Annals of the Institute of Statistical Mathematics (2022) 74:233–270
https://doi.org/10.1007/s10463-021-00802-6

Whittle estimation for continuous-time stationary state space models with finite second moments

Vicky Fasen-Hartmann[1] · **Celeste Mayer**[1]

Received: 22 February 2020 / Revised: 18 December 2020 / Accepted: 22 March 2021 /
Published online: 31 July 2021

Abstract
We consider Whittle estimation for the parameters of a stationary solution of a continuous-time linear state space model sampled at low frequencies. In our context, the driving process is a Lévy process which allows flexible margins of the underlying model. The Lévy process is supposed to have finite second moments. Then, the classes of stationary solutions of linear state space models and of multivariate CARMA processes coincide. We prove that the Whittle estimator, which is based on the periodogram, is strongly consistent and asymptotically normal. A comparison with ARMA models shows that in the continuous-time setting the limit covariance matrix of the estimator has an additional term for non-Gaussian models. Thereby, we investigate the asymptotic normality of the integrated periodogram. Furthermore, for univariate processes we introduce an adjusted version of the Whittle estimator and derive its asymptotic properties. The practical applicability of our estimators is demonstrated through a simulation study.

Keywords Asymptotic normality · CARMA process · Consistency · Identifiability · Periodogram · Quasi-maximum-likelihood estimator · State space model · Whittle estimator

1 Introduction

Continuous-time linear state space models are widely used in diverse fields as, e.g., in signal processing and control, high-frequency financial econometrics and financial mathematics. The advantages of continuous-time models are that they allow to

✉ Celeste Mayer
celeste.mayer@kit.edu

Vicky Fasen-Hartmann
vicky.fasen@kit.edu

1 Institute of Stochastics, Karlsruhe Institute of Technology, Englerstraße 2, 76131 Karlsruhe, Germany

model high-frequency data as in finance and in turbulence but as well irregularly spaced data, missing observations or situations when estimation and inference at various frequencies have to be carried out.

In this paper, we investigate stationary solutions of continuous-time linear state space models driven by a Lévy process. A one-sided d-dimensional Lévy process $(L_t)_{t\geq 0}$ is a stochastic process with stationary and independent increments satisfying $L_0 = 0$ almost surely and having continuous in probability sample paths. For matrices $A \in \mathbb{R}^{N\times N}$, $B \in \mathbb{R}^{N\times d}$, $C \in \mathbb{R}^{m\times N}$ and a d-dimensional centered Lévy process $L = (L_t)_{t\geq 0}$ a continuous-time linear state space model (A, B, C, L) is defined by

$$\begin{aligned} \mathrm{d}X_t &= AX_t\mathrm{d}t + B\mathrm{d}L_t, \\ Y_t &= CX_t, \quad t \geq 0. \end{aligned} \tag{1}$$

The processes $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ in the state space representation (1) are called state and output process, respectively.

In the case of a finite second moment of the driving Lévy process the classes of stationary linear state space models and multivariate continuous-time ARMA (MCARMA) models are equivalent (see Schlemm and Stelzer 2012b, Corollary 3.4). This means that for every output process $(Y_t)_{t\geq 0}$ of the state space model (1) there exist an autoregressive polynomial $P^{(C)}(z) := I_m z^p + P_1 z^{p-1} + \cdots + P_{p-1}z + P_p$ with $P_1, \dots, P_p \in \mathbb{R}^{m\times m}$ and a moving average polynomial $Q^{(C)}(z) := Q_0 z^q + Q_1 z^{q-1} + \ldots + Q_{q-1}z + Q_q$ with $Q_0, \dots, Q_q \in \mathbb{R}^{m\times d}$ such that $(Y_t)_{t\geq 0}$ can be interpreted as solution of the differential equation

$$P^{(C)}(\mathsf{D})Y_t = Q^{(C)}(\mathsf{D})\mathsf{D}L_t, \quad t \geq 0, \tag{2}$$

where D is the differential operator with respect to t. Since the orders of the autoregressive polynomial and the moving average polynomial are p and q, Y is called MCARMA(p, q) process. Formally, MCARMA processes were introduced as linear state space models with special matrices A, B, C, see Marquardt and Stelzer (2007). Since the parametrization of a general linear state space model (1) is more flexible than the parametrization of an MCARMA model (2), it is advantageous to use (1) and estimate the parameters within this representation.

The defining differential Eq. (2) of an MCARMA process reminds of the defining difference equation of a discrete-time vector ARMA (VARMA) process. A VARMA process $(Z_n)_{n\in\mathbb{N}}$ is the m-dimensional solution of a difference equation of the form

$$P^{(D)}(\mathsf{B})Z_n = Q^{(D)}(\mathsf{B})e_n, \quad n \in \mathbb{N}, \tag{3}$$

where B is the Backshift-operator $\mathsf{B}Z_n = Z_{n-1}$ and $(e_n)_{n\in\mathbb{Z}}$ is a d-dimensional white noise, see, e.g., the monographs of Brockwell and Davis (1991) and Lütkepohl (2005). From Thornton and Chambers (2017), see Brockwell and Lindner (2009) for the univariate case, it is well known that a discretely sampled MCARMA process admits a VARMA representation with a weak white noise $(e_n)_{n\in\mathbb{Z}}$. The covariance matrix of e_n depends on the parameters of the polynomial P and Q in the MCARMA representation, respectively on the parameters of (A, B, C) in the state space model (1). For Lévy driven models the white noise of the sampled process is neither a strong white noise nor a martingale difference in general. Since the results concerning the asymptotic behavior of the quasi maximum likelihood estimator and the Whittle estimator for VARMA models require the white noise to be a martingale difference, see Dunsmuir and Hannan (1976), Deistler et al. (1978) and Dahlhaus and Pötscher (1989), they are not transferable to non-Gaussian Lévy driven state space models.

In the econometric literature there are several papers using the Kalman filter approach for maximum likelihood estimation of Gaussian possibly non-stationary MCARMA processes as, e.g., Harvey and Stock (1985, 1988, 1989), Zadrozny (1988) and Thornton and Chambers (2017). The rigorous mathematical derivation of the asymptotic properties of the quasi-maximum likelihood estimator for stationary Lévy driven state space and MCARMA models was given recently in Schlemm and Stelzer (2012a) and for non-stationary models in Fasen-Hartmann and Scholz (2019). In the case of univariate MCARMA processes with $d = m = 1$, which are called CARMA processes, there exist some further estimation methods. An indirect estimation procedure for CARMA models, which is robust against outliers, is the topic of Fasen-Hartmann and Kimmig (2020). To the best of our knowledge Fasen and Fuchs (2013) present the only frequency domain estimator for high-frequency sampled CARMA processes.

In this paper, we investigate a frequency domain estimator, the Whittle estimator, for a low-frequency sampled state space model (1) with stationary observations $Y_\Delta, \dots, Y_{n\Delta}$ ($\Delta > 0$ fixed). The Whittle estimator dates back to Whittle (1951, 1953), Walker (1964) and is very well investigated for different time series models in discrete time. If the autocovariance function of $Y^{(\Delta)} := (Y_{k\Delta})_{k\in\mathbb{N}_0}$ is denoted by $\Gamma^{(\Delta)}(h) = \mathrm{Cov}(Y_{(h+1)\Delta}, Y_\Delta)$ and $\Gamma^{(\Delta)}(-h) = \Gamma^{(\Delta)}(h)^\top$, $h \in \mathbb{N}_0$, the spectral density $f_{Y^{(\Delta)}}$ of $Y^{(\Delta)}$ is defined as the Fourier transform of the autocovariance function

$$f_{Y^{(\Delta)}}(\omega) = \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} \Gamma^{(\Delta)}(h) e^{-ih\omega}, \quad \omega \in [-\pi, \pi]. \tag{4}$$

Conversely, using the inverse Fourier transform, yields

$$\Gamma^{(\Delta)}(h) = \int_{-\pi}^{\pi} f_{Y^{(\Delta)}}(\omega) e^{ih\omega} \mathrm{d}\omega, \quad h \in \mathbb{Z}. \tag{5}$$

The empirical version of the spectral density is the periodogram $I_n : [-\pi, \pi] \to \mathbb{R}^{m\times m}$ defined as

$$I_n(\omega) = \frac{1}{2\pi n} \left(\sum_{j=1}^{n} Y_{j\Delta} e^{-ij\omega} \right) \left(\sum_{k=1}^{n} Y_{k\Delta} e^{ik\omega} \right)^\top = \frac{1}{2\pi} \sum_{h=-n+1}^{n-1} \overline{\Gamma}_n^{(\Delta)}(h) e^{-ih\omega}, \tag{6}$$

where

$$\overline{\Gamma}_n^{(\Delta)}(h) := \frac{1}{n} \sum_{k=1}^{n-h} Y_{(k+h)\Delta} Y_{k\Delta}^\top \quad \text{and} \quad \overline{\Gamma}_n^{(\Delta)}(-h) := \overline{\Gamma}_n^{(\Delta)}(h)^\top, \quad h = 0, \dots, n,$$

is the empirical autocovariance function. For different frequencies the periodogram behaves asymptotically like independent exponentially distributed random variables, see Fasen (2013), and is not a consistent estimator for the spectral density. However, the periodogram is the basic part of the Whittle estimator.

Let $\Theta \subseteq \mathbb{R}^r$ be a parameter space and for any $\vartheta \in \Theta$ let $f_{Y^{(\Delta)}}(\omega, \vartheta)$ be the spectral density of a stationary equidistant sampled state space process $Y^{(\Delta)}(\vartheta)$. Then, the Whittle function W_n is defined by

$$W_n(\vartheta) = \frac{1}{2n} \sum_{j=-n+1}^{n} \left[\operatorname{tr}\left(f_{Y^{(\Delta)}}(\omega_j, \vartheta)^{-1} I_n(\omega_j)\right) + \log\left(\det\left(f_{Y^{(\Delta)}}(\omega_j, \vartheta)\right)\right) \right],$$

with $\omega_j = \frac{\pi j}{n}$ for $j = -n+1, \ldots, n$ and the Whittle estimator is

$$\widehat{\vartheta}_n^{(\Delta)} := \arg\min_{\vartheta \in \Theta} W_n(\vartheta).$$

In the definition of the Whittle function it is also possible to replace the term $\log(\det(f_{Y^{(\Delta)}}(\omega_j, \vartheta)))$ by $\log(\det V^{(\Delta)}(\vartheta))$ where $V^{(\Delta)}(\vartheta)$ is the covariance matrix of the one-step linear prediction error. Therefore, if the covariance matrix $V^{(\Delta)}(\vartheta)$ of the linear prediction error does not depend on ϑ, we can neglect the penalty term $\log(\det V^{(\Delta)}(\vartheta))$ completely since it is constant for all ϑ. However, in the case of state space models, $V^{(\Delta)}(\vartheta)$ depends on ϑ and has to be computed additionally (cf. Proposition 1). Conversely, for VARMA models, $V^{(\Delta)}(\vartheta)$ is the covariance matrix of the white noise. Hence, the Whittle function for VARMA models with penalty function $\log(\det V^{(\Delta)}(\vartheta))$ in Dunsmuir and Hannan (1976) differs from our Whittle function. That paper is also one of the few papers using the Whittle estimator for the estimation of a multivariate model.

Empirical spectral processes indexed by a class of functions are applied to derive the asymptotic properties of frequency domain estimators as the Whittle estimator. The asymptotic behavior of empirical spectral processes is very well investigated but unfortunately the known results cannot be utilized to our setting. The empirical spectral process theory usually requires some exponential inequality and therefore some stronger model assumptions are necessary. For example, Mikosch and Norvaiša (1997) investigate empirical spectral processes for linear models with i.i.d. (independent and identically distributed) noise having finite fourth moments; similarly Dahlhaus and Polonik (2009), Dahlhaus (1988) assume some exponential moment condition for the stationary time series model and Dahlhaus and Polonik (2006) study Gaussian locally stationary processes. The recent paper of Bardet et al. (2008) assumes some weak dependence on the stationary time series and that the one-step linear prediction error variance, which corresponds to the variance of the white noise in the ARMA representation of the discrete sampled process, does not depend on the model parameters. However, in our case, the parameters of (A, B, C) affect this variance. Whittle estimation for continuous-time fractionally integrated CAR processes, where the driving process is a fractional Brownian motion, is studied in Tsai and Chan (2005). But essential for the proofs in that paper is again that the driving process is Gaussian such that the techniques cannot be used for Lévy driven models. Moreover, all of these papers only analyze univariate models, whereas we consider a multivariate model.

The paper is structured in the following way. We start by stating the basic facts on discrete-time sampled linear state space models in Sect. 2. Then, the main results of this paper are presented. In Sect. 3, we derive the consistency and the asymptotic normality of the Whittle estimator. Interesting is that for non-Gaussian state space models the limit covariance matrix of the Whittle estimator differs from the covariance matrix in the Gaussian case. As a contrast to Whittle estimation for VARMA models, this confirms that for the proofs standard techniques cannot be applied as well. An advantage of the Whittle estimator over the quasi-maximum likelihood estimator of Schlemm and Stelzer (2012a) is that we have an analytic representation of the limit covariance matrix which can be used for the determination of confidence bands. For the proof of the asymptotic normality of the Whittle estimator we show as well the asymptotic normality of the integrated periodogram. This result lays the basis for goodness-of-fit tests for state space models which can be written as continuous functionals of the integrated periodogram as, e.g., the Grenander and Rosenblatt test or Bartlett's test for the integrated periodogram, Bartlett's T_p test or the Cramér-von Mises test (cf. Priestley 1981), and is topic of some future research. Furthermore, results of this type are typically used for bootstraps in the frequency domain. In Sect. 4, we motivate the definition of the adjusted Whittle estimator, which works only for univariate state space models with $d = m = 1$, and present the consistency and the asymptotic normality for this estimator as well. Finally, the applicability of the Whittle and the adjusted Whittle estimator is demonstrated through a simulation study in Sect. 5 and compared to the quasi maximum likelihood estimator of Schlemm and Stelzer (2012a). For the Whittle estimator, the detailed proofs are given in Sect. 6 and since the proofs for the adjusted Whittle estimator are very similar, they are moved to Sect. 7 in the Supplementary Material. Some further simulation studies are presented there as well.

Notation For some matrix A, $\operatorname{tr}(A)$ stands for the trace of A, $\det(A)$ for its determinant, $A^\top$ for its transpose and A^H for the Hermitian of A. Further, $A[i,j]$ denotes the (i,j)-th component of A. We write $\operatorname{vec}(A)$ for the vectorization of A and $A \otimes B$ for the Kronecker product of A and B where B is any matrix. The N-dimensional identity matrix is denoted as I_N. For a matrix function $g(\vartheta)$ in $\mathbb{R}^{m\times s}$ with ϑ in $\mathbb{R}^r$ the gradient with respect to the parameter vector ϑ is denoted by $\nabla_\vartheta g(\vartheta) = \frac{\partial \operatorname{vec}(g(\vartheta))}{\partial \vartheta} \in \mathbb{R}^{ms\times r}$ and $\nabla_\vartheta g(\vartheta_0)$ is the shorthand for $\nabla_\vartheta g(\vartheta)|_{\vartheta=\vartheta_0}$. If $g : \mathbb{R}^r \to \mathbb{R}$, then $\nabla^2_\vartheta g(\vartheta) \in \mathbb{R}^{r\times r}$ denotes the Hessian matrix of $g(\vartheta)$. For the real and the imaginary part of a complex valued z, we use the notation $\mathfrak{R}(z)$ and $\mathfrak{I}(z)$, respectively. Throughout the article, $\|\cdot\|$ denotes an arbitrary sub-multiplicative matrix norm. Finally, $\mathfrak{C} > 0$ is a constant which may change from line to line.

2 Preliminaries

Let $\Theta \subset \mathbb{R}^r$ be a parameter space, and suppose that for any $\vartheta \in \Theta$, $A(\vartheta) \in \mathbb{R}^{N\times N}$ has eigenvalues with strictly negative real parts, $B(\vartheta) \in \mathbb{R}^{N\times d}$, $C(\vartheta) \in \mathbb{R}^{m\times N}$ and $L(\vartheta) := (L_t(\vartheta))_{t\in\mathbb{R}}$ is an $\mathbb{R}^d$-valued Lévy process with existing covariance matrix $\Sigma_L(\vartheta)$. A two-sided Lévy process can be constructed from two independent one-sided Lévy processes $(L_t^{(1)}(\vartheta))_{t\geq 0}$ and $(L_t^{(2)}(\vartheta))_{t\geq 0}$ through $L_t(\vartheta) = L_t^{(1)}(\vartheta)\mathbf{1}_{\{t\geq 0\}} - \lim_{s\uparrow -t} L_s^{(2)}(\vartheta)\mathbf{1}_{\{t<0\}}$. Details on Lévy processes can be found in Sato (1999). The stationary solution of the state space model

$$Y_t(\vartheta) = C(\vartheta)X_t(\vartheta) \quad \text{and} \quad \mathrm{d}X_t(\vartheta) = A(\vartheta)X_t(\vartheta)\mathrm{d}t + B(\vartheta)\mathrm{d}L_t(\vartheta), \quad t \geq 0,$$

has the representation

$$Y_t(\vartheta) = C(\vartheta)X_t(\vartheta) \quad \text{and} \quad X_t(\vartheta) = \int_{-\infty}^{t} \mathrm{e}^{A(\vartheta)(t-s)}B(\vartheta)\,\mathrm{d}L_s(\vartheta), \quad t \geq 0.$$

The true parameter of the output process Y of our observations $Y_\Delta, \dots, Y_{n\Delta}$ is denoted by ϑ_0 and is supposed to be in Θ. Since we only observe the output process of the state space model at discrete time points with distance $\Delta > 0$, we are interested in the probabilistic properties of $Y^{(\Delta)}(\vartheta) := (Y_k^{(\Delta)}(\vartheta))_{k\in\mathbb{N}_0} := (Y_{k\Delta}(\vartheta))_{k\in\mathbb{N}_0}$ as well. The discrete-time process $Y^{(\Delta)}(\vartheta)$ has the discrete-time state space representation

$$Y_k^{(\Delta)}(\vartheta) = C(\vartheta)X_k^{(\Delta)}(\vartheta) \quad \text{and} \quad X_k^{(\Delta)}(\vartheta) = e^{A(\vartheta)\Delta}X_{k-1}^{(\Delta)}(\vartheta) + N_k^{(\Delta)}(\vartheta), \quad k \in \mathbb{N}_0,$$

where

$$N_k^{(\Delta)}(\vartheta) = \int_{(k-1)\Delta}^{k\Delta} e^{A(\vartheta)(k\Delta-u)}B(\vartheta)\mathrm{d}L_u(\vartheta), \quad k \in \mathbb{N}_0, \tag{7}$$

is an i.i.d. sequence with mean zero and covariance matrix

$$\Sigma_N^{(\Delta)}(\vartheta) = \int_0^{\Delta} e^{A(\vartheta)u}B(\vartheta)\Sigma_L(\vartheta)B(\vartheta)^\top e^{A(\vartheta)^\top u}\mathrm{d}u$$

(see Schlemm and Stelzer 2012a, Proposition 3.6). Furthermore, $Y^{(\Delta)}(\vartheta)$ has the vector MA(∞) representation

$$Y_k^{(\Delta)}(\vartheta) = \sum_{j=0}^{\infty} \Phi_j(\vartheta)N_{k-j}^{(\Delta)}(\vartheta), \quad k \in \mathbb{N}_0,$$

where $\Phi_j(\vartheta) = C(\vartheta)e^{A(\vartheta)\Delta j} \in \mathbb{R}^{m\times N}$. Defining $\Phi(z,\vartheta) := \sum_{j=0}^{\infty} \Phi_j(\vartheta)z^j$, $z \in \mathbb{C}$, an application of Brockwell and Davis (1991), Theorem 11.8.3, gives the spectral density

$$
\begin{aligned}
f_{Y^{(\Delta)}}(\omega,\vartheta) &= \frac{1}{2\pi}\Phi(e^{-i\omega},\vartheta)\Sigma_N^{(\Delta)}(\vartheta)\Phi(e^{i\omega},\vartheta)^\top \\
&= \frac{1}{2\pi}C(\vartheta)\big(e^{i\omega}I_N - e^{A(\vartheta)\Delta}\big)^{-1}\Sigma_N^{(\Delta)}(\vartheta)\Big(e^{-i\omega}I_N - e^{A(\vartheta)^\top\Delta}\Big)^{-1}C(\vartheta)^\top,
\end{aligned}
\tag{8}
$$

of $Y^{(\Delta)}(\vartheta)$. For better readability, we will omit the true parameter ϑ_0 whenever possible and write $Y_k^{(\Delta)}, X_k^{(\Delta)}, f_{Y^{(\Delta)}}(\cdot), \ldots$ instead of $Y_k^{(\Delta)}(\vartheta_0), X_k^{(\Delta)}(\vartheta_0), f_{Y^{(\Delta)}}(\cdot,\vartheta_0), \ldots$.

To define the adjusted Whittle estimator and for the proof of the consistency of the Whittle estimator we introduce the linear innovations of $Y^{(\Delta)}(\vartheta)$.

Definition 1 The linear innovations $\varepsilon^{(\Delta)}(\vartheta) := (\varepsilon_k^{(\Delta)}(\vartheta))_{k\in\mathbb{N}}$ of $Y^{(\Delta)}(\vartheta)$ are defined by

$$
\begin{aligned}
&\varepsilon_k^{(\Delta)}(\vartheta) = Y_k^{(\Delta)}(\vartheta) - \mathrm{Pr}_{k-1}(\vartheta)Y_k^{(\Delta)}(\vartheta), \qquad \text{where} \\
&\mathrm{Pr}_k(\vartheta) = \text{ orthogonal projection onto } \mathcal{M}_k(\vartheta) := \overline{\mathrm{span}}\{Y_\nu^{(\Delta)}(\vartheta) : -\infty < \nu \le k\},
\end{aligned}
$$

where the closure is taken in the Hilbert space of random vectors with square-integrable components and inner product $(X,Y) \to \mathbb{E}\big[X^\top Y\big]$.

Adjusted to our notation, Proposition 2.1 of Schlemm and Stelzer (2012a) gives the following representation of the linear innovations of $Y^{(\Delta)}(\vartheta)$.

Proposition 1 *Suppose that the eigenvalues of $A(\vartheta)$ have strictly negative real parts and $\Sigma_L(\vartheta)$ is positive definite. Then, the following holds*:

(a) *The Riccati equation*

$$
\begin{aligned}
\Omega^{(\Delta)}(\vartheta) &= e^{A(\vartheta)\Delta}\Omega^{(\Delta)}(\vartheta)\big(e^{A(\vartheta)\Delta}\big)^\top + \Sigma_N^{(\Delta)}(\vartheta) \\
&\quad - \big(e^{A(\vartheta)\Delta}\Omega^{(\Delta)}(\vartheta)C(\vartheta)^\top\big)\big(C(\vartheta)\Omega^{(\Delta)}(\vartheta)C(\vartheta)^\top\big)^{-1}\big(e^{A(\vartheta)\Delta}\Omega^{(\Delta)}(\vartheta)C(\vartheta)^\top\big)^\top
\end{aligned}
$$

has a unique positive semidefinite solution $\Omega^{(\Delta)}(\vartheta)$.

(b) *Let*

$$
K^{(\Delta)}(\vartheta) = \big(e^{A(\vartheta)\Delta}\Omega^{(\Delta)}(\vartheta)C(\vartheta)^\top\big)\big(C(\vartheta)\Omega^{(\Delta)}(\vartheta)C(\vartheta)^\top\big)^{-1}
$$

be the Kalman gain matrix. Furthermore, define the polynomial Π as

$$
\begin{aligned}
\Pi(z,\vartheta) &:= \Pi^{(\Delta)}(z,\vartheta) \\
&:= \Big(I_m - C(\vartheta)\big(I_N - (e^{A(\vartheta)\Delta} - K^{(\Delta)}(\vartheta)C(\vartheta))z\big)^{-1}K^{(\Delta)}(\vartheta)z\Big).
\end{aligned}
$$

Then, the linear innovations are

$$
\varepsilon_k^{(\Delta)}(\vartheta) = \Pi(\mathsf{B},\vartheta)Y_k^{(\Delta)}(\vartheta), \quad k\in\mathbb{N}.
$$

Furthermore, the absolute value of any eigenvalue of $e^{A(\vartheta)\Delta} - K^{(\Delta)}(\vartheta)C(\vartheta)$ is less than one and $Y^{(\Delta)}(\vartheta)$ has the moving average representation

$$
\begin{aligned}
Y_k^{(\Delta)}(\vartheta) &= \varepsilon_k^{(\Delta)}(\vartheta) + C(\vartheta) \sum_{j=1}^{\infty} \left(e^{A(\vartheta)\Delta}\right)^{j-1} K^{(\Delta)}(\vartheta) \varepsilon_{k-j}^{(\Delta)}(\vartheta) \\
&=: \Pi^{-1}(\mathsf{B}, \vartheta) \varepsilon_k^{(\Delta)}(\vartheta).
\end{aligned} \tag{9}
$$

(c) *The covariance matrix $V^{(\Delta)}(\vartheta)$ of the linear innovations $\varepsilon^{(\Delta)}(\vartheta)$ has the representation $V^{(\Delta)}(\vartheta) = C(\vartheta)\Omega^{(\Delta)}(\vartheta)C(\vartheta)^\top$. If $\Omega^{(\Delta)}(\vartheta)$ is positive definite and $C(\vartheta)$ has full rank, $V^{(\Delta)}(\vartheta)$ is invertible.*

Note that $\operatorname{tr}(V^{(\Delta)}(\vartheta)) = \min_{X \in \mathcal{M}_{k-1}(\vartheta)} \mathbb{E}[(Y_k^{(\Delta)}(\vartheta) - X)^\top (Y_k^{(\Delta)}(\vartheta) - X)]$. An application of Brockwell and Davis (1991), Theorem 11.8.3, and (9) yield the representation

$$
f_{Y^{(\Delta)}}(\omega, \vartheta) = \Pi^{-1}(e^{-i\omega}, \vartheta) \frac{V^{(\Delta)}(\vartheta)}{2\pi} \Pi^{-1}(e^{i\omega}, \vartheta)^\top, \quad \omega \in [-\pi, \pi], \tag{10}
$$

for the spectral density of $Y^{(\Delta)}(\vartheta)$.

3 The Whittle estimator

3.1 Consistency of the Whittle estimator

Assumption A For all $\vartheta \in \Theta$ the following holds:

(A1) The parameter space Θ is a compact subset of $\mathbb{R}^r$.

(A2) $L(\vartheta) = (L_t(\vartheta))_{t \in \mathbb{R}}$ is a centered Lévy process with positive definite covariance matrix $\Sigma_L(\vartheta)$.

(A3) The eigenvalues of $A(\vartheta)$ have strictly negative real parts.

(A4) The functions $\vartheta \mapsto \Sigma_L(\vartheta)$, $\vartheta \mapsto A(\vartheta)$, $\vartheta \mapsto B(\vartheta)$ and $\vartheta \mapsto C(\vartheta)$ are continuous. In addition, $C(\vartheta)$ has full rank.

(A5) The linear state space model $(A(\vartheta), B(\vartheta), C(\vartheta), L(\vartheta))$ is minimal with McMillan degree N, i.e., there exist no integer $\widetilde{N} < N$ and matrices $\tilde{A} \in \mathbb{R}^{\widetilde{N} \times \widetilde{N}}$, $\widetilde{B} \in \mathbb{R}^{\widetilde{N} \times d}$ and $\widetilde{C} \in \mathbb{R}^{m \times \widetilde{N}}$ with $C(\vartheta)(zI_N - A(\vartheta))^{-1}B(\vartheta) = \widetilde{C}(zI_{\widetilde{N}} - \widetilde{A})^{-1}\widetilde{B}$ for all $z \in \mathbb{R}$.

(A6) For any $\vartheta_1, \vartheta_2 \in \Theta$ with $\vartheta_1 \neq \vartheta_2$ there exists an $\omega \in [-\pi, \pi]$ such that $f_Y(\omega, \vartheta_1) \neq f_Y(\omega, \vartheta_2)$, where $f_Y(\omega, \vartheta)$ is the spectral density of $Y(\vartheta)$.

(A7) The spectrum of $A(\vartheta) \in \mathbb{R}^{N \times N}$ is a subset of $\left\{ z \in \mathbb{C} : -\frac{\pi}{\Delta} < \mathfrak{I}(z) < \frac{\pi}{\Delta} \right\}$.

Remark 1

(a) Note that Assumptions (A2) and (A3) allow us to calculate the linear innovations. Furthermore, the covariance matrix $V^{(\Delta)}(\vartheta)$ of the linear innovations is non-singular (cf. Lemma 3.14 in Schlemm and Stelzer 2012a).

(b) Theorem 2.3.4 in Hannan and Deistler (1988) shows that (A5) guarantees the uniqueness of the state space representation $(A(\vartheta), B(\vartheta), C(\vartheta), L(\vartheta))$ up to a change of basis. Hence, (A5) reduces redundancies in the continuous-time model. In addition, Schlemm and Stelzer (2012a), Theorem 3.13, proved that Assumptions (A2)–(A7) provide Δ-identifiability of the collection of output processes $(Y(\vartheta), \vartheta \in \Theta)$, i.e., for fixed $\Delta > 0$ and arbitrary $\vartheta_1, \vartheta_2 \in \Theta$ with $\vartheta_1 \neq \vartheta_2$, there exists an $\omega \in [-\pi, \pi]$ with $f_{Y^{(\Delta)}}(\omega, \vartheta_1) \neq f_{Y^{(\Delta)}}(\omega, \vartheta_2)$. Hence, Assumption (A7) is an anti-aliasing condition, i.e., it prevents redundancies due to the sampling process. For more details on overcoming the aliasing effect, we refer to Phillips (1973), Hansen and Sergant (1983), Chambers et al. (2018) and Blevins (2017).

(c) Assumptions (A2) and (A5) imply that $\Sigma_N^{(\Delta)}(\vartheta)$ has full rank.

(d) Under Assumption A and representation (8) of the spectral density, the inverse $f_{Y^{(\Delta)}}(\omega, \vartheta)^{-1}$ of the spectral density exists and the mapping $(\vartheta, \omega) \mapsto f_{Y^{(\Delta)}}(\omega, \vartheta)^{-1}$ is continuous.

We start to prove some auxiliary results which we need for the proof of the consistency of Whittle‘s estimator. The following proposition states that the Whittle function W_n converges almost surely uniformly.

Proposition 2 *Let Assumptions* (A1)–(A4) *hold and*

$$W(\vartheta) := \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{tr}\left(f_{Y^{(\Delta)}}(\omega, \vartheta)^{-1} f_{Y^{(\Delta)}}(\omega)\right) + \log\left(\det\left(f_{Y^{(\Delta)}}(\omega, \vartheta)\right)\right) d\omega, \quad \vartheta \in \Theta.$$

Then,

$$\sup_{\vartheta \in \Theta} \left| W_n(\vartheta) - W(\vartheta) \right| \xrightarrow{n \to \infty} 0 \quad \mathbb{P}\text{-a.s.}$$

Obviously, it is necessary that ϑ_0 is a global minimum of W to guarantee the consistency of the Whittle estimator.

Proposition 3 *Let Assumptions* (A1)–(A4) *and* (A6) *hold. Then*, *W has a unique global minimum in* ϑ_0.

The proof is based on an alternative representation of W. Namely, the function W is exactly the limit function of the quasi maximum likelihood estimator of Schlemm and Stelzer (2012a).

Lemma 1 *Let Assumptions* (A1)–(A4) *hold and let* $\xi_k^{(\Delta)}(\vartheta) = \Pi(\mathsf{B}, \vartheta) Y_k^{(\Delta)}$ *with* $\Pi(z, \vartheta)$ *as given in Proposition 1. Furthermore, define for* $\vartheta \in \Theta$

$$\mathcal{L}(\vartheta) := \mathbb{E}\left[\operatorname{tr}\left(\xi_1^{(\Delta)}(\vartheta)^\top V^{(\Delta)}(\vartheta)^{-1} \xi_1^{(\Delta)}(\vartheta)\right)\right] + \log(\det(V^{(\Delta)}(\vartheta))) - m \log(2\pi).$$

Then, $W(\vartheta) = \mathcal{L}(\vartheta)$ *for* $\vartheta \in \Theta$.

Finally, we are able to state the first main result of this paper, which gives the consistency of the Whittle estimator.

Theorem 1 *Let Assumption A hold. Then, as $n \to \infty$,*

$$\widehat{\vartheta}_n^{(\Delta)} \xrightarrow{\text{a.s.}} \vartheta_0.$$

3.2 Asymptotic normality of the Whittle estimator

For the asymptotic normality of the Whittle estimator some further assumptions are required.

Assumption B

(B1) The true parameter value ϑ_0 is in the interior of Θ.
(B2) $\mathbb{E}\|L_1\|^4 < \infty$.
(B3) The functions $\vartheta \mapsto A(\vartheta)$, $\vartheta \mapsto B(\vartheta)$, $\vartheta \mapsto C(\vartheta)$ and $\vartheta \mapsto \Sigma_L(\vartheta)$ are three times continuously differentiable.
(B4) For any $c \in \mathbb{C}^r$, there exists an $\omega^* \in [-\pi, \pi]$ such that $\nabla_\vartheta f_{Y^{(\Delta)}}(\omega^*, \vartheta_0)c \neq 0_{m^2}$.

Remark 2

(a) Due to representation (8) of the spectral density, under Assumption A and (B3) the mapping $\vartheta \mapsto f_{Y^{(\Delta)}}(\omega, \vartheta)$ is three times continuously differentiable.
(b) Note that for the proof of the asymptotic normality of the quasi maximum likelihood estimator, Schlemm and Stelzer applied a covariance inequality which requires that the driving process has more than four moments, see Lemma 2.13 of Schlemm and Stelzer (2012a). In contrast, we only need the existence of the fourth moment of the driving process. This is needed, since we apply a central limit theorem for the autocovariance function of the white noise.

The proof of the asymptotic normality of the Whittle estimator is based on a Taylor expansion of $\nabla_\vartheta W_n$ around $\widehat{\vartheta}_n^{(\Delta)}$ in ϑ_0, i.e.,

$$\sqrt{n}\left[\nabla_\vartheta W_n(\vartheta_0)\right] = \sqrt{n}\left[\nabla_\vartheta W_n(\widehat{\vartheta}_n^{(\Delta)})\right] - \sqrt{n}(\widehat{\vartheta}_n^{(\Delta)} - \vartheta_0)^\top \left[\nabla_\vartheta^2 W_n(\vartheta_n^*)\right] \tag{11}$$

for an appropriate $\vartheta_n^* \in \Theta$ with $\|\vartheta_n^* - \vartheta_0\| \leq \|\widehat{\vartheta}_n^{(\Delta)} - \vartheta_0\|$. Since $\widehat{\vartheta}_n^{(\Delta)}$ minimizes W_n and converges almost surely to ϑ_0, which is in the interior of Θ (Assumption (B1)), $\nabla_\vartheta W_n(\widehat{\vartheta}_n^{(\Delta)}) = 0$. Hence, in the case of an invertible matrix $\nabla_\vartheta^2 W_n(\vartheta_n^*)$ we can rewrite (11) and obtain

$$\sqrt{n}(\widehat{\vartheta}_n^{(\Delta)} - \vartheta_0)^\top = -\sqrt{n}\left[\nabla_\vartheta W_n(\vartheta_0)\right]\left[\nabla_\vartheta^2 W_n(\vartheta_n^*)\right]^{-1}. \tag{12}$$

Therefore, we receive the asymptotic normality of the Whittle estimator from the asymptotic behavior of the individual components in (12).

First, we investigate the asymptotic behavior of the Hessian matrix $\nabla_\vartheta^2 W_n(\vartheta_n^*)$.

Proposition 4 *Let Assumptions* $(A1)$–$(A4)$ *and* $(B3)$ *hold and*

$$\Sigma_{\nabla^2 W} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \nabla_\vartheta f_{Y^{(\Delta)}}(-\omega,\vartheta_0)^\top \left[f_{Y^{(\Delta)}}(-\omega)^{-1} \otimes f_{Y^{(\Delta)}}(\omega)^{-1}\right] \nabla_\vartheta f_{Y^{(\Delta)}}(\omega,\vartheta_0)\mathrm{d}\omega. \tag{13}$$

Furthermore, let $(\vartheta_n^*)_{n\in\mathbb{N}}$ *be a sequence in* Θ *with* $\vartheta_n^* \xrightarrow{\text{a.s.}} \vartheta_0$ *as* $n\to\infty$. *Then, as* $n\to\infty$,

$$\nabla_\vartheta^2 W_n(\vartheta_n^*) \xrightarrow{\text{a.s.}} \Sigma_{\nabla^2 W}.$$

Further, we require that for large n the random matrix $\nabla_\vartheta^2 W_n(\vartheta_n^*)$ is invertible. Therefore, we show the positive definiteness of the limit matrix $\Sigma_{\nabla^2 W}$.

Lemma 2 *Let Assumptions A and* $(B4)$ *hold. Then,* $\Sigma_{\nabla^2 W}$ *is positive definite.*

Remark 3 For Gaussian state space processes

$$\begin{aligned} J = \Bigg[2\mathbb{E}\left[\left(\frac{\partial}{\partial\vartheta_i}\varepsilon_1^{(\Delta)}(\vartheta_0)\right)^\top V^{(\Delta)-1}\left(\frac{\partial}{\partial\vartheta_j}\varepsilon_1^{(\Delta)}(\vartheta_0)\right)\right] \\ +\operatorname{tr}\left(\left(\frac{\partial}{\partial\vartheta_i}V^{(\Delta)}(\vartheta_0)\right)V^{(\Delta)-1}\left(\frac{\partial}{\partial\vartheta_j}V^{(\Delta)}(\vartheta_0)\right)V^{(\Delta)-1}\right)\Bigg]_{i,j=1,\ldots,r} \end{aligned}$$

is the Fisher information matrix (cf. Schlemm and Stelzer 2012a). Since $W(\vartheta) = \mathcal{L}(\vartheta)$ due to Lemma 1, and $\nabla_\vartheta f_{Y^{(\Delta)}}(\omega,\vartheta)$ is uniformly bounded by an integrable dominant, we get by some straightforward applications of dominated convergence and some arguments of the proof of Schlemm and Stelzer (2012a), Lemma 2.17, that

$$\begin{aligned} J[i,j] &= \lim_{n\to\infty}\mathbb{E}\left[\frac{\partial}{\partial\vartheta_i}\frac{\partial}{\partial\vartheta_j}\mathcal{L}_n(\vartheta_0)\right] = \frac{\partial}{\partial\vartheta_i}\frac{\partial}{\partial\vartheta_j}\lim_{n\to\infty}\mathbb{E}[\mathcal{L}_n(\vartheta_0)] \\ &= \frac{\partial}{\partial\vartheta_i}\frac{\partial}{\partial\vartheta_j}W(\vartheta_0) = \lim_{n\to\infty}\mathbb{E}\left[\frac{\partial}{\partial\vartheta_i}\frac{\partial}{\partial\vartheta_j}W_n(\vartheta_0)\right] = \Sigma_{\nabla^2 W}[i,j], \end{aligned}$$

where $\mathcal{L}_n(\vartheta)$ is the quasi-Gaussian likelihood function. Furthermore, Schlemm and Stelzer (2012a), Lemma 2.17, show that if Assumption A holds and if there exists an $j_0\in\mathbb{N}$ such that the $((j_0+2)m^2)\times r$-matrix

$$\nabla\begin{bmatrix} \left[I_{j_0+1}\otimes K^{(\Delta)}(\vartheta_0)^\top \otimes C(\vartheta_0)\right]\left[\left(\operatorname{vec}\left(e^{I_N\Delta}\right)\right)^\top\left(\operatorname{vec}\left(e^{A(\vartheta_0)\Delta}\right)\right)^\top\cdots\left(\operatorname{vec}\left(e^{A^j(\vartheta_0)\Delta}\right)\right)^\top\right]^\top \\ \operatorname{vec}\left(V^{(\Delta)}(\vartheta_0)\right)\end{bmatrix}$$

has rank r, then the matrix J is positive definite. Thus, our assumption (B4) can be replaced by this condition.

Next, we investigate the asymptotic behavior of the second term in (12). Since the components of the score $\nabla_\vartheta W_n(\vartheta_0)$ can be written as an integrated periodogram, we first derive the asymptotic behavior of the integrated periodogram and state the asymptotic normality afterward.

Proposition 5 *Let Assumptions* (A2)–(A4) *and* (B2) *hold. Suppose* $\eta : [-\pi,\pi] \to \mathbb{C}^{m\times m}\eta : [-\pi,\pi] \to \mathbb{C}^{m\times m}$ *is a matrix-valued continuous function with* $\eta(\omega) = \eta(\omega)^H, \omega \in [-\pi,\pi]$, *and Fourier coefficients* $(\mathfrak{f}_u)_{u\in\mathbb{Z}}$ *satisfying* $\sum_{u=-\infty}^{\infty} \|\mathfrak{f}_u\| |u|^{1/2} < \infty$. *Then, as* $n \to \infty$,

$$\frac{1}{2\sqrt{n}} \sum_{j=-n+1}^{n} \operatorname{tr}\big(\eta(\omega_j) I_n(\omega_j) - \eta(\omega_j) f_{Y^{(\Delta)}}(\omega_j)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_\eta),$$

where

$$\begin{aligned}\Sigma_\eta = {} & \frac{1}{\pi} \int_{-\pi}^{\pi} \operatorname{tr}\big(\eta(\omega) f_{Y^{(\Delta)}}(\omega) \eta(\omega) f_{Y^{(\Delta)}}(\omega)\big) \mathrm{d}\omega \\ & + \frac{1}{16\pi^4} \int_{-\pi}^{\pi} \operatorname{vec}\big(\Phi(e^{-i\omega})^\top \eta(\omega)^\top \Phi(e^{i\omega})\big)^\top \mathrm{d}\omega \\ & \cdot \left(\mathbb{E}\left[N_1^{(\Delta)} N_1^{(\Delta)\top} \otimes N_1^{(\Delta)} N_1^{(\Delta)\top}\right] - 3\Sigma_N^{(\Delta)} \otimes \Sigma_N^{(\Delta)}\right) \\ & \cdot \int_{-\pi}^{\pi} \operatorname{vec}\big(\Phi(e^{i\omega})^\top \eta(\omega) \Phi(e^{-i\omega})\big) \mathrm{d}\omega.\end{aligned}$$

The asymptotic behavior of the integrated periodogram is interesting for its own. It can be modified to derive goodness-of-fit tests for state space models which are continuous functionals of the integrated periodogram (cf. Priestley 1981).

Remark 4 Let the driving Lévy process be a Brownian motion. Since the fourth moment of a centered normal distribution is equal to three times its second moment and $N_1^{(\Delta)} \overset{\mathcal{D}}{\sim} \mathcal{N}(0, \Sigma_N^{(\Delta)})$, we get $\mathbb{E}[N_1^{(\Delta)} N_1^{(\Delta)\top} \otimes N_1^{(\Delta)} N_1^{(\Delta)\top}] = 3\Sigma_N^{(\Delta)} \otimes \Sigma_N^{(\Delta)}$. Therefore, the matrix Σ_η in Proposition 5 reduces to

$$\Sigma_\eta = \frac{1}{\pi} \int_{-\pi}^{\pi} \operatorname{tr}\big(\eta(\omega) f_{Y^{(\Delta)}}(\omega) \eta(\omega) f_{Y^{(\Delta)}}(\omega)\big) \mathrm{d}\omega,$$

which is for $m = 1$ equal to $\Sigma_\eta = \frac{1}{\pi} \int_{-\pi}^{\pi} \eta(\omega)^2 f_{Y^{(\Delta)}}(\omega)^2 \mathrm{d}\omega$.

Finally, we obtain the asymptotic behavior of the score function.

Proposition 6 *Let Assumptions* (A2)–(A4) *and* (B2)–(B3) *hold. Define*

$$
\begin{aligned}
\Sigma_{\nabla W} &= \frac{1}{\pi}\int_{-\pi}^{\pi} \nabla_{\vartheta} f_{Y^{(\Delta)}}(-\omega,\vartheta_0)^{\top}\left[f_{Y^{(\Delta)}}(-\omega)^{-1}\otimes f_{Y^{(\Delta)}}(\omega)^{-1}\right]\nabla_{\vartheta} f_{Y^{(\Delta)}}(\omega,\vartheta_0)\mathrm{d}\omega \\
&\quad + \frac{1}{16\pi^4}\left[\int_{-\pi}^{\pi}\left[\Phi(e^{i\omega})^{\top} f_{Y^{(\Delta)}}(\omega)^{-1}\otimes \Phi(e^{-i\omega})^{\top} f_{Y^{(\Delta)}}(-\omega)^{-1}\right]\nabla_{\vartheta} f_{Y^{(\Delta)}}(-\omega,\vartheta_0)\mathrm{d}\omega\right]^{\top} \\
&\quad \cdot\left[\mathbb{E}\left[N_1^{(\Delta)}N_1^{(\Delta)\top}\otimes N_1^{(\Delta)}N_1^{(\Delta)\top}\right]-3\Sigma_N^{(\Delta)}\otimes\Sigma_N^{(\Delta)}\right] \\
&\quad \cdot\left[\int_{-\pi}^{\pi}\left[\Phi(e^{-i\omega})^{\top} f_{Y^{(\Delta)}}(-\omega)^{-1}\otimes \Phi(e^{i\omega})^{\top} f_{Y^{(\Delta)}}(\omega)^{-1}\right]\nabla_{\vartheta} f_{Y^{(\Delta)}}(\omega,\vartheta_0)\mathrm{d}\omega\right].
\end{aligned}
\tag{14}
$$

Then, as $n \to \infty$,

$$
\sqrt{n}\left[\nabla_{\vartheta} W_n(\vartheta_0)\right]^{\top} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma_{\nabla W}).
$$

Now, we are able to present the main result of this paper, the asymptotic normality of the Whittle estimator.

Theorem 2 *Let Assumptions A and B hold. Furthermore, let $\Sigma_{\nabla W}$ be defined as in* (14) *and $\Sigma_{\nabla^2 W}$ be defined as in* (13). *Then, as $n \to \infty$,*

$$
\sqrt{n}\left(\widehat{\vartheta}_n^{(\Delta)}-\vartheta_0\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma_W),
$$

where Σ_W has the representation $\Sigma_W = [\Sigma_{\nabla^2 W}]^{-1}\Sigma_{\nabla W}[\Sigma_{\nabla^2 W}]^{-1}$.

In contrast to the quasi maximum likelihood estimator of Schlemm and Stelzer (2012a), the limit covariance matrix of the Whittle estimator has an analytic representation. It can be used for the calculation of confidence bands.

Remark 5 We want to compare our outcome with an analog result for stationary discrete-time VARMA(p,q) processes $(Z_n)_{n\in\mathbb{N}}$ of the form (3) with finite fourth moments. In our setting we have the drawback that the autoregressive and the moving average polynomial influence the covariance matrix $\Sigma_N^{(\Delta)}$ of $(N_k^{(\Delta)})_{k\in\mathbb{N}_0}$. In the setting of stationary VARMA(p,q) processes of Dunsmuir and Hannan (1976) the covariance matrix Σ_e of the white noise $(e_n)_{n\in\mathbb{Z}}$ is not affected by the AR and MA polynomials. It was shown in Dunsmuir and Hannan (1976) that under very general assumptions for $d = m$ the resulting limit covariance matrix of the Whittle estimator for the VARMA parameters has the representation

$$
\begin{aligned}
\Sigma_W^{\mathrm{VARMA}} &= \left[\frac{1}{4\pi}\int_{-\pi}^{\pi}\nabla_{\vartheta} f_Z(-\omega,\vartheta_0)^{\top}\left[f_Z(-\omega)^{-1}\otimes f_Z(\omega)^{-1}\right]\nabla_{\vartheta} f_Z(\omega,\vartheta_0)\mathrm{d}\omega\right]^{-1} \\
&= 2\cdot[\Sigma_{\nabla^2 W}^{\mathrm{VARMA}}]^{-1},
\end{aligned}
$$

which is simpler than our Σ_W. This can be traced back to $\Sigma_{\nabla W}^{\mathrm{VARMA}} = 2\cdot\Sigma_{\nabla^2 W}^{\mathrm{VARMA}}$, which is motivated on p. 38. In particular, for a Gaussian VARMA model, $\Sigma_W^{\mathrm{VARMA}}$ is the inverse of the Fisher information matrix.

Remark 6

(a) Let the driving Lévy process be a Brownian motion. Due to Remark 4, the matrix $\Sigma_{\nabla W}$ reduces to

$$\Sigma_{\nabla W} = \frac{1}{\pi}\int_{-\pi}^{\pi} \nabla_{\vartheta} f_{Y^{(\Delta)}}(-\omega,\vartheta_0)^{\top}\left[f_{Y^{(\Delta)}}(-\omega)^{-1}\otimes f_{Y^{(\Delta)}}(\omega)^{-1}\right]\nabla_{\vartheta} f_{Y^{(\Delta)}}(\omega,\vartheta_0)\mathrm{d}\omega$$
$$= 2\cdot[\Sigma_{\nabla^2 W}]^{-1},$$

and hence, $\Sigma_W = 2\cdot[\Sigma_{\nabla W}]^{-1}$ is the inverse of the Fisher information matrix and corresponds to $\Sigma_W^{\mathrm{VARMA}}$ as in the previously mentioned discrete-time VARMA setting.

(b) Let $d=m=N$ and $C(\vartheta)=I_m$. Then, the state space model is a multivariate Ornstein–Uhlenbeck process (MCAR(1) process). In this example, $\Sigma_{\nabla W}=2\cdot[\Sigma_{\nabla^2 W}]^{-1}$ holds as well. Because of $\Phi(z,\vartheta)=\sum_{j=0}^{\infty} e^{A(\vartheta)\Delta j}z^j=(1-e^{A(\vartheta)\Delta}z)^{-1}=\Pi^{-1}(z,\vartheta)$, the arguments are very similar to the arguments for VARMA models in Remark 5.

4 The adjusted Whittle estimator

In the following, we solely consider state space models where Y and L are one-dimensional, i.e., $A\in\mathbb{R}^{N\times N}$, $B\in\mathbb{R}^{N\times 1}$ and $C\in\mathbb{R}^{1\times N}$. This includes, in particular, univariate CARMA processes, see, e.g., Brockwell and Lindner (2009) and Brockwell (2014) for the explicit definition and existence criteria. Further, we assume that the variance parameter σ_L^2 of the driving Lévy process does not depend on ϑ and has not to be estimated. In this context, we consider an adjusted Whittle estimator which takes into account that we do not have to estimate the variance. Such adjusted Whittle estimators are useful for the estimation of heavy tailed CARMA models with infinite variance. For example, Mikosch et al. (1995) estimate the parameters of ARMA models in discrete time whose noise has a symmetric stable distribution. In some future work we will investigate such an adjusted Whittle estimator for heavy tailed models as well.

Now, the Whittle function is adapted in a way which makes it independent of the variance of the driving Lévy process. Therefore, we use the representation of the spectral density in (10). Although the variance σ_L^2 goes linearly in $\Omega^{(\Delta)}(\vartheta)$ and $V^{(\Delta)}(\vartheta)$, both $K^{(\Delta)}(\vartheta)$ and $\Pi(z,\vartheta)$ do not depend on σ_L^2 anymore. The second summand of the Whittle function W_n is removed and the first term is adjusted so that we obtain the *adjusted Whittle function*

$$W_n^{(A)}(\vartheta)=\frac{\pi}{n}\sum_{j=-n+1}^{n}|\Pi(e^{i\omega_j},\vartheta)|^2 I_n(\omega_j)=\frac{V^{(\Delta)}(\vartheta)}{2n}\sum_{j=-n+1}^{n} f_{Y^{(\Delta)}}(\omega_j,\vartheta)^{-1}I_n(\omega_j).$$

The corresponding minimizer

$$\widehat{\vartheta}_n^{(\Delta,A)} = \arg\min_{\vartheta\in\Theta} W_n^{(A)}(\vartheta)$$

is the *adjusted Whittle estimator*.

4.1 Consistency of the adjusted Whittle estimator

Since the estimation procedure is different to that of the previous sections, we have to adjust Assumption A.

Assumption $\widetilde{A}$ Let Assumptions (A1)–(A5) and (A7) hold. Furthermore, assume
($\widetilde{A}$6) For any ϑ_1, $\vartheta_2 \in \Theta$, $\vartheta_1 \neq \vartheta_2$, there exists some $z \in \mathbb{C}$ with $|z| = 1$ and $\Pi(z, \vartheta_1) \neq \Pi(z, \vartheta_2)$.

It is needless to say that conditions as those for the function $\vartheta \to \sigma_L^2$ are fulfilled naturally. In addition to Remark 1, which is still valid, we stress that, under Assumption $\widetilde{A}$, Π^{-1} as defined in (9) exists for all $\vartheta \in \Theta$ and that the mapping $(\omega, \vartheta) \to \Pi^{-1}(e^{i\omega}, \vartheta)$ is continuous.

Theorem 3 *Let Assumption $\widetilde{A}$ hold. Then, as $n \to \infty$,*

$$\widehat{\vartheta}_n^{(\Delta,A)} \xrightarrow{\text{a.s.}} \vartheta_0.$$

The proof follows the same steps as the proof for the consistency of the Whittle estimator in Theorem 1.

4.2 Asymptotic normality of the adjusted Whittle estimator

For the asymptotic normality of the adjusted Whittle estimator we have to adapt Assumption B.

Assumption $\widetilde{B}$ Let Assumptions (B1)–(B3) hold. Furthermore, assume
($\widetilde{B}$4) For any $c \in \mathbb{C}^r$ there exists an $\omega^* \in [-\pi, \pi]$ such that $\nabla_\vartheta |\Pi(e^{i\omega^*}, \vartheta_0)|^{-2} c \neq 0$.

Remark 7 Under Assumption $\widetilde{A}$ and Assumption $\widetilde{B}$ the mapping $\vartheta \to \Pi(e^{i\omega}, \vartheta)$ is three times continuously differentiable. Similarly to Lemma 2, ($\widetilde{B}$4) guarantees the invertibility of

$$\Sigma_{\nabla^2 W^{(\Delta)}} := \frac{V^{(\Delta)}}{2\pi}\int_{-\pi}^{\pi}\nabla_\vartheta \log\left(|\Pi(e^{i\omega},\vartheta_0)|^{-2}\right)^\top \nabla_\vartheta \log\left(|\Pi(e^{i\omega},\vartheta_0)|^{-2}\right)\mathrm{d}\omega. \quad (15)$$

Theorem 4 *Let Assumption $\widetilde{A}$ and $\widetilde{B}$ hold. Further, let $\Sigma_{\nabla^2 W^{(\Delta)}}$ be defined as in (15) and*

$$\begin{aligned}
\Sigma_{\nabla W^{(\Delta)}} = {} & \frac{V^{(\Delta)2}}{\pi}\int_{-\pi}^{\pi}\nabla_\vartheta \log\left(|\Pi(e^{i\omega},\vartheta_0)|^{-2}\right)^\top \nabla_\vartheta \log\left(|\Pi(e^{i\omega},\vartheta_0)|^{-2}\right)\mathrm{d}\omega \\
& + \frac{1}{4\pi^2}\left[\int_{-\pi}^{\pi}\nabla_\vartheta |\Pi(e^{i\omega},\vartheta_0)|^{2\top}\left[\Phi(e^{i\omega})\otimes\Phi(e^{-i\omega})\right]\mathrm{d}\omega\right] \\
& \cdot\left[\mathbb{E}\left[N_1^{(\Delta)}N_1^{(\Delta)\top}\otimes N_1^{(\Delta)}N_1^{(\Delta)\top}\right] - 3\Sigma_N^{(\Delta)}\otimes\Sigma_N^{(\Delta)}\right] \\
& \cdot\left[\int_{-\pi}^{\pi}\nabla_\vartheta |\Pi(e^{i\omega},\vartheta_0)|^{2\top}\left[\Phi(e^{-i\omega})\otimes\Phi(e^{i\omega})\right]\mathrm{d}\omega\right]^\top .
\end{aligned}$$

Then, as $n\to\infty$,

$$\sqrt{n}\left(\widehat{\vartheta}_n^{(\Delta,A)} - \vartheta_0\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma_{W^{(\Delta)}}),$$

where $\Sigma_{W^{(\Delta)}}$ has the representation $\Sigma_{W^{(\Delta)}} = [\Sigma_{\nabla^2 W^{(\Delta)}}]^{-1}\Sigma_{\nabla W^{(\Delta)}}[\Sigma_{\nabla^2 W^{(\Delta)}}]^{-1}$.

Remark 8 For the one-dimensional CAR(1) (Ornstein–Uhlenbeck) process, for which $m=d=N=1$ and $C(\vartheta)=B(\vartheta)=1$ holds, the limit covariance matrix $\Sigma_{W^{(\Delta)}}$ of Theorem 4 reduces due to Remark 9 in the Supplementary Material and Theorem 3''', Chapter 3, of Hannan (2009) to

$$\begin{aligned}
\Sigma_{W^{(\Delta)}} &= 4\pi\left[\int_{-\pi}^{\pi}\nabla_\vartheta \log\left(|\Pi(e^{i\omega},\vartheta_0)|^{-2}\right)^\top \nabla_\vartheta \log\left(|\Pi(e^{i\omega},\vartheta_0)|^{-2}\right)\mathrm{d}\omega\right]^{-1} \\
&= 4\pi\left[\int_{-\pi}^{\pi}\nabla_\vartheta \log(f_{Y^{(\Delta)}}(\omega,\vartheta_0))^\top \nabla_\vartheta \log(f_{Y^{(\Delta)}}(\omega,\vartheta_0))\mathrm{d}\omega\right. \\
&\qquad \left.-\frac{1}{2\pi}\left(\int_{-\pi}^{\pi}\nabla_\vartheta \log(f_{Y^{(\Delta)}}(\omega,\vartheta_0))\mathrm{d}\omega\right)^\top\left(\int_{-\pi}^{\pi}\nabla_\vartheta \log(f_{Y^{(\Delta)}}(\omega,\vartheta_0))\mathrm{d}\omega\right)\right]^{-1}.
\end{aligned}$$

Due to Remark 6(b)

$$\Sigma_W = 2\cdot[\Sigma_{\nabla^2 W}]^{-1} = 4\pi\left[\int_{-\pi}^{\pi}\nabla_\vartheta \log(f_{Y^{(\Delta)}}(\omega,\vartheta_0))^\top \nabla_\vartheta \log(f_{Y^{(\Delta)}}(\omega,\vartheta_0))\mathrm{d}\omega\right]^{-1}$$

and hence, $\Sigma_{W^{(\Delta)}} \geq \Sigma_W$. Thus, the adjusted Whittle estimator has a higher variance than the Whittle estimator. Let $\vartheta_0 < 0$ be the zero of the AR polynomial in the CAR(1) model, i.e., $A(\vartheta_0)=\vartheta_0$. Simple calculations show that $\Sigma_{W^{(\Delta)}} = e^{-2\vartheta_0}-1$ which is equal to the asymptotic variance of the maximum likelihood estimator of Brockwell and Lindner (2019). However, it is not possible to make this conclusion for general CARMA processes. There exist CARMA processes for which the adjusted Whittle estimator has a different asymptotic variance than the maximum likelihood estimator of Brockwell and Lindner (2019).

5 Simulation

In this section, we show the practical applicability of the Whittle and the adjusted Whittle estimator. We simulate continuous-time state space models with an Euler-Maruyama scheme for differential equations with initial value $X(0) = Y(0) = 0$ and step size 0.01. Using $\Delta = 1$ and the interval [0, 500], we therefore get $n_1 = 500$ discrete observations. Furthermore, we investigate how the results change qualitatively when we consider the intervals [0, 2000] and [0, 5000], which imply $n_2 = 2000$ and $n_3 = 5000$ observations, respectively. In each sample, we use 500 replicates. We investigate the estimation procedure based on two different driving Lévy processes. Since the Brownian motion is the most common Lévy process, we examine Whittle‘s estimation based on a Brownian motion. As a second case, we analyze the performance based on a bivariate normal-inverse Gaussian (NIG) Lévy process, which is often used in modeling stochastic volatility or stock returns, see Barndorff-Nielsen (1997). The resulting increments of this process are characterized by the density

$$f(x, \mu, \alpha, \beta, \delta_{\mathrm{NIG}}, \Delta_{\mathrm{NIG}}) = \frac{\delta_{\mathrm{NIG}}}{2\pi} \frac{(1 + \alpha g(x))}{g(x)^3} \exp(\delta_{\mathrm{NIG}}\kappa + \beta^\top x - \alpha g(x)), \quad x \in \mathbb{R}^2,$$

with

$$g(x) = \sqrt{\delta_{\mathrm{NIG}}^2 + \langle x - \mu, \Delta_{\mathrm{NIG}}(x - \mu)\rangle}, \ \kappa^2 = \alpha^2 - \langle \beta, \Delta_{\mathrm{NIG}}\beta\rangle > 0.$$

Thereby, $\beta \in \mathbb{R}^2$ is a symmetry parameter, $\delta_{\mathrm{NIG}} \geq 0$ is a scale parameter and the positive definite matrix Δ_{NIG} models the dependency between the two components of the bivariate Lévy process $(L_t)_{t\in\mathbb{R}}$. We set $\mu = -(\delta_{\mathrm{NIG}}\Delta_{\mathrm{NIG}}\beta)/\kappa$ to guarantee that the resulting Lévy process is centered, see, e.g., Øigård et al. (2005) or Barndorff-Nielsen (1997) for more details. For better comparability of the Brownian motion case and the NIG Lévy process case, we choose the parameters of the NIG Lévy process in a way that the resulting covariance matrices of the Lévy processes are the same.

The performances of the Whittle and the adjusted Whittle estimator are compared with the well known quasi maximum likelihood estimator (QMLE) presented in Schlemm and Stelzer (2012a). The assumptions concerning the QMLE of Schlemm and Stelzer (2012a) are the same as ours. Therefore, the Echelon canonical form given in Schlemm and Stelzer (2012a), Section 4, is used as parametrization (cf. Guidorzi 1975) which is standard for state space and VARMA models (cf. Hannan and Deistler 1988). In particular, Assumptions (A1)–(A7) and (B1)–(B3) are satisfied.

In the multivariate setting, we consider bivariate MCARMA(2,1) processes of the form

$$\mathrm{d}X_t(\vartheta) = A(\vartheta)X_t(\vartheta)\mathrm{d}t + B(\vartheta)\mathrm{d}L_t(\vartheta) \quad \text{and} \quad Y_t(\vartheta) = C(\vartheta)X_t(\vartheta), \quad t \geq 0,$$

with

$$A(\vartheta) = \begin{pmatrix} \vartheta_1 & \vartheta_2 & 0 \\ 0 & 0 & 1 \\ \vartheta_3 & \vartheta_4 & \vartheta_5 \end{pmatrix}, \qquad B(\vartheta) = \begin{pmatrix} \vartheta_1 & \vartheta_2 \\ \vartheta_6 & \vartheta_7 \\ \vartheta_3 + \vartheta_5\vartheta_6 & \vartheta_6 + \vartheta_5\vartheta_7 \end{pmatrix},$$
$$C(\vartheta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \Sigma_L(\vartheta) = \begin{pmatrix} \vartheta_8 & \vartheta_9 \\ \vartheta_9 & \vartheta_{10} \end{pmatrix}.$$

This parametrization is given in Table 1 of Schlemm and Stelzer (2012a) and the representations of the corresponding AR polynomial P and MA polynomial Q are given in Table 2 of that paper. Furthermore, we get the order (2,1) of the MCARMA process from there as well. In our example, the true parameter value is

$$\vartheta_0^{(1)} = (-1, -2, 1, -2, -3, 1, 2, 0.4751, -0.1622, 0.3708).$$

To generate a NIG Lévy process with the same covariance matrix, we rely on the parameters

$$\delta_{\mathrm{NIG}}^{(1)} = 1, \quad \alpha^{(1)} = 3, \quad \beta^{(1)} = (1, 1)^{\top}, \quad \Delta_{\mathrm{NIG}}^{(1)} = \begin{pmatrix} 5/4 & -1/2 \\ -1/2 & 1 \end{pmatrix}.$$

The estimation results are summarized in Tables 1 and 2 for the Brownian motion driven model and the NIG driven model, respectively. The consistency can be observed in all simulations, namely the bias and the standard deviations are decreasing for increasing sample size for both the Whittle estimator and the quasi maximum likelihood estimator. The performance of the estimators is very similar.

Since we introduced an alternative estimator for the univariate setting, we perform an additional simulation study concerning one dimensional CARMA processes. In accordance with Assumption $\widetilde{A}$, the variance parameter σ_L^2 of the Lévy process is fixed in this study and has not to be estimated. We consider a CARMA(2, 1) model where

$$A(\vartheta) = \begin{pmatrix} 0 & 1 \\ \vartheta_1 & \vartheta_2 \end{pmatrix}, \qquad B(\vartheta) = \begin{pmatrix} \vartheta_3 \\ \vartheta_1 + \vartheta_2\vartheta_3 \end{pmatrix} \quad \text{and} \quad C(\vartheta) = (1\ 0).$$

Since the output process $Y(\vartheta)$ of this minimal state space model is of dimension one, the order of the AR polynomial p is equal to $N = 2$ and the order of the MA polynomial is $q = p - 1 = 1$. This means we have a CARMA(2,1) process. For more details on CARMA processes we refer to Brockwell and Lindner (2009) and Brockwell (2014). In our simulation study the true parameter is

$$\vartheta_0^{(2)} = (-2, -2, -1).$$

The simulation results for the Brownian motion driven and the NIG driven CARMA(2,1) process are given in Tables 3 and 4, respectively. For all sample sizes,

Table 1 Estimation results for a Brownian motion driven bivariate MCARMA(2,1) process with parameter $\vartheta_0^{(1)}$

	Whittle			QMLE		
ϑ_0	Mean	Bias	Std.	Mean	Bias	Std.
$n_1 = 500$						
−1	−0.9969	0.0031	0.0325	−1.0012	0.0012	0.0572
−2	−2.0218	0.0218	0.0582	−2.0128	0.0128	0.0689
1	0.9980	0.0020	0.0520	1.0075	0.0075	0.0722
−2	−2.0498	0.0498	0.1060	−1.9797	0.0203	0.0758
−3	−2.9840	0.0160	0.0498	−2.9913	0.0087	0.0907
1	1.0062	0.0062	0.1309	0.8034	0.1966	0.3896
2	1.9983	0.0017	0.0532	2.0036	0.0036	0.0768
0.4751	0.4746	0.0005	0.0407	0.4693	0.0048	0.0691
−0.1622	−0.1629	0.0007	0.0134	−0.1624	0.0002	0.0405
0.3708	0.3706	0.0002	0.0064	0.3712	0.0004	0.0328
$n_2 = 2000$						
−1	−0.9970	0.0030	0.0155	−0.9957	0.0043	0.0260
−2	−2.0062	0.0062	0.0252	−2.0047	0.0047	0.0350
1	0.9909	0.0091	0.0266	1.0038	0.0038	0.0399
−2	−2.0394	0.0394	0.0501	−2.0122	0.0122	0.0481
−3	−2.9857	0.0143	0.0371	−3.0350	0.0350	0.0583
1	1.0775	0.0775	0.1030	0.9572	0.0428	0.2583
2	2.0033	0.0033	0.0205	2.0452	0.0452	0.0463
0.4751	0.4731	0.0020	0.0092	0.4719	0.0032	0.0321
−0.1622	−0.1620	0.0002	0.0059	−0.1632	0.0010	0.0197
0.3708	0.3708	0	0.0037	0.3731	0.0023	0.0167
$n_3 = 5000$						
−1	−1.0028	0.0028	0.0172	−0.9960	0.0040	0.0174
−2	−1.9954	0.0146	0.0041	−2.0059	0.0059	0.0196
1	0.9972	0.0028	0.0133	1.0052	0.0052	0.0268
−2	−2.0202	0.0202	0.0210	−2.0043	0.0043	0.0284
−3	−3.0091	0.0091	0.0441	−3.0013	0.0013	0.0261
1	1.0585	0.0585	0.0409	1.0253	0.0253	0.1249
2	2.0109	0.0109	0.0318	2.0479	0.0479	0.0346
0.4751	0.4759	0.0008	0.0100	0.4735	0.0016	0.0200
−0.1622	−0.1652	0.0030	0.0088	−0.1634	0.0012	0.0135
0.3708	0.3904	0.0196	0.0079	0.3727	0.0019	0.0109

the Whittle estimator and the QMLE behave very similar and give excellent estimation results. Whereas for small sample sizes the adjusted Whittle estimator is remarkably worse, for increasing sample sizes it performs much better and seems to converge. Further simulations for a bivariate MCAR(1) process and an univariate CAR(3) process showing a similar pattern as the simulations of this section are presented in Section 9 in the Supplementary Material.

Table 2 Estimation results for a NIG driven bivariate MCARMA(2,1) process with parameter $\vartheta_0^{(1)}$

ϑ_0	Whittle			QMLE		
	Mean	Bias	Std.	Mean	Bias	Std.
$n_1 = 500$						
−1	−0.9555	0.0445	0.1559	−0.9651	0.0349	0.1854
−2	−1.8822	0.1178	0.2653	−1.6978	0.3022	0.3452
1	0.8746	0.1254	0.1888	1.1479	0.1479	0.2526
−2	−2.0981	0.0981	0.2273	−2.0066	0.0066	0.2962
−3	−3.1833	0.1833	0.2517	−3.0578	0.0578	0.4076
1	1.0533	0.0533	0.3614	1.0272	0.0272	1.2301
2	2.0461	0.0461	0.5710	2.0490	0.0490	1.6673
0.4751	0.4992	0.0241	0.1061	0.4645	0.0106	0.8220
−0.1622	−0.1520	0.0102	0.1130	−0.1669	0.0047	0.3317
0.3708	0.4100	0.0392	0.1081	0.3748	0.0040	0.6100
$n_2 = 2000$						
−1	−1.0351	0.0351	0.1224	−0.9673	0.0327	0.0243
−2	−1.8779	0.1221	0.1894	−1.0564	0.0426	0.0713
1	0.9457	0.0543	0.2620	1.1331	0.1331	0.1214
−2	−1.9586	0.0414	0.2573	−1.9494	0.0506	0.0827
−3	−3.1682	0.1682	0.2238	−3.1990	0.1990	0.4911
1	1.1234	0.1234	0.3120	1.1720	0.1720	0.5933
2	2.0842	0.0842	0.4842	2.0432	0.0432	0.1817
0.4751	0.5010	0.0259	0.1000	0.5237	0.0486	0.2726
−0.1622	−0.1740	0.0118	0.0992	−0.0856	0.0766	0.1413
0.3708	0.3908	0.0200	0.0758	0.3220	0.0488	0.0049
$n_3 = 5000$						
−1	−1.0238	0.0238	0.1182	−0.9844	0.0156	0.0194
−2	−1.9954	0.0046	0.2048	−2.0139	0.0139	0.0246
1	0.9942	0.0058	0.1517	1.0102	0.0102	0.0299
−2	−2.2202	0.2202	0.2210	−2.0043	0.0043	0.0284
−3	−3.0104	0.0104	0.2463	−3.0015	0.0015	0.2291
1	1.0585	0.0585	0.2409	1.0655	0.0655	0.1347
2	2.1169	0.1169	0.0866	2.0400	0.0400	0.0355
0.4751	0.4855	0.0104	0.1180	0.4737	0.0018	0.0206
−0.1622	−0.1682	0.0060	0.0408	−0.1634	0.0012	0.0145
0.3708	0.3908	0.0200	0.0842	0.3730	0.0022	0.0139

6 Proofs for the Whittle estimator in Sect. 3

6.1 Proofs of Sect. 3.1

Proof of Proposition 2 We divide W_n in two parts and investigate them separately. Therefore, define

Table 3 Estimation results for a Brownian motion driven CARMA(2,1) process with parameter $\vartheta_0^{(2)}$

	Whittle			Adjusted Whittle			QMLE		
ϑ_0	Mean	Bias	Std.	Mean	Bias	Std.	Mean	Bias	Std.
$n_1 = 500$									
−2	−2.0951	0.0951	0.7766	−3.1063	1.1063	3.4195	−2.0880	0.0880	0.7628
−2	−2.0482	0.0482	0.6500	−2.9233	0.9233	2.9957	−2.0449	0.0449	0.5889
−1	−0.9731	0.0269	0.1186	−0.9028	0.0972	0.3710	−0.9729	0.0271	0.1779
$n_2 = 2000$									
−2	−2.0204	0.0204	0.0755	−2.0816	0.0816	1.0399	−2.0015	0.0015	0.1926
−2	−1.9975	0.0025	0.0637	−2.0732	0.0732	0.9199	−1.9948	0.0052	0.1466
−1	−0.9933	0.0067	0.0547	−0.9965	0.0035	0.1267	−0.9993	0.0007	0.0674
$n_3 = 5000$									
−2	−2.0046	0.0046	0.0117	−1.9854	0.0146	0.0860	−2.0068	0.0068	0.0997
−2	−1.9914	0.0086	0.0149	−1.9840	0.0160	0..0821	−1.9942	0.0058	0.0772
−1	−1.0004	0.0004	0.0153	−1.0070	0.0070	0.0488	−1.0009	0.0009	0.0408

Table 4 Estimation results for a NIG driven CARMA(2,1) process with parameter $\vartheta_0^{(2)}$

	Whittle			adjusted Whittle			QMLE		
ϑ_0	Mean	Bias	Std.	Mean	Bias	Std.	Mean	Bias	Std.
$n_1 = 500$									
−2	−2.3278	0.3278	1.7598	−3.0174	1.0174	3.2090	−2.3175	0.3175	1.0862
−2	−2.2612	0.2612	1.4892	−2.8550	0.8550	2.8684	−2.2047	0.2047	0.8023
−1	−0.9855	0.0145	0.1652	−0.9445	0.0555	0.3376	−0.9243	0.0757	0.2938
$n_2 = 2000$									
−2	−2.0261	0.0261	0.1038	−1.9996	0.0004	0.5351	−2.0122	0.0122	0.2526
−2	−1.9977	0.0023	0.0784	−1.9988	0.0012	0.4552	−2.0034	0.0034	0.1845
−1	−0.9968	0.0032	0.0607	−1.0153	0.0153	0.0961	−1.0037	0.0037	0.0848
$n_3 = 5000$									
−2	−2.0138	0.0138	0.0575	−1.9842	0.0158	0.0902	−1.9938	0.0062	0.1093
−2	−1.9948	0.0052	0.0466	−1.9866	0.0134	0.0825	−1.9917	0.0083	0.0906
−1	−0.9991	0.0009	0.0339	−1.0097	0.0097	0.0508	−1.0059	0.0059	0.0415

$$W_n^{(1)}(\vartheta) := \frac{1}{2n} \sum_{j=-n+1}^{n} \operatorname{tr}\left(f_{Y^{(\Delta)}}(\omega_j, \vartheta)^{-1} I_n(\omega_j)\right)$$

and

$$W_n^{(2)}(\vartheta) = \frac{1}{2n} \sum_{j=-n+1}^{n} \log\left(\det\left(f_{Y^{(\Delta)}}(\omega_j, \vartheta)\right)\right),$$

such that $W_n(\vartheta) = W_n^{(1)}(\vartheta) + W_n^{(2)}(\vartheta)$. By Assumption (A1) and (A4) and the representation (8), we can apply Lemma 7 of the Supplementary Material, which gives the uniform convergence

$$\sup_{\vartheta\in\Theta}\left|W_n^{(2)}(\vartheta) - \frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(\det\left(f_{Y^{(\Delta)}}(\omega,\vartheta)\right)\right)\mathrm{d}\omega\right| \xrightarrow{n\to\infty} 0. \tag{16}$$

It remains to prove the appropriate convergence of $W_n^{(1)}$. Therefore, it is sufficient to show that

$$\sup_{\vartheta\in\Theta}\left\|\frac{1}{2n}\sum_{j=-n+1}^{n} f_{Y^{(\Delta)}}(\omega_j,\vartheta)^{-1}I_n(\omega_j) - \frac{1}{2\pi}\int_{-\pi}^{\pi} f_{Y^{(\Delta)}}(\omega,\vartheta)^{-1}f_{Y^{(\Delta)}}(\omega)\mathrm{d}\omega\right\| \xrightarrow{\text{a.s.}} 0 \tag{17}$$

holds. We approximate $f_{Y^{(\Delta)}}(\omega_j,\vartheta)^{-1}$ by the Cesáro sum of its Fourier series of size M for M sufficiently large. Define

$$q_M(\omega,\vartheta) := \frac{1}{M}\sum_{j=0}^{M-1}\left(\sum_{|k|\le j} b_k(\vartheta)e^{-ik\omega}\right) = \sum_{|k|<M}\left(1-\frac{|k|}{M}\right)b_k(\vartheta)e^{-ik\omega} \quad \text{with}$$
$$b_k(\vartheta) := \frac{1}{2\pi}\int_{-\pi}^{\pi} f_{Y^{(\Delta)}}(\omega,\vartheta)^{-1}e^{ik\omega}\mathrm{d}\omega.$$

The inverse $f_{Y^{(\Delta)}}(\omega,\vartheta)^{-1}$ exists, is continuous and 2π-periodic in the first component. Thus, it follows from Féjer's Theorem (see, e.g., Theorem 2.11.1 of Brockwell and Davis 1991) that for any $\epsilon > 0$ there exists an $M_0(\epsilon)\in\mathbb{N}$ such that for $M \ge M_0(\epsilon)$

$$\sup_{\omega\in[-\pi,\pi]}\sup_{\vartheta\in\Theta}\left\|f_{Y^{(\Delta)}}(\omega,\vartheta)^{-1} - q_M(\omega,\vartheta)\right\| < \epsilon. \tag{18}$$

Let $\epsilon > 0$. In view of (18), we get

$$\begin{aligned}&\left\|\frac{1}{2n}\sum_{j=-n+1}^{n} f_{Y^{(\Delta)}}(\omega_j,\vartheta)^{-1}I_n(\omega_j) - \frac{1}{2n}\sum_{j=-n+1}^{n} q_M(\omega_j,\vartheta)I_n(\omega_j)\right\|\\ &\quad\le \frac{\epsilon}{2n}\sum_{j=-n+1}^{n}\left\|I_n(\omega_j)\right\|.\end{aligned} \tag{19}$$

Since all matrix norms are equivalent, using the 1-norm yields

$$\frac{\epsilon}{2n}\sum_{j=-n+1}^{n}\left\|I_n(\omega_j)\right\| \le \frac{\epsilon\mathfrak{C}}{2n}\sum_{j=-n+1}^{n}\sum_{k=1}^{m}\sum_{\ell=1}^{m}|I_n(\omega_j)[k,\ell]|. \tag{20}$$

Trivially, $I_n(\omega_j)$ is positive semidefinite and Hermitian. Therefore, for $k,\ell\in\{1,\dots,m\}$, $j\in\{-n+1,\dots,n\}$,

$$\det\begin{pmatrix} I_n(\omega_j)[k,k] & I_n(\omega_j)[k,\ell]\\ I_n(\omega_j)[\ell,k] & I_n(\omega_j)[\ell,\ell]\end{pmatrix} \ge 0,$$

which implies

$$\left|I_n(\omega_j)[k,\ell]\right| \leq \sqrt{I_n(\omega_j)[k,k]I_n(\omega_j)[\ell,\ell]} \leq I_n(\omega_j)[k,k] + I_n(\omega_j)[\ell,\ell]. \tag{21}$$

Combining (19), (20), (21) and Lemma 4 of the Supplementary Material gives for $M \geq M_0(\epsilon)$

$$\begin{aligned}
&\left\| \frac{1}{2n} \sum_{j=-n+1}^{n} f_{Y^{(\Delta)}}(\omega_j,\vartheta)^{-1} I_n(\omega_j) - \frac{1}{2n} \sum_{j=-n+1}^{n} q_M(\omega_j,\vartheta) I_n(\omega_j) \right\| \\
&\quad \leq \frac{\epsilon \mathfrak{C}}{2n} \sum_{j=-n+1}^{n} \sum_{k=1}^{m} \sum_{\ell=1}^{m} \left[I_n(\omega_j)[k,k] + I_n(\omega_j)[\ell,\ell] \right] \\
&\quad \leq \frac{\epsilon \mathfrak{C} m}{n} \sum_{j=-n+1}^{n} \sum_{k=1}^{m} I_n(\omega_j)[k,k] \\
&\quad \leq 2\epsilon \mathfrak{C} m \sum_{k=1}^{m} \overline{\Gamma}_n^{(\Delta)}(0)[k,k].
\end{aligned}$$

Since $\sum_{k=1}^{m} \overline{\Gamma}_n^{(\Delta)}(0)[k,k] \xrightarrow{\text{a.s.}} \sum_{k=1}^{m} \Gamma^{(\Delta)}(0)[k,k] < \infty$ due to Lemma 5 in the Supplementary Material, we obtain for $M \geq M_0(\epsilon)$ and n large

$$\sup_{\vartheta \in \Theta} \left\| \frac{1}{2n} \sum_{j=-n+1}^{n} \left(f_{Y^{(\Delta)}}(\omega_j,\vartheta)^{-1} I_n(\omega_j) \right) - \frac{1}{2n} \sum_{j=-n+1}^{n} q_M(\omega_j,\vartheta) I_n(\omega_j) \right\| \leq \varepsilon \mathfrak{C}$$

almost surely. Consequently, for the proof of (17) it is sufficient to show that

$$\sup_{\vartheta \in \Theta} \left\| \frac{1}{2n} \sum_{j=-n+1}^{n} q_M(\omega_j,\vartheta) I_n(\omega_j) - \frac{1}{2\pi} \int_{-\pi}^{\pi} f_{Y^{(\Delta)}}(\omega,\vartheta)^{-1} f_{Y^{(\Delta)}}(\omega) \mathrm{d}\omega \right\| \xrightarrow{\text{a.s.}} 0. \tag{22}$$

On the one hand, Lemma 4 of the Supplementary Material yields

$$\begin{aligned}
&\frac{1}{2n} \sum_{j=-n+1}^{n} q_M(\omega_j,\vartheta) I_n(\omega_j) \\
&= \frac{1}{2\pi} \sum_{|k|<M} \sum_{|h|<n} \left(\left(1 - \frac{|k|}{M}\right) b_k(\vartheta) \overline{\Gamma}_n^{(\Delta)}(h) \left(\frac{1}{2n} \sum_{j=-n+1}^{n} e^{-i(k+h)\omega_j} \right) \right) \\
&= \frac{1}{2\pi} \sum_{|k|<M} \left(1 - \frac{|k|}{M}\right) b_k(\vartheta) \overline{\Gamma}_n^{(\Delta)}(-k) \\
&\xrightarrow{\text{a.s.}} \frac{1}{2\pi} \sum_{|k|<M} \left(1 - \frac{|k|}{M}\right) b_k(\vartheta) \Gamma^{(\Delta)}(-k)
\end{aligned} \tag{23}$$

uniformly in ϑ, since $b_k(\vartheta)$ is uniformly bounded in ϑ for all k. The reason is that $f_{Y^{(\Delta)}}(\omega,\vartheta)^{-1}$ is continuous on the compact set $[-\pi,\pi]\times\Theta$ and

$$\sup_{\substack{\vartheta\in\Theta\\ k\in\mathbb{Z}}}\|b_k(\vartheta)\|=\sup_{\substack{\vartheta\in\Theta\\ k\in\mathbb{Z}}}\left\|\frac{1}{2\pi}\int_{-\pi}^{\pi}f_{Y^{(\Delta)}}(\omega,\vartheta)^{-1}e^{ik\omega}\mathrm{d}\omega\right\|\le\max_{\vartheta\in\Theta}\max_{\omega\in[-\pi,\pi]}\|f_{Y^{(\Delta)}}(\omega,\vartheta)^{-1}\|.$$

On the other hand, due to (5), we get

$$\begin{aligned}
&\left\|\frac{1}{2\pi}\sum_{|h|<M}\left(1-\frac{|h|}{M}\right)b_{-h}(\vartheta)\Gamma^{(\Delta)}(h)-\frac{1}{2\pi}\int_{-\pi}^{\pi}f_{Y^{(\Delta)}}(\omega,\vartheta)^{-1}f_{Y^{(\Delta)}}(\omega)\mathrm{d}\omega\right\|\\
&\quad=\left\|\frac{1}{2\pi}\sum_{|h|<M}\left(1-\frac{|h|}{M}\right)b_{-h}(\vartheta)\int_{-\pi}^{\pi}f_{Y^{(\Delta)}}(\omega)e^{ih\omega}\mathrm{d}\omega-\frac{1}{2\pi}\int_{-\pi}^{\pi}f_{Y^{(\Delta)}}(\omega,\vartheta)^{-1}f_{Y^{(\Delta)}}(\omega)\mathrm{d}\omega\right\|\\
&\quad=\left\|\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(q_M(\omega,\vartheta)-f_{Y^{(\Delta)}}(\omega,\vartheta)^{-1}\right)f_{Y^{(\Delta)}}(\omega)\mathrm{d}\omega\right\|\\
&\quad\le\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\|q_M(\omega,\vartheta)-f_{Y^{(\Delta)}}(\omega,\vartheta)^{-1}\right\|\left\|f_{Y^{(\Delta)}}(\omega)\right\|\mathrm{d}\omega\le\epsilon\mathfrak{C},
\end{aligned}\tag{24}$$

where we used (18) and the continuity of $f_{Y^{(\Delta)}}(\omega)$ for the last inequality. Combining (23) and (24) gives (22). □

Proof of Lemma 1 In view of Proposition 1, we express the linear innovations as

$$\varepsilon_k^{(\Delta)}(\vartheta)=\Pi(\mathsf{B},\vartheta)Y_k^{(\Delta)}(\vartheta),\quad k\in\mathbb{N},$$

and define the pseudo-innovations as

$$\xi_k^{(\Delta)}(\vartheta):=\Pi(\mathsf{B},\vartheta)Y_k^{(\Delta)}(\vartheta_0),\quad k\in\mathbb{N}.$$

An application of Theorem 11.8.3 of Brockwell and Davis (1991) leads to the spectral densities of $(\varepsilon_k^{(\Delta)}(\vartheta))_{k\in\mathbb{N}}$ and $(\xi_k^{(\Delta)}(\vartheta))_{k\in\mathbb{N}}$ as

$$\begin{aligned}
f_{\varepsilon^{(\Delta)}}(\omega,\vartheta)&=\Pi(e^{-i\omega},\vartheta)f_{Y^{(\Delta)}}(\omega,\vartheta)\Pi(e^{i\omega},\vartheta)^\top,\quad\omega\in[-\pi,\pi],\\
f_{\xi^{(\Delta)}}(\omega,\vartheta)&=\Pi(e^{-i\omega},\vartheta)f_{Y^{(\Delta)}}(\omega)\Pi(e^{i\omega},\vartheta)^\top,\quad\omega\in[-\pi,\pi],
\end{aligned}$$

respectively. Consequently,

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left(f_{Y^{(\Delta)}}(\omega,\vartheta)^{-1}f_{Y^{(\Delta)}}(\omega)\right)\mathrm{d}\omega\\
&\quad=\frac{1}{2\pi}\mathrm{tr}\left(\int_{-\pi}^{\pi}2\pi\Pi(e^{i\omega},\vartheta)^\top V^{(\Delta)}(\vartheta)^{-1}\Pi(e^{-i\omega},\vartheta)f_{Y^{(\Delta)}}(\omega)\mathrm{d}\omega\right)\\
&\quad=\mathrm{tr}\left(V^{(\Delta)}(\vartheta)^{-1}\int_{-\pi}^{\pi}f_{\xi^{(\Delta)}}(\omega,\vartheta)\mathrm{d}\omega\right)\\
&\quad=\mathbb{E}\left[\mathrm{tr}\left(\xi_1^{(\Delta)}(\vartheta)^\top V^{(\Delta)}(\vartheta)^{-1}\xi_1^{(\Delta)}(\vartheta)\right)\right]
\end{aligned}$$

holds. Finally,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(\det\left(f_{Y^{(\Delta)}}(\omega,\vartheta)\right)\right)\mathrm{d}\omega=\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(\det\left(2\pi f_{Y^{(\Delta)}}(\omega,\vartheta)\right)\right)\mathrm{d}\omega-m\log(2\pi),$$

and an application of Theorem 3''' of Chapter 3 of Hannan (2009) results in

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(\det\left(2\pi f_{Y^{(\Delta)}}(\omega,\vartheta)\right)\right)\mathrm{d}\omega-m\log(2\pi)=\log(\det V^{(\Delta)}(\vartheta))-m\log(2\pi), \tag{25}$$

which completes the proof. □

Proof of Proposition 3 Considering Lemma 1 we get $W(\vartheta)=\mathcal{L}(\vartheta)$. Schlemm and Stelzer (2012a), Lemma 2.10, proved that $\mathcal{L}$ has a unique global minimum in ϑ_0 under conditions which are fulfilled in our setting (see Lemma 2.3 and Lemma 3.14 of Schlemm and Stelzer 2012a). □

Proof of Theorem 1 Note that by Proposition 3 W has a unique global minimum in ϑ_0. By Proposition 2, W_n converges almost surely to W. Since $f_{Y^{(\Delta)}}$ is continuous under Assumption A and therefore W as well, the assertion follows directly from Theorem 2.1 of Newey and McFadden (1994) and the discussion below. □

6.2 Proofs of Sect. 3.2

Proof of Proposition 4 Under the Assumptions (A1)–(A4) and (B3) the spectral density $f_{Y^{(\Delta)}}(\omega,\vartheta)$ and its inverse $f_{Y^{(\Delta)}}(\omega,\vartheta)^{-1}$ are three times continuously differentiable in ϑ (see Remarks 1 and 2). Furthermore,

$$\frac{\partial^2}{\partial\vartheta_k\partial\vartheta_\ell}\mathrm{tr}\left(f_{Y^{(\Delta)}}(\omega,\vartheta)^{-1}I_n(\omega)\right)=\mathrm{tr}\left(\frac{\partial^2}{\partial\vartheta_k\partial\vartheta_\ell}\left(f_{Y^{(\Delta)}}(\omega,\vartheta)^{-1}\right)I_n(\omega)\right),$$

$k,\ell\in\{1,\dots,r\}$. Therefore, the proof of

$$\sup_{\vartheta\in\Theta}\left\|\nabla^2_\vartheta W_n(\vartheta)-\nabla^2_\vartheta W(\vartheta)\right\|\xrightarrow{\text{a.s.}}0$$

goes in the same way as the proof of Proposition 2. It remains to show that $\nabla^2_\vartheta W(\vartheta_0)=\Sigma_{\nabla^2 W}$.

First, note that

$$\nabla^2_\vartheta W(\vartheta_0)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\nabla^2_\vartheta\mathrm{tr}(f_{Y^{(\Delta)}}(\omega,\vartheta_0)^{-1}f_{Y^{(\Delta)}}(\omega))+\nabla^2_\vartheta\log\left(\det\left(f_{Y^{(\Delta)}}(\omega,\vartheta_0)\right)\right)\mathrm{d}\omega. \tag{26}$$

On the one hand,

$$
\begin{aligned}
&\frac{1}{2\pi}\int_{-\pi}^{\pi}\operatorname{tr}\left(\frac{\partial^2}{\partial\vartheta_k\partial\vartheta_l}\big(f_{Y^{(\Delta)}}(\omega,\vartheta_0)^{-1}\big)f_{Y^{(\Delta)}}(\omega)\right)\mathrm{d}\omega\\
&\quad=\frac{1}{2\pi}\int_{-\pi}^{\pi}\operatorname{tr}\left(2f_{Y^{(\Delta)}}(\omega)^{-1}\left(\frac{\partial}{\partial\vartheta_k}f_{Y^{(\Delta)}}(\omega,\vartheta_0)\right)f_{Y^{(\Delta)}}(\omega)^{-1}\left(\frac{\partial}{\partial\vartheta_\ell}f_{Y^{(\Delta)}}(\omega,\vartheta_0)\right)\mathrm{d}\omega\right.\\
&\qquad\left.-\int_{-\pi}^{\pi}f_{Y^{(\Delta)}}(\omega)^{-1}\left(\frac{\partial^2}{\partial\vartheta_k\partial\vartheta_\ell}f_{Y^{(\Delta)}}(\omega,\vartheta_0)\right)\right)\mathrm{d}\omega
\end{aligned}
\tag{27}
$$

holds. On the other hand, Jacobi's formula leads to

$$
\begin{aligned}
&\frac{\partial^2}{\partial\vartheta_k\partial\vartheta_\ell}\log(\det(f_{Y^{(\Delta)}}(\omega,\vartheta_0)))\\
&\quad=\operatorname{tr}\left(-f_{Y^{(\Delta)}}(\omega)^{-1}\left(\frac{\partial}{\partial\vartheta_k}f_{Y^{(\Delta)}}(\omega,\vartheta_0)\right)f_{Y^{(\Delta)}}(\omega)^{-1}\left(\frac{\partial}{\partial\vartheta_\ell}f_{Y^{(\Delta)}}(\omega,\vartheta_0)\right)\right)\\
&\qquad+\operatorname{tr}\left(f_{Y^{(\Delta)}}(\omega)^{-1}\left(\frac{\partial^2}{\partial\vartheta_k\partial\vartheta_\ell}f_{Y^{(\Delta)}}(\omega,\vartheta_0)\right)\right).
\end{aligned}
\tag{28}
$$

Combining (26), (27), (28) and the property

$$
\operatorname{vec}\left(A^\top\right)^\top\left(B^\top\otimes C\right)\operatorname{vec}(D)=\operatorname{tr}(BACD)
\tag{29}
$$

for appropriate matrices A, B, C, D (see Brewer 1978, properties T2.4, T3.4 and T3.8) give

$$
\begin{aligned}
\nabla^2_\vartheta W(\vartheta_0)&=\frac{1}{2\pi}\int_{-\pi}^{\pi}\nabla_\vartheta f_{Y^{(\Delta)}}(-\omega,\vartheta_0)^\top\left[f_{Y^{(\Delta)}}(-\omega)^{-1}\otimes f_{Y^{(\Delta)}}(\omega)^{-1}\right]\nabla_\vartheta f_{Y^{(\Delta)}}(\omega,\vartheta_0)\mathrm{d}\omega\\
&=\Sigma_{\nabla^2W}.
\end{aligned}
$$

□

Proof of Lemma 2 Let $c\in\mathbb{C}^r$ be fixed and ω^* as in (B4). The continuity of $f_{Y^{(\Delta)}}(\omega)$ and its regularity imply for any ω in a neighborhood of ω^* that

$$
\left\|\left(f_{Y^{(\Delta)}}(-\omega)^{-1/2}\otimes f_{Y^{(\Delta)}}(\omega)^{-1/2}\right)\nabla_\vartheta f_{Y^{(\Delta)}}(\omega,\vartheta_0)c\right\|_2>0
$$

where $\|\cdot\|_2$ is the Euclidean norm. Consequently,

$$
\begin{aligned}
c^\top\Sigma_{\nabla^2W}c&=\frac{1}{2\pi}\int_{-\pi}^{\pi}c^\top\nabla_\vartheta f_{Y^{(\Delta)}}(\omega,\vartheta_0)^H\left[f_{Y^{(\Delta)}}(-\omega)^{-1}\otimes f_{Y^{(\Delta)}}(\omega)^{-1}\right]\nabla_\vartheta f_{Y^{(\Delta)}}(\omega,\vartheta_0)c\mathrm{d}\omega\\
&=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\|\left(f_{Y^{(\Delta)}}(-\omega)^{-1/2}\otimes f_{Y^{(\Delta)}}(\omega)^{-1/2}\right)\nabla_\vartheta f_{Y^{(\Delta)}}(\omega,\vartheta_0)c\right\|_2^2\mathrm{d}\omega>0.
\end{aligned}
$$

Therefore, Σ_{∇^2W} is positive definite. □

For the proof of Proposition 5 we require some auxiliary result. Therefore, we denote the periodogram and the sample covariance corresponding to $N_1^{(\Delta)}, \dots, N_n^{(\Delta)}$ as defined in (7) as $I_{n,N}$ and $\overline{\Gamma}_{n,N}$, respectively.

Lemma 3 *Let Assumptions (A2)–(A4) hold and $\eta : [-\pi, \pi] \to \mathbb{C}^{m\times m}$ be a matrix-valued continuous function with $\eta(\omega) = \eta(\omega)^H, \omega \in [-\pi, \pi]$, and Fourier coefficients $(\mathfrak{f}_u)_{u\in\mathbb{Z}}$ satisfying $\sum_{u=-\infty}^{\infty} \|\mathfrak{f}_u\| < \infty$. Then,*

$$\lim_{n\to\infty} \mathbb{E}\left| \frac{1}{2\sqrt{n}} \sum_{j=-n+1}^{n} \operatorname{tr}\Big(\eta(\omega_j)I_n(\omega_j) - \eta(\omega_j)\Phi(e^{-i\omega_j})I_{n,N}(\omega_j)\Phi(e^{i\omega_j})^\top\Big) \right| = 0.$$

Proof Define $R_n(\omega) = I_n(\omega) - \Phi(e^{-i\omega})I_{n,N}(\omega)\Phi(e^{i\omega})^\top$ for $\omega \in [-\pi, \pi]$. We get

$$\begin{aligned}
R_n(\omega_j) &= \frac{1}{2\pi n}\left(\sum_{k=1}^{n}\sum_{s=0}^{\infty}\Phi_s N_{k-s}{}^{(\Delta)}\right)\left(\sum_{\ell=1}^{n}\sum_{t=0}^{\infty}\Phi_t N_{\ell-t}{}^{(\Delta)}\right)^\top e^{-i(k-\ell)\omega_j} \\
&\quad - \frac{1}{2\pi n}\left(\sum_{k=1}^{n}\sum_{s=0}^{\infty}\Phi_s N_k{}^{(\Delta)}\right)\left(\sum_{\ell=1}^{n}\sum_{t=0}^{\infty}\Phi_t N_\ell{}^{(\Delta)}\right)^\top e^{-i(k+s-\ell-t)\omega_j} \\
&= \frac{1}{2\pi n}\left(\sum_{s=0}^{\infty}\sum_{t=0}^{\infty}\Phi_s\left(\left(\sum_{k=1}^{n}\sum_{\ell=1-t}^{0} - \sum_{k=1}^{n}\sum_{\ell=n-t+1}^{n} + \sum_{k=1-s}^{0}\sum_{\ell=1}^{n} + \sum_{k=1-s}^{0}\sum_{\ell=1-t}^{0}\right.\right.\right. \\
&\quad \left. - \sum_{k=1-s}^{0}\sum_{\ell=n-t+1}^{n} - \sum_{k=n-s+1}^{n}\sum_{\ell=1}^{n} - \sum_{k=n-s+1}^{n}\sum_{\ell=1-t}^{0} + \sum_{k=n-s+1}^{n}\sum_{\ell=n-t+1}^{n}\right) \\
&\quad \left.\left. N_k{}^{(\Delta)}N_\ell^{(\Delta)\top} e^{-i(k+s-\ell-t)\omega_j}\right)\Phi_t^\top\right) \\
&=: \sum_{i=1}^{8} R_n^{(i)}(\omega_j).
\end{aligned}$$

Thus,

$$\mathbb{E}\left|\frac{1}{2\sqrt{n}}\sum_{j=-n+1}^{n}\operatorname{tr}\Big(\eta(\omega_j)R_n(\omega_j)\Big)\right| \le \sum_{i=1}^{8}\mathbb{E}\left|\frac{1}{2\sqrt{n}}\sum_{j=-n+1}^{n}\operatorname{tr}\Big(\eta(\omega_j)R_n^{(i)}(\omega_j)\Big)\right|.$$

We have to show that these 8 components converge to zero. Since we can treat each component similarly, we only give the detailed proof for the convergence of the first term.

Due to $\operatorname{tr}(A) \le \|A\|_1$ for all quadratic matrices A, we get an upper bound for the trace of any quadratic matrix. Once again, the equivalence of all matrix norms and $\eta(\omega_j) = \sum_{u=-\infty}^{\infty}\mathfrak{f}_u e^{-i\omega_j u}$ yield

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{2\sqrt{n}}\sum_{j=-n+1}^{n}\operatorname{tr}\big(R_n^{(1)}(\omega_j)\eta(\omega_j)\big)\right| \\
&\quad\leq \mathfrak{C}\mathbb{E}\left\|\sum_{j=-n+1}^{n}\frac{1}{\sqrt{n}}\frac{1}{n}\sum_{s=0}^{\infty}\sum_{t=0}^{\infty}\boldsymbol{\Phi}_s\sum_{k=1}^{n}\sum_{\ell=1-t}^{0}N_k^{(\Delta)}N_\ell^{(\Delta)\top}\boldsymbol{\Phi}_t^\top\sum_{u=-\infty}^{\infty}\mathfrak{f}_u e^{-i(k+s-\ell-t+u)\omega_j}\right\| \\
&\quad\leq \mathfrak{C}\frac{1}{\sqrt{n}}\sum_{s=0}^{\infty}\sum_{t=0}^{\infty}\|\boldsymbol{\Phi}_s\|\sum_{k=1}^{n}\sum_{\ell=1-t}^{0}\mathbb{E}\|N_1^{(\Delta)}\|^2\|\boldsymbol{\Phi}_t\| \\
&\qquad\cdot\sum_{u=-\infty}^{\infty}\|\mathfrak{f}_u\|\frac{1}{n}\left\|\sum_{j=-n+1}^{n}e^{-i(k+s-\ell-t+u)\omega_j}\right\|.
\end{aligned}$$

Due to (A2), $\mathbb{E}\|N_1^{(\Delta)}\|^2 < \infty$. Furthermore, by Assumption (A3) the coefficients of $\boldsymbol{\Phi}$ are exponentially decreasing which implies $\sum_{t=0}^{\infty} t\|\boldsymbol{\Phi}_t\| < \infty$. Along with an application of Lemma 4 of the Supplementary Material, it follows

$$\mathbb{E}\left|\frac{1}{2\sqrt{n}}\sum_{j=-n+1}^{n}\operatorname{tr}\big(R_n^{(1)}(\omega_j)\eta(\omega_j)\big)\right| \leq \mathfrak{C}\frac{1}{\sqrt{n}}\sum_{s=0}^{\infty}\|\boldsymbol{\Phi}_s\|\sum_{t=0}^{\infty}t\|\boldsymbol{\Phi}_t\|\sum_{u=-\infty}^{\infty}\|\mathfrak{f}_u\| \xrightarrow{n\to\infty} 0.$$

□

This lemma helps to deduce Proposition 5, which can be seen as the main part of the proof of the asymptotic normality of the Whittle estimator.

Proof of Proposition 5 Due to Lemma 3, we get

$$\begin{aligned}
&\frac{1}{2\sqrt{n}}\sum_{j=-n+1}^{n}\operatorname{tr}\big(\eta(\omega_j)I_n(\omega_j)-\eta(\omega_j)f_{Y^{(\Delta)}}(\omega_j)\big) \\
&\quad=\frac{1}{2\sqrt{n}}\sum_{j=-n+1}^{n}\left\{\operatorname{tr}\big(I_{n,N}(\omega_j)\boldsymbol{\Phi}(e^{i\omega_j})^\top\eta(\omega_j)\boldsymbol{\Phi}(e^{-i\omega_j})\big)-\operatorname{tr}\big(\eta(\omega_j)f_{Y^{(\Delta)}}(\omega_j)\big)\right\}+o_{\mathbb{P}}(1).
\end{aligned}$$

We define

$$q(\omega) := \boldsymbol{\Phi}(e^{i\omega})^\top\eta(\omega)\boldsymbol{\Phi}(e^{-i\omega}), \quad \omega\in[-\pi,\pi],$$

and approximate q by its Fourier series of degree M, namely,

$$q_M(\omega) = \sum_{|k|\leq M} b_k e^{ik\omega} \quad \text{where} \quad b_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ik\omega}q(\omega)\mathrm{d}\omega, \quad k\in\mathbb{Z}. \tag{30}$$

The coefficients b_k satisfy

$$\begin{aligned}
\sum_{k=-\infty}^{\infty} \|b_k\| |k|^{1/2} &= \sum_{k=-\infty}^{\infty} \left\| \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ik\omega} \boldsymbol{\Phi}(e^{i\omega})^\top \eta(\omega) \boldsymbol{\Phi}(e^{-i\omega}) \mathrm{d}\omega \right\| |k|^{1/2} \\
&= \sum_{k=-\infty}^{\infty} \left\| \frac{1}{2\pi} \sum_{j=0}^{\infty} \sum_{\ell=0}^{\infty} \sum_{u=-\infty}^{\infty} \boldsymbol{\Phi}_j^\top \mathfrak{f}_u \boldsymbol{\Phi}_\ell \int_{-\pi}^{\pi} e^{-i(k-j+u+\ell)\omega} \mathrm{d}\omega \right\| |k|^{1/2} \\
&\leq \sum_{j=0}^{\infty} \sum_{\ell=0}^{\infty} \sum_{u=-\infty}^{\infty} \|\boldsymbol{\Phi}_j\| \|\mathfrak{f}_u\| \|\boldsymbol{\Phi}_\ell\| |j-u-\ell|^{1/2} \\
&\leq \mathfrak{C} \sum_{j=0}^{\infty} \|\boldsymbol{\Phi}_j\| (\max\{1,|j|\})^{1/2} \sum_{u=-\infty}^{\infty} \|\mathfrak{f}_u\| (\max\{1,|u|\})^{1/2} \sum_{\ell=0}^{\infty} \|\boldsymbol{\Phi}_\ell\| (\max\{1,|\ell|\})^{1/2} \\
&< \infty,
\end{aligned} \tag{31}$$

and therefore $\sum_{k=-\infty}^{\infty} \|b_k\| < \infty$ as well. It follows from Körner (1989), Theorem 3.1, that

$$q_M(\omega) \overset{M\to\infty}{\longrightarrow} q(\omega) \quad \text{uniformly in } \omega \in [-\pi, \pi].$$

Step 1: We show

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\frac{1}{\sqrt{n}} \left| \sum_{j=-n+1}^{n} \operatorname{tr}(I_{n,N}(\omega_j)(q(\omega_j) - q_M(\omega_j)) \right| > \epsilon \right) = 0 \quad \forall\, \epsilon > 0. \tag{32}$$

Consider

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{j=-n+1}^{n} \operatorname{tr}\big(I_{n,N}(\omega_j)(q(\omega_j) - q_M(\omega_j))\big) \\
&\quad = \frac{\sqrt{n}}{\pi} \sum_{|k|>M} \operatorname{tr}\left(\sum_{h=-n+1}^{n-1} \overline{\Gamma}_{n,N}(h) b_k \left(\frac{1}{2n} \sum_{j=-n+1}^{n} e^{-i(h-k)\omega_j} \right) \right).
\end{aligned} \tag{33}$$

We investigate the terms with $h = 0$ and $h \neq 0$ separately. For $h = 0$ and $n > M$ we get

$$\begin{aligned}
&\left| \frac{\sqrt{n}}{\pi} \sum_{|k|>M} \operatorname{tr}\left(\overline{\Gamma}_{n,N}(0) b_k \mathbf{1}_{\{\exists z \in \mathbb{Z}\setminus\{0\} : k = 2nz\}} \right) \right| \\
&\qquad \leq \mathfrak{C} \sqrt{n} \|\overline{\Gamma}_{n,N}(0)\| \sum_{|k|\geq 2n} \|b_k\| \overset{n\to\infty}{\longrightarrow} 0 \quad \mathbb{P}\text{-a.s.},
\end{aligned} \tag{34}$$

since Remark 10 in the Supplementary Material and the continuous mapping theorem imply $\|\overline{\Gamma}_{n,N}(0)\| \overset{\text{a.s.}}{\to} \|\Sigma_N^{(\Delta)}\|$.

Now, we investigate the terms with $h \neq 0$. The independence of the sequence $(N_k^{(\Delta)})_{k\in\mathbb{N}_0}$ leads to

$$\mathbb{E}\left[\overline{\Gamma}_{n,N}(h)\right] = 0 \quad \text{for} \quad h \neq 0$$

and therefore,

$$\mathbb{E}\left[\sqrt{n}\sum_{|k|>M}\operatorname{tr}\left(\left(\sum_{h=1}^{n-1}\overline{\Gamma}_{n,N}(h)+\sum_{h=-n+1}^{-1}\overline{\Gamma}_{n,N}(h)\right)b_k\mathbf{1}_{\{\exists z\in\mathbb{Z}:h=k+2nz\}}\right)\right] = 0. \quad (35)$$

Due to (33)–(35) and the Tschebycheff inequality, for the proof of (32) it is sufficient to show that

$$\lim_{M\to\infty}\lim_{n\to\infty}\operatorname{Var}\left(\sqrt{n}\sum_{|k|>M}\operatorname{tr}\left(\left(\sum_{h=1}^{n-1}\overline{\Gamma}_{n,N}(h)+\sum_{h=-n+1}^{-1}\overline{\Gamma}_{n,N}(h)\right)b_k\mathbf{1}_{\{\exists z\in\mathbb{Z}:h=k+2nz\}}\right)\right) = 0. \quad (36)$$

First, property (29) and $\mathbb{E}\left\|\operatorname{vec}\left(\overline{\Gamma}_{n,N}(h)\right)\operatorname{vec}\left(\overline{\Gamma}_{n,N}(h)\right)^{\top}\right\| \leq \frac{\mathfrak{C}}{n}$ result in

$$\begin{aligned}
&\operatorname{Var}\left(\sqrt{n}\sum_{|k|>M}\operatorname{tr}\left(\left(\sum_{h=1}^{n-1}\overline{\Gamma}_{n,N}(h)+\sum_{h=-n+1}^{-1}\overline{\Gamma}_{n,N}(h)\right)b_k\mathbf{1}_{\{\exists z\in\mathbb{Z}:h=k+2nz\}}\right)\right)\\
&\quad= \operatorname{Var}\left(2\sqrt{n}\sum_{h=1}^{n-1}\operatorname{vec}\left(\sum_{|k|>M}b_k^{\top}\mathbf{1}_{\{\exists z\in\mathbb{Z}:h=k+2nz\}}\right)^{\top}\left(I_N\otimes I_N\right)\operatorname{vec}\left(\overline{\Gamma}_{n,N}(h)\right)\right)\\
&\quad\leq 4n\sum_{h=1}^{n-1}\left\|\operatorname{vec}\left(\sum_{|k|>M}b_k^{\top}\mathbf{1}_{\{\exists z\in\mathbb{Z}:h=k+2nz\}}\right)\right\|^2\left\|\left(I_N\otimes I_N\right)\right\|^2\\
&\qquad\cdot\left\|\mathbb{E}\left[\operatorname{vec}\left(\overline{\Gamma}_{n,N}(h)\right)\operatorname{vec}\left(\overline{\Gamma}_{n,N}(h)\right)^{\top}\right]\right\|\\
&\quad\leq \mathfrak{C}\sum_{h=1}^{n-1}\left\|\sum_{|k|>M}b_k\mathbf{1}_{\{\exists z\in\mathbb{Z}:h=k+2nz\}}\right\|^2\\
&\quad\leq \mathfrak{C}\left(\sum_{|k|>M}\|b_k\|\right)^2 \xrightarrow{M\to\infty} 0.
\end{aligned}$$

Step 2: We show

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{j=-n+1}^{n}\left(\operatorname{tr}\left(I_{n,N}(\omega_j)q_M(\omega_j)\right)-\operatorname{tr}\left(\eta(\omega_j)f_{Y^{(\Delta)}}(\omega_j)\right)\right)\\
&\quad= \frac{\sqrt{n}}{\pi}\operatorname{tr}\left(\sum_{h=-M}^{M}\left(\overline{\Gamma}_{n,N}(h)-\Gamma_N(h)\right)b_h\right)+o(1).
\end{aligned} \quad (37)$$

Let $M > n$. Then, due to Lemma 8 of the Supplementary Material and Parseval's equality, we receive

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{j=-n+1}^{n}\Big(\mathrm{tr}\big(I_{n,N}(\omega_j)q_M(\omega_j)\big)-\mathrm{tr}\big(\eta(\omega_j)f_{Y^{(\Delta)}}(\omega_j)\big)\Big)\\
&\quad=\frac{\sqrt{n}}{\pi}\mathrm{tr}\left(\sum_{h=-M}^{M}\overline{\Gamma}_{n,N}(h)b_h\right)-\frac{\sqrt{n}}{\pi}\int_{-\pi}^{\pi}\mathrm{tr}\big(\eta(\omega)f_{Y^{(\Delta)}}(\omega)\big)\mathrm{d}\omega\\
&\qquad+\frac{\sqrt{n}}{\pi}\int_{-\pi}^{\pi}\mathrm{tr}\big(\eta(\omega)f_{Y^{(\Delta)}}(\omega)\big)\mathrm{d}\omega-\mathrm{tr}\left(\frac{1}{\sqrt{n}}\sum_{j=-n+1}^{n}\eta(\omega_j)f_{Y^{(\Delta)}}(\omega_j)\right)\\
&\quad=\frac{\sqrt{n}}{\pi}\mathrm{tr}\left(\sum_{h=-M}^{M}\overline{\Gamma}_{n,N}(h)b_h\right)-\frac{\sqrt{n}}{\pi}\int_{-\pi}^{\pi}\mathrm{tr}\big(\eta(\omega)f_{Y^{(\Delta)}}(\omega)\big)\mathrm{d}\omega+o(1).
\end{aligned}
\tag{38}
$$

Taking $\Gamma_N(h)=0_{N\times N}$ for $h\neq 0$ into account, we receive

$$
\begin{aligned}
&\frac{\sqrt{n}}{\pi}\mathrm{tr}\left(\sum_{h=-M}^{M}\overline{\Gamma}_{n,N}(h)b_h\right)-\frac{\sqrt{n}}{\pi}\int_{-\pi}^{\pi}\mathrm{tr}\big(\eta(\omega)f_{Y^{(\Delta)}}(\omega)\big)\mathrm{d}\omega\\
&\quad=\frac{\sqrt{n}}{\pi}\mathrm{tr}\left(\sum_{h=-M}^{M}\Big(\overline{\Gamma}_{n,N}(h)-\Gamma_N(h)\Big)b_h\right)\\
&\qquad+\frac{\sqrt{n}}{\pi}\left(\mathrm{tr}(\Gamma_N(0)b_0)-\int_{-\pi}^{\pi}\mathrm{tr}\big(\eta(\omega)f_{Y^{(\Delta)}}(\omega)\big)\mathrm{d}\omega\right).
\end{aligned}
\tag{39}
$$

Using the representation $f_{Y^{(\Delta)}}(\omega)=\frac{1}{2\pi}\Phi(e^{-i\omega})\Sigma_N^{(\Delta)}\Phi(e^{i\omega})^\top$ and $q(\omega)=\Phi(e^{i\omega})^\top\eta(\omega)\Phi(e^{-i\omega})$ for $\omega\in[-\pi,\pi]$, yield

$$
\begin{aligned}
&\frac{\sqrt{n}}{\pi}\mathrm{tr}\big(\Gamma_N(0)b_0\big)-\mathrm{tr}\left(\int_{-\pi}^{\pi}\eta(\omega)f_{Y^{(\Delta)}}(\omega)\mathrm{d}\omega\right)\\
&\quad=\frac{\sqrt{n}}{\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left(\frac{1}{2\pi}\Sigma_N^{(\Delta)}q(\omega)\right)-\mathrm{tr}\big(\eta(\omega)f_{Y^{(\Delta)}}(\omega)\big)\mathrm{d}\omega\\
&\quad=\frac{\sqrt{n}}{\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left(\eta(\omega)\frac{1}{2\pi}\Phi(e^{-i\omega})\Sigma_N^{(\Delta)}\Phi(e^{i\omega})^\top-\eta(\omega)f_{Y^{(\Delta)}}(\omega)\right)\mathrm{d}\omega\\
&\quad=0.
\end{aligned}
\tag{40}
$$

Then, (38)–(40) result in (37).

Step 3: Next, we prove the asymptotic normality

$$
\frac{\sqrt{n}}{2\pi}\mathrm{tr}\left(\sum_{h=-M}^{M}\Big(\overline{\Gamma}_{n,N}(h)-\Gamma_N(h)\Big)b_h\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,\Sigma_\eta(M)),
\tag{41}
$$

where $\Sigma_\eta(M)$ is defined as

$$\Sigma_\eta(M) = \frac{1}{\pi^2}\sum_{h=1}^{M}\mathrm{tr}\Big(b_h\Sigma_N^{(\Delta)}b_h^H\Sigma_N^{(\Delta)}\Big) \\ + \frac{1}{4\pi^2}\mathrm{vec}(b_0^\top)^\top\Big(\mathbb{E}\Big[N_1^{(\Delta)}N_1^{(\Delta)\top}\otimes N_1^{(\Delta)}N_1^{(\Delta)\top}\Big] - \Sigma_N^{(\Delta)}\otimes\Sigma_N^{(\Delta)}\Big)\mathrm{vec}(b_0^H).$$

Therefore, we consider

$$\frac{\sqrt{n}}{2\pi}\mathrm{tr}\Bigg(\sum_{h=-M}^{M}\Big(\overline{\Gamma}_{n,N}(h)-\Gamma_N(h)\Big)b_h\Bigg) \\ = \frac{1}{\pi}\sum_{h=1}^{M}\sqrt{n}\mathrm{tr}\Big(\Big(\overline{\Gamma}_{n,N}(h)-\Gamma_N(h)\Big)b_h\Big) + \frac{\sqrt{n}}{2\pi}\mathrm{tr}\Big(\Big(\overline{\Gamma}_{n,N}(0)-\Gamma_N(0)\Big)b_0\Big). \tag{42}$$

Writing

$$\sqrt{n}\mathrm{tr}\Big(\Big(\overline{\Gamma}_{n,N}(h)-\Gamma_N(h)\Big)b_h\Big) = \sqrt{n}\mathrm{vec}(b_h^\top)^\top\mathrm{vec}\Big(\overline{\Gamma}_{n,N}(h)-\Gamma_N(h)\Big),$$

an application of Lemma 6 of the Supplementary Material leads to

$$\sqrt{n}\mathrm{tr}\Big(\Big(\overline{\Gamma}_{n,N}(h)-\Gamma_N(h)\Big)b_h\Big)\xrightarrow{\mathcal{D}}\mathcal{N}_h,$$

where $(\mathcal{N}_h)_{h\in\mathbb{N}_0}$ is an independent centered normally distributed sequence of random vectors with covariance matrix

$$\Sigma_{\mathcal{N}_h} := \mathrm{vec}(b_h^\top)^\top\Big(\Sigma_N^{(\Delta)}\otimes\Sigma_N^{(\Delta)}\Big)\mathrm{vec}(b_h^H) = \mathrm{tr}\Big(b_h\Sigma_N^{(\Delta)}b_h^H\Sigma_N^{(\Delta)}\Big) \quad \text{for } h\neq 0$$

and

$$\Sigma_{\mathcal{N}_0} := \mathrm{vec}(b_0^\top)^\top\Big(\mathbb{E}\Big[N_1^{(\Delta)}N_1^{(\Delta)\top}\otimes N_1^{(\Delta)}N_1^{(\Delta)\top}\Big] - \Sigma_N^{(\Delta)}\otimes\Sigma_N^{(\Delta)}\Big)\mathrm{vec}(b_0^H).$$

Finally,

$$\frac{\sqrt{n}}{2\pi}\mathrm{tr}\Bigg(\sum_{h=-M}^{M}\Big(\overline{\Gamma}_{n,N}(h)-\Gamma_N(h)\Big)b_h\Bigg)\xrightarrow{\mathcal{D}}\mathcal{N}\Bigg(0, \frac{1}{\pi^2}\sum_{h=1}^{M}\mathrm{tr}\Big(b_h\Sigma_N^{(\Delta)}b_h^H\Sigma_N^{(\Delta)}\Big) \\ + \frac{1}{4\pi^2}\mathrm{vec}(b_0^\top)^\top\Big(\mathbb{E}\Big[N_1^{(\Delta)}N_1^{(\Delta)\top}\otimes N_1^{(\Delta)}N_1^{(\Delta)\top}\Big] - \Sigma_N^{(\Delta)}\otimes\Sigma_N^{(\Delta)}\Big)\mathrm{vec}(b_0^H)\Bigg).$$

Step 4: We show

$$
\begin{aligned}
&\frac{1}{\pi^2}\sum_{h=1}^{M}\operatorname{tr}\Big(b_h\Sigma_N^{(\Delta)}b_h^H\Sigma_N^{(\Delta)}\Big)\\
&\quad+\frac{1}{4\pi^2}\operatorname{vec}(b_0^\top)^\top\Big(\mathbb{E}\Big[N_1^{(\Delta)}N_1^{(\Delta)\top}\otimes N_1^{(\Delta)}N_1^{(\Delta)\top}\Big]-\Sigma_N^{(\Delta)}\otimes\Sigma_N^{(\Delta)}\Big)\operatorname{vec}(b_0^H)\\
&\xrightarrow{M\to\infty}\frac{1}{\pi}\int_{-\pi}^{\pi}\operatorname{tr}\big(\eta(\omega)f_{Y^{(\Delta)}}(\omega)\eta(\omega)f_{Y^{(\Delta)}}(\omega)\big)\mathrm{d}\omega\\
&\quad+\frac{1}{16\pi^4}\int_{-\pi}^{\pi}\operatorname{vec}\big(\boldsymbol{\Phi}(e^{-i\omega})^\top\eta(\omega)^\top\boldsymbol{\Phi}(e^{i\omega})\big)^\top\mathrm{d}\omega\\
&\quad\cdot\Big(\mathbb{E}\Big[N_1^{(\Delta)}N_1^{(\Delta)\top}\otimes N_1^{(\Delta)}N_1^{(\Delta)\top}\Big]-3\Sigma_N^{(\Delta)}\otimes\Sigma_N^{(\Delta)}\Big)\int_{-\pi}^{\pi}\operatorname{vec}\big(\boldsymbol{\Phi}(e^{i\omega})^\top\eta(\omega)\boldsymbol{\Phi}(e^{-i\omega})\big)\mathrm{d}\omega.
\end{aligned}
\tag{43}
$$

Therefore, note that

$$
\begin{aligned}
&\frac{1}{\pi^2}\sum_{h=1}^{M}\operatorname{tr}\Big(b_h\Sigma_N^{(\Delta)}b_h^H\Sigma_N^{(\Delta)}\Big)\\
&\qquad+\frac{1}{4\pi^2}\operatorname{vec}(b_0^\top)^\top\Big(\mathbb{E}\Big[N_1^{(\Delta)}N_1^{(\Delta)\top}\otimes N_1^{(\Delta)}N_1^{(\Delta)\top}\Big]-\Sigma_N^{(\Delta)}\otimes\Sigma_N^{(\Delta)}\Big)\operatorname{vec}(b_0^H)\\
&\quad\xrightarrow{M\to\infty}\frac{1}{\pi^2}\sum_{h=1}^{\infty}\operatorname{tr}\Big(b_h\Sigma_N^{(\Delta)}b_h^H\Sigma_N^{(\Delta)}\Big)\\
&\qquad+\frac{1}{4\pi^2}\operatorname{vec}(b_0^\top)^\top\Big(\mathbb{E}\Big[N_1^{(\Delta)}N_1^{(\Delta)\top}\otimes N_1^{(\Delta)}N_1^{(\Delta)\top}\Big]-\Sigma_N^{(\Delta)}\otimes\Sigma_N^{(\Delta)}\Big)\operatorname{vec}(b_0^H).
\end{aligned}
$$

But

$$
\begin{aligned}
\frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\operatorname{tr}\Big(b_h\Sigma_N^{(\Delta)}b_h^H\Sigma_N^{(\Delta)}\Big)&=\frac{1}{4\pi^2}\sum_{h=-\infty}^{\infty}\sum_{\ell=-\infty}^{\infty}\operatorname{tr}\Big(b_h\Sigma_N^{(\Delta)}b_\ell^H\Sigma_N^{(\Delta)}\Big)\int_{-\pi}^{\pi}e^{i(h-\ell)\omega}\mathrm{d}\omega\\
&=\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\operatorname{tr}\Big(q(\omega)\Sigma_N^{(\Delta)}q(\omega)^H\Sigma_N^{(\Delta)}\Big)\mathrm{d}\omega\\
&=\int_{-\pi}^{\pi}\operatorname{tr}\big(\eta(\omega)f_{Y^{(\Delta)}}(\omega)\eta(\omega)^Hf_{Y^{(\Delta)}}(\omega)\big)\mathrm{d}\omega,
\end{aligned}
$$

where we plugged in the definition of q in the last equality. Eventually, due to the representation of b_0, we receive

$$\begin{aligned}
&\frac{1}{\pi^2}\sum_{h=1}^{\infty}\operatorname{tr}\Big(b_h\Sigma_N^{(\Delta)}b_h^H\Sigma_N^{(\Delta)}\Big)\\
&\quad+\frac{1}{4\pi^2}\operatorname{vec}(b_0^\top)^\top\Big(\mathbb{E}\Big[N_1^{(\Delta)}N_1^{(\Delta)\top}\otimes N_1^{(\Delta)}N_1^{(\Delta)\top}\Big]-\Sigma_N^{(\Delta)}\otimes\Sigma_N^{(\Delta)}\Big)\operatorname{vec}(b_0^H)\\
&=\frac{1}{\pi}\int_{-\pi}^{\pi}\operatorname{tr}\Big(\eta(\omega)f_{Y^{(\Delta)}}(\omega)\eta(\omega)f_{Y^{(\Delta)}}(\omega)\Big)\mathrm{d}\omega\\
&\quad+\frac{1}{16\pi^4}\int_{-\pi}^{\pi}\operatorname{vec}\Big(\Phi(e^{-i\omega})^\top\eta(\omega)^\top\Phi(e^{i\omega})\Big)^\top\mathrm{d}\omega\\
&\quad\cdot\Big(\mathbb{E}\Big[N_1^{(\Delta)}N_1^{(\Delta)\top}\otimes N_1^{(\Delta)}N_1^{(\Delta)\top}\Big]-3\Sigma_N^{(\Delta)}\otimes\Sigma_N^{(\Delta)}\Big)\\
&\quad\cdot\int_{-\pi}^{\pi}\operatorname{vec}\Big(\Phi(e^{i\omega})^\top\eta(\omega)\Phi(e^{-i\omega})\Big)\mathrm{d}\omega.
\end{aligned}$$

Finally, Step 3, Step 4 and a multivariate version of Problem 6.16 of Brockwell and Davis (1991) give

$$\frac{\sqrt{n}}{2\pi}\operatorname{tr}\left(\sum_{h=-M}^{M}\Big(\overline{\Gamma}_{n,N}(h)-\Gamma_N(h)\Big)b_h\right)\overset{\mathcal{D},n\to\infty}{\longrightarrow}\mathcal{N}(0,\Sigma_\eta(M))\overset{\mathcal{D},M\to\infty}{\longrightarrow}\mathcal{N}(0,\Sigma_\eta).$$

Along with Step 1, Step 2 and Proposition 6.3.9 of Brockwell and Davis (1991), the statement follows. □

Proof of Proposition 6 The proof is based on the Cramér Wold Theorem and Proposition 5. Therefore, let $\lambda=(\lambda_1,\dots,\lambda_r)^\top\in\mathbb{R}^r$. We obtain

$$\begin{aligned}
\sqrt{n}\big[\nabla_\vartheta W_n(\vartheta_0)\big]\lambda&=\frac{1}{2\sqrt{n}}\sum_{j=-n+1}^{n}\nabla_\vartheta\Big[\operatorname{tr}\Big(f_{Y^{(\Delta)}}(\omega_j,\vartheta_0)^{-1}I_n(\omega_j)\Big)+\log(\det(f_{Y^{(\Delta)}}(\omega_j,\vartheta_0)))\Big]\lambda\\
&=\frac{1}{2\sqrt{n}}\sum_{j=-n+1}^{n}\left[\sum_{t=1}^{r}\operatorname{tr}\Big(-\lambda_t f_{Y^{(\Delta)}}(\omega_j)^{-1}\Big(\frac{\partial}{\partial\vartheta_t}f_{Y^{(\Delta)}}(\omega_j,\vartheta_0)\Big)f_{Y^{(\Delta)}}(\omega_j)^{-1}I_n(\omega_j)\Big)\right]\\
&\quad+\frac{1}{2\sqrt{n}}\sum_{j=-n+1}^{n}\nabla_\vartheta[\operatorname{tr}(\log(f_{Y^{(\Delta)}}(\omega_j,\vartheta_0)))]\lambda.
\end{aligned}$$

We define the matrix function $\eta_\lambda:[-\pi,\pi]\to\mathbb{C}^{m\times m}$ as

$$\eta_\lambda(\omega)=-\sum_{t=1}^{r}\lambda_t f_{Y^{(\Delta)}}(\omega)^{-1}\left(\frac{\partial}{\partial\vartheta_t}f_{Y^{(\Delta)}}(\omega,\vartheta_0)\right)f_{Y^{(\Delta)}}(\omega)^{-1},\quad\omega\in[-\pi,\pi].\tag{44}$$

Furthermore,

$$\operatorname{tr}\left(\frac{\partial}{\partial\vartheta_t}\log\big(f_{Y^{(\Delta)}}(\omega,\vartheta_0)\big)\right)=\operatorname{tr}\left(f_{Y^{(\Delta)}}(\omega)^{-1}\left(\frac{\partial}{\partial\vartheta_t}f_{Y^{(\Delta)}}(\omega,\vartheta_0)\right)\right).$$

Then,

$$\sqrt{n}\big[\nabla_\vartheta W_n(\vartheta_0)\big]\lambda = \frac{1}{2\sqrt{n}}\sum_{j=-n+1}^{n}\mathrm{tr}\big(\eta_\lambda(\omega_j)\big(I_n(\omega_j)-f_{Y^{(\Delta)}}(\omega_j)\big)\big).$$

Apparently, η_λ is two times continuously differentiable by Remark 2 and 2π periodic. Moreover, every component of the Fourier coefficients $(\mathfrak{f}_{\lambda,u})_{u\in\mathbb{Z}}$ of η_λ satisfies $\sum_{u=-\infty}^{\infty}|\mathfrak{f}_{\lambda,u}[k,\ell]||u|^{1/2}<\infty, k,\ell\in\{1,\ldots,m\}$ (see Brockwell and Davis 1991, Exercise 2.22 applied to η_λ and its derivative η_λ'), and therefore, $\sum_{u=-\infty}^{\infty}\|\mathfrak{f}_{\lambda,u}\||u|^{1/2}<\infty$ follows. Then, due to Proposition 5, we get as $n\to\infty$,

$$\sqrt{n}\big[\nabla_\vartheta W_n(\vartheta_0)\big]\lambda \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma_{\nabla_\vartheta W\lambda}),$$

where

$$\begin{aligned}\Sigma_{\nabla_\vartheta W\lambda} &= \frac{1}{\pi}\int_{-\pi}^{\pi}\mathrm{tr}\big(\eta_\lambda(\omega)f_{Y^{(\Delta)}}(\omega)\eta_\lambda(\omega)f_{Y^{(\Delta)}}(\omega)\big)\mathrm{d}\omega\\ &\quad+\frac{1}{16\pi^4}\int_{-\pi}^{\pi}\mathrm{vec}\big(\Phi(e^{-i\omega})^\top\eta_\lambda(\omega)^\top\Phi(e^{i\omega})\big)^\top\mathrm{d}\omega\\ &\quad\cdot\Big(\mathbb{E}\Big[N_1^{(\Delta)}N_1^{(\Delta)\top}\otimes N_1^{(\Delta)}N_1^{(\Delta)\top}\Big]-3\Sigma_N^{(\Delta)}\otimes\Sigma_N^{(\Delta)}\Big)\\ &\quad\cdot\int_{-\pi}^{\pi}\mathrm{vec}\big(\Phi(e^{i\omega})^\top\eta_\lambda(\omega)\Phi(e^{-i\omega})\big)\mathrm{d}\omega\\ &=:\ \Sigma_{\lambda,1}+\Sigma_{\lambda,2}+\Sigma_{\lambda,3}.\end{aligned}$$

We investigate the three terms separately. With (29), the first term fulfills the representation

$$\begin{aligned}\Sigma_{\lambda,1} &= \frac{1}{\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left(\left(\sum_{t=1}^{r}\lambda_t\frac{\partial}{\partial\vartheta_t}f_{Y^{(\Delta)}}(\omega,\vartheta_0)\right)f_{Y^{(\Delta)}}(\omega)^{-1}\left(\sum_{s=1}^{r}\lambda_s\frac{\partial}{\partial\vartheta_s}f_{Y^{(\Delta)}}(\omega,\vartheta_0)\right)f_{Y^{(\Delta)}}(\omega)^{-1}\right)\mathrm{d}\omega\\ &= \lambda^\top\left[\frac{1}{\pi}\int_{-\pi}^{\pi}\big(\nabla_\vartheta f_{Y^{(\Delta)}}(-\omega,\vartheta_0)\big)^\top\big(f_{Y^{(\Delta)}}(-\omega)^{-1}\otimes f_{Y^{(\Delta)}}(\omega)^{-1}\big)\nabla_\vartheta f_{Y^{(\Delta)}}(\omega,\vartheta_0)\right]\lambda.\end{aligned}$$

Similarly, we get the representation

$$\begin{aligned}\Sigma_{\lambda,2} = \frac{\lambda^\top}{16\pi^4}\Big[\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}&\mathbb{E}\big[\nabla_\vartheta f_{Y^{(\Delta)}}(-\omega,\vartheta_0)^\top\big(f_{Y^{(\Delta)}}(-\omega)^{-1}\Phi(e^{i\omega})\ \otimes f_{Y^{(\Delta)}}(\omega)^{-1}\Phi(e^{-i\omega})\big)N_1^{(\Delta)}N_1^{(\Delta)\top}\\ &\otimes N_1^{(\Delta)}N_1^{(\Delta)\top}\big(\Phi(e^{-i\tau})^\top f_{Y^{(\Delta)}}(-\tau)^{-1}\otimes\Phi(e^{i\tau})^\top f_{Y^{(\Delta)}}(\tau)^{-1}\big)\ \nabla_\vartheta f_{Y^{(\Delta)}}(\tau,\vartheta_0)\big]\mathrm{d}\omega\mathrm{d}\tau\Big]\lambda\end{aligned}$$

for the second term, and analogously

$$\Sigma_{\lambda,3} = -\frac{3\lambda^\top}{16\pi^4}\Big[\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \nabla_\vartheta f_{Y^{(\Delta)}}(-\omega,\vartheta_0)^\top \cdot \Big(f_{Y^{(\Delta)}}(-\omega)^{-1}\Phi(e^{i\omega})\Sigma_N^{(\Delta)}\Phi(e^{-i\tau})^\top f_{Y^{(\Delta)}}(-\tau)^{-1}\Big) \otimes\Big(f_{Y^{(\Delta)}}(\omega)^{-1}\Phi(e^{-i\omega})\ \Sigma_N^{(\Delta)}\Phi(e^{i\tau})^\top f_{Y^{(\Delta)}}(\tau)^{-1}\Big) \cdot \nabla_\vartheta f_{Y^{(\Delta)}}(\tau,\vartheta_0)\mathrm{d}\omega\mathrm{d}\tau\Big]\lambda$$

for the third term. □

Proof of Theorem 2 Since $\hat{\vartheta}_n^{(\Delta)} \xrightarrow{\text{a.s.}} \vartheta_0$ (see Theorem 1) and $\Sigma_{\nabla^2 W}$ is positive definite (see Theorem 2) the conclusion follows from (12), Propositions 4 and 6. □

Sketch of the proof of Remark 5 Let Φ_Z be the polynomial of the (existing) VAR(∞) of the VARMA(p, q) process. Proposition 5 can be formulated for VARMA processes. As in the proof of Theorem 2 we have to plug in there for η the function η_λ as given in (44). Then, b_0 in (30) has for the VARMA process $(Z_n)_{n\in\mathbb{N}}$ the form

$$\begin{aligned} b_0 &= \int_{-\pi}^{\pi} -2\pi\sum_{t=1}^{r}\lambda_t\Sigma_e^{-1}\Phi_Z(e^{-i\omega})^{-1}\left(\frac{\partial}{\partial\vartheta_t}f_Z(\omega,\vartheta_0)\right)\Phi_Z(e^{i\omega})^{\top-1}\Sigma_e^{-1}\mathrm{d}\omega \\ &= -\Sigma_e^{-1}\int_{-\pi}^{\pi}\sum_{t=1}^{r}\lambda_t\frac{\partial}{\partial\vartheta_t}\log\left(\Phi_Z(e^{-i\omega},\vartheta_0)\right)\mathrm{d}\omega \\ &\quad -\int_{-\pi}^{\pi}\left(\sum_{t=1}^{r}\lambda_t\frac{\partial}{\partial\vartheta_t}\log\left(\Phi_Z(e^{i\omega},\vartheta_0)\right)\right)^\top \mathrm{d}\omega\ \Sigma_e^{-1}. \end{aligned}$$

If Φ_Z is two times differentiable, the Leibniz rule yield

$$\begin{aligned} b_0 = &-\Sigma_e^{-1}\sum_{t=1}^{r}\lambda_t\frac{\partial}{\partial\vartheta_t}\int_{-\pi}^{\pi}\log\left(\Phi_Z(e^{-i\omega},\vartheta_0)\right)\mathrm{d}\omega \\ &-\left[\sum_{t=1}^{r}\lambda_t\frac{\partial}{\partial\vartheta_t}\int_{-\pi}^{\pi}\log\left(\Phi_Z(e^{i\omega},\vartheta_0)\right)\mathrm{d}\omega\right]^\top\Sigma_e^{-1}. \end{aligned}$$

Similarly to the proof of Theorem 5.8.1 of Brockwell and Davis (1991), one can show that the integrals are constant and therefore, that $b_0 = 0$. For a more detailed approach, we refer to Dunsmuir and Hannan (1976).

Supplementary Information The online version supplementary material is available at https://doi.org/10.1007/s10463-021-00802-6.

References

Bardet, J.-M., Doukhan, P., León, J. R. (2008). Uniform limit theorems for the integrated periodogram of weakly dependent time series and their applications to Whittle's estimate. *Journal of Time Series Analysis*, *29*(5), 906–945.

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, *24*(1), 1–13.

Blevins, J. (2017). Identifying restrictions for finite parameter continuous time models with discrete time data. *Econometric Theory*, *33*, 739–754.

Brewer, J. (1978). Kronecker products and matrix calculus in system theory. *IEEE Transactions on Circuits and Systems. I. Regular Papers*, *25*(9), 772–781.

Brockwell, P. (2014). Recent results in the theory and applications of CARMA processes. *Annals of the Institute of Statistical Mathematics*, *66*(4), 647–685.

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

Brockwell, P. J., Lindner, A. (2009). Existence and uniqueness of stationary Lévy-driven CARMA processes. *Stochastic Processes and their Applications*, *119*(8), 2660–2681.

Brockwell, P. J., Lindner, A. (2019). Sampling, embedding and inference for CARMA processes. *Journal of Time Series Analysis*, *40*(2), 163–181.

Chambers, M. J., McCrorie, J. R., Thornton, M. A. (2018). Continuous time modelling based on an exact discrete time representation. In K. van Montfort, J. Oud & M. Voelkle (Eds.), *Continuous time modeling in the behavioral and related sciences* (pp. 317–357). Springer.

Dahlhaus, R. (1988). Empirical spectral processes and their applications to time series analysis. *Stochastic Processes and their Applications*, *30*(1), 69–83.

Dahlhaus, R., Polonik, W. (2006). Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes. *The Annals of Statistics*, *34*(6), 2790–2824.

Dahlhaus, R., Polonik, W. (2009). Empirical spectral processes for locally stationary time series, *Bernoulli*, *15*, 1–39.

Dahlhaus, R., Pötscher, B. (1989). Convergence results for maximum likelihood type estimators in multivariable ARMA models II. *Journal of Multivariate Analysis*, *30*(2), 241–244.

Deistler, M., Dunsmuir, W., Hannan, E. J. (1978). Vector linear time series models: Corrections and extensions. *Advances in Applied Probability*, *10*(2), 360–372.

Dunsmuir, W., Hannan, E. J. (1976). Vector linear time series models. *Advances in Applied Probability*, *8*(2), 339–364.

Fasen, V. (2013). Statistical inference of spectral estimation for continuous-time MA processes with finite second moments. *Mathematical Methods of Statistics*, *22*(4), 283–309.

Fasen, V., Fuchs, F. (2013). Spectral estimates for high-frequency sampled continuous-time autoregressive moving average processes. *Journal of Time Series Analysis*, *34*(5), 532–551.

Fasen-Hartmann, V., Kimmig, S. (2020). Robust estimation of continuous-time ARMA models via indirect inference. *Journal of Time Series Analysis*, *41*, 620–651.

Fasen-Hartmann, V., Scholz, M. (2019). Quasi-maximum likelihood estimation for cointegrated solutions of continuous-time state space models observed at discrete time points. *Electronic Journal of Statistics*, *13*(2), 5151–5212.

Guidorzi, R. (1975). Canonical structures in the identification of multivariable systems. *Automatica*, *11*(4), 361–374.

Hannan, E. J. (2009). *Multiple time series*. Wiley series in probability and mathematical statistics. Wiley.

Hannan, E. J., Deistler, M. (1988). *The statistical theory of linear systems*. Wiley.

Hansen, H. P., Sergant, T. (1983). The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica*, *51*, 377–387.

Harvey, A. C., Stock, J. H. (1985). The estimation of higher-order continuous time autoregressive models. *Econometric Theory*, *1*(1), 97–117.

Harvey, A. C., Stock, J. H. (1988). Continuous time autoregressive models with common stochastic trends. *Journal of Economic Dynamics, Control*, *12*(2–3), 365–384.

Harvey, A. C., Stock, J. H. (1989). Estimating integrated higher-order continuous time autoregressions with an application to money-income causality. *Journal of Econometrics*, *42*(3), 319–336.

Körner, T. W. (1989). *Fourier analysis*. Cambridge University Press.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer.

Marquardt, T., Stelzer, R. (2007). Multivariate CARMA processes. *Stochastic Processes and their Applications*, *117*(1), 96–120.
Mikosch, T., Norvaiša, R. (1997). Uniform convergence of the empirical spectral distribution function. *Stochastic Processes and their Applications*, *70*(1), 85–114.
Mikosch, T., Gadrich, T., Klüppelberg, C., Adler, R. J. (1995). Parameter estimation for ARMA models with infinite variance innovations. *The Annals of Statistics*, *23*(1), 305–326.
Newey, K., McFadden, D. (1994). Large sample estimation and hypothesis. In R. F. Engle & D. McFadden (Eds.), *Handbook of econometrics* (Vol. 4, pp. 2112–2245). Elsevier.
Øigård, T. A., Hanssen, A., Hansen, R. E., Godtliebsen, F. (2005). EM-estimation and modeling of heavy-tailed processes with the multivariate normal inverse Gaussian distribution. *Signal Processing*, *85*(8), 1655–1673.
Phillips, P. C. (1973). The problem of identification in finite parameter continuous time models. *Journal of Econometrics*, *1*, 351–362.
Priestley, M. B. (1981). *Spectral analysis and time series*. Academic Press Inc.
Sato, K.-I. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge University Press.
Schlemm, E., Stelzer, R. (2012a). Quasi maximum likelihood estimation for strongly mixing state space models and multivariate Lévy-driven CARMA processes. *Electronic Journal of Statistics*, *6*, 2185–2234.
Schlemm, E., Stelzer, R. (2012b). Multivariate CARMA processes, continuous-time state space models and complete regularity of the innovations of the sampled processes. *Bernoulli*, *18*(1), 46–63.
Thornton, M. A., Chambers, M. J. (2017). Continuous time ARMA processes: Discrete time representation and likelihood evaluation. *Journal of Economic Dynamics & Control*, *79*, 48–65.
Tsai, H., Chan, K. (2005). Quasi-maximum likelihood estimation for a class of continuous-time long-memory processes. *Journal of Time Series Analysis*, *26*(5), 691–713.
Walker, A. M. (1964). Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *Journal of the Australian Mathematical Society*, *4*, 363–384.
Whittle, P. (1951). *Hypothesis testing in time series analysis*. PhD thesis, Uppsala University, Uppsala.
Whittle, P. (1953). Estimation and information in stationary time series. *Arkiv för Matematik*, *2*(5), 423–434.
Zadrozny, P. (1988). Gaussian likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies. *Econometric Theory*, *4*(1), 108–124.