FREQUENCY DOMAIN MINIMUM DISTANCE INFERENCE FOR POSSIBLY NONINVERTIBLE AND NONCAUSAL ARMA MODELS

BY CARLOS VELASCO[1] AND IGNACIO N. LOBATO[2]

Universidad Carlos III de Madrid and Centro de Investigación Económica, ITAM

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 propose a minimum distance estimator that optimally combines the information contained in second, third, and fourth moments. Contrary to existing estimators, the proposed one is consistent under general assumptions, and may improve on the efficiency of estimators based on only second order moments. Our procedures are also applicable for processes for which either the third or the fourth order spectral density is the zero function.

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 noncausal or noninvertible representations. Hence, for estimation of Gaussian ARMA processes causality and invertibility need to be imposed. For non-Gaussian ARMA models the causality and invertibility assumptions are not necessary and not always justified, and in fact, noncausal or noninvertible ARMA models have been employed in many areas such as economics, seismology, engineering or astronomy; see references in Breidt, Davis and Trindade (2001). For some examples in economics, see Alessi, Barigozzi and Capasso (2011), Hansen and Sargent (1980, 1991), Huang and Pawitan (2000), Leeper, Walker and Yang (2013), and Mountford and Uhlig (2009).

The broad literature devoted to estimating general nonstandard ARMA models can be classified according to two criteria. The first criteria is whether the distri-

bution for the innovations is assumed to be known or unknown, so either ML 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 noncausal noninvertible ARMA models irrespective 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.

Some references are briefly summarized next. Lii and Rosenblatt (1992) investigate the properties of a one-step approximate maximum likelihood procedure for the possibly noninvertible or nonminimum phase 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 first order conditions of the Whittle (1953) estimate. Anh, Leonenko and Sakhno (2007) follow Kumon's approach, and propose estimates based on related estimating equations using weighted higher order periodograms. Note that in both references, identification is achieved by introducing *ad hoc* conditions on the sign of the higher order cumulants of the innovation (Kumon) or on a weighting function that controls the sign of the cumulant (Anh *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 the asymptotic properties of the final estimates recovered from these two steps are unclear. In addition, the second step is only meaningful when the first step residuals are nonindependent 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 are 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) 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) and Gospodinov and Ng (2015)], but these time-domain minimum distance estimators are inefficient since they only employ a finite number of arbitrary 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 noncausal noninvertible ARMA model using higher order spectral densities as long as some higher order cumulant of the innovations is not zero. Identification comes from establishing that an objective function (OF) based on the L^2 distance of the higher order spectral densities has a unique global minimum at the correct parameter values. This result motivates the proposal of an efficient estimator that optimally combines the information contained in the second, third and fourth spectral densities. This minimum distance estimator optimally weights the scores of the corresponding sample analogues of the objective functions employed in the identification result.

The second contribution of this article is to establish the consistency (as long as the third or the fourth cumulant of the innovations is not zero) and the asymptotic normality (even for the case where the third or the fourth cumulant of the innovations is zero) of the proposed estimator 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 estimator studied in this paper weights optimally 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 unfeasible Whittle estimator, the Gaussian quasi ML estimator (QMLE) that assumes correct knowledge of the location of the ARMA roots. Using the information contained in higher order moments does not only achieve identification, but also can improve on the efficiency with respect to the Whittle estimator.

Although the article focus on identification and estimation of ARMA models, a 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 consists of centered powers of the innovations, which is applied to weighted averages of higher order periodograms.

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, auxiliary lemmas and additional material are contained in Appendices A, B and C, respectively, in the Supplementary Material [Velasco and Lobato (2018)].

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

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

and the jth 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

$$\text{cum}(Y_1, \ldots, Y_k)$$
$$= \sum (-1)^{p-1} (p-1)! E\left(\prod_{j \in v_1} Y_j\right) \cdots E\left(\prod_{j \in v_p} Y_j\right), \qquad k = 1, 2, \ldots$$

assuming $E[|Y_t|^k] < \infty$, and where $v_1, \ldots, v_p$ is a partition of $(1, 2, \ldots, k)$, and the sum runs over all these partitions; see Brillinger (1975) or Rosenblatt (1985), page 34. Hence, the first and second cumulants are the mean and the variance, respectively.

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

$$f_k(\boldsymbol{\lambda}) = f_k(\lambda_1, \ldots, \lambda_{k-1})$$

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

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 $k = 2$; also note

that 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

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

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

$$I_k(\boldsymbol{\lambda}) = I_k(\lambda_1, \ldots, \lambda_{k-1}) = \frac{1}{(2\pi)^{k-1} T} \prod_{j=1}^{k-1} w(\lambda_j) w\left(-\sum_{s=1}^{k-1} \lambda_s\right),$$

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 $\boldsymbol{\lambda}$, $k = 2, 3, \ldots$,

(3) $$E[I_k(\boldsymbol{\lambda})] = f_k(\boldsymbol{\lambda}) + o(1),$$

and when $\lambda_s \neq 0 \bmod 2\pi$, $s = 1, \ldots, k-1$,

(4)
$$T^{2-k} \operatorname{Var}[I_k(\boldsymbol{\lambda})]$$
$$= (k-1)! f(\lambda_1) f(\lambda_2) \cdots f(\lambda_{k-1}) f(\lambda_1 + \lambda_2 + \cdots + \lambda_{k-1}) + o(1),$$

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), pages 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 evaluate these 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.

$$\operatorname{Cov}[I_k(\boldsymbol{\lambda}), I_\ell(\boldsymbol{\lambda}')] = o(T^{(k+\ell-4)/2}), \qquad k, \ell = 2, 3, \ldots,$$

as T tends to infinity for almost all λ and λ' (not satisfying some particular linear restrictions); see, for instance, Lemma 1 in page 172 in Rosenblatt (1985), and Theorems 2 and 4 in Alekseev (1993).

2.2. *Model.*   We assume that Y_t is given by

$$(5) \qquad Y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j},$$

where ε_t is an independent identically distributed (i.i.d.) sequence with zero mean, variance κ_2^0, and with bounded moments of order k, with $k \geq 3$; in addition $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$ (see Assumptions 1 and 2 below). Note that the double side summation in equation (5) allows Y_t to be noncausal or noninvertible.

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

$$(6) \qquad \alpha(L)Y_t = \beta(L)\varepsilon_t,$$

and 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 it, and do not have any common roots. Denote the model parameters by $\theta = (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)' \in \Gamma_{p,q} = \{\theta \in \mathbb{R}^{p+q} : \alpha(z)\beta(z) \neq 0$ for $|z| = 1, |\alpha(z)| + |\beta(z)| > 0$ for all $z \in \mathbb{C}, \alpha_p \neq 0, \beta_q \neq 0\}$. 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

$$\mathcal{A}_{p,q} = \left\{ \boldsymbol{\varphi} = (a_1, \ldots, a_p, b_1, \ldots, b_q)' \in \bigcup_{r=0}^{\lfloor p/2 \rfloor + \lfloor q/2 \rfloor} \left(\mathbb{R}^{p+q-2r} \cup \mathbb{C}_*^{2r}\right) : \right.$$

$$\left. |a_i| \neq 0, 1; |b_j| \neq 0, 1; a_i \neq b_j, i = 1, \ldots, p, j = 1, \ldots, q \right\},$$

where $\bar{a}$ denotes the complex conjugate of a, $\mathbb{C}_*^2 = \{(a, b) \in \mathbb{C}^2 : b = \bar{a}\}$ denotes the space of pairs of complex conjugate numbers, guaranteing that $\theta(\boldsymbol{\varphi})$, $\boldsymbol{\varphi} \in \mathcal{A}_{p,q}$, is real, and r accounts for the number of such pairs.

Model (5) 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 jth order marginal cumulant of ε_t, so that κ_2 is its variance, and

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

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

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

For instance, for the ARMA model (6),

(7)
$$\phi(\theta; \lambda) = \frac{1 + \sum_{j=1}^{q} \beta_j \exp(-ij\lambda)}{1 - \sum_{j=1}^{p} \alpha_j \exp(-ij\lambda)}.$$

In addition, model (5) establishes that $f_k(\boldsymbol{\lambda}) = f_k(\theta, \kappa_k; \boldsymbol{\lambda})$ where

(8)
$$f_k(\theta, \kappa_k; \boldsymbol{\lambda}) = \frac{\kappa_k}{(2\pi)^{k-1}} \phi_k(\theta; \boldsymbol{\lambda})$$

and

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

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,

(9)
$$|f_k(\theta, \kappa_k; \boldsymbol{\lambda})|^2$$
$$= \frac{\nu_k^2}{(2\pi)^{k-2}} f(\theta, \kappa_2; \lambda_1) \cdots f(\theta, \kappa_2; \lambda_{k-1}) f(\theta, \kappa_2; \lambda_1 + \cdots + \lambda_{k-1}),$$

where

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

is 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 (6). Denote $\Pi = [-\pi, \pi]$ and $d\boldsymbol{\lambda} = d\lambda_1 \cdots d\lambda_{k-1}$.

ASSUMPTION 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

$$\mathcal{A}_{p,q}(\eta) = \left\{ \boldsymbol{\varphi} = (a_1, \ldots, a_p, b_1, \ldots, b_q)' \in \bigcup_{r=0}^{\lfloor p/2 \rfloor + \lfloor q/2 \rfloor} (\mathbb{R}^{p+q-2r} \cup \mathbb{C}_*^{2r}) : \right.$$

$$\min\{|a_i|, |b_j|, ||a_i| - 1|, ||b_j| - 1|, |a_i - b_j|\} \geq \eta > 0,$$

$$\left. i = 1, \ldots, p, j = 1, \ldots, q \right\}.$$

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 $\theta \in \Theta$.

THEOREM 1.   *Consider an* ARMA(p, q) *model* (6) *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*

$$(10) \quad \inf_{\theta \in \Theta : \|\theta - \theta_0\| > \eta, \kappa_k \in \mathbb{R}} \int_{\Pi^{k-1}} |f_k(\theta, \kappa_k; \boldsymbol{\lambda}) - f_k(\theta_0, \kappa_k^0; \boldsymbol{\lambda})|^2 g(\theta; \boldsymbol{\lambda}) \, d\boldsymbol{\lambda} \geq \epsilon > 0.$$

Proofs and technical results are contained in Appendices A and B of the Supplementary Material [Velasco and Lobato (2018)]. The weighting function $g(\theta; \boldsymbol{\lambda})$ 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 proposed in Terdik (1999), equation (4.3). 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 a noninvertible or noncausal 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.

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 $\mathcal{S} \subseteq \mathcal{A}_{p,q}$ the discrete set of all potential roots obtained by inversion (of at least one) of the elements of $\boldsymbol{\varphi}_0$ so that complex roots always appear in conjugate pairs. Note that the cardinality of $\mathcal{S}$ is $2^n - 1$ where $n = p + q - r$ and r is the number of pairs of complex roots determined by $\boldsymbol{\varphi}_0$. Then, $f_k(\theta(\boldsymbol{\varphi}), \kappa_k; \boldsymbol{\lambda})$ is the spectral density of order k of the ARMA process (6) calculated from (7) and (8) 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 of $\boldsymbol{\varphi} \in \mathcal{A}_{p,q}$.

LEMMA 1.   *Consider an* ARMA(p, q) *model* (6) *with true roots* φ_0 *satisfying Assumption* A(p, q) *and* $\kappa_k^0 \neq 0$ *for some* $k \geq 3$. *Then,*

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

Lemma 1 shows that for any $\varphi \neq \varphi_0$ that belongs to the set $\mathcal{S}$, the resulting spectral density is different from $f_k(\theta(\varphi_0), \kappa_k^0; \lambda)$ in $g(\theta; \lambda)$-L^2 distance in Π^{k-1} for 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 $f_k(\theta(\varphi), \kappa_k; \lambda)$ match simultaneously those of $f_k(\theta(\varphi_0), \kappa_k^0; \lambda)$ for all λ. Note, however, that this is indeed possible for $k = 2$ because $f = f_2$ is real and it is always possible for any $\varphi \in \mathcal{S}$ to find a value $\kappa_2 = \kappa_2(\varphi)$ that satisfies $f_2(\theta(\varphi), \kappa_2; \lambda) = f_2(\theta(\varphi_0), \kappa_2^0; \lambda)$ for all $\lambda \in \Pi$. An equivalent 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 from Lemma 1 it only remains to consider elements φ 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).

4. Minimum distance estimation.

4.1. *Minimum distance estimation based on the spectral density of order k.* Consider initially the problem of estimating the parameter vector θ using the information contained in the spectral density of order k. The idea is to use the information provided by the normalized distance between the data, reflected in the higher-order periodogram, $I_k(\lambda)$, and the model, reflected in the parameterization of the spectral density of order k, $f_k(\theta, \kappa_k; \lambda)$. In particular, using (3) and (4) we define

$$(11) \quad L_k(\theta, \kappa_k) = \int_{\Pi^{k-1}} \frac{E|I_k(\lambda) - f_k(\theta, \kappa_k; \lambda)|^2}{T^{k-2} f(\lambda_1) \cdots f(\lambda_{k-1}) f(\lambda_1 + \lambda_2 + \cdots + \lambda_{k-1})} \, d\lambda.$$

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 $I_k(\lambda)$ and $f_k(\theta, \kappa_k; \lambda)$ are complex. 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 L_k comparable for different k. Note that the scaling is well defined when unit roots are excluded in the parameter space, so that $0 < f(\lambda) < \infty$ for all λ.

Hence, the OF based on $f_k(\theta, \kappa_k; \lambda)$ is the sample analogue of (11) where in the denominator we replace the normalizing spectral densities by their traditional Whittle estimators (under invertibility and causality), $\overline{f}_T(\lambda) = f(\overline{\theta}_T, \overline{\kappa}_{2T}; \lambda)$, say,

$$(12) \quad L_{kT}(\theta, \kappa_k) = \frac{1}{T} \sum_{j=1}^{T-1} \frac{|I_k(\lambda_j) - f_k(\theta, \kappa_k; \lambda_j)|^2}{T^{k-2} \overline{f}_T(\lambda_{j_1}) \cdots \overline{f}_T(\lambda_{j_{k-1}}) \overline{f}_T(\lambda_{j_1} + \cdots + \lambda_{j_{k-1}})}.$$

Here we have simplified the notation by introducing the general $(k - 1)$-dimensional vector of Fourier frequencies $\lambda_j = (\lambda_{j_1}, \ldots, \lambda_{j_{k-1}})$ and by writing

$\sum_{j=1}^{T-1} = \sum_{j_1=1}^{T-1} \cdots \sum_{j_{k-1}=1}^{T-1}$, where we discard all combinations of frequencies such that $\lambda_{j_a} + \lambda_{j_b} = T$, $a \neq b$, and $\lambda_{j_1} + \lambda_{j_2} + \lambda_{j_3} = T \bmod T$. This avoids referring in the notation to specific sets of $\boldsymbol{\lambda}_j \in \Pi^{k-1}$ where the periodogram can be defined uniquely; see the discussion in Anh, Leonenko and Sakhno (2007). Note that plugging in these consistent estimators $\overline{f}_T(\lambda) = f(\overline{\theta}_T, \overline{\kappa}_{2T}; \lambda)$ gives always the appropriate scaling in (12) since they are invariant to any potential inversion of the polynomials roots implied by $\overline{\theta}_T$ if $\overline{\kappa}_{2T}$ is adjusted accordingly to estimate the corresponding innovation variance. Further, the normalization by preliminary estimates of f does not have any effect on the asymptotic properties of the parameter estimates, and greatly simplifies the analysis when compared to the case where scaling is simultaneously estimated; cf. Section 4 in Terdik (1999).

In the minimization of the previous OF (12), the cumulant κ_k is a nuisance parameter that can be concentrated out. Then, recalling (9), expression (12) can be written as

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

Also note that $\overline{\kappa}_{2T}$ is only a scaling factor in $L_{kT}(\theta, \kappa_k)$, but it is useful in order to make comparisons among the contribution from different k. Considering the first order conditions (FOC) of minimizing $L_{kT}(\theta, \kappa_k)$ respect to κ_k, after straightforward algebra using the definition of f_k, it is simple to solve for κ_k, and define

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

The fact that $|\phi_k(\overline{\theta}_T; \boldsymbol{\lambda}_j)|^2$ is a consistent estimator for $|\phi_k(\theta_0; \boldsymbol{\lambda}_j)|^2$ up to scale implies that, for $\theta \approx \theta_0$ and large T, the ratio in parenthesis can be approximated by a constant that cancels out with the ratio involving I_k, so we propose to replace $\kappa_{kT}(\theta)$ by the following simpler estimate of κ_k:

$$(13) \qquad \kappa_{kT}^{\dagger}(\theta) = \left(\frac{2\pi}{T} \right)^{k-1} \sum_{j=1}^{T-1} \mathrm{Re}\left(\frac{I_k(\boldsymbol{\lambda}_j)}{\phi_k(\theta; \boldsymbol{\lambda}_j)} \right),$$

that does not depend on $\overline{\theta}_T$. Then a consistent estimator for κ_k is obtained by plugging in a consistent estimator of θ into (13) for $k = 2, 3, 4$; see Theorem 2 in the next section. The modified concentrated OF is defined as $L_{kT}(\theta, \kappa_{kT}^{\dagger}(\theta))$ and, for simplicity in the analytical derivation, it is of interest to work with a rescaled

version of this OF, namely,

$$
\begin{aligned}
L_{kT}^*(\theta) &= \frac{(2\pi)^{k-2}}{2k} L_{kT}\big(\theta, \kappa_{kT}^\dagger(\theta)\big) \\
&= \frac{(2\pi)^{2(k-1)}}{2k\overline{\kappa}_{2T}^k T^{k-1}} \sum_{j=1}^{T-1} \frac{|I_k(\boldsymbol{\lambda}_j) - f_{kT}^\dagger(\theta; \boldsymbol{\lambda}_j)|^2}{|\phi_k(\overline{\theta}_T; \boldsymbol{\lambda}_j)|^2},
\end{aligned}
$$

(14)

where $f_{kT}^\dagger(\theta; \boldsymbol{\lambda}_j) = f_k(\theta, \kappa_{kT}^\dagger(\theta); \boldsymbol{\lambda}_j)$.

4.2. *Asymptotic properties of estimators based on minimizing a weighted linear combination of L_{kT}^*.*   The OF $L_{kT}^*(\theta)$ extracts the information on θ contained in f_k only. For identification, we need to consider f_k for which $\kappa_k \neq 0$, and $k > 2$. In addition, although f_2 does not help for global identification, it is very important for efficiency purposes. In this article, we aim to combine the information contained in f_2, f_3 and f_4 (note that, when both $\kappa_3 = \kappa_4 = 0$, additional higher order f_k should be considered). In order to combine these sources of information, we initially consider general minimum distance estimators based on minimizing the weighted sum

(15) $$L_{wT}^*(\theta) = w_2 L_{2T}^*(\theta) + w_3 L_{3T}^*(\theta) + w_4 L_{4T}^*(\theta),$$

where $\underline{w} = (w_2, w_3, w_4)'$ are some nonnegative weights that can give more emphasis to information from a particular cumulant. Then define the general family of minimum distance frequency domain estimators

$$\theta_{wT} = \arg\min_{\theta \in \Theta} L_{wT}^*(\theta).$$

Including both $L_{3T}^*(\theta)$ and $L_{4T}^*(\theta)$ in (15), guarantees that θ_{wT} is consistent as long as $\kappa_3^0 \neq 0$ or $\kappa_4^0 \neq 0$. In addition, even in the case where either $\kappa_3^0 = 0$ or $\kappa_4^0 = 0$, the presence in (15) of the particular $L_{kT}^*(\theta)$ for which $\kappa_k^0 = 0$, does not affect the consistency of θ_{wT}, since when $\kappa_k^0 = 0$ it holds that, as $T \to \infty$,

$$\sup_{\theta \in \Theta} |L_{kT}^*(\theta) - C_T| \to_p 0,$$

where C_T depends on the data, but does not depend on θ.

Note also that, by setting $w_k = 1$ and $w_j = 0$, $j \neq k$, the estimator θ_{wT} includes, as particular cases of interest, the marginal estimators $\theta_T^{(k)}$, $k = 2, 3, 4$, which minimize only $L_{kT}^*(\theta)$.

Although specific examples of θ_{wT} have been proposed previously in the literature [see, for instance, Chapter 4 in Terdik (1999)], the class of estimators θ_{wT} presents the problem that it is not possible to derive closed form optimal weights $\underline{w}$ for the general noninvertible/noncausal case. Despite that, for a given set of weights $\underline{w}$, θ_{wT} can be used as a robust initial estimate for developing a feasible

efficient estimator, which we analyze in the next section. Hence, we next establish the asymptotic properties of θ_{wT} under general conditions.

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

ASSUMPTION 1. Y_t follows an ARMA(p, q) model (6) with true roots φ_0 satisfying Assumption A(p, q) with $\theta_0 \in \Theta \subset \Gamma_{p,q}$, Θ compact.

ASSUMPTION 2.   The process ε_t is an i.i.d. sequence with zero mean, variance $\kappa_2^0 > 0$, $|\kappa_3^0| + |\kappa_4^0| > 0$, and $E[\varepsilon_t^8] < \infty$.

Under these conditions the results (3) and (4) hold because the ARMA model implies smooth spectral densities. The next two theorems establish the asymptotic properties of the estimators θ_{wT}.

THEOREM 2.   *Under Assumptions 1 and 2, with* $\min\{w_3, w_4\} > 0$, *as* $T \to \infty$,

$$\theta_{wT} \to_p \theta_0$$

and

$$\kappa_{kT}^{\dagger}(\theta_{wT}) \to_p \kappa_k^0, \qquad k = 2, 3, 4.$$

We emphasize that the previous consistency result just requires that either κ_3^0 or κ_4^0 is nonzero. In practice, we set both w_3 and w_4 positive since we might not have a priori information on which one is nonnegligible. The value of w_2 is irrelevant for consistency of θ_{wT} since second moments are of no use for identification, and hence L_{2T}^* does not distort the identification provided from higher order cumulants. The fact that $\kappa_{kT}^{\dagger}(\theta_{wT})$ is able to recover the true value of κ_k^0, even if this is zero, will be employed in the next subsection to propose an efficient estimator for θ, valid for general linear processes.

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

$$(16) \qquad\qquad \Phi_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi^1(\theta_0; \lambda) \varphi^1(\theta_0; -\lambda)' d\lambda,$$

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

$$(17) \qquad \varphi(\theta_0; \lambda) = \frac{\dot{\phi}(\theta_0; \lambda)}{\phi(\theta_0; \lambda)}, \qquad \mu(\theta_0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(\theta_0; \lambda) d\lambda,$$

and

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

Φ_0 is positive definite under Assumption A excluding common and unit roots in the AR and MA polynomials. 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. Also call $\bar{\mu}_4 = \mu_4/(\kappa_2^0)^2 = \nu_4 + 3$, $\bar{\mu}_5 = \mu_5/(\kappa_2^0)^{5/2} = \nu_5 + 10\nu_3$ and $\bar{\mu}_6 = \mu_6/(\kappa_2^0)^3 = \nu_6 + 10\nu_3^2 + 15\nu_4 + 15$ to the standardized fourth, fifth and sixth moments, respectively.

Finally, define the matrix $\mathbb{V} = \mathbb{V}(\theta_0, \nu_3, \nu_4)$ with blocks $\mathbb{V}_{j,k}$, $j, k = 2, 3, 4$, as

$$(18) \quad \mathbb{V} = \begin{pmatrix} \Phi_0 + \Phi_0^* & \nu_3^2(\Phi_0 + \Phi_0^*) & \bar{\mu}_4\nu_4(\Phi_0 + \Phi_0^*) \\ \nu_3^2(\Phi_0 + \Phi_0^*) & \nu_3^2\{(2+\nu_4)\Phi_0 + \nu_3^2\Phi_0^*\} & \nu_3\nu_4\{(\bar{\mu}_5 - \nu_3)\Phi_0 + \bar{\mu}_4\nu_3\Phi_0^*\} \\ \bar{\mu}_4\nu_4(\Phi_0 + \Phi_0^*) & \nu_3\nu_4\{(\bar{\mu}_5 - \nu_3)\Phi_0 + \bar{\mu}_4\nu_3\Phi_0^*\} & \nu_4^2\{(\bar{\mu}_6 - \nu_3^2)\Phi_0 + \bar{\mu}_4^2\Phi_0^*\} \end{pmatrix}.$$

THEOREM 3.   *Under Assumptions 1 and 2, with* $\min\{w_3, w_4\} > 0$, $\theta_0 \in$ Int(Θ), *as* $T \to \infty$,

$$(19) \qquad\qquad \sqrt{T}(\theta_{wT} - \theta) \to_d N\big(0, \Sigma_0^{-1}\Omega_0\Sigma_0^{-1}\big),$$

where

$$\Sigma_0 = \Sigma_0(w) = (w_2 + w_3\nu_3^2 + w_4\nu_4^2)\Phi_0 + w_2\Phi_0^*$$

is positive definite and

$$\Omega_0 = \Omega_0(w) = \sum_{j,k=2}^{4} w_j w_k \mathbb{V}_{j,k}.$$

Notice again that the previous theorem covers the case when one higher order cumulant (κ_3^0 or κ_4^0, but not both) is zero. In any of these situations the asymptotic distribution is derived using the Cramér–Wold device (see Lemma A.4 in the Appendix), with the implicit assumption that X_T converges in distribution to a $N(0, \Sigma)$ distribution with Σ singular as $T \to \infty$ if and only if $c'X_T$ converges to $c'X$, which is a $N(0, c'\Sigma c)$, for any $c \in R^k$, so, for instance, the $N(0, 0)$ is the 0-point mass distribution.

The expressions for Σ_0 and Ω_0 illustrate that the (relatively) larger are w_3 or ν_3^2 and w_4 or ν_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_{wT}

on higher order moments is reflected in the elements of $\mathbb{V}$. From Theorem 3, it is immediate to derive the asymptotic behavior of the individual $\theta_T^{(k)}$ (obviously for $\theta_T^{(2)}$ we have to presume identification).

COROLLARY 1. *Under the conditions of Theorem* 3, *and assuming that* Θ *for* $\theta_T^{(2)}$ *correctly identifies the location of the roots of* θ_0, $\kappa_3^0 \neq 0$ *and* $\kappa_4^0 \neq 0$, *as* $T \to \infty$,

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

where

$$(20) \quad \mathbb{W} = \begin{pmatrix} (\Phi_0 + \Phi_0^*)^{-1} & \Phi_0^{-1} & \frac{\bar{\mu}_4}{\nu_4}\Phi_0^{-1} \\ \Phi_0^{-1} & \frac{\bar{\mu}_4 - 1}{\nu_3^2}\Phi_0^{-1} + \Phi_0^{-1}\Phi_0^*\Phi_0^{-1} & \frac{\bar{\mu}_5 - \nu_3}{\nu_3\nu_4}\Phi_0^{-1} + \frac{\bar{\mu}_4}{\nu_4}\Phi_0^{-1}\Phi_0^*\Phi_0^{-1} \\ \frac{\bar{\mu}_4}{\nu_4}\Phi_0^{-1} & \frac{\bar{\mu}_5 - \nu_3}{\nu_3\nu_4}\Phi_0^{-1} + \frac{\bar{\mu}_4}{\nu_4}\Phi_0^{-1}\Phi_0^*\Phi_0^{-1} & \frac{\bar{\mu}_6 - \nu_3^2}{\nu_4^2}\Phi_0^{-1} + \left(\frac{\bar{\mu}_4}{\nu_4}\right)^2\Phi_0^{-1}\Phi_0^*\Phi_0^{-1} \end{pmatrix}.$$

Our results can be compared with previous CLTs in the related literature. A particular case of interest, covered by Theorem 3, occurs when both κ_3^0 or κ_4^0 are zero. For example, 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 asymptotic variance is a particular case of ours. For establishing the CLT, the i.i.d. restriction could be relaxed to martingale difference sequence with constant conditional homoskedasticity (and additional higher order conditional moment restrictions) at cost of further notational complexity. The extension to the conditional heteroskedasticity case is more challenging since this will affect the parametric form of f_4 and left open for future research.

Kumon (1992) 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 $\theta_T^{(3)}$ differs from ours due to two important aspects: he is not centering by $\mu(\theta_0)$ in expression (16), which is needed for noninvertible processes, and the scaling in his expression (3.9) is adjusted by the fact that we only use the real part of the biperiodogram in the FOC.

Terdik [(1999), Theorem 76] considers a loss function that weights second and third moments, with simultaneous normalization of the periodograms, although he does not analyze the contribution of this modification to the asymptotic variance. The results in Terdik can be compared to ours. In particular, for the invertible case setting $w_2 = 2p_1$, $w_3 = 3q_1$ and $w_4 = 0$, the expression for the asymptotic

variance is

$$\Sigma_0^{-1} \Omega_0 \Sigma_0^{-1} = \frac{w_2^2 + \{2w_3 w_2 + w_3^2(2 + v_4)\}v_3^2}{(w_2 + w_3 v_3^2)^2} \Phi_0^{-1},$$

which in Terdik's notation would be

$$(21) \qquad \frac{4p_1^2 + \{9q_1^2 v_4 + 18q_1^2 + 12p_1 q_1\}v_3^2}{(2p_1 + 3q_1 v_3^2)^2} \Phi_0^{-1}.$$

Notice that (21) differs from Terdik's expression (4.21) in the factor $12p_1 q_1$ (versus $6p_1 q_1$), which corresponds to the term $pq\varsigma^2$ in his expression for $\Sigma_{R1}(\vartheta_0)$. This discrepancy seems to originate from ignoring the 2 factor in front of the definition of $\Sigma_{R23}(\vartheta_0)$ (page 167). Note that $p_1 + q_1 = 1$ is imposed in Terdik (1999) in terms of a geometric motivation, but this normalization is not needed if $\min\{p_1, q_1\} \geq 0$, as it happens with our weights w_j when aggregating loss functions to be minimized.

Corollary 1 shows that the asymptotic variances of $\theta_T^{(3)}$ and $\theta_T^{(4)}$ are smaller the larger are v_3^2 and v_4^2, respectively, that is, the more information is contained in these particular higher order moments. In addition, notice that for pure invertible and causal processes, $\mathbb{W}$ simplifies to

$$\mathbb{W} = \begin{pmatrix} 1 & 1 & \dfrac{\bar{\mu}_4}{v_4} \\[2mm] 1 & \dfrac{\bar{\mu}_4 - 1}{v_3^2} & \dfrac{\bar{\mu}_5 - v_3}{v_3 v_4} \\[2mm] \dfrac{\bar{\mu}_4}{v_4} & \dfrac{\bar{\mu}_5 - v_3}{v_3 v_4} & \dfrac{\bar{\mu}_6 - v_3^2}{v_4^2} \end{pmatrix} \otimes \Phi_0^{-1},$$

using that in this case $\Phi_0^* = 0$. Then, the asymptotic variance of $\theta_T^{(2)}$ for invertible and causal models equals the one of the Whittle estimate because

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

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 of the Supplementary Material [Velasco and Lobato (2018)] we show that these estimates are also asymptotically equivalent in this case. In addition, it is simple to show that both $(\bar{\mu}_4 - 1)/v_3^2$ and $(\bar{\mu}_6 - v_3^2)/v_4^2$ are larger than one. Hence, both $\theta_T^{(3)}$ and $\theta_T^{(4)}$ would be less efficient estimators than $\theta_T^{(2)}$, in case this estimator were consistent, which is not in general. Finally, the fact that the matrix $\mathbb{W}$ is nondiagonal implies that a consistent, and more efficient than $\theta_T^{(2)}$, estimator can be constructed by linearly combining $\theta_T^{(2)}$, $\theta_T^{(3)}$ and $\theta_T^{(4)}$.

REMARK 1.   Note that for invertible processes μ_0, defined in (17), 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 2.   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

$$f_2(\beta_1, \alpha_1, \kappa_2^0; \lambda) = \frac{\kappa_2^0}{2\pi} \frac{|1 + \beta_1 e^{i\lambda}|^2}{|1 - \alpha_1 e^{i\lambda}|^2} = \frac{\kappa_2^0}{2\pi} \frac{|1 + \theta_0 e^{i\lambda}|^2}{|1 + \theta_0^{-1} e^{i\lambda}|^2} = \frac{\kappa_2^0}{2\pi} \theta_0^2.$$

Then,

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

since

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

REMARK 3.   In the case of restricted estimation of an all-pass model of order r with

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

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

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

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 $\theta_T^{(3)}$ and $\theta_T^{(4)}$ are, respectively,

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

4.3. *Efficient estimation based on second, third and fourth moments.* Implementation of the estimator θ_{wT} requires selecting optimally the weights $\underline{w}$. In the absence of a priori information, the weights should be selected using sample information. Given the complex structure of the asymptotic covariance matrix in (19), determining the optimal weights for the general non causal/non invertible case is an impossible task. In the working paper version of the article,[3] we address optimally selecting the weights $\underline{w}$ in the invertible and causal case. That result is obviously of limited interest for the general problem studied in this article.

In this subsection, we propose an alternative approach to construct a feasible estimator that is consistent and optimally weights the information contained in L_{2T}^*, L_{3T}^* and L_{4T}^*. The motivation for this estimator comes from observing that the FOC that defines the estimator θ_{wT} is an arbitrary linear combination of the scores of L_{2T}^*, L_{3T}^* and L_{4T}^*. Hence, a more efficient estimator than θ_{wT} can be constructed by optimally combining the information contained in these scores. In particular, denote the gradient vector by

$$\mathbb{S}_T(\theta) = \begin{pmatrix} \dfrac{\partial}{\partial \theta} L_{2T}^*(\theta) \\[2mm] \dfrac{\partial}{\partial \theta} L_{3T}^*(\theta) \\[2mm] \dfrac{\partial}{\partial \theta} L_{4T}^*(\theta) \end{pmatrix},$$

and recall that the asymptotic variance of $T^{1/2}\mathbb{S}_T(\theta_0)$ is the matrix $\mathbb{V}$ defined in (18). We consider the objective function

(22) $$Q_T(\theta) = \mathbb{S}_T(\theta)' \hat{\mathbb{V}}_T^- \mathbb{S}_T(\theta),$$

---

[3]Available at https://sites.google.com/site/ignacionlobato/.

where $\hat{\mathbb{V}}_T^-$ is a consistent estimator of the matrix $\mathbb{V}^-$, which is a reflexive generalized inverse of $\mathbb{V}$, and hence, it satisfies $\mathbb{V}\mathbb{V}^-\mathbb{V} = \mathbb{V}$ and $\mathbb{V}^-\mathbb{V}\mathbb{V}^- = \mathbb{V}^-$. A leading example of $\mathbb{V}^-$ is the Moore–Penrose inverse, $\mathbb{V}^+$. We employ generalized inverses to take account of cases where the asymptotic variance of $\mathbb{S}_T(\theta)$ is default rank due to some higher order cumulant being zero. The OF (22) optimally weights the information on θ_0 contained in the score vector $\mathbb{S}_T(\theta)$. Hence, the proposed consistent and efficient estimator of θ_0 is

$$(23) \qquad \hat{\theta}_T = \tilde{\theta}_T - \big(\mathbb{H}_T(\tilde{\theta}_T)'\hat{\mathbb{V}}_T^-\mathbb{H}_T(\tilde{\theta}_T)\big)^{-1}\mathbb{H}_T(\tilde{\theta}_T)'\hat{\mathbb{V}}_T^-\mathbb{S}_T(\tilde{\theta}_T),$$

where $\mathbb{H}_T(\theta) = (\partial/\partial\theta')\mathbb{S}_T(\theta)$, and $\tilde{\theta}_T$ satisfies

$$(24) \qquad\qquad\qquad \tilde{\theta}_T - \theta_0 = O_p(T^{-1/2}).$$

Finally, define $\mathbb{H} = (\Phi_0 + \Phi_0^*, \nu_3^2\Phi_0, \nu_4^2\Phi_0)'$, and notice that, using the same arguments as for Theorem 3,

$$\mathbb{H}_T(\tilde{\theta}_T) = \frac{\partial}{\partial\theta'}\mathbb{S}_T(\theta_0) + o_p(1) = \mathbb{H} + o_p(1).$$

For implementing (23), we can employ for $\tilde{\theta}_T$ some version of θ_{wT}, which, as we have seen in the previous section, verifies condition (24) and is robust to a zero higher order cumulant. For instance, in the simulations in the next section we employ as $\tilde{\theta}_T$ the estimator that minimizes $L_{3T}^* + L_{4T}^*$.

Given (24), the consistency of $\hat{\theta}_T$ is trivial. The next theorem states its asymptotic distribution.

THEOREM 4. *Under Assumptions* 1 *and* 2 (*irrespective of the values of κ_3^0 and κ_4^0*), (24) *and* $\hat{\mathbb{V}}_T^- \to_p \mathbb{V}^-$ *as* $T \to \infty$,

$$\sqrt{T}(\hat{\theta}_T - \theta) \to_d N\big(0, (\mathbb{H}'\mathbb{V}^-\mathbb{H})^{-1}\big).$$

REMARK 4. The asymptotic variance matrix $(\mathbb{H}'\mathbb{V}^-\mathbb{H})^{-1}$ is positive definite because Φ_0 is so, even when $\kappa_3^0 = \kappa_4^0 = 0$. Note that the matrices $\mathbb{V}^-$ and $\mathbb{H}'\mathbb{V}^-\mathbb{H}$ are not unique in general. However, $(\mathbb{H}'\mathbb{V}^-\mathbb{H})^{-1}$ is invariant to the particular choice of $\mathbb{V}^-$ when $\mathbb{H}$ belongs to the column space of $\mathbb{V}$; see, for example, Rao and Mitra (1972), page 615. Furthermore, in this case, using the same argument as in Theorem 5 of Schneeweiss (2014), it can be showed that $\hat{\theta}_T$ is efficient in the class of estimators that minimize $\mathbb{S}_T(\theta)'\mathbb{W}_T\mathbb{S}_T(\theta)$, where $\mathbb{W}_T$ is any weighting matrix sequence that preserves identification asymptotically, that is, for which $\mathbb{H}'\mathbb{W}\mathbb{H}$ is positive definite, where $\mathbb{W} = \text{plim}_{T\to\infty}\mathbb{W}_T$.

REMARK 5. It is immediate to show that $\mathbb{H}$ belongs to the column space of $\mathbb{V}$ for invertible and causal processes, since for these processes $\Phi_0^* = 0$ and both $\mathbb{H}$ and $\mathbb{V}$ are multiples of Φ_0, irrespective of whether κ_3^0 and/or κ_4^0 are zero or not.

In this situation, it is also simple to show, using the Cauchy–Schwarz inequality, that, for any choice of weights, $\Sigma_0^{-1}\Omega_0\Sigma_0^{-1} - (\mathbb{H}'\mathbb{V}^-\mathbb{H})^{-1}$ is a positive semi-definite matrix, so that $\hat{\theta}_T$ is asymptotically at least as efficient as θ_{wT} for any $\underline{w}$. The efficiency condition is also satisfied when $\kappa_3^0 = \kappa_4^0 = 0$ and when $p + q = 1$, because the structure or the dimensions of $\mathbb{H}$ and $\mathbb{V}$ guarantee that there always exists a matrix A so that $\mathbb{V}A = \mathbb{H}$. Hence, the estimator $\hat{\theta}_T$ is consistent, efficient for a large class of processes, and simple to calculate since it avoids the optimal selection of the weights required for implementation of θ_{wT}.

REMARK 6.   Note that Theorem 4 allows for the case where both higher order cumulants are zero, $\kappa_3^0 = \kappa_4^0 = 0$, given (24). In this case, all blocks of $\mathbb{V}$ are equal to zero except the top left one, and if we choose the Moore–Penrose inverse $\mathbb{V}^+$ for $\mathbb{V}^-$, it has the same shape in terms of the inverse of this block, so that the asymptotic variance is just given by

$$(\mathbb{H}'\mathbb{V}^+\mathbb{H})^{-1} = (\Phi_0 + \Phi_0^*)^{-1},$$

which is the one obtained by using only L_{2T}^* under the assumption of the correct knowledge of the exact location with respect to the unit circle of the roots of the lag polynomials implied by θ_0; cf. Corollary 1.

REMARK 7.   Notice that Theorem 4 requires that $\hat{\mathbb{V}}_T^- \to_p \mathbb{V}^-$ as $T \to \infty$. For constructing estimators $\hat{\mathbb{V}}_T^-$ that satisfy this condition, it is not sufficient to consider the generalized inverse of any consistent estimate $\hat{\mathbb{V}}_T$ of $\mathbb{V}$, since generalized inverses are not necessarily continuous. There are several approaches to guarantee that $\hat{\mathbb{V}}_T^- \to_p \mathbb{V}^-$. One simple solution is to require that $\Pr(\text{rank}(\hat{\mathbb{V}}_T) = \text{rank}(\mathbb{V})) \to 1$ as $T \to \infty$, when using the Moore–Penrose inverse $\mathbb{V}^+$; cf. Theorem 2 in Andrews (1987). Notice that in our problem $\mathbb{V}$ becomes singular only when $\kappa_k^0 = 0$, $k = 3$ and/or 4, and similarly for $\hat{\mathbb{V}}_T$ in terms of the estimates of the standardized cumulants ν_k. Hence, for our case it is simple to provide explicitly an estimator $\hat{\mathbb{V}}_T$ that guarantees the previous rank condition. In particular, we could choose as $\hat{\mathbb{V}}_T$ the sample analogue of (18), where we replace ν_k by

$$\tilde{\nu}_{kT} = \frac{\kappa_{kT}^\dagger(\tilde{\theta}_T)}{\kappa_{2T}^\dagger(\tilde{\theta}_T)^{k/2}} 1\big(\big|\kappa_{kT}^\dagger(\tilde{\theta}_T)\big| > \zeta_T\big), \qquad k = 3, 4,$$

for some threshold parameter ζ_T such that $\zeta_T + T^{-1/2}\zeta_T^{-1} \to 0$ as $T \to \infty$. Similarly for the other components for $\hat{\mathbb{V}}_T$, population moments would be replaced by the corresponding expressions using consistent cumulants estimates $\kappa_{kT}^\dagger(\tilde{\theta}_T)$, $k = 2, \ldots, 6$ (which would entail strengthening the moment conditions in Assumption 2), and Φ_0 and Φ_0^* would be replaced by the same expressions using $\tilde{\theta}_T$ instead

of θ_0. Then, using the methods in the proof of Theorem 3 it is simple to prove that under (24)

$$T^{1/2}(\kappa_{kT}^{\dagger}(\tilde{\theta}_T) - \kappa_k^0) = O_p(1), \qquad k = 2, 3, 4,$$

so that $\zeta_T^{-1}\kappa_{kT}^{\dagger}(\tilde{\theta}_T) = o_p(1)$ when $\kappa_k^0 = 0$, while $\zeta_T^{-1}\kappa_{kT}^{\dagger}(\tilde{\theta}_T) = \zeta_T^{-1}\kappa_k^0 + o_p(1)$ when $\kappa_k^0 \neq 0$, $k = 3, 4$, and therefore,

$$\Pr(\text{rank}(\hat{\mathbb{V}}_T) \neq \text{rank}(\mathbb{V}))$$

$$\leq \Pr\left(\bigcap_{k:\kappa_k^0=0} \{|\kappa_{kT}^{\dagger}(\tilde{\theta}_T)| > \zeta_T\} \right) + \Pr\left(\bigcap_{k:\kappa_k^0\neq0} \{|\kappa_{kT}^{\dagger}(\tilde{\theta}_T)| \leq \zeta_T\} \right)$$

$$\to 0,$$

as $T \to \infty$ and therefore $\hat{\mathbb{V}}_T^+ \to_p \mathbb{V}^+$.

REMARK 8.   The analytic expressions for the components of the score vector are given by

$$\frac{\partial}{\partial\theta}L_{kT}^*(\theta) = -\frac{(2\pi)^{2(k-1)}}{k\overline{\kappa}_{2T}^k T^{k-1}} \sum_{j=1}^{T-1} \frac{\text{Re}\{(I_k(\lambda_j) - f_{kT}^{\dagger}(\theta;\lambda_j))\frac{\partial}{\partial\theta}f_{kT}^{\dagger}(\theta;-\lambda_j)\}}{|\phi_k(\bar{\theta}_T;\lambda_j)|^2},$$

where here we can set $\bar{\theta}_T = \tilde{\theta}_T$ and $\overline{\kappa}_{2T} = \kappa_{2T}^{\dagger}(\tilde{\theta}_T)$, while $\mathbb{H}_T(\tilde{\theta}_T)$ can be replaced by its probability limit $\mathbb{H}$. Hence, we can avoid the analytical evaluation of the second order derivatives by using consistent estimates of Φ_0 and Φ_0^* as in the estimation of $\mathbb{V}$. In addition, in the Appendix B we proportionate approximating expressions for $\mathbb{S}_T(\theta)$ in terms of residuals. These approximations can be used to obtain direct expressions of all its components in the time domain.

5. Simulations.   In this section we report a short simulation exercise to assess two issues: first, the ability of $L_{3T}^* + L_{4T}^*$ to identify the proper location of the roots, and second, the finite sample performance (bias and root mean squared error) of the efficient estimator (23). We focus on a simple MA(1) process, and consider three innovation distributions: an exponential, a Student's t with 5 degrees of freedom, and a Uniform. The last two distributions are symmetric, so $\nu_3 = 0$, whereas the exponential distribution is highly asymmetric, $\nu_3 = 2$. The first two distributions have the same positive kurtosis $\nu_4 = 6$, while the uniform has negative kurtosis, $\nu_4 = -6/5$. Moments higher than the fourth are not defined for the Student's t case, and so the theoretical results are not readily applicable for this example, but otherwise Assumption 2 holds in all cases.

Table 1 reports the percentage of replications where the estimator based on minimizing $L_{3T}^* + L_{4T}^*$ identifies correctly the location of the root of the MA component, that is, $L_{3T}^*(\theta) + L_{4T}^*(\theta)$ is minimized for a value θ_T so that $\mathbb{I}\{|\theta_T| < 1\} =$

TABLE 1
Percentage correct root identification using $L_{3T}^* + L_{4T}^*$

Distribution	T	θ_0			
		0.5	**0.9**	**0.9^{-1}**	**0.5^{-1}**
Exp(1)	100	96.3	79.8	80.7	96.0
	200	99.5	86.0	87.1	98.9
t_5	100	85.0	70.4	69.8	86.5
	200	93.0	77.9	79.5	93.6
Uniform	100	73.6	55.9	56.3	73.9
	200	93.7	59.2	59.4	93.2

Note: Percentage of replications in which $L_{3T}^*(\theta) + L_{4T}^*(\theta)$ is minimized for a value θ_T so that $\mathbb{I}\{|\theta_T| < 1\} = \mathbb{I}\{|\theta_0| < 1\}$.

$\mathbb{I}\{|\theta_0| < 1\}$. We consider two sample sizes ($T = 100$ and 200), four values for θ_0 (0.5, 0.9 and their inverse values) and the number of replications is 3000. We just report the results for the estimator that minimizes $L_{3T}^* + L_{4T}^*$ since that is the one that we employ as $\tilde{\theta}_T$ when computing (23) in the second part of the simulations results. Obviously, alternative versions of θ_{wT} could be employed as $\tilde{\theta}_T$, for instance L_{2T}^* could also be added, but the inclusion of L_{2T}^* does not theoretically add any benefit for the correct identification of the location of the roots. From Table 1, it is clear that correct identification is easiest for the exponential case, which could be expected given that both the third and fourth cumulants are non zero. Obviously, L_{3T}^* does not provide any identification information for the uniform and the Student t distributions since the third cumulant is zero, $\nu_3 = 0$, for these two cases, as was confirmed in the working paper version of the article where the results using L_{3T}^* and L_{4T}^* individually were reported. Note also that for moderate sample sizes the uniform case is more complicated than the Student's t case, most likely due to the much smaller value of ν_4^2. In addition, as could be expected, identification is always easier when the parameter is further away of the unit root circle.

For the cases where the estimator based on minimizing $L_{3T}^* + L_{4T}^*$ properly identifies the location of the MA root, we can compare the bias and the RMSE of the proposed estimator $\hat{\theta}_T$ and of the unfeasible version of the $\theta_T^{(2)}$ estimator, which is asymptotically as efficient as the Whittle estimator. This version imposes correct knowledge of the location of the roots (since identification is impossible using just L_{2T}^*). Implementation of the proposed estimator (23) has employed plugging in estimators for the unknown cumulants for constructing $\hat{\mathbb{V}}_T$ and $\mathbb{H}_T$, and a numerical approximation for the score vector $\mathbb{S}_T$. In addition, the usual inverse of $\hat{\mathbb{V}}_T$ was employed for constructing (23) in all cases. Implementing the estimator $\hat{\mathbb{V}}_T^-$ described in Remark 7 is hampered by the selection of the threshold parameter ζ_T, and we leave this matter for future research. Table 2 reports the bias and the RMSE

TABLE 2
Bias and RMSE for $\theta_T^{(2)}$ and $\hat{\theta}_T$

Distribution	T		θ_0			
			0.5	0.9	0.9^{-1}	0.5^{-1}
			Bias			
Exp(1)	100	$\theta_T^{(2)}$	−0.008	−0.108	0.093	0.093
		$\hat{\theta}_T$	0.004	−0.030	0.083	0.080
	200	$\theta_T^{(2)}$	0.001	−0.072	0.073	0.037
		$\hat{\theta}_T$	0.002	−0.002	0.023	0.027
t_5	100	$\theta_T^{(2)}$	−0.010	−0.106	0.092	0.111
		$\hat{\theta}_T$	−0.008	−0.037	0.076	0.101
	200	$\theta_T^{(2)}$	−0.002	−0.056	0.072	0.029
		$\hat{\theta}_T$	0.001	−0.027	0.034	0.023
Uniform	100	$\theta_T^{(2)}$	−0.014	−0.093	0.108	0.102
		$\hat{\theta}_T$	0.003	−0.085	0.079	0.081
	200	$\theta_T^{(2)}$	−0.001	−0.059	0.065	0.039
		$\hat{\theta}_T$	−0.000	−0.022	0.041	0.022
			RMSE			
Exp(1)	100	$\theta_T^{(2)}$	0.089	0.158	0.288	0.423
		$\hat{\theta}_T$	0.098	0.146	0.284	0.438
	200	$\theta_T^{(2)}$	0.068	0.125	0.170	0.268
		$\hat{\theta}_T$	0.068	0.120	0.164	0.264
t_5	100	$\theta_T^{(2)}$	0.109	0.147	0.219	0.510
		$\hat{\theta}_T$	0.103	0.140	0.221	0.498
	200	$\theta_T^{(2)}$	0.064	0.106	0.150	0.275
		$\hat{\theta}_T$	0.066	0.102	0.143	0.271
Uniform	100	$\theta_T^{(2)}$	0.090	0.115	0.218	0.416
		$\hat{\theta}_T$	0.082	0.100	0.172	0.386
	200	$\theta_T^{(2)}$	0.069	0.094	0.141	0.267
		$\hat{\theta}_T$	0.059	0.082	0.119	0.221

Note: $\theta_T^{(2)}$ is obtained searching only for $|\theta| < 1$ or $|\theta| > 1$ as appropriate given the true value θ_0, while $\hat{\theta}_T$ uses as $\tilde{\theta}_T$ the estimate obtained minimizing $L_{3T}^* + L_{4T}^*$.

for both estimators for two sample sizes $T = 100$ and 200, and 1000 replications. Table 2 shows that bias is larger for values of θ_0 closer to one, and that the sign of the bias depends on θ_0: for values of θ_0 less than one bias is typically negative, while for θ_0 larger than one the bias is positive. In general, the estimator $\hat{\theta}_T$ exhibits less bias than the $\theta_T^{(2)}$ estimator. In addition, in terms of RMSE, the bias plays a minor role. The RMSE is increasing with the values for θ_0 (though roughly constant in relative terms) and appears to be somewhat smaller for the uniform distribution. For the uniform platykurtic distribution, the $\hat{\theta}_T$ efficient estimator clearly dominates the infeasible $\theta_T^{(2)}$ estimator in terms of RMSE for the considered sample sizes. This might indicate that at least for symmetric platykurtic distributions there are potential efficiency improvements, which the working paper version of the article confirmed theoretically for the causal and invertible case. However, for the other two, leptokurtic, distributions $\hat{\theta}_T$ only presents slightly smaller RMSE than the unfeasible $\theta_T^{(2)}$ estimator for most of the cases.

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, as long as the third or the fourth order cumulant is different from zero, the proposed estimator is consistent under general assumptions. Furthermore, we provide one-step valid inference irrespective of exact knowledge of the location of the ARMA roots and even for the case when the third or the fourth or both cumulants are zero. In addition, our procedures are efficient for a large class of processes, which include the case where both the third and fourth cumulants are zero. In that case, we have shown that our procedures are asymptotically as efficient as the unfeasible Whittle estimator that imposes correct knowledge of the location of the roots.

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 particularly relevant for this case, an example would be ARMA-GARCH models. Additional extensions of interest are the use of the bootstrap to estimate the standard errors of the estimates, the analysis of the multivariate case, which is of special interest in Economics, and the study of the implementation of automatic criteria, such as AIC or BIC, to select the order of the ARMA model.

SUPPLEMENTARY MATERIAL

Technical Appendices to "Frequency domain minimum distance inference for possibly noninvertible and noncausal ARMA models" (DOI: 10.1214/17-AOS1560SUPP; .pdf). This Supplementary Material contains three appendices with proofs of main results, technical lemmas and comparison with Whittle estimation.

REFERENCES

ALEKSEEV, V. G. (1993). Asymptotic properties of higher-order periodograms. *Theory Probab. Appl.* **40** 409–419.

ALESSI, L., BARIGOZZI, M. and CAPASSO, M. (2011). Non-fundamentalness in structural econometric models: A review. *Int. Stat. Rev.* **79** 16–47.

ANDREWS, D. W. (1987). Asymptotic results for generalized Wald tests. *Econometric Theory* **3** 348–358.

ANDREWS, B., DAVIS, R. A. and BREIDT, F. J. (2007). Rank-based estimation for all-pass time series models. *Ann. Statist.* **35** 844–869. MR2336871

ANH, V. V., LEONENKO, N. N. and SAKHNO, L. M. (2007). Minimum contrast estimation of random processes based on information of second and third orders. *J. Statist. Plann. Inference* **137** 1302–1331. MR2301481

BREIDT, F. J., DAVIS, R. A. and TRINDADE, A. A. (2001). Least absolute deviation estimation for all-pass time series models. *Ann. Statist.* **29** 919–946. MR1869234

BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory*. Holden-Day, San Francisco, CA. MR0595684

BRILLINGER, D. R. (1985). Fourier inference: Some methods for the analysis of array and nongaussian series data. *Water Resour. Bull.* **21** 744–756.

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York. MR1093459

GOSPODINOV, N. and NG, S. (2015). Minimum distance estimation of possibly noninvertible moving average models. *J. Bus. Econom. Statist.* **33** 403–417. MR3372667

HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York. MR0279952

HANSEN, L. P. and SARGENT, T. J. (1980). Formulating and estimating dynamic linear rational expectations models. *J. Econom. Dynam. Control* **2** 7–46. MR0601128

HANSEN, L. P. and SARGENT, T. J. (1991). Two difficulties in interpreting vector autoregressions. In *Rational Expectations Econometrics* 77–119. Westview Press, Boulder, CO.

HUANG, J. and PAWITAN, Y. (2000). Quasi-likelihood estimation of non-invertible moving average processes. *Scand. J. Stat.* **27** 689–702. MR1804170

KUMON, M. (1992). Identification of non-minimum phase transfer function using higher-order spectrum. *Ann. Inst. Statist. Math.* **44** 239–260. MR1177462

LANNE, M. and SAIKKONEN, P. (2011). Noncausal autoregressions for economic time series. *J. Time Ser. Econom.* **3** Art. 2, 32. MR2928654

LEEPER, E. M., WALKER, T. B. and YANG, S.-C. S. (2013). Fiscal foresight and information flows. *Econometrica* **81** 1115–1145. MR3064062

LII, K.-S. and ROSENBLATT, M. (1992). An approximate maximum likelihood estimation for non-Gaussian non-minimum phase moving average processes. *J. Multivariate Anal.* **43** 272–299. MR1193615

MOUNTFORD, A. and UHLIG, H. (2009). What are the effects of fiscal policy shocks? *J. Appl. Econometrics* **24** 960–992. MR2750186

RAMSEY, J. B. and MONTENEGRO, A. (1992). Identification and estimation of non-invertible non-Gaussian MA(q) processes. *J. Econometrics* **54** 301–320.

RAO, C. R. and MITRA, S. K. (1972). Generalized inverse of a matrix and its applications. In *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* 601–620. Univ. California Press, Berkeley, CA. MR0403093

ROSENBLATT, M. (1985). *Stationary Sequences and Random Fields*. Birkhäuser, Boston, MA. MR0885090

SCHNEEWEISS, H. (2014). The linear GMM model with singular covariance matrix due to the elimination of a nuisance parameter. Technical Report 165, Dept. Statistics, Univ. Munich.

TERDIK, G. (1999). *Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis*: *A Frequency Domain Approach*. *Lecture Notes in Statistics* **142**. Springer, New York. MR1702281

VELASCO, C. and LOBATO, I. N. (2018). Supplement to "Frequency domain minimum distance inference for possibly noninvertible and noncausal ARMA models." DOI:10.1214/17-AOS1560SUPP.

WHITTLE, P. (1953). The analysis of multiple stationary time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **15** 125–139. MR0056902

DEPARTAMENTO DE ECONOMÍA
UNIVERSIDAD CARLOS III DE MADRID
CALLE MADRID 126
28903 GETAFE (MADRID)
SPAIN
E-MAIL: carlos.velasco@uc3m.es

CENTRO DE INVESTIGACIÓN ECONÓMICA
DEPARTMENT OF ECONOMICS
ITAM
RÍO HONDO 1, CIUDAD DE MÉXICO 01080
MÉXICO
E-MAIL: ilobato@itam.mx