Reconciling the Gaussian and Whittle Likelihood with an application to estimation in the frequency domain

Suhasini Subba Rao[*] and Junho Yang[†]

September 28, 2020

Abstract

In time series analysis there is an apparent dichotomy between time and frequency domain methods. The aim of this paper is to draw connections between frequency and time domain methods. Our focus will be on reconciling the Gaussian likelihood and the Whittle likelihood. We derive an exact, 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 discrete Fourier transform 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. We show that the difference between the Gaussian and Whittle likelihood is due to the omission of the best linear predictions outside the domain of observation in the periodogram associated with the Whittle likelihood. Based on this result, we obtain an approximation for the difference between the Gaussian and Whittle likelihoods in terms of the best fitting, finite order autoregressive parameters. These approximations are used to define two new frequency domain quasi-likelihoods criteria. We show that these new criteria can yield a better approximation of the spectral divergence criterion, as compared to both the Gaussian and Whittle likelihoods. In simulations, we show that the proposed estimators have satisfactory finite sample properties.

Keywords and phrases: Biorthogonal transforms, discrete Fourier transform, periodogram, quasi-likelihoods and second order stationary time series.

1 Introduction

In his seminal work, Whittle (1951, 1953) introduced the Whittle likelihood as an approximation of the Gaussian likelihood. A decade later, the asymptotic sampling properties of moving average models fitted using the Whittle likelihood were derived in Walker (1964). Subsequently, the

---

[*]Texas A&M University, College Station, Texas, TX 77845, U.S.A.

[†]Authors ordered alphabetically

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 technology, interest in the Whittle likelihood has not abated. The Whittle likelihood has gained further traction as a quasi-likelihood (or as an information criterion, see Parzen (1983)) between the periodogram and the spectral density. Several diverse applications of the Whittle likelihood can be found in Fox and Taqqu (1986), Dahlhaus and Künsch (1987) (for spatial processes), Robinson (1995), Dahlhaus (2000), Hurvich and Chen (2000), Giraitis and Robinson (2001), Choudhuri et al. (2004), Abadir et al. (2007), Shao and Wu (2007), Giraitis et al. (2012) (long memory time series and local Whittle methods), Panaretos and Tavakoli (2013), Kirch et al. (2019) (Bayesian spectral methods), and van Delft and Eichler (2020) (functional time series), to name but a few.

Despite its advantages, it is well known that for small samples the Whittle likelihood can give rise to estimators with a substantial bias (see Priestley (1981) and Dahlhaus (1988)). Dahlhaus (1988) shows that the finite sample bias in the periodogram impacts the performance of the Whittle likelihood. Motivated by this discrepancy, Sykulski et al. (2019) proposes the debiased Whittle likelihood, which fits directly to the expectation of the periodogram rather than the limiting spectral density. Alternatively, Dahlhaus (1988) shows that the tapered periodogram is better at capturing the features in the spectral density, such as peaks, than the regular periodogram. He uses this as the basis of the tapered Whittle likelihood. Empirical studies show that the tapered Whittle likelihood yields a smaller bias than the regular Whittle likelihood. As a theoretical justification, Dahlhaus (1988, 1990) uses 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 of the Toeplitz matrix. These results can be used to approximate the Gaussian likelihood.

However, as far as we are aware, there are no results which explain what is lost when using the Whittle likelihood rather than the Gaussian likelihood. The objective of this paper is to address some of these issues. The benefits of such insights are not only of theoretical interest but also lead to the development of computationally simple frequency domain methods which are comparable with the Gaussian likelihood.

We first recall the definition of the Gaussian and Whittle likelihood. Our aim is to fit a parametric second order stationary model with spectral density $f_\theta(\omega)$ and corresponding auto-covariance function $\{c_{f_\theta}(r)\}_{r \in \mathbb{Z}}$ to the observed time series $\{X_t\}_{t=1}^n$. The (quasi) log-Gaussian likelihood 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 $\Gamma_n(f_\theta)_{s,t} = c_{f_\theta}(s-t)$, $|A|$ 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 candidate spectral density. There are two subtly different methods for defining this contrast, one is with an integral the other is to use the Riemann sum. In this paper, we focus on the Whittle likelihood defined in terms of the Riemann sum over the fundamental frequencies

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

where $J_n(\omega_{k,n}) = n^{-1/2} \sum_{t=1}^{n} X_t e^{it\omega_{k,n}}$ is the discrete Fourier transform (DFT) of the observed time series. To compare the Gaussian and Whittle likelihood, we rewrite the Whittle likelihood in matrix form. We define the $n \times n$ circulant matrix $C_n(f_\theta)$ with entries $(C_n(f_\theta))_{s,t} = n^{-1} \sum_{k=1}^{n} f_\theta(\omega_{k,n}) e^{-i(s-t)\omega_{k,n}}$. The Whittle likelihood $K_n(\theta; \underline{X}_n)$ can be written as

$$K_n(\theta; \underline{X}_n) = n^{-1} \left(\underline{X}'_n C_n(f_\theta^{-1}) \underline{X}_n + \sum_{k=1}^{n} \log f_\theta(\omega_{k,n}) \right). \tag{1.3}$$

To obtain an exact expression for $\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1})$ and $\underline{X}'_n[\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1})]\underline{X}_n$, we focus on the DFT of the time series. The idea is to obtain the linear transformation of the observed time series $\{X_t\}_{t=1}^{n}$ which is biorthogonal to the regular DFT, $\{J_n(\omega_{k,n})\}_{k=1}^{n}$. The biorthogonal transform, when coupled with the regular DFT, exactly decorrelates the time series. In Section 2.3, we show that the biorthogonal transform corresponding to the regular DFT contains the regular DFT *plus* the Fourier transform of the best linear predictors of the time series outside the domain of observation. Since this transformation completes the information not found in the regular DFT, we call it the complete DFT. It is common to use the Cholesky decomposition to decompose the inverse of a Toeplitz matrix. An interesting aspect of the biorthogonal transformation is that it provides an alternative decomposition of the inverse of a Toeplitz matrix.

In Section 2.4, we show that the complete DFT, together with the regular DFT, allows us to rewrite the Gaussian likelihood within the frequency domain (which, as far as we are aware, is new). Further, it is well known that the Whittle likelihood has a bias due to the boundary effect. By rewriting the Gaussian likelihood within the frequency domain we show that the Gaussian likelihood avoids the boundary effect problem by predicting the time series outside the domain of observation. Precisely, the approximation error between the Gaussian and Whittle likelihood is due to the omission of these linear predictors in the regular DFT. From this result, we observe that the greater the persistence in the time series model (which corresponds to a more peaked spectral density) the larger the loss in approximating the complete DFT with the regular DFT. In order to obtain a better approximation of the Gaussian likelihood in the frequency domain, it is of interest to approximate the difference of the two likelihoods $\mathcal{L}_n(\theta; \underline{X}_n) - K_n(\theta; \underline{X}_n)$. For autoregressive processes of finite order, we obtain an analytic expression for the difference in the two likelihoods in terms of the AR parameters (see equation (2.18)). For general second order

3

stationary models, the expression is more complex. In Section 3, we obtain an approximation for $\mathcal{L}_n(\theta; \underline{X}_n) - K_n(\theta; \underline{X}_n)$ in terms of the infinite order (causal/minimum phase) autoregressive factorisation of $f_\theta(\omega) = \sigma^2 |1 - \sum_{j=1}^\infty \phi_j e^{-ij\omega}|^{-2}$. We show that this approximation is the first order term in a series expansion of the inverse of the Toeplitz matrix, $\Gamma_n(f)^{-1}$. More precisely, in Section 3.2, we show that $\Gamma_n(f)^{-1}$ can be expressed in terms of $C_n(f_\theta)^{-1}$ plus a polynomial-type series expansion of the AR(∞) coefficients.

In Section 4, we obtain an approximation for the difference $\mathcal{L}_n(\theta; \underline{X}_n) - K_n(\theta; \underline{X}_n)$ in terms of a finite order autoregressive process. We use this to define two spectral divergence criteria which are "almost" unbiased estimators of the spectral divergence between the true (underlying spectral) density and the parametric spectral density. We use these criteria to define two new frequency domain estimators. In Section 5, we obtain the asymptotic sampling properties of the new likelihood estimators including the asymptotic bias and variance. Finally, in Section 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 and misspecified.

The proofs can be found in the Supplementary material. The main proofs can be found in Appendix A, B, D and E. Baxter type inequalities for derivatives of finite predictors can be found in Appendix C. These results are used to obtain an approximation for the difference between the derivatives of the Gaussian and Whittle likelihood. In Appendix E we derive an expression for the asymptotic bias of the Gaussian, Whittle likelihoods, and the new frequency domain likelihoods, described above. In Appendix F, G and H we present additional simulations.

2 The Gaussian likelihood in the frequency domain

2.1 Preliminaries

In this section, we introduce most of the notation used in the paper, it can be skipped on first reading. To reduce notation, we omit the symbol $\underline{X}_n$ in the Gaussian and Whittle likelihood. Moreover, since the focus in this paper 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:

$$\mathcal{L}_n(\theta) = n^{-1}\underline{X}_n' \Gamma_n(f_\theta)^{-1}\underline{X}_n \quad \text{and} \quad K_n(\theta) = n^{-1}\underline{X}_n' C_n(f_\theta^{-1})\underline{X}_n. \tag{2.1}$$

Let A^* denote the conjugate transpose of the matrix A. We recall that the circulant matrix $C_n(g)$ can be written as $C_n(g) = F_n^* \Delta_n(g) F_n$, where $\Delta_n(g) = \text{diag}(g(\omega_{1,n}), \ldots, g(\omega_{n,n}))$ (diagonal matrix) and F_n is the $n \times n$ DFT matrix with entries $(F_n)_{k,t} = n^{-1/2} e^{it\omega_{k,n}}$. 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.

In general, 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 $\text{cov}_f(X_t, X_{t+r}) = \mathbb{E}_f[X_t X_{t+r}] = c_f(r)$. 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 entrywise p-norm for $p \geq 1$, 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 2π-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 $\sum_{j=0}^{K+2} \sup_\omega |g^{(j)}(\omega)| < \infty$ then $\|g\|_K < \infty$, where $g^{(j)}(\cdot)$ denotes the jth 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 π. By using the classical factorisation results in Szegö (1921) and Baxter (1962) 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 σ_g, $\phi_g(\cdot)$, and $\psi_g(\cdot)$ are defined similarly. We use these expansions in Sections 3 and 4, where we require the following notation

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

for some $0 < \varepsilon < 1$.

For postive sequences $\{a_n\}$ and $\{b_n\}$, we denote $a_n \sim b_n$ if there exist $0 < C_1 \leq C_2 < \infty$ such that $C_1 \leq a_n / b_n \leq C_2$ for all n. Lastly, we denote Re and Im as the real and imaginary part of a complex variable respectively.

2.2 Motivation

In order 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 1, we plot the bias in the estimator of ϕ in the AR(1) model $X_t = \phi X_{t-1} + \varepsilon_t$ for different values of ϕ (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). Let $\{X_t\}_{t=1}^n$ denote the observed time series. Straightforward calculations (based on expressions for $\Gamma_n(f_\phi)^{-1}$ and $C_n(f_\phi^{-1})$) show that the difference between the Gaussian and Whittle likelihoods for an AR(1) model is

$$
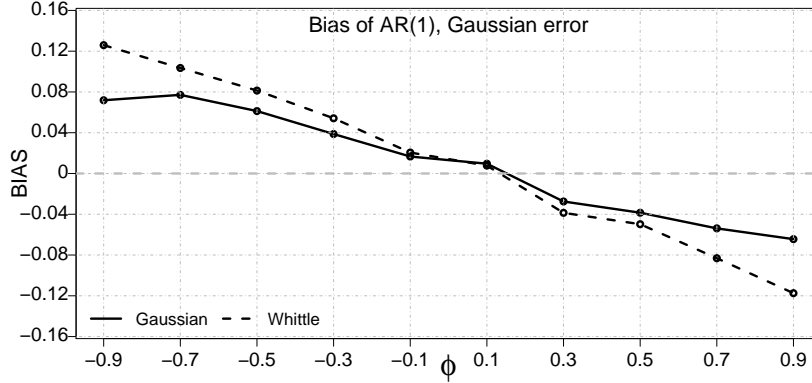\mathcal{L}_n(\phi) - K_n(\phi) = n^{-1} \left[2\phi X_1 X_n - \phi^2 (X_1^2 + X_n^2) \right] \tag{2.2}
$$

5

Figure 1: *The model $X_t = \phi X_{t-1} + \varepsilon_t$ with independent standard normal errors is simulated. The bias of the estimator of ϕ based on sample size $n = 20$ over 1000 replications.*

Thus we observe that the closer $|\phi|$ is to one, the larger the expected difference between the likelihoods. Using (2.2) 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 E.2) Generalisations 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. In the remainder of this section, we derive an exact expression for the Gaussian likelihood in the frequency domain. Using these derivations, we obtain a simple expression for the difference between the Whittle and Gaussian likelihood for $AR(p)$ models. In subsequent sections, we obtain approximations for this difference for general time series models.

2.3   The biorthogonal transform to the discrete Fourier transform

In order to obtain an exact bound, we start with the Whittle likelihood and recall that the DFT of the time series plays a fundamental role in its formulation. With this in mind, our approach is based on deriving the transformation $\{Z_{k,n}\}_{k=1}^{n} \subset \mathrm{sp}(\underline{X}_n)$ (where $\mathrm{sp}(\underline{X}_n)$ denotes the linear space over a complex field spanned by $\underline{X}_n = \{X_t\}_{t=1}^{n}$), 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

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

where $\delta_{k_1,k_2} = 1$ if $k_1 = k_2$ (and zero otherwise). Since $\underline{Z}'_n = (Z_{1,n}, \ldots, Z_{n,n}) \in \text{sp}(\underline{X}_n)^n$, 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 $\text{cov}_f(U_n \underline{X}_n, F_n \underline{X}_n) = \Delta_n(f)$. The benefit of biorthogonality is that it leads to the following simple identity on the inverse of the variance matrix.

Lemma 2.1 *Suppose that U_n and V_n are invertible matrices which are biorthogonal with respect to the variance matrix $\text{var}(\underline{X}_n)$. That is $\text{cov}(U_n \underline{X}_n, V_n \underline{X}_n) = \Delta_n$, where Δ_n is a diagonal matrix. Then*

$$\text{var}(\underline{X}_n)^{-1} = V_n^* \Delta_n^{-1} U_n. \tag{2.3}$$

PROOF. It follows immediately from $\text{cov}(U_n \underline{X}_n, V_n \underline{X}_n) = U_n \text{var}(\underline{X}_n) V_n^* = \Delta_n$ and $\text{var}(\underline{X}_n) = U_n^{-1} \Delta_n (V_n^*)^{-1}$. □

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_τ for $\tau \leq 0$ and $\tau > n$ given $\{X_t\}_{t=1}^n$ as

$$\widehat{X}_{\tau,n} = \sum_{t=1}^n \phi_{t,n}(\tau; f) X_t, \tag{2.4}$$

where $\{\phi_{t,n}(\tau; f)\}_{t=1}^n$ are the coefficients which minimize the L_2-distance $\mathbb{E}_f[X_\tau - \sum_{t=1}^n \phi_{t,n}(\tau; f) X_t]^2$. Using this notation we obtain the following theorem.

Theorem 2.1 (The biorthogonal transform) *Let $\{X_t\}$ be a second order stationary, zero mean time series with spectral density f which is bounded away from zero and whose autocovariance satisfies $\sum_{r \in \mathbb{Z}} |r c_f(r)| < \infty$. Let $\widehat{X}_{\tau,n}$ denote the best linear predictors of X_τ as defined in (2.4) and $\{\phi_{t,n}(\tau; f)\}_{t=1}^n$ the corresponding coefficients. Then*

$$\text{cov}_f\big((F_n + D_n(f))\underline{X}_n, F_n \underline{X}_n\big) = \Delta_n(f), \tag{2.5}$$

where $D_n(f)$ has entries

$$D_n(f)_{k,t} = n^{-1/2} \sum_{\tau \leq 0} \big(\phi_{t,n}(\tau; f) e^{i\tau \omega_{k,n}} + \phi_{n+1-t,n}(\tau; f) e^{-i(\tau-1)\omega_{k,n}}\big), \tag{2.6}$$

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

$$\text{cov}_f\left(\widetilde{J}_n(\omega_{k_1,n}; f), J_n(\omega_{k_2,n})\right) = f(\omega_{k_1,n}) \delta_{k_1,k_2} \tag{2.7}$$

where $\widetilde{J}_n(\omega; f) = J_n(\omega) + \widehat{J}_n(\omega; f)$ and

$$\widehat{J}_n(\omega; f) = n^{-1/2} \sum_{\tau \leq 0} \widehat{X}_{\tau,n} e^{i\tau\omega} + n^{-1/2} \sum_{\tau > n} \widehat{X}_{\tau,n} e^{i\tau\omega}. \tag{2.8}$$

PROOF. See Appendix A (note that identity (2.7) can be directly verified using results on best linear predictors). $\qquad\square$

Corollary 2.1 (Inverse Toeplitz identity) *Let* $\Gamma_n(f)$ *denote an* $n \times n$ *Toeplitz matrix generated by the spectral density* f. *Then equations (2.3) and (2.5) yield the following identity*

$$\Gamma_n(f)^{-1} = F_n^* \Delta_n(f^{-1})(F_n + D_n(f)), \tag{2.9}$$

where $D_n(f)$ *is defined in (2.6). 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*

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

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 visualisation of the observations and the predictors that are involved in the construction of $\widetilde{J}_n(\omega; f)$ is given in Figure 2.
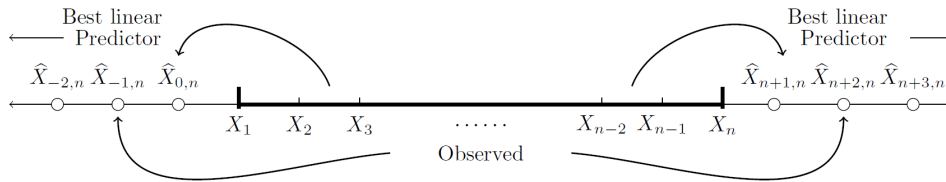

Figure 2: $\widetilde{J}_n(\omega; f)$ *is the Fourier transform over both the observed time series and its predictors outside this domain.*

It is quite surprising that only a small modification of the regular DFT leads to its biorthogonal transformation. Furthermore, the contribution of the additional DFT term is $\widehat{J}_n(\omega_{k,n}; f) = O_p(n^{-1/2})$. 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 Lahiri (2003) and Brillinger (2001). For future reference, we will use the following definitions.

Definition 2.1 *We refer to* $\widehat{J}_n(\omega; f)$ *as the predictive DFT (as it is the Fourier transform of all*

the linear predictors), noting that basic algebra yields the expression

$$\widehat{J}_n(\omega; f) = n^{-1/2} \sum_{t=1}^{n} X_t \sum_{\tau \leq 0} (\phi_{t,n}(\tau; f) e^{i\tau\omega} + e^{in\omega} \phi_{n+1-t,n}(\tau; f) e^{-i(\tau-1)\omega}). \qquad (2.10)$$

Note that when $\omega = \omega_{k,n}$, the term $e^{in\omega}$ in (2.10) 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.

Example 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 $\tau \leq 0$, $\widehat{X}_{\tau,n} = \phi^{|\tau|+1} X_1$ and for $\tau > n$ $\widehat{X}_{\tau,n} = \phi^{\tau-n} X_n$. An illustration is given in Figure 3.*
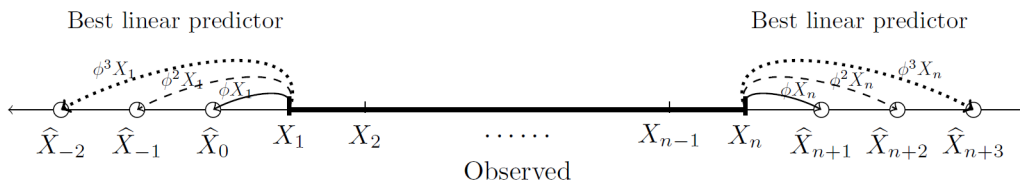

Figure 3: *The past and future best linear predictors based on a AR(1) model.*

Then the predictive DFT for the AR(1) model is

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

In other words, a small adjustment of the boundary leads to $\widetilde{J}_n(\omega; f_\phi)\overline{J_n(\omega)}$ being an unbiased estimator of $f(\omega) = \sigma^2 |\phi(\omega)|^{-2}$.

Remark 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 $\operatorname{cov}_f(\underline{\widetilde{X}}_n, \underline{X}_n) = I_n$). 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 (2.9) 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 (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 we have*

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

2.4 The Gaussian likelihood in the frequency domain

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 2.1.

Theorem 2.2 (A frequency domain representation of the Gaussian likelihood) *Suppose the spectral density f_θ is bounded away from zero, and the corresponding autocovariance is such that $\sum_r |rc_{f_\theta}(r)| < \infty$. Let $\mathcal{L}_n(\theta)$ and $K_n(\theta)$ be defined as in (2.1). Then we have*

$$\mathcal{L}_n(\theta) = \frac{1}{n} \underline{X}'_n \Gamma_n(f_\theta)^{-1} \underline{X}_n = \frac{1}{n} \sum_{k=1}^{n} \frac{\widetilde{J}_n(\omega_{k,n}; f_\theta) \overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})}. \tag{2.11}$$

Further

$$\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1}) = F_n^* \Delta_n(f_\theta^{-1}) D_n(f_\theta). \tag{2.12}$$

This yields the difference between the Gaussian and Whittle likelihood

$$
\begin{aligned}
\mathcal{L}_n(\theta) - K_n(\theta) &= n^{-1} \underline{X}'_n \left[\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1}) \right] \underline{X}_n \\
&= \frac{1}{n} \sum_{k=1}^{n} \frac{\widehat{J}_n(\omega_{k,n}; f_\theta) \overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})}.
\end{aligned}
\tag{2.13}
$$

PROOF. (2.12) follows immediately from Corollary 2.1. 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 (2.11), 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 (2.13). $\qquad\square$

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_θ 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.

In the remainder of this section and the subsequent section, we study the difference between the two likelihoods and corresponding matrices. This will allow us to develop methods that better capture the Gaussian likelihood within the frequency domain. By using Theorem 2.2, 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

$$(F_n^* \Delta_n(f_\theta^{-1})D_n(f_\theta))_{s,t}$$
$$= \sum_{\tau \leq 0} [\phi_{t,n}(\tau; f_\theta)G_{1,n}(s, \tau; f_\theta) + \phi_{n+1-t,n}(\tau; f_\theta)G_{2,n}(s, \tau; f_\theta)] \qquad (2.14)$$

with

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

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

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 in the following theorem, where we consider the special case of AR(p) models. We showed in Example 2.1 that for AR(1) processes,

the predictive DFT has a simple form. In the following theorem, we obtain an analogous result for $AR(p)$ models (where $p \leq n$).

Theorem 2.3 (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 $p \leq n$. The predictive DFT has the analytic form*

$$\widehat{J}_n(\omega; f_\theta) =$$
$$\frac{n^{-1/2}}{\phi_p(\omega)} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{-is\omega} + e^{in\omega} \frac{n^{-1/2}}{\overline{\phi_p(\omega)}} \sum_{\ell=1}^{p} X_{n+1-\ell} \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{i(s+1)\omega}. \qquad (2.15)$$

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

$$D_n(f_\theta)$$
$$= n^{-1/2} \begin{pmatrix} \phi_{1,p}(\omega_{1,n}) & \cdots & \phi_{p,p}(\omega_{1,n}) \ 0 \ \cdots \ 0 & e^{i\omega_{1,n}}\overline{\phi_{p,p}(\omega_{1,n})} & \cdots & e^{i\omega_{1,n}}\overline{\phi_{1,p}(\omega_{1,n})} \\ \phi_{1,p}(\omega_{2,n}) & \cdots & \phi_{p,p}(\omega_{2,n}) \ 0 \ \cdots \ 0 & e^{i\omega_{2,n}}\overline{\phi_{p,p}(\omega_{2,n})} & \cdots & e^{i\omega_{2,n}}\overline{\phi_{1,p}(\omega_{2,n})} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \phi_{1,p}(\omega_{n,n}) & \cdots & \phi_{p,p}(\omega_{n,n}) \ 0 \ \cdots \ 0 & e^{i\omega_{n,n}}\overline{\phi_{p,p}(\omega_{n,n})} & \cdots & e^{i\omega_{n,n}}\overline{\phi_{1,p}(\omega_{n,n})} \end{pmatrix} \qquad (2.16)$$

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

$$\left(\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1}) \right)_{s,t} = (F_n^* \Delta_n(f_\theta^{-1}) D_n(f_\theta))_{s,t}$$
$$= \begin{cases} \sigma^{-2} \sum_{\ell=0}^{p-t} \phi_{\ell+t} \widetilde{\phi}_{(\ell+s) \bmod n} & 1 \leq t \leq p \\ \sigma^{-2} \sum_{\ell=1}^{p-(n-t)} \phi_{\ell+(n-t)} \widetilde{\phi}_{(\ell-s) \bmod n} & n-p+1 \leq t \leq n \\ 0 & otherwise \end{cases} \qquad (2.17)$$

PROOF. In Appendix A. $\qquad \square$

Theorem 2.3 shows that for $AR(p)$ models, the predictive DFT only involves the p observations on each side of the observational boundary $X_1, \ldots, X_p$ and $X_{n-p+1}, \ldots, X_n$, 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 (2.17).

By using (2.15) 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

$$\mathcal{L}_n(\phi) = \frac{\sigma^{-2}}{n} \sum_{k=1}^{n} |J_n(\omega_{k,n})|^2 |\phi_p(\omega_{k,n})|^2$$

$$+ \frac{\sigma^{-2}}{n} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s} \left(X_{(-s) \bmod n} - \sum_{j=1}^{p} \phi_j X_{(j-s) \bmod n} \right)$$

$$+ \frac{\sigma^{-2}}{n} \sum_{\ell=1}^{p} X_{n+1-\ell} \sum_{s=0}^{p-\ell} \phi_{\ell+s} \left(X_{(s+1) \bmod n} - \sum_{j=1}^{p} \phi_j X_{(s+1-j) \bmod n} \right), \qquad (2.18)$$

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 Appendix A. Equation (2.18) 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.

3 Frequency domain approximations of the Gaussian likelihood

In Theorem 2.2 we rewrote the Gaussian likelihood within the frequency domain. This allowed us to obtain an expression for the difference between the Gaussian and Whittle likelihoods for AR(p) models (see (2.18)). This is possible because the predictive DFT $\widehat{J}_n(\cdot; f_\theta)$ has a simple analytic form.

It would be of interest to generalize this result to general time series models. However, for infinite order autoregressive models, the predictions across the boundary and the predictive DFT given in (2.10) do not have a simple, analytic form. In Section 3.1 we show that we can obtain an approximation of the predictive DFT in terms of the AR(∞) coefficients corresponding to f_θ. In turn, this allows us to obtain an approximation for $\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1})$, which is analogous to equation (2.17) for AR(p) models. Such a result proves to be very useful from both a theoretical and practical perspective. Theoretically, we use this result to show that the difference between the Whittle and Gaussian likelihood is of order $O(n^{-1})$. Furthermore, in Section 3.2 we show that the approximation described in Section 3.1 is the first order term of a polynomial-type series expansion of $\Gamma_n(f_\theta)^{-1}$ in terms of the AR(∞) parameters. From a practical perspective, the approximations are used in Section 4 to motivate alternative quasi-likelihoods defined within the frequency domain.

First, we require the following set of assumptions on the spectral density f_θ.

Assumption 3.1   *(i) The spectral density f is bounded away from zero.*

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

13

Under the above assumptions, we can write $f(\omega) = \sigma^2|\psi(\omega; f)|^2 = \sigma^2|\phi(\omega; f)|^{-2}$ where

$$\psi(\omega; f) = 1 + \sum_{j=1}^{\infty} \psi_j(f)e^{-ij\omega} \quad \text{and} \quad \phi(\omega; f) = 1 - \sum_{j=1}^{\infty} \phi_j(f)e^{-ij\omega}. \tag{3.1}$$

Further, under Assumption 3.1 we have $\sum_{r=1}^{\infty} |r^K\psi_r(f)|$ and $\sum_{r=1}^{\infty} |r^K\phi_r(f)|$ are both finite (see Kreiss et al. (2011)). Thus if f satisfies Assumption 3.1 with some $K > 1$, then $\|\psi_f\|_K < \infty$ and $\|\phi_f\|_K < \infty$.

3.1 The first order approximation

In order to obtain a result analogous to Theorem 2.3, we replace $\phi_{s,n}(\tau; f_\theta)$ in $D_n(f_\theta)$ with $\phi_s(\tau; f_\theta)$ which are the coefficients of the best linear predictor of X_τ (for $\tau \leq 0$) given $\{X_t\}_{t=1}^{\infty}$ i.e. $\widehat{X}_\tau = \sum_{t=1}^{\infty} \phi_t(\tau; f_\theta)X_t$. This gives the matrix $D_{\infty,n}(f_\theta)$, where

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

It can be shown that for $1 \leq k, t \leq n$,

$$(D_{\infty,n}(f_\theta))_{k,t} = n^{-1/2}\frac{\phi_t^{\infty}(\omega_{k,n}; f_\theta)}{\phi(\omega_{k,n}; f_\theta)} + n^{-1/2}e^{i\omega_{k,n}}\frac{\overline{\phi_{n+1-t}^{\infty}(\omega_{k,n}; f_\theta)}}{\overline{\phi(\omega_{k,n}; f_\theta)}}, \tag{3.2}$$

where $\phi_t^{\infty}(\omega; f_\theta) = \sum_{s=0}^{\infty} \phi_{t+s}(f_\theta)e^{-is\omega}$. The proof of the above identity can be found in Appendix B.1. Using the above we can show that $(D_{\infty,n}(f_\theta)\underline{X}_n)_k = \widehat{J}_{\infty,n}(\omega_{k,n}; f_\theta)$ where

$$\widehat{J}_{\infty,n}(\omega; f_\theta)$$
$$= \frac{n^{-1/2}}{\phi(\omega; f_\theta)} \sum_{t=1}^{n} X_t\phi_t^{\infty}(\omega; f_\theta) + e^{i(n+1)\omega}\frac{n^{-1/2}}{\overline{\phi(\omega; f_\theta)}} \sum_{t=1}^{n} X_{n+1-t}\overline{\phi_t^{\infty}(\omega; f_\theta)}. \tag{3.3}$$

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

Theorem 3.1 (An AR(∞) approximation for general processes) *Suppose f satisfies Assumption 3.1, f_θ 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 (2.6) and (3.2) and (3.3) respectively. Then we have*

$$\underline{X}_n' F_n^* \Delta_n(f_\theta^{-1}) \left(D_n(f) - D_{\infty,n}(f) \right) \underline{X}_n$$
$$= \sum_{k=1}^{n} \frac{\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} \left(\widehat{J}_n(\omega_{k,n}; f) - \widehat{J}_{\infty,n}(\omega_{k,n}; f) \right) \tag{3.4}$$

14

and

$$\left\|F_n^*\Delta_n(f_\theta^{-1})\left(D_n(f) - D_{\infty,n}(f)\right)\right\|_1 \leq \frac{C_{f,0}\rho_{n,K}(f)}{n^{K-1}}A_K(f, f_\theta). \tag{3.5}$$

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

$$n^{-1}\left\|\underline{X}_n'F_n^*\Delta_n(f_\theta^{-1})\left(D_n(f) - D_{\infty,n}(f)\right)\underline{X}_n\right\|_{\mathbb{E},q}$$
$$\leq \frac{C_{f,0}\rho_{n,K}(f)}{n^K}A_K(f, f_\theta)\|X\|_{\mathbb{E},2q}^2. \tag{3.6}$$

PROOF. See Appendix B.1. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

We mention that we state the above theorem in 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_θ. This is to allow generalisations of the Whittle and Gaussian likelihoods, which we discuss in Section 4.

Applying the above theorem to the Gaussian likelihood gives an approximation which is analogous to (2.18)

$$\begin{aligned}
\mathcal{L}_n(\theta) &= K_n(\theta) + \frac{1}{n}\sum_{k=1}^n \frac{\widehat{J}_{\infty,n}(\omega_{k,n}; f_\theta)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} + O_p(n^{-K}) \\
&= K_n(\theta) + \frac{1}{n}\sum_{s,t=1}^n X_s X_t \frac{1}{n}\sum_{k=1}^n e^{-is\omega_{k,n}}\varphi_{t,n}(\omega_{k,n}; f_\theta) + O_p(n^{-K}),
\end{aligned} \tag{3.7}$$

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_θ 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 3.1 *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 (2.14).

In the following theorem, we obtain a bound between the Gaussian and Whittle likelihood.

Theorem 3.2 (The difference in the likelihoods) *Suppose f_θ satisfies Assumption 3.1. Let*

$D_n(f_\theta)$ and $D_{\infty,n}(f_\theta)$ be defined as in (2.6) and (3.2) respectively. Then we have

$$\left\| F_n^* \Delta_n(f_\theta^{-1}) D_{\infty,n}(f_\theta) \right\|_1 \leq A_1(f_\theta, f_\theta) \tag{3.8}$$

and

$$\left\| \Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1}) \right\|_1 \leq \left(A_1(f_\theta, f_\theta) + \frac{C_{f_\theta,0} \rho_{n,K}(f_\theta)}{n^{K-1}} A_K(f_\theta, f_\theta) \right). \tag{3.9}$$

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

$$\|\mathcal{L}_n(\theta) - K_n(\theta)\|_{\mathbb{E},q} \leq n^{-1} \left(A_1(f_\theta, f_\theta) + \frac{C_{f_\theta,0} \rho_{n,K}(f_\theta)}{n^{K-1}} A_K(f_\theta, f_\theta) \right) \|X\|_{\mathbb{E},2q}^2. \tag{3.10}$$

PROOF. See Appendix B.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 $O(n^{-1})$. We conclude this section by obtaining a higher order expansion of $\Gamma_n(f_\theta)^{-1}$.

3.2   A series expansion

Theorem 3.1 gives an approximation of the predictive DFT $\widehat{J}_n(\omega; f)$ in terms of $\widehat{J}_{\infty,n}(\omega; f)$, which is comprised of the AR(∞) coefficients corresponding to f. In the following lemma we show that it is possible to obtain a series expansion of $\widehat{J}_n(\omega; f)$ and $\Gamma_n(f)^{-1} - C_n(f)^{-1}$ in terms of the products of AR(∞) coefficients. The proof of the results in this section hinge on applying von Neumann's alternative projection theorem to stationary time series. This technique was first developed for time series in Inoue and Kasahara (2006). We make use of Theorem 2.5, Inoue and Kasahara (2006), where an expression for the coefficients of the finite predictors $\phi_{t,n}(\tau)$ is given.

We define the function $\zeta_{t,n}^{(1)}(\omega; f) = \phi_t^\infty(\omega)$ and for $s \geq 2$

$$\begin{aligned}
\zeta_{t,n}^{(s)}(\omega; f) &= \frac{1}{(2\pi)^{s-1}} \int_{[0,2\pi]^s} \left(\prod_{a=1}^{s-1} \phi(\lambda_{a+1}; f)^{-1} \Phi_n(\lambda_a, \lambda_{a+1}) \right) \times \\
&\quad \left(\phi_t^\infty(\lambda_s; f) \delta_{s \equiv 1(\bmod\ 2)} + \phi_{n+1-t}^\infty(\lambda_s; f) \delta_{s \equiv 0(\bmod\ 2)} \right) \delta_{\lambda_1 = \omega} d\underline{\lambda}_s,
\end{aligned} \tag{3.11}$$

where $d\underline{\lambda}_s = d\lambda_1 \cdots d\lambda_s$ denotes the s-dimensional Lebesgue measure,

$$\Phi_n(\lambda_1, \lambda_2) = \sum_{u=0}^\infty \phi_{n+1+u}^\infty(\lambda_1; f) e^{iu\lambda_2} = \sum_{u=0}^\infty \sum_{s=0}^\infty \phi_{n+1+u+s}(f) e^{-is\lambda_1} e^{iu\lambda_2}$$

and δ denotes the indicator variable. In the following lemma, we show that $\zeta_{t,n}^{(s)}(\omega; f)$ plays the same role as $\phi_t^\infty(\omega; f)$ in the predictive DFT approximation given in equation (3.3). It will be used to approximate $\widehat{J}_n(\omega; f)$ to a greater degree of accuracy.

Theorem 3.3 *Suppose f satisfies Assumption 3.1, where $f(\omega) = \sigma^2 |\phi(\omega; f)|^{-2}$. Let $D_n(f)$ and $\zeta_{j,n}^{(s)}(\omega; f)$ be defined as in (2.6) and (3.11) respectively. Define the s-order predictive DFT*

$$\widehat{J}_n^{(s)}(\omega; f) = \frac{n^{-1/2}}{\phi(\omega; f)} \sum_{t=1}^n X_t \zeta_{t,n}^{(s)}(\omega; f) + e^{i(n+1)\omega} \frac{n^{-1/2}}{\phi(\omega; f)} \sum_{t=1}^n X_{n+1-t} \overline{\zeta_{t,n}^{(s)}(\omega; f)}.$$

Then

$$\widehat{J}_n(\omega, f) = \sum_{s=1}^\infty \widehat{J}_n^{(s)}(\omega; f) \tag{3.12}$$

and $D_n(f) = \sum_{s=1}^\infty D_n^{(s)}(f)$, where

$$(D_n^{(s)}(f))_{k,t} = n^{-1/2} \frac{\zeta_{t,n}^{(s)}(\omega_{k,n}; f)}{\phi(\omega_{k,n}; f)} + n^{-1/2} e^{i\omega_{k,n}} \frac{\overline{\zeta_{n+1-t,n}^{(s)}(\omega_{k,n}; f)}}{\phi(\omega_{k,n}; f)}. \tag{3.13}$$

Further, for a sufficiently large n we have

$$\widehat{J}_n(\omega; f) = \sum_{s=1}^m \widehat{J}_n^{(s)}(\omega; f) + O_p\left(\frac{1}{n^{m(K-1)+1/2}}\right). \tag{3.14}$$

PROOF. See Appendix B.2. □

In the case $s = 1$, it is straightforward to show that

$$\widehat{J}_n^{(1)}(\omega; f) = \widehat{J}_{\infty,n}(\omega; f) \quad \text{and} \quad D_n^{(1)}(f) = D_{\infty,n}(f).$$

Therefore, the first term in the expansion of $\widehat{J}_n(\omega, f)$ and $D_n(f)$ is the AR(∞) approximation $\widehat{J}_{\infty,n}(\omega; f)$ and $D_{\infty,n}(f)$ respectively. We mention, that it is simple to check that if f corresponds to an AR(p) spectral density for some $p \leq n$, then $\widehat{J}_n^{(s)}(\omega; f) = 0$ for all $s \geq 2$. For general spectral densities, the higher order expansion gives a higher order approximation of $\widehat{J}_n(\omega; f)$ and $\Gamma_n(f)^{-1}$ in terms of products of the AR(∞) coefficients. Using the above result we have the expansions

$$\Gamma_n(f)^{-1} = C_n(f^{-1}) + \sum_{s=1}^\infty F_n^* \Delta_n(f^{-1}) D_n^{(s)}(f)$$

and

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

It is interesting to note that $\zeta_{t,n}^{(s)}(\omega; f)$ can be evaluated recursively using

$$\zeta_{t,n}^{(s+2)}(\omega; f)$$
$$= \frac{1}{(2\pi)^2} \int_{[0,2\pi]^2} \phi(y_1; f)^{-1}\phi(y_2; f)^{-1}\Phi_n(\omega, y_1)\Phi_n(y_1, y_2)\zeta_{t,n}^{(s)}(y_2; f)dy_1 dy_2. \qquad (3.15)$$

In a similar vein, both the s-order predictive DFT $\widehat{J}_n^{(s)}(\omega; f)$ and $D_n^{(s)}(f)$ can be evaluated recursively using a recursion similar to the above (see Appendix B.2 for the details).

The above results show that it is possible to obtain an analytic expression for $\widehat{J}_n(\omega; f)$ and $\Gamma_n(f)^{-1}$ in terms of the products of the AR(∞) coefficients. This expression for the inverse of a Toeplitz matrix may have applications outside time series. However, from the perspective of estimation, the first order approximation $\widehat{J}_n^{(1)}(\omega; f) = \widehat{J}_{\infty,n}(\omega; f)$ is sufficient. We discuss some applications in the next section.

4 New frequency domain quasi-likelihoods

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

To motivate the criteria, we recall from Theorem 2.2 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 spectral density f_θ, 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 (this is discussed further in Section 5.2).

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_θ. But since $-\log f - 1$ does not depend on θ we ignore this term.

4.1 The boundary corrected Whittle likelihood

In order to address some of the issues raised above, we recall from Theorem 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, we recall that $\widetilde{J}_n(\omega_{k,n}; f)$ is comprised of the best linear predictors based on the unobserved time series. The coefficients of the best linear predictors can be replaced with the h-step ahead predictors evaluated with the best fitting autoregressive parameters of order p (the so called plug-in estimators; see Bhansali (1996) and Kley et al. (2019)). This is equivalent 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 (2.15). 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 the proof of Lemma 4.1, below, 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 *Suppose f satisfies Assumption 3.1, f_θ 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 $1 \le p < n$, then we have*

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

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

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

PROOF. See Appendix B.1. □

Remark 4.1 *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 $p \ge p_0$, 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 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 $0 \le \rho < 1$ we have*

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

20

(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 $1 \le p < n$, $\mathbb{E}_f[W_{p,n}(\theta)] = I_n(f; f_\theta) + O((np^{K-1})^{-1})$ 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. But f_p is easily estimated 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. We denote this estimator as $\widehat{\underline{\phi}}_p$ and the corresponding spectral density as $\widehat{f}_p$. Using this we define $\widehat{J}_n(\omega_{k,n}; \widehat{f}_p)$ where

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

and $\widehat{\phi}_p(\omega) = 1 - \sum_{u=1}^{p} \widehat{\phi}_{u,p} e^{-iu\omega}$. This estimator allows us to replace $\widehat{J}_n(\omega_{k,n}; f_p)$ in $W_{p,n}(\theta)$ with $\widehat{J}_n(\omega_{k,n}; \widehat{f}_p)$ to give the "observed" 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 θ, $\widehat{\theta}_n = \arg\min \widehat{W}_{p,n}(\theta)$. 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 θ. However, due to rounding errors it is prudent to use $\mathrm{Re}\, \widehat{W}_{p,n}(\theta)$ in the minimisation 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 6).

In this paper, we focus on estimating $\widehat{J}_n(\omega_{k,n}; f_p)$ using the Yule-Walker estimator. However, as pointed out by two referees, 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 Appendix H, 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 Appendix H together with the results of some simulations.

4.2   The hybrid Whittle likelihood

The simulations in Section 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 $\underline{h}_n = \{h_{t,n}\}_{t=1}^n$ is a data taper, where the weights $\{h_{t,n}\}$ are non-negative and $\sum_{t=1}^n h_{t,n} = n$. We define the tapered DFT as

$$J_{n,\underline{h}_n}(\omega_{k,n}) = n^{-1/2} \sum_{t=1}^n h_{t,n} X_t e^{it\omega_{k,n}}.$$

Suppose f is the best fitting spectral density function. Using that $\sum_{t=1}^n h_{t,n} = n$ and $\mathrm{cov}_f(X_t, \widehat{X}_{\tau,n}) = c_f(t - \tau)$ we have

$$\mathbb{E}_f[\widetilde{J}_n(\omega; f)\overline{J_{n,\underline{h}_n}(\omega)}] = f(\omega), \tag{4.9}$$

which is analogous to the non-tapered result $\mathbb{E}_f[\widetilde{J}_n(\omega; f)\overline{J_n(\omega)}] = f(\omega)$. Based on the above result we define the infeasible hybrid Whittle likelihood which combines the regular DFT of the tapered time series and the complete DFT (which is not tapered)

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

Using (4.9), it can be shown that $\mathbb{E}_f[H_n(\theta)] = I_n(f; f_\theta)$. Thus $H_n(\theta)$ is an unbiased estimator of $I_n(f; f_\theta)$. Clearly, it is not possible to estimate θ 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

$$\widehat{H}_{p,n}(\theta) = \frac{1}{n} \sum_{k=1}^{n} \frac{\widetilde{J}_n(\omega_{k,n}; \widehat{f}_p) \overline{J_{n,\underline{h}_n}(\omega_{k,n})}}{f_\theta(\omega_{k,n})} + \frac{1}{n} \sum_{k=1}^{n} \log f_\theta(\omega_{k,n}). \tag{4.11}$$

We then use as an estimator of θ, $\widehat{\theta}_n = \arg\min \widehat{H}_{p,n}(\theta)$. An illustration which visualises and compares the boundary corrected Whittle likelihood and hybrid Whittle likelihood is given in Figure 4.
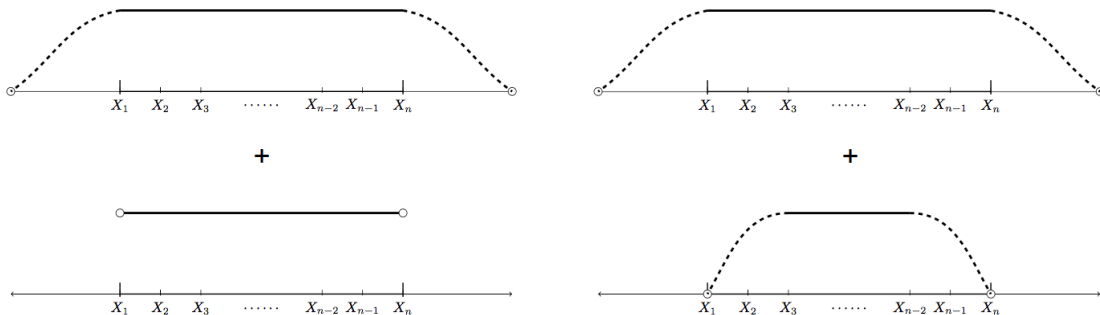

Figure 4: *Left: The estimated complete DFT and the regular DFT which yields the boundary corrected Whittle likelihood. Right: The estimated complete DFT and the tapered DFT which forms the hybrid Whittle likelihood.*

5 The sampling properties of the hybrid Whittle likelihood

In this section, we 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 $h_{t,n} = 1$ for $1 \le t \le n$. In Das et al. (2020) we study the sampling properties of the estimated complete periodogram $\widetilde{J}_n(\omega; \widehat{f}_p) \overline{J_{n,\underline{h}_n}(\omega)}$. Using these results and the results in Appendix D and E, we obtain the bias and variance of the boundary corrected and hybrid Whittle likelihood.

Suppose we fit the spectral density $f_\theta(\omega)$ (where θ 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_{θ_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 $\widehat{\theta}_n = \arg\min \widehat{H}_{p,n}(\theta)$.

5.1 Assumptions

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

$$h_{t,n} = c_n h_n(t/n), \tag{5.1}$$

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 $H_{2,n}/H_{1,n}^2 \sim n^{-1}$. Under this condition, the hybrid Whittle is $n^{1/2}$–consistency and the equivalence result in Theorem 5.1 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 $H_{2,n}/H_{1,n}^2 \sim n^{-1}$ will lead to a change of rate in Theorem 5.1.

Assumption 5.1 (Assumptions on the parameter space)   *(i) The parameter space $\Theta \subset \mathbb{R}^d$ is compact, $0 < \inf_{\theta \in \Theta} \inf_\omega f_\theta(\omega) \le \sup_{\theta \in \Theta} \sup_\omega f_\theta(\omega) < \infty$ and θ_n lies in the interior of Θ.*

(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 θ.

(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_θ respectively. Then for all $\theta \in \Theta$ and $0 \le s \le \kappa$ (for some $\kappa \ge 4$), 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 ath order partial derivative of g with respect to θ, 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 5.1(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 5.1(ii, iii-a) the asymptotic bias of the new-frequency domain likelihood estimators would contain some additional terms. Assumption 5.1(iii-b) is used to bound the sth derivative of the spectral density.

Assumption 5.2 (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 $\mathrm{cum}(X_0, X_{t_1}, \ldots, X_{t_{\ell-1}})$.*

Then for all $1 \le j \le \ell \le 12$,

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

(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

$$I(\theta) = -\frac{1}{2\pi} \int_0^{2\pi} [\nabla_\theta^2 f_\theta(\omega)^{-1}] f(\omega) d\omega. \tag{5.2}$$

We require Assumption 5.2(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 5.2(i,ii), Theorem 3.1 in Das et al. (2020), we can show 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 5.1(i,iii) the above error is uniform over the parameter space. If the model is an AR(p_0) and $p_0 \leq p$, then the term $O((np^{K-1})^{-1})$ in the above disappears.

To obtain a bound for the mean and variance of $\widehat{\theta}_n = (\widehat{\theta}_{1,n}, \ldots, \widehat{\theta}_{d,n})$ we require the following quantities. Let

$$\begin{aligned}
V(g,h) &= \frac{2}{2\pi} \int_0^{2\pi} g(\omega) h(\omega) f(\omega)^2 d\omega \\
&+ \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} g(\omega_1) h(\omega_2) f_4(\omega_1, -\omega_1, \omega_2) d\omega_1 d\omega_2 \\
\text{and} \qquad J(g) &= \frac{1}{2\pi} \int_0^{2\pi} g(\omega) f(\omega) d\omega,
\end{aligned} \tag{5.3}$$

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 (5.2)) as $I^{(s,r)}$, and define

$$\begin{aligned}
G_r(\theta) &= \sum_{s_1,s_2=1}^d I^{(s_1,s_2)} V \left(\frac{\partial f_\theta^{-1}}{\partial \theta_{s_2}}, \frac{\partial^2 f_\theta^{-1}}{\partial \theta_{s_1} \partial \theta_r} \right) \\
&+ \frac{1}{2} \sum_{s_1,s_2,s_3,s_4=1}^d I^{(s_1,s_3)} I^{(s_2,s_4)} V \left(\frac{\partial f_\theta^{-1}}{\partial \theta_{s_3}}, \frac{\partial f_\theta^{-1}}{\partial \theta_{s_4}} \right) J \left(\frac{\partial^3 f_\theta^{-1}}{\partial \theta_{s_1} \partial \theta_{s_2} \partial \theta_r} \right).
\end{aligned} \tag{5.4}$$

5.2 The asymptotic sampling properties

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

Theorem 5.1 (Equivalence of feasible and infeasible estimators) *Suppose Assumptions 5.1 and 5.2 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 $p \geq 1$ 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 $p_0 \leq p$, then the $O((np^{K-1})^{-1})$ term is zero.

PROOF. In Appendix D. □

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.

5.2.1 The bias and variance of the hybrid Whittle likelihood

The expressions in this section are derived under Assumptions 5.1 and 5.2.

<u>The bias</u> We show in Appendix E.4, that the asymptotic bias (in the sense of Bartlett) for $\widehat{\theta}_n = (\widehat{\theta}_{1,n}, \ldots, \widehat{\theta}_{d,n})$ is

$$\mathbb{E}[\widehat{\theta}_{j,n} - \theta_{j,n}] = \frac{H_{2,n}}{H_{1,n}^2} \sum_{r=1}^d I^{(j,r)} G_r(\theta_n) + O\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right) \quad 1 \leq j \leq d, \tag{5.5}$$

where $I^{(j,r)}$ and $G_r(\theta_n)$ is defined in (5.4). 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 E.4 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 straightfoward "Bartlett correction" can be made to estimate the bias in (5.5). 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

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

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,

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

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 (5.5) does not fully depict what we observe in the simulations in Section 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 $(H_{2,n}/H_{1,n}^2) \sum_{r=1}^{d} I^{(j,r)} G_r(\theta_n) \geq n^{-1} \sum_{r=1}^{d} I^{(j,r)} G_r(\theta_n)$. This would suggest that the hybrid Whittle likelihood should have a larger bias than the boundary corrected Whittle likelihood. But the simulations (see Section 6) suggest this is not necessarily true and the hybrid likelihood tends to have a smaller bias.

The variance We show in Corollary 3.1, Das et al. (2020) 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 $H_{2,n}/H_{1,n}^2 \sim n^{-1}$) Using this observation, standard Taylor expansion methods and Corollary 3.1 in Das et al. (2020), the asymptotic variance of $\widehat{\theta}_n$ is

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

where $V(\cdot)$ is defined in (5.3).

5.2.2 The role of order estimation on the rates

The order in the AR(p) approximation is selected using the AIC, where $\widehat{p} = \arg\min \text{AIC}(p)$ with

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

$\widehat{\sigma}_{p,n}^2 = \frac{1}{n-K_n} \sum_{t=K_n}^{n} (X_t - \sum_{j=1}^{p} \widehat{\phi}_{j,p} X_{t-j})^2$, K_n is such that $K_n^{2+\delta} \sim n$ for some $\delta > 0$. Ing and Wei (2005) assume that the underlying time series is a linear, stationary time series with an AR(∞) that satisfies Assumption K.1$-$K.4 in Ing and Wei (2005). They show that under the condition that the AR(∞) coefficients satisfy $(\sum_{j=p+1}^{\infty} |\phi_j|)^2 = O(p^{-2K})$, then $\widehat{p} = O_p(n^{1/(1+2K)})$ (see Example 2 in Ing and Wei (2005)). Thus, if $K > 5/2$, then $\widehat{p}^3/n^{1/2} \overset{\mathcal{P}}{\to} 0$ (where $\widehat{p} = O_p(n^{1/(1+2K)})$) and $\widehat{p} \overset{\mathcal{P}}{\to} \infty$ 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.

5.2.3   The computational cost of the estimators

We now discuss some of the implementation issues of the new 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 $\widehat{\underline{\phi}}_p = (\widehat{\phi}_{1,p}, ..., \widehat{\phi}_{p,p})'$ 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 $\widehat{\underline{\phi}}_p$ for $1 \le p \le K_n$ requires in total $O(K_n^2)$ arithmetic operations. Given the estimated AR coefficients $\widehat{\underline{\phi}}_{\widehat{p}}$, the predictive DFT $\{\widehat{J}_n(\omega_{k,n}; \widehat{f}_p)\}_{k=1}^n$ can be calculated in $O(\min(n \log n, n\widehat{p}))$ arithmetic operations (the details of the algorithm for optimal calculation can be found in Appendix A.1). 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 3.1). Therefore, we conclude that the computational cost of the new likelihoods is of the same order as the Whittle likelihood.

6   Empirical results

To substantiate our theoretical results, we conduct some simulations (further simulations can be found in Appendix F, G and H). To compare different methods, we evaluate six different quasi-likelihoods: the Gaussian likelihood (equation (1.1)), the Whittle likelihood (equation (1.3)), the boundary corrected Whittle likelihood (equation (4.8)), the hybrid Whittle likelihood (equation

(4.11)), 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 (also known as the cosine-bell taper) where

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

We set the proportion of tapering at each end of the time series is 0.1, i.e. $d = n/10$ (the default in R).

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, Re $\widetilde{J}_n(\omega_{k,n}; \widehat{f}_p)\overline{J_n(\omega_{k,n})}$ and 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 Re $\widetilde{J}_n(\omega_{k,n}; \widehat{f}_p)\overline{J_n(\omega_{k,n})}$ and 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 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 in parenthesis 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).

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 F.2, 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 5 gives the bias (first row) and the RMSE (second row) of each estimated parameter θ for both AR(1) and MA(1) models. We focus on positive θ, similar results are obtained for negative θ. The results are also summarized in Table 3 in Appendix F.1.

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 ϕ is large). A caveat is that for the AR(1) model the bias of the boundary corrected and hybrid Whittle 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 E.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 boundary corrected and hybrid Whittle it is $O(p^3 n^{-3/2})$. Indeed, the $O(p^3 n^{-3/2})$ 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 $\theta = 0.9$. This observation is noteworthy, as the computational cost of the Gaussian likelihood is greater than the computational cost of the new likelihoods (see Section 5.2). Thus the simulations suggest that in certain situations the new estimator may outperform the Gaussian likelihood at a lower computational cost.

In summary, the new likelihoods perform well compared with the standard methods, including
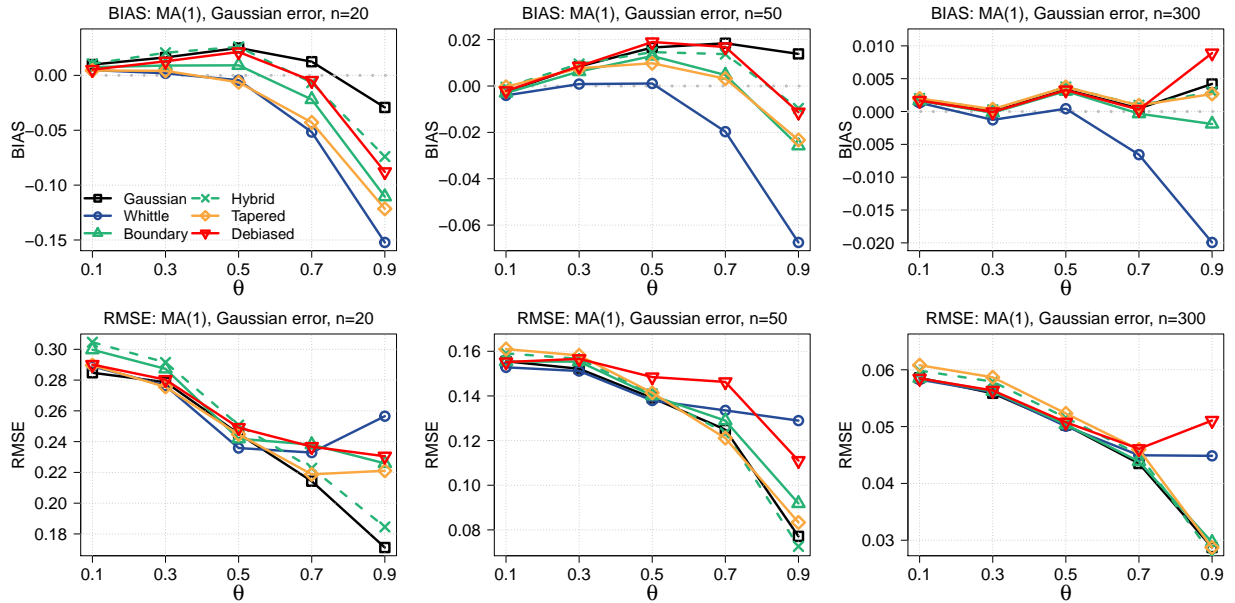
AR(1) model
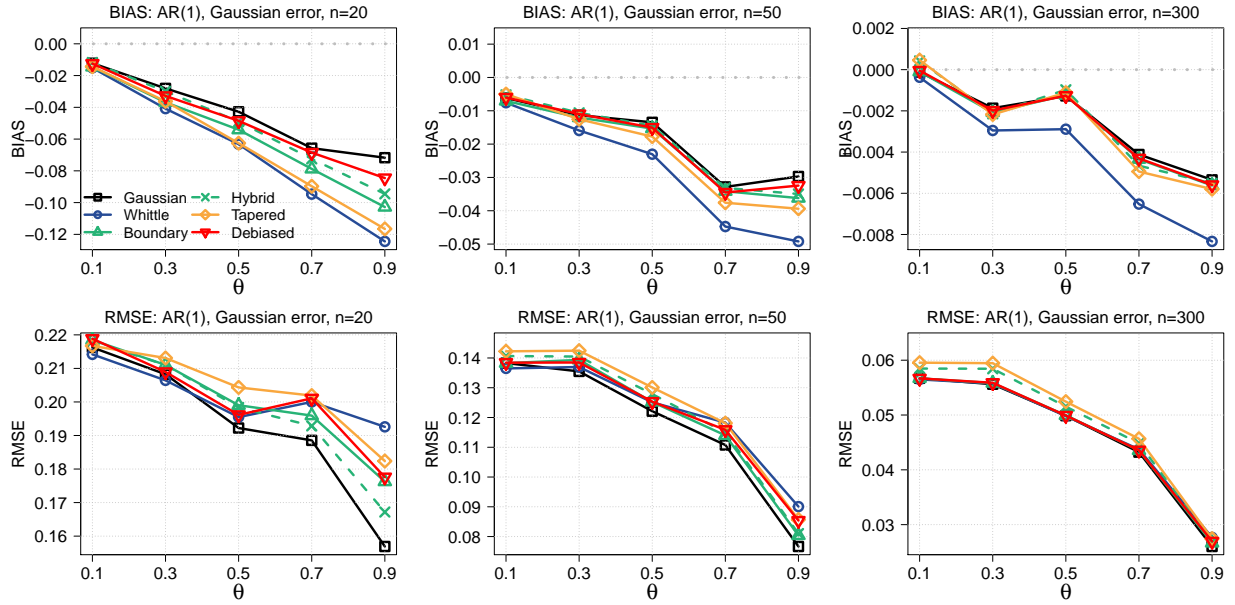
MA(1) model

Figure 5: *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).*

the benchmark Gaussian likelihood. As expected, for large sample sizes the performance of all the estimators improves considerably. And for some models, the new likelihood is able to outperform the Gaussian likelihood estimator. Though there is no clear rule when this will happen.

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 6 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 Θ 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.

For each simulation, we calculate the six different parameter estimators and the spectral
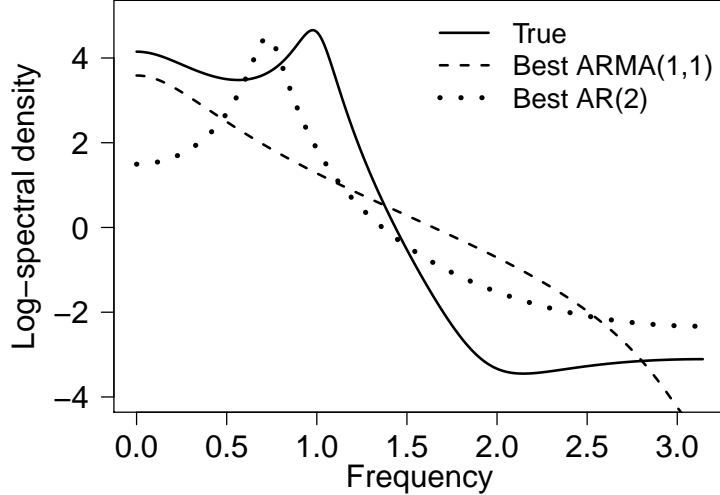
Figure 6: *Plot of* $\log f_Z(\omega)$ *and* $\log f_{\theta^{Best}}(\omega)$*; Theoretical ARMA(3,2) spectral density (solid), best fitting ARMA(1,1) spectral density (dashed), and best fitting AR(2) spectral density (dotted) for* $n = 20$.

divergence. The result of the estimators using the six different quasi-likelihoods is given in Table 1 (for ARMA(1,1)) and Table 2 (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 E.4), 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 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. Overall (in the simulations) the hybrid Whittle likelihood tends to outperform the Gaussian likelihood.

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 F.4 we compare the the likelihoods constructed using the predictive DFT based

n	Parameter	Gaussian	Whittle	Boundary	Hybrid	Tapered	Debiased
	ϕ	0.031(0.1)	-0.095(0.16)	-0.023(0.12)	-0.006(0.1)	-0.080(0.13)	0.187(0.11)
20	ψ	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)
	ϕ	0.012(0.07)	-0.054(0.09)	-0.006(0.07)	0.004(0.07)	-0.005(0.07)	0.154(0.11)
50	ψ	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)
	ϕ	0.002(0.03)	-0.014(0.03)	0(0.03)	0.001(0.03)	0(0.03)	0.093(0.08)
300	ψ	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 1: *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 *to denote the smallest RMSE and* blue *to denote the second smallest RMSE.*

n	Parameter	Gaussian	Whittle	Boundary	Hybrid	Tapered	Debiased
	ϕ_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)
20	ϕ_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)
	ϕ_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)
50	ϕ_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)
	ϕ_1	0.004(0.04)	-0.013(0.04)	0(0.04)	0.001(0.04)	0.001(0.04)	0.014(0.04)
300	ϕ_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(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 2: *Same as in Table 1 but fitting AR(2) model.*

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$(1,1)$ 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).

7 Concluding remarks and discussion

In this paper we have derived an exact expression for the differences $\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1})$ and $\underline{X}'_n[\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1})]\underline{X}_n$. These expressions are simple, with an intuitive interpretation, in terms of predicting outside the boundary of observation. They also provide a new perspective to the Whittle likelihood as an approximation based on a biorthogonal transformation. 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 Whittle 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 for this and one future aim is to investigate these differences.

We believe that it is possible to use a similar construction to obtain an expression for the difference between the Gaussian likelihood of a multivariate time series and the corresponding multivariate Whittle likelihood. The construction we use in this paper hinges on past and future predictions. In the univariate set-up there is an elegant symmetry for the predictors in the past and future. In the multivariate set-up there are some important differences. This leads to interesting, but different expressions for the predictive DFT. To prove analogous results to those in this paper, we will require 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.

The emphasis of this paper is on short memory time series. But we conclude by briefly discussing extensions to long memory time series. The fundamental feature (in the frequency domain) that distinguishes a short memory time series from a long memory time series is that the spectral density of a long memory time series is not bounded at the origin. However, we conjecture that the complete DFT described in Theorem 2.1 can have applications within this setting too. Suppose f is the spectral density corresponding to a long memory time series. And let $\widehat{J}_n(\omega_{k,n}; f)$ be defined as in (2.8). We assume that for $1 \leq k \leq n-1$ that $\widehat{J}_n(\omega_{k,n}; f)$ is a well defined random variable. Under these conditions, simple calculations show that equation (2.7) in Theorem 2.1 applies for $1 \leq k_1, k_2 \leq n-1$, but not when $k = n$ (the frequency at the origin). Using this, a result analogous to Corollary 2.1 can be obtained for long memory time series.

Theorem 7.1 (Inverse Toeplitz identity for long memory time series) *Suppose that the spectral density of a time series, f, is bounded away from zero and bounded on $[\varepsilon, \pi]$ for any $\varepsilon > 0$ and satisfies $f(\omega) \sim c_f |\omega|^{-2d}$ as $\omega \to 0$ for some $d \in (0, 1/2)$ and $c_f > 0$. Define the $(n-1) \times n$ submatrix $(\widetilde{D}_n(f))_{k,t} = (D_n(f))_{k,t}$ for $1 \leq k \leq (n-1)$, where $D_n(f)$ is defined in Theorem 3.3, equation (3.13). We assume that the entries of $\widetilde{D}_n(f)$ are finite. Let $\widetilde{F}_n$ denotes $(n-1) \times n$ submatrix of F_n, where $(\widetilde{F}_n)_{k,t} = n^{-1/2} e^{ik\omega_{t,n}}$. Let $\widetilde{\Delta}_n(f) = \mathrm{diag}(f(\omega_{1,n}), \ldots, f(\omega_{n-1,n}))$. Then*

$$(I_n - n^{-1} \mathbf{1}_n \mathbf{1}_n')\Gamma_n(f)^{-1} = \widetilde{F}_n^* \widetilde{\Delta}_n(f)^{-1}(\widetilde{F}_n + \widetilde{D}_n(f)) \tag{7.1}$$

where $\mathbf{1}_n = (1, \ldots, 1)'$ is an n-dimension vector

PROOF. See Appendix A. □

We note that an expansion of $D_n(f)_{k,t}$ and $\widehat{J}_n(\omega_{k,n}; f)$ is given in Theorem 3.3 in terms of the

AR(∞) coefficients corresponding to the spectral density f. The theorem above is contingent on these quantities being finite. We conjecture that this holds for invertible ARFIMA(p, d, q) time series where $d \in (0, 1/2)$. Unfortunately, a precise proof of this condition uses a different set of tools to those developed in this paper. Thus we leave it for future research.

The plug-in Gaussian likelihood replaces the population mean in the Gaussian likelihood with the sample mean. We now apply the above result to representing the long memory plug-in Gaussian likelihood in the frequency domain. We define the demeaned time series $\underline{X}_n^{(c)} = \underline{X}_n - \overline{X}\mathbf{1}_n$, where $\bar{X} = n^{-1}\sum_{t=1}^n X_t$. Then by using Theorem 7.1 we have

$$\frac{1}{n}\underline{X}_n^{(c)\prime}\Gamma_n(f_\theta)^{-1}\underline{X}_n^{(c)} \quad = \quad \frac{1}{n}\sum_{k=1}^{n-1}\frac{|J_n(\omega_{k,n})|^2}{f_\theta(\omega_{k,n})} + \frac{1}{n}\sum_{k=1}^{n-1}\frac{\widehat{J}_n^{(c)}(\omega_{k,n}; f_\theta)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})},$$

where $\widehat{J}_n^{(c)}(\cdot; f_\theta)$ denotes the predictive DFT of the demeaned time series $\underline{X}_n^{(c)}$. It would be of interest to show that the new likelihoods defined in this paper (and a local frequency version of them) could be used in the analysis of long memory time series. In Appendix G we have presented some preliminary simulations for long memory time series. The results are not conclusive, but they do suggest that the new likelihoods can, in some settings, reduce the bias for long memory parameter estimators.

In summary, the notion of biorthogonality and its application to the inversion of certain variance matrices may be of value in future research.

Acknowledgements

References

K. M. Abadir, W. Distaso, and L. Giraitis. Nonstationarity-extended local Whittle estimation. *J. Econometrics*, 141(2):1353–1384, 2007.

M. S. Bartlett. Approximate confidence intervals. II. More than one unknown parameter. *Biometrika*, 40(3/4):306–317, 1953.

G. Baxter. An asymptotic result for the finite predictor. *Math. Scand.*, 10:137–144, 1962.

G. Baxter. A norm inequality for a "finite-section" Wiener-Hopf equation. *Illinois J. Math.*, 7 (1):97–103, 1963.

R. J. Bhansali. The evaluation of certain quadratic forms occurring in autoregressive model fitting. *Ann. Statist.*, 10(1):121–131, 1982.

R. J. Bhansali. Asymptotically efficient autoregressive model selection for multistep prediction. *Ann. Inst. Statist. Math.*, 48(3):577–602, 1996.

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 of *Classics Appl. Math.* SIAM, Philadelphia, PA, 2001.

W. W. Chen and C. M. Hurvich. Semiparametric estimation of multivariate fractional cointegration. *J. Amer. Statist. Assoc.*, 98(463):629–642, 2003.

R. Cheng and M. Pourahmadi. Baxter's inequality and convergence of finite predictors of multivariate stochastic processes. *Probab. Theory Related Fields*, 95:115–124, 1993.

N. Choudhuri, S. Ghosal, and A. Roy. Bayesian estimation of the spectral density of a time series. *J. Amer. Statist. Assoc.*, 99(468):1050–1059, 2004.

J. Coursol and D. Dacunha-Castelle. Remarques sur l'approximation de la vraisemblance d'un processus Gaussien stationnaire. *Teor. Veroyatnost. i Primenen. (Theory of Probability and its Applications)*, 27(1):155–160, 1982.

D. R. Cox and E. J. Snell. A general definition of residuals. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 30(2):248–265, 1968.

R. Dahlhaus. Spectral analysis with tapered data. *J. Time Series Anal.*, 4(3):163–175, 1983.

R. Dahlhaus. Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *Ann. Statist.*, 16(2):808–841, 1988.

R. Dahlhaus. Nonparametric high resolution spectral estimation. *Probability Theory and Related Fields*, 85(2):147–180, 1990.

R. Dahlhaus. A likelihood approximation for locally stationary processes. *Ann. Statist.*, 28(6): 1762–1794, 2000.

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. *arXiv preprint arXiv:2007.00363*, 2020.

R. Fox and M. S. Taqqu. Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.*, 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. *J. Appl. Probab.*, 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.

C. M. Hurvich and W. W. Chen. An efficient taper for potentially overdifferenced long-memory time series. *J. Time Series Anal.*, 21(2):155–180, 2000.

C.-K. Ing and C.-Z. Wei. Order selection for same-realization predictions in autoregressive processes. *Ann. Statist.*, 33(5):2423–2474, 2005.

A. Inoue and Y. Kasahara. Explicit representation of finite predictor coefficients and its applications. *Ann. Statist.*, 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. *Statist. Probab. Lett.*, 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 Anal.*, 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. *Electron. J. Stat.*, 13(2):3710–3774, 2019.

J. Krampe, J.-P. Kreiss, and E. Paparoditis. Estimated Wold representation and spectral-density-driven bootstrap for time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 80:703–726, 2018.

J.-P. Kreiss, E. Paparoditis, and D. N. Politis. On the range of validity of the autoregressive sieve bootstrap. *Ann. Statist.*, 39(4):2103–2130, 2011.

H. R. Künsch. Statistical aspects of self-similar processes. In *Proceedings of the 1st World Congress of the Bernoulli Society*, volume 1, pages 67–74. VNU Science Press, 1987.

S. N. Lahiri. A necessary and sufficient condition for asymptotic independence of discrete Fourier transforms under short- and long-range dependence. *Ann. Statist.*, 31(2):613–641, 2003.

O. Lieberman. On plug-in estimation of long memory models. *Econometric Theory*, 21(2): 431–454, 2005.

M. Meyer, C. Jentsch, and J.-P. Kreiss. Baxter's inequality and sieve bootstrap for random fields. *Bernoulli*, 23(4B):2988–3020, 2017.

V. M. Panaretos and S. Tavakoli. Fourier analysis of stationary time series in function space. *Ann. Statist.*, 41(2):568–603, 2013.

Emanuel Parzen. Autoregressive spectral estimation. In *Time series in the frequency domain*, volume 3 of *Handbook of Statist.*, pages 221–247. North-Holland, Amsterdam, 1983.

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. Vol. 2.* Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London-New York, 1981. Multivariate series, prediction and control, Probability and Mathematical Statistics.

P. M. Robinson. Gaussian semiparametric estimation of long range dependence. *Ann. Statist.*, 23(5):1630–1661, 1995.

P. Shaman. An approximate inverse for the covariance matrix of moving average and autoregressive processes. *Ann. Statist.*, 3(2):532–538, 1975.

P. Shaman. Approximations for stationary covariance matrices and their inverses with application to ARMA models. *Ann. Statist.*, 4(2):292–301, 1976.

P. Shaman and R. A. Stine. The bias of autoregressive coefficient estimators. *J. Amer. Statist. Assoc.*, 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. *Ann. Math. Statist.*, 29(2):585–588, 1958.

S. Subba Rao. Orthogonal samples for estimators in time series. *J. Time Series Anal.*, 39: 313–337, 2018.

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. *Math. Ann.*, 84:232–244, 1921.

K. Tanaka. An asymptotic expansion associated with the maximum likelihood estimators in arma models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 46(1):58–67, 1984.

M. Taniguchi. On the second order asymptotic efficiency of estimators of gaussian arma processes. *Ann. Statist.*, 11:157–169, 1983.

D. Tjøstheim and J. Paulsen. Bias of some commonly-used time series estimates. *Biometrika*, 70(2):389–399, 1983.

A. van Delft and M. Eichler. A note on Herglotz's theorem for time series on functional spaces. *Stochastic Process. Appl.*, 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. *J. Aust. Math. Soc.*, 4:363–384, 1964.

P. Whittle. The analysis of multiple stationary time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 15:125–139, 1953.

Peter Whittle. *Hypothesis Testing in Time Series Analysis*. Thesis, Uppsala University, 1951.

H.-C. Zhang. Reduction of the asymptotic bias of autoregressive and spectral estimators by tapering. *J. Time Series Anal.*, 13(5):451–469, 1992.

Summary of results in the Supplementary material

To navigate the supplementary material, we briefly summarize the contents of each section.

1. In Appendix A we prove the results stated in Section 2 (which concern representing the Gaussian likelihood in the frequency domain and obtaining an explicit expression for the predictive DFT for AR models). Some of these proofs will use results from Appendix A.1.

2. In Appendix B.1 we prove both the first order approximation and higher order approximation results stated in Sections 3. The proof of Theorem 3.1, uses the extended Baxter's inequality (for completeness we prove this result in Appendix C.1, though we believe the result is well known). The proof Theorem 3.3 uses an explicit expression for finite predictors.

3. In Appendix B.3 we prove Lemma 4.1. The proof is similar to the proof of Theorem 3.1 but with some subtle differences.

4. Appendix C mainly deals with Baxter-type inequalities. In Appendix C we give a proof of the extended Baxter inequality. In Appendix C.2 we obtain Baxter type bounds for the derivatives of the finite predictors is stated (with respect to the unknown parameter). These results are used in Appendix C.3 to bound the difference between the derivatives of the Gaussian and Whittle likelihood.

5. In Appendix D we show consistency of the new likelihood estimators (defined in Section 4). We also state and prove the necessary lemmas for proving the asymptotic equivalence between the feasible and infeasible estimators (proof of Theorem 5.1).

6. In Appendix E we obtain the bias of the Gaussian, Whittle, boundary corrected and hybrid Whittle likelihoods under quite general assumptions on the underlying time series $\{X_t\}$. In particular, in Appendix E.1 we state the results in the one-parameter case (the proof is given in Appendix E.3). The results for the special case of the AR(1) is given in Appendix E.2. The bias for multi-parameters is given in Appendix E.4.

7. In Appendix F, we supplement the simulations in Section 6.

8. In Appendix G we apply the new likelihood estimation methods to long memory time series. We consider both parametric methods and also variants on the local Whittle likelihood estimator for the long memory parameter. In Appendix H we construct alternative estimators of the predictive DFT, which are used to build the new likelihoods. Through simulations we compare the estimators based on these different new likelihoods.

A Proof of Theorems 2.1, 7.1 and 2.3

We start this section by giving the proof of Theorems 2.1. This result is instrumental to the subsequent results in the paper.

PROOF of Theorem 2.1 First, to prove Theorem 2.1, we recall that it entails obtaining a transform $U_n\underline{X}_n$ where $\text{cov}_f(U_n\underline{X}_n, F_n\underline{X}_n) = \Delta_n(f)$. Pre and post multiplying this covariance with F_n^* and F_n gives

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

Thus our objective is to find the transform $\underline{Y}_n = F_n^*U_n\underline{X}_n$ such that $\text{cov}_f(\underline{Y}_n, \underline{X}_n) = C_n(f)$. 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 $\text{cov}_f(Y_{u,n}, X_v) = \sum_{\ell\in\mathbb{Z}} c_f(u-v+\ell n)$ and $Y_u \in \text{sp}(\underline{X}_n)$. Since $\text{cov}_f(X_{u+\ell n}, X_v) = c_f(u-v+\ell n)$, at least "formally" $\text{cov}_f(\sum_{\ell\in\mathbb{Z}} X_{u+\ell n}, X_v) = \sum_{\ell\in\mathbb{Z}} c_f(u-v+\ell n)$. However, $\sum_{\ell\in\mathbb{Z}} X_{u+\ell n}$ is neither a well defined random variable nor does not it belong to $\text{sp}(\underline{X}_n)$. We replace each element in the sum $\sum_{\ell\in\mathbb{Z}} X_{u+\ell n}$ with an element that belongs to $\text{sp}(\underline{X}_n)$ 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 $\text{sp}(\underline{X})$, i.e., the best linear predictor of Z given $\underline{X}$, then $\text{cov}_f(Z, \underline{X}) = \text{cov}_f(P_{\underline{X}}(Z), \underline{X})$. Let $\widehat{X}_{\tau,n}$ denote best linear predictor of X_τ given $\underline{X}_n = (X_1, \ldots, X_n)$ (as defined in (2.4)). $\widehat{X}_{\tau,n}$ retains the pertinent properties of X_τ in the sense that $\text{cov}_f(\widehat{X}_{\tau,n}, X_t) = c_f(\tau-t)$ for all $\tau\in\mathbb{Z}$ and $1\le t\le n$. Define

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

where we note that $Y_{u,n}$ a well defined random variable, since by using Lemma B.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

$$\text{cov}_f(Y_{u,n}, X_v) = \sum_{\ell\in\mathbb{Z}} c_f(u-v+\ell n) = (C_n(f))_{u,v}, \tag{A.1}$$

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

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

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 $1 \leq \tau \leq n$, $\phi_{t,n}(\tau; f) = \delta_{\tau,t}$, 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

$$
\phi_{t,n}(m; f) = \phi_{n+1-t,n}(n+1-m; f) \qquad \text{for } m > n. \quad (A.3)
$$

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

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

It immediately follows from the above decomposition that $U_n = F_n + D_n(f)$ where $D_n(f)$ is defined in (2.6). Thus proving (2.5).

To prove (2.7), we first observe that (2.5) implies

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

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 (A.2) we have

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

This immediately proves (2.7). $\qquad \square$

Note that equation (2.7) can be verified directly by using the properties of linear predictors and covariances discussed in the above proof.

To prove Theorem 2.3 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

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

where $\phi_1(\omega) = 1 - \phi e^{-i\omega}$. In order to prove Theorem 2.3, 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

$$
A_p(\underline{\phi}) = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}. \tag{A.4}
$$

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

$$
\widehat{J}_{n,L}(\omega; f_\theta) = n^{-1/2} \sum_{\tau=-\infty}^{0} \left[A_p(\underline{\phi})^{|\tau|+1} \underline{X}_p\right]_{(1)} e^{i\tau\omega}. \tag{A.5}
$$

Lemma A.1 Let $\widehat{J}_{n,L}(\omega)$ be defined as in (A.5), where the parameters $\underline{\phi}$ are such that the roots of $\phi(z) = 1 - \sum_{j=1}^p \phi_j z^j$ lie outside the unit circle. Then an analytic expression for $\widehat{J}_{n,L}(\omega; f_\theta)$ is

$$
\widehat{J}_{n,L}(\omega; f_\theta) = \frac{n^{-1/2}}{\phi_p(\omega)} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{-is\omega}. \tag{A.6}
$$

where $\phi_p(\omega) = 1 - \sum_{s=1}^{p} \phi_s e^{-is\omega}$.

PROOF. By using (A.11) 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 (A.5) 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

$$\widehat{J}_{n,L}(\omega; f_\theta) = \frac{n^{-1/2}}{\phi_p(\omega)} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{-is\omega}, \tag{A.7}$$

Thus we obtain the desired result. $\qquad\square$

PROOF of Theorem 2.3 To prove (2.15), we note that the same proof as that above can be used to prove that the right hand side predictive DFT $\widehat{J}_{n,R}(\omega; f_\theta)$ has the representation

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

Since $\widehat{J}_n(\omega; f_\theta) = \widehat{J}_{n,L}(\omega; f_\theta) + \widehat{J}_{n,R}(\omega; f_\theta)$, Lemma A.1 and the above give an explicit expression for $\widehat{J}_n(\omega; f_\theta)$, thus proving equation (2.15).

To prove (2.16) we use that

$$(\widehat{J}_n(\omega_{1,n}; f_\theta), \dots, \widehat{J}_n(\omega_{n,n}; f_\theta))' = D_n(f_\theta)\underline{X}_n.$$

Now by using (2.15) together with the above we immediately obtain (2.16).

Finally, we prove (2.17). 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 $p+1 \leq r \leq n$. For $1 \leq t \leq p$ 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 $1 \leq t \leq p$,

$$
\begin{aligned}
(F_n^* \Delta_n(f_\theta^{-1}) D_n(f_\theta))_{s,n-t+1} &= \frac{1}{n} \sum_{k=1}^{n} \frac{\overline{\phi_{t,p}(\omega_{k,n})}}{f_\theta(\omega_{k,n})} \exp(i(1-s)\omega_{k,n}) \\
&= \frac{\sigma^{-2}}{n} \sum_{k=1}^{n} \phi_p(\omega_{k,n}) \sum_{\ell=0}^{p-t} \phi_{\ell+t} \exp(i\ell\omega_{k,n}) \exp(i(1-s)\omega_{k,n}) \\
&= \sigma^{-2} \sum_{\ell=0}^{p-t} \phi_{\ell+t} \frac{1}{n} \sum_{k=1}^{n} \phi_p(\omega_{k,n}) \exp(i(\ell+1-s)\omega_{k,n}) \\
&= \sigma^{-2} \sum_{\ell=0}^{p-t} \phi_{\ell+t} \widetilde{\phi}_{(\ell+1-s) \bmod n}.
\end{aligned}
$$

$\square$

PROOF of Equation (2.18) 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

$$II = \frac{\sigma^{-2}}{n^{3/2}} \sum_{k=1}^{n} \frac{\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} \frac{1}{\overline{\phi_p(\omega_{k,n})}} \sum_{\ell=1}^{p} X_{n+1-\ell} \sum_{s=0}^{p-\ell} \phi_{\ell+s} e^{i(s+1)\omega_{k,n}}. \tag{A.8}$$

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

$$\begin{aligned}
I &= -\frac{\sigma^{-2}}{n} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \sum_{j=0}^{p} \phi_j \phi_{s+\ell} \frac{1}{n^{1/2}} \sum_{k=1}^{n} \overline{J_n(\omega_{k,n})} e^{-i(s-r)\omega_{k,n}} \quad (\text{set } \phi_0 = -1) \\
&= -\frac{\sigma^{-2}}{n} \sum_{\ell=1}^{p} X_\ell \sum_{s=0}^{p-\ell} \sum_{j=0}^{p} \phi_j \phi_{s+\ell} X_{-(s-j) \bmod n}.
\end{aligned}$$

The proof of II is similar. Altogether this proves the result. $\qquad\square$

Finally, in this section we prove Theorem 7.1 (from Section 7). This result is analogous to Theorem 2.1.

PROOF of Theorem 7.1 Under the stated condition that $\widetilde{D}_n$ is a finite matrix, then $(\widetilde{F}_n + \widetilde{D}_n)\underline{X}_n$ is a well defined random variable. Thus

$$\mathrm{cov}((\widetilde{F}_n + \widetilde{D}_n)\underline{X}_n, \widetilde{F}_n\underline{X}_n) = (\widetilde{F}_n + \widetilde{D}_n)\Gamma_n \widetilde{F}_n^* = \widetilde{\Delta}_n. \tag{A.9}$$

Therefore, using (A.9) and $F_n^* = [\widetilde{F}_n^*, \overline{e_n}] = [\widetilde{F}_n^*, e_n]$ gives

$$\begin{aligned}
(\widetilde{F}_n + \widetilde{D}_n)\Gamma_n F_n^* = (\widetilde{F}_n + \widetilde{D}_n)\Gamma_n[\widetilde{F}_n^*, e_n] &= [(\widetilde{F}_n + \widetilde{D}_n)\Gamma_n \widetilde{F}_n^*, (\widetilde{F}_n + \widetilde{D}_n)\Gamma_n e_n] \\
&= [\widetilde{\Delta}_n, (\widetilde{F}_n + \widetilde{D}_n)\Gamma_n e_n].
\end{aligned}$$

Note that for $1 \le k \le n-1$,

$$(e_k + d_k)'\Gamma_n e_n = \mathrm{cov}(\widetilde{J}_n(\omega_{k,n}; f), J_n(0)) = 0.$$

Therefore, we have

$$(\widetilde{F}_n + \widetilde{D}_n)\Gamma_n F_n^* = [\widetilde{\Delta}_n, \mathbf{0}_{n-1}].$$

where $\mathbf{0}_{n-1} = (0, ..., 0)'$ is a $(n-1)$-dimension zero vector. Right multipling the above with $F_n \Gamma_n^{-1}$ gives

$$(\widetilde{F}_n + \widetilde{D}_n) = [\widetilde{\Delta}_n, \mathbf{0}_{n-1}] F_n \Gamma_n^{-1}.$$

Left multiplying the above with $[\widetilde{\Delta}_n^{-1}, \mathbf{0}_{n-1}]^T$ gives

$$
\begin{aligned}
\begin{bmatrix} \widetilde{\Delta}_n^{-1} \\ \mathbf{0}'_{n-1} \end{bmatrix} (\widetilde{F}_n + \widetilde{D}_n) = \begin{bmatrix} \widetilde{\Delta}_n^{-1} \\ \mathbf{0}'_{n-1} \end{bmatrix} [\widetilde{\Delta}_n, \mathbf{0}_{n-1}] F_n \Gamma_n^{-1} &= \begin{bmatrix} I_{n-1} & \mathbf{0}_{n-1} \\ \mathbf{0}'_{n-1} & 0 \end{bmatrix} F_n \Gamma_n^{-1} \\
&= \begin{bmatrix} I_{n-1} & \mathbf{0}_{n-1} \\ \mathbf{0}'_{n-1} & 0 \end{bmatrix} \begin{bmatrix} \widetilde{F}_n \\ e'_n \end{bmatrix} \Gamma_n^{-1} \\
&= \begin{bmatrix} \widetilde{F}_n \\ \mathbf{0}'_{n-1} \end{bmatrix} \Gamma_n^{-1}
\end{aligned}
$$

Left multiplying the above with $[\widetilde{F}_n^*, \mathbf{0}_{n-1}]$ gives

$$
\widetilde{F}_n^* \widetilde{\Delta}_n^{-1} (\widetilde{F}_n + \widetilde{D}_n) = [\widetilde{F}_n^*, \mathbf{0}_{n-1}] \begin{bmatrix} \widetilde{\Delta}_n^{-1} \\ \mathbf{0}'_{n-1} \end{bmatrix} (\widetilde{F}_n + \widetilde{D}_n) = [\widetilde{F}_n^*, \mathbf{0}_{n-1}] \begin{bmatrix} \widetilde{F}_n \\ \mathbf{0}'_{n-1} \end{bmatrix} \Gamma_n^{-1} = \widetilde{F}_n^* \widetilde{F}_n \Gamma_n^{-1}. \quad \text{(A.10)}
$$

Finally, using that $e_n = n^{-1/2} \mathbf{1}_n$

$$
I_n = F_n^* F_n = [\widetilde{F}_n^*, n^{-1/2} \mathbf{1}_n] \begin{bmatrix} \widetilde{F}_n \\ n^{-1/2} \mathbf{1}'_n \end{bmatrix} = \widetilde{F}_n^* \widetilde{F}_n + n^{-1} \mathbf{1}_n \mathbf{1}'_n.
$$

Substituting the above into (A.10) gives

$$
\widetilde{F}_n^* \widetilde{\Delta}_n^{-1} (\widetilde{F}_n + \widetilde{D}_n) = (I_n - n^{-1} \mathbf{1}_n \mathbf{1}'_n) \Gamma_n^{-1}.
$$

This proves the identity (7.1). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

A.1 Some auxiliary lemmas

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 $\tau \leq 0$. 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 A.2 *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 (2.4). Then* $\phi_j(\tau; f) = \sum_{s=0}^{p-j} \phi_{j+s} \psi_{|\tau|-s}$.

$$
[A_p(\underline{\phi})^{|\tau|+1} \underline{X}_p]_{(1)} = \sum_{\ell=1}^p X_\ell \sum_{s=0}^{p-\ell} \phi_{\ell+s} \psi_{|\tau|-s}. \quad \text{(A.11)}
$$

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 jth row of A^m ($m \geq 1$) 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

$$
\begin{pmatrix}
a_{1,m} & a_{2,m} & \cdots & a_{p,m} \\
a_{1,m-1} & a_{2,m-1} & \cdots & a_{p,m-1} \\
\vdots & \vdots & \ddots & \vdots \\
a_{1,m-p+1} & a_{2,m-p+1} & \cdots & a_{p,m-p+1}
\end{pmatrix}
=
\begin{pmatrix}
\phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0
\end{pmatrix}
\begin{pmatrix}
a_{1,m-1} & a_{2,m-1} & \cdots & a_{p,m-1} \\
a_{1,m-2} & a_{2,m-2} & \cdots & a_{p,m-2} \\
\vdots & \vdots & \ddots & \vdots \\
a_{1,m-p} & a_{2,m-p} & \cdots & a_{p,m-p}
\end{pmatrix}.
$$

From the above we observe that $a_{\ell,m}$ satisfies the system of equations

$$
\begin{aligned}
a_{\ell,m} &= \phi_\ell a_{1,m-1} + a_{\ell+1,m-1} & 1 \leq \ell \leq p-1 \\
a_{p,m} &= \phi_p a_{1,m-1}.
\end{aligned}
\tag{A.12}
$$

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 $|z| \leq 1$ and has the power series expansion $(1 - \sum_{i=1}^p \phi_i z)^{-1} = \sum_{i=0}^\infty \psi_i z^i$ for $|z| \leq 1$. 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 $\{a_{\ell,m}; 2 \leq \ell \leq p\}$ in terms of $\{\phi_i\}$ and $\{\psi_i\}$. Solving the system of equations in (A.12), starting with $a_{1,1} = \psi_1$ and recursively solving for $a_{p,m}, \ldots, a_{2,m}$ we have

$$
\begin{aligned}
a_{p,r} &= \phi_p \psi_{r-1} & m - p \leq r \leq m \\
a_{\ell,r} &= \phi_\ell a_{1,r-1} + a_{\ell+1,r-1} & 1 \leq \ell \leq p-1, \quad m - p \leq r \leq m
\end{aligned}
$$

This gives $a_{p,m} = \phi_p \psi_{m-1}$, for $\ell = p-1$

$$
\begin{aligned}
a_{p-1,m} &= \phi_{p-1} a_{1,m-1} + a_{p,m-1} \\
&= \phi_{p-1} \psi_{m-1} + \psi_p \psi_{m-2}
\end{aligned}
$$

$$
\begin{aligned}
a_{p-2,m} &= \phi_{p-2} a_{1,m-1} + a_{p-1,m-1} \\
&= \phi_{p-2} \psi_{m-1} + \phi_{p-1} \psi_{m-2} + \psi_p \psi_{m-3}
\end{aligned}
$$

49

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

$$
a_{p-r,m} = \sum_{s=0}^{r} \phi_{p-r+s} \psi_{m-1-s} \qquad 0 \le r \le p-1.
$$

In the last line of the above we change variables with $\ell = p - r$ to give for $m \ge 1$

$$
a_{\ell,m} = \sum_{s=0}^{p-\ell} \phi_{\ell+s} \psi_{m-1-s} \qquad 1 \le \ell \le p,
$$

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. $\qquad\square$

In the proof of Lemma A.1 we obtained an expression in terms for the best linear predictor based on the parameters of an $AR(p)$ process. For completeness, we obtain an expression for the left hand side predictive DFT for a general second order stationary time series which is based on infinite future. This expression can be used to prove the identity in equation (3.3).

For $\tau \le 0$, let $\widehat{X}_\tau$ be the best linear predictor of X_τ given the infinite future $\{X_t\}_{t=1}^{\infty}$ i.e.

$$
\widehat{X}_\tau = \sum_{s=1}^{\infty} \phi_s(\tau; f) X_s. \tag{A.13}
$$

The left hand side predictive DFT given the infinite future is defined as

$$
\widehat{J}_{\infty,L}(\omega; f) = n^{-1/2} \sum_{\tau=-\infty}^{0} \widehat{X}_\tau e^{i\tau\omega}. \tag{A.14}
$$

Corollary A.1 *Suppose that f satisfies Assumption 3.1. Let $\{\phi_j(f)\}_{j=1}^{\infty}$ denote the $AR(\infty)$ coefficients associated with f. Let $\widehat{J}_{\infty,L}(\omega; f)$ be defined as in (A.14). Then, the best linear predictor of X_τ given $\{X_t\}_{t=1}^{\infty}$ where $\tau \le 0$ (defined in (A.13)) can be evaluated using the recursion $\widehat{X}_\tau =$*

$\sum_{s=1}^{\infty} \phi_s(f)\widehat{X}_{\tau+s}$, where we set $\widehat{X}_t = X_t$ for $t \geq 1$. Further, $\phi_\ell(\tau; f) = \sum_{s=0}^{\infty} \phi_{\ell+s}(f)\psi_{|\tau|-s}(f)$ and

$$\widehat{J}_{\infty,L}(\omega; f) = \frac{n^{-1/2}}{\phi(\omega; f)} \sum_{\ell=1}^{\infty} X_\ell \sum_{s=0}^{\infty} \phi_{\ell+s}(f)e^{-is\omega}. \tag{A.15}$$

PROOF. We recall that for general processes X_t with f bounded away from 0 has the AR(∞) representation $X_t = \sum_{j=1}^{\infty} \phi_j X_{j+t} + \varepsilon_t$ where $\{\varepsilon_t\}$ are uncorrelated random variables. This immediately implies that the best linear prediction of X_τ given $\{X_t\}_{t=1}^{\infty}$ can be evaluated using the recursion $\widehat{X}_\tau = \sum_{s=1}^{\infty} \phi_s(f)\widehat{X}_{\tau+s}$. By using (A.11) where we let $p \to \infty$ we have $\phi_\ell(\tau; f) = \sum_{s=0}^{\infty} \phi_{\ell+s}(f)\psi_{|\tau|-s}(f)$. This gives the first part of the result.

To obtain an expression for $\widehat{J}_L(\cdot; f)$ we use (A.6) where we let $p \to \infty$ to obtain the desired result. □

By a similar argument we can show that

$$\widehat{J}_{\infty,R}(\omega; f) = \sum_{\tau=n+1}^{\infty} \widehat{X}_\tau e^{i\tau\omega} = e^{i(n+1)\omega} \frac{n^{-1/2}}{\phi(\omega; f_\theta)} \sum_{\ell=1}^{n} X_{n+1-\ell} \overline{\phi_t^{\infty}(\omega; f_\theta)}. \tag{A.16}$$

Since $\widehat{J}_{\infty,n}(\omega; f) = \widehat{J}_{\infty,L}(\omega; f) + \widehat{J}_{\infty,R}(\omega; f)$, by using (A.15) and (A.16) we immediately obtain the identity (3.3).

A.1.1 A fast algorithm for computing the predictive DFT

To end this section, for $p \leq n$, we provide an $O(\min(n \log n, np))$ algorithm to compute the predictive DFT of an AR(p) spectral density at frequencies $\omega = \omega_{k,n}$, $1 \leq k \leq n$. Recall from (2.15),

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

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

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

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

By direct calculation $\{\phi_p(\omega_{k,n}) : 0 \le k \le n-1\}$ and $\{\sum_{s=0}^{p-1} Y_s e^{-is\omega_{k,n}} : 0 \le k \le n-1\}$ 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 kth 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 $\{\phi_p(\omega_{k,n}) : 0 \le k \le n-1\}$ and $\{\sum_{s=0}^{p-1} Y_s e^{-is\omega_{k,n}} : 0 \le k \le n-1\}$ 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 $\{\phi_p(\omega_{k,n}) : 0 \le k \le n-1\}$ and $\{\sum_{s=0}^{p-1} Y_s e^{-is\omega_{k,n}} : 0 \le k \le n-1\}$ is $O(\min(n \log n, np))$.

Therefore, the overall computational complexity is $O(p \log p + n \log n \wedge np) = O(\min(n \log n, np))$.

B Proof of results in Sections 3 and 4.1

B.1 Proof of Theorems 3.1 and 3.2

Many of the results below hinge on a small generalisation of Baxter's inequality which we summarize below.

Lemma B.1 (Extended Baxter's inequality) *Suppose $f(\cdot)$ is a spectral density function which satisfies Assumption 3.1. Let $\psi(\cdot)$ and $\phi(\cdot)$ be defined as in (3.1) (for the simplicity, we omit the notation f inside the $\psi(\cdot)$ and $\phi(\cdot)$). 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_τ given $\underline{X}_n = \{X_t\}_{t=1}^n$ and $\{\phi_s(\tau)\}$ the corresponding the coefficients in the best linear predictor of X_τ given $\underline{X}_\infty = \{X_t\}_{t=1}^\infty$, where $\tau \le 0$. Suppose p is large enough such that $\left\|\phi_p^\infty\right\|_K \left\|\psi\right\|_K \le \varepsilon < 1$. Then for all $n > p$ we have*

$$\sum_{s=1}^n (2^K + s^K) |\phi_{s,n}(\tau) - \phi_s(\tau)| \le C_{f,K} \sum_{s=n+1}^\infty (2^K + s^K) |\phi_s(\tau)|, \tag{B.1}$$

where $C_{f,K} = \frac{3-\varepsilon}{1-\varepsilon} \left\|\phi\right\|_K^2 \left\|\psi\right\|_K^2$ and $\phi_s(\tau) = \sum_{j=0}^\infty \phi_{s+j} \psi_{|\tau|-j}$ (we set $\psi_0 = 1$ and $\psi_j = 0$ for $j < 0$).

PROOF. For completeness we give the proof in Appendix C. □

PROOF of Equation (3.2) Since

$$(D_{\infty,n}(f_\theta))_{k,t} = n^{-1/2} \sum_{\tau \le 0} \left(\phi_t(\tau) e^{i\tau\omega_{k,n}} + \phi_{n+1-t}(\tau) e^{-i(\tau-1)\omega_{k,n}} \right) \tag{B.2}$$

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 (B.2) gives

$$
\begin{aligned}
n^{-1/2}\sum_{\tau\leq 0}\phi_t(\tau)e^{i\tau\omega_{k,n}} &= n^{-1/2}\sum_{\tau\leq 0}\sum_{s=0}^{\infty}\phi_{t+s}\psi_{-\tau-s}e^{i\tau\omega_{k,n}} \\
&= n^{-1/2}\sum_{s=0}^{\infty}\phi_{t+s}e^{-is\omega_{k,n}}\sum_{\tau\leq 0}\psi_{-\tau-s}e^{-i(-\tau-s)\omega_{k,n}} \\
&= n^{-1/2}\psi(\omega_{k,n})\sum_{s=0}^{\infty}\phi_{t+s}e^{-is\omega_{k,n}} \\
&= n^{-1/2}\phi(\omega_{k,n})^{-1}\phi_t^{\infty}(\omega_{k,n}),
\end{aligned}
$$

which gives the first term in (3.2). The second term follows similarly. Thus giving the identity in equation (3.2). □

Next we prove Theorem 3.1. To do this we note that the entries of $F_n^*\Delta_n(f_\theta^{-1})D_{\infty,n}(f)$ are

$$
\begin{aligned}
(F_n^*&\Delta_n(f_\theta^{-1})D_{\infty,n}(f))_{s,t} \\
&= \sum_{\tau\leq 0}\left[\phi_t(\tau;f)G_{1,n}(s,\tau;f_\theta) + \phi_{n+1-t}(\tau;f)G_{2,n}(s,\tau;f_\theta)\right],
\end{aligned} \qquad (B.3)
$$

where $G_{1,n}$ and $G_{2,n}$ are defined as in (2.14). Thus

$$
\begin{aligned}
\left(F_n^*\Delta_n(f_\theta^{-1})\left[D_n(f) - D_{\infty,n}(f)\right]\right)_{s,t} \\
= \sum_{\tau\leq 0}[\{\phi_{t,n}(\tau;f) - \phi_t(\tau;f)\}G_{1,n}(s,\tau;f_\theta) \\
+ \{\phi_{n+1-t,n}(\tau;f) - \phi_{n+1-t}(\tau;f)\}G_{2,n}(s,\tau;f_\theta)].
\end{aligned} \qquad (B.4)
$$

To prove Theorem 3.1 we bound the above terms.

PROOF of Theorem 3.1. To simplify notation we only emphasis the coefficients associated with f_θ 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 (3.4) simply follows from the definitions of $D_n(f)$ and $D_{\infty,n}(f)$.

Next we prove (3.5). By using (B.4) we have

$$
\left\|F_n^*\Delta_n(f_\theta^{-1})D_n(f) - F_n^*\Delta_n(f_\theta^{-1})D_{\infty,n}(f)\right\|_1 \leq T_{1,n} + T_{2,n},
$$

53

where

$$T_{1,n} = \sum_{s,t=1}^{n} \sum_{\tau=-\infty}^{0} |\phi_{s,n}(\tau) - \phi_s(\tau)||G_{1,n}(t,\tau;f_\theta)|$$

$$T_{2,n} = \sum_{s,t=1}^{n} \sum_{\tau=-\infty}^{0} |\phi_{n+1-s,n}(\tau) - \phi_{n+1-s}(\tau)||G_{2,n}(t,\tau;f_\theta)|.$$

We focus on $T_{1,n}$, noting that the method for bounding $T_{2,n}$ is similar. Exchanging the summands we have

$$T_{1,n} \leq \sum_{\tau=-\infty}^{0} \sum_{t=1}^{n} |G_{1,n}(t,\tau;f_\theta)| \sum_{s=1}^{n} |\phi_{s,n}(\tau) - \phi_s(\tau)|.$$

To bound $\sum_{s=1}^{n} |\phi_{s,n}(\tau) - \phi_s(\tau)|$ we require the generalized Baxter's inequality stated in Lemma B.1. Substituting the bound in Lemma B.1 into the above (and for a sufficiently large n) we have

$$T_{1,n} \leq C_{f,0} \sum_{\tau=-\infty}^{0} \sum_{t=1}^{n} |G_{1,n}(t,\tau;f_\theta)| \sum_{s=n+1}^{\infty} |\phi_s(\tau)|.$$

Using that $G_{1,n}(t,\tau) = \sum_{a \in \mathbb{Z}} K_{f_\theta^{-1}}(\tau - t + an)$ we have the bound

$$\begin{aligned}
T_{1,n} &\leq C_{f,0} \sum_{\tau=-\infty}^{0} \sum_{t=1}^{n} \sum_{a \in \mathbb{Z}} |K_{f_\theta^{-1}}(t - \tau + an)| \sum_{s=n+1}^{\infty} |\phi_s(\tau)| \\
&= C_{f,0} \sum_{r \in \mathbb{Z}} |K_{f_\theta^{-1}}(r)| \sum_{\tau=-\infty}^{0} \sum_{s=n+1}^{\infty} |\phi_s(\tau)|.
\end{aligned}$$

Therefore,

$$\begin{aligned}
T_{1,n} &\leq C_{f,0} \sum_{r \in \mathbb{Z}} |K_{f_\theta^{-1}}(r)| \sum_{\tau=-\infty}^{0} \sum_{s=n+1}^{\infty} |\phi_s(\tau)| \\
&\leq C_{f,0} \sum_{r \in \mathbb{Z}} |K_{f_\theta^{-1}}(r)| \sum_{\tau=-\infty}^{0} \sum_{s=n+1}^{\infty} \sum_{j=0}^{\infty} |\phi_{s+j}||\psi_{-\tau-j}| \quad (\text{use } \phi_s(\tau) = \sum_{j=0}^{\infty} \phi_{s+j}\psi_{|\tau|-j}) \\
&= C_{f,0} \sum_{r \in \mathbb{Z}} |K_{f_\theta^{-1}}(r)| \sum_{\tau=0}^{\infty} |\psi_{\tau-j}| \sum_{s=n+1}^{\infty} \sum_{j=0}^{\infty} |\phi_{s+j}| \quad (\text{change limits of } \sum_{\tau}) \\
&\leq C_{f,0} \sum_{r \in \mathbb{Z}} |K_{f_\theta^{-1}}(r)| \sum_{\ell} |\psi_\ell| \sum_{u=n+1}^{\infty} |u\phi_u| \quad (\text{change of variables } u = s + j).
\end{aligned}$$

Next we use Assumption 3.1(i) to give

$$
\begin{aligned}
T_{1,n} \;&\leq\; C_{f,0}\sum_{r\in\mathbb{Z}}|K_{f_\theta^{-1}}(r)|\sum_\ell|\psi_\ell|\sum_{s=n+1}^{\infty}\frac{s^K}{s^{K-1}}|\phi_s| \\
&\leq\; \frac{C_{f,0}}{n^{K-1}}\sum_{r\in\mathbb{Z}}|K_{f_\theta^{-1}}(r)|\sum_\ell|\psi_\ell|\sum_{s=n+1}^{\infty}|s^K\phi_s| \\
&\leq\; \frac{C_{f,0}}{n^{K-1}}\rho_{n,K}(f)\|\psi\|_0\|\phi\|_K\sum_{r\in\mathbb{Z}}|K_{f_\theta^{-1}}(r)|.
\end{aligned}
$$

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

$$
\sum_{r=-\infty}^{\infty}|K_{f_\theta}(r)|\leq\sigma_{f_\theta}^{-2}\|\phi_{f_\theta}\|_0^2. \tag{B.5}
$$

Substituting this into the above yields the bound

$$
T_{1,n}\leq\frac{C_{f,0}}{\sigma_{f_\theta}^2 n^{K-1}}\rho_{n,K}(f)\|\psi\|_0\|\phi_{f_\theta}\|_0^2\|\phi\|_K.
$$

The same bound holds for $T_{2,n}$. Together the bounds for $T_{1,n}$ and $T_{2,n}$ give

$$
\left\|F_n^*\Delta_n(f_\theta^{-1})D_n(f_\theta)-F_n^*\Delta_n(f_\theta^{-1})D_{\infty,n}(f_\theta)\right\|_1\leq\frac{2C_{f,0}}{\sigma_{f_\theta}^2 n^{K-1}}\rho_{n,K}(f)\|\psi\|_0\|\phi_{f_\theta}\|_0^2\|\phi\|_K.
$$

Replacing $\|\psi_f\|_0=\|\psi\|_0$ and $\|\phi_f\|_K=\|\phi\|_K$, this proves (3.5).

To prove (3.6) we recall

$$
\|X_tX_s\|_{\mathbb{E},q}=(\mathbb{E}|X_tX_s|^q)^{1/q}\leq\left(\mathbb{E}|X_t|^{2q}\right)^{1/2q}\left(\mathbb{E}|X_s|^{2q}\right)^{1/2q}\leq\|X\|_{\mathbb{E},2q}^2.
$$

Therefore,

$$
\begin{aligned}
n^{-1}&\left\|\underline{X}_n'F_n^*\Delta_n(f_\theta^{-1})\left(D_n(f)-D_{\infty,n}(f)\right)\underline{X}_n\right\|_{\mathbb{E},q} \\
&\leq n^{-1}\sum_{s,t=1}^{n}\left|\left(F_n^*\Delta_n(f_\theta^{-1})\left(D_n(f)-D_{\infty,n}(f)\right)\right)_{s,t}\right|\|X_tX_s\|_{\mathbb{E},q} \\
&\leq n^{-1}\left\|F_n^*\Delta_n(f_\theta^{-1})\left(D_n(f)-D_{\infty,n}(f)\right)\right\|_1\|X\|_{\mathbb{E},2q}^2 \\
&\leq\frac{2C_{f,0}}{\sigma_{f_\theta}^2 n^K}\rho_{n,K}(f)\|\psi_f\|_0\|\phi_{f_\theta}\|_0^2\|\phi_f\|_K\|X\|_{\mathbb{E},2q}^2,
\end{aligned}
$$

where the last line follows from the inequality in (3.5). This proves (3.6). □

PROOF of Theorem 3.2 For notational simplicity, we omit the parameter dependence on f_θ. We first prove (3.8). We observe that

$$\left\|F_n^* \Delta_n(f_\theta^{-1}) D_{\infty,n}(f_\theta)\right\|_1 \leq \sum_{s,t=1}^{n} \sum_{\tau=-\infty}^{0} \left(|\phi_s(\tau)||G_{1,n}(t,\tau)| + |\phi_{n+1-s}(\tau)||G_{2,n}(t,\tau)|\right)$$

$$= S_{1,n} + S_{2,n}.$$

As in the proof of Theorem 3.1, we bound each term separately. Using a similar set of bounds to those used in the proof of Theorem 3.1 we have

$$S_{1,n} \leq \sum_{r \in \mathbb{Z}} |K_{f_\theta^{-1}}(r)| \sum_{\ell} |\psi_\ell| \sum_{s=1}^{n} \sum_{j=0}^{\infty} |\phi_{s+j}|$$

$$\leq \sum_{r \in \mathbb{Z}} |K_{f_\theta^{-1}}(r)| \sum_{\ell} |\psi_\ell| \sum_{s=1}^{\infty} |s\phi_s| \leq \frac{1}{\sigma_{f_\theta}^2} \|\psi_{f_\theta}\|_0 \|\phi_{f_\theta}\|_0^2 \|\phi_{f_\theta}\|_1,$$

where the bound $\sum_{r \in \mathbb{Z}} |K_{f_\theta^{-1}}(r)| \leq \sigma_{f_\theta}^{-2} \|\phi_{f_\theta}\|_0^2$ follows from (B.5). Using a similar method we obtain the bound $S_{2,n} \leq \sigma_{f_\theta}^{-2} \|\psi_{f_\theta}\|_0 \|\phi_{f_\theta}\|_0^2 \|\phi_{f_\theta}\|_1$. Altogether the bounds for $S_{1,n}$ and $S_{2,n}$ give

$$\left\|F_n^* \Delta_n(f_\theta^{-1}) D_{\infty,n}(f_\theta)\right\|_1 \leq \frac{2}{\sigma_{f_\theta}^2} \|\psi_{f_\theta}\|_0 \|\phi_{f_\theta}\|_0^2 \|\phi_{f_\theta}\|_1,$$

this proves (3.8).

The proof of (3.9) uses the triangle inequality

$$\left\|\Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1})\right\|_1 = \left\|F_n^* \Delta_n(f_\theta^{-1}) D_n(f_\theta)\right\|_1$$

$$\leq \left\|F_n^* \Delta_n(f_\theta^{-1}) \left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right)\right\|_1$$

$$+ \left\|F_n^* \Delta_n(f_\theta^{-1}) D_{\infty,n}(f_\theta)\right\|_1.$$

Substituting the bound Theorem 3.1 (equation (3.5)) and (3.8) into the above gives (3.9).

The proof of (3.10) uses the bound in (3.9) together with similar arguments to those in the proof of Theorem 3.1, we omit the details. □

B.2 Proof of results in Section 3.2

To prove Theorem 3.3 we recall some notation. Let $P_{[1,\infty)}X$ and $P_{(-\infty,n]}X$ denote the projection of X onto $\text{sp}(X_t; t \geq 1)$ or $\text{sp}(X_t; t \leq n)$. Then using the definition of $\{\phi_j(\tau)\}$ we have

$$P_{[1,\infty)}X_\tau = \sum_{j=1}^{\infty} \phi_j(\tau)X_j \text{ for } \tau \leq 1 \quad \text{and} \quad P_{(-\infty,n]}X_\tau = \sum_{j=1}^{\infty} h_j(\tau)X_j \text{ for } \tau > n.$$

By stationarity we have $h_j(\tau) = \phi_{n+1-j}(n+1-\tau)$. We use this notation below.

PROOF of Theorem 3.3 We first focus on the prediction coefficient associated with X_j where $1 \le j \le n$ and show that

$$\sum_{\tau=-\infty}^{0} \phi_{j,n}(\tau)e^{i\tau\omega} = \phi(\omega)^{-1} \sum_{s=1}^{\infty} \zeta_{j,n}^{(s)}(\omega;f),$$

where $\zeta_{j,n}^{(s)}(\omega;f)$ is defined in (3.11). This will lead to the series expansion of $\widehat{J}_n(\omega;f)$.

Let us assume $\tau \le 0$. By using Inoue and Kasahara (2006), Theorem 2.5, we obtain an expression for the difference between the coefficients of X_j for the finite predictor (based on the space $\mathrm{sp}(X_1, \ldots, X_n)$) and the infinite predictor (based on the space $\mathrm{sp}(X_1, X_2, \ldots)$):

$$\phi_{j,n}(\tau) - \phi_j(\tau)$$
$$= \sum_{u_1=n+1}^{\infty} \phi_{u_1}(\tau)h_j(u_1) + \sum_{u_1=n+1}^{\infty}\sum_{v_1=-\infty}^{0} \phi_{u_1}(\tau)h_{v_1}(u_1)\phi_j(v_1) +$$
$$\sum_{u_1=n+1}^{\infty}\sum_{v_1=-\infty}^{0}\sum_{u_2=n+1}^{\infty} \phi_{u_1}(\tau)h_{v_1}(u_1)\phi_{u_2}(v_1)h_j(u_2) + \ldots$$

Replacing $h_j(\tau) = \phi_{n+1-j}(n+1-\tau)$ we have

$$\phi_{j,n}(\tau) - \phi_j(\tau)$$
$$= \sum_{u_1=n+1}^{\infty} \phi_{u_1}(\tau)\phi_{n+1-j}(n+1-u_1) + \sum_{u_1=n+1}^{\infty}\sum_{v_1=-\infty}^{0} \phi_{u_1}(\tau)\phi_{n+1-v_1}(n+1-u_1)\phi_j(v_1) +$$
$$\sum_{u_1=n+1}^{\infty}\sum_{v_1=-\infty}^{0}\sum_{u_2=n+1}^{\infty} \phi_{u_1}(\tau)\phi_{n+1-v_1}(n+1-u_1)\phi_{u_2}(v_1)\phi_{n+1-j}(n+1-u_2) + \ldots$$

Changing variables with $u_1 \to u_1 - n - 1$, $v_1 \to -v_1, \ldots$ gives

$$\phi_{j,n}(\tau) - \phi_j(\tau)$$
$$= \sum_{u_1=0}^{\infty} \phi_{n+1+u_1}(\tau)\phi_{n+1-j}(-u_1) + \sum_{u_1=0}^{\infty}\sum_{v_1=0}^{\infty} \phi_{n+1+u_1}(\tau)\phi_{n+1+v_1}(-u_1)\phi_j(-v_1) +$$
$$\sum_{u_1=0}^{\infty}\sum_{v_1=0}^{\infty}\sum_{u_2=0}^{\infty} \phi_{n+1+u_1}(\tau)\phi_{n+1+v_1}(-u_1)\phi_{n+1+u_2}(-v_1)\phi_{n+1-j}(-u_2) + \ldots$$
$$= \sum_{s=2}^{\infty} \sum_{u_1,\ldots,u_s=0}^{\infty} \phi_{n+1+u_1}(\tau) \left[\prod_{a=1}^{s-1} \phi_{n+1+u_{a+1}}(-u_a)\right] [\phi_{n+1-j}(-u_s)\delta_{s \bmod 2=1} + \phi_j(-u_s)\delta_{s \bmod 2=0}].$$

Next our focus will be on the term inside the sum $\sum_{s=2}^{\infty}$, where we define

$$\phi_{j,n}^{(s)}(\tau) = \sum_{u_1,\dots,u_s=0}^{\infty} \phi_{n+1+u_1}(\tau) \left[\prod_{a=1}^{s-1} \phi_{n+1+u_{a+1}}(-u_a) \right] \left[\phi_{n+1-j}(-u_s)\delta_{s \bmod 2=1} + \phi_j(-u_s)\delta_{s \bmod 2=0} \right].$$

Using this notation we have

$$\phi_{j,n}(\tau) - \phi_j(\tau) = \sum_{s=2}^{\infty} \phi_{j,n}^{(s)}(\tau).$$

We will rewrite $\phi_{j,n}^{(s)}(\tau)$ as a convolution. To do so, we first note that from Lemma A.1

$$\phi_j(\tau) = \sum_{s=0}^{\infty} \phi_{j+s}\psi_{|\tau|-s} = \sum_{\ell=0}^{|\tau|} \psi_\ell \phi_{j+|\tau|-\ell} \quad \text{for } \tau \le 0.$$

This can be written as an integral

$$\phi_j(\tau) = \frac{1}{2\pi} \int_0^{2\pi} \phi(\lambda)^{-1}\phi_j^{\infty}(\lambda)e^{-i\tau\lambda}d\lambda, \quad \tau \le 0$$

where $\phi(\lambda)^{-1} = \sum_{s=0}^{\infty} \psi_s e^{-is\lambda}$ and $\phi_j^{\infty}(\lambda) = \sum_{s=0}^{\infty} \phi_{j+s}e^{-is\lambda}$. Using this representation we observe that for $\tau \ge 1$

$$\frac{1}{2\pi} \int_0^{2\pi} \phi(\lambda)^{-1}\phi_j^{\infty}(\lambda)e^{-i\tau\lambda}d\lambda = 0.$$

Based on the above we define the "extended" coefficient

$$\widetilde{\phi}_j(\tau) = \frac{1}{2\pi} \int_0^{2\pi} \phi(\lambda)^{-1}\phi_j^{\infty}(\lambda)e^{-i\tau\lambda}d\lambda = \begin{cases} \phi_j(\tau) & \tau \le 0 \\ 0 & \tau \ge 1. \end{cases} \tag{B.6}$$

$\phi_j(\tau)$ can be treated as the first term in the expansion $\phi_{j,n}(\tau) = \phi_j(\tau) + \sum_{s=2}^{\infty} \phi_{j,n}^{(s)}(\tau)$. In the same way we have extended the definition of $\phi_{j,n}(\tau)$ over $\tau \in \mathbb{Z}$ we do the same for the higher order terms and define

$$\widetilde{\phi}_{j,n}^{(s)}(\tau) = \sum_{u_1,\dots,u_s=0}^{\infty} \widetilde{\phi}_{n+1+u_1}(\tau) \left[\prod_{a=1}^{s-1} \phi_{n+1+u_{a+1}}(-u_a) \right] \left[\phi_{n+1-j}(-u_s)\delta_{s \bmod 2=1} + \phi_j(-u_s)\delta_{s \bmod 2=0} \right].$$

Observe from the above definition that $\widetilde{\phi}_{j,n}^{(s)}(\tau) = \phi_{j,n}^{(s)}(\tau)$ for $\tau \le 0$ and $\widetilde{\phi}_{j,n}^{(s)}(\tau) = 0$ for $\tau \ge 1$.

Substituting (B.6) into $\widetilde{\phi}_{j,n}^{(s)}(\tau)$ for $\tau \in \mathbb{Z}$ gives rise to a convolution for $\widetilde{\phi}_{j,n}^{(s)}(\tau)$

$$
\begin{aligned}
\widetilde{\phi}_{j,n}^{(s)}(\tau) &= \frac{1}{(2\pi)^s} \int_{[0,2\pi]^s} e^{-i\tau\lambda_1} \left[\prod_{a=1}^{s} \phi(\lambda_a)^{-1}\right] \left[\prod_{a=1}^{s-1} \sum_{u_a=0}^{\infty} \phi_{n+1+u_a}^{\infty}(\lambda_a) e^{iu_a\lambda_{a+1}}\right] \\
&\qquad \left[\phi_{n+1-j}^{\infty}(\lambda_s)\delta_{s \bmod 2=0} + \phi_j^{\infty}(\lambda_s)\delta_{s \bmod 2=1}\right] d\underline{\lambda}_s \\
&= \frac{1}{(2\pi)^s} \int_{[0,2\pi]^s} e^{-i\tau\lambda_1} \left[\prod_{a=1}^{s} \phi(\lambda_a)^{-1}\right] \prod_{a=1}^{s-1} \Phi_n(\lambda_a, \lambda_{a+1}) \\
&\qquad \times \left[\phi_{n+1-j}^{\infty}(\lambda_s)\delta_{s \bmod 2=0} + \phi_j^{\infty}(\lambda_s)\delta_{s \bmod 2=1}\right] d\underline{\lambda}_s. \quad\quad (B.7)
\end{aligned}
$$

Evaluating the Fourier transform of $\{\widetilde{\phi}_{j,n}^{(s)}(\tau)\}_{\tau \in \mathbb{Z}}$ gives

$$
\begin{aligned}
\sum_{\tau \leq 0} \phi_{j,n}^{(s)}(\tau)e^{i\tau\omega} &= \sum_{\tau \in \mathbb{Z}} \widetilde{\phi}_{j,n}^{(s)}(\tau)e^{i\tau\omega} \\
&= \sum_{\tau \in \mathbb{Z}} \frac{1}{(2\pi)^s} \int_{[0,2\pi]^s} e^{-i\tau(\lambda_1-\omega)} \left[\prod_{a=1}^{s} \phi(\lambda_a)^{-1}\right] \prod_{a=1}^{s-1} \Phi_n(\lambda_a, \lambda_{a+1}) \\
&\qquad \times \left[\phi_{n+1-j}^{\infty}(\lambda_s)\delta_{s \bmod 2=0} + \phi_j^{\infty}(\lambda_s)\delta_{s \bmod 2=1}\right] d\underline{\lambda}_s \\
&= \frac{1}{(2\pi)^s} \int_{[0,2\pi]^s} \left(\sum_{\tau \in \mathbb{Z}} e^{-i\tau(\lambda_1-\omega)}\right) \left[\prod_{a=1}^{s} \phi(\lambda_a)^{-1}\right] \prod_{a=1}^{s-1} \Phi_n(\lambda_a, \lambda_{a+1}) \\
&\qquad \times \left[\phi_{n+1-j}^{\infty}(\lambda_s)\delta_{s \bmod 2=0} + \phi_j^{\infty}(\lambda_s)\delta_{s \bmod 2=1}\right] d\underline{\lambda}_s \\
&= \frac{1}{(2\pi)^{s-1}} \int_{[0,2\pi]^s} \left[\prod_{a=1}^{s} \phi(\lambda_a)^{-1}\right] \prod_{a=1}^{s-1} \Phi_n(\lambda_a, \lambda_{a+1}) \\
&\qquad \times \left[\phi_{n+1-j}^{\infty}(\lambda_s)\delta_{s \bmod 2=0} + \phi_j^{\infty}(\lambda_s)\delta_{s \bmod 2=1}\right] \delta_{\lambda_1=\omega} d\underline{\lambda}_s \\
&= \phi(\omega)^{-1} \frac{1}{(2\pi)^{s-1}} \int_{[0,2\pi]^s} \prod_{a=1}^{s-1} \phi(\lambda_{a+1})^{-1} \Phi_n(\lambda_a, \lambda_{a+1}) \\
&\qquad \times \left[\phi_{n+1-j}^{\infty}(\lambda_s)\delta_{s \bmod 2=0} + \phi_j^{\infty}(\lambda_s)\delta_{s \bmod 2=1}\right] \delta_{\lambda_1=\omega} d\underline{\lambda}_s \\
&= \phi(\omega)^{-1} \zeta_{j,n}^{(s)}(\omega; f),
\end{aligned}
$$

where

$$
\begin{aligned}
\zeta_{j,n}^{(s)}(\omega; f) &= \frac{1}{(2\pi)^{s-1}} \int_{[0,2\pi]^s} \prod_{a=1}^{s-1} \phi(\lambda_{a+1})^{-1} \Phi_n(\lambda_a, \lambda_{a+1}) \\
&\qquad \times \left[\phi_{n+1-j}^{\infty}(\lambda_s)\delta_{s \bmod 2=0} + \phi_j^{\infty}(\lambda_s)\delta_{s \bmod 2=1}\right] \delta_{\lambda_1=\omega} d\underline{\lambda}_s.
\end{aligned}
$$

The above holds for $s \geq 2$. But a similar representation also holds for the first term, $\phi_j(\tau)$, in

the expansion of $\phi_{j,n}(\tau)$. Using the same argument as above we have

$$\sum_{\tau \leq 0} \phi_j(\tau) e^{i\tau\omega} = \phi(\lambda_1)^{-1} \zeta_{j,n}^{(1)}(\omega; f) = \int_0^{2\pi} \phi(\lambda_1)^{-1} \phi_j^\infty(\lambda_1) \delta_{\lambda_1 = \omega} d\lambda_1.$$

Altogether this gives an expression for the Fourier transform of the predictor coefficients corresponding to X_j ($1 \leq j \leq n$) at all lags $\tau \leq 0$:

$$\sum_{\tau \leq 0} \phi_{j,n}(\tau) e^{i\tau\omega} = \sum_{\tau \leq 0} \phi_j(\tau) e^{i\tau\omega} + \sum_{s=2}^{\infty} \sum_{\tau \leq 0} \phi_{j,n}^{(s)}(\tau) e^{i\tau\omega} = \phi(\omega)^{-1} \sum_{s=1}^{\infty} \zeta_{j,n}^{(s)}(\omega; f). \tag{B.8}$$

Using a similar set of arguments we have

$$\sum_{\tau \geq n+1} \phi_{j,n}(n+1-\tau) e^{i\tau\omega} = \overline{\phi(\omega)}^{-1} e^{i(n+1)\omega} \sum_{s=1}^{\infty} \overline{\zeta_{n+1-j,n}^{(s)}(\omega; f)}. \tag{B.9}$$

Therefore by using the above and setting $j = t$ we have the series expansion

$$\begin{aligned}
\widehat{J}_n(\omega; f) &= \frac{1}{n} \sum_{t=1}^{n} X_t \left(\sum_{\tau \leq 0} \phi_{t,n}(\tau) e^{i\tau\omega} + \sum_{\tau \geq n+1} \phi_{n+1-t,n}(n+1-\tau) e^{i\tau\omega} \right) \\
&= \sum_{s=1}^{\infty} \widehat{J}_n^{(s)}(\omega; f),
\end{aligned}$$

where $\widehat{J}_n^{(s)}(\omega; f)$ is defined in Theorem 3.3. Thus proving (3.12). To prove (3.13) we use that

$$(D_n(f))_{k,t} = n^{-1/2} \sum_{\tau \leq 0} \phi_{t,n}(\tau) e^{i\tau\omega_{k,n}} + n^{-1/2} e^{i\omega_{k,n}} \sum_{\tau \leq 0} \phi_{n+1-t,n}(\tau) e^{-i\tau\omega_{k,n}}.$$

This together with (B.8) and (B.9) proves the (3.13).

Finally, we will obtain a bound for $\zeta_{t,n}^{(s)}(\omega; f)$, which results in bound for $\widehat{J}_n^{(s)}(\omega; f)$ (which we use to prove (3.14)). By using (3.11) we have

$$\begin{aligned}
|\zeta_{t,n}^{(s)}(\omega; f)| &\leq \left(\sup_\lambda |\phi(\lambda; f)|^{-1} \sup_{\lambda_1, \lambda_2} |\Phi_n(\lambda_1, \lambda_2)| \right)^{s-1} \times \\
&\quad \left(\|\phi_t^\infty(\lambda_s; f)\|_0 \delta_{s \equiv 1(\bmod 2)} + \|\phi_{n+1-t}^\infty(\lambda_s; f)\|_0 \delta_{s \equiv 0(\bmod 2)} \right). \tag{B.10}
\end{aligned}$$

To bound the terms above we note that

$$\sup_{\lambda_1, \lambda_2} |\Phi_n(\lambda_1, \lambda_2)| \leq \sum_{j=1}^{\infty} |j\phi_{n+j}(f)| = O(n^{-K+1}),$$

where the above follows from Assumption 3.1. Further

$$|\phi(\lambda; f)|^{-1} \leq \sum_{j=0}^{\infty} |\psi_j(f)|,$$

where $\{\psi_j(j)\}$ are the MA(∞) coefficients corresponding to the spectral density f. Substituting these bounds into (B.10) gives

$$|\zeta_{t,n}^{(s)}(\omega; f)| \leq \left(\sum_{j=0}^{\infty} |\psi_j(f)| \sum_{j=1}^{\infty} |j\phi_{n+j}(f)| \right)^{s-1} \left(\sum_{j=t}^{\infty} |\phi_j(f)| + \sum_{j=n+1-t}^{\infty} |\phi_j(f)| \right).$$

Substituting the above into $\widehat{J}_n^{(s)}(\omega; f)$ gives the bound

$$|\widehat{J}_n^{(s)}(\omega; f)| \leq \sum_{j=0}^{\infty} |\psi_j(f)| \left(\sum_{j=0}^{\infty} |\psi_j(f)| \sum_{j=1}^{\infty} |j\phi_{n+j}(f)| \right)^{s-1}$$
$$\times \frac{2}{\sqrt{n}} \sum_{t=1}^{n} |X_t| \left(\sum_{j=t}^{\infty} |\phi_j(f)| + \sum_{j=n+1-t}^{\infty} |\phi_j(f)| \right).$$

Therefore taking expectation of $|\widehat{J}_n^{(s)}(\omega; f)|$ gives the bound

$$\mathbb{E}|\widehat{J}_n^{(s)}(\omega; f)| \leq \mathbb{E}|X_0| \sum_{j=0}^{\infty} |\psi_j(f)| \left(\sum_{j=0}^{\infty} |\psi_j(f)| \sum_{j=1}^{\infty} |j\phi_{n+j}(f)| \right)^{s-1} \frac{4}{\sqrt{n}} \sum_{j=1}^{\infty} |j\phi_j(f)|.$$

For large enough n, $\sum_{j=0}^{\infty} |\psi_j(f)| \cdot \sum_{j=1}^{\infty} |j\phi_{n+j}(f)| \leq Cn^{-K+1} < 1$. Therefore

$$\sum_{s=m+1}^{\infty} \mathbb{E}|\widehat{J}_n^{(s)}(\omega; f)| \leq n^{-1/2} \sum_{s=m+1}^{\infty} \left(\frac{C}{n^{K-1}} \right)^{s-1} = O(n^{-m(K-1)-1/2})$$

thus

$$\sum_{s=m+1}^{\infty} \widehat{J}_n^{(s)}(\omega; f) = O_p(n^{-m(K-1)-1/2}).$$

This proves the approximation in (3.14). Thus we have proved the result. □

Proof of equation (3.15) Our aim is to show for $s \geq 3$,

$$\zeta_{t,n}^{(s)}(\omega; f) = \frac{1}{(2\pi)^2} \int_{[0,2\pi]^2} \phi(y_1; f)^{-1} \phi(y_2; f)^{-1} \Phi_n(\omega, y_1) \Phi_n(y_1, y_2) \zeta_{t,n}^{(s-2)}(y_2; f) dy_1 dy_2. \quad \text{(B.11)}$$

61

We recall the definition of $\zeta_{t,n}^{(s)}(\omega; f)$ in equation (3.11)

$$\zeta_{t,n}^{(s)}(\omega; f) = \frac{1}{(2\pi)^{s-1}} \int_{[0,2\pi]^{s-1}} \Phi_n(\omega, \lambda_2)\phi(\lambda_2; f)^{-1} \left(\prod_{a=2}^{s-1} \phi(\lambda_{a+1}; f)^{-1}\Phi_n(\lambda_a, \lambda_{a+1}) \right) \times$$
$$\left(\phi_t^\infty(\lambda_s; f)\delta_{s\equiv 1(\bmod 2)} + \phi_{n+1-t}^\infty(\lambda_s; f)\delta_{s\equiv 0(\bmod 2)} \right) d\lambda_2 d\lambda_3 \ldots \lambda_s.$$

We change notation and set $u_s = \omega, u_{s-1} = \lambda_2 \ldots, u_2 = \lambda_{s-1}$ This gives

$$\zeta_{t,n}^{(s)}(u_s; f) = \frac{1}{(2\pi)^{s-1}} \int_{[0,2\pi]^{s-1}} \Phi_n(u_s, u_{s-1})\phi(u_{s-1}; f)^{-1} \left(\prod_{a=s-1}^{2} \phi(u_{s-a-1}; f)^{-1}\Phi_n(u_{s-a}, u_{s-a-1}) \right)$$
$$\times \left(\phi_t^\infty(u_1; f)\delta_{s\equiv 1(\bmod 2)} + \phi_{n+1-t}^\infty(u_1; f)\delta_{s\equiv 0(\bmod 2)} \right) du_1 du_2 \ldots du_{s-1}.$$

Again we write the above as (by rearranging the product term)

$$\zeta_{t,n}^{(s)}(u_s; f) = \frac{1}{(2\pi)^{s-1}} \int_{[0,2\pi]^{s-1}} \Phi_n(u_s, u_{s-1})\phi(u_{s-1}; f)^{-1} \left(\prod_{a=2}^{s-1} \phi(u_a; f)^{-1}\Phi_n(u_{a+1}, u_a) \right) \times$$
$$\left(\phi_t^\infty(u_1; f)\delta_{s\equiv 1(\bmod 2)} + \phi_{n+1-t}^\infty(u_1; f)\delta_{s\equiv 0(\bmod 2)} \right) du_1 du_2 \ldots du_{s-2} du_{s-1}$$
$$= \frac{1}{(2\pi)} \int_{[0,2\pi]} \Phi_n(u_s, u_{s-1})\phi(u_{s-1}; f)^{-1} \frac{1}{(2\pi)^{s-2}} \left\{ \int_{[0,2\pi]^{s-2}} \left(\prod_{a=2}^{s-1} \phi(u_a; f)^{-1}\Phi_n(u_{a+1}, u_a) \right) \right.$$
$$\times \left. \left(\phi_t^\infty(u_1; f)\delta_{s\equiv 1(\bmod 2)} + \phi_{n+1-t}^\infty(u_1; f)\delta_{s\equiv 0(\bmod 2)} \right) du_1 du_2 \ldots du_{s-2} \right\} du_{s-1}.$$

The term inside the integral is analogus to $\zeta_{t,n}^{(s-1)}(u_{s-1}; f)$ (though it is not this). To obtain the exactly expression we apply the same procedure to that described above to the inner integral of the above. This proves (B.11). $\qquad\square$

It is worth mentioning that analogous to the recursion for $\zeta_{t,n}^{(s)}(\omega; f)$ a recursion can also be obtained for

$$\widehat{J}_n^{(s)}(\omega; f) = \frac{n^{-1/2}}{\phi(\omega; f)} \sum_{t=1}^n X_t \zeta_{t,n}^{(s)}(\omega; f) + e^{i(n+1)\omega} \frac{n^{-1/2}}{\overline{\phi(\omega; f)}} \sum_{t=1}^n X_{n+1-t}\overline{\zeta_{t,n}^{(s)}(\omega; f)}$$
$$= \widehat{J}_{L,n}^{(s)}(\omega; f) + e^{i(n+1)\omega}\overline{\widehat{J}_{R,n}^{(s)}(\omega; f)},$$

where

$$\widehat{J}_{L,n}^{(s)}(\omega; f) = \frac{n^{-1/2}}{\phi(\omega; f)} \sum_{t=1}^{n} X_t \zeta_{t,n}^{(s)}(\omega; f)$$

$$\text{and} \quad \widehat{J}_{R,n}^{(s)}(\omega; f) = \frac{n^{-1/2}}{\phi(\omega; f)} \sum_{t=1}^{n} X_{n+1-t} \zeta_{t,n}^{(s)}(\omega; f).$$

By using the recursion for $\zeta_{t,n}^{(s)}(\omega; f)$ we observe that for $s \geq 3$ we can write $\widehat{J}_{L,n}^{(s)}(\omega; f)$ and $\widehat{J}_{R,n}^{(s)}(\omega; f)$ as

$$\widehat{J}_{L,n}^{(s)}(\omega; f) = \frac{1}{(2\pi)^2} \int_{[0,2\pi]^2} \frac{\Phi_n(\omega, y_1)}{\phi(\omega; f)} \frac{\Phi_n(y_1, y_2)}{\phi(y_1; f)} \widehat{J}_{L,n}^{(s-2)}(y_2; f) dy_1 dy_2$$

and

$$\widehat{J}_{R,n}^{(s)}(\omega; f) = \frac{1}{(2\pi)^2} \int_{[0,2\pi]^2} \frac{\Phi_n(\omega, y_1)}{\phi(\omega; f)} \frac{\Phi_n(y_1, y_2)}{\phi(y_1; f)} \widehat{J}_{R,n}^{(s-2)}(y_2; f) dy_1 dy_2.$$

B.3 Proof of Lemma 4.1

We now prove Lemma 4.1. The proof is similar to the proof of Theorem 3.1, 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.

PROOF of Lemma 4.1 We first prove (4.4). By using the triangular inequality we have

$$\left\| F_n^* \Delta_n(f_\theta^{-1}) \left(D_n(f) - D_n(f_p) \right) \right\|_1$$
$$\leq \left\| F_n^* \Delta_n(f_\theta^{-1}) \left(D_n(f) - D_{\infty,n}(f) \right) \right\|_1 + \left\| F_n^* \Delta_n(f_\theta^{-1}) \left(D_{\infty,n}(f) - D_n(f_p) \right) \right\|_1$$
$$\leq \frac{C_{f,0}\rho_{n,K}(f)}{n^{K-1}} A_K(f, f_\theta) + \left\| F_n^* \Delta_n(f_\theta^{-1}) \left(D_{\infty,n}(f) - D_n(f_p) \right) \right\|_1, \quad \text{(B.12)}$$

where the first term of the right hand side of the above follows from (3.5). 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 func-

tion 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 (3.2) we have

$$\left[F_n^* \Delta_n(f_\theta^{-1}) \left(D_{\infty,n}(f) - D_{\infty,n}(f_p) \right) \right]_{t,j} = U_{1,n}^{j,t} + U_{2,n}^{j,t}$$

where

$$U_{1,n}^{j,t} = \frac{1}{n} \sum_{k=1}^{n} \frac{e^{-it\omega_{k,n}}}{f_\theta(\omega_{k,n})} \left(\frac{\phi_j^{\infty}(\omega_{k,n})}{\phi(\omega_{k,n})} - \frac{a_{j,p}^{\infty}(\omega_{k,n})}{a_p(\omega_{k,n})} \right)$$

$$U_{2,n}^{j,t} = \frac{1}{n} \sum_{k=1}^{n} \frac{e^{-i(t-1)\omega_{k,n}}}{f_\theta(\omega_{k,n})} \left(\frac{\overline{\phi_{n+1-j}^{\infty}(\omega_{k,n})}}{\overline{\phi(\omega_{k,n})}} - \frac{\overline{a_{n+1-j,p}^{\infty}(\omega_{k,n})}}{\overline{a_p(\omega_{k,n})}} \right).$$

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

$$U_{1,n,2}^{j,t} = \frac{1}{n} \sum_{k=1}^{n} \frac{e^{-it\omega_{k,n}} a_{j,p}^{\infty}(\omega_{k,n})}{f_\theta(\omega_{k,n})} \left(\phi(\omega_{k,n})^{-1} - a_p(\omega_{k,n})^{-1} \right)$$

$$= \frac{1}{n} \sum_{k=1}^{n} \frac{e^{-it\omega_{k,n}} a_{j,p}^{\infty}(\omega_{k,n})}{f_\theta(\omega_{k,n})} \left(\psi(\omega_{k,n}) - \psi_p(\omega_{k,n}) \right).$$

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

$$U_{1,n,1}^{j,t} = \sum_{s=0}^{\infty} (\phi_{j+s} - a_{j+s}(p)) \frac{1}{n} \sum_{k=1}^{n} \frac{e^{-i(t+s)\omega_{k,n}}}{\phi(\omega_{k,n})f_\theta(\omega_{k,n})}$$

$$= \sum_{s=0}^{\infty} (\phi_{j+s} - a_{j+s}(p)) \sum_{\ell=0}^{\infty} \psi_\ell \frac{1}{n} \sum_{k=1}^{n} f_\theta(\omega_{k,n})^{-1} e^{-i(t+\ell+s)\omega_{k,n}}$$

$$= \sum_{s=0}^{\infty} (\phi_{j+s} - a_{j+s}(p)) \sum_{\ell=0}^{\infty} \psi_\ell \sum_{r \in \mathbb{Z}} K_{f_\theta^{-1}}(t + \ell + s + rn),$$

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

$$
\begin{aligned}
\sum_{j,t=1}^{n} |U_{1,n,1}^{j,t}| &\leq \sum_{j,t=1}^{n} \sum_{s=0}^{\infty} |\phi_{j+s} - a_{j+s}(p)| \sum_{\ell=0}^{\infty} |\psi_\ell| \sum_{r\in\mathbb{Z}} |K_{f_\theta^{-1}}(t + \ell + s + rn)| \\
&= \sum_{j=1}^{n} \sum_{s=0}^{\infty} |\phi_{j+s} - a_{j+s}(p)| \sum_{\ell=0}^{\infty} |\psi_\ell| \sum_{t=1}^{n} \sum_{r\in\mathbb{Z}} |K_{f_\theta^{-1}}(t + \ell + s + rn)| \\
&\leq \left(\sum_{j=1}^{n} \sum_{s=0}^{\infty} |\phi_{j+s} - a_{j+s}(p)| \right) \|\psi_f\|_0 \sum_{\tau\in\mathbb{Z}} |K_{f_\theta^{-1}}(\tau)| \\
&\leq \left(\sum_{s=1}^{\infty} s|\phi_s - a_s(p)| \right) \|\psi_f\|_0 \sum_{\tau\in\mathbb{Z}} |K_{f_\theta^{-1}}(\tau)|.
\end{aligned}
$$

By using (B.5) we have $\sum_{\tau\in\mathbb{Z}} |K_{f_\theta^{-1}}(\tau)| \leq \sigma_{f_\theta}^{-2} \|\phi_{f_\theta}\|_0^2$. Further, by using the regular Baxter inequality we have

$$
\sum_{s=1}^{\infty} s|\phi_s - a_s(p)| \leq (1 + C_{f,1}) \sum_{s=p+1}^{\infty} s|\phi_s| \leq (1 + C_{f,1}) p^{-K+1} \rho_{p,K}(f) \|\phi_f\|_K.
$$

Substituting these two bounds into $\sum_{j,t=1}^{n} |U_{1,n,1}^{j,t}|$ yields

$$
\sum_{j,t=1}^{n} |U_{1,n,1}^{j,t}| \leq \frac{(1 + C_{f,1})}{\sigma_{f_\theta}^2 p^{K-1}} \rho_{p,K}(f) \|\phi_f\|_K \|\psi_f\|_0 \|\phi_{f_\theta}\|_0^2.
$$

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

$$
\begin{aligned}
\sum_{j,t=1}^{n} |U_{1,n,2}^{j,t}| &\leq \sum_{j,t=1}^{n} \sum_{s=0}^{\infty} |\psi_s - \psi_{s,p}| \sum_{\ell=0}^{\infty} |a_{j+\ell}(p)| \sum_{r\in\mathbb{Z}} |K_{f_\theta^{-1}}(t+s+\ell+rn)| \\
&= \sum_{s=0}^{\infty} |\psi_s - \psi_{s,p}| \sum_{j=1}^{n} \sum_{\ell=0}^{\infty} |a_{j+\ell}(p)| \sum_{r\in\mathbb{Z}} |K_{f_\theta^{-1}}(r)| \quad \text{(apply the bound (B.5))} \\
&\leq \sigma_{f_\theta}^{-2} \|\phi_{f_\theta}\|_0^2 \left(\sum_{s=0}^{\infty} |\psi_s - \psi_{s,p}| \right) \sum_{u=0}^{\infty} |u a_u(p)| \\
&\leq \sigma_{f_\theta}^{-2} \|\phi_{f_\theta}\|_0^2 \|a_p\|_1 \sum_{s=0}^{\infty} |\psi_s - \psi_{s,p}|.
\end{aligned}
$$

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(∞) process). Then by applying Baxter's inequality, it is straightforward to show that

$$
\|a_p\|_1 \leq \|\phi_p\|_1 + \|a_p - \phi_p\|_1 \leq (C_{f,1}+1)\|\phi_f\|_1. \tag{B.13}
$$

To bound $\sum_{s=0}^{\infty} |\psi_s - \psi_{s,p}|$ we use the inequality in Kreiss et al. (2011), page 2126

$$
\sum_{s=0}^{\infty} |\psi_s - \psi_{s,p}| \leq \frac{\|\psi_f\|_0^2 \cdot \sum_{j=1}^{\infty} |\phi_j - a_j(p)|}{1 - \|\psi_f\| \cdot \|a_p - \phi\|_0}.
$$

Applying Baxter's inequality to the numerator of the above gives

$$
\sum_{s=0}^{\infty} |\psi_s - \psi_{s,p}| \leq \frac{\|\psi_f\|_0^2 (C_{f,0}+1)\rho_{p,K}(f)\|\phi_f\|_K}{p^K(1 - \|\psi_f\|_0 \cdot \|a_p - \phi\|_0)} \tag{B.14}
$$

Substituting the bound in (B.13) and (B.14) into $\sum_{j,t=1}^{n} |U_{1,n,2}^{j,t}|$ gives

$$
\sum_{j,t=1}^{n} |U_{1,n,2}^{j,t}| \leq \frac{(C_{f,1}+1)^2}{\sigma_{f_\theta}^2 p^K} \cdot \frac{\|\psi_f\|_0^2 \|\phi_f\|_1 \|\phi_f\|_K \|\phi_{f_\theta}\|_0^2 \rho_{p,K}(f)}{1 - \|\psi_f\|_0 \|a_p - \phi\|_0}
$$

Altogether, for sufficiently large p, where $\|\psi_f\|_0 \cdot \|a_p - \phi\|_0 \leq 1/2$ we have

$$
\begin{aligned}
\sum_{t,j=1}^{n} |U_{1,n}^{j,t}| &\leq \frac{(1+C_{f,1})}{\sigma_{f_\theta}^2 p^{K-1}} \rho_{p,K}(f)\|\phi_f\|_K \|\psi_f\|_0 \|\phi_{f_\theta}\|_0^2 \\
&\quad + \frac{2(C_{f,1}+1)^2}{\sigma_{f_\theta}^2 p^K} \|\psi_f\|_0^2 \|\phi_f\|_1 \|\phi_f\|_K \|\phi_{f_\theta}\|_0^2 \rho_{p,K}(f) \\
&\leq \frac{(C_{f,1}+1)}{\sigma_{f_\theta}^2 p^{K-1}} \rho_{p,K}(f)\|\phi_f\|_K \|\phi_{f_\theta}\|_0^2 \|\psi_f\|_0 \left(1 + \frac{2(1+C_{f,1})}{p}\|\psi_f\|_0 \|\phi_f\|_1\right)
\end{aligned}
$$

66

The same bound holds for $\sum_{t,j=1}^{n} |U_{2,n}^{j,t}|$, thus using (B.12) and $\rho_{n,K}(f) \leq \rho_{p,K}(f)$ gives

$$\left\| F_n^* \Delta_n(f_\theta^{-1}) \left(D_{\infty,n}(f) - D_n(f_p) \right) \right\|_1$$
$$\leq \rho_{p,K}(f) A_K(f, f_\theta) \left(\frac{(C_{f,1} + 1)}{p^{K-1}} + \frac{2(C_{f,1} + 1)^2}{p^K} \|\psi_f\|_0 \|\phi_f\|_1 \right).$$

Substituting the above into (B.12) gives (4.4).

The proof of (4.5) is similar to the proof of Theorem 3.1, we omit the details. $\qquad \square$

C An extension of Baxter's inequalities

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 $\mathrm{AR}(\infty)$ and $\mathrm{MA}(\infty)$ coefficients respectively and $\psi(\omega) = \phi(\omega)^{-1}$. To simplify notation we have ignored the variance of the innovation.

C.1 Proof of the extended Baxter inequality

Let $\{\phi_{s,p}(\tau)\}_{s=1}^{p}$ denote the the coefficients of the best linear predictor of $X_{t+\tau}$ (for $\tau \geq 0$) given $\{X_s\}_{t-p}^{t-1}$

$$\mathbb{E}\left[\left(X_{t+\tau} - \sum_{s=1}^{p} \phi_{s,p}(\tau) X_{t-s} \right) X_{t-k} \right] = 0 \text{ for } k = 1, \dots, p. \qquad (\text{C.1})$$

and $\{\phi_s(\tau)\}$ denote the coefficients of the best linear predictor of $X_{t+\tau}$ given the infinite past $\{X_s\}_{s=-\infty}^{t-1}$

$$\mathbb{E}\left[\left(X_{t+\tau} - \sum_{s=1}^{\infty} \phi_s(\tau) X_{t-s} \right) X_{t-k} \right] = 0 \text{ for } k = 1, 2, \dots \qquad (\text{C.2})$$

Before we begin, we define an appropriate norm on the subspace of $L_2[0, 2\pi]$.

Definition C.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 $n \geq 0$; (2) $v(n + m) \leq v(n)v(m)$ 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 C.1 (Properties of $\|\cdot\|$) *Suppose the sequence $\{v(k)\}_{k \in \mathbb{Z}}$ satisfies 2 conditions in Definition C.1, and define the norm $\|\cdot\|$ with respect to $\{v(k)\}$. Then, beside the triangle inequality, this norm also satisfies $\|1\| = v(0) \leq 1$, $\|f\| = \|\overline{f}\|$, and $\|fg\| \leq \|f\|\|g\|$ (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 kth Fourier coefficient of f and g. Thus*

$$
\begin{aligned}
\|fg\| &\leq \sum_{k \in \mathbb{Z}} v(k) \left| \sum_{r \in \mathbb{Z}} f_r g_{k-r} \right| \\
&\leq \sum_{k \in \mathbb{Z}} v(r) v(k-r) \left| \sum_{r \in \mathbb{Z}} f_r g_{k-r} \right| \leq \sum_{k,r \in \mathbb{Z}} v(r) v(k-r) |f_r||g_{k-r}| = \|f\|\|g\|.
\end{aligned}
$$

Examples of weights include $v(r) = (2^q + |r|^q)$ or $v(r) = (1 + |r|)^q$ for some $q \geq 0$. In these two examples, when $q = K$, under Assumption 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)).

We believe that Lemma B.1 is well known. But as we could not find a prove we give a proof. The proof below follows closely the proof of Baxter (1962, 1963).

PROOF of Lemma B.1 We use the same proof as Baxter, which is based on rewriting the normal equations in (C.1) 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

$$
\frac{1}{2\pi} \int_0^{2\pi} \left(e^{i\tau\omega} - \sum_{s=1}^{\infty} \phi_s(\tau) e^{-is\omega} \right) f(\omega) e^{-ik\omega} d\omega = 0, \text{ for } k \geq 1.
$$

Thus taking differences of the above two equations for $k = 1, \ldots, p$ gives

$$
\begin{aligned}
&\frac{1}{2\pi} \int_0^{2\pi} \left(\sum_{s=1}^{p} [\phi_{s,p}(\tau) - \phi_s(\tau)] e^{-is\omega} \right) f(\omega) e^{-ik\omega} d\omega \\
&= \frac{1}{2\pi} \int_0^{2\pi} \left(\sum_{s=p+1}^{\infty} \phi_s(\tau) e^{-is\omega} \right) f(\omega) e^{-ik\omega} d\omega \quad 1 \leq k \leq p. \quad\quad (C.3)
\end{aligned}
$$

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 $\left(\sum_{s=1}^{p} [\phi_{s,p}(\tau) - \phi_s(\tau)] e^{-is\omega}\right) f(\omega)$ 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

$$h_p(\omega) = \sum_{s=1}^{p} [\phi_{s,p}(\tau) - \phi_s(\tau)] e^{-is\omega} \quad \text{and} \quad g_p(\omega) = \sum_{k=1}^{p} g_{k,p} e^{ik\omega} \tag{C.4}$$

where

$$g_{k,p} = (2\pi)^{-1} \int_0^{2\pi} \left(\sum_{s=p+1}^{\infty} \phi_s(\tau) e^{-is\omega}\right) f(\omega) e^{-ik\omega} d\omega. \tag{C.5}$$

For the general norm $\| \cdot \|$ defined in Definition C.1, will show that for a sufficiently large p, $\|h_p\| \le C_f \|g_p\|$, 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 (C.3) for $1 \le k \le p$, $\widetilde{g}_{k,p} = g_{k,p}$ (where $g_{k,p}$ is defined in (C.5)). Thus

$$h_p(\omega) f(\omega) = G_{-\infty}^0(\omega) + g_p(\omega) + G_{p+1}^{\infty}(\omega) \tag{C.6}$$

where

$$G_{-\infty}^0(\omega) = \sum_{k=-\infty}^{0} \widetilde{g}_{k,p} e^{ik\omega} \quad \text{and} \quad G_{p+1}^{\infty}(\omega) = \sum_{s=p+1}^{\infty} \widetilde{g}_{k,p} e^{ik\omega}.$$

Dividing by $f^{-1} = \phi\overline{\phi}$ and taking the $\| \cdot \|$-norm we have

$$\begin{aligned}
\|h_p\| &\le \left\|f^{-1} G_{-\infty}^0\right\| + \left\|f^{-1} g_p\right\| + \left\|f^{-1} G_{p+1}^{\infty}\right\| \\
&\le \left\|f^{-1} G_{-\infty}^0\right\| + \left\|f^{-1}\right\| \|g_p\| + \left\|f^{-1} G_{p+1}^{\infty}\right\| \\
&\le \left\|\overline{\phi}\right\| \left\|\phi G_{-\infty}^0\right\| + \left\|f^{-1}\right\| \|g_p\| + \|\phi\| \left\|\overline{\phi} G_{p+1}^{\infty}\right\|.
\end{aligned} \tag{C.7}$$

First we obtain bounds for $\left\|\phi G_{-\infty}^0\right\|$ and $\left\|\overline{\phi} G_{p+1}^{\infty}\right\|$ in terms of $\|g_p\|$. We will show that for a sufficiently large p

$$\begin{aligned}
\left\|\phi G_{-\infty}^0\right\| &\le \|\phi\| \|g_p\| + \varepsilon \left\|\overline{\phi} G_{p+1}^{\infty}\right\| \\
\left\|\overline{\phi} G_{p+1}^{\infty}\right\| &\le \left\|\overline{\phi}\right\| \|g_p\| + \varepsilon \left\|\phi G_{-\infty}^0\right\|.
\end{aligned}$$

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 pth 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 (C.6) using that $f = \psi(\omega)\overline{\psi(\omega)}$ we multiply (C.6) by $\psi(\omega)^{-1} = \phi(\omega)$ to give

$$h_p(\omega)\overline{\psi(\omega)} = \phi(\omega)G^0_{-\infty}(\omega) + \phi(\omega)g_p(\omega) + \phi(\omega)G^{\infty}_{p+1}(\omega). \tag{C.8}$$

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

$$\begin{aligned} &-\phi(\omega)G^0_{-\infty}(\omega) \\ &\quad = \left\{-\phi(\omega)G^0_{-\infty}(\omega)\right\}_- = \left\{\phi(\omega)g_p(\omega)\right\}_- + \left\{\phi(\omega)G^{\infty}_{p+1}(\omega)\right\}_-. \end{aligned} \tag{C.9}$$

We further observe that G^{∞}_{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^{∞}_{p+1}. Therefore

$$-\phi(\omega)G^{-1}_{-\infty}(\omega) = \left\{\phi(\omega)g_p(\omega)\right\}_- + \left\{\phi^{\infty}_{p+1}(\omega)G^{\infty}_{p+1}(\omega)\right\}_-,$$

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

$$\begin{aligned} \left\|\phi G^0_{-\infty}\right\| &\leq \|\phi\|\,\|g_p\| + \left\|\phi^{\infty}_{p+1}G^{\infty}_{p+1}\right\| \\ &\leq \|\phi\|\,\|g_p\| + \left\|\phi^{\infty}_{p+1}\right\|\,\left\|\overline{\psi}\right\|\,\left\|\overline{\phi}G^{\infty}_{p+1}\right\| \quad \text{since } \overline{\psi(\omega)}\phi(\psi) = 1. \end{aligned}$$

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 (C.6), $f(\omega) = \psi(\omega)\overline{\psi(\omega)}$, but this time multiplying (C.6) by $\overline{\psi(\omega)}^{-1} = \overline{\phi(\omega)}$, we

have

$$h_p(\omega)\psi(\omega) = \overline{\phi(\omega)}G_{-\infty}^0(\omega) + \overline{\phi(\omega)}g_p(\omega) + \overline{\phi(\omega)}G_{p+1}^\infty(\omega).$$

Rearranging the above gives

$$\overline{\phi(\omega)}G_{p+1}^\infty(\omega) = h_p(\omega)\psi(\omega) - \overline{\phi(\omega)}G_{-\infty}^0(\omega) - \overline{\phi(\omega)}g_p(\omega).$$

We observe that $\overline{\phi(\omega)}G_{p+1}^\infty(\omega)$ 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

$$e^{-ip\omega}\overline{\phi(\omega)}G_{p+1}^\infty(\omega)$$
$$- \left\{e^{-ip\omega}\overline{\phi(\omega)}G_{-\infty}^0(\omega)\right\}_+ - \left\{e^{-ip\omega}\overline{\phi(\omega)}g_p(\omega)\right\}_+, \tag{C.10}$$

By the similar technique from the previous, it is easy to show

$$\left\{e^{-ip\omega}\overline{\phi(\omega)}G_{-\infty}^0(\omega)\right\}_+ = \left\{e^{-ip\omega}\overline{\phi_{p+1}^\infty(\omega)}G_{-\infty}^0(\omega)\right\}_+. \tag{C.11}$$

Multiplying $e^{ip\omega}$ and evaluating the $\|\cdot\|$-norm of the above yields the inequality

$$\begin{aligned}
\left\|\overline{\phi}G_{p+1}^\infty\right\| &\leq \left\|\overline{\phi}g_p\right\| + \left\|\overline{\phi_{p+1}^\infty}G_{-\infty}^0\right\| \\
&\leq \left\|\overline{\phi}\right\|\|g_p\| + \left\|\overline{\phi_{p+1}^\infty}\right\|\|\psi\|\left\|\phi G_{-\infty}^0\right\|.
\end{aligned}$$

We note that $\|\phi_{p+1}^\infty\| = \|\overline{\phi_{p+1}^\infty}\|$. For $\phi \in A_v$ (see Definition C.1 and Remark C.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 $\|\phi_{p+1}^\infty(\omega)\|\|\psi(\omega)\| \leq \varepsilon < 1$, then we have the desired bounds

$$\begin{aligned}
\left\|\phi G_{-\infty}^0\right\| &\leq \|\phi\|\|g_p\| + \varepsilon\left\|\overline{\phi}G_{p+1}^\infty\right\| \\
\left\|\overline{\phi}G_{p+1}^\infty\right\| &\leq \left\|\overline{\phi}\right\|\|g_p\| + \varepsilon\left\|\phi G_{-\infty}^0\right\|.
\end{aligned}$$

The above implies that $\left\|\phi G_{-\infty}^0\right\| + \left\|\overline{\phi}G_{p+1}^\infty\right\| \leq 2(1-\varepsilon)^{-1}\|\phi\|\|g_p\|$. Substituting the above in (C.7), and using that $\|\phi\| \geq 1$ (since $\phi = 1 - \sum_{s=1}^\infty \phi_s e^{-is\omega}$, $\|\phi\| \geq \|1\| = v(0) \geq 1$) we have

$$\begin{aligned}
\|h_p\| &\leq \frac{2\|\phi\|\|g_p\|}{1-\varepsilon} + \|f^{-1}\|\|g_p\| \\
&\leq (1-\varepsilon)^{-1}\left(2\|\phi\| + (1-\varepsilon)\|\phi\|^2\right)\|g_p\| \leq \frac{3-\varepsilon}{1-\varepsilon}\|\phi\|^2\|g_p\|.
\end{aligned}$$

Thus based on the above we have

$$\|h_p\| \leq \frac{3-\varepsilon}{1-\varepsilon}\|\phi\|^2\|g_p\|. \tag{C.12}$$

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 (C.4). 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

$$\|g_p\| \le \|\widetilde{g}_p\| = \left\| \phi_{p+1}^{\infty}(\tau) f \right\| \le \left\| \phi_{p+1}^{\infty}(\tau) \right\| \|f\| \le \left\| \phi_{p+1}^{\infty} \right\| \|\psi\|^2. \tag{C.13}$$

where $\phi_{p+1}^{\infty}(\tau)(\omega) = \sum_{s=p+1}^{\infty} \phi_s(\tau) e^{-is\omega}$. Finally, substituting (C.13) into (C.12), implies that if p is large enough such that $\left\| \phi_{p+1}^{\infty} \right\| \|\psi\| \le \varepsilon < 1$, then

$$\|h_p\| \le \frac{3-\varepsilon}{1-\varepsilon} \|\phi\|^2 \|\psi\|^2 \left\| \phi_{p+1}^{\infty}(\tau) \right\|.$$

Thus, if the weights in the norm are $v(m) = (2^K + m^K)$ (it is well-defined weights, see Remark C.1) we have

$$\sum_{s=1}^{p} (2^K + s^K) \left| \phi_{s,p}(\tau) - \phi_s(\tau) \right|$$

$$\le \frac{3-\varepsilon}{1-\varepsilon} \|\phi\|_K^2 \|\psi\|_K^2 \sum_{s=p+1}^{\infty} (2^K + s^K) \left| \phi_s(\tau) \right|. \tag{C.14}$$

Using Corollary A.1 we have for $\tau \le 0$ $\phi_s(\tau) = \sum_{j=0}^{\infty} \phi_{s+j} \psi_{|\tau|-j}$ (noting that $\psi_j = 0$ for $j < 0$), and the desired result. $\qquad \square$

C.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

$$\underline{\varphi}_n(\tau; f_\theta) = (\phi_{1,n}(\tau; f_\theta), \ldots, \phi_{n,n}(\tau; f_\theta))' \quad \text{(best linear finite future predictor)} \tag{C.15}$$

$$\underline{\phi}_n(\tau; f_\theta) = (\phi_1(\tau; f_\theta), \ldots, \phi_n(\tau; f_\theta))' \quad \text{(truncated best linear infinite future predictor)}.$$

Lemma C.1 *Let θ 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_θ. For*

all $\theta \in \Theta$ and for $0 \le i \le \kappa$ we assume

$$\sum_{j=1}^{\infty} \|j^K \nabla_\theta^i \phi_j(f_\theta)\|_1 < \infty \quad \sum_{j=1}^{\infty} \|j^K \nabla_\theta^i \psi_j(f_\theta)\|_1 < \infty, \tag{C.16}$$

where $K > 1$. Let $\underline{\varphi}_n(\tau; f_\theta)$ and $\underline{\phi}_n(\tau; f_\theta)$, be defined as in (C.15). We assume that $\tau \le 0$. Then for all $0 \le i \le \kappa$, we have

$$\left\| \frac{\partial^i}{\partial \theta_{r_1} \dots \partial \theta_{r_i}} \left[\underline{\varphi}_n(\tau; f_\theta) - \underline{\phi}_n(\tau; f_\theta) \right] \right\|_2 \le f_0 \left(\sum_{\substack{a_1+a_2=i \\ a_2 \ne i}} C_{a_1} \binom{i}{a_1} \left\| \nabla_\theta^{a_2} \left[\underline{\varphi}_n(\tau; f_\theta) - \underline{\phi}_n(\tau; f_\theta) \right] \right\|_2 \right.$$

$$\left. + \sum_{b_1+b_2=i} C_{b_1} \binom{i}{b_1} \sum_{j=n+1}^{\infty} \left\| \nabla_\theta^{b_2} \phi_j(\tau; f_\theta) \right\|_1 \right),$$

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 ath order partial derivative of g with respect to $\theta = (\theta_1, \dots, \theta_d)$ and $\|\nabla_\theta^a g(f_\theta)\|_p$ denotes the ℓ_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

$$\underline{c}_{n,\tau} = (c(\tau-1), c(\tau-2), \dots, c(\tau-n))' \quad \text{(covariances from lag } \tau - 1 \text{ to lag } \tau - n\text{)}.$$

To simplify notation we drop the f_θ notation from the prediction coefficients $\phi_{j,n}(\tau; f_\theta)$ and $\phi_j(\tau; f_\theta)$.

<u>Proof for the case $i = 0$</u> This is the regular Baxter inequality but with the ℓ_2-norm rather than ℓ_1-norm. We recall that for $\tau \le 0$ 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 $1 \le r \le n$ gives the sequence of r normal equation, which can be written in matrix form

$$\Gamma_n(f_\theta) \underline{\varphi}_n(\tau) = \underline{c}_{n,\tau} \quad \text{and} \quad \Gamma_n(f_\theta) \underline{\phi}_n(\tau) + \sum_{j=n+1}^{\infty} \phi_j(\tau) \underline{c}_{n,-j+\tau} = \underline{c}_{n,\tau}.$$

Taking differences of the above gives

$$\Gamma_n(f_\theta)\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] = \sum_{j=n+1}^{\infty} \phi_j(\tau)\underline{c}_{n,-j+\tau}$$

$$\Rightarrow \quad \left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] = \Gamma_n(f_\theta)^{-1}\sum_{j=n+1}^{\infty} \phi_j(\tau)\underline{c}_{n,-j+\tau} \qquad (C.17)$$

The ℓ_2-norm of the above gives

$$\left\|\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right\|_2 \leq \|\Gamma_n(f_\theta)^{-1}\|_{spec}\sum_{j=n+1}^{\infty} |\phi_j(\tau)| \cdot \|\underline{c}_{n,-j+\tau}\|_2.$$

To bound the above we use the well known result $\|\Gamma_n(f_\theta)^{-1}\|_{spec} \leq 1/\inf_\omega f(\omega) = f_0$ and $\|\underline{c}_{n,-j+\tau}\|_2 \leq \sum_{r\in\mathbb{Z}}|c_\theta(r)| = C_0$. This gives the bound

$$\left\|\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right\|_2 \leq f_0 C_0 \sum_{j=n+1}^{\infty} |\phi_j(\tau)|.$$

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 (C.17) with respective to θ_r and isolate $\partial[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)]/\partial\theta_r$. Differentiating both sides of (C.17) with respect to θ_r gives

$$\frac{\partial\Gamma_n(f_\theta)}{\partial\theta_r}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] + \Gamma_n(f_\theta)\frac{\partial}{\partial\theta_r}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right]$$

$$= \sum_{j=n+1}^{\infty}\left[\frac{\partial\phi_j(\tau)}{\partial\theta_r}\underline{c}_{n,-j+\tau} + \phi_j(\tau)\frac{\partial\underline{c}_{n,-j+\tau}}{\partial\theta_r}\right]. \qquad (C.18)$$

Isolating $\partial[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)]/\partial\theta_r$ gives

$$\frac{\partial}{\partial\theta_r}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right] = -\Gamma_n(f_\theta)^{-1}\frac{\partial\Gamma_n(f_\theta)}{\partial\theta_r}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right]$$

$$+\Gamma_n(f_\theta)^{-1}\sum_{j=n+1}^{\infty}\left[\frac{\partial\phi_j(\tau)}{\partial\theta_r}\underline{c}_{n,-j+\tau} + \phi_j(\tau)\frac{\partial\underline{c}_{n,-j+\tau}}{\partial\theta_r}\right]. \qquad (C.19)$$

Evaluating the ℓ_2 norm of the above and using $\|AB\underline{x}\|_2 \leq \|A\|_{spec}\|B\|_{spec}\|\underline{x}\|_2$ gives the bound

$$
\left\|\frac{\partial}{\partial \theta_r}\left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right]\right\|_2
$$

$$
\leq \|\Gamma_n(f_\theta)^{-1}\|_{spec}\left(\left\|\frac{\partial \Gamma_n(f_\theta)}{\partial \theta_r}\right\|_{spec}\left\|\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right\|_2\right.
$$

$$
\left.+ \sum_{j=n+1}^{\infty}\left|\frac{\partial \phi_j(\tau)}{\partial \theta_r}\right|\|\underline{c}_{n,-j+\tau}\|_2 + \sum_{j=n+1}^{\infty}|\phi_j(\tau)|\left\|\frac{\partial \underline{c}_{n,-j+\tau}}{\partial \theta_r}\right\|_2\right)
$$

$$
\leq f_0\left(\left\|\frac{\partial \Gamma_n(f_\theta)}{\partial \theta_r}\right\|_{spec}\left\|\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right\|_2\right.
$$

$$
\left.+ C_0\sum_{j=n+1}^{\infty}\left|\frac{\partial \phi_j(\tau)}{\partial \theta_r}\right| + \left(\sum_{r\in\mathbb{Z}}\|\nabla_\theta c_\theta(r)\|_2\right)\sum_{j=n+1}^{\infty}|\phi_j(\tau)|\right)
$$

$$
\leq f_0\left(\left\|\frac{\partial \Gamma_n(f_\theta)}{\partial \theta_r}\right\|_{spec}\left\|\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)\right\|_2 + C_0\sum_{j=n+1}^{\infty}\left|\frac{\partial \phi_j(\tau)}{\partial \theta_r}\right| + C_1\sum_{j=n+1}^{\infty}|\phi_j(\tau)|\right) \quad \text{(C.20)}
$$

where the last line in the above uses the bound $\sum_{r\in\mathbb{Z}}\|\nabla_\theta^a c_\theta(r)\|_2 \leq \sum_{r\in\mathbb{Z}}\|\nabla_\theta^a c_\theta(r)\|_1 = C_a$ (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

$$
\left\|\frac{\partial \Gamma_n(f_\theta)}{\partial \theta_r}\right\|_{spec} = \sup_{\|x\|_2=1}\left|\sum_{s,t=1}^{n}x_s x_t\frac{\partial c(s-t;f_\theta)}{\partial \theta_r}\right| = \sup_{\|x\|_2=1}\left|\frac{1}{2\pi}\int_0^{2\pi}|\sum_{s=1}^{n}x_s e^{is\omega}|^2\frac{\partial f_\theta(\omega)}{\partial \theta_r}d\omega\right|
$$

$$
\leq \sup_\omega\left|\frac{\partial f_\theta(\omega)}{\partial \theta_r}\right|\sup_{\|x\|_2=1}\frac{1}{2\pi}\int_0^{2\pi}\left|\sum_{s=1}^{n}x_s e^{is\omega}\right|^2 d\omega = \sup_\omega\left|\frac{\partial f_\theta(\omega)}{\partial \theta_r}\right|.
$$

By using the same argument one can show that the ath derivative is

$$
\left\|\frac{\partial^a \Gamma_n(f_\theta)}{\partial \theta_{r_1}\dots\partial \theta_{r_a}}\right\|_{spec} \leq \sup_\omega\left|\frac{\partial^a f_\theta(\omega)}{\partial \theta_{r_1}\dots\partial \theta_{r_a}}\right| = \sum_{r\in\mathbb{Z}}\left|\frac{\partial^a c(r;f_\theta)}{\partial \theta_{r_1}\dots\partial \theta_{r_a}}\right| \leq C_a. \quad \text{(C.21)}
$$

This general bound will be useful when evaluating the higher order derivatives below. Substi-

tuting (C.21) into (C.20) gives

$$\left\| \frac{\partial}{\partial \theta_r} \left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right] \right\|_2 \leq f_0 \left(C_1 \left\| \underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right\|_2 + C_1 \sum_{j=n+1}^{\infty} |\phi_j(\tau)| + C_0 \sum_{j=n+1}^{\infty} \|\nabla_\theta \phi_j(\tau)\|_1 \right).$$

This proves the result for $i = 1$.

<u>Proof for the case $i = 2$</u> We differentiate both sides of (C.18) with respect to θ_{r_2} to give the second derivative

$$\left(\frac{\partial^2 \Gamma_n(f_\theta)}{\partial \theta_{r_1} \partial \theta_{r_2}} \right) \left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right] + \frac{\partial \Gamma_n(f_\theta)}{\partial \theta_{r_2}} \frac{\partial}{\partial \theta_{r_1}} \left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right] +$$

$$\frac{\partial \Gamma_n(f_\theta)}{\partial \theta_{r_1}} \frac{\partial}{\partial \theta_{r_2}} \left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right] + \Gamma_n(f_\theta) \left(\frac{\partial^2}{\partial \theta_{r_2} \partial \theta_{r_1}} \left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right] \right)$$

$$= \sum_{j=n+1}^{\infty} \left[\frac{\partial^2 \phi_j(\tau)}{\partial \theta_{r_2} \partial \theta_{r_1}} \underline{c}_{n,-j+\tau} + \frac{\partial \phi_j(\tau)}{\partial \theta_{r_1}} \frac{\partial \underline{c}_{n,-j+\tau}}{\partial \theta_{r_2}} + \frac{\partial \phi_j(\tau)}{\partial \theta_{r_2}} \frac{\partial \underline{c}_{n,-j+\tau}}{\partial \theta_{r_1}} + \phi_j(\tau) \frac{\partial^2 \underline{c}_{n,-j+\tau}}{\partial \theta_{r_2} \partial \theta_{r_1}} \right].$$

(C.22)

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

$$\frac{\partial^2}{\partial \theta_{r_1} \partial \theta_{r_2}} \left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right]$$

$$= -\Gamma_n(f_\theta)^{-1} \left(\frac{\partial^2 \Gamma_n(f_\theta)}{\partial \theta_{r_1} \partial \theta_{r_2}} \right) \left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right] - \Gamma_n(f_\theta)^{-1} \frac{\partial \Gamma_n(f_\theta)}{\partial \theta_{r_2}} \frac{\partial}{\partial \theta_{r_1}} \left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right]$$

$$- \Gamma_n(f_\theta)^{-1} \frac{\partial \Gamma_n(f_\theta)}{\partial \theta_{r_1}} \frac{\partial}{\partial \theta_{r_2}} \left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right]$$

$$+ \Gamma_n(f_\theta)^{-1} \sum_{j=n+1}^{\infty} \left[\frac{\partial^2 \phi_j(\tau)}{\partial \theta_{r_2} \partial \theta_{r_1}} \underline{c}_{n,-j+\tau} + \frac{\partial \phi_j(\tau)}{\partial \theta_{r_1}} \frac{\partial \underline{c}_{n,-j+\tau}}{\partial \theta_{r_2}} \right.$$

$$\left. + \frac{\partial \phi_j(\tau)}{\partial \theta_{r_2}} \frac{\partial \underline{c}_{n,-j+\tau}}{\partial \theta_{r_1}} + \phi_j(\tau) \frac{\partial^2 \underline{c}_{n,-j+\tau}}{\partial \theta_{r_2} \partial \theta_{r_1}} \right].$$

Taking the ℓ_2-norm of $\frac{\partial^2}{\partial \theta_{r_1} \partial \theta_{r_2}} \left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right]$ and using (C.21) gives

$$\left\| \frac{\partial^2}{\partial \theta_{r_1} \partial \theta_{r_2}} \left[\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right] \right\|_2$$

$$\leq f_0 \left(C_2 \left\| \underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau) \right\|_2 + 2C_1 \left\| \nabla_\theta [\underline{\varphi}_n(\tau) - \underline{\phi}_n(\tau)] \right\|_2 \right.$$

$$\left. + C_0 \sum_{j=n+1}^{\infty} \|\nabla^2 \phi_j(\tau)\|_1 + 2C_1 \sum_{j=n+1}^{\infty} \|\nabla \phi_j(\tau)\|_1 + C_2 \sum_{j=n+1}^{\infty} |\phi_j(\tau)| \right).$$

This proves the result for $i = 2$. The proof for $i > 2$ follows using a similar argument (we omit the details). $\qquad\square$

The above result gives an ℓ_2-bound between the derivatives of the finite and infinite predictors. However, for our purposes an ℓ_1-bound is more useful. Thus we use the Cauchy-Schwarz inequality and norm inequality $\|\cdot\|_2 \le \|\cdot\|_1$ to give the ℓ_1-bound

$$
\left\| \frac{\partial^i}{\partial\theta_{r_1}\ldots\partial\theta_{r_i}} \left[\underline{\varphi}_n(\tau; f_\theta) - \underline{\phi}_n(\tau; f_\theta) \right] \right\|_1
$$
$$
\le n^{1/2} f_0 \left(\sum_{\substack{a_1+a_2=i \\ a_2 \neq i}} \binom{i}{a_1} C_{a_1} \left\| \nabla_\theta^{a_2} [\underline{\varphi}_n(\tau; f_\theta) - \underline{\phi}_n(\tau; f_\theta)] \right\|_1 + \right.
$$
$$
\left. \sum_{b_1+b_2=i} \binom{i}{b_1} C_{b_1} \sum_{j=n+1}^{\infty} \left\| \nabla_\theta^{b_2} \phi_j(\tau; f_\theta) \right\|_1 \right), \tag{C.23}
$$

this incurs an additional $n^{1/2}$ term. Next, considering all the partial derivatives with respect to θ of order i and using (C.15) we have

$$
\sum_{t=1}^{n} \| \nabla_\theta^i [\phi_{t,n}(\tau; f_\theta) - \phi_t(\tau; f_\theta)] \|_1
$$
$$
\le d^i n^{1/2} f_0 \left(\sum_{\substack{a_1+a_2=i \\ a_2 \neq i}} \binom{i}{a_1} C_{a_1} \left\| \nabla_\theta^{a_2} [\underline{\varphi}_n(\tau; f_\theta) - \underline{\phi}_n(\tau; f_\theta)] \right\|_1 + \right.
$$
$$
\left. \sum_{b_1+b_2=i} \binom{i}{b_1} C_{b_1} \sum_{j=n+1}^{\infty} \left\| \nabla_\theta^{b_2} \phi_j(\tau; f_\theta) \right\|_1 \right), \tag{C.24}
$$

where d is the dimension of the vector θ. The above gives a bound in terms of the infinite predictors. We now obtain a bound in terms of the corresponding $\mathrm{AR}(\infty)$ and $\mathrm{MA}(\infty)$ coefficients. To do this, we recall that for $\tau \le 0$, $\phi_j(\tau; f_\theta) = \sum_{s=0}^{\infty} \phi_{s+j}(f_\theta) \psi_{|\tau|-j}(f_\theta)$. Thus the partial derivatives of $\phi_j(\tau; f_\theta)$ give the bound

$$
\sum_{j=n+1}^{\infty} \| \nabla_\theta \phi_j(\tau; f_\theta) \|_1 \le \sum_{s=0}^{\infty} \sum_{j=n+1}^{\infty} \left(|\psi_{|\tau|-j}(f_\theta)| \cdot \| \nabla_\theta \phi_{s+j}(f_\theta) \|_1 + |\phi_{s+j}(f_\theta)| \cdot \| \nabla_\theta \psi_{|\tau|-j}(f_\theta) \|_1 \right).
$$

Substituting the above bound into (C.24) and using Lemma B.1 for the case $i = 1$ gives

$$\sum_{s=1}^{n} \|\nabla_\theta [\phi_{s,n}(\tau; f_\theta) - \phi_s(\tau; f_\theta)]\|_1$$

$$\leq n^{1/2} df_0 \bigg\{ C_1(C_0 + 1) \sum_{j=n+1}^{\infty} |\phi_j(\tau; f_\theta)| +$$

$$C_0 \sum_{s=0}^{\infty} \sum_{j=n+1}^{\infty} \left(|\psi_{|\tau|-j}(f_\theta)| \cdot \|\nabla_\theta \phi_{s+j}(f_\theta)\|_1 + |\phi_{s+j}(f_\theta)| \cdot \|\nabla_\theta \psi_{|\tau|-j}(f_\theta)\|_1 \right) \bigg\}. \quad \text{(C.25)}$$

The above results are used to obtain bounds between the derivatives of the Whittle and Gaussian likelihood in Appendix C.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(∞) and AR(∞) coefficients.

Remark C.2 *It is interesting to note that it is possible to prove analgous bounds to those in Lemma C.1 using the derivatives of the series expansion in the proof of Theorem 3.3, that is*

$$\frac{\partial}{\partial \theta_r} [\phi_{j,n}(\tau; f_\theta) - \phi_j(\tau; f_\theta)] = \sum_{s=2}^{\infty} \frac{\partial}{\partial \theta_r} \phi_{j,n}^{(s)}(\tau; f_\theta).$$

This would give bounds on

$$\sum_{\tau \leq 0} \frac{\partial}{\partial \theta_r} [\phi_{j,n}(\tau; f_\theta) - \phi_j(\tau; f_\theta)] e^{i\tau\omega} = \phi(\omega)^{-1} \sum_{s=2}^{\infty} \frac{\partial}{\partial \theta_r} \left[\phi(\omega; f_\theta)^{-1} \zeta_{j,n}^{(s)}(\omega; f_\theta) \right].$$

This approach would immediately give bounds based on the derivatives of the AR(∞) coefficients.

We now state and prove a lemma which will be useful in a later section (it is not directly related to Baxter's inequality).

Lemma C.2 *Let $\{\phi_j(f_\theta)\}$ and $\{\psi_j(f_\theta)\}$ denote the AR(∞), and MA(∞) coefficients corresponding to the spectral density f_θ. Suppose the same set of Assumptions in Lemma C.1 holds. Let $\{\alpha_j(f_\theta)\}$ denote the Fourier coefficients in the one-sided expansion*

$$\log(\sum_{j=0}^{\infty} \psi_j(f_\theta) z^j) = \sum_{j=1}^{\infty} \alpha_j(f_\theta) z^j \quad \text{for } |z| < 1, \quad \text{(C.26)}$$

Then for all $\theta \in \Theta$ and for $0 \leq s \leq \kappa$ 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 (C.26) 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

$$\sum_{j=1}^{\infty} j\alpha_j(f_\theta) z^{j-1} = \left(\sum_{j=1}^{\infty} j\psi_j(f_\theta) z^{j-1} \right) \left(\sum_{j=0}^{\infty} \psi_j(f_\theta) z^j \right)^{-1}$$

$$= \left(\sum_{j=1}^{\infty} j\psi_j(f_\theta) z^{j-1} \right) \left(\sum_{j=0}^{\infty} \widetilde{\phi}_j(f_\theta) z^j \right), \qquad (C.27)$$

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

$$j\alpha_j(f_\theta) = \sum_{\ell=0}^{j-1} (j - \ell)\psi_{j-\ell}(f_\theta)\widetilde{\phi}_\ell(f_\theta)$$

$$\Rightarrow \quad \alpha_j(f_\theta) = j^{-1} \sum_{\ell=0}^{j-1} (j - \ell)\psi_{j-\ell}(f_\theta)\widetilde{\phi}_\ell(f_\theta) \quad \text{for} \quad j \geq 1. \qquad (C.28)$$

Therefore, using the above and taking the absolute into the summand we have

$$\sum_{j=1}^{\infty} j^K |\alpha_j(f_\theta)| \leq \sum_{j=1}^{\infty} \sum_{\ell=0}^{j-1} j^{K-1}(j - \ell)|\psi_{j-\ell}(f_\theta)||\widetilde{\phi}_\ell(f_\theta)|$$

$$= \sum_{\ell=1}^{\infty} |\widetilde{\phi}_\ell(f_\theta)| \sum_{j=\ell+1}^{\infty} j^{K-1}(j - \ell)|\psi_{j-\ell}(f_\theta)| \qquad \text{(exchange summation)}$$

$$= \sum_{\ell=1}^{\infty} |\widetilde{\phi}_\ell(f_\theta)| \sum_{s=1}^{\infty} (s + \ell)^{K-1} s|\psi_s(f_\theta)| \qquad \text{(change of variable } s = j + \ell)$$

$$\leq \sum_{\ell=1}^{\infty} |\widetilde{\phi}_\ell(f_\theta)| \sum_{s=1}^{\infty} (s + \ell)^K |\psi_s(f_\theta)|. \qquad ((s + \ell)^{-1} \leq s^{-1})$$

Since $K \geq 1$, using inequality $(a + b)^K \leq 2^{K-1}(a^K + b^K)$ for $a, b > 0$, we have

$$\sum_{j=1}^{\infty} j^K |\alpha_j(f_\theta)| \leq \sum_{\ell=1}^{\infty} |\widetilde{\phi}_\ell(f_\theta)| \sum_{s=1}^{\infty} (s + \ell)^K |\psi_s(f_\theta)|$$

$$\leq 2^{K-1} \sum_{\ell=1}^{\infty} |\widetilde{\phi}_\ell(f_\theta)| \sum_{s=1}^{\infty} (s^K + \ell^K)|\psi_s(f_\theta)|$$

$$= 2^{K-1} \left(\sum_{\ell=1}^{\infty} \ell^K |\widetilde{\phi}_\ell(f_\theta)| \cdot \sum_{s=1}^{\infty} |\psi_s(f_\theta)| + \sum_{\ell=1}^{\infty} |\widetilde{\phi}_\ell(f_\theta)| \cdot \sum_{s=1}^{\infty} s^K |\psi_s(f_\theta)| \right) \leq \infty$$

and this proves the lemma when $s = 0$.

To prove lemma for $s = 1$, we differentiate (C.28) with θ, then, by Assumption 5.1 (iii),

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

$$\leq \sum_{j=1}^{\infty} \sum_{\ell=0}^{j-1} \|j^{K-1}(j-\ell)\nabla_\theta \psi_{j-\ell}(f_\theta)\widetilde{\phi}_\ell(f_\theta)\|_1 + \sum_{j=1}^{\infty} \sum_{\ell=0}^{j-1} \|j^{K-1}(j-\ell)\psi_{j-\ell}(f_\theta)\nabla_\theta \widetilde{\phi}_\ell(f_\theta)\|_1$$

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 $s \geq 2$ is similar (we omit the detail). $\qquad\square$

C.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. 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 2.2 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 θ, the second term is the additional term due to the Gaussian likelihood.

For the simplicity, assume θ 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 3.2, but for the derivatives. We will use this result to prove Theorem E.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 3.2, the derivative of this term with respect to θ does not (usually) have a simple analytic form. Therefore, analogous to Theorem 3.1 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 C.3 *Suppose Assumption 5.1(i),(iii) holds and g is a bounded function. Then for $1 \leq i \leq 3$ we have*

$$\left\| F_n^* \Delta_n(g) \frac{d^i}{d\theta^i} \left(D_n(f_\theta) - D_{\infty,n}(f_\theta) \right) \right\|_1 = O(n^{-K+3/2}), \tag{C.29}$$

and

$$\left\| F_n^* \sum_{k=0}^{i} \binom{i}{k} \frac{d^k \Delta_n \left(f_\theta^{-1} \right)}{d\theta^k} \frac{d^{k-i} D_{\infty,n}(f_\theta)}{d\theta^{k-i}} \right\|_1 = O(1). \tag{C.30}$$

PROOF. To bound (C.29), we use the expression for $F_n^* \Delta_n(g^{-1}) \left(D_n(f_\theta) - D_{\infty,n}(f_\theta) \right)$ given in (B.4)

$$\left(F_n^* \Delta_n(g^{-1}) \left[D_n(f_\theta) - D_{\infty,n}(f_\theta) \right] \right)_{s,t} = \sum_{\tau \leq 0} [\{ \phi_{t,n}(\tau; f_\theta) - \phi_t(\tau; f_\theta) \} G_{1,n}(s, \tau; g)$$
$$+ \{ \phi_{n+1-t,n}(\tau; f_\theta) - \phi_{n+1-t}(\tau; f_\theta) \} G_{2,n}(s, \tau; g)].$$

Differentiating the above with respect to θ gives

$$\left[F_n^* \Delta_n(g) \frac{d}{d\theta} \left(D_n(f_\theta) - D_{\infty,n}(f_\theta) \right) \right]_{s,t}$$
$$= \sum_{\tau \leq 0} \left[G_{1,n}(s, \tau) \frac{d}{d\theta} [\phi_{t,n}(\tau) - \phi_t(\tau)] + G_{2,n}(s, \tau) \frac{d}{d\theta} [\phi_{n+1-t}(\tau) - \phi_{n+1-t}(\tau)] \right]$$
$$= T_{s,t,1} + T_{s,t,2}.$$

We recall that equation (C.25) gives the bound

$$\sum_{s=1}^{n} \left| \frac{d}{d\theta} [\phi_{s,n}(\tau; f_\theta) - \phi_s(\tau; f_\theta)] \right|$$

$$\leq n^{1/2} f_0 \left\{ C_1(C_{f,0} + 1) \sum_{j=n+1}^{\infty} |\phi_j(\tau; f_\theta)| + \right.$$

$$\left. C_0 \sum_{s=0}^{\infty} \sum_{j=n+1}^{\infty} \left(|\psi_{|\tau|-j}(f_\theta)| \cdot \left| \frac{d}{d\theta} \phi_{s+j}(f_\theta) \right| + |\phi_{s+j}(f_\theta)| \cdot \left| \frac{d}{d\theta} \nabla_\theta \psi_{|\tau|-j}(f_\theta) \right| \right) \right\}.$$

Substituting this into $T_{s,t,1}$ gives the bound

$$|T_{s,t,1}| \leq Cn^{1/2} \sum_{\tau \leq 0} G_{1,n}(s, \tau) \left(\sum_{j=n+1}^{\infty} \sum_{s=0}^{\infty} |\phi_{s+j}||\psi_{|\tau|-j}| + \sum_{s=0}^{\infty} \sum_{j=n+1}^{\infty} \left| \frac{d\phi_{s+j}}{d\theta} \psi_{|\tau|-j} + \phi_{s+j} \frac{d\psi_{|\tau|-j}}{d\theta} \right| \right).$$

81

Using the same techniques used to prove Theorem 3.1 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 (C.29) for the case $i = 1$. The proof for the cases $i = 2, 3$ is similar.

To prove (C.30) we use the same method used to prove Theorem 3.1, equation (3.5). But with $\frac{d^k f_\theta^{-1}}{d\theta^k}$ replacing f_θ 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 C.1 *Suppose Assumptions 5.1 (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*

$$\left\|\frac{1}{n}\underline{X}_n'\left[F_n^*\Delta_n(g)\frac{d^i}{d\theta^i}\left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right)\right]\underline{X}_n\right\|_{\mathbb{E},q} = O(n^{-K+1/2}), \qquad (C.31)$$

and

$$\left\|\frac{1}{n}\underline{X}_n'F_n^*\sum_{\ell=0}^{i}\binom{i}{\ell}\frac{d^\ell\Delta_n\left(f_\theta^{-1}\right)}{d\theta^\ell}\frac{d^{\ell-i}D_{\infty,n}(f_\theta)}{d\theta^{\ell-i}}\underline{X}_n\right\|_{\mathbb{E},q} = O(n^{-1}) \qquad (C.32)$$

for $i = 1, 2$ and 3.

PROOF. To prove (C.31), we observe that

$$\left\|\frac{1}{n}\underline{X}_n'\left[F_n^*\Delta_n(g)\frac{d^i}{d\theta^i}\left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right)\right]\underline{X}_n\right\|_{\mathbb{E},q}$$

$$\leq \frac{1}{n}\sum_{s,t=1}^{n}\left|\left[F_n^*\Delta_n(g)\frac{d^i}{d\theta^i}\left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right)\right]_{s,t}\right|\|X_sX_t\|_{\mathbb{E},q}$$

$$= \frac{1}{n}\sup_t\|X_t\|_{\mathbb{E},2q}^2\sum_{s,t=1}^{n}\left|\left[F_n^*\Delta_n(g)\frac{d^i}{d\theta^i}\left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right)\right]_{s,t}\right|$$

$$= \frac{1}{n}\|X\|_{\mathbb{E},2q}^2\left\|F_n^*\Delta_n(g)\frac{d^i}{d\theta^i}\left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right)\right\|_1 = O(n^{-K+1/2})$$

where the above follows from Lemma C.3, equation (C.29). This proves (C.31).

To prove (C.32) we use the the bound in (C.30) together with a similar proof to that described

above. This immediately proves (C.32). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We now apply the above result to the difference in the derivatives of the Gaussian and Whittle likelihood. It is straightforward to show that

$$\underline{X}'_n F^*_n \frac{d^i}{d\theta^i}\Delta_n(f_\theta^{-1})D_n(f_\theta)\underline{X}_n \qquad\qquad\qquad\text{(C.33)}$$

$$= \underline{X}'_n F^*_n \left[\sum_{\ell=0}^{i}\binom{i}{\ell}\frac{d^\ell\Delta_n\left(f_\theta^{-1}\right)}{d\theta^\ell}\frac{d^{\ell-i}D_n(f_\theta)}{d\theta^{\ell-i}}\right]\underline{X}_n$$

$$= \underline{X}'_n F^*_n \left[\sum_{\ell=0}^{i}\binom{i}{\ell}\frac{d^\ell\Delta_n\left(f_\theta^{-1}\right)}{d\theta^\ell}\frac{d^{\ell-i}D_{\infty,n}(f_\theta)}{d\theta^{\ell-i}}\right]\underline{X}_n$$

$$+\underline{X}'_n F^*_n \left(\sum_{\ell=0}^{i}\binom{i}{\ell}\frac{d^\ell\Delta_n\left(f_\theta^{-1}\right)}{d\theta^\ell}\left[\frac{d^{\ell-i}D_n(f_\theta)}{d\theta^{\ell-i}} - \frac{d^{\ell-i}D_{\infty,n}(f_\theta)}{d\theta^{\ell-i}}\right]\right)\underline{X}_n. \qquad\text{(C.34)}$$

First we study the second term on the right hand side of the above. By applying Corollary C.1 (and under Assumption 5.1) for $1 \leq i \leq 3$ we have

$$\left\|n^{-1}\underline{X}'_n F^*_n\left(\frac{d^i}{d\theta^i}\Delta_n(f_\theta^{-1})\left[D_n(f_\theta) - D_{\infty,n}(f_\theta)\right]\right)\underline{X}_n\right\|_{\mathbb{E},1}$$

$$= \left\|n^{-1}\underline{X}'_n F^*_n\sum_{\ell=0}^{i}\binom{i}{\ell}\frac{d^\ell\Delta_n\left(f_\theta^{-1}\right)}{d\theta^\ell}\left[\frac{d^{\ell-i}D_n(f_\theta)}{d\theta^{\ell-i}} - \frac{d^{\ell-i}D_{\infty,n}(f_\theta)}{d\theta^{\ell-i}}\right]\underline{X}_n\right\|_{\mathbb{E},1}$$

$$= O(n^{-K+1/2}). \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{(C.35)}$$

On the other hand, the first term on the right hand side of (C.33) has the bound

$$\left\|n^{-1}\underline{X}'_n F^*_n\frac{d^i}{d\theta^i}\left[\Delta_n(f_\theta^{-1})D_n(f_\theta)\right]\underline{X}_n\right\|_{\mathbb{E},1} = O(n^{-1}). \qquad\text{(C.36)}$$

D Rates of convergence of the new likelihood estimators

In this section we study the sampling properties of the new criteria.

D.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_θ 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

$$\theta_n = \arg\min_\theta I_n(f, f_\theta). \tag{D.1}$$

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 5.1(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(∞) 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

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

and since $\log f_\theta$ is real and symmetric about π, $\alpha_{-j}(f_\theta) = \alpha_j(f_\theta) \in \mathbb{R}$. This allows us to obtain coefficients $\{\alpha_j(f_\theta)\}$ in terms of the MA(∞) coefficients (it is interesting to note that Pourahmadi (2001) gives a recursion for $\alpha_j(f_\theta)$ in terms of the MA(∞) coefficient). The result is given in Lemma C.2, but we summarize it below. Under Assumption 5.1(iii) we have for $0 \le s \le \kappa$ (for some $\kappa \ge 4$)

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

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

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

The sth-order derivative ($s \ge 1$) with respect to θ (and using Assumption 5.1(ii) that σ^2 does

not depend on θ) we have

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

By using Lemma C.2 for $0 \leq s \leq \kappa$ we have

$$\|\sum_{r\in\mathbb{Z}\backslash\{0\}}\nabla_{\theta}^{s}\alpha_{rn}(f_{\theta})\|_{1} \leq 2\sum_{j\geq n}\|\nabla_{\theta}^{s}\alpha_{j}(f_{\theta})\|_{1} = O(n^{-K}). \tag{D.3}$$

Substituting the bound in (D.3) (for $s = 0$) into (D.2) gives

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

Using (D.3) for $1 \leq s \leq \kappa$ 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 deterministic 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 E.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

$$
\begin{aligned}
\mathcal{L}_n(\theta) &= n^{-1}\underline{X}_n'\Gamma_n(f_\theta)^{-1}\underline{X}_n = \frac{1}{n}\sum_{k=1}^{n}\frac{\widetilde{J}_n(\omega_{k,n};f_\theta)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} \\
K_n(\theta) &= \frac{1}{n}\sum_{k=1}^{n}\frac{|J_n(\omega_{k,n})|^2}{f_\theta(\omega_{k,n})} \\
\widehat{W}_{p,n}(\theta) &= \frac{1}{n}\sum_{k=1}^{n}\frac{\widetilde{J}_n(\omega_{k,n};\widehat{f}_p)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} \\
\widehat{H}_{p,n}(\theta) &= \frac{1}{n}\sum_{k=1}^{n}\frac{\widetilde{J}_n(\omega_{k,n};\widehat{f}_p)\overline{J_{n,\underline{h}_n}(\omega_{k,n})}}{f_\theta(\omega_{k,n})}.
\end{aligned}
\tag{D.4}
$$

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

$$
\begin{aligned}
\widehat{\theta}_n^{(G)} &= \arg\min \mathcal{L}_n(\theta), \quad \widehat{\theta}_n^{(K)} = \arg\min K_n(\theta), \\
\widehat{\theta}_n^{(W)} &= \arg\min \widehat{W}_{p,n}(\theta), \text{ and } \widehat{\theta}_n^{(H)} = \arg\min \widehat{H}_{p,n}(\theta)
\end{aligned}
\tag{D.5}
$$

D.2 Asymptotic equivalence to the infeasible criteria

In this section we analyze the feasible estimators $\widehat{\theta}_n^{(W)} = \arg\min \widehat{W}_{p,n}(\theta)$ and $\widehat{\theta}_n^{(H)} = \arg\min \widehat{H}_{p,n}(\theta)$. We show that they are asymptotically equivalent to the corresponding infeasible estimators $\widetilde{\theta}_n^{(W)} = \arg\min W_n(\theta)$ and $\widetilde{\theta}_n^{(H)} = \arg\min H_n(\theta)$, in the sense that

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

where $|\underline{a}|_1 = \sum_{j=1}^{d}|a_j|$, where $\underline{a} = (a_1, \ldots, a_d)$. We will make the assumption that that the ratio of tapers satisfy the condition $H_{2,n}/H_{1,n}^2 \sim n^{-1}$. 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 Corollary 2.1 and Theorem 3.1 in Das et al. (2020) (under Assumption 5.2) we have

$$
[\widehat{J}_n(\omega;\widehat{f}_p) - \widehat{J}_n(\omega;f)]\overline{J_{n,\underline{h}_n}(\omega)} = O_p\left(\frac{p^2}{n} + \frac{p^3}{n^{3/2}}\right)
\tag{D.6}
$$

and

$$
\widehat{H}_{p,n}(\theta) = H_n(\theta) + O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).
\tag{D.7}
$$

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 $H_{2,n}/H_{1,n}^2 \sim n^{-1}$, then the $n^{1/2}$–rate and the rates in (D.6) and (D.7) would change. This will make the proofs more technical. Thus for ease of notation and presentation we will assume that $H_{2,n}/H_{1,n}^2 \sim n^{-1}$.

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

Lemma D.1 *Suppose that Assumptions 5.1(i,iii) and 5.2(i,ii) hold. Then for $0 \leq s \leq \kappa$ (for some $\kappa \geq 4$) 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. 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

$$\sup_{\theta \in \Theta} |\widehat{W}_{p,n}(\theta) - W_n(\theta)| \leq \sup_{\theta, \omega} f_\theta(\omega)^{-1} \times \frac{1}{n} \sum_{k=1}^n \left| \left[\widehat{J}_n(\omega_{k,n}; \widehat{f}_p) - \widehat{J}_n(\omega_{k,n}; f) \right] \overline{J_n(\omega_{k,n})} \right|.$$

By using Corollary 2.1 in Das et al. (2020) (under Assumption 5.2) 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 Corollary 2.1 in Das et al. (2020), $O_p\left(p^3 n^{-3/2}\right)$ bound is uniform all frequencies, $\sup_\omega \mathbb{E}[\Delta(\omega)] = O((np^{K-1})^{-1} + p^3/n^2)$ and $\sup_\omega \operatorname{var}[\Delta(\omega)] = O(p^4/n^2)$. 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 5.1(iii)) and $\widehat{H}_{p,n}(\theta)$, we omit the details. □

87

Lemma D.2 *Suppose that Assumptions 5.1(i,iii) and 5.2(i,ii) hold. Then*

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

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

PROOF. 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.,

$$\sup_{\theta \in \Theta} |W_n(\theta) - \mathcal{W}_n(\theta)| \overset{\mathcal{P}}{\to} 0. \tag{D.8}$$

Using Das et al. (2020), Theorem A.1 and the classical result $\mathrm{var}[K_n(\theta)] = O(n^{-1})$ we have

$$
\begin{aligned}
\mathrm{var}[W_n(\theta)] &= \mathrm{var}\left(K_n(\theta) + n^{-1}\sum_{k=1}^{n} \frac{\widehat{J}_n(\omega_{k,n}; f)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})}\right) \\
&\leq 2\mathrm{var}[K_n(\theta)] + \frac{2}{n}\sum_{k=1}^{n} \mathrm{var}[\widehat{J}_n(\omega_{k,n}; f)\overline{J_n(\omega_{k,n})}]/f_\theta(\omega_{k,n})^2 \\
&\leq 2\mathrm{var}[K_n(\theta)] + O(n^{-2}) = O(n^{-1}).
\end{aligned}
$$

Therefore, by Markov's inequality, $W_n(\theta) \overset{\mathcal{P}}{\to} \mathcal{W}_n(\theta)$ for each $\theta \in \Theta$. To show a uniform convergence, since Θ 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

$$
\begin{aligned}
I_1(\theta_1,\theta_2) &= n^{-1}\sum_{k=1}^{n}\left(f_{\theta_1}^{-1}(\omega_{k,n}) - f_{\theta_2}^{-1}(\omega_{k,n})\right)\widetilde{J}_n(\omega_{k,n}; f)\overline{J_n(\omega_{k,n})} \\
&= n^{-1}\sum_{k=1}^{n}\nabla_\theta f_\theta^{-1}(\omega_{k,n})\rfloor'_{\theta=\bar\theta_k}(\theta_1 - \theta_2)\widetilde{J}_n(\omega_{k,n}; f)\overline{J_n(\omega_{k,n})} \\
&= \mathcal{K}_n(\bar\theta)'(\theta_1 - \theta_2),
\end{aligned}
$$

where $\mathcal{K}_n(\bar\theta) = n^{-1}\sum_{k=1}^{n}\widetilde{J}_n(\omega_{k,n}; f)\overline{J_n(\omega_{k,n})}\nabla_\theta f_\theta^{-1}(\omega_{k,n})\rfloor_{\theta=\bar\theta_k}$ and $\bar\theta_1, ..., \bar\theta_n$ are convex combina-

tions of θ_1 and θ_2. It is clear that

$$\|\mathcal{K}_n(\bar{\theta})\|_1 \le \sup_{\theta,\omega} \|\nabla_\theta f_\theta^{-1}(\omega)\|_1 \frac{1}{n} \sum_{k=1}^{n} \left| \widetilde{J}_n(\omega_{k,n}; f) \overline{J_n(\omega_{k,n})} \right| = K_n.$$

Thus

$$|I_1(\theta_1, \theta_2)| \le K_n |\theta_1 - \theta_2|_1. \tag{D.9}$$

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

$$\begin{aligned}
\mathbb{E} \left| \widetilde{J}_n(\omega_{k,n}; f) \overline{J_n(\omega_{k,n})} \right| &\le \mathbb{E}|J_n(\omega_{k,n})|^2 + \mathbb{E} \left| \widehat{J}_n(\omega_{k,n}; f) \overline{J_n(\omega_{k,n})} \right| \\
&\le f(\omega_{k,n}) + \mathrm{var}(\widehat{J}_n(\omega_{k,n}; f))^{1/2} \mathrm{var}(J_n(\omega_{k,n}))^{1/2} \\
&\le f(\omega_{k,n})(1 + O(n^{-1})).
\end{aligned}$$

Using above and Assumption 5.1(iii-a) gives

$$\sup_n \mathbb{E}[K_n] \le \sup_{\theta,\omega} \|\nabla_\theta f_\theta^{-1}(\omega)\|_1 \cdot \sup_n \frac{1}{n} \sum_{k=1}^{n} f(\omega_{k,n})(1 + O(n^{-1})) < \infty.$$

Therefore, $K_n = O_p(1)$ and from (D.9), $I_1(\theta_1, \theta_2)$ 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 (D.8).

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

$$W_n(\widetilde{\theta}_n^{(W)}) - \mathcal{W}_n(\widetilde{\theta}_n^{(W)}) \le W_n(\widetilde{\theta}_n^{(W)}) - \mathcal{W}_n(\theta_n) \le W_n(\theta_n) - \mathcal{W}_n(\theta_n).$$

Thus

$$|W_n(\widetilde{\theta}_n^{(W)}) - \mathcal{W}_n(\theta_n)| \le \sup_\theta |W_n(\theta) - \mathcal{W}_n(\theta)| \xrightarrow{\mathcal{P}} 0.$$

If θ_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

$$\sup_\theta |\widehat{W}_{p,n}(\theta) - \mathcal{W}_n(\theta)| \xrightarrow{\mathcal{P}} 0. \tag{D.10}$$

Now by using the triangular inequality, together with (D.8) and Lemma D.1, (D.10) 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 $|\widehat{\theta}_n^{(H)} - \theta_n|_1 \xrightarrow{\mathcal{P}} 0$. $\qquad\qquad\square$

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

Lemma D.3 *Suppose Assumptions 5.1(i,iii) and 5.2 hold. Then for $i = 1, 2$ we have*

$$\frac{d^i \widehat{W}_{p,n}(\theta)}{d\theta^i}\rfloor_{\theta=\theta_n} = \frac{d^i W_n(\theta)}{d\theta^i}\rfloor_{\theta=\theta_n} + O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right) \tag{D.11}$$

and

$$\left|\frac{d^3 \widehat{W}_{p,n}(\theta)}{d\theta^3}\rfloor_{\theta=\bar{\theta}_n} - \frac{d^3 W_n(\theta)}{d\theta^3}\rfloor_{\theta=\theta_n}\right| = O_p\left(\frac{p^2}{n}\right) + |\widehat{\theta}_n^{(W)} - \theta_n|O_p(1), \tag{D.12}$$

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

$$(\widehat{\theta}_n^{(W)} - \theta_n) = -\left[\mathbb{E}\left[\frac{d^2 W_n(\theta_n)}{d\theta_n^2}\right]\right]^{-1} \frac{dW_n(\theta_n)}{d\theta_n} + O_p\left(\frac{1}{n} + \frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right) \tag{D.13}$$

and

$$\frac{dW_n(\theta)}{d\theta}\rfloor_{\theta=\theta_n} + (\widehat{\theta}_n^{(W)} - \theta_n)\frac{d^2 W_n(\theta)}{d\theta^2}\rfloor_{\theta=\theta_n} + \frac{1}{2}(\widehat{\theta}_n^{(W)} - \theta_n)^2\frac{d^3 W_n(\theta)}{d\theta^3}\rfloor_{\theta=\theta_n}$$

$$= O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right). \tag{D.14}$$

PROOF. By using Theorem 3.1, Das et al. (2020) 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 (D.11). Let $\bar{\theta}_n$ denote a convex combination of θ_n and $\widehat{\theta}_n^{(W)}$ (note that $\widehat{\theta}_n^{(W)}$ is a consistent estimator of θ_n). To evaluate $\frac{d^3 \widehat{W}_n(\theta)}{d\theta^3}$ at the (consistent) estimator $\bar{\theta}_n$, a slightly different approach is required (due to the additional random parameter $\bar{\theta}_n$). By using

triangular inequality and Lemma D.1 we have

$$\left| \frac{d^3 \widehat{W}_{p,n}(\theta)}{d\theta^3} \right|_{\theta=\bar{\theta}_n} - \frac{d^3 W_n(\theta)}{d\theta^3} \right|_{\theta=\theta_n} \right|$$

$$\leq \left| \frac{d^3 \widehat{W}_{p,n}(\theta)}{d\theta^3} \right|_{\theta=\bar{\theta}_n} - \frac{d^3 W_n(\theta)}{d\theta^3} \right|_{\theta=\bar{\theta}_n} \right|$$

$$+ \frac{1}{n} \sum_{k=1}^{n} \left| \frac{d^3}{d\theta^3} \left[f_{\bar{\theta}_n}(\omega_{k,n})^{-1} - f_{\theta_n}(\omega_{k,n})^{-1} \right] \right| \left| \widetilde{J}_n(\omega_{k,n}; f) \overline{J_n(\omega_{k,n})} \right|$$

$$= O_p \left(\frac{p^2}{n} \right) + \frac{1}{n} \sum_{k=1}^{n} \left| \frac{d^3}{d\theta^3} \left[f_{\bar{\theta}_n}(\omega_{k,n})^{-1} - f_{\theta_n}(\omega_{k,n})^{-1} \right] \right| \left| \widetilde{J}_n(\omega_{k,n}; f) \overline{J_n(\omega_{k,n})} \right|.$$

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

$$\left| \frac{d^3}{d\theta^3} (f_{\bar{\theta}_n}^{-1} - f_{\theta_n}^{-1}) \right| \leq \sup_\theta \left| \frac{d^4}{d\theta^4} f_\theta^{-1} \right| \cdot |\bar{\theta}_n - \theta_n| \leq \sup_\theta \left| \frac{d^4}{d\theta^4} f_\theta^{-1} \right| \cdot |\widehat{\theta}_n^{(W)} - \theta_n|,$$

note that to bound the fourth derivation we require Assumption 5.1(iii) for $\kappa = 4$. Substituting this into the previous inequality gives

$$\left| \frac{d^3 \widehat{W}_{p,n}(\theta)}{d\theta^3} \right|_{\theta=\bar{\theta}_n} - \frac{d^3 W_n(\theta)}{d\theta^3} \right|_{\theta=\theta_n} \right|$$

$$\leq O_p \left(\frac{p^2}{n} \right) + \frac{1}{n} \sum_{k=1}^{n} \left| \frac{d^3}{d\theta^3} \left[f_{\bar{\theta}_n}(\omega_{k,n})^{-1} - f_{\theta_n}(\omega_{k,n})^{-1} \right] \right| \left| \widetilde{J}_n(\omega_{k,n}; f) \overline{J_n(\omega_{k,n})} \right|$$

$$= O_p \left(\frac{p^2}{n} \right) + |\widehat{\theta}_n^{(W)} - \theta_n| O_p(1).$$

The above proves (D.12).

Using (D.11) and (D.12) we now obtain the first and second order expansions in (D.13) and (D.14). In order to prove (D.13), 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 θ_n and assuming that $\widehat{\theta}_n^{(W)}$ lies inside the parameter space we have

$$0 = \frac{d\widehat{W}_{p,n}(\widehat{\theta}_n^{(W)})}{d\widehat{\theta}_n^{(W)}} = \frac{d\widehat{W}_{p,n}(\theta_n)}{d\theta_n} + (\widehat{\theta}_n^{(W)} - \theta_n) \frac{d^2 \widehat{W}_{p,n}(\theta_n)}{d\theta_n^2} + \frac{1}{2} (\widehat{\theta}_n^{(W)} - \theta_n)^2 \frac{d^3 \widehat{W}_{p,n}(\bar{\theta}_n)}{d\theta_n^3}$$

where $\bar{\theta}_n$ is a convex combination of θ_n and $\widehat{\theta}_n^{(W)}$. Now by using (D.11) and (D.12) we can replace

91

in the above $\widehat{W}_{p,n}(\theta_n)$ and its derivatives with $W_n(\theta_n)$ and its derivatives. Therefore,

$$\frac{dW_n(\theta_n)}{d\theta_n} + (\widehat{\theta}_n^{(W)} - \theta_n)\frac{d^2W_n(\theta_n)}{d\theta_n^2} + \frac{1}{2}(\widehat{\theta}_n^{(W)} - \theta_n)^2\frac{d^3W_n(\theta_n)}{d\theta_n^3}$$

$$= O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right) + (\widehat{\theta}_n^{(W)} - \theta_n)^2 O_p\left(\frac{p^2}{n}\right) + |\widehat{\theta}_n^{(W)} - \theta_n|^3 O_p(1). \qquad \text{(D.15)}$$

Rearranging the above gives

$$(\widehat{\theta}_n^{(W)} - \theta_n) = -\left[\frac{d^2W_n(\theta_n)}{d\theta_n^2}\right]^{-1}\frac{dW_n(\theta_n)}{d\theta_n} - \frac{1}{2}\left[\frac{d^2W_n(\theta_n)}{d\theta_n^2}\right]^{-1}\frac{d^3W_n(\theta_n)}{d\theta_n^3}(\widehat{\theta}_n^{(W)} - \theta_n)^2$$

$$+ O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right) + (\widehat{\theta}_n^{(W)} - \theta_n)^2 O_p\left(\frac{p^2}{n}\right) + |\widehat{\theta}_n^{(W)} - \theta_n|^3 O_p(1). \text{(D.16)}$$

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 (D.3)) and $\text{var}[\frac{dW_n(\theta_n)}{d\theta_n}] = O_p(n^{-1})$ we have $\frac{dW_n(\theta_n)}{d\theta_n} = O_p(n^{-1/2})$. Substituting this into (D.16) 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

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

$$\text{(D.17)}$$

Thus, from the above and the consistency result in Lemma D.2 ($|\widehat{\theta}_n^{(W)} - \theta_n| = o_p(1)$) we have[1]

$$(\widehat{\theta}_n^{(W)} - \theta_n) = O_p\left(\frac{1}{n^{1/2}} + \frac{p^3}{n^{3/2}}\right). \tag{D.18}$$

We use the above bound to obtain an exact expression for the dominating rate $O_p(n^{-1/2})$. Returning to equation (D.16) and substituting this bound into the quadratic term in (D.16) gives

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

Using that $\frac{d^2 W_n(\theta_n)}{d\theta_n^2} = \mathbb{E}\left(\frac{d^2 W_n(\theta_n)}{d\theta_n^2}\right) + O_p(n^{-1/2})$ and under Assumption 5.2(iii) we have

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

This proves (D.13).

To prove (D.14) we return to (D.15). By substituting (D.18) into (D.15) we have

$$\frac{dW_n(\theta)}{d\theta}\Big|_{\theta=\theta_n} + (\widehat{\theta}_n^{(W)} - \theta_n)\frac{d^2 W_n(\theta)}{d\theta^2}\Big|_{\theta=\theta_n} + \frac{1}{2}(\widehat{\theta}_n^{(W)} - \theta_n)^2\frac{d^3 W_n(\theta)}{d\theta^3}\Big|_{\theta=\theta_n} = O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).$$

This proves (D.14). □

The second order expansion (D.14) is instrumental in proving the equivalence result Theorem 5.1. By following a similar set of arguments to those in Lemma D.3 for the multivariate parameter

---

[1]The precise proof for (D.18): By using (D.17) we have

$$(\widehat{\theta}_n^{(W)} - \theta_n) = A_n + (\widehat{\theta}_n^{(W)} - \theta_n)^2 B_n + (\widehat{\theta}_n^{(W)} - \theta_n)^3 C_n$$

where the random variables A_n, B_n and C_n are such that $A_n = O_p\left(\frac{1}{n^{1/2}} + \frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right)$, $B_n = O_p\left(\frac{p^2}{n} + 1\right)$, and $C_n = O_p(1)$. Then, moving the second and third term in the RHS to LHS

$$(\widehat{\theta}_n^{(W)} - \theta_n)[1 + (\widehat{\theta}_n^{(W)} - \theta_n)B_n + (\widehat{\theta}_n^{(W)} - \theta_n)^2 C_n] = O_p\left(\frac{1}{n^{1/2}} + \frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).$$

Using consistency, $(\widehat{\theta}_n^{(W)} - \theta_n)B_n = o_p(1)O_p\left(\frac{p^2}{n} + 1\right) = o_p(\frac{p^2}{n} + 1)$ and $(\widehat{\theta}_n^{(W)} - \theta_n)^2 C_n = o_p(1)$. Finally, we use that $(1 + o_p(1))^{-1} = O_p(1)$, then,

$$(\widehat{\theta}_n^{(W)} - \theta_n) = O_p(1)O_p\left(\frac{1}{n^{1/2}} + \frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).$$

Thus giving the required probabilistic rate.

$\theta = (\theta_1, \ldots, \theta_d)$, the feasible estimator satisfies the expansion

$$\frac{\partial W_n(\theta)}{\partial \theta_r} + \sum_{s=1}^d (\widehat{\theta}_{s,n}^{(W)} - \theta_{s,n}) \frac{\partial^2 W_n(\theta)}{\partial \theta_s \partial \theta_r} \Big|_{\theta=\theta_n} +$$

$$\frac{1}{2} \sum_{s_1,s_2=1}^d (\widehat{\theta}_{s_1,n}^{(W)} - \theta_{s_1,n})(\widehat{\theta}_{s_2,n}^{(W)} - \theta_{s_2,n}) \frac{\partial^3 W_n(\theta)}{\partial \theta_{s_1} \partial \theta_{s_2} \partial \theta_r} \Big|_{\theta=\theta_n} = \left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}} \right). \quad \text{(D.19)}$$

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

$$(\widehat{\theta}_n^{(H)} - \theta_n) = - \left[\mathbb{E} \left[\frac{d^2 H_n(\theta_n)}{d\theta_n^2} \right] \right]^{-1} \frac{d H_n(\theta_n)}{d\theta_n} + O_p \left(\frac{1}{n} + \frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}} \right) \quad \text{(D.20)}$$

and

$$\frac{dH_n(\theta)}{d\theta} \Big|_{\theta=\theta_n} + (\widehat{\theta}_n^{(W)} - \theta_n) \frac{d^2 H_n(\theta)}{d\theta^2} \Big|_{\theta=\theta_n} + \frac{1}{2} (\widehat{\theta}_n^{(H)} - \theta_n)^2 \frac{d^3 H_n(\theta)}{d\theta^3} \Big|_{\theta=\theta_n}$$

$$= \left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}} \right). \quad \text{(D.21)}$$

PROOF of Theorem 5.1. We first prove the result for the one parameter case when $p \geq 1$. By using (D.14) for the feasible estimator $\widehat{\theta}_n^{(W)} = \arg\min \widehat{W}_{p,n}(\theta)$ we have

$$\frac{dW_n(\theta)}{d\theta} \Big|_{\theta=\theta_n} + (\widehat{\theta}_n^{(W)} - \theta_n) \frac{d^2 W_n(\theta)}{d\theta^2} \Big|_{\theta=\theta_n} + \frac{1}{2} (\widehat{\theta}_n^{(W)} - \theta_n)^2 \frac{d^3 W_n(\theta)}{d\theta^3} \Big|_{\theta=\theta_n} = O_p \left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}} \right).$$

Whereas for the infeasible estimator $\widetilde{\theta}_n^{(W)} = \arg\min \widehat{W}_{p,n}(\theta)$ we have

$$\frac{dW_n(\theta)}{d\theta} \Big|_{\theta=\theta_n} + (\widetilde{\theta}_n^{(W)} - \theta_n) \frac{d^2 W_n(\theta)}{d\theta^2} \Big|_{\theta=\theta_n} + \frac{1}{2} (\widetilde{\theta}_n^{(W)} - \theta_n)^2 \frac{d^3 W_n(\theta)}{d\theta^3} \Big|_{\theta=\theta_n} = O_p \left(\frac{1}{n^{3/2}} \right).$$

Taking differences for the two expansions above we have

$$(\widetilde{\theta}_n^{(W)} - \widehat{\theta}_n^{(W)}) \frac{d^2 W_n(\theta)}{d\theta^2} \Big|_{\theta=\theta_n} + \frac{1}{2} (\widetilde{\theta}_n^{(W)} - \widehat{\theta}_n^{(W)}) \left[(\widetilde{\theta}_n^{(W)} - \theta_n) + (\widehat{\theta}_n^{(W)} - \theta_n) \right] \frac{d^3 W_n(\theta)}{d\theta^3} \Big|_{\theta=\theta_n}$$

$$= O_p \left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}} \right).$$

Now replacing $\frac{d^2 W_n(\theta)}{d\theta^2} \Big|_{\theta=\theta_n}$ 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

$$(\widetilde{\theta}_n^{(W)} - \widehat{\theta}_n^{(W)}) \mathbb{E} \left(\frac{d^2 W_n(\theta)}{d\theta^2} \Big|_{\theta=\theta_n} \right) + o_p(1)(\widetilde{\theta}_n^{(W)} - \widehat{\theta}_n^{(W)}) = O_p \left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}} \right).$$

Since $\mathbb{E}\left(\frac{d^2 W_n(\theta)}{d\theta^2}\big|_{\theta=\theta_n}\right)$ 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

$$\frac{dK_n(\theta)}{d\theta}\big|_{\theta=\theta_n} + (\widehat{\theta}_n^{(K)} - \theta_n)\frac{d^2 K_n(\theta)}{d\theta^2}\big|_{\theta=\theta_n} + \frac{1}{2}(\widehat{\theta}_n^{(K)} - \theta_n)^2\frac{d^3 K_n(\theta)}{d\theta^3}\big|_{\theta=\theta_n} = O_p\left(\frac{1}{n^{3/2}}\right)$$

and

$$\frac{dW_n(\theta)}{d\theta}\big|_{\theta=\theta_n} + (\widetilde{\theta}_n^{(W)} - \theta_n)\frac{d^2 W_n(\theta)}{d\theta^2}\big|_{\theta=\theta_n} + \frac{1}{2}(\widetilde{\theta}_n^{(W)} - \theta_n)^2\frac{d^3 W_n(\theta)}{d\theta^3}\big|_{\theta=\theta_n} = O_p\left(\frac{1}{n^{3/2}}\right).$$

Taking differences of the two expansions

$$\frac{dC_n(\theta)}{d\theta}\big|_{\theta=\theta_n} + (\widehat{\theta}_n^{(K)} - \widetilde{\theta}_n^{(W)})\frac{d^2 K_n(\theta)}{d\theta^2}\big|_{\theta=\theta_n} - (\widetilde{\theta}_n^{(W)} - \theta_n)\frac{d^2 C_n(\theta)}{d\theta^2}\big|_{\theta=\theta_n}$$
$$+\frac{1}{2}(\widehat{\theta}_n^{(K)} - \widetilde{\theta}_n^{(W)})\left[(\widehat{\theta}_n^{(K)} - \theta_n) + (\widetilde{\theta}_n^{(W)} - \theta_n)\right]\frac{d^3 K_n(\theta)}{d\theta^3}\big|_{\theta=\theta_n}$$
$$-\frac{1}{2}(\widetilde{\theta}_n^{(W)} - \theta_n)^2\frac{d^3 C_n(\theta)}{d\theta^3}\big|_{\theta=\theta_n} = O_p\left(\frac{1}{n^{3/2}}\right). \tag{D.22}$$

To bound the above we use that

$$|\widehat{\theta}_n^{(K)} - \theta_n| = O_p(n^{-1/2}) \quad \text{and} \quad |\widetilde{\theta}_n^{(W)} - \theta_n| = O_p(n^{-1/2}).$$

In addition by using a proof analogous to the proves of Theorem 3.2, equation (3.10) we have

$$\frac{d^s C_n(\theta)}{d\theta^s} = \frac{1}{n}\sum_{k=1}^{n}\widehat{J}_n(\omega_{k,n}; f)\overline{J_n(\omega_{k,n})}\frac{d^s}{d\theta^s}f_\theta(\omega_{k,n})^{-1} = O_p(n^{-1}) \quad \text{for} \quad 0 \le s \le 3.$$

Substituting the above bounds into (D.22) 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 (D.19) and the same argument to give

$$\sum_{s=1}^{d}(\widehat{\theta}_{s,n}^{(W)} - \widetilde{\theta}_{s,n}^{(W)})\frac{\partial^2 W_n(\theta)}{\partial\theta_s\partial\theta_r}\rfloor_{\theta=\theta_n} +$$

$$\frac{1}{2}\sum_{s_1,s_2=1}^{d}\left[(\widehat{\theta}_{s_1,n}^{(W)} - \widetilde{\theta}_{s_1,n}^{(W)})(\widehat{\theta}_{s_2,n}^{(W)} - \theta_{s_2,n}) + (\widehat{\theta}_{s_2,n}^{(W)} - \widetilde{\theta}_{s_2,n}^{(W)})(\widehat{\theta}_{s_1,n}^{(W)} - \theta_{s_1,n})\right] \times \frac{\partial^3 W_n(\theta)}{\partial\theta_{s_1}\partial\theta_{s_2}\partial\theta_r}\rfloor_{\theta=\theta_n}$$

$$= O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right).$$

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 $\mathbb{E}\left[\nabla_\theta^2 W_n(\theta)\rfloor_{\theta=\theta_n}\right]$ 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. $\square$

E   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).

E.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

$$
\begin{aligned}
V(g, h) &= \frac{2}{2\pi} \int_0^{2\pi} g(\omega) h(\omega) f(\omega)^2 d\omega \\
&+ \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} g(\omega_1) h(\omega_2) f_4(\omega_1, -\omega_1, \omega_2) d\omega_1 d\omega_2,
\end{aligned}
\tag{E.1}
$$

where f_4 denotes the fourth order cumulant density of the time series $\{X_t\}$. Further, we define

$$
\begin{aligned}
B_{G,n}(\theta) &= \operatorname{Re} \frac{2}{n} \sum_{s,t=1}^n c(s-t) \frac{1}{n} \sum_{k=1}^n e^{-is\omega_{k,n}} \frac{d}{d\theta} \left[\overline{\phi(\omega_{k,n}; f_\theta)} \phi_t^\infty(\omega_{k,n}; f_\theta) \right] \\
B_{K,n}(\theta) &= \frac{1}{n} \sum_{k=1}^n f_n(\omega_{k,n}) \frac{df_\theta(\omega_{k,n})^{-1}}{d\theta}
\end{aligned}
$$

where $\phi(\omega; f_\theta)$ and $\phi_t^\infty(\omega; f_\theta)$ are defined in Section 3, $c(r) = \operatorname{cov}(X_0, X_r)$ and $f_n(\omega_k) = \int F_n(\omega - \lambda) f(\lambda) d\lambda$ and $F_n(\cdot)$ is the Fejér kernel of order n.

Theorem E.1 *Suppose that the parametric spectral densities $\{f_\theta; \theta \in \Theta\}$ satisfy Assumptions 5.1. Suppose the underlying time series $\{X_t\}$ is a stationary time series with spectral density f and satisfies Assumption 5.2. Let $\widehat{\theta}_n^{(G)}$, $\widehat{\theta}_n^{(K)}$ and $\widehat{\theta}_n^{(W)}$, and $\widehat{\theta}_n^{(H)}$ be defined as in (D.5). Then the asymptotic bias is*

$$
\begin{aligned}
\mathbb{E}_\theta[\widehat{\theta}_n^{(G)} - \theta_n] &= I(\theta)^{-1} \left(B_{K,n}(\theta_n) + B_{G,n}(\theta_n) \right) + n^{-1} G(\theta_n) + O(n^{-3/2}) \\
\mathbb{E}_\theta[\widehat{\theta}_n^{(K)} - \theta_n] &= I(\theta)^{-1} B_{K,n}(\theta_n) + n^{-1} G(\theta_n) + O(n^{-3/2}) \\
\mathbb{E}_\theta[\widehat{\theta}_n^{(W)} - \theta_n] &= n^{-1} G(\theta_n) + O\left(p^3 n^{-3/2} + n^{-1} p^{-K+1} \right) \\
\text{and} \quad \mathbb{E}_\theta[\widehat{\theta}_n^{(H)} - \theta_n] &= \frac{H_{2,n}}{H_{1,n}^2} G(\theta_n) + O\left(p^3 n^{-3/2} + n^{-1} p^{-K+1} \right)
\end{aligned}
$$

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 (E.1).

PROOF. See Supplementary E.3. □

Remark E.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 σ^2 and κ_4 is the 2nd and 4th order cumulant of the innovation in the model.*

Furthermore, in the case the model is correct specification and linear, we can show that Assumption 5.1(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.

E.2   The bias for the AR(1) model

In general, it is difficult to obtain a simple expression for the bias defined in Theorem E.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 θ 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 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 $O(p^3 n^{-3/2})$ (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

$$I(\theta) = -\frac{1}{2\pi}\int_0^{2\pi}\frac{d^2 f_\theta(\omega)^{-1}}{d\theta^2}f(\omega)d\omega = -\frac{1}{\pi\sigma^2}\int_0^{2\pi}f(\omega)d\omega = -\frac{2}{\sigma^2}c(0). \tag{E.2}$$

Next we calculate $B_{G,n}$, since $\phi(\omega) = 1 - \theta e^{-i\omega}$ it is is easy to show

$$\phi_1^\infty(\omega) = \theta \quad\text{and}\quad \phi_j^\infty(\omega) = 0 \text{ for } j \geq 2.$$

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,

$$
B_{G,n}(\theta)=\frac{2\sigma^{-2}}{n}c(n-1)-\frac{4\sigma^{-2}\theta}{n}c(0). \tag{E.3}
$$

To calculate $B_{K,n}$ we have

$$
\begin{aligned}
B_{K,n}(\theta) &= \frac{1}{n}\sum_{k=1}^{n}f_n(\omega_{k,n})\frac{df_\theta(\omega_{k,n})^{-1}}{d\theta}\\
&= \frac{2\sigma^{-2}}{n}\sum_{k=1}^{n}f_n(\omega_{k,n})\left(\theta-\cos(\omega_{k,n})\right)\\
&= 2\sigma^{-2}\left[\theta c(0)-\left(\frac{n-1}{n}\right)c(1)-\frac{1}{n}c(1-n)\right]\\
&= \frac{2\sigma^{-2}}{n}\left[c(1)-c(n-1)\right]. \tag{E.4}
\end{aligned}
$$

Altogether this gives

$$
\begin{aligned}
I(\theta)^{-1}\left(B_{K,n}(\theta)+B_{G,n}(\theta)\right) &= -\frac{\sigma^2}{2nc(0)}\left(2\sigma^{-2}\left[c(1)-c(n-1)\right]+2\sigma^{-2}c(n-1)-4\sigma^{-2}\theta c(0)\right)\\
&= -\frac{1}{nc(0)}\left(c(1)-2\theta c(0)\right)=\frac{\theta}{n} \tag{E.5}
\end{aligned}
$$

and

$$
I(\theta)^{-1}B_{K,n}(\theta)=-\frac{1}{nc(0)}\left[c(1)-c(n-1)\right]=-\frac{1}{n}(\theta-\theta^{n-1}). \tag{E.6}
$$

Next, we calculate $G(\theta)$. Since the third derivative of f_θ^{-1} with respect to θ 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

$$
\begin{aligned}
V \left(\frac{d}{d\theta} f_\theta^{-1}, \frac{d^2}{d\theta^2} f_\theta^{-1} \right) &= \frac{1}{\pi} \int_0^{2\pi} \left(\frac{2\theta - 2\cos(\omega)}{\sigma^2} \right) \left(\frac{2}{\sigma^2} \right) f(\omega)^2 d\omega \\
&= \frac{4}{\sigma^4} \frac{1}{\pi} \int_0^{2\pi} [\theta - \cos(\omega)] f(\omega)^2 d\omega \\
&= \frac{8}{\sigma^4} \left(\theta c_2(0) - c_2(1) \right)
\end{aligned}
$$

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

$$G(\theta) = \frac{\sigma^4}{4c(0)^2} \frac{8}{\sigma^4} \left(\theta c_2(0) - c_2(1) \right) = \frac{2}{c(0)^2} \left(\theta c_2(0) - c_2(1) \right). \tag{E.7}$$

Putting (E.7) with (E.5) 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 $A_nR_n = B_n$, 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,

$$\mathbb{E}[\widetilde{\theta}^G - \theta] \approx I(\theta)^{-1}\left(B_{K,n}(\theta) + B_{G,n}(\theta) + \frac{d}{d\theta}n^{-1}\log|\Gamma_n(f_\theta)|\right) + n^{-1}G(\theta)$$

$$= I(\theta)^{-1}\frac{1}{n}(2\sigma^{-2}c(1) - 4\sigma^{-2}\theta c(0) + 2\sigma^{-2}c(1)) - \frac{2\theta}{n} = -\frac{2\theta}{n},$$

which proves the results. □

E.3 Proof of Theorem E.1

In Theorem 5.1 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

$$(\widehat{\theta} - \theta) \approx U(\theta)^{-1}\frac{dL_n(\theta)}{d\theta}, \tag{E.8}$$

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

$$\mathbb{E}\left[\frac{dL_n(\theta)}{d\theta}\right] + \mathbb{E}\left[(\widehat{\theta} - \theta)\frac{d^2L_n(\theta)}{d\theta^2}\right] + \frac{1}{2}\mathbb{E}\left[(\widehat{\theta} - \theta)^2\frac{d^3L_n(\theta)}{d\theta^3}\right]$$

$$= \mathbb{E}\left[\frac{dL_n(\theta)}{d\theta}\right] + \mathbb{E}\left[(\widehat{\theta} - \theta)\right]\mathbb{E}\left[\frac{d^2L_n(\theta)}{d\theta^2}\right] + \text{cov}\left[(\widehat{\theta} - \theta), \frac{d^2L_n(\theta)}{d\theta^2}\right]$$

$$+ \frac{1}{2}\mathbb{E}\left[(\widehat{\theta} - \theta)^2\right]\mathbb{E}\left[\frac{d^3L_n(\theta)}{d\theta^3}\right] + \frac{1}{2}\text{cov}\left[(\widehat{\theta} - \theta)^2, \frac{d^3L_n(\theta)}{d\theta^3}\right].$$

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

$$\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right) - U(\theta)\mathbb{E}(\widehat{\theta} - \theta) + U(\theta)^{-1}\text{cov}\left(\frac{dL_n(\theta)}{d\theta}, \frac{d^2L_n(\theta)}{d\theta^2}\right)$$

$$+ 2^{-1}U(\theta)^{-2}\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right)^2\mathbb{E}\left(\frac{d^3L_n(\theta)}{d\theta^3}\right)$$

$$+ 2^{-1}U(\theta)^{-2}\text{cov}\left(\left(\frac{dL_n(\theta)}{d\theta}\right)^2, \frac{d^3L_n(\theta)}{d\theta^3}\right) \approx 0.$$

Using the above to solve for $\mathbb{E}(\widehat{\theta} - \theta)$ gives

$$\begin{aligned}
\mathbb{E}(\widehat{\theta} - \theta) &= U(\theta)^{-1}\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right) + U(\theta)^{-2}\text{cov}\left(\frac{dL_n(\theta)}{d\theta}, \frac{d^2L_n(\theta)}{d\theta^2}\right) \\
&\quad + 2^{-1}U(\theta)^{-3}\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right)^2\mathbb{E}\left(\frac{d^3L_n(\theta)}{d\theta^3}\right) + 2^{-1}U(\theta)^{-3}\text{cov}\left(\left(\frac{dL_n(\theta)}{d\theta}\right)^2, \frac{d^3L_n(\theta)}{d\theta^3}\right) \\
&= U(\theta)^{-1}\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right) + U(\theta)^{-2}\text{cov}\left(\frac{dL_n(\theta)}{d\theta}, \frac{d^2L_n(\theta)}{d\theta^2}\right) \\
&\quad + 2^{-1}U(\theta)^{-3}\left[\text{var}\left(\frac{dL_n(\theta)}{d\theta}\right) + \left\{\mathbb{E}\left[\frac{dL_n(\theta)}{d\theta}\right]\right\}^2\right]\mathbb{E}\left(\frac{d^3L_n(\theta)}{d\theta^3}\right) \\
&\quad + 2^{-1}U(\theta)^{-3}\text{cov}\left(\left(\frac{dL_n(\theta)}{d\theta}\right)^2, \frac{d^3L_n(\theta)}{d\theta^3}\right).
\end{aligned}$$

Thus

$$\mathbb{E}(\widehat{\theta} - \theta) = I_0 + I_1 + I_2 + I_3 + I_4 \tag{E.9}$$

where

$$I_0 = U(\theta)^{-1}\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right)$$

$$I_1 = U(\theta)^{-2}\text{cov}\left(\frac{dL_n(\theta)}{d\theta}, \frac{d^2L_n(\theta)}{d\theta^2}\right)$$

$$I_2 = 2^{-1}U(\theta)^{-3}\text{var}\left(\frac{dL_n(\theta)}{d\theta}\right)\mathbb{E}\left(\frac{d^3L_n(\theta)}{d\theta^3}\right)$$

$$I_3 = 2^{-1}U(\theta)^{-3}\left\{\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right)\right\}^2\mathbb{E}\left(\frac{d^3L_n(\theta)}{d\theta^3}\right)$$

$$I_4 = 2^{-1}U(\theta)^{-3}\text{cov}\left(\left(\frac{dL_n(\theta)}{d\theta}\right)^2, \frac{d^3L_n(\theta)}{d\theta^3}\right).$$

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.

<u>The first derivative</u> 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 (C.33) we have

$$\mathbb{E}\left[\frac{d}{d\theta}\mathcal{L}_n(\theta)\right] = \mathbb{E}\left[\frac{d}{d\theta}K_n(\theta)\right] + \mathbb{E}\left[n^{-1}\underline{X}'_n F_n^* \frac{d}{d\theta}\Delta_n(f_\theta^{-1})D_{\infty,n}(f_\theta)\underline{X}_n\right]$$

$$+ \mathbb{E}\left[n^{-1}\underline{X}'_n F_n^* \frac{d}{d\theta}\Delta_n\left(f_\theta^{-1}\right)\left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right)\underline{X}_n\right]$$

$$= \mathbb{E}\left[\frac{d}{d\theta}K_n(\theta)\right] + n^{-1}\sum_{s,t=1}^n c(s-t)\frac{1}{n}\sum_{k=1}^n e^{-is\omega_{k,n}}\frac{d}{d\theta}\varphi_{t,n}(\omega_{k,n}; f_\theta)$$

$$+ n^{-1}\mathbb{E}\left[\underline{X}'_n F_n^* \frac{d}{d\theta}\Delta_n\left(f_\theta^{-1}\right)\left(D_n(f_\theta) - D_{\infty,n}(f_\theta)\right)\underline{X}_n\right]$$

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

$$
n^{-1} \sum_{s,t=1}^{n} c(s-t) \frac{1}{n} \sum_{k=1}^{n} e^{-is\omega_{k,n}} \frac{d}{d\theta} \varphi_{t,n}(\omega_{k,n}; f_\theta)
$$

$$
= n^{-1} \sum_{s,t=1}^{n} c(s-t) \frac{1}{n} \sum_{k=1}^{n} \frac{d}{d\theta} \left[e^{-is\omega_{k,n}} \overline{\phi(\omega_{k,n}; f_\theta)} \phi_t^\infty(\omega_{k,n}; f_\theta) + e^{-i(s-1)\omega_{k,n}} \phi(\omega_{k,n}; f_\theta) \overline{\phi_{n+1-t}^\infty(\omega_{k,n}; f_\theta)} \right]
$$

$$
= n^{-1} \sum_{s,t=1}^{n} c(s-t) \frac{1}{n} \sum_{k=1}^{n} e^{-is\omega_{k,n}} \frac{d}{d\theta} \overline{\phi(\omega_{k,n}; f_\theta)} \phi_t^\infty(\omega_{k,n}; f_\theta)
$$

$$
+ n^{-1} \sum_{s,t'=1}^{n} c(s-n-1+t') \frac{1}{n} \sum_{k=1}^{n} e^{-i(s-1)\omega_{k,n}} \frac{d}{d\theta} \phi(\omega_{k,n}; f_\theta) \overline{\phi_{t'}^\infty(\omega_{k,n}; f_\theta)} \qquad (\text{let} \quad t' = n+1-t)
$$

$$
= \operatorname{Re} \frac{2}{n} \sum_{s,t=1}^{n} c(s-t) \frac{1}{n} \sum_{k=1}^{n} e^{-is\omega_{k,n}} \frac{d}{d\theta} \overline{\phi(\omega_{k,n}; f_\theta)} \phi_t^\infty(\omega_{k,n}; f_\theta) = B_{G,n}(\theta).
$$

Finally, by using Corollary C.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

$$
\mathbb{E}\left(\frac{dW_n(\theta)}{d\theta} \right) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[\widetilde{J}_n(\omega_{k,n}; f) \overline{J_n(\omega_{k,n})}] \frac{d}{d\theta} f_\theta(\omega_{k,n})^{-1}
$$

$$
= \frac{1}{n} \sum_{k=1}^{n} f(\omega_{k,n}) \frac{d}{d\theta} f_\theta(\omega_{k,n})^{-1}.
$$

Finally, the analysis of $H_n(\theta)$ is identical to the analysis of $W_n(\theta)$ and we obtain

$$
\mathbb{E}\left(\frac{dH_n(\theta)}{d\theta} \right) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[\widetilde{J}_n(\omega_{k,n}; f) \overline{J_{\underline{h}_n,n}(\omega_{k,n})}] \frac{d}{d\theta} f_\theta(\omega_{k,n})^{-1}
$$

$$
= \frac{1}{n} \sum_{k=1}^{n} f(\omega_{k,n}) \frac{d}{d\theta} f_\theta(\omega_{k,n})^{-1}.
$$

In summary, evaluating the above at the best fitting parameter θ_n and by Assumption 5.1(ii) gives

$$\mathbb{E}\left(\frac{dK_n(\theta)}{d\theta}\right)\big|_{\theta=\theta_n} = B_{K,n}(\theta_n)$$

$$\mathbb{E}\left(\frac{d\mathcal{L}_n(\theta)}{d\theta}\right)\big|_{\theta=\theta_n} = B_{K,n}(\theta_n) + B_{G,n}(\theta_n) + O(n^{-K+1/2})$$

$$\text{and} \quad \mathbb{E}\left(\frac{dW_n(\theta)}{d\theta}\right)\big|_{\theta=\theta_n} = \mathbb{E}\left(\frac{dH_n(\theta)}{d\theta}\right)\big|_{\theta=\theta_n} = 0. \qquad (E.10)$$

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).

<u>The second and third order derivatives</u> 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 derivatives of the Gaussian likelihood in terms of the Whittle likelihood

$$\frac{d^i \mathcal{L}_n(\theta)}{d\theta^i} = \frac{d^i K_n(\theta)}{d\theta^i} + \mathbb{E}\left[n^{-1}\underline{X}'_n F_n^* \frac{d^i}{d\theta^i}\Delta_n(f_\theta^{-1})D_n(f_\theta)\underline{X}_n\right].$$

By using (C.36), for $1 \leq i \leq 3$ we have

$$\left\|n^{-1}\underline{X}'_n F_n^* \frac{d^i}{d\theta^i}\Delta_n(f_\theta^{-1})D_n(f_\theta)\underline{X}_n\right\|_{\mathbb{E},1} = O(n^{-1}). \qquad (E.11)$$

Similarly, we represent the derivatives of $W_n(\theta)$ and $H_{p,n}(\theta)$ in terms of the derivatives of $K_n(\theta)$

$$\frac{d^i W_n(\theta)}{d\theta^i} = \frac{d^i K_n(\theta)}{d\theta^i} + C_{i,n}$$

$$\frac{d^i H_n(\theta)}{d\theta^i} = \frac{d^i K_{n,\underline{h}_n}(\theta)}{d\theta^i} + D_{i,n}$$

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

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

$$D_{i,n} = \frac{1}{n}\sum_{k=1}^n \frac{d^i}{d\theta^i}\frac{\widehat{J}_n(\omega_{k,n};f)\overline{J_{n,\underline{h}_n}(\omega_{k,n})}}{f_\theta(\omega_{k,n})} = \frac{1}{n}\underline{X}'_n H_n F_n^*\Delta_n(\frac{d^i}{d\theta^i}f_\theta^{-1})D_n(f)\underline{X}_n,$$

where $H_n = \text{diag}(h_{1,n}, \ldots, h_{n,n})$. 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

$$\left\| F_n^* \Delta_n \left(\frac{d^i}{d\theta^i} f_\theta^{-1} \right) D_n(f) \right\|_1$$

$$\leq \left\| F_n^* \Delta_n \left(\frac{d^i}{d\theta^i} f_\theta^{-1} \right) [D_n(f) - D_{\infty,n}(f)] \right\|_1 + \left\| F_n^* \Delta_n \left(\frac{d^i}{d\theta^i} f_\theta^{-1} \right) D_{\infty,n}(f) \right\|_1.$$

We use a similar method to the proof of Theorem 3.1, equation (3.5) and Theorem 3.2, equation (3.8) 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 5.1 and 5.2. By using the proof of Theorem 3.1, equation (3.5), 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 3.2, equation (3.8) 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

$$(\mathbb{E}|C_{i,n}|^2)^{1/2} = n^{-1} \left\| \underline{X}_n' F_n^* \Delta_n \left(\frac{d^i}{d\theta^i} f_\theta^{-1} \right) D_n(f) \underline{X}_n \right\|_{\mathbb{E},2} = O(n^{-1}). \tag{E.12}$$

For the hybrid likelihood, we use that $\sup_{t,n} |h_{t,n}| < \infty$, this gives

$$\|H_n F_n^* \Delta_n \left(\frac{d^i}{d\theta^i} f_\theta^{-1} \right) D_n(f)\|_1 \leq (\sup_t h_{t,n}) \times \|F_n^* \Delta_n \left(\frac{d^i}{d\theta^i} f_\theta^{-1} \right) D_n(f)\|_1 = O(1).$$

Therefore, under the condition that $\{h_{t,n}\}$ is a bounded sequence

$$(\mathbb{E}|D_{i,n}|^2)^{1/2} = n^{-1} \left\| H_n \underline{X}_n' F_n^* \Delta_n \left(\frac{d^i}{d\theta^i} f_\theta^{-1} \right) D_n(f) \underline{X}_n \right\|_{\mathbb{E},2} = O\left(n^{-1} \right). \tag{E.13}$$

Thus the expectations of the derivatives are

$$\mathbb{E} \left(\frac{d^i \mathcal{L}_n(\theta)}{d\theta^i} \right) = \mathbb{E} \left(\frac{d^i K_n(\theta)}{d\theta^i} \right) + O(n^{-1})$$

$$\mathbb{E} \left(\frac{d^i W_n(\theta)}{d\theta^i} \right) = \mathbb{E} \left(\frac{d^i K_n(\theta)}{d\theta^i} \right) + O(n^{-1})$$

$$\mathbb{E} \left(\frac{d^i H_n(\theta)}{d\theta^i} \right) = \mathbb{E} \left(\frac{d^i K_n(\theta)}{d\theta^i} \right) + O\left(n^{-1} \right).$$

This gives the expectation of the second and third derivatives of all likelihoods in terms of $I(\theta)$ and $J(\frac{d^3 f_\theta^{-1}}{d\theta^3})$:

$$\mathbb{E} \left(\frac{d^2 L_n(\theta)}{d\theta^2} \right) = -I(\theta) + O(n^{-1}), \quad \text{and} \quad \mathbb{E} \left(\frac{d^3 L_n(\theta)}{d\theta^3} \right) = J(\frac{d^3 f_\theta^{-1}}{d\theta^3}) + O(n^{-1}).$$

Bounds for the covariances between the derivatives The terms I_1, I_2 and I_4 all contain the co-

variance between various likelihoods and its derivatives. Thus to obtain expression and bounds for these terms we use that

$$\text{var}\left(\frac{d^i}{d\theta^i}K_n(\theta)\right) = O(n^{-1}), \tag{E.14}$$

where the above can be proved using Brillinger (2001), 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

$$\text{var}\left(\frac{d^i}{d\theta^i}K_{n,\underline{h}_n}(\theta)\right) = O\left(\frac{H_{2,n}}{H_{1,n}^2}\right). \tag{E.15}$$

By using (E.11), (E.12), and (E.14) we have

$$\begin{aligned}
\text{cov}\left(\frac{d\mathcal{L}_n(\theta)}{d\theta}, \frac{d^2\mathcal{L}_n(\theta)}{d\theta^2}\right) &= \text{cov}\left(\frac{dK_n(\theta)}{d\theta}, \frac{d^2K_n(\theta)}{d\theta^2}\right) + O(n^{-3/2}) \\
\text{cov}\left(\frac{dW_n(\theta)}{d\theta}, \frac{d^2W_n(\theta)}{d\theta^2}\right) &= \text{cov}\left(\frac{dK_n(\theta)}{d\theta}, \frac{d^2K_n(\theta)}{d\theta^2}\right) + O(n^{-3/2}) \\
\text{var}\left(\frac{d\mathcal{L}_n(\theta)}{d\theta}\right) &= \text{var}\left(\frac{dK_n(\theta)}{d\theta}\right) + O(n^{-3/2}) \\
\text{var}\left(\frac{dW_n(\theta)}{d\theta}\right) &= \text{var}\left(\frac{dK_n(\theta)}{d\theta}\right) + O(n^{-3/2}).
\end{aligned}$$

For the hybrid Whittle likelihood, by using (E.13) and (E.15)

$$\begin{aligned}
\text{cov}\left(\frac{dH_n(\theta)}{d\theta}, \frac{d^2H_n(\theta)}{d\theta^2}\right) &= \text{cov}\left(\frac{dK_{n,\underline{h}_n}(\theta)}{d\theta}, \frac{d^2K_{n,\underline{h}_n}(\theta)}{d\theta^2}\right) + O\left(\frac{H_{2,n}^{1/2}}{nH_{1,n}}\right) \\
\text{var}\left(\frac{dH_n(\theta)}{d\theta}\right) &= \text{var}\left(\frac{dK_{n,\underline{h}_n}(\theta)}{d\theta}\right) + O\left(\frac{H_{2,n}^{1/2}}{nH_{1,n}}\right).
\end{aligned}$$

Using that $H_{2,n}/H_{1,n}^2 \sim n^{-1}$, 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 (2001), Theorem 4.3.2 and well-known cumulant arguments we can show that

$$\begin{aligned}
\text{cov}\left(\frac{dK_n(\theta)}{d\theta}, \frac{d^2K_n(\theta)}{d\theta^2}\right) &= n^{-1}V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{d^2f_\theta^{-1}}{d\theta^2}\right) + O(n^{-2}) \\
\text{and } \text{var}\left(\frac{dK_n(\theta)}{d\theta}\right) &= n^{-1}V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{df_\theta^{-1}}{d\theta}\right) + O(n^{-2}).
\end{aligned}$$

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

$$\text{cov}\left(\frac{dK_{n,\underline{h}_n}(\theta)}{d\theta}, \frac{d^2 K_{n,\underline{h}_n}(\theta)}{d\theta^2}\right) = \frac{H_{2,n}}{H_{1,n}^2} V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{d^2 f_\theta^{-1}}{d\theta^2}\right) + O\left(\frac{H_{2,n}}{nH_{1,n}^2}\right)$$

$$\text{and var}\left(\frac{dK_{n,\underline{h}_n}(\theta)}{d\theta}\right) = \frac{H_{2,n}}{H_{1,n}^2} V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{df_\theta^{-1}}{d\theta}\right) + O\left(\frac{H_{2,n}}{nH_{1,n}^2}\right).$$

These results yield expressions for I_1 and I_2 (we obtain these below).

<u>Expression for I_0 and a bound for I_3.</u> Using the results above we have

- The Gaussian likelihood

$$I_0 = I(\theta_n)^{-1}\left[B_{K,n}(\theta_n) + B_{G,n}(\theta_n)\right] + O(n^{-2}) \tag{E.16}$$

- 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 $\mathbb{E}[\frac{dL_n(\theta)}{d\theta}|_{\theta=\theta_n}] = O(n^{-1})$, this implies that for all the likelihoods the term I_3 is

$$I_3 = 2^{-1}U(\theta)^{-3}\left\{\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right)\right\}^2 \mathbb{E}\left(\frac{d^3 L_n(\theta)}{d\theta^3}\right) = O(n^{-2}).$$

<u>Expression for I_1 and I_2.</u> For the Gaussian, Whittle, and boundary corrected Whittle likelihoods we have

$$I_1 = n^{-1} I(\theta_n)^{-2} V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{d^2 f_\theta^{-1}}{d\theta^2}\right) + O(n^{-3/2})$$

$$I_2 = n^{-1} 2^{-1} I(\theta_n)^{-3} V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{df_\theta^{-1}}{d\theta}\right) J\left(\frac{d^3 f_\theta^{-1}}{d\theta^3}\right) + O(n^{-3/2}).$$

For the hybrid Whittle likelihood we obtain a similar expression

$$I_1 = \frac{H_{2,n}}{H_{1,n}^2} I(\theta_n)^{-2} V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{d^2 f_\theta^{-1}}{d\theta^2}\right) + O\left(n^{-3/2}\right)$$

$$I_2 = \frac{H_{2,n}}{H_{1,n}^2} 2^{-1} I(\theta_n)^{-3} V\left(\frac{df_\theta^{-1}}{d\theta}, \frac{df_\theta^{-1}}{d\theta}\right) J\left(\frac{d^3 f_\theta^{-1}}{d\theta^3}\right) + O\left(n^{-3/2}\right).$$

<u>A bound for I_4</u> 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

$$I_4 = 2^{-1} U(\theta)^{-3} \mathrm{cov}\left(\left(\frac{dL_n(\theta)}{d\theta}\right)^2, \frac{d^3 L_n(\theta)}{d\theta^3}\right).$$

To bound the above we focus on $\mathrm{cov}\left(\left(\frac{dL_n(\theta)}{d\theta}\right)^2, \frac{d^3 L_n(\theta)}{d\theta^3}\right)$. By using indecomposable partitions we have

$$\mathrm{cov}\left(\left(\frac{dL_n(\theta)}{d\theta}\right)^2, \frac{d^3 L_n(\theta)}{d\theta^3}\right) = 2\mathrm{cov}\left(\frac{dL_n(\theta)}{d\theta}, \frac{d^3 L_n(\theta)}{d\theta^3}\right) \mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right)$$

$$+ \mathrm{cum}\left(\frac{dL_n(\theta)}{d\theta}, \frac{dL_n(\theta)}{d\theta} \frac{d^3 L_n(\theta)}{d\theta^3}\right)$$

$$+ \left[\mathbb{E}\left(\frac{dL_n(\theta)}{d\theta}\right)\right]^2 \mathbb{E}\left(\frac{d^3 L_n(\theta)}{d\theta^3}\right).$$

We use (E.14), (E.11) and (E.12) 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 (2001), Theorem 4.3.2 we have that for the non-hybrid likelihoods

$$I_4 = O(n^{-2})$$

and for the hybrid Whittle likelihood

$$I_4 = O\left(\frac{H_{2,n}}{n H_{1,n}^2}\right).$$

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

$$I_1 + I_2 = n^{-1} \left[I(\theta_n)^{-2} V \left(\frac{df_\theta^{-1}}{d\theta}, \frac{d^2 f_\theta^{-1}}{d\theta^2} \right) + 2^{-1} I(\theta_n)^{-3} V \left(\frac{df_\theta^{-1}}{d\theta}, \frac{df_\theta^{-1}}{d\theta} \right) J \left(\frac{d^3 f_\theta^{-1}}{d\theta^3} \right) \right] + O(n^{-3/2})$$

$$= n^{-1} G(\theta_n) + O(n^{-3/2})$$

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 (E.16). This proves the result. $\qquad\square$

E.4 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 (D.4). Let $\widehat{\theta}_n^{(G)}$, $\widehat{\theta}_n^{(W)}$, $\widehat{\theta}_n^{(W)}$ and $\widehat{\theta}_n^{(H)}$ be the corresponding estimators defined in (D.5) and $\theta_n = (\theta_{1,n}, \ldots, \theta_{d,n})$ is the best fitting parameter defined as in (D.1). Then under Assumption 5.1 and 5.2 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

$$\mathbb{E}[\widehat{\theta}_{j,n}^{(W)} - \theta_{j,n}] = n^{-1} \sum_{r=1}^{d} I^{(j,r)} G_r(\theta) + O\left(p^3 n^{-3/2} + (n p^{K-1})^{-1} \right).$$

- The hybrid Whittle likelihood has bias

$$\mathbb{E}[\widehat{\theta}_{j,n}^{(H)} - \theta_{j,n}] = \frac{H_{2,n}}{H_{1,n}^2} \sum_{r=1}^{d} I^{(j,r)} G_r(\theta) + O\left(p^3 n^{-3/2} + (n p^{K-1})^{-1} \right). \qquad (E.17)$$

Where $I^{(j,r)}$, $B_{r,G,n}(\cdot)$, $B_{r,K,n}(\cdot)$, and $G_r(\cdot)$ are defined as in Section 5.

PROOF. Let $L_n(\theta)$ be the criterion and $\widehat{\theta}_n = \arg\min L_n(\theta)$ and θ_n the best fitting parameter. We use a similar technique used to prove Theorem E.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 $1 \le r \le d$ 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

$$\mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial\theta_r}\right] + \sum_{s=1}^d\mathbb{E}[\widehat{\theta}_s - \theta_s]\mathbb{E}\left[\frac{\partial^2 L_n(\theta)}{\partial\theta_s\partial\theta_r}\right] + \sum_{s=1}^d\text{cov}\left[\widehat{\theta}_s - \theta_s, \frac{\partial^2 L_n(\theta)}{\partial\theta_s\partial\theta_r}\right]$$

$$+\frac{1}{2}\sum_{s_1,s_2=1}^d\text{cov}\left(\widehat{\theta}_{s_1} - \theta_{s_1}, \widehat{\theta}_{s_2} - \theta_{s_2}\right)\mathbb{E}\left[\frac{\partial^3 L_n(\theta)}{\partial\theta_{s_1}\partial\theta_{s_2}\partial\theta_r}\right]$$

$$+\frac{1}{2}\sum_{s_1,s_2=1}^d\mathbb{E}[\widehat{\theta}_{s_1} - \theta_{s_1}]\mathbb{E}[\widehat{\theta}_{s_2} - \theta_{s_2}]\mathbb{E}\left[\frac{\partial^3 L_n(\theta)}{\partial\theta_{s_1}\partial\theta_{s_2}\partial\theta_r}\right]$$

$$+\frac{1}{2}\sum_{s_1,s_2=1}^d\text{cov}\left[(\widehat{\theta}_{s_1} - \theta_{s_1})(\widehat{\theta}_{s_2} - \theta_{s_2}), \frac{\partial^3 L_n(\theta)}{\partial\theta_{s_1}\partial\theta_{s_2}\partial\theta_r}\right] \approx 0.$$

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

$$
\mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial \theta_r}\right] - \sum_{s=1}^{d} \mathbb{E}[\widehat{\theta}_s - \theta_s]U_{s,r} + \sum_{s_1,s_2=1}^{d} U^{(s_1,s_2)}\mathrm{cov}\left[\frac{\partial L_n(\theta)}{\partial \theta_{s_2}}, \frac{\partial^2 L_n(\theta)}{\partial \theta_{s_1}\partial \theta_r}\right]
$$

$$
+\frac{1}{2}\sum_{s_1,s_2,s_3,s_4=1}^{d} U^{(s_1,s_3)}U^{(s_2,s_4)}\mathrm{cov}\left(\frac{\partial L_n(\theta)}{\partial \theta_{s_3}}, \frac{\partial L_n(\theta)}{\partial \theta_{s_4}}\right)\mathbb{E}\left[\frac{\partial^3 L_n(\theta)}{\partial \theta_{s_1}\partial \theta_{s_2}\partial \theta_r}\right]
$$

$$
+\frac{1}{2}\sum_{s_1,s_2,s_3,s_4=1}^{d} U^{(s_1,s_3)}U^{(s_2,s_4)}\mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial \theta_{s_3}}\right]\mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial \theta_{s_4}}\right]\mathbb{E}\left[\frac{\partial^3 L_n(\theta)}{\partial \theta_{s_1}\partial \theta_{s_2}\partial \theta_r}\right]
$$

$$
+\frac{1}{2}\sum_{s_1,s_2,s_3,s_4=1}^{d} U^{(s_1,s_3)}U^{(s_2,s_4)}\mathrm{cov}\left[\frac{\partial L_n(\theta)}{\partial \theta_{s_3}}\frac{\partial L_n(\theta)}{\partial \theta_{s_4}}, \frac{\partial^3 L_n(\theta)}{\partial \theta_{s_1}\partial \theta_{s_2}\partial \theta_r}\right] \approx 0,
$$

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 E.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

$$
\mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial \theta_r}\right] - \sum_{s=1}^{d} \mathbb{E}[\widehat{\theta}_s - \theta_s]U_{s,r} + \sum_{s_1,s_2=1}^{d} U^{(s_1,s_2)}\mathrm{cov}\left[\frac{\partial L_n(\theta)}{\partial \theta_{s_2}}, \frac{\partial^2 L_n(\theta)}{\partial \theta_{s_1}\partial \theta_r}\right]
$$

$$
+\frac{1}{2}\sum_{s_1,s_2,s_3,s_4=1}^{d} U^{(s_1,s_3)}U^{(s_2,s_4)}\mathrm{cov}\left(\frac{\partial L_n(\theta)}{\partial \theta_{s_3}}, \frac{\partial L_n(\theta)}{\partial \theta_{s_4}}\right)\mathbb{E}\left[\frac{\partial^3 L_n(\theta)}{\partial \theta_{s_1}\partial \theta_{s_2}\partial \theta_r}\right] \approx 0.
$$

Let

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

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

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 E.1 we have

$$
\mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial \theta_r}\right] - \sum_{s=1}^{d} I_{s,r}\mathbb{E}[\widehat{\theta}_s - \theta_s] + \frac{H_{2,n}}{H_{1,n}^2}\sum_{s_1,s_2=1}^{d} I^{(s_1,s_2)}V\left(\frac{\partial f_\theta^{-1}}{\partial \theta_{s_2}}, \frac{\partial^2 f_\theta^{-1}}{\partial \theta_{s_1}\partial \theta_r}\right)
$$

$$
+\frac{H_{2,n}}{2H_{1,n}^2}\sum_{s_1,s_2,s_3,s_4=1}^{d} I^{(s_1,s_3)}I^{(s_2,s_4)}V\left(\frac{\partial f_\theta^{-1}}{\partial \theta_{s_3}}, \frac{\partial f_\theta^{-1}}{\partial \theta_{s_4}}\right)J\left(\frac{\partial^3 f_\theta^{-1}}{\partial \theta_{s_1}\partial \theta_{s_2}\partial \theta_r}\right) \approx 0.
$$

114

Thus

$$\sum_{s=1}^{d} I_{s,r} \mathbb{E}[\widehat{\theta}_s - \theta_s] \approx \mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial \theta_r}\right] + \frac{H_{2,n}}{H_{1,n}^2} \sum_{s_1,s_2=1}^{d} I^{(s_1,s_2)} V\left(\frac{\partial f_\theta^{-1}}{\partial \theta_{s_2}}, \frac{\partial^2 f_\theta^{-1}}{\partial \theta_{s_1}\partial \theta_r}\right)$$

$$+ \frac{H_{2,n}}{2H_{1,n}^2} \sum_{s_1,s_2,s_3,s_4=1}^{d} I^{(s_1,s_3)} I^{(s_2,s_4)} V\left(\frac{\partial f_\theta^{-1}}{\partial \theta_{s_3}}, \frac{\partial f_\theta^{-1}}{\partial \theta_{s_4}}\right) J\left(\frac{\partial^3 f_\theta^{-1}}{\partial \theta_{s_1}\partial \theta_{s_2}\partial \theta_r}\right).$$

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

$$D_r \approx \mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial \theta_r}\right] + \frac{H_{2,n}}{H_{1,n}^2} \sum_{s_1,s_2=1}^{d} I^{(s_1,s_2)} V\left(\frac{\partial f_\theta(\omega)^{-1}}{\partial \theta_{s_2}}, \frac{\partial^2 f_\theta(\omega)^{-1}}{\partial \theta_{s_1}\partial \theta_r}\right)$$

$$+ \frac{H_{2,n}}{2H_{1,n}^2} \sum_{s_1,s_2,s_3,s_4=1}^{d} I^{(s_1,s_3)} I^{(s_2,s_4)} V\left(\frac{\partial f_\theta^{-1}}{\partial \theta_{s_3}}, \frac{\partial f_\theta^{-1}}{\partial \theta_{s_4}}\right) J\left(\frac{\partial^3 f_\theta^{-1}}{\partial \theta_{s_1}\partial \theta_{s_2}\partial \theta_r}\right).$$

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 $\widehat{\theta}_j$

$$\mathbb{E}[\widehat{\theta}_j - \theta_j] \approx \sum_{r=1}^{d} I^{(j,r)} \left[\mathbb{E}\left[\frac{\partial L_n(\theta)}{\partial \theta_r}\right] + \frac{H_{2,n}}{H_{1,n}^2} \sum_{s_1,s_2=1}^{d} I^{(s_1,s_2)} V\left(\frac{\partial f_\theta(\omega)^{-1}}{\partial \theta_{s_2}}, \frac{\partial^2 f_\theta(\omega)^{-1}}{\partial \theta_{s_1}\partial \theta_r}\right)\right.$$

$$\left. + \frac{H_{2,n}}{2H_{1,n}^2} \sum_{s_1,s_2,s_3,s_4=1}^{d} I^{(s_1,s_3)} I^{(s_2,s_4)} V\left(\frac{\partial f_\theta^{-1}}{\partial \theta_{s_3}}, \frac{\partial f_\theta^{-1}}{\partial \theta_{s_4}}\right) J\left(\frac{\partial^3 f_\theta^{-1}}{\partial \theta_{s_1}\partial \theta_{s_2}\partial \theta_r}\right)\right]. \quad \text{(E.18)}$$

The above is a general result. We now obtain the bias for the different criteria. Let

$$B_{r,G,n}(\theta) = \text{Re}\,\frac{2}{n}\sum_{s,t=1}^{n} c(s-t)\frac{1}{n}\sum_{k=1}^{n} e^{-is\omega_{k,n}}\frac{\partial}{\partial \theta_r}\left[\overline{\phi(\omega_{k,n}; f_\theta)}\phi_t^\infty(\omega_{k,n}; f_\theta)\right]$$

$$B_{r,K,n}(\theta) = \frac{1}{n}\sum_{k=1}^{n} f_n(\omega_{k,n})\frac{\partial f_\theta(\omega_{k,n})^{-1}}{\partial \theta_r}$$

$$\text{and} \quad G_r(\theta) = \sum_{s_1,s_2=1}^{d} I^{(s_1,s_2)} V\left(\frac{\partial f_\theta(\omega)^{-1}}{\partial \theta_{s_2}}, \frac{\partial^2 f_\theta(\omega)^{-1}}{\partial \theta_{s_1}\partial \theta_r}\right)$$

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

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 (E.18) gives the four difference biases in (E.17). Thus we have proved the result. $\qquad\square$

F Additional Simulations

F.1 Table of results for the AR(1) and MA(1) for a Gaussian time series

Likelihoods	θ									
	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 3: *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 *to denote the smallest RMSE and* blue *to denote the second smallest RMSE.*

F.2 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 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.

AR(1) model
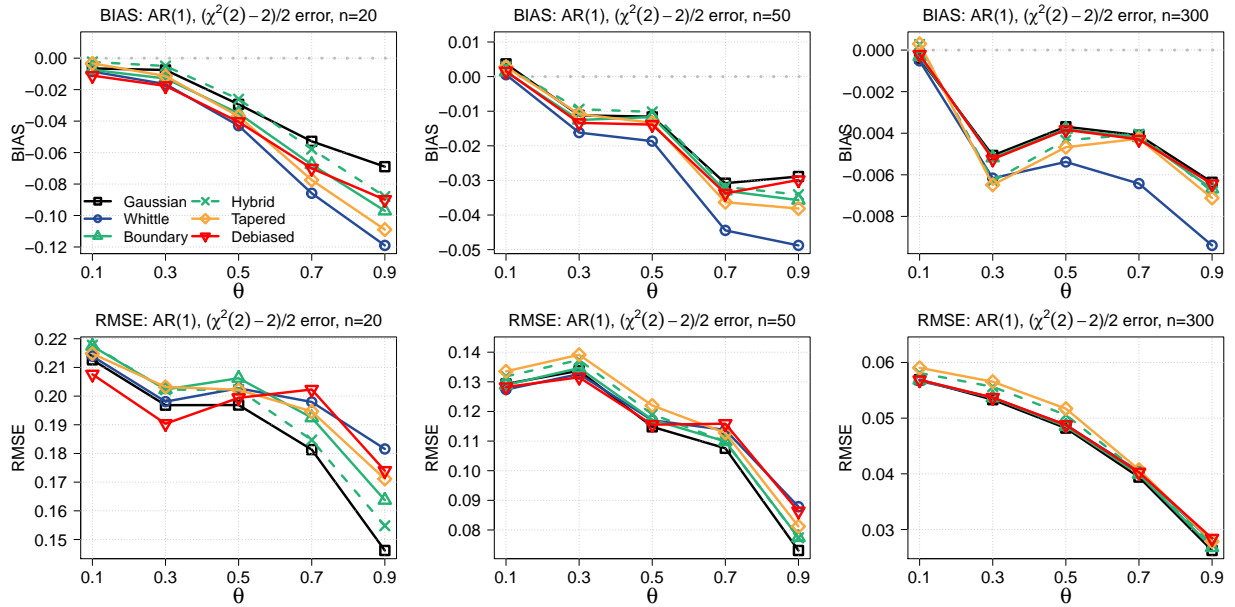
MA(1) model
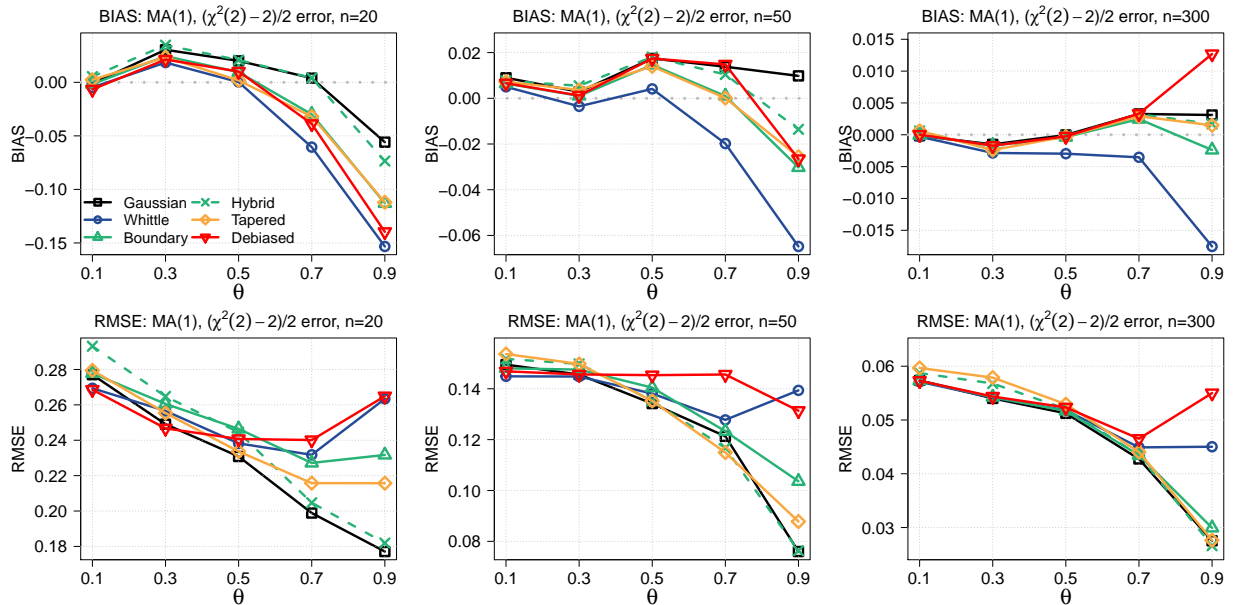
Figure 7: *Bias (first row) and the RMSE (second row) of the parameter estimates for the AR(1) and MA(1) models where the innovations follow the standardized chi-squared distribution with 2 degrees of freedom. Length of the time series $n = 20$(left), 50(middle), and 300(right).*

Likelihoods	θ									
	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 4: *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 to denote the smallest RMSE and blue to denote the second smallest RMSE.*

F.3 Misspecified model for a non-Gaussian time series

In this section, we provide figures and table of the results in Section 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 5 and 6.

n	Parameter	Gaussian	Whittle	Boundary	Hybrid	Tapered	Debiased
	ϕ	0.029(0.1)	-0.102(0.16)	-0.032(0.12)	-0.001(0.1)	-0.088(0.13)	0.170(0.12)
20	ψ	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)
	ϕ	0.014(0.07)	-0.051(0.10)	-0.004(0.07)	0.007(0.07)	-0.003(0.07)	0.143(0.11)
50	ψ	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)
	ϕ	0.001(0.03)	-0.015(0.03)	-0.002(0.03)	0(0.03)	-0.001(0.03)	0.090(0.08)
300	ψ	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 5: *Best fitting (bottom lines) and the bias of estimated coefficients for six different methods for the ARMA(3, 1) misspecified case fitting ARMA(1, 1) model for the standardized chi-squared innovations. Standard deviations are in the parentheses. We use* red *to denote the smallest RMSE and* blue *to denote the second smallest RMSE.*

n	Parameter	Gaussian	Whittle	Boundary	Hybrid	Tapered	Debiased
	ϕ_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)
20	ϕ_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)
	ϕ_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)
50	ϕ_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)
	ϕ_1	0.002(0.04)	-0.015(0.04)	-0.002(0.04)	0(0.04)	-0.001(0.04)	0.012(0.04)
300	ϕ_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 6: *Same as in Table 6 but fitting an AR(2) model.*

F.4 Comparing the 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 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 7 and 8.

		ϕ	ψ	$I_n(f; f_\theta)$
	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 7: *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.*

		ϕ_1	ϕ_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 8: *Same as in Table 7, but fitting an AR(2).*

G   Simulations: Estimation for long memory time series

G.1   Parametric estimation for long memory Gaussian time series

We conduct some simulations for time series whose spectral density, f, does not satisfies Assumption 3.1. We focus on the ARFIMA$(0, d, 0)$ model where

$$(1 - B)^d W_t = \varepsilon_t,$$

B is the backshift operator, $-1/2 < d < 1/2$ is a fractional differencing parameter, and $\{\varepsilon_t\}$ is an i.i.d. standard normal random variable. Let $\Gamma(x)$ denote the gamma function. The spectral density and autocovariance of the ARFIMA$(0, d, 0)$ model (where the variance of the innovations

122

is set to $\sigma^2 = 1$) is

$$f_W(\omega) = (1 - e^{-i\omega})^{-2d} = (2\sin(\omega/2))^{-2d} \quad \text{and} \quad c_W(k) = \frac{\Gamma(k+d)\Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-d)\Gamma(d)} \quad \text{(G.1)}$$

respectively (see Giraitis et al. (2012), Chapter 7.2). Observe that for $-1/2 < d < 0$, the $f_W(0) = 0$, this is called antipersistence. On the other hand, if $0 < d < 1/2$, then $f_W(0) = \infty$ and W_t has long memory.

We generate ARFIMA$(0, d, 0)$ models with $d = -0.4, -0.2, 0.2$ and 0.4 and Gaussian innovations. We fit both the ARFIMA$(0, d, 0)$ model, with d unknown (specified case) and the AR(2) model (with unknown parameters $\theta = (\phi_1, \phi_2)$) (misspecified case) to the data. To do so, we first demean the time series. We evaluate the (plug-in) Gaussian likelihood using the autocovariance function in (G.1) and the autocovariance function of AR(2) model. For the other 5 frequency domain likelihoods, we evaluate the likelihoods at all the fundamental frequencies with the exception of the zero frequency $\omega_{n,n} = 0$. We fit using the spectral density in (G.1) or the spectral density $f_\theta(\omega) = |1 - \phi_1 e^{-i\omega} - \phi_2 e^{-2i\omega}|^{-2}$ where $\theta = (\phi_1, \phi_2)$ (depending on whether the model is specified or misspecified). For each simulation, we calculate six different parameter estimators. For the misspecified case, we also calculate the spectral divergence

$$\widetilde{I}_n(f; f_\theta) = \frac{1}{n-1} \sum_{k=1}^{n-1} \left(\frac{f(\omega_{k,n})}{f_\theta(\omega_{k,n})} + \log f_\theta(\omega_{k,n}) \right),$$

where we omit the zero frequency. The best fitting AR(2) model is $\theta^{Best} = \arg\min_{\theta \in \Theta} \widetilde{I}_n(f; f_\theta)$. In Tables 9 and 10 we give a bias and standard deviation for the parameter estimators for the correctly specified and misspecified model (in the misspecified case we also give the spectral divergence).

Correctly specified model From Table 9 we observe that the bias of both new likelihood estimators is consistently the smallest over all sample sizes and all d except for $d = -0.2$. The new likelihoods have the smallest or second smallest RMSE for n=300, but not for the small sample sizes (e.g. n=20 and 50). This is probably due to increased variation in the new likelihood estimators caused by the estimation of the AR parameter for the predictive DFT. Since the time series has a long memory, the AIC is likely to choose a large order autoregressive order p, which will increase the variance in the estimator (recall that the second order error of the boundary corrected Whittle is $O(p^3 n^{-3/2})$). The (plug-in) Gaussian likelihood has a relatively large bias for all d, which matches the observations in Lieberman (2005), Table 1. However, it has the smallest variance and this results in the smallest RMSE for almost all n when d is negative. The debiased Whittle also has a larger bias than most of the other estimators. However, it has a smaller variance and thus, having the smallest RMSE for almost all n and positive d.

Likelihoods	d			
	-0.4	-0.2	0.2	0.4
	$n = 20$			
Gaussian	-0.097(0.23)	-0.148(0.22)	-0.240(0.23)	-0.289(0.22)
Whittle	0.027(0.3)	0.006(0.28)	-0.008(0.29)	-0.016(0.29)
Boundary	0.014(0.31)	0(0.29)	-0.005(0.30)	-0.007(0.30)
Hybrid	0.009(0.31)	-0.007(0.3)	0.005(0.30)	-0.001(0.30)
Tapered	0.026(0.3)	0(0.3)	0.006(0.30)	0.003(0.29)
Debiased	0.015(0.29)	-0.003(0.27)	-0.029(0.26)	-0.044(0.27)
	$n = 50$			
Gaussian	-0.042(0.13)	-0.073(0.14)	-0.097(0.14)	-0.123(0.12)
Whittle	0.006(0.15)	-0.016(0.15)	-0.013(0.15)	-0.005(0.15)
Boundary	-0.005(0.15)	-0.020(0.16)	-0.011(0.16)	0.001(0.16)
Hybrid	-0.011(0.15)	-0.021(0.15)	-0.012(0.16)	0(0.16)
Tapered	-0.007(0.15)	-0.019(0.16)	-0.011(0.16)	0.010(0.16)
Debiased	-0.008(0.16)	-0.020(0.16)	-0.019(0.15)	-0.021(0.14)
	$n = 300$			
Gaussian	-0.006(0.05)	-0.013(0.05)	-0.020(0.05)	-0.083(0.02)
Whittle	0.006(0.05)	-0.001(0.05)	-0.004(0.05)	0.002(0.05)
Boundary	0.002(0.05)	-0.003(0.05)	-0.003(0.05)	0.002(0.05)
Hybrid	0(0.05)	-0.004(0.05)	-0.004(0.05)	0.001(0.05)
Tapered	0(0.05)	-0.004(0.05)	-0.004(0.05)	0.003(0.05)
Debiased	-0.001(0.05)	-0.003(0.05)	-0.007(0.05)	-0.060(0.02)

Table 9: *Bias and the standard deviation (in the parentheses) of six different quasi-likelihoods for ARFIMA$(0, d, 0)$ model for the standard normal innovations. Length of the time series $n = 20, 50$, and 300. We use* red *to denote the smallest RMSE and* blue *to denote the second smallest RMSE.*

Misspecified model We now compare the estimator when we fit the misspecified AR(2) model to the data. From Table 10 we observe that the Gaussian likelihood performs uniformly well for all d and n, it usually has the smallest bias and RMSE for the positive d. The tapered Whittle also performs uniformly well for all d and n, especially for the negative d. In comparison, the new likelihood estimators do not perform that well as compared with Gaussian and Whittle likelihood. As mentioned above, this may be due to the increased variation caused by estimating many AR parameters. However, it is interesting to note that when $d = 0.4$ and $n = 300$, the estimated spectral divergence outperforms the Gaussian likelihood. We leave the theoretical development of the sampling properties of the new likelihoods and long memory time series for future research.

d	n	Par.	Best	Bias					
				Gaussian	Whittle	Boundary	Hybrid	Tapered	Debiased
	20	ϕ_1	-0.300	-0.028(0.22)	-0.015(0.22)	-0.022(0.23)	-0.026(0.23)	-0.010(0.21)	-0.026(0.22)
		ϕ_2	-0.134	-0.067(0.19)	-0.058(0.19)	-0.064(0.2)	-0.068(0.2)	-0.062(0.18)	-0.067(0.19)
		$I_n(f;f_\theta)$	1.141	0.103(0.1)	0.099(0.1)	0.110(0.12)	0.108(0.11)	0.095(0.1)	0.106(0.1)
-0.4	50	ϕ_1	-0.319	-0.006(0.14)	0.02(0.14)	-0.004(0.15)	-0.004(0.15)	0.003(0.15)	-0.006(0.15)
		ϕ_2	-0.152	-0.028(0.13)	-0.022(0.13)	-0.027(0.13)	-0.029(0.13)	-0.022(0.13)	-0.029(0.13)
		$I_n(f;f_\theta)$	1.092	0.043(0.05)	0.043(0.05)	0.045(0.05)	0.046(0.05)	0.043(0.05)	0.045(0.05)
	300	ϕ_1	-0.331	-0.003(0.06)	-0.002(0.06)	-0.003(0.06)	-0.002(0.06)	-0.001(0.06)	-0.003(0.06)
		ϕ_2	-0.164	-0.005(0.06)	-0.004(0.05)	-0.005(0.06)	-0.005(0.06)	-0.004(0.06)	-0.005(0.06)
		$I_n(f;f_\theta)$	1.062	0.008(0.01)	0.008(0.01)	0.008(0.01)	0.008(0.01)	0.008(0.01)	0.008(0.01)
	20	ϕ_1	-0.157	-0.027(0.23)	-0.024(0.23)	-0.028(0.24)	-0.029(0.23)	-0.020(0.22)	-0.028(0.23)
		ϕ_2	-0.066	-0.084(0.21)	-0.085(0.2)	-0.088(0.21)	-0.089(0.21)	-0.084(0.19)	-0.089(0.21)
		$I_n(f;f_\theta)$	1.066	0.107(0.1)	0.104(0.1)	0.111(0.11)	0.110(0.11)	0.096(0.1)	0.108(0.1)
-0.2	50	ϕ_1	-0.170	-0.020(0.15)	-0.015(0.15)	-0.018(0.15)	-0.020(0.15)	-0.016(0.15)	-0.019(0.15)
		ϕ_2	-0.079	-0.035(0.14)	-0.032(0.14)	-0.034(0.14)	-0.035(0.14)	-0.031(0.13)	-0.036(0.14)
		$I_n(f;f_\theta)$	1.038	0.043(0.04)	0.043(0.04)	0.045(0.04)	0.045(0.04)	0.043(0.04)	0.045(0.04)
	300	ϕ_1	-0.001	0(0.06)	-0.001(0.06)	0(0.06)	0.001(0.06)	0.001(0.06)	-0.006(0.06)
		ϕ_2	-0.088	-0.007(0.05)	-0.007(0.05)	-0.007(0.05)	-0.007(0.06)	-0.007(0.06)	-0.007(0.05)
		$I_n(f;f_\theta)$	1.019	0.007(0.01)	0.007(0.01)	0.007(0.01)	0.008(0.01)	0.008(0.01)	0.007(0.01)
	20	ϕ_1	0.167	-0.066(0.24)	-0.072(0.24)	-0.071(0.25)	-0.068(0.25)	-0.080(0.23)	-0.070(0.24)
		ϕ_2	0.057	-0.098(0.2)	-0.106(0.19)	-0.107(0.19)	-0.108(0.2)	-0.111(0.18)	-0.105(0.19)
		$I_n(f;f_\theta)$	0.938	0.1(0.11)	0.1(0.11)	0.103(0.12)	0.105(0.12)	0.098(0.11)	0.1(0.1)
0.2	50	ϕ_1	0.186	-0.025(0.15)	-0.027(0.15)	-0.026(0.15)	-0.027(0.15)	-0.034(0.15)	-0.025(0.15)
		ϕ_2	0.075	-0.040(0.15)	-0.043(0.15)	-0.043(0.15)	-0.042(0.15)	-0.047(0.15)	-0.042(0.15)
		$I_n(f;f_\theta)$	0.971	0.046(0.05)	0.044(0.05)	0.046(0.05)	0.048(0.05)	0.047(0.05)	0.046(0.05)
	300	ϕ_1	0.208	-0.007(0.06)	-0.007(0.06)	-0.007(0.06)	-0.007(0.06)	-0.008(0.06)	-0.007(0.06)
		ϕ_2	0.097	-0.006(0.06)	-0.007(0.06)	-0.006(0.06)	-0.007(0.07)	-0.008(0.07)	-0.006(0.06)
		$I_n(f;f_\theta)$	1.002	0.008(0.01)	0.008(0.01)	0.008(0.01)	0.008(0.01)	0.009(0.01)	0.008(0.01)
	20	ϕ_1	0.341	-0.072(0.25)	-0.082(0.25)	-0.075(0.26)	-0.074(0.26)	-0.103(0.24)	-0.077(0.25)
		ϕ_2	0.094	-0.116(0.21)	-0.134(0.2)	-0.135(0.2)	-0.133(0.2)	-0.133(0.19)	-0.130(0.2)
		$I_n(f;f_\theta)$	0.877	0.111(0.12)	0.113(0.12)	0.116(0.12)	0.116(0.13)	0.114(0.12)	0.113(0.11)
0.4	50	ϕ_1	0.378	-0.024(0.15)	-0.030(0.15)	-0.027(0.15)	-0.029(0.15)	-0.039(0.15)	-0.027(0.15)
		ϕ_2	0.129	-0.058(0.15)	-0.069(0.14)	-0.067(0.15)	-0.064(0.15)	-0.072(0.15)	-0.066(0.15)
		$I_n(f;f_\theta)$	0.944	0.051(0.06)	0.053(0.06)	0.053(0.06)	0.054(0.06)	0.055(0.06)	0.054(0.06)
	300	ϕ_1	0.428	-0.004(0.06)	-0.004(0.06)	-0.004(0.06)	-0.007(0.06)	-0.009(0.06)	-0.004(0.06)
		ϕ_2	0.178	-0.003(0.06)	-0.005(0.06)	-0.004(0.06)	-0.004(0.06)	-0.006(0.06)	-0.004(0.06)
		$I_n(f;f_\theta)$	1.010	0.009(0.01)	0.009(0.01)	0.009(0.01)	0.010(0.01)	0.010(0.01)	0.009(0.01)

Table 10: *Best fitting and the bias of estimated coefficients using the six different methods for misspecified Gaussian ARFIMA$(0,d,0)$ case fitting AR(2) model. Standard deviations are in the parentheses. We use red to denote the smallest RMSE and blue to denote the second smallest RMSE.*

G.2 Parametric estimation for long memory non-Gaussian time series

We fit the parametric models described in Appendix G.1. However, the underlying time series is non-Gaussian and generated from the ARFIMA$(0,d,0)$

$$(1-B)^d W_t = \varepsilon_t,$$

where $\{\varepsilon_t\}$ are i.i.d. standardized chi-square random variables with two-degrees of freedom i.e. $\varepsilon_t \sim (\chi^2(2) - 2)/2$. The results in the specified setting are given in Table 11 and from the non-specified setting in Table 12.

Likelihoods	d			
	-0.4	-0.2	0.2	0.4
	$n = 20$			
Gaussian	-0.077(0.22)	-0.130(0.23)	-0.219(0.21)	-0.277(0.20)
Whittle	0.089(0.34)	0.092(0.36)	0.077(0.32)	0.058(0.31)
Boundary	0.077(0.34)	0.086(0.37)	0.079(0.33)	0.070(0.32)
Hybrid	0.077(0.35)	0.084(0.37)	0.087(0.33)	0.086(0.32)
Tapered	0.092(0.35)	0.096(0.37)	0.089(0.33)	0.078(0.31)
Debiased	0.057(0.33)	0.051(0.31)	0.009(0.25)	-0.021(0.26)
	$n = 50$			
Gaussian	-0.047(0.13)	-0.065(0.14)	-0.097(0.13)	-0.130(0.12)
Whittle	0.008(0.15)	0.004(0.16)	-0.001(0.15)	0.001(0.16)
Boundary	-0.002(0.16)	0(0.16)	0.002(0.15)	0.006(0.16)
Hybrid	-0.005(0.15)	-0.001(0.16)	0.004(0.16)	0.006(0.16)
Tapered	0.001(0.15)	0.002(0.16)	0.006(0.16)	0.016(0.17)
Debiased	-0.004(0.16)	-0.001(0.16)	-0.013(0.14)	-0.025(0.14)
	$n = 300$			
Gaussian	-0.011(0.05)	-0.012(0.05)	-0.017(0.05)	-0.085(0.02)
Whittle	0.002(0.05)	0(0.05)	-0.001(0.05)	-0.001(0.05)
Boundary	-0.003(0.05)	-0.001(0.05)	0(0.05)	0.001(0.05)
Hybrid	-0.006(0.05)	-0.003(0.05)	-0.001(0.05)	-0.003(0.05)
Tapered	-0.006(0.05)	-0.003(0.05)	-0.001(0.05)	-0.001(0.05)
Debiased	-0.006(0.05)	-0.002(0.05)	-0.004(0.05)	-0.061(0.03)

Table 11: *Same as in Table 9 but for the chi-squared innovations.*

G.3 Semi-parametric estimation for Gaussian time series

Suppose the time series $\{X_t\}$ has a spectral density $f(\cdot)$ with $\lim_{\omega \to 0+} f(\omega) \sim C\omega^{-2d}$ for some $d \in (-1/2, 1/2)$. The local Whittle (LW) estimator is an estimation method for estimating d without using assuming any parametric structure on d. It was first proposed in Künsch (1987); Robinson (1995); Chen and Hurvich (2003), see also Giraitis et al. (2012), Chapter 8). The LW estimator is defined as $\widehat{d} = \arg\min R(d)$ where

$$R(d) = \log \left(M^{-1} \sum_{k=1}^{M} \frac{|J_n(\omega_{k,n})|^2}{\omega_{k,n}^{-2d}} \right) - \frac{2d}{M} \sum_{k=1}^{M} \log \omega_{k,n}, \tag{G.2}$$

and $M = M(n)$ is an integer such that $M^{-1} + M/n \to 0$ as $n \to \infty$. The objective function can be viewed as "locally" fitting a spectral density of form $f_\theta(\omega) = C\omega^{-2d}$ where $\theta = (C, d)$ using the Whittle likelihood.

Since $\widetilde{J}_n(\omega_{k,n}; f)\overline{J_n(\omega_{k,n})}$ is an unbiased estimator of true spectral density $f(\omega_{k,n})$, it is possible that replacing the periodogram with the (feasible) complete periodogram my lead to a better

d	n	Par.	Best	Bias					
				Gaussian	Whittle	Boundary	Hybrid	Tapered	Debiased
	20	ϕ_1	-0.300	-0.017(0.21)	-0.004(0.21)	-0.012(0.22)	-0.018(0.22)	0.001(0.21)	-0.009(0.22)
		ϕ_2	-0.134	-0.073(0.18)	-0.071(0.18)	-0.076(0.18)	-0.077(0.18)	-0.064(0.16)	-0.074(0.18)
		$I_n(f;f_\theta)$	1.141	0.096(0.11)	0.097(0.11)	0.105(0.11)	0.102(0.11)	0.088(0.11)	0.100(0.11)
-0.4	50	ϕ_1	-0.319	-0.008(0.14)	0(0.14)	-0.005(0.14)	-0.007(0.15)	0.001(0.14)	-0.006(0.14)
		ϕ_2	-0.152	-0.035(0.12)	-0.030(0.12)	-0.035(0.12)	-0.035(0.12)	-0.026(0.12)	-0.035(0.12)
		$I_n(f;f_\theta)$	1.092	0.042(0.05)	0.040(0.04)	0.043(0.05)	0.043(0.05)	0.041(0.04)	0.043(0.05)
	300	ϕ_1	-0.331	-0.004(0.06)	-0.002(0.06)	-0.004(0.06)	-0.004(0.06)	-0.003(0.06)	-0.004(0.06)
		ϕ_2	-0.164	-0.007(0.05)	-0.006(0.05)	-0.007(0.05)	-0.008(0.05)	-0.007(0.05)	-0.007(0.05)
		$I_n(f;f_\theta)$	1.062	0.007(0.01)	0.007(0.01)	0.007(0.01)	0.008(0.01)	0.008(0.01)	0.007(0.01)
	20	ϕ_1	-0.157	-0.036(0.22)	-0.034(0.22)	-0.037(0.23)	-0.039(0.23)	-0.028(0.21)	-0.034(0.22)
		ϕ_2	-0.066	-0.083(0.19)	-0.080(0.18)	-0.082(0.19)	-0.083(0.19)	-0.076(0.18)	-0.082(0.19)
		$I_n(f;f_\theta)$	1.066	0.099(0.11)	0.095(0.11)	0.100(0.11)	0.101(0.11)	0.086(0.1)	0.097(0.11)
-0.2	50	ϕ_1	-0.170	-0.018(0.15)	-0.016(0.14)	-0.019(0.15)	-0.017(0.15)	-0.012(0.15)	-0.019(0.15)
		ϕ_2	-0.079	-0.034(0.13)	-0.033(0.13)	-0.034(0.13)	-0.035(0.13)	-0.031(0.13)	-0.034(0.13)
		$I_n(f;f_\theta)$	1.038	0.042(0.05)	0.041(0.05)	0.043(0.05)	0.043(0.05)	0.041(0.04)	0.043(0.05)
	300	ϕ_1	-0.179	-0.001(0.06)	0(0.06)	-0.001(0.06)	0(0.06)	0(0.06)	-0.001(0.06)
		ϕ_2	-0.088	-0.008(0.05)	-0.007(0.05)	-0.008(0.06)	-0.008(0.06)	-0.007(0.06)	-0.008(0.06)
		$I_n(f;f_\theta)$	1.019	0.007(0.01)	0.007(0.01)	0.007(0.01)	0.007(0.01)	0.007(0.01)	0.007(0.01)
	20	ϕ_1	0.167	-0.053(0.23)	-0.064(0.22)	-0.062(0.23)	-0.055(0.23)	-0.071(0.21)	-0.066(0.21)
		ϕ_2	0.057	-0.091(0.2)	-0.100(0.2)	-0.100(0.2)	-0.101(0.2)	-0.099(0.19)	-0.095(0.19)
		$I_n(f;f_\theta)$	0.938	0.094(0.1)	0.091(0.09)	0.094(0.1)	0.097(0.1)	0.086(0.09)	0.086(0.08)
0.2	50	ϕ_1	0.186	-0.023(0.15)	-0.026(0.15)	-0.025(0.15)	-0.024(0.15)	-0.030(0.15)	-0.024(0.15)
		ϕ_2	0.075	-0.044(0.14)	-0.051(0.14)	-0.051(0.14)	-0.047(0.14)	-0.051(0.13)	-0.050(0.14)
		$I_n(f;f_\theta)$	0.971	0.043(0.04)	0.042(0.04)	0.043(0.04)	0.044(0.04)	0.042(0.04)	0.043(0.04)
	300	ϕ_1	0.208	-0.003(0.06)	-0.004(0.06)	-0.003(0.06)	-0.003(0.06)	-0.005(0.06)	-0.003(0.06)
		ϕ_2	0.097	-0.008(0.06)	-0.009(0.06)	-0.008(0.06)	-0.010(0.07)	-0.011(0.06)	-0.008(0.06)
		$I_n(f;f_\theta)$	1.002	0.007(0.01)	0.007(0.01)	0.007(0.01)	0.007(0.01)	0.008(0.01)	0.007(0.01)
	20	ϕ_1	0.341	-0.073(0.24)	-0.080(0.23)	-0.074(0.24)	-0.070(0.25)	-0.103(0.23)	-0.085(0.23)
		ϕ_2	0.094	-0.106(0.21)	-0.129(0.19)	-0.128(0.2)	-0.126(0.2)	-0.122(0.19)	-0.123(0.19)
		$I_n(f;f_\theta)$	0.877	0.103(0.11)	0.103(0.11)	0.107(0.12)	0.108(0.12)	0.103(0.11)	0.101(0.11)
0.4	50	ϕ_1	0.378	-0.021(0.14)	-0.027(0.14)	-0.023(0.15)	-0.026(0.15)	-0.037(0.15)	-0.026(0.14)
		ϕ_2	0.129	-0.044(0.14)	-0.054(0.14)	-0.051(0.14)	-0.049(0.14)	-0.059(0.14)	-0.051(0.14)
		$I_n(f;f_\theta)$	0.944	0.044(0.05)	0.045(0.05)	0.046(0.05)	0.046(0.05)	0.047(0.05)	0.045(0.05)
	300	ϕ_1	0.428	-0.003(0.06)	-0.003(0.06)	-0.003(0.06)	-0.004(0.06)	-0.006(0.06)	-0.002(0.06)
		ϕ_2	0.178	-0.004(0.06)	-0.006(0.06)	-0.004(0.06)	-0.006(0.06)	-0.008(0.06)	-0.005(0.06)
		$I_n(f;f_\theta)$	1.010	0.010(0.01)	0.009(0.01)	0.009(0.01)	0.010(0.01)	0.010(0.01)	0.009(0.01)

Table 12: *Best fitting and the bias of estimated coefficients using the six different methods for misspecified ARFIMA$(0,d,0)$ case fitting AR(2) model for the chi-squared innovations. Standard deviations are in the parentheses. We use* red *to denote the smallest RMSE and* blue *to denote the second smallest RMSE.*

estimator of d. Based on this we define the (feasible) hybrid LW criterion,

$$Q(d) = \log\left(M^{-1}\sum_{k=1}^{M}\frac{\widetilde{J}_n(\omega_{k,n};\widehat{f}_p)\overline{J_{n,\underline{h}_n}(\omega_{k,n})}}{\omega_{k,n}^{-2d}}\right) - \frac{2d}{M}\sum_{k=1}^{M}\log\omega_{k,n}.$$

In a special case that the data taper $h_{t,n} \equiv 1$, we call it the boundary corrected LW criterion.

To empirically assess the validity of the above estimation scheme, we generate a Gaussian ARFIMA$(0,d,0)$ model from Section G.1 for $d = -0.4, -0.2, 0.2$ and 0.4 and evaulate the LW, tapered LW (using tapered DFT in (G.2)), boundary corrected LW, and hybrid LW. We set $M \approx n^{0.65}$ where n is a length of the time series and we use Tukey taper with 10% of the taper on each end of the time series. For each simulation, we obtain four different LW estimators.

Table 13 summarizes the bias and standard deviation (in the parentheses) of LW estimators. We observe that the bondary corrected LW has a smaller bias than the regular Local Whittle

likelihood except when $d = -0.2$ and $n = 50, 300$. However, the standard error tends to be larger (this is probably because of the additional error caused by estimating the $\text{AR}(p)$ parameters in the new likelihoods). Despite the larger standard error, in terms of RMSE, the boundary corrected LW (or hybrid) tends to have overall at least the second smallest RSME for most d and n.

Local likelihoods	d			
	-0.4	-0.2	0.2	0.4
	$n = 20$			
Whittle	0.283(0.46)	0.111(0.5)	-0.075(0.48)	-0.226(0.43)
Boundary	0.282(0.46)	0.109(0.51)	-0.075(0.48)	-0.220(0.44)
Hybrid	0.275(0.46)	0.115(0.52)	-0.067(0.49)	-0.209(0.44)
Tapered	0.279(0.46)	0.121(0.52)	-0.068(0.49)	-0.204(0.43)
	$n = 50$			
Whittle	0.060(0.26)	-0.056(0.33)	-0.089(0.37)	-0.109(0.32)
Boundary	0.045(0.26)	-0.063(0.34)	-0.088(0.38)	-0.106(0.32)
Hybrid	0.033(0.25)	-0.069(0.34)	-0.090(0.38)	-0.110(0.32)
Tapered	0.035(0.25)	-0.068(0.34)	-0.085(0.38)	-0.085(0.31)
	$n = 300$			
Whittle	0.056(0.12)	-0.014(0.11)	-0.010(0.12)	0.004(0.11)
Boundary	0.052(0.12)	-0.017(0.11)	-0.009(0.12)	0.003(0.11)
Hybrid	0.054(0.12)	-0.018(0.12)	-0.011(0.12)	-0.002(0.11)
Tapered	0.057(0.12)	-0.018(0.12)	-0.011(0.12)	0.003(0.12)

Table 13: *Bias and the standard deviation (in the parentheses) of four different Local Whittle estimators for ARFIMA$(0, d, 0)$ model for the standard normal innovations. Length of the time series $n = 20, 50$, and 300. We use* red *to denote the smallest RMSE and* blue *to denote the second smallest RMSE.*

G.4   Semi-parametric estimation for long memory non-Gaussian time series

Once again we consider the semi-parametric Local Whittle estimator described in Appendix G.3. However, this time we assess the estimation scheme for non-Gaussian time series. We generate the ARFIMA$(0, d, 0)$

$$(1 - B)^d W_t = \varepsilon_t,$$

where $\{\varepsilon_t\}$ are i.i.d. standardarized chi-square random variables with two-degrees of freedom i.e. $\varepsilon_t \sim (\chi^2(2) - 2)/2$. The results are summarized in Table 14.

Local likelihoods	d			
	-0.4	-0.2	0.2	0.4
	$n = 20$			
Whittle	$0.354_{(0.52)}$	$0.235_{(0.54)}$	$0.002_{(0.51)}$	$-0.154_{(0.41)}$
Boundary	$0.349_{(0.52)}$	$0.231_{(0.54)}$	$0.003_{(0.51)}$	$-0.146_{(0.41)}$
Hybrid	$0.347_{(0.53)}$	$0.229_{(0.55)}$	$0.016_{(0.52)}$	$-0.133_{(0.4)}$
Tapered	$0.351_{(0.52)}$	$0.229_{(0.54)}$	$0.023_{(0.51)}$	$-0.144_{(0.41)}$
	$n = 50$			
Whittle	$0.019_{(0.25)}$	$-0.080_{(0.34)}$	$-0.125_{(0.38)}$	$-0.149_{(0.33)}$
Boundary	$0.007_{(0.25)}$	$-0.086_{(0.35)}$	$-0.123_{(0.39)}$	$-0.146_{(0.33)}$
Hybrid	$-0.003_{(0.24)}$	$-0.098_{(0.35)}$	$-0.128_{(0.39)}$	$-0.153_{(0.34)}$
Tapered	$0.001_{(0.24)}$	$-0.100_{(0.35)}$	$-0.125_{(0.39)}$	$-0.133_{(0.33)}$
	$n = 300$			
Whittle	$0.104_{(0.14)}$	$-0.006_{(0.14)}$	$-0.015_{(0.14)}$	$-0.011_{(0.13)}$
Boundary	$0.101_{(0.15)}$	$-0.008_{(0.14)}$	$-0.014_{(0.14)}$	$-0.012_{(0.13)}$
Hybrid	$0.100_{(0.15)}$	$-0.011_{(0.15)}$	$-0.017_{(0.14)}$	$-0.020_{(0.13)}$
Tapered	$0.104_{(0.15)}$	$-0.012_{(0.15)}$	$-0.017_{(0.15)}$	$-0.016_{(0.13)}$

Table 14: *Same as in Table 13 but for the chi-square innovations.*

H   Simulations: 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 remedy to reduce the bias is via data tapering (Dahlhaus (1988) and Zhang (1992)). Therefore, we define the **b**oundary **c**orrected Whittle likelihood using **t**apered **Y**ule-**W**alker (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 3.1 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(∞) 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 de-

fine a new feasible criterion which we call the **b**oundary **c**orrected Whittle likelihood using **N**on**p**arametric estimation (BC-NP).

H.1 Alternative methods for estimating the predictive DFT results for a Gaussian time series

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,

$$\phi_U(z) = \prod_{j=1}^{4}(1 - r_j e^{i\lambda_j}z)(1 - r_j e^{-i\lambda_j}z) = 1 - \sum_{j=1}^{8}\phi_j z^j \tag{H.1}$$

$\underline{r} = (r_1, r_2, r_3, r_4) = (0.95, 0.95, 0.95, 0.95)$ 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 15 summarizes the bias and standard derivation of the estimators and the last row is an average ℓ_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 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 15: *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.

H.2   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 as above with

$$V_t = \phi_U(B)\varepsilon_t$$

where $\{\varepsilon_t\}$ are i.i.d. standardarized chi-square random variables with two-degrees of freedom i.e. $\varepsilon_t \sim (\chi^2(2) - 2)/2$ and $\phi_U(z)$ is defined as in (H.1). 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 16.

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 16: *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 *to denote the smallest RMSE and* blue *to denote the second smallest RMSE.*