The Asymptotic CRLB for the Spectrum of ARMA Processes

Brett Ninness*

Abstract

This paper addresses the issue of quantifying the frequency domain accuracy of ARMA spectral estimates as dictated by the Cramér–Rao Lower Bound (CRLB). Classical work in this area has led to expressions that are asymptotically exact as both data length and model order tend to infinity, although they are commonly used in finite model order and finite data length settings as approximations. More recent work has established quantifications which, for AR models, are exact for finite model order. By employing new analysis methods based on rational orthonormal parameterisations, together with the ideas of reproducing kernel Hilbert spaces, this paper develops quantifications that extend this previous work by being exact for finite model order in all of the AR, MA and ARMA system cases. These quantifications, via their explicit dependence on poles and zeros of the underlying spectral factor, reveal certain fundamental aspects of the accuracy achievable by spectral estimates of ARMA processes.

Technical Report EE2003012, Department of Electrical and Computer Engineering,
University of Newcastle, AUSTRALIA

1 Introduction

In a wide variety of applications including adaptive filtering, acoustics, econometrics, array processing, radar and speech processing it is necessary to estimate the correlation structure of a signal that can be modelled as a stationary stochastic process $\{y_t\}$.

This correlation structure is completely described by the spectral density $\Phi_y(\omega)$ of the process, and in turn this is often of interest in its own right. While there is a very large variety of methods available to estimate such a spectral density [12, 20, 4, 19], when it has a finite order rational form, the approach of using an ARMA model structure together with a Maximum–Likelihood criterion is well known to offer optimal accuracy, in the sense that the Cramér–Rao lower bound (CRLB) on parameter space variability is asymptotically achieved as the data length tends to infinity.

Via a first order Taylor series argument, this also implies that the associated estimate of the spectral density $\Phi_y(\omega)$ also asymptotically achieves its Cramér–Rao bound. This opens the question of quantifying what this bound on the estimate of $\Phi_y(\omega)$ is, both in order to inform what problem aspects might limit or enhance estimation accuracy, and also to actually quantify that accuracy.

Recognising this importance, several prior works [2, 15, 26, 10] have sought to find expressions for it. A central motif of those contributions has been to simplify what appear to be quite complex

*This work was supported by the Australian Research Council. This author is with the School of Electrical Engineering & Computer Science, University of Newcastle, Australia and can be contacted at email:`brett@ee.newcastle.edu.au` or FAX: +61 49 21 69 93.

expressions by a strategy of allowing the model order to tend to infinity, and then using the ensuing asymptotic in model order result as an approximate quantification applying for finite model order.

Very recently, in the special case of AR modelling, the work [24] has developed expressions that are not asymptotic in the model order, and the key idea in achieving the results there is to treat the afore-mentioned complex expression, which is a certain frequency dependent quadratic form, by introducing a so-called 'virtual' time series which allows the problem to be re-expressed with respect to a particular orthonormal basis examined in [17].

The paper here, while employing the same orthonormal basis as in [24], develops a completely different and new analysis technique which depends on recognising that the quadratic form quantifying the CRLB is, in fact, the reproducing kernel for a particular well defined space and hence, perhaps surprisingly, can be very simply expressed with respect to the afore-mentioned basis. The advantage of this approach is that it allows the simple accommodation of the full ARMA situation, with the AR and MA scenarios subsumed as special cases.

In particular, amongst other things, the work here establishes that with $\Phi_y(\omega,\widehat{\theta}_N)$ denoting an estimate of the spectrum $\Phi_y(\omega)$ based on an intervening estimate $\widehat{\theta}_N$ of the ARMA parameters, then

$$\lim_{N\to\infty} N\mathrm{Var}\left\{\frac{\Phi_y(\omega,\widehat{\theta}_N)}{\Phi_y(\omega)}\right\} = 2\left[\sum_{k=0}^{2m-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2} + \mathrm{Re}\left\{\sum_{k=0}^{2m-1}\frac{1-\xi_k^2}{(e^{j\omega}-\xi_k)^2}\prod_{\ell=0}^{k-1}\left(\frac{1-\xi_\ell e^{j\omega}}{e^{j\omega}-\xi_\ell}\right)^2\right\}\right]. \tag{1}$$

Here N is the length of the data record used to generate the parameter estimate $\widehat{\theta}_N$, and $\{\xi_0,\cdots,\xi_{2m-1}\}$ are the poles and zeros of the m'th order ARMA representation which here (but not later) are assumed all real valued and also here (but not later) it has been assumed that the minimal one-step ahead prediction error variance σ^2 is known.

This establishes, for example, that the spectral estimate will be less accurate at frequencies close to any poles *or* zeros of the underlying ARMA process. As well, it establishes a 'waterbed' effect, in that since the integral of the right hand side of (1) over $\omega\in[-\pi,\pi]$ equals $8m\pi$, then any increase in relative spectral estimation error at certain frequencies (eg. near poles or zeros close to the unit circle) must be balanced by commensurate decreases at other frequencies.

It should be emphasised that there is a fundamental difference between the concentration in this paper on the variability of *functions of* the parameter estimates and many well known previous works that have addressed the variability of the parameters themselves, and which are not the focus of this contribution.

For example [20, Appendix B.5], [12, pp293-296] and [8] all detail how the CRLB on ARMA *parameter* estimates may be reliably computed, and then indicate how the CRLB of functions of the parameters may then be numerically evaluated via computation of certain quadratic forms.

The paper here extends this pre-existing work to show how *closed form* expressions for these quadratic forms may be derived. Apart from adding new insight into the parametric spectral estimation problem, the closed forms presented here also provides alternative means for evaluating the associated CRLB that may be more reliable than previous methods.

Finally, the author believes that the reproducing kernel methods introduced here to derive the closed forms may have interest in their own right as a potential analysis tool for problems not considered in this paper.

2 Problem Formulation and Background

Suppose that $\{y_t\}$ is a wide sense stationary and zero mean stochastic process with spectral density $\Phi_y(\omega)$ which is assumed to be bounded away from zero so that the Paley–Wiener condition is satisfied, and hence $\{y_t\}$ is a regular process that possesses a Wold decomposition devoid of deterministic component as follows

$$y_t = e_t + \sum_{n=1}^{\infty} h_n e_{t-n}. \tag{2}$$

Here $\{e_t\}$ is a zero mean i.i.d. process of variance $\mathbf{E}\left\{e_t^2\right\} = \sigma^2$ and the spectral factor

$$H(z) \quad = \quad 1 + \sum_{n=1}^{\infty} h_n z^{-n} = \exp\left\{\frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\frac{z+e^{j\omega}}{z-e^{j\omega}}\right)\log\Phi_y(\omega)\,\mathrm{d}\omega\right\} \tag{3}$$

and its inverse $H^{-1}(z)$ are both analytic in $|z| \geq 1$. This permits an alternative expression for the power spectral density $\Phi_y(\omega)$ of $\{y_t\}$ in terms of this spectral factor $H(z)$ as

$$\Phi_y(\omega) = \sigma^2 |H(e^{j\omega})|^2.$$

Now, as mentioned in the introduction, it is often of interest to estimate this spectral density from observations of a realisation of $\{y_t\}$. Considering that the class of rational $|H(e^{j\omega})|^2$ are dense within the space of all continuous ones (with respect to the supremum norm) then a common strategy for estimating $\Phi_y(\omega)$ is to express (2) according to the so-called Auto-Regressive Moving Average (ARMA) model structure [13, 5, 21]

$$y_t = H(q,\theta)e_t = \frac{C(q,\theta)}{D(q,\theta)}e_t \tag{4}$$

where the numerator and denominator polynomials are of the form

$$D(q,\theta) \quad = \quad q^m + d_{m-1}q^{m-1} + \cdots + d_1 q + d_0, \tag{5}$$
$$C(q,\theta) \quad = \quad q^m + c_{m-1}q^{m-1} + \cdots + c_1 q + c_0 \tag{6}$$

and the parameter vector $\theta \in \mathbf{R}^n$ (with $n = 2m$) is defined as the vector of real valued co-efficients

$$\theta = [d_0, c_0, d_1, c_1, \cdots, d_{m-1}, c_{m-1}]^T.$$

There are two important sub-classes of this model structure; the Autoregressive (AR) and Moving Average (MA) cases which occur when (respectively) $C(q,\theta) = q^m$ and $D(q,\theta) = q^m$ are specified.

For all these AR, MA and ARMA cases, the mean-square optimal one-step ahead predictor $\widehat{y}_{t|t-1}(\theta)$ based on the model structure (4) is [13]

$$\widehat{y}_{t|t-1}(\theta) = \left[1 - H^{-1}(q,\theta)\right] y_t$$

with associated prediction error

$$\varepsilon_t(\theta) \triangleq y_t - \widehat{y}_{t|t-1}(\theta) = H^{-1}(q,\theta)y_t. \tag{7}$$

Therefore, if $\{e_t\}$ has a Gaussian distribution, then the Maximum–Likelihood estimates $\widehat{\theta}_N$ and $\widehat{\sigma}^2_N$ of θ and σ^2 are given as

$$\widehat{\theta}_N \triangleq \arg\min_{\theta\in\mathbf{R}} V_N(\theta), \qquad \widehat{\sigma}^2_N = \frac{1}{N}\sum_{t=1}^{N}\varepsilon_t^2(\widehat{\theta}_N) \tag{8}$$

where V_N is a quadratic estimation criterion defined as

$$V_N(\theta) = \frac{1}{2N}\sum_{t=1}^{N} \varepsilon_t^2(\theta). \tag{9}$$

This then leads to an estimate $H(z, \widehat{\theta}_N)$ of the spectral factor $H(z)$ in (3) and thereby also of the spectral density; viz.

$$\Phi_y(\omega, \widehat{\theta}_N) = \widehat{\sigma}_N^2 \left| H(e^{j\omega}, \widehat{\theta}_N) \right|^2. \tag{10}$$

It is known that this Maximum–Likelihood approach leads to an estimate variability $\mathrm{Var}\left\{\Phi_y(\omega, \widehat{\theta}_N)\right\}$ that asymptotically in the data length achieves the Cramér–Rao lower bound [13, 21, 5]. The focus of this paper is to provide an explicit formula for this bound, since it also quantifies the asymptotic variability of the spectral estimate (10) formed via (8) for cases in which $\{e_t\}$ is not Gaussian [16].

The importance of this evaluation of the CRLB was first recognised in [2] which established that for AR model structures

$$\lim_{m\to\infty} \lim_{N\to\infty} \frac{N}{m} \mathrm{Var}\left\{\Phi_y(\omega, \widehat{\theta}_N)\right\} = 2\Phi_y^2(\omega), \qquad \omega \neq 0, \pi \tag{11}$$

which suggests the approximate quantification

$$\mathrm{Var}\left\{\Phi_y(\omega, \widehat{\theta}_N)\right\} \approx \frac{2m}{N} \cdot \Phi_y^2(\omega), \qquad \omega \neq 0, \pi. \tag{12}$$

Later, the result (11) (and hence (12)) was shown to also be applicable in the case of ARMA modelling [15].

The motivation for allowing the model order to tend to infinity in (11) is to facilitate the derivation of a simple expression, such as the right hand side of (11). The clear drawback of this strategy is that it requires approximate convergence in (11) to have occurred in order for the ensuing quantification (12) to be accurate, and it is difficult to ensure that this convergence holds in practise.

The contribution of this paper is to present a new reproducing kernel based analysis method, which quantifies the CRLB on $\mathrm{Var}\left\{\Phi_y(\omega, \widehat{\theta}_N)\right\}$ in closed form and without requiring that $m \to \infty$, while also addressing all of the parametric AR, MA and ARMA modelling cases.

3 Main Results

Given the estimation scheme (8)–(10), then although it can be motivated via a Maximum–Likelihood perspective, under the much more general conditions on $\{e_t\}$ that it simply satisfy

$$\mathbf{E}\left\{e_t^2\right\} = \sigma^2 < \infty, \qquad \mathbf{E}\left\{|e_t|^{4+\epsilon}\right\} < \infty \tag{13}$$

for some $\epsilon > 0$, then as has been established in [14, 13] the estimate $\widehat{\theta}_N$ converges with increasing N and with probability one as

$$\lim_{N\to\infty} [\widehat{\theta}_N, \widehat{\sigma}_N^2] = [\theta_\circ, \sigma_\circ^2] \triangleq \underset{\theta\in\mathbf{R}, \sigma^2\in\mathbf{R}}{\arg\min} \lim_{N\to\infty} \mathbf{E}\left\{\frac{1}{2}\log\sigma^2 + \frac{1}{\sigma^2}V_N(\theta)\right\}. \tag{14}$$

As well, it also holds that if the underlying spectral factor $H(z)$ is, in fact, of rational form $C(z)/D(z)$ of order m, then as N increases the estimates $[\widehat{\theta}_N, \widehat{\sigma}^2_N]$ converge in law to a Normally distributed random variable with mean value $\theta_\circ$ such that [13, pg 258],[7] $|H(e^{j\omega}, \theta_\circ)| = |H(e^{j\omega})|$, and according to [16, 5, 13]

$$\sqrt{N}\begin{pmatrix} \widehat{\theta}_N - \theta_\circ \\ \widehat{\sigma}^2_N - \sigma^2 \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \begin{bmatrix} P & 0 \\ 0 & \mu \end{bmatrix}\right) \qquad \text{as } N \to \infty \tag{15}$$

where

$$\mu = \mathbf{E}\left\{\left(e_t^2 - \sigma^2\right)^2\right\} \tag{16}$$

and

$$P^{-1} \triangleq \lim_{N\to\infty} \frac{1}{\sigma^2}\mathbf{E}\left\{ \frac{\mathrm{d}^2}{\mathrm{d}\theta \mathrm{d}\theta^T} V_N(\theta)\Big|_{\theta=\theta_\circ}\right\} = \frac{1}{\sigma^2} \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \mathbf{E}\left\{\psi_t(\theta_\circ)\psi_t^T(\theta_\circ)\right\} \tag{17}$$

with $\psi_t(\theta)$ being the prediction error gradient given by

$$\psi_t(\theta_\circ) \triangleq \frac{\mathrm{d}\widehat{y}_{t|t-1}(\theta)}{\mathrm{d}\theta}\bigg|_{\theta=\theta_\circ} = H^{-1}(q, \theta_\circ) \cdot \frac{\mathrm{d}H(q, \theta_\circ)}{\mathrm{d}\theta}\varepsilon_t(\theta_\circ). \tag{18}$$

Furthermore, by Taylor expansion about $\theta = \theta_\circ$ and recognising that $H(z, \theta_\circ) = H(z)$

$$H(e^{j\omega}, \widehat{\theta}_N) - H(e^{j\omega}) = \left[\frac{\mathrm{d}H(e^{j\omega}, \theta)}{\mathrm{d}\theta}\bigg|_{\theta=\theta_\circ}\right]^T (\widehat{\theta}_N - \theta_\circ) + o(\|\widehat{\theta}_N - \theta_\circ\|^2). \tag{19}$$

Therefore, since $\|\widehat{\theta}_N - \theta_\circ\| \to 0$ as $N \to \infty$, a consequence of (15) is that

$$\sqrt{N}\left[\ H(e^{j\omega}, \widehat{\theta}_N) - H(e^{j\omega})\ \right] \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Delta(\omega)), \qquad \text{as } N \to \infty \tag{20}$$

where ($\cdot^\star$ denotes 'conjugate transpose')

$$\Delta(\omega) \triangleq \left[\frac{\mathrm{d}H(e^{j\omega}, \theta)}{\mathrm{d}\theta}\bigg|_{\theta=\theta_\circ}\right]^\star P \left[\frac{\mathrm{d}H(e^{j\omega}, \theta)}{\mathrm{d}\theta}\bigg|_{\theta=\theta_\circ}\right]. \tag{21}$$

In order to further analyse this expression for the purpose of realising a simple quantification of $\Delta(\omega)$, note that with the following definition of notation (q is the forward shift operator)

$$\Lambda_m(q) \triangleq [1, q, q^2, \cdots, q^{m-1}]^T, \tag{22}$$

and

$$Z(q, \theta) \triangleq \begin{bmatrix} -H(q, \theta) \\ 1 \end{bmatrix}, \tag{23}$$

then with $\otimes$ representing Kronecker tensor product [3], and I_ℓ being an $\ell \times \ell$ identity matrix

$$\frac{\mathrm{d}H(q, \theta)}{\mathrm{d}\theta} = [\Lambda_m(q) \otimes I_2]\frac{1}{D(q, \theta)} Z(q, \theta). \tag{24}$$

Therefore, by Parseval's Theorem, and noting that $\mathbf{E}\{\varepsilon_t^2(\theta_\circ)\} = \sigma^2$

$$
\begin{aligned}
P^{-1} &= \frac{1}{\sigma^2}\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\mathbf{E}\left\{\psi_t(\theta_\circ)\psi_t^T(\theta_\circ)\right\} \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}[\Lambda_m(e^{j\omega})\otimes I_2]\frac{Z(e^{j\omega},\theta_\circ)Z^\star(e^{j\omega},\theta_\circ)}{|C(e^{j\omega},\theta_\circ)|^2}[\Lambda_m^\star(e^{j\omega})\otimes I_2]\mathrm{d}\omega \qquad (25)\\
&= T_n\left(\frac{Z(e^{j\omega},\theta_\circ)Z(e^{j\omega},\theta_\circ)^\star}{|C(e^{j\omega},\theta_\circ)|^2}\right). \qquad (26)
\end{aligned}
$$

Here, the $T_n(F)$ notation denotes an $n\times n$ block Toeplitz matrix that is completely determined by the matrix valued function $F(\omega)$ as indicated by the passage from (25) to (26). Consequently, substituting (26) into (21) provides a frequency domain formulation for the variance $\Delta(\omega)$ as follows

$$
\Delta(\omega) \triangleq \frac{1}{|D(e^{j\omega},\theta_\circ)|^2}Z_\circ^\star(e^{j\omega})[\Lambda_m^\star(e^{j\omega})\otimes I_2]T_n^{-1}\left(\frac{Z_\circ Z_\circ^\star}{|C_\circ|^2}\right)[\Lambda_m(e^{j\omega})\otimes I_2]Z_\circ(e^{j\omega}). \qquad (27)
$$

Here, there has been some obvious compactification of notation involving subscripting with $\circ$. Now, as recently remarked in [24], it would initially appear that an expression such as this is prima facie '*hard to convert into a simple analytic form*'. It is this apparent difficulty that inspired the idea of allowing m to tend to infinity in previous works [2, 15, 25, 26] as a means for deriving simple expressions.

The first main contribution of this paper is to illustrate that, perhaps surprisingly, it is in fact relatively straightforward to provide an exact analytical expression for (27) which, as is obviously clear, depends only on the poles and zeros of $H(z)$. This is done in the following Theorem 3.1 for the full ARMA case, but in order to convey the essential ideas as clearly as possible, consider the simpler AR modelling situation addressed in [24] for which (27) reduces to

$$
\Delta_{\mathrm{AR}}(\omega) \triangleq |H(e^{j\omega})|^2\varphi_m(\omega,\omega) \qquad (28)
$$

where

$$
\varphi_m(\lambda,\omega) \triangleq \frac{1}{D(e^{-j\omega})D(e^{j\lambda})}\Lambda_m^\star(e^{j\omega})T_n^{-1}\left(\frac{1}{|D|^2}\right)\Lambda_m(e^{j\lambda}). \qquad (29)
$$

Now, by simple computation using the block Toeplitz matrix definition (26)

$$
\begin{aligned}
\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\Lambda_m(e^{j\lambda})}{D(e^{j\lambda})}\overline{\varphi_m(\lambda,\omega)}\,\mathrm{d}\lambda &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\Lambda_m(e^{j\lambda})\frac{1}{|D(e^{j\lambda})|^2}\Lambda_m^\star(e^{j\lambda})T_n^{-1}\left(\frac{1}{|D|^2}\right)\frac{\Lambda_m(e^{j\omega})}{D(e^{j\omega})}\,\mathrm{d}\lambda \\
&= \underbrace{\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\Lambda_m(e^{j\lambda})\Lambda_m^\star(e^{j\lambda})}{|D(e^{j\lambda})|^2}\,\mathrm{d}\lambda}_{T_n(1/|D|^2)}\cdot T_n^{-1}\left(\frac{1}{|D|^2}\right)\frac{\Lambda_m(e^{j\omega})}{D(e^{j\omega})} \\
&= \frac{\Lambda_m(e^{j\omega})}{D(e^{j\omega})} \qquad (30)
\end{aligned}
$$

and therefore, considering the definition of $\Lambda_m(z)$, and with the inner product $\langle\cdot,\cdot\rangle$ being the usual one for the space L_2 on the complex unit circle [11], equation (30) implies that

$$
f(e^{j\omega}) = \langle f(e^{j\lambda}),\varphi_m(\lambda,\omega)\rangle \qquad (31)
$$

for any $f\in X_m$ where

$$
X_m \triangleq \text{Span}\left\{\frac{1}{D(z)},\frac{z}{D(z)},\cdots,\frac{z^{m-1}}{D(z)}\right\}. \qquad (32)
$$

By virtue of the property (31), the function $\varphi_m(\lambda,\omega)$ is known as the 'Reproducing Kernel' for the space X_m [1].

Furthermore, a key point is that the reproducing kernel for a given space X_m is unique, and if $\{\mathcal{B}_0(z),\mathcal{B}_1(z),\cdots,\mathcal{B}_{m-1}(z)\}$ is an *orthonormal* basis for X_m, then a trivial construction for this kernel $\varphi_m(\lambda,\omega)$ is given as [6]

$$\varphi_m(\lambda,\omega) = \sum_{k=0}^{m-1} \overline{\mathcal{B}_k(e^{j\omega})}\mathcal{B}_k(e^{j\lambda}). \tag{33}$$

Since, as shown in [17], the formulation

$$\mathcal{B}_k(z) \triangleq \frac{\sqrt{1-|\xi_k|^2}}{z-\xi_k}\cdot\Pi_k(z), \qquad \Pi_k(z) \triangleq \prod_{\ell=0}^{k-1}\left(\frac{1-\overline{\xi_\ell}z}{z-\xi_\ell}\right), \qquad \Pi_{-1}(z)\triangleq 1 \tag{34}$$

where the $\{\xi_k\}$ are the zeros of $D(z)$ is such an orthonormal basis for X_m, then substituting this explicit construction into (33), and thence into (28) leads to a simple analytic formulation (note that $|\Pi_k(e^{j\omega})| = 1$)

$$\Delta_{\mathrm{AR}}(\omega) \triangleq |H(e^{j\omega})|^2 \sum_{k=1}^{m-1} \frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}. \tag{35}$$

While this quantification of spectral factor variability is identical to that obtained via Theorem 2.1 of [24], which then further leads to (48), it is achieved by a quite different means than employed in [24], and which has the advantage of being easily extended to cover the more general ARMA modelling situation. This is considered in the following main result, which subsumes the preceding AR-based illustrational discussion in a more formal manner.

Theorem 3.1. *Suppose that $H(z) = C(z)/D(z)$ is minimal, and define the functions*

$$\varphi_m(\lambda,\omega) \triangleq \frac{Z_\circ^\star(e^{j\omega})}{C(e^{-j\omega})}[\Lambda_m^\star(e^{j\omega})\otimes I_2]T_n^{-1}\left(\frac{Z_\circ Z_\circ^\star}{|C|^2}\right)[\Lambda_m(e^{j\lambda})\otimes I_2]\frac{Z_\circ(e^{j\lambda})}{C(e^{j\lambda})}, \tag{36}$$

$$\overline{\psi_m(\lambda,\omega)} \triangleq \frac{Z_\circ^T(e^{j\omega})}{C(e^{j\omega})}[\Lambda_m^T(e^{j\omega})\otimes I_2]T_n^{-1}\left(\frac{Z_\circ Z_\circ^\star}{|C|^2}\right)[\Lambda_m(e^{j\lambda})\otimes I_2]\frac{Z_\circ(e^{j\lambda})}{C(e^{j\lambda})}. \tag{37}$$

Furthermore, define the zeros $\{\xi_k\}$ via

$$C(z)D(z) = (z-\xi_0)(z-\xi_1)\cdots(z-\xi_{2m-1}). \tag{38}$$

Then the following expressions hold

$$\varphi_m(\lambda,\omega) = \sum_{k=0}^{2m-1} \frac{1-|\xi_k|^2}{(e^{j\lambda}-\xi_k)(e^{-j\omega}-\overline{\xi_k})}\prod_{\ell=0}^{k-1}\left(\frac{1-e^{j\lambda}\overline{\xi_\ell}}{e^{j\lambda}-\xi_\ell}\right)\left(\frac{1-e^{-j\omega}\xi_\ell}{e^{-j\omega}-\overline{\xi_\ell}}\right) \tag{39}$$

$$\psi_m(\lambda,\omega) = \varphi_{m-\rho}(\lambda,-\omega) + \sum_{\tau=0}^{\rho-1}\zeta_\tau(e^{j\lambda})\zeta_\tau(e^{j\omega}) \tag{40}$$

where

$$\zeta_\tau(z) \triangleq \frac{\sqrt{1-|\xi_{r(\tau)}|^4}(z-\alpha)}{(z-\xi_{r(\tau)})(z-\overline{\xi_{r(\tau)}})}\prod_{\ell=0}^{r(\tau)-1}\left(\frac{1-\overline{\xi_\ell}z}{z-\xi_\ell}\right) \tag{41}$$

$$r(\tau) \triangleq 2(m-\rho+\tau), \qquad \alpha \triangleq \frac{\xi_r(\tau) + \overline{\xi}_r(\tau) - \sqrt{(1-\xi^2_{r(\tau)})(1-\overline{\xi}^2_{r(\tau)})}}{1+|\xi_{r(\tau)}|^2}. \tag{42}$$

In (40)– (42) it has been assumed (without loss of generality) that the zeros defined by (38) are arranged so that the first $2(m-\rho)$ of them are purely real valued, and the remaining 2ρ then occur in complex conjugate pairs.

Proof. See Appendix A. □

The second of these expressions (40) involving $\psi_m(\lambda,\omega)$ will have application later in quantifying the variability of spectral density estimate $\Phi_y(\omega,\widehat{\theta}_N)$, while the first involving $\varphi_m(\lambda,\omega)$ has immediate application now in quantifying the asymptotic Cramér–Rao lower bound for the estimate of the spectral factor $H(z)$.

Corollary 3.1. *Suppose that $\widehat{\theta}_N$ is calculated via (4)– (9) using the m'th order ARMA model structure (4), and that the data $\{y_t\}$ has true underlying spectral factor of $H(z) = C(z)/D(z)$ of minimal order equal to m. Suppose further that the zeros $\{\xi_k\}$ defined by*

$$C(z)D(z) = (z-\xi_0)(z-\xi_1)\cdots(z-\xi_{2m-1}) \tag{43}$$

are all strictly within the open unit disk $\mathbf{D}$, and that $\{e_t\}$ satisfies the conditions (13). Then

$$\sqrt{N}\begin{bmatrix} H(e^{j\omega},\widehat{\theta}_N) - H(e^{j\omega}) \\ H(e^{j\lambda},\widehat{\theta}_N) - H(e^{j\lambda}) \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma(\omega,\lambda))$$

as $N\to\infty$ where, for $\omega\neq\lambda$

$$\Sigma(\omega,\lambda) = \begin{bmatrix} H(e^{j\omega}) & 0 \\ 0 & H(e^{j\lambda}) \end{bmatrix} \begin{bmatrix} \overline{\varphi_m(\omega,\omega)} & \varphi_m(\lambda,\omega) \\ \overline{\varphi_m(\lambda,\omega)} & \varphi_m(\lambda,\lambda) \end{bmatrix} \begin{bmatrix} \overline{H(e^{j\omega})} & 0 \\ 0 & \overline{H(e^{j\lambda})} \end{bmatrix} \tag{44}$$

with $\varphi_m(\lambda,\omega)$ given by (39). The result also holds for the AR and MA cases with the following modifications:

1. *The model order m can be greater than an underlying true one ℓ;*
2. *The substitutions $C(z)=1$ and $D(z)=1$ in (43) are made for the AR or MA cases (respectively);*
3. *The zeros $\{\xi_{\ell+1},\cdots,\xi_m\}$ in (43) are set to zero.*

Proof. This follows directly from the argument used leading to equation (27) combined with the application of Theorem 3.1. The AR speciality has already been discussed between equations (28)–(35), but in that development, and contrary to the ARMA case, the associated block Toeplitz matrix is invertible for model orders m greater than the underlying true one ℓ, and hence the constraint $m=\ell$ can be discarded. However, the unique asymptotic value $\theta_\circ$ is given as

$$\theta_\circ = \arg\min_{\theta\in\mathbf{R}n} \lim_{N\to\infty} \mathbf{E}\{V_N(\theta)\} = \arg\min_{\theta\in\mathbf{R}n} \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|\frac{H(e^{j\omega})}{H(e^{j\omega},\theta)}\right|^2 \mathrm{d}\omega \tag{45}$$

which [13, pg 258],[7] implies that $|H(e^{j\omega})| = |H(e^{j\omega},\theta_\circ)|$, and hence that any poles of $H(z,\theta_\circ)$ (zeros of $D(z,\theta_\circ)$) in excess of the ℓ underlying true ones in this AR case must be located at the origin. The MA case follows by an identical argument. □

This result is an extension over previous work such as [2, 15, 25, 26] in that, not only does it consider a wider class of model structures, it also quantifies the finite model order covariance between estimates at different frequencies. Again, this is made directly possible by virtue of the reproducing kernel approach taken here.

In turn, this result may then be applied to quantify the asymptotic variability for the parametric spectral density estimate (10) as follows.

Corollary 3.2. *Suppose that $\widehat{\theta}_N$ is calculated via (4)–(9) using the m'th order ARMA model structure (4), and that the data $\{y_t\}$ has true underlying spectral factor of $H(z) = C(z)/D(z)$ of minimal order equal to m. Suppose further that the zeros $\{\xi_k\}$ defined by*

$$C(z)D(z) = (z - \xi_0)(z - \xi_1) \cdots (z - \xi_{2m-1}) \tag{46}$$

are all strictly within the open unit disk $\mathbf{D}$, and that $\{e_t\}$ satisfies the conditions (13). Then

$$\sqrt{N} \begin{bmatrix} \Phi_y(\omega, \widehat{\theta}_N) - \Phi_y(\omega) \\ \Phi_y(\lambda, \widehat{\theta}_N) - \Phi_y(\lambda) \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma(\omega, \lambda))$$

as $N \to \infty$ where

$$\Sigma(\omega, \lambda) = \begin{bmatrix} |H(e^{j\omega})|^2 & 0 \\ 0 & |H(e^{j\lambda})|^2 \end{bmatrix} \left[\mu I_2 + 2\sigma^4 \Gamma(\omega, \lambda)\right] \begin{bmatrix} |H(e^{j\omega})|^2 & 0 \\ 0 & |H(e^{j\lambda})|^2 \end{bmatrix}$$

$$\Gamma(\omega, \lambda) \triangleq \mathit{Re} \left\{ \begin{bmatrix} \varphi_m(\omega, \omega) + \psi_m(\omega, \omega) & \varphi_m(\lambda, \omega) + \psi_m(\lambda, \omega) \\ \varphi_m(\lambda, \omega) + \psi_m(\lambda, \omega) & \varphi_m(\lambda, \lambda) + \psi_m(\lambda, \lambda) \end{bmatrix} \right\}.$$

The functions φ_m and ψ_m are defined in (39), (40). The result also holds for the AR and MA cases with the following modifications:

1. *The model order m can be greater than an underlying true one ℓ;*
2. *The substitutions $C(z) = 1$ and $D(z) = 1$ in (43) are made for the AR or MA cases (respectively);*
3. *The zeros $\{\xi_{\ell+1}, \cdots, \xi_m\}$ in (43) are set to zero.*

Proof. See Appendix B. □

The most important consequence of this corollary is that it establishes the result

$$\lim_{N\to\infty} N \mathrm{Var} \left\{ \frac{\Phi_y(\omega, \widehat{\theta}_N)}{\Phi_y(\omega)} \right\} = \frac{\mu}{\sigma^4} + 2\mathrm{Re}\left\{\varphi_m(\omega, \omega) + \psi_m(\omega, \omega)\right\}. \tag{47}$$

The first key point about (47) is that, via the formulae (39)-(42) established in Theorem 3.1, then (47) is a *closed form* expression for the asymptotic in N variability for all of the cases of AR, MA and ARMA parametric spectral estimates $\Phi_y(\omega, \widehat{\theta}_N)$. This is in contrast to previous work such as [20],[12],[8] which established our departure point (27) as a formulation of the spectral estimate variability, but did not provide a closed form expression for it such as (47).

The second key point about this closed form expression (51) is that, again in contrast previous work such as [2, 13], it is *not* derived via an argument that is asymptotic in the model order m.

Therefore, the approximation this paper now proposes for the practical case of finite data length N *and* finite model order settings of

$$\mathrm{Var}\left\{\frac{\Phi_y(\omega,\widehat{\theta}_N)}{\Phi_y(\omega)}\right\} \approx \frac{1}{N}\left[\frac{\mu}{\sigma^4} + 2\mathrm{Re}\left\{\varphi_m(\omega,\omega)+\psi_m(\omega,\omega)\right\}\right] \tag{48}$$

is likely to be far more accurate that ones such as (12) derived from previous results such as (11) which require $m \to \infty$. This is illustrated via simulation example in the following section. Balancing this enhanced accuracy arising from an analysis that applies for finite m is the disadvantage that, in the ARMA case, the model order chosen must equal a true underlying one. While it can be argued that this rarely occurs in practice, it is equally true that an approximately correct model order is usually chosen such that a residual whiteness is test is passed. Furthermore, the case in which the model order is greater than an underlying true one is considered in the following Theorem 3.2.

The third key point is that since via (39), (40)

$$\varphi_m(\omega,\omega) = \sum_{k=0}^{2m-1} \frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2} \tag{49}$$

and

$$\psi_m(\omega,\omega) = \sum_{k=0}^{2(m-\rho)-1} \frac{1-\xi_k^2}{(e^{j\omega}-\xi_k)^2} \prod_{\ell=0}^{k-1}\left(\frac{1-\xi_\ell e^{j\omega}}{e^{j\omega}-\xi_\ell}\right)^2 + \sum_{\tau=0}^{\rho-1} \frac{(1-|\xi_{r(\tau)}|^4)(e^{j\omega}-\alpha)^2}{(e^{j\omega}-\xi_{r(\tau)})^2(e^{j\omega}-\overline{\xi_{r(\tau)}})^2} \prod_{\ell=0}^{r(\tau)-1}\left(\frac{1-\overline{\xi_\ell}e^{j\omega}}{e^{j\omega}-\xi_\ell}\right)^2 \tag{50}$$

where $r(\tau)$ and α are defined in (42), then the closed form expressions (47), (48) highlight that because all the denominators in (49), (50) are small when $e^{j\omega}$ is close to any of the $\{\xi_k\}$, then the relative estimation error is likely to be larger at those frequencies near both the poles *and* zeros of the underlying spectral factor $H(z)$. Furthermore, this relative estimation error is likely to be larger when those poles or zeros are very close to the unit circle, then when they are not.

To explore this even more closely, consider for the moment the simplest case of all the poles and zeros $\{\xi_k\}$ being real valued, in which case use of (49), (50) permits (47) to be evaluated as

$$\lim_{N\to\infty} N\mathrm{Var}\left\{\frac{\Phi_y(\omega,\widehat{\theta}_N)}{\Phi_y(\omega)}\right\} = \frac{\mu}{\sigma^4} + 4\sum_{k=0}^{2m-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}\cos^2\left(\phi_k(\omega) + \sum_{\ell=0}^{k-1}[2\phi_\ell(\omega)-\omega]\right) \tag{51}$$

where the notation $\phi_k(\omega) \triangleq \measuredangle(e^{j\omega}-\xi_k)$ has been used. Simple geometry then indicates that poles and zeros near the origin lead to small and smooth variations in the subtended angle $\phi_k(\omega)$, and hence via (51) smooth variations in spectral estimate variability with changing ω, and vice-versa for poles and zeros near the boundary of the unit circle.

Finally, returning to (47) which applies for any real or complex value of $\{\xi_k\}$, it can be used to establish that the average relative estimation error over all frequencies is given as

$$\lim_{N\to\infty}\frac{N}{2\pi}\int_{-\pi}^{\pi}\mathrm{Var}\left\{\frac{\Phi_y(\omega,\widehat{\theta}_N)}{\Phi_y(\omega)}\right\}\,\mathrm{d}\omega = \frac{\mu}{\sigma^4} + 4m. \tag{52}$$

This follows by a simple contour integral argument in which, via the orthonormality of the $\{\mathcal{B}_k(e^{j\omega})\}$ functions, the first frequency dependent term $\varphi_m(\omega,\omega)$ on the right hand side of (47) integrates to $2m$

for any choice of $\{\xi_k\}$, while the second term $\psi_m(\omega,\omega)$ integrates to zero since it is analytic within the unit circle; see (50).

This illustrates a "waterbed effect" in that, although as just discussed, the expression (47) indicates increased relative error near poles and zeros, with increased effect according to distance from the unit circle, these effects must be balanced by a commensurate *decrease* in relative error at other frequencies, since the average (over frequency) relative error depends only on the model order.

As far as the author is aware, these qualitative and quantitative conclusions which relate to fundamental limitations of spectral estimate accuracy are new, and illustrate the practical utility arising from the availability of closed form expressions for the CRLB on spectral estimates.

As a final extension to the preceding results, we now establish that the caveats imposed in Corollaries 3.1 and 3.2 of the true model order m being equal to an underlying true one ℓ in the ARMA modelling case may be dropped to allow for the case of $m > \ell$. However, a procedural difficulty in this case of over-modelling is that the value of $\theta_\circ$ defined by (14) is not unique. Instead, with probability one, $\widehat{\theta}_N$ converges to a set Θ as follows

$$\lim_{N\to\infty} \widehat{\theta}_N \in \Theta \triangleq \left\{ \theta : \lim_{N\to\infty} \mathbf{E}\left\{V_N(\theta)\right\} \le \lim_{N\to\infty} \mathbf{E}\left\{V_N(\beta)\right\} \ \forall\beta \right\}. \tag{53}$$

To circumvent this, consider the so-called *regularised* refinement of (9) which, for some regularising parameter $\delta > 0$, is defined as

$$V_N(\theta) = \frac{1}{2N}\sum_{t=1}^{N} \varepsilon_t^2(\theta) + \frac{\delta}{2}\|\theta - \theta_\circ\|^2. \tag{54}$$

Here, the norm $\|\cdot\|$ is the Euclidean one, and $\theta_\circ$ is fixed as that defined in (14). With respect to this cost the following asymptotic variance result applies for ARMA modelling with model order m possibly greater than an underlying true one ℓ.

Theorem 3.2. *Suppose that $\widehat{\theta}_N$ is calculated (4)– (8) and the regularised criterion (54) using the m'th order ARMA model structure (4), and that the data $\{y_t\}$ has true underlying spectral factor of $H(z) = C(z)/D(z)$ of minimal order equal to $\ell \le m$. Let this order be used to define polynomials $C_\ell(z)$, $D_\ell(z)$ and $T(z)$ by requiring that $C_\ell(z)/D_\ell(z) = C(z,\theta_\circ)/D(z,\theta_\circ)$ is a minimal representation of $H(z,\theta_\circ) = H(z)$ and that $T(z)$ contains the common zeros in $C(z,\theta_\circ)/D(z,\theta_\circ)$ as follows:*

$$\frac{C(z,\theta_\circ)}{D(z,\theta_\circ)} = \frac{C_\ell(z)}{D_\ell(z)} \cdot \frac{T(z)}{T(z)}. \tag{55}$$

Furthermore, define the zeros $\{\xi_0,\cdots,\xi_{m+\ell-1}\}$ via the factorisation

$$C_\ell(z)D_\ell(z)T(z) = (z-\xi_0)(z-\xi_1)\cdots(z-\xi_{m+\ell-1}). \tag{56}$$

Then all the ARMA related results of Corollary 3.1 and Corollary 3.2 apply with the substitution of the above zeros into the formulation of $\varphi_m(\lambda,\omega)$ and $\psi_m(\lambda,\omega)$ and in the limit as $\delta \to 0$ in the regularised criterion (54).

Proof. See Appendix C. $\square$

Note that although this theorem relies on the regularised criterion (54) in order to specify a point $\theta_\circ$, and not a set Θ, that $\widehat{\theta}_N$ is convergent to, it asserts quantifications that are valid only in the limit as the regularising parameter δ tends to zero. Certainly, in the limit as $N \to \infty$, the criterion (54) has a unique minimum at $\theta_\circ$ for an arbitrarily small δ. Therefore, this strategy is simply a technical artifice to allow the specification of a asymptotic value for $\widehat{\theta}_N$, about which distributional results may be derived. The same technique has been employed in other works, such as [15].

An important feature of this result is that it allows an investigation of the rapprochement of the work here with afore-mentioned pre-existing results that apply asymptotically as the model order $m \to \infty$. In particular, according to [15, 13],

$$\lim_{m\to\infty}\lim_{N\to\infty}\frac{N}{m}\mathrm{Var}\{H(e^{j\omega},\widehat{\theta}_N)\} = |H(e^{j\omega})|^2. \tag{57}$$

On the other hand, if the regularised criterion (54) is used with $\widehat{\theta}_\circ$ chosen so that any pole-zero cancellations in $H(z,\theta_\circ)$ occur at the origin, then for model orders m greater than or equal to an underlying true model order ℓ, Theorem 3.2 asserts that

$$\lim_{N\to\infty}\frac{N}{m}\mathrm{Var}\{H(e^{j\omega},\widehat{\theta}_N)\} = |H(e^{j\omega})|^2\left[\left(\frac{1}{m}\sum_{k=0}^{2\ell-1}\frac{1-|\xi_k|^2}{|e^{j\omega}-\xi_k|^2}\right)+1-\frac{\ell}{m}\right]. \tag{58}$$

The term within the square brackets is simply $\varphi_m(\omega,\omega)/m$, and since ℓ is fixed, then in the limit as $m \to \infty$, it will tend to one. In this case, (58) provides the same limiting result as $m \to \infty$ as (57), However, with different regularisation choices such that pole zero cancellations in $H(z,\theta_\circ)$ are not at the origin, this agreement is lost, as will be demonstrated in the following section.

4 Simulation Examples

In order to provide concrete illustration of the results presented here, consider the case of a true ARMA system with spectral factor

$$H(z) = \frac{z^3 - 1.9235z^2 + 1.5910z - 0.5203}{z^3 - 1.9464z^2 + 1.5155z - 0.5368} \tag{59}$$

and suppose that the innovations driving this are Gaussian distributed with variance $\sigma^2 = 1$. Then according to (47), the variability of a Maximum–Likelihood estimate of the associated spectral density $\Phi_y(\omega)$ should be quantifiable via the CRLB for this estimation problem according to (48). This can be compared with the previous asymptotic results [2, 15, 25, 26], which are asymptotic in both data length N *and* model order m according to (11), and which have led to the pre-existing approximation

$$\mathrm{Var}\left\{\Phi_y(\omega,\widehat{\theta}_N)\right\} \approx \frac{1}{N}\Phi_y^2(\omega)\left[\frac{\mu}{\sigma^4}+2m\right]. \tag{60}$$

Note that the first term within accounts for the possibility of estimating the value of σ^2, while previous work leading to (11) assumed this was known - see Appendix B for complete detail on this point.

With this in mind, the new expression (48), whose accuracy does not depend on m is essentially different from (60) according to the φ_m and ψ_m terms, which are determined by the zeros of $C(z)$ and $D(z)$ in $H(z) = C(z)/D(z)$, which in this case are given as

$$\{\xi_k\} = \left\{0.7165, 0.9429, 0.852e^{\pm j0.784}, 0.7545e^{\pm j0.8433}\right\}. \tag{61}$$

The utility of the ensuing new quantification (48) is illustrated in figure 1, where it is profiled as a dashed line together with the ‘true’ variability $\mathrm{Var}\left\{\Phi_y(\omega,\widehat{\theta}_N)\right\}$ which is estimated in a Monte–Carlo

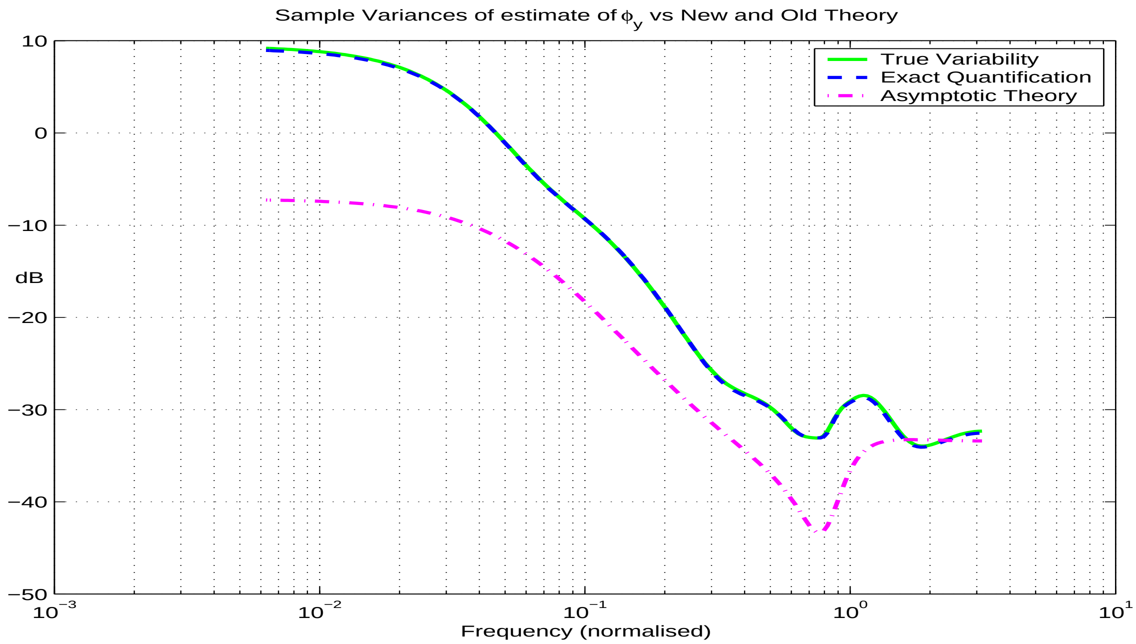


Figure 1: *Variability of* $\Phi_y(e^{j\omega}, \widehat{\theta}_N)$*. The solid line is the true variability, as estimated via averaging over Monte–Carlo trial, the dashed line exactly matching it is the new quantification (48) of this paper, while the dash-dot line is the pre-existing quantification (60) which depends on an asymptotic in model order argument.*

fashion by computing the sample variance over 1000 simulation experiments, each of which involves $N = 10000$ data points. Clearly, the agreement is excellent, and certainly superior to the pre-existing quantification (60), which is shown as the dash-dot line.

Furthermore, in order to illustrate the application of Theorem 3.2 with the attendant conclusions about the influence of regularisation point on asymptotic variability, consider the case of a true underlying first order spectral factor of

$$H(z) = \frac{z}{z - 0.5} \tag{62}$$

which clearly implies a true model order of $\ell = 1$. Suppose that an $m = 3$'rd order model is then fitted via the use of the regularised criterion (54) with $\delta = 0.01$ to a realisation of $N = 10000$ data points generated according to the spectral factor (62) combined with white innovations variance of $\sigma^2 = 1$,

In the case of the regularisation point $\theta_\circ$ representing $H(z, \theta_\circ) = H(z)$ equal to (62), and also involving pole-zero cancellations at $z = \{0, 0\}$ then the results of Theorem 3.2 together with (estimated) true variability computed by averaging across 1000 data realisations is shown in the left plot of figure 2. As expected the agreement is essentially exact, and again superior to the pre-existing quantification (60), which is shown as the dash-dot line.

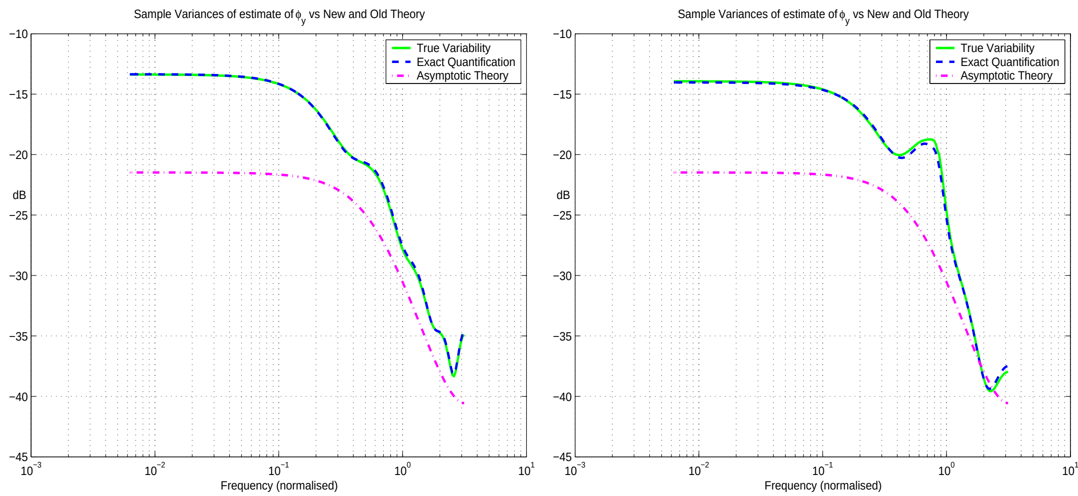
On the other hand, if the regularisation point $\theta_\circ$ is altered to imply pole zero cancellations at $z = \{0.85e^{\pm j\pi/4}\}$, but with $H(z, \theta_\circ)$ still equal to (62), then the variability results are as illustrated in the right plot of figure 2.

Clearly they are affected by the choice of regularisation point, and it would appear that the existence of this phenomenon has not been previously recognised in the literature.

5 Conclusion

The contribution of this paper has been to quantify the asymptotic in data-length variability of ARMA based spectral estimates and their AR and MA specialisations, which also quantify the CRLB for these estimation problems in the case of Gaussian innovations.

The key feature discriminating the work here from previous ones has been the invention of new analysis techniques based on reproducing kernel theory that has the dividend of providing expressions that are valid for finite model orders. There are many applications for these sorts of results, and we refer the reader to previous works such as [2, 15, 26, 10, 22, 23, 18] for more discussion on this point. However, a particular feature of the results here applying for finite model order is that they allow new phenomena to be discovered, such as the dependence of spectral estimate variability on regularisation point, the variance increasing effect of spectral factor poles and zeros near the unit circle, and ‘conservation of uncertainty’ results via an afore-mentioned ‘waterbed effect’ on the relative spectral estimation error.



(a) Regularisation about $\xi = 0$ (b) Regularisation about $\xi = 0.85^{\pm j\pi/4}$

Figure 2: *Variability of $\Phi_y(e^{j\omega}, \widehat{\theta}_N)$ for the case of over-modelling $m = 3 > \ell = 1$ and regularised estimate found via the criterion (54). This illustrates the effect of regularisation point $\theta_\circ$ on variability. The line legend is as per previous figure, and the left plot involves regularisation about a point implying pole-zero cancellations at the origin, while the right plot pertains to pole-zero cancellations at* $0.85e^{\pm j\pi/4}$.

A Proof of Theorem 3.1

Proof. First, via the assumption of $C(z)$ and $D(z)$ containing no common zeros, then by Lemma D.1 the associated block Toeplitz matrices in (36) and (37) are in fact invertible, and hence $\varphi_m(\omega,\lambda)$ and $\psi_m(\omega,\lambda)$ are well defined. Next, note that by the definition (26)

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}[\Lambda_m(e^{j\lambda})\otimes I_2]\frac{Z_\circ(e^{j\lambda})Z_\circ^\star(e^{j\lambda})}{|C(e^{j\lambda})|^2}[\Lambda_m^\star(e^{j\lambda})\otimes I_2]\times$$

$$T_n^{-1}\left(\frac{Z_\circ Z_\circ^\star}{|C|^2}\right)[\Lambda_m(e^{j\omega})\otimes I_2]\frac{Z_\circ(e^{j\omega})}{C(e^{j\omega})}\,\mathrm{d}\lambda = [\Lambda_m(e^{j\omega})\otimes I_2]\frac{Z_\circ(e^{j\omega})}{C(e^{j\omega})} \tag{A.63}$$

Furthermore,

$$[\Lambda_m(z)\otimes I_2]\,\frac{Z(z,\theta_\circ)}{C(z)} = \left[-\frac{1}{D(z)},\frac{1}{C(z)},\cdots-\frac{z^{m-1}}{D(z)},\frac{z^{m-1}}{C(z)}\right]^T. \tag{A.64}$$

Therefore, with the definition of the space

$$X_m \triangleq \left\{f : f = \sum_{k=0}^{m-1}\left(\alpha_k\frac{z^k}{C(z)} + \beta_k\frac{z^k}{D(z)}\right) \quad ;\alpha_k,\beta_k\in\mathbf{C}\right\}$$

the relationship (A.63) indicates that $\varphi(\lambda,\omega)$ satisfies

$$f(\omega) = \langle f(\lambda), \varphi(\lambda,\omega)\rangle \tag{A.65}$$

for any $f \in X_m$ and with respect to the usual inner product on the Hilbert space $L_2(\mathbf{T})$. Hence, $\varphi(\lambda,\omega)$ is the reproducing kernel for the space X_m. However, X_m can also be defined as

$$X_m \triangleq \left\{ f : f = \sum_{k=0}^{2m-1} \gamma_k \mathcal{B}_k(z) \quad ; \gamma_k \in \mathbf{C} \right\}$$

where the functions $\{\mathcal{B}_k(z)$ are defined in (34) and are orthonormal with respect to the above mentioned inner product and hence provide an alternative formulation for the reproducing kernel as

$$\varphi(\lambda,\omega) = \sum_{k=0}^{2m-1} \overline{\mathcal{B}_k(e^{j\omega})}\mathcal{B}_k(e^{j\lambda}). \tag{A.66}$$

However, any reproducing kernel $\varphi_m(\lambda,\omega)$ is 'Hermitian Symmetric' in that, since for any fixed ω the kernel $\varphi_m(\lambda,\omega) \in X_m$, then

$$\varphi_m(\lambda,\omega) = \langle \varphi_m(\zeta,\omega), \varphi_m(\zeta,\lambda)\rangle = \overline{\langle \varphi_m(\zeta,\lambda), \varphi_m(\zeta,\omega)\rangle} = \overline{\varphi_m(\omega,\lambda)}.$$

and therefore, $\varphi_m(\lambda,\omega)$ is the *unique* element in X_m that has the property (A.65), since if another function $H_m(\lambda,\omega)$ also satisfied (A.65), then it would hold that

$$H_m(\lambda,\omega) = \overline{H_m(\omega,\lambda)} = \overline{\langle H_m(\zeta,\lambda), \varphi_m(\zeta,\omega)\rangle} = \langle \varphi_m(\zeta,\omega), H_m(\zeta,\lambda)\rangle = \varphi_m(\lambda,\omega).$$

Therefore, (A.66) must be an alternative formulation for (36). Similarly, with the definition of the space $\widetilde{X}_m$ as

$$\widetilde{X}_m \triangleq \left\{ f : f = \sum_{k=0}^{m-1} \left(\alpha_k \frac{z^k}{C(z)} + \beta_k \frac{z^k}{D(z)} \right) \quad ; \alpha_k, \beta_k \in \mathbf{R} \right\}$$

then $f(\omega) = \langle \psi_m(\lambda,\omega), f(\lambda)\rangle$ for any $f \in \widetilde{X}_m$. Note that because $\psi_m(\lambda,\omega)$ appears as the first element in the above inner product, it is only capable of reproducing functions f with corresponding real valued impulse response, as embodied in the definition of $\widetilde{X}_m$. Finally, with the assumptions made in the theorem on the ordering of the zeros $\{\xi_k\}$, the space $\widetilde{X}_m$ can also be written as

$$\widetilde{X}_m = \left\{ f : f = \sum_{k=0}^{2(m-\rho)-1} \gamma_k \, \mathcal{B}_k(z) + \sum_{k=2(m-\rho)}^{2m-1} \gamma_k \, \mathcal{B}'_k(z) \quad ; \gamma_k \in \mathbf{R} \right\}.$$

where the particular formulation for the $\mathcal{B}'_k(z)$ to ensure real valued impulse response, while still being orthonormal, was presented in [17]. Therefore, via an identical argument as employed above, an alternative formulation for this 'conjugate' reproducing kernel is

$$\psi_m(\lambda,\omega) = \sum_{k=0}^{2(m-\rho)-1} \mathcal{B}_k(e^{j\lambda})\mathcal{B}_k(e^{j\omega}) + \sum_{k=2(m-\rho)}^{2m-1} \mathcal{B}'_k(e^{j\lambda})\mathcal{B}'_k(e^{j\omega}).$$

Expanding and then simplifying this expression and (A.66) according to the definitions given in [17] then completes the proof. $\square$

B Proof of Corollary 3.2

Proof. With the following definitions

$$\widehat{H}_r(\omega) \triangleq \text{Re}\left\{H(e^{j\omega},\widehat{\theta}_N)\right\}, \qquad \widehat{H}_i(\omega) \triangleq \text{Im}\{H(e^{j\omega},\widehat{\theta}_N)\} \tag{B.67}$$

then

$$\Phi_y(\omega,\widehat{\theta}_N) = \widehat{\sigma}_N^2\left[\widehat{H}_r^2(\omega)+\widehat{H}_i^2(\omega)\right]$$

and hence, via Taylor series expansion about $\theta_\circ$, σ^2

$$\begin{aligned}\Phi_y(\omega,\widehat{\theta}_N)-\Phi_y(\omega) &= |H(e^{j\omega})|^2(\widehat{\sigma}_N^2-\sigma^2)+2\sigma^2\left[H_r(\omega),H_i(\omega)\right]\begin{bmatrix}\widehat{H}_r(\omega)-H_r(\omega)\\ \widehat{H}_i(\omega)-H_i(\omega)\end{bmatrix}+\\ &\qquad o\left(|H(e^{j\omega},\widehat{\theta}_N)-H(e^{j\omega})|^2\right)+o(|\widehat{\sigma}_N^2-\sigma^2|).\end{aligned}$$

Furthermore, also by Taylor expansion as in (48)

$$\widehat{H}_r(\omega)-H_r(\omega) = \left[\left.\frac{\mathrm{d}H_r(e^{j\omega},\theta)}{\mathrm{d}\theta}\right|_{\theta=\theta_\circ}\right]^T(\widehat{\theta}_N-\theta_\circ)+o(\|\widehat{\theta}_N-\theta_\circ\|). \tag{B.68}$$

Therefore, by application of (24), (26) and Theorem 3.1

$$\begin{aligned}&\lim_{N\to\infty} N\mathbf{E}\left\{\left|\widehat{H}_r(\omega)-H_r(\omega)\right|^2\right\} = \\ &\frac{1}{4}\left[\frac{1}{D(e^{j\omega})}Z_\circ^T(e^{j\omega})[\Lambda_m^T(e^{j\omega})\otimes I_2]+\frac{1}{D(e^{-j\omega})}Z_\circ^\star(e^{j\omega})[\Lambda_m^\star(e^{j\omega})\otimes I_2]\right]T_n^{-1}\left(\frac{Z_\circ Z_\circ^\star}{|C|^2}\right)\times\\ &\left[\frac{1}{D(e^{j\omega})}[\Lambda_m(e^{j\omega})\otimes I_2]Z_\circ(e^{j\omega})+\frac{1}{D(e^{-j\omega})}[\Lambda_m(e^{-j\omega})\otimes I_2]Z_\circ(e^{-j\omega})\right]\\ =\ &\frac{1}{2}\left[|H(e^{j\omega})|^2\varphi(\omega,\omega)+\text{Re}\left\{H^2(e^{j\omega})\overline{\psi_m(\omega,\omega)}\right\}\right].\end{aligned}$$

Similarly

$$\lim_{N\to\infty} N\mathbf{E}\left\{\left|\widehat{H}_i(\omega)-H_i(\omega)\right|^2\right\} = \frac{1}{2}\left[|H(e^{j\omega})|^2\varphi(\omega,\omega)-\text{Re}\left\{H^2(e^{j\omega})\overline{\psi_m(\omega,\omega)}\right\}\right]$$

and

$$\lim_{N\to\infty} N\mathbf{E}\left\{\left(\widehat{H}_r(\omega)-H_r(\omega)\right)\left(\widehat{H}_i(\omega)-H_i(\omega)\right)\right\} = \frac{1}{2}\text{Im}\left\{H^2(e^{j\omega})\overline{\psi_m(\omega,\omega)}\right\}.$$

Therefore

$$\begin{aligned}&\lim_{N\to\infty} N\mathbf{E}\left\{\left|\Phi_y(\omega,\widehat{\theta}_N)-\Phi_y(\omega)\right|^2\right\} = \mu|H(e^{j\omega})|^4+2\sigma^4[H_r(\omega),H_i(\omega)]\times\\ &\begin{bmatrix}|H(e^{j\omega})|^2\varphi(\omega,\omega)+\text{Re}\left\{H^2(e^{j\omega})\overline{\psi_m(\omega,\omega)}\right\} & \text{Im}\left\{H^2(e^{j\omega})\overline{\psi_m(\omega,\omega)}\right\}\\ \text{Im}\left\{H^2(e^{j\omega})\overline{\psi_m(\omega,\omega)}\right\} & |H(e^{j\omega})|^2\varphi(\omega,\omega)-\text{Re}\left\{H^2(e^{j\omega})\overline{\psi_m(\omega,\omega)}\right\}\end{bmatrix}\begin{bmatrix}H_r(\omega)\\ H_i(\omega)\end{bmatrix}\\ =\ &\mu|H(e^{j\omega})|^4+2\sigma^4|H(e^{j\omega})|^4\left[\varphi_m(\omega,\omega)+\text{Re}\left\{\psi_m(\omega,\omega)\right\}\right].\end{aligned}$$

This quantifies the block diagonal entries of $\Sigma(\omega,\lambda)$. For the off-diagonal entries we follow a similar procedure and note that

$$\lim_{N\to\infty} N\mathbf{E}\left\{\left(\widehat{H}_r(\omega)-H_r(\omega)\right)\left(\widehat{H}_r(\lambda)-H_r(\lambda)\right)\right\} = \frac{1}{2}\mathrm{Re}\left\{\overline{H(e^{j\omega})}H(e^{j\lambda})\varphi_m(\lambda,\omega)+H(e^{j\omega})H(e^{j\lambda})\overline{\psi_m(\lambda,\omega)}\right\}$$

$$\lim_{N\to\infty} N\mathbf{E}\left\{\left(\widehat{H}_i(\omega)-H_i(\omega)\right)\left(\widehat{H}_i(\lambda)-H_i(\lambda)\right)\right\} = \frac{1}{2}\mathrm{Re}\left\{\overline{H(e^{j\omega})}H(e^{j\lambda})\varphi_m(\lambda,\omega)-H(e^{j\omega})H(e^{j\lambda})\overline{\psi_m(\lambda,\omega)}\right\}$$

and

$$\lim_{N\to\infty} N\mathbf{E}\left\{\left(\widehat{H}_r(\omega)-H_r(\omega)\right)\left(\widehat{H}_i(\lambda)-H_i(\lambda)\right)\right\} = \frac{1}{2}\mathrm{Im}\left\{\overline{H(e^{j\omega})}H(e^{j\lambda})\varphi_m(\lambda,\omega)+H(e^{j\omega})H(e^{j\lambda})\overline{\psi_m(\lambda,\omega)}\right\}$$

so that

$$\lim_{N\to\infty} N\mathbf{E}\left\{\left(\Phi_y(\omega,\widehat{\theta}_N)-\Phi_y(\omega)\right)\left(\Phi_y(\lambda,\widehat{\theta}_N)-\Phi_y(\lambda)\right)\right\} = \mu|H(e^{j\omega})|^2|H(e^{j\lambda})|^2+2\sigma^4[H_r(\omega),H_i(\omega)]\times$$

$$\begin{bmatrix} \mathrm{Re}\left\{\overline{H(e^{j\omega})}H(e^{j\lambda})\varphi_m+H(e^{j\omega})H(e^{j\lambda})\overline{\psi_m}\right\}, \mathrm{Im}\left\{\overline{H(e^{j\omega})}H(e^{j\lambda})\varphi_m+H(e^{j\omega})H(e^{j\lambda})\overline{\psi_m}\right\} \\ \mathrm{Im}\left\{H(e^{j\omega})H(e^{j\lambda})\overline{\psi_m}-H(e^{j\lambda})\overline{H(e^{j\omega})}\varphi_m+\right\}, \mathrm{Re}\left\{\overline{H(e^{j\omega})}H(e^{j\lambda})\varphi_m-H(e^{j\omega})H(e^{j\lambda})\overline{\psi_m}\right\} \end{bmatrix}\begin{bmatrix} H_r(\lambda) \\ H_i(\lambda) \end{bmatrix}$$

$$= \quad |H(e^{j\omega})|^2|H(e^{j\lambda})|^2\left[\mu+2\sigma^4\mathrm{Re}\left\{\varphi_m(\lambda,\omega)+\psi_m(\lambda,\omega)\right\}\right].$$

□

C Proof of Theorem 3.2

Proof. In what follows, the notation $\cdot'$ and $\cdot''$ will be used to denote differentiation, and double differention (respectively) with respect to θ. With this in mind, by the definition of $\widehat{\theta}_N$

$$\left.\frac{\mathrm{d}V_N(\theta)}{\mathrm{d}\theta}\right|_{\theta=\widehat{\theta}_N} = 0 \quad \text{w.p.1}$$

Now choose some $\theta_\circ\in\Theta$. Then using the Mean Value Theorem, for large enough N $\exists\alpha\in[0,1]$ such that

$$\frac{\mathrm{d}V_N(\theta_\circ)}{\mathrm{d}\theta} = R_N(\beta)(\theta_\circ-\widehat{\theta}_N) \quad \text{w.p.1} \tag{C.69}$$

where

$$R_N(\beta)\triangleq\left.\frac{\mathrm{d}^2V_N(\theta)}{\mathrm{d}\theta\mathrm{d}\theta^T}\right|_{\theta=\beta}, \qquad \beta\triangleq\alpha\widehat{\theta}_N+(1-\alpha)\theta_\circ. \tag{C.70}$$

Furthermore, as established in [9, 5, 16, 13], for any $\theta_\circ\in\Theta$

$$\sqrt{N}\frac{\mathrm{d}V_N(\theta_\circ)}{\mathrm{d}\theta}\xrightarrow{\mathcal{D}}\mathcal{N}(0,M) \qquad \text{as } N\to\infty \tag{C.71}$$

where

$$M \quad = \quad \lim_{N\to\infty}\mathbf{E}\left\{\frac{\mathrm{d}}{\mathrm{d}\theta}V_N(\theta_\circ)\left[\frac{\mathrm{d}}{\mathrm{d}\theta}V_N(\theta_\circ)\right]^T\right\} = \sigma^2T_n\left(\frac{Z_\circ(e^{j\omega})Z_\circ^\star(e^{j\omega})}{|C(e^{j\omega})|^2}\right). \tag{C.72}$$

Now, as established in Lemma D.1, this matrix is singular if the model order m is greater than an underlying true one ℓ. In this case, T_n will have a spectral decomposition

$$T_n\left(\frac{Z_\circ(e^{j\omega})Z_\circ^\star(e^{j\omega})}{|C(e^{j\omega})|^2}\right) = [V_1, V_2]\left[\begin{array}{cc} S_1 & 0 \\ 0 & 0 \end{array}\right]\left[\begin{array}{c} V_1^T \\ V_2^T \end{array}\right] = V_1 S_1 V_1^T \tag{C.73}$$

where S_1 is a diagonal matrix formed from the non-zero eigenvalues of T_n. This allows the definition of the pseudo-inverse $T_n^\dagger$ as

$$T_n^\dagger\left(\frac{Z_\circ(e^{j\omega})Z_\circ^\star(e^{j\omega})}{|C(e^{j\omega})|^2}\right) \triangleq V_1 S_1^{-1} V_1^T. \tag{C.74}$$

In this case (C.71) implies that

$$\sqrt{N}\left[\frac{H'(e^{j\omega},\theta_\circ)}{H(e^{j\omega})}\right]^\star T_n^\dagger\left(\frac{Z_\circ(e^{j\omega})Z_\circ^\star(e^{j\omega})}{|C(e^{j\omega})|^2}\right) V_N'(\theta_\circ) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2\phi_m(\omega,\omega)\right), \qquad \text{as } N\to\infty \tag{C.75}$$

where

$$\overline{\phi_m(\lambda,\omega)} \triangleq \left[\frac{H'(e^{j\lambda},\theta_\circ)}{H(e^{j\lambda})}\right]^\star T_n^\dagger\left(\frac{Z_\circ Z_\circ^\star}{|C|^2}\right)\frac{H'(e^{j\omega},\theta_\circ)}{H(e^{j\omega})}. \tag{C.76}$$

Furthermore, as established in the proof of Lemma D.1, a vector x is in the kernel of $T_n(Z_\circ Z_\circ^\star/|C|^2)$ if, and only if

$$x^\star\left[\frac{1}{H(e^{j\omega})}\frac{\mathrm{d}H(e^{j\omega})}{\mathrm{d}\theta}\right] = 0, \qquad \omega\in[-\pi,\pi].$$

Therefore, $H'(e^{j\omega})/H(e^{j\omega})$ is orthogonal to this kernel for all ω, and hence using (24)

$$\begin{aligned}\left\langle \frac{H'(e^{j\lambda},\theta_\circ)}{H(e^{j\lambda})}, \phi_m(\lambda,\omega)\right\rangle &= \left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{H'(e^{j\lambda},\theta_\circ)}{H(e^{j\lambda})}\left[\frac{H'(e^{j\lambda},\theta_\circ)}{H(e^{j\lambda})}\right]^\star \mathrm{d}\lambda\right] T_n^\dagger\left(\frac{Z_\circ Z_\circ^\star}{|C|^2}\right)\frac{H'(e^{j\omega},\theta_\circ)}{H(e^{j\omega})} \\ &= \frac{H'(e^{j\omega},\theta_\circ)}{H(e^{j\omega})}.\end{aligned} \tag{C.77}$$

Now, suppose that a minimal version of the m'th order $H(z) = C(z)/D(z)$ can be written as the ℓ'th order $H(z) = C_\ell(z)/D_\ell(z)$. Then the linear span of the columns of $H'(z)/H(z)$ is the space Z_m defined as

$$Z_m \triangleq \left\{f : f = \sum_{k=0}^{m-1}\frac{\alpha_k z^k D_\ell(z) + \beta_k z^k C_\ell(z)}{C_\ell(z)D_\ell(z)T(z)},\ \alpha_k,\beta_k\in\mathbf{C}\right\}. \tag{C.78}$$

Therefore, (C.77) implies that $\overline{\phi_m(\lambda,\omega)}$ is the reproducing kernel for the space Z_m. According to the argument used previously in the proof of Corollary 3.1, this kernel is unique and also expressible as

$$\phi_m(\lambda,\omega) = \sum_{k=0}^{m+\ell-1}\mathcal{B}_k(e^{j\lambda})\overline{\mathcal{B}_k(e^{j\omega})} \tag{C.79}$$

with the factorisation (56) defining the zeros $\{\xi_k\}$ which further determine the orthonormal basis functions $\{\mathcal{B}_k(z)\}$ above.

Furthermore, returning to (C.69), note that as established in [9, 5, 16, 13] $\lim_{N\to\infty} R_N(\beta) = \lim_{N\to\infty} \mathbf{E}\{R_N(\theta_\circ)\}$ element-wise and with probability one. Finally, using the block Toeplitz matrix definition (25) and again using (24)

$$\left[\frac{H'(e^{j\omega},\theta_\circ)}{H(e^{j\omega})}\right]^\star T_n^\dagger\left(\frac{Z_\circ Z_\circ^\star}{|C|^2}\right)\mathbf{E}\{R_N(\theta_\circ)\}(\theta_\circ - \widehat{\theta}_N) =$$

$$\begin{aligned}
& \sigma^2\left[\frac{H'(e^{j\omega},\theta_\circ)}{H(e^{j\omega})}\right]^\star T_n^\dagger\left(\frac{Z_\circ Z_\circ^\star}{|C|^2}\right)T_n\left(\frac{Z_\circ Z_\circ^\star}{|C|^2}+\delta I_2\right)(\theta_\circ - \widehat{\theta}_N)\\
= \; & \sigma^2\left\langle \phi_m(\lambda,\omega),(\theta_\circ - \widehat{\theta}_N)^T\frac{H'(e^{-j\lambda},\theta_\circ)}{H(e^{-j\lambda})}\right\rangle + \sigma^2\delta\left[\frac{H'(e^{j\omega},\theta_\circ)}{H(e^{j\omega})}\right]^\star T_n^\dagger\left(\frac{Z_\circ Z_\circ^\star}{|C|^2}\right)(\theta_\circ - \widehat{\theta}_N)\\
= \; & \sigma^2\left[\frac{H'(e^{j\omega},\theta_\circ)}{H(e^{j\omega})}\right]^\star(\theta_\circ - \widehat{\theta}_N) + \sigma^2\delta\left[\frac{H'(e^{j\omega},\theta_\circ)}{H(e^{j\omega})}\right]^\star T_n^\dagger\left(\frac{Z_\circ Z_\circ^\star}{|C|^2}\right)(\theta_\circ - \widehat{\theta}_N)
\end{aligned}$$

with the progression to the last line following from the reproducing kernel property of $\phi_m(\lambda,\omega)$. Collectively then, this result together with (C.75) (C.71) and the afore-mentioned fact that $R_N(\beta) \to \mathbf{E}\{R_N(\theta_\circ)\}$ with probability one implies that with δ arbitrarily small

$$\sqrt{N}\left[\frac{H'(e^{j\omega},\theta_\circ)}{H(e^{j\omega})}\right]^\star(\theta_\circ - \widehat{\theta}_N) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\phi_m(\omega,\omega)\right), \qquad \text{as } N\to\infty \tag{C.80}$$

and hence via the Taylor expansion (19)

$$\sqrt{N}\left[\; H(e^{j\omega},\widehat{\theta}_N) - H(e^{j\omega}) \;\right] \xrightarrow{\mathcal{D}} \mathcal{N}(0,|H(e^{j\omega})|^2\phi_m(\omega,\omega)), \qquad \text{as } N\to\infty. \tag{C.81}$$

The remainder of the proof is now identical to the argument employed to prove Corollary 3.1 and Corollary 3.1 upon recognising that with the indicated modification of the zeros $\{\xi_k\}$ according to (56), then $\phi_m(\lambda,\omega) = \varphi_m(\lambda,\omega)$. □

D Technical Lemma

Lemma D.1. *The symmetric block Toeplitz matrix*

$$T_n\left(\frac{Z(e^{j\omega},\theta_\circ)Z^\star(e^{j\omega},\theta_\circ)}{|C(e^{j\omega})|^2}\right) \tag{D.1}$$

is positive definite, and hence invertible, if and only if there are no pole zero cancellations in $H(z)$.

Proof. By definition

$$T_n\left(\frac{Z_\circ Z_\circ^\star}{|C|^2}\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{|H(e^{j\omega})|^2}\left[\Lambda_m(e^{j\omega})\otimes I_2\right]\frac{Z_\circ(e^{j\omega})Z_\circ^\star(e^{j\omega})}{|D(e^{j\omega})|^2}\left[\Lambda_m^\star(e^{j\omega})\otimes I_2\right]\mathrm{d}\omega$$

Therefore suppose that the matrix (D.1) is rank deficient. Then via (24) there exists a non-zero $x \in \mathbf{C}$ such that

$$0 \;=\; x^\star T_n\left(\frac{Z(e^{j\omega},\theta_\circ)Z^\star(e^{j\omega},\theta_\circ)}{|C|^2}\right)x = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|x^\star\frac{\mathrm{d}H(e^{j\omega})}{\mathrm{d}\theta}\right|^2\frac{1}{|H(e^{j\omega}|^2}\,\mathrm{d}\omega.$$

Since the integrand is non-negative, the integral is zero if, and only if

$$x^\star \frac{\mathrm{d}H(e^{j\omega})}{\mathrm{d}\theta} = 0 \tag{D.2}$$

for all $\omega \in [-\pi, \pi]$. However, according to the model structure (4)

$$x^\star \frac{\mathrm{d}H(z,\theta_\circ)}{\mathrm{d}\theta} = \frac{1}{D^2(z,\theta_\circ)}\left[p_1(z)D(z,\theta_\circ) + p_2(z)C(z,\theta_\circ)\right]$$

for some polynomials $p_1(z), p_2(z)$ that are determined by x. With $\partial p(z)$ used to denote the order of a polynomial $p(z)$ then

$$\partial p_1(z) = \partial p_2(z) = m-1, \qquad \partial D(z,\theta_\circ) = m, \qquad \partial C(z,\theta_\circ) = m.$$

Therefore, in seeking to set

$$p_1(z)D(z,\theta_\circ) + p_2(z)C(z,\theta_\circ) = 0 \tag{D.3}$$

if there are no common zeros between $C(z,\theta_\circ)$ and $D(z,\theta_\circ)$, then the left hand side is an order $2m-1$ polynomial, while there there are only $2(m-1)$ degrees of freedom in the choices of $p_1(z)$ and $p_2(z)$. Therefore, (D.2) is impossible and the matrix (D.1) is positive definite. On the other hand, any pole-zero cancellation between $C(z,\theta_\circ)$ and $D(z,\theta_\circ)$ will reduce the order of the left hand side of (D.3) to $2m-2$ or less, and hence (D.3) can be achieved for some $p_1(z)$ and $p_2(z)$, implying that (D.1) is not full rank. □

References

[1] N. Aronszajn, *Theory of reproducing kernels*, Acta Mathematica, (1950), pp. 337–404.

[2] K. Berk, *Consistent autoregressive spectral estimates*, Annals of Statistics, 2 (1974), pp. 489–502.

[3] J. W. Brewer, *Kronecker products and matrix calculus in system theory*, IEEE Transactions on Circuits and Systems, 25 (1978), pp. 772–781.

[4] D. R. Brillinger, *Time Series: Data Analysis and Theory*, Holden-Day, 1981.

[5] P. Caines, *Linear Stochastic Systems*, John Wiley and Sons, New York, 1988.

[6] P. Davis, *Interpolation and Approximation*, Blaisdell Publishing Company, 1963.

[7] U. Forsell and L. Ljung, *Closed-loop identification revisited*, Automatica, 53 (1999), pp. 1215–1241.

[8] B. Friedlander and B. Porat, *On the computation of an asymptotic bound for estimating autoregressive signals in white noise*, Signal Process., 8 (1985), pp. 291–302.

[9] E. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*, John Wiley and Sons, New York, 1988.

[10] E. Hannan and B. Wahlberg, *Convergence rates for inverse Toeplitz matrix forms*, Journal of Multivariate Analysis, 31 (1989), pp. 127–135.

[11] K. HOFFMAN, *Banach Spaces of Analytic Functions*, Prentice-Hall, Inc., New Jersey, 1962.

[12] S. M. KAY, *Modern Spectral Estimation: Theory and Application*, Prentice Hall, 1987.

[13] L. LJUNG, *System Identification: Theory for the User, (2nd edition)*, Prentice-Hall, Inc., New Jersey, 1999.

[14] L.LJUNG, *Convergence analysis of parametric identification methods*, IEEE Transactions on Automatic Control, AC-23 (1978), pp. 770–783.

[15] ——, *Asymptotic variance expressions for identified black-box transfer function models*, IEEE Transactions on Automatic Control, AC-30 (1985), pp. 834–844.

[16] L.LJUNG AND P.E.CAINES, *Asymptotic Normality of prediction error estimators for approximate system models*, Stochastics, 3 (1979), pp. 29–46.

[17] B. NINNESS AND F. GUSTAFSSON, *A unifying construction of orthonormal bases for system identification*, IEEE Transactions on Automatic Control, 42 (1997), pp. 515–521.

[18] P.M.J. VAN DEN HOF, P.S.C. HEUBERGER, AND J. BOKOR, *System identification with generalized orthonormal basis functions*, Automatica, 31 (1995), pp. 1821–1834.

[19] M. PRIESTLEY, *Spectral Analysis and Time Series*, Academic Press, 1981.

[20] P. STOICA AND R. MOSES, *Introduction to Spectral Analysis*, Prentice Hall, 1997.

[21] T.SÖDERSTRÖM AND P.STOICA, *System Identification*, Prentice Hall, New York, 1989.

[22] B. WAHLBERG, *System identification using Laguerre models*, IEEE Transactions on Automatic Control, AC-36 (1991), pp. 551–562.

[23] B. WAHLBERG, *System identification using Kautz models*, IEEE Transactions on Automatic Control, AC-39 (1994), pp. 1276–1282.

[24] L. L. XIE AND L. LJUNG, *Variance expressions for spectra estimated using auto-regression*, Journal of Econometrics, (2002), p. To appear.

[25] Z. YUAN AND L. LJUNG, *Black box identification of multivariable transfer functions:Asymptotic properties and optimal input design*, International Journal of Control, 40 (1984), pp. 233–256.

[26] Y. ZHU, *Estimation of transfer functions:Asymptotic theory and a bound of model uncertainty*, International Journal of Control, 49 (1989), pp. 2241–2258.