.1(i,iii) and 4.2.2(i,ii) hold. Then*

$$|\widehat{\theta}_n^{(W)} - \theta_n|_1 \xrightarrow{\mathcal{P}} 0 \qquad \textit{and} \qquad |\widehat{\theta}_n^{(H)} - \theta_n|_1 \xrightarrow{\mathcal{P}} 0$$

*with* $p^2/n \to 0$ *as* $p, n \to \infty$.

PROOF. See Section 4.7.2. □

For the simplicity, we assume $\theta$ is univariate and state the following lemma. It can be easily generalized to the multivariate case.

**Lemma 4.3.3.** *Suppose Assumptions 4.2.1(i,iii) and 4.2.2 hold. Then for $i = 1, 2$ we have*

$$\frac{d^i \widehat{W}_{p,n}(\theta)}{d\theta^i}\rfloor_{\theta=\theta_n} = \frac{d^i W_n(\theta)}{d\theta^i}\rfloor_{\theta=\theta_n} + O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right) \tag{4.20}$$

*and*

$$\left|\frac{d^3 \widehat{W}_{p,n}(\theta)}{d\theta^3}\rfloor_{\theta=\bar{\theta}_n} - \frac{d^3 W_n(\theta)}{d\theta^3}\rfloor_{\theta=\theta_n}\right| = O_p\left(\frac{p^2}{n}\right) + |\widehat{\theta}_n^{(W)} - \theta_n| O_p(1), \tag{4.21}$$

*where $\overline{\theta}_n$ is a convex combination of $\widehat{\theta}_n^{(W)}$ and $\theta_n$. This gives rise to the first order and second expansions*

$$(\widehat{\theta}_n^{(W)} - \theta_n) = -\left[\mathbb{E}\left[\frac{d^2 W_n(\theta_n)}{d\theta_n^2}\right]\right]^{-1} \frac{dW_n(\theta_n)}{d\theta_n} + O_p\left(\frac{1}{n} + \frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right) \tag{4.22}$$

*and*

$$\begin{aligned} &\frac{dW_n(\theta)}{d\theta}\rfloor_{\theta=\theta_n} + (\widehat{\theta}_n^{(W)} - \theta_n)\frac{d^2 W_n(\theta)}{d\theta^2}\rfloor_{\theta=\theta_n} + \frac{1}{2}(\widehat{\theta}_n^{(W)} - \theta_n)^2 \frac{d^3 W_n(\theta)}{d\theta^3}\rfloor_{\theta=\theta_n} \\ &\qquad = O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right). \end{aligned} \tag{4.23}$$

PROOF. See Section 4.7.2. □

The second order expansion (4.23) is instrumental in proving the equivalence result Theorem 4.3.1. By following a similar set of arguments to those in Lemma 4.3.3 for the multivariate parameter $\theta = (\theta_1, \ldots, \theta_d)$, the feasible estimator satisfies the expansion

$$\begin{aligned} &\frac{\partial W_n(\theta)}{\partial \theta_r} + \sum_{s=1}^{d} (\widehat{\theta}_{s,n}^{(W)} - \theta_{s,n}) \frac{\partial^2 W_n(\theta)}{\partial \theta_s \partial \theta_r}\rfloor_{\theta=\theta_n} + \\ &\frac{1}{2}\sum_{s_1,s_2=1}^{d} (\widehat{\theta}_{s_1,n}^{(W)} - \theta_{s_1,n})(\widehat{\theta}_{s_2,n}^{(W)} - \theta_{s_2,n}) \frac{\partial^3 W_n(\theta)}{\partial \theta_{s_1} \partial \theta_{s_2} \partial \theta_r}\rfloor_{\theta=\theta_n} = \left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right). \end{aligned} \tag{4.24}$$

By using the same set of arguments we can obtain a first and second order expansion for the hybrid

Whittle estimator

$$(\widehat{\theta}_n^{(H)} - \theta_n) = -\left[\mathbb{E}\left[\frac{d^2 H_n(\theta_n)}{d\theta_n^2}\right]\right]^{-1} \frac{dH_n(\theta_n)}{d\theta_n} + O_p\left(\frac{1}{n} + \frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right) \tag{4.25}$$

and

$$\begin{aligned}\frac{dH_n(\theta)}{d\theta}\rfloor_{\theta=\theta_n} + (\widehat{\theta}_n^{(W)} - \theta_n)\frac{d^2 H_n(\theta)}{d\theta^2}\rfloor_{\theta=\theta_n} + \frac{1}{2}(\widehat{\theta}_n^{(H)} - \theta_n)^2 \frac{d^3 H_n(\theta)}{d\theta^3}\rfloor_{\theta=\theta_n} \\ = \left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).\end{aligned} \tag{4.26}$$

Using the assumptions above we obtain a bound between the feasible and infeasible estimators.

**Theorem 4.3.1** (Equivalence of feasible and infeasible estimators)**.** *Suppose Assumptions 4.2.1 and 4.2.2 hold. Define the feasible and infeasible estimators as* $\widetilde{\theta}_n = \arg\min H_n(\theta)$ *and* $\widehat{\theta}_n = \arg\min \widehat{H}_{p,n}(\theta)$ *respectively. Then for* $p \geqslant 1$ *we have*

$$|\widehat{\theta}_n - \widetilde{\theta}_n|_1 = O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right),$$

*where* $|a|_1 = \sum_{j=1}^d |a_j|$*. For the case* $p = 0$*,* $\widehat{\theta}_n$ *is the parameter estimator based on the Whittle likelihood using the one-sided tapered periodogram* $J_n(\omega_{k,n})\overline{J_{n,\underline{h}_n}(\omega_{k,n})}$ *rather than the regular tapered periodogram. In this case,* $|\widehat{\theta}_n - \widetilde{\theta}_n|_1 = O_p(n^{-1})$*.*

*Note if the true spectral density of the time series is that of an AR*$(p_0)$ *where* $p_0 \leqslant p$*, then the* $O((np^{K-1})^{-1})$ *term is zero.*

PROOF. See Section 4.7.2. □

The implication of the equivalence result is if $p^3/n^{1/2} \to 0$ as $p \to \infty$ and $n \to \infty$, then $n|\widehat{\theta}_n - \widetilde{\theta}_n|_1 \to 0$ and asymptotically the properties of the infeasible estimator (such as bias and variance) transfer to the feasible estimator.

## 4.4 The bias and variance of the hybrid Whittle likelihood

The expressions in this section are derived under Assumptions 4.2.1 and 4.2.2. To obtain an expression for the mean and variance of $\widehat{\theta}_n = (\widehat{\theta}_{1,n}, \ldots, \widehat{\theta}_{d,n})$ we require the following quantities. Let

$$\begin{aligned} V(g,h) &= \frac{2}{2\pi}\int_0^{2\pi} g(\omega)h(\omega)f(\omega)^2 d\omega \\ &\quad + \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} g(\omega_1)h(\omega_2)f_4(\omega_1,-\omega_1,\omega_2)d\omega_1 d\omega_2 \\ \text{and} \qquad J(g) &= \frac{1}{2\pi}\int_0^{2\pi} g(\omega)f(\omega)d\omega, \end{aligned} \tag{4.27}$$

where $f_4$ denotes the fourth order cumulant density of the time series $\{X_t\}$. We denote the $(s,r)$th element of $I(\theta_n)^{-1}$ (where $I(\theta_n)$ is defined in (4.12)) as $I^{(s,r)}$, and define

$$\begin{aligned} G_r(\theta) &= \sum_{s_1,s_2=1}^{d} I^{(s_1,s_2)} V\left(\frac{\partial f_\theta^{-1}}{\partial\theta_{s_2}}, \frac{\partial^2 f_\theta^{-1}}{\partial\theta_{s_1}\partial\theta_r}\right) \\ &\quad + \frac{1}{2}\sum_{s_1,s_2,s_3,s_4=1}^{d} I^{(s_1,s_3)}I^{(s_2,s_4)} V\left(\frac{\partial f_\theta^{-1}}{\partial\theta_{s_3}}, \frac{\partial f_\theta^{-1}}{\partial\theta_{s_4}}\right) J\left(\frac{\partial^3 f_\theta^{-1}}{\partial\theta_{s_1}\partial\theta_{s_2}\partial\theta_r}\right). \end{aligned} \tag{4.28}$$

### 4.4.1 The bias

We show in Appendix B.3, that the asymptotic bias for $\widehat{\theta}_n = (\widehat{\theta}_{1,n}, \ldots, \widehat{\theta}_{d,n})$ is

$$\mathbb{E}[\widehat{\theta}_{j,n} - \theta_{j,n}] = \frac{H_{2,n}}{H_{1,n}^2}\sum_{r=1}^{d} I^{(j,r)} G_r(\theta_n) + O\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right) \quad 1 \leqslant j \leqslant d, \tag{4.29}$$

where $I^{(j,r)}$ and $G_r(\theta_n)$ is defined in (4.28). We note that if no tapering were used then $H_{2,n}/H_{1,n}^2 = n^{-1}$. The Gaussian and Whittle likelihood have a bias which includes the above term (where $H_{2,n}/H_{1,n}^2 = n^{-1}$) plus an additional term of the form $\sum_{r=1}^{d} I^{(j,r)}\mathbb{E}[\nabla_\theta L_n(\theta_n)]$, where $L_n(\cdot)$ is the Gaussian or Whittle likelihood (see Appendix B.3 for the details).

Theoretically, it is unclear which criteria has the smallest bias (since the inclusion of additional

terms does not necessarily increase the bias). However, for the hybrid Whittle likelihood estimator, a straightforward "Bartlett correction" can be made to estimate the bias in (4.29). We briefly outline how this can be done. We observe that the bias is built of $I(\cdot)$, $J(\cdot)$ and $V(\cdot,\cdot)$. Both $I(\cdot)$ and $J(\cdot)$ can easily be estimated with their sample means. The term $V(\cdot,\cdot)$ can also be estimated by using an adaption of orthogonal samples (see Subba Rao (2018)), which we now describe. Define the random variable

$$h_r(g;f) = \frac{1}{n}\sum_{k=1}^{n} g(\omega_{k,n})\widetilde{J}_n(\omega_{k+r,n};f)\overline{J_n(\omega_{k,n})} \quad \text{for} \quad r \geqslant 1,$$

where $g$ is a continuous and bounded function. Suppose $g_1$ and $g_2$ are continuous and bounded functions. If $r \neq n\mathbb{Z}$, then $\mathbb{E}_f[h_r(g_j;f)] = 0$ (for $j = 1$ and 2). But interestingly, if $r << n$, then $n\text{cov}_f[h_r(g_1;f), h_r(g_2;f)] = n\mathbb{E}_f[h_r(g_1;f)\overline{h_r(g_2;f)}] = V(g_1,g_2) + O(r/n)$. Using these results, we estimate $V(g_1,g_2)$ by replacing $h_r(g_j;f)$ with $h_r(g_j;\widehat{f}_p)$ and defining the "sample covariance"

$$\widehat{V}_M(g_1,g_2) = \frac{n}{M}\sum_{r=1}^{M} h_r(g_1;\widehat{f}_p)\overline{h_r(g_2;\widehat{f}_p)}$$

where $M << n$. Thus, $\widehat{V}_M(g_1,g_2)$ is an estimator of $V(g_1,g_2)$. Based on this construction,

$$\widehat{V}_M\left(\frac{\partial}{\partial\theta_{s_2}}f^{-1}_{\widehat{\theta}_n}, \frac{\partial^2}{\partial\theta_{s_1}\partial\theta_r}f^{-1}_{\widehat{\theta}_n}\right) \quad \text{and} \quad \widehat{V}_M\left(\frac{\partial}{\partial\theta_{s_3}}f^{-1}_{\widehat{\theta}_n}, \frac{\partial}{\partial\theta_{s_4}}f^{-1}_{\widehat{\theta}_n}\right)$$

are estimators of $V\left(\frac{\partial f_\theta^{-1}}{\partial\theta_{s_2}}, \frac{\partial^2 f_\theta^{-1}}{\partial\theta_{s_1}\partial\theta_r}\right)$ and $V\left(\frac{\partial f_\theta^{-1}}{\partial\theta_{s_3}}, \frac{\partial f_\theta^{-1}}{\partial\theta_{s_4}}\right)$ respectively. This estimation scheme yields a consistent estimate of the bias even when the model is misspecified. In contrast, it is unclear how a bias correction would work for the Gaussian and Whittle likelihood under misspecification, as they also involve the term $\mathbb{E}_f[\nabla_\theta L_n(\theta_n)]$. In the case of misspecification, $\mathbb{E}_f[\nabla_\theta L_n(\theta_n)] \neq 0$ and is of order $O(n^{-1})$.

It is worth mentioning that the asymptotic expansion in (4.29) does not fully depict what we observe in the simulations in Section 4.6. A theoretical comparison of the biases of both new likelihoods show that for the boundary corrected Whittle likelihood, the bias is asymptotically

$n^{-1}\sum_{r=1}^{d} I^{(j,r)}G_r(\theta_n)$, whereas when tapering is used the bias is $(H_{2,n}/H_{1,n}^2)\sum_{r=1}^{d} I^{(j,r)}G_r(\theta_n) \geqslant n^{-1}\sum_{r=1}^{d} I^{(j,r)}G_r(\theta_n)$. This would suggest that the hybrid Whittle likelihood should have a larger bias than the boundary corrected Whittle likelihood. But the simulations (see Section 4.6) suggest this is not necessarily true and the hybrid likelihood tends to have a smaller bias.

### 4.4.2 The variance

We show in Section 3.2.1 that the inclusion of the prediction DFT in the hybrid Whittle likelihood has a variance which asymptotically is small as compared with the main Whittle term if $p^3/n \to 0$ as $p, n \to \infty$ (under the condition $H_{2,n}/H_{1,n}^2 = O(n^{-1})$) Using this observation, standard Taylor expansion methods give the asymptotic variance of $\widehat{\theta}_n$ is

$$\frac{H_{1,n}^2}{H_{2,n}}\mathrm{var}(\widehat{\theta}_n) = I(\theta_n)^{-1}V\left(\nabla_\theta f_\theta^{-1}, \nabla_\theta f_\theta^{-1}\right)\rfloor_{\theta=\theta_n} I(\theta_n)^{-1} + o(1),$$

where $V(\cdot)$ is defined in (4.27).

## 4.5 Order selection and computational cost

### 4.5.1 The role of order estimation on the rates

Note that the order in the AR($p$) approximation is selected using the AIC. We assume that the underlying time series is a linear, stationary time series with an AR($\infty$) that satisfies Assumption K.1−K.4 in Ing and Wei (2005). Then, using the same augument in the end of Section 2.3.2, we show that the AIC order $\widehat{p} = O_p(n^{1/(1+2K)})$. Thus, if $K > 5/2$, then $\widehat{p}^3/n^{1/2} \xrightarrow{\mathcal{P}} 0$ and $\widehat{p} \xrightarrow{\mathcal{P}} \infty$ as $n \to \infty$. These rates ensure that the difference between the feasible and infeasible estimator is $|\widehat{\theta}_n - \widetilde{\theta}_n|_1 = o_p(n^{-1})$. Thus the feasible estimator, constructed using the AIC, and the infeasible estimator are equivalent and the bias and variance derived above are valid for this infeasible estimator.

### 4.5.2 The computational cost of the estimators

The Durbin-Levinson algorithm is often used to maximize the Gaussian likelihood. If this is employed, then the computational cost of the algorithm is $O(n^2)$. On the other hand, by using the

FFT, the computational cost of the Whittle likelihood is $O(n \log n)$.

For the boundary corrected Whittle and hybrid Whittle likelihood algorithm, there is an additional cost over the Whittle likelihood due to the estimation of $\{\widehat{J}_n(\omega_{k,n}; \widehat{f}_p)\}_{k=1}^n$. We recall that $\widehat{f}_p$ is constructed using the Yule-Walker estimator $\underline{\widehat{\phi}}_p = (\widehat{\phi}_{1,p}, ..., \widehat{\phi}_{p,p})'$ where $p$ is selected with the AIC. We now calculate the complexity of calculating $\{\widehat{J}_n(\omega_{k,n}; \widehat{f}_p)\}_{k=1}^n$.

The sample autocovariances, $\{\widehat{c}_n(r)\}_{r=0}^{n-1}$ (which are required in the Yule-Walker estimator) can be calculated in $O(n \log n)$ operations. Let $K_n$ denote the maximum order used for the evaluation of the AIC. If we implement the Durbin-Levinson algorithm, then evaluating $\underline{\widehat{\phi}}_p$ for $1 \leqslant p \leqslant K_n$ requires in total $O(K_n^2)$ arithmetic operations.

Suppose that the AR coefficients $\underline{\widehat{\phi}}_{\hat{p}}$ are given and compute the predictive DFT $\{\widehat{J}_n(\omega_{k,n}; \widehat{f}_p)\}_{k=1}^n$. Recall from (2.8),

$$\begin{aligned}
&\widehat{J}_n(\omega_{k,n}; f_p) \\
&= \frac{n^{-1/2}}{\phi_p(\omega_{k,n})} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{-is\omega_{k,n}} + \frac{n^{-1/2}}{\overline{\phi_p(\omega_{k,n})}} \sum_{\ell=1}^{p} X_{n+1-\ell} \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{i(s+1)\omega_{k,n}},
\end{aligned}$$

where $f_p(\cdot) = |\phi_p(\cdot)|^2$ and $\phi_p(\omega_{k,n}) = 1 - \sum_{j=1}^p \phi_j e^{-ij\omega_{k,n}}$. We focus on the first term of $\widehat{J}_n(\omega_{k,n}; f_p)$ since the second term is almost identical. Interchange the summation, the first term is

$$\begin{aligned}
&\frac{n^{-1/2}}{\phi_p(\omega_{k,n})} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{-is\omega_{k,n}} \\
&= \frac{n^{-1/2}}{\phi_p(\omega_{k,n})} \sum_{s=0}^{p-1} \left( \sum_{\ell=1}^{p-s} X_\ell \phi_{\ell+s} \right) e^{-is\omega_{k,n}} = \frac{n^{-1/2}}{\phi_p(\omega_{k,n})} \sum_{s=0}^{p-1} Y_s e^{-is\omega_{k,n}}
\end{aligned}$$

where $Y_s = \sum_{\ell=1}^{p-s} X_\ell \phi_{\ell+s}$ for $0 \leqslant s \leqslant p-1$. Note that $Y_s$ can be viewed as a convolution between $(X_1, ..., X_p)$ and $(0, 0, \ldots, 0, \phi_1, ..., \phi_p, 0, \ldots, 0)$. Based on this observation, the FFT can be utilized to evaluate $\{Y_s : 0 \leqslant s \leqslant p-1\}$ in $O(p \log p)$ operations.

By direct calculation $\{\phi_p(\omega_{k,n}) : 0 \leqslant k \leqslant n-1\}$ and $\{\sum_{s=0}^{p-1} Y_s e^{-is\omega_{k,n}} : 0 \leqslant k \leqslant n-1\}$ has $O(np)$ complexity. An alternative method of calculation is based on the observation that both $\phi_p(\omega_{k,n})$ and $\sum_{s=0}^{p-1} Y_s e^{-is\omega_{k,n}}$ can be viewed as the $k$th component of the DFT of length $n$

sequences $(1, -\phi_1, ..., -\phi_p, 0, ..., 0)$ and $(Y_0, ..., Y_{p-1}, 0, .., 0)$ respectively. Thus the FFT can be used to evaluate both $\{\phi_p(\omega_{k,n}) : 0 \leqslant k \leqslant n-1\}$ and $\{\sum_{s=0}^{p-1} Y_s e^{-is\omega_{k,n}} : 0 \leqslant k \leqslant n-1\}$ in $O(n \log n)$ operations. Therefore, since either method can be used to evaluate these terms the total number of operations for evaluation of $\{\phi_p(\omega_{k,n}) : 0 \leqslant k \leqslant n-1\}$ and $\{\sum_{s=0}^{p-1} Y_s e^{-is\omega_{k,n}} : 0 \leqslant k \leqslant n-1\}$ is $O(\min(n \log n, np))$.

Therefore, the overall computational cost of implementing both the boundary corrected Whittle and hybrid Whittle likelihood algorithms is $O(n \log n + K_n^2)$. Using Ing and Wei (2005) Example 2, for consistent order selection $K_n$ should be such that $K_n \sim n^{1/(2K+1)+\varepsilon}$ for some $\varepsilon > 0$ (where $K$ is defined in Assumption 2.3.1). Therefore, we conclude that the computational cost of the new likelihoods is of the same order as the Whittle likelihood.

### 4.6 Simulations

To substantiate our theoretical results, we conduct some simulations (further simulations can be found in Appendix C). To compare different methods, we evaluate six different quasi-likelihoods: the Gaussian likelihood (equation (1.1)), the Whittle likelihood (equation (2.23)), the boundary corrected Whittle likelihood (equation (4.8)), the hybrid Whittle likelihood (equation (4.10)), the tapered Whittle likelihood (p.810 of Dahlhaus (1988)) and the debiased Whittle likelihood (equation (7) in Sykulski et al. (2019)).

The tapered and hybrid Whittle likelihoods require the use of data tapers. We use a Tukey taper defined as in (3.2). We set the proportion of tapering at each end of the time series is $0.1$, i.e. $d = n/10$.

When evaluating the boundary corrected Whittle likelihood and hybrid Whittle likelihood, the order $p$ is selected with the AIC and $\widehat{f}_p$ is estimated using the Yule-Walker estimator.

Unlike the Whittle, the tapered Whittle and debiased Whittle likelihood, $\mathrm{Re}\, \widetilde{J}_n(\omega_{k,n}; \widehat{f}_p)\overline{J_n(\omega_{k,n})}$ and $\mathrm{Re}\, \widetilde{J}_n(\omega_{k,n}; \widehat{f}_p)\overline{J_{n,\underline{h}_n}(\omega_{k,n})}$ can be negative. To avoid negative values, we apply the thresholding function $f(t) = \max(t, 10^{-3})$ to $\mathrm{Re}\, \widetilde{J}_n(\omega_{k,n}; \widehat{f}_p)\overline{J_n(\omega_{k,n})}$ and $\mathrm{Re}\, \widetilde{J}_n(\omega_{k,n}; \widehat{f}_p)\overline{J_{n,\underline{h}_n}(\omega_{k,n})}$ over all the frequencies. Thresholding induces an additional (small) bias to the new criteria. The proportion of times that $\mathrm{Re}\, \widetilde{J}_n(\omega_{k,n}; \widehat{f}_p)\overline{J_n(\omega_{k,n})}$ drops below the threshold increases for spectral density

functions with large peaks and when the spectral density is close to zero. However, at least for the models that we studied in the simulations, the bias due to the thresholding is negligible.

All simulations are conducted over 1000 replications with sample sizes $n = 20$, $50$, and $300$. In all the tables below and Appendix, the bias of the estimates are reported in the table and the standard deviation are in parenthesis. The ordering of the performance of the estimators is colour coded and is based on their squared root of the mean squared error (RMSE).

### 4.6.1 Estimation with correctly specified models

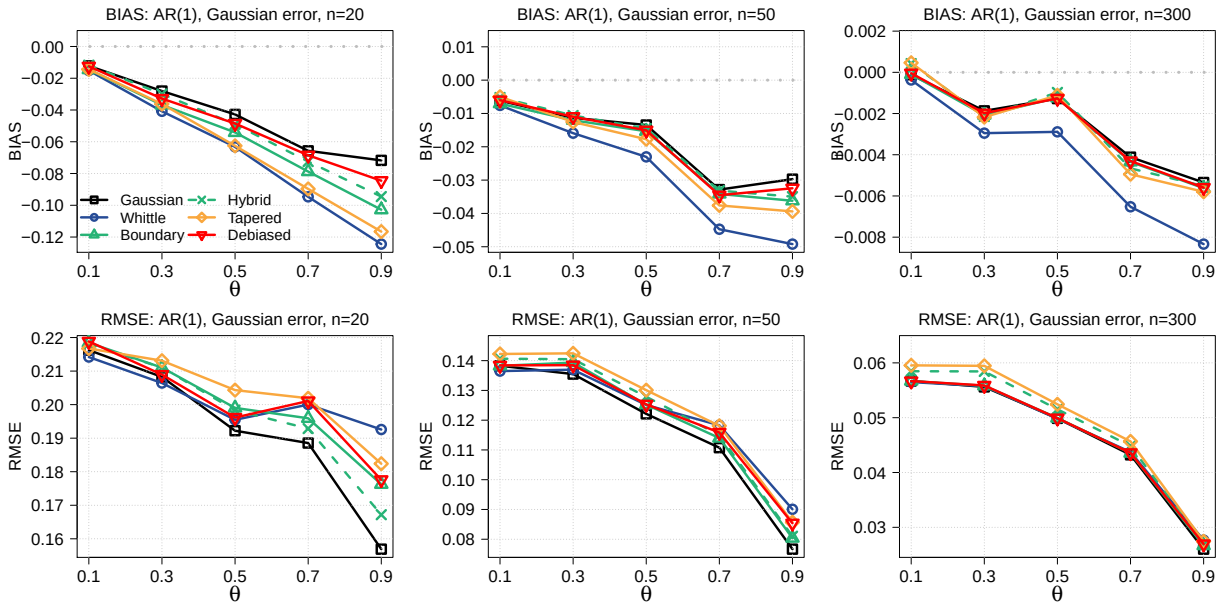
We first study the AR(1) and MA(1) parameter estimates when the models are correctly specified. We generate two types of time series models $\underline{X}_n$ and $\underline{Y}_n$, which satisfy the following recursions

$$\begin{aligned} \mathbf{AR(1)}: &\quad X_t = \theta X_{t-1} + e_t; \quad \phi_X(\omega) = 1 - \theta e^{-i\omega} \\ \mathbf{MA(1)}: &\quad Y_t = e_t + \theta e_{t-1}; \quad \phi_Y(\omega) = (1 + \theta e^{-i\omega})^{-1}, \end{aligned}$$

where $|\theta| < 1$, $\{e_t\}$ are independent, identically distributed Gaussian random variables with mean 0 and variance 1. Note that the Gaussianity of the innovations is not required to obtain the theoretical properties of the estimations. In Appendix C.1, we include simulations when the innovations follow a standardized chi-squared distribution with two degrees of freedom. The results are similar to those with Gaussian innovations. We generate the AR$(1)$ and MA$(1)$ models with parameters $\theta = 0.1, 0.3, 0.5, 0.7$ and $0.9$. For the time series generated by an AR$(1)$ process, we fit an AR$(1)$ model, similarly, for the time series generated by a MA$(1)$ process we fit a MA$(1)$ model.

For each simulation, we evaluate the six different parameter estimators. The empirical bias and standard deviation are calculated. Figures 4.2 gives the bias (first row) and the RMSE (second row) of each estimated parameter $\theta$ for both AR$(1)$ and MA$(1)$ models. We focus on positive $\theta$, similar results are obtained for negative $\theta$. The results are also summarized in Table 4.1.

**AR(1) model**




**MA(1) model**

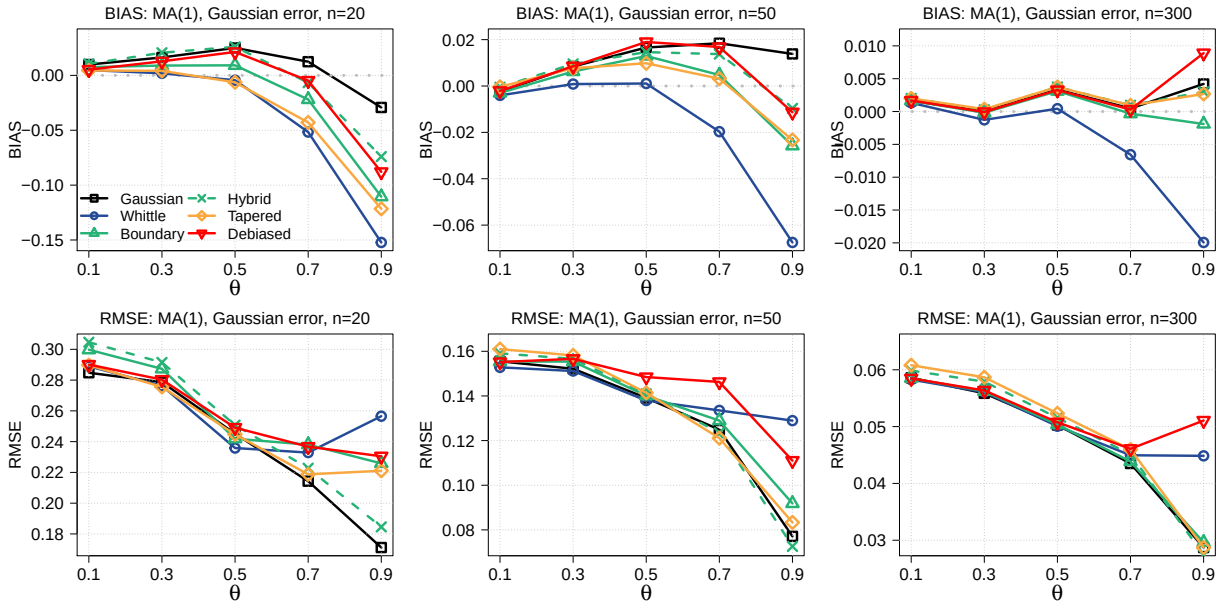



Figure 4.2: Bias (first row) and the RMSE (second row) of the parameter estimates for the Gaussian AR(1) models and Gaussian MA(1) models. Length of the time series $n = 20$(left), $50$(middle), and $300$(right).

| *Likelihoods* | $\theta$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
| | **AR(1)**, $\{e_t\} \sim \mathcal{N}(0,1)$, $n = 20$ | | | | | **MA(1)**, $\{e_t\} \sim \mathcal{N}(0,1)$, $n = 20$ | | | | |
| Gaussian | -0.012(0.22) | -0.028(0.21) | -0.043(0.19) | -0.066(0.18) | -0.072(0.14) | 0.010(0.28) | 0.016(0.28) | 0.025(0.24) | 0.012(0.21) | 0.029(0.17) |
| Whittle | -0.015(0.21) | -0.041(0.20) | -0.063(0.19) | -0.095(0.18) | -0.124(0.15) | 0.005(0.29) | 0.002(0.28) | -0.004(0.24) | -0.052(0.23) | -0.152(0.21) |
| Boundary | -0.015(0.22) | -0.037(0.21) | -0.054(0.19) | -0.079(0.18) | -0.103(0.14) | 0.007(0.30) | 0.009(0.29) | 0.009(0.24) | -0.022(0.24) | -0.111(0.20) |
| Hybrid | -0.012(0.22) | -0.030(0.21) | -0.049(0.19) | -0.072(0.18) | -0.095(0.14) | 0.011(0.30) | 0.021(0.29) | 0.026(0.25) | -0.007(0.22) | -0.074(0.17) |
| Tapered | -0.014(0.22) | -0.036(0.21) | -0.063(0.19) | -0.090(0.18) | -0.117(0.14) | 0.004(0.29) | 0.004(0.28) | -0.006(0.24) | -0.043(0.21) | -0.122(0.18) |
| Debiased | -0.013(0.22) | -0.033(0.21) | -0.049(0.19) | -0.069(0.19) | -0.085(0.16) | 0.005(0.29) | 0.013(0.28) | 0.021(0.25) | -0.005(0.24) | -0.088(0.21) |
| | **AR(1)**, $\{e_t\} \sim \mathcal{N}(0,1)$, $n = 50$ | | | | | **MA(1)**, $\{e_t\} \sim \mathcal{N}(0,1)$, $n = 50$ | | | | |
| Gaussian | -0.006(0.14) | -0.011(0.14) | -0.013(0.12) | -0.033(0.11) | -0.030(0.07) | -0.002(0.16) | 0.008(0.15) | 0.017(0.14) | 0.018(0.12) | 0.014(0.08) |
| Whittle | -0.008(0.14) | -0.016(0.14) | -0.023(0.12) | -0.045(0.11) | -0.049(0.08) | -0.004(0.15) | 0.001(0.15) | 0.001(0.14) | -0.020(0.13) | -0.067(0.11) |
| Boundary | -0.007(0.14) | -0.012(0.14) | -0.015(0.12) | -0.034(0.11) | -0.036(0.07) | -0.003(0.16) | 0.006(0.16) | 0.013(0.14) | 0.005(0.13) | -0.026(0.09) |
| Hybrid | -0.005(0.14) | -0.011(0.14) | -0.015(0.13) | -0.033(0.11) | -0.035(0.07) | -0.001(0.16) | 0.010(0.16) | 0.015(0.14) | 0.014(0.12) | -0.010(0.07) |
| Tapered | -0.005(0.14) | -0.013(0.14) | -0.018(0.13) | -0.038(0.11) | -0.039(0.08) | 0(0.16) | 0.008(0.16) | 0.010(0.14) | 0.003(0.12) | -0.023(0.08) |
| Debiased | -0.006(0.14) | -0.011(0.14) | -0.015(0.12) | -0.035(0.11) | -0.032(0.08) | -0.002(0.16) | 0.009(0.16) | 0.019(0.15) | 0.017(0.15) | -0.011(0.11) |
| | **AR(1)**, $\{e_t\} \sim \mathcal{N}(0,1)$, $n = 300$ | | | | | **MA(1)**, $\{e_t\} \sim \mathcal{N}(0,1)$, $n = 300$ | | | | |
| Gaussian | 0(0.06) | -0.002(0.06) | -0.001(0.05) | -0.004(0.04) | -0.005(0.03) | 0.002(0.06) | 0(0.06) | 0.003(0.05) | 0(0.04) | 0.004(0.03) |
| Whittle | 0(0.06) | -0.003(0.06) | -0.003(0.05) | -0.007(0.04) | -0.008(0.03) | 0.001(0.06) | -0.001(0.06) | 0(0.05) | -0.007(0.04) | -0.020(0.04) |
| Boundary | 0(0.06) | -0.002(0.06) | -0.001(0.05) | -0.004(0.04) | -0.006(0.03) | 0.002(0.06) | 0(0.06) | 0.003(0.05) | 0(0.04) | -0.002(0.03) |
| Hybrid | 0(0.06) | -0.002(0.06) | -0.001(0.05) | -0.005(0.04) | -0.006(0.03) | 0.002(0.06) | 0(0.06) | 0.004(0.05) | 0.001(0.05) | 0.003(0.03) |
| Tapered | 0(0.06) | -0.002(0.06) | -0.001(0.05) | -0.005(0.05) | -0.006(0.03) | 0.002(0.06) | 0(0.06) | 0.004(0.05) | 0.001(0.05) | 0.003(0.03) |
| Debiased | 0(0.06) | -0.002(0.06) | -0.001(0.05) | -0.004(0.04) | -0.006(0.03) | 0.002(0.06) | 0(0.06) | 0.003(0.05) | 0(0.05) | 0.009(0.05) |

Table 4.1: Bias and the standard deviation (in the parentheses) of six different quasi-likelihoods for an AR(1) (left) and MA(1) (right) model for the standard normal innovations. Length of the time series $n = 20, 50$, and 300. We use red text to denote the smallest RMSE and blue text to denote the second smallest RMSE.

For both AR$(1)$ and MA$(1)$ models, we observe a stark difference between the bias of the Whittle likelihood estimator (blue line) and the other five other methods, which in most cases have a lower bias. The Gaussian likelihood performs uniformly well for both models and all sample sizes. Whereas, the tapered Whittle estimator performs very well for the MA$(1)$ model but not quite as well for the AR$(1)$ model. The debiased Whittle likelihood performs quite well for both models, especially when the parameter values are small (e.g. $\theta = 0.1, 0.3$, and $0.5$).

The simulations suggest that the boundary corrected and hybrid Whittle likelihoods (referred from now on as the new likelihoods) are competitive with the benchmark Gaussian likelihood for both AR$(1)$ and MA$(1)$ models. For the AR$(1)$ model the new likelihoods tend to have the smallest or second smallest RMSE (over all sample sizes and more so when $\phi$ is large). A caveat is that for the AR$(1)$ model the bias of the new likelihoods tends to be a little larger than the bias of the Gaussian likelihood (especially for the smaller sample sizes). This is interesting, because in Appendix B.2 we show that if the AR$(1)$ model is correctly specified, the first order bias of the boundary corrected Whittle likelihood and the Gaussian likelihood are the same (both are $-2\theta/n$). The bias of the hybrid Whittle likelihood is slightly large, due to the data taper. However, there are differences in the second order expansions. Specifically, for the Gaussian likelihood, it is $O(n^{-3/2})$, whereas, for the new likelihoods it is $O(p^3n^{-3/2})$. Indeed, the $O(p^3n^{-3/2})$ term arises because of the parameter estimation in the predictive DFT. This term is likely to dominate the $O(n^{-3/2})$ in the Gaussian likelihood. Therefore, for small sample sizes, the second order terms can impact the bias. It is this second order term that may be causing the larger bias seen in the boundary corrected Whittle likelihood as compared with the Gaussian likelihood.

On the other hand, the bias for the MA$(1)$ model tends to be smaller for the new likelihoods, including the benchmark Gaussian likelihood. Surprisingly, there appears to be examples where the new likelihood does better (in terms of RMSE) than the Gaussian likelihood. This happens when $n \in \{50, 300\}$ for $\theta = 0.9$.

In summary, the new likelihoods perform well compared with the standard methods, including the benchmark Gaussian likelihood. As expected, for large sample sizes the performance of all the

estimators improves considerably.

### 4.6.2 Estimation under misspecification

Next, we turn into our attention to the case that the model is misspecified (which is more realistic for real data). As we mentioned above, the estimation of the AR parameters in the predictive DFT of the new likelihoods leads to an additional error of order $O(p^3 n^{-3/2})$. The more complex the model, the larger $p$ will be, leading to a larger $O(p^3 n^{-3/2})$. To understand the effect this may have for small sample sizes, in this section we fit a simple model to a relatively complex process.

For the "true" data generating process we use an ARMA$(3, 2)$ Gaussian time series with spectral density $f_Z(\omega) = |\psi_Z(e^{-i\omega})|^2/|\phi_Z(e^{-i\omega})|^2$, where AR and MA characteristic polynomials are

$$\phi_Z(z) = (1 - 0.7z)(1 - 0.9e^i z)(1 - 0.9e^{-i} z) \quad \text{and} \quad \psi_Z(z) = (1 + 0.5z + 0.5z^2).$$

This spectral density has some interesting characteristics: a pronounced peak, a large amount of power at the low frequencies, and a sudden drop in power at the higher frequencies. We consider sample sizes $n = 20, 50$ and $300$, and fit a model with fewer parameters. Specifically, we fit two different ARMA models with the same number of unknown parameters. The first is the ARMA(1,1) model with spectral density

$$f_\theta(\omega) = |1 + \psi e^{-i\omega}|^2 |1 - \phi e^{-i\omega}|^{-2} \qquad \theta = (\phi, \psi).$$

The second is the AR(2) model with spectral density

$$f_\theta(\omega) = |1 - \phi_1 e^{-i\omega} - \phi_2 e^{-2i\omega}|^{-2} \qquad \theta = (\phi_1, \phi_2).$$

Figure 4.3 shows the logarithm of the theoretical ARMA(3,2) spectral density (solid line, $f_Z$) and the corresponding log spectral densities of the best fitting ARMA(1,1) (dashed line) and AR(2) (dotted line) processes for $n = 20$. The best fitting models are obtained by minimizing the spectral divergence $\theta^{Best} = \arg\min_{\theta \in \Theta} I_n(f; f_\theta)$, where $I_n(f, f_\theta)$ is defined in (4.1) and $\Theta$ is the parameter

space. The best fitting models for $n = 50$ and $300$ are similar. We observe that neither of the misspecified models capture all of the features of the true spectral density. The best fitting ARMA(1,1) model has a large amount of power at the low frequencies and the power declines for the higher frequencies. The best fitting AR(2) model peaks around frequency 0.8, but the power at the low frequencies is small. Overall, the spectral divergence between the true and the best fitting AR(2) model is smaller than the spectral divergence between the true and the best ARMA(1,1) model.

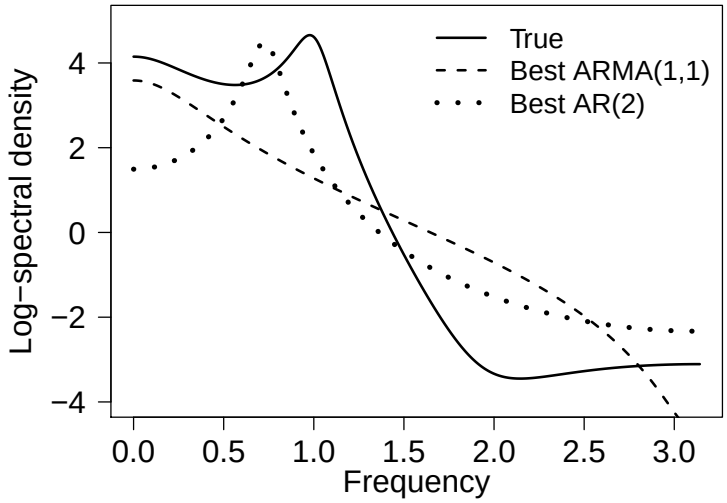



Figure 4.3: Plot of $\log f_Z(\omega)$ and $\log f_{\theta^{Best}}(\omega)$; Theoretical ARMA(3,2) spectral density (solid), best fitting ARMA(1,1) spectral density (dashed), and best fitting AR(2) spectral density (dotted) for $n = 20$.

For each simulation, we calculate the six different parameter estimators and the spectral divergence. The result of the estimators using the six different quasi-likelihoods is given in Table 4.2 (for ARMA(1,1)) and Table 4.3 (for AR(2)).

We first discuss the parameter estimates. Comparing the asymptotic bias of the Gaussian likelihood with the boundary corrected Whittle likelihood (see Appendix B.3), the Gaussian likelihood has an additional bias term of form $\sum_{r=1}^{d} I^{(j,r)} \mathbb{E}[\frac{\partial \mathcal{L}_n}{\partial \theta_r}]]_{\theta_n}$. But there is no guarantee that the inclusion of this term increases or decreases the bias. This is borne out in the simulations, where we

| $n$ | Parameter | Gaussian | Whittle | Boundary | Hybrid | Tapered | Debiased |
|---|---|---|---|---|---|---|---|
| 20 | $\phi$ | 0.031(0.1) | -0.095(0.16) | -0.023(0.12) | -0.006(0.1) | -0.080(0.13) | 0.187(0.11) |
| | $\psi$ | 0.069(0.08) | -0.172(0.18) | -0.026(0.14) | 0.028(0.1) | -0.068(0.12) | 0.093(0.06) |
| | $I_n(f;f_\theta)$ | 1.653(0.81) | 1.199(1.57) | 0.945(0.84) | 1.024(0.89) | 0.644(0.61) | 2.727(0.73) |
| 50 | $\phi$ | 0.012(0.07) | -0.054(0.09) | -0.006(0.07) | 0.004(0.07) | -0.005(0.07) | 0.154(0.11) |
| | $\psi$ | 0.029(0.06) | -0.116(0.12) | -0.008(0.08) | 0.009(0.07) | 0.011(0.06) | 0.093(0) |
| | $I_n(f;f_\theta)$ | 0.354(0.34) | 0.457(0.46) | 0.292(0.3) | 0.235(0.28) | 0.225(0.26) | 1.202(0.34) |
| 300 | $\phi$ | 0.002(0.03) | -0.014(0.03) | 0(0.03) | 0.001(0.03) | 0(0.03) | 0.093(0.08) |
| | $\psi$ | 0.005(0.03) | -0.033(0.05) | 0.001(0.03) | 0.003(0.03) | 0.003(0.03) | 0.092(0.01) |
| | $I_n(f;f_\theta)$ | 0.027(0.05) | 0.064(0.09) | 0.029(0.05) | 0.026(0.04) | 0.027(0.05) | 0.752(0.22) |

Best fitting ARMA$(1,1)$ coefficients $\theta = (\phi, \psi)$ and spectral divergence:
- $\theta_{20} = (0.693, 0.845)$, $\theta_{50} = (0.694, 0.857)$, $\theta_{300} = (0.696, 0.857)$.
- $I_{20}(f;f_\theta) = 3.773$, $I_{50}(f;f_\theta) = 3.415$, $I_{300}(f;f_\theta) = 3.388$.

Table 4.2: The bias of estimated coefficients for six different estimation methods for the Gaussian ARMA$(3,2)$ misspecified case fitting ARMA$(1,1)$ model. Standard deviations are in the parentheses. We use red text to denote the smallest RMSE and blue text to denote the second smallest RMSE.

observe that overall the Gaussian likelihood or the new likelihoods tend to have a smaller parameter bias (there is no clear winner). The tapered likelihood is a close contender, performing very well for the moderate sample sizes $n = 50$. Similarly, in terms of the RMSE, again there is no clear winner between the Gaussian and the new likelihoods.

We next turn our attention to the estimated spectral divergence $I_n(f, f_{\hat{\theta}})$. For the fitted ARMA$(1,1)$ model, the estimated spectral divergence of the new likelihood estimators tends to be the smallest or second smallest in terms of the RMSE (its nearest competitor is the tapered likelihood). On the other hand, for the AR$(2)$ model the spectral divergence of Gaussian likelihood has the smallest RMSE for all the sample sizes. The new likelihood comes in second for sample sizes $n = 20$ and 300.

In the simulations above we select $p$ using the AIC. As mention at the start of the section, this leads to an additional error of $O(p^3 n^{-3/2})$ in the new likelihoods. Thus, if a large $p$ is selected the error $O(p^3 n^{-3/2})$ will be large. In order to understand the impact $p$ has on the estimator, in Appendix 4.6.3 we compare the the likelihoods constructed using the predictive DFT based on the AIC with the likelihoods constructed using the predictive DFT based on the best fitting estimated AR$(1)$ model. We simulate from the ARMA$(3,2)$ model described above and fit an ARMA$(1,1)$

| $n$ | Parameter | Gaussian | Whittle | Boundary | Hybrid | Tapered | Debiased |
|---|---|---|---|---|---|---|---|
| 20 | $\phi_1$ | 0.028(0.14) | -0.162(0.22) | -0.032(0.16) | 0.003(0.14) | -0.123(0.16) | 0.069(0.15) |
| | $\phi_2$ | -0.004(0.09) | 0.169(0.18) | 0.052(0.14) | 0.025(0.12) | 0.132(0.12) | -0.034(0.11) |
| | $I_n(f;f_\theta)$ | 0.679(0.72) | 1.203(1.46) | 0.751(0.85) | 0.684(0.8) | 0.862(0.97) | 0.686(0.81) |
| 50 | $\phi_1$ | 0.019(0.09) | -0.077(0.12) | -0.009(0.09) | 0.003(0.09) | -0.017(0.09) | 0.156(0.15) |
| | $\phi_2$ | -0.024(0.06) | 0.066(0.1) | 0.006(0.07) | -0.003(0.06) | 0.013(0.06) | -0.121(0.06) |
| | $I_n(f;f_\theta)$ | 0.275(0.33) | 0.382(0.45) | 0.283(0.37) | 0.283(0.37) | 0.283(0.36) | 0.65(0.7) |
| 300 | $\phi_1$ | 0.004(0.04) | -0.013(0.04) | 0(0.04) | 0.001(0.04) | 0.001(0.04) | 0.014(0.04) |
| | $\phi_2$ | -0.005(0.02) | 0.011(0.03) | -0.001(0.02) | -0.001(0.03) | -0.001(0.03) | 0.016(0.04) |
| | $I_n(f;f_\theta)$ | 0.049(0.07) | 0.053(0.07) | 0.049(0.07) | 0.053(0.07) | 0.054(0.08) | 0.058(0.08) |

Best fitting AR(2) coefficients $\theta = (\phi_1, \phi_2)$ and spectral divergence:
- $\theta_{20} = (1.367, -0.841)$, $\theta_{50} = (1.364, -0.803)$, $\theta_{300} = (1.365, -0.802)$.
- $I_{20}(f;f_\theta) = 2.902$, $I_{50}(f;f_\theta) = 2.937$, $I_{300}(f;f_\theta) = 2.916$.

Table 4.3: The bias of estimated coefficients for six different estimation methods for the Gaussian ARMA$(3,2)$ misspecified case fitting AR$(2)$ model. Standard deviations are in the parentheses. We use red text to denote the smallest RMSE and blue text to denote the second smallest RMSE.

and AR$(2)$ model. As is expected, the bias tends to be a little larger when the order is fixed to $p = 1$. But even when fixing $p = 1$, we do observe an improvement over the Whittle likelihood (in some cases an improvement over the Gaussian likelihood).

### 4.6.3 Comparing the new likelihoods constructed with the predictive DFT with AR$(1)$ coefficients and AIC order selected AR$(p)$ coefficients

In this section we compare the performance of new likelihoods where the order of the AR model used in the predictive DFT is determined using the AIC with a fixed choice of order with the AR model (set to $p = 1$). We use ARMA$(3,2)$ model considered in Section 4.6.2 and fit the the ARMA$(1,1)$ and AR$(2)$ to the data. We compare the new likelihoods with the Gaussian likelihood and the Whittle likelihood. The results are given in Tables 4.4 and 4.5.

### 4.6.4 Alternative methods for estimating the predictive DFT

As pointed out by the referees, using the Yule-Walker estimator to estimate the prediction coefficients in the predictive DFT may in certain situations be problematic. We discuss the issues and potential solutions below.

The first issue is that Yule-Walker estimator suffers a finite sample bias, especially when the spectral density has a root close to the unit circle (see, e.g., Tjøstheim and Paulsen (1983)). One

| | | $\phi$ | $\psi$ | $I_n(f;f_\theta)$ |
|---|---|---|---|---|
| Best | | 0.694 | 0.857 | 3.415 |
| Bias | Gaussian | 0.012(0.07) | 0.029(0.06) | 0.354(0.34) |
| | Whittle | -0.054(0.09) | -0.116(0.12) | 0.457(0.46) |
| | Boundary(AIC) | -0.006(0.07) | -0.008(0.08) | 0.292(0.3) |
| | Boundary($p$=1) | -0.020(0.08) | -0.045(0.09) | 0.299(0.29) |
| | Hybrid(AIC) | 0.004(0.07) | 0.009(0.07) | 0.235(0.28) |
| | Hybrid($p$=1) | 0.003(0.07) | 0.010(0.07) | 0.261(0.3) |

Table 4.4: Best fitting (top row) and the bias of estimated coefficients for six different methods for the Gaussian ARMA$(3,2)$ misspecified case fitting ARMA$(1,1)$ model. Length of the time series $n$=50. Standard deviations are in the parentheses. (AIC): an order $p$ is chosen using AIC; ($p$=1): an order $p$ is set to 1.

| | | $\phi_1$ | $\phi_2$ | $I_n(f;f_\theta)$ |
|---|---|---|---|---|
| Best | | 1.364 | -0.803 | 2.937 |
| Bias | Gaussian | 0.019(0.09) | -0.024(0.06) | 0.275(0.33) |
| | Whittle | -0.077(0.12) | 0.066(0.1) | 0.382(0.45) |
| | Boundary(AIC) | -0.009(0.09) | 0.006(0.07) | 0.283(0.37) |
| | Boundary($p$=1) | -0.030(0.1) | 0.032(0.07) | 0.295(0.35) |
| | Hybrid(AIC) | 0.003(0.09) | -0.006(0.07) | 0.283(0.37) |
| | Hybrid($p$=1) | -0.003(0.09) | 0.003(0.06) | 0.276(0.35) |

Table 4.5: Best fitting (top row) and the bias of estimated coefficients for six different methods for the Gaussian ARMA$(3,2)$ misspecified case fitting AR$(2)$ model. Length of the time series $n$=50. Standard deviations are in the parentheses. (AIC): an order $p$ is chosen using AIC; ($p$=1): an order $p$ is set to 1.

remedy to reduce the bias is via data tapering (Dahlhaus (1988) and Zhang (1992)). Therefore, we define the boundary corrected Whittle likelihood using tapered Yule-Walker (BC-tYW) replace $\widehat{f}_p$ with $\widetilde{f}_p$ in (4.2) where $\widetilde{f}_p$ is a spectral density of AR$(p)$ process where the AR coefficients are estimated using Yule-Walker with tapered time series. In the simulations we use the Tukey taper with $d = n/10$ and select the order $p$ using the AIC.

The second issue is if the underlying time series is complicated in the sense that the underlying AR representation has multiple roots. Then fitting a large order AR$(p)$ model may result in a loss of efficiency. As an alternative, we consider a fully nonparametric estimator of $\widehat{J}_n(\omega;f)$ based on the estimated spectral density function. To do so, we recall from Section 2.4.1 the first order

approximation of $\widehat{J}_n(\omega; f)$ is $\widehat{J}_{\infty,n}(\omega; f)$ where

$$\begin{aligned}\widehat{J}_{\infty,n}(\omega; f) &= \frac{n^{-1/2}}{\phi(\omega; f)} \sum_{t=1}^{n} X_t \phi_t^{\infty}(\omega; f) + e^{i(n+1)\omega} \frac{n^{-1/2}}{\overline{\phi(\omega; f)}} \sum_{t=1}^{n} X_{n+1-t} \overline{\phi_t^{\infty}(\omega; f)} \\ &= \frac{\psi(\omega; f)}{\sqrt{n}} \sum_{t=1}^{n} X_t \sum_{s=0}^{\infty} \phi_{s+t}(f) e^{-is\omega} + e^{i(n+1)\omega} \frac{\overline{\psi(\omega; f)}}{\sqrt{n}} \sum_{t=1}^{n} X_{n+1-t} \sum_{s=0}^{\infty} \phi_{s+t}(f) e^{is\omega},\end{aligned}$$

where $\psi(\omega; f) = \sum_{j=0}^{\infty} \psi_j(f) e^{-ij\omega}$ be an MA transfer function. Our goal is to estimate $\psi(\omega; f)$ and $\{\phi_j(f)\}$ based on the observed time series. We use the method proposed in Section 2.2. of Krampe et al. (2018). We first start from the well known Szegö's identity

$$\log f(\cdot) = \log \sigma^2 |\psi(\cdot; f)|^2 = \log \sigma^2 + \log \psi(\cdot; f) + \log \overline{\psi(\cdot; f)}.$$

Next, let $\alpha_k(f)$ be the $k$-th Fourier coefficient of $\log f$, i.e., $\alpha_k(f) = (2\pi)^{-1} \int_{-\pi}^{\pi} \log f(\lambda) e^{-ik\lambda} d\lambda$. Then, since $\log f$ is real, $\alpha_{-k}(f) = \overline{\alpha_k(f)}$. Plug in the expansion of $\log f$ to the above identity gives

$$\log \psi(\omega; f) = \sum_{j=1}^{\infty} \alpha_j(f) e^{-ij\omega}.$$

Using above identity, we estimator $\psi(\cdot; f)$. let $\widehat{f}$ be a spectral density estimator and let $\widehat{\alpha}_k$ be the estimated $k$-th Fourier coefficient of $\log \widehat{f}$. Then define

$$\widehat{\psi}(\omega; \widehat{f}) = \exp\left( \sum_{j=1}^{M} \widehat{\alpha}_j e^{-ij\omega} \right)$$

for some large enough $M$. To estimate the AR($\infty$) coefficients we use the recursive formula in equation (2.7) in Krampe et al. (2018),

$$\widehat{\phi}_{k+1} = -\sum_{j=0}^{k} \left(1 - \frac{j}{k+1}\right) \widehat{\alpha}_{k+1-j} \widehat{\phi}_j \qquad k = 0, 1, ..., M-1$$

where $\widehat{\phi}_0 = -1$. Based on this a nonparametric estimator of $\widehat{J}_n(\omega; f)$ is

$$\widehat{J}_n(\omega; \widehat{f}) = \frac{\widehat{\psi}(\omega; \widehat{f})}{\sqrt{n}} \sum_{t=1}^{n \wedge M} X_t \sum_{s=0}^{M-t} \widehat{\phi}_{s+t} e^{-is\omega} + e^{i(n+1)\omega} \frac{\overline{\widehat{\psi}(\omega; \widehat{f})}}{\sqrt{n}} \sum_{t=1}^{n \wedge M} X_{n+1-t} \sum_{s=0}^{M-t} \widehat{\phi}_{s+t} e^{is\omega}$$

where $n \wedge M = \min(n, M)$. In the simulations we estimate $\widehat{f}$ using `iospecden` function in R (smoothing with infinite order Flat-top kernel) and set $M$=30.

By replacing $\widehat{J}_n(\omega; f)$ with its nonparametric estimator $\widehat{J}_n(\omega; \widehat{f})$ in (4.2) leads us to define a new feasible criterion which we call the boundary corrected Whittle likelihood using nonparametric estimation (BC-NP).

To access the performance of all the different likelihoods (with different estimates of the predictive DFT), we generate the AR$(8)$ model

$$U_t = \phi_U(B)\varepsilon_t$$

where $\{\varepsilon_t\}$ are i.i.d. normal random variables,

$$\phi_U(z) = \prod_{j=1}^{4}(1 - r_j e^{i\lambda_j} z)(1 - r_j e^{-i\lambda_j} z) = 1 - \sum_{j=1}^{8} \phi_j z^j \tag{4.30}$$

$\underline{r} = (r_1, r_2, r_3, r_4) = (0.95, 0.95, 0.95, 0.95)$ and $\underline{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0.5, 1, 2, 2.5)$. We observe that corresponding spectral density $f_U(\omega) = |\phi_U(e^{-i\omega})|^{-2}$ has pronounced peaks at $\omega = 0.5, 1, 1.5$ and 2. For all the simulations below we use $n = 100$.

For each simulation, we fit AR$(8)$ model, evaluate six likelihoods from the previous sections plus two likelihoods (BC-tYW and BC-NP), and calculate the parameter estimators. Table 4.6 summarizes the bias and standard derivation of the estimators and the last row is an average $\ell_2$-distance between the true and estimator scaled with $n$. The Gaussian likelihood has the smallest bias and the smallest RMSE. As mentioned in Section 4.6.1, our methods still need to estimate AR coefficients which has an additional error of order $O(p^3 n^{-3/2})$ and it could potentially increase the bias compared to the Gaussian likelihood. The boundary corrected Whittle and hybrid Whittle

have smaller bias than the Whittle, tapered, and debiased Whittle. Especially, the hybrid Whittle usually has the second smallest RMSE.

| Par. | Bias | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Gaussian | Whittle | Boundary | Hybrid | Tapered | Debiased | BC-tYW | BC-NP |
| $\phi_1$(0.381) | -0.008(0.08) | -0.025(0.09) | -0.009(0.08) | -0.006(0.09) | -0.012(0.09) | -0.008(0.09) | -0.008(0.08) | -0.005(0.12) |
| $\phi_2$(-0.294) | 0.002(0.09) | 0.024(0.1) | 0.005(0.09) | 0.002(0.09) | 0.010(0.09) | 0.003(0.1) | 0.003(0.09) | 0.002(0.13) |
| $\phi_3$(0.315) | -0.009(0.08) | -0.038(0.09) | -0.011(0.09) | -0.009(0.09) | -0.023(0.09) | -0.010(0.09) | -0.009(0.09) | -0.010(0.12) |
| $\phi_4$(-0.963) | 0.031(0.09) | 0.108(0.1) | 0.042(0.09) | 0.034(0.09) | 0.075(0.09) | 0.043(0.1) | 0.037(0.09) | 0.076(0.12) |
| $\phi_5$(0.285) | -0.015(0.08) | -0.049(0.09) | -0.020(0.09) | -0.016(0.08) | -0.029(0.08) | -0.017(0.1) | -0.018(0.09) | -0.022(0.12) |
| $\phi_6$(-0.240) | 0.010(0.08) | 0.040(0.09) | 0.014(0.09) | 0.010(0.09) | 0.024(0.08) | 0.012(0.1) | 0.011(0.09) | 0.022(0.11) |
| $\phi_7$(0.280) | -0.017(0.08) | -0.053(0.09) | -0.021(0.09) | -0.020(0.09) | -0.039(0.08) | -0.022(0.09) | -0.020(0.09) | -0.027(0.1) |
| $\phi_8$(-0.663) | 0.049(0.08) | 0.116(0.08) | 0.059(0.08) | 0.055(0.08) | 0.096(0.08) | 0.061(0.09) | 0.056(0.08) | 0.101(0.1) |
| $n\|\underline{\phi} - \widehat{\underline{\phi}}\|_2$ | 6.466 | 18.607 | 8.029 | 7.085 | 13.611 | 8.164 | 7.470 | 13.280 |

Table 4.6: Bias and the standard deviation (in the parenthesis) of eight different quasi-likelihoods for the Gaussian AR(8) model. Length of time series $n$=100. True AR coefficients are in the parenthesis of the first column.

Bear in mind that neither of the two new criteria uses a hybrid method (tapering on the actual DFT), the BC-tYW significantly reduces the bias than the boundary corrected Whittle and it is comparable with the hybrid Whittle. This gives some credence to the referee's claim that the bias due to the Yule-Walker estimation can be alleviated using tapered Yule-Walker estimation. Whereas, BC-NP reduces the bias for the first few coefficients but overall, has a larger bias than the boundary corrected Whittle. Also, the standard deviation of BC-NP is quite large than other methods. We suspect that the nonparametric estimator $\widehat{J}(\omega; \widehat{f})$ is sensitive to the choice of the tuning parameters (e.g. bandwidth, kernel function, etc). Moreover, since the true model follows a finite autoregressive process, other methods (boundary corrected Whittle, BC-tYW, and hybrid Whittle) have an advantage over the nonparametric method. Therefore, by choosing appropriate tuning parameters under certain underlying process (e.g., seasonal ARMA model) can improve the estimators, and this will be investigated in future research.

## 4.7 Proofs

In this section, we give a proof of Sections 2.4 and 4.

### 4.7.1 Proof of Section 2.4

PROOF of Theorem 2.4.2 We use the same notation in the proof of Theorem 2.2.2 in Section 3.6.1. To prove (2.33) we use that

$$(\widehat{J}_n(\omega_{1,n};f_\theta),\ldots,\widehat{J}_n(\omega_{n,n};f_\theta))' = D_n(f_\theta)\underline{X}_n.$$

Now by using (2.8) together with the above we immediately obtain (2.33).

Finally, we prove (2.34). We use the result $n^{-1}\sum_{k=1}^n \phi_p(\omega_{k,n})\exp(is\omega_{k,n}) = \widetilde{\phi}_{s \bmod n}$ where $\phi_p(\omega) = \sum_{r=0}^{n-1}\widetilde{\phi}_r e^{-ir\omega}$ and $\widetilde{\phi}_r = 0$ for $p+1 \leqslant r \leqslant n$. For $1 \leqslant t \leqslant p$ we use have

$$\begin{aligned}
(F_n^*\Delta_n(f_\theta^{-1})D_n(f_\theta))_{s,t} &= \frac{1}{n}\sum_{k=1}^n \frac{\phi_{t,p}(\omega_{k,n})}{f_\theta(\omega_{k,n})}\exp(-is\omega_{k,n}) \\
&= \frac{\sigma^{-2}}{n}\sum_{k=1}^n \overline{\phi_p(\omega_{k,n})}\sum_{\ell=0}^{p-t}\phi_{\ell+t}\exp(-i\ell\omega_{k,n})\exp(-is\omega_{k,n}) \\
&= \sigma^{-2}\sum_{\ell=0}^{p-t}\phi_{\ell+t}\frac{1}{n}\sum_{k=1}^n \overline{\phi_p(\omega_{k,n})}\exp(-i(\ell+s)\omega_{k,n}) \\
&= \sigma^{-2}\sum_{\ell=0}^{p-t}\phi_{\ell+t}\widetilde{\phi}_{(\ell+s) \bmod n}.
\end{aligned}$$

Similarly, for $1 \leqslant t \leqslant p$,

$$\begin{aligned}
(F_n^*\Delta_n(f_\theta^{-1})D_n(f_\theta))_{s,n-t+1} &= \frac{1}{n}\sum_{k=1}^n \frac{\overline{\phi_{t,p}(\omega_{k,n})}}{f_\theta(\omega_{k,n})}\exp(i(1-s)\omega_{k,n}) \\
&= \frac{\sigma^{-2}}{n}\sum_{k=1}^n \phi_p(\omega_{k,n})\sum_{\ell=0}^{p-t}\phi_{\ell+t}\exp(i\ell\omega_{k,n})\exp(i(1-s)\omega_{k,n}) \\
&= \sigma^{-2}\sum_{\ell=0}^{p-t}\phi_{\ell+t}\frac{1}{n}\sum_{k=1}^n \phi_p(\omega_{k,n})\exp(i(\ell+1-s)\omega_{k,n}) \\
&= \sigma^{-2}\sum_{\ell=0}^{p-t}\phi_{\ell+t}\widetilde{\phi}_{(\ell+1-s) \bmod n}.
\end{aligned}$$

□

PROOF of (2.35) We use that $\frac{1}{\phi_p(\omega)} f_\theta(\omega)^{-1} = \sigma^{-2}\overline{\phi_p(\omega)}$. This gives

$$\mathcal{L}_n(\phi) - K_n(\phi) = I + II$$

where

$$\begin{aligned} I &= \frac{1}{n^{3/2}} \sum_{k=1}^{n} \frac{\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} \left\{ \frac{1}{\phi_p(\omega_{k,n})} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{-is\omega_{k,n}} \right\} \\ &= \frac{\sigma^{-2}}{n^{3/2}} \sum_{k=1}^{n} \overline{J_n(\omega_{k,n})\phi_p(\omega_{k,n})} \sum_{\ell=1}^{n} \sum_{s=0}^{p-\ell} \phi_{s+k} e^{-is\omega_{k,n}} \phi_{\ell+s} e^{-is\omega_{k,n}} \\ &= \frac{\sigma^{-2}}{n} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{s+k} \frac{1}{n^{1/2}} \sum_{k=1}^{n} \overline{J_n(\omega_{k,n})\phi_p(\omega_{k,n})} e^{-is\omega_{k,n}} \end{aligned}$$

and

$$II = \frac{\sigma^{-2}}{n^{3/2}} \sum_{k=1}^{n} \frac{\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} \frac{1}{\overline{\phi_p(\omega_{k,n})}} \sum_{\ell=1}^{p} X_{n+1-\ell} \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{i(s+1)\omega_{k,n}}. \tag{4.31}$$

We first consider $I$. Using that $\overline{\phi_p(\omega_{k,n})} = 1 - \sum_{j=1}^{p} \phi_j e^{ij\omega_{k,n}}$ and $n^{-1/2} \sum_{k=1}^{n} \overline{J_n(\omega_{k,n})} e^{is\omega_{k,n}} = X_{s \bmod n}$, gives

$$\begin{aligned} I &= -\frac{\sigma^{-2}}{n} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \sum_{j=0}^{p} \phi_j \phi_{s+\ell} \frac{1}{n^{1/2}} \sum_{k=1}^{n} \overline{J_n(\omega_{k,n})} e^{-i(s-r)\omega_{k,n}} \text{ (set } \phi_0 = -1) \\ &= -\frac{\sigma^{-2}}{n} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \sum_{j=0}^{p} \phi_j \phi_{s+\ell} X_{-(s-j) \bmod n}. \end{aligned}$$

The proof of $II$ is similar. Altogether this proves the result. □

PROOF of (2.37) Since

$$(D_{\infty,n}(f_\theta))_{k,t} = n^{-1/2} \sum_{\tau \leqslant 0} \left( \phi_t(\tau) e^{i\tau\omega_{k,n}} + \phi_{n+1-t}(\tau) e^{-i(\tau-1)\omega_{k,n}} \right) \tag{4.32}$$

we replace $\phi_t(\tau)$ in the above with the coefficients of the MA and AR infinity expansions; $\phi_t(\tau) =$

$\sum_{s=0}^{\infty}\phi_{t+s}\psi_{|\tau|-s}$. Substituting this into the first term in (4.32) gives

$$\begin{aligned} n^{-1/2}\sum_{\tau\leqslant 0}\phi_t(\tau)e^{i\tau\omega_{k,n}} &= n^{-1/2}\sum_{\tau\leqslant 0}\sum_{s=0}^{\infty}\phi_{t+s}\psi_{-\tau-s}e^{i\tau\omega_{k,n}} \\ &= n^{-1/2}\sum_{s=0}^{\infty}\phi_{t+s}e^{-is\omega_{k,n}}\sum_{\tau\leqslant 0}\psi_{-\tau-s}e^{-i(-\tau-s)\omega_{k,n}} \\ &= n^{-1/2}\psi(\omega_{k,n})\sum_{s=0}^{\infty}\phi_{t+s}e^{-is\omega_{k,n}} \\ &= n^{-1/2}\phi(\omega_{k,n})^{-1}\phi_t^{\infty}(\omega_{k,n}), \end{aligned}$$

which gives the first term in (2.37). The second term follows similarly. Thus giving the identity in equation (2.37). □

PROOF of Theorem 2.4.3. We note that the entries of $F_n^*\Delta_n(f_\theta^{-1})D_{\infty,n}(f)$ are

$$\begin{aligned} &(F_n^*\Delta_n(f_\theta^{-1})D_{\infty,n}(f))_{s,t} \\ &\quad=\sum_{\tau\leqslant 0}\left[\phi_t(\tau;f)G_{1,n}(s,\tau;f_\theta)+\phi_{n+1-t}(\tau;f)G_{2,n}(s,\tau;f_\theta)\right], \end{aligned} \tag{4.33}$$

where $G_{1,n}$ and $G_{2,n}$ are defined as in (2.32). Thus

$$\begin{aligned} &\left(F_n^*\Delta_n(f_\theta^{-1})\left[D_n(f)-D_{\infty,n}(f)\right]\right)_{s,t} \\ &\quad=\sum_{\tau\leqslant 0}[\left\{\phi_{t,n}(\tau;f)-\phi_t(\tau;f)\right\}G_{1,n}(s,\tau;f_\theta) \\ &\qquad+\left\{\phi_{n+1-t,n}(\tau;f)-\phi_{n+1-t}(\tau;f)\right\}G_{2,n}(s,\tau;f_\theta)]. \end{aligned} \tag{4.34}$$

To prove Theorem 2.4.3 we bound the above terms.

To simplify notation we only emphasis the coefficients associated with $f_\theta$ and not the coefficients associated with $f$. I.e. we set $\phi_{s,n}(\tau;f)=\phi_{s,n}(\tau)$, $\phi_s(\tau;f)=\phi_s(\tau)$, $\phi_f=\phi$ and $\psi_f=\psi$.

The proof of (2.39) simply follows from the definitions of $D_n(f)$ and $D_{\infty,n}(f)$.

Next we prove (2.40). By using (4.34) we have

$$\left\|F_n^*\Delta_n(f_\theta^{-1})D_n(f) - F_n^*\Delta_n(f_\theta^{-1})D_{\infty,n}(f)\right\|_1 \leqslant T_{1,n} + T_{2,n},$$

where

$$\begin{aligned} T_{1,n} &= \sum_{s,t=1}^{n}\sum_{\tau=-\infty}^{0} |\phi_{s,n}(\tau) - \phi_s(\tau)||G_{1,n}(t,\tau;f_\theta)| \\ T_{2,n} &= \sum_{s,t=1}^{n}\sum_{\tau=-\infty}^{0} |\phi_{n+1-s,n}(\tau) - \phi_{n+1-s}(\tau)||G_{2,n}(t,\tau;f_\theta)|. \end{aligned}$$

We focus on $T_{1,n}$, noting that the method for bounding $T_{2,n}$ is similar. Exchanging the summands we have

$$T_{1,n} \leqslant \sum_{\tau=-\infty}^{0}\sum_{t=1}^{n}|G_{1,n}(t,\tau;f_\theta)|\sum_{s=1}^{n}|\phi_{s,n}(\tau) - \phi_s(\tau)|.$$

To bound $\sum_{s=1}^{n}|\phi_{s,n}(\tau) - \phi_s(\tau)|$ we require the generalized Baxter's inequality stated in Lemma A.1.1. Substituting the bound in Lemma A.1.1 into the above (and for a sufficiently large $n$) we have

$$T_{1,n} \leqslant C_{f,0}\sum_{\tau=-\infty}^{0}\sum_{t=1}^{n}|G_{1,n}(t,\tau;f_\theta)|\sum_{s=n+1}^{\infty}|\phi_s(\tau)|.$$

Using that $G_{1,n}(t,\tau) = \sum_{a\in\mathbb{Z}} K_{f_\theta^{-1}}(\tau - t + an)$ we have the bound

$$\begin{aligned} T_{1,n} &\leqslant C_{f,0}\sum_{\tau=-\infty}^{0}\sum_{t=1}^{n}\sum_{a\in\mathbb{Z}}|K_{f_\theta^{-1}}(t-\tau+an)|\sum_{s=n+1}^{\infty}|\phi_s(\tau)| \\ &= C_{f,0}\sum_{r\in\mathbb{Z}}|K_{f_\theta^{-1}}(r)|\sum_{\tau=-\infty}^{0}\sum_{s=n+1}^{\infty}|\phi_s(\tau)|. \end{aligned}$$

Therefore,

$$\begin{aligned}
T_{1,n} &\leqslant C_{f,0}\sum_{r\in\mathbb{Z}}|K_{f_\theta^{-1}}(r)|\sum_{\tau=-\infty}^{0}\sum_{s=n+1}^{\infty}|\phi_s(\tau)| \\
&\leqslant C_{f,0}\sum_{r\in\mathbb{Z}}|K_{f_\theta^{-1}}(r)|\sum_{\tau=-\infty}^{0}\sum_{s=n+1}^{\infty}\sum_{j=0}^{\infty}|\phi_{s+j}||\psi_{-\tau-j}| \quad (\text{use } \phi_s(\tau)=\sum_{j=0}^{\infty}\phi_{s+j}\psi_{|\tau|-j}) \\
&= C_{f,0}\sum_{r\in\mathbb{Z}}|K_{f_\theta^{-1}}(r)|\sum_{\tau=0}^{\infty}|\psi_{\tau-j}|\sum_{s=n+1}^{\infty}\sum_{j=0}^{\infty}|\phi_{s+j}| \qquad (\text{change limits of } \sum_{\tau}) \\
&\leqslant C_{f,0}\sum_{r\in\mathbb{Z}}|K_{f_\theta^{-1}}(r)|\sum_{\ell}|\psi_\ell|\sum_{u=n+1}^{\infty}|u\phi_u| \quad (\text{change of variables } u=s+j).
\end{aligned}$$

Next we use Assumption 2.3.1(i) to give

$$\begin{aligned}
T_{1,n} &\leqslant C_{f,0}\sum_{r\in\mathbb{Z}}|K_{f_\theta^{-1}}(r)|\sum_{\ell}|\psi_\ell|\sum_{s=n+1}^{\infty}\frac{s^K}{s^{K-1}}|\phi_s| \\
&\leqslant \frac{C_{f,0}}{n^{K-1}}\sum_{r\in\mathbb{Z}}|K_{f_\theta^{-1}}(r)|\sum_{\ell}|\psi_\ell|\sum_{s=n+1}^{\infty}|s^K\phi_s| \\
&\leqslant \frac{C_{f,0}}{n^{K-1}}\rho_{n,K}(f)\|\psi\|_0\|\phi\|_K\sum_{r\in\mathbb{Z}}|K_{f_\theta^{-1}}(r)|.
\end{aligned}$$

We note that the inverse covariance $K_{f_\theta^{-1}}(r) = \int_0^{2\pi} f_\theta^{-1}(\omega)e^{ir\omega}d\omega = \sigma_{f_\theta}^{-2}\int_0^{2\pi}|\phi_{f_\theta}(\omega)|^2e^{ir\omega}d\omega = \sigma_{f_\theta}^{-2}\sum_j \phi_j(f_\theta)\phi_{j+r}(f_\theta)$. Therefore

$$\sum_{r=-\infty}^{\infty}|K_{f_\theta}(r)| \leqslant \sigma_{f_\theta}^{-2}\|\phi_{f_\theta}\|_0^2. \tag{4.35}$$

Substituting this into the above yields the bound

$$T_{1,n} \leqslant \frac{C_{f,0}}{\sigma_{f_\theta}^2 n^{K-1}}\rho_{n,K}(f)\|\psi\|_0\|\phi_{f_\theta}\|_0^2\|\phi\|_K.$$

The same bound holds for $T_{2,n}$. Together the bounds for $T_{1,n}$ and $T_{2,n}$ give

$$\left\|F_n^*\Delta_n(f_\theta^{-1})D_n(f_\theta) - F_n^*\Delta_n(f_\theta^{-1})D_{\infty,n}(f_\theta)\right\|_1 \leqslant \frac{2C_{f,0}}{\sigma_{f_\theta}^2 n^{K-1}}\rho_{n,K}(f)\|\psi\|_0\|\phi_{f_\theta}\|_0^2\|\phi\|_K.$$

Replacing $\|\psi_f\|_0 = \|\psi\|_0$ and $\|\phi_f\|_K = \|\phi\|_K$, this proves (2.40).

To prove (2.41) we recall

$$\|X_t X_s\|_{\mathbb{E},q} = (\mathbb{E}|X_t X_s|^q)^{1/q} \leqslant \left(\mathbb{E}|X_t|^{2q}\right)^{1/2q} \left(\mathbb{E}|X_s|^{2q}\right)^{1/2q} \leqslant \|X\|^2_{\mathbb{E},2q}.$$

Therefore,

$$\begin{aligned}
&n^{-1} \left\| \underline{X}'_n F_n^* \Delta_n(f_\theta^{-1}) \left(D_n(f) - D_{\infty,n}(f)\right) \underline{X}_n \right\|_{\mathbb{E},q} \\
&\quad \leqslant n^{-1} \sum_{s,t=1}^{n} \left| \left(F_n^* \Delta_n(f_\theta^{-1}) \left(D_n(f) - D_{\infty,n}(f)\right)\right)_{s,t} \right| \|X_t X_s\|_{\mathbb{E},q} \\
&\quad \leqslant n^{-1} \left\| F_n^* \Delta_n(f_\theta^{-1}) \left(D_n(f) - D_{\infty,n}(f)\right) \right\|_1 \|X\|^2_{\mathbb{E},2q} \\
&\quad \leqslant \frac{2C_{f,0}}{\sigma^2_{f_\theta} n^K} \rho_{n,K}(f) \|\psi_f\|_0 \|\phi_{f_\theta}\|_0^2 \|\phi_f\|_K \|X\|^2_{\mathbb{E},2q},
\end{aligned}$$

where the last line follows from the inequality in (2.40). This proves (2.41). □

PROOF of Theorem 2.4.4 For notational simplicity, we omit the parameter dependence on $f_\theta$. We first prove (2.43). We observe that

$$\begin{aligned}
\left\| F_n^* \Delta_n(f_\theta^{-1}) D_{\infty,n}(f_\theta) \right\|_1 &\leqslant \sum_{s,t=1}^{n} \sum_{\tau=-\infty}^{0} \left(|\phi_s(\tau)||G_{1,n}(t,\tau)| + |\phi_{n+1-s}(\tau)||G_{2,n}(t,\tau)|\right) \\
&= S_{1,n} + S_{2,n}.
\end{aligned}$$

As in the proof of Theorem 2.4.3, we bound each term separately. Using a similar set of bounds to those used in the proof of Theorem 2.4.3 we have

$$\begin{aligned}
S_{1,n} &\leqslant \sum_{r\in\mathbb{Z}} |K_{f_\theta^{-1}}(r)| \sum_{\ell} |\psi_\ell| \sum_{s=1}^{n} \sum_{j=0}^{\infty} |\phi_{s+j}| \\
&\leqslant \sum_{r\in\mathbb{Z}} |K_{f_\theta^{-1}}(r)| \sum_{\ell} |\psi_\ell| \sum_{s=1}^{\infty} |s\phi_s| \leqslant \frac{1}{\sigma^2_{f_\theta}} \|\psi_{f_\theta}\|_0 \|\phi_{f_\theta}\|_0^2 \|\phi_{f_\theta}\|_1,
\end{aligned}$$

where the bound $\sum_{r\in\mathbb{Z}} |K_{f_\theta^{-1}}(r)| \leqslant \sigma_{f_\theta}^{-2} \|\phi_{f_\theta}\|_0^2$ follows from (4.35). Using a similar method we

obtain the bound $S_{2,n} \leqslant \sigma_{f_\theta}^{-2}\|\psi_{f_\theta}\|_0\|\phi_{f_\theta}\|_0^2\|\phi_{f_\theta}\|_1$. Altogether the bounds for $S_{1,n}$ and $S_{2,n}$ give

$$\left\|F_n^*\Delta_n(f_\theta^{-1})D_{\infty,n}(f_\theta)\right\|_1 \leqslant \frac{2}{\sigma_{f_\theta}^2}\|\psi_{f_\theta}\|_0\|\phi_{f_\theta}\|_0^2\|\phi_{f_\theta}\|_1,$$

this proves (2.43).

The proof of (2.44) uses the triangle inequality

$$\begin{aligned}\left\|\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1})\right\|_1 &= \left\|F_n^*\Delta_n(f_\theta^{-1})D_n(f_\theta)\right\|_1 \\ &\leqslant \left\|F_n^*\Delta_n(f_\theta^{-1})\left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right)\right\|_1 \\ &\quad + \left\|F_n^*\Delta_n(f_\theta^{-1})D_{\infty,n}(f_\theta)\right\|_1.\end{aligned}$$

Substituting the bound Theorem 2.4.3 (equation (2.40)) and (2.43) into the above gives (2.44).

The proof of (2.45) uses the bound in (2.44) together with similar arguments to those in the proof of Theorem 2.4.3, we omit the details. □

### 4.7.2 Proof of Section 4

PROOF of Lemma 4.1.1 The proof is similar to the proof of Theorem 2.4.3, but with some subtle differences. Rather than bounding the best finite predictors with the best infinite predictors, we bound the best infinite predictors with the plug-in estimators based on the best fitting AR($p$) parameters. For example, the bounds use the regular Baxter's inequality rather than the generalized Baxter's inequality.

We first prove (4.4). By using the triangular inequality we have

$$\begin{aligned}&\left\|F_n^*\Delta_n(f_\theta^{-1})\left(D_n(f) - D_n(f_p)\right)\right\|_1 \\ \leqslant\ &\left\|F_n^*\Delta_n(f_\theta^{-1})\left(D_n(f) - D_{\infty,n}(f)\right)\right\|_1 + \left\|F_n^*\Delta_n(f_\theta^{-1})\left(D_{\infty,n}(f) - D_n(f_p)\right)\right\|_1 \\ &\leqslant \frac{C_{f,0}\rho_{n,K}(f)}{n^{K-1}}A_K(f,f_\theta) + \left\|F_n^*\Delta_n(f_\theta^{-1})\left(D_{\infty,n}(f) - D_n(f_p)\right)\right\|_1, \quad (4.36)\end{aligned}$$

where the first term of the right hand side of the above follows from (2.40). Now we bound the

second term on the right hand side of the above. We observe that since the AR($p$) process only uses the first and last $p$ observations for the predictions that $D_n(f_p) = D_{\infty,n}(f_p)$, thus we can write the second term as

$$F_n^* \Delta_n(f_\theta^{-1}) \left( D_{\infty,n}(f) - D_n(f_p) \right) = F_n^* \Delta_n(f_\theta^{-1}) \left( D_{\infty,n}(f) - D_{\infty,n}(f_p) \right).$$

Recall that $\{a_j(p)\}_{j=1}^p$ are the best fitting AR($p$) parameters based on the autocovariance function associated with the spectral density $f$. Let $a_p(\omega) = 1 - \sum_{s=1}^p a_s(p) e^{-is\omega}$, $a_{j,p}^\infty(\omega) = 1 - \sum_{s=1}^{p-j} a_{s+j}(p) e^{-is\omega}$ and $a_p(\omega)^{-1} = \psi_p(\omega) = \sum_{j=0}^\infty \psi_{j,p} e^{-ij\omega}$. By using the expression for $D_{\infty,n}(f)$ given in (2.37) we have

$$\left[ F_n^* \Delta_n(f_\theta^{-1}) \left( D_{\infty,n}(f) - D_{\infty,n}(f_p) \right) \right]_{t,j} \quad = \quad U_{1,n}^{j,t} + U_{2,n}^{j,t}$$

where

$$\begin{aligned} U_{1,n}^{j,t} &= \frac{1}{n} \sum_{k=1}^n \frac{e^{-it\omega_{k,n}}}{f_\theta(\omega_{k,n})} \left( \frac{\phi_j^\infty(\omega_{k,n})}{\phi(\omega_{k,n})} - \frac{a_{j,p}^\infty(\omega_{k,n})}{a_p(\omega_{k,n})} \right) \\ U_{2,n}^{j,t} &= \frac{1}{n} \sum_{k=1}^n \frac{e^{-i(t-1)\omega_{k,n}}}{f_\theta(\omega_{k,n})} \left( \frac{\overline{\phi_{n+1-j}^\infty(\omega_{k,n})}}{\overline{\phi(\omega_{k,n})}} - \frac{\overline{a_{n+1-j,p}^\infty(\omega_{k,n})}}{\overline{a_p(\omega_{k,n})}} \right). \end{aligned}$$

We focus on $U_{1,n}^{j,t}$, and partition it into two terms $U_{1,n}^{j,t} = U_{1,n,1}^{j,t} + U_{1,n,2}^{j,t}$, where

$$U_{1,n,1}^{j,t} = \frac{1}{n} \sum_{k=1}^n \frac{e^{-it\omega_{k,n}}}{\phi(\omega_{k,n}) f_\theta(\omega_{k,n})} \left( \phi_j^\infty(\omega_{k,n}) - a_{j,p}^\infty(\omega_{k,n}) \right)$$

and

$$\begin{aligned} U_{1,n,2}^{j,t} &= \frac{1}{n} \sum_{k=1}^n \frac{e^{-it\omega_{k,n}} a_{j,p}^\infty(\omega_{k,n})}{f_\theta(\omega_{k,n})} \left( \phi(\omega_{k,n})^{-1} - a_p(\omega_{k,n})^{-1} \right) \\ &= \frac{1}{n} \sum_{k=1}^n \frac{e^{-it\omega_{k,n}} a_{j,p}^\infty(\omega_{k,n})}{f_\theta(\omega_{k,n})} \left( \psi(\omega_{k,n}) - \psi_p(\omega_{k,n}) \right). \end{aligned}$$

We first consider $U_{1,n,1}^{j,t}$. We observe $\phi(\omega_{k,n})^{-1} = \psi(\omega_{k,n}) = \sum_{\ell=0}^{\infty} \psi_\ell e^{-i\ell\omega_{k,n}}$. Substituting this into $U_{1,n,1}^{j,t}$ gives

$$\begin{aligned} U_{1,n,1}^{j,t} &= \sum_{s=0}^{\infty} (\phi_{j+s} - a_{j+s}(p)) \frac{1}{n} \sum_{k=1}^{n} \frac{e^{-i(t+s)\omega_{k,n}}}{\phi(\omega_{k,n}) f_\theta(\omega_{k,n})} \\ &= \sum_{s=0}^{\infty} (\phi_{j+s} - a_{j+s}(p)) \sum_{\ell=0}^{\infty} \psi_\ell \frac{1}{n} \sum_{k=1}^{n} f_\theta(\omega_{k,n})^{-1} e^{-i(t+\ell+s)\omega_{k,n}} \\ &= \sum_{s=0}^{\infty} (\phi_{j+s} - a_{j+s}(p)) \sum_{\ell=0}^{\infty} \psi_\ell \sum_{r\in\mathbb{Z}} K_{f_\theta^{-1}}(t+\ell+s+rn), \end{aligned}$$

where $K_{f_\theta^{-1}}(r) = \int_0^{2\pi} f_\theta(\omega)^{-1} e^{ir\omega} d\omega$. Therefore, the absolute sum of the above gives

$$\begin{aligned} \sum_{j,t=1}^{n} |U_{1,n,1}^{j,t}| &\leqslant \sum_{j,t=1}^{n} \sum_{s=0}^{\infty} |\phi_{j+s} - a_{j+s}(p)| \sum_{\ell=0}^{\infty} |\psi_\ell| \sum_{r\in\mathbb{Z}} |K_{f_\theta^{-1}}(t+\ell+s+rn)| \\ &= \sum_{j=1}^{n} \sum_{s=0}^{\infty} |\phi_{j+s} - a_{j+s}(p)| \sum_{\ell=0}^{\infty} |\psi_\ell| \sum_{t=1}^{n} \sum_{r\in\mathbb{Z}} |K_{f_\theta^{-1}}(t+\ell+s+rn)| \\ &\leqslant \left( \sum_{j=1}^{n} \sum_{s=0}^{\infty} |\phi_{j+s} - a_{j+s}(p)| \right) \|\psi_f\|_0 \sum_{\tau\in\mathbb{Z}} |K_{f_\theta^{-1}}(\tau)| \\ &\leqslant \left( \sum_{s=1}^{\infty} s|\phi_s - a_s(p)| \right) \|\psi_f\|_0 \sum_{\tau\in\mathbb{Z}} |K_{f_\theta^{-1}}(\tau)|. \end{aligned}$$

By using (4.35) we have $\sum_{\tau\in\mathbb{Z}} |K_{f_\theta^{-1}}(\tau)| \leqslant \sigma_{f_\theta}^{-2} \|\phi_{f_\theta}\|_0^2$. Further, by using the regular Baxter inequality we have

$$\sum_{s=1}^{\infty} s|\phi_s - a_s(p)| \leqslant (1 + C_{f,1}) \sum_{s=p+1}^{\infty} s|\phi_s| \leqslant (1 + C_{f,1}) p^{-K+1} \rho_{p,K}(f) \|\phi_f\|_K.$$

Substituting these two bounds into $\sum_{j,t=1}^{n} |U_{1,n,1}^{j,t}|$ yields

$$\sum_{j,t=1}^{n} |U_{1,n,1}^{j,t}| \leqslant \frac{(1 + C_{f,1})}{\sigma_{f_\theta}^2 p^{K-1}} \rho_{p,K}(f) \|\phi_f\|_K \|\psi_f\|_0 \|\phi_{f_\theta}\|_0^2.$$

Next we consider the second term $U_{1,n,2}^{j,t}$. Using that $\psi(\omega_{k,n}) = \sum_{s=0}^{\infty} \psi_s e^{-is\omega}$ and $\psi_p(\omega_{k,n}) =$

$\sum_{s=0}^{\infty} \psi_{s,p} e^{-is\omega}$ we have

$$\begin{aligned} U_{1,n,2}^{j,t} &= \frac{1}{n}\sum_{k=1}^{n} \frac{e^{-it\omega_{k,n}} a_{j,p}^{\infty}(\omega_{k,n})}{f_\theta(\omega_{k,n})} \left(\psi(\omega_{k,n}) - \psi_p(\omega_{k,n})\right) \\ &= \sum_{s=0}^{\infty} (\psi_s - \psi_{s,p}) \frac{1}{n}\sum_{k=1}^{n} \frac{e^{-i(t+s)\omega_{k,n}} a_{j,p}^{\infty}(\omega_{k,n})}{f_\theta(\omega_{k,n})} \\ &= \sum_{s=0}^{\infty} (\psi_s - \psi_{s,p}) \sum_{\ell=0}^{\infty} a_{j+\ell}(p) \frac{1}{n}\sum_{k=1}^{n} \frac{e^{-i(t+s+\ell)\omega_{k,n}}}{f_\theta(\omega_{k,n})} \\ &= \sum_{s=0}^{\infty} (\psi_s - \psi_{s,p}) \sum_{\ell=0}^{\infty} a_{j+\ell}(p) \sum_{r\in\mathbb{Z}} K_{f_\theta^{-1}}(t+s+\ell+rn). \end{aligned}$$

Taking the absolute sum of the above gives

$$\begin{aligned} \sum_{j,t=1}^{n} |U_{1,n,2}^{j,t}| &\leqslant \sum_{j,t=1}^{n}\sum_{s=0}^{\infty} |\psi_s - \psi_{s,p}| \sum_{\ell=0}^{\infty} |a_{j+\ell}(p)| \sum_{r\in\mathbb{Z}} |K_{f_\theta^{-1}}(t+s+\ell+rn)| \\ &= \sum_{s=0}^{\infty} |\psi_s - \psi_{s,p}| \sum_{j=1}^{n}\sum_{\ell=0}^{\infty} |a_{j+\ell}(p)| \sum_{r\in\mathbb{Z}} |K_{f_\theta^{-1}}(r)| \quad \text{(apply the bound (4.35))} \\ &\leqslant \sigma_{f_\theta}^{-2} \|\phi_{f_\theta}\|_0^2 \left(\sum_{s=0}^{\infty} |\psi_s - \psi_{s,p}|\right) \sum_{u=0}^{\infty} |u a_u(p)| \\ &\leqslant \sigma_{f_\theta}^{-2} \|\phi_{f_\theta}\|_0^2 \|a_p\|_1 \sum_{s=0}^{\infty} |\psi_s - \psi_{s,p}|. \end{aligned}$$

Next we bound $\|a_p\|_1$ and $\sum_{s=0}^{\infty} |\psi_s - \psi_{s,p}|$. Let $\phi_p(\omega) = 1 - \sum_{j=1}^{p} \phi_j e^{ij\omega}$ (the truncated AR($\infty$) process). Then by applying Baxter's inequality, it is straightforward to show that

$$\|a_p\|_1 \leqslant \|\phi_p\|_1 + \|a_p - \phi_p\|_1 \leqslant (C_{f,1} + 1)\|\phi_f\|_1. \tag{4.37}$$

To bound $\sum_{s=0}^{\infty} |\psi_s - \psi_{s,p}|$ we use the inequality in Kreiss et al. (2011), page 2126

$$\sum_{s=0}^{\infty} |\psi_s - \psi_{s,p}| \leqslant \frac{\|\psi_f\|_0^2 \cdot \sum_{j=1}^{\infty} |\phi_j - a_j(p)|}{1 - \|\psi_f\| \cdot \|a_p - \phi\|_0}.$$

Applying Baxter's inequality to the numerator of the above gives

$$\sum_{s=0}^{\infty}|\psi_s-\psi_{s,p}|\leqslant\frac{\|\psi_f\|_0^2(C_{f,0}+1)\rho_{p,K}(f)\|\phi_f\|_K}{p^K(1-\|\psi_f\|_0\cdot\|a_p-\phi\|_0)} \tag{4.38}$$

Substituting the bound in (4.37) and (4.38) into $\sum_{j,t=1}^{n}|U_{1,n,2}^{j,t}|$ gives

$$\sum_{j,t=1}^{n}|U_{1,n,2}^{j,t}|\leqslant\frac{(C_{f,1}+1)^2}{\sigma_{f_\theta}^2p^K}\cdot\frac{\|\psi_f\|_0^2\|\phi_f\|_1\|\phi_f\|_K\|\phi_{f_\theta}\|_0^2\rho_{p,K}(f)}{1-\|\psi_f\|_0\|a_p-\phi\|_0}$$

Altogether, for sufficiently large $p$, where $\|\psi_f\|_0\cdot\|a_p-\phi\|_0\leqslant 1/2$ we have

$$\begin{aligned}\sum_{t,j=1}^{n}|U_{1,n}^{j,t}| &\leqslant \frac{(1+C_{f,1})}{\sigma_{f_\theta}^2p^{K-1}}\rho_{p,K}(f)\|\phi_f\|_K\|\psi_f\|_0\|\phi_{f_\theta}\|_0^2\\ &\quad+\frac{2(C_{f,1}+1)^2}{\sigma_{f_\theta}^2p^K}\|\psi_f\|_0^2\|\phi_f\|_1\|\phi_f\|_K\|\phi_{f_\theta}\|_0^2\rho_{p,K}(f)\\ &\leqslant \frac{(C_{f,1}+1)}{\sigma_{f_\theta}^2p^{K-1}}\rho_{p,K}(f)\|\phi_f\|_K\|\phi_{f_\theta}\|_0^2\|\psi_f\|_0\left(1+\frac{2(1+C_{f,1})}{p}\|\psi_f\|_0\|\phi_f\|_1\right)\end{aligned}$$

The same bound holds for $\sum_{t,j=1}^{n}|U_{2,n}^{j,t}|$, thus using (4.36) and $\rho_{n,K}(f)\leqslant\rho_{p,K}(f)$ gives

$$\begin{aligned}&\left\|F_n^*\Delta_n(f_\theta^{-1})\left(D_{\infty,n}(f)-D_n(f_p)\right)\right\|_1\\ &\quad\leqslant\rho_{p,K}(f)A_K(f,f_\theta)\left(\frac{(C_{f,1}+1)}{p^{K-1}}+\frac{2(C_{f,1}+1)^2}{p^K}\|\psi_f\|_0\|\phi_f\|_1\right).\end{aligned}$$

Substituting the above into (4.36) gives (4.4).

The proof of (4.5) is similar to the proof of Theorem 2.4.3, we omit the details. □

PROOF of Lemma 4.3.1 We first prove the result in the case that $s=0$ and for $\widehat{W}_{p,n}(\cdot)$. In this case

$$\widehat{W}_{p,n}(\theta)-W_n(\theta)=\frac{1}{n}\sum_{k=1}^{n}f_\theta(\omega_{k,n})^{-1}\left[\widehat{J}_n(\omega_{k,n};\widehat{f}_p)-\widehat{J}_n(\omega_{k,n};f)\right]\overline{J_n(\omega_{k,n})}.$$

Thus

$$\sup_{\theta\in\Theta}|\widehat{W}_{p,n}(\theta)-W_n(\theta)| \leqslant \sup_{\theta,\omega} f_\theta(\omega)^{-1}\times\frac{1}{n}\sum_{k=1}^{n}\left|\left[\widehat{J}_n(\omega_{k,n};\widehat{f}_p)-\widehat{J}_n(\omega_{k,n};f)\right]\overline{J_n(\omega_{k,n})}\right|.$$

By using Theorem 2.3.4 (under Assumption 4.2.2) we have

$$\left[\widehat{J}_n(\omega;\widehat{f}_p)-\widehat{J}_n(\omega;f)\right]\overline{J_n(\omega)}=\Delta(\omega)+O_p\left(\frac{p^3}{n^{3/2}}\right)$$

where $\Delta(\omega)$ is defined in Theorem 2.3.4, $O_p\left(p^3n^{-3/2}\right)$ bound is uniform all frequencies, $\sup_\omega \mathbb{E}[\Delta(\omega)]=O((np^{K-1})^{-1}+p^3/n^2)$ and $\sup_\omega \mathrm{var}[\Delta(\omega)]=O(p^4/n^2)$. Thus using this we have

$$\begin{aligned}\sup_{\theta\in\Theta}|\widehat{W}_{p,n}(\theta)-W_n(\theta)| &= \sup_{\theta,\omega} f_\theta(\omega)^{-1}\times\frac{1}{n}\sum_{k=1}^{n}|\Delta(\omega_{k,n})|+O_p\left(\frac{p^3}{n^{3/2}}\right)\\ &= O_p\left(\frac{p^2}{n}+\frac{p^3}{n^{3/2}}\right)=O_p\left(\frac{p^2}{n}\right).\end{aligned}$$

This proves the result for $s=0$. A similar argument applies for the derivatives of $\widehat{W}_{p,n}(\theta)$ (together with Assumption 4.2.1(iii)) and $\widehat{H}_{p,n}(\theta)$, we omit the details. □

PROOF of Lemma 4.3.2 We start with the infeasible criterion $W_n(\theta)$. Let $\mathbb{E}[W_n(\theta)]=\mathcal{W}_n(\theta)$. We first show the uniformly convergence of $W_n(\theta)$, i.e.,

$$\sup_{\theta\in\Theta}|W_n(\theta)-\mathcal{W}_n(\theta)|\xrightarrow{\mathcal{P}}0. \tag{4.39}$$

Using Das et al. (2021), Theorem A.1 and the classical result $\mathrm{var}[K_n(\theta)]=O(n^{-1})$ we have

$$\begin{aligned}\mathrm{var}[W_n(\theta)] &= \mathrm{var}\left(K_n(\theta)+n^{-1}\sum_{k=1}^{n}\frac{\widehat{J}_n(\omega_{k,n};f)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})}\right)\\ &\leqslant 2\mathrm{var}[K_n(\theta)]+\frac{2}{n}\sum_{k=1}^{n}\mathrm{var}[\widehat{J}_n(\omega_{k,n};f)\overline{J_n(\omega_{k,n})}]/f_\theta(\omega_{k,n})^2\\ &\leqslant 2\mathrm{var}[K_n(\theta)]+O(n^{-2})=O(n^{-1}).\end{aligned}$$

Therefore, by Markov's inequality, $W_n(\theta) \xrightarrow{\mathcal{P}} \mathcal{W}_n(\theta)$ for each $\theta \in \Theta$. To show a uniform convergence, since $\Theta$ is compact, it is enough to show that $\{W_n(\theta); \theta \in \Theta\}$ is equicontinuous in probability. For arbitrary $\theta_1, \theta_2 \in \Theta$,

$$\begin{aligned} W_n(\theta_1) - W_n(\theta_2) &= n^{-1} \sum_{k=1}^{n} (f_{\theta_1}^{-1}(\omega_{k,n}) - f_{\theta_2}^{-1}(\omega_{k,n})) \widetilde{J}_n(\omega_{k,n}; f) \overline{J_n(\omega_{k,n})} \\ &+ n^{-1} \sum_{k1}^{n} (\log f_{\theta_1}(\omega_{k,n}) - \log f_{\theta_2}(\omega_{k,n})) = I_1(\theta_1, \theta_2) + I_2(\theta_1, \theta_2). \end{aligned}$$

To (uniformly) bound $I_1(\theta_1, \theta_2)$, we use the mean value theorem

$$\begin{aligned} I_1(\theta_1, \theta_2) &= n^{-1} \sum_{k=1}^{n} \left( f_{\theta_1}^{-1}(\omega_{k,n}) - f_{\theta_2}^{-1}(\omega_{k,n}) \right) \widetilde{J}_n(\omega_{k,n}; f) \overline{J_n(\omega_{k,n})} \\ &= n^{-1} \sum_{k=1}^{n} \nabla_\theta f_\theta^{-1}(\omega_{k,n}) \big]'_{\theta=\overline{\theta}_k} (\theta_1 - \theta_2) \widetilde{J}_n(\omega_{k,n}; f) \overline{J_n(\omega_{k,n})} \\ &= \mathcal{K}_n(\overline{\theta})'(\theta_1 - \theta_2), \end{aligned}$$

where $\mathcal{K}_n(\overline{\theta}) = n^{-1} \sum_{k=1}^{n} \widetilde{J}_n(\omega_{k,n}; f) \overline{J_n(\omega_{k,n})} \nabla_\theta f_\theta^{-1}(\omega_{k,n}) \big]_{\theta=\overline{\theta}_k}$ and $\overline{\theta}_1, ..., \overline{\theta}_n$ are convex combinations of $\theta_1$ and $\theta_2$. It is clear that

$$\|\mathcal{K}_n(\overline{\theta})\|_1 \leqslant \sup_{\theta,\omega} \|\nabla_\theta f_\theta^{-1}(\omega)\|_1 \frac{1}{n} \sum_{k=1}^{n} \left| \widetilde{J}_n(\omega_{k,n}; f) \overline{J_n(\omega_{k,n})} \right| = K_n.$$

Thus

$$|I_1(\theta_1, \theta_2)| \leqslant K_n |\theta_1 - \theta_2|_1. \tag{4.40}$$

We need to show that $K_n = O_p(1)$ (it is enough to show that $\sup_n \mathbb{E}[K_n] < \infty$). To show this, we

use the classical results on DFT

$$\begin{aligned}\mathbb{E}\left|\widetilde{J}_n(\omega_{k,n};f)\overline{J_n(\omega_{k,n})}\right| &\leqslant \mathbb{E}|J_n(\omega_{k,n})|^2 + \mathbb{E}\left|\widehat{J}_n(\omega_{k,n};f)\overline{J_n(\omega_{k,n})}\right| \\ &\leqslant f(\omega_{k,n}) + \mathrm{var}(\widehat{J}_n(\omega_{k,n};f))^{1/2}\mathrm{var}(J_n(\omega_{k,n}))^{1/2} \\ &\leqslant f(\omega_{k,n})(1+O(n^{-1})).\end{aligned}$$

Using above and Assumption 4.2.1(iii-a) gives

$$\sup_n \mathbb{E}[K_n] \leqslant \sup_{\theta,\omega} \|\nabla_\theta f_\theta^{-1}(\omega)\|_1 \cdot \sup_n \frac{1}{n}\sum_{k=1}^{n} f(\omega_{k,n})(1+O(n^{-1})) < \infty.$$

Therefore, $K_n = O_p(1)$ and from (4.40), $I_1(\theta_1,\theta_2)$ is equicontinuous in probability. Using similar argument, we can show that $I_2(\theta_1,\theta_2)$ is equicontinous in probability and thus, $\{W_n(\theta);\theta\in\Theta\}$ is equicontinous in probability. This imples $\sup_{\theta\in\Theta}|W_n(\theta)-\mathcal{W}_n(\theta)| \xrightarrow{\mathcal{P}} 0$, thus we have shown (4.39).

Next, let $\widetilde{\theta}_n^{(W)} = \arg\min_{\theta\in\Theta} W_n(\theta)$. Since $\theta_n = \arg\min_{\theta\in\Theta}\mathcal{W}_n(\theta)$ we have

$$W_n(\widetilde{\theta}_n^{(W)}) - \mathcal{W}_n(\widetilde{\theta}_n^{(W)}) \leqslant W_n(\widetilde{\theta}_n^{(W)}) - \mathcal{W}_n(\theta_n) \leqslant W_n(\theta_n) - \mathcal{W}_n(\theta_n).$$

Thus

$$|W_n(\widetilde{\theta}_n^{(W)}) - \mathcal{W}_n(\theta_n)| \leqslant \sup_\theta |W_n(\theta)-\mathcal{W}_n(\theta)| \xrightarrow{\mathcal{P}} 0.$$

If $\theta_n$ uniquely minimises $I_n(f,f_\theta)$, then by using the above we have that $|\widetilde{\theta}_n^{(W)} - \theta_n|_1 \xrightarrow{\mathcal{P}} 0$. However, $W_n(\theta)$ is an infeasible criterion. To show consistency we need to obtain a uniform bound on the feasible criterion $\widehat{W}_{p,n}(\theta)$. That is

$$\sup_\theta |\widehat{W}_{p,n}(\theta) - \mathcal{W}_n(\theta)| \xrightarrow{\mathcal{P}} 0. \tag{4.41}$$

Now by using the triangular inequality, together with (4.39) and Lemma 4.3.1, (4.41) immediately follows. Therefore, by using the same arguments those given above we have $|\widehat{\theta}_n^{(W)} - \theta_n|_1 \xrightarrow{\mathcal{P}} 0$, which is the desired result.

By the same set of arguments we have $|\widehat{\theta}_n^{(H)} - \theta_n|_1 \xrightarrow{\mathcal{P}} 0$. □

PROOF of Lemma 4.3.3 By using Theorem 3.2.1 we have for $i = 1$ and 2

$$\frac{d^i \widehat{W}_{p,n}(\theta)}{d\theta^i}\rfloor_{\theta=\theta_n} = \frac{d^i W_n(\theta)}{d\theta^i}\rfloor_{\theta=\theta_n} + O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right),$$

this immediately gives (4.20). Let $\overline{\theta}_n$ denote a convex combination of $\theta_n$ and $\widehat{\theta}_n^{(W)}$ (note that $\widehat{\theta}_n^{(W)}$ is a consistent estimator of $\theta_n$). To evaluate $\frac{d^3\widehat{W}_n(\theta)}{d\theta^3}$ at the (consistent) estimator $\overline{\theta}_n$, a slightly different approach is required (due to the additional random parameter $\overline{\theta}_n$). By using triangular inequality and Lemma 4.3.1 we have

$$\begin{aligned}
&\left|\frac{d^3 \widehat{W}_{p,n}(\theta)}{d\theta^3}\rfloor_{\theta=\overline{\theta}_n} - \frac{d^3 W_n(\theta)}{d\theta^3}\rfloor_{\theta=\theta_n}\right| \\
&\leqslant \left|\frac{d^3 \widehat{W}_{p,n}(\theta)}{d\theta^3}\rfloor_{\theta=\overline{\theta}_n} - \frac{d^3 W_n(\theta)}{d\theta^3}\rfloor_{\theta=\overline{\theta}_n}\right| \\
&\quad + \frac{1}{n}\sum_{k=1}^{n}\left|\frac{d^3}{d\theta^3}\left[f_{\overline{\theta}_n}(\omega_{k,n})^{-1} - f_{\theta_n}(\omega_{k,n})^{-1}\right]\right|\left|\widetilde{J}_n(\omega_{k,n};f)\overline{J_n(\omega_{k,n})}\right| \\
&= O_p\left(\frac{p^2}{n}\right) + \frac{1}{n}\sum_{k=1}^{n}\left|\frac{d^3}{d\theta^3}\left[f_{\overline{\theta}_n}(\omega_{k,n})^{-1} - f_{\theta_n}(\omega_{k,n})^{-1}\right]\right|\left|\widetilde{J}_n(\omega_{k,n};f)\overline{J_n(\omega_{k,n})}\right|.
\end{aligned}$$

For the second term in the above, we apply the mean value theorem to $\frac{d^3}{d\theta^3} f_\theta^{-1}$ to give

$$\left|\frac{d^3}{d\theta^3}(f_{\overline{\theta}_n}^{-1} - f_{\theta_n}^{-1})\right| \leqslant \sup_\theta\left|\frac{d^4}{d\theta^4} f_\theta^{-1}\right| \cdot |\overline{\theta}_n - \theta_n| \leqslant \sup_\theta\left|\frac{d^4}{d\theta^4} f_\theta^{-1}\right| \cdot |\widehat{\theta}_n^{(W)} - \theta_n|,$$

note that to bound the fourth derivation we require Assumption 4.2.1(iii) for $\kappa = 4$. Substituting

this into the previous inequality gives

$$\begin{aligned}
&\left|\frac{d^3\widehat{W}_{p,n}(\theta)}{d\theta^3}\rfloor_{\theta=\overline{\theta}_n} - \frac{d^3W_n(\theta)}{d\theta^3}\rfloor_{\theta=\theta_n}\right| \\
&\leqslant O_p\left(\frac{p^2}{n}\right) + \frac{1}{n}\sum_{k=1}^{n}\left|\frac{d^3}{d\theta^3}\left[f_{\overline{\theta}_n}(\omega_{k,n})^{-1} - f_{\theta_n}(\omega_{k,n})^{-1}\right]\right|\left|\widetilde{J}_n(\omega_{k,n};f)\overline{J_n(\omega_{k,n})}\right| \\
&= O_p\left(\frac{p^2}{n}\right) + |\widehat{\theta}_n^{(W)} - \theta_n|O_p(1).
\end{aligned}$$

The above proves (4.21).

Using (4.20) and (4.21) we now obtain the first and second order expansions in (4.22) and (4.23). In order to prove (4.22), we will show that

$$(\widehat{\theta}_n^{(W)} - \theta_n) \;=\; O_p\left(\frac{1}{n^{1/2}} + \frac{p^3}{n^{3/2}}\right).$$

if $p^2/n \to 0$ we make a second order expansion of $\frac{d\widehat{W}_{p,n}(\widehat{\theta}_n^{(W)})}{d\theta}$ about $\theta_n$ and assuming that $\widehat{\theta}_n^{(W)}$ lies inside the parameter space we have

$$0 = \frac{d\widehat{W}_{p,n}(\widehat{\theta}_n^{(W)})}{d\widehat{\theta}_n^{(W)}} = \frac{d\widehat{W}_{p,n}(\theta_n)}{d\theta_n} + (\widehat{\theta}_n^{(W)} - \theta_n)\frac{d^2\widehat{W}_{p,n}(\theta_n)}{d\theta_n^2} + \frac{1}{2}(\widehat{\theta}_n^{(W)} - \theta_n)^2\frac{d^3\widehat{W}_{p,n}(\overline{\theta}_n)}{d\theta_n^3}$$

where $\overline{\theta}_n$ is a convex combination of $\theta_n$ and $\widehat{\theta}_n^{(W)}$. Now by using (4.20) and (4.21) we can replace in the above $\widehat{W}_{p,n}(\theta_n)$ and its derivatives with $W_n(\theta_n)$ and its derivatives. Therefore,

$$\begin{aligned}
&\frac{dW_n(\theta_n)}{d\theta_n} + (\widehat{\theta}_n^{(W)} - \theta_n)\frac{d^2W_n(\theta_n)}{d\theta_n^2} + \frac{1}{2}(\widehat{\theta}_n^{(W)} - \theta_n)^2\frac{d^3W_n(\theta_n)}{d\theta_n^3} \\
=\;& O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right) + (\widehat{\theta}_n^{(W)} - \theta_n)^2O_p\left(\frac{p^2}{n}\right) + |\widehat{\theta}_n^{(W)} - \theta_n|^3O_p(1). \qquad (4.42)
\end{aligned}$$

Rearranging the above gives

$$(\widehat{\theta}_n^{(W)} - \theta_n) = -\left[\frac{d^2W_n(\theta_n)}{d\theta_n^2}\right]^{-1}\frac{dW_n(\theta_n)}{d\theta_n} - \frac{1}{2}\left[\frac{d^2W_n(\theta_n)}{d\theta_n^2}\right]^{-1}\frac{d^3W_n(\theta_n)}{d\theta_n^3}(\widehat{\theta}_n^{(W)} - \theta_n)^2 + O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right) + (\widehat{\theta}_n^{(W)} - \theta_n)^2 O_p\left(\frac{p^2}{n}\right) + |\widehat{\theta}_n^{(W)} - \theta_n|^3 O_p(1) \quad (4.43)$$

Next we obtain a bound for $\frac{dW_n(\theta_n)}{d\theta_n}$ (to substitute into the above). Since $\mathbb{E}[\frac{dW_n(\theta_n)}{d\theta_n}] = O(n^{-K})$ (from equation (4.15)) and $\mathrm{var}[\frac{dW_n(\theta_n)}{d\theta_n}] = O_p(n^{-1})$ we have $\frac{dW_n(\theta_n)}{d\theta_n} = O_p(n^{-1/2})$. Substituting this into (4.43) gives

$$(\widehat{\theta}_n^{(W)} - \theta_n) = \frac{1}{2}\left[\frac{d^2W_n(\theta_n)}{d\theta_n^2}\right]^{-1}\frac{d^3W_n(\theta_n)}{d\theta_n^3}(\widehat{\theta}_n^{(W)} - \theta_n)^2 + O_p(n^{-1/2}) + O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right) + (\widehat{\theta}_n^{(W)} - \theta_n)^2 O_p\left(\frac{p^2}{n}\right) + |\widehat{\theta}_n^{(W)} - \theta_n|^3 O_p(1).$$

Using that $\left[\frac{d^2W_n(\theta_n)}{d\theta_n^2}\right]^{-1}\frac{d^3W_n(\theta_n)}{d\theta_n^3} = O_p(1)$ and substituting this into the above gives

$$(\widehat{\theta}_n^{(W)} - \theta_n) = O_p\left(\frac{1}{n^{1/2}} + \frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right) + (\widehat{\theta}_n^{(W)} - \theta_n)^2 O_p\left(\frac{p^2}{n} + 1\right) + |\widehat{\theta}_n^{(W)} - \theta_n|^3 O_p(1). \quad (4.44)$$

Thus, from the above and the consistency result in Lemma 4.3.2 ($|\widehat{\theta}_n^{(W)} - \theta_n| = o_p(1)$) we have

$$(\widehat{\theta}_n^{(W)} - \theta_n) = O_p\left(\frac{1}{n^{1/2}} + \frac{p^3}{n^{3/2}}\right). \quad (4.45)$$

We use the above bound to obtain an exact expression for the dominating rate $O_p(n^{-1/2})$. Returning to equation (4.43) and substituting this bound into the quadratic term in (4.43) gives

$$(\widehat{\theta}_n^{(W)} - \theta_n) = -\left[\frac{d^2W_n(\theta_n)}{d\theta_n^2}\right]^{-1}\frac{dW_n(\theta_n)}{d\theta_n} + O_p\left(\frac{1}{n} + \frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).$$

Using that $\frac{d^2W_n(\theta_n)}{d\theta_n^2} = \mathbb{E}\left(\frac{d^2W_n(\theta_n)}{d\theta_n^2}\right) + O_p(n^{-1/2})$ and under Assumption 4.2.2(iii) we have

$$(\widehat{\theta}_n^{(W)} - \theta_n) = -\left[\mathbb{E}\left[\frac{d^2W_n(\theta_n)}{d\theta_n^2}\right]\right]^{-1}\frac{dW_n(\theta_n)}{d\theta_n} + O_p\left(\frac{1}{n} + \frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).$$

This proves (4.22).

To prove (4.23) we return to (4.42). By substituting (4.45) into (4.42) we have

$$\frac{dW_n(\theta)}{d\theta}\rfloor_{\theta=\theta_n} + (\widehat{\theta}_n^{(W)} - \theta_n)\frac{d^2W_n(\theta)}{d\theta^2}\rfloor_{\theta=\theta_n} + \frac{1}{2}(\widehat{\theta}_n^{(W)} - \theta_n)^2\frac{d^3W_n(\theta)}{d\theta^3}\rfloor_{\theta=\theta_n} = O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).$$

This proves (4.23). □

PROOF of Theorem 4.3.1 We first prove the result for the one parameter case when $p \geqslant 1$. By using (4.23) for the feasible estimator $\widehat{\theta}_n^{(W)} = \arg\min \widehat{W}_{p,n}(\theta)$ we have

$$\frac{dW_n(\theta)}{d\theta}\rfloor_{\theta=\theta_n} + (\widehat{\theta}_n^{(W)} - \theta_n)\frac{d^2W_n(\theta)}{d\theta^2}\rfloor_{\theta=\theta_n} + \frac{1}{2}(\widehat{\theta}_n^{(W)} - \theta_n)^2\frac{d^3W_n(\theta)}{d\theta^3}\rfloor_{\theta=\theta_n} = O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).$$

Whereas for the infeasible estimator $\widetilde{\theta}_n^{(W)} = \arg\min \widehat{W}_{p,n}(\theta)$ we have

$$\frac{dW_n(\theta)}{d\theta}\rfloor_{\theta=\theta_n} + (\widetilde{\theta}_n^{(W)} - \theta_n)\frac{d^2W_n(\theta)}{d\theta^2}\rfloor_{\theta=\theta_n} + \frac{1}{2}(\widetilde{\theta}_n^{(W)} - \theta_n)^2\frac{d^3W_n(\theta)}{d\theta^3}\rfloor_{\theta=\theta_n} = O_p\left(\frac{1}{n^{3/2}}\right).$$

Taking differences for the two expansions above we have

$$\begin{aligned}(\widetilde{\theta}_n^{(W)} - \widehat{\theta}_n^{(W)})\frac{d^2W_n(\theta)}{d\theta^2}\rfloor_{\theta=\theta_n} + \frac{1}{2}(\widetilde{\theta}_n^{(W)} - \widehat{\theta}_n^{(W)})\left[(\widetilde{\theta}_n^{(W)} - \theta_n) + (\widehat{\theta}_n^{(W)} - \theta_n)\right]\frac{d^3W_n(\theta)}{d\theta^3}\rfloor_{\theta=\theta_n} \\ = O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).\end{aligned}$$

Now replacing $\frac{d^2W_n(\theta)}{d\theta^2}\rfloor_{\theta=\theta_n}$ with its expectation and using that $|\widetilde{\theta}_n^{(W)} - \theta_n| = o_p(1)$ and $|\widehat{\theta}_n^{(W)} - \theta_n| = o_p(1)$ we have

$$(\widetilde{\theta}_n^{(W)} - \widehat{\theta}_n^{(W)})\mathbb{E}\left(\frac{d^2W_n(\theta)}{d\theta^2}\rfloor_{\theta=\theta_n}\right) + o_p(1)(\widetilde{\theta}_n^{(W)} - \widehat{\theta}_n^{(W)}) = O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).$$

Since $\mathbb{E}\left(\frac{d^2W_n(\theta)}{d\theta^2}\rfloor_{\theta=\theta_n}\right)$ is greater than 0, the above implies

$$(\widetilde{\theta}_n^{(W)} - \widehat{\theta}_n^{(W)}) = O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).$$

Now we prove the result for the case $p = 0$. If $p = 0$, then $\widehat{W}_{p,n}(\theta) = K_n(\theta)$ (the Whittle likelihood). Let

$$\widehat{\theta}_n^{(K)} = \arg\min K_n(\theta) \quad \text{and} \quad \widetilde{\theta}_n^{(W)} = \arg\min W_n(\theta).$$

Our aim is to show that $|\widehat{\theta}_n^{(K)} - \widetilde{\theta}_n^{(W)}| = O_p(n^{-1})$. Note that $W_n(\theta) = K_n(\theta) + C_n(\theta)$, where

$$C_n(\theta) = \frac{1}{n}\sum_{k=1}^{n} \frac{\widehat{J}_n(\omega; f)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})}.$$

Using a Taylor expansion, similar to the above, we have

$$\frac{dK_n(\theta)}{d\theta}\rfloor_{\theta=\theta_n} + (\widehat{\theta}_n^{(K)} - \theta_n)\frac{d^2K_n(\theta)}{d\theta^2}\rfloor_{\theta=\theta_n} + \frac{1}{2}(\widehat{\theta}_n^{(K)} - \theta_n)^2\frac{d^3K_n(\theta)}{d\theta^3}\rfloor_{\theta=\theta_n} = O_p\left(\frac{1}{n^{3/2}}\right)$$

and

$$\frac{dW_n(\theta)}{d\theta}\rfloor_{\theta=\theta_n} + (\widetilde{\theta}_n^{(W)} - \theta_n)\frac{d^2W_n(\theta)}{d\theta^2}\rfloor_{\theta=\theta_n} + \frac{1}{2}(\widetilde{\theta}_n^{(W)} - \theta_n)^2\frac{d^3W_n(\theta)}{d\theta^3}\rfloor_{\theta=\theta_n} = O_p\left(\frac{1}{n^{3/2}}\right).$$

Taking differences of the two expansions

$$\begin{aligned}
&\frac{dC_n(\theta)}{d\theta}\rfloor_{\theta=\theta_n} + (\widehat{\theta}_n^{(K)} - \widetilde{\theta}_n^{(W)})\frac{d^2K_n(\theta)}{d\theta^2}\rfloor_{\theta=\theta_n} - (\widehat{\theta}_n^{(W)} - \theta_n)\frac{d^2C_n(\theta)}{d\theta^2}\rfloor_{\theta=\theta_n} \\
&+\frac{1}{2}(\widehat{\theta}_n^{(K)} - \widetilde{\theta}_n^{(W)})\left[(\widehat{\theta}_n^{(K)} - \theta_n) + (\widetilde{\theta}_n^{(W)} - \theta_n)\right]\frac{d^3K_n(\theta)}{d\theta^3}\rfloor_{\theta=\theta_n} \\
&-\frac{1}{2}(\widetilde{\theta}_n^{(W)} - \theta_n)^2\frac{d^3C_n(\theta)}{d\theta^3}\rfloor_{\theta=\theta_n} = O_p\left(\frac{1}{n^{3/2}}\right). \qquad (4.46)
\end{aligned}$$

To bound the above we use that

$$|\widehat{\theta}_n^{(K)} - \theta_n| = O_p(n^{-1/2}) \text{ and } |\widetilde{\theta}_n^{(W)} - \theta_n| = O_p(n^{-1/2}).$$

In addition by using a proof analogous to the proves of Theorem 2.4.4, equation (2.45) we have

$$\frac{d^s C_n(\theta)}{d\theta^s} = \frac{1}{n}\sum_{k=1}^{n} \widehat{J}_n(\omega_{k,n};f)\overline{J_n(\omega_{k,n})}\frac{d^s}{d\theta^s} f_\theta(\omega_{k,n})^{-1} = O_p(n^{-1}) \quad \text{for} \quad 0 \leqslant s \leqslant 3.$$

Substituting the above bounds into (4.46) gives

$$(\widehat{\theta}_n^{(K)} - \widetilde{\theta}_n^{(W)})\frac{d^2 K_n(\theta)}{d\theta^2}\rfloor_{\theta=\theta_n} + \frac{1}{2}(\widehat{\theta}_n^{(K)} - \widetilde{\theta}_n^{(W)})O_p(n^{-1/2}) = -\frac{dC_n(\theta)}{d\theta}\rfloor_{\theta=\theta_n} + O_p\left(\frac{1}{n^{3/2}}\right).$$

Since $[\frac{d^2 K_n(\theta)}{d\theta^2}\rfloor_{\theta=\theta_n}]^{-1} = O_p(1)$ we have

$$|\widehat{\theta}_n^{(K)} - \widetilde{\theta}_n^{(W)}| = O_p(n^{-1}),$$

thus giving the desired rate.

For the multiparameter case we use (4.24) and the same argument to give

$$\begin{aligned} &\sum_{s=1}^{d}(\widehat{\theta}_{s,n}^{(W)} - \widetilde{\theta}_{s,n}^{(W)})\frac{\partial^2 W_n(\theta)}{\partial\theta_s\partial\theta_r}\rfloor_{\theta=\theta_n} + \\ &\frac{1}{2}\sum_{s_1,s_2=1}^{d}\Big[(\widehat{\theta}_{s_1,n}^{(W)} - \widetilde{\theta}_{s_1,n}^{(W)})(\widehat{\theta}_{s_2,n}^{(W)} - \theta_{s_2,n}) + (\widehat{\theta}_{s_2,n}^{(W)} - \widetilde{\theta}_{s_2,n}^{(W)})(\widetilde{\theta}_{s_1,n}^{(W)} - \theta_{s_1,n})\Big] \times \frac{\partial^3 W_n(\theta)}{\partial\theta_{s_1}\partial\theta_{s_2}\partial\theta_r}\rfloor_{\theta=\theta_n} \\ = \;& O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right). \end{aligned}$$

Replacing $\frac{\partial^2 W_n(\theta)}{\partial\theta_s\partial\theta_r}\rfloor_{\theta=\theta_n}$ with its expectation gives

$$(\widehat{\theta}_n^{(W)} - \widetilde{\theta}_n^{(W)})'\mathbb{E}\left[\nabla_\theta^2 W_n(\theta)\rfloor_{\theta=\theta_n}\right] + |\widehat{\theta}_n^{(W)} - \widetilde{\theta}_n^{(W)}|_1 o_p(1) \;=\; O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).$$

Thus under the assumption that $\mathbb{E}\left[\nabla_\theta^2 W_n(\theta)\right]_{\theta=\theta_n}]$ is invertible we have

$$|\widetilde{\theta}_n^{(W)} - \widehat{\theta}_n^{(W)}|_1 = O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).$$

By a similar argument we have

$$|\widetilde{\theta}_n^{(H)} - \widehat{\theta}_n^{(H)}|_1 = O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).$$

The case when $p = 0$ is analogous to the uniparameter case and we omit the details. This concludes the proof. □

# 5. CONCLUDING REMARKS AND DISCUSSION *

In this dissertation, we have proposed a new approach to overcome the notorious bias issue of spectral analysis. The key idea behind the method is to obtaining a linear transform, denotes $\widetilde{J}_n(\cdot; f)$, that is biorthogonal to the regular DFT. We named it the complete DFT. The complete DFT is an extension of the regular DFT by predicting the time series in the unobserved domain on the top of the original (observed) time series. Unlike other existing methods, the complete DFT, together with the regular DFT, fully decorrelates the second order stationary time series (equation (2.4)). Therefore, we obtain an unbiased estimator of the spectral density $I_n(\cdot; f) = \widetilde{J}_n(\cdot; f)\overline{J_n(\cdot)}$ , so called the complete periodogram. For finite order autoregressive models, the complete periodogram has a finite term analytic expression in terms of the corresponding autoregressive coefficients. This observation shows that estimating the complete periodogram of the finite order autoregressive models boils down to estimate the autoregressive coefficients. In general processes, we have provided steps of approximation to estimate the complete periodogram using data. Both theoretically and empirically, the estimated complete DFT outperforms the ordinary periodogram. It is interesting to note that in simulations, the complete periodogram tends to have a better local and global performance than the tapered periodogram especially when the spectral density has a large peak.

Using the notion of the complete DFT, we have derived an exact expression for the differences $\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1})$ and $\mathcal{L}_n(\theta) - K_n(\theta)$. These expressions are simple, with an intuitive interpretation, in terms of predicting outside the boundary of observation. We have used these expansions and approximations to define two new spectral divergence criteria (in the frequency domain). Our simulations show that both new estimators (termed the boundary corrected and hybrid Whittle) tend to outperform the Whittle likelihood. Intriguingly, the hybrid Whittle likelihood tends to outperform the boundary corrected Whittle likelihood. Currently, we have no theoretical justification

*Parts of this section have been modified with permission from [S. Subba Rao and J. Yang. Reconciling the Gaussian and Whittle likelihood with an application to estimation in the frequency domain. *Annals of Statistics (To appear), arXiv:2001.06966*, 2021.]

for this and one future aim is to investigate these differences.

We believe that it is possible to use a similar construction to obtain expressions for the complete DFT and the difference between the Gaussian likelihood and the Whittle likelihood of a multivariate time series. The construction we use in this dissertation (for an univariate case) relies on the Wold-type AR($\infty$) and MA($\infty$) representation of a time series and makes heavy use of the commutativity property of these expansions. In the multivariate situation, we lose the commutativity property. So the expressions in the past and future predictions are asymmetric. To prove analogous results to those in this dissertation, we will require the Baxter-type inequalities for the multivariate framework. The bounds derived in Cheng and Pourahmadi (1993) and Inoue et al. (2018) may be useful in this context.

An issue in the spatial grid framework is more complicated. This is because the edge effects are accumulated as the dimension increases. Guyon (1982) Section 3.3, showed that the bias caused by using the classical periodogram in the Whittle likelihood is not asymptotically negligible. Therefore, some preprocessors on the time series are necessary. Examples are Guyon (1982) (edge-correction) and Dahlhaus and Künsch (1987) (data taper). Moreover, the AR($p$) approximations for spatial random fields and corresponding Baxter's inequality described in Meyer et al. (2017) may be useful to obtain the bound between the feasible and infeasible estimator (if possible).

Lastly, the emphasis of this dissertation is on short memory time series. But we conclude by briefly discussing extensions to long memory time series. The fundamental feature (in the time domain) that distinguishes a short memory time series from a long memory time series is that the autocovariance function of a long memory time series is not absolutely summable. Proof of results heavily replies on interchanging the order of summation which is guaranteed by the absolutely summable autocovariances when the time series has a short memory. Therefore, for long memory time series, a more careful argument on the interchangeability of summation is required. Series expansion of finite predictor coefficients and Baxter's inequality for long memory time series in Inoue and Kasahara (2006) may be a useful tool to tackle this problem. Moreover, in the frequency domain, the spectral density of a long memory time series is not bounded on the origin. Therefore,

the frequency domain representation of the Gaussian likelihood in Theorem 2.4.1 is not well-defined at $k = n$. However, in our unpublished manuscript Subba Rao and Yang (2021b), we show that the complete DFT is well-defined in a much larger class of second order stationary time series which includes the long memory time series in certain setting. Also, in Subba Rao and Yang (2021a) Appendix A.1, we showed a version of Theorem 2.4.1 for a long memory time series. We state the result without proof. Suppose that $\underline{X}_n^{(c)} = \underline{X}_n - \overline{X}\mathbf{1}_n$, where $\overline{X} = n^{-1}\sum_{t=1}^n X_t$ is a demeaned time series. Then,

$$\frac{1}{n}\underline{X}_n^{(c)\prime}\Gamma_n(f_\theta)^{-1}\underline{X}_n^{(c)} = \frac{1}{n}\sum_{k=1}^{n-1}\frac{|J_n(\omega_{k,n})|^2}{f_\theta(\omega_{k,n})} + \frac{1}{n}\sum_{k=1}^{n-1}\frac{\widehat{J}_n^{(c)}(\omega_{k,n};f_\theta)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})},$$

where $\widehat{J}_n^{(c)}(\cdot;f_\theta)$ denotes the predictive DFT of the demeaned time series $\underline{X}_n^{(c)}$. The results are not conclusive, but they do suggest that the new likelihoods, in some settings, can be used to reduce the bias for long memory parameter estimators.

In summary, a new spectral method using the complete DFT may be of value in future research.

# REFERENCES


K. M. Abadir, W. Distaso, and L. Giraitis. Nonstationarity-extended local Whittle estimation. *Journal of Econometrics*, 141(2):1353–1384, 2007.

J.-M. Bardet, P. Doukhan, and J. R. León. Uniform limit theorems for the integrated periodogram of weakly dependent time series and their applications to Whittle's estimate. *Journal of Time Series Analysis*, 29(5):906–945, 2008.

M. S. Bartlett. Approximate confidence intervals II. *Biometrika*, 40:306–317, 1953.

G. Baxter. An asymptotic result for the finite predictor. *Mathematica Scandinavica*, 10:137–144, 1962.

G. Baxter. A norm inequality for a "finite-section" Wiener-Hopf equation. *Illinois Journal of Mathematics*, 7(1):97–103, 1963.

K. I. Beltrão and P. Bloomfield. Determining the bandwidth of a kernel spectrum estimate. *Journal of Time Series Analysis*, 8(1):21–38, 1987.

R. J. Bhansali. The evaluation of certain quadratic forms occurring in autoregressive model fitting. *Annals of Statistics*, 10(1):121–131, 1982.

R. J. Bhansali. Asymptotically efficient autoregressive model selection for multistep prediction. *Annals of the Institute of Statistical Mathematics*, 48(3):577–602, 1996.

Peter Bloomfield. *Fourier analysis of time series: an introduction*. John Wiley & Sons, 2004.

Albrecht Böttcher and Bernd Silbermann. *Analysis of Toeplitz operators*. Springer Science & Business Media, 2013.

David R Brillinger. *Time series: Data Analysis and Theory*, volume 36. SIAM, 1981.

Peter J. Brockwell and Richard A. Davis. *Time series: theory and methods*. Springer Series in Statistics. Springer, New York, 2006. Reprint of the second (1991) edition.

R. Cheng and M. Pourahmadi. Baxter's inequality and convergence of finite predictors of multivariate stochastic processes. *Probability Theory and Related Fields*, 95:115–124, 1993.

N. Choudhuri, S. Ghosal, and A. Roy. Bayesian estimation of the spectral density of a time series.

*Journal of the American Statistical Association*, 99(468):1050–1059, 2004.

J. Coursol and D. Dacunha-Castelle. Remarques sur l'approximation de la vraisemblance d'un processus Gaussien stationnaire. *Teoriya Veroyatnosteĭ i ee Primeneniya (Theory of Probability and its Applications)*, 27(1):155–160, 1982.

D. R. Cox and E. J. Snell. A general definition of residuals. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 30(2):248–265, 1968.

R. Dahlhaus. Spectral analysis with tapered data. *Journal of Time Series Analysis*, 4(3):163–175, 1983.

R. Dahlhaus. Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *Annals of Statistics*, 16(2):808–841, 1988.

R. Dahlhaus. Nonparametric high resolution spectral estimation. *Probability Theory and Related Fields*, 85(2):147–180, 1990.

R. Dahlhaus and D. Janas. A frequency domain bootstrap for ratio statistics in time series analysis. *Annals of Statistics*, 24(5):1934–1963, 1996.

R. Dahlhaus and H. Künsch. Edge effects and efficient parameter estimation for stationary random fields. *Biometrika*, 74(4):877–882, 1987.

S. Das, S. Subba Rao, and J. Yang. Spectral methods for small sample time series: A complete periodogram approach. *Journal of Time Series Analysis (To appear)*, 2021.

M. Eichler. Testing nonparametric and semi-parametric hypothesis in vector stationary processes. *Journal of Multivariate Analysis*, 99:968–1009, 2008.

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

R. F. Galbraith and J. I. Galbraith. On the inverses of some patterned matrices arising in the theory of stationary time series. *Journal of Applied Probability*, 11(1):63–71, 1974.

L. Giraitis and P. M. Robinson. Whittle estimation of ARCH models. *Econometric Theory*, 17(3): 608–631, 2001.

Liudas Giraitis, Hira L. Koul, and Donatas Surgailis. *Large sample inference for long memory*

*processes*. Imperial College Press, London, 2012.

J. Guinness. Spectral density estimation for random fields via periodic embeddings. *Biometrika*, 106:267–286, 2019.

X. Guyon. Parameter estimation for a stationary process on a $d$-dimensional lattice. *Biometrika*, 69(1):95–105, 1982.

C. M. Hurvich. Data-driven choice of a spectrum estimate: extending the applicability of cross-validation methods. *Journal of the American Statistical Association*, 80(392):933–940, 1985.

C. M. Hurvich. A mean squared error criterion for time series data windows. *Biometrika*, 75(3): 485–490, 1988.

C. M. Hurvich and W. W. Chen. An efficient taper for potentially overdifferenced long-memory time series. *Journal of Time Series Analysis*, 21(2):155–180, 2000.

C.-K. Ing and C.-Z. Wei. Order selection for same-realization predictions in autoregressive processes. *Annals of Statistics*, 33(5):2423–2474, 2005.

A. Inoue and Y. Kasahara. Explicit representation of finite predictor coefficients and its applications. *Annals of Statistics*, 34(2):973–993, 2006.

A. Inoue, Y. Kasahara, and M. Pourahmadi. Baxter's inequality for finite predictor coefficients of multivariate long-memory stationary processes. *Bernoulli*, 24(2):1202–1232, 2018.

Y. Kasahara, M. Pourahmadi, and A. Inoue. Duals of random vectors and processes with applications to prediction problems with missing values. *Statistics & Probability Letters*, 79(14): 1637–1646, 2009.

C. Kirch, M. C. Edwards, A. Meier, and R. Meyer. Beyond Whittle: Nonparametric correction of a parametric likelihood with a focus on bayesian time series analysis. *Bayesian Analysis*, 14(4): 1037–1073, 2019.

T. Kley, P. Preuß, and P. Fryzlewicz. Predictive, finite-sample model choice for time series under stationarity and non-stationarity. *Electronic Journal of Statistics*, 13(2):3710–3774, 2019.

J. Krampe, J.-P. Kreiss, and E. Paparoditis. Estimated Wold representation and spectral-density-driven bootstrap for time series. *Journal of the Royal Statistical Society. Series B. Statistical*

*Methodology*, 80(4):703–726, 2018.

J.-P. Kreiss, E. Paparoditis, and D. N. Politis. On the range of validity of the autoregressive sieve bootstrap. *Annals of Statistics*, 39(4):2103–2130, 2011.

S. N. Lahiri. A necessary and sufficient condition for asymptotic independence of discrete Fourier transforms under short- and long-range dependence. *Annals of Statistics*, 31(2):613–641, 2003.

T. C. Lee and Z. Zhu. Nonparametric spectral density estimation with missing observations. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3041–3044. IEEE, 2009.

T. L. McMurry and D. N. Politis. High-dimensional autocovariance matrices and optimal linear prediction. *Electronic Journal of Statistics*, 9(1):753–788, 2015.

M. Meyer, C. Jentsch, and J.-P. Kreiss. Baxter's inequality and sieve bootstrap for random fields. *Bernoulli*, 23(4B):2988–3020, 2017.

T. Mikosch and Y. Zhao. The integrated periodogram of a dependent extremal event sequence. *Stochastic Processes and their Applications*, 125(8):3126–3169, 2015.

A. Milhøj. A test of fit in time series models. *Biometrika*, 68:177–187, 1981.

T. Niebuhr and J.-P. Kreiss. Asymptotics for autocovariances and integrated periodograms for linear processes observed at lower frequencies. *International Statistical Review*, 82(1):123–140, 2014.

H. C. Ombao, J. A. Raz, R. L. Strawderman, and R. Von Sachs. A simple generalised crossvalidation method of span selection for periodogram smoothing. *Biometrika*, 88(4):1186–1192, 2001.

V. M. Panaretos and S. Tavakoli. Fourier analysis of stationary time series in function space. *Annals of Statistics*, 41(2):568–603, 2013.

M. Pourahmadi. Exact factorization of the spectral density and its application to forecasting and time series analysis. *Communications in Statistics. A. Theory and Methods*, 12(18):2085–2094, 1983.

M. Pourahmadi. Taylor expansion of $\exp(\sum_{k=0}^{\infty} a_k z^k)$ and some applications. *American Mathe-*

*matical Monthly*, 91(5):303–307, 1984.

Mohsen Pourahmadi. *Foundations of time series analysis and prediction theory*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York, 2001.

Maurice B. Priestley. *Spectral Analysis and Time Series*. Academic Press, London, 1981.

R. B. Randall and J. Antoni. Rolling element bearing diagnostics—a tutorial. *Mechanical systems and signal processing*, 25(2):485–520, 2011.

P. M. Robinson. Gaussian semiparametric estimation of long range dependence. *Journal of the American Statistical Association*, 23(5):1630–1661, 1995.

A. Schuster. On lunar and solar periodicities of earthquakes. *Proceedings of the Royal Society of London*, 61:455–465, 1897.

A. Schuster. On the periodicities of sunspots. *Philosophical transactions of the royal society A*, 206:69–100, 1906.

P. Shaman. An approximate inverse for the covariance matrix of moving average and autoregressive processes. *Annals of Statistics*, 3(2):532–538, 1975.

P. Shaman. Approximations for stationary covariance matrices and their inverses with application to ARMA models. *Annals of Statistics*, 4(2):292–301, 1976.

P. Shaman and R. A. Stine. The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association*, 83(403):842–848, 1988.

X. Shao and W. B. Wu. Local whittle estimation of fractional integration for nonlinear processes. *Econometric Theory*, 23(5):899–929, 2007.

M. M. Siddiqui. On the inversion of the sample covariance matrix in a stationary autoregressive process. *Annals of Mathematical Statistics*, 29(2):585–588, 1958.

W. A. Smith and R. B. Randall. Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mechanical Systems and Signal Processing*, 64:100–131, 2015.

S. Subba Rao. Orthogonal samples for estimators in time series. *Journal of Time Series Analysis*,

39:313–337, 2018.

S. Subba Rao and J. Yang. Reconciling the Gaussian and Whittle likelihood with an application to estimation in the frequency domain. *Annals of Statistics (To appear), arXiv:2001.06966*, 2021a.

S. Subba Rao and J. Yang. Frequency domain representation of the Gaussian likelihood for the long memory time series. *In progress*, 2021b.

A. M. Sykulski, S. C. Olhede, A. P. Guillaumin, J. M. Lilly, and J. J. Early. The debiased Whittle likelihood. *Biometrika*, 106(2):251–266, 2019.

G. Szegö. Über die randwerte einer analytischen funktion. *Mathematische Annalen*, 84:232–244, 1921.

K. Tanaka. An asymptotic expansion associated with the maximum likelihood estimators in ARMA models. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 46(1):58–67, 1984.

M. Taniguchi. On the second order asymptotic efficiency of estimators of gaussian ARMA processes. *Annals of Statistics*, 11:157–169, 1983.

D. Tjøstheim and J. Paulsen. Bias of some commonly-used time series estimates. *Biometrika*, 70 (2):389–399, 1983.

J. W. Tukey. An introduction to the calculations of numerical spectrum analysis. In *Spectral Analysis of Time Series (Proc. Advanced Sem., Madison, Wis., 1966)*, pages 25–46. John Wiley, New York, 1967.

A. van Delft and M. Eichler. A note on Herglotz's theorem for time series on functional spaces. *Stochastic Processes and their Applications*, 130(6):3687–3710, 2020.

A. M. Walker. Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *Journal of the Australian Mathematical Society*, 4: 363–384, 1964.

P. Whittle. The analysis of multiple stationary time series. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 15:125–139, 1953.

Peter Whittle. *Hypothesis Testing in Time Series Analysis*. Thesis, Uppsala University, 1951.

G. T. Wilson. The factorization of matricial spectral densities. *SIAM Journal on Applied Mathematics*, 23:420–426, 1972.

H.-C. Zhang. Reduction of the asymptotic bias of autoregressive and spectral estimators by tapering. *Journal of Time Series Analysis*, 13(5):451–469, 1992.

# APPENDIX A

# THE BAXTER'S INEQUALITY *

## A.1 An extension of Baxter's inequality

Let $\{X_t\}$ be a second order stationary time series with absolutely summable autocovariance and spectral density $f$. We can represent $f$ as $f(\omega) = \psi(\omega)\overline{\psi(\omega)} = 1/\left(\phi(\omega)\overline{\phi(\omega)}\right)$ where

$$\phi(\omega) = 1 - \sum_{s=1}^{\infty} \phi_s e^{-is\omega} \text{ and } \psi(\omega) = 1 + \sum_{s=1}^{\infty} \psi_s e^{-is\omega}.$$

Note that $\{\phi_s\}$ and $\{\psi_s\}$ are the corresponding AR($\infty$) and MA($\infty$) coefficients respectively and $\psi(\omega) = \phi(\omega)^{-1}$. To simplify notation we have ignored the variance of the innovation.

Many of the results in this disseration hinge on a generalization of Baxter's inequality which we summarize below.

**Lemma A.1.1** (Extended Baxter's inequality)**.** *Suppose $f(\cdot)$ is a spectral density function which satisfies Assumption 2.3.1. Let $\psi(\cdot)$ and $\phi(\cdot)$ AR and MA characteristic function respectively. Let $\phi_{p+1}^{\infty}(\omega) = \sum_{s=p+1}^{\infty} \phi_s e^{-is\omega}$. Further, let $\{\phi_{s,n}(\tau)\}$ denote the coefficients in the best linear predictor of $X_\tau$ given $\underline{X}_n = \{X_t\}_{t=1}^n$ and $\{\phi_s(\tau)\}$ the corresponding the coefficients in the best linear predictor of $X_\tau$ given $\underline{X}_\infty = \{X_t\}_{t=1}^\infty$, where $\tau \leqslant 0$. Suppose $p$ is large enough such that $\left\|\phi_p^\infty\right\|_K \|\psi\|_K \leqslant \varepsilon < 1$. Then for all $n > p$ we have*

$$\sum_{s=1}^{n} (2^K + s^K) \left|\phi_{s,n}(\tau) - \phi_s(\tau)\right| \leqslant C_{f,K} \sum_{s=n+1}^{\infty} (2^K + s^K) \left|\phi_s(\tau)\right|, \tag{A.1}$$

*where $C_{f,K} = \frac{3-\varepsilon}{1-\varepsilon} \|\phi\|_K^2 \|\psi\|_K^2$ and $\phi_s(\tau) = \sum_{j=0}^{\infty} \phi_{s+j}\psi_{|\tau|-j}$ (we set $\psi_0 = 1$ and $\psi_j = 0$ for $j < 0$).*

Before we give a proof, we define an appropriate norm on the subspace of $L_2[0, 2\pi]$.

*Parts of this section have been modified with permission from [S. Subba Rao and J. Yang. Reconciling the Gaussian and Whittle likelihood with an application to estimation in the frequency domain. *Annals of Statistics (To appear), arXiv:2001.06966*, 2021.]

**Definition A.1.1** (Norm on the subspace of $L_2[0, 2\pi]$)**.** *Suppose the sequence of positive weights* $\{v(k)\}_{k\in\mathbb{Z}}$ *satisfies 2 conditions: (1)* $v(n)$ *is even, i.e.,* $v(-n) = v(n)$ *for all* $n \geqslant 0$*; (2)* $v(n+m) \leqslant v(n)v(m)$ *for all* $n, m \in \mathbb{Z}$*.*

*Given* $\{v(k)\}$ *satisfies 2 conditions above, define a subspace* $A_v$ *of* $L_2[0, 2\pi]$ *by*

$$A_v = \{f \in L_2[0, 2\pi] : \sum_{k\in\mathbb{Z}} v(k)|f_k| < \infty\}.$$

*where,* $f(\omega) = \sum_{k\in\mathbb{Z}} f_k e^{ik\omega}$*. We define a norm* $\|f\|$ *on* $A_v$ *by* $\|f\| = \sum_{k\in\mathbb{Z}} v(k)|f_k|$*, then it is easy to check this is a valid norm.*

**Remark A.1.1** (Properties of $\|\cdot\|$)**.** *Suppose the sequence* $\{v(k)\}_{k\in\mathbb{Z}}$ *satisfies 2 conditions in Definition A.1.1, and define the norm* $\|\cdot\|$ *with respect to* $\{v(k)\}$*. Then, beside the triangle inequality, this norm also satisfies* $\|1\| = v(0) \leqslant 1$*,* $\|f\| = \|\overline{f}\|$*, and* $\|fg\| \leqslant \|f\|\|g\|$ *(which does not hold for all norms but is an important component of the (extended) Baxter's proof), i.e.,* $(A_v, \|\cdot\|)$ *is a Banach algebra with involution operator. The proof for the multiplicative inequality follows from the fact that* $(fg)_k = \sum_r f_r g_{k-r}$*, where* $f_k$ *and* $g_k$ *are* $k$*th Fourier coefficient of* $f$ *and* $g$*. Thus*

$$\begin{aligned}\|fg\| &\leqslant \sum_{k\in\mathbb{Z}} v(k)\left|\sum_{r\in\mathbb{Z}} f_r g_{k-r}\right| \\ &\leqslant \sum_{k\in\mathbb{Z}} v(r)v(k-r)\left|\sum_{r\in\mathbb{Z}} f_r g_{k-r}\right| \leqslant \sum_{k,r\in\mathbb{Z}} v(r)v(k-r)|f_r||g_{k-r}| = \|f\|\|g\|.\end{aligned}$$

*Examples of weights include* $v(r) = (2^q + |r|^q)$ *or* $v(r) = (1 + |r|)^q$ *for some* $q \geqslant 0$*. In these two examples, when* $q = K$*, under Assumption 2.3.1,* $\psi(\omega),\ \phi(\omega) \in A_v$ *where* $\psi(\omega) = 1 + \sum_{j=1}^{\infty} \psi_j e^{-ij\omega}$ *and* $\phi(\omega) = 1 - \sum_{j=1}^{\infty} \phi_j e^{-ij\omega}$ *(see Kreiss et al. (2011)).*

PROOF. The proof below follows closely the proof Baxter (1962, 1963). Let $\{\phi_{s,p}(\tau)\}_{s=1}^{p}$ denote the the coefficients of the best linear predictor of $X_{t+\tau}$ (for $\tau \geqslant 0$) given $\{X_s\}_{t-p}^{t-1}$

$$\mathbb{E}\left[\left(X_{t+\tau} - \sum_{s=1}^{p} \phi_{s,p}(\tau)X_{t-s}\right)X_{t-k}\right] = 0 \text{ for } k = 1, \ldots, p. \tag{A.2}$$

and $\{\phi_s(\tau)\}$ denote the coefficients of the best linear predictor of $X_{t+\tau}$ given the infinite past $\{X_s\}_{s=-\infty}^{t-1}$

$$\mathbb{E}\left[\left(X_{t+\tau} - \sum_{s=1}^{\infty} \phi_s(\tau) X_{t-s}\right) X_{t-k}\right] = 0 \text{ for } k = 1, 2, \ldots \tag{A.3}$$

We use the same proof as Baxter, which is based on rewriting the normal equations in (A.2) within the frequency domain to yield

$$\frac{1}{2\pi}\int_0^{2\pi}\left(e^{i\tau\omega} - \sum_{s=1}^{p}\phi_{s,p}(\tau)e^{-is\omega}\right) f(\omega)e^{-ik\omega}d\omega = 0, \text{ for } k = 1, \ldots, p$$

Similarly, using the infinite past to do prediction yields the normal equations

$$\frac{1}{2\pi}\int_0^{2\pi}\left(e^{i\tau\omega} - \sum_{s=1}^{\infty}\phi_s(\tau)e^{-is\omega}\right) f(\omega)e^{-ik\omega}d\omega = 0, \text{ for } k \geqslant 1.$$

Thus taking differences of the above two equations for $k = 1, \ldots, p$ gives

$$\begin{aligned}\frac{1}{2\pi}\int_0^{2\pi}&\left(\sum_{s=1}^{p}\left[\phi_{s,p}(\tau) - \phi_s(\tau)\right]e^{-is\omega}\right) f(\omega)e^{-ik\omega}d\omega \\ &= \frac{1}{2\pi}\int_0^{2\pi}\left(\sum_{s=p+1}^{\infty}\phi_s(\tau)e^{-is\omega}\right) f(\omega)e^{-ik\omega}d\omega \quad 1 \leqslant k \leqslant p.\end{aligned} \tag{A.4}$$

These $p$-equations give rise to Baxter's Weiner-Hopf equations and allow one to find a bound for $\sum_{s=1}^{p}|\phi_{s,p}(\tau) - \phi_s(\tau)|$ in terms of $\sum_{s=p+1}^{\infty}|\phi_s(\tau)|$. Interpreting the above, we have two different functions $\left(\sum_{s=1}^{p}\left[\phi_{s,p}(\tau) - \phi_s(\tau)\right]e^{-is\omega}\right) f(\omega)$ and $\left(\sum_{s=p+1}^{\infty}\phi_s(\tau)e^{-is\omega}\right) f(\omega)$ whose first $p$ Fourier coefficients are the same.

Define the polynomials

$$h_p(\omega) = \sum_{s=1}^{p}\left[\phi_{s,p}(\tau) - \phi_s(\tau)\right]e^{-is\omega} \quad \text{and} \quad g_p(\omega) = \sum_{k=1}^{p} g_{k,p}e^{ik\omega} \tag{A.5}$$

where

$$g_{k,p} = (2\pi)^{-1}\int_0^{2\pi}\left(\sum_{s=p+1}^{\infty}\phi_s(\tau)e^{-is\omega}\right) f(\omega)e^{-ik\omega}d\omega. \tag{A.6}$$

For the general norm $\|\cdot\|$ defined in Definition A.1.1, will show that for a sufficiently large $p$, $\|h_p\| \leqslant C_f \|g_p\|$, where the constant $C_f$ is a function of the spectral density (that we will derive).

The Fourier expansion of $h_p f$ is

$$h_p(\omega) f(\omega) = \sum_{k=-\infty}^{\infty} \widetilde{g}_{k,p} e^{ik\omega},$$

where $\widetilde{g}_{k,p} = (2\pi)^{-1} \int_0^{2\pi} h_p(\omega) f(\omega) e^{-ik\omega} d\omega$. Then, by (A.4) for $1 \leqslant k \leqslant p$, $\widetilde{g}_{k,p} = g_{k,p}$ (where $g_{k,p}$ is defined in (A.6)). Thus

$$h_p(\omega) f(\omega) = G^0_{-\infty}(\omega) + g_p(\omega) + G^{\infty}_{p+1}(\omega) \tag{A.7}$$

where

$$G^0_{-\infty}(\omega) = \sum_{k=-\infty}^{0} \widetilde{g}_{k,p} e^{ik\omega} \quad \text{and} \quad G^{\infty}_{p+1}(\omega) = \sum_{s=p+1}^{\infty} \widetilde{g}_{k,p} e^{ik\omega}.$$

Dividing by $f^{-1} = \phi\overline{\phi}$ and taking the $\|\cdot\|$-norm we have

$$\begin{aligned} \|h_p\| &\leqslant \left\|f^{-1} G^0_{-\infty}\right\| + \left\|f^{-1} g_p\right\| + \left\|f^{-1} G^{\infty}_{p+1}\right\| \\ &\leqslant \left\|f^{-1} G^0_{-\infty}\right\| + \left\|f^{-1}\right\| \|g_p\| + \left\|f^{-1} G^{\infty}_{p+1}\right\| \\ &\leqslant \left\|\overline{\phi}\right\| \left\|\phi G^0_{-\infty}\right\| + \left\|f^{-1}\right\| \|g_p\| + \|\phi\| \left\|\overline{\phi} G^{\infty}_{p+1}\right\|. \end{aligned} \tag{A.8}$$

First we obtain bounds for $\left\|\phi G^0_{-\infty}\right\|$ and $\left\|\overline{\phi} G^{\infty}_{p+1}\right\|$ in terms of $\|g_p\|$. We will show that for a sufficiently large $p$

$$\begin{aligned} \left\|\phi G^0_{-\infty}\right\| &\leqslant \|\phi\| \|g_p\| + \varepsilon \left\|\overline{\phi} G^{\infty}_{p+1}\right\| \\ \left\|\overline{\phi} G^{\infty}_{p+1}\right\| &\leqslant \left\|\overline{\phi}\right\| \|g_p\| + \varepsilon \left\|\phi G^0_{-\infty}\right\|. \end{aligned}$$

The bound for these terms hinges on the Fourier coefficients of a function being unique, which allows us to compare coefficients across functions. Some comments are in order that will help in

the bounding of the above. We recall that $f(\omega)^{-1} = \phi(\omega)\overline{\phi(\omega)}$, where

$$\phi(\omega) = 1 - \sum_{s=1}^{\infty} \phi_s e^{-is\omega} \qquad \overline{\phi(\omega)} = 1 - \sum_{s=1}^{\infty} \phi_s e^{is\omega}.$$

Thus $\phi(\omega)G^0_{-\infty}(\omega)$ and $\overline{\phi(\omega)}G^{\infty}_{p+1}(\omega)$ have Fourier expansions with only less than the first and greater than the $p$th frequencies respectively. This observation gives the important insight into the proof. Suppose $b(\omega) = \sum_{j=-\infty}^{\infty} b_j e^{ij\omega}$, we will make the use of the notation $\{b(\omega)\}_+ = \sum_{j=1}^{\infty} b_j e^{ij\omega}$ and $\{b(\omega)\}_- = \sum_{j=-\infty}^{0} b_j e^{ij\omega}$, thus $b(\omega) = \{b(\omega)\}_- + \{b(\omega)\}_+$.

We now return to (A.7) using that $f = \psi(\omega)\overline{\psi(\omega)}$ we multiply (A.7) by $\psi(\omega)^{-1} = \phi(\omega)$ to give

$$h_p(\omega)\overline{\psi(\omega)} = \phi(\omega)G^0_{-\infty}(\omega) + \phi(\omega)g_p(\omega) + \phi(\omega)G^{\infty}_{p+1}(\omega). \tag{A.9}$$

Rearranging the above gives

$$-\phi(\omega)G^0_{-\infty}(\omega) = -h_p(\omega)\overline{\psi(\omega)} + \phi(\omega)g_p(\omega) + \phi(\omega)G^{\infty}_{p+1}(\omega).$$

We recall that $h_p(\omega)\overline{\psi(\omega)}$ only contain positive frequencies, whereas $\phi(\omega)G^0_{-\infty}(\omega)$ only contains non-positive frequencies. Based on these observations we have

$$\begin{aligned} &-\phi(\omega)G^0_{-\infty}(\omega) \\ &\quad = \left\{-\phi(\omega)G^0_{-\infty}(\omega)\right\}_- = \{\phi(\omega)g_p(\omega)\}_- + \left\{\phi(\omega)G^{\infty}_{p+1}(\omega)\right\}_-. \end{aligned} \tag{A.10}$$

We further observe that $G^{\infty}_{p+1}$ only contains non-zero coefficients for positive frequencies of $p$+1 and greater, thus only the coefficients of $\phi(\omega)$ with frequencies less or equal to $-(p+1)$ will give non-positive frequencies when multiplied with $G^{\infty}_{p+1}$. Therefore

$$-\phi(\omega)G^{-1}_{-\infty}(\omega) = \{\phi(\omega)g_p(\omega)\}_- + \left\{\phi^{\infty}_{p+1}(\omega)G^{\infty}_{p+1}(\omega)\right\}_-,$$

where $\phi^{\infty}_{p+1}(\omega) = \sum_{s=p+1}^{\infty} \phi_s e^{-is\omega}$. Evaluating the norm of the above (using both the triangle and

the multiplicative inequality) we have

$$\begin{aligned}\left\|\phi G^0_{-\infty}\right\| &\leqslant \|\phi\|\,\|g_p\| + \left\|\phi^\infty_{p+1} G^\infty_{p+1}\right\| \\ &\leqslant \|\phi\|\,\|g_p\| + \left\|\phi^\infty_{p+1}\right\|\left\|\overline{\psi}\right\|\left\|\overline{\phi} G^\infty_{p+1}\right\| \quad \text{since } \overline{\psi(\omega)\phi(\psi)} = 1.\end{aligned}$$

This gives a bound for $\left\|\phi G^0_{-\infty}\right\|$ in terms of $\|g_p\|$ and $\left\|\overline{\phi} G^\infty_{p+1}\right\|$. Next we obtain a similar bound for $\left\|\overline{\phi} G^\infty_{p+1}\right\|$ in terms of $\|g_p\|$ and $\left\|\phi G^0_{-\infty}\right\|$.

Again using (A.7), $f(\omega) = \psi(\omega)\overline{\psi(\omega)}$, but this time multiplying (A.7) by $\overline{\psi(\omega)}^{-1} = \overline{\phi(\omega)}$, we have

$$h_p(\omega)\psi(\omega) = \overline{\phi(\omega)} G^0_{-\infty}(\omega) + \overline{\phi(\omega)} g_p(\omega) + \overline{\phi(\omega)} G^\infty_{p+1}(\omega).$$

Rearranging the above gives

$$\overline{\phi(\omega)} G^\infty_{p+1}(\omega) = h_p(\omega)\psi(\omega) - \overline{\phi(\omega)} G^0_{-\infty}(\omega) - \overline{\phi(\omega)} g_p(\omega).$$

We observe that $\overline{\phi(\omega)} G^\infty_{p+1}(\omega)$ contains frequencies greater than $p$ whereas $h_p(\omega)\psi(\omega)$ only contains frequencies less or equal to the order $p$ (since $h_p$ is a polynomial up to order $p$). Therefore multiply $e^{-ip\omega}$ on both side and take $\{\}_+$ gives

$$\begin{aligned}&e^{-ip\omega}\overline{\phi(\omega)} G^\infty_{p+1}(\omega) \\ &\quad - \left\{e^{-ip\omega}\overline{\phi(\omega)} G^0_{-\infty}(\omega)\right\}_+ - \left\{e^{-ip\omega}\overline{\phi(\omega)} g_p(\omega)\right\}_+,\end{aligned} \tag{A.11}$$

By the similar technique from the previous, it is easy to show

$$\left\{e^{-ip\omega}\overline{\phi(\omega)} G^0_{-\infty}(\omega)\right\}_+ = \left\{e^{-ip\omega}\overline{\phi^\infty_{p+1}(\omega)} G^0_{-\infty}(\omega)\right\}_+. \tag{A.12}$$

Multiplying $e^{ip\omega}$ and evaluating the $\|\cdot\|$-norm of the above yields the inequality

$$\begin{aligned}\left\|\overline{\phi}G_{p+1}^{\infty}\right\| &\leqslant \left\|\overline{\phi}g_p\right\| + \left\|\overline{\phi_{p+1}^{\infty}}G_{-\infty}^{0}\right\| \\ &\leqslant \left\|\overline{\phi}\right\|\|g_p\| + \left\|\overline{\phi_{p+1}^{\infty}}\right\|\|\psi\|\left\|\phi G_{-\infty}^{0}\right\|.\end{aligned}$$

We note that $\|\phi_{p+1}^{\infty}\| = \|\overline{\phi_{p+1}^{\infty}}\|$. For $\phi \in A_v$ (see Definition A.1.1 and Remark A.1.1), $\|\overline{\phi_{p+1}^{\infty}}\| = \sum_{s=p+1}^{\infty} v(s)|\phi_s| \to 0$ as $p \to \infty$, for a large enough $p$, $\|\psi(\omega)\| \cdot \|\phi_{p+1}^{\infty}\| < 1$. Suppose that $p$ is such that $\left\|\phi_{p+1}^{\infty}(\omega)\right\|\|\psi(\omega)\| \leqslant \varepsilon < 1$, then we have the desired bounds

$$\begin{aligned}\left\|\phi G_{-\infty}^{0}\right\| &\leqslant \|\phi\|\|g_p\| + \varepsilon\left\|\overline{\phi}G_{p+1}^{\infty}\right\| \\ \left\|\overline{\phi}G_{p+1}^{\infty}\right\| &\leqslant \left\|\overline{\phi}\right\|\|g_p\| + \varepsilon\left\|\phi G_{-\infty}^{0}\right\|.\end{aligned}$$

The above implies that $\left\|\phi G_{-\infty}^{0}\right\| + \left\|\overline{\phi}G_{p+1}^{\infty}\right\| \leqslant 2(1-\varepsilon)^{-1}\|\phi\|\|g_p\|$. Substituting the above in (A.8), and using that $\|\phi\| \geqslant 1$ (since $\phi = 1 - \sum_{s=1}^{\infty}\phi_s e^{-is\omega}$, $\|\phi\| \geqslant \|1\| = v(0) \geqslant 1$) we have

$$\begin{aligned}\|h_p\| &\leqslant \frac{2\|\phi\|\|g_p\|}{1-\varepsilon} + \left\|f^{-1}\right\|\|g_p\| \\ &\leqslant (1-\varepsilon)^{-1}\left(2\|\phi\| + (1-\varepsilon)\|\phi\|^2\right)\|g_p\| \leqslant \frac{3-\varepsilon}{1-\varepsilon}\|\phi\|^2\|g_p\|.\end{aligned}$$

Thus based on the above we have

$$\|h_p\| \leqslant \frac{3-\varepsilon}{1-\varepsilon}\|\phi\|^2\|g_p\|. \tag{A.13}$$

Finally, we obtain a bound for $\|g_p\|$ in terms of $\sum_{s=p+1}^{\infty}|\phi_s(\tau)|$. We define an extended version of the function $g_p(\omega)$. Let $\widetilde{g}_p(\omega) = \sum_{k\in\mathbb{Z}} g_{k,p}e^{ik\omega}$ where $g_{k,p}$ is as in (A.5). By definition, $\widetilde{g}_p(\omega) = \left(\sum_{s=p+1}^{\infty}\phi_s(\tau)e^{-is\omega}\right)f(\omega)$ and the Fourier coefficients of $g_p(\omega)$ are contained within $\widetilde{g}_p(\omega)$, which implies

$$\|g_p\| \leqslant \|\widetilde{g}_p\| = \left\|\phi_{p+1}^{\infty}(\tau)f\right\| \leqslant \left\|\phi_{p+1}^{\infty}(\tau)\right\|\|f\| \leqslant \left\|\phi_{p+1}^{\infty}\right\|\|\psi\|^2. \tag{A.14}$$

where $\phi_{p+1}^{\infty}(\tau)(\omega) = \sum_{s=p+1}^{\infty}\phi_s(\tau)e^{-is\omega}$. Finally, substituting (A.14) into (A.13), implies that if $p$

is large enough such that $\left\|\phi_{p+1}^{\infty}\right\| \|\psi\| \leqslant \varepsilon < 1$, then

$$\|h_p\| \leqslant \frac{3-\varepsilon}{1-\varepsilon} \|\phi\|^2 \|\psi\|^2 \left\|\phi_{p+1}^{\infty}(\tau)\right\|.$$

Thus, if the weights in the norm are $v(m) = (2^K + m^K)$ (it is well-defined weights, see Remark A.1.1) we have

$$\begin{aligned}
&\sum_{s=1}^{p}(2^K + s^K) \left|\phi_{s,p}(\tau) - \phi_s(\tau)\right| \\
&\qquad \leqslant \frac{3-\varepsilon}{1-\varepsilon} \|\phi\|_K^2 \|\psi\|_K^2 \sum_{s=p+1}^{\infty} (2^K + s^K) \left|\phi_s(\tau)\right|. \qquad \text{(A.15)}
\end{aligned}$$

## A.2 Baxter's inequality on the derivatives of the coefficients

Our aim is to obtain a Baxter-type inequality for the derivatives of the linear predictors. These bounds will be used when obtaining expression for the bias of the Gaussian and Whittlelikelihoods. However, they may also be of independent interest. It is interesting to note that the following result can be used to show that the Gaussian and Whittle likelihood estimators are asymptotically equivalent in the sense that $\sqrt{n}|\widehat{\theta}_n^{(G)} - \widehat{\theta}_n^{(K)}|_1 \xrightarrow{\mathcal{P}} 0$ as $n \to \infty$.

The proof of the result is based on the novel proof strategy developed in Theorem 3.2 of Meyer et al. (2017) (for spatial processes). We require the following definitions. Define the two $n$-dimension vectors

$$\begin{aligned}
\underline{\varphi}_n(\tau; f_\theta) &= (\phi_{1,n}(\tau; f_\theta), \ldots, \phi_{n,n}(\tau; f_\theta))' \quad \text{(best linear finite future predictor)} \qquad \text{(A.16)} \\
\underline{\phi}_n(\tau; f_\theta) &= (\phi_1(\tau; f_\theta), \ldots, \phi_n(\tau; f_\theta))' \quad \text{(truncated best linear infinite future predictor)}.
\end{aligned}$$

**Lemma A.2.1.** *Let $\theta$ be a $d$-dimension vector. Let $\{c_\theta(r)\}$, $\{\phi_j(f_\theta)\}$ and $\{\psi_j(f_\theta)\}$ denote the autocovariances, AR($\infty$), and MA($\infty$) coefficients corresponding to the spectral density $f_\theta$. For all*

*$\theta \in \Theta$ and for $0 \leqslant i \leqslant \kappa$ we assume*

$$\sum_{j=1}^{\infty} \|j^K \nabla_\theta^i \phi_j(f_\theta)\|_1 < \infty \quad \sum_{j=1}^{\infty} \|j^K \nabla_\theta^i \psi_j(f_\theta)\|_1 < \infty, \tag{A.17}$$

*where $K > 1$. Let $\underline{\varphi}_n(\tau; f_\theta)$ and $\underline{\phi}_n(\tau; f_\theta)$, be defined as in (A.16). We assume that $\tau \leqslant 0$. Then for all $0 \leqslant i \leqslant \kappa$, we have*

$$\begin{aligned}\left\| \frac{\partial^i}{\partial\theta_{r_1} \dots \partial\theta_{r_i}} \left[ \underline{\varphi}_n(\tau; f_\theta) - \underline{\phi}_n(\tau; f_\theta) \right] \right\|_2 \leqslant\ & f_0 \Bigg( \sum_{\substack{a_1+a_2=i \\ a_2 \neq i}} C_{a_1} \binom{i}{a_1} \left\| \nabla_\theta^{a_2} [\underline{\varphi}_n(\tau; f_\theta) - \underline{\phi}_n(\tau; f_\theta)] \right\|_2 \\ & + \sum_{b_1+b_2=i} C_{b_1} \binom{i}{b_1} \sum_{j=n+1}^{\infty} \left\| \nabla_\theta^{b_2} \phi_j(\tau; f_\theta) \right\|_1 \Bigg),\end{aligned}$$

*where $f_0 = (\inf_\omega f_\theta(\omega))^{-1}$ and $C_a = \sum_r \|\nabla_\theta^a c_\theta(r)\|_1$, $\nabla_\theta^a g(f_\theta)$ is the $a$th order partial derivative of $g$ with respect to $\theta = (\theta_1, \dots, \theta_d)$ and $\|\nabla_\theta^a g(f_\theta)\|_p$ denotes the $\ell_p-$norm of the matrix with elements containing all the partial derivatives in $\nabla_\theta^a g(f_\theta)$.*

PROOF. To prove the result, we define the $n$-dimension vector

$$\underline{c}_{n,\tau} = (c(\tau-1), c(\tau-2), \dots, c(\tau-n))' \quad \text{(covariances from lag } \tau-1 \text{ to lag } \tau-n).$$

To simplify notation we drop the $f_\theta$ notation from the prediction coefficients $\phi_{j,n}(\tau; f_\theta)$ and $\phi_j(\tau; f_\theta)$.

Proof for the case $i = 0$ This is the regular Baxter inequality but with the $\ell_2$-norm rather than $\ell_1$-norm. We recall that for $\tau \leqslant 0$ we have the best linear predictors

$$\widehat{X}_{\tau,n} = \sum_{j=1}^{n} \phi_{j,n}(\tau) X_j \quad \text{and} \quad X_\tau = \sum_{j=1}^{\infty} \phi_j(\tau) X_j.$$

Thus by evaluating the covariance of the above with $X_r$ for all $1 \leqslant r \leqslant n$ gives the sequence of $r$ normal equation, which can be written in matrix form

$$\Gamma_n(f_\theta) \underline{\varphi}_n(\tau) = \underline{c}_{n,\tau} \quad \text{and} \quad \Gamma_n(f_\theta) \underline{\phi}_n(\tau) + \sum_{j=n+1}^{\infty} \phi_j(\tau) \underline{c}_{n,j} = \underline{c}_{n,\tau}.$$

Taking differences of the above gives

$$\begin{aligned} \Gamma_n(f_\theta)\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] &= \sum_{j=n+1}^{\infty} \phi_j(\tau)\underline{c}_{n,j} \\ \Rightarrow \quad \left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] &= \Gamma_n(f_\theta)^{-1} \sum_{j=n+1}^{\infty} \phi_j(\tau)\underline{c}_{n,j} \end{aligned} \tag{A.18}$$

The $\ell_2$-norm of the above gives

$$\left\|\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right\|_2 \leqslant \|\Gamma_n(f_\theta)^{-1}\|_{spec} \sum_{j=n+1}^{\infty} |\phi_j(\tau)| \cdot \|\underline{c}_{n,j}\|_2.$$

To bound the above we use the well known result $\|\Gamma_n(f_\theta)^{-1}\|_{spec} \leqslant 1/\inf_\omega f(\omega) = f_0$ and $\|\underline{c}_{n,j}\|_2 \leqslant \sum_{r\in\mathbb{Z}} |c_\theta(r)| = C_0$. This gives the bound

$$\left\|\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right\|_2 \leqslant f_0 C_0 \sum_{j=n+1}^{\infty} |\phi_j(\tau)|.$$

Proof for the case $i = 1$. As our aim is to bound the derivative of the difference $\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)$, we evaluate the partial derivative of (A.18) with respective to $\theta_r$ and isolate $\partial[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)]/\partial\theta_r$. Differentiating both sides of (A.18) with respect to $\theta_r$ gives

$$\begin{aligned} &\frac{\partial \Gamma_n(f_\theta)}{\partial \theta_r}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] + \Gamma_n(f_\theta)\frac{\partial}{\partial \theta_r}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] \\ &= \sum_{j=n+1}^{\infty}\left[\frac{\partial \phi_j(\tau)}{\partial \theta_r}\underline{c}_{n,j} + \phi_j(\tau)\frac{\partial \underline{c}_{n,j}}{\partial \theta_r}\right]. \end{aligned} \tag{A.19}$$

Isolating $\partial[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)]/\partial\theta_r$ gives

$$\begin{aligned} \frac{\partial}{\partial \theta_r}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] &= -\Gamma_n(f_\theta)^{-1}\frac{\partial \Gamma_n(f_\theta)}{\partial \theta_r}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] \\ &\quad +\Gamma_n(f_\theta)^{-1}\sum_{j=n+1}^{\infty}\left[\frac{\partial \phi_j(\tau)}{\partial \theta_r}\underline{c}_{n,j} + \phi_j(\tau)\frac{\partial \underline{c}_{n,j}}{\partial \theta_r}\right]. \end{aligned} \tag{A.20}$$

Evaluating the $\ell_2$ norm of the above and using $\|AB\underline{x}\|_2 \leqslant \|A\|_{spec}\|B\|_{spec}\|\underline{x}\|_2$ gives the bound

$$\begin{aligned}
&\left\| \frac{\partial}{\partial \theta_r} \left[ \underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right] \right\|_2 \\
&\leqslant \|\Gamma_n(f_\theta)^{-1}\|_{spec} \Bigg( \left\| \frac{\partial \Gamma_n(f_\theta)}{\partial \theta_r} \right\|_{spec} \left\| \underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right\|_2 \\
&\qquad + \sum_{j=n+1}^{\infty} \left| \frac{\partial \phi_j(\tau)}{\partial \theta_r} \right| \|\underline{c}_{n,j}\|_2 + \sum_{j=n+1}^{\infty} |\phi_j(\tau)| \left\| \frac{\partial \underline{c}_{n,j}}{\partial \theta_r} \right\|_2 \Bigg) \\
&\leqslant f_0 \Bigg( \left\| \frac{\partial \Gamma_n(f_\theta)}{\partial \theta_r} \right\|_{spec} \left\| \underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right\|_2 \\
&\qquad + C_0 \sum_{j=n+1}^{\infty} \left| \frac{\partial \phi_j(\tau)}{\partial \theta_r} \right| + \left( \sum_{r \in \mathbb{Z}} \|\nabla_\theta c_\theta(r)\|_2 \right) \sum_{j=n+1}^{\infty} |\phi_j(\tau)| \Bigg) \\
&\leqslant f_0 \Bigg( \left\| \frac{\partial \Gamma_n(f_\theta)}{\partial \theta_r} \right\|_{spec} \left\| \underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right\|_2 + C_0 \sum_{j=n+1}^{\infty} \left| \frac{\partial \phi_j(\tau)}{\partial \theta_r} \right| + C_1 \sum_{j=n+1}^{\infty} |\phi_j(\tau)| \Bigg) \qquad \text{(A.21)}
\end{aligned}$$

where the last line in the above uses the bound $\sum_{r \in \mathbb{Z}} \|\nabla_\theta^a c_\theta(r)\|_2 \leqslant \sum_{r \in \mathbb{Z}} \|\nabla_\theta^a c_\theta(r)\|_1 = C_a$ (for $a = 0$ and 1). We require a bound for $\|\partial \Gamma_n(f_\theta)/\partial \theta_r\|_{spec}$. Since $\Gamma_n(f_\theta)$ is a symmetric Toeplitz matrix, then $\partial \Gamma_n(f_\theta)/\partial \theta_r$ is also a symmetric Toeplitz matrix (though not necessarily positive definite) with entries

$$\left[ \frac{\partial \Gamma_n(f_\theta)}{\partial \theta_r} \right]_{s,t} = \frac{1}{2\pi} \int_0^{2\pi} \frac{\partial f_\theta(\omega)}{\partial \theta_r} \exp(i(s-t)\omega) d\omega.$$

We mention that the symmetry is clear, since $\frac{\partial c(s-t;f_\theta)}{\partial \theta_r} = \frac{\partial c(t-s;f_\theta)}{\partial \theta_r}$. Since the matrix is symmetric the spectral norm is the spectral radius. This gives

$$\begin{aligned}
\left\| \frac{\partial \Gamma_n(f_\theta)}{\partial \theta_r} \right\|_{spec} &= \sup_{\|x\|_2=1} \left| \sum_{s,t=1}^{n} x_s x_t \frac{\partial c(s-t;f_\theta)}{\partial \theta_r} \right| = \sup_{\|x\|_2=1} \left| \frac{1}{2\pi} \int_0^{2\pi} |\sum_{s=1}^{n} x_s e^{is\omega}|^2 \frac{\partial f_\theta(\omega)}{\partial \theta_r} d\omega \right| \\
&\leqslant \sup_{\omega} \left| \frac{\partial f_\theta(\omega)}{\partial \theta_r} \right| \sup_{\|x\|_2=1} \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{s=1}^{n} x_s e^{is\omega} \right|^2 d\omega = \sup_{\omega} \left| \frac{\partial f_\theta(\omega)}{\partial \theta_r} \right|.
\end{aligned}$$

By using the same argument one can show that the $a$th derivative is

$$\left\| \frac{\partial^a \Gamma_n(f_\theta)}{\partial\theta_{r_1}\dots\partial\theta_{r_a}} \right\|_{spec} \leqslant \sup_\omega \left| \frac{\partial^a f_\theta(\omega)}{\partial\theta_{r_1}\dots\partial\theta_{r_a}} \right| = \sum_{r\in\mathbb{Z}} \left| \frac{\partial^a c(r;f_\theta)}{\partial\theta_{r_1}\dots\partial\theta_{r_a}} \right| \leqslant C_a. \tag{A.22}$$

This general bound will be useful when evaluating the higher order derivatives below. Substituting (A.22) into (A.21) gives

$$\left\| \frac{\partial}{\partial\theta_r}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] \right\|_2 \leqslant f_0\left( C_1 \left\| \underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right\|_2 + C_1 \sum_{j=n+1}^{\infty} |\phi_j(\tau)| + C_0 \sum_{j=n+1}^{\infty} \|\nabla_\theta \phi_j(\tau)\|_1 \right).$$

This proves the result for $i = 1$.

Proof for the case $i = 2$ We differentiate both sides of (A.19) with respect to $\theta_{r_2}$ to give the second derivative

$$\begin{aligned}
&\left( \frac{\partial^2 \Gamma_n(f_\theta)}{\partial\theta_{r_1}\partial\theta_{r_2}} \right)\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] + \frac{\partial\Gamma_n(f_\theta)}{\partial\theta_{r_2}}\frac{\partial}{\partial\theta_{r_1}}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] + \\
&\frac{\partial\Gamma_n(f_\theta)}{\partial\theta_{r_1}}\frac{\partial}{\partial\theta_{r_2}}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] + \Gamma_n(f_\theta)\left( \frac{\partial^2}{\partial\theta_{r_2}\partial\theta_{r_1}}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] \right) \\
&= \sum_{j=n+1}^{\infty} \left[ \frac{\partial^2\phi_j(\tau)}{\partial\theta_{r_2}\partial\theta_{r_1}}\underline{c}_{n,j} + \frac{\partial\phi_j(\tau)}{\partial\theta_{r_1}}\frac{\partial\underline{c}_{n,j}}{\partial\theta_{r_2}} + \frac{\partial\phi_j(\tau)}{\partial\theta_{r_2}}\frac{\partial\underline{c}_{n,j}}{\partial\theta_{r_1}} + \phi_j(\tau)\frac{\partial^2\underline{c}_{n,j}}{\partial\theta_{r_2}\partial\theta_{r_1}} \right].
\end{aligned} \tag{A.23}$$

Rearranging the above to isolate $\frac{\partial^2}{\partial\theta_{r_2}\partial\theta_{r_1}}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right]$ gives

$$\begin{aligned}
&\frac{\partial^2}{\partial\theta_{r_1}\partial\theta_{r_2}}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] \\
&= -\Gamma_n(f_\theta)^{-1}\left( \frac{\partial^2\Gamma_n(f_\theta)}{\partial\theta_{r_1}\partial\theta_{r_2}} \right)\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] - \Gamma_n(f_\theta)^{-1}\frac{\partial\Gamma_n(f_\theta)}{\partial\theta_{r_2}}\frac{\partial}{\partial\theta_{r_1}}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] \\
&\quad -\Gamma_n(f_\theta)^{-1}\frac{\partial\Gamma_n(f_\theta)}{\partial\theta_{r_1}}\frac{\partial}{\partial\theta_{r_2}}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] \\
&\quad +\Gamma_n(f_\theta)^{-1}\sum_{j=n+1}^{\infty}\left[ \frac{\partial^2\phi_j(\tau)}{\partial\theta_{r_2}\partial\theta_{r_1}}\underline{c}_{n,j} + \frac{\partial\phi_j(\tau)}{\partial\theta_{r_1}}\frac{\partial\underline{c}_{n,j}}{\partial\theta_{r_2}} + \frac{\partial\phi_j(\tau)}{\partial\theta_{r_2}}\frac{\partial\underline{c}_{n,j}}{\partial\theta_{r_1}} + \phi_j(\tau)\frac{\partial^2\underline{c}_{n,j}}{\partial\theta_{r_2}\partial\theta_{r_1}} \right].
\end{aligned}$$

Taking the $\ell_2$-norm of $\frac{\partial^2}{\partial\theta_{r_1}\partial\theta_{r_2}}\Big[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\Big]$ and using (A.22) gives

$$\begin{aligned}
&\left\|\frac{\partial^2}{\partial\theta_{r_1}\partial\theta_{r_2}}\Big[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\Big]\right\|_2 \\
&\leqslant f_0\bigg(C_2\left\|\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right\|_2 + 2C_1\left\|\nabla_\theta[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)]\right\|_2 \\
&\quad + C_0\sum_{j=n+1}^{\infty}\left\|\nabla^2\phi_j(\tau)\right\|_1 + 2C_1\sum_{j=n+1}^{\infty}\left\|\nabla\phi_j(\tau)\right\|_1 + C_2\sum_{j=n+1}^{\infty}|\phi_j(\tau)|\bigg).
\end{aligned}$$

This proves the result for $i = 2$. The proof for $i > 2$ follows using a similar argument (we omit the details). □

The above result gives an $\ell_2$-bound between the derivatives of the finite and infinite predictors. However, for our purposes an $\ell_1$-bound is more useful. Thus we use the Cauchy-Schwarz inequality and norm inequality $\|\cdot\|_2 \leqslant \|\cdot\|_1$ to give the $\ell_1$-bound

$$\begin{aligned}
&\left\|\frac{\partial^i}{\partial\theta_{r_1}\ldots\partial\theta_{r_i}}\Big[\underline{\varphi}_n(\tau; f_\theta) - \underline{\phi}_n(\tau; f_\theta)\Big]\right\|_1 \\
&\quad\leqslant n^{1/2}f_0\bigg(\sum_{\substack{a_1+a_2=i\\ a_2\neq i}}\binom{i}{a_1}C_{a_1}\left\|\nabla_\theta^{a_2}[\underline{\varphi}_n(\tau; f_\theta) - \underline{\phi}_n(\tau; f_\theta)]\right\|_1 + \\
&\qquad\sum_{b_1+b_2=i}\binom{i}{b_1}C_{b_1}\sum_{j=n+1}^{\infty}\left\|\nabla_\theta^{b_2}\phi_j(\tau; f_\theta)\right\|_1\bigg),
\end{aligned} \tag{A.24}$$

this incurs an additional $n^{1/2}$ term. Next, considering all the partial derivatives with respect to $\theta$ of order $i$ and using (A.16) we have

$$\begin{aligned}
&\sum_{t=1}^{n}\|\nabla_\theta^i[\phi_{t,n}(\tau; f_\theta) - \phi_t(\tau; f_\theta)]\|_1 \\
&\quad\leqslant d^i n^{1/2}f_0\bigg(\sum_{\substack{a_1+a_2=i\\ a_2\neq i}}\binom{i}{a_1}C_{a_1}\left\|\nabla_\theta^{a_2}[\underline{\varphi}_n(\tau; f_\theta) - \underline{\phi}_n(\tau; f_\theta)]\right\|_1 + \\
&\qquad\sum_{b_1+b_2=i}\binom{i}{b_1}C_{b_1}\sum_{j=n+1}^{\infty}\left\|\nabla_\theta^{b_2}\phi_j(\tau; f_\theta)\right\|_1\bigg),
\end{aligned} \tag{A.25}$$

where $d$ is the dimension of the vector $\theta$. The above gives a bound in terms of the infinite predictors. We now obtain a bound in terms of the corresponding AR($\infty$) and MA($\infty$) coefficients. To do this, we recall that for $\tau \leqslant 0$, $\phi_j(\tau; f_\theta) = \sum_{s=0}^{\infty} \phi_{s+j}(f_\theta)\psi_{|\tau|-j}(f_\theta)$. Thus the partial derivatives of $\phi_j(\tau; f_\theta)$ give the bound

$$\sum_{j=n+1}^{\infty} \|\nabla_\theta \phi_j(\tau; f_\theta)\|_1 \leqslant \sum_{s=0}^{\infty} \sum_{j=n+1}^{\infty} \left( |\psi_{|\tau|-j}(f_\theta)| \cdot \|\nabla_\theta \phi_{s+j}(f_\theta)\|_1 + |\phi_{s+j}(f_\theta)| \cdot \|\nabla_\theta \psi_{|\tau|-j}(f_\theta)\|_1 \right).$$

Substituting the above bound into (A.25) and using Lemma A.1.1 for the case $i = 1$ gives

$$\begin{aligned} & \sum_{s=1}^{n} \|\nabla_\theta[\phi_{s,n}(\tau; f_\theta) - \phi_s(\tau; f_\theta)]\|_1 \\ \leqslant \; & n^{1/2} d f_0 \Bigg\{ C_1(C_0+1) \sum_{j=n+1}^{\infty} |\phi_j(\tau; f_\theta)| + \\ & C_0 \sum_{s=0}^{\infty} \sum_{j=n+1}^{\infty} \left( |\psi_{|\tau|-j}(f_\theta)| \cdot \|\nabla_\theta \phi_{s+j}(f_\theta)\|_1 + |\phi_{s+j}(f_\theta)| \cdot \|\nabla_\theta \psi_{|\tau|-j}(f_\theta)\|_1 \right) \Bigg\}. \quad \text{(A.26)} \end{aligned}$$

The above results are used to obtain bounds between the derivatives of the Whittle and Gaussian likelihood in Appendix A.3. Similar bounds can also be obtained for the higher order derivatives $\sum_{s=1}^{n} \|\nabla_\theta^i[\phi_{s,n}(\tau; f_\theta) - \phi_s(\tau; f_\theta)]\|_1$ in terms of the derivatives of the MA($\infty$) and AR($\infty$) coefficients.

### A.3 The difference between the derivatives of the Gaussian and Whittle likelihoods

We now obtain an expression for the difference between the derivatives of the Gaussian likelihood and the Whittle likelihood using the variant of Baxter's inequality. These expression will be used later for obtaining the bias of the Gaussian likelihood (as compared with the Whittle likelihood).

For the Gaussian likelihood, we have shown in Theorem 2.4.1 that

$$\underline{X}_n' \Gamma_n(\theta)^{-1} \underline{X}_n = \underline{X}_n' F_n^* \Delta_n(f_\theta^{-1}) F_n \underline{X}_n + \underline{X}_n' F_n^* \Delta_n(f_\theta^{-1}) D_n(f_\theta) \underline{X}_n,$$

where the first term is the Whittle likelihood and the second term the additional term due to the Gaussian likelihood. Clearly the derivative with respect to $\theta' = (\theta_1, \ldots, \theta_d)$ is

$$\underline{X}_n' \nabla_\theta^i \Gamma_n(\theta)^{-1} \underline{X}_n = \underline{X}_n' F_n^* \nabla_\theta^i \Delta_n(f_\theta^{-1}) F_n \underline{X}_n + \underline{X}_n' F_n^* \nabla_\theta^i [\Delta_n(f_\theta^{-1}) D_n(f_\theta)] \underline{X}_n.$$

The first term on the right hand side is the derivative of the Whittle likelihood with respect to $\theta$, the second term is the additional term due to the Gaussian likelihood.

For the simplicity, assume $\theta$ is univariate. Our objective in the next few lemmas is to show that

$$\left\| \underline{X}_n' F_n^* \frac{d^i}{d\theta^i} \Delta_n(f_\theta^{-1}) D_n(f_\theta) \underline{X}_n \right\|_1 = O(1),$$

which is a result analogous to Theorem 2.4.4, but for the derivatives. We will use this result to prove Theorem B.1.1, in particular to show the derivatives of the Whittle likelihood and the Gaussian likelihood (after normalization by $n^{-1}$) differ by $O(n^{-1})$.

Just as in the proof of Theorem 2.4.4, the derivative of this term with respect to $\theta$ does not (usually) have a simple analytic form. Therefore, analogous to Theorem 2.4.3 it is easier to replace the derivatives of $D_n(f_\theta)$ with the derivatives of $D_{\infty,n}(f_\theta)$, and show that the replacement error is "small".

**Lemma A.3.1.** *Suppose Assumption 4.2.1(i),(iii) holds and $g$ is a bounded function. Then for $1 \leqslant i \leqslant 3$ we have*

$$\left\| F_n^* \Delta_n(g) \frac{d^i}{d\theta^i} \left( D_n(f_\theta) - D_{\infty,n}(f_\theta) \right) \right\|_1 = O(n^{-K+3/2}), \tag{A.27}$$

*and*

$$\left\| F_n^* \sum_{k=0}^{i} \binom{i}{k} \frac{d^k \Delta_n \left( f_\theta^{-1} \right)}{d\theta^k} \frac{d^{k-i} D_{\infty,n}(f_\theta)}{d\theta^{k-i}} \right\|_1 = O(1). \tag{A.28}$$

PROOF. To bound (A.27), we use the expression for $F_n^* \Delta_n(g^{-1}) \left( D_n(f_\theta) - D_{\infty,n}(f_\theta) \right)$ given in

(4.34)

$$\left(F_n^*\Delta_n(g^{-1})\left[D_n(f_\theta) - D_{\infty,n}(f_\theta)\right]\right)_{s,t} = \sum_{\tau\leqslant 0}[\{\phi_{t,n}(\tau;f_\theta) - \phi_t(\tau;f_\theta)\}\, G_{1,n}(s,\tau;g) + \{\phi_{n+1-t,n}(\tau;f_\theta) - \phi_{n+1-t}(\tau;f_\theta)\}\, G_{2,n}(s,\tau;g)].$$

Differentiating the above with respect to $\theta$ gives

$$\begin{aligned}&\left[F_n^*\Delta_n(g)\frac{d}{d\theta}\left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right)\right]_{s,t}\\ &= \sum_{\tau\leqslant 0}\left[G_{1,n}(s,\tau)\frac{d}{d\theta}[\phi_{t,n}(\tau) - \phi_t(\tau)] + G_{2,n}(s,\tau)\frac{d}{d\theta}\left[\phi_{n+1-t}(\tau) - \phi_{n+1-t}(\tau)\right]\right]\\ &= T_{s,t,1} + T_{s,t,2}.\end{aligned}$$

We recall that equation (A.26) gives the bound

$$\begin{aligned}&\sum_{s=1}^{n}\left|\frac{d}{d\theta}[\phi_{s,n}(\tau;f_\theta) - \phi_s(\tau;f_\theta)]\right|\\ \leqslant\ & n^{1/2}f_0\Bigg\{C_1(C_{f,0}+1)\sum_{j=n+1}^{\infty}|\phi_j(\tau;f_\theta)| +\\ & C_0\sum_{s=0}^{\infty}\sum_{j=n+1}^{\infty}\left(|\psi_{|\tau|-j}(f_\theta)|\cdot\left|\frac{d}{d\theta}\phi_{s+j}(f_\theta)\right| + |\phi_{s+j}(f_\theta)|\cdot\left|\frac{d}{d\theta}\nabla_\theta\psi_{|\tau|-j}(f_\theta)\right|\right)\Bigg\}.\end{aligned}$$

Substituting this into $T_{s,t,1}$ gives the bound

$$|T_{s,t,1}| \leqslant Cn^{1/2}\sum_{\tau\leqslant 0}G_{1,n}(s,\tau)\left(\sum_{j=n+1}^{\infty}\sum_{s=0}^{\infty}|\phi_{s+j}||\psi_{|\tau|-j}| + \sum_{s=0}^{\infty}\sum_{j=n+1}^{\infty}\left|\frac{d\phi_{s+j}}{d\theta}\psi_{|\tau|-j} + \phi_{s+j}\frac{d\psi_{|\tau|-j}}{d\theta}\right|\right).$$

Using the same techniques used to prove Theorem 2.4.3 yields

$$\sum_{s,t=1}^{n}|T_{s,t,1}| = O\left(n^{1/2}n^{-K+1}\right) = O(n^{-K+3/2}).$$

Similarly, we can show that $\sum_{s,t=1}^{n} |T_{s,t,2}| = O\left(n^{1/2}n^{-K+1}\right) = O(n^{-K+3/2})$. Altogether this gives

$$\|F_n^* \Delta_n(g) \left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right)\|_1 = O(n^{-K+3/2}).$$

This proves (A.27) for the case $i = 1$. The proof for the cases $i = 2, 3$ is similar.

To prove (A.28) we use the same method used to prove Theorem 2.4.3, equation (2.40). But with $\frac{d^k f_\theta^{-1}}{d\theta^k}$ replacing $f_\theta$ in $\Delta_n(\cdot)$ and $\frac{d^{i-k}}{d\theta^k}\phi_j(\tau; f_\theta) = \frac{d^{i-k}}{d\theta^k}\sum_{s=0}^{\infty} \phi_{s+j}\psi_{|\tau|-j}$ replacing $\phi_j(\tau; f_\theta) = \sum_{s=0}^{\infty} \phi_{s+j}\psi_{|\tau|-j}$ in $D_n(f_\theta)$. We omit the details. □

We now apply the above results to quadratic forms of random variables.

**Corollary A.3.1.** *Suppose Assumptions 4.2.1 (i),(iii) hold and $g$ is a bounded function. Further, if $\{X_t\}$ is a time series where $\sup_t \|X_t\|_{\mathbb{E},2q} = \|X\|_{\mathbb{E},2q} < \infty$ (for some $q > 1$), then*

$$\left\| \frac{1}{n} \underline{X}_n' \left[ F_n^* \Delta_n(g) \frac{d^i}{d\theta^i} \left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right) \right] \underline{X}_n \right\|_{\mathbb{E},q} = O(n^{-K+1/2}), \tag{A.29}$$

*and*

$$\left\| \frac{1}{n} \underline{X}_n' F_n^* \sum_{\ell=0}^{i} \binom{i}{\ell} \frac{d^\ell \Delta_n\left(f_\theta^{-1}\right)}{d\theta^\ell} \frac{d^{\ell-i} D_{\infty,n}(f_\theta)}{d\theta^{\ell-i}} \underline{X}_n \right\|_{\mathbb{E},q} = O(n^{-1}) \tag{A.30}$$

*for $i = 1, 2$ and 3.*

PROOF. To prove (A.29), we observe that

$$\begin{aligned}
&\left\| \frac{1}{n} \underline{X}_n' \left[ F_n^* \Delta_n(g) \frac{d^i}{d\theta^i} \left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right) \right] \underline{X}_n \right\|_{\mathbb{E},q} \\
&\quad \leqslant \frac{1}{n} \sum_{s,t=1}^{n} \left| \left[ F_n^* \Delta_n(g) \frac{d^i}{d\theta^i} \left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right) \right]_{s,t} \right| \|X_s X_t\|_{\mathbb{E},q} \\
&\quad = \frac{1}{n} \sup_t \|X_t\|_{\mathbb{E},2q}^2 \sum_{s,t=1}^{n} \left| \left[ F_n^* \Delta_n(g) \frac{d^i}{d\theta^i} \left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right) \right]_{s,t} \right| \\
&\quad = \frac{1}{n} \|X\|_{\mathbb{E},2q}^2 \left\| F_n^* \Delta_n(g) \frac{d^i}{d\theta^i} \left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right) \right\|_1 = O(n^{-K+1/2})
\end{aligned}$$

where the above follows from Lemma A.3.1, equation (A.27). This proves (A.29).

To prove (A.30) we use the the bound in (A.28) together with a similar proof to that described above. This immediately proves (A.30). □

We now apply the above result to the difference in the derivatives of the Gaussian and Whittle likelihood. It is straightforward to show that

$$\begin{aligned}
&\underline{X}_n' F_n^* \frac{d^i}{d\theta^i} \Delta_n(f_\theta^{-1}) D_n(f_\theta) \underline{X}_n \qquad (A.31)\\
&= \underline{X}_n' F_n^* \left[ \sum_{\ell=0}^{i} \binom{i}{\ell} \frac{d^\ell \Delta_n\left(f_\theta^{-1}\right)}{d\theta^\ell} \frac{d^{\ell-i} D_n(f_\theta)}{d\theta^{\ell-i}} \right] \underline{X}_n \\
&= \underline{X}_n' F_n^* \left[ \sum_{\ell=0}^{i} \binom{i}{\ell} \frac{d^\ell \Delta_n\left(f_\theta^{-1}\right)}{d\theta^\ell} \frac{d^{\ell-i} D_{\infty,n}(f_\theta)}{d\theta^{\ell-i}} \right] \underline{X}_n \\
&+ \underline{X}_n' F_n^* \left( \sum_{\ell=0}^{i} \binom{i}{\ell} \frac{d^\ell \Delta_n\left(f_\theta^{-1}\right)}{d\theta^\ell} \left[ \frac{d^{\ell-i} D_n(f_\theta)}{d\theta^{\ell-i}} - \frac{d^{\ell-i} D_{\infty,n}(f_\theta)}{d\theta^{\ell-i}} \right] \right) \underline{X}_n. \qquad (A.32)
\end{aligned}$$

First we study the second term on the right hand side of the above. By applying Corollary A.3.1 (and under Assumption 4.2.1) for $1 \leqslant i \leqslant 3$ we have

$$\begin{aligned}
&\left\| n^{-1} \underline{X}_n' F_n^* \left( \frac{d^i}{d\theta^i} \Delta_n(f_\theta^{-1}) \left[ D_n(f_\theta) - D_{\infty,n}(f_\theta) \right] \right) \underline{X}_n \right\|_{\mathbb{E},1} \\
&= \left\| n^{-1} \underline{X}_n' F_n^* \sum_{\ell=0}^{i} \binom{i}{\ell} \frac{d^\ell \Delta_n\left(f_\theta^{-1}\right)}{d\theta^\ell} \left[ \frac{d^{\ell-i} D_n(f_\theta)}{d\theta^{\ell-i}} - \frac{d^{\ell-i} D_{\infty,n}(f_\theta)}{d\theta^{\ell-i}} \right] \underline{X}_n \right\|_{\mathbb{E},1} \\
&= O(n^{-K+1/2}). \qquad (A.33)
\end{aligned}$$

On the other hand, the first term on the right hand side of (A.31) has the bound

$$\left\| n^{-1} \underline{X}_n' F_n^* \frac{d^i}{d\theta^i} \left[ \Delta_n(f_\theta^{-1}) D_n(f_\theta) \right] \underline{X}_n \right\|_{\mathbb{E},1} = O(n^{-1}). \qquad (A.34)$$

APPENDIX B

THE BIAS OF THE DIFFERENT CRITERIA *

In this section, we derive the approximate bias of the Gaussian, Whittle, boundary corrected and hybrid Whittle likelihoods under quite general assumptions on the underlying time series $\{X_t\}$. The bias we evaluate will be in the sense of Bartlett (1953) and will be based on the second order expansion of the loss function. We mention that for certain specific models (such as the specified AR, or certain MA or ARMA) the bias of the least squares, Whittle likelihood or maximum likeihood estimators are given in Taniguchi (1983); Tanaka (1984); Shaman and Stine (1988).

## B.1 Bias for the estimator of one unknown parameter

In order to derive the limiting bias, we require the following definitions

$$I(\theta) = -\frac{1}{2\pi}\int_0^{2\pi}\left(\frac{d^2 f_\theta(\omega)^{-1}}{d\theta^2}\right) f(\omega)d\omega \quad \text{and} \quad J(g) = \frac{1}{2\pi}\int_0^{2\pi} g(\omega)f(\omega)d\omega.$$

For real functions $g, h \in L^2[0, 2\pi]$ we define

$$\begin{aligned} V(g,h) &= \frac{2}{2\pi}\int_0^{2\pi} g(\omega)h(\omega)f(\omega)^2 d\omega \\ &\quad + \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} g(\omega_1)h(\omega_2)f_4(\omega_1, -\omega_1, \omega_2)d\omega_1 d\omega_2, \end{aligned} \tag{B.1}$$

where $f_4$ denotes the fourth order cumulant density of the time series $\{X_t\}$. Further, we define

$$\begin{aligned} B_{G,n}(\theta) &= \mathrm{Re}\,\frac{2}{n}\sum_{s,t=1}^{n} c(s-t)\frac{1}{n}\sum_{k=1}^{n} e^{-is\omega_{k,n}}\frac{d}{d\theta}\left[\overline{\phi(\omega_{k,n}; f_\theta)}\phi_t^\infty(\omega_{k,n}; f_\theta)\right] \\ B_{K,n}(\theta) &= \frac{1}{n}\sum_{k=1}^{n} f_n(\omega_{k,n})\frac{df_\theta(\omega_{k,n})^{-1}}{d\theta} \end{aligned}$$

*Parts of this section have been modified with permission from [S. Subba Rao and J. Yang. Reconciling the Gaussian and Whittle likelihood with an application to estimation in the frequency domain. *Annals of Statistics (To appear), arXiv:2001.06966*, 2021.]

where $f_n(\omega_k) = \int F_n(\omega - \lambda) f(\lambda) d\lambda$ and $F_n(\cdot)$ is the Fejér kernel of order $n$.

**Theorem B.1.1.** *Suppose that the parametric spectral densities $\{f_\theta; \theta \in \Theta\}$ satisfy Assumptions 4.2.1. Suppose the underlying time series $\{X_t\}$ is a stationary time series with spectral density $f$ and satisfies Assumption 4.2.2. Let $\widehat{\theta}_n^{(G)}$, $\widehat{\theta}_n^{(K)}$ and $\widehat{\theta}_n^{(W)}$, and $\widehat{\theta}_n^{(H)}$ be defined as in (4.17). Then the asymptotic bias is*

$$\begin{aligned}
\mathbb{E}_\theta[\widehat{\theta}_n^{(G)} - \theta_n] &= I(\theta)^{-1} \left(B_{K,n}(\theta_n) + B_{G,n}(\theta_n)\right) + n^{-1} G(\theta_n) + O(n^{-3/2}) \\
\mathbb{E}_\theta[\widehat{\theta}_n^{(K)} - \theta_n] &= I(\theta)^{-1} B_{K,n}(\theta_n) + n^{-1} G(\theta_n) + O(n^{-3/2}) \\
\mathbb{E}_\theta[\widehat{\theta}_n^{(W)} - \theta_n] &= n^{-1} G(\theta_n) + O\left(p^3 n^{-3/2} + n^{-1} p^{-K+1}\right) \\
\textit{and} \qquad \mathbb{E}_\theta[\widehat{\theta}_n^{(H)} - \theta_n] &= \frac{H_{2,n}}{H_{1,n}^2} G(\theta_n) + O\left(p^3 n^{-3/2} + n^{-1} p^{-K+1}\right)
\end{aligned}$$

*where $H_{q,n} = \sum_{t=1}^n h_n(t/n)^q$,*

$$G(\theta) = I(\theta)^{-2} V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{d^2 f_\theta^{-1}}{d\theta^2}\right) + 2^{-1} I(\theta)^{-3} V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{df_\theta^{-1}}{d\theta}\right) J\left(\frac{d^3 f_\theta^{-1}}{d\theta^3}\right),$$

*and $V(g,h)$ is defined in (B.1).*

PROOF. In Theorem 4.3.1 we showed that

$$|\widehat{\theta}_n^{(W)} - \widetilde{\theta}_n^{(W)}| = O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right) \quad \text{and} \quad |\widehat{\theta}_n^{(H)} - \widetilde{\theta}_n^{(H)}| = O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right),$$

where $\widehat{\theta}_n^{(W)} = \arg\min \widehat{W}_{p,n}(\theta)$, $\widetilde{\theta}_n^{(W)} = \arg\min W_n(\theta)$, $\widehat{\theta}_n^{(H)} = \arg\min \widehat{H}_{p,n}(\theta)$, and $\widetilde{\theta}_n^{(H)} = \arg\min H_n(\theta)$. We will show that the asymptotic bias of $\widetilde{\theta}_n^{(W)}$ and $\widetilde{\theta}_n^{(H)}$ (under certain conditions on the taper) are of order $O(n^{-1})$, thus if $p^3 n^{-1/2} \to 0$ as $n, p \to \infty$, then the infeasible estimators and feasible estimators share the same asymptotic bias. Therefore in the proof we obtain the bias of the infeasible estimators.

Now we obtain a general expansion (analogous to the Bartlett correction). Let $L_n(\cdot)$ denote the

general minimization criterion (it can be $\mathcal{L}_n(\theta)$, $K_n(\theta)$, $W_n(\theta)$, or $H_n(\theta)$) and $\widehat{\theta} = \arg\min L_n(\theta)$. For all the criteria, it is easily shown that

$$(\widehat{\theta} - \theta) = U(\theta)^{-1} \frac{dL_n(\theta)}{d\theta} + O_p(n^{-1})$$

where $U(\theta) = -\mathbb{E}[\frac{d^2 L_n}{d\theta^2}]$ and

$$\frac{dL_n(\theta)}{d\theta} + (\widehat{\theta} - \theta)\frac{d^2 L_n(\theta)}{d\theta^2} + \frac{1}{2}(\widehat{\theta} - \theta)^2 \frac{d^3 L_n(\theta)}{d\theta^3} = O_p(n^{-3/2}).$$

Ignoring the probabilistic error, the first and second order expansions are

$$(\widehat{\theta} - \theta) \approx U(\theta)^{-1} \frac{dL_n(\theta)}{d\theta}, \tag{B.2}$$

and

$$\frac{dL_n(\theta)}{d\theta} + (\widehat{\theta} - \theta)\frac{d^2 L_n(\theta)}{d\theta^2} + \frac{1}{2}(\widehat{\theta} - \theta)^2 \frac{d^3 L_n(\theta)}{d\theta^3} \approx 0.$$

The method described below follows the Bartlett correction described in Bartlett (1953) and Cox and Snell (1968). Taking expectation of the above we have

$$\begin{aligned}
&\mathbb{E}\left[\frac{dL_n(\theta)}{d\theta}\right] + \mathbb{E}\left[(\widehat{\theta} - \theta)\frac{d^2 L_n(\theta)}{d\theta^2}\right] + \frac{1}{2}\mathbb{E}\left[(\widehat{\theta} - \theta)^2 \frac{d^3 L_n(\theta)}{d\theta^3}\right] \\
&= \mathbb{E}\left[\frac{dL_n(\theta)}{d\theta}\right] + \mathbb{E}\left[(\widehat{\theta} - \theta)\right]\mathbb{E}\left[\frac{d^2 L_n(\theta)}{d\theta^2}\right] + \operatorname{cov}\left[(\widehat{\theta} - \theta), \frac{d^2 L_n(\theta)}{d\theta^2}\right] \\
&\quad + \frac{1}{2}\mathbb{E}\left[(\widehat{\theta} - \theta)^2\right]\mathbb{E}\left[\frac{d^3 L_n(\theta)}{d\theta^3}\right] + \frac{1}{2}\operatorname{cov}\left[(\widehat{\theta} - \theta)^2, \frac{d^3 L_n(\theta)}{d\theta^3}\right].
\end{aligned}$$

Substituting $(\widehat{\theta} - \theta) \approx U(\theta)^{-1} \frac{dL_n(\theta)}{d\theta}$ into the last three terms on the right hand side of the above

gives

$$\begin{aligned}
&\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right) - U(\theta)\mathbb{E}(\widehat{\theta}-\theta) + U(\theta)^{-1}\mathrm{cov}\left(\frac{dL_n(\theta)}{d\theta}, \frac{d^2L_n(\theta)}{d\theta^2}\right) \\
&+2^{-1}U(\theta)^{-2}\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right)^2\mathbb{E}\left(\frac{d^3L_n(\theta)}{d\theta^3}\right) \\
&+2^{-1}U(\theta)^{-2}\mathrm{cov}\left(\left(\frac{dL_n(\theta)}{d\theta}\right)^2, \frac{d^3L_n(\theta)}{d\theta^3}\right) \approx 0.
\end{aligned}$$

Using the above to solve for $\mathbb{E}(\widehat{\theta}-\theta)$ gives

$$\begin{aligned}
\mathbb{E}(\widehat{\theta}-\theta) &= U(\theta)^{-1}\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right) + U(\theta)^{-2}\mathrm{cov}\left(\frac{dL_n(\theta)}{d\theta}, \frac{d^2L_n(\theta)}{d\theta^2}\right) \\
&\quad +2^{-1}U(\theta)^{-3}\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right)^2\mathbb{E}\left(\frac{d^3L_n(\theta)}{d\theta^3}\right) + 2^{-1}U(\theta)^{-3}\mathrm{cov}\left(\left(\frac{dL_n(\theta)}{d\theta}\right)^2, \frac{d^3L_n(\theta)}{d\theta^3}\right) \\
&= U(\theta)^{-1}\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right) + U(\theta)^{-2}\mathrm{cov}\left(\frac{dL_n(\theta)}{d\theta}, \frac{d^2L_n(\theta)}{d\theta^2}\right) \\
&\quad +2^{-1}U(\theta)^{-3}\left[\mathrm{var}\left(\frac{dL_n(\theta)}{d\theta}\right) + \left\{\mathbb{E}\left[\frac{dL_n(\theta)}{d\theta}\right]\right\}^2\right]\mathbb{E}\left(\frac{d^3L_n(\theta)}{d\theta^3}\right) \\
&\quad +2^{-1}U(\theta)^{-3}\mathrm{cov}\left(\left(\frac{dL_n(\theta)}{d\theta}\right)^2, \frac{d^3L_n(\theta)}{d\theta^3}\right).
\end{aligned}$$

Thus

$$\mathbb{E}(\widehat{\theta}-\theta) = I_0 + I_1 + I_2 + I_3 + I_4 \tag{B.3}$$

where

$$\begin{aligned}
I_0 &= U(\theta)^{-1}\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right) \\
I_1 &= U(\theta)^{-2}\mathrm{cov}\left(\frac{dL_n(\theta)}{d\theta}, \frac{d^2L_n(\theta)}{d\theta^2}\right) \\
I_2 &= 2^{-1}U(\theta)^{-3}\mathrm{var}\left(\frac{dL_n(\theta)}{d\theta}\right)\mathbb{E}\left(\frac{d^3L_n(\theta)}{d\theta^3}\right) \\
I_3 &= 2^{-1}U(\theta)^{-3}\left\{\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right)\right\}^2\mathbb{E}\left(\frac{d^3L_n(\theta)}{d\theta^3}\right) \\
I_4 &= 2^{-1}U(\theta)^{-3}\mathrm{cov}\left(\left(\frac{dL_n(\theta)}{d\theta}\right)^2, \frac{d^3L_n(\theta)}{d\theta^3}\right).
\end{aligned}$$

Note that the term $\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right)$ will be different for the four quasi-likelihoods (and will be of order $O(n^{-1})$). However the remaining terms are asymptotically the same for three quasi-likelihoods and will be slightly different for the hybrid Whittle likelihood.

The first derivative We first obtain expressions for $\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right)$ for the four quasi-likelihoods:

$$\mathbb{E}\left(\frac{dK_n(\theta)}{d\theta}\right) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[|J_n(\omega_{k,n})|^2]\frac{d}{d\theta}f_\theta(\omega_{k,n})^{-1} = \frac{1}{n}\sum_{k=1}^{n}f_n(\omega_{k,n})\frac{d}{d\theta}f_\theta(\omega_{k,n})^{-1} = B_{K,n}(\theta),$$

where $f_n(\omega) = \int F_n(\omega-\lambda)f(\lambda)d\lambda$ and $F_n$ is the Fejér kernel of order $n$.

To obtain the expected derivative of $\mathcal{L}_n(\theta)$ we recall that

$$\mathbb{E}\left[\frac{d}{d\theta}\mathcal{L}_n(\theta)\right] = \mathbb{E}\left[\frac{d}{d\theta}K_n(\theta)\right] + \mathbb{E}\left[n^{-1}\underline{X}_n'F_n^*\frac{d}{d\theta}\Delta_n(f_\theta^{-1})D_n(f_\theta)\underline{X}_n\right].$$

Now by replacing $D_n(f_\theta)$ with $D_{\infty,n}(f_\theta)$ and using (A.31) we have

$$\begin{aligned}
\mathbb{E}\left[\frac{d}{d\theta}\mathcal{L}_n(\theta)\right] &= \mathbb{E}\left[\frac{d}{d\theta}K_n(\theta)\right] + \mathbb{E}\left[n^{-1}\underline{X}_n' F_n^* \frac{d}{d\theta}\Delta_n(f_\theta^{-1}) D_{\infty,n}(f_\theta)\underline{X}_n\right] \\
&\quad + \mathbb{E}\left[n^{-1}\underline{X}_n' F_n^* \frac{d}{d\theta}\Delta_n\left(f_\theta^{-1}\right)\left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right)\underline{X}_n\right] \\
&= \mathbb{E}\left[\frac{d}{d\theta}K_n(\theta)\right] + n^{-1}\sum_{s,t=1}^{n} c(s-t)\frac{1}{n}\sum_{k=1}^{n} e^{-is\omega_{k,n}}\frac{d}{d\theta}\varphi_{t,n}(\omega_{k,n}; f_\theta) \\
&\quad + n^{-1}\mathbb{E}\left[\underline{X}_n' F_n^* \frac{d}{d\theta}\Delta_n\left(f_\theta^{-1}\right)\left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right)\underline{X}_n\right]
\end{aligned}$$

where $\varphi_{t,n}(\omega; f_\theta) = \sigma^{-2}\left[\overline{\phi(\omega; f_\theta)}\phi_t^\infty(\omega; f_\theta) + e^{i\omega}\phi(\omega; f_\theta)\overline{\phi_{n+1-t}^\infty(\omega; f_\theta)}\right]$. The first term on the RHS of the above is $B_{K,n}(\theta)$. Using the change of variables $t' = n+1-t$, the second term in RHS above can be written as

$$\begin{aligned}
&n^{-1}\sum_{s,t=1}^{n} c(s-t)\frac{1}{n}\sum_{k=1}^{n} e^{-is\omega_{k,n}}\frac{d}{d\theta}\varphi_{t,n}(\omega_{k,n}; f_\theta) \\
&= n^{-1}\sum_{s,t=1}^{n} c(s-t)\frac{1}{n}\sum_{k=1}^{n}\frac{d}{d\theta}\left[e^{-is\omega_{k,n}}\overline{\phi(\omega_{k,n}; f_\theta)}\phi_t^\infty(\omega_{k,n}; f_\theta) + e^{-i(s-1)\omega_{k,n}}\phi(\omega_{k,n}; f_\theta)\overline{\phi_{n+1-t}^\infty(\omega_{k,n}; f_\theta)}\right] \\
&= n^{-1}\sum_{s,t=1}^{n} c(s-t)\frac{1}{n}\sum_{k=1}^{n} e^{-is\omega_{k,n}}\frac{d}{d\theta}\overline{\phi(\omega_{k,n}; f_\theta)}\phi_t^\infty(\omega_{k,n}; f_\theta) \\
&\quad + n^{-1}\sum_{s,t'=1}^{n} c(s-n-1+t')\frac{1}{n}\sum_{k=1}^{n} e^{-i(s-1)\omega_{k,n}}\frac{d}{d\theta}\phi(\omega_{k,n}; f_\theta)\overline{\phi_{t'}^\infty(\omega_{k,n}; f_\theta)} \qquad (\text{let} \quad t' = n+1-t) \\
&= \mathrm{Re}\,\frac{2}{n}\sum_{s,t=1}^{n} c(s-t)\frac{1}{n}\sum_{k=1}^{n} e^{-is\omega_{k,n}}\frac{d}{d\theta}\overline{\phi(\omega_{k,n}; f_\theta)}\phi_t^\infty(\omega_{k,n}; f_\theta) = B_{G,n}(\theta).
\end{aligned}$$

Finally, by using Corollary A.3.1 we have

$$n^{-1}\left\|\underline{X}_n' F_n^* \frac{d}{d\theta}\Delta_n(f_\theta^{-1})\left[D_n(f_\theta) - D_{\infty,n}(f_\theta)\right]\underline{X}_n\right\|_{\mathbb{E},1} = O(n^{-K+1/2}).$$

Thus the derivative of the Gaussian likelihood is

$$\mathbb{E}\left(\frac{d\mathcal{L}_n(\theta)}{d\theta}\right) = B_{K,n}(\theta) + B_{G,n}(\theta) + O(n^{-K+1/2}).$$

Next we consider the boundary corrected Whittle likelihood. By using that

$$\mathbb{E}[\widetilde{J}_n(\omega_{k,n};f)\overline{J_n(\omega_{k,n})}] = f(\omega_{k,n})$$

we have

$$\begin{aligned}\mathbb{E}\left(\frac{dW_n(\theta)}{d\theta}\right) &= \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\widetilde{J}_n(\omega_{k,n};f)\overline{J_n(\omega_{k,n})}]\frac{d}{d\theta}f_\theta(\omega_{k,n})^{-1} \\ &= \frac{1}{n}\sum_{k=1}^{n}f(\omega_{k,n})\frac{d}{d\theta}f_\theta(\omega_{k,n})^{-1}.\end{aligned}$$

Finally, the analysis of $H_n(\theta)$ is identical to the analysis of $W_n(\theta)$ and we obtain

$$\begin{aligned}\mathbb{E}\left(\frac{dH_n(\theta)}{d\theta}\right) &= \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\widetilde{J}_n(\omega_{k,n};f)\overline{J_{\underline{h}_n,n}(\omega_{k,n})}]\frac{d}{d\theta}f_\theta(\omega_{k,n})^{-1} \\ &= \frac{1}{n}\sum_{k=1}^{n}f(\omega_{k,n})\frac{d}{d\theta}f_\theta(\omega_{k,n})^{-1}.\end{aligned}$$

In summary, evaluating the above at the best fitting parameter $\theta_n$ and by Assumption 4.2.1(ii) gives

$$\begin{aligned}\mathbb{E}\left(\frac{dK_n(\theta)}{d\theta}\right)\rfloor_{\theta=\theta_n} &= B_{K,n}(\theta_n) \\ \mathbb{E}\left(\frac{d\mathcal{L}_n(\theta)}{d\theta}\right)\rfloor_{\theta=\theta_n} &= B_{K,n}(\theta_n) + B_{G,n}(\theta_n) + O(n^{-K+1/2}) \\ \text{and} \qquad \mathbb{E}\left(\frac{dW_n(\theta)}{d\theta}\right)\rfloor_{\theta=\theta_n} &= \mathbb{E}\left(\frac{dH_n(\theta)}{d\theta}\right)\rfloor_{\theta=\theta_n} = 0. \qquad \text{(B.4)}\end{aligned}$$

It can be shown that $B_{K,n}(\theta_n) = O(n^{-1})$ and $B_{G,n}(\theta_n) = O(n^{-1})$. These terms could be negative or positive so there is no clear cut answer as to whether $B_{K,n}(\theta_n)$ or $B_{K,n}(\theta_n) + B_{G,n}(\theta_n)$ is larger (our simulations results suggest that often $B_{K,n}(\theta_n)$ tends to be larger).

The second and third order derivatives The analysis of all the higher order terms will require comparisons between the derivatives of $\mathcal{L}_n(\theta), K_n(\theta), W_n(\theta)$ and $H_n(\theta)$. We first represent the deriva-

tives of the Gaussian likelihood in terms of the Whittle likelihood

$$\frac{d^i\mathcal{L}_n(\theta)}{d\theta^i} = \frac{d^iK_n(\theta)}{d\theta^i} + \mathbb{E}\left[n^{-1}\underline{X}_n'F_n^*\frac{d^i}{d\theta^i}\Delta_n(f_\theta^{-1})D_n(f_\theta)\underline{X}_n\right].$$

By using (A.34), for $1 \leqslant i \leqslant 3$ we have

$$\left\|n^{-1}\underline{X}_n'F_n^*\frac{d^i}{d\theta^i}\Delta_n(f_\theta^{-1})D_n(f_\theta)\underline{X}_n\right\|_{\mathbb{E},1} = O(n^{-1}). \tag{B.5}$$

Similarly, we represent the derivatives of $W_n(\theta)$ and $H_{p,n}(\theta)$ in terms of the derivatives of $K_n(\theta)$

$$\begin{aligned}\frac{d^iW_n(\theta)}{d\theta^i} &= \frac{d^iK_n(\theta)}{d\theta^i} + C_{i,n}\\ \frac{d^iH_n(\theta)}{d\theta^i} &= \frac{d^iK_{n,\underline{h}_n}(\theta)}{d\theta^i} + D_{i,n}\end{aligned}$$

where $K_{n,\underline{h}_n}(\theta) = n^{-1}\sum_{k=1}^n \frac{J_n(\omega_{k,n})\overline{J_{n,\underline{h}_n}(\omega_{k,n})}}{f_\theta(\omega_{k,n})}$ and

$$\begin{aligned}C_{i,n} &= \frac{1}{n}\sum_{k=1}^n \frac{d^i}{d\theta^i}\frac{\widehat{J}_n(\omega_{k,n};f)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} = \frac{1}{n}\underline{X}_n'F_n^*\Delta_n(\frac{d^i}{d\theta^i}f_\theta^{-1})D_n(f)\underline{X}_n\\ D_{i,n} &= \frac{1}{n}\sum_{k=1}^n \frac{d^i}{d\theta^i}\frac{\widehat{J}_n(\omega_{k,n};f)\overline{J_{n,\underline{h}_n}(\omega_{k,n})}}{f_\theta(\omega_{k,n})} = \frac{1}{n}\underline{X}_n'H_nF_n^*\Delta_n(\frac{d^i}{d\theta^i}f_\theta^{-1})D_n(f)\underline{X}_n,\end{aligned}$$

where $H_n = \mathrm{diag}(h_{1,n},\ldots,h_{n,n})$. In the analysis of the first order derivative obtaining an exact bound between each "likelihood" and the Whittle likelihood was important. However, for the higher order derivatives we simply require a moment bound on the difference. To bound $C_{i,n}$, we use that

$$\begin{aligned}&\left\|F_n^*\Delta_n(\frac{d^i}{d\theta^i}f_\theta^{-1})D_n(f)\right\|_1\\ &\quad\leqslant \left\|F_n^*\Delta_n(\frac{d^i}{d\theta^i}f_\theta^{-1})[D_n(f)-D_{\infty,n}(f)]\right\|_1 + \left\|F_n^*\Delta_n(\frac{d^i}{d\theta^i}f_\theta^{-1})D_{\infty,n}(f)\right\|_1.\end{aligned}$$

We use a similar method to the proof of Theorem 2.4.3, equation (2.40) and Theorem 2.4.4, equa-

tion (2.43) with $\Delta_n(\frac{d^i}{d\theta^i}f_\theta^{-1})$ and $D_{\infty,n}(f)$ replacing $\Delta_n(f_\theta^{-1})$ and $D_{\infty,n}(f_\theta)$ respectively together with Assumption 4.2.1 and 4.2.2. By using the proof of Theorem 2.4.3, equation (2.40), we have $\|F_n^*\Delta_n(\frac{d^i}{d\theta^i}f_\theta^{-1})[D_n(f) - D_{\infty,n}(f)]\|_1 = O(n^{-K+1})$. Similarly, by using the proof of Theorem 2.4.4, equation (2.43) we have $\|F_n^*\Delta_n(\frac{d^i}{d\theta^i}f_\theta^{-1})D_{\infty,n}(f)\|_1 = O(1)$. Altogether this gives

$$(\mathbb{E}|C_{i,n}|^2)^{1/2} = n^{-1}\left\|\underline{X}_n' F_n^*\Delta_n(\frac{d^i}{d\theta^i}f_\theta^{-1})D_n(f)\underline{X}_n\right\|_{\mathbb{E},2} = O(n^{-1}). \tag{B.6}$$

For the hybrid likelihood, we use that $\sup_{t,n}|h_{t,n}| < \infty$, this gives

$$\|H_n F_n^*\Delta_n(\frac{d^i}{d\theta^i}f_\theta^{-1})D_n(f)\|_1 \leqslant (\sup_t h_{t,n}) \times \|F_n^*\Delta_n(\frac{d^i}{d\theta^i}f_\theta^{-1})D_n(f)\|_1 = O(1).$$

Therefore, under the condition that $\{h_{t,n}\}$ is a bounded sequence

$$(\mathbb{E}|D_{i,n}|^2)^{1/2} = n^{-1}\left\|H_n\underline{X}_n' F_n^*\Delta_n(\frac{d^i}{d\theta^i}f_\theta^{-1})D_n(f)\underline{X}_n\right\|_{\mathbb{E},2} = O\left(n^{-1}\right). \tag{B.7}$$

Thus the expectations of the derivatives are

$$\begin{aligned}
\mathbb{E}\left(\frac{d^i\mathcal{L}_n(\theta)}{d\theta^i}\right) &= \mathbb{E}\left(\frac{d^iK_n(\theta)}{d\theta^i}\right) + O(n^{-1}) \\
\mathbb{E}\left(\frac{d^iW_n(\theta)}{d\theta^i}\right) &= \mathbb{E}\left(\frac{d^iK_n(\theta)}{d\theta^i}\right) + O(n^{-1}) \\
\mathbb{E}\left(\frac{d^iH_n(\theta)}{d\theta^i}\right) &= \mathbb{E}\left(\frac{d^iK_n(\theta)}{d\theta^i}\right) + O\left(n^{-1}\right).
\end{aligned}$$

This gives the expectation of the second and third derivatives of all likelihoods in terms of $I(\theta)$ and $J(\frac{d^3f_\theta^{-1}}{d\theta^3})$:

$$\mathbb{E}\left(\frac{d^2L_n(\theta)}{d\theta^2}\right) = -I(\theta) + O(n^{-1}), \quad \text{and} \quad \mathbb{E}\left(\frac{d^3L_n(\theta)}{d\theta^3}\right) = J(\frac{d^3f_\theta^{-1}}{d\theta^3}) + O(n^{-1}).$$

Bounds for the covariances between the derivatives The terms $I_1, I_2$ and $I_4$ all contain the covariance between various likelihoods and its derivatives. Thus to obtain expression and bounds for

these terms we use that

$$\operatorname{var}\left(\frac{d^i}{d\theta^i}K_n(\theta)\right) = O(n^{-1}), \tag{B.8}$$

where the above can be proved using Brillinger (1981), Theorem 4.3.2. Further, if the data taper $\{h_{t,n}\}$ is such that $h_{t,n} = c_n h_n(t/n)$ where $c_n = n/H_{1,n}$ and $h_n : [0,1] \to \mathbb{R}$ is a sequence of taper functions which satisfy the taper conditions in Section 5, Dahlhaus (1988), then

$$\operatorname{var}\left(\frac{d^i}{d\theta^i}K_{n,\underline{h}_n}(\theta)\right) = O\left(\frac{H_{2,n}}{H_{1,n}^2}\right). \tag{B.9}$$

By using (B.5), (B.6), and (B.8) we have

$$\begin{aligned}
\operatorname{cov}\left(\frac{d\mathcal{L}_n(\theta)}{d\theta}, \frac{d^2\mathcal{L}_n(\theta)}{d\theta^2}\right) &= \operatorname{cov}\left(\frac{dK_n(\theta)}{d\theta}, \frac{d^2K_n(\theta)}{d\theta^2}\right) + O(n^{-3/2}) \\
\operatorname{cov}\left(\frac{dW_n(\theta)}{d\theta}, \frac{d^2W_n(\theta)}{d\theta^2}\right) &= \operatorname{cov}\left(\frac{dK_n(\theta)}{d\theta}, \frac{d^2K_n(\theta)}{d\theta^2}\right) + O(n^{-3/2}) \\
\operatorname{var}\left(\frac{d\mathcal{L}_n(\theta)}{d\theta}\right) &= \operatorname{var}\left(\frac{dK_n(\theta)}{d\theta}\right) + O(n^{-3/2}) \\
\operatorname{var}\left(\frac{dW_n(\theta)}{d\theta}\right) &= \operatorname{var}\left(\frac{dK_n(\theta)}{d\theta}\right) + O(n^{-3/2}).
\end{aligned}$$

For the hybrid Whittle likelihood, by using (B.7) and (B.9)

$$\begin{aligned}
\operatorname{cov}\left(\frac{dH_n(\theta)}{d\theta}, \frac{d^2H_n(\theta)}{d\theta^2}\right) &= \operatorname{cov}\left(\frac{dK_{n,\underline{h}_n}(\theta)}{d\theta}, \frac{d^2K_{n,\underline{h}_n}(\theta)}{d\theta^2}\right) + O\left(\frac{H_{2,n}^{1/2}}{nH_{1,n}}\right) \\
\operatorname{var}\left(\frac{dH_n(\theta)}{d\theta}\right) &= \operatorname{var}\left(\frac{dK_{n,\underline{h}_n}(\theta)}{d\theta}\right) + O\left(\frac{H_{2,n}^{1/2}}{nH_{1,n}}\right).
\end{aligned}$$

Using that $H_{2,n}/H_{1,n}^2 = O(n^{-1})$, we show that the above error terms $O(H_{2,n}^{1/2}/(nH_{1,n}))$ (for the hybrid Whittle likelihood) is the same as the other likelihoods. Next, having reduced the above covariances to those of the derivatives of $K_n(\theta)$ and $K_{n,\underline{h}_n}(\theta)$. We first focus on $K_n(\theta)$. By using the expressions for cumulants of DFTs given in Brillinger (1981), Theorem 4.3.2 and well-known

cumulant arguments we can show that

$$\begin{aligned} \text{cov}\left(\frac{dK_n(\theta)}{d\theta}, \frac{d^2K_n(\theta)}{d\theta^2}\right) &= n^{-1}V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{d^2f_\theta^{-1}}{d\theta^2}\right) + O(n^{-2}) \\ \text{and var}\left(\frac{dK_n(\theta)}{d\theta}\right) &= n^{-1}V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{df_\theta^{-1}}{d\theta}\right) + O(n^{-2}). \end{aligned}$$

To obtain expressions for the covariance involving $K_{n,\underline{h}_n}(\theta)$, we apply similar techniques as those developed in Dahlhaus (1983), Lemma 6 together with cumulant arguments. This gives

$$\begin{aligned} \text{cov}\left(\frac{dK_{n,\underline{h}_n}(\theta)}{d\theta}, \frac{d^2K_{n,\underline{h}_n}(\theta)}{d\theta^2}\right) &= \frac{H_{2,n}}{H_{1,n}^2}V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{d^2f_\theta^{-1}}{d\theta^2}\right) + O\left(\frac{H_{2,n}}{nH_{1,n}^2}\right) \\ \text{and var}\left(\frac{dK_{n,\underline{h}_n}(\theta)}{d\theta}\right) &= \frac{H_{2,n}}{H_{1,n}^2}V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{df_\theta^{-1}}{d\theta}\right) + O\left(\frac{H_{2,n}}{nH_{1,n}^2}\right). \end{aligned}$$

These results yield expressions for $I_1$ and $I_2$ (we obtain these below).

Expression for $I_0$ and a bound for $I_3$. Using the results above we have

- The Gaussian likelihood

$$I_0 = I(\theta_n)^{-1}\left[B_{K,n}(\theta_n) + B_{G,n}(\theta_n)\right] + O(n^{-2}) \tag{B.10}$$

- The Whittle likelihood

$$I_0 = I(\theta_n)^{-1}B_{K,n}(\theta_n) + O(n^{-2})$$

- The boundary corrected Whittle and hybrid Whittle likelihood

$$I_0 = 0$$

However, since for all the likelihoods $\mathbb{E}[\frac{dL_n(\theta)}{d\theta}\rfloor_{\theta=\theta_n}] = O(n^{-1})$, this implies that for all the likeli-

hoods the term $I_3$ is

$$I_3 = 2^{-1}U(\theta)^{-3}\left\{\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right)\right\}^2\mathbb{E}\left(\frac{d^3L_n(\theta)}{d\theta^3}\right) = O(n^{-2}).$$

Expression for $I_1$ and $I_2$. For the Gaussian, Whittle, and boundary corrected Whittle likelihoods we have

$$\begin{aligned} I_1 &= n^{-1}I(\theta_n)^{-2}V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{d^2f_\theta^{-1}}{d\theta^2}\right) + O(n^{-3/2}) \\ I_2 &= n^{-1}2^{-1}I(\theta_n)^{-3}V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{df_\theta^{-1}}{d\theta}\right)J\left(\frac{d^3f_\theta^{-1}}{d\theta^3}\right) + O(n^{-3/2}). \end{aligned}$$

For the hybrid Whittle likelihood we obtain a similar expression

$$\begin{aligned} I_1 &= \frac{H_{2,n}}{H_{1,n}^2}I(\theta_n)^{-2}V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{d^2f_\theta^{-1}}{d\theta^2}\right) + O\left(n^{-3/2}\right) \\ I_2 &= \frac{H_{2,n}}{H_{1,n}^2}2^{-1}I(\theta_n)^{-3}V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{df_\theta^{-1}}{d\theta}\right)J\left(\frac{d^3f_\theta^{-1}}{d\theta^3}\right) + O\left(n^{-3/2}\right). \end{aligned}$$

A bound for $I_4$ We now show that $I_4$ has a lower order term than the dominating terms $I_0, I_1$ and $I_2$. We recall that

$$I_4 = 2^{-1}U(\theta)^{-3}\mathrm{cov}\left(\left(\frac{dL_n(\theta)}{d\theta}\right)^2, \frac{d^3L_n(\theta)}{d\theta^3}\right).$$

To bound the above we focus on $\mathrm{cov}\left(\left(\frac{dL_n(\theta)}{d\theta}\right)^2, \frac{d^3L_n(\theta)}{d\theta^3}\right)$. By using indecomposable partitions

we have

$$\begin{aligned}\operatorname{cov}\left(\left(\frac{dL_n(\theta)}{d\theta}\right)^2, \frac{d^3L_n(\theta)}{d\theta^3}\right) &= 2\operatorname{cov}\left(\frac{dL_n(\theta)}{d\theta}, \frac{d^3L_n(\theta)}{d\theta^3}\right)\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right) \\ &\quad +\operatorname{cum}\left(\frac{dL_n(\theta)}{d\theta}, \frac{dL_n(\theta)}{d\theta}\frac{d^3L_n(\theta)}{d\theta^3}\right) \\ &\quad +\left[\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right)\right]^2\mathbb{E}\left(\frac{d^3L_n(\theta)}{d\theta^3}\right).\end{aligned}$$

We use (B.8), (B.5) and (B.6) to replace $L_n(\theta)$ with $K_n(\theta)$ or $K_{\underline{h},n}(\theta)$. Finally by using the expressions for cumulants of DFTs given in Brillinger (1981), Theorem 4.3.2 we have that for the non-hybrid likelihoods

$$I_4 = O(n^{-2})$$

and for the hybrid Whittle likelihood

$$I_4 = O\left(\frac{H_{2,n}}{nH_{1,n}^2}\right).$$

Thus, altogether for all the estimators we have that

$$(\widehat{\theta}_n - \theta_n) = I_0 + I_1 + I_2 + O(n^{-2}),$$

where for the Gaussian, Whittle and boundary corrected Whittle likelihoods

$$\begin{aligned}I_1 + I_2 &= n^{-1}\left[I(\theta_n)^{-2}V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{d^2f_\theta^{-1}}{d\theta^2}\right) + 2^{-1}I(\theta_n)^{-3}V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{df_\theta^{-1}}{d\theta}\right)J\left(\frac{d^3f_\theta^{-1}}{d\theta^3}\right)\right] + O(n^{-3/2}) \\ &= n^{-1}G(\theta_n) + O(n^{-3/2})\end{aligned}$$

and for the hybrid Whittle likelihood

$$I_1 + I_2 = \frac{H_{2,n}}{H_{1,n}^2}G(\theta_n) + O\left(n^{-3/2}\right).$$

The terms for $I_0$ are given in (B.10). This proves the result. □

**Remark B.1.1.** *In the case that the model is linear, then* $f_4(\omega_1, -\omega_1, \omega_2) = (\kappa_4/\sigma^4)f(\omega_1)f(\omega_2)$ *where* $\sigma^2$ *and* $\kappa_4$ *is the* 2*nd and* 4*th order cumulant of the innovation in the model.*

*Furthermore, in the case the model is correct specification and linear, we can show that Assumption 4.2.1(ii) implies that fourth order cumulant term in* $V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{d^2 f_\theta^{-1}}{d\theta^2}\right)$ *and* $V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{df_\theta^{-1}}{d\theta}\right)$ *is zero. This results in the fourth order cumulant term in* $G(\cdot)$ *being zero.*

## B.2 The bias for the AR(1) model

In general, it is difficult to obtain a simple expression for the bias defined in Theorem B.1.1, but in the special case a model $AR(1)$ is fitted to the data the bias can be found. In the calculation below let $\theta$ denote the AR$(1)$ coefficient for the best fitting AR$(1)$ parameter. We assume Gaussianity, which avoids dealing with the fourth order spectral density.

If the true model is a Gaussian AR$(1)$ the bias for the various criteria is

- The Gaussian likelihood

$$\mathbb{E}[\widehat{\theta}_n^G - \theta] = -\frac{1}{n}\theta + O(n^{-3/2})$$

- The Whittle likelihood

$$\mathbb{E}[\widehat{\theta}_n^K - \theta] = -\frac{3}{n}\theta + \frac{1}{n}\theta^{n-1} + O(n^{-3/2})$$

- The boundary corrected Whittle likelihood

$$\mathbb{E}[\widehat{\theta}_n^W - \theta] = -\frac{2}{n}\theta + O(p^3 n^{-3/2} + (np^{K-1})^{-1})$$

- The hybrid Whittle likelihood

$$\mathbb{E}[\widehat{\theta}_n^H - \theta] = -2\frac{H_{2,n}}{H_{1,n}^2}\theta + O(p^3 n^{-3/2} + (np^{K-1})^{-1}).$$

Moreover, if the Gaussian likelihood included the determinant term in the Gaussian likelihood, i.e. $\widetilde{\theta}_n^G = \arg\min_\theta[\mathcal{L}_n(\theta) + n^{-1}\log|\Gamma_n(f_\theta)|]$, then

$$\mathbb{E}[\widetilde{\theta}_n^G - \theta] = -\frac{2}{n}\theta + O(n^{-3/2}).$$

We observe for the AR$(1)$ model (when the true time series is Gaussian with an AR$(1)$ representation) that the "true" Gaussian likelihood with the log-determinant term has a larger bias than the Gaussian likelihood without the Gaussian determinant term.

The above bounds show that the Gaussian likelihood with the log-determinant term and the boundary corrected Whittle likelihood have the same asymptotic bias. This is substantiated in the simulations. However, in the simulations in Section 4.6.1, we do observe that the bias of the Gaussian likelihood is a little less than the boundary corrected Whittle. The difference between two likelihoods is likely due to differences in the higher order terms which are of order $O(n^{-3/2})$ (for the Gaussian likelihood) and $O(p^3n^{-3/2})$ (for the boundary corrected Whittle likelihood, due to additional estimation of the predictive DFT).

PROOF. The inverse of the spectral density function and autocovariance function is

$$f_\theta(\omega)^{-1} = \sigma^{-2}\left(1 + \theta^2 - 2\theta\cos(\omega)\right) \quad \text{and} \quad c(r) = \frac{\sigma^2\theta^{|r|}}{1-\theta^2}.$$

Thus

$$\frac{d}{d\theta}f_\theta(\omega)^{-1} = 2\sigma^{-2}(\theta - \cos\omega) \quad \text{and} \quad \frac{d^2}{d\theta^2}f_\theta(\omega)^{-1} = 2\sigma^{-2}.$$

This gives

$$I(\theta) = -\frac{1}{2\pi}\int_0^{2\pi}\frac{d^2 f_\theta(\omega)^{-1}}{d\theta^2}f(\omega)d\omega = -\frac{1}{\pi\sigma^2}\int_0^{2\pi}f(\omega)d\omega = -\frac{2}{\sigma^2}c(0). \tag{B.11}$$

Next we calculate $B_{G,n}$, since $\phi(\omega) = 1 - \theta e^{-i\omega}$ it is is easy to show

$$\phi_1^\infty(\omega) = \theta \quad \text{and} \quad \phi_j^\infty(\omega) = 0 \text{ for } j \geqslant 2.$$

Therefore,

$$\begin{aligned} B_{G,n}(\theta) &= \operatorname{Re} \frac{2}{n} \sum_{t,j=1}^{n} c(t-j) \frac{1}{n} \sum_{k=1}^{n} e^{-it\omega_{k,n}} \frac{d}{d\theta} \left[ \overline{\phi(\omega_{k,n}; f_\theta)} \phi_j^\infty(\omega_{k,n}; f_\theta) \right] \\ &= \frac{2\sigma^{-2}}{n} \sum_{t=1}^{n} c(t-1) \frac{1}{n} \sum_{k=1}^{n} e^{-it\omega_{k,n}} \frac{d}{d\theta} \left[ (1 - \theta e^{i\omega_{k,n}})\theta \right] \\ &= \frac{2\sigma^{-2}}{n} \sum_{t=1}^{n} c(t-1) \frac{1}{n} \sum_{k=1}^{n} \left( e^{-it\omega_{k,n}} - 2\theta e^{-i(t-1)\omega_{k,n}} \right) \\ &= \frac{2\sigma^{-2}}{n} \sum_{t=1}^{n} c(t-1) \left[ \frac{1}{n} \sum_{k=1}^{n} e^{-it\omega_{k,n}} - \frac{2\theta}{n} \sum_{k=1}^{n} e^{-i(t-1)\omega_{k,n}} \right] \end{aligned}$$

The second summation (over $k$) is 0 unless $t \in \{1, n\}$. Therefore,

$$B_{G,n}(\theta) = \frac{2\sigma^{-2}}{n} c(n-1) - \frac{4\sigma^{-2}\theta}{n} c(0). \tag{B.12}$$

To calculate $B_{K,n}$ we have

$$\begin{aligned} B_{K,n}(\theta) &= \frac{1}{n} \sum_{k=1}^{n} f_n(\omega_{k,n}) \frac{df_\theta(\omega_{k,n})^{-1}}{d\theta} \\ &= \frac{2\sigma^{-2}}{n} \sum_{k=1}^{n} f_n(\omega_{k,n}) \left( \theta - \cos(\omega_{k,n}) \right) \\ &= 2\sigma^{-2} \left[ \theta c(0) - \left( \frac{n-1}{n} \right) c(1) - \frac{1}{n} c(1-n) \right] \\ &= \frac{2\sigma^{-2}}{n} \left[ c(1) - c(n-1) \right]. \end{aligned} \tag{B.13}$$

Altogether this gives

$$\begin{aligned} I(\theta)^{-1}\left(B_{K,n}(\theta)+B_{G,n}(\theta)\right) &= -\frac{\sigma^2}{2nc(0)}\left(2\sigma^{-2}\left[c(1)-c(n-1)\right]+2\sigma^{-2}c(n-1)-4\sigma^{-2}\theta c(0)\right) \\ &= -\frac{1}{nc(0)}\left(c(1)-2\theta c(0)\right)=\frac{\theta}{n} \end{aligned} \tag{B.14}$$

and

$$I(\theta)^{-1}B_{K,n}(\theta) = -\frac{1}{nc(0)}\left[c(1)-c(n-1)\right] = -\frac{1}{n}(\theta-\theta^{n-1}). \tag{B.15}$$

Next, we calculate $G(\theta)$. Since the third derivative of $f_\theta^{-1}$ with respect to $\theta$ is zero we have

$$G(\theta) = I(\theta)^{-2}V\left(\frac{d}{d\theta}f_\theta^{-1}, \frac{d^2}{d\theta^2}f_\theta^{-1}\right)$$

where

$$\begin{aligned} V\left(\frac{d}{d\theta}f_\theta^{-1}, \frac{d^2}{d\theta^2}f_\theta^{-1}\right) &= \frac{1}{\pi}\int_0^{2\pi}\left(\frac{2\theta-2\cos(\omega)}{\sigma^2}\right)\left(\frac{2}{\sigma^2}\right)f(\omega)^2 d\omega \\ &= \frac{4}{\sigma^4}\frac{1}{\pi}\int_0^{2\pi}\left[\theta-\cos(\omega)\right]f(\omega)^2 d\omega \\ &= \frac{8}{\sigma^4}\left(\theta c_2(0)-c_2(1)\right) \end{aligned}$$

where $\{c_2(r)\}$ is the autocovariance function associated with $f(\omega)^2$, it is the convolution of $c(r)$ with itself;

$$c_2(r) = \sum_{\ell\in\mathbb{Z}} c(\ell)c(\ell+r)$$

Using this expansion we have

$$V\left(\frac{d}{d\theta}f_\theta^{-1}, \frac{d^2}{d\theta^2}f_\theta^{-1}\right) = \frac{8}{\sigma^4}\sum_{\ell\in\mathbb{Z}} c(\ell)\left[\theta c(\ell)-c(\ell+1)\right]$$

and

$$G(\theta) = \frac{\sigma^4}{4c(0)^2}\frac{8}{\sigma^4}\left(\theta c_2(0) - c_2(1)\right) = \frac{2}{c(0)^2}\left(\theta c_2(0) - c_2(1)\right). \tag{B.16}$$

Putting (B.16) with (B.14) gives

$$\begin{aligned}
\mathbb{E}[\widehat{\theta}_n^G - \theta] &\approx I(\theta)^{-1}\left(B_{K,n}(\theta) + B_{G,n}(\theta)\right) + n^{-1}G(\theta) \\
&= \frac{\theta}{n} + \frac{2}{nc(0)^2}\left(\theta c_2(0) - c_2(1)\right), \\
\mathbb{E}[\widehat{\theta}_n^K - \theta] &\approx I(\theta)^{-1}B_{K,n}(\theta) + n^{-1}G(\theta) \\
&= -\frac{1}{n}(\theta - \theta^{n-1}) + \frac{2}{nc(0)^2}\left(\theta c_2(0) - c_2(1)\right), \\
\mathbb{E}[\widehat{\theta}_n^W - \theta] &\approx \frac{2}{nc(0)^2}\left(\theta c_2(0) - c_2(1)\right), \\
\mathbb{E}[\widehat{\theta}_n^H - \theta] &\approx \frac{2}{c(0)^2}\frac{H_{2,n}}{H_{1,n}^2}\left(\theta c_2(0) - c_2(1)\right).
\end{aligned}$$

It is not entirely clear how to access the above. So now we consider the case that the model is fully specified. Under correct specification we have

$$\begin{aligned}
V\left(\frac{d}{d\theta}f_\theta^{-1}, \frac{d^2}{d\theta^2}f_\theta^{-1}\right) &= \frac{2\sigma^2}{\pi}\int_0^{2\pi} f(\omega)^2\frac{df_\theta(\omega)^{-1}}{d\theta}d\omega \\
&= -\frac{2\sigma^2}{\pi}\int_0^{2\pi} f(\omega)^2\left(\frac{1}{f_\theta(\omega)^2}\right)\frac{df_\theta(\omega)}{d\theta}d\omega \\
&= -\frac{2\sigma^2}{\pi}\int_0^{2\pi}\frac{df_\theta(\omega)}{d\theta}d\omega = -\frac{2\sigma^2}{\pi}\frac{d}{d\theta}\int_0^{2\pi} f_\theta(\omega)d\omega \\
&= -4\sigma^2\frac{d}{d\theta}c(0) = -\frac{8\sigma^4\theta}{(1-\theta^2)^2}.
\end{aligned}$$

Thus $I(\theta)^{-2}V\left(\frac{d}{d\theta}f_\theta^{-1}, \frac{d^2}{d\theta^2}f_\theta^{-1}\right) = -2\theta/n$. Substituting this into the above we have

$$\begin{aligned}
\mathbb{E}[\widehat{\theta}_n^{(G)} - \theta] &= \frac{1}{n}\theta - \frac{2}{n}\theta + O(n^{-3/2}) = -\frac{1}{n}\theta + O(n^{-3/2}), \\
\mathbb{E}[\widehat{\theta}_n^{(K)} - \theta] &= -\frac{1}{n}\left[\theta - \theta^{n-1}\right] - \frac{2}{n}\theta + o(n^{-1}) \approx -\frac{3}{n}\theta + \frac{1}{n}\theta^{n-1} + O(n^{-3/2}), \\
\mathbb{E}[\widehat{\theta}_n^{(W)} - \theta] &= -\frac{2}{n}\theta + O(n^{-3/2}), \\
\mathbb{E}[\widehat{\theta}_n^{(H)} - \theta] &= -2\frac{H_{2,n}}{H_{1,n}^2}\theta + O(n^{-3/2}).
\end{aligned}$$

This proves the main part of the assertion. To compare the above bias with the "true" Gaussian likelihood, we consider the Gaussian likelihood with the log determinant term. First, consider the correlation of AR(1) matrix $(A_n)_{s,t} = \theta^{|s-t|}$. Then,

$$A_{n+1} = \begin{pmatrix} A_n & B_n \\ B_n' & 1 \end{pmatrix}, \qquad B_n = (\theta^n, ..., \theta)'.$$

Therefore, using block matrix determinant identity, $|A_{n+1}| = |A_n|(1 - B_n' A_n^{-1} B_n)$. Moreover, it is easy to show $A_n R_n = B_n$, where $R_n = (0, ..., 0, \theta)'$. Thus

$$|A_{n+1}| = |A_n|(1 - B_n' R_n) = |A_n|(1 - \theta^2).$$

Using iteration, $|A_n| = (1-\theta^2)^{n-2}|A_2| = (1-\theta^2)^{n-1}$ and thus,

$$|\Gamma_n(f_\theta)| = \left|\frac{\sigma^2}{1-\theta^2}A_n\right| = \left(\frac{\sigma^2}{1-\theta^2}\right)^n (1-\theta^2)^{n-1} = \frac{(\sigma^2)^n}{1-\theta^2}.$$

Then, by simple calculus,

$$\frac{d}{d\theta} n^{-1} \log|\Gamma_n(f_\theta)| = \frac{2\theta}{n(1-\theta^2)} = \frac{2\sigma^{-2}}{n} c(1).$$

and thus,

$$\begin{aligned}\mathbb{E}[\widetilde{\theta}^G - \theta] &\approx I(\theta)^{-1}\left(B_{K,n}(\theta) + B_{G,n}(\theta) + \frac{d}{d\theta}n^{-1}\log|\Gamma_n(f_\theta)|\right) + n^{-1}G(\theta) \\ &= I(\theta)^{-1}\frac{1}{n}(2\sigma^{-2}c(1) - 4\sigma^{-2}\theta c(0) + 2\sigma^{-2}c(1)) - \frac{2\theta}{n} = -\frac{2\theta}{n},\end{aligned}$$

which proves the results. □

### B.3 Bias for estimators of multiple parameters

We now generalize the ideas above to multiple unknown parameters. Suppose we fit the spectral density $f_\theta(\omega)$ to the time series $\{X_t\}$ where $\theta = (\theta_1, \ldots, \theta_d)$ are the unknown parameters in $\Theta \subset \mathbb{R}^d$. $\mathcal{L}_n(\theta)$, $K_n(\theta)$, $\widehat{W}_{p,n}(\theta)$ and $\widehat{H}_{p,n}(\theta)$ denote the Gaussian likelihood, Whittle likelihood, boundary corrected Whittle and hybrid Whittle likelihood defined in (4.16). Let $\widehat{\theta}_n^{(G)}$, $\widehat{\theta}_n^{(W)}$, $\widehat{\theta}_n^{(W)}$ and $\widehat{\theta}_n^{(H)}$ be the corresponding estimators defined in (4.17) and $\theta_n = (\theta_{1,n}, ..., \theta_{d,n})$ is the best fitting parameter defined as in (4.13). Then under Assumption 4.2.1 and 4.2.2 we have the following asymptotic bias:

- The Gaussian likelihood (excluding the term $n^{-1}\log|\Gamma_n(\theta)|$)

$$\mathbb{E}[\widehat{\theta}_{j,n}^{(G)} - \theta_{j,n}] = \sum_{r=1}^{d} I^{(j,r)}\left[B_{r,K,n}(\theta) + B_{r,G,n}(\theta) + n^{-1}G_r(\theta)\right] + O\left(n^{-3/2}\right)$$

- The Whittle likelihood has bias

$$\mathbb{E}[\widehat{\theta}_{j,n}^{(K)} - \theta_{j,n}] = \sum_{r=1}^{d} I^{(j,r)}\left[B_{r,K,n}(\theta) + n^{-1}G_r(\theta)\right] + O\left(n^{-3/2}\right).$$

- The boundary corrected Whittle likelihood has bias

$$\mathbb{E}[\widehat{\theta}_{j,n}^{(W)} - \theta_{j,n}] = n^{-1}\sum_{r=1}^{d} I^{(j,r)}G_r(\theta) + O\left(p^3n^{-3/2} + (np^{K-1})^{-1}\right).$$

- The hybrid Whittle likelihood has bias

$$\mathbb{E}[\widehat{\theta}_{j,n}^{(H)} - \theta_{j,n}] \;=\; \frac{H_{2,n}}{H_{1,n}^2} \sum_{r=1}^{d} I^{(j,r)} G_r(\theta) + O\left(p^3 n^{-3/2} + (np^{K-1})^{-1}\right). \qquad \text{(B.17)}$$

PROOF. Let $L_n(\theta)$ be the criterion and $\widehat{\theta}_n = \arg\min L_n(\theta)$ and $\theta_n$ the best fitting parameter. We use a similar technique used to prove Theorem B.1.1. The first order expansion is

$$\widehat{\theta}_n - \theta_n = U(\theta_n)^{-1} \nabla_\theta L_n(\theta_n)$$

where $U(\theta)$ is the $d \times d$ matrix

$$U(\theta) = -\mathbb{E}\left[\nabla_\theta^2 L_n(\theta)\right].$$

Thus entrywise we have

$$\widehat{\theta}_{r,n} - \theta_{r,n} = \sum_{s=1}^{d} U^{r,s} \frac{\partial L_n(\theta)}{\partial \theta_s}$$

where $U^{(r,s)}$ denotes the $(r,s)$-entry of the $d \times d$ matrix $U(\theta_n)^{-1}$. To obtain the "bias" we make a second order expansion. For the simplicity, we omit the subscript $n$ from $\widehat{\theta}_{r,n}$ and $\theta_{r,n}$. For $1 \leqslant r \leqslant d$ we evaluate the partial derivative

$$\frac{\partial L_n(\theta)}{\partial \theta_r} + \sum_{s=1}^{d} (\widehat{\theta}_s - \theta_s) \frac{\partial^2 L_n(\theta)}{\partial \theta_s \partial \theta_r} + \frac{1}{2} \sum_{s_1,s_2=1}^{d} (\widehat{\theta}_{s_1} - \theta_{s_1})(\widehat{\theta}_{s_2} - \theta_{s_2}) \frac{\partial^3 L_n(\theta)}{\partial \theta_{s_1} \partial \theta_{s_2} \partial \theta_r} \approx 0.$$

Taking expectation of the above gives

$$\mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial \theta_r}\right] + \sum_{s=1}^{d} \mathbb{E}\left[(\widehat{\theta}_s - \theta_s) \frac{\partial^2 L_n(\theta)}{\partial \theta_s \partial \theta_r}\right] + \frac{1}{2} \sum_{s_1,s_2=1}^{d} \mathbb{E}\left[(\widehat{\theta}_{s_1} - \theta_{s_1})(\widehat{\theta}_{s_2} - \theta_{s_2}) \frac{\partial^3 L_n(\theta)}{\partial \theta_{s_1} \partial \theta_{s_2} \partial \theta_r}\right] \approx 0.$$

We now replace the product of random variables with their covariances

$$\mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial \theta_r}\right] + \sum_{s=1}^{d} \mathbb{E}[\widehat{\theta}_s - \theta_s]\mathbb{E}\left[\frac{\partial^2 L_n(\theta)}{\partial \theta_s \partial \theta_r}\right] + \sum_{s=1}^{d} \operatorname{cov}\left[\widehat{\theta}_s - \theta_s, \frac{\partial^2 L_n(\theta)}{\partial \theta_s \partial \theta_r}\right]$$
$$+\frac{1}{2}\sum_{s_1,s_2=1}^{d} \operatorname{cov}\left(\widehat{\theta}_{s_1} - \theta_{s_1}, \widehat{\theta}_{s_2} - \theta_{s_2}\right)\mathbb{E}\left[\frac{\partial^3 L_n(\theta)}{\partial \theta_{s_1}\partial \theta_{s_2}\partial \theta_r}\right]$$
$$+\frac{1}{2}\sum_{s_1,s_2=1}^{d} \mathbb{E}[\widehat{\theta}_{s_1} - \theta_{s_1}]\mathbb{E}[\widehat{\theta}_{s_2} - \theta_{s_2}]\mathbb{E}\left[\frac{\partial^3 L_n(\theta)}{\partial \theta_{s_1}\partial \theta_{s_2}\partial \theta_r}\right]$$
$$+\frac{1}{2}\sum_{s_1,s_2=1}^{d} \operatorname{cov}\left[(\widehat{\theta}_{s_1} - \theta_{s_1})(\widehat{\theta}_{s_2} - \theta_{s_2}), \frac{\partial^3 L_n(\theta)}{\partial \theta_{s_1}\partial \theta_{s_2}\partial \theta_r}\right] \approx 0.$$

With the exception of $\mathbb{E}[\widehat{\theta}_s - \theta_s]$, we replace $\widehat{\theta}_s - \theta_s$ in the above with their first order expansions $\sum_{j=1}^{d} U^{(s,j)}\frac{\partial L_n(\theta)}{\partial \theta_j}$. This gives

$$\mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial \theta_r}\right] - \sum_{s=1}^{d} \mathbb{E}[\widehat{\theta}_s - \theta_s]U_{s,r} + \sum_{s_1,s_2=1}^{d} U^{(s_1,s_2)}\operatorname{cov}\left[\frac{\partial L_n(\theta)}{\partial \theta_{s_2}}, \frac{\partial^2 L_n(\theta)}{\partial \theta_{s_1}\partial \theta_r}\right]$$
$$+\frac{1}{2}\sum_{s_1,s_2,s_3,s_4=1}^{d} U^{(s_1,s_3)}U^{(s_2,s_4)}\operatorname{cov}\left(\frac{\partial L_n(\theta)}{\partial \theta_{s_3}}, \frac{\partial L_n(\theta)}{\partial \theta_{s_4}}\right)\mathbb{E}\left[\frac{\partial^3 L_n(\theta)}{\partial \theta_{s_1}\partial \theta_{s_2}\partial \theta_r}\right]$$
$$+\frac{1}{2}\sum_{s_1,s_2,s_3,s_4=1}^{d} U^{(s_1,s_3)}U^{(s_2,s_4)}\mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial \theta_{s_3}}\right]\mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial \theta_{s_4}}\right]\mathbb{E}\left[\frac{\partial^3 L_n(\theta)}{\partial \theta_{s_1}\partial \theta_{s_2}\partial \theta_r}\right]$$
$$+\frac{1}{2}\sum_{s_1,s_2,s_3,s_4=1}^{d} U^{(s_1,s_3)}U^{(s_2,s_4)}\operatorname{cov}\left[\frac{\partial L_n(\theta)}{\partial \theta_{s_3}}\frac{\partial L_n(\theta)}{\partial \theta_{s_4}}, \frac{\partial^3 L_n(\theta)}{\partial \theta_{s_1}\partial \theta_{s_2}\partial \theta_r}\right] \approx 0,$$

where $U_{s,r}$ denotes the $(s,r)$-entry of the $d \times d$ matrix $U(\theta_n)$

Now we consider concrete examples of likelihoods. Using the same arguments as those used in the proof of Theorem B.1.1 we have the last two terms of the above are of order $O(n^{-2})$ or $O(H_{2,n}/(nH_{1,n}^2))$ depending on the likelihood used. This implies that

$$\mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial \theta_r}\right] - \sum_{s=1}^{d} \mathbb{E}[\widehat{\theta}_s - \theta_s]U_{s,r} + \sum_{s_1,s_2=1}^{d} U^{(s_1,s_2)}\operatorname{cov}\left[\frac{\partial L_n(\theta)}{\partial \theta_{s_2}}, \frac{\partial^2 L_n(\theta)}{\partial \theta_{s_1}\partial \theta_r}\right]$$
$$+\frac{1}{2}\sum_{s_1,s_2,s_3,s_4=1}^{d} U^{(s_1,s_3)}U^{(s_2,s_4)}\operatorname{cov}\left(\frac{\partial L_n(\theta)}{\partial \theta_{s_3}}, \frac{\partial L_n(\theta)}{\partial \theta_{s_4}}\right)\mathbb{E}\left[\frac{\partial^3 L_n(\theta)}{\partial \theta_{s_1}\partial \theta_{s_2}\partial \theta_r}\right] \approx 0.$$

Let

$$\begin{aligned} J(g) &= \frac{1}{2\pi}\int_0^{2\pi} g(\omega)f(\omega)d\omega \\ I(\theta) &= -\frac{1}{2\pi}\int_0^{2\pi}[\nabla_\theta^2 f_\theta(\omega)^{-1}]f(\omega)d\omega \end{aligned}$$

and $I_{s,r}$ (and $I^{(s,r)}$) corresponds to the $(s,r)$-th element of $I(\theta_n)$ (and $I^{-1}(\theta_n)$). So far, we have no specified the likelihood $L_n(\theta)$. But to write a second order expansion for all four likelihoods we set $H_{2,n}/H_{1,n}^2 = n^{-1}$ for the Gaussian, Whittle, and boundary corrected Whittle likelihood and using the notation a similar proof to Theorem B.1.1 we have

$$\begin{aligned} &\mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial\theta_r}\right] - \sum_{s=1}^d I_{s,r}\mathbb{E}[\widehat{\theta}_s - \theta_s] + \frac{H_{2,n}}{H_{1,n}^2}\sum_{s_1,s_2=1}^d I^{(s_1,s_2)}V\left(\frac{\partial f_\theta^{-1}}{\partial\theta_{s_2}}, \frac{\partial^2 f_\theta^{-1}}{\partial\theta_{s_1}\partial\theta_r}\right) \\ &+\frac{H_{2,n}}{2H_{1,n}^2}\sum_{s_1,s_2,s_3,s_4=1}^d I^{(s_1,s_3)}I^{(s_2,s_4)}V\left(\frac{\partial f_\theta^{-1}}{\partial\theta_{s_3}}, \frac{\partial f_\theta^{-1}}{\partial\theta_{s_4}}\right)J\left(\frac{\partial^3 f_\theta^{-1}}{\partial\theta_{s_1}\partial\theta_{s_2}\partial\theta_r}\right) \approx 0. \end{aligned}$$

Thus

$$\begin{aligned} \sum_{s=1}^d I_{s,r}\mathbb{E}[\widehat{\theta}_s - \theta_s] &\approx \mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial\theta_r}\right] + \frac{H_{2,n}}{H_{1,n}^2}\sum_{s_1,s_2=1}^d I^{(s_1,s_2)}V\left(\frac{\partial f_\theta^{-1}}{\partial\theta_{s_2}}, \frac{\partial^2 f_\theta^{-1}}{\partial\theta_{s_1}\partial\theta_r}\right) \\ &\quad +\frac{H_{2,n}}{2H_{1,n}^2}\sum_{s_1,s_2,s_3,s_4=1}^d I^{(s_1,s_3)}I^{(s_2,s_4)}V\left(\frac{\partial f_\theta^{-1}}{\partial\theta_{s_3}}, \frac{\partial f_\theta^{-1}}{\partial\theta_{s_4}}\right)J\left(\frac{\partial^3 f_\theta^{-1}}{\partial\theta_{s_1}\partial\theta_{s_2}\partial\theta_r}\right). \end{aligned}$$

In the final stage, to extract $\mathbb{E}[\widehat{\theta}_s - \theta_s]$ from the above we define the $d$-dimensional column vector $\underline{D}' = (D_1, \ldots, D_d)$, where $D_r = \sum_{s=1}^d I_{s,r}\mathbb{E}[\widehat{\theta}_s - \theta_s] = [I(\theta_n)(\widehat{\theta}_n - \theta_n)]_r$. Substituting this in the above gives

$$\begin{aligned} D_r &\approx \mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial\theta_r}\right] + \frac{H_{2,n}}{H_{1,n}^2}\sum_{s_1,s_2=1}^d I^{(s_1,s_2)}V\left(\frac{\partial f_\theta(\omega)^{-1}}{\partial\theta_{s_2}}, \frac{\partial^2 f_\theta(\omega)^{-1}}{\partial\theta_{s_1}\partial\theta_r}\right) \\ &\quad +\frac{H_{2,n}}{2H_{1,n}^2}\sum_{s_1,s_2,s_3,s_4=1}^d I^{(s_1,s_3)}I^{(s_2,s_4)}V\left(\frac{\partial f_\theta^{-1}}{\partial\theta_{s_3}}, \frac{\partial f_\theta^{-1}}{\partial\theta_{s_4}}\right)J\left(\frac{\partial^3 f_\theta^{-1}}{\partial\theta_{s_1}\partial\theta_{s_2}\partial\theta_r}\right). \end{aligned}$$

Using that $\mathbb{E}[\widehat{\theta}_n - \theta_n] \approx I(\theta_n)^{-1}\underline{D}$ and substituting this into the above gives the bias for $\widehat{\theta}_j$

$$\begin{aligned}\mathbb{E}[\widehat{\theta}_j - \theta_j] &\approx \sum_{r=1}^{d} I^{(j,r)}\Bigg[\mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial\theta_r}\right] + \frac{H_{2,n}}{H_{1,n}^2}\sum_{s_1,s_2=1}^{d} I^{(s_1,s_2)}V\left(\frac{\partial f_\theta(\omega)^{-1}}{\partial\theta_{s_2}}, \frac{\partial^2 f_\theta(\omega)^{-1}}{\partial\theta_{s_1}\partial\theta_r}\right) \\ &\quad + \frac{H_{2,n}}{2H_{1,n}^2}\sum_{s_1,s_2,s_3,s_4=1}^{d} I^{(s_1,s_3)}I^{(s_2,s_4)}V\left(\frac{\partial f_\theta^{-1}}{\partial\theta_{s_3}}, \frac{\partial f_\theta^{-1}}{\partial\theta_{s_4}}\right)J\left(\frac{\partial^3 f_\theta^{-1}}{\partial\theta_{s_1}\partial\theta_{s_2}\partial\theta_r}\right)\Bigg] \quad \text{(B.18)}\end{aligned}$$

The above is a general result. We now obtain the bias for the different criteria. Let

$$\begin{aligned}B_{r,G,n}(\theta) &= \mathrm{Re}\,\frac{2}{n}\sum_{s,t=1}^{n} c(s-t)\frac{1}{n}\sum_{k=1}^{n} e^{-is\omega_{k,n}}\frac{\partial}{\partial\theta_r}\left[\overline{\phi(\omega_{k,n};f_\theta)}\phi_t^\infty(\omega_{k,n};f_\theta)\right] \\ B_{r,K,n}(\theta) &= \frac{1}{n}\sum_{k=1}^{n} f_n(\omega_{k,n})\frac{\partial f_\theta(\omega_{k,n})^{-1}}{\partial\theta_r} \\ \text{and} \qquad G_r(\theta) &= \sum_{s_1,s_2=1}^{d} I^{(s_1,s_2)}V\left(\frac{\partial f_\theta(\omega)^{-1}}{\partial\theta_{s_2}}, \frac{\partial^2 f_\theta(\omega)^{-1}}{\partial\theta_{s_1}\partial\theta_r}\right) \\ &\quad + \frac{1}{2}\sum_{s_1,s_2,s_3,s_4=1}^{d} I^{(s_1,s_3)}I^{(s_2,s_4)}V\left(\frac{\partial f_\theta^{-1}}{\partial\theta_{s_3}}, \frac{\partial f_\theta^{-1}}{\partial\theta_{s_4}}\right)J\left(\frac{\partial^3 f_\theta^{-1}}{\partial\theta_{s_1}\partial\theta_{s_2}\partial\theta_r}\right).\end{aligned}$$

Then, using similar technique from the univariate case, we can show

- The Gaussian likelihood: $\mathbb{E}\left[\partial\mathcal{L}_n(\theta)/\partial\theta_r\right] = B_{r,G,n}(\theta) + B_{r,K,n}(\theta)$.
- The Whittle likelihood: $\mathbb{E}\left[\partial K_n(\theta)/\partial\theta_r\right] = B_{r,K,n}(\theta)$
- The boundary corrected Whittle and hybrid Whittle likelihood:
  $\mathbb{E}\left[\partial W_n(\theta)/\partial\theta_r\right] = \mathbb{E}\left[\partial H_n(\theta)/\partial\theta_r\right] = 0$.

Substituting the above into (B.18) gives the four difference biases in (B.17). Thus we have proved the result. □

# APPENDIX C

# ADDITIONAL SIMULTIONS *

## C.1 Figures and Table of results for the AR$(1)$ and MA$(1)$ for a non-Gaussian time series

In this section, we provide figures and table of the results in Section 4.6.1 when the innovations follow a standardized chi-squared distribution two degrees of freedom, i.e. $\varepsilon_t \sim (\chi^2(2) - 2)/2$ (this time the asymptotic bias will contain the fourth order cumulant term). The results are very similar to the Gaussian innovations.

*Parts of this section have been modified with permission from [S. Subba Rao and J. Yang. Reconciling the Gaussian and Whittle likelihood with an application to estimation in the frequency domain. *Annals of Statistics (To appear), arXiv:2001.06966*, 2021.]

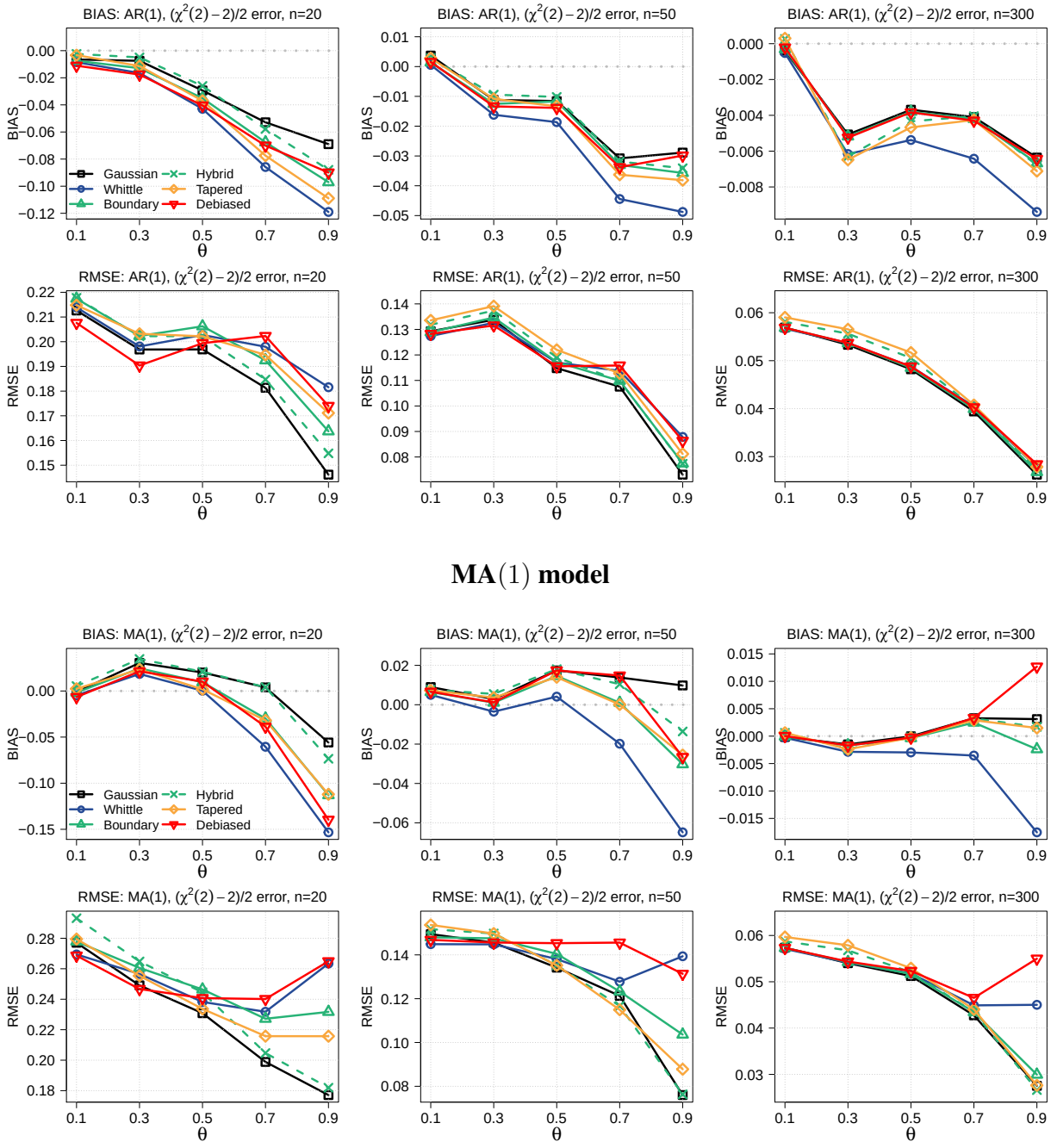



Figure C.1: Bias (first row) and the RMSE (second row) of the parameter estimates for the AR(1) and MA(1) models where the innovations follow the standardized chi-squared distribution with 2 degrees of freedom. Length of the time series $n = 20$(left), $50$(middle), and $300$(right).

| *Likelihoods* | $\theta$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
| | **AR(1)**, $\{e_t\} \sim (\chi^2(2)-2)/2$, $n = 20$ | | | | | **MA(1)**, $\{e_t\} \sim (\chi^2(2)-2)/2$, $n = 20$ | | | | |
| Gaussian | -0.007(0.21) | -0.007(0.20) | -0.029(0.19) | -0.053(0.17) | -0.069(0.13) | -0.001(0.28) | 0.030(0.25) | 0.020(0.23) | 0.004(0.20) | 0.056(0.17) |
| Whittle | -0.009(0.21) | -0.016(0.20) | -0.043(0.20) | -0.086(0.18) | -0.119(0.14) | -0.005(0.27) | 0.018(0.26) | 0(0.24) | -0.061(0.22) | -0.153(0.21) |
| Boundary | -0.007(0.22) | -0.013(0.20) | -0.035(0.20) | -0.068(0.18) | -0.097(0.13) | -0.002(0.28) | 0.024(0.26) | 0.009(0.25) | -0.030(0.23) | -0.113(0.20) |
| Hybrid | -0.002(0.22) | -0.005(0.20) | -0.026(0.20) | -0.058(0.18) | -0.088(0.13) | 0.005(0.29) | 0.035(0.26) | 0.021(0.24) | 0.004(0.20) | -0.074(0.17) |
| Tapered | -0.003(0.21) | -0.011(0.20) | -0.037(0.20) | -0.077(0.18) | -0.109(0.13) | 0.002(0.28) | 0.023(0.25) | 0.002(0.23) | -0.032(0.21) | -0.112(0.18) |
| Debiased | -0.011(0.21) | -0.018(0.19) | -0.040(0.20) | -0.070(0.19) | -0.090(0.15) | -0.007(0.27) | 0.021(0.25) | 0.010(0.24) | -0.039(0.24) | -0.140(0.23) |
| | **AR(1)**, $\{e_t\} \sim (\chi^2(2)-2)/2$, $n = 50$ | | | | | **MA(1)**, $\{e_t\} \sim (\chi^2(2)-2)/2$, $n = 50$ | | | | |
| Gaussian | 0.004(0.13) | -0.011(0.13) | -0.012(0.11) | -0.031(0.10) | -0.029(0.07) | 0.009(0.15) | 0.003(0.15) | 0.017(0.13) | 0.014(0.12) | 0.010(0.08) |
| Whittle | 0.001(0.13) | -0.016(0.13) | -0.019(0.12) | -0.044(0.10) | -0.049(0.07) | 0.005(0.14) | -0.004(0.14) | 0.004(0.14) | -0.020(0.13) | -0.065(0.12) |
| Boundary | 0.001(0.13) | -0.013(0.13) | -0.012(0.12) | -0.033(0.10) | -0.036(0.07) | 0.006(0.15) | 0.001(0.15) | 0.015(0.14) | 0.001(0.12) | -0.030(0.10) |
| Hybrid | 0.003(0.13) | -0.009(0.14) | -0.010(0.12) | -0.032(0.11) | -0.034(0.07) | 0.008(0.15) | 0.005(0.15) | 0.018(0.13) | 0.010(0.12) | -0.014(0.09) |
| Tapered | 0.003(0.13) | -0.011(0.14) | -0.013(0.12) | -0.036(0.11) | -0.038(0.07) | 0.007(0.15) | 0.004(0.15) | 0.014(0.13) | 0(0.11) | -0.026(0.08) |
| Debiased | 0.002(0.13) | -0.013(0.13) | -0.014(0.11) | -0.034(0.11) | -0.030(0.08) | 0.007(0.15) | 0.001(0.15) | 0.017(0.14) | 0.015(0.14) | -0.027(0.13) |
| | **AR(1)**, $\{e_t\} \sim (\chi^2(2)-2)/2$, $n = 300$ | | | | | **MA(1)**, $\{e_t\} \sim (\chi^2(2)-2)/2$, $n = 300$ | | | | |
| Gaussian | 0(0.06) | -0.005(0.05) | -0.004(0.05) | -0.004(0.04) | -0.006(0.03) | 0(0.06) | -0.002(0.05) | 0(0.05) | 0.003(0.04) | 0.003(0.03) |
| Whittle | -0.001(0.06) | -0.006(0.05) | -0.005(0.05) | -0.006(0.04) | -0.009(0.03) | 0(0.06) | -0.003(0.05) | -0.003(0.05) | -0.004(0.04) | -0.018(0.04) |
| Boundary | 0(0.06) | -0.005(0.05) | -0.004(0.05) | -0.004(0.04) | -0.007(0.03) | 0(0.06) | -0.002(0.05) | 0(0.05) | 0.002(0.04) | -0.002(0.03) |
| Hybrid | 0(0.06) | -0.006(0.06) | -0.004(0.05) | -0.004(0.04) | -0.007(0.03) | 0.001(0.06) | -0.002(0.06) | 0(0.05) | 0.003(0.04) | 0.002(0.03) |
| Tapered | 0(0.06) | -0.006(0.06) | -0.005(0.05) | -0.004(0.04) | -0.007(0.03) | 0.001(0.06) | -0.002(0.06) | 0(0.05) | 0.003(0.04) | 0.001(0.03) |
| Debiased | 0(0.06) | -0.005(0.05) | -0.004(0.05) | -0.004(0.04) | -0.006(0.03) | 0(0.06) | -0.002(0.05) | 0(0.05) | 0.003(0.05) | 0.013(0.05) |

Table C.1: Bias and the standard deviation (in the parentheses) of six different quasi-likelihoods for an AR(1) (left) and MA(1) (right) model for the standardized chi-squared innovations. Length of the time series $n = 20, 50$, and $300$. We use red text to denote the smallest RMSE and blue text to denote the second smallest RMSE.

## C.2 Misspecified model for a non-Gaussian time series

In this section, we provide figures and table of the results in Section 4.6.2 when the innovations follow a standardized chi-squared distribution two degrees of freedom, i.e. $\varepsilon_t \sim (\chi^2(2) - 2)/2$. The results are given in Tables C.2 and C.3.

| $n$ | Parameter | Gaussian | Whittle | Boundary | Hybrid | Tapered | Debiased |
|---|---|---|---|---|---|---|---|
| 20 | $\phi$ | 0.029(0.1) | -0.102(0.16) | -0.032(0.12) | -0.001(0.1) | -0.088(0.13) | 0.170(0.12) |
| | $\psi$ | 0.066(0.08) | -0.184(0.20) | -0.039(0.15) | 0.030(0.09) | -0.064(0.12) | 0.086(0.09) |
| | $I_n(f;f_\theta)$ | 1.573(0.82) | 1.377(3.11) | 0.952(0.91) | 1.006(0.84) | 0.675(0.63) | 2.618(0.84) |
| 50 | $\phi$ | 0.014(0.07) | -0.051(0.10) | -0.004(0.07) | 0.007(0.07) | -0.003(0.07) | 0.143(0.11) |
| | $\psi$ | 0.027(0.06) | -0.118(0.13) | -0.013(0.09) | 0.008(0.07) | 0.009(0.06) | 0.090(0.03) |
| | $I_n(f;f_\theta)$ | 0.342(0.34) | 0.478(0.53) | 0.298(0.32) | 0.230(0.27) | 0.222(0.27) | 1.158(0.37) |
| 300 | $\phi$ | 0.001(0.03) | -0.015(0.03) | -0.002(0.03) | 0(0.03) | -0.001(0.03) | 0.090(0.08) |
| | $\psi$ | 0.006(0.03) | -0.033(0.05) | 0.002(0.03) | 0.003(0.03) | 0.003(0.03) | 0.091(0.02) |
| | $I_n(f;f_\theta)$ | 0.029(0.05) | 0.067(0.10) | 0.034(0.06) | 0.027(0.04) | 0.028(0.04) | 0.747(0.23) |

Best fitting ARMA$(1,1)$ coefficients $\theta = (\phi, \psi)$ and spectral divergence:
- $\theta_{20} = (0.693, 0.845)$, $\theta_{50} = (0.694, 0.857)$, $\theta_{300} = (0.696, 0.857)$.
- $I_{20}(f;f_\theta) = 3.773$, $I_{50}(f;f_\theta) = 3.415$, $I_{300}(f;f_\theta) = 3.388$.

Table C.2: Best fitting (bottom lines) and the bias of estimated coefficients for six different methods for the ARMA$(3,2)$ misspecified case fitting ARMA$(1,1)$ model for the standardized chi-squared innovations. Standard deviations are in the parentheses. We use red text to denote the smallest RMSE and blue text to denote the second smallest RMSE.

## C.3 Alternative methods for estimating the predictive DFT results for a non-Gaussian time series

This time we assess the different estimation schemes for non-Gaussian time series. We generate the same AR$(8)$ model in Section 4.6.4 but the innovations $\{\varepsilon_t\}$ are i.i.d. standardarized chi-square random variables with two-degrees of freedom i.e. $\varepsilon_t \sim (\chi^2(2) - 2)/2$. For each simulation, we fit AR$(8)$ model, evaluate six likelihoods from the previous sections plus two likelihoods (BC-tYW and BC-NP), and calculate the parameter estimators. The results are summarized in Table C.4.

| $n$ | Parameter | Gaussian | Whittle | Boundary | Hybrid | Tapered | Debiased |
|---|---|---|---|---|---|---|---|
| 20 | $\phi_1$ | 0.017(0.13) | -0.178(0.23) | -0.047(0.17) | -0.006(0.14) | -0.134(0.15) | 0.044(0.14) |
| | $\phi_2$ | 0.002(0.09) | 0.176(0.2) | 0.057(0.16) | 0.023(0.12) | 0.135(0.13) | -0.019(0.13) |
| | $I_n(f;f_\theta)$ | 0.652(0.72) | 1.3073(1.46) | 0.788(0.85) | 0.671(0.8) | 0.887(0.97) | 0.658(0.81) |
| 50 | $\phi_1$ | 0.018(0.09) | -0.079(0.12) | -0.010(0.09) | 0.002(0.09) | -0.018(0.09) | 0.140(0.15) |
| | $\phi_2$ | -0.018(0.06) | 0.072(0.11) | 0.012(0.07) | 0.001(0.06) | 0.016(0.06) | -0.1(0.09) |
| | $I_n(f;f_\theta)$ | 0.287(0.36) | 0.406(0.52) | 0.302(0.39) | 0.298(0.39) | 0.293(0.38) | 0.631(0.7) |
| 300 | $\phi_1$ | 0.002(0.04) | -0.015(0.04) | -0.002(0.04) | 0(0.04) | -0.001(0.04) | 0.012(0.04) |
| | $\phi_2$ | -0.005(0.02) | 0.011(0.03) | -0.001(0.02) | -0.001(0.02) | -0.001(0.02) | -0.016(0.04) |
| | $I_n(f;f_\theta)$ | 0.050(0.07) | 0.056(0.07) | 0.051(0.07) | 0.052(0.07) | 0.054(0.08) | 0.061(0.08) |

Best fitting AR(1) coefficients $\theta = (\phi_1, \phi_2)$ and spectral divergence:
- $\theta_{20} = (1.367, -0.841)$, $\theta_{50} = (1.364, -0.803)$, $\theta_{300} = (1.365, -0.802)$.
- $I_{20}(f;f_\theta) = 2.902$, $I_{50}(f;f_\theta) = 2.937$, $I_{300}(f;f_\theta) = 2.916$.

Table C.3: Best fitting (bottom lines) and the bias of estimated coefficients for six different methods for the ARMA$(3,2)$ misspecified case fitting AR$(2)$ model for the standardized chi-squared innovations. Standard deviations are in the parentheses. We use red text to denote the smallest RMSE and blue text to denote the second smallest RMSE.

| Par. | Bias | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Gaussian | Whittle | Boundary | Hybrid | Tapered | Debiased | BC-tYW | BC-NP |
| $\phi_1$(0.381) | 0.001(0.08) | -0.013(0.09) | -0.002(0.09) | 0.001(0.09) | -0.003(0.09) | 0.004(0.09) | 0(0.09) | 0.001(0.12) |
| $\phi_2$(-0.294) | -0.001(0.09) | 0.014(0.1) | -0.001(0.09) | -0.002(0.09) | 0.006(0.09) | -0.008(0.11) | -0.002(0.09) | -0.010(0.13) |
| $\phi_3$(0.315) | -0.004(0.09) | -0.027(0.1) | -0.005(0.09) | -0.003(0.09) | -0.015(0.09) | 0(0.1) | -0.003(0.09) | -0.005(0.12) |
| $\phi_4$(-0.963) | 0.034(0.09) | 0.097(0.09) | 0.040(0.09) | 0.034(0.09) | 0.073(0.09) | 0.038(0.11) | 0.036(0.09) | 0.068(0.12) |
| $\phi_5$(0.285) | -0.007(0.09) | -0.032(0.09) | -0.009(0.09) | -0.005(0.09) | -0.018(0.09) | -0.004(0.1) | -0.007(0.09) | -0.005(0.12) |
| $\phi_6$(-0.240) | 0.007(0.09) | 0.029(0.09) | 0.009(0.09) | 0.006(0.09) | 0.018(0.09) | 0.003(0.1) | 0.007(0.09) | 0.006(0.12) |
| $\phi_7$(0.280) | -0.019(0.08) | -0.047(0.09) | -0.021(0.09) | -0.018(0.09) | -0.034(0.09) | -0.020(0.1) | -0.019(0.09) | -0.026(0.11) |
| $\phi_8$(-0.663) | 0.058(0.08) | 0.114(0.08) | 0.062(0.09) | 0.059(0.09) | 0.098(0.08) | 0.065(0.1) | 0.060(0.08) | 0.107(0.1) |
| $n\|\underline{\phi} - \widehat{\underline{\phi}}\|_2$ | 7.006 | 16.607 | 7.728 | 7.107 | 13.054 | 7.889 | 7.319 | 13.001 |

Table C.4: Bias and the standard deviation (in the parenthesis) of eight different quasi-likelihoods for the AR$(8)$ model for the standardized chi-squared innovations. Length of time series $n$=100. True AR coefficients are in the parenthesis of the first column. We use red text to denote the smallest RMSE and blue text to denote the second smallest RMSE.

APPENDIX D

TECHNICAL LEMMAS *

## D.1 Technical lemmas in Sections 2.3 and 3

The purpose of this section is to prove the main two lemmas which are required to prove Theorems 2.3.3 and 3.2.1.

**Lemma D.1.1.** *Suppose Assumption 2.3.2 holds. Let*

$$\breve{\mu}_\ell(\omega) = n^{-1}\sum_{t=1}^{n}(X_tX_\ell - \mathbb{E}[X_tX_\ell])e^{it\omega} \quad \textit{and} \quad \breve{c}_j = \widehat{c}_{j,n} - \mathbb{E}[\widehat{c}_{j,n}],$$

*where* $\widehat{c}_{j,n} = n^{-1}\sum_{t=1}^{n-|j|} X_tX_{t+|j|}$. *Then for any* $I$ *and* $J$ *of size* $r$ *and* $s$ *with* $r = 0, 1, 2$ *and* $r + s = m \geqslant 2$

$$\operatorname{cum}\left(\breve{\mu}_I^{\otimes^0}, \breve{c}_J^{\otimes^m}\right) = O\left(n^{-m+1}\right) \qquad r = 0, m \geqslant 2 \tag{D.1}$$

$$\operatorname{cum}\left(\breve{\mu}_I^{\otimes^1}, \breve{c}_J^{\otimes^{m-1}}\right) = \begin{cases} O\left(n^{-m}\right) & r = 1, m = 2 \\ O\left(n^{-m+1}\right) & r = 1, m \geqslant 3 \end{cases} \tag{D.2}$$

$$\operatorname{cum}\left(\breve{\mu}_I^{\otimes^2}, \breve{c}_J^{\otimes^{m-2}}\right) = \begin{cases} O\left(n^{-m+1}\right) & r = 2, m = 2, 3 \\ O\left(n^{-m+2}\right) & r = 2, m \geqslant 4 \end{cases} \tag{D.3}$$

*The next result is a little different to the above and concerns the bias of* $\widehat{c}_{j,n}$. *Suppose Assumption 2.3.1 (ii) holds. Then,*

$$\sup_{0\leqslant j\leqslant n} |\mathbb{E}[\widehat{c}_{j,n}] - c_j| = O(n^{-1}). \tag{D.4}$$

*Parts of this section have been modified with permission from [S. Das, S. Subba Rao, and J. Yang. Spectral methods for small sample time series: A complete periodogram approach. *Journal of Time Series Analysis (To appear)*, 2021, https://doi.org/10.1111/jtsa.12584.] and [S. Subba Rao and J. Yang. Reconciling the Gaussian and Whittle likelihood with an application to estimation in the frequency domain. *Annals of Statistics (To appear), arXiv:2001.06966*, 2021.]

PROOF. By assumption 2.3.1 *(ii)*, $\sup_{0\leqslant j\leqslant n} n|\mathbb{E}[\widehat{c}_j] - c_j| = \sup_{0\leqslant j\leqslant n} |jc_j| = O(1)$ as $n \to \infty$, thus (D.4) holds.

Before we show (D.1)$\sim$(D.3), it is interesting to observe the differences in rates. We first consider the very simple case and from this, we sketch how to generalize it. When $m = 2$,

$$\begin{aligned}
|\mathrm{cum}\,(\breve{\mu}_i, \breve{c}_j)\,| &\leqslant \frac{1}{n^2}\sum_{t=1}^{n}\sum_{\tau=1}^{n-|j|} |\mathrm{cov}(X_t X_i e^{it\omega}, X_\tau X_{\tau+j})| \\
&\leqslant \frac{1}{n^2}\sum_{t=1}^{n}\sum_{\tau=1}^{n-|j|} \big|\mathrm{cov}(X_t, X_\tau)\mathrm{cov}(X_i, X_{\tau+r}) \\
&\qquad +\mathrm{cov}(X_t, X_{\tau+j})\mathrm{cov}(X_i, X_\tau) + \mathrm{cum}(X_t, X_i, X_\tau, X_{\tau+j})\big|.
\end{aligned}$$

Under Assumption 2.3.2,

$$\sum_{t=1}^{n}\sum_{\tau=1}^{n-|j|} (|\kappa_2(t-\tau)\kappa_2(i-\tau+j)| + |\kappa_2(t-\tau-j)\kappa_2(i-\tau)| + |\kappa_4(i-t, \tau-t, \tau+j-t)|) < \infty$$

for all $n$. Thus

$$|\mathrm{cum}\,(\breve{\mu}_i, \breve{c}_j)\,| = O(n^{-2}).$$

This is in contrast to

$$\begin{aligned}
\mathrm{cum}\,(\breve{c}_{j_1}, \breve{c}_{j_2}) &= \frac{1}{n^2}\sum_{t=1}^{n-|j_1|}\sum_{\tau=1}^{n-|j_2|} \mathrm{cov}(X_t X_{t+j_1}, X_\tau X_{\tau+j_2}) \\
&= \frac{1}{n^2}\sum_{t=1}^{n-|j_1|}\sum_{\tau=1}^{n-|j_2|} \big[\mathrm{cov}(X_t, X_\tau)\mathrm{cov}(X_{t+j_1}, X_{\tau+j_2}) + \mathrm{cov}(X_t, X_{\tau+j_2})\mathrm{cov}(X_{t+j_1}, X_\tau) \\
&\qquad +\mathrm{cum}(X_t, X_{t+j_1}, X_\tau, X_{\tau+j_2})\big] \\
&= n^{-2}\sum_{t=1}^{n-|j_1|}\sum_{\tau=1}^{n-|j_2|} \big[\kappa_2(t-\tau)\kappa_2(t-\tau+j_1-j_2) + \kappa_2(t-\tau-j_2)\kappa_2(t-\tau+j_1) \\
&\qquad +\kappa_4(j_1, \tau-t, \tau-t+j_2)\big].
\end{aligned}$$

and

$$\begin{aligned}\operatorname{cum}(\breve{\mu}_{i_1},\breve{\mu}_{i_2}) &= \frac{1}{n^2}\sum_{t,\tau=1}^{n}\operatorname{cov}(X_tX_{i_1}e^{it\omega},X_\tau X_{i_2}e^{i\tau\omega})\\ &= n^{-2}\sum_{t,\tau=1}^{n}e^{i(t-\tau)\omega}\big[\kappa_2(t-\tau)\kappa_2(i_1-i_2)+\kappa_2(t-i_2)\kappa_2(\tau-i_1)\\ &\qquad +\kappa_4(i_1-t,\tau-t,i_2-t)\big].\end{aligned}$$

Unlike $\operatorname{cum}(\breve{\mu}_i,\breve{c}_j)$, there is a term that contains $(t-\tau)$ which cannot be separable. Thus

$$|\operatorname{cum}(\breve{c}_{j_1},\breve{c}_{j_2})| = O(n^{-1}),\quad |\operatorname{cum}(\breve{\mu}_{i_1},\breve{\mu}_{i_2})| = O(n^{-1}).$$

From the above examples, it is important to find the number of "free" parameters in each term of the indecomposable partition. For example, in $\operatorname{cum}(\breve{\mu}_i,\breve{c}_j)$ there are 3 possible indecomposable partitions, and for the first term, $|\kappa_2(t-\tau)\kappa_2(i-\tau+j)|$, we can reparametrize

$$z_1 = t-\tau,\quad z_2=\tau$$

then by the assumption,

$$n^{-2}\sum_{t=1}^{n}\sum_{\tau=1}^{n-|j|}|\kappa_2(t-\tau)\kappa_2(i-\tau+j)| \leqslant n^{-2}\sum_{z_1,z_2\in\mathbb{Z}}|\kappa_2(z_1)\kappa_2(i+j-z_2)| < Cn^{-2}.$$

However, for the first term of $\operatorname{cum}(\breve{c}_{j_1},\breve{c}_{j_2})$, $\kappa_2(t-\tau)\kappa_2(t-\tau+j_1-j_2)$, there is only one free parameter which is $(t-\tau)$ and thus gives a lower order, $O(n^{-1})$.

Lets consider the general order when $m>2$. To show (D.1), it is equivalent to show the number of "free" parameters in each indecomposable partition are at least $m-1$, then, gives an order at least $O(n^{-m+1})$ which proves (D.1). To show this, we use a mathmatical induction for $m$. We have shown above that (D.1) holds when $m=2$. Next, assume that (D.1) holds for $m$, and

consider

$$\operatorname{cum}\left(\breve{c}_J^{\otimes^m}, \breve{c}_j\right) = n^{-1} \sum_{t=1}^{n-|j|} \operatorname{cum}\left(\breve{c}_J^{\otimes^m}, X_t X_{t+j}\right) = n^{-(m+1)} \sum_{t=1}^{n-|j|} \sum_{v \in \Gamma} \operatorname{cum}_v \left(\breve{c}_J^{\otimes^m}, X_t X_{t+j}\right)$$

where $\Gamma$ is a set of all indecomposable partitions, and $\operatorname{cum}_v$ is a product of joint cumulants characterized by the partition $v$. Then, we can separate $\Gamma$ into 2 cases.

- The first case, $\Gamma_1$, is that the partition it still be an indecomposable partition for $\breve{c}_J^{\otimes^m}$ after removing $\{t, t+j\}$. In this case, by the induction hypothesis, there are at least $m-1$ free parameters in the partition, plus "$t$", thus at least $m$ free parameters.

- The second case, $\Gamma_2$, is that the partition becomes a decomposable partition for $\breve{c}_J^{\otimes^m}$ after removing $\{t, t+j\}$. Then, it is easy to show that $\Gamma_2 \backslash \{t, t+j\} = A \cup B$ where $A$ and $B$ are indecomposable partitions with elements $2a$ and $2b$ respectively where $a+b=m$. Moreover,$t$ and $t+j$ are in the different indecomposable partitions $A$ and $B$. In this case,

$$\sum_{t=1}^{n-|j|} \sum_{v \in \Gamma_2} \operatorname{cum}_v \left(\breve{c}_J^{\otimes^m}, X_t X_{t+j}\right) = \sum_{t=1}^{n-|j|} \sum_{v_1 \in A} \operatorname{cum}_{v_1} \left(\breve{c}_{J_A}^{\otimes^a}, X_t\right) \sum_{v_1 \in B} \operatorname{cum}_{v_2} \left(\breve{c}_{J_B}^{\otimes^b}, X_{t+j}\right).$$

In the first term $\left(\breve{c}_{J_A}^{\otimes^a}, X_t\right)$, there are at least $a-1$ free parameters plus "$t$", and thus $\sum_{v_1 \in A} \operatorname{cum}_{v_1} \left(\breve{c}_{J_A}^{\otimes^a}, X_t\right) = O(1)$, thus

$$n^{-m+1} \sum_{t=1}^{n-|j|} \sum_{v_1 \in A} \operatorname{cum}_{v_1} \left(\breve{c}_{J_A}^{\otimes^a}, X_t\right) \sum_{v_1 \in B} \operatorname{cum}_{v_2} \left(\breve{c}_{J_B}^{\otimes^b}, X_{t+j}\right) \leqslant C n^{-m+1} \sum_{t=1}^{n-|j|} 1 = O(n^{-m}).$$

Therefore, by induction (D.1) is true. For (D.2), when $m > 2$, it loses an order of one. For example, when $m = 3$

$$|\operatorname{cum}(\breve{\mu}_i, \breve{c}_{j_1}, \breve{c}_{j_2})| \leqslant \frac{1}{n^3} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n-j_1} \sum_{t_3=1}^{n-j_2} |\operatorname{cum}(X_{t_1} X_i, X_{t_2} X_{t_2+j_1}, X_{t_3} X_{t_3+j_2})|.$$

Then, above contains an indecomposable partition (see left panel of Figure D.1)

$$n^{-3}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n-j_1}\sum_{t_3=1}^{n-j_2}|\mathrm{cum}(X_{t_1},X_{t_2},X_{t_2+j_1})\mathrm{cum}(X_i,X_{t_3},X_{t_3+j_2})|$$
$$=n^{-3}\left(\sum_{t_1=1}^{n}\sum_{t_2=1}^{n-j_1}|\kappa_3(t_2-t_1,t_2-t_1+j_1)|\right)\left(\sum_{t_3=1}^{n-j_2}|\kappa_3(t_3,t_3+j_2-i)|\right)=O(n^{-2}).$$

Similarly, for (D.3), when $m=4$, $\mathrm{cum}(\breve{\mu}_{i_1},\breve{\mu}_{i_2},\breve{c}_{j_1},\breve{c}_{j_2})$ contains an indecomposable partition

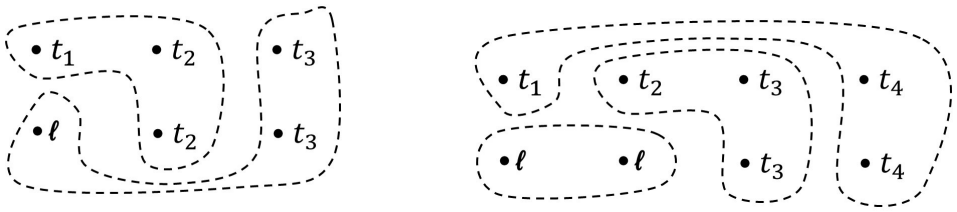



Figure D.1: Left: indecomposable partition of $\mathrm{cum}(\breve{\mu}_i,\breve{c}_{j_1},\breve{c}_{j_2})$. Right: indecomposable partition of $\mathrm{cum}(\breve{\mu}_{i_1},\breve{\mu}_{i_2},\breve{c}_{j_1},\breve{c}_{j_2})$

(see right panel of Figure D.1)

$$n^{-4}\sum_{t_1,t_2,t_3,t_4}|\mathrm{cum}(X_{t_1},X_{t_4},X_{t_4+j_2})\mathrm{cum}(X_{t_2},X_{t_3},X_{t_3+j_1})\mathrm{cum}(X_{i_1},X_{i_2})|$$
$$\leqslant Cn^{-4}\left(\sum_{t_1,t_4=1}^{n}|\kappa_3(t_4-t_1,t_4-t_1+j_2)|\right)\left(\sum_{t_2,t_3=1}^{n}|\kappa_3(t_3-t_2,t_3-t_2+j_1)|\right)=O(n^{-2}).$$

thus loses an order of two. Proof for (D.2) and (D.3) in a general case uses a similar induction argument from the above but we omit the proof. □

We now need to prove that the derivative of the random function $g(\cdot)$ defined in Theorem 2.3.3, equation (3.21) is bounded in probability. We recall these bounds are required to show that the final term in the Taylor expansion of $\widehat{J}_n(\omega;f_p)-\widehat{J}_n(\omega;f_p)$ with respect to $\{c_j\}_{j=0}^{p}$ is bounded in probability.

To do so, we define the following notation. Let $\underline{\widetilde{c}}_p = (\widetilde{c}_0, \widetilde{c}_1, \ldots, \widetilde{c}_p)'$ be a random vector such that $\underline{\widetilde{c}}_p$ is a convex combination of the true covariance vector $\underline{c}_p = (c_0, \ldots, c_p)'$ and the sample covariance vector $\underline{\widehat{c}}_p = (\widehat{c}_{0,n}, \ldots, \widehat{c}_{p,n})'$. Thus $\widetilde{c}_s$ is also a sample covariance that inherits many of the properties of the original sample covariance $\widehat{c}_{s,n}$. Based on these definitions we define the matrix and vector $\widetilde{R}_{p,n}$and $\underline{\widetilde{r}}_{p,n}$ where $(\widetilde{R}_{p,n})_{s,t} = \widetilde{c}_{s-t}$ and $(\underline{\widetilde{r}}_{p,n})_s = \widetilde{c}_s$. As our aim is to bound the derivatives in the proof Theorem 2.3.3, using (3.21) and (3.22) we define the random function

$$g_{\ell,p}(\omega, \underline{\widetilde{c}}_{p,n}) = \frac{\underline{\widetilde{r}}'_{p,n}\widetilde{R}^{-1}_{p,n}\underline{e}_\ell(\omega)}{1 - \underline{\widetilde{r}}'_p\widetilde{R}^{-1}_{p,n}\underline{e}_0(\omega)} = \frac{\widetilde{a}_{\ell,p}(\omega)}{1 - \widetilde{a}_{0,p}(\omega)} \tag{D.5}$$

where

$$\widetilde{a}_{\ell,p}(\omega) = \sum_{s=0}^{p-\ell} \widetilde{a}_{\ell+s,n} e^{-is\omega}, \quad a_0 \equiv 0,$$

$\underline{\widetilde{a}}_{p,n} = \widetilde{R}^{-1}_{p,n}\underline{\widetilde{r}}_{p,n}$ and $\underline{e}_\ell(\omega)$ is defined in (3.23). In the following lemma we show that the derivatives of $g_{\ell,p}(\omega, \underline{\widetilde{c}}_{p,n})$ are uniformly bounded in probability.

**Lemma D.1.2.** *Suppose Assumptions 2.3.1 and 2.3.2 hold with $m = 2$. For $1 \leqslant \ell \leqslant p$, let $g_{\ell,p}(\omega, \underline{\widetilde{c}}_p)$ be defined as in (D.5), where we recall $\underline{\widetilde{c}}_p$ denote a convex combination of the true covariances $\underline{c}_p = (c_0, \ldots, c_p)'$ and the sample autocovariances $\underline{\widehat{c}}_p = (\widehat{c}_{0,n}, \ldots, \widehat{c}_{p,n})'$.*

*If $p^{3/2}n^{-1/2} \to 0$ as $p, n \to \infty$, then for $k \in \mathbb{N}^+$ we have*

$$\sup_\omega \sup_{1 \leqslant \ell \leqslant p} \sup_{0 \leqslant j_1, \ldots, j_k \leqslant p} \left| \frac{\partial^k g_{\ell,p}(\omega, \underline{\widetilde{c}}_p)}{\partial \widetilde{c}_{j_1} \cdots \partial \widetilde{c}_{j_k}} \right| = O_p(1).$$

PROOF. First some simple preliminary comments are in order. We observe that $\widetilde{a}_{\ell,p}(\omega)$ is a linear function of $\underline{\widetilde{a}}_p = (\widetilde{a}_{1,p}, ..., \widetilde{a}_{p,p})'$ and $\underline{\widetilde{a}}_p = \widetilde{R}^{-1}_p\underline{\widetilde{r}}_p$. Therefore

$$g_{\ell,p}(\omega, \underline{\widetilde{c}}_{p,n}) = \frac{\underline{\widetilde{r}}'_{p,n}\widetilde{R}^{-1}_{p,n}\underline{e}_\ell(\omega)}{1 - \underline{\widetilde{r}}'_p\widetilde{R}^{-1}_{p,n}\underline{e}_0(\omega)} = \frac{\widetilde{a}_{\ell,p}(\omega)}{1 - \widetilde{a}_{0,p}(\omega)}$$

is an analytic function of $\underline{\widetilde{c}}_p$, thus for all $k$ we can evaluate its $k$ order partial derivative.

Since $g_{\ell,p}(\omega, \widetilde{\underline{c}}_{p,n})$ is a function of $\widetilde{\underline{a}}_p$ we require some consistency results on $\widetilde{\underline{a}}_p$. By Lemma D.1.1 (here we use Assumptions 2.3.1(ii) and 2.3.2), it is easy to show $\sup_s \mathbb{E}[\widehat{c}_{s,n} - c_s]^2 = O(n^{-1/2})$ and $\widetilde{c}_s$ is a convex combination of $\widehat{c}_{s,n}$ and $c_s$, then $\sup_s \mathbb{E}[\widetilde{c}_s - c_s]^2 = O(n^{-1/2})$. Thus since $\widetilde{\underline{a}}_p = \widetilde{R}_p^{-1}\widetilde{\underline{r}}_p$ we have

$$\left|\widetilde{\underline{a}}_p - \underline{a}_p\right|_1 = O_p(pn^{-1/2}). \tag{D.6}$$

where $|\cdot|_p$ is an $\ell_p$-norm. With this in hand, we can prove that the derivatives of $g_{\ell,p}(\omega, \widetilde{\underline{c}}_p)$ are uniformly bounded in probability. We give the precise details below.

In order to prove the result, we first consider the first derivative of $g_{\ell,p}(\omega, \widetilde{\underline{c}}_{p,n})$. By the chain rule, we have

$$\frac{\partial g_{\ell,p}}{\partial \widetilde{c}_j} = \sum_{r=1}^{p} \frac{\partial g_{\ell,p}}{\partial \widetilde{a}_{r,p}} \frac{\partial \widetilde{a}_{r,p}}{\partial \widetilde{c}_j} \tag{D.7}$$

where basic algebra gives

$$\frac{\partial g_{\ell,p}}{\partial \widetilde{a}_{r,p}} = \frac{e^{-ir\omega}}{(1-\widetilde{a}_{0,p}(\omega))^2} \times \begin{cases} \widetilde{a}_{\ell,p}(\omega) & r < \ell \\ e^{i\ell\omega}(1 - \sum_{s=1}^{\ell-1} \widetilde{a}_{s,p} e^{-is\omega}) & r \geqslant \ell \end{cases} \tag{D.8}$$

and

$$\left(\frac{\partial \widetilde{a}_{1,p}}{\partial \widetilde{c}_j}, \ldots, \frac{\partial \widetilde{a}_{p,p}}{\partial \widetilde{c}_j}\right)' = \frac{\partial \widetilde{\underline{a}}_p}{\partial \widetilde{c}_j} = \frac{\partial}{\partial \widetilde{c}_j} \widetilde{R}_p^{-1} \widetilde{\underline{r}}_p = \widetilde{R}_p^{-1} \left(\frac{\partial \widetilde{R}_p}{\partial \widetilde{c}_j}\right) \widetilde{\underline{a}}_p + \widetilde{R}_p^{-1} \frac{\partial \widetilde{\underline{r}}_p}{\partial \widetilde{c}_j}. \tag{D.9}$$

Therefore to bound (D.7) we take its absolute. We will bound the left hand side of an inequality below

$$\left|\frac{\partial g_{\ell,p}}{\partial \widetilde{c}_j}\right| \leqslant \sup_{\omega,s,\ell} \left|\frac{\partial g_{\ell,p}}{\partial \widetilde{a}_{s,p}}\right| \sum_{r=1}^{p} \left|\frac{\partial \widetilde{a}_{r,p}}{\partial \widetilde{c}_j}\right|, \tag{D.10}$$

which will prove the result for the first derivative. Therefore, we bound each term:

$\sup_{\omega,s,\ell} |\partial g_{\ell,p}/\partial \widetilde{a}_{s,p}|$ and $(\partial \widetilde{a}_{1,p}/\partial \widetilde{c}_j, \ldots, \partial \widetilde{a}_{p,p}/\partial \widetilde{c}_j)$.

A bound for $\sup_{\omega,s,\ell} |\partial g_{\ell,p}/\partial \widetilde{a}_{s,p}|$ Using (D.8) gives

$$\begin{aligned}\sup_{1\leqslant \ell,r\leqslant p}\sup_{\omega}\left|\frac{\partial g_{\ell,p}(\omega,\widetilde{c})}{\partial \widetilde{a}_{r,p}}\right| &\leqslant \sup_{\omega}\sup_{1\leqslant \ell\leqslant p}\frac{1}{|1-\sum_{s=1}^{p}\widetilde{a}_{s,p}e^{-is\omega}|^2}\left(\sum_{s=0}^{p-\ell}\left|\widetilde{a}_{s+\ell,p}e^{-is\omega}\right|+|1-\sum_{s=1}^{p}\widetilde{a}_{s,p}e^{-is\omega}|\right)\\ &\leqslant \sup_{\omega}\frac{1}{|1-\sum_{s=1}^{p}\widetilde{a}_{s,p}e^{-is\omega}|^2}\left(1+2\sum_{s=1}^{p}|\widetilde{a}_{s,p}|\right). \qquad \text{(D.11)}\end{aligned}$$

We first bound the denominator of the above. It is clear that

$$\begin{aligned}\inf_{\omega}|1-\sum_{s=1}^{p}\widetilde{a}_{s,p}e^{-is\omega}| &\geqslant \inf_{\omega}\left(|1-\sum_{s=1}^{p}a_{s,p}e^{-is\omega}|-|\sum_{s=1}^{p}(a_{s,p}-\widetilde{a}_{s,p})e^{-is\omega}|\right)\\ &\geqslant \inf_{\omega}\left(|1-\sum_{s=1}^{p}a_{s,p}e^{-is\omega}|-\sum_{s=1}^{p}|a_{s,p}-\widetilde{a}_{s,p}|\right).\end{aligned}$$

By using (D.6), we have $|\underline{a}_p - \underline{\widetilde{a}}_p|_1 = O_p(pn^{-1/2})$ thus for $pn^{-1/2} \to 0$, we have that $\sum_{s=1}^{p}|a_{s,p} - \widetilde{a}_{s,p}| = o_p(1)$. Moreover, by Assumption 2.3.1(i) (and the Baxter's inequality), the first term is bounded away from 0 for large $p$. Therefore, we conclude that $\inf_\omega |1-\sum_{s=1}^{p}\widetilde{a}_{s,p}e^{-is\omega}|$ is bounded away in probability from zero, thus giving

$$\frac{1}{|1-\sum_{s=1}^{p}\widetilde{a}_{s,p}e^{-is\omega}|^2} = O_p(1) \qquad \text{(D.12)}$$

as $n, p \to \infty$ and $pn^{-1/2} \to 0$. This bounds the denominator of (D.11). Next to bound the numerator in (D.11) we use again (D.6)

$$\sum_{s=1}^{p}|\widetilde{a}_{s,p}| \leqslant \sum_{s=1}^{p}|a_{s,p}| + \sum_{s=1}^{p}|\widetilde{a}_{s,p}-a_{s,p}| = O_p(1+pn^{-1/2}). \qquad \text{(D.13)}$$

Therefore, by (D.12) and (D.13) we have

$$\sup_{\omega}\sup_{1\leqslant \ell\leqslant p}\sup_{1\leqslant k\leqslant p}\left|\frac{\partial g_{\ell,p}(\omega,\widetilde{c})}{\partial \widetilde{a}_{k,p}}\right| = O_p(1). \qquad \text{(D.14)}$$

A bound for $(\partial\widetilde{a}_{1,p}/\partial\widetilde{c}_j, \ldots, \partial\widetilde{a}_{p,p}/\partial\widetilde{c}_j)$ We use the expansion in (D.9):

$$\left(\frac{\partial a_{1,p}}{\partial c_j}, \ldots, \frac{\partial a_{p,p}}{\partial c_j}\right)' = \frac{\partial \underline{a}_p}{\partial c_j} = \frac{\partial}{\partial c_j} R^{-1}\underline{r}_p = R_p^{-1}\left(\frac{\partial R_p}{\partial c_j}\right)\underline{a}_p + R_p^{-1}\frac{\partial \underline{r}_p}{\partial c_j}.$$

We observe that the structure of Toeplitz matrix of $R_p$ means that $\partial R_p/\partial c_j$ has ones on the lower and upper $j$th diagonal and is zero elsewhere and $\partial\widetilde{\underline{r}}_p/\partial c_j$ is one at the $j$th entry and zero elsewhere. Using these properties we have

$$\sup_{0\leqslant j\leqslant p}\left|\frac{\partial R_p}{\partial c_j}\underline{a}_p\right|_1 \leqslant 2\sum_{s=1}^{p}|a_{s,p}| \quad \text{and} \quad \sup_{0\leqslant j\leqslant p}\left|R_p^{-1}\frac{\partial \underline{r}_p}{\partial c_j}\right|_1 \leqslant \|R_p^{-1}\|_1$$

where $\|A\|_p$ is an operator norm induced by the vector $\ell_p$-norm. Therefore, using the above and the inequality $|A\underline{x}|_1 \leqslant \|A\|_1|\underline{x}|_1$ gives

$$\begin{aligned}\sup_{0\leqslant j\leqslant p}\sum_{s=1}^{p}\left|\frac{\partial\widetilde{a}_{s,p}}{\partial\widetilde{c}_j}\right| &\leqslant \left|\widetilde{R}_p^{-1}\frac{\partial\widetilde{R}_p}{\partial\widetilde{c}_j}\widetilde{\underline{a}}_p\right|_1 + \left|\widetilde{R}_p^{-1}\frac{\partial\widetilde{\underline{r}}_p}{\partial\widetilde{c}_j}\right|_1 \\ &\leqslant 2\|\widetilde{R}_p^{-1}\|_1\sum_{s=1}^{p}|\widetilde{a}_{s,p}| + \|\widetilde{R}_p^{-1}\|_1 \leqslant \|\widetilde{R}_p^{-1}\|_1\left(2\sum_{s=1}^{p}|\widetilde{a}_{s,p}| + 1\right), \quad \text{(D.15)}\end{aligned}$$

where we note that in (D.13) we have shown that $\sum_{s=1}^{p}|\widetilde{a}_{s,p}| = O_p(1 + pn^{-1/2})$. Next we show $\|\widetilde{R}_p^{-1}\|_1 = O_p(1)$. To do this we define the circulant matrix $C_p(f^{-1})$ where

$$(C_p(f^{-1}))_{u,v} = n^{-1}\sum_{k=1}^{p} f^{-1}\left(\frac{2\pi k}{p}\right)\exp\left(-i(u-v)\frac{2\pi k}{p}\right) = \sum_{r\in\mathbb{Z}} K_{f^{-1}}(u-v+rp)$$

with $K_{f^{-1}}(r) = (2\pi)^{-1}\int_0^{2\pi} f^{-1}(\omega)e^{-ir\omega}d\omega$. By using Theorem 3.2 in SY20,

$$\|R_p^{-1}\|_1 \leqslant \|C_p(f^{-1})\|_1 + \|R_p^{-1} - C_p(f^{-1})\|_1 \leqslant \|C_p(f^{-1})\|_1 + A(f)$$

where $A(f)$ is a finite constant that does not depend on $p$ (the exact form is given in SY20).

Furthermore we have

$$\|C_p(f^{-1})\|_1 = \max_{1\leqslant v\leqslant p}\sum_{u=1}^{p}|C_p(f^{-1})_{u,v}| \leqslant \max_{1\leqslant v\leqslant p}\sum_{u=1}^{p}\sum_{r\in\mathbb{Z}}|K_{f^{-1}}(u-v+rp)| = \sum_{r\in\mathbb{Z}}|K_{f^{-1}}(r)| < \infty.$$

altogether this gives $\|R_p^{-1}\|_1 = O(1)$. To bound the random matrix $\|\widetilde{R}_p^{-1}\|_1$ we use that

$$\|\widetilde{R}_p^{-1}\|_1 \leqslant \|R_p^{-1}\|_1 + \|\widetilde{R}_p^{-1} - R_p^{-1}\|_1 \leqslant \|R_p^{-1}\|_1 + \sqrt{p}\|\widetilde{R}_p^{-1} - R_p^{-1}\|_2.$$

By using similar argument to Corollary 1 in McMurry and Politis (2015), we have $\|\widetilde{R}_p^{-1} - R_p^{-1}\|_2 = O(pn^{-1/2})$. Thus, if $p^{3/2}n^{-1/2} \to 0$ as $p$ and $n \to \infty$, then $\|\widetilde{R}_p^{-1}\|_1 = O_p(1)$. Substituting this into (D.15) gives

$$\sup_{0\leqslant j\leqslant p}\sum_{s=1}^{p}\left|\frac{\partial\widetilde{a}_{s,p}}{\partial\widetilde{c}_j}\right| \;\leqslant\; \|\widetilde{R}_p^{-1}\|_1\left(2\sum_{s=1}^{p}|\widetilde{a}_{s,p}| + 1\right) = O_p(1). \tag{D.16}$$

• Bound for the first derivatives Substituting the two bounds above into (D.10), gives the bound for the first derivative:

$$\sup_{\omega}\left|\frac{\partial g_{\ell,p}(\omega,\underline{\widetilde{c}}_p)}{\partial\widetilde{c}_j}\right| \leqslant \sup_{\omega}\sup_{1\leqslant\ell\leqslant p}\sup_{1\leqslant k\leqslant p}\left|\frac{\partial g_{\ell,p}(\omega,\widetilde{c}_p)}{\partial\widetilde{a}_{k,p}}\right|\|\widetilde{R}_p^{-1}\|_1\left(2\sum_{s=1}^{p}|\widetilde{a}_{s,p}| + 1\right) = O_p\left(1\right).$$

• Bound for the second derivatives To simplify notation, we drop the subscript $p$ in $a_{k,p}$ (though we should keep in mind it is a function of $p$). Using the chain rule we have

$$\frac{\partial^2 g_{\ell,p}}{\partial c_i\partial c_j} = \sum_{r=1}^{p}\frac{\partial g_{\ell,p}}{\partial a_r}\cdot\frac{\partial^2 a_r}{\partial c_i\partial c_j} + \sum_{r_1,r_2=1}^{p}\frac{\partial^2 g_{\ell,p}}{\partial a_{r_1}\partial a_{r_2}}\cdot\frac{\partial a_{r_1}}{\partial c_i}\cdot\frac{\partial a_{r_2}}{\partial c_j}.$$

Thus taking absolute of the above gives

$$\left|\frac{\partial^2 g_{\ell,p}}{\partial c_i\partial c_j}\right| \leqslant \sup_{k,\omega}\left|\frac{\partial g_{\ell,p}}{\partial a_k}\right|\sum_{k=1}^{p}\left|\frac{\partial^2 a_k}{\partial c_i\partial c_j}\right| + \sup_{r_1,r_2,\omega}\left|\frac{\partial^2 g_{\ell,p}}{\partial a_{r_1}\partial a_{r_2}}\right|\left(\sum_{r=1}^{p}\left|\frac{\partial a_r}{\partial c_i}\right|\right)^2. \tag{D.17}$$

We now bound the terms in (D.17). We first consider the term $\partial^2 g_{\ell,p}/\partial a_k \partial a_t$, which by using (D.8) is

$$\frac{\partial^2 g_{\ell,p}}{\partial a_k \partial a_t} = \frac{e^{-i(k+t)\omega}}{(1-a_{0,p}(\omega))^3} \times \begin{cases} a_{\ell,p}(\omega) & k,t < \ell \\ e^{i\ell\omega}(1-\sum_{s=1}^{\ell-1} a_s e^{-is\omega}) + a_{\ell,p}(\omega) & k < \ell \leqslant t \\ e^{i\ell\omega}(1-\sum_{s=1}^{\ell-1} a_s e^{-is\omega}) & k,t \geqslant \ell \end{cases}$$

Therefore, using a similar argument as used to bound (D.14), we have

$$\sup_{\omega} \sup_{1\leqslant \ell,k,t\leqslant p} \left| \frac{\partial^2 g_{\ell,p}(\omega, \widetilde{\underline{c}}_p)}{\partial \widetilde{a}_k \partial \widetilde{a}_t} \right| = O_p(1) \tag{D.18}$$

with $pn^{-1/2} \to 0$ as $p \to \infty$ and $n \to \infty$

Next, we obtain a probabilistic bound for $|\partial^2 \widetilde{\underline{a}}/\partial \widetilde{c}_i \partial \widetilde{c}_j|_1$. Note that by (D.9)

$$\begin{aligned} \frac{\partial^2 \underline{a}_p}{\partial c_i \partial c_j} &= R_p^{-1}\left(\frac{\partial R_p}{\partial c_i}\right) R_p^{-1}\left(\frac{\partial R_p}{\partial c_j}\right) \underline{a}_p + R_p^{-1}\left(\frac{\partial R_p}{\partial c_i}\right) R_p^{-1} \frac{\partial \underline{r}_p}{\partial c_j} \\ &\quad + R_p^{-1}\left(\frac{\partial R_p}{\partial c_j}\right) R_p^{-1}\left(\frac{\partial R_p}{\partial c_i}\right) \underline{a}_p + R_p^{-1}\left(\frac{\partial R_p}{\partial c_j}\right) R_p^{-1} \frac{\partial \underline{r}_p}{\partial c_i} \\ &= R_p^{-1}\left(\frac{\partial R_p}{\partial c_i}\right) \frac{\partial \underline{a}_p}{\partial c_j} + R_p^{-1}\left(\frac{\partial R_p}{\partial c_j}\right) \frac{\partial \underline{a}_p}{\partial c_i}. \end{aligned}$$

Our focus will be on the first term of right hand side of the above. By symmetry, bound for the second term is the same. Using the submultiplicative of the operator norm we have

$$\left| R_p^{-1}\left(\frac{\partial R_p}{\partial c_i}\right) \frac{\partial \underline{a}_p}{\partial c_j} \right|_1 \leqslant \|R_p^{-1}\left(\frac{\partial R_p}{\partial c_i}\right)\|_1 \left|\frac{\partial \underline{a}_p}{\partial c_j}\right|_1 \leqslant \|R_p^{-1}\|_1 \|\frac{\partial R_p}{\partial c_i}\|_1 \left|\frac{\partial \underline{a}_p}{\partial c_j}\right|_1 \leqslant 2\|R_p^{-1}\|_1 \left|\frac{\partial \underline{a}_p}{\partial c_j}\right|_1.$$

Therefore by (D.16),

$$\sup_{0\leqslant i,j\leqslant p} \left| \frac{\partial^2 \widetilde{\underline{a}}_p}{\partial \widetilde{c}_i \partial \widetilde{c}_j} \right|_1 \leqslant 4\|\widetilde{R}_p^{-1}\|_1 \sup_{0\leqslant j\leqslant p} \left| \frac{\partial \widetilde{\underline{a}}_p}{\partial \widetilde{c}_j} \right|_1 = O_p(1). \tag{D.19}$$

The bounds in (D.18) and (D.19) gives bounds for two of the terms in (D.17). The remaining two terms in (D.17) involve only first derivatives and bounds for these terms are given in equations

(D.14) and (D.16). Thus by using (D.17) and the four bounds described above we have

$$\sup_{\omega}\sup_{1\leqslant\ell\leqslant p}\sup_{0\leqslant j_1,j_2\leqslant p}\left|\frac{\partial^2 g_{\ell,p}}{\partial\widetilde{c}_{j_1}\partial\widetilde{c}_{j_2}}(\omega,\underline{\widetilde{c}}_{p,n})\right| = O_p(1),$$

which gives a bound for the second derivative.

• Bounds for the higher order derivatives The bounds for the higher order derivatives follows a similar pattern. We bound the $m$th order derivatives

$$\frac{\partial^m g_{\ell,p}}{\partial\widetilde{a}_{t_1}\partial\widetilde{a}_{t_2}\ldots\partial\widetilde{a}_{t_m}} \quad \text{and} \quad \frac{\partial^m\underline{\widetilde{a}}_p}{\partial\widetilde{c}_{i_1}\partial\widetilde{c}_{i_1}\ldots\partial\widetilde{c}_{i_m}},$$

using the methods described above. In particular, we can show that

$$\left|\frac{\partial^m g_{\ell,p}}{\partial\widetilde{a}_{t_1}\partial\widetilde{a}_{t_2}\ldots\partial\widetilde{a}_{t_m}}\right| = O_p(1) \quad \text{and} \quad \left|\frac{\partial^m\underline{\widetilde{a}}}{\partial\widetilde{c}_{j_1}\partial\widetilde{c}_{j_1}\ldots\partial\widetilde{c}_{j_m}}\right| = O_p(1).$$

Since these bounds hold for $1 \leqslant m \leqslant k$, by using the chain rule we have

$$\sup_{\omega}\sup_{1\leqslant\ell\leqslant p}\sup_{0\leqslant j_1,\ldots,j_k\leqslant p}\left|\frac{\partial^k g_{\ell,p}}{\partial\widetilde{c}_{j_1}\partial\widetilde{c}_{j_2}\ldots\partial\widetilde{c}_{j_k}}(\omega,\underline{\widetilde{c}}_p)\right| = O_p(1).$$

This proves the lemma. □

Finally, we state the following lemma which is required to prove Theorem 2.3.3

**Lemma D.1.3.** *Suppose the same set of Assumptions in Theorem 2.3.3 holds. Let $E_{11}(\cdot),...,E_{32}(\cdot)$ is defined as in (3.25). Then, the following error bounds hold:*

*The first order expansion yields the bounds*

$$\begin{aligned}\mathbb{E}[E_{11}(\omega)] &= O\left(\frac{p^2}{n^2}\right), \quad \text{var}[E_{11}(\omega)] = O\left(\frac{p^4}{n^2}\right),\\ \mathbb{E}[E_{12}(\omega)] &= O\left(\frac{p^2}{n^2}\right), \quad \text{var}[E_{12}(\omega)] = O\left(\frac{p^4}{n^3}\right).\end{aligned}$$

*The second order expansion yields the bounds*

$$\begin{aligned}\mathbb{E}[E_{21}(\omega)] &= O\left(\frac{p^3}{n^2}\right), \quad \mathrm{var}[E_{21}(\omega)] = O\left(\frac{p^6}{n^3}\right), \\ \mathbb{E}[E_{22}(\omega)] &= O\left(\frac{p^3}{n^2}\right), \quad \mathrm{var}[E_{22}(\omega)] = O\left(\frac{p^6}{n^4}\right).\end{aligned}$$

*Altogether, the third order expansion yields the probablistic bounds*

$$E_{31}(\omega) = O_p\left(\frac{p^4}{n^2}\right) \qquad E_{32}(\omega) = O_p\left(\frac{p^4}{n^{5/2}}\right).$$

PROOF. Bound for $E_{11}(\omega)$ and $E_{12}(\omega)$

• Bound for $E_{11}(\omega)$: We partition $E_{11}(\omega)$ into the two terms

$$E_{11}(\omega) = \sum_{j=0}^{p}\sum_{\ell=1}^{p} \breve{\mu}_\ell \left(\widehat{c}_j - c_j\right) \frac{\partial g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_j} = E_{111}(\omega) + E_{112}(\omega)$$

where

$$E_{111}(\omega) = \sum_{j=0}^{p}\sum_{\ell=1}^{p} \breve{\mu}_\ell \breve{c}_j \frac{\partial g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_j} \quad \text{and} \quad E_{112}(\omega) = \sum_{j=0}^{p}\sum_{\ell=1}^{p} \breve{\mu}_\ell \left(\mathbb{E}[\widehat{c}_j] - c_j\right) \frac{\partial g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_j}.$$

We first bound $E_{111}(\omega)$;

$$\mathbb{E}[E_{111}(\omega)] = \sum_{j=0}^{p}\sum_{\ell=1}^{p} \mathrm{cum}(\breve{\mu}_\ell, \breve{c}_j) \frac{\partial g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_j}$$

Thus by using Lemma D.1.1 and D.1.2 we have

$$|\mathbb{E}[E_{111}(\omega)]| \leqslant \frac{C}{n^2} \sum_{j=0}^{p}\sum_{\ell=1}^{p} \left|\frac{\partial g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_j}\right| = O\left(\frac{p^2}{n^2}\right).$$

Next we consider the variance

$$\operatorname{var}[E_{111}(\omega)] \leqslant \sum_{j_1,j_2=0}^{p} \sum_{\ell_1,\ell_2=1}^{p} |\operatorname{cov}(\breve{\mu}_{\ell_1}\breve{c}_{j_1}, \breve{\mu}_{\ell_2}\breve{c}_{j_2})| \left| \frac{\partial g_{\ell_1,p}(\omega, \underline{c}_p)}{\partial c_{j_1}} \frac{\partial g_{\ell_2,p}(\omega, \underline{c}_p)}{\partial c_{j_2}} \right|.$$

Splitting the covariance gives

$$\begin{aligned}
&\operatorname{cov}(\breve{\mu}_{\ell_1}\breve{c}_{j_1}, \breve{\mu}_{\ell_2}\breve{c}_{j_2}) \\
&\quad = \operatorname{cov}(\breve{\mu}_{\ell_1}, \breve{\mu}_{\ell_2})\operatorname{cov}(\breve{c}_{j_1}, \breve{c}_{j_2}) + \operatorname{cov}(\breve{\mu}_{\ell_1}, \breve{c}_{j_2})\operatorname{cov}(\breve{\mu}_{\ell_2}, \breve{c}_{j_1}) + \operatorname{cum}(\breve{\mu}_{\ell_2}, \breve{c}_{j_1}, \breve{\mu}_{\ell_2}, \breve{c}_{j_2}).
\end{aligned}$$

By using Lemma D.1.1, the above is

$$|\operatorname{cov}(\breve{\mu}_{\ell_1}\breve{c}_{j_1}, \breve{\mu}_{\ell_2}\breve{c}_{j_2})| = O(n^{-2}),$$

thus by Lemma D.1.2

$$\operatorname{var}[E_{111}(\omega)] = \frac{C}{n^2} \sum_{j_1,j_2=1}^{p} \sum_{\ell_1,\ell_2=1}^{p} \left| \frac{\partial g_{\ell_1,p}(\omega, \underline{c}_p)}{\partial c_{j_1}} \frac{\partial g_{\ell_2,p}(\omega, \underline{c}_p)}{\partial c_{j_2}} \right| = O\left(\frac{p^4}{n^2}\right).$$

Next we consider $E_{112}(\omega)$:

$$\mathbb{E}[E_{112}(\omega)] = \sum_{j=1}^{p} \sum_{\ell=1}^{p} \underbrace{\mathbb{E}[\breve{\mu}_\ell]}_{=0} \left(\mathbb{E}[\widehat{c}_j] - c_j\right) \frac{\partial g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_j} = 0$$

and

$$\operatorname{var}[E_{112}(\omega)] = \sum_{j_1,j_2=1}^{p} \sum_{\ell_1,\ell_2=1}^{p} \operatorname{cov}(\breve{\mu}_{\ell_1}, \breve{\mu}_{\ell_2}) \left(\mathbb{E}[\widehat{c}_{j_1}] - c_{j_1}\right) \left(\mathbb{E}[\widehat{c}_{j_2}] - c_{j_2}\right) \frac{\partial g_{\ell_1,p}(\omega, \underline{c}_p)}{\partial c_{j_1}} \frac{\partial g_{\ell_2,p}(\omega, \underline{c}_p)}{\partial c_{j_2}}.$$

Again by using Lemma D.1.1 and D.1.2 (which gives $|\operatorname{cov}(\breve{\mu}_{\ell_1}, \breve{\mu}_{\ell_2})| \leqslant C/n$ and $|\mathbb{E}[\widehat{c}_{j_1}] - c_{j_1}| \leqslant$

$C/n$), a bound for the above is

$$\mathrm{var}[E_{112}(\omega)] \leqslant \frac{C}{n^3} \sum_{j_1,j_2=1}^{p} \sum_{\ell_1,\ell_2=1}^{p} \left| \frac{\partial g_{\ell_1,p}(\omega, \underline{c}_p)}{\partial c_{j_1}} \frac{\partial g_{\ell_2,p}(\omega, \underline{c}_p)}{\partial c_{j_2}} \right| = O\left(\frac{p^4}{n^3}\right).$$

Thus altogether we have

$$\mathbb{E}[E_{11}(\omega)] = O\left(\frac{p^2}{n^2}\right), \quad \mathrm{var}[E_{11}(\omega)] = O\left(\frac{p^4}{n^2}\right). \tag{D.20}$$

• Bound for $E_{12}(\omega)$. We partition $E_{12}(\omega)$ into the two terms

$$\begin{aligned} E_{12}(\omega) &= \sum_{j=0}^{p} \sum_{\ell=1}^{p} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}[X_t X_\ell] e^{it\omega} \left(\widehat{c}_j - c_j\right) \frac{\partial g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_j} \\ &= \frac{1}{n} \sum_{j=0}^{p} \sum_{\ell=1}^{p} f_{\ell,n}(\omega) \left(\widehat{c}_j - c_j\right) \frac{\partial g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_j} \\ &= E_{121}(\omega) + E_{122}(\omega) \end{aligned}$$

where

$$\begin{aligned} E_{121}(\omega) &= \frac{1}{n} \sum_{j=0}^{p} \sum_{\ell=1}^{p} f_{\ell,n}(\omega) \breve{c}_j \frac{\partial g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_j} \\ E_{122}(\omega) &= \frac{1}{n} \sum_{j=0}^{p} \sum_{\ell=1}^{p} f_{\ell,n}(\omega) (\mathbb{E}[\widehat{c}_j] - c_j) \frac{\partial g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_j}. \end{aligned}$$

We first bound $E_{121}(\omega)$:

$$\mathbb{E}[E_{121}(\omega)] = \frac{1}{n} \sum_{j=0}^{p} \sum_{\ell=1}^{p} f_{\ell,n}(\omega) \mathbb{E}[\breve{c}_j] \frac{\partial g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_j} = 0$$

and

$$\mathrm{var}[E_{121}(\omega)] = \frac{1}{n^2} \sum_{j_1,j_2=0}^{p} \sum_{\ell_1,\ell_2=1}^{p} f_{\ell_1,n}(\omega) f_{\ell_2,n}(\omega) \mathrm{cov}(\breve{c}_{j_1}, \breve{c}_{j_2}) \frac{\partial g_{\ell_1,p}(\omega, \underline{c}_p)}{\partial c_{j_1}} \frac{\partial g_{\ell_2,p}(\omega, \underline{c}_p)}{\partial c_{j_2}}.$$

By using Lemma D.1.1 and D.1.2, and (3.24) we have

$$\mathrm{var}[E_{122}(\omega)] = \frac{C}{n^3} \sum_{j_1,j_2=0}^{p} \sum_{\ell_1,\ell_2=1}^{p} \left| \frac{\partial g_{\ell_1,p}(\omega, \underline{c}_p)}{\partial c_{j_1}} \frac{\partial g_{\ell_2,p}(\omega, \underline{c}_p)}{\partial c_{j_2}} \right| = O\left(\frac{p^4}{n^3}\right).$$

Next we consider $E_{122}(\omega)$ (which is non-random), using (3.24) we have

$$|E_{122}(\omega)| \leqslant \frac{C}{n^2} \sum_{j=1}^{p} \sum_{\ell=1}^{p} \left| \frac{\partial g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_j} \right| = O\left(\frac{p^2}{n^2}\right).$$

Thus we have

$$\mathbb{E}[E_{12}(\omega)] = O\left(\frac{p^2}{n^2}\right) \qquad \mathrm{var}(E_{12}(\omega)) = O\left(\frac{p^4}{n^3}\right). \tag{D.21}$$

This gives a bound for the first order expansion. The bound for the second order expansion given below is similar.

Bound for $E_{21}(\omega)$ and $E_{22}(\omega)$ The proof closely follows the bounds for $E_{11}(\omega)$ and $E_{12}(\omega)$ but requires higher order moment conditions.

• Bound for $E_{21}(\omega)$: We have

$$\begin{aligned}
E_{21}(\omega) &= \frac{1}{2} \sum_{j_1,j_2=0}^{p} \sum_{\ell=1}^{p} \breve{\mu}_\ell \left(\widehat{c}_{j_1} - c_{j_1}\right) \left(\widehat{c}_{j_2} - c_{j_2}\right) \frac{\partial^2 g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_{j_1} \partial c_{j_2}} \\
&= \frac{1}{2} \sum_{j_1,j_2=0}^{p} \sum_{\ell=1}^{p} \breve{\mu}_\ell \left(\breve{c}_{j_1} + (\mathbb{E}[\widehat{c}_{j_1}] - c_{j_1})\right) \left(\breve{c}_{j_2} + (\mathbb{E}[\widehat{c}_{j_2}] - c_{j_2})\right) \frac{\partial^2 g_{\ell,p}(\omega, \underline{c}_p)}{\partial c_{j_1} \partial c_{j_2}} \\
&= E_{211}(\omega) + E_{212}(\omega)
\end{aligned}$$

where

$$
\begin{aligned}
E_{211}(\omega) &= \frac{1}{2}\sum_{j_1,j_2=0}^{p}\sum_{\ell=1}^{p}\breve{\mu}_\ell\breve{c}_{j_1}\breve{c}_{j_2}\frac{\partial^2 g_{\ell,p}(\omega,\underline{c}_p)}{\partial c_{j_1}\partial c_{j_2}} \\
E_{212}(\omega) &= \frac{1}{2}\sum_{j_1,j_2=0}^{p}\sum_{\ell=1}^{p}\breve{\mu}_\ell\breve{c}_{j_1}(\mathbb{E}[\widehat{c}_{j_2}]-c_{j_2})\frac{\partial^2 g_{\ell,p}(\omega,\underline{c}_p)}{\partial c_{j_1}\partial c_{j_2}}+\frac{1}{2}\sum_{j_1,j_2=0}^{p}\sum_{\ell=1}^{p}\breve{\mu}_\ell\breve{c}_{j_2}(\mathbb{E}[\widehat{c}_{j_1}]-c_{j_1})\frac{\partial^2 g_{\ell,p}(\omega,\underline{c}_p)}{\partial c_{j_1}\partial c_{j_2}} \\
&\quad+\frac{1}{2}\sum_{j_1,j_2=0}^{p}\sum_{\ell=1}^{p}\breve{\mu}_\ell(\mathbb{E}[\widehat{c}_{j_1}]-c_{j_1})(\mathbb{E}[\widehat{c}_{j_2}]-c_{j_2})\frac{\partial^2 g_{\ell,p}(\omega,\underline{c}_p)}{\partial c_{j_1}\partial c_{j_2}}.
\end{aligned}
$$

Comparing $E_{212}(\omega)$ with $E_{111}(\omega)$, we observe that $E_{212}(\omega)$ is the same order as $(p/n)E_{111}(\omega)$, i.e.

$$
\mathbb{E}[E_{212}(\omega)] = O\left(\frac{p^3}{n^3}\right) \quad \text{var}[E_{212}(\omega)] = O\left(\frac{p^6}{n^4}\right).
$$

Now we can evaluate the mean and variance of the "lead" term $E_{211}(\omega)$. To bound the mean and variance, we use the following decompositions together with Lemma D.1.1

$$
\mathbb{E}[\breve{\mu}_\ell\breve{c}_{j_1}\breve{c}_{j_2}] = \text{cum}(\breve{\mu}_\ell,\breve{c}_{j_1},\breve{c}_{j_2}) = O(n^{-2})
$$

and

$$
\text{cov}[\breve{\mu}_{\ell_1}\breve{c}_{j_1}\breve{c}_{j_2},\breve{\mu}_{\ell_2}\breve{c}_{j_3}\breve{c}_{j_4}] = \text{cov}(\breve{\mu}_{\ell_1},\breve{\mu}_{\ell_2})\text{cov}(\breve{c}_{j_1},\breve{c}_{j_3})\text{cov}(\breve{c}_{j_2},\breve{c}_{j_4}) + (\textit{lower order}) = O\left(n^{-3}\right).
$$

Therefore, using Lemma D.1.2 we get $\mathbb{E}[E_{211}(\omega)] = O(p^3n^{-2})$ and $\text{var}[E_{211}(\omega)] = O(p^6n^{-3})$. Thus combining the bounds for $E_{211}(\omega)$ and $E_{212}(\omega)$ we have

$$
\mathbb{E}[E_{21}(\omega)] = O\left(\frac{p^3}{n^2}\right) \qquad \text{var}[E_{21}(\omega)] = O\left(\frac{p^6}{n^3}\right). \tag{D.22}
$$

• Bound for $E_{22}(\omega)$ Next we consider $E_{22}(\omega)$

$$\begin{aligned} E_{22}(\omega) &= \frac{1}{2n}\sum_{j_1,j_2=0}^{p}\sum_{\ell=1}^{p} f_{\ell,n}(\omega)\left(\widehat{c}_{j_1}-c_{j_1}\right)\left(\widehat{c}_{j_2}-c_{j_2}\right)\frac{\partial^2 g_{\ell,p}(\omega,\underline{c}_p)}{\partial c_{j_1}\partial c_{j_2}} \\ &= \frac{1}{2n}\sum_{j_1,j_2=0}^{p}\breve{c}_{j_1}\breve{c}_{j_2}\sum_{\ell=1}^{p} f_{\ell,n}(\omega)\frac{\partial^2 g_{\ell,p}(\omega,\underline{c}_p)}{\partial c_{j_1}\partial c_{j_2}} + \text{(lower order term)}. \end{aligned}$$

By using Lemma D.1.1 we have

$$\mathbb{E}[E_{22}(\omega)] = O\left(\frac{p^3}{n^2}\right) \qquad \text{var}[E_{22}(\omega)] = O\left(\frac{p^6}{n^4}\right).$$

Probabilistic bounds for $E_{31}(\omega)$, $E_{32}(\omega)$. Unlike the first four terms, evaluating the mean and variance of $E_{31}(\omega)$ and $E_{32}(\omega)$ is extremely difficult, due to the random third order derivative $\partial^3 g_{\ell,p}(\omega,\widetilde{\underline{c}}_{p,n})/\partial\widetilde{c}_{j_1}\partial\widetilde{c}_{j_2}\partial\widetilde{c}_{j_3}$. Instead we obtain probabilistic rates.

• Probabilistic bound for $E_{31}(\omega)$: Using Lemma D.1.2, we have $\sup_{\omega,\ell,j_1,j_2,j_2}|\frac{\partial^3 g_{\ell,p}(\omega,\widetilde{\underline{c}}_{p,n})}{\partial\widetilde{c}_{j_1}\partial\widetilde{c}_{j_2}\partial\widetilde{c}_{j_3}}| = O_p(1)$ this allows us to take the term out of the summand:

$$\begin{aligned} |E_{31}(\omega)| &\leqslant \sup_{\omega,\ell,j_1,j_2,j_2}\left|\frac{\partial^3 g_{\ell,p}(\omega,\widetilde{\underline{c}}_{p,n})}{\partial\widetilde{c}_{j_1}\partial\widetilde{c}_{j_2}\partial\widetilde{c}_{j_3}}\right|\frac{1}{3!}\sum_{j_1,j_2,j_3=0}^{p}\sum_{\ell=1}^{p}|\breve{\mu}_\ell\left(\widehat{c}_{j_1}-c_{j_1}\right)\left(\widehat{c}_{j_2}-c_{j_2}\right)\left(\widehat{c}_{j_3}-c_{j_3}\right)| \\ &= O_p(1)\sum_{j_1,j_2,j_3=0}^{p}\sum_{\ell=1}^{p}|\breve{\mu}_\ell\left(\widehat{c}_{j_1}-c_{j_1}\right)\left(\widehat{c}_{j_2}-c_{j_2}\right)\left(\widehat{c}_{j_3}-c_{j_3}\right)| \end{aligned}$$

Thus the analysis of the above hinges on obtaining a bound for $\mathbb{E}\,|\breve{\mu}_\ell\left(\widehat{c}_{j_1}-c_{j_1}\right)\left(\widehat{c}_{j_2}-c_{j_2}\right)\left(\widehat{c}_{j_3}-c_{j_3}\right)|$, whose leading term is $\mathbb{E}\,|\breve{\mu}_\ell\breve{c}_{j_1}\breve{c}_{j_2}\breve{c}_{j_3}|$. We use that $\mathbb{E}|A| \leqslant var[A]^{1/2} + |\mathbb{E}[A]|$ to bound this term by deriving bounds for its mean and variance. By using Lemma D.1.1, expanding $\mathbb{E}\left[\breve{\mu}_\ell\breve{c}_{j_1}\breve{c}_{j_2}\breve{c}_{j_3}\right]$ in terms of covariances and cumulants gives

$$\mathbb{E}\left[\breve{\mu}_\ell\breve{c}_{j_1}\breve{c}_{j_2}\breve{c}_{j_3}\right] = \sum_{\{a,b,c\}=\{1,2,3\}}\text{cov}(\breve{\mu}_\ell,\breve{c}_{j_a})\text{cov}(\breve{c}_{j_b},\breve{c}_{j_b}) + \text{cum}\left[\breve{\mu}_\ell,\breve{c}_{j_1},\breve{c}_{j_2},\breve{c}_{j_3}\right] = O(n^{-3})$$

and

$$\mathrm{var}[\breve{\mu}_\ell \breve{c}_{j_1}\breve{c}_{j_2}\breve{c}_{j_3}] = \mathrm{var}(\breve{\mu}_\ell)\prod_{s=1}^{3}\mathrm{var}(\breve{c}_{j_s}) + \ldots + \mathrm{cum}\left(\breve{\mu}_I^{\otimes^2}, \breve{c}_J^{\otimes^6}\right) + \mathrm{cum}(\breve{\mu}_\ell, \breve{c}_{j_1}, \breve{c}_{j_2}, \breve{c}_{j_3})^2 = O(n^{-4}).$$

This gives $\mathbb{E}\left|\breve{\mu}_\ell\left(\widehat{c}_{j_1} - c_{j_1}\right)\left(\widehat{c}_{j_2} - c_{j_2}\right)\left(\widehat{c}_{j_3} - c_{j_3}\right)\right| = O(n^{-2})$, therefore

$$E_{31}(\omega) \;=\; O_p\left(\frac{p^4}{n^2}\right).$$

• Probabilistic bound for $E_{32}(\omega)$: Again taking the third order derivaive out of the summand gives

$$\begin{aligned} E_{32}(\omega) &\leqslant \sup_{\omega,\ell,j_1,j_2,j_2}\left|\frac{\partial^3 g_{\ell,p}(\omega,\widetilde{\underline{c}}_{p,n})}{\partial\widetilde{c}_{j_1}\partial\widetilde{c}_{j_2}\partial\widetilde{c}_{j_3}}\right|\frac{1}{3!n}\sum_{j_1,j_2,j_3=0}^{p}\sum_{\ell=1}^{p}|f_{\ell,n}(\omega)|\left|\left(\widehat{c}_{j_1} - c_{j_1}\right)\left(\widehat{c}_{j_2} - c_{j_2}\right)\left(\widehat{c}_{j_3} - c_{j_3}\right)\right| \\ &= O_p(n^{-1})\sum_{j_1,j_2,j_3=0}^{p}\sum_{\ell=1}^{p}\left|\left(\widehat{c}_{j_1} - c_{j_1}\right)\left(\widehat{c}_{j_2} - c_{j_2}\right)\left(\widehat{c}_{j_3} - c_{j_3}\right)\right|. \end{aligned}$$

Using Lemma D.1.1 to evaluate the mean and variance of $\breve{c}_{j_1}\breve{c}_{j_2}\breve{c}_{j_3}$ we have

$$\mathbb{E}[\breve{c}_{j_1}\breve{c}_{j_2}\breve{c}_{j_3}] = O(n^{-2}) \quad \text{and} \quad \mathrm{var}[\breve{c}_{j_1}\breve{c}_{j_2}\breve{c}_{j_3}] = O(n^{-3}),$$

thus, $E_{32}(\omega) = O_p\left(\frac{p^4}{n^{5/2}}\right)$. □

### D.2 Technical lemmas in Sections 2.4 and 4

In the case that the spectral density $f$ corresponds to an AR($p$) model, $\phi_j(\tau; f) = \sum_{s=0}^{p-j}\phi_{j+s}\psi_{|\tau|-s}$ for $\tau \leqslant 0$. This result is well known (see Inoue and Kasahara (2006), page 980). However we could not find the proof, thus for completeness we give the proof below.

**Lemma D.2.1.** *Suppose $f_\theta(\omega) = \sigma^2|1 - \sum_{j=1}^{p}\phi_j e^{-ij\omega}|^{-2} = \sigma^2|\sum_{j=0}^{\infty}\psi_j e^{-ij\omega}|^2$, where $\{\phi_j\}_{j=1}^{p}$ correspond to the causal AR($p$) representation. Let $\phi_j(\tau; f)$ be defined as in (2.1). Then $\phi_j(\tau; f) =$*

$\sum_{s=0}^{p-j} \phi_{j+s}\psi_{|\tau|-s}$.

$$[A_p(\underline{\phi})^{|\tau|+1}\underline{X}_p]_{(1)} = \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s}\psi_{|\tau|-s}. \tag{D.23}$$

*where we set* $\psi_j = 0$ *for* $j < 0$.

PROOF. To simplify notation let $A = A_p(\underline{\phi})$. The proof is based on the observation that the $j$th row of $A^m$ ($m \geqslant 1$) is the $(j-1)$th row of $A^{m-1}$ (due to the structure of $A$). Let $(a_{1,m}, \ldots, a_{p,m})$ denote the first row of $A^m$. Using this notation we have

$$\begin{pmatrix} a_{1,m} & a_{2,m} & \ldots & a_{p,m} \\ a_{1,m-1} & a_{2,m-1} & \ldots & a_{p,m-1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,m-p+1} & a_{2,m-p+1} & \ldots & a_{p,m-p+1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \ldots & \phi_{p-1} & \phi_p \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \begin{pmatrix} a_{1,m-1} & a_{2,m-1} & \ldots & a_{p,m-1} \\ a_{1,m-2} & a_{2,m-2} & \ldots & a_{p,m-2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,m-p} & a_{2,m-p} & \ldots & a_{p,m-p} \end{pmatrix}.$$

From the above we observe that $a_{\ell,m}$ satisfies the system of equations

$$\begin{aligned} a_{\ell,m} &= \phi_\ell a_{1,m-1} + a_{\ell+1,m-1} \qquad 1 \leqslant \ell \leqslant p-1 \\ a_{p,m} &= \phi_p a_{1,m-1}. \end{aligned} \tag{D.24}$$

Our aim is to obtain an expression for $a_{\ell,m}$ in terms of $\{\phi_j\}_{j=1}^p$ and $\{\psi_j\}_{j=0}^\infty$ which we now define. Since the roots of $\phi(\cdot)$ lies outside the unit circle the function $(1-\sum_{j=1}^p \phi_j z^j)^{-1}$ is well defined for $|z| \leqslant 1$ and has the power series expansion $(1-\sum_{i=1}^p \phi_i z)^{-1} = \sum_{i=0}^\infty \psi_i z^i$ for $|z| \leqslant 1$. We use the well know result $[A^m]_{1,1} = a_{1,m} = \psi_m$ (which can be proved by induction). Using this we obtain an expression for the coefficients $\{a_{\ell,m}; 2 \leqslant \ell \leqslant p\}$ in terms of $\{\phi_i\}$ and $\{\psi_i\}$. Solving the system

of equations in (D.24), starting with $a_{1,1} = \psi_1$ and recursively solving for $a_{p,m}, \ldots, a_{2,m}$ we have

$$\begin{aligned} a_{p,r} &= \phi_p \psi_{r-1} && m-p \leqslant r \leqslant m \\ a_{\ell,r} &= \phi_\ell a_{1,r-1} + a_{\ell+1,r-1} && 1 \leqslant \ell \leqslant p-1, \quad m-p \leqslant r \leqslant m \end{aligned}$$

This gives $a_{p,m} = \phi_p \psi_{m-1}$, for $\ell = p-1$

$$\begin{aligned} a_{p-1,m} &= \phi_{p-1} a_{1,m-1} + a_{p,m-1} \\ &= \phi_{p-1}\psi_{m-1} + \psi_p \psi_{m-2} \end{aligned}$$

$$\begin{aligned} a_{p-2,m} &= \phi_{p-2} a_{1,m-1} + a_{p-1,m-1} \\ &= \phi_{p-2}\psi_{m-1} + \phi_{p-1}\psi_{m-2} + \psi_p \psi_{m-3} \end{aligned}$$

up to

$$\begin{aligned} a_{1,m} &= \phi_1 a_{1,m-1} + a_{2,m-1} \\ &= \sum_{s=0}^{p-1} \phi_{1+s}\psi_{m-1-s} = (\psi_m). \end{aligned}$$

This gives the general expression

$$a_{p-r,m} = \sum_{s=0}^{r} \phi_{p-r+s}\psi_{m-1-s} \qquad 0 \leqslant r \leqslant p-1.$$

In the last line of the above we change variables with $\ell = p - r$ to give for $m \geqslant 1$

$$a_{\ell,m} = \sum_{s=0}^{p-\ell} \phi_{\ell+s}\psi_{m-1-s} \qquad 1 \leqslant \ell \leqslant p,$$

where we set $\psi_0 = 1$ and for $t < 0$, $\psi_t = 0$. Therefore

$$[A^{|\tau|+1}\underline{X}_p]_{(1)} = \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s}\psi_{|\tau|-s}.$$

Thus we obtain the desired result. □

**Lemma D.2.2.** *Let $\{\phi_j(f_\theta)\}$ and $\{\psi_j(f_\theta)\}$ denote the AR($\infty$), and MA($\infty$) coefficients corresponding to the spectral density $f_\theta$. Suppose the same set of Assumptions in Lemma A.2.1 holds. Let $\{\alpha_j(f_\theta)\}$ denote the Fourier coefficients in the one-sided expansion*

$$\log(\sum_{j=0}^{\infty} \psi_j(f_\theta)z^j) = \sum_{j=1}^{\infty} \alpha_j(f_\theta)z^j \quad \textit{for } |z| < 1, \tag{D.25}$$

*Then for all $\theta \in \Theta$ and for $0 \leqslant s \leqslant \kappa$ we have*

$$\sum_{j=1}^{\infty} \|j^K \nabla_\theta^s \alpha_j(f_\theta)\|_1 < \infty,$$

PROOF. We first consider the case $s = 0$. The derivative of (D.25) with respect to $z$ together with $\psi(z; f_\theta)^{-1} = \phi(z; f_\theta) = 1 - \sum_{j=1}^{\infty} \phi_j(f_\theta)z^j$ gives

$$\begin{aligned} \sum_{j=1}^{\infty} j\alpha_j(f_\theta)z^{j-1} &= \left(\sum_{j=1}^{\infty} j\psi_j(f_\theta)z^{j-1}\right)\left(\sum_{j=0}^{\infty} \psi_j(f_\theta)z^j\right)^{-1} \\ &= \left(\sum_{j=1}^{\infty} j\psi_j(f_\theta)z^{j-1}\right)\left(\sum_{j=0}^{\infty} \widetilde{\phi}_j(f_\theta)z^j\right), \end{aligned} \tag{D.26}$$

where $\sum_{j=0}^{\infty} \widetilde{\phi}_j(f_\theta)z^j = 1 - \sum_{j=1}^{\infty} \phi_j(f_\theta)z^j$. Comparing the coefficients of $z^{j-1}$ from both side of above yields the identity

$$\begin{aligned} j\alpha_j(f_\theta) &= \sum_{\ell=0}^{j-1} (j-\ell)\psi_{j-\ell}(f_\theta)\widetilde{\phi}_\ell(f_\theta) \\ \Rightarrow \quad \alpha_j(f_\theta) &= j^{-1}\sum_{\ell=0}^{j-1} (j-\ell)\psi_{j-\ell}(f_\theta)\widetilde{\phi}_\ell(f_\theta) \quad \text{for} \quad j \geqslant 1. \end{aligned} \tag{D.27}$$

Therefore, using the above and taking the absolute into the summand we have

$$\begin{aligned}
\sum_{j=1}^{\infty} j^K|\alpha_j(f_\theta)| &\leqslant \sum_{j=1}^{\infty}\sum_{\ell=0}^{j-1} j^{K-1}(j-\ell)|\psi_{j-\ell}(f_\theta)||\widetilde{\phi}_\ell(f_\theta)| \\
&= \sum_{\ell=1}^{\infty}|\widetilde{\phi}_\ell(f_\theta)|\sum_{j=\ell+1}^{\infty} j^{K-1}(j-\ell)|\psi_{j-\ell}(f_\theta)| \qquad \text{(exchange summation)} \\
&= \sum_{\ell=1}^{\infty}|\widetilde{\phi}_\ell(f_\theta)|\sum_{s=1}^{\infty}(s+\ell)^{K-1}s|\psi_s(f_\theta)| \qquad \text{(change of variable } s=j+\ell) \\
&\leqslant \sum_{\ell=1}^{\infty}|\widetilde{\phi}_\ell(f_\theta)|\sum_{s=1}^{\infty}(s+\ell)^{K}|\psi_s(f_\theta)|. \qquad ((s+\ell)^{-1}\leqslant s^{-1})
\end{aligned}$$

Since $K \geqslant 1$, using inequality $(a+b)^K \leqslant 2^{K-1}(a^K+b^K)$ for $a, b > 0$, we have

$$\begin{aligned}
\sum_{j=1}^{\infty} j^K|\alpha_j(f_\theta)| &\leqslant \sum_{\ell=1}^{\infty}|\widetilde{\phi}_\ell(f_\theta)|\sum_{s=1}^{\infty}(s+\ell)^{K}|\psi_s(f_\theta)| \\
&\leqslant 2^{K-1}\sum_{\ell=1}^{\infty}|\widetilde{\phi}_\ell(f_\theta)|\sum_{s=1}^{\infty}(s^K+\ell^K)|\psi_s(f_\theta)| \\
&= 2^{K-1}\bigg(\sum_{\ell=1}^{\infty}\ell^K|\widetilde{\phi}_\ell(f_\theta)|\cdot\sum_{s=1}^{\infty}|\psi_s(f_\theta)| + \sum_{\ell=1}^{\infty}|\widetilde{\phi}_\ell(f_\theta)|\cdot\sum_{s=1}^{\infty}s^K|\psi_s(f_\theta)|\bigg) \leqslant \infty
\end{aligned}$$

and this proves the lemma when $s = 0$.

To prove lemma for $s = 1$, we differentiate (D.27) with $\theta$, then, by Assumption 4.2.1 (iii),

$$\begin{aligned}
&\sum_{j=1}^{\infty} j^K\|\nabla_\theta\alpha_j(f_\theta)\|_1 \\
&\quad\leqslant \sum_{j=1}^{\infty}\sum_{\ell=0}^{j-1}\|j^{K-1}(j-\ell)\nabla_\theta\psi_{j-\ell}(f_\theta)\widetilde{\phi}_\ell(f_\theta)\|_1 + \sum_{j=1}^{\infty}\sum_{\ell=0}^{j-1}\|j^{K-1}(j-\ell)\psi_{j-\ell}(f_\theta)\nabla_\theta\widetilde{\phi}_\ell(f_\theta)\|_1
\end{aligned}$$

Using similar technique to prove $s = 0$, we show $\sum_{j=1}^{\infty}\|j^K\nabla_\theta\alpha_j(f_\theta)\|_1 < \infty$ and the proof for $s \geqslant 2$ is similar (we omit the detail). □