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-likelihood 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.

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 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 Dahlhaus and Künsch (1987) (for spatial processes), Fox and Taqqu (1986), Robinson (1995) Hurvich and Chen (2000), Giraitis and Robinson (2001), Abadir, Distaso and Giraitis (2007), Shao and Wu (2007), Giraitis, Koul and Surgailis (2012) (long memory time series and local Whittle methods), Choudhuri, Ghosal and Roy (2004), Kirch et al. (2019) (Bayesian spectral methods), and Panaretos and Tavakoli (2013), 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) proved that the tapered periodogram has a smaller bias than the regular periodogram and is better at capturing the features in the spectral density, such as peaks. 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 autocovariance 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

\[
\mathcal{L}_n(\theta; X_n) = n^{-1} \left( \frac{X_n^t \Gamma_n(f_\theta)^{-1} X_n + \log |\Gamma_n(f_\theta)|}{2} \right)
\]

where \( \Gamma_n(f_\theta)_s,t = c_{f_\theta}(s-t), |A| \) denotes the determinant of the matrix \( A \) and \( X_n = (X_1, \ldots, X_n) \). 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

\[
K_n(\theta; X_n) = n^{-1} \sum_{k=1}^n \left( \frac{|J_n(\omega_{k,n})|^2}{f_\theta(\omega_{k,n})} + \log f_\theta(\omega_{k,n}) \right)
\]

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 \( K_n(\theta; X_n) \) can be written as

\[
K_n(\theta; X_n) = n^{-1} \left( X_n^t C_n(f_\theta)^{-1} X_n + \sum_{k=1}^n \log f_\theta(\omega_{k,n}) \right).
\]

To obtain an exact expression for \( \Gamma_n(f_\theta)^{-1} - C_n(f_\theta)^{-1} \) and \( X_n^t [\Gamma_n(f_\theta)^{-1} - C_n(f_\theta)^{-1}] X_n \), 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 \( L_m \). 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 stationary models, the expression is more complex. In Section 3, we obtain an approximation for \( L_n(\theta; X_n) - K_n(\theta; X_n) \) 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(\( \infty \)) coefficients.

In Section 4, we obtain an approximation for the difference \( L_n(\theta; X_n) - K_n(\theta; X_n) \) 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 (which we call the Appendix from now onwards). 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 \( X_n \) 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 \( L_n(\theta) \) and \( K_n(\theta) \) to denote only these terms:

\[
L_n(\theta) = n^{-1} X_n' \Gamma_n(f_\theta)^{-1} X_n \quad \text{and} \quad K_n(\theta) = n^{-1} X_n' C_n(f_\theta)^{-1} X_n.
\]

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 $(F_n)_{k,t} = n^{-1/2}e^{it\omega n}$. 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 $\{e^{i\omega k,n}\} = (e^{i\omega 1,n}, \ldots, e^{i\omega n,n})_{k=1}^n$ 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 $\text{cov}_f(X_t, X_{t+r}) = \mathbb{E}[X_tX_{t+r}] = c_f(r)$. 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 $\|A\|_{\text{spec}}$ denote the spectral norm. Let $\|X\|_{p,q} = (\mathbb{E} |X|^p)^{1/q}$, where $X$ is a random variable. For the $2\pi$-periodic square integrable function $g(\omega) = \sum_{r \in \mathbb{Z}} g_r e^{ir\omega}$, we use the sub-multiplicative norm $\|g\|_K = \sum_{r \in \mathbb{Z}} (2^K + |r|^K)|g_r|$. Note that if $\sum_{j=0}^{K} \sup_{\omega} |g^{(j)}(\omega)| < \infty$ then $\|g\|_K < \infty$, where $g^{(j)}(\cdot)$ denotes the $j$th derivative of $g$.

Suppose $f, g : [0, 2\pi] \to \mathbb{R}$ are bounded functions, that are strictly larger than zero and are symmetric about $\pi$. By using the classical 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 $\sigma_g, \phi_g(\cdot)$, and $\psi_g(\cdot)$ are defined similarly. We use these expansions in Sections 3 and 4, where we require the following notation

$$
\rho_{n,K}(f) = \sum_{r=n+1}^{\infty} |r^K \phi_r(f)|,
$$

$$
A_{K}(f,g) = 2\sigma_g^{-2} \|\psi_f\|_o \|\phi_g\|_0^2 \|\phi_f\|_K,
$$

and

$$
C_{f,K} = \frac{3 - \varepsilon}{1 - \varepsilon} \|\phi_f\|_K^2 \|\psi_f\|_K^2
$$

for some $0 < \varepsilon < 1$.

For positive 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 $\phi$ in the AR(1) model $X_t = \phi X_{t-1} + \varepsilon_t$ for different values of $\phi$ (based on sample size $n = 20$). We observe that the difference between the bias of the two estimators increases as $|\phi|$ approaches one. Further, the Gaussian likelihood clearly has a smaller bias than the Whittle likelihood (which is more pronounced when $|\phi|$ is close to one). 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

$$
L_n(\phi) - K_n(\phi) = n^{-1} \left[2\phi X_1X_n - \phi^2(X_1^2 + X_n^2)\right]
$$

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 $L_n(\theta) - K_n(\theta)$ 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 \{$Z_{k,n}$\}$^n_{k=1}$ \subset sp($X_n$) (where sp($X_n$) denotes the linear space over a complex field spanned by $X_n = \{X_i\}^n_{i=1}$), which is biorthogonal to \{$J_n(\omega_{k,n})\}$$_{k=1}^n$. That is, we derive a transformation \{$Z_{k,n}\}$$_{n}^{n}$ which when coupled with \{$J_n(\omega_{k,n})\}$$_{k=1}^n$ satisfies the following condition

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

where $\delta_{k_1, k_2} = 1$ if $k_1 = k_2$ (and zero otherwise). Since $Z'_n = (Z_{1,n}, \ldots, Z_{n,n}) \in \text{sp}(X_n)^n$, there exists an $n \times n$ complex matrix $U_n$, such that $Z_n = U_n X_n$. Since $(J_n(\omega_{k_1}), \ldots, J_n(\omega_{n,n}))' = F_n X_n$, the biorthogonality of $U_n X_n$ and $F_n X_n$ gives $\text{cov}_f \left(U_n X_n, F_n X_n \right) = \Delta_n(f)$. 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 $\text{var}(X_n)$. That is $\text{cov}(U_n X_n, V_n X_n) = \Delta_n$, where $\Delta_n$ is a diagonal matrix. Then

$$\text{var}(X_n)^{-1} = V_n^* \Delta_n^{-1} U_n.$$  \hspace{1cm} (2.3)

**Proof.** It follows immediately from $\text{cov}(U_n X_n, V_n X_n) = U_n \text{var}(X_n) V_n^* = \Delta_n$ and $\text{var}(X_n) = U_n^{-1} \Delta_n (V_n^*)^{-1}$. \hfill $\square$

To understand how $U_n X_n$ is related to $F_n X_n$ we rewrite $U_n = F_n + D_n(f)$. We show in the following theorem that $D_n(f)$ has a specific form with an intuitive interpretation. In order to
develop these ideas, we use methods from linear prediction. In particular, we define the best linear predictor of $X_\tau$ for $\tau \leq 0$ and $\tau > n$ given $\{X_t\}_{t=1}^n$ as

\begin{equation}
\hat{X}_{\tau,n} = \sum_{t=1}^{n} \phi_{t,n}(\tau; f) X_t,
\end{equation}

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 $\sum_{r \in \mathbb{Z}} |rc_f(r)| < \infty$. Let $\hat{X}_{\tau,n}$ denote the best linear predictors of $X_\tau$ as defined in (2.4) and $\{\phi_{t,n}(\tau; f)\}_{t=1}^n$ the corresponding coefficients. Then

\begin{equation}
\text{cov}_f((F_n + D_n(f))\Sigma_n, F_n\Sigma_n) = \Delta_n(f),
\end{equation}

where $D_n(f)$ has entries

\begin{equation}
D_n(f)_{k,t} = 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)\omega_{k,n}} \right),
\end{equation}

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

\begin{equation}
\text{cov}_f(J_n(\omega_{k_1,n}; f), J_n(\omega_{k_2,n})) = f(\omega_{k_1,n})\delta_{k_1,k_2}
\end{equation}

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

\begin{equation}
\tilde{J}_n(\omega; f) = n^{-1/2} \sum_{\tau > n} \hat{X}_{\tau,n}e^{i\tau \omega} + n^{-1/2} \sum_{\tau < 0} \hat{X}_{\tau,n}e^{i\tau \omega}.
\end{equation}

**Proof.** See Appendix A (note that identity (2.7) can be directly verified using results on best linear predictors). $\Box$

**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

\begin{equation}
\Gamma_n(f)^{-1} = F_n^*\Delta_n(f^{-1})(F_n + D_n(f)),
\end{equation}

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

\begin{equation}
\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}.
\end{equation}

What we observe is that the biorthogonal transformation $(F_n + D_n(f))\Sigma_n$ 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 $\tilde{J}_n(\omega; f)$ is given in Figure 2.

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 $\tilde{J}_n(\omega_{k,n}; f) = O_p(n^{-1/2})$. This is why the regular DFT satisfies the well known “near” orthogonal property

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

see Lahiri (2003) and Brillinger (2001). For future reference, we will use the following definitions.
Fig 2. \( \tilde{J}_n(\omega; f) \) is the Fourier transform over both the observed time series and its predictors outside this domain.

**Definition 2.1.** We refer to \( \hat{J}_n(\omega; f) \) as the predictive DFT (as it is the Fourier transform of all the linear predictors), noting that basic algebra yields the expression

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

Note that when \( \omega = \omega_{k,n} \), the term \( e^{in\omega} \) in (2.10) vanishes. Further, we refer to \( \tilde{J}_n(\omega; f) \) as the complete DFT (as it contains the classical DFT of the time series together with the predictive DFT). Note that both \( \hat{J}_n(\omega; f) \) and \( \tilde{J}_n(\omega; f) \) 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 \( X_t = \phi X_{t-1} + \epsilon_t \ (|\phi| < 1) \). Then the best linear predictors are simply a function of the observations at the two endpoints. That is for \( \tau \leq 0 \), \( \hat{X}_{\tau,n} = \phi^{1+|\tau|} X_1 \) and for \( \tau > n \), \( \hat{X}_{\tau,n} = \phi^{-n} X_n \). An illustration is given in Figure 3.

![Diagram of best linear predictors for AR(1) process](image)

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

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

\[
\tilde{J}_n(\omega; f_\phi) = \frac{\phi}{\sqrt{n}} \left( \frac{1}{\phi(\omega)} X_1 + \frac{e^{i(n+1)\omega}}{\phi(\omega)} X_n \right)
\]

where \( \phi(\omega) = 1 - \phi e^{-i\omega} \).

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

**Remark 2.1.** Bioorthogonality of random variables is rarely used in statistics. An interesting exception is Kasahara, Pourahmadi and Inoue (2009). They apply the notion of bioorthogonality to problems in prediction. In particular they consider the biorthogonal transform of \( \tilde{X}_n \), which is the random vector \( \tilde{X}_n = \Gamma_n(f)^{-1} \hat{X}_n \) (since \( \text{cov}_f(\tilde{X}_n, \hat{X}_n) = I_n \)). They obtain an expression for the entries of \( \tilde{X}_n \) in terms of the Cholesky decomposition of \( \Gamma_n(f)^{-1} \). However, there is an interesting duality between \( \hat{X}_n \) and \( \tilde{J}_n = \)
(\tilde{J}_n(\omega_1,n; f), \ldots, \tilde{J}_n(\omega_n,n; f))'. In particular, applying identity (2.9) to the DFT of \(\tilde{X}_n\) gives

\[ F_n \tilde{X}_n = F_n \Gamma_n(f)^{-1} X_n = \Delta_n(f^{-1}) \tilde{J}_n. \]

This shows that the DFT of the biorthogonal transform of \(X_n\) 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 \(\tilde{X}_n\) are

\[ \tilde{X}_{j,n} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \tilde{J}_n(\omega_{k,n}; f) \frac{f(\omega_{k,n})}{f(\omega_{j,n})} e^{-i\omega_{k,n}}. \]

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

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

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 \(f_\theta\) is bounded away from zero, and the corresponding autocovariance is such that \(\sum_r |r \gamma_{f_\theta}(r)| < \infty\). Let \(L_n(\theta)\) and \(K_n(\theta)\) be defined as in (2.1). Then we have

\[ L_n(\theta) = \frac{1}{n} \tilde{X}_n^t \Gamma_n(f_\theta)^{-1} \tilde{X}_n = \frac{1}{n} \sum_{k=1}^{n} \tilde{J}_n(\omega_{k,n}; f_\theta) \frac{\Gamma_n(\omega_{k,n})}{f_\theta(\omega_{k,n})}. \]

Further

\[ \Gamma_n(f_\theta)^{-1} - C_n(f_\theta^{-1}) = F_n^* \Delta_n(f_{\theta}^{-1}) D_n(f_\theta). \]

This yields the difference between the Gaussian and Whittle likelihood

\[ L_n(\theta) - K_n(\theta) = n^{-1} \tilde{X}_n^t [\Gamma_n(f_\theta)^{-1} - C_n(f_{\theta}^{-1})] \tilde{X}_n \]

\[ = n^{-1} \sum_{k=1}^{n} \tilde{J}_n(\omega_{k,n}; f_\theta) \frac{\Gamma_n(\omega_{k,n})}{f_\theta(\omega_{k,n})}. \]

PROOF. (2.12) follows immediately from Corollary 2.1. Next, we note that \(F_n \tilde{X}_n = J_n\) and \((F_n + D_n(f_\theta)) \tilde{X}_n = \tilde{J}_n\), thus we immediately obtain equation (2.11), and since \(\tilde{J}_n(\omega_{k,n}; f_\theta) = J_n(\omega_{k,n}) + J_n(\omega_{k,n}; f_\theta)\), it proves (2.13). \(\square\)

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

\[ L_n(\theta) = \frac{1}{n} \sum_{k=1}^{n} \frac{|J_n(\omega_{k,n})|^2}{f_\theta(\omega_{k,n})} + \frac{1}{n} \sum_{k=1}^{n} \tilde{J}_n(\omega_{k,n}; f_\theta) \frac{\Gamma_n(\omega_{k,n})}{f_\theta(\omega_{k,n})}. \]

To summarize, the Gaussian likelihood compensates for the well known boundary effect in the Whittle likelihood, by predicting outside the domain of observation. The Whittle likelihood estimator selects the spectral density \(f_\theta\) which best fits the periodogram. On the other
hand, since $\mathbb{E}_f \overline{\mathcal{J}_n(\omega_k, n; f_\theta) \mathcal{J}_n(\omega_k, n)} = f_\theta(\omega_k, n)$, the Gaussian likelihood estimator selects the spectral density which best fits $\mathcal{J}_n(\omega_k, n; f_\theta) \mathcal{J}_n(\omega_k, n)$ by simultaneously predicting and fitting. Therefore, the “larger” the level of “persistence” in the time series, the greater the predictive DFT $\mathcal{J}_n(\omega_k, n; f_\theta)$, 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

$$
\mathcal{L}_n(\theta) - K_n(\theta) = \frac{1}{n} \sum_{k=1}^{n} \overline{\mathcal{J}_n(\omega_k, n; f_\theta) \mathcal{J}_n(\omega_k, n)} f_\theta(\omega_k, n) = n^{-1} \sum_{t} F_n^* \Delta_n(f_\theta^{-1}) D_n(f_\theta) \Xi_n,
$$

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

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

with

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

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

and $K_{f_\theta^{-1}}(\tau) = \int_{0}^{2\pi} f_\theta(\omega)^{-1} e^{i\omega \tau} d\omega$. 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 $\phi_p(\omega) = 1 - \sum_{u=1}^{p} \phi_u e^{-i\omega u}$ (the roots of the corresponding characteristic polynomial lie outside the unit circle) and $p \leq n$. The predictive DFT has the analytic form

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

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

$$
D_n(f_\theta) = n^{-1/2} \begin{pmatrix}
\phi_{1,p}(\omega_1, n) & \cdots & \phi_{1,p}(\omega_n, n) & 0 & e^{i\omega_1} \phi_{1,p}(\omega_1, n) & \cdots & e^{i\omega_n} \phi_{1,p}(\omega_1, n) \\
\phi_{2,p}(\omega_2, n) & \cdots & \phi_{2,p}(\omega_n, n) & 0 & e^{i\omega_1} \phi_{2,p}(\omega_2, n) & \cdots & e^{i\omega_n} \phi_{2,p}(\omega_2, n) \\
\ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\phi_{p,p}(\omega_1, n) & \cdots & \phi_{p,p}(\omega_n, n) & 0 & e^{i\omega_1} \phi_{p,p}(\omega_n, n) & \cdots & e^{i\omega_n} \phi_{p,p}(\omega_n, n)
\end{pmatrix}.
$$
and $\phi_{j,p}(\omega) = \phi_p(\omega)^{-1} \sum_{s=0}^{p-j} \phi_{j+s} e^{-i\omega}$. Note, if $n/2 < p \leq n$, then the entries of $D_n(f_\theta)$ will overlap. Let $\phi_0 = 1$ and for $1 \leq s \leq p$, $\phi_s = -\phi_s$ (zero otherwise), then if $1 \leq p \leq n/2$ we have

$$
(\Gamma_n(f_\theta) - C_n(f_\theta^{-1}))_{s,t} = (F_n^* D_n(f_\theta))_{s,t}
$$

(2.17)

$$
= \begin{cases} 
\sigma^{-2} \sum_{t=0}^{p-t} \phi_{t+s} \phi_{(t+s) \mod n} & 1 \leq t \leq p \\
\sigma^{-2} \sum_{t=1}^{p-(n-t)} \phi_{t+(n-t)} \phi_{(t-s) \mod n} \mod n - p + 1 \leq t \leq n \\
0 & \text{otherwise}
\end{cases}
$$

PROOF. In Appendix A.

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 $F_n^* D_n(f_\theta)$ is non-zero only at the $(p \times p)$ submatrices located in the four corners of $F_n^* D_n(f_\theta)$ 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

$$
L_n(\phi) = \frac{\sigma^{-2}}{n} \left| \sum_{k=1}^{n} J_n(\omega_{k,n}) \phi_p(\omega_{k,n}) \right|^2 \\
+ \frac{\sigma^{-2}}{n} \left[ \sum_{\ell=1}^{p} X_{\ell} \sum_{s=0}^{p-\ell} \phi_{\ell+s} \left( X_{(s+1) \mod n} - \sum_{j=1}^{p} \phi_j X_{(s+1-j) \mod n} \right) \right] \\
+ \frac{\sigma^{-2}}{n} \left[ \sum_{\ell=1}^{p} X_{n+1-\ell} \sum_{s=0}^{p-\ell} \phi_{\ell+s} \left( X_{(s+1) \mod n} - \sum_{j=1}^{p} \phi_j X_{(s+1-j) \mod n} \right) \right],
$$

(2.18)

where $\phi = (\phi_1, \ldots, \phi_p)'$ 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 $\hat{J}_n(: f_\theta)$ 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($\infty$) coefficients corresponding to $f_\theta$. 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 approximation of the predictive DFT.
term of a polynomial-type series expansion of $\Gamma_n(f_0)^{-1}$ in terms of the AR($\infty$) 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 $f_0$.

**Assumption 3.1.**

(i) The spectral density $f$ is bounded away from zero.

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

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

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

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, Paparoditis and Politis (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 $\phi_{s,n}(\tau; f_0)$ in $D_n(f_0)$ with $\phi_{s}(\tau; f_0)$ which are the coefficients of the best linear predictor of $X_\tau$ (for $\tau \leq 0$) given $\{X_t\}_{t=1}^{\infty}$ i.e. $\hat{X}_\tau = \sum_{t=1}^{\infty} \phi_{t}(\tau; f_0) X_t$. This gives the matrix $D_{\infty,n}(f_0)$, where

$$(D_{\infty,n}(f_0))_{k,t} = n^{-1/2} \sum_{\tau \leq 0} \left( \phi_{t}(\tau; f_0) e^{i \tau \omega_{k,n}} + \phi_{n+1-t}(\tau; f_0) e^{-i(\tau-1) \omega_{k,n}} \right).$$

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

$$
(3.2) \quad (D_{\infty,n}(f_0))_{k,t} = n^{-1/2} \frac{\phi_{t} \omega_{k,n}; f_0}{\phi(\omega_{k,n}; f_0)} + n^{-1/2} e^{i \omega_{k,n}} \frac{\phi_{n+1-t} \omega_{k,n}; f_0}{\phi(\omega_{k,n}; f_0)},
$$

where $\phi_t \omega_{k,n} = \sum_{s=0}^{\infty} \phi_{t+s}(f_0) e^{-i s \omega}$. The proof of the above identity can be found in Appendix B.1. Using the above we can show that $(D_{\infty,n}(f_0) X_n)_k = \hat{J}_{\infty,n}(\omega_{k,n}; f_0)$ where

$$
(3.3) \quad \hat{J}_{\infty,n}(\omega_{k,n}; f_0) = \frac{n^{-1/2}}{\phi(\omega_{k,n}; f_0)} \sum_{t=1}^{n} X_t \phi_t \omega_{k,n}; f_0 + e^{i(n+1) \omega} \frac{n^{-1/2}}{\phi(\omega_{k,n}; f_0)} \sum_{t=1}^{n} X_{n+1-t} \phi_t \omega_{k,n}; f_0.
$$

We show below that $\hat{J}_{\infty,n}(\omega_{k,n}; f_0)$ is an approximation of $\hat{J}_{n}(\omega_{k,n}; f_0)$.

**Theorem 3.1 (An AR($\infty$) approximation for general processes).** Suppose $f$ satisfies Assumption 3.1, $f_0$ is bounded away from zero and $\|f_0\|_0 < \infty$ (with $f_0(\omega) = \sigma^2 |\phi_0(\omega)|^{-2}$). Let $D_n(f)$, $D_{\infty,n}(f)$ and $\hat{J}_{\infty,n}(\omega_{k,n}; f)$ be defined as in (2.6) and (3.2) and (3.3) respectively. Then we have

$$
X_n^T F^*_n \Delta_n(f_0^{-1}) (D_n(f) - D_{\infty,n}(f)) X_n
$$

$$
= \sum_{k=1}^{n} \frac{J_n(\omega_{k,n})}{f_0(\omega_{k,n})} (\hat{J}_n(\omega_{k,n}; f) - \hat{J}_{\infty,n}(\omega_{k,n}; f))
$$

and

$$
\|F^*_n \Delta_n(f_0^{-1}) (D_n(f) - D_{\infty,n}(f))\|_1 \leq \frac{C_{f,0} \rho_n K(f)}{n K-1} A_K(f, f_0).
$$
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
\[
n^{-1} \left\| \sum_{s,t=1}^n X_s X_t \left( \frac{1}{n} \sum_{k=1}^n \frac{\hat{J}_{\infty,n}(\omega_k,n; f_0) J_n(\omega_k,n)}{f_0(\omega_k,n)} \right) \right\|_{\mathbb{E},q} \leq \frac{C_{f,0} \rho_{n,K}(f)}{n^K} \left\| A_K(f, f_0) \right\|_{\mathbb{E},2q}.
\] (3.6)

PROOF. See Appendix B.1.

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 \( f_0 \). 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)
\[
\mathcal{L}_n(\theta) = K_n(\theta) + \frac{1}{n} \sum_{k=1}^n \frac{\hat{J}_{\infty,n}(\omega_k,n; f_0) J_n(\omega_k,n)}{f_0(\omega_k,n)} + O_p(n^{-K})
\]
(3.7)
where \( \varphi_{t,n}(\omega; f_0) = \sigma^{-2} \left( \phi(\omega; f_0) \phi_k^\infty(\omega; f_0) + e^{i\omega t} \phi(\omega; f_0) \phi_{n+1-t}^\infty(\omega; f_0) \right) \). The above approximation shows that if the autocovariance function, corresponding to \( f_0 \) decays sufficiently fast (in the sense that \( \sum_{n \in \mathbb{Z}} |r^K \phi_0(n)| < \infty \) 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
\[
(\Gamma_n(f_0)^{-1} - C_n(f_0^{-1}))_{s,t} \approx (F_n^* \Delta_n(f_0^{-1}) D_{\infty,n}(f_0))_{s,t} = \frac{1}{n} \sum_{k=1}^n e^{-i\omega_k,n \varphi_{t,n}(\omega_k,n; f_0)},
\]
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 \( f_0 \) satisfies Assumption 3.1. Let \( D_n(f_0) \) and \( D_{\infty,n}(f_0) \) be defined as in (2.6) and (3.2) respectively. Then we have
\[
\| F_n^* \Delta_n(f_0^{-1}) D_{\infty,n}(f_0) \|_1 \leq A_1(f_0, f_0)
\]
(3.8)
and
\[
\| \Gamma_n(f_0)^{-1} - C_n(f_0^{-1}) \|_1 \leq \left( A_1(f_0, f_0) + \frac{C_{f_0,0} \rho_{n,K}(f_0)}{n^{K-1}} A_K(f_0, f_0) \right).
\]
(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
\[
\| \mathcal{L}_n(\theta) - K_n(\theta) \|_{\mathbb{E},q} \leq \frac{1}{n} \left( A_1(f_0, f_0) + \frac{C_{f_0,0} \rho_{n,K}(f_0)}{n^{K-1}} A_K(f_0, f_0) \right) \|X\|_{\mathbb{E},2q}^2.
\]
(3.10)
PROOF. See Appendix B.1.

The above result shows that under the stated conditions
\[ n^{-1} \| \Gamma_n(f_\theta)^{-1} - C_n(f_\theta)^{-1} \|_1 = O(n^{-1}), \]
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 \( \hat{J}_n(\omega; f) \) in terms of \( \hat{J}_{\infty,n}(\omega; f) \), which is comprised of the AR(\( \infty \)) coefficients corresponding to \( f \). In the following lemma we show that it is possible to obtain a series expansion of \( \hat{J}_n(\omega; f) \) and \( \Gamma_n(f)^{-1} - C_n(f)^{-1} \) in terms of the products of AR(\( \infty \)) 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 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 \( \zeta^{(s)}_{t,n}(\omega; f) = \phi^{\infty}_t(\omega) \) and for \( s \geq 2 \)
\[
\zeta^{(s)}_{t,n}(\omega; f) = \frac{1}{(2\pi)^{s-1}} \int_{[0,2\pi]^s} \left( \prod_{a=1}^{s-1} \phi(\lambda_{a+1}; f)^{-1} \Phi_n(\lambda_a, \lambda_{a+1}) \right) \times
\]
\[
\left( \phi^{\infty}_t(\lambda_s; f) \delta_{s=1(\text{mod} \ 2)} + \phi^{\infty}_{n+1-t}(\lambda_s; f) \delta_{s=0(\text{mod} \ 2)} \right) \delta_{\lambda_1=\omega} d\lambda_s,
\]
where \( d\lambda_s = d\lambda_1 \cdots d\lambda_s \) denotes the \( s \)-dimensional Lebesgue measure,
\[
\Phi_n(\lambda_1, \lambda_2) = \sum_{\omega=0}^{\infty} \phi^{\infty}_{n+1+\omega}(\lambda_1; f) e^{iu\lambda_2} = \sum_{\omega=0}^{\infty} \sum_{s=0}^{\infty} \phi_{n+1+s}(f) e^{-is\lambda_1} e^{iu\lambda_2}
\]
and \( \delta \) denotes the indicator variable. In the following lemma, we show that \( \zeta^{(s)}_{t,n}(\omega; f) \) plays the same role as \( \phi^{\infty}_t(\omega; f) \) in the predictive DFT approximation given in equation (3.3). It will be used to approximate \( \hat{J}_n(\omega; f) \) 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 \( \zeta_{t,n}^{(s)}(\omega; f) \) be defined as in (2.6) and (3.11) respectively. Define the \( s \)-order predictive DFT
\[
\hat{J}^{(s)}_n(\omega; f) = \frac{n^{-1/2}}{\phi(\omega; f)} \sum_{t=1}^{n} X_t \zeta_{t,n}^{(s)}(\omega; f) + e^{i(n+1)\omega} \frac{n^{-1/2}}{\phi(\omega; f)} \sum_{t=1}^{n} X_{n+1-t} \zeta_{t,n}^{(s)}(\omega; f).
\]

Then
\[
\hat{J}_n(\omega; f) = \sum_{s=1}^{\infty} \hat{J}^{(s)}_n(\omega; f)
\]
and \( D_n(f) = \sum_{s=1}^{\infty} D_n^{(s)}(f) \), where
\[
(D_n^{(s)}(f))_{k,t} = n^{-1/2} \frac{\zeta_{t,n}^{(s)}(\omega_{k,n}; f)}{\phi(\omega_{k,n}; f)} + n^{-1/2} e^{i\omega_{k,n}} \frac{\zeta_{n+1-t,n}^{(s)}(\omega_{k,n}; f)}{\phi(\omega_{k,n}; f)}.
\]

Further, for a sufficiently large \( n \) we have
\[
\hat{J}_n(\omega; f) = \sum_{s=1}^{m} \hat{J}^{(s)}_n(\omega; f) + O_p \left( \frac{1}{n^{m(K-1)+1/2}} \right).
\]
In the case $s = 1$, it is straightforward to show that
\[ \hat{J}_n^{(1)}(ω; f) = \hat{J}_{∞, n}(ω; f) \] and \[ D_n^{(1)}(f) = D_{∞, n}(f). \]

Therefore, the first term in the expansion of \( \hat{J}_n(ω, f) \) and \( D_n(f) \) is the \( AR(∞) \) approximation \( \hat{J}_{∞, n}(ω; f) \) and \( D_{∞, n}(f) \) respectively. We mention, that it is simple to check that if \( f \) corresponds to an \( AR(p) \) spectral density for some \( p ≤ n \), then \( \hat{J}_n(ω; f) = 0 \) for all \( s ≥ 2 \).

For general spectral densities, the higher order expansion gives a higher order approximation of \( \hat{J}_n(ω; f) \) and \( Γ_n(f)^{-1} \) in terms of products of the \( AR(∞) \) coefficients. Using the above result we have the expansions

\[ Γ_n(f)^{-1} = C_n(f^{-1}) + \sum_{s=1}^{∞} F_n^s Δ_n(f^{-1}) D_n^{(s)}(f) \]

and

\[ \mathcal{L}_n(θ) = K_n(θ) + \sum_{s=1}^{∞} \frac{1}{n} \sum_{k=1}^{n} \frac{\hat{J}_n^{(s)}(ω_k, n; f_θ) Φ_n(ω_k, n)}{f_θ(ω_k, n)}. \]

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

\[ ζ_{t, n}^{(s+2)}(ω; f) = \frac{1}{(2π)^2} \int_{[0,2π]^2} φ(y_1; f)^{-1} φ(y_2; f)^{-1} Φ_n(ω, y_1) Φ_n(y_1, y_2) ζ_{t, n}^{(s)}(y_2; f) dy_1 dy_2. \]

In a similar vein, both the \( s \)-order predictive DFT \( \hat{J}_n^{(s)}(ω; f) \) 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 \( \hat{J}_n(ω; f) \) and \( Γ_n(f)^{-1} \) 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 \( \hat{J}_n^{(1)}(ω; f) = \hat{J}_{∞, n}(ω; f) \) 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 \( J_n(ω; f_θ) \hat{J}_n(ω) \) and \( f_θ(ω) \). 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 \( θ ∈ Θ \) where \( f = f_θ \) (and \( f \) is the true spectral density). Then

\[ E_{f_θ}[\hat{J}_n(ω; f_θ) \hat{J}_n(ω)] = f_θ(ω) \]

and the Gaussian criterion has a clear interpretation. However, if the model is misspecified (which for real data is likely), \( E_f[\hat{J}_n(ω; f_θ) \hat{J}_n(ω)] \) has no clear interpretation. Instead, to understand what the Gaussian likelihood is estimating, we use that \( E_f[\hat{J}_n(ω; f_θ) \hat{J}_n(ω)] = O(n^{-1}) \), which leads to the approximation \( E_f[\hat{J}_n(ω; f_θ) \hat{J}_n(ω)] = f(ω) + O(n^{-1}) \). From this, we observe that the expected negative log Gaussian likelihood is

\[ n^{-1}E_f[X_n Γ_n(f_θ)^{-1} X_n] + n^{-1} \log |Γ_n(f_θ)| = I(f, f_θ) + O(n^{-1}), \]
where

\begin{equation}
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).
\end{equation}

Since \( I_n(f; f_\theta) \) is the spectral divergence between the true spectral \( f \) density and parametric spectral density \( f_\theta \), asymptotically the misspecified Gaussian likelihood estimator has a meaningful interpretation. However, there is still a finite sample bias in the Gaussian likelihood of order \( O(n^{-1}) \). This can have a knock-on effect, by increasing the finite sample bias in the resulting Gaussian likelihood estimator. To remedy this, in the following section, we obtain a frequency domain criterion which approximates the spectral divergence \( I_n(f; f_\theta) \) to a greater degree of accuracy. This may lead to estimators which may give a more accurate fit of the underlying spectral density. We should emphasis at this point, that reducing the bias in the likelihood, does not necessarily translate to a provable reduction in the bias of the resulting estimators (this is discussed further in Section 5.2).

It is worth noting that, strictly, the spectral divergence is defined as
\[
\sum_{k=1}^{n} \left( \frac{f(\omega_{k,n})}{f_\theta(\omega_{k,n})} - \log f_\theta(\omega_{k,n}) - 1 \right).
\]
It is zero when \( f_\theta = f \) and positive for other values of \( f_\theta \). But since \( -\log f - 1 \) does not depend on \( \theta \) we ignore this term.

### 4.1. The boundary corrected Whittle likelihood

In order to address some of the issues raised above, we recall from Theorem 2.1 that \( \mathbb{E}_f[\tilde{J}_n(\omega; f)\tilde{J}_n(\omega)] = f(\omega) \). In other words, by predicting over the boundary using the (unobserved) spectral density which generates the data, the “complete periodogram” \( \tilde{J}_n(\omega; f)\tilde{J}_n(\omega) \) is an inconsistent but unbiased of the true spectral density \( f \). This motivates the (infeasible) boundary corrected Whittle likelihood

\begin{equation}
W_n(\theta) = \frac{1}{n} \sum_{k=1}^{n} \frac{\tilde{J}_n(\omega_{k,n}; f)\tilde{J}_n(\omega_{k,n})}{f_\theta(\omega_{k,n})} + \frac{1}{n} \sum_{k=1}^{n} \log f_\theta(\omega_{k,n}).
\end{equation}

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 \( \tilde{J}_n(\omega_{k,n}; f) \) are unknown. However, we recall that \( \tilde{J}_n(\omega_{k,n}; f) \) 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, Preuß and Fryzlewicz (2019)). This is equivalent to replacing \( f \) in \( \tilde{J}_n(\omega_{k,n}; f) \) with the spectral density function corresponding to the best fitting AR(\( p \)) process \( \tilde{J}_n(\omega_{k,n}; f_p) \), where an analytic form is given in (2.15). Since we have replaced \( f \) with \( f_p \), the “periodogram” \( \tilde{J}_n(\omega_{k,n}; f_p)\tilde{J}_n(\omega_{k,n}) \) 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

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

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

\begin{equation}
W_{p,n}(\theta) = \frac{1}{n} \sum_{k=1}^{n} \frac{\tilde{J}_n(\omega_{k,n}; f_p)\tilde{J}_n(\omega_{k,n})}{f_\theta(\omega_{k,n})} + \frac{1}{n} \sum_{k=1}^{n} \log f_\theta(\omega_{k,n}).
\end{equation}

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, \( f_\theta \) is bounded away from zero and \( \|f_\theta\|_0 < \infty \). Let \( \{a_j(p)\} \) denote the coefficients of the best fitting AR(p) model corresponding to the spectral density \( f \) and define \( f_p(\omega) = |1 - \sum_{j=1}^p a_j(p)e^{-ij\omega}|^{-2} \). Suppose \( 1 \leq p < n \), then we have
\[
\left\| \frac{F_n^* \Delta_n(f_\theta^{-1})(D_n(f) - D_n(f_p))}{1} \right\|_1 \leq \rho_{p,K}(f)A_K(f, f_\theta)
\]
Further, if \( \{X_t\} \) is a time series where \( \sup_t \|X_t\|_{E,2q} = \|X\|_{E,2q} < \infty \) (for some \( q > 1 \)), then
\[
\|W_n(\theta) - W_{p,n}(\theta)\|_{E,q} \leq \rho_{p,K}(f)A_K(f, f_\theta) \times
\]
(4.5)
\[
\left( \frac{C_{f,1} + 1}{np^{K-1}} + \frac{2(C_{f,1} + 1)^2}{np^{K}} \right) \|\psi_f\|_0 \|\phi_f\|_1 + \frac{C_{f,0}}{n^{K-1}} \|X\|^2_{E,2q}.
\]
PROOF. See Appendix B.1. \( \square \)

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). If \( p \geq p_0 \), then
\[
\left\| \frac{F_n^* \Delta_n(f_\theta^{-1})(D_n(f) - D_n(f_p))}{1} \right\|_1 = 0 \text{ and } \|W_n(\theta) - W_{p,n}(\theta)\|_{E,q} = 0. \text{ On the other hand, if } p < p_0 \text{ we replace the } p^K \text{ and } p^{K-1} \text{ terms in Lemma 4.1 with } \sum_{j=p+1}^{p_0} |\phi_j| \text{ and } \sum_{j=p+1}^{p_0} |\phi_j| \text{ respectively, where } \{\phi_j\}_{j=1}^p \text{ are the AR(p) coefficients corresponding to } f.
\]

(ii) If the autocovariances corresponding to \( f \) decay geometrically fast to zero (for example an ARMA processes), then for some \( 0 \leq \rho < 1 \) we have
\[
\|W_n(\theta) - W_{p,n}(\theta)\|_{E,q} = O \left( \frac{\rho^p}{n} + \rho^n \right).
\]

(iii) If the autocovariances corresponding to \( f \) decay to zero at a polynomial rate with \( \sum_r |r^K c(r)| < \infty \), then
\[
\|W_n(\theta) - W_{p,n}(\theta)\|_{E,q} = O \left( \frac{1}{np^{K-1}} \right).
\]

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 \( 1 \leq p < n \), \( \mathbb{E}_f [W_{p,n}(\theta)] = I_n(f; f_\theta) + O((np^{K-1})^{-1}) \) 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 \( \hat{f}_p \) and the corresponding spectral density as \( \hat{f}_p \). Using this we define \( \hat{f}_n(\omega; \hat{f}_p) \) where
\[
\hat{f}_n(\omega; \hat{f}_p) = \frac{n^{-1/2}}{\hat{f}_p(\omega)} \sum_{\ell=1}^{p-\ell} X_{\ell} \sum_{s=0}^p \hat{\phi}_{\ell+\omega, p} e^{-is\omega} + e^{im\omega} \frac{n^{-1/2}}{\hat{f}_p(\omega)} \sum_{\ell=1}^{p-\ell} X_{n+1-\ell} \sum_{s=0}^p \hat{\phi}_{\ell+\omega, p} e^{is(1)\omega},
\]
and \( \hat{\phi}_p(\omega) = 1 - \sum_{u=1}^p \hat{\phi}_{u,p} e^{-i u \omega} \). This estimator allows us to replace \( \tilde{J}_n(\omega_k; f_p) \) in \( W_{p,n}(\theta) \) with \( \tilde{J}_n(\omega_k; f_p) \) to give the “observed” boundary corrected Whittle likelihood

\[
\tilde{W}_{p,n}(\theta) = \frac{1}{n} \sum_{k=1}^n \frac{\tilde{J}_n(\omega_k; f_p) J_n(\omega_k)}{f_\theta(\omega_k)} + \frac{1}{n} \sum_{k=1}^n \log f_\theta(\omega_k).
\]

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

\[
\text{Im} \frac{\tilde{J}_n(\omega_k; f_p) J_n(\omega_k)}{f_\theta(\omega_k)} = - \text{Im} \frac{\tilde{J}_n(\omega_k - \kappa; f_p) J_n(\omega_k - \kappa)}{f_\theta(\omega_k - \kappa)}
\]

thus \( \tilde{W}_{p,n}(\theta) \) is real for all \( \theta \). However, due to rounding errors it is prudent to use \( \text{Re} \tilde{W}_{p,n}(\theta) \) in the minimisation algorithm. Sometimes \( \text{Re} \tilde{J}_n(\omega_k; f_p) J_n(\omega_k) \) 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 \( \tilde{J}_n(\omega_k; f_p) \) 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 \( \tilde{J}_{\infty,n}(\omega_k; f) \), 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 \( h_{t,n} = \{h_{t,n}\}_{t=1}^n \) 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

\[
J_{n,h_{t,n}}(\omega_k) = n^{-1/2} \sum_{t=1}^n h_{t,n} X_te^{i t \omega_k}.
\]

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

\[
\mathbb{E}_f[\tilde{J}_n(\omega; f) J_{n,h_{t,n}}(\omega)] = f(\omega),
\]

which is analogous to the non-tapered result \( \mathbb{E}_f[\tilde{J}_n(\omega; f) J_n(\omega)] = f(\omega) \). 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)

\[
H_n(\theta) = \frac{1}{n} \sum_{k=1}^n \frac{\tilde{J}_n(\omega_k; f) J_{n,h_{t,n}}(\omega_k)}{f_\theta(\omega_k)} + \frac{1}{n} \sum_{k=1}^n \log f_\theta(\omega_k).
\]
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 $\theta$ using the (unobserved) criterion $H_n(\theta)$. Instead we replace $\tilde{J}_n(\omega_k, n; f)$ with its estimator $\tilde{J}_n(\omega_k, n; \hat{f}_p)$ and define

$$
\hat{H}_{p,n}(\theta) = \frac{1}{n} \sum_{k=1}^{n} \frac{\tilde{J}_n(\omega_k, n; \hat{f}_p)J_{n,h_n}(\omega_k, n)}{f_\theta(\omega_k, n)} + \frac{1}{n} \sum_{k=1}^{n} \log f_\theta(\omega_k, n).
$$

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

![Figure 4](image)

**FIG 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 $1 \leq t \leq n$. In Das, Subba Rao and Yang (2020) we study the sampling properties of the estimated complete periodogram $\tilde{J}_n(\omega; \hat{f}_p)J_{n,h_n}(\omega)$. 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 $\theta$ is an unknown $d$-dimension parameter vector) to the stationary time series $\{X_t\}_{t=1}^n$ whose true spectral density is $f$. The best fitting spectral density is $f_{\hat{\theta}_n}$, where $\hat{\theta}_n = \arg \min I_n(f; f_\theta)$. Let $\hat{\theta}_n = (\hat{\theta}_{1,n}, \ldots, \hat{\theta}_{d,n})$ be its estimator, where $\hat{\theta}_n = \arg \min \hat{H}_{p,n}(\theta)$.

5.1. Assumptions. To derive the sampling properties of $\hat{\theta}_n$ we assume the data taper has the following form

$$
h_{t,n} = c_nh_n(t/n),
$$

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 $H_{\theta,n} = \sum_{t=1}^{n} h_n(t/n)^q$. 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, $0 < \inf_{\theta \in \Theta} \inf_{\omega} f_\theta(\omega) \leq \sup_{\theta \in \Theta} \sup_{\omega} f_\theta(\omega) < \infty$ and $\theta_n$ lies in the interior of $\Theta$.
(ii) The one-step ahead prediction error $\sigma^2 = \exp((2\pi)^{-1} \int_0^{2\pi} \log f_\theta(\omega)d\omega)$ is not a function of the parameter $\theta$.
(iii) Let $\{\phi_j(f_\theta)\}$ and $\{\psi_j(f_\theta)\}$ denote the AR($\infty$) and MA($\infty$) coefficients corresponding to the spectral density $f_\theta$ respectively. Then for all $\theta \in \Theta$ and $0 \leq s \leq \kappa$ (for some $\kappa \geq 4$), we have
\[
(a) \sup_{\theta \in \Theta} \sum_{j=1}^{\infty} \|j^K \nabla^g \phi_j(f_\theta)\| < \infty \quad (b) \sup_{\theta \in \Theta} \sum_{j=1}^{\infty} \|j^K \nabla^g \psi_j(f_\theta)\| < \infty,
\]
where $K > 3/2$, $\nabla^a g(f_\theta)$ is the $a$th order partial derivative of $g$ with respect to $\theta$, and $\|\nabla^g f_\theta\|$ denotes the absolute sum of all the partial derivatives in $\nabla^a g(f_\theta)$.

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 to 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 $s$th 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 $\text{cum}(X_{t_1}, X_{t_2}, \ldots, X_{t_{\ell-1}})$.
Then for all $1 \leq j \leq \ell \leq 12$,
\[
\sum_{t_1, \ldots, t_{\ell-1}} |(1 + t_j)\kappa_\ell(t_1, \ldots, t_{\ell-1})| < \infty.
\]
(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
\[
I(\theta) = -\frac{1}{2\pi} \int_0^{2\pi} [\nabla^2 f_\theta(\omega)^{-1}] f(\omega)d\omega.
\]

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 $\hat{H}_{p,n}(\theta)$ and its infeasible counterparts $H_n(\theta)$. Under Assumption 5.2(ii,iii), Theorem 3.1 in Das, Subba Rao and Yang (2020), we can show that
\[
\hat{H}_{p,n}(\theta) = H_n(\theta) + O_p \left( \frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}} \right).
\]

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 $\hat{\theta}_n = (\hat{\theta}_{1,n}, \ldots, \hat{\theta}_{d,n})$ we require the following quantities. Let
\[
V(g, h) = \frac{2}{2\pi} \int_0^{2\pi} g(\omega)h(\omega)f(\omega)^2d\omega
\]
\[= \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} g(\omega_1)h(\omega_2)f_4(\omega_1, -\omega_1, \omega_2)d\omega_1d\omega_2
\]
(5.3) and
\[
J(g) = \frac{1}{2\pi} \int_0^{2\pi} g(\omega)f(\omega)d\omega.
\]

RECONCILING THE GAUSSIAN AND WHITTLE LIKELIHOOD
where \( f_t \) 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

\[
G_v(\theta) = \sum_{s_1, s_2=1}^d I^{(s_1, s_2)} V \left( \frac{\partial f_{\theta}^{-1}}{\partial \theta_{s_2}}, \frac{\partial^2 f_{\theta}^{-1}}{\partial \theta_{s_1} \partial \theta_r} \right) + \frac{1}{2} \sum_{s_1, s_2, s_3, s_4=1}^d I^{(s_1, s_2)} I^{(s_2, s_4)} V \left( \frac{\partial f_{\theta}^{-1}}{\partial \theta_{s_3}}, \frac{\partial f_{\theta}^{-1}}{\partial \theta_{s_4}} \right) J \left( \frac{\partial^3 f_{\theta}^{-1}}{\partial \theta_{s_2} \partial \theta_{s_2} \partial \theta_r} \right).
\]

\[(5.4)\]

### 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 \( \hat{\theta}_n = \arg\min H_n(\theta) \) and \( \tilde{\theta}_n = \arg\min \tilde{H}_{p,n}(\theta) \) respectively. Then for \( p \geq 1 \) we have

\[
|\hat{\theta}_n - \tilde{\theta}_n|_1 = O_P \left( \frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}} \right),
\]

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

Note if the true spectral density of the time series is that of an AR(p0) 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 \( n|\hat{\theta}_n - \theta|_1 \to 0 \) 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.

**The bias** We show in Appendix E.4, that the asymptotic bias (in the sense of Bartlett) for \( \hat{\theta}_n = (\hat{\theta}_{1,n}, \ldots, \hat{\theta}_{d,n}) \) is

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

\[(5.5)\]

where \( I^{(j,r)} \) and \( G_v(\theta_n) \) 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 straightforward “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
We show in Corollary 3.1, Das, Subba Rao and Yang (2020) that the inclusion of the prediction DFT in the hybrid Whittle likelihood has a variance which asymptotically is 

\[ V(\hat{\theta}_n) = I(\hat{\theta}_n)^{-1}V(\nabla_\theta f_g^{-1}, \nabla_\theta f_{g^{-1}}) \big|_{\theta = \hat{\theta}_n} I(\hat{\theta}_n)^{-1} + o(1), \]

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 \( \hat{p} = \arg\min_{p} \text{AIC}(p) \) with

\[ \text{AIC}(p) = \log \hat{\sigma}_{p,n}^2 + \frac{2p}{n}, \]

\[ \hat{\sigma}_{p,n}^2 = \frac{1}{n-n_i} \sum_{t=K_n}^{T-n_i} (X_t - \hat{\varphi}_{j,p} X_{t-j})^2, \]

where \( K_n \) 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(\( \infty \)) 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 \(\hat{p} = O_p(n^{1/(1+2K)})\) (see Example 2 in Ing and Wei (2005)). Thus, if \(K > 5/2\), then \(\hat{p}^3 / n^{1/2} P \to 0\) (where \(\hat{p} = O_p(n^{1/(1+2K)})\)) and \(\hat{p} \to \infty\) as \(n \to \infty\). These rates ensure that the difference between the feasible and infeasible estimator is \(|\hat{\theta}_n - \tilde{\theta}_n| = O_p(n^{-1})\). 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 \((\hat{J}_n(\omega_k; \hat{f}_p))^n_{k=1}\). We recall that \(\hat{f}_p\) is constructed using the Yule-Walker estimator \(\hat{\phi}_n = (\hat{\phi}_{1,p}, \ldots, \hat{\phi}_{p,p})'\) where \(p\) is selected with the AIC. We now calculate the complexity of calculating \((\hat{J}_n(\omega_k; \hat{f}_p))^n_{k=1}\).

The sample autocovariances, \((c_n(r))^{n-1}_{r=0}\) (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 \(\hat{\phi}_n\) for \(1 \leq p \leq K_n\) requires in total \(O(K_n^2)\) arithmetic operations. Given the estimated AR coefficients \(\hat{\phi}_n\), the predictive DFT \((\hat{J}_n(\omega_k; \hat{f}_p))^n_{k=1}\) can be calculated in \(O(\min(n \log n, n \hat{p}))\) 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

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

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 \(\hat{f}_p\) is estimated using the Yule-Walker estimator.

Unlike the Whittle, the tapered Whittle and debiased Whittle likelihood, \(\text{Re} \, \hat{J}_n(\omega_k; \hat{f}_p) \hat{J}_n(\omega_k; \hat{f}_p)\) and \(\text{Re} \, \hat{J}_n(\omega_k; \hat{f}_p) J_n,\hat{\theta}_n(\omega_k; \hat{f}_p)\) can be negative. To avoid negative
values, we apply the thresholding function \( f(t) = \max(t, 10^{-3}) \) to \( \Re \tilde{J}_n(\omega_k; f_p) \tilde{J}_n(\omega_k; f_p) \) and \( \Re \tilde{J}_n(\omega_k; f_p) \tilde{J}_n(\omega_k; f_p) \) over all the frequencies. Thresholding induces an additional (small) bias to the new criteria. The proportion of times that \( \Re \tilde{J}_n(\omega_k; f_p) \tilde{J}_n(\omega_k; f_p) \) drops below the threshold increases for spectral density functions with large peaks and when the spectral density is close to zero. However, at least for the models that we studied in the simulations, the bias due to the thresholding is negligible.

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

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 \( X_n \) and \( Y_n \), which satisfy the following recursions

\[
\begin{align*}
\text{AR}(1) : \quad X_t &= \theta X_{t-1} + e_t; \quad \phi_X(\omega) = 1 - \theta e^{-i\omega} \\
\text{MA}(1) : \quad Y_t &= e_t + \theta e_{t-1}; \quad \phi_Y(\omega) = (1 + \theta e^{-i\omega})^{-1},
\end{align*}
\]

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 \( \theta \) for both AR(1) and MA(1) models. We focus on positive \( \theta \), similar results are obtained for negative \( \theta \). The results are also summarized in Table 1 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 \( \phi \) is large). A caveat is that for the AR(1) model the bias of the new likelihoods tends to be a little larger than the bias of the Gaussian likelihood (especially for the smaller sample sizes). This is interesting, because in Appendix 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 new likelihoods it is \( O(p^3n^{-3/2}) \). Indeed, the \( O(p^3n^{-3/2}) \) term arises because of the parameter estimation in
the predictive DFT. This term is likely to dominate the $O(n^{-3/2})$ in the Gaussian likelihood. Therefore, for small sample sizes, the second order terms can impact the bias. It is this second order term that may be causing the larger bias seen in the boundary corrected Whittle likelihood as compared with the Gaussian likelihood.

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

In summary, the new likelihoods seem to perform quite well compared with the standard methods. For large sample sizes the performance of all the estimators improves considerably.
6.2. Estimation under misspecification. Next, we turn 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

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

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

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

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

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

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 $\theta_{\text{Best}} = \arg\min_{\theta \in \Theta} I_n(f; f_{\theta})$, where $I_n(f, f_{\theta})$ is defined in (4.1) and $\Theta$ is the appropriate 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 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)).

![Table 1](image1)

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.

![Table 2](image2)

Same as in Table 1 but fitting AR(2) model.

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 $\sum_{r=1}^{d} I^{(r)}[E[(\partial L)_{\theta}^{(r)}]]_{\theta_{0}}$. 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.

We next turn our attention to the estimated spectral divergence $I_{n}(f; \hat{f})$. 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 mentioned at the start of the section, this leads to an additional error of $O(p^3n^{-3/2})$ in the new likelihoods. Thus, if a large $p$ is selected the error $O(p^3n^{-3/2})$ 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 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.

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 $X_n[X_n'\Gamma_n(f_{\theta})^{-1} - C_n(f_{\theta}^{-1})]X_n$. 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, Kasahara and Pourahmadi (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. We give a version of this result in Appendix A. In Appendix G we present some preliminary simulations for long memory time series.

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

Acknowledgements. The research was partially conducted while SSR was visiting the Universität Heidelberg, SSR is extremely grateful to the hospitality of everyone at Mathematik. SSR is grateful to Thomas Hotz for several fruitful discussions. Finally, this paper is dedicated to SSR’s Father, Tata Subba Rao, who introduced the Whittle likelihood to (the young and rather confused) SSR many years ago. SSR and JY gratefully acknowledge the support of the National Science Foundation (grant DMS-1812054).

The authors wish to thank three anonymous referees and editors for their insightful observations and suggestions, which substantially improved all aspects of the paper.

SUPPLEMENTARY MATERIAL

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

(doi: TBA; .pdf). The supplement contains the proofs of the results and the additional results.
REFERENCES

ABADIR, K. M., DISTASO, W. and GIRAITTIS, L. (2007). Nonstationarity-extended local Whittle estimation. J. Econometrics 141 1353-1384.

BARTLETT, M. S. (1953). Approximate confidence intervals. II. More than one Unknown Parameter. Biometrika 40 306-317.

BAXTER, G. (1962). An asymptotic result for the finite predictor. Math. Scand. 10 137–144.

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

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

BRILLINGER, D. R. (2001). Time series: Data Analysis and theory. Classics Appl. Math. 36. SIAM, Philadelphia, PA.

CHENG, R. and POURAHMADI, M. (1993). Baxter’s inequality and convergence of finite predictors of multivariate stochastic processes. Probab. Theory Related Fields 95 115–124.

CHOU, N., GHOSAL, S. and ROY, A. (2004). Bayesian estimation of the spectral density of a time series. J. Amer. Statist. Assoc. 99 1050–1059.

COURSOL, J. and DACUNHA-CASTELLE, D. (1982). Remarques sur l’approximation de la vraisemblance d’un processus Gaussien stationnaire. Teor. Veroyatnost. i Primenen. (Theory Probab. Appl.) 27 155–160.

COX, D. R. and SNELL, E. J. (1968). A general definition of residuals. J. R. Stat. Soc. Ser. B. Stat. Methodol. 30 248–265.

DAHLHAUS, R. (1983). Spectral analysis with tapered data. J. Time Series Anal. 4 163–175.

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

DAHLHAUS, R. (1990). Nonparametric high resolution spectral estimation. Probab. Theory Related Fields 85 147–180.

DAHLHAUS, R. and KUNSCHE, H. (1987). Edge effects and efficient parameter estimation for stationary random fields. Biometrika 74 877–882.

DAS, S., SUBBA RAO, S. and YANG, J. (2020). Spectral methods for small sample time series: A complete periodogram approach. To appear in J. Time Series Anal.

FOX, R. and TAQUÉ, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. Ann. Statist. 14 517–532.

GALBRAITH, R. F. and GALBRAITH, J. I. (1974). On the inverses of some patterned matrices arising in the theory of stationary time series. J. Appl. Probab. 11 63–71.

GIRAITTIS, L., KOUL, H. L. and SURGAILIS, D. (2012). Large sample inference for long memory processes. Imperial College Press, London.

GIRAITTIS, L. and ROBINSON, P. M. (2001). Whittle estimation of ARCH models. Econometric Theory 17 608–631.

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

ING, C. K. and WEI, C. Z. (2005). Order selection for same-realization predictions in autoregressive processes. Ann. Statist. 33 2423-2474.

INOUÉ, A. and KASAHARA, Y. (2006). Explicit representation of finite predictor coefficients and its applications. Ann. Statist. 34 973–993.

INOUÉ, A., KASAHARA, Y. and POURAHMADI, M. (2018). Baxter’s inequality for finite predictor coefficients of multivariate long-memory stationary processes. Bernoulli 24 1202–1232.

KASAHARA, Y., POURAHMADI, M. and INOUÉ, A. (2009). Duals of random vectors and processes with applications to prediction problems with missing values. Statist. Probab. Lett. 79 1637–1646.

KIRCH, C., EDWARDS, M. C., MEIER, A. and MEYER, R. (2019). Beyond Whittle: Nonparametric correction of a parametric likelihood with a focus on Bayesian time series analysis. Bayesian Anal. 14 1037–1073.

KLEY, T., PREUSS, P. and FRIZQUELEWICZ, P. (2019). Predictive, finite-sample model choice for time series under stationarity and non-stationarity. Electron. J. Stat. 13 3710–3774.

KREISS, J. P., PAPARODITIS, E. and POLITIS, D. N. (2011). On the range of validity of the autoregressive sieve bootstrap. Ann. Statist. 39 2103–2130.

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

PANARETOS, V. M. and TAVAKOLI, S. (2013). Fourier analysis of stationary time series in function space. Ann. Statist. 41 568–603.

PARZEN, E. (1983). Autoregressive spectral estimation. In Time series in the frequency domain. Handbook of Statist. 3 221–247. North-Holland, Amsterdam.
PRIESTLEY, M. B. (1981). *Spectral analysis and time series. Vol. 2.* Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London-New York Multivariate series, prediction and control, Probability and Mathematical Statistics.

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

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

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

SHAO, X. and WU, W. B. (2007). Local Whittle estimation of fractional integration for nonlinear processes. *Econometric Theory* 23 899-929.

SIDDQUI, M. M. (1958). On the inversion of the sample covariance matrix in a stationary autoregressive process. *Ann. Math. Statist.* 29 585–588.

SUBBA RAO, S. (2018). Orthogonal samples for estimators in time series. *J. Time Series Anal.* 39 313-337.

SYKULSKI, A. M., OLHEDE, S. C., GUILLAUMIN, A. P., LILLY, J. M. and EARLY, J. J. (2019). The debiased Whittle likelihood. *Biometrika* 106 251–266.

SZEGŐ, G. (1921). Über die Randwerte einer analytischen Funktion. *Math. Ann.* 84 232-244.

VAN DELFT, A. and EICHLER, M. (2020). A note on Herglotz’s theorem for time series on functional spaces. *Stochastic Process. Appl.* 130 3687–3710.

WALKER, A. M. (1964). 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.

WHITTLE, P. (1951). *Hypothesis Testing in Time Series Analysis.* Thesis, Uppsala University.

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

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