On the CRLB for Combined Model and Model-Order Estimation of Stationary Stochastic Processes

Brett Ninness

Abstract—This letter is concerned with quantifying the Cramér-Rao lower bound for model-based spectral density estimation in the case of joint model and model-order estimation. In particular, the results here extend previous work by providing closed-form frequency domain expressions that, among other things, highlight the effect of order estimation bias on the total accuracy of model-based spectral density estimation.

Index Terms—Autoregressive moving-average (ARMA) modeling, Cramér-Rao bound, maximum-likelihood estimation, spectrum estimation.

I. INTRODUCTION

CONSIDER THE CASE of obtaining N observations

$$Y_N \triangleq [y_0, y_1, \ldots, y_{N-1}]^T \qquad (1)$$

of a wide-sense stationary stochastic process with spectral density $\Phi_y(\omega)$, and on the basis of these forming an estimate of that spectrum via a model-based approach that involves the intermediate estimation $\widehat{\theta}_n^N$ of a parameter vector $\theta_n \in \mathbf{R}^n$.

A fundamental measure of the nature of this estimation problem is provided by the Fisher Information matrix $\mathcal{I}_n$

$$\mathcal{I}_n = \mathbf{E}_{\theta_n^\circ} \left\{ \frac{dp(Y_N|\theta_n)}{d\theta_n} \cdot \left(\frac{dp(Y_N|\theta_n)}{d\theta_n} \right)^T \right\} \qquad (2)$$

and the associated Cramér–Rao lower bound (CRLB) on estimate variability that arises as the inverse of this matrix. In (2), $p(Y_N|\theta_n)$ is the likelihood of the observed data conditional on $\theta_n \in \mathbf{R}^n$, and $\mathbf{E}_{\theta_n^\circ}\{\cdot\}$ denotes expectation with respect to $p(Y_N|\theta_n^\circ)$ where θ_n° is the true parameter vector that characterizes the process $\{y_t\}$.

An interesting recent contribution [1] recognizes the fact that if the order n true parameter vector θ_n° is unknown, then $\mathcal{I}_n^{-1}$ no longer quantifies the CRLB for the spectral density estimation problem. That work then establishes the following key result.

Theorem 1.1: For the case of a true underlying process of order n°, a maximal considered model order $\overline{n}$, a scheme that derives a model-order estimate $\widehat{n}$ that satisfies (for $N \to \infty$)

$$p(\widehat{n} = n) = \begin{cases} p_n, & n \geq n^\circ \\ 0, & n < n^\circ \end{cases} \qquad (3)$$

and a situation in which $|\mathcal{I}_n| > 0$ for all $n \in [n^\circ, \overline{n}]$, then the CRLB matrix V involved with joint estimation of order $n \leq \overline{n}$ and model parameterization θ_n is given by

$$V = \sum_{n=n^\circ}^{\overline{n}} p_n V_n \qquad (4)$$

where each $V_n \in \mathbf{R}^{\overline{n} \times \overline{n}}$ is defined via (2) as

$$V_n \triangleq \begin{bmatrix} \mathcal{I}_n^{-1} & \varnothing \\ \varnothing & \varnothing \end{bmatrix}. \qquad (5)$$

Proof: See [1]. □

Although the proof presented in [1] does not depend on assumptions of the true process type, the assumption of $|\mathcal{I}_n| > 0$ for $n > n^\circ$ is only satisfied by processes of AR or MA type, and hence [1] concentrates on models of that type. In particular, that work focuses on evaluating the above main result by virtue of its implied CRLB for the estimated spectral density $\Phi_y(\omega, \widehat{\theta}_n^n)$, viz. [1, eq. (16)]

$$\text{CRLB} \left\{ \Phi\left(\omega, \widehat{\theta}_n^N\right) \right\} = \frac{d\Phi(\omega, \theta_n^\circ)}{d\theta_n} \cdot V \cdot \left[\frac{d\Phi(\omega, \theta_n^\circ)}{d\theta_n} \right]^T. \quad (6)$$

The right-hand side of (6) is then numerically evaluated in [1] over several different simulation examples in order to illustrate the significance of the parameter space result of Theorem 1.1.

The contribution of this letter is to extend the work of [1] by establishing a closed-form expression for the right-hand side of (6) which provides a theoretical basis to explain certain phenomena observed empirically in [1], such as the effects on the variance of $\Phi_y(\omega, \widehat{\theta}_n^N)$ according to the distance of AR process poles (MA process zeros) to the origin, and the variance increasing effects of order estimation bias.

II. MAIN RESULTS

To begin with, this letter is concerned with situations in which the observed data $\{y_t\}$ is a realization of an order $n = n^\circ$ autoregressive (AR) or moving average (MA) process. That is, with $D(q, \theta_n^\circ)$ being an n°th-order polynomial in the forward shift operator q, and with $\{e_t\}$ being a zero-mean independently and identically distributed (i.i.d.) process, either

$$q^{-n} D\left(q, \theta_n^\circ\right) y_t = e_t \qquad (7)$$

for the AR case or, in the MA case

$$y_t = q^{-n} D\left(q, \theta_n^\circ\right) e_t. \qquad (8)$$

For these two scenarios of an observed realization of N data points, we will denote the asymptotic *normalized* CRLB for a model of *fixed* order n as $\phi_n(\omega)$, i.e.,

$$\phi_n(\omega) \triangleq \lim_{N \to \infty} N \cdot \text{CRLB} \left\{ \frac{\Phi_y\left(\omega, \widehat{\theta}_n^N\right)}{\Phi_y(\omega)} \right\}$$

$$\leq \lim_{N \to \infty} N \text{Var} \left\{ \frac{\Phi_y\left(\omega, \widehat{\theta}_n^N\right)}{\Phi_y(\omega)} \right\}. \tag{9}$$

The associated CRLB for the *joint* estimation of normalized spectral estimate $\Phi_y(\omega, \widehat{\theta}_n^N)/\Phi_y(\omega)$ and model-order estimate $\widehat{n}$ will be denoted by $\phi(\omega)$. The first main result of this letter is that $\phi(\omega)$ can be quantified in closed form as follows.

Theorem 2.1: Under the assumptions of Theorem 1.1 and with the definition of the zeros $\{\xi_k\}$ via

$$D\left(z, \theta_n^\circ\right) = (z - \xi_0)(z - \xi_1) \cdots (z - \xi_{n^\circ - 1}) \tag{10}$$

then for either of the AR or MA cases

$$\frac{\phi(\omega)}{2} = \mathbf{E}\{\widehat{n} - n^\circ\} + \varphi_{n^\circ}(\omega) + Re\{\psi_{n^\circ}(\omega)\}$$
$$+ Re\left\{ \beta_{n^\circ}^2(\omega) \cdot \sum_{n=n^\circ+1}^{\overline{n}} p_n \frac{e^{-j(n-n^\circ+1)\omega} \sin(n-n^\circ)\omega}{\sin\omega} \right\} \tag{11}$$

where the following expressions hold

$$\varphi_{n^\circ}(\omega) \triangleq \sum_{k=0}^{n^\circ-1} \frac{1 - |\xi_k|^2}{|e^{j\omega} - \xi_k|^2}$$

$$\beta_k(\omega) \triangleq \prod_{\ell=0}^{k-1} \left(\frac{1 - e^{j\omega}\overline{\xi_\ell}}{e^{j\omega} - \xi_\ell} \right)$$

$$\beta_0(\omega) = 1 \tag{12}$$

$$\psi_{n^\circ}(\omega) = \sum_{k=0}^{n^\circ-2\rho-1} \frac{1 - \xi_k^2}{(e^{j\omega} - \xi_k)^2} \cdot \beta_k^2(\omega)$$
$$+ \sum_{\tau=0}^{\rho-1} \frac{\left(1 - |\xi_{r(\tau)}|^4\right)(e^{j\omega} - \alpha)^2}{\left(e^{j\omega} - \xi_{r(\tau)}\right)^2 \left(e^{j\omega} - \overline{\xi_{r(\tau)}}\right)^2} \cdot \beta_{r(\tau)}^2(\omega) \tag{13}$$

with

$$r(\tau) \triangleq n^\circ - 2\rho + 2\tau$$

$$\alpha \triangleq \frac{\xi_r(\tau) + \overline{\xi}_r(\tau) - \sqrt{\left(1 - \xi_{r(\tau)}^2\right)\left(1 - \overline{\xi}_{r(\tau)}^2\right)}}{1 + \left|\xi_{r(\tau)}\right|^2}. \tag{14}$$

In (13) and (14), it has been assumed (without loss of generality) that the zeros defined by (10) are arranged so that the first $n_\circ - 2\rho$ of them are purely real valued, and the remaining 2ρ then occur in complex conjugate pairs.

Proof: First, note that according to the results of [2, Cor. 3.2], then for $n \geq n^\circ$

$$\frac{\phi_n(\omega)}{2} = (n - n^\circ) + \varphi_{n^\circ}(\omega) + Re\{\psi_{n^\circ}(\omega)\}$$
$$+ Re\left\{ \beta_{n^\circ}^2(\omega) \cdot \sum_{k=1}^{n-n^\circ} e^{-j2k\omega} \right\} \tag{15}$$

where the sum in the last term should be taken as zero for the case $n = n^\circ$. Furthermore, according to (6) and the results of Theorem 1.1

$$\phi(\omega) = \sum_{n=n^\circ}^{\overline{n}} p_n \phi_n(\omega). \tag{16}$$

Substitution of (15) in (16) and using the facts that

$$\sum_{n=n^\circ}^{\overline{n}} n p_n = \mathbf{E}\{\widehat{n}\} \qquad \sum_{n=n^\circ}^{\overline{n}} p_n = 1 \tag{17}$$

then completes the proof. □

There are several important conclusions to be drawn from this result. To elucidate them, first note that the quantification (11) for joint model and model-order estimation CRLB $\phi(\omega)$ is dominated by the first two terms. This is because the final two terms are the real part of the sum of quantities which themselves have both positive and negative real parts, and they therefore tend to remain small relative to the preceding two terms of (11) which being the sum of strictly positive quantities, are generically larger.

With this in mind, further note that the component $\varphi_{n^\circ}(\omega) + Re\{\psi_{n^\circ}(\omega)\}$ is the CRLB for the spectral density estimation for known *and true* model order $n = n^\circ$ and is independent of any model-order selection rule used. The additional $\mathbf{E}\{\widehat{n} - n^\circ\}$ component of (11) is the bias of the model-order selection rule used and is independent of the nature of the associated spectral density.

The new quantification (11), therefore, highlights that order estimation bias produces a constant additive offset $\mathbf{E}\{\widehat{n} - n^\circ\}$ on the CRLB relative to the known order case. The phenomenon was observed empirically in [1], but was not be explained or quantified there.

Furthermore, with the exception of the very last term of (11) (which has just been argued to be negligible in comparison to earlier terms), Theorem 2.1 indicates that there is a decoupled nature to the contribution of order estimation bias and underlying true spectral density on the CRLB of joint density and order estimation.

In addition, the closed-form expression (11) for the CRLB highlights that because all the denominators in (12) and (13), are small when $e^{j\omega}$ is close to any of the zeros $\{\xi_k\}$, then the relative estimation error is likely to be larger at those frequencies near the zeros of $D(z, \theta_n^\circ)$. 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 (12) and (13) permits (11), with the last $\sin(n - n^\circ)\omega/\sin\omega$ term neglected to be evaluated as

$$\lim_{N \to \infty} \frac{N}{2} \text{Var} \left\{ \frac{\Phi_y(\omega, \widehat{\theta}_N)}{\Phi_y(\omega)} \right\} = \mathbf{E}\{\widehat{n} - n^\circ\} + \sum_{k=0}^{n^\circ-1} \frac{1 - |\xi_k|^2}{|e^{j\omega} - \xi_k|^2}$$
$$\cdot \cos^2\left(\gamma_k(\omega) + \sum_{\ell=0}^{k-1} [2\gamma_\ell(\omega) - \omega] \right) \tag{18}$$

where the notation $\gamma_\ell(\omega) \triangleq \measuredangle(e^{j\omega} - \xi_\ell)$ has been used. Simple geometry then indicates that poles and zeros near the origin lead to small and smooth variations in the subtended angle $\gamma_\ell(\omega)$, and hence via (18) smooth variations in spectral estimate variability

with changing ω, and vice-versa for poles and zeros near the boundary of the unit circle.

Again, this provides a theoretical background for explaining phenomena that were first observed empirically in [1], such as the fact that the variance increasing effect of combined model and model-order estimation is "*more pronounced when the poles are close to the origin.*" According to the results of this letter and the above discussion, this is clearly because poles away from the origin produce higher magnitude peaks in the first component of (11) which tend to swamp the constant offset produced by the second component of (11). This reasoning is also consistent with the figures provided in [1].

The additional effect of the last term in (11) is one which, since $|\beta_{n^\circ}(\omega)| = 1$, is dominated by the nature of the sinc function term $\sin(n - n^\circ)\omega / \sin \omega$ which adds an "oscillatory" component to the CRLB, of frequency $(n - n^\circ)/2\pi$, i.e., faster oscillatory variation as n increases beyond n°.

Finally, returning again to (11) which applies for any real or complex value of the $\{\xi_k\}$, it can be used to establish the following final result of this section

Corollary 2.1: Under the same conditions imposed in Theorem 2.1

$$\lim_{N \to \infty} \frac{N}{2\pi} \int_{-\pi}^{\pi} \text{Var} \left\{ \frac{\Phi_y(\omega, \widehat{\theta}_N)}{\Phi_y(\omega)} \right\} d\omega = 2\mathbf{E}\{\widehat{n}\}. \quad (19)$$

Proof: First, by the formulation (12)

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi_{n^\circ}(\omega) d\omega = n^\circ \quad (20)$$

which can be simply established by rewriting the integral as a contour one using the substitution $z = e^{j\omega}$ and then using Cauchy's Residue Theorem [3]. Furthermore, using the substitution $z = e^{-j\omega}$ (for ease of notation, redefine ψ_{n° as taking argument $e^{j\omega}$ rather that ω)

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \psi_{n^\circ}(e^{j\omega}) d\omega = \frac{1}{2\pi j} \oint_{|z|=1} \psi_{n^\circ}(z^{-1}) \frac{dz}{z} = 0 \quad (21)$$

where the latter equality to zero holds according to Cauchy's Integral Theorem since by the formulation (13), $z^{-1}\psi_{n^\circ}(z^{-1})$ is analytic within the unit disk $|z| < 1$. Finally, via a similar argument and for any $k \geq 1$ (and again for ease of notation, redefine β_{n° as taking argument $e^{j\omega}$ rather that ω)

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \beta_{n^\circ}^2(e^{j\omega}) e^{-j2k\omega} d\omega = \frac{1}{2\pi j} \oint_{|z|=1} \beta_{n^\circ}^2(z^{-1}) z^{2k} \frac{dz}{z} = 0 \quad (22)$$

again since the integrand in the contour integral is, by the formulation (12), analytic within the unit disk. Substitution of (20)–(22) into the integral with respect to ω of (11) then completes the proof after recognising that the rightmost term in (11) can be written as the rightmost summation term in (15). $\square$

This surprisingly simple relation between the average relative estimation error over all frequencies and the expected value of the order estimate $\widehat{n}$ illustrates a "waterbed effect." Specifically, although as just discussed, the expression (11) indicates increased relative error near zeros of $D(z, \theta_n^\circ)$, with increased effect according to distance from the unit circle, the quantifi-

cation (19) establishes that 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 expected value $\mathbf{E}\{\widehat{n}\}$ of the order estimate.

III. ARMA EXTENSIONS

These results can be extended to the ARMA case in which the model structure is

$$D(q, \theta_n^\circ) y_t = C(q, \theta_n^\circ) e_t \quad (23)$$

where now both $D(q, \theta_{2n}^\circ)$ and $C(q, \theta_{2n}^\circ)$ are nth-order polynomials in the forward shift operator q, and they jointly require a vector θ_{2n}° of size $2n$ to parameterize them.

However, an initial difficulty is that the information matrix $\mathcal{I}_n$ in this ARMA situation becomes singular for $n > n^\circ$ [4], [5]. In order to circumvent this, the achieved variance of a particular, but widely used, estimation scheme can be considered, instead of the CRLB underbounding the variance of all (unbiased) estimation schemes. In particular, consider the "regularized least squares" estimation scheme

$$\widehat{\theta}_{2n}^N \triangleq \arg\min_{\theta_{2n} \in \mathbf{R}^{2n}} V_N(\theta_{2n})$$

$$V_N(\theta_{2n}) \triangleq \frac{1}{2N} \sum_{t=1}^{N} \varepsilon_t^2(\theta_{2n}) + \frac{\delta}{2} \|\theta_{2n} - \theta_{2n}^\circ\|^2 \quad (24)$$

where $\varepsilon_t(\theta_{2n})$ is the error associated with the mean square optimal one step ahead predictor $\widehat{y}_{t|t-1(\theta_{2n})}$ associated with the model parameterized by θ_{2n}, viz.

$$\varepsilon_t(\theta_{2n}) = y_t - \widehat{y}_{t|t-1}(\theta_{2n}) = \frac{D(q, \theta_{2n})}{C(q, \theta_{2n})} y_t. \quad (25)$$

In (24), the second term in the definition of $V_N(\theta_{2n})$, with $\delta > 0$ a user chosen parameter, is a so-called "regularization" term [6] that ensures that $\widehat{\theta}_{2n}^N$ is uniquely defined. In this case, let us define the asymptotic in N normalized variance of the associated nth-order ARMA spectral density estimate as

$$\phi_n(\omega, \delta) \triangleq \lim_{N \to \infty} N \cdot \text{Var} \left\{ \frac{\Phi_y\left(\omega, \widehat{\theta}_{2n}^N\right)}{\Phi_y(\omega)} \right\}. \quad (26)$$

The variance associated with *joint* density estimation and model-order estimation will be denoted by $\phi(\omega, \delta)$, and may be quantified by the following final main result.

Theorem 3.1: Under the assumptions of Theorem 2.1, and with the zeros $\{\xi_k\}$ being defined according to

$$C(z, \theta_n^\circ) D(z, \theta_n^\circ) T(z) = (z - \xi_0)(z - \xi_1) \cdots (z - \xi_{n+n^\circ-1}) \quad (27)$$

with the left-to-right ordering of the zeros being preserved on both sides of (27), and where $T(z)$ is an $n - n^\circ$th-order polynomial that contains the common zeros of $C(z, \theta_n)$ and $D(z, \theta_n)$, i.e.,

$$\frac{C(z, \theta_{2n})}{D(z, \theta_{2n})} = \frac{C(z, \theta_{2n}^\circ)}{D(z, \theta_{2n}^\circ)} \cdot \frac{T(z)}{T(z)} \quad (28)$$

then

$$\lim_{\delta \to 0} \frac{\phi(\omega, \delta)}{2} = \mathbf{E}\{n - n^\circ\} + \varphi_{n^\circ}(\omega) + Re\{\psi_{n^\circ}(\omega)\}$$

$$+ \sum_{n=n^\circ+1}^{\overline{n}} p_n Re\{\psi_n(\omega) - \psi_{n^\circ}(\omega)\}. \quad (29)$$

Sample Variances of estimate of ϕ_y vs theory

(a)

Sample Variances of estimate of ϕ_y vs theory

(b)
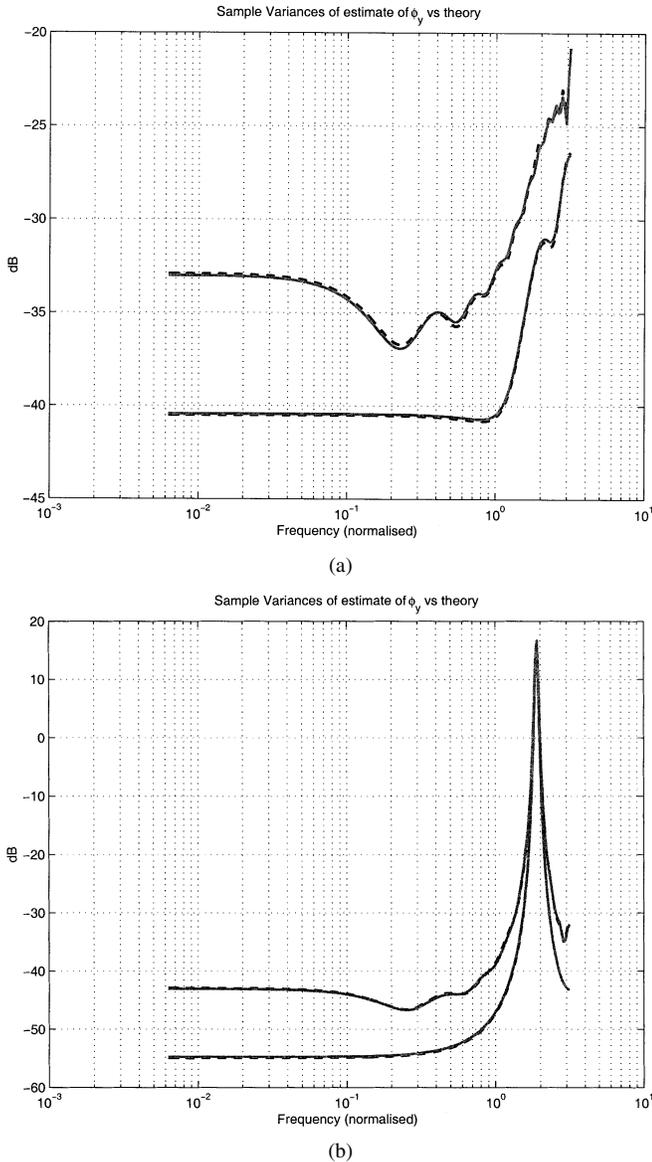
Fig. 1. Variability of $\Phi_y(e^{j\omega}, \widehat{\theta}_n^N)$. The solid line is the true variability, as estimated via averaging over Monte Carlo trial, the dashed line exactly matching it is the new quantification (11) of this letter. The bottom curves corresponds to $n = n^\circ$, and the ones above to $n > n^\circ$.

Proof: In a manner similar to the proof of Theorem 2.1, but this time using the results of [2, Corollary 3.2], for $n \geq n^\circ$

$$\frac{\phi_n(\omega, \delta)}{2} = (n - n^\circ) + \varphi_{n^\circ}(\omega) + \text{Re}\left\{\psi_n(\omega)\right\}. \quad (30)$$

The proof is then completed in the same manner as that of Theorem 2.1. $\square$

It is important to note that the $n - n^\circ$ zeros $\{\xi_{2n^\circ-2}, \ldots, \xi_{n+n^\circ-1}\}$ of $T(z)$ are not uniquely defined by the underlying ARMA process $\{y_t\}$, but instead are uniquely defined by the regularized estimation criterion (24), according to the definition θ_{2n}°.

To expand on this, clearly there are $2n^\circ$ degrees of freedom in θ_{2n}° which are constrained so that $C(z, \theta_{2n}^\circ)/D(z, \theta_{2n}^\circ)$ is the associated n°th-order spectral factor for the ARMA process. However, this still leaves $n - n^\circ$ degrees of freedom available. Indeed, it is this freedom that implies a singular information matrix for $n > n^\circ$, and necessitates the use of the regularized cri-

terion (24) in order for $\widehat{\theta}_{2n}^N$, and hence its variance, to be asymptotically (in N) uniquely defined.

If these spare degrees of freedom are used to constrain the excess zeros of $T(z)$ to all lie at the origin, then the last term in (29) is identical to the last term in (11) and the results for the asymptotic variance of AR, MA and regularized ARMA cases are identical.

However, Theorem 3.1 highlights the further new result that the choice of where the regularized zeros lie has an effect on the achieved asymptotic variability of the spectral density estimate $\Phi_y(\omega, \widehat{\theta}_{2n}^N)$ according to the last term of (30). Additionally, Theorem 3.1 also allows the new insights into AR and MA spectral estimate variability that are provided by the results of Theorem 1.1 (as just discussed) to also be applied to the ARMA case.

IV. SIMULATION EXAMPLES

In order to illustrate the results just presented, this section presents a brief simulation example by considering the same scenario as considered in [1, Fig. 1] where a Gaussian second-order AR process process with zeros

$$\{\xi_0, \xi_1\} = 0.3e^{\pm j0.7\pi} \quad (31)$$

is considered. This situation is analyzed here by estimating the true variability of $\Phi_y(\omega, \widehat{\theta}_n^N)$ (which under Gaussian assumptions achieves the CRLB) via sample average over 1000 different data realizations, each of length $N = 20\,000$ samples.

This is done for two cases of $n = n^\circ = 2$, and $n = \overline{n} = 10$ with the resultant CRLB estimates being shown as the solid lines in Fig. 1. Also shown there is the theoretical quantification of the variability as given by (11).

Fig. 1(a) illustrates the case of the zeros being as in (31). Note that, as explained in the discussion after Theorem 1.1, the effect of biased estimation of order is to add a constant offset to the CRLB equal to the bias $\mathbf{E}\{\widehat{n} - n^\circ\}$, plus to add an oscillatory variation to the CRLB with rate of oscillation again proportional to the bias.

To complete the simulation illustration, Fig. 1(b) shows the case considered in [1, Fig. 2] in which

$$\{\xi_0, \xi_1\} = 0.95e^{\pm j0.6\pi} \quad (32)$$

are chosen for the zeros of the AR process. Again, as predicted by Theorem 1.1, the closeness of the zeros to the unit circle lead to large peaks in the CRLB, which at the peaking frequency swamp the effects of order estimation bias.

REFERENCES

[1] S. Sando, A. Mitra, and P. Stoica, "On the Cramér-Rao bound for model-based spectral analysis," *IEEE Signal Processing Lett.*, vol. 9, pp. 68–71, Feb. 2002.
[2] B. Ninness, "The asymptotic CRLB for the spectrum of ARMA processes," *IEEE Trans. Signal Processing*, vol. 51, pp. 1520–1531, June 2003.
[3] W. Rudin, *Real and Complex Analysis*. New York: McGraw-Hill, 1966.
[4] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1997.
[5] S. M. Kay, *Modern Spectral Estimation: Theory and Application*. Upper Saddle River, NJ: Prentice-Hall, 1987.
[6] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.
[7] S. Kay and S. Marple, "Spectrum analysis-a modern perspective," *Proc. IEEE*, vol. 69, pp. 1380–1416, Nov. 1981.