Frequency Domain Minimum Distance Estimation of Possibly Noninvertible and Noncausal ARMA models*

Carlos Velasco and Ignacio N. Lobato

Universidad Carlos III de Madrid and Instituto Tecnológico Autónomo de México

September 14, 2015

Abstract

This article introduces frequency domain minimum distance procedures for performing inference in general possibly non causal and/or noninvertible autoregressive moving average (ARMA) models. We use information from higher order moments to achieve identification on the location of the roots of the AR and MA polynomials for non-Gaussian time series. We study minimum distance estimation that combines the information contained in second, third, and fourth moments. Contrary to existing estimators, the proposed estimator is consistent under general assumptions, and can improve on the efficiency of the estimates based on second order moments only.

Keywords and Phrases: nonfundamentalness; higher-order moments; higher-order spectra; noninvertible moving average; minimum phase.

1. INTRODUCTION

Estimation of autoregressive-moving average (ARMA) models is typically performed under causality and invertibility assumptions using second-order procedures, such as least squares or some variant of the Gaussian maximum likelihood (ML) estimator. Causality and invertibility are crucial assumptions when using second-order estimation procedures since these cannot identify non-causal or non-invertible representations. Hence, for estimation of Gaussian ARMA processes causality and invertibility need to be imposed. For non-Gaussian ARMA models the causality and invertibility

*This research was carried out while Lobato was visiting Universidad Carlos III de Madrid thanks to the Cátedras de Excelencia del Banco de Santander program. Lobato acknowledges financial support from Asociación Mexicana de Cultura and from the Mexican Consejo Nacional de Ciencia y Tecnología (CONACYT) under project grant 151624. Velasco Research acknowledges financial support from the Spanish Plan Nacional de I+D+I (SEJ2007-62908, ECO2012-31748 and ECO2014-57007).

assumptions are not necessary and not always justified, and, in fact, non-causal or non-invertible ARMA models have been employed in many areas such as economics, seismology, engineering or astronomy; for some examples in economics see Alessi, Barigozzi, and Capasso (2011), Hansen and Sargent (1980, 1991), Huang and Pawitan (2000), Leeper, Walter, and Yang (2013), and Montford and Uhlig (2009).

The literature devoted to estimating general non-standard ARMA models can be classified according to two criteria. The first criteria is whether the distribution for the innovations is assumed to be known or unknown, so either Maximum likelihood or alternative approaches, such as those based on methods of moments, are used. The second criteria is whether estimation is performed in one or in two steps. Estimation in one step attempts to estimate the general possibly non-causal non-invertible ARMA models irrespectively of whether the absolute values of the roots of AR and MA polynomials are larger or smaller than one. Estimation in two steps consists of performing an initial first step where a causal and invertible model is estimated using standard procedures, and then, in a second step, an ARMA all-pass model, which is a model where all the roots of the AR part are the inverse of the roots of the MA part, is fitted to the first-step white noise residuals.

Lii and Rosenblatt (1992) investigate the properties of a one-step approximate maximum likelihood procedure for the possibly noninvertible moving average case where the exact distribution of the innovations is known. Breidt, Davis and Trindade (2001) propose a two-step estimation where the least absolute deviation (LAD) estimator is applied to the first step white noise all pass residuals. This LAD estimator coincides with the maximum likelihood estimator for a Laplacian distribution of the innovations. Andrews, Davis and Breidt (2007) extend the previous article to rank-estimators so they can dispose of the Laplacian assumption. Using also a two step approach, Kumon (1992) proposes estimates based on the extension to higher order spectral densities of the frequency domain Whittle (1953) estimate first order conditions. In a spirit similar to Kumon, Ahn, Leonenko and Sakhno (2007) propose estimates based on related estimating equations using weighted higher order periodograms. Note that in both references, identification is achieved by introducing arbitrary conditions on the sign of the higher order cumulants of the innovation (Kumon) or on some weighting function that controls the sign of the cumulant (Ahn et al.). Lanne and Saikkonen (2011) employ a two-step strategy for estimating general noncausal AR models by maximum likelihood.

ML procedures are subject to the usual arbitrariness criticism, whereas the two step approach presents the obvious problem of independently estimating twice the same parameters, so it is unclear the asymptotic properties of the final estimates recovered from these two steps. In addition, the second step is only meaningful when the first step residuals are non-independent white noise, hence an independence test should be implemented between the two steps, which is not typically considered. These criticisms lead to the conclusion that, ideally, one would like to employ a one step estimation procedure without restricting the distribution of the innovations.

This article proposes a one-step estimator that does not rely on arbitrary distributional assumptions nor on arbitrary identification conditions. The only requisites we need for identification is that some higher order cumulant (either the third or the fourth) is non zero and that the innovations are independent up to this order. In particular, we follow the approach of Brillinger (1985), and propose minimum distance (MD, hereinfater) procedures based on second and higher order information in the frequency domain. Minimum distance procedures have also been employed in the time domain, see Ramsey and Montenegro (1992), Amano and Lobato (2012) and Gospodinov and Ng (2012), but these time-domain minimum distance estimators are inefficient since they only employ a finite number of moment conditions. An additional advantage of carrying out the analysis in the frequency domain is the useful asymptotic properties of the frequency domain statistics when evaluated at Fourier frequencies, see details in Section 2.

The first contribution of this article is to establish global identification of the parameters of a possibly non-causal non-invertible ARMA model using higher order spectral densities as long as some higher order cumulant of the innovations is not zero. This theorem motivates a general minimum distance frequency domain objective function (OF, hereafter) that, similar to Brillinger (1985) or Terdik (1999), employs the information in second, third and fourth moments. The second contribution of this article is to establish the consistency and the asymptotic normality of estimators based on this objective function for ARMA models irrespective of noncausality or noninvertibility. Hence, for general ARMA models with no distributional assumptions, this article is the first one to establish a one step estimation method with rigorously established statistical properties. Also note that the proposed procedures overcome the need of using tests for causality or invertibility, tests for independence, and our theory also cover all-pass models, both ignoring or using such configuration.

The general objective function studied in this paper weights the information coming from second, third and fourth moments. The particular case of ignoring third and fourth moments leads to a minimum distance estimator, based just on second moments, which is shown to be asymptotically equivalent to the efficient Whittle estimator, the Gaussian quasi ML estimator (QMLE). Using the information contained in higher order moments does not only achieve identification, but also, under certain conditions, may improve on the efficiency with respect to the Whittle estimator.

Although the article focus on identification and estimation of ARMA models,we emphasize that great part of the technical proofs for the asymptotic properties of parameter estimates are established for general linear models where the parametric filter function is smooth enough. A technical contribution of the article, of independent interest, is a central limit theorem (CLT) for martingales where the leading term are centered powers of the innovations.

The article is structured as follows. Section 2 introduces the notation and the model, and recalls the basics properties of the (higher-order) periodograms. Section 3 studies the identification of noncausal/noninvertible ARMA models. Section 4 presents the proposed minimum distance estimator, and establishes its asymptotic properties. Section 5 contains simulations, and Section 6 concludes. Proofs and two auxiliary lemmas are contained in the appendices.

2. NOTATION, MODEL AND BASICS OF PERIODOGRAMS

2.1 Notation

Consider a stationary stochastic process $\{Y_t\}_{t\in Z}$ with $E[Y_t^2] < \infty$ and call $\mu = E[Y_t]$. Define the autocovariance of order j as

$$\gamma_j = Cov(Y_t, Y_{t-j}) = E[(Y_t - \mu)(Y_{t-j} - \mu)], \qquad \text{for } j = 0, \pm 1, \ldots,$$

and the j-th order autocorrelation as $\rho_j = \gamma_j/\gamma_0$. The spectral density, $f(\lambda)$, is defined implicitly as

$$\gamma_j = \int_{-\pi}^{\pi} f(\lambda)\exp(-ij\lambda)d\lambda.$$

The autocovariance sequence and the spectral density are measures of the dependence of the stochastic process based on second moments, hence they are the objects of interest of usual time series analysis. The dependence contained in higher order moments can also be described by the cumulants which are defined in terms of higher order moments as

$$\text{cum}(Y_1, \ldots, Y_k) = \sum (-1)^{p-1}(p-1)! E(\Pi_{j\in\nu_1} Y_j) \cdots E(\Pi_{j\in\nu_p} Y_j), \quad k = 1, 2, \ldots$$

assuming $E[|Y_t|^k] < \infty$ and where $\nu_1, \ldots, \nu_p$ is a partition of $(1, 2, \ldots, k)$ and the sum runs over all these partitions, see Brillinger (1975) or Rosenblatt (1985, p. 34). Hence, the first and second cumulants are the mean and the variance, respectively.

We also define the k-th order cumulant spectral density $k = 2, 3, \ldots$, which is the Fourier transform of the k-th order cumulant, as

$$f_k(\boldsymbol{\lambda}) = f_k(\lambda_1, \ldots, \lambda_{k-1}) = \frac{1}{(2\pi)^{k-1}} \sum_{j_1,\ldots,j_{k-1}=-\infty}^{\infty} \text{cum}(Y_t, Y_{t+j_1}, \ldots, Y_{t+j_{k-1}}) \exp\left(-\sum_{s=1}^{k-1} ij_s\lambda_s\right), \tag{1}$$

introducing for simplicity, when there is no ambiguity, the notation $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{k-1})$. Note that the usual spectral density is then recovered for $k = 2$, while f_k can be complex valued for $k > 2$, unlike $f = f_2$ which is always real valued.

From a sample of size T, one can consistently estimate the higher order moments and cumulants by their sample analogs. In order to address estimation of the higher order spectral densities, first we recall the definitions of the finite Fourier transform as

$$w(\lambda) = \sum_{t=1}^{T} Y_t \exp(-it\lambda),$$

and of the standard second order periodogram

$$I(\lambda) = \frac{1}{2\pi T} w(\lambda) w(-\lambda) = \frac{1}{2\pi T} |w(\lambda)|^2 . \tag{2}$$

Expression (2) can be easily extended to define the higher order periodogram of order k as

$$I_k(\boldsymbol{\lambda}) = I_k(\lambda_1, \ldots, \lambda_{k-1}) = \frac{1}{(2\pi)^{k-1} T} \prod_{j=1}^{k-1} w(\lambda_j) w(-\sum_{i=1}^{k-1} \lambda_i), \tag{3}$$

where we use the notation in (1). In particular, the statistic $I_3(\lambda_1, \lambda_2)$ is called the biperiodogram, which is the natural (although inconsistent) estimator of the bispectral density, $f_3(\lambda_1, \lambda_2)$, and the triperiodogram, $I_4(\lambda_1, \lambda_2, \lambda_3)$ that is the natural estimator of the trispectral density, $f_4(\lambda_1, \lambda_2, \lambda_3)$.

Similar to the periodogram, both biperiodogram and triperiodogram are asymptotically unbiased estimators but inconsistent. In particular, we have that, under Assumption 1 below, the following properties hold for fixed frequencies $\boldsymbol{\lambda}$, $k = 2, 3, \ldots$,

$$E\left[I_k(\boldsymbol{\lambda})\right] = f_k(\boldsymbol{\lambda}) + o(1), \tag{4}$$

and when $\boldsymbol{\lambda} \neq \mathbf{0} \bmod 2\pi$,

$$\text{Var}\left[I_k(\boldsymbol{\lambda})\right] = T^{k-2}\left((k-1)! f(\lambda_1) f(\lambda_2) \cdots f(\lambda_{k-1}) f(\lambda_1 + \lambda_2 + \cdots + \lambda_{k-1}) + o(1)\right) \tag{5}$$

as $T \to \infty$. Note that these properties hold under a variety of weak dependence conditions, for instance, mixing or summability of cumulants, see Brillinger (1975), Rosenblatt (1985, p. 172-173) or Alekseev (1993). Note also that by tapering these variances can be reduced, see Alekseev (1993).

In this article, as it is common in time series analysis, we are going to evaluate the statistics at the Fourier frequencies defined as $\lambda_j = \frac{2\pi j}{T}$, for $j = 1, \ldots, T-1$. The main reason is that when evaluated at (different) Fourier frequencies ($\boldsymbol{\lambda} \neq \boldsymbol{\lambda}'$) the higher order spectra are asymptotically uncorrelated, that is

$$\text{Cov}(I_k(\boldsymbol{\lambda}), I_\ell(\boldsymbol{\lambda}')) = o\left(T^{(k+\ell-4)/2}\right), \quad k, \ell = 2, \ldots \tag{6}$$

as T tends to infinity for almost all $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$ (not satisfying some particular linear restrictions), where we use the notation in (1), see for instance Lemma 1 in p. 172 in Rosenblatt (1985), and Theorems 2 and 4 in Alekseev (1993).

2.2 Model

We assume that Y_t is given by

$$Y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \tag{7}$$

where ε_t is iid$(0, \kappa_2)$ with bounded moments of order $k \geq 3$ and $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$ (see Assumptions 1 and 2 below).

A model establishes a structure on the $\psi_j' s$ in terms of some parameter vector $\theta \in \mathbb{R}$:

$$\psi_j = \psi_j(\theta),$$

and the target is the estimation of θ. This article focuses on ARMA(p, q) models where

$$\alpha(L)Y_t = \beta(L)\varepsilon_t \tag{8}$$

where the polynomials $\alpha(L) = 1 - \sum_{j=1}^{p} \alpha_j L$ and $\beta(L) = 1 + \sum_{j=1}^{q} \beta_j L$ are of order p and q respectively, have all their roots away the unit circle, inside or outside, and do not have any common roots. Denote the model parameters by $\theta = (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)' \in \Gamma_{p,q} = \{\theta \in \mathbb{R}^{p+q} : \alpha(z) \neq 0,\ \beta(z) \neq 0 \text{ for } |z| = 1, \alpha_p \neq 0,\ \beta_q \neq 0\}$. Since θ can be expressed as a continuous function $\theta(\boldsymbol{\varphi})$, $\boldsymbol{\varphi} = (a_1, \ldots, a_p, b_1, \ldots, b_q)'$ of the zeroes $a_1, \ldots, a_p$ of $\alpha(\cdot)$ and $b_1, \ldots, b_q$ of $\beta(\cdot)$, the parameter set $\Gamma_{p,q}$ is the image under $\theta(\cdot)$ of the set

$$\mathcal{A}_{p,q} = \left\{ \begin{array}{c} \boldsymbol{\varphi} = (a_1, \ldots, a_p, b_1, \ldots, b_q)' \in \bigcup_{r=0}^{\lfloor p/2 \rfloor + \lfloor q/2 \rfloor} \left(\mathbb{R}^{p+q-2r} \cup \mathbb{C}_*^{2r}\right) : \\ |a_i| \neq 0, 1, |b_j| \neq 0, 1,\ a_i \neq b_j, i = 1, \ldots, p, j = 1, \ldots, q \end{array} \right\},$$

where $\mathbb{C}_*^2 = \left\{(a, b) \in \mathbb{C}^2 : b = \bar{a}\right\}$ denotes the space of pairs of complex conjugate numbers to guarantee that θ is real.

Model (7) establishes that $f(\lambda) = f(\theta, \kappa_2; \lambda)$ where

$$f(\theta, \kappa_2; \lambda) = \frac{\kappa_2}{2\pi} \phi_2(\theta; \lambda),$$

and we employ κ_j to denote the j-th order marginal cumulant of ε_t, so that κ_2 is its variance, with

$$\phi_2(\theta; \lambda) = \phi(\theta; \lambda)\phi(\theta; -\lambda),$$

where we denote the transfer function of the filter $\left\{\psi_j\right\}_{j=-\infty}^{\infty}$ by

$$\phi(\theta; \lambda) = \sum_{j=-\infty}^{\infty} \psi_j(\theta) \exp(-ij\lambda).$$

For instance, for the ARMA model (8)

$$\phi(\theta; \lambda) = \frac{1 + \sum_{j=1}^{q} \beta_j \exp(-ij\lambda)}{1 - \sum_{j=1}^{p} \alpha_j \exp(-ij\lambda)}. \tag{9}$$

In addition, model (7) establishes that $f_k(\boldsymbol{\lambda}) = f_k(\theta, \kappa_k; \boldsymbol{\lambda})$ where

$$f_k(\theta, \kappa_k; \boldsymbol{\lambda}) = \frac{\kappa_k}{(2\pi)^{k-1}} \phi_k(\theta; \boldsymbol{\lambda}), \tag{10}$$

where

$$\phi_k(\theta; \boldsymbol{\lambda}) = \phi(\theta; \lambda_1) \cdots \phi(\theta; \lambda_{k-1})\phi(\theta; -\lambda_1 - \cdots - \lambda_{k-1}).$$

Then, it is simple to show the following relation that will be used later for understanding the role that higher order terms have in our final OF,

$$|f_k(\theta, \kappa_k; \boldsymbol{\lambda})|^2 = \frac{\nu_k^2}{(2\pi)^{k-2}} f(\theta, \kappa_2; \lambda_1) \cdots f(\theta, \kappa_2; \lambda_{k-1}) f(\theta, \kappa_2; \lambda_1 + \cdots + \lambda_{k-1}), \tag{11}$$

where

$$\nu_k^2 = \frac{\kappa_k^2}{\kappa_2^k}$$

is the square of the standardized cumulant of order k, in particular ν_3 and ν_4 are the skewness and kurtosis coefficients, respectively.

3. IDENTIFICATION OF NONCAUSAL/NONINVERTIBLE ARMA MODELS USING HIGHER ORDER SPECTRA

Although the standard spectral density $f = f_2$, based on second moments, cannot identify the parameters in the noninvertible/noncausal case, identification can be achieved by using the information about these parameters contained in the higher order spectral densities. The next theorem shows that the L_2 distance of the higher order spectral density identifies the correct values of the parameters for an ARMA model, defined in (8). Denote $\Pi = [-\pi, \pi]$ and $d\boldsymbol{\lambda} = d\lambda_1 \cdots d\lambda_{k-1}$.

ASSUMPTION $\mathrm{A}(p, q)$: *The polynomials $\alpha(L)$ and $\beta(L)$ of order p and q respectively have roots $\varphi \in A_{p,q}(\eta)$ for some $\eta > 0$, where*

$$\mathcal{A}_{p,q}(\eta) = \left\{ \begin{array}{c} \boldsymbol{\varphi} = (a_1, \ldots, a_p, b_1, \ldots, b_q)' \in \bigcup_{r=0}^{\lfloor p/2 \rfloor + \lfloor q/2 \rfloor} \left(\mathbb{R}^{p+q-2r} \cup \mathbb{C}_*^{2r}\right) : \\ \min\{|a_i|, |b_j|, ||a_i| - 1|, ||b_j| - 1|, |a_i - b_j|\} \geq \eta > 0, \ i = 1, \ldots, p, j = 1, \ldots, q \end{array} \right\}.$$

Consider a positive weighting function $g(\theta; \boldsymbol{\lambda})$, possibly depending on θ, which is uniformly bounded away from zero and from above for all $\boldsymbol{\lambda} \in \Pi^{k-1}$and $\boldsymbol{\theta} \in \Theta$.

THEOREM 1: *Consider an ARMA(p, q) model (8) with true roots φ_0 satisfying Assumption $A(p, q)$ and $\kappa_k^0 \neq 0$ for some $k \geq 3$. Then for all $\theta \in \Theta \subset \Gamma_{p,q}$, Θ compact, and all $\eta > 0$ there exists an $\epsilon > 0$ such that*

$$\inf_{\boldsymbol{\theta} \in \Theta : \|\theta - \theta_0\| > \eta, \kappa_k \in \mathbb{R}} \int_{\Pi^{k-1}} \left| f_k(\theta, \kappa_k; \boldsymbol{\lambda}) - f_k(\theta_0, \kappa_k^0; \boldsymbol{\lambda}) \right|^2 g(\theta; \boldsymbol{\lambda}) \, d\boldsymbol{\lambda} \geq \epsilon > 0. \tag{12}$$

The weighting function $g(\boldsymbol{\theta};\lambda)$ could be for instance $|\phi_k(\theta; \boldsymbol{\lambda})|^{-2}$ to allow for periodogram optimal normalization, since MA unit roots are excluded in the definition of Θ, as is proposed in Terdik (1999, equation (4.3)). The identification provided in Theorem 1 relies on the next lemma, which will be proved for $k = 3$ only, the extension to $k \geq 4$ is straightforward. Denote by $\mathcal{S}$ the discrete set of all potential roots obtained by inversion of the elements of $\boldsymbol{\varphi}_0$ (whose cardinality is $2^n - 1$ where $n = p + q - r$ and r is the number of pairs of complex roots determined by $\boldsymbol{\varphi}_0$) so that complex roots always appear in conjugate pairs and $\mathcal{S} \subseteq \mathcal{A}_{p,q}$. Then $f_k(\theta(\boldsymbol{\varphi}), \kappa_k; \boldsymbol{\lambda})$ is the spectral density of order k of the ARMA process (8) calculated from (9) and (10) expressed in terms of $\boldsymbol{\varphi}$, the roots of the AR and MA polynomials, rather than in terms of the parameters θ. Note that $\theta(\boldsymbol{\varphi})$ is a continuous function $\boldsymbol{\varphi} \in \mathcal{A}_{p,q}$.

LEMMA 1: *Consider an ARMA(p,q) model (8) with true roots φ_0 satisfying Assumption $A(p,q)$ and $\kappa_k^0 \neq 0$ for some $k \geq 3$. Then*

$$\inf_{\boldsymbol{\varphi} \in \mathcal{S}, \kappa_k \in \mathbb{R}} \int_{\Pi^{k-1}} \left| f_k(\theta(\boldsymbol{\varphi}), \kappa_k; \boldsymbol{\lambda}) - f_k(\theta(\boldsymbol{\varphi}_0), \kappa_k^0; \boldsymbol{\lambda}) \right|^2 g(\theta; \boldsymbol{\lambda})\, d\boldsymbol{\lambda} > 0.$$

Proofs and technical results are contained in the Appendices A and B. Lemma 1 shows that for any $\boldsymbol{\varphi} \neq \boldsymbol{\varphi}_0$ that belongs to the set $\mathcal{S}$, the resulting spectral density is different of $f_k(\theta(\boldsymbol{\varphi}_0), \kappa_k^0; \boldsymbol{\lambda})$ for all $\boldsymbol{\lambda} \in \Pi^{k-1}$ whatever choice of κ_k if $k \geq 3$. The proof relies on the fact that it is not possible to choose κ_k so that both the real and imaginary parts of $f_k(\theta(\boldsymbol{\varphi}), \kappa_k; \boldsymbol{\lambda})$ match simultaneously those of $f_k(\theta(\boldsymbol{\varphi}_0), \kappa_k^0; \boldsymbol{\lambda})$. Note, however, that this is indeed possible for $k = 2$ because $f = f_2$ is real and it is always possible for any $\boldsymbol{\varphi} \in \mathcal{S}$ to find a value κ_2 that satisfies $f_2(\theta(\boldsymbol{\varphi}), \kappa_2; \lambda) = f_2(\theta(\boldsymbol{\varphi}_0), \kappa_2^0; \lambda)$ for all $\lambda \in \Pi$. This result holds also when allowing for unit roots in the MA polynomial, since this fact will not affect the integrability of f_k, noting that in case of real unit roots (i.e. ± 1) the inversion leads to the same solution, so this case should be excluded from $\mathcal{S}$. Then, in order to prove Theorem 1 only remains to consider elements $\boldsymbol{\varphi}$ outside $\mathcal{S}$, which can not replicate the transfer function $\phi(\theta; \lambda)$ under the identification of the orders (p,q) in Assumption A(p,q).

In particular, for $k = 3$ Theorem 1 shows that, when the third order cumulant κ_3 is different from 0, the bispectral density can be used to identify the parameters of an noninvertible linear model. By a similar reasoning, the trispectral density, f_4, can be used for identification when the fourth order cumulant is different from 0.

4. MINIMUM DISTANCE ESTIMATION

4.1 Minimum distance approach

In order to make inference on the parameter vector θ, we propose to employ minimum distance estimators based on the normalized distance between the data, reflected in the higher-order periodogram, $I_k(\boldsymbol{\lambda})$, and the model, reflected in the parameterization of the spectral density of order k, $f_k(\theta, \kappa_k; \boldsymbol{\lambda})$. In particular, using (4) and (5) we employ

$$L_k(\theta, \kappa_k) = \int \frac{E\left| I_k(\boldsymbol{\lambda}) - f_k(\theta, \kappa_k; \boldsymbol{\lambda}) \right|^2}{T^{k-2} f(\lambda_1) \cdots f(\lambda_{k-1}) f(\lambda_1 + \lambda_2 + \cdots + \lambda_{k-1})} d\boldsymbol{\lambda}.$$

Note that in the $k = 2$ case the modulus is not needed since both the periodogram and the spectral density are real. However, for $k > 2$, both $I_k(\boldsymbol{\lambda})$ and $f_k(\theta, \kappa_k; \boldsymbol{\lambda})$ are complex. Note also that using

(11), $L_k(\theta, \kappa_k)$ can be written as

$$L_k(\theta, \kappa_k) = \frac{\nu_k^2}{(2\pi)^{k-2}} \int \frac{E\left|I_k(\boldsymbol{\lambda}) - f_k(\theta, \kappa_k; \boldsymbol{\lambda})\right|^2}{T^{k-2}\left|f_k(\boldsymbol{\lambda})\right|^2} d\boldsymbol{\lambda},$$

this expression is relevant to show the role played by the higher order cumulants when adding up the sample analogues of $L_k(\theta, \kappa_k)$ in Subsection 4.3.

Then, the L_2 distance estimator of θ is based on minimizing the empirical analogue of $L_k(\theta, \kappa_k)$, scaling each periodogram ordinate by its variance, which can also be interpreted in terms of data standardization to make comparable L_k for different k. Note that the scaling is well defined when unit roots and long memory are excluded in the parameter space, so that $0 < f(\lambda) < \infty$ for all λ. Since the denominator depends on the true unknown spectral density, we propose to employ the traditional Whittle estimators under invertibility and causality identification as preliminary estimators of θ and κ_2, denoted by $\overline{\theta}_T$ and $\overline{\kappa}_{2T}$ so that $f(\overline{\theta}_T, \overline{\kappa}_{2T}; \lambda_{j_1})$ is invariant if the roots implied by $\overline{\theta}$ are inverted and $\overline{\kappa}_2$ is adjusted accordingly. Then, the objective function based on $f_k(\theta, \kappa_k; \boldsymbol{\lambda})$ is

$$L_{kT}(\theta, \kappa_k) = \frac{1}{T} \sum_{\mathbf{j}=1}^{T-1} \frac{\left|I_k(\boldsymbol{\lambda}_\mathbf{j}) - f_k(\theta, \kappa_k; \boldsymbol{\lambda}_\mathbf{j})\right|^2}{T^{k-2} f(\overline{\theta}_T, \overline{\kappa}_{2T}; \lambda_{j_1}) \cdots f(\overline{\theta}_T, \overline{\kappa}_{2T}; \lambda_{j_{k-1}}) f(\overline{\theta}_T, \overline{\kappa}_{2T}; \lambda_{j_1} + \lambda_{j_2} + \cdots + \lambda_{j_{k-1}})}, \tag{13}$$

where we have simplified the notation by introducing the general k-dimensional vector of Fourier frequencies $\boldsymbol{\lambda}_\mathbf{j} = (\lambda_{j_1}, \ldots, \lambda_{j_{k-1}})$ and by writing $\sum_{\mathbf{j}=1}^{T-1} = \sum_{j_1=1}^{T-1} \cdots \sum_{j_{k-1}=1}^{T-1}$ where we discard all combinations of frequencies such that $\lambda_{j_a} + \lambda_{j_b} = T$, $a \neq b$. This avoids referring in the notation to specific sets of $\boldsymbol{\lambda}_\mathbf{j} \in \Pi^{k-1}$ where the periodogram can be defined uniquely, see the discussion in Aleeksev (1993) and Appendix B. The normalization by preliminary estimates of f does not affect asymptotic properties of parameter estimates and greatly simplifies the analysis when compared to the case where scaling is simultaneously estimated, cf. Section 4 in Terdik (1999).

4.2 Concentrating the Cumulants out of the Objective Function

Note that in the minimization of the previous objective function (13), the cumulant κ_k is a nuisance parameter that only appears in L_{kT}, hence, although eventually the final objective function is a weighted average of the L_{kT}'s, we can focus on L_{kT} to concentrate out κ_k. Also note that $\overline{\kappa}_{2T}$ is only a scaling factor in $L_{kT}(\theta, \kappa_k)$, but it is needed to make comparisons among different $k's$. First, recalling (11), then (13) can be written as

$$L_{kT}(\theta, \kappa_k) = \frac{(2\pi)^k}{\overline{\kappa}_{2T}^k} \frac{1}{T} \sum_{\mathbf{j}=1}^{T-1} \frac{\left|I_k(\boldsymbol{\lambda}_j) - f_k(\theta, \kappa_k; \boldsymbol{\lambda}_\mathbf{j})\right|^2}{T^{k-2}\left|\phi_k(\overline{\theta}_T; \boldsymbol{\lambda}_\mathbf{j})\right|^2}.$$

Considering the FOC of minimizing $L_{kT}(\theta, \kappa_k)$ respect to κ_k, after straightforward algebra, it is simple to concentrate out κ_k, and define

$$\kappa_{kT}(\theta) = (2\pi)^{k-1} \left(\sum_{\mathbf{j}=1}^{T-1} \frac{|\phi_k(\theta; \boldsymbol{\lambda}_\mathbf{j})|^2}{|\phi_k(\bar{\theta}_T; \boldsymbol{\lambda}_\mathbf{j})|^2} \right)^{-1} \sum_{\mathbf{j}=1}^{T-1} \frac{\mathrm{Re}\,(I_k(\boldsymbol{\lambda}_\mathbf{j})\phi_k(\theta; -\boldsymbol{\lambda}_\mathbf{j}))}{|\phi_k(\bar{\theta}_T; \boldsymbol{\lambda}_\mathbf{j})|^2}.$$

The fact that $|\phi_k(\bar{\theta}_T; \boldsymbol{\lambda}_\mathbf{j})|^2$ is a consistent estimator for $|\phi_k(\theta; \boldsymbol{\lambda}_\mathbf{j})|^2$ up to scale, even if obtained from Whittle estimation under (possibly wrong) invertibility and causality assumptions, motivates the simpler estimate

$$\kappa_{kT}^{\dagger}(\theta) = \left(\frac{2\pi}{T}\right)^{k-1} \sum_{\mathbf{j}=1}^{T-1} \mathrm{Re}\left(\frac{I_k(\boldsymbol{\lambda}_\mathbf{j})}{\phi_k(\theta; \boldsymbol{\lambda}_\mathbf{j})}\right). \tag{14}$$

So, a consistent estimator for κ_k is obtained by plugging in a consistent estimator of θ into (14) for $k = 2, 3, 4$, and the concentrated objective function is defined as

$$\widetilde{L}_{kT}(\theta) = L_{kT}(\theta, \kappa_{kT}^{\dagger}(\theta)).$$

4.3 Weighted Objective Function

In principle, frequency domain minimum distance estimators would combine the information contained in f_2, f_3 and f_4, but higher order f_k could be considered when $\kappa_3 = \kappa_4 = 0$. Hence, general minimum distance estimators are based on minimizing the weighted sum

$$\tilde{w}_2 \widetilde{L}_{2T}(\theta) + \tilde{w}_3 \widetilde{L}_{3T}(\theta) + \tilde{w}_4 \widetilde{L}_{4T}(\theta)$$

where $(\tilde{w}_2, \tilde{w}_3, \tilde{w}_4)$ are some positive weights that can give more emphasis to information from a particular moment or cumulant. In Subsection 4.5 we address the issue of optimally selecting the weights. Note that the individual objective functions lead to first order conditions that are correlated, despite the individual (scaled) periodograms are uncorrelated, leading to contributions in the asymptotic variance of the estimates using simultaneously more than one $\widetilde{L}_{kT}(\theta)$.

An additional insight for this objective function can be gained by calling $\widetilde{v}_{kT}^2 = \widetilde{\kappa}_{kT}^2/\bar{\kappa}_2^k$ where $\widetilde{\kappa}_{kT} = \kappa_{kT}^{\dagger}(\widetilde{\theta}_T)$ and $\widetilde{\theta}_T$ is a preliminary consistent estimate of θ, and defining

$$\widetilde{L}_{kT}^0(\theta) = \frac{(2\pi)^{k-2}}{\widetilde{v}_{kT}^2} \widetilde{L}_{kT}(\theta) = \frac{1}{T^{k-1}} \sum_{\mathbf{j}=1}^{T-1} \frac{\left|I_k(\boldsymbol{\lambda}_\mathbf{j}) - f_k^{\dagger}(\theta; \boldsymbol{\lambda}_\mathbf{j})\right|^2}{\left|f_k(\bar{\theta}, \widetilde{\kappa}_{kT}; \boldsymbol{\lambda}_\mathbf{j})\right|^2},$$

so that

$$\widetilde{L}_{kT}(\theta) = (2\pi)^{2-k}\, \widetilde{v}_{kT}^2 \widetilde{L}_{kT}^0(\theta), \tag{15}$$

then we define the general family of minimum distance frequency domain estimators

$$\theta_T = \arg\min_{\theta \in \Theta} \widetilde{L}_{wT}(\theta),$$

where we rewrite the loss function as

$$\widetilde{L}_{wT}(\theta) = w_2 \widetilde{L}_{2T}^0(\theta) + w_3 \widetilde{v}_{3T}^2 \widetilde{L}_{3T}^0(\theta) + w_4 \widetilde{v}_{4T}^2 \widetilde{L}_{4T}^0(\theta), \tag{16}$$

noting that $\widetilde{v}_{kT}^2 \widetilde{L}_{kT}^0(\theta)$ does not depend on $\widetilde{\kappa}_{kT}^2$, while the factors $(2\pi)^{k-2}$ are absorbed by the constants w_3 and w_4. This version of the objective function makes evident that the role played by the higher order cumulant spectral densities is larger the larger are the skewness and the kurtosis for fixed (w_1, w_2, w_3), because the second derivative of $\widetilde{L}_{kT}^0(\theta)$ evaluated at θ_0 converges to the same scale free constant matrix for all $k = 2, 3, \ldots$ up to a scalar factor that depends only on k. Further, the variance of the score of $\widetilde{L}_{kT}^0(\theta_0)$ varies inversely with ν_k^2 for $k > 2$.

4.4 Asymptotic Theory

Given the linear nature of the model, the dependence condition we employ is just based on restricting the moments of ε_t together with an i.i.d. assumption, whereas conditions on the summability of the $\psi' s$ are directly implied by the ARMA parametrization since unit roots are excluded.

We introduce Assumption 1 on the parameter space for identification and Assumption 2 on the innovations.

ASSUMPTION 1: *Y_t follows an ARMA(p, q) model (8) with true roots φ_0 satisfying Assumption $A(p, q)$ with $\Theta \subset \Gamma_{p,q}$ compact.*

ASSUMPTION 2: *The process ε_t is an i.i.d. sequence with zero mean, variance $\kappa_2 > 0$, $\kappa_k^0 \neq 0$ for $k = 3, 4$, and $E\left[\varepsilon_t^8\right] < \infty$.*

Under these conditions the results (4) and (5) hold because the ARMA model implies smooth spectral densities. The next two theorems establish the asymptotic properties of the estimator of θ that minimizes (16).

THEOREM 2: *Under Assumptions 1 and 2, as $T \to \infty$*

$$\theta_T \to_p \theta_0,$$

and for $k = 2, 3, 4$,

$$\kappa_{kT}^{\dagger}(\theta_T) \to_p \kappa_k.$$

In addition to consistency, the next theorem establishes the asymptotic normality. Define

$$\Phi_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi^1(\theta_0; \lambda) \varphi^1(\theta_0; -\lambda)' d\lambda \tag{17}$$

with $\varphi^1(\theta_0; \lambda) = \varphi(\theta_0; \lambda) - \mu(\theta_0)$,

$$\varphi(\theta_0; \lambda) = \frac{\dot{\phi}\,(\theta_0; \lambda)}{\phi\,(\theta_0; \lambda)}, \quad \mu(\theta_0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(\theta_0; \lambda) d\lambda, \tag{18}$$

and with

$$\dot{\phi}(\theta_0;\lambda) = \frac{\partial}{\partial\theta}\phi(\theta_0;\lambda).$$

Define also

$$\Phi_0^* = \frac{1}{2\pi}\int_{-\pi}^{\pi}\varphi^1(\theta_0;\lambda)\varphi^1(\theta_0;\lambda)'d\lambda,$$

and note that Φ_0^* is zero for pure invertible and causal processes because in this case $\varphi^1 = \varphi$ has no constant term in its Fourier expansion. Note that both Φ_0 and Φ_0^* are real because imaginary parts of the integrands are odd and cancel out in the integration. Then, define the symmetric matrix

$$V = \begin{bmatrix} \Phi_0 + \Phi_0^* & \frac{3}{2}\left(\Phi_0 + \Phi_0^*\right) & 2\frac{\bar{\mu}_4}{\nu_4}\left(\Phi_0 + \Phi_0^*\right) \\ & \frac{9}{4}\left\{\frac{\bar{\mu}_4 - 1}{\nu_3^2}\Phi_0 + \Phi_0^*\right\} & 3\left\{\frac{\bar{\mu}_5 - \nu_3}{\nu_3\nu_4}\Phi_0 + \frac{\bar{\mu}_4}{\nu_4}\Phi_0^*\right\} \\ & & 4\left\{\frac{\bar{\mu}_6 - \nu_3^2}{\nu_4^2}\Phi_0 + \left(\frac{\bar{\mu}_4}{\nu_4}\right)^2\Phi_0^*\right\} \end{bmatrix},$$

where $\bar{\mu}_4 = \mu_4/\kappa_2^2 = \nu_4 + 3$, $\bar{\mu}_5 = \mu_5/\kappa_2^{5/2} = \nu_5 + 10\nu_3$ and $\bar{\mu}_6 = \mu_6/\kappa_2^3 = \nu_6 + 10\nu_3^2 + 15\nu_4 + 15 > 0$ are the standardized fourth, fifth and sixth moments, respectively.

Theorem 3: *Under Assumptions 1 and 2, as $T \to \infty$,*

$$\sqrt{T}(\theta_T - \theta) \to_d N(0, \Sigma_0^{-1}\Omega_0\Sigma_0^{-1})$$

where

$$\Sigma_0 = \left(w_2 + \frac{3}{2}w_3\nu_3^2 + 2w_4\nu_4^2\right)\Phi_0 + w_2\Phi_0^*$$

is assumed positive definite and

$$\Omega_0 = \sum_{j=2}^{4}\sum_{k=2}^{4} w_j w_k \nu_j^2 \nu_k^2 V_{j-1,k-1}.$$

The expressions for Σ_0 and Ω_0 illustrate that the larger, relatively, are $w_3\nu_3^2$ and $w_4\nu_4^2$, the more important is the contribution from these particular moments in the estimation method, while the dependence of the asymptotic variance of θ_T on higher order moments is reflected in the elements of V.

From this result we can obtain the properties of many specific estimates using a single spectral density setting $w_j = 0, 1$, as we do in the next Corollary. Define $\theta_T^{(k)}$ as the L_2-distance estimate using only spectral densities of order $k = 2, 3, 4$, i.e., $w_k = 1,\ w_j = 0,\ j \neq k$.

Corollary 1: *Under the conditions of Theorem 3, assuming that for $\hat{\theta}_T^{(2)}$ Θ correctly identifies the location of the roots of θ_0, as $T \to \infty$,*

$$\sqrt{T}\begin{pmatrix} \theta_T^{(2)} - \theta_0 \\ \theta_T^{(3)} - \theta_0 \\ \theta_T^{(4)} - \theta_0 \end{pmatrix} \to_d N\left(0, W\right)$$

where

$$W = \begin{bmatrix} (\Phi_0 + \Phi_0^*)^{-1} & \Phi_0^{-1} & \frac{\bar{\mu}_4}{\nu_4}\Phi_0^{-1} \\ & \frac{\bar{\mu}_4 - 1}{\nu_3^2}\Phi_0^{-1} + \Phi_0^{-1}\Phi_0^*\Phi_0^{-1} & \frac{\bar{\mu}_5 - \nu_3}{\nu_3\nu_4}\Phi_0^{-1} + \frac{\bar{\mu}_4}{\nu_4}\Phi_0^{-1}\Phi_0^*\Phi_0^{-1} \\ & & \frac{\bar{\mu}_6 - \nu_3^2}{\nu_4^2}\Phi_0^{-1} + \left(\frac{\bar{\mu}_4}{\nu_4}\right)^2 \Phi_0^{-1}\Phi_0^*\Phi_0^{-1} \end{bmatrix}.$$

Remark 1: Corollary 1 shows that the asymptotic variances of $\theta_T^{(3)}$ and $\theta_T^{(4)}$ are smaller the larger are ν_3 and ν_4, respectively, i.e. the more information is contained in these particular higher order moments.

Remark 2: For pure invertible and causal processes, W simplifies to

$$W = \begin{bmatrix} 1 & 1 & \frac{\bar{\mu}_4}{\nu_4} \\ & \frac{\bar{\mu}_4 - 1}{\nu_3^2} & \frac{\bar{\mu}_5 - \nu_3}{\nu_3\nu_4} \\ & & \frac{\bar{\mu}_6 - \nu_3^2}{\nu_4^2} \end{bmatrix} \otimes \Phi_0^{-1}.$$

Then the asymptotic variance of $\hat{\theta}_T^{(2)}$ equals the usual Whittle result for invertible and causal models, using that in this case $\Phi_0^* = 0$, so

$$(\Phi_0 + \Phi_0^*)^{-1} = 4\pi \left(\int_{-\pi}^{\pi} \varphi_2(\theta_0; \lambda)\varphi_2(\theta_0; -\lambda)' d\lambda \right)^{-1},$$

since $\varphi^1 = \varphi$ and $\int_\Pi \varphi_2(\theta_0; \lambda)\varphi_2(\theta_0; -\lambda)' d\lambda = 2\int_\Pi \varphi(\theta_0; \lambda)\varphi(\theta_0; -\lambda)' d\lambda$. In Appendix C we show that the estimates are also asymptotically equivalent in this case. We also note that the asymptotic variance of $\hat{\theta}_T^{(3)}$ is positive definite (given that Φ_0 is positive definite) since $\bar{\mu}_4 - 1 = \bar{\mu}_4 - \bar{\mu}_2^2 = 2 + \nu_4 \geq 0$, so that $\nu_4 \geq -2$ while $\nu_3^2 > 0$ by assumption. Similarly $\bar{\mu}_6 - \nu_3^2 = \bar{\mu}_6 - \bar{\mu}_3^2 \geq 0$ and the variance of $\hat{\theta}_T^{(4)}$ is also positive semidefinite.

Remark 3: Note that for invertible processes μ_0, defined in (18), equals 0, so that Φ_0 simplifies to

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(\theta_0; \lambda)\varphi(\theta_0; -\lambda)' d\lambda,$$

but for general noninvertible cases the expression for μ_0 has to be considered. For instance, for a simple noninvertible MA(1) process, $|\theta_0| > 1$, from Cauchy formula, it is simple to show that

$$\mu_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i\lambda}}{1 + \theta_0 e^{i\lambda}} d\lambda = \frac{1}{2\pi i} \oint \frac{z}{1 + \theta_0 z} dz = \frac{1}{\theta_0},$$

and also that in this case with $q = 1$

$$\Phi_0^* = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{e^{i\lambda}}{1 + \theta_0 e^{i\lambda}} \right)^2 d\lambda - \frac{1}{\theta_0^2} = 0,$$

but this will not be generally true when $q > 1$.

Remark 4: Further, it can be showed that for this MA(1) process the asymptotic variance of the (unrestricted) parameter estimate using L_{2T}^0 is the same for non-invertible parameter values as the one obtained using the restricted estimate under invertibility and then inverting the roots appropriately. To see this, the "invertible" asymptotic variance of the Whittle and of the $\theta_0^{(2)}$ estimates, $|\theta_0| < 1$, is given by

$$\Phi_0^{-1} = \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\left|1+\theta_0 e^{i\lambda}\right|^2} d\lambda \right)^{-1} = 1 - \theta_0^2,$$

while the "noninvertible" asymptotic variance for $\theta_0^* = 1/\theta_0$ provided by the L_{2T}^0 loss function is

$$\Omega_0 \left(\theta_0^*\right)^{-1} = \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\left|1+\theta_0^* e^{i\lambda}\right|^2} d\lambda - \mu_0^2 \right)^{-1} = \left(\theta_0^{*2} - 1\right) \theta_0^{*2},$$

since $\Phi_0^* = 0$ in this case (cf. Remark 3). Using the delta method, the AVar of the estimate of $\theta_0^* = \theta_0^{-1}$ from estimation of θ_0 in an invertible model is equal to that of $\hat{\theta}_0^{(2)}$ times $\theta_0^{-4} = \theta_0^{*4} > 1$, that is, $\left(1-\theta_0^2\right) * \theta_0^{*4} = \left(1-\theta_0^{-*2}\right) * \theta_0^{*4} = \left(\theta_0^{*2}-1\right)\theta_0^{*2} = \Omega_0\left(\theta_0^*\right)^{-1} > 0$ as if he had estimated θ_0^* directly. The same results holds for estimates using higher order information since the AVar of these estimates is always proportional to Φ_0^{-1}.

Remark 5: For all-pass models with constant spectral density we have that $\Phi_0 + \Phi_0^*$ is singular, indicating that second moments cannot identify the parameters in absence of further information. Consider the case of unrestricted estimation of an ARMA$(1,1)$ with (invertible) MA parameter, $\beta_1 = \theta_0$, $|\theta_0| < 1$ and with (noncausal) AR parameter $\alpha_1 = -\theta_0^{-1}$ so the roots of both polynomials are $-\theta_0^{-1}$ and $-\theta_0$ respectively and

$$f_2\left(\beta_1, \alpha_1; \lambda\right) = \frac{\kappa_2}{2\pi} \frac{\left|1+\beta_1 e^{i\lambda}\right|^2}{\left|1-\alpha_1 e^{i\lambda}\right|^2} = \frac{\kappa_2}{2\pi} \frac{\left|1+\theta_0 e^{i\lambda}\right|^2}{\left|1+\theta_0^{-1} e^{i\lambda}\right|^2} = \frac{\kappa_2}{2\pi}\theta_0^2.$$

Then

$$\Phi_0 = \frac{1}{1-\theta_0^2} \begin{pmatrix} 1 & 0 \\ 0 & \theta_0^4 \end{pmatrix}, \quad \Phi_0^* = \frac{1}{1-\theta_0^2} \begin{pmatrix} 0 & -\theta_0^2 \\ -\theta_0^2 & 0 \end{pmatrix}$$

since

$$\varphi^1(\beta_1, \alpha_1; \lambda) = \begin{pmatrix} \frac{e^{i\lambda}}{1+\theta_0 e^{i\lambda}} \\ -\frac{e^{i\lambda}}{1+\theta_0^{-1} e^{i\lambda}} + \theta_0 \end{pmatrix} = \begin{pmatrix} \frac{e^{i\lambda}}{1+\theta_0 e^{i\lambda}} \\ \frac{\theta_0^2 e^{-i\lambda}}{1+\theta_0 e^{-i\lambda}} \end{pmatrix}.$$

Remark 6: In the case of restricted estimation of an all pass model of order r with

$$\phi(\theta; \lambda) = \frac{e^{i\lambda r}\theta\left(-\lambda\right)}{-\theta_r \theta\left(\lambda\right)},$$

and $\theta = (\theta_1, \ldots, \theta_r)'$, $\theta(\lambda) = 1 - \theta_1 e^{i\lambda} \cdots - \theta_r e^{ir\lambda}$, we can obtain that

$$\varphi_j^1(\beta_1, \alpha_1; \lambda) = -\frac{e^{-i\lambda j}}{\theta(-\lambda)} + \frac{e^{i\lambda j}}{\theta(\lambda)}, \quad j = 1, \ldots, r$$

so that

$$\Phi_{0,j,k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{2\cos(j-k)\lambda}{|\theta(\lambda)|^2} d\lambda$$

(which is twice the value for typical AR or MA coefficients) and $\Phi_0^* = -\Phi_0$, so $\Phi_0 + \Phi_0^* = 0$. Hence, second moments do not provide information at all as f_2 is constant for all θ, while the asymptotic variances of $\hat{\theta}_0^{(3)}$ and $\hat{\theta}_0^{(4)}$ are respectively

$$\frac{\bar{\mu}_4 - 1 - \nu_3^2}{\nu_3^2} \Phi_0^{-1} \text{ and } \frac{\bar{\mu}_6 - \nu_3^2 - \bar{\mu}_4^2}{\nu_4^2} \Phi_0^{-1}.$$

Remark 7: Theorem 3 can be compared with previous CLTs in the literature. In particular, classical textbooks from Hannan (1970) or Brillinger (1975) to Brockwell and Davis (1991) provide CLT for the QMLE for causal and invertible ARMA models. Their asyvar is a particular case of our asyvar. Terdik (1999, Theorem 76) considers a loss function similar to $w_2 \widetilde{L}_{2T}^0(\theta) + w_3 \widetilde{v}_{3T}^2 \widetilde{L}_{3T}^0(\theta)$, but with simultaneous normalization of the periodograms, although he does not analyze the contribution of this modification to the asymptotic variance. Kumon provides a CLT for Z-estimators based on higher order cumulants, which are asymptotically equivalent to the corresponding MD estimators. Notice that Kumon's formula (3.9), which provides the asymptotic variance of $\theta_T^{(k)}$ cannot be compared to ours for two reasons: he is not centering by $\mu(\theta_0)$ in expression (17), which is needed for noninvertible processes, and the constants in his expression (3.9) are not adjusted by the fact that one uses only the real part of the biperiodogram in the first order conditions. Similarly, the results of Ahn, Leonenko and Sakhno (2007) are not directly comparable since their loss function depends on a weighting function a a different identification strategy.

Remark 8: For establishing the CLT the iid restriction could be relaxed to mds with constant conditional homoskedasticity (and additional moment restrictions) at cost of further notational complexity. The extension to the conditional heteroskedasticity case is more challenging and left open for future research.

4.5 Optimal Selection of Weights

Given Theorem 3 and Corollary 1 it is possible to derive the optimal weights (w_2, w_3, w_4) in the sense of minimizing the asymptotic variance of θ_T for a given value of higher order cumulants $(\kappa_3, \kappa_4, \kappa_5, \kappa_6)$ when the model is invertible and causal, $\Phi_0^* = 0$, since in this case the problem reduces to a univariate one. This analysis is not reported since the expressions for the optimal weights are complicated and unintuitive, apart from the fact that large ν_3^2 and ν_4^2 improve the efficiency of estimates based on third and fourth order cumulants. If $\Phi_0^* \neq 0$, then the analysis cannot be reduced to a univariate optimization and it does not seem possible to give general results.

Here, instead, we will provide three lemmas with particular cases of interest for which relatively simple and intuitive solutions are available for estimates based on only two moments for invertible and causal models. Lemma 2 considers the case of the optimal selection of w_2 and w_3 when w_4 is restricted to be zero. This result is of interest since it shows that adding $\widetilde{L}_{3T}^0(\theta)$ to $\widetilde{L}_{2T}^0(\theta)$ can not make any improvement from an efficiency point of view. Lemma 3 addresses the case where w_3 is restricted to 0, for instance, due to anticipation of symmetric innovations**.** This case is of interest because when $\nu_4 < 0$, employing the optimal w_4 delivers an estimator whose asymptotic variance is lower than the one for the Whittle estimator. The gain in efficiency depends on the specific values for the cumulants. For instance, for the uniform distribution the decrease in the asymptotic variance is around 20%. Lemma 4 considers the case where w_2 is restricted to be 0. This case is of interest since, in terms of identification, $\widetilde{L}_{2T}^0(\theta)$ does not help to $\widetilde{L}_{wT}(\theta)$ since second order moments do not globally identify the parameters., Hence, Lemma 4 provides the optimal weights from an identification point of view (that is, imposing $w_2 = 0$). Proofs are in Appendix D.

LEMMA 2: Restricting to the weights $(w_2, w_3, w_4) = (1, w_3, 0)$, the optimal weights are $w_3^* = 0$.

Note that these weights would preclude identification.

LEMMA 3: Restricting to the weights $(w_2, w_3, w_4) = (1, 0, w_4)$, the optimal weights are

$$w_4^* = \max\left\{\frac{3}{2}\frac{1}{\nu_4^2\bar{\mu}_4 - \nu_4\left(\bar{\mu}_6 - \nu_3^2\right)}, 0\right\}.$$

It is immediate to see that w_4^* is positive when $\nu_4 < 0$ because $\bar{\mu}_4 > 0$ and $\bar{\mu}_6 - \nu_3^2 > 0$, which implies that fourth order cumulants are informative if the distribution tails of the innovations are lighter than the Gaussian ones. In this case the asymptotic variance of θ_T is lower than that of the Whittle estimator. The gain in efficiency varies according to the distribution of e_t, as commented above.

LEMMA 4: Restricting to the weights $(w_2, w_3, w_4) = (0, 1, w_4)$, the optimal weights are

$$w_4^\dagger = \max\left\{\frac{3}{4}\frac{\nu_3\nu_4\left(\bar{\mu}_4 - 1\right) - \nu_3^2\left(\bar{\mu}_5 - \nu_3\right)}{\nu_3\nu_4\left(\bar{\mu}_6 - \nu_3^2\right) - \nu_4^2\left(\bar{\mu}_5 - \nu_3\right)}, 0\right\}.$$

In principle $w_4^\dagger$ can be positive when either $\nu_4 < 0$ or $\nu_4 > 0$, depending also on the values of ν_3, ν_5 and ν_6. For instance, if ν_3 and ν_5 have the same sign, then $w_4^\dagger > 0$ when $\nu_4 < 0$, as shown in the proof of Lemma 6, confirming that fourth cumulants are also useful for identification in this situation.

Given that estimating the optimal weights is a complicated task, it is a sensible approach to start with a consistent estimate, and then perform a Newton-Raphson step on $\widetilde{L}_{2T}^0(\theta)$, so that the resulting estimator is consistent and its asymptotic variance is asymptotically equivalent to the one for the Whittle estimator. The initial consistent estimator could be constructed by minimizing just $\widetilde{L}^0_{3T}(\theta)$, or adding $\widetilde{L}^0_{4T}(\theta)$ in case a negative kurtosis is suspected or found after this initial estimation. If symmetry of innovations is suspected, the initial consistent estimator could just be based on minimizing $\widetilde{L}^0_{4T}(\theta)$.

5. SIMULATIONS

In this section we report a short simulation exercise to assess the ability of $\tilde{L}_{3T}$ and $\tilde{L}_{4T}$ to identify the proper location of the roots, complementing the theoretical analysis of Lemma 6. We focus on a simple $MA(1)$ process, and consider three innovation distributions: an exponential, and a Student's t with 5 degrees of freedom and a Uniform. The last two distributions are symmetric, so $\kappa_3 = 0$ and $\tilde{L}_{3T}$ should not provide identification information, but the exponential distribution is highly asymmetric, $v_3 = 2$. The first two distributions have the same positive kurtosis $\nu_4 = 6$, while the uniform has negative kurtosis, $\nu_4 = -6/5$ so we will expect $\tilde{L}_{4T}$ to be much more informative for the latter.

In sum, we would expect that $\tilde{L}_{3T}$ (over $\tilde{L}_{3T} + \tilde{L}_{4T}$) should be the best (in relative terms) criterion to identify the location of the roots for exponential distributions because kurtosis is positive, while $\tilde{L}_{4T}$ should be the best option for t and uniform distributions, but the identification of the t should be much more difficult than that of the uniform, everything else the same.

Tables 1-3 reports the percentage of proper identification for 3 sample sizes, 50, 100, and 200, respectively, and four values for θ (0.5, 0.9 and their inverse values) and 1000 replications.

The previous predictions for the exponential distribution are confirmed, $\tilde{L}_{3T}$ is the most informative in most cases, $\tilde{L}_{3T} + \tilde{L}_{4T}$ is getting closer the larger T, while $\tilde{L}_{4T}$ is always doing worse. For the t distribution we find that $\tilde{L}_{3T}$, provides some marginal information (since κ_3 be would expect just 50% of right identifications) we should not be able to identify models with symmetric innovations, and $\tilde{L}_{4T}$ performs similarly to $\tilde{L}_{3T} + \tilde{L}_{4T}$. For the uniform, $\tilde{L}_{3T}$ is very close to 50% as expected, while $\tilde{L}_{4T}$ improves over $\tilde{L}_{3T} + \tilde{L}_{4T}$ as T gets larger. Overall, we can conclude that $\tilde{L}_{3T} + \tilde{L}_{4T}$ provides reasonable identification results, despite one of the elements $\tilde{L}_{kT}$ might not be well defined or not improve on the information of the other element.

In terms of sample size, we can check that it always helps identification expect for $\tilde{L}_{3T}$ for symmetric distributions, while the distance to the unit circle helps to identify the right location of the roots, everything else the same.

Table 1. Identification with $\tilde{L}_{3T} + \tilde{L}_{4T}$

		θ			
T	Distribution	**0.5**	**0.9**	$\mathbf{0.9}^{-1}$	$\mathbf{0.5}^{-1}$
50	$\varepsilon_t \sim$ Exp(1)	86.6	69.2	77.2	88.2
	$\varepsilon_t \sim t_5$	59.9	47.8	49.9	59.6
	$\varepsilon_t \sim$ Unif	63.6	54.4	61.2	70.2
100	$\varepsilon_t \sim$ Exp(1)	94.7	75.1	87.6	94.7
	$\varepsilon_t \sim t_5$	71.9	56.4	63.8	75.7
	$\varepsilon_t \sim$ Unif	81.0	55.8	63.9	79.7
200	$\varepsilon_t \sim$ Exp(1)	98.5	76.2	93.6	99.2
	$\varepsilon_t \sim t_5$	85.1	59.5	68.0	84.2
	$\varepsilon_t \sim$ Unif	96.4	53.8	66.3	95.7

Table 2. Identification with $\tilde{L}_{3T}$

		θ			
T	Distribution	**0.5**	**0.9**	$\mathbf{0.9}^{-1}$	$\mathbf{0.5}^{-1}$
50	$\varepsilon_t \sim$ Exp(1)	95.4	75.4	81.8	96.1
	$\varepsilon_t \sim t_5$	64.1	57.3	57.8	65.9
	$\varepsilon_t \sim$ Unif	46.4	47.4	45.2	48.0
100	$\varepsilon_t \sim$ Exp(1)	98.8	78.2	91.0	99.1
	$\varepsilon_t \sim t_5$	67.3	55.3	61.0	68.7
	$\varepsilon_t \sim$ Unif	46.0	44.3	51.5	46.9
200	$\varepsilon_t \sim$ Exp(1)	99.9	73.7	95.6	100.0
	$\varepsilon_t \sim t_5$	70.5	56.8	62.4	70.4
	$\varepsilon_t \sim$ Unif	46.3	42.3	54.8	44.5

Table 3. Identification with $\tilde{L}_{4T}$

		θ			
T	Distribution	**0.5**	**0.9**	$\mathbf{0.9}^{-1}$	$\mathbf{0.5}^{-1}$
50	$\varepsilon_t \sim$ Exp(1)	77.5	68.4	77.0	80.7
	$\varepsilon_t \sim t_5$	67.0	56.9	60.2	65.7
	$\varepsilon_t \sim$ Unif	61.4	47.7	49.7	60.0
100	$\varepsilon_t \sim$ Exp(1)	89.2	74.6	87.2	91.0
	$\varepsilon_t \sim t_5$	71.6	56.0	63.3	75.2
	$\varepsilon_t \sim$ Unif	81.6	59.2	61.0	81.3
200	$\varepsilon_t \sim$ Exp(1)	95.6	76.7	93.5	97.0
	$\varepsilon_t \sim t_5$	85.2	59.6	68.1	83.3
	$\varepsilon_t \sim$ Unif	96.6	54.2	65.8	96.1

6. CONCLUSIONS AND EXTENSIONS

This article introduces frequency domain minimum distance procedures for performing inference in general time series linear models that may be noncausal and noninvertible. We propose a minimum distance approach that combines the information contained in second, third, and fourth moments. Contrary to existing estimators, the proposed estimator is consistent under general assumptions, and can be more efficient that the Whittle estimator by a careful selection of the weights given to the higher order contributions.

This article has focused in ARMA models with independent innovations, but the analysis can be extended to other linear models, such as Bloomfield, and, especially to nonlinear models since information contained in higher order spectra is particulary relevant for this case, an example would be ARMA-GARCH models, so we could compare our methods with Ahn et al. (2007). Additional extensions of interest are the following: use and justification of the bootstrap to estimate the standard errors of the estimates, use and justification of the two-step estimator commented at the end of Subsection 4.5, use and justification of automatic criteria, such as AIC or BIC, to select the order of the ARMA model.

APPENDICES

APPENDIX A: MAIN PROOFS

Proof of Theorem 1. First, by Lemma 1 and with $\theta_0 = \theta(\boldsymbol{\varphi}_0)$,

$$\inf_{\theta=\theta(\boldsymbol{\varphi}),\boldsymbol{\varphi}\in\mathcal{S},\kappa_k\in\mathbb{R}} \int_{\Pi^{k-1}} \left| f_k(\theta(\boldsymbol{\varphi}), \kappa_k; \boldsymbol{\lambda}) - f_k(\theta_0, \kappa_k^0; \boldsymbol{\lambda}) \right|^2 g(\theta; \boldsymbol{\lambda}) d\boldsymbol{\lambda} \geq \epsilon_1 > 0,$$

for some $\epsilon_1 > 0$, so that under Assumption A, by continuity of the mapping $\theta : \bigcup_{r=0}^{\lfloor p/2\rfloor+\lfloor q/2\rfloor} \mathbb{R}^{p+q-2r} \cup \mathbb{C}_*^{2r} \to \mathbb{R}^{p+q}$ and f_k and for a ball of radius ρ small enough around $\theta = \theta(\boldsymbol{\varphi})$ for $\boldsymbol{\varphi} \in\mathcal{S}$, $B(\rho, \theta) = \left\{\theta^\dagger : \left\|\theta^\dagger - \theta\right\| \leq \rho\right\}$ and any $\eta > 0$,

$$\inf_{\theta\in B(\rho,\theta(\boldsymbol{\varphi}))\cap B(\eta,\theta_0)^c,\boldsymbol{\varphi}\in\mathcal{S},\kappa_k\in\mathbb{R}} \int_{\Pi^{k-1}} \left| f_k(\theta, \kappa_k; \boldsymbol{\lambda}) - f_k(\theta_0, \kappa_k^0; \boldsymbol{\lambda}) \right|^2 g(\theta; \boldsymbol{\lambda}) d\boldsymbol{\lambda} \geq \epsilon_2 > 0,$$

for some $\epsilon_2 > 0$.

Then, compactness of Θ and usual identification arguments for $\phi(\theta; \lambda_a)$, $a = 1, \ldots, k-1$, imply that for parameter values which are away of those whose roots $\boldsymbol{\varphi} \in \mathcal{S}\cup\{\boldsymbol{\varphi}_0\}$, it holds that

$$\inf_{\theta\in B(\rho,\theta(\boldsymbol{\varphi}))^c\cap B(\eta,\theta_0)^c,\boldsymbol{\varphi}\in\mathcal{S},\kappa_k\in\mathbb{R}} \int_{\Pi^{k-1}} \left| f_k(\theta, \kappa_k; \boldsymbol{\lambda}) - f_k(\theta_0, \kappa_k^0; \boldsymbol{\lambda}) \right|^2 g(\theta; \boldsymbol{\lambda}) d\boldsymbol{\lambda} \geq \epsilon_3 > 0,$$

for some $\epsilon_3 > 0$. Then the proof follows setting $\epsilon = \min\{\epsilon_1, \epsilon_2, \epsilon_3\}$. $\square$

Proof of Lemma 1: We only give the proof for $k = 3$ and $g = 1$. Consider first the case of a pure $MA(q)$ model where all roots are real. In this case, we can write the model for f_3 in terms of the roots $\boldsymbol{\varphi}_0 = (\varphi_1, \ldots, \varphi_q)$ as

$$f_3(\theta(\boldsymbol{\varphi}_0), \kappa_3; \lambda_1, \lambda_2) = \frac{\kappa_3}{(2\pi)^2} \prod_{j=1}^{q} (a(\varphi_j; \lambda_1, \lambda_2) + ib(\varphi_j; \lambda_1, \lambda_2))$$

where

$$\begin{aligned} a(\varphi_j; \lambda_1, \lambda_2) &= -\left(\varphi_j - 1\right)\left(\varphi_j - \varphi_j \cos(\lambda_1 + \lambda_2) + \varphi_j^2 - \varphi_j \cos\lambda_1 - \varphi_j \cos\lambda_2 + 1\right) \\ b(\varphi_j; \lambda_1, \lambda_2) &= -\varphi_j\left(\varphi_j + 1\right)(\sin\lambda_1 - \sin(\lambda_1 + \lambda_2) + \sin\lambda_2). \end{aligned}$$

Note that when $\varphi_j = 1$ the bispectrum is purely imaginary, whereas when φ_j takes the values 0 or -1 it is purely real.

Now, consider the typical element of $\mathcal{S}$, where (without generality, the first) $q_1 = 1, \ldots, q$ roots are inverted and the last $q - q_1$ are kept the same, and denote these new roots as $\boldsymbol{\varphi}^*$ and any possible value of the third order cumulant by κ_3^*. In that case

$$f_3(\theta(\boldsymbol{\varphi}^*), \kappa_3^*; \lambda_1, \lambda_2) = \frac{\kappa_3^*}{(2\pi)^2} \prod_{j=1}^{q_1} \left(a(\frac{1}{\varphi_j}; \lambda_1, \lambda_2) + ib(\frac{1}{\varphi_j}; \lambda_1, \lambda_2) \right) \prod_{j=q_1+1}^{q} \left(a(\varphi_j; \lambda_1, \lambda_2) + ib(\varphi_j; \lambda_1, \lambda_2) \right).$$

Since, for $\varphi_j \neq 0$,

$$a(\frac{1}{\varphi_j}; \lambda_1, \lambda_2) = \frac{-1}{\varphi_j^3} a(\varphi_j; \lambda_1, \lambda_2), \text{ and } b(\frac{1}{\varphi_j}; \lambda_1, \lambda_2) = \frac{1}{\varphi_j^2} b(\varphi_j; \lambda_1, \lambda_2)), \tag{19}$$

then the LHS of the inequality in (12) is

$$\frac{1}{(2\pi)^4} \int_{\Pi^2} \left| \prod_{j=1}^{q_1} ((\kappa_3 + \frac{\kappa_3^*}{\varphi_j^3}) a(\varphi_j; \lambda_1, \lambda_2) + i(\kappa_3 - \frac{\kappa_3^*}{\varphi_j^2}) b(\varphi_j; \lambda_1, \lambda_2)) \right|^2 C(\boldsymbol{\varphi}_0; \lambda_1, \lambda_2) d\lambda_1 d\lambda_2$$

writing $C(\boldsymbol{\varphi}_0; \lambda_1, \lambda_2) = \Pi_{j=q_1+1}^{q} \left| (a(\varphi_j; \lambda_1, \lambda_2) + ib(\varphi_j; \lambda_1, \lambda_2) \right|^2$, with a and b orthogonal in Π^2, this can be written as

$$\frac{1}{(2\pi)^4} \int_{\Pi^2} \prod_{j=1}^{q_1} \left(\left(\kappa_3 + \frac{\kappa_3^*}{\varphi_j^3} \right)^2 a(\varphi_j; \lambda_1, \lambda_2)^2 + \left(\kappa_3 - \frac{\kappa_3^*}{\varphi_j^2} \right)^2 b(\varphi_j; \lambda_1, \lambda_2)^2 \right) C(\boldsymbol{\varphi}_0; \lambda_1, \lambda_2) d\lambda_1 d\lambda_2 \tag{20}$$

$$\begin{aligned} &= \frac{1}{(2\pi)^4} \prod_{j=1}^{q_1} \left(\kappa_3 + \frac{\kappa_3^*}{\varphi_j^3} \right)^2 \int_{\Pi^2} a(\varphi_j; \lambda_1, \lambda_2)^2 C(\boldsymbol{\varphi}_0; \lambda_1, \lambda_2) d\lambda_1 d\lambda_2 \\ &\quad + \frac{1}{(2\pi)^4} \prod_{j=1}^{q_1} \left(\kappa_3 - \frac{\kappa_3^*}{\varphi_j^2} \right)^2 \int_{\Pi^2} b(\varphi_j; \lambda_1, \lambda_2)^2 C(\boldsymbol{\varphi}_0; \lambda_1, \lambda_2) d\lambda_1 d\lambda_2 \\ &> \prod_{j=1}^{q_1} \left\{ C_1 \left(\kappa_3 + \frac{\kappa_3^*}{\varphi_j^3} \right)^2 + C_2 \left(\kappa_3 - \frac{\kappa_3^*}{\varphi_j^2} \right)^2 \right\} > 0, \end{aligned}$$

for any choice of κ_3^* and all possible κ_3 because $C_1, C_2 > 0$ for all $\varphi_j \neq \pm 1$, $j = 1, \ldots, q_1$.

The previous discussion has considered the case where all roots are real. The basic mechanism to achieve identification is through relations (19) that show the change in the bispectrum when the roots are inverted. Note that the complex roots are dealt with considering simultaneously the pairs of conjugated roots, guaranteeing that the observed process is real (and not complex). In the case that some pairs of conjugated complex roots appear there is also a relation between the contributions to the bispectrum when the roots are inverted. Note that for a complex φ_j the contribution to the bispectrum is

$$\left(1 - \varphi_j e^{-ij\lambda_1}\right) \left(1 - \bar{\varphi}_j e^{-ij\lambda_1}\right) \left(1 - \varphi_j e^{-ij\lambda_2}\right) \left(1 - \bar{\varphi}_j e^{-ij\lambda_2}\right) \left(1 - \varphi_j e^{ij(\lambda_1+\lambda_2)}\right) \left(1 - \bar{\varphi}_j e^{ij(\lambda_1+\lambda_2)}\right), \tag{21}$$

which, introducing the polar representation of φ_j, $\varphi_j = Me^{i\omega}$, so that M is the modulus and ω the phase, (21) can be written as, $\rho = \cos(\omega)$,

$$\left(1 - 2M\rho e^{-ij\lambda_1} + M^2 e^{-ij2\lambda_1}\right)\left(1 - 2M\rho e^{-ij\lambda_2} + M^2 e^{-ij2\lambda_2}\right)\left(1 - 2M\rho e^{ij(\lambda_1+\lambda_2)} + M^2 e^{ij2(\lambda_1+\lambda_2)}\right). \tag{22}$$

This expression is useful since inverting the roots just involves now inverting M since $\cos(\omega) = \cos(-\omega)$. Then, it is simple to show that the real and imaginary parts of (22) are respectively

$$\begin{aligned} c(M,\omega,\lambda_1,\lambda_2) \quad : \quad &= 1 - 8M^3\rho^3 + M^6 + (M^2 + M^4)(\cos 2j\lambda_1 + \cos 2j\lambda_2 + \cos 2(j\lambda_1 + j\lambda_2)) \\ &+ (4M^4\rho^2 + 4M^2\rho^2 - 2M^5\rho - 2M\rho)(\cos j\lambda_1 + \cos j\lambda_2 + \cos(j\lambda_1 + j\lambda_2)) \\ &- 4M^3\rho(\cos(j\lambda_1 - j\lambda_2) + \cos(j\lambda_1 + 2j\lambda_2) + \cos(2j\lambda_1 + j\lambda_2)), \end{aligned}$$

and

$$\begin{aligned} d(M,\omega,\lambda_1,\lambda_2) \quad : \quad &= (M^4 - M^2)(\sin 2j\lambda_1 + \sin 2j\lambda_2 - \sin 2(j\lambda_1 + j\lambda_2)) \\ &+ (4M^2\rho^2 - 2M^5\rho - 4M^4\rho^2 + 2M\rho)(\sin j\lambda_1 + \sin j\lambda_2 - \sin((j\lambda_1 + j\lambda_2)). \end{aligned}$$

Now the contribution to the bispectrum when a complex root is inverted is $M^{-6}(c(M,\omega,\lambda_1,\lambda_2) - id(M,\omega,\lambda_1,\lambda_2))$. So, the contribution of one pair of conjugated complex roots to the LHS of the inequality in (12) is given by

$$\left|\left(\kappa_3 - \frac{\kappa_3^*}{M^6}\right)c(M,\omega,\lambda_1,\lambda_2) + i\left(\kappa_3 + \frac{\kappa_3^*}{M^6}\right)d(M,\omega,\lambda_1,\lambda_2)\right|^2,$$

so identification is achieved similarly given that c and d are orthogonal in Π^2.

Finally, the extension when considering the inverse roots of an autoregressive component is also straightforward. Notice that in the general ARMA case the bispectrum has the form

$$f_3(\boldsymbol{\varphi}, \kappa_3; \lambda_1, \lambda_2) = \frac{\kappa_3}{(2\pi)^2}\frac{\Pi_{j=1}^q h(\varphi_j;\lambda_1,\lambda_2)}{\Pi_{j=1}^p h(\phi_j;\lambda_1,\lambda_2)},$$

where $h = a + ib$ for real roots and $h(\varphi_j)h(\bar{\varphi}_j) = c + id$ for a pair of complex conjugate roots. For simplicity in the notation, suppose we invert the AR roots and keep the MA component. In particular, if p_1 roots are inverted and the last $p - p_1$ are kept the same, then

$$f_3(\boldsymbol{\varphi}^*, \kappa_3^*; \lambda_1, \lambda_2) = \frac{\kappa_3^*}{(2\pi)^2}\frac{\Pi_{j=1}^q h(\varphi_j;\lambda_1,\lambda_2)}{\Pi_{j=1}^{p_1} h(\phi_j^{-1};\lambda_1,\lambda_2)\Pi_{j=p_1+1}^p h(\phi_j;\lambda_1,\lambda_2)},$$

so that the LHS of the inequality in (12) takes the form

$$\frac{1}{(2\pi)^4}\int_{\Pi^2}\left|\frac{\Pi_{j=1}^q h(\varphi_j;\lambda_1,\lambda_2)\Pi_{j=1}^p h(\phi_j^{-1};\lambda_1,\lambda_2)}{\Pi_{j=p_1+1}^p h(\phi_j;\lambda_1,\lambda_2)\Pi_{j=1}^p h(\phi_j;\lambda_1,\lambda_2)}\left\{\kappa_3\prod_{j=1}^p h(\phi_j^{-1};\lambda_1,\lambda_2) - \kappa_3^*\prod_{j=1}^p h(\phi_j;\lambda_1,\lambda_2)\right\}\right|^2 d\lambda_1 d\lambda_2,$$

and the analysis follows similarly to equation (20) where now

$$C(\varphi;\lambda_1,\lambda_2) = \frac{\Pi_{j=1}^q h(\varphi_j;\lambda_1,\lambda_2)\Pi_{j=1}^p h(\phi_j^{-1};\lambda_1,\lambda_2)}{\Pi_{j=p_1+1}^p h(\phi_j;\lambda_1,\lambda_2)\Pi_{j=1}^p h(\phi_j;\lambda_1,\lambda_2)}.$$

Similarly the same argument follows for the case of autoregressive complex roots, covering as well the case of all-pass models, and for any $k \geq 4$. $\square$

Proof of Theorem 2. It follows from the identification provided by Theorem 1 for the objective functions with $k = 3$ and 4, and the uniform converge of the objective function, exploiting Lemmas A1 and A2 in Appendix B and the smoothness of $f_k(\theta, \kappa_k; \boldsymbol{\lambda})$ for ARMA models under Assumption 1. $\square$

Proof of Theorem 3. We work with (re-scaled) concentrated objective functions using $\kappa_{kT}^{\dagger}(\theta)$ but not depending on any preliminary $\widetilde{\kappa}_k$ as $\widetilde{L}_{wT}(\theta)$,

$$L_{kT}^{\dagger}(\theta) = \frac{\widetilde{\kappa}_k^2}{(2\pi)^{2(k-1)}} \widetilde{L}_{kT}^{0}(\theta, \kappa_{kT}^{\dagger}(\theta)) = \frac{1}{T^{k-1}} \sum_{\mathbf{j}=1}^{T-1} \frac{\left| I_k(\boldsymbol{\lambda}_{\mathbf{j}}) - f_k(\theta, \kappa_{kT}^{\dagger}(\theta); \boldsymbol{\lambda}_{\mathbf{j}}) \right|^2}{\left| \phi_k(\bar{\theta}_T; \boldsymbol{\lambda}_{\mathbf{j}}) \right|^2}.$$

Then

$$\frac{\partial}{\partial \theta} L_{kT}^{\dagger}(\theta) = \frac{-2}{T^{k-1}} \sum_{\mathbf{j}=1}^{T-1} \frac{\mathrm{Re}\left\{ I_k(\boldsymbol{\lambda}_{\mathbf{j}}) - f_k(\theta, \kappa_{kT}^{\dagger}(\theta); \boldsymbol{\lambda}_{\mathbf{j}}) \right\} \frac{\partial}{\partial \theta} f_k(\theta, \kappa_{kT}^{\dagger}(\theta); -\boldsymbol{\lambda}_{\mathbf{j}})}{\left| \phi_k(\bar{\theta}_T; \boldsymbol{\lambda}_{\mathbf{j}}) \right|^2}$$

while

$$\begin{aligned} \frac{\partial^2}{\partial \theta \partial \theta'} L_{kT}^{\dagger}(\theta) &= \frac{-2}{T^{k-1}} \sum_{\mathbf{j}=1}^{T-1} \frac{\mathrm{Re}\left\{ I_k(\boldsymbol{\lambda}_{\mathbf{j}}) - f_k(\theta, \kappa_{kT}^{\dagger}(\theta); \boldsymbol{\lambda}_{\mathbf{j}}) \right\} \frac{\partial^2}{\partial \theta \partial \theta'} f_k(\theta, \kappa_{kT}^{\dagger}(\theta); -\boldsymbol{\lambda}_{\mathbf{j}})}{\left| \phi_k(\bar{\theta}_T; \boldsymbol{\lambda}_{\mathbf{j}}) \right|^2} \\ &\quad + \frac{2}{T^{k-1}} \sum_{\mathbf{j}=1}^{T-1} \frac{\mathrm{Re}\left\{ \frac{\partial}{\partial \theta} f_k(\theta, \kappa_{kT}^{\dagger}(\theta); \boldsymbol{\lambda}_{\mathbf{j}}) \frac{\partial}{\partial \theta} f_k(\theta, \kappa_{kT}^{\dagger}(\theta); -\boldsymbol{\lambda}_{\mathbf{j}})' \right\}}{\left| \phi_k(\bar{\theta}_T; \boldsymbol{\lambda}_{\mathbf{j}}) \right|^2}. \end{aligned}$$

Now

$$\frac{\partial}{\partial \theta} f_k(\theta, \kappa_{kT}^{\dagger}(\theta); \boldsymbol{\lambda}_{\mathbf{j}}) = \frac{\dot{\kappa}_{kT}^{\dagger}(\theta)}{(2\pi)^{k-1}} \phi_k(\theta; \boldsymbol{\lambda}_{\mathbf{j}}) + \frac{\kappa_{kT}^{\dagger}(\theta)}{(2\pi)^{k-1}} \dot{\phi}_k(\theta; \boldsymbol{\lambda}_{\mathbf{j}})$$

where $\dot{\phi}_k(\theta; \lambda_{\mathbf{j}}) = \frac{\partial}{\partial \theta} \phi_k(\theta; \lambda_{\mathbf{j}})$ so that

$$\dot{\kappa}_{kT}^{\dagger}(\theta) = \frac{\partial}{\partial \theta} \kappa_{kT}^{\dagger}(\theta) = -\left(\frac{2\pi}{T} \right)^{k-1} \sum_{\mathbf{j}=1}^{T-1} \mathrm{Re}\left(\frac{I_k(\lambda_{\mathbf{j}})}{\phi_k(\theta; \lambda_{\mathbf{j}})} \varphi_k(\theta; \lambda_{\mathbf{j}}) \right)$$

with $\varphi_k(\theta; \lambda_{\mathbf{j}}) = \frac{\partial}{\partial \theta} \log \phi_k(\theta; \lambda_{\mathbf{j}})$.

In particular, from Theorem 2 we have that the estimate θ_T minimizing $\tilde{L}_{wT}(\theta)$ satisfies

$$\theta_T - \theta_0 = -\left(\frac{\partial^2}{\partial \theta \partial \theta'} \tilde{L}_{wT}(\theta_0) + o_p(1) \right)^{-1} \frac{\partial}{\partial \theta} \tilde{L}_{wT}(\theta_0) \tag{23}$$

where we argue, that for any $\tilde{\theta}_T$ such that $\left\|\tilde{\theta}_T - \theta_0\right\| \leq \|\theta_T - \theta_0\|$, using $\theta_T - \theta_0 = o_p(1)$,

$$\begin{aligned}
\frac{\partial^2}{\partial\theta\partial\theta'} L_{kT}^{\dagger}(\tilde{\theta}_T) &= \frac{\partial^2}{\partial\theta\partial\theta'} L_{kT}^{\dagger}(\theta_0) + o_p(1) \qquad (24)\\
&= \frac{2}{T^{k-1}} \sum_{\mathbf{j}=1}^{T-1} \frac{\operatorname{Re}\left\{\frac{\partial}{\partial\theta} f_k(\theta, \kappa_{kT}^{\dagger}(\theta_0); \boldsymbol{\lambda}_{\mathbf{j}}) \frac{\partial}{\partial\theta} f_k(\theta, \kappa_{kT}^{\dagger}(\theta_0); -\boldsymbol{\lambda}_{\mathbf{j}})'\right\}}{\left|\phi_k(\bar{\theta}_T; \boldsymbol{\lambda}_{\mathbf{j}})\right|^2} + o_p(1)\\
&= \frac{2\kappa_k^2}{(2\pi)^{3(k-1)}} \int_{\Pi^{k-1}} \operatorname{Re}\left\{\varphi_k^1(\theta_0; \boldsymbol{\lambda})\varphi_k^1(\theta_0; -\boldsymbol{\lambda})'\right\} d\boldsymbol{\lambda} + o_p(1)\\
&= \frac{2\kappa_k^2}{(2\pi)^{2(k-1)}} \Phi_{k,0} + o_p(1),
\end{aligned}$$

where $\varphi_k^1(\theta_0; \lambda_{\mathbf{j}}) := \varphi_k(\theta_0; \lambda_{\mathbf{j}}) - \operatorname{Av}[\varphi_k(\theta_0)]$, $\operatorname{Av}[\varphi_k(\theta_0)] = (2\pi)^{1-k} \int_{\Pi^{k-1}} \varphi_k(\theta_0; \boldsymbol{\lambda}) d\boldsymbol{\lambda}$, and

$$\begin{aligned}
\Phi_{k,0} &= (2\pi)^{1-k} \int_{\Pi^{k-1}} \operatorname{Re}\left\{\varphi_k^1(\theta_0; \boldsymbol{\lambda})\varphi_k^1(\theta_0; -\boldsymbol{\lambda})'\right\} d\boldsymbol{\lambda} = k\Phi_0, \qquad k = 3, 4\\
\Phi_{2,0} &= (2\pi)^{-1} \int_{\Pi} \operatorname{Re}\left\{\varphi_2^1(\theta_0; \lambda)\varphi_2^1(\theta_0; \lambda)'\right\} d\lambda = 2\left(\Phi_0 + \Phi_0^*\right), \qquad k = 2,
\end{aligned}$$

using from immediate application of Lemmas A1 and A2,

$$\kappa_{kT}^{\dagger}(\theta_0) = \kappa_k + O_p\left(T^{-1/2}\right) \qquad (25)$$

and

$$\dot{\kappa}_{kT}^{\dagger}(\theta_0) = -\frac{\kappa_k}{(2\pi)^{k-1}} \int_{\Pi^{k-1}} \operatorname{Re}\left\{\varphi_k(\theta_0; \boldsymbol{\lambda})\right\} d\boldsymbol{\lambda} + O_p\left(T^{-1/2}\right) = -\kappa_k \operatorname{Av}\left[\varphi_k(\theta_0)\right] + O_p\left(T^{-1/2}\right), \qquad (26)$$

since $\int_{\Pi^{k-1}} \operatorname{Im}\left\{\varphi_k(\theta_0; \boldsymbol{\lambda})\right\} d\boldsymbol{\lambda} = 0$, and, uniformly in $\mathbf{j}$,

$$\begin{aligned}
\frac{\partial}{\partial\theta} f_k(\theta_0, \kappa_{kT}^{\dagger}(\theta_0); \boldsymbol{\lambda}_{\mathbf{j}}) &= \frac{-\kappa_k \operatorname{Av}\left[\varphi_k(\theta_0)\right]}{(2\pi)^{k-1}} \phi_k(\theta_0; \boldsymbol{\lambda}_{\mathbf{j}}) + \frac{\kappa_k}{(2\pi)^{k-1}} \dot{\phi}_k(\theta_0; \boldsymbol{\lambda}_{\mathbf{j}}) + o_p(1)\\
&= \frac{\kappa_k}{(2\pi)^{k-1}} \phi_k(\theta_0; \boldsymbol{\lambda}_{\mathbf{j}}) \left\{\varphi_k(\theta_0; \boldsymbol{\lambda}_{\mathbf{j}}) - \operatorname{Av}\left[\varphi_k(\theta_0)\right]\right\} + o_p(1)\\
&= f_k(\theta_0, \kappa_k; \boldsymbol{\lambda}_{\mathbf{j}})\varphi_k^1(\theta_0; \boldsymbol{\lambda}_{\mathbf{j}}) + o_p(1).
\end{aligned}$$

This gives the expression for Σ_0 by just normalizing $\tilde{L}_{wT}(\theta_0)$ by $\frac{1}{4}$ and eliminating the factors $\frac{\widetilde{\kappa}_k^2}{(2\pi)^{2(k-1)}}$ in the definition of $\tilde{L}_{kT}$.

Denoting $f_k^0(\boldsymbol{\lambda}_{\mathbf{j}}) = f_k(\theta_0, \kappa_k; \boldsymbol{\lambda}_{\mathbf{j}})$, the terms inside the summation in $\frac{\partial}{\partial\theta} L_{kT}^{\dagger}(\theta_0)$ can be written as

$$\frac{\operatorname{Re}\left\{\left(I_k(\boldsymbol{\lambda}_{\mathbf{j}}) - f_k^0(\boldsymbol{\lambda}_{\mathbf{j}})\right) f_k^0(-\boldsymbol{\lambda}_{\mathbf{j}})\varphi_k^1(\theta_0; -\boldsymbol{\lambda}_{\mathbf{j}})\right\}}{\left|\phi_k(\theta_0; \boldsymbol{\lambda}_{\mathbf{j}})\right|^2} \qquad (27)$$

$$-\frac{\operatorname{Re}\left\{\left(I_k(\boldsymbol{\lambda}_{\mathbf{j}}) - f_k^0(\boldsymbol{\lambda}_{\mathbf{j}})\right) f_k^0(-\boldsymbol{\lambda}_{\mathbf{j}})\varphi_k^1(\theta_0; -\boldsymbol{\lambda}_{\mathbf{j}})\right\}}{\left|\phi_k(\theta_0; \boldsymbol{\lambda}_{\mathbf{j}})\right|^2} \left\{1 - \frac{\left|\phi_k(\theta_0; \boldsymbol{\lambda}_{\mathbf{j}})\right|^2}{\left|\phi_k(\bar{\theta}_T; \boldsymbol{\lambda}_{\mathbf{j}})\right|^2}\right\} \qquad (28)$$

$$-\frac{\operatorname{Re}\left\{\left(I_k(\boldsymbol{\lambda}_{\mathbf{j}}) - f_k^0(\boldsymbol{\lambda}_{\mathbf{j}})\right)\left(f_k^0(-\boldsymbol{\lambda}_{\mathbf{j}})\varphi_k^1(\theta_0; -\boldsymbol{\lambda}_{\mathbf{j}}) - \frac{\partial}{\partial\theta} f_k(\theta_0, \kappa_{kT}^{\dagger}(\theta_0); -\boldsymbol{\lambda}_{\mathbf{j}})\right)\right\}}{\left|\phi_k(\bar{\theta}_T; \boldsymbol{\lambda}_{\mathbf{j}})\right|^2} \qquad (29)$$

$$-\frac{\operatorname{Re}\left\{\left(f_k(\theta_0, \kappa_{kT}^{\dagger}(\theta_0); \boldsymbol{\lambda}_{\mathbf{j}}) - f_k^0(\boldsymbol{\lambda}_{\mathbf{j}})\right) \frac{\partial}{\partial\theta} f_k(\theta_0, \kappa_{kT}^{\dagger}(\theta_0); -\boldsymbol{\lambda}_{\mathbf{j}})\right\}}{\left|\phi_k(\bar{\theta}_T; \boldsymbol{\lambda}_{\mathbf{j}})\right|^2} \qquad (30)$$

Using that

$$\frac{\left|\phi_k(\theta_0;\boldsymbol{\lambda}_{\mathbf{j}})\right|^2}{\left|\phi_k(\bar{\theta};\boldsymbol{\lambda}_{\mathbf{j}})\right|^2} = 1 - 2\operatorname{Re}\left\{\varphi_k(\theta_0;\boldsymbol{\lambda}_{\mathbf{j}})\right\}'\left(\bar{\theta}-\theta_0\right),$$

plus $O_p\left(T^{-1}\right)$ terms uniform in $\mathbf{j}$, the contribution from (28) to $\frac{\partial}{\partial\theta}L^{\dagger}_{kT}(\theta_0)$ is showed to be $O_p\left(T^{-1}\right)$ because $I_k(\boldsymbol{\lambda}_{\mathbf{j}}) - f_k(\theta_0,\kappa_k;\boldsymbol{\lambda}_{\mathbf{j}})$ is centered up to a factor $o\left(T^{-1/2}\right)$, so that the average is $O_p\left(T^{-1/2}\right)$ and $\bar{\theta}-\theta_0 = O_p\left(T^{-1/2}\right)$.

Next, we note that

$$\begin{aligned}
& f^0_k(-\boldsymbol{\lambda}_{\mathbf{j}})\varphi^1_k(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}}) - \frac{\partial}{\partial\theta}f_k(\theta_0,\kappa^{\dagger}_{kT}(\theta_0);-\boldsymbol{\lambda}_{\mathbf{j}}) \\
= \; & \frac{\kappa_k}{(2\pi)^2}\phi_k(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}})\varphi_k(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}}) - \frac{\kappa_k}{(2\pi)^2}\phi_k(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}})\mathrm{Av}\left[\varphi_k(\theta_0)\right] \\
& -\frac{\dot{\kappa}^{\dagger}_{kT}(\theta_0)}{(2\pi)^2}\phi_k(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}}) - \frac{\kappa^{\dagger}_{kT}(\theta_0)}{(2\pi)^2}\dot{\phi}_k(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}}) \\
= \; & -\frac{\left(\dot{\kappa}^{\dagger}_{kT}(\theta_0)+\kappa_k\mathrm{Av}\left[\varphi_k(\theta_0)\right]\right)}{(2\pi)^2}\phi_k(\theta;-\boldsymbol{\lambda}_{\mathbf{j}}) - \frac{\left(\kappa^{\dagger}_{kT}(\theta_0)-\kappa_k\right)}{(2\pi)^2}\dot{\phi}_k(\theta;-\boldsymbol{\lambda}_{\mathbf{j}})
\end{aligned}$$

so using (25) and (26) the contribution of the term (29) to $\frac{\partial}{\partial\theta}L^{\dagger}_{kT}(\theta_0)$ is $O_p\left(T^{-1}\right)$.

Finally, using (25),

$$f_k(\theta_0,\kappa^{\dagger}_{kT}(\theta_0);\boldsymbol{\lambda}_{\mathbf{j}}) - f^0_k(\boldsymbol{\lambda}_{\mathbf{j}}) = \left(\frac{\kappa^{\dagger}_{kT}(\theta_0)}{\kappa_k}-1\right)f_k(\theta_0,\kappa_k;\boldsymbol{\lambda}_{\mathbf{j}})$$

and

$$\frac{\partial}{\partial\theta}f_k(\theta_0,\kappa^{\dagger}_{kT}(\theta_0);-\boldsymbol{\lambda}_{\mathbf{j}}) = f^0_k(-\boldsymbol{\lambda}_{\mathbf{j}})\varphi^1_k(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}})\left(1+O_p\left(T^{-1/2}\right)\right),$$

where the $O_p\left(T^{-1/2}\right)$ term is uniform in $\mathbf{j}$, we find that the term coming from (30) in $\frac{\partial}{\partial\theta}L^{\dagger}_{kT}(\theta_0)$ is

$$\begin{aligned}
& \frac{2}{T^{k-1}}\frac{\left(\kappa^{\dagger}_{kT}(\theta_0)-\kappa_k\right)}{(2\pi)^{k-1}}\sum_{\mathbf{j}=1}^{T-1}\frac{\operatorname{Re}\left\{\phi_k(\theta_0;\boldsymbol{\lambda}_{\mathbf{j}})f^0_k(-\boldsymbol{\lambda}_{\mathbf{j}})\varphi^1_k(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}})\right\}}{\left|\phi_k(\theta_0;\boldsymbol{\lambda}_{\mathbf{j}})\right|^2} \\
& +O_p\left(T^{-1}T^{1-k}\sum_{\mathbf{j}=1}^{T-1}\frac{f_k(\theta_0,\kappa_k;\boldsymbol{\lambda}_{\mathbf{j}})f^0_k(-\boldsymbol{\lambda}_{\mathbf{j}})\left\|\varphi^1_k(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}})\right\|^2}{\left|\phi_k(\theta_0;\boldsymbol{\lambda}_{\mathbf{j}})\right|^2}\right)
\end{aligned} \tag{31}$$

which is $O_p\left(T^{-1}\right)$ because the first average in (31) converges to a zero value integral,

$$\begin{aligned}
& \frac{1}{T^{k-1}}\sum_{\mathbf{j}=1}^{T-1}\frac{\operatorname{Re}\left\{\phi_k(\theta_0;\boldsymbol{\lambda}_{\mathbf{j}})f^0_k(-\boldsymbol{\lambda}_{\mathbf{j}})\varphi^1_k(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}})\right\}}{\left|\phi_k(\theta_0;\boldsymbol{\lambda}_{\mathbf{j}})\right|^2} \\
= \; & \frac{1}{T^{k-1}}\frac{\kappa_k}{(2\pi)^{k-1}}\sum_{\mathbf{j}=1}^{T-1}\operatorname{Re}\left\{\varphi^1_k(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}})\right\} \\
= \; & \frac{1}{(2\pi)^{2(k-1)}}\int_{\Pi^{k-1}}\varphi^1_k(\theta_0;-\boldsymbol{\lambda})d\boldsymbol{\lambda} + O\left(T^{-1}\right) = O\left(T^{-1}\right)
\end{aligned}$$

as φ^1_k is centered in Π^2, and the second average in (31) converges to a bounded integral so this term is also negligible

So finally, considering the terms (27),

$$T^{1/2}\frac{\partial}{\partial\theta}\tilde{L}_{kT}^{\dagger}(\theta_0)=T^{1/2}\frac{-2}{T^{1-k}}\left(\frac{\kappa_k}{(2\pi)^{k-1}}\right)^2\sum_{\mathbf{j}=1}^{T-1}\mathrm{Re}\left\{\left(\frac{I_k(\boldsymbol{\lambda}_{\mathbf{j}})}{f_k(\boldsymbol{\lambda}_{\mathbf{j}})}-1\right)\varphi_k^1(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}})\right\}+o_p(1),$$

From Lemma A1 in Appendix B, we can write

$$T^{1/2}\frac{\partial}{\partial\theta}\tilde{L}_{kT}^{\dagger}(\theta_0)=T^{1/2}\frac{-2}{T^{1-k}}\frac{\kappa_k}{(2\pi)^{2(k-1)}}\sum_{\mathbf{j}=1}^{T-1}\mathrm{Re}\left\{\left((2\pi)^{k-1}I_k^{\varepsilon}(\boldsymbol{\lambda}_{\mathbf{j}})-\kappa_k\right)\varphi_k^1(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}})\right\}+o_p(1).$$

Now we analyze each of the elements in $\frac{\partial}{\partial\theta}\tilde{L}_{wT}(\theta_0)$ separately for $k=2,3,4$. In particular for $k=3$, $I_3^{\varepsilon}(\boldsymbol{\lambda}_{\mathbf{j}})=I_3^{\varepsilon}(\lambda_{j_1},\lambda_{j_2})$ is the innovations biperiodogram and we can write

$$\begin{aligned}(2\pi)^2 I_3^{\varepsilon}(\boldsymbol{\lambda}_{\mathbf{j}})-\kappa_3 &= \frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t^3-\kappa_3\right)+\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2\sum_{r=1}^{t-1}\varepsilon_r A_{t,r}(\boldsymbol{\lambda}_{\mathbf{j}}) \\ &\quad+\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\sum_{r,s=1}^{t-1}\varepsilon_r\varepsilon_s B_{t,r,s}(\boldsymbol{\lambda}_{\mathbf{j}}),\end{aligned}\tag{32}$$

where

$$\begin{aligned}A_{t,r}(\boldsymbol{\lambda}_{\mathbf{j}}) &= \exp\{-i((r-t)\lambda_{j_1})\}+\exp\{-i((r-t)\lambda_{j_2})\}+\exp\{-i((t-r)(\lambda_{j_1}+\lambda_{j_2}))\} \\ B_{t,r,s}(\boldsymbol{\lambda}_{\mathbf{j}}) &= \exp\{-i((r-t)\lambda_{j_1}+(s-t)\lambda_{j_2})\}+\exp\{-i((r-s)\lambda_{j_1}+(t-s)\lambda_{j_2})\} \\ &\quad+\exp\{-i((t-s)\lambda_{j_1}+(r-s)\lambda_{j_2})\}.\end{aligned}$$

The first term in (32) does not contribute asymptotically to the distribution of $T^{1/2}\frac{\partial}{\partial\theta}\tilde{L}_{3T}^{\dagger}(\theta_0)$ because it is $O_p\left(T^{-1/2}\right)$ and not depending on $\boldsymbol{\lambda}_{\mathbf{j}}$, with

$$\frac{1}{T^2}\sum_{\mathbf{j}=1}^{T-1}\mathrm{Re}\left\{\varphi_3^1(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}})\right\}=O\left(T^{-1}\right)\tag{33}$$

because $\mathrm{Av}\left[\varphi_3^1(\theta_0)\right]=(2\pi)^{-2}\int_{\Pi^2}\mathrm{Re}\left\{\varphi_3^1(\theta_0;\boldsymbol{\lambda})\right\}d\boldsymbol{\lambda}=0$ by construction.

We can write the second term in (32) as

$$\frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t^2-\sigma^2\right)\sum_{r=1}^{t-1}\varepsilon_r A_{t,r}(\boldsymbol{\lambda}_{\mathbf{j}})+\frac{\sigma^2}{T}\sum_{t=1}^{T}\sum_{r=1}^{t-1}\varepsilon_r A_{t,r}(\boldsymbol{\lambda}_{\mathbf{j}}),\tag{34}$$

where the elements in the average in t of the first term in (34) are a MDS, and noting that $j_1+j_2\neq T$ the second term in (34) can be written as

$$\frac{\sigma^2}{T}\sum_{t=1}^{T}\varepsilon_t\sum_{r=t+1}^{T}A_{r,t}(\boldsymbol{\lambda}_{\mathbf{j}})=-\frac{\sigma^2}{T}\sum_{t=1}^{T}\varepsilon_t\sum_{r=1}^{t-1}A_{r,t}(\boldsymbol{\lambda}_{\mathbf{j}})-\frac{3\sigma^2}{T}\sum_{t=1}^{T}\varepsilon_t\tag{35}$$

using that $\sum_{t=1}^{T}\exp(it\lambda_j)=0$ if $j\neq 0 \bmod T$ and that restriction on $j_1+j_2\neq 0 \bmod T$. Note that the second term in (35) is $O_p\left(T^{-1/2}\right)$ and does not contribute to the asymptotic distribution of $T^{1/2}\frac{\partial}{\partial\theta}\tilde{L}_{3T}(\theta_0)$ by (33).

The third term in (32) can be written as

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\sum_{r,s=1}^{t-1}\left(\varepsilon_r\varepsilon_s-\sigma^2\{r=s\}\right)B_{t,r,s}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right) \tag{36}$$

$$+\frac{\sigma^2}{T}\sum_{t=1}^{T}\varepsilon_t\sum_{r=1}^{t-1}B_{t,r,r}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right) \tag{37}$$

where (36) is the average of a MDS and (37) is equal to the opposite of the first term in the rhs of (35) noting that $B_{t,r,r}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)=A_{r,t}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)$, so they cancel out.

Therefore, we can write

$$T^{1/2}\frac{\partial}{\partial\theta}\tilde{L}_{3T}^{\dagger}(\theta_0)=\sum_{t=1}^{T}z_{3,t}+o_p(1),$$

where $z_{3,t}$ is a MDS,

$$\begin{aligned} z_{3,t} &= \frac{1}{T^{1/2}}\left(\varepsilon_t^2-\sigma^2\right)\sum_{r=1}^{t-1}\varepsilon_r c_{T,t-r} \\ &\quad+\frac{1}{T^{1/2}}\varepsilon_t\sum_{r,s=1}^{t-1}\left(\varepsilon_r\varepsilon_s-\sigma^2\{r=s\}\right)g_{T,t-r,t-s} \end{aligned}$$

supressing dependence on T in the notation, and

$$\begin{aligned} c_{T,t-r} &= \frac{-2\kappa_3}{(2\pi)^4}\frac{1}{T^2}\sum_{\mathbf{j}=1}^{T-1}\operatorname{Re}\left\{A_{t,r}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)\varphi_3^1(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}})\right\} \\ g_{T,t-r,t-s} &= \frac{-2\kappa_3}{(2\pi)^4}\frac{1}{T^2}\sum_{\mathbf{j}=1}^{T-1}\operatorname{Re}\left\{B_{t,r,s}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)\varphi_3^1(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}})\right\}. \end{aligned}$$

The variance of $T^{1/2}z_{3,t}$ is

$$\begin{aligned} &\sigma^2\left(2\sigma^4+\kappa_4\right)\sum_{r=1}^{t-1}c_{T,t-r}c_{t-r}'+\sigma^2\left(2\sigma^4+\kappa_4\right)\sum_{r=1}^{t-1}g_{T,t-r,t-r}g_{T,t-r,t-r}' \\ &+\sigma^6\sum_{r\neq s=1}^{t-1}\left\{g_{T,t-r,t-s}g_{T,t-r,t-s}'+g_{T,t-r,t-s}g_{T,t-s,t-r}'\right\}+\kappa_3^2\sum_{r=1}^{t-1}\left\{c_{T,t-r}g_{T,t-r,t-r}'+g_{T,t-r,t-r}c_{T,t-r}'\right\} \end{aligned}$$

so that the variance of $\sum_{t=1}^{T}z_{3,t}$ converges to

$$\begin{aligned} V_3 &= \sigma^2\left(2\sigma^4+\kappa_4\right)\sum_{j=0}^{\infty}c_jc_j'+\sigma^2\left(2\sigma^4+\kappa_4\right)\sum_{j=0}^{\infty}g_{j,j}g_{j,j}' \\ &\quad+\sigma^6\sum_{j=0}^{\infty}\sum_{\substack{k=0\\k\neq j}}^{\infty}\left\{g_{j,k}g_{j,k}'+g_{j,k}g_{k,j}'\right\}+\kappa_3^2\sum_{j=0}^{\infty}\left\{c_jg_{j,j}'+g_{j,j}c_j'\right\}, \end{aligned}$$

using that the Fourier coefficients of φ^1, $\varphi_j^1=(2\pi)^{-1}\int_\Pi\exp(-ij\lambda)\,\varphi^1(\theta_0;\lambda)d\lambda=(2\pi)^{-1}\int_\Pi\operatorname{Re}\left\{\exp(ij\lambda)\,\varphi^1(\theta_0;-\lambda)\right\}$ (because the parameters are real and the complex contribution is eliminated by the symmetric integral), decay exponentially fast for ARMA models without unit roots, and

$$\begin{aligned} c_j &= \frac{-2\kappa_3}{(2\pi)^6}\int_{\Pi^2}\mathrm{Re}\left\{\left[\begin{array}{c}\exp\{ij\lambda_1\}+\exp\{ij\lambda_2\}\\ +\exp\{-ij(\lambda_1+\lambda_2)\}\end{array}\right]\varphi_3^1(\theta_0;-\boldsymbol{\lambda})\right\}d\boldsymbol{\lambda} \\ &= \frac{-6\kappa_3}{(2\pi)^4}\varphi_j^1 \\ g_{j,k} &= \frac{-2\kappa_3}{(2\pi)^6}\int_{\Pi^2}\mathrm{Re}\left\{\left[\begin{array}{c}\exp\{i(j\lambda_1+k\lambda_2)\}\\ +\exp\{-i((k-j)\lambda_1+k\lambda_2)\}\\ +\exp\{-i(j\lambda_1+(j-k)\lambda_2)\}\end{array}\right]\varphi_3^1(\theta_0;-\boldsymbol{\lambda})\right\}d\boldsymbol{\lambda} \qquad (38) \\ &= c_{-j}\delta(j=k), \end{aligned}$$

with $c_0=0$ because φ^1 has zero mean and $c_{-j}=0$ for $j>0$ for causal and invertible processes since the Fourier power series of $\varphi_3^1(\theta_0;\lambda_1,\lambda_2)$ has only terms in positive powers of $e^{i\lambda_i}$. Then

$$\sum_{j=-\infty}^{\infty}c_jc_j' = \frac{36\kappa_3^2}{(2\pi)^{10}}2\pi\int_\Pi\varphi^1(\theta_0;\lambda)\varphi^1(\theta_0;-\lambda)'d\lambda = \frac{36\kappa_3^2}{(2\pi)^8}\Phi_0$$

and

$$\sum_{j=-\infty}^{\infty}c_{-j}c_j' = \frac{36\kappa_3^2}{(2\pi)^{10}}2\pi\int_\Pi\varphi^1(\theta_0;\lambda)\varphi^1(\theta_0;\lambda)'d\lambda = \frac{36\kappa_3^2}{(2\pi)^8}\Phi_0^*$$

where $\Phi_0^*=0$ for pure causal and invertible processes.

Therefore

$$\begin{aligned} V_3 &= \sigma^2\left(2\sigma^4+\kappa_4\right)\sum_{j=-\infty}^{\infty}c_jc_j'+\kappa_3^2\sum_{j=-\infty}^{\infty}c_jc_{-j}' \\ &= \frac{36\kappa_3^2}{(2\pi)^8}\left\{\sigma^2\left(2\sigma^4+\kappa_4\right)\Phi_0+\kappa_3^2\Phi_0^*\right\}. \end{aligned}$$

Proof for $\tilde{L}_{2T}^\dagger$. We can write

$$T^{1/2}\frac{\partial}{\partial\theta}\tilde{L}_{2T}^\dagger(\theta_0) = T^{1/2}\frac{-2}{T}\frac{\sigma^2}{(2\pi)^2}\sum_{j=1}^{T-1}\left((2\pi)^2 I_2^\varepsilon(\lambda_j)-\sigma^2\right)\varphi_2^1(\theta_0;-\lambda_j)+o_p(1),$$

where $I_2^\varepsilon(\lambda_j)$ is the innovations (usual) periodogram with

$$\begin{aligned} (2\pi)^2 I_2^\varepsilon(\lambda_j)-\sigma^2 &= \frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t^2-\sigma^2\right) \qquad (39) \\ &\quad +\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\sum_{r=1}^{t-1}\varepsilon_r\left[\exp\{-i((r-t)\lambda_j)\}+\exp\{-i((t-r)\lambda_j)\}\right]. \end{aligned}$$

The first term in (39) does not contribute asymptotically because it is $O_p\left(T^{-1/2}\right)$ and not depending on λ_j, cf. (33).

Therefore we can write

$$T^{1/2}\frac{\partial}{\partial\theta}\tilde{L}_{2T}^\dagger(\theta_0) = \sum_{t=1}^{T}z_{2,t}+o_p(1),$$

where $z_{2,t}$ is a MDS,

$$z_{2,t} = \frac{1}{T^{1/2}} \varepsilon_t \sum_{r=1}^{t-1} \varepsilon_r b_{T,t-r},$$

and

$$b_{T,t-r} = \frac{-2\sigma^2}{(2\pi)^2} \frac{1}{T} \sum_{j=1}^{T-1} \mathrm{Re} \left\{ \left[\begin{array}{c} \exp\{-i((r-t)\lambda_j)\} \\ +\exp\{-i((t-r)\lambda_j)\} \end{array} \right] \varphi_2^1(\theta_0; -\lambda_j) \right\}$$

so that the variance of $T^{1/2} z_{2,t}$ is

$$\sigma^4 \sum_{r=1}^{t-1} b_{t-r} b'_{t-r}$$

and the variance of $\sum_{t=1}^{T} z_{2,t}$ converges to $V_2 = \sigma^4 \sum_{j=0}^{\infty} b_j b'_j$ where

$$\begin{aligned} b_j &= \frac{-2\sigma^2}{(2\pi)^3} \int_\Pi [\exp\{ij\lambda\} + \exp\{-ij\lambda\}] \varphi_2^1(\theta_0; -\lambda) d\lambda \\ &= \frac{-4\sigma^2}{(2\pi)^3} \int_\Pi [\exp\{ij\lambda\} + \exp\{-ij\lambda\}] \varphi^1(\theta_0; -\lambda) d\lambda \\ &= \frac{-4\sigma^2}{(2\pi)^2} \left\{ \varphi_j^1 + \varphi_{-j}^1 \right\}, \end{aligned}$$

$b_0 = 0$, so that,

$$V_2 = \frac{16\sigma^8}{(2\pi)^6} (2\pi)^2 \sum_{j=-\infty}^{\infty} \left(\varphi_j^1 \varphi_j^{1\prime} + \varphi_j^1 \varphi_{-j}^{1\prime} \right) = \frac{16\sigma^8}{(2\pi)^4} \left(\Phi_0 + \Phi_0^* \right).$$

Proof for $\tilde{L}_{4T}^{\dagger}$. We can write

$$T^{1/2} \frac{\partial}{\partial \theta} \tilde{L}_{4T}^{\dagger}(\theta_0) = T^{1/2} \frac{-2}{T^3} \frac{\kappa_4}{(2\pi)^6} \sum_{\mathbf{j}=\mathbf{1}}^{T-1} \mathrm{Re} \left\{ \left((2\pi)^3 I_4^{\varepsilon}(\boldsymbol{\lambda}_{\mathbf{j}}) - \kappa_4 \right) \varphi_4^1(\theta_0; -\boldsymbol{\lambda}_{\mathbf{j}}) \right\} + o_p(1),$$

where $\boldsymbol{\lambda}_{\mathbf{j}} = (\lambda_{j_1}, \lambda_{j_2}, \lambda_{j_3})$ and $I_4^{\varepsilon}(\boldsymbol{\lambda}_{\mathbf{j}})$ is the innovations triperiodogram and

$$\begin{aligned} (2\pi)^3 I_4^{\varepsilon}(\boldsymbol{\lambda}_{\mathbf{j}}) - \kappa_4 &= \frac{1}{T} \sum_{t=1}^{T} \left(\varepsilon_t^4 - \kappa_4 \right) + \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^3 \sum_{r=1}^{t-1} \varepsilon_r C_{t,r}(\boldsymbol{\lambda}_{\mathbf{j}}) \\ &+ \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 \sum_{r,s=1}^{t-1} \varepsilon_r \varepsilon_s D_{t,r,s}(\boldsymbol{\lambda}_{\mathbf{j}}) + \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t \sum_{r,s,u=1}^{t-1} \varepsilon_r \varepsilon_s \varepsilon_u F_{t,r,s,u}(\boldsymbol{\lambda}_{\mathbf{j}}), \end{aligned} \tag{40}$$

where

$$C_{t,r}(\boldsymbol{\lambda}_{\mathbf{j}}) = \left[\begin{array}{c} \exp\{-i((r-t)\lambda_{j_1})\} + \exp\{-i((r-t)\lambda_{j_2})\} \\ +\exp\{-i((r-t)\lambda_{j_3})\} + \exp\{-i((t-r)(\lambda_{j_1}+\lambda_{j_2}+\lambda_{j_3}))\} \end{array}\right]$$

$$D_{t,r,s}(\boldsymbol{\lambda}_{\mathbf{j}}) = \left[\begin{array}{c} \exp\{-i((t-s)\lambda_{j_1}+(t-s)\lambda_{j_2}+(r-s)\lambda_{j_3})\} \\ +\exp\{-i((t-s)\lambda_{j_1}+(r-s)\lambda_{j_2}+(t-s)\lambda_{j_3})\} \\ +\exp\{-i((r-t)\lambda_{j_2}+(s-t)\lambda_{j_3})\} \\ +\exp\{-i((r-s)\lambda_{j_1}+(t-s)\lambda_{j_2}+(t-s)\lambda_{j_3})\} \\ +\exp\{-i((r-t)\lambda_{j_1}+(s-t)\lambda_{j_3})\} \\ +\exp\{-i((r-t)\lambda_{j_1}+(s-t)\lambda_{j_2})\} \end{array}\right]$$

$$F_{t,r,s,u}(\boldsymbol{\lambda}_{\mathbf{j}}) = \left[\begin{array}{c} \exp\{-i((t-u)\lambda_{j_1}+(r-u)\lambda_{j_2}+(s-u)\lambda_{j_3})\} \\ \exp\{-i((r-u)\lambda_{j_1}+(t-u)\lambda_{j_2}+(s-u)\lambda_{j_3})\} \\ \exp\{-i((s-u)\lambda_{j_1}+(r-u)\lambda_{j_2}+(t-u)\lambda_{j_3})\} \\ \exp\{-i((u-t)\lambda_{j_1}+(r-t)\lambda_{j_2}+(s-t)\lambda_{j_3})\} \end{array}\right].$$

The first term in (40) can be written as

$$\frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t^4 - 3\sigma^4 - \kappa_4\right) + 3\sigma^4, \tag{41}$$

whose first element does not contribute asymptotically because it is $O_p\left(T^{-1/2}\right)$ and not depending on $\boldsymbol{\lambda}_{\mathbf{j}}$.

We can write the second term in (40) as

$$\begin{aligned} & \frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t^3-\kappa_3\right)\sum_{r=1}^{t-1}\varepsilon_r C_{t,r}(\boldsymbol{\lambda}_{\mathbf{j}}) + \frac{\kappa_3}{T}\sum_{t=1}^{T}\sum_{r=1}^{t-1}\varepsilon_r C_{t,r}(\boldsymbol{\lambda}_{\mathbf{j}}) \\ = & \frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t^3-\kappa_3\right)\sum_{r=1}^{t-1}\varepsilon_r C_{t,r}(\boldsymbol{\lambda}_{\mathbf{j}}) + \frac{\kappa_3}{T}\sum_{t=1}^{T}\varepsilon_t\sum_{r=t+1}^{T}C_{r,t}(\boldsymbol{\lambda}_{\mathbf{j}}), \end{aligned} \tag{42}$$

where the summands of the two terms are MDS.

The third term in (40) is

$$\begin{aligned} & \frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t^2-\sigma^2\right)\sum_{r,s=1}^{t-1}\left(\varepsilon_r\varepsilon_s-\sigma^2\delta_{r=s}\right)D_{t,r,s}(\boldsymbol{\lambda}_{\mathbf{j}}) \\ & +\frac{\sigma^2}{T}\sum_{t=1}^{T}\sum_{r,s=1}^{t-1}\left(\varepsilon_r\varepsilon_s-\sigma^2\delta_{r=s}\right)D_{t,r,s}(\boldsymbol{\lambda}_{\mathbf{j}}) \\ & +\frac{\sigma^2}{T}\sum_{t=1}^{T}\left(\varepsilon_t^2-\sigma^2\right)\sum_{r=1}^{t-1}D_{t,r,r}(\boldsymbol{\lambda}_{\mathbf{j}}) + \frac{\sigma^4}{T}\sum_{t=1}^{T}\sum_{r=1}^{t-1}D_{t,r,r}(\boldsymbol{\lambda}_{\mathbf{j}}). \end{aligned} \tag{43}$$

where the first element in (43) is an average of MDS, and discarding the cases when $\lambda_{j_a}+\lambda_{j_b}=T$, $a \neq b$, the second element in (43) can be further decomposed as

$$\begin{aligned} & \frac{\sigma^2}{T}\sum_{t=1}^{T}\sum_{r=1}^{t-1}\left(\varepsilon_t^2-\sigma^2\right)D_{t,r,r}(\boldsymbol{\lambda}_{\mathbf{j}}) + \frac{\sigma^2}{T}\sum_{t=1}^{T}\sum_{r=1}^{t-1}\varepsilon_r\sum_{s\neq r=1}^{t-1}\varepsilon_s D_{t,r,s}(\boldsymbol{\lambda}_{\mathbf{j}}) \\ = & \frac{\sigma^2}{T}\sum_{t=1}^{T}\left(\varepsilon_r^2-\sigma^2\right)\sum_{r=t+1}^{T}D_{t,r,r}(\boldsymbol{\lambda}_{\mathbf{j}}) + \left\{\frac{\sigma^2}{T}\sum_{t=1}^{T}\left\{\sum_{r=1}^{t-1}\varepsilon_r\sum_{s=1}^{r-1}\varepsilon_s + \sum_{s=1}^{t-1}\varepsilon_s\sum_{r=1}^{s-1}\varepsilon_r\right\}\right\}D_{t,r,s}(\boldsymbol{\lambda}_{\mathbf{j}}) \end{aligned} \tag{44}$$

$$= -\frac{3\sigma^2}{T}\sum_{t=1}^{T}\left(\varepsilon_t^2-\sigma^2\right)+\frac{\sigma^2}{T}\sum_{t=1}^{T}\varepsilon_t\sum_{s=1}^{t-1}\varepsilon_s\sum_{r=t+1}^{T}\left\{D_{r,t,s}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)+D_{r,s,t}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)\right\}$$

because $\sum_{r=1}^{T} D_{t,r,r}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)=0$ and $D_{t,r,r}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)=D_{r,t,t}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)$, so the first term in (44) is not going to contribute as does not depend on λ_i and is $O_p\left(T^{-1/2}\right)$ while the second is an average of a MDS.

Using the same reasoning the third element in (43) is

$$\frac{\sigma^2}{T}\sum_{t=1}^{T}\left(\varepsilon_t^2-\sigma^2\right)\sum_{\substack{r=1\\ r\neq t}}^{T} D_{t,r,r}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)=-\frac{3\sigma^2}{T}\sum_{t=1}^{T}\left(\varepsilon_t^2-\sigma^2\right)$$

and does not contribute because is $O_p\left(T^{-1/2}\right)$ and not depending on λ_{j_i} since we have discarded that $j_1+j_2=T$ or $j_2+j_3=T$ or $j_2+j_3=T$.

Finally the fourth term in (43) is

$$\frac{\sigma^4}{T}\sum_{t=1}^{T}\sum_{\substack{r=1\\ r\neq t}}^{T} D_{t,r,r}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)=-\frac{\sigma^4}{T}\sum_{t=1}^{T}3=-3\sigma^4,$$

which cancels out with the second element in (41) corresponding to the first term of (40).

The fourth term in (40) can be written as

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\sum_{r,s,u=1}^{t-1}\left(\varepsilon_r\varepsilon_s\varepsilon_u-\kappa_3\delta_{r=s=u}\right)F_{t,r,s,u}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)+\frac{\kappa_3}{T}\sum_{t=1}^{T}\varepsilon_t\sum_{r=1}^{t-1}F_{t,r,r,r}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right),$$

where the first term is MDS, and, noting that $F_{t,r,s,u}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)=C_{r,t}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)$ the second term plus the second term in (42) is equal to

$$-\frac{\kappa_3}{T}\sum_{t=1}^{T}\varepsilon_t,$$

which does not contribute since it does not depend on λ_{j_i} and is $O_p\left(T^{-1/2}\right)$.

Therefore we can write

$$T^{1/2}\frac{\partial}{\partial\theta}\tilde{L}_{4T}^{\dagger}(\theta_0)=\sum_{t=1}^{T}z_{4,t}+o_p(1),$$

where $z_{4,t}$ is a MDS with

$$\begin{aligned}
z_{4,t}T^{1/2} \quad = \quad & \left(\varepsilon_t^3-\kappa_3\right)\sum_{r=1}^{t-1}\varepsilon_r c_{T,t-r}^{(4)}+\left(\varepsilon_t^2-\sigma^2\right)\sum_{r,s=1}^{t-1}\left(\varepsilon_r\varepsilon_s-\sigma^2\delta_{r=s}\right)g_{T,t-r,t-s}^{(4)}\\
&+\sigma^2\varepsilon_t\sum_{s=1}^{t-1}\varepsilon_s\sum_{r=t+1}^{T}d_{T,t-r,t-s}^{(4)}+\varepsilon_t\sum_{r,s,u=1}^{t-1}\left(\varepsilon_r\varepsilon_s\varepsilon_u-\kappa_3\delta_{r=s=u}\right)b_{T,t-r,t-s,t-u}^{(4)}
\end{aligned}$$

and

$$
\begin{aligned}
c_{T,t-r}^{(4)} &= \frac{-2}{T^3}\frac{\kappa_4}{(2\pi)^6}\sum_{\mathbf{j}=\mathbf{1}}^{T-1}\mathrm{Re}\left\{C_{t,r}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)\varphi_4^1(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}})\right\}\\
g_{T,t-r,t-s}^{(4)} &= \frac{-2}{T^3}\frac{\kappa_4}{(2\pi)^6}\sum_{\mathbf{j}=\mathbf{1}}^{T-1}\mathrm{Re}\left\{D_{t,r,s}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)\varphi_4^1(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}})\right\}\\
d_{T,t-r,t-s}^{(4)} &= \frac{-2}{T^3}\frac{\kappa_4}{(2\pi)^6}\sum_{\mathbf{j}=\mathbf{1}}^{T-1}\mathrm{Re}\left\{\left[D_{r,t,s}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)+D_{r,s,t}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)\right]\varphi_4^1(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}})\right\}\\
b_{T,t-r,t-s,t-u}^{(4)} &= \frac{-2}{T^3}\frac{\kappa_4}{(2\pi)^6}\sum_{\mathbf{j}=\mathbf{1}}^{T-1}\mathrm{Re}\left\{F_{t,r,s,u}\left(\boldsymbol{\lambda}_{\mathbf{j}}\right)\varphi_4^1(\theta_0;-\boldsymbol{\lambda}_{\mathbf{j}})\right\},
\end{aligned}
$$

so that the variance of $T^{1/2}z_{4,t}$ is

$$
\begin{aligned}
&\left(\mu_6-\kappa_3^2\right)\sigma^2\sum_{r=1}^{t-1}c_{T,t-r}^{(4)}c_{T,t-r}^{(4)\prime}+\left(2\sigma^4+\kappa_4\right)\left\{\begin{array}{c}\left(2\sigma^4+\kappa_4\right)\sum_{r=1}^{t-1}g_{T,t-r,t-r}^{(4)}g_{T,t-r,t-r}^{(4)\prime}\\ +\sigma^4\sum_{r\neq s=1}^{t-1}\left[g_{T,t-r,t-s}^{(4)}g_{T,t-r,t-s}^{(4)\prime}+g_{T,t-r,t-s}^{(4)}g_{T,t-s,t-r}^{(4)\prime}\right]\end{array}\right\}\\
&+\sigma^8\sum_{s=1}^{t-1}\sum_{r=t+1}^{T}\sum_{r'=t+1}^{T}d_{T,t-r,t-s}^{(4)}d_{T,t-r',t-s}^{(4)}\\
&+\sigma^2\sum_{r,s,u=1}^{t-1}\sum_{r',s',u'=1}^{t-1}E\left[\left(\varepsilon_r\varepsilon_s\varepsilon_u-\kappa_3\delta_{r=s=u}\right)\left(\varepsilon_{r'}\varepsilon_{s'}\varepsilon_{u'}-\kappa_3\delta_{r'=s'=u'}\right)\right]b_{T,t-r,t-s,t-u}^{(4)}b_{T,t-r',t-s',t-u'}^{(4)}\\
&+\left(3\sigma^4+\kappa_4\right)\sum_{a=1}^{t-1}\sum_{r,s,u=1}^{t-1}E\left[\varepsilon_a\left(\varepsilon_r\varepsilon_s\varepsilon_u-\kappa_3\delta_{r=s=u}\right)\right]\left\{c_{T,t-a}^{(4)}b_{T,t-r,t-s,t-u}^{(4)\prime}+b_{T,t-r,t-s,t-u}^{(4)}c_{T,t-a}^{(4)\prime}\right\}
\end{aligned}
$$

where $\mu_6=\kappa_6+15\kappa_4\sigma^2+9\kappa_3^2+15\sigma^6$ and the variance of $\sum_{t=1}^{T}z_t^{(4)}$ converges to

$$
\begin{aligned}
V_4 &= \left(\mu_6-\kappa_3^2\right)\sigma^2\sum_{j=0}^{\infty}c_j^{(4)}c_j^{(4)\prime}+\left(\mu_6-\kappa_3^2\right)\sigma^2\sum_{j=0}^{\infty}b_{j,j,j}^{(4)}b_{j,j,.j}^{(4)\prime}+\left(3\sigma^4+\kappa_4\right)^2\sum_{j=0}^{\infty}\left\{c_j^{(4)}b_{j,j,j}^{(4)\prime}+b_{j,j,j}^{(4)}c_j^{(4)\prime}\right\}\\
&= \left(\mu_6-\kappa_3^2\right)\sigma^2\sum_{j=-\infty}^{\infty}c_j^{(4)}c_j^{(4)\prime}+\left(3\sigma^4+\kappa_4\right)^2\sum_{j=-\infty}^{\infty}c_j^{(4)}c_{-j}^{(4)\prime}\\
&= \left(\mu_6-\kappa_3^2\right)\sigma^2\frac{64\kappa_4^2}{(2\pi)^{12}}\Phi_0+\left(3\sigma^4+\kappa_4\right)^2\frac{64\kappa_4^2}{(2\pi)^{12}}\Phi_0^*,
\end{aligned}
$$

where $c_j^{(4)}=0$ for $j<0$ for invertible processes, $c_0^{(4)}=0$,

$$
\begin{aligned}
c_j^{(4)} &= \frac{-2\kappa_4}{(2\pi)^9}\int_{\Pi^3}\mathrm{Re}\left\{\left[\begin{array}{c}\exp\left\{ij\lambda_1\right\}+\exp\left\{ij\lambda_2\right\}+\exp\left\{ij\lambda_3\right\}\\ +\exp\left\{-ij\left(\lambda_1+\lambda_2+\lambda_3\right)\right\}\end{array}\right]\varphi_4^1(\theta_0;-\boldsymbol{\lambda})\right\}d\boldsymbol{\lambda}\\
&= \frac{-8\kappa_4}{(2\pi)^7}\int_{\Pi}\mathrm{Re}\left\{\exp\left(ij\lambda\right)\varphi^1(\theta_0;-\lambda)\right\}d\lambda\\
&= \frac{-8\kappa_4}{(2\pi)^6}\varphi_j^1\\
b_{j,k,\ell}^{(4)} &= \frac{-2\kappa_4}{(2\pi)^9}\int_{\Pi^3}\mathrm{Re}\left\{\left[\begin{array}{c}\exp\left\{-i\left(\ell\lambda_1+(\ell-j)\lambda_2+(\ell-k)\lambda_3\right)\right\}\\ +\exp\left\{-i\left((\ell-j)\lambda_1+\ell\lambda_2+(\ell-k)\lambda_3\right)\right\}\\ +\exp\left\{-i\left((\ell-k)\lambda_1+(\ell-j)\lambda_2+\ell\lambda_3\right)\right\}\\ +\exp\left\{-i\left(-\ell\lambda_1-j\lambda_2-k\lambda_3\right)\right\}\end{array}\right]\varphi_4^1(\theta_0;-\boldsymbol{\lambda})\right\}d\boldsymbol{\lambda}\\
&= c_{-j}^{(4)}*\delta\left(j=k=\ell\right),
\end{aligned}
$$

while the limits of $g_{T,j,k}^{(4)}$ and $d_{T,j,k}^{(4)}$ are, respectively,

$$\begin{aligned}
g_{j,k}^{(4)} &= \frac{-2\kappa_4}{(2\pi)^9}\int_{\Pi^3}\mathrm{Re}\left\{\left[\begin{array}{c}\exp\{-i(k\lambda_1+k\lambda_2+(k-j)\lambda_3)\}\\ +\exp\{-i(k\lambda_1+(k-j)\lambda_2+k\lambda_3)\}\\ +\exp\{-i(-j\lambda_2-k\lambda_3)\}\\ +\exp\{-i((k-j)\lambda_1+k\lambda_2+k\lambda_3)\}\\ +\exp\{-i(-j\lambda_1-k\lambda_3)\}\\ +\exp\{-i(-j\lambda_1-k\lambda_2)\}\end{array}\right]\varphi_4^1(\theta_0;-\boldsymbol{\lambda})\right\}d\boldsymbol{\lambda}\\
&= \frac{-2\kappa_4}{(2\pi)^9}\int_{\Pi^3}\mathrm{Re}\left\{\left[\begin{array}{c}\exp\{-i(k\lambda_1+k\lambda_2+(k-j)\lambda_3)\}\\ +\exp\{-i(k\lambda_1+(k-j)\lambda_2+k\lambda_3)\}\\ +\exp\{-i((k-j)\lambda_1+k\lambda_2+k\lambda_3)\}\end{array}\right]\varphi_4^1(\theta_0;-\boldsymbol{\lambda})\right\}d\boldsymbol{\lambda}\\
&= 0 \ \text{ if } j,k\neq 0,
\end{aligned}$$

$$\begin{aligned}
d_{j,k}^{(4)} &= \frac{-2\kappa_4}{(2\pi)^9}\int_{\Pi^3}\mathrm{Re}\left\{\begin{array}{c}\left[\begin{array}{c}\exp\{-i((k-j)\lambda_1+(k-j)\lambda_2+k\lambda_3)\}\\ +\exp\{-i((k-j)\lambda_1+k\lambda_2+(k-j)\lambda_3)\}\\ +\exp\{-i(j\lambda_2+(j-k)\lambda_3)\}\\ +\exp\{-i(k\lambda_1+(k-j)\lambda_2+(k-j)\lambda_3)\}\\ +\exp\{-i(j\lambda_1+(j-k)\lambda_3)\}\\ +\exp\{-i(j\lambda_1+k\lambda_2)\}\end{array}\right]\\ +\left[\begin{array}{c}\exp\{-i(-j\lambda_1-j\lambda_2-k\lambda_3)\}\\ +\exp\{-i(-j\lambda_1-k\lambda_2-j\lambda_3)\}\\ +\exp\{-i((j-k)\lambda_2+j\lambda_3)\}\\ +\exp\{-i(-k\lambda_1-j\lambda_2 j\lambda_3)\}\\ +\exp\{-i((j-k)\lambda_1+j\lambda_3)\}\\ +\exp\{-i((j-k)\lambda_1+j\lambda_2)\}\end{array}\right]\end{array}\right\}\varphi_4^1(\theta_0;-\boldsymbol{\lambda})d\boldsymbol{\lambda}\\
&= 0 \text{ if } \ j,k\neq 0
\end{aligned}$$

so these terms do not contribute, and

$$\begin{aligned}
\sum_{j=-\infty}^{\infty}c_j^{(4)}c_j^{(4)\prime} &= \frac{64\kappa_4^2}{(2\pi)^{12}}\sum_{j=-\infty}^{\infty}\varphi_j\varphi_j' = \frac{64\kappa_4^2}{(2\pi)^{12}}\Phi_0\\
\sum_{j=-\infty}^{\infty}c_j^{(4)}c_{-j}^{(4)\prime} &= \frac{64\kappa_4^2}{(2\pi)^{12}}\sum_{j=-\infty}^{\infty}\varphi_j\varphi_{-j}' = \frac{64\kappa_4^2}{(2\pi)^{12}}\Phi_0^*.
\end{aligned}$$

The covariance terms, $V_{j,k} = \lim_{T\to\infty} Cov\left(\sum_{t=1}^T z_{j,t}, \sum_{t=1}^T z_{k,t}\right)$, $j\neq k$, can be obtained by similar arguments,

$$\begin{aligned}
V_{2,3} &= \kappa_3\sigma^2\sum_{t=1}^{T}\sum_{r=1}^{t-1}\left\{b_{T,t-r}c_{T,t-r}' + b_{T,t-r}g_{T,t-r,t-r}'\right\} + o(1)\\
&= \frac{24}{(2\pi)^6}\left(\sigma^2\kappa_3\right)^2(\Phi_0+\Phi_0^*) + o(1).
\end{aligned}$$

Next, in a similar way,

$$\begin{aligned}
V_{2,4} &= \mu_4\sigma^2\sum_{t=1}^{T}\sum_{r=1}^{t-1}\left\{b_{T,t-r}c_{T,t-r}^{(4)\prime} + b_{T,t-r}b_{T,r-t}^{(4)\prime}\right\} + o(1)\\
&= \frac{32}{(2\pi)^8}\sigma^4\mu_4\kappa_4(\Phi_0+\Phi_0^*) + o(1),
\end{aligned}$$

and

$$\begin{aligned}
V_{3,4} &= \sum_{t=1}^{T}\sum_{r=1}^{t-1}\left\{\begin{array}{c}\sigma^2\left(\mu_5-\kappa_3\sigma^2\right)\left\{c_{T,t-r}^{(3)}c_{T,t-r}^{(4)\prime}+g_{T,t-r,t-r}^{(3)}b_{T,t-r,t-r,t-r}^{(4)\prime}\right\}\\ +\mu_4\kappa_3\left\{c_{T,t-r}^{(3)}b_{T,t-r,t-r,t-r}^{(4)\prime}+g_{T,t-r,t-r}^{(3)}c_{T,t-r}^{(4)\prime}\right\}\end{array}\right\}+o\left(1\right)\\
&= \frac{48\kappa_3\kappa_4}{(2\pi)^{10}}\left\{\sigma^2\left(\mu_5-\kappa_3\sigma^2\right)\Phi_0+\mu_4\kappa_3\Phi_0^*\right\}+o\left(1\right).
\end{aligned}$$

Collecting all terms in the covariance matrix of $z_t = \left(z_{2,t}^{\prime}, z_{3,t}^{\prime}, z_{4,t}^{\prime}\right)^{\prime}$ we obtain the asymptotic variance of $\frac{\partial}{\partial\theta}\tilde{L}_{wT}(\theta_0)$ and V after adjustment by the factors in the defintion of $\tilde{L}_{kT}$.

To complete the proof of the martingale central limit theorem for the vector z_t we have to check that for any linear combination of elements of z_t,

$$\sum_{t=1}^{T}E\left[\left.\left(\sum_{a}p_a z_{a,t}\right)^2\right|\mathcal{F}_{t-1}\right] \rightarrow_p \sum_{a,b}p_a p_b V_{a,b} \tag{45}$$

$$\sum_{t=1}^{T}E\left[\left(\sum_{a}p_a z_{a,t}\right)^2 1\left\{\left(\sum_{a}p_a z_{a,t}\right)^2>\delta\right\}\right] \rightarrow 0 \text{ for all } \delta>0. \tag{46}$$

to obtain the asymptotic normality of the parameter estimates, where $\mathcal{F}_{t-1}=\sigma\left\{z_{t-1},z_{t-2},\ldots\right\}$. We give the proof just for $z_{3,t}$, pretending that is scalar to simplify notation. First we note that

$$\begin{aligned}
\sum_{t=1}^{T}E\left[\left.z_{3,t}^2\right|\mathcal{F}_{t-1}\right] &= T^{-1}\left(2\sigma^4+\kappa_4\right)\sum_{t=1}^{T}\left(\sum_{r=1}^{t-1}\varepsilon_r c_{T,t-r}\right)^2 \qquad (47)\\
&\quad +\sigma^2 T^{-1}\sum_{t=1}^{T}\left(\sum_{r,s=1}^{t-1}\left(\varepsilon_r\varepsilon_s-\sigma^2\left\{r=s\right\}\right)g_{T,t-r,t-s}\right)^2\\
&\quad +2\kappa_3 T^{-1}\sum_{t=1}^{T}\sum_{a=1}^{t-1}\varepsilon_a c_{T,t-a}\left(\sum_{r,s=1}^{t-1}\left(\varepsilon_r\varepsilon_s-\sigma^2\left\{r=s\right\}\right)g_{T,t-r,t-s}\right),
\end{aligned}$$

where the expectation of the rhs of (47) converges to V_3. Then, checking the variance of each of the three terms in $\sum_{t=1}^{T}E\left[\left.z_{3,t}^2\right|\mathcal{F}_{t-1}\right]$ from (47) we find that the variance of the first one is, up to a constant, equal to

$$\begin{aligned}
&E\left[T^{-2}\left(\sum_{t=1}^{T}\left(\sum_{r=1}^{t-1}\varepsilon_r c_{T,t-r}\right)^2\right)^2\right]-E\left[T^{-1}\sum_{t=1}^{T}\left(\sum_{r=1}^{t-1}\varepsilon_r c_{T,t-r}\right)^2\right]^2 \qquad (48)\\
&= 2\sigma^4 T^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{r=1}^{t-1\wedge t'-1}c_{T,t-r}c_{T,t'-r}\sum_{s=1}^{t-1\wedge t'-1}c_{T,t-s}c_{T,t'-s}\\
&\quad +\left(3\sigma^4+\kappa_4\right)T^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{r=1}^{t-1\wedge t'-1}c_{T,t-r}^2 c_{T,t'-r}^2.
\end{aligned}$$

Now the first term in the rhs (48) is not larger than

$$2\sigma^4\sup_{T,j}\left\|c_{T,j}\right\|T^{-2}\sum_{r=1}^{T}\sum_{t=1}^{T}\left\|c_{T,t}\right\|\sum_{t'=-T}^{T}\left\|c_{T,t'}\right\|\sum_{s=-T}^{T}\left\|c_{T,s}\right\|=O\left(T^{-1}\right)$$

because $\sum_{s=-T}^{T}\|c_{T,s}\| = \sum_{s=-\infty}^{\infty}\|c_s\| + O(1) < \infty$, and the second term in (48) can be showed similarly to be $O\left(T^{-1}\right)$.

The variance of the second term on the rhs of (47) is, up to a constant,

$$E\left[T^{-2}\left\{\sum_{t=1}^{T}\left(\sum_{r,s=1}^{t-1}\left(\varepsilon_r\varepsilon_s-\sigma^2\{r=s\}\right)g_{T,t-r,t-s}\right)^2\right\}^2\right]-E\left[T^{-1}\sum_{t=1}^{T}\left(\sum_{r,s=1}^{t-1}\left(\varepsilon_r\varepsilon_s-\sigma^2\{r=s\}\right)g_{T,t-r,t-s}\right)^2\right]^2$$

where the first term is

$$\begin{aligned}&T^{-2}\sum_{t,t'=1}^{T}\sum_{r,s=1}^{t-1}\sum_{a,b=1}^{t-1}\sum_{r',s'=1}^{t'-1}\sum_{a',b'=1}^{t'-1}g_{T,t-r,t-s}g_{T,t-a,t-b}g_{T,t'-r',t'-s'}g_{T,t'-a',t'-b'}\\&\times E\left[\left(\varepsilon_r\varepsilon_s-\sigma^2_{r=s}\right)\left(\varepsilon_a\varepsilon_b-\sigma^2_{a=b}\right)\left(\varepsilon_{r'}\varepsilon_{s'}-\sigma^2_{r'=s'}\right)\left(\varepsilon_{a'}\varepsilon_{b'}-\sigma^2_{a'=b'}\right)\right],\end{aligned}$$

and the expectation introduces at least 4 restrictions on the 8 indexes r,s,a,b,r',s',a',b' because to contribute to the variance they must involve simultaneously primed and not primed indexes because of the substraction of the expectation squared. Then these eight summations are reduced to at most four, such as in the particular example where $r=r', s=s', a=a', b=b'$, which leads to a term whose contribution is bounded for some positive $C<\infty$ by

$$\begin{aligned}&\sigma^8T^{-2}\sum_{t,t'=1}^{T}\sum_{r,s=1}^{t-1\wedge t'-1}\sum_{a,b=1}^{t-1\wedge t'-1}\|g_{T,t-r,t-s}\|\,\|g_{T,t-a,t-b}\|\,\|g_{T,t'-r,t'-s}\|\,\|g_{T,t'-a,t'-b}\|\\ \leq\;& C\sigma^8T^{-2}\sum_{t,t'=1}^{T}\sum_{r=1}^{t-1\wedge t'-1}\sum_{a=1}^{t-1\wedge t'-1}\|c_{r-t}\|\,\|c_{a-t}\|\,\|c_{r-t'}\|\,\|c_{a-t'}\| = O\left(T^{-1}\right),\end{aligned}$$

because writing $\varphi_3^1(\theta_0;-\lambda_1,-\lambda_2) = (2\pi)^{-2}\sum_{k_1,k_2=-\infty}^{\infty}\exp\{i(k_1\lambda_{j_1}+k_2\lambda_{j_2})\}\varphi^1_{3,k_1,k_2}$ and noting that up to a multiplicative constant,

$$\begin{aligned}g_{T,j,k} &= \frac{1}{(2\pi)^2T^2}\sum_{j_1,j_2=1}^{T-1}\mathrm{Re}\left\{\begin{bmatrix}\exp\{i(j\lambda_{j_1}+k\lambda_{j_2})\}\\+\exp\{-i((k-j)\lambda_{j_1}+k\lambda_{j_2})\}\\+\exp\{-i(j\lambda_{j_1}+(j-k)\lambda_{j_2})\}\end{bmatrix}\sum_{k_1,k_2=-\infty}^{\infty}\exp\{i(k_1\lambda_{j_1}+k_2\lambda_{j_2})\}\varphi^1_{3,k_1,k_2}\right\}\\
&= \frac{(T-1)^2}{T^2}\sum_{k_1,k_2=-\infty}^{\infty}\varphi^1_{3,k_1,k_2}\left\{\begin{bmatrix}\delta(k_1=-j\bmod T,k_2=-k\bmod T)\\+\delta(k_1=(k-j)\bmod T,k_2=k\bmod T)\\+\delta(k_1=j\bmod T,k_2=(j-k)\bmod T)\end{bmatrix}\right\}\\
&= \frac{(T-1)^2}{T^2}\left\{\begin{bmatrix}\left(\varphi^1_{3,-j,-k}+\varphi^1_{3,-j\pm T,k}+\varphi^1_{3,-j,-k\pm T}+\varphi^1_{3,-j\pm T,-k\pm T}+\cdots\right)\\+\left(\varphi^1_{3,k-j,k}+\varphi^1_{3,k-j\pm T,k}+\varphi^1_{3,k-j,k\pm T}+\varphi^1_{3,k-j\pm T,-k\pm T}+\cdots\right)\\+\left(\varphi^1_{3,j,j-k}+\varphi^1_{3,j\pm T,j-k}+\varphi^1_{3,j,j-k\pm T}+\varphi^1_{3,j\pm T,j-k\pm T}+\cdots\right)\end{bmatrix}\right\}\\
&= \frac{(T-1)^2}{T^2}\left(\varphi^1_{3,-j,-j}+\varphi^1_{3,-j+T,-j+T}+\varphi^1_{3,-j-T,-j-T}+\cdots\right)\delta(j=k)\\
&= \frac{(T-1)^2}{T^2}\left(\varphi^1_{3,-j,-j}+\varphi^1_{3,-j+T,-j+T}+\varphi^1_{3,-j-T,-j-T}\right)\left(1+O\left(T^{-1}\right)\right)\delta(j=k)\\
&= O(c_{-j})\,\delta(j=k)\end{aligned}$$

uniformly for $|j|,|k|<T$, since from (38) $\varphi^1_{3,j,j}$ is proportional to $g_{-j,-j}=c_j$, $g_{jk}=0$ for $j\neq k$ and $\|c_j\|$ decays in j exponentially fast since φ^1_3 is analytic.

The variance of the third term in (47) can be showed to be also $O\left(T^{-1}\right)$, so (45) follows.

To check the second condition (46), we just check the sufficient condition

$$\sum_{t=1}^{T} E\left[z_{3,t}^4\right] < \infty.$$

Just considering the first term of $z_{3,t}$, its contribution to $\sum_{t=1}^{T} E\left[z_{3,t}^4\right]$ is

$$\begin{aligned}
& \frac{1}{T^2}\sum_{t=1}^{T} E\left[\left(\varepsilon_t^2-\sigma^2\right)^4\right]\sum_{r_1=1}^{t-1}\cdots\sum_{r_4=1}^{t-1} E\left[\varepsilon_{r_1}c_{T,t-r_1}\cdots\varepsilon_{r_4}c_{T,t-r_4}\right] \\
= \;& \frac{3\sigma^4 E\left[\left(\varepsilon_t^2-\sigma^2\right)^4\right]}{T^2}\sum_{t=1}^{T}\sum_{r_1=1}^{t-1}\sum_{r_2=1}^{t-1} c_{T,t-r_1}^2 c_{T,t-r_2}^2 + \frac{\kappa_4 E\left[\left(\varepsilon_t^2-\sigma^2\right)^4\right]}{T^2}\sum_{t=1}^{T}\sum_{r_1=1}^{t-1} c_{T,t-r_1}^4 \\
= \;& O\left(T^{-1}\right)
\end{aligned}$$

and a similar proof holds for the other components in $z_{3,t}^4$, and in z_t, completing the proof of the theorem. □

Proof of Corollary 1. It follows from the proof of Theorem 2 noting that from (23) and (24) the asymptotic variance of $\theta_T^{(3)}$ is

$$\begin{aligned}
\left(6\frac{\kappa_3^2}{(2\pi)^4}\Phi_0\right)^{-1} V_3 \left(6\frac{\kappa_3^2}{(2\pi)^4}\Phi_0\right)^{-1} &= \frac{\sigma^6}{\kappa_3^2}\left\{2+\nu_4\right\}\Phi_0^{-1} + \Phi_0^{-1}\Phi_0^*\Phi_0^{-1} \\
&= \left\{\frac{2+\nu_4}{\nu_3^2}\right\}\Phi_0^{-1} + \Phi_0^{-1}\Phi_0^*\Phi_0^{-1},
\end{aligned}$$

$2+\nu_4 = \bar{\mu}_4 - 1$, which is positive definite noting that Φ_0 is positive definite, Φ_0^* is positive semidefinite, $\nu_4 \geq -2$ and $\nu_3^2 > 0$ by assumption. Then the asymptotic variance of $\theta_T^{(2)}$ follows as

$$\left(4\frac{\sigma^4}{(2\pi)^2}\Phi_0\right)^{-1} V_2 \left(4\frac{\sigma^4}{(2\pi)^2}\Phi_0\right)^{-1} = \Phi_0^{-1}\left(\Phi_0+\Phi_0^*\right)\Phi_0^{-1},$$

and the asymptotic variance of ${}_T^{(4)}$ is

$$\left(8\frac{\kappa_4^2}{(2\pi)^6}\Phi_0\right)^{-1} V_4 \left(8\frac{\kappa_4^2}{(2\pi)^6}\Phi_0\right)^{-1} = \frac{\bar{\mu}_6-\nu_3^2}{\nu_4^2}\Phi_0^{-1} + \left(3\nu_4^{-1}+1\right)^2\Phi_0^{-1}\Phi_0^*\Phi_0^{-1},$$

where $\bar{\mu}_6 = \nu_6 + 10\nu_3^2 + 15\nu_4 + 15$, while the covariances follow by immediate calculations,

$$ACov\left(\theta_T^{(2)},\theta_T^{(3)}\right) = \left(\frac{4\sigma^4}{(2\pi)^2}\left(\Phi_0+\Phi_0^*\right)\right)^{-1} V_{2,3}\left(\frac{6\kappa_3^2}{(2\pi)^4}\Phi_0\right)^{-1} = \Phi_0^{-1},$$

just sustituting the value of $V_{2,3}$, and similarly,

$$\begin{aligned}
ACov\left(\theta_T^{(2)},\theta_T^{(4)}\right) &= \left(\frac{4\sigma^4}{(2\pi)^2}\left(\Phi_0+\Phi_0^*\right)\right)^{-1} V_{2,4}\left(\frac{8\kappa_4^2}{(2\pi)^6}\Phi_0\right)^{-1} \\
&= \frac{\mu_4}{\kappa_4}\Phi_0^{-1} = \frac{\bar{\mu}_4}{\nu_4}\Phi_0^{-1},
\end{aligned}$$

and

$$
\begin{aligned}
ACov\left(\theta_T^{(3)}, \theta_T^{(4)}\right) &= \left(\frac{6\kappa_3^2}{(2\pi)^4}\Phi_0\right)^{-1} V_{3,4} \left(\frac{8\kappa_4^2}{(2\pi)^6}\Phi_0\right)^{-1} \\
&= \sigma^2 \left(\frac{\mu_5 - \kappa_3\sigma^2}{\kappa_3\kappa_4}\right)\Phi_0^{-1} + \frac{\mu_4}{\kappa_4}\Phi_0^{-1}\Phi_0^*\Phi_0^{-1} \\
&= \left(\frac{\nu_5 + 9\nu_3}{\nu_3\nu_4}\right)\Phi_0^{-1} + \frac{\bar{\mu}_4}{\nu_4}\Phi_0^{-1}\Phi_0^*\Phi_0^{-1}.
\end{aligned}
$$

□

APPENDIX B: AUXILIARY LEMMAS

Define for $k = 3, 4$ the following fundamental sets Δ_k of frequencies in $\Pi^{k-1} = [-\pi, \pi]^{k-1}$ where higher order periodograms can be defined uniquely and usual orthogonality properties of discrete Fourier Transforms are preserved. The set Δ_3 is the closure of anyone of the 12 equal-area triangles in which the hexagon Q_3 defined by the restrictions $|\lambda_1 - \lambda_2| \le 2\pi$ and $|2\lambda_a + \lambda_b| \le 2\pi$ for $a, b = 1, 2$, $a \neq b$ can be cut by the lines $\lambda_1 \pm \lambda_2 = 0$, $\lambda_1 + \lambda_2 \pm \lambda_1 = 0$ and $\lambda_1 + \lambda_2 \pm \lambda_2 = 0$. We can take the one defined by $0 < \lambda_1 < \pi, \lambda_2 < \lambda_1, \lambda_2 < 2(\pi - \lambda_1)$.

The set Δ_4 is the closure of anyone of the 48 equal-volume tetrahedrons in which the 12-face polyhedron (rhombic dodecahedron) Q_4 with volume $8\pi^3$, defined by the system of inequalities

$$
\begin{aligned}
|\lambda_a - \lambda_b| &\le 2\pi, \ a, b = 1, 2, 3, \ a \neq b, \\
|\lambda_a + \lambda_b + 2\lambda_c| &\le 2\pi, \ \ a \neq b \neq c \neq a,
\end{aligned}
$$

can be cut by the planes

$$
\begin{aligned}
\lambda_a \pm \lambda_b &= 0, \ \ a, b = 1, 2, 3, \ \ a \neq b, \\
\lambda_1 + \lambda_2 + \lambda_3 + \lambda_a &= 0, \ \ a = 1, 2, 3.
\end{aligned}
$$

For $k = 2$, Δ_2 is composed of just $[-\pi, 0]$ and $[0, \pi]$. See Alekseev (2010) for details.

Let for any bounded function g and for $k = 2, 3, 4$,

$$
\alpha_{k,T} = T^{1/2}\frac{1}{T^{k-1}} \sum_{j_1, \cdots, j_{k-1}=1}^{T-1} I_k(\lambda_{j_1}, \cdots, \lambda_{j_{k-1}}) g(\lambda_{j_1}, \cdots, \lambda_{j_{k-1}})
$$

and for any bounded scalar function g and the periodogram I_k^ε of order k of the innovations ε_t,

$$
\beta_{k,T} = T^{1/2}\frac{1}{T^{k-1}} \sum_{j_1, \cdots, j_{k-1}=1}^{T-1} I_k^\varepsilon(\lambda_{j_1}, \cdots, \lambda_{j_{k-1}}) \phi_k(\lambda_{j_1}, \cdots, \lambda_{j_{k-1}}) g(\lambda_{j_1}, \cdots, \lambda_{j_{k-1}}).
$$

Lemma A1: Under Assumptions 1 and 2,

$$E\left|\alpha_{k,T}-\beta_{k,T}\right|=O\left(T^{-1/2}\log^{k}T\right).$$

Proof of Lemma A1. Write for $k=3$,

$$(2\pi)^{2}T\cdot I_{3}^{\varepsilon}(\lambda_{j_1},\lambda_{j_2})\phi_3(\lambda_{j_1},\lambda_{j_2})=v\left(\lambda_{j_1}\right)v\left(\lambda_{j_2}\right)v\left(\lambda_{-j_1-j_2}\right)$$

where $v\left(\lambda\right)=w^{\varepsilon}\left(\lambda\right)\phi(\lambda)$ and $w_{T}^{\varepsilon}\left(\lambda\right)$ is the innovations discrete Fourier transform so that by the triangle inequality we can see that $E\left|\alpha_{T}-\beta_{T}\right|$ can be bounded by

$$\frac{T^{1/2}}{(2\pi)^{2}T^{3}}E\left|\sum_{j_1,j_2=1}^{T-1}\left\{w\left(\lambda_{j_1}\right)-v\left(\lambda_{j_1}\right)\right\}w\left(\lambda_{j_2}\right)w\left(-\lambda_{j_1+j_2}\right)g(\lambda_{j_1},\lambda_{j_2})\right| \tag{49}$$

plus two similar terms depending on $\left\{w\left(\lambda_{j_2}\right)-v\left(\lambda_{j_2}\right)\right\}v\left(\lambda_{j_1}\right)w\left(-\lambda_{j_1+j_2}\right)$ and $\left\{w\left(-\lambda_{j_1+j_2}\right)-v\left(-\lambda_{j_1+j_2}\right)\right\}v\left(\lambda_{j_2}\right)v\left(\lambda_{j_1}\right)$. By the triangle inequality we can restrict ourselves to frequencies within one of the fundamental triangles Δ_3. Expression (49) then is bounded by

$$\frac{CT^{1/2}}{(2\pi)^{2}T^{3}}\left(\sum_{j_1,j_2\in\Delta_3}\sum_{k_1,k_2\in\Delta_3}E\left[\begin{array}{c}\left\{w\left(\lambda_{j_1}\right)-v\left(\lambda_{j_1}\right)\right\}w\left(\lambda_{j_2}\right)w\left(-\lambda_{j_1+j_2}\right)\\ \times\left\{w\left(-\lambda_{k_1}\right)-v\left(-\lambda_{k_1}\right)\right\}w\left(-\lambda_{k_2}\right)w\left(\lambda_{k_1+k_2}\right)\end{array}\right]g(\lambda_{j_1},\lambda_{j_2})g(\lambda_{k_1},\lambda_{k_2})\right)^{1/2} \tag{50}$$

Then, expanding the product $\left\{w\left(\lambda_{j_1}\right)-v\left(\lambda_{j_1}\right)\right\}\left\{w\left(-\lambda_{k_1}\right)-v\left(-\lambda_{k_1}\right)\right\}$ and multiplying the four terms by $w\left(\lambda_{j_2}\right)w\left(-\lambda_{j_1+j_2}\right)w\left(-\lambda_{k_2}\right)w\left(\lambda_{k_1+k_2}\right)$ we can calculate the expectations using Lemma A2 for both Y_t data and ε_t with flat spectrum of all orders. Then the main terms of non negligible contributions cancel out up to approximation errors, which require two restrictions among the indexes j_1,j_2,k_1,k_2 in the products of moments of w involving cumulants κ_2^3, just one restriction for contributions involving $\kappa_2\kappa_4$ or κ_3^2, and none for terms involving κ_6. Then the magnitude of (50) is

$$\begin{aligned}&\leq\frac{CT^{1/2}}{(2\pi)^{2}T^{3}}\left(\sum_{j_1,j_2\in\Delta_3}\sum_{k_1,k_2\in\Delta_3}T^{3}\frac{\log T}{T}\delta\left\{2\text{ restrictions}\right\}+T^{2}\frac{\log^{3}T}{T}\delta\left\{1\text{ restriction}\right\}+T\frac{\log^{5}T}{T}\right)^{1/2}\\&=O\left(T^{-5/2}\left(T^{4}\log^{5}T\right)^{1/2}\right)=O\left(T^{-1}\log^{5}T\right)^{1/2},\end{aligned}$$

which proofs the lemma, since other terms could be dealt with similarly. The proof for $k=2,4$ follows similarly. □

Lemma A2: Under Assumptions 1 and 2, uniformly for $\left(\lambda_{j_1},\ldots,\lambda_{j_{k-1}}\right)\in\Delta_k$, $k=2,3,4$,

$$E\left[I_k\left(\lambda_{j_1},\ldots,\lambda_{j_{k-1}}\right)\right]=f_k\left(\lambda_{j_1},\ldots,\lambda_{j_{k-1}}\right)+O\left(T^{-1}\log^{k-1}T\right)$$

and for $\lambda_{j_1}+\cdots+\lambda_{j_k}\neq 0\bmod 2\pi$

$$\frac{1}{(2\pi)^{k-1}T}\mathrm{cum}_k\left[w\left(\lambda_{j_1}\right),\ldots,w\left(\lambda_{j_k}\right)\right]=O\left(T^{-1}\log^{k-1}T\right)$$

If $v(\lambda) = w^{\varepsilon}(\lambda)\,\phi(\lambda)$, and denoting by u either v or w, for $\left(\lambda_{j_1}, \ldots, \lambda_{j_{k-1}}\right) \in \Delta_k$,

$$\frac{1}{(2\pi)^{k-1} T} E\left[u\left(\lambda_{j_1}\right) \cdots u\left(-\lambda_{j_1+\cdots+j_k}\right)\right] = f_k\left(\lambda_{j_1}, \ldots, \lambda_{j_{k-1}}\right) + O\left(T^{-1} \log^{k-1} T\right)$$

and for $\lambda_{j_1} + \cdots + \lambda_{j_k} \neq 0 \bmod 2\pi$,

$$\frac{1}{(2\pi)^{k-1} T} \operatorname{cum}_k\left[u\left(\lambda_{j_1}\right), \ldots, u\left(\lambda_{j_k}\right)\right] = O\left(T^{-1} \log^{k-1} T\right).$$

□

Proof of Lemma A2. Noting that for ARMA processes without unit roots all existing spectral densities are differentiable and (finite) cumulants satisfy any summability condition, these results are an extension of Lemma 2 in Delgado, Hidalgo and Velasco (2005) using arguments in Theorem 2 in Robinson (1995) to higher order cumulants, cf. Theorem 4.3.2 of Brillinger (1975). □

APPENDIX C: ASYMPTOTIC EQUIVALENCE BETWEEN MD ESTIMATION WITH $\mathbf{L}_{2T}$ AND WHITTLE ESTIMATION

In this Appendix we show that the estimator based on minimizing $L_{2T}(\theta, \kappa_2)$ is asymptotically equivalent to the Whittle estimator under invertibility and causality, which is asymptotically equivalent to the maximum likelihood estimator in the Gaussian case. In addition, our estimator is efficient in the sense of Kumon (1992) (that is, in the invertible case, it achieves the minimum variance that an estimator that belongs to the generalized Whittle-class proposed by Kumon can achieve).

Next, we show that for a particular case, when $\kappa_3 = \kappa_4 = 0$ and the process is invertible, the proposed L_2-estimator has optimal asymptotic properties in the sense that it is asymptotically equivalent to the efficient Whittle estimator defined by minimizing the Whittle Likelihood (an approximation to the Gaussian likelihood), see Whittle (1953). Recall the Whittle function

$$\begin{aligned} L_{WT}(\theta, \kappa_2) &= \frac{1}{T}\sum_{j=1}^{T-1} \log f(\theta, \kappa_2; \lambda_j) + \frac{1}{T}\sum_{j=1}^{T-1} \frac{I(\lambda_j)}{f(\theta, \kappa_2; \lambda_j)} \qquad (51) \\ &= \log\left(\frac{\kappa_2}{2\pi}\right) + \frac{1}{T}\sum_{j=1}^{T-1} \frac{I(\lambda_j)}{f(\theta, \kappa_2; \lambda_j)}, \end{aligned}$$

where we assume that the second equality holds in finite samples using that $\int_{-\pi}^{\pi} \log \phi_2(\theta; \lambda) d\lambda = 0$ for all θ under invertibility and causality of the ARMA parameterization.

Note that κ_2 can be also concentrated out of the objective function, so that

$$\kappa_{2T}(\theta) = \frac{2\pi}{T}\sum_{j=1}^{T-1} \frac{I(\lambda_j)}{\phi_2(\theta; \lambda_j)},$$

that coincides with (14) for the $k = 2$ case so that estimation of θ and κ_2 can be conducted then separately because and

$$\dot{\kappa}_{2T}(\theta_0) = \frac{\partial}{\partial\theta}\kappa_{2T}(\theta_0) = -\frac{2\pi}{T}\sum_{j=1}^{T-1}\frac{I(\lambda_j)}{\phi_2(\theta_0;\lambda_j)}\varphi_2(\theta_0;\lambda_j) = o_p(1)$$

since $\varphi_2(\theta_0, \lambda) = \frac{\partial}{\partial\theta}\log\phi_2(\theta_0, \lambda)$ integrates to zero in Π and the properties of the periodogram, cf. Lemma A2.

The asymptotic equivalence between estimates is simply shown by checking that the score and Hessian of both problems are asymptotically equivalent. Then, on the one hand the normalized score with respect to θ corresponding to L_W is

$$\begin{aligned} S_{WT}(\theta_0, \kappa_2) &= -\frac{1}{T}\sum_{j=1}^{T-1}\frac{I(\lambda_j)}{f(\theta_0, \kappa_2; \lambda_j)}\frac{(\partial/\partial\theta) f(\theta_0, \kappa_2; \lambda_j)}{f(\theta_0, \kappa_2; \lambda_j)} + o_P\left(T^{-1/2}\right) \\ &= -\frac{1}{T}\sum_{j=1}^{T-1}\frac{I(\lambda_j)}{f(\theta_0, \kappa_2; \lambda_j)}\varphi_2(\theta_0; \lambda_j) + o_P\left(T^{-1/2}\right). \end{aligned}$$

On the other hand, the normalized score with respect to θ corresponding to the minimum distance L_2-estimator based on minimizing $\tilde{L}_{2T}(\theta, \kappa_2)$ is

$$\begin{aligned} S_{2T}(\theta_0) &= -\frac{1}{T}\sum_{j=1}^{T-1}\left(\frac{I(\lambda_j) - f(\theta_0, \kappa_{2T}^{\dagger}(\theta_0); \lambda_j)}{f(\bar{\theta}_T, \bar{\kappa}_{2T}; \lambda_j)^2}\right)\frac{\partial}{\partial\theta}f(\theta, \kappa_{2T}^{\dagger}(\theta), \lambda_j) + o_P\left(T^{-1/2}\right) \qquad (52) \\ &= -\frac{1}{T}\sum_{j=1}^{T-1}\left(\frac{I(\lambda_j) - f(\theta_0, \kappa_2; \lambda_j)}{f(\lambda_j)}\right)\varphi_2(\theta_0; \lambda_j) + o_P\left(T^{-1/2}\right) \\ &= -\frac{1}{T}\sum_{j=1}^{T-1}\frac{I(\lambda_j)}{f(\lambda_j)}\varphi_2(\theta_0; \lambda_j) + o_P\left(T^{-1/2}\right) \end{aligned}$$

using that $\kappa_{2T}(\theta_0) = \kappa_2 + O_P\left(T^{-1/2}\right)$, $\dot{\kappa}_{2T}(\theta_0) = O_P\left(T^{-1/2}\right)$ and preliminary consistent estimates $\bar{\theta}_T$ and $\bar{\kappa}_{2T}$, following the proof of Theorem 2.

This shows that both objective functions have asymptotically the same score up to error $T^{-1/2}$ when evaluated at the parameter true value. Then, the same arguments can be used to show the asymptotic equivalence of the respective Hessian functions in a neighborhood of θ_0 and of the corresponding estimates up to terms $o_P\left(T^{-1/2}\right)$.

APPENDIX D: PROOFS OF LEMMAS 4, 5 AND 6

Proof of Lemma 4. For pure invertible and causal models, $\Phi_0^* = 0$, the AVar of an estimate using only second and third moments with $(w_2, w_3, w_4) = (1, w_3, 0)$ w.l.o.g., is

$$\frac{1 + 3w_3\nu_3^2 + \frac{9}{4}w_3^2\nu_3^4\left(\frac{\bar{\mu}_4 - 1}{\nu_3^2}\right)}{1 + 3w_3\nu_3^2 + \frac{9}{4}w_3^2\nu_3^4}\Phi_0^{-1}$$

where the factor in front of Φ_0^{-1} is always larger than 1 if $w_3 > 0$ because $\bar{\mu}_4 - 1 = 2 + \nu_4 \geq \nu_3^2 > 0$, see Rao (1973, p.143). $\square$

Proof of Lemma 5. For an estimate using only second and fourth moments we could check for $(w_2, w_3, w_4) = (1, 0, w_4)$ w.l.o.g. that the AVar is

$$\frac{1 + 4w_4\nu_4^2\frac{\bar{\mu}_4}{\nu_4} + 4w_4^2\nu_4^4\frac{\bar{\mu}_6-\nu_3^2}{\nu_4^2}}{1 + 4w_4\nu_4^2 + 4w_4^2\nu_4^4}\Phi_0^{-1} = \frac{1 + 4w_4\nu_4\bar{\mu}_4 + 4w_4^2\nu_4^2\left(\bar{\mu}_6 - \nu_3^2\right)}{1 + 4w_4\nu_4^2 + 4w_4^2\nu_4^4}\Phi_0^{-1}$$

where $\frac{\bar{\mu}_4}{\nu_4} = \frac{\nu_4+3}{\nu_4} = 1 + \frac{3}{\nu_4} > 1$ if $\nu_4 > 0$; < -0.5 if $v_4 < 0$, and $\bar{\mu}_6 - \nu_3^2 \geq \bar{\mu}_4^2 > \nu_4^2$. Therefore if $\nu_4 > 0$, then using L_4 reduces efficiency, but if $v_4 < 0$ it can increase for some values of w_4.

The problem of minimizing the variance can be written as

$$\min_{w_4\in[0,\infty)} \frac{1 + 4w_4\nu_4^2\left(1 + \frac{3}{\nu_4}\right) + 4w_4^2\nu_4^4\frac{\bar{\mu}_6-\nu_3^2}{\nu_4^2}}{1 + 4w_4\nu_4^2 + 4w_4^2\nu_4^4}$$

so the FOC is

$$\begin{aligned} 0 \quad = \quad & \left(1 + 4w_4\nu_4^2 + 4w_4^2\nu_4^4\right)\left(4\nu_4^2\left(1 + \frac{3}{\nu_4}\right) + 8w_4\nu_4^4\frac{\bar{\mu}_6 - \nu_3^2}{\nu_4^2}\right) \\ & - \left(4\nu_4^2 + 8w_4\nu_4^4\right)\left(1 + 4w_4\nu_4^2\left(1 + \frac{3}{\nu_4}\right) + 4w_4^2\nu_4^4\frac{\bar{\mu}_6 - \nu_3^2}{\nu_4^2}\right), \end{aligned}$$

whose solutions for $\nu_4 \neq 0$ are

$$\begin{array}{cl} \left\{\frac{3}{6\nu_4^2+2\nu_4^3-2\nu_4\mu_6+2\nu_3^2\nu_4}, -\frac{1}{2\nu_4^2}\right\} & \text{if} \quad 6\nu_4^2 + 2\nu_4^3 - 2\nu_4\mu_6 + 2\nu_3^2\nu_4 \neq 0 \\ \left\{-\frac{1}{2\nu_4^2}\right\} & \text{if} \quad 6\nu_4^2 + 2\nu_4^3 - 2\nu_4\mu_6 + 2\nu_3^2\nu_4 = 0. \end{array}$$

So if we disregard the simple solution $-\frac{1}{2\nu_4^2}$ because is always negative, we obtain the solution provided given the denominator is not zero.

If we assume that $\nu_4 < 0$, then $w_4^\dagger > 0$, then the denominator is always positive since it is equal to

$$2\nu_4^2\left(3 + \nu_4\right) - 2\nu_4\left(\bar{\mu}_6 - \nu_3^2\right) = 2\nu_4^2\left(3 - |\nu_4|\right) + 2|\nu_4|\left(\bar{\mu}_6 - \nu_3^2\right)$$

and all factors are positive, where $\bar{\mu}_6 - \nu_3^2 \geq \bar{\mu}_4^2 > \nu_4^2 > 0$ and $3 - |\nu_4| > 1 > 0$ since $-2 < \nu_4 < 0$.

To explore whether $\nu_4 > 0$ is compatible with efficiency gains we can alternatively consider linear combinations of estimates, $\gamma\hat{\theta}_T^{(2)} + (1-\gamma)\,\hat{\theta}_T^{(4)}$, $\gamma \in [0, 1]$, whose asymptotic variance is proportional to

$$\gamma^2 + (1-\gamma)^2\,\frac{\bar{\mu}_6 - \nu_3^2}{\nu_4^2} + 2\gamma\,(1-\gamma)\,\frac{\bar{\mu}_4}{\nu_4}$$

which is minimized by

$$\gamma^* = \frac{\frac{\bar{\mu}_6-\nu_3^2}{\nu_4^2} - \frac{\bar{\mu}_4}{\nu_4}}{\frac{\bar{\mu}_6-\nu_3^2}{\nu_4^2} - 2\frac{\bar{\mu}_4}{\nu_4} + 1} = \frac{\left\{\frac{\bar{\mu}_6-\nu_3^2}{\nu_4^2} - 1\right\} - \frac{3}{\nu_4}}{\left\{\frac{\bar{\mu}_6-\nu_3^2}{\nu_4^2} - 1\right\} - 2\frac{3}{\nu_4}}$$

where the terms in brackets are always positive. If $\nu_4 < 0$ then $\gamma^* \in (0,1)$, so there are efficiency gains by using L_2, but if $\nu_4 > 0$ then $\gamma^* > 1$. □

Proof of Lemma 6. Consider estimates which only use higher order moments, $(w_2, w_3, w_4) = (0, 1, w_4)$, whose asymptotic variance is

$$\frac{\frac{9}{4}\nu_3^4\frac{\bar{\mu}_4-1}{\nu_3^2}+4w_4^2\nu_4^4\frac{\bar{\mu}_6-\nu_3^2}{\nu_4^2}+6w_4\nu_3^2\nu_4^2\frac{\bar{\mu}_5-\nu_3}{\nu_3\nu_4}}{\left(\frac{3}{2}\nu_3^2+2w_4\nu_4^2\right)^2}\Phi_0^{-1}$$

and the foc of its minimization with respect to w_4 is

$$\begin{aligned} 0 &= \left(\frac{3}{2}\nu_3^2+2w_4\nu_4^2\right)\left(8w_4\nu_4^4\frac{\bar{\mu}_6-\nu_3^2}{\nu_4^2}+6\nu_3^2\nu_4^2\frac{\bar{\mu}_5-\nu_3}{\nu_3\nu_4}\right) \\ &\quad -4\nu_4^2\left(\frac{9}{4}\nu_3^4\frac{\bar{\mu}_4-1}{\nu_3^2}+4w_4^2\nu_4^4\frac{\bar{\mu}_6-\nu_3^2}{\nu_4^2}+6w_4\nu_3^2\nu_4^2\frac{\bar{\mu}_5-\nu_3}{\nu_3\nu_4}\right), \end{aligned}$$

with solution

$$w_4^\dagger = \frac{3\nu_3^3-3\nu_3\nu_4-3\nu_3^2\mu_5+3\mu_4\nu_3\nu_4}{4\nu_3\nu_4^2-4\mu_5\nu_4^2-4\nu_3^3\nu_4+4\nu_3\nu_4\mu_6},$$

when the denominator is different from zero. The numerator is

$$\begin{aligned} 3\left(\nu_3^3-\nu_3\nu_4-\nu_3^2\bar{\mu}_5+\bar{\mu}_4\nu_3\nu_4\right) &= 3\left\{\nu_3\nu_4\left(\bar{\mu}_4-1\right)-\nu_3^2\left(\bar{\mu}_5-\nu_3\right)\right\} \\ &= 3\nu_3\left\{\nu_4\left(\bar{\mu}_4-1\right)-9\nu_3^2\right\}-3\nu_3^2\nu_5 \end{aligned}$$

because $\bar{\mu}_5 = \nu_5 + 10\nu_3$, while the denominator is

$$\begin{aligned} 4\left(\nu_3\nu_4^2-\bar{\mu}_5\nu_4^2-\nu_3^3\nu_4+\nu_3\nu_4\bar{\mu}_6\right) &= 4\left\{\nu_3\nu_4\left(\bar{\mu}_6-\nu_3^2\right)-\nu_4^2\left(\bar{\mu}_5-\nu_3\right)\right\} \\ &= 4\nu_3\left\{\nu_4\left(\bar{\mu}_6-\nu_3^2\right)-9\nu_4^2\right\}-4\nu_4^2\nu_5. \end{aligned}$$

so that when $\nu_5 > 0$ and $\nu_3 > 0$ w.l.o.g., both terms are negative because $\bar{\mu}_6 - \nu_3^2 > 0$ by Cauchy-Swartz inequality, and $w_4^\dagger > 0$. □

REFERENCES

Ahn, V.V., N.N. Leonenko, L.M. Sakhno (2007). "Minimum contrast estimation of random processes based on information of second and third orders". *Journal of Statistical Planning and Inference* 137, 1302-1331.

Alekseev, V.G. (1993), "Asymptotic Properties of Higher-Order Periodograms", *Theory of Probability and its Applications*, 40, 409-419.

Alekseev, V.G. (2010). "On Nonparametric estimates of higher order spectral densities". *Theory of Probability and its Applications*, 54.3, 499-04.

Alessi L., Barigozzi, M. and M. Capasso (2011), "Non-Fundamentalness in Structural Econometric Models: A Review", *International Statistical Review*, 79, 16–47.

Amano, N. and I.N. Lobato (2012), "GMM estimation of noninvertible linear models", working paper, ITAM.

Andrews, B., Davis, R.A. and Breidt, J. (2007). "Rank-based estimation for all-pass time series models". *Annals of Statistics*, 35.2, 844-869.

Breidt, F.J., Davis, R.A. and Trindade, A.A. (2001). "Least absolute deviation estimation for all-pass time series models". *Annals of Statistics*, 29.4, 919-946.

Brillinger, D.R. (1975). *Time Series: Data Analysis and Theory*, Holden Day, San Francisco.

Brillinger, D.R. (1985). "Fourier inference: some methods for the analysis of array and non-gaussian series data". *Water Resources Bulletin*, 21.5, 744-756.

Delgado, M.A., Hidalgo, F.J. and Velasco, C. (2005). "Distribution free goodness-of-fit tests for linear processes". *Annals of Statistics*, 33.6, 2568-2609.

Gospodinov, N. and Ng, S. (2013). "Minimum distance estimation of possibly non-invertible moving average models," Working Paper 2013-11, Federal Reserve Bank of Atlanta.

Hannan, E.J. (1970), *Multiple Time Series*, John Wiley, New York.

Hansen, L. P. and T. J. Sargent, (1980). "Formulating and estimating dynamic linear rational expectations models," *Journal of Economic Dynamics and Control*, Elsevier, 2(1), 7-46.

Hansen, L. P. and T. J. Sargent, (1991). "Two Difficulties in Interpreting Vector Autoregressions," *Rational Expectations Econometrics*. Boulder and Oxford: Westview Press, 77-119.

Huang, J. and Pawitan, Y. (2000). Quasi-likelihood estimation of non-invertible moving average processes. *Scandinavian Journal of Statistic*s, 27:689{702.

Kumon, M. (1992), "Identification of non-minimum phase transfer function using higher-order spectrum", A*nn. Inst. Statist. Math.*, 44, 239-260.

Leeper, E.M., Walter, T.B. and S. Yang (2013). "Fiscal Foresight and Information Flows", *Econometrica*, 81,1115-1145.

Lanne, M. and Saikkonen, P. (2011). "Noncausal Autoregressions for Economic Time Series," *Journal of Time Series Econometrics*, Berkeley Electronic Press, vol. 3(3), 2.

Lii, K.S., and M. Rosenblatt, (1992). "An approximate maximum likelihood estimation for non-Gaussian non-minimum phase moving average processes." *Journal of Multivariate Analysis*, 43, 272-299.

Mountford, A. and Uhlig, H. (2009). What are the effects of scal policy shocks? Journal of Applied Econometrics, 24:960-992.

Ramsey, J. B., A. Montenegro (1992). "Identification and Estimation of Non-invertible Non-Gaussian MA(q) Processes," *Journal of Econometrics*, 54, 301-320.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, Wiley, New York.

Robinson, P.M. (1995). "Gaussian semiparametric estimation of long range dependence." *Annals* *of Statistics*, 23.5, 1630-1661.

Rosenblatt, M. (1985). *Stationary Sequences and Random Fields*, Springer, New York.

Terdik, G. (1999), *Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis*, Springer, New York.

Whittle, P. (1953). "The Analysis of Multiple Stationary Time Series", *Journal of the Royal Statistical Society Series B,* 15, 125–139.