Order Estimation and Discrimination Between Stationary and Time-Varying (TVAR) Autoregressive Models

Yuri I. Abramovich, *Senior Member, IEEE*, Nicholas K. Spencer, and Michael D. E. Turley

***Abstract*—For a set of T independent observations of the same N-variate correlated Gaussian process, we derive a method of estimating the order of an autoregressive (AR) model of this process, regardless of its stationary or time-varying nature. We also derive a test to discriminate between stationary AR models of order m, $\mathrm{AR}(m)$, and time-varying autoregressive models of order m, $\mathrm{TVAR}(m)$. We demonstrate that within this technique the number T of independent identically distributed data samples required for order estimation and discrimination just exceeds the maximum possible order $m_{\max}$, which in many cases is significantly fewer than the dimension of the problem N.**

***Index Terms*—Adaptive processing, autoregressive (AR), nonstationary interference, time-varying.**

I. Introduction

METHODS for order estimation and parameter estimation of a stationary autoregressive (AR) model of order m, $\mathrm{AR}(m)$, given a set of T independent identically distributed (i.i.d.) N-variate Gaussian samples, are well established [1]–[3]. These methods usually require a complete identification of the model, which can be achieved only approximately in the maximum-likelihood (ML) sense [2]. Indeed, the globally optimal ML solution of the Toeplitz covariance matrix estimation problem is still unknown [4]. Meanwhile, in some cases the strict stationarity of the observed training data is questionable, and so the general problem is to select between a stationary and a time-varying model. For time-varying autoregressive (TVAR) models of order m, $\mathrm{TVAR}(m)$, order-estimation methods have not been reported yet.

Since a stationary $\mathrm{AR}(m)$ model can be considered as a special case of the more general $\mathrm{TVAR}(m)$ model, we expect that such order-estimation techniques would also be applicable to stationary models. In fact, a similar "embedding" was introduced by Wax and Kailath [5], where the problem of estimating the number of independent sources impinging upon an antenna array was substituted by the more general problem of testing the equality of noise-subspace eigenvalues. Instead of the complicated joint detection-estimation problem, therefore, this approach allowed two relatively simple separate routines to be applied: detection via the Wax–Kailath information-theoretic criteria (ITC) technique [5], and estimation (via MUSIC, for example).

A similar approach can be applied for identifying a stationary $\mathrm{AR}(m)$ model provided that there exists a simple and accurate procedure for selecting the order of a $\mathrm{TVAR}(m)$ model. Moreover, when we wish to discriminate between a stationary and a time-varying model, this approach can be used to first estimate the order of a general $\mathrm{TVAR}(m)$ model, and then to test the suitability of the more restrictive stationary $\mathrm{AR}(m)$ model. In essence, we wish to find a process that discriminates between $\mathrm{AR}(\mu)$ and $\mathrm{TVAR}(m)$ models, and if both acceptably match the input data, one that selects the $\mathrm{TVAR}(m)$ model if $\mu > m$. To meet this requirement, therefore, our proposal is to first estimate the minimum possible order of the $\mathrm{TVAR}(m)$ model, and then to decide whether or not a stationary $\mathrm{AR}(m)$ model *of the same order* has an acceptable degradation in terms of the LF. For any given sample volume T and data dimension N, it is clear that some level of nonstationarity exists that cannot be reliably discriminated against the $\mathrm{AR}(m)$ model. For any proposed technique, this "fidelity" needs to be investigated. The purpose of this study is to develop techniques for $\mathrm{TVAR}(m)$ model order estimation and then to test whether the selected ML $\mathrm{TVAR}(m)$ model can be substituted by a stationary (suboptimal in the ML sense) $\mathrm{AR}(m)$ model of the same order.

This paper is organized as follows. Section II describes our technique for $\mathrm{TVAR}(m)$ model order estimation, which relies upon certain properties of the $\mathrm{TVAR}(m)$ ML covariance matrix estimate that we have previously established in [6]. In Section III, we propose the subsequent test for deciding whether the ML $\mathrm{TVAR}(m)$ model can be acceptably replaced by a certain sub-ML stationary $\mathrm{AR}(m)$ model of the same order. In addition, a technique for generating an $\mathrm{AR}(m)$ model (that is ML-suboptimal) is described. Section IV details results of Monte Carlo simulations that support these new techniques and demonstrate their high sensitivity. Our summary and conclusions are presented in Section V, while probability density function (p.d.f.) derivations appear in the Appendices.

II. Order Estimation of a TVAR Model $\mathrm{TVAR}(m)$

In [6], we demonstrated that the necessary and sufficient condition for a vector $\boldsymbol{x} \equiv [x_1, \ldots, x_N]^T$ to be an N-variate sample of the $\mathrm{TVAR}(m)$ process

$$x_j = \sum_{k=1}^{m} a_{kj}^* x_{j-k} + \eta_j, \quad \text{for } j = m+1, \ldots, N \tag{1}$$

Manuscript received September 28, 2005; revised June 27, 2006. The associate editor coordinating the review of this paper and approving it for publication was Dr. Ta-Hsin Li. Some results of this work were presented at the 39th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2005.

Y. I. Abramovich and M. D. E. Turley are with the Intelligence, Surveillance, and Reconnaissance Division, Defence Science and Technology Organization (DSTO), Adelaide 5111, Australia (e-mail: yuri.abramovich@dsto.defence.gov.au; mike.turley@dsto.defence.gov.au).

N. K. Spencer is with Adelaide Research and Innovation Pty Ltd. (ARI), Adelaide 5000, Australia (e-mail: nick.spencer@adelaide.edu.au).

Digital Object Identifier 10.1109/TSP.2007.893966

$$\mathcal{E}\{\eta_j\eta_k^*\} = \sigma_0^2\delta_{jk}, \quad a_0 = 1 \tag{2}$$

is that its positive definite (p.d.) Hermitian covariance matrix

$$R_N^{(m)} \equiv \mathcal{E}\{\boldsymbol{x}\boldsymbol{x}^H\} \tag{3}$$

satisfies the "band-inverse" property

$$\left\{\left[R_N^{(m)}\right]^{-1}\right\}_{jk} = 0, \quad \text{for } |j-k| > m \tag{4}$$

i.e., the elements of its inverse are zero outside the $(2m+1)$-wide diagonal band. Since the p.d. Toeplitz covariance matrix of the stationary $\mathrm{AR}(m)$ model has this same band-inverse property, some test to check the validity of this property may be introduced as a unified test for order m estimation, irrespective of the stationary or time-varying nature of the model.

In this regard, let us consider a set of T i.i.d. N-variate training data

$$\boldsymbol{x}_j \equiv \left[x_1^{(j)}, \ldots, x_N^{(j)}\right]^T, \quad \text{for } j = 1, \ldots, T \tag{5}$$

that are samples of a complex Gaussian random process whose p.d.f. is $\mathcal{CN}(0, R_N)$, where R_N is an N-variate p.d. Hermitian matrix. The sample covariance matrix

$$\hat{R} = \frac{1}{T}\sum_{j=1}^{T}\boldsymbol{x}_j\boldsymbol{x}_j^H \tag{6}$$

is rank deficient for $T < N$, and the matrix $T\hat{R}$ is described by the anti-Wishart complex distribution $\mathcal{ACW}(T < N, N, R_N)$ [7]. Yet, for $T \geq m+1$, all $(m+1)$-variate central block matrices $\hat{R}_q^{(m)}$ of $\hat{R}$ are p.d. [8], i.e.,

$$\hat{R}_q^{(m)} \equiv \begin{bmatrix} \hat{r}_{qq} & \cdots & \hat{r}_{q,q+m} \\ \vdots & \ddots & \vdots \\ \hat{r}_{q+m,q} & \cdots & \hat{r}_{q+m,q+m} \end{bmatrix} > 0, \quad \text{for } q = 1, \ldots, N-m. \tag{7}$$

In [6], we demonstrated that this condition is necessary and sufficient for the existence of an accurate nondegenerate ML estimate of a $\mathrm{TVAR}(m)$ covariance matrix that is calculated directly from the blocks $\hat{R}_q^{(m)}$ using the Dym–Gohberg formula [9]–[12]

$$\hat{R}_{\mathrm{TVAR}}^{(m)} = \left[\hat{V}^{(m)H}\right]^{-1}\left[\hat{V}^{(m)}\right]^{-1} \tag{8}$$

where $\hat{V}^{(m)}$ is a lower-triangular matrix whose elements are defined as

$$\hat{V}_{ij}^{(m)} \equiv \begin{cases} \hat{v}_{ij}^{(m)}\hat{v}_{jj}^{(m)-1/2}, & \text{for } j \leq i \leq L(j) \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where

$$\begin{bmatrix} \hat{v}_{qq}^{(m)} \\ \vdots \\ \hat{v}_{L(q),q}^{(m)} \end{bmatrix} = \begin{bmatrix} \hat{r}_{qq} & \cdots & \hat{r}_{q,L(q)} \\ \vdots & \ddots & \vdots \\ \hat{r}_{L(q),q} & \cdots & \hat{r}_{L(q),L(q)} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \equiv \left[\hat{R}_q^{(L)}\right]^{-1}\boldsymbol{e}_{L(q)-q+1} \tag{10}$$

are the time-varying AR coefficients, with $L(q) \equiv \min\{N, q+m\}$, and $\boldsymbol{e}_z$ is the z-variate unit vector.

This ML $\mathrm{TVAR}(m)$ covariance matrix is uniquely specified by the remarkable properties

$$\left\{\hat{R}_{\mathrm{TVAR}}^{(m)}\right\}_{ij} = \hat{r}_{ij}, \quad \text{for } |i-j| \leq m$$
$$\left\{\left[\hat{R}_{\mathrm{TVAR}}^{(m)}\right]^{-1}\right\}_{ij} = 0, \quad \text{for } |i-j| > m. \tag{11}$$

For a stationary $\mathrm{AR}(m)$ model, the inverse of its ML-optimal Toeplitz covariance matrix estimate is also a band matrix, like $[\hat{R}_{\mathrm{TVAR}}^{(m)}]^{-1}$, and of the same bandwidth, but its elements cannot be directly and simply obtained from $\hat{R}$; only numerical solutions are currently available for ML Toeplitz covariance matrix estimation, and so suboptimal solutions are usually suggested [4], [13], [14].

For the ML estimate $\hat{R}_{\mathrm{TVAR}}^{(m)}$, the Gaussian LF

$$\mathrm{LF}(X, R) = \frac{\exp[-\mathrm{tr}(TR^{-1}\hat{R})]}{[\det R]^T} \tag{12}$$

evaluates to

$$\max_R \mathrm{LF}(X, R) = \mathrm{LF}\left(X, \hat{R}_{\mathrm{TVAR}}^{(m)}\right) = \exp[-NT]\left(\prod_{q=1}^{N}\hat{v}_{qq}^{(m)}\right)^T \tag{13}$$

where

$$\hat{v}_{qq}^{(m)} \equiv \boldsymbol{e}_{L(q)-q+1}^T\left[\hat{R}_q^{(L)}\right]^{-1}\boldsymbol{e}_{L(q)-q+1}. \tag{14}$$

In fact, (11) follows directly from the ML equation $\partial \log \mathrm{LF}(X,R)/\partial(R^{-1})_{ij} = 0$ subject to the $\mathrm{TVAR}(m)$ constraint $(R^{-1})_{ij} = 0$ for $|i-j| > m$. According to the properties (11), for $\hat{B} \equiv [\hat{R}_{\mathrm{TVAR}}^{(m)}]^{-1}$, we get

$$\mathrm{tr}[\hat{B}\hat{R}] = \mathrm{tr}\left[\hat{B}\hat{R}_{\mathrm{TVAR}}^{(m)}\right] = NT \tag{15}$$

and by (8)

$$\det\hat{B} = \det\left[\hat{V}^{(m)H}\hat{V}^{(m)}\right] = \prod_{q=1}^{N}\hat{v}_{qq}^{(m)}. \tag{16}$$

Let $m_{\max}$ be the maximum admissible order of a $\mathrm{TVAR}(m)$ model that is identifiable for the sample volume T, then $m_{\max}+1 \leq T$. From the "nested" property of the model-order testing problem, and directly from (13), it is evident that

$$\mathrm{LF}\left(X, \hat{R}_{\mathrm{TVAR}}^{(m_1)}\right) \geq \mathrm{LF}\left(X, \hat{R}_{\mathrm{TVAR}}^{(m_2)}\right), \quad \text{for } m_1 > m_2 \tag{17}$$

and so our hypothesis test for the $\mathrm{TVAR}(m)$ order can be based on the likelihood ratio (LR)

$$\mathrm{LR}(m) = \frac{\mathrm{LF}\left[X, \hat{R}^{(m)}_{\mathrm{TVAR}}\right]}{\max\limits_{\mu \le m_{\max}} \mathrm{LF}\left[X, \hat{R}^{(\mu)}_{\mathrm{TVAR}}\right]} = \frac{\mathrm{LF}\left[X, \hat{R}^{(m)}_{\mathrm{TVAR}}\right]}{\mathrm{LF}\left[X, \hat{R}^{(m_{\max})}_{\mathrm{TVAR}}\right]} = \left(\prod_{q=1}^{N} \frac{\hat{v}^{(m)}_{qq}}{\hat{v}^{(m_{\max})}_{qq}}\right)^{T} \tag{18}$$

where, according to (10) and (14)

$$\hat{v}^{(m_{\max})}_{qq} \equiv \boldsymbol{e}^{T}_{K(q)-q+1}\left[\hat{R}^{(K)}_{q}\right]^{-1}\boldsymbol{e}_{K(q)-q+1} \tag{19}$$

with $K(q) \equiv \min\{N, q+m_{\max}\}$. Since

$$\hat{v}^{(m)}_{qq} = \hat{v}^{(m_{\max})}_{qq}, \quad \text{for } q \ge N-m \tag{20}$$

the LR is specified by

$$\mathrm{LR}(m) = \left(\prod_{q=1}^{N-m-1} \frac{\hat{v}^{(m)}_{qq}}{\hat{v}^{(m_{\max})}_{qq}}\right)^{T}. \tag{21}$$

Note that the dimension of the matrix $\hat{R}^{(K)}_{q}$ is

$$\dim \hat{R}^{(K)}_{q} = \begin{cases} m_{\max}+1, & \text{for } q \le N-m_{\max} \\ m+2, & \text{for } q = N-m-1. \end{cases} \tag{22}$$

Let us introduce the notation

$$\mu \equiv \begin{cases} m_{\max}, & \text{for } q < N-m_{\max} \\ N-q, & \text{for } N-m_{\max} \le q \le N-m-1. \end{cases} \tag{23}$$

Instead of $\mathrm{LR}(m)$, we can deal with $\mathrm{LR}_0(m) \equiv [\mathrm{LR}(m)]^{1/T}$. In what follows, we therefore investigate the properties of

$$\mathrm{LR}_0(m) = \prod_{q=1}^{N-m-1} \frac{\hat{v}^{(m)}_{qq}}{\hat{v}^{(\mu)}_{qq}}. \tag{24}$$

Theorem 1: Let m_0 be the true order of AR or TVAR input data, then for all $m \ge m_0$, the p.d.f. of $\mathrm{LR}_0(m)$ does not depend on scenario, and can be expressed as the p.d.f. of a product of $(N-m-1)$ independent random numbers β_q

$$\mathrm{LR}_0(m) = \prod_{q=1}^{N-m-1} \beta_q \tag{25}$$

with

$$\beta_q \sim \frac{\beta_q^{(T-\mu-1)}(1-\beta_q)^{(\mu-m-1)}}{B[\mu-m, T-\mu]}. \tag{26}$$

This p.d.f. is completely specified by the parameters $N, T, m_{\max}$ and m

$$f(x) = C(N,T,m_{\max},m)x^{(T-m_{\max}-1)} \times G^{(N-m-1),0}_{(N-m-1),(N-m-1)}\left(x \,\middle|\, {}^{m_{\max}-m,\ldots,m_{\max}-m}_{m_{\max}-m-1,\ldots,0,\ldots,0}\right) \tag{27}$$

where $G^{a,b}_{c,d}(\,\cdot\,)$ is Meijer's G-function [15], and

$$C(N,T,m_{\max},m) = \prod_{q=1}^{N-m-1} \frac{\Gamma(T-m)}{\Gamma(T-\mu)}. \tag{28}$$

The pth moment of $\mathrm{LR}_0(m \ge m_0)$ is

$$\mathcal{E}\{x^p\} = \prod_{q=1}^{N-m-1} \frac{\Gamma(T-m)\Gamma(T-\mu+p)}{\Gamma(T-\mu)\Gamma(T-m+p)}. \tag{29}$$

See the Appendix I for the proof.

For $N \gg m$, it is computationally preferable to deal with another monotonic transformation of the LR [see (18) and (24)]

$$\mathrm{LR}_1(m) \equiv [\mathrm{LR}_0(m)]^{1/N} = [\mathrm{LR}(m)]^{1/NT}. \tag{30}$$

The importance of Theorem 1's analytical expression (27), or statistical equivalent (25), is that it enables us to precalculate (for any given N, T and $m_{\max}$, and for each hypothesis on $m < m_{\max}$) the threshold value that corresponds to any given probability of order overestimation ("false alarm"). Our current approach is to declare the minimal m that exceeds the threshold to be the generalized likelihood-ratio test (GLRT) estimate for the AR/TVAR model order.

In effect, this LR test determines whether or not the bandwidth of the inverse covariance matrix (which is the same for the $\mathrm{AR}(m)$ and $\mathrm{TVAR}(m)$ models) is equal to m.

Another approach is to use ITC

$$\hat{m} = \arg\min_{m}\left[-\log \mathrm{LR}_1(m) + \frac{\nu_m}{NT}\right] \tag{31}$$

where the penalty term is chosen from [16]

$$\nu_m = \begin{cases} \nu_{\mathrm{AIC}} \equiv d_m, & \text{Akaike information criterion} \\ \nu_{\mathrm{MDL}} \equiv \frac{1}{2} d_m \log T, & \text{minimum description length} \\ \nu_{\mathrm{MAP}} \equiv \frac{5}{6} d_m \log T, & \text{maximum a posteriori probability} \end{cases} \tag{32}$$

where d_m is the (unknown) number of real-valued parameters that define the $\mathrm{TVAR}(m)$ model

$$d^{\mathrm{TVAR}}_{m} = N(2m+1) - m(m+1). \tag{33}$$

III. Discriminating Between Stationary and Time-Varying Models

Since a stationary $\mathrm{AR}(m)$ model is a special case of the more general $\mathrm{TVAR}(m)$ model, the ML $\mathrm{TVAR}(m)$ estimate $\hat{R}^{(m)}_{\mathrm{TVAR}}$ is a sufficient statistic for the N-variate stationary covariance matrix $T_m = R^{(m)}_{\mathrm{AR}}$. Thus, the ML estimate $\hat{T}_m$ is the one that yields a maximum in the LR $\Lambda(T_m)$ [17]

$$\hat{T}_m = \arg\max_{T_m > 0} \frac{\det\left[T_m^{-1}\hat{R}^{(m)}_{\mathrm{TVAR}}\right]\exp N}{\exp\left(\mathrm{tr}\left[T_m^{-1}\hat{R}^{(m)}_{\mathrm{TVAR}}\right]\right)} \equiv \arg\max_{T_m>0} \Lambda(T_m) \tag{34}$$

where

$$\left\{T_m^{-1}\right\}_{jk} = 0, \quad \text{for } |j-k| > m. \tag{35}$$

This maximized LR originates from the hypothesis test

$$\begin{aligned} H_0 &: \mathcal{E}\left\{\hat{R}_{\mathrm{TVAR}}^{(m)}\right\} = T_m \quad \text{against} \\ H_1 &: \mathcal{E}\left\{\hat{R}_{\mathrm{TVAR}}^{(m)}\right\} \neq T_m \end{aligned} \tag{36}$$

which is nondegenerate for $m+1 \leq T$. Since with probability one the sample $\mathrm{TVAR}(m)$ covariance matrix estimate $\hat{R}_{\mathrm{TVAR}}^{(m)}$ is never strictly Toeplitz, i.e., $\hat{R}_{\mathrm{TVAR}}^{(m)} \overset{\mathrm{Prob}}{\neq} T_m$, for any Toeplitz matrix in (34) we have

$$\Lambda(T_m) = \frac{\exp N \det\left[T_m^{-1}\hat{R}_{\mathrm{TVAR}}^{(m)}\right]}{\exp\left(\mathrm{tr}\left[T_m^{-1}\hat{R}_{\mathrm{TVAR}}^{(m)}\right]\right)} < 1. \tag{37}$$

Since $\Lambda(\hat{R}_{\mathrm{TVAR}}^{(m)}) = 1$, the LR $\Lambda(T_m)$ quantifies the degradation due to the stationarity restriction on the admissible covariance matrix set in comparison to the ML $\mathrm{TVAR}(m)$ estimate $\hat{R}_{\mathrm{TVAR}}^{(m)}$.

We wish to specify what degradations are acceptable for a truly stationary model, given a certain probability of "false alarm" when a truly stationary model is wrongly identified as nonstationary. Hence, for any given P_{FA} we have to find the threshold p such that

$$1 - \int_p^1 f[\Lambda(\hat{T}_m \,|\, T_{m_0})]\, d\Lambda = P_{\mathrm{FA}} \tag{38}$$

where $\hat{T}_m$ is the ML estimate of the stationary covariance matrix and T_{m_0} is the true covariance matrix, whereas the actual discrimination is performed by the thresholding

$$\max_{T_m} \Lambda(T_m) \underset{\hat{T}_m}{\overset{\hat{R}_{\mathrm{TVAR}}^{(m)}}{\lesseqgtr}} p. \tag{39}$$

Obviously it is difficult to specify the p.d.f. $f[\Lambda(\hat{T}_m)]$, since the analytic solution for $\hat{T}_m$ is currently unknown; however, the following straightforward observation plays a crucial role:

$$\Lambda(T_{m_0}) \leq \max_{\substack{T_m > 0 \\ m \geq m_0}} \Lambda(T_m) < 1 \tag{40}$$

where m_0 is the true order of the stationary AR input data, since the set $T_m > 0$ for $m \geq m_0$ includes the true covariance matrix T_{m_0}. Of course, T_{m_0} is unknown in practical applications, but we will shortly see that the p.d.f. $f[\Lambda(T_{m_0})]$ *does not depend on the scenario* T_{m_0}, similarly to some other LRs that have this remarkable property [18], [19]. This means that we can precalculate, for any N, T and m, the scenario-free threshold $p_0(N,T,m)$ such that

$$1 - \int_{p_0(N,T,m)}^1 f[\Lambda(T_{m_0})]\, d\Lambda = P_{\mathrm{FA}} \tag{41}$$

and use $p_0(N,T,m)$ instead of p in (39). Property (40) ensures that the actual false-alarm probability for the optimized ML estimate $\hat{T}_m$ in (38) is somewhat smaller than P_{FA} in (41), provided that $\hat{T}_m$ "is better than" the true covariance matrix in terms of LR.

Theorem 2: Let m_0 be the true order of some stationary AR input data whose Toeplitz covariance matrix is T_{m_0}, then for all $m \geq m_0$, the p.d.f. of $\Lambda(T_{m_0})$ does not depend on scenario, and can be expressed as the p.d.f. of a product of $2N$ independent random numbers α_q and Ω_q

$$\Lambda(T_{m_0}) = \exp[N] \prod_{q=1}^{N} \Omega_q \alpha_q \tag{42}$$

where

$$\alpha_q \sim \frac{\alpha_q^{(T-\nu-1)}(1-\alpha_q)^{(\nu-1)}}{B[\nu, T-\nu]}, \quad \Omega_q \equiv \frac{C_{qq}}{T}\exp[-C_{qq}/T] \tag{43}$$

$$C_{qq} \sim \frac{C_{qq}^{T-1}}{\Gamma(T)}\exp[-C_{qq}], \quad 1 \leq \nu \equiv L(q) - q \leq m. \tag{44}$$

This p.d.f. is completely specified by the parameters N, T, and m. The pth moment of $\Lambda(T_{m_0})$ is

$$\mathcal{E}\{\Lambda^p(T_{m_0})\} = \frac{T^{NT}\exp[pN]}{(T+p)^{N(T+p)}} \prod_{q=1}^{N} \frac{\Gamma(T-\nu-p)}{\Gamma(T-\nu)}. \tag{45}$$

See Appendix II for the proof.

By applying an inverse Mellin transform to this moment function, we can express the p.d.f. in a serial form, similarly to [20].

Rather than the LR $\Lambda(T_m)$ in (37), it is more appropriate computationally to deal with its Nth root

$$\Lambda_0(T_m) \equiv [\Lambda(T_m)]^{(1/N)} = \frac{\exp[1]\det^{(1/N)}\left(T_m^{-1}\hat{R}_{\mathrm{TVAR}}^{(m)}\right)}{\exp\left[\mathrm{tr}\left(T_m^{-1}\hat{R}_{\mathrm{TVAR}}^{(m)}\right)\Big/ N\right]}. \tag{46}$$

Even when $\Lambda_0(T_m)$ is close to its ultimate maximum of unity, $\Lambda(T_m)$ is an extremely small quantity, due to the power N.

Let us consider the numerator of $\Lambda(T_m)$ in (37) more carefully. Similarly to (18), we can evaluate the determinant as a product (see Appendix II)

$$\det\left[T_m^{-1}\hat{R}_{\mathrm{TVAR}}^{(m)}\right] = \prod_{q=1}^{N} \frac{w_{qq}^{\mathrm{Toep}}}{\hat{w}_{qq}^{\mathrm{TVAR}}} \tag{47}$$

where w_{qq}^{Toep} and $\hat{w}_{qq}^{\mathrm{TVAR}}$ are the diagonal elements of the triangular decomposition of the Toeplitz matrix T_m and the ML TVAR matrix $\hat{R}_{\mathrm{TVAR}}^{(m)}$, respectively. For $q \leq N-m$ [see (10)]

$$\theta_q \equiv \frac{w_{qq}^{\mathrm{Toep}}}{\hat{w}_{qq}^{\mathrm{TVAR}}} = \frac{\boldsymbol{e}_{m+1}^T T^{-1}\boldsymbol{e}_{m+1}}{\boldsymbol{e}_{m+1}^T\left[\hat{R}_q^{(L)}\right]^{-1}\boldsymbol{e}_{m+1}} \tag{48}$$

where T comprises the first $(m+1)$ rows and columns of T_m. For $q \leq N-m$, we have $w_{qq}^{\mathrm{Toep}} = v_{qq}^{(m)}$, and in fact for a truly stationary process we expect all the θ_q to be close to unity. Thus, we may introduce a test that is solely based on θ_q

$$\Lambda_1(\hat{T}_m) \equiv \frac{\exp[N-m]\prod_{q=1}^{N-m}\theta_q}{\exp\left[\sum_{q=1}^{N-m}\theta_q\right]}. \quad (49)$$

Again, we have

$$\max_{T_m} \Lambda_1(T_m) \geq \Lambda_1(T_{m_0}), \quad \text{for } m \geq m_0 \quad (50)$$

and for truly stationary $\mathrm{AR}(m_0)$ input data with true covariance matrix T_{m_0}, the p.d.f. for $\Lambda_1(T_{m_0})$ does not depend on scenario, but is fully specified by N, T, and $m \geq m_0$.

Theorem 3: Let m_0 be the true order of some stationary AR input data whose Toeplitz covariance matrix is T_{m_0}, then for all $m \geq m_0$, the p.d.f. of $\Lambda_1(T_{m_0})$ does not depend on scenario, and can be expressed as the p.d.f. of a product of $(N-m)$ independent random numbers

$$\Lambda_1(T_{m_0}) = \prod_{q=1}^{N-m} \frac{\gamma_q}{T}\exp[1-\gamma_q/T] \quad (51)$$

where

$$\gamma_q \sim \frac{\gamma_q^{T-m-1}\exp[-\gamma_q]}{\Gamma(T-m)}. \quad (52)$$

This p.d.f. is completely specified by the parameters N, T, and m. The pth moment of $\Lambda_1(T_{m_0})$ is

$$\mathcal{E}\left\{\Lambda_1^p(T_{m_0})\right\} = \left[\frac{T^{T-m}\exp[p]\Gamma(T-m+p)}{(T+p)^{T-m+p}\Gamma(T-m)}\right]^{N-m} \quad (53)$$

with

$$\mu = \left[\frac{(T-m)e}{T(1+1/T)^T}\right]^{N-m}. \quad (54)$$

See Appendices I and II for the proof.

Similarly to [20], this p.d.f. can be expressed in a serial form by applying an inverse Mellin transform.

Again, it is computationally preferable to deal with the $(N-m)$th root of the LR (49), whose p.d.f. for T_{m_0} does not depend on N

$$\Lambda_{10}(\hat{T}_m) = \left[\prod_{q=1}^{N-m}\theta_q\exp[1-\theta_q]\right]^{\frac{1}{N-m}}. \quad (55)$$

Finally, we need to specify a method for stationary $\mathrm{AR}(m)$ model identification. Whereas it is possible to numerically optimize the LR $\Lambda(T_m)$ or $\Lambda_1(T_m)$ over the set of admissible Toeplitz matrices, we choose to employ a more practical approach. We selected one routine that finds the approximate ML estimate of $\hat{T}_m$ given the sufficient statistic $\hat{R}_{\mathrm{TVAR}}^{(m)}$. The efficiency of this technique in terms of proximity to the true ML solution will then be analyzed by Monte Carlo simulations by direct comparison (40) and also by noting the actual false-alarm rates for the thresholds calculated by the scenario-free p.d.f.'s $\Lambda(T_{m_0})$ or $\Lambda_1(T_{m_0})$. If the observed false-alarm rate does not exceed the threshold probability, then our technique gives covariance matrix estimates that are statistically "as good as" the true covariance matrix (in the likelihood metric).

This $\mathrm{AR}(m)$ estimation technique was introduced in [21] and [22]. The first step is to find the standard persymmetric $(m+1)$-variate covariance matrix ML estimate [23]

$$\hat{R}_m = \sum_{q=1}^{N-m} \hat{R}_q^{(m)} + J\,\mathrm{conj}\left[\hat{R}_q^{(m)}\right]^* J \quad (56)$$

where * represents conjugation and J is the exchange matrix

$$J \equiv \begin{bmatrix} & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \end{bmatrix}. \quad (57)$$

Then, we take the $(m+1)$-variate vector

$$\boldsymbol{a} \equiv [a_0, \ldots, a_m]^T = \hat{R}_m^{-1}\boldsymbol{e}_{m+1} \quad (58)$$

and find the roots of the associated polynomial $a(z) = \sum_{k=0}^{m} a_k z^k$ to form the new polynomial

$$p(z) \equiv \sum_{k=0}^{m} p_k z^k = \exp[i\delta]\prod_{j=1}^{p}\frac{1-\mathrm{conj}\,(z_j)z}{z-z_j}a(z) \quad (59)$$

where

$$\delta = \pi p + \sum_{j=1}^{p}\arg z_j \quad (60)$$

and z_j are the roots of the polynomial $a(z)$ located inside the unit circle $|z| < 1$, taking multiplicity into account. The new $(m+1)$-variate vector $\boldsymbol{p} \equiv [p_0 > 0, \ldots, p_m]^T$ has no zeroes inside the unit circle, and so can be presented as

$$\begin{bmatrix}\boldsymbol{p} \\ 0\end{bmatrix} = \hat{T}_m^{-1}\boldsymbol{e}_N \quad (61)$$

where $\hat{T}_m$ is the $\mathrm{AR}(m)$ N-variate p.d. Toeplitz covariance matrix, given by $\boldsymbol{p}$ and the Gohberg–Semencul formula [24], [2].

IV. Simulation Results

We consider the case where the TVAR model arises due to Doppler frequency modulation (FM) over the coherent integration time (CIT) of some stationary $\mathrm{AR}(m)$ "carrier." This model is supported by the ionospheric phase-path variation phenomenology [25]. Specifically, as a stationary "carrier," we consider the AR(2) model that has already been used as a simple high-frequency (HF) over-the-horizon radar (OTHR) sea-clutter model in [26], [27]

$$y_j = -\sum_{\ell=1}^{2} a_\ell y_{j-\ell} + \sigma_0\eta_j, \quad \text{for } j = 3, \ldots, N \quad (62)$$

$$a_1 = -1.9359, \quad a_2 = 0.998, \quad \sigma_0 = 0.009675. \quad (63)$$

We simulate FM over the CIT by the diagonal matrix

$$D(k) = \mathrm{diag}\left[\exp\left(\frac{i2\pi k}{N}\left[1-\cos\frac{2\pi pj}{N\ell}\right]\right)\right], \quad \text{for } j = 1, \ldots, N \quad (64)$$

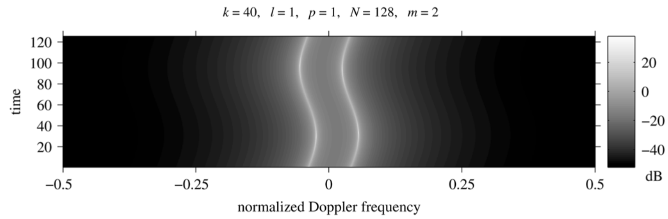

Fig. 1. Exact TVAR(2) time-frequency function for the sea-clutter model (62) for $k = 40$.

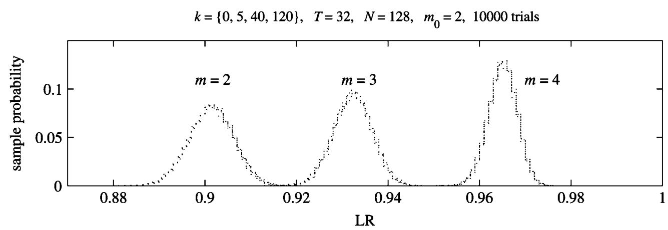

Fig. 2. Sample p.d.f. for $\mathrm{LR}_1(m)$ (30) for various m and k.

where k is the index of the periodic FM, and p/ℓ is its (relative) frequency. If $\boldsymbol{x} = D(k)\boldsymbol{y}$, then

$$R_x = D(k) R_y D^*(k). \tag{65}$$

For $k = 40, N = 128$, and $p/\ell = 1$, Fig. 1 illustrates the exact TVAR(2) time-frequency function of this process [6]

$$F_j^{(2)}(w) = \frac{\sigma_0^2}{\left|\sum_{\ell=0}^{2} a_{j\ell} \exp[-iw\ell]\right|^2}, \quad \text{for } a_{0j} = 1 \quad j = 1, \ldots, N. \tag{66}$$

We shall simulate for the sample volume $\mathrm{T} = 32$ and maximum order $m_{\max} = 5$.

First, we analyze the sample p.d.f.'s of $\mathrm{LR}_1(m)$ (30) for $N = 128$, $k = \{0, 5, 40, 120\}$, and $2 = m_0 \leq m < m_{\max} = 5$. In fact, rather than using

$$\mathrm{LR}_1(m) = \left(\det\left[\hat{R}_{\mathrm{TVAR}}^{(m)-1} \hat{R}_{\mathrm{TVAR}}^{(m_{\max})}\right]\right)^{1/N} \tag{67}$$

we avoided significant computational errors by calculating it as

$$\mathrm{LR}_1(m) = \exp\left[\frac{1}{N}\sum_{q=1}^{N} \log \frac{\hat{v}_q^{(m)}}{\hat{v}_q^{(m_{\max})}}\right]. \tag{68}$$

Fig. 2 presents sample p.d.f.'s calculated over 10^4 Monte Carlo trials. Only one set of curves is shown because the p.d.f.'s for different k are indistinguishable, as expected.

Rather than calculating the rather cumbersome Meijer G-function, we conducted 10^6 Monte Carlo trials to obtain $\mathrm{LR}_1(m)$ for $1 \leq m < m_{\max} = 5$ using Theorem 1. These theoretical representation results (see Fig. 3) agree remarkably with the direct Monte Carlo simulation results for $m_0 = 2$ (see Fig. 2). The advantage of dealing with $\mathrm{LR}_1(m)$ (30), instead of $\mathrm{LR}_0(m)$ (24) and especially $\mathrm{LR}(m)$ (18), is made more obvious by these results; otherwise, we would be dealing with tiny numbers of the order of 0.9^{128} or even $0.9^{128\times 32}$. Nevertheless, in order to prove the accuracy of our derivations, Table I lists mean values of $\mathrm{LR}_0(m)$ calculated by the

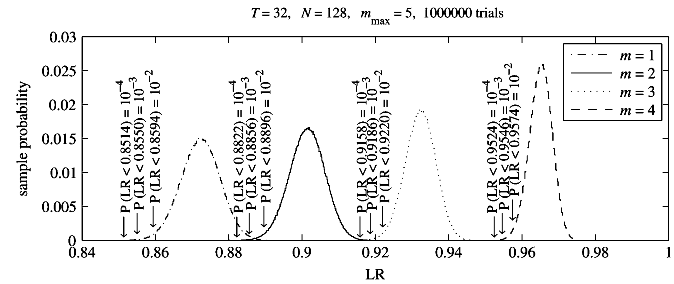

Fig. 3. Theoretical model for $\mathrm{LR}_1(m)$ (30) with corresponding "false-alarm" thresholds.

theoretical formula (29) compared with simulation results for $m_0 = 2$. For $m \geq m_0$, theoretical and experimental mean values coincide with the accuracy expected for 10^4 trials.

Apart from validating Theorem 1, Fig. 3 provides the required threshold values for our order-estimation routine: for each hypothesis regarding the order of the model $m = 1, \ldots, 4$ ($m_{\max} = 5$), we found the thresholds for a probability of order overestimation equal to 10^{-2}, 10^{-3} and 10^{-4}. Our method is to estimate the AR order by comparing the obtained $\mathrm{LR}_0(m)$ with the desired threshold and select the smallest m that exceeds the threshold. Table II (left) compares results obtained for our test scenario for this GLRT approach with the ITC methods (31), for 10^3 Monte Carlo trials. We also considered the Hannan–Quinn information criteria (HQIC) [28] with its penalty function $\nu_{\mathrm{HQ}} \equiv d_m \log\log T$. Here, all four ITC criteria have ideal order-estimation performance, as does the GLRT approach for probabilities of 10^{-3} and 10^{-4}. For a probability of 10^{-2}, however, the thresholds were exceeded in 99.2% of trials; in the other eight trials, the thresholds were not exceeded for all three admissible hypotheses. If we allow $m = 5$ to be admissible, with every trial resulting in $\mathrm{LR}_0(m_{\max}) = 1$, then the eight trials may be treated as being for order overestimation.

We next consider a more challenging scenario that highlights differences in performance between the GLRT (threshold) and ITC methods. A stationary AR(4) model has been created from a mixture of a single source in white noise in a five-sensor uniform linear antenna array. The five-variate Toeplitz covariance matrix is

$$T_5 = \sigma_0^2 I_5 + \sigma_s^2 \mathbf{1}_5 \mathbf{1}_5^T \tag{69}$$

where $\mathbf{1}_5 \equiv [1, 1, 1, 1, 1]^T$, and

$$T_5^{-1} = \sigma_0^{-2}\left[I_5 - \frac{q^2 \mathbf{1}_5 \mathbf{1}_5^T}{1 + 5q^2}\right] \tag{70}$$

with $q \equiv \sigma_s/\sigma_0$ (=40 dB) and $N = 128$. The AR(4) parameters are

$$\sigma^2 = \sigma_0^2\left[\frac{1 + 5q^2}{1 + 4q^2}\right], \quad a_0 = 1$$
$$a_j = -\frac{q^2}{1 + 4q^2}, \quad \text{for } j = 1, \ldots, 4. \tag{71}$$

By augmenting these AR parameters by $N - 5 = 123$ zeroes and applying the Gohberg–Semencul formula [2], [24], we are able to reconstruct the N-variate Toeplitz matrix whose nine main

TABLE I
COMPARISON OF THEORETIC AND SAMPLE MEANS FOR $\mathrm{LR}_1(m)$ (30)

m	1	2	3	4
theoretical	3.4107×10^{-8}	2.1239×10^{-6}	1.4708×10^{-4}	1.1441×10^{-2}
simulated	0	2.1275×10^{-6}	1.4745×10^{-4}	1.1436×10^{-2}

TABLE II
ORDER ESTIMATION RESULTS FOR THE SEA-CLUTTER MODEL (62) FOR 1000 TRIALS WITH $m_0 = 2$ (LEFT), AND THE SINGLE-SOURCE MODEL (69) FOR 10 000 TRIALS WITH $m_0 = 4$ (CENTER) AND $m_0 = 3$ (RIGHT)

m	1	2	3	4	1	2	3	4	1	2	3	4
AIC	0	1000	0	0	0	0	0	10000	0	0	10000	0
MDL	0	1000	0	0	0	0	9837	163	0	0	10000	0
MAP	0	1000	0	0	0	10000	0	0	0	10000	0	0
HQIC	0	1000	0	0	0	0	43	9957	0	0	10000	0
$P(10^{-4})$	0	1000	1000	1000	0	0	0	9999	0	0	10000	10000
$P(10^{-3})$	0	1000	1000	998	0	0	0	9987	0	0	9986	9996
$P(10^{-2})$	0	992	992	992	0	0	0	9899	0	0	9907	9907

diagonal bands are specified by the elements of T_5 and the band inverse

$$\{T_N^{-1}\}_{jk} = 0, \quad \text{for } |j-k| > 4. \tag{72}$$

By the Dym-Gohberg decomposition (8), $T_N = (V_N^H)^{-1}(V_N)^{-1}$, where V_N is a lower-triangular matrix, we simulated the AR(4) input data as

$$Y_N = V_N^H X_N, \quad \text{for } X_N \sim \mathcal{CN}(0, I_N). \tag{73}$$

This example illustrates the greatest distinction between the AR order ($m_0 = 4$) and signal-subspace dimension ($\mu = 1$) for the same covariance matrix T_5.

Table II (center) displays the results of Monte Carlo simulations over 10^4 trials. Here, the ITC techniques have different performance, unlike for the previous scenario where each of them unmistakably estimated the true order. While the Akaike criterion has a tendency to overestimate, here it has perfect performance. On the contrary, the MDL criterion here underestimates badly, while MAP always incorrectly identifies the order as $\hat{m} = 2$. Interestingly, in no trial was the AR model estimate unity as might be expected for an $\mathrm{AR}(m)$ model generated by a single (rank one) source in noise. In no trial has the GLRT method underestimated the AR(4) order. The GLRT results for the number of trials that did not exceed that theoretical threshold show that these thresholds are accurate and do not depend on scenario. As in the previous scenario, if we decide $m_{\max} = 5$ is admissible, then all those trials would be treated as order-overestimation cases.

We can see that the GLRT method is significantly better than the traditional ITC approach in this case. We also similarly processed the case $m_0 = 3$ (see right-hand side of Table II). For a severe sample support shortage ($T \ll N$), we see that the performance of the ITCs do not correspond to their asymptotic properties ($T \to \infty$). We could consider different scenarios with, say, $|a_4| \ll |a_j|$, for $j = 1, 2, 3$, where a smaller probability of overestimation should lead to a larger (scenario-dependent) probability of order underestimation. However, the scenario-independent probability of overestimation or no estimation remains a distinct advantage of our method.

The demonstrated efficiency (and its relative computational simplicity) of the GLRT method means that it may be considered for purely stationary scenarios as an alternative algorithm for other well-known techniques [2], [28].

By the property of the covariance matrix R_x in (65), the order estimation results are identical to the previous results for any k in this specific TVAR model. Thus, the AR model order estimation problem has a simple and reasonably accurate solution that includes both stationary and time-varying models, and does not require model identification. The number of i.i.d. training samples that just exceeds the maximum tested order is sufficient, regardless of the problem dimension N.

Now, we turn our attention to simulation results on discriminating between stationary $\mathrm{AR}(m)$ and time-varying $\mathrm{TVAR}(m)$ autoregressive models.

First, Fig. 4(a) shows the results of 40 000 Monte Carlo trials when the truly stationary AR(2) data (46) was used to calculate the sample LR p.d.f.'s $\Lambda_0(T_{m_0})$ [see (62)] ("lower bound") and $\Lambda_0(\hat{T}_m)$ [see (56)–(61)]. Fig. 4(b) has the corresponding results for $\Lambda_{10}(T_{m_0})$ and $\Lambda_{10}(\hat{T}_m)$ (55). We also show the threshold values for $P_{\mathrm{FA}} = 10^{-2}, 10^{-3}, 10^{-4}$ that are calculated from the lower bound distribution. We see that our technique for stationary $\mathrm{AR}(m)$ model identification (56)–(61) generates solutions that are statistically indistinguishable from the true covariance matrix in terms of LR. In this case, 54% of trials gave $\Lambda(\hat{T}_m) > \Lambda(T_{m_0})$. Of course, similarly to [29], [30], we could have used the results of the routine (56)–(61) to initialize a further direct numerical LR optimization so as to obtain $\Lambda(\hat{T}_m) \geq \Lambda(T_{m_0})$ in every trial. In practical applications, where T_{m_0} is unknown, we can assess the "quality" of any estimate by comparing its LR with the lower-bound p.d.f. Statistically though, the lower bound and the LR p.d.f.'s are indistinguishable, and the thresholds for the lower bound $\Lambda(T_{m_0})$ are even slightly smaller than for the estimate $\hat{T}_m$. Note that other existing tech-

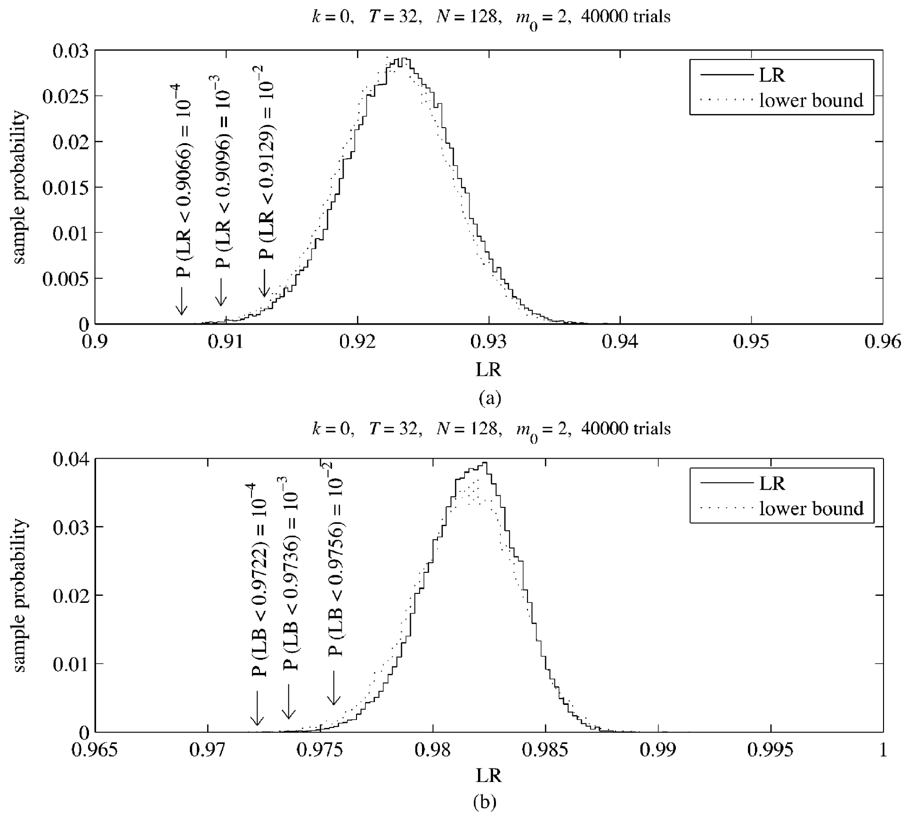

Fig. 4. Monte Carlo results for (a) $\Lambda_0(T_m)$ (46) and (b) $\Lambda_{10}(\hat{T}_m)$ (55) for both the true covariance matrix T_{m_0} ("lower bound") and the estimate $\hat{T}_m$ (61) ("LR").

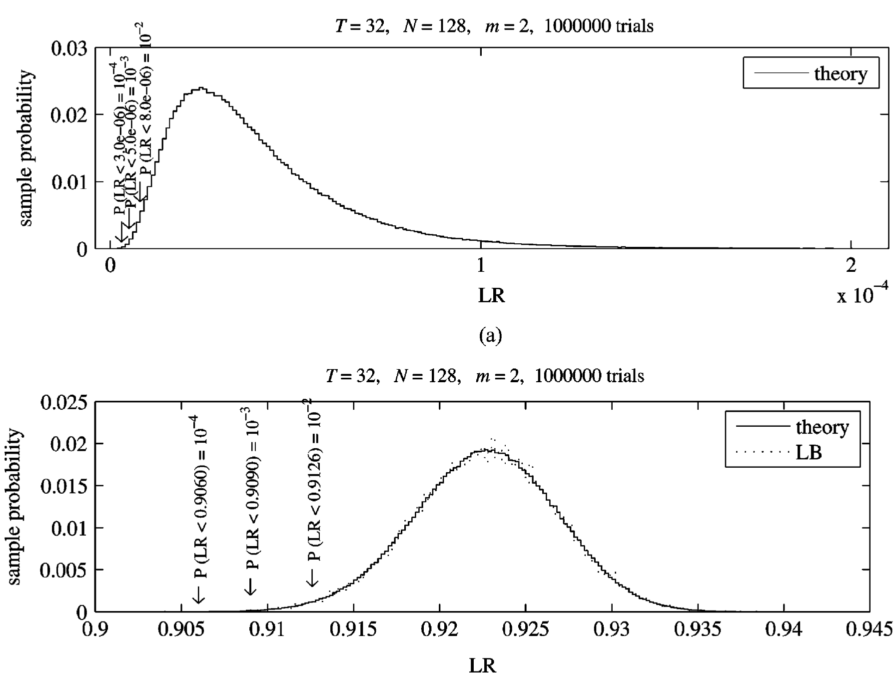

Fig. 5. Theoretical model (42) for (a) $\Lambda(T_{m_0})$ (37) and (b) $\Lambda_0(T_{m_0})$ (46) compared with the Monte Carlo results from Fig. 4(a).

niques could be tested in the same way, and could well demonstrate similar proximity to the ML-optimal solution.

In order to prove the validity of Theorem 2, we generated the lower-bound p.d.f.'s for the LRs $\Lambda(T_{m_0})$ (37) and $\Lambda_0(T_{m_0})$ (46) using the scenario-free representation (42). Fig. 5 shows the results for one million Monte Carlo trials. Fig. 5(b) also repeats the sample p.d.f. from Fig. 4(a). While there are now 10^6 trials instead of 4×10^4 trials, the theoretical and observed p.d.f.'s coincide. Moreover, the threshold values in Fig. 5(b) that are calculated using the scenario-free representation (42)

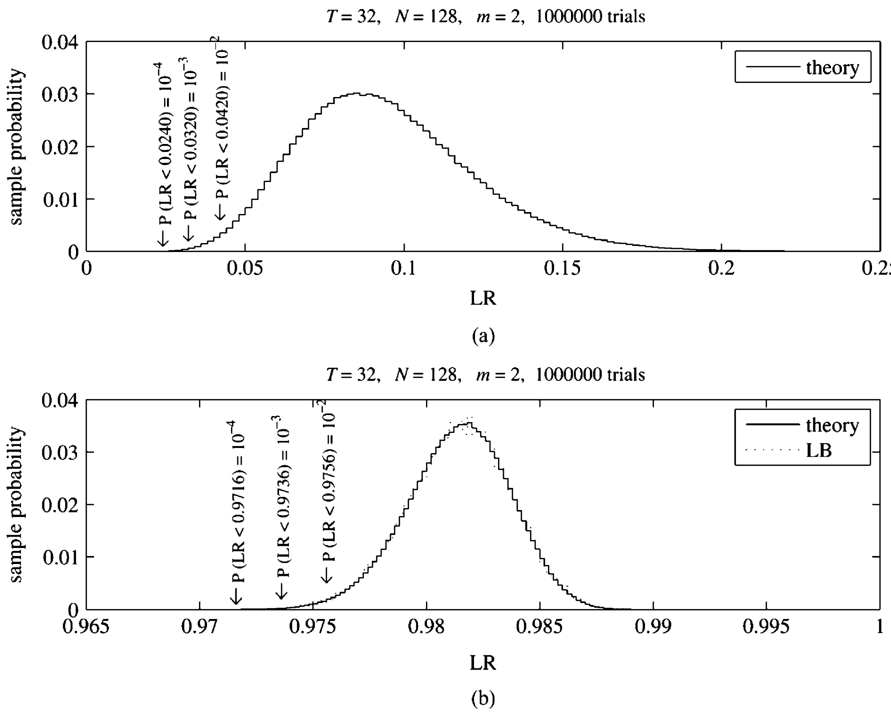

Fig. 6. Theoretical model (51) for (a) $\Lambda_1(T_{m_0})$ (49) and (b) $\Lambda_{10}(T_{m_0})$ (55) compared with the Monte Carlo results from Fig. 4(b).

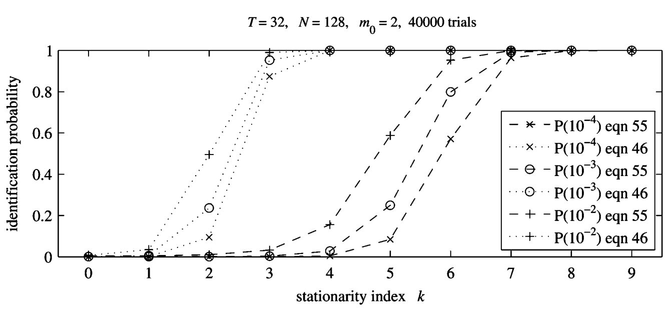

Fig. 7. Sample probabilities of correctly identifying AR versus TVAR for the LR tests $\Lambda_0(\hat{T}_m)$ (46) and $\Lambda_{10}(\hat{T}_m)$ (55).

are virtually the same as in Fig. 4(a). This proves our assertion that scenario-independent thresholds can be precalculated for any given $P_{\rm FA}$, i.e., when a truly stationary process is wrongly identified as a time-varying one.

Fig. 6 in the same format demonstrates the accurate correspondence between the theoretical scenario-free representation of the LR $\Lambda_{10}(T_{m_0})$ [the $(N-m)$th root of (51)] and the direct $\Lambda_1(T_{m_0})$ calculations of (55) in Fig. 4(b).

In Fig. 7, the more accurate million-trial theoretical thresholds have been used to analyze probabilities of correct discrimination, and are illustrated for the tests $\Lambda_0(\hat{T}_m)$ and $\Lambda_{10}(\hat{T}_m)$ as a function of the FM nonstationarity parameter k. We see that the "ML" test $\Lambda_0(T_m)$ (46) has extremely high sensitivity. For example, even at $P_{\rm FA}=10^{-4}$, the probability of correct TVAR identification is as high as 87% for the minimal nonstationarity $k=3$.

Fig. 9 shows a sample time-frequency function for $k=4$, which is visually almost indistinguishable from the introduced

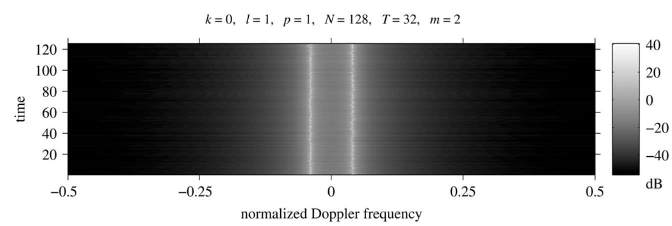

Fig. 8. Sample TVAR(2) time-frequency function for $k=0$.

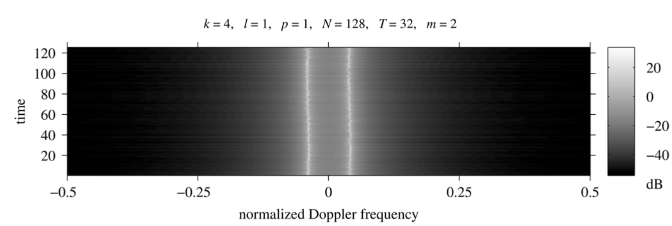

Fig. 9. Sample TVAR(2) time-frequency function for $k=4$.

stationary case $(k=0)$ as shown in Fig. 8. Our simplified LR test $\Lambda_1(\hat{T}_m)$ or $\Lambda_{10}(\hat{T}_m)$ demonstrates a certain loss in sensitivity with respect to the test $\Lambda_0(\hat{T}_m)$ (46). Indeed, this test only responds to a degradation in clutter mitigation efficiency due to the "stationary" filter $\boldsymbol{p}=\hat{T}^{-1}\boldsymbol{e}_{m+1}$ compared with the TVAR-based filter, whereas the ML test (46) also accounts for more subtle differences in the covariance matrix estimate. Yet, even the simplified test shows reliable discrimination for the small FM parameter $k\leq 7$.

These discrimination simulations have been performed for the correct model order $m=m_0=2$, which is consistent with the extremely high probability of correct order estimation for this "sea-clutter" scenario (see left-side of Table II). Still, in order to ensure sufficient robustness with respect to order overestimation, we performed the same simulations for the misspecified order $m = 3$. The results (not shown here) are negligibly degraded in terms of the probability of correct discrimination.

V. Conclusion

We have proposed GLRT-based tests for order estimation and discrimination between stationary and time-varying autoregressive models. These tests stem from the ML $\mathrm{TVAR}(m)$ estimation technique developed in [6], that itself was based on Dym–Gohberg band matrix extensions [9].

In our two-stage approach, we first estimate the order of the AR model, regardless of its stationary or time-varying nature, by testing the bandwidth of the inverse covariance matrix. This test requires the i.i.d. sample volume T to exceed the maximum possible order $m_{\max}$. In most practical cases, both these quantities are significantly smaller than the data dimension N. For OTHR applications, for example, the coherent integration time typically encompasses 256 or 512 repetition periods, with only a few available training ranges containing homogeneous clutter. Our test has an important invariance property: for any tested order m that is not less than the true order, the LR p.d.f. is scenario-independent, since it is only a function of N, T, m and $m_{\max}$. This means that we can precalculate the threshold for each hypothesis involving the order m, given any desired probability of order overestimation. Our order estimate $\hat{m}$ is the minimal order that exceeds this threshold. Since this threshold is precalculated, our technique is no more demanding than the traditional ITC methods. Whereas the ITC methods can be applied to the introduced LR, they all have inferior performance compared with GLRT thresholding.

Having obtained $\hat{m}$, the second stage of our approach is to test the appropriateness of the ML stationary $\mathrm{AR}(m)$ model. We used the test developed in [6] that exploits the analytic ML estimate of the time-varying covariance matrix as a sufficient statistic for any hypothesis test involving the structure of the true covariance matrix. In particular, we test here the hypothesis that a certain Toeplitz matrix with a banded inverse is the mean of the ML $\mathrm{TVAR}(\hat{m})$ covariance matrix estimate. This Toeplitz matrix must be the ML estimate, but since the analytic derivation of it is unknown, we used one technique (among many other possibilities) that is expected to give solutions that are reasonably close to the ML estimate.

Fortunately, we found that the LR p.d.f. of this introduced test for the true covariance matrix does not depend on this matrix (is "scenario-free"). This invariance, together with the observation that the maximized LR must exceed the LR for the true covariance matrix T_{m_0}, has been used in two ways. First, during each Monte Carlo trial with its known stationary covariance matrix, we compared the LR of the $\mathrm{AR}(m)$ estimate $\Lambda(\hat{T}_m)$ with $\Lambda(T_{m_0})$. This assesses the quality of our suboptimal estimation in terms of proximity to the ML optimal solution. We found that all trials generated an LR that was statistically indistinguishable from $\Lambda(T_{m_0})$, hence, we treat our $\mathrm{AR}(m)$ estimation routine as being statistically equivalent to the (unknown) ML optimal estimation.

Second, and most importantly, this invariance property again lets us precalculate thresholds for any desired probability of misidentifying a stationary model as a time-varying one. The scenario-free p.d.f. for $\Lambda(T_{m_0})$ is again fully specified by N, T, and m (provided $m \geq m_0$). Note that the minimum sample volume required to discriminate between $\mathrm{AR}(m)$ and $\mathrm{TVAR}(m)$ is only $T = m+1$, unlike the traditional hypothesis testing for a covariance matrix [17] that needs $T \geq N$ samples. We also proposed a slightly simplified test that has the same invariance properties as the ML one, but with somewhat inferior performance.

All scenario-free LR p.d.f.'s for order estimation and $\mathrm{AR}(m)$ versus $\mathrm{TVAR}(m)$ tests have been derived. Analytic expressions have been given for them and their moments, and the inverse Mellin transform in one instance led to an explicit expression for the p.d.f. in terms of Meijer's G-function. For the two other p.d.f.'s, we described how they can be expressed in terms of a convergent series, similarly to [20]. We then presented them as products of independent random variables with standard (F, β, and χ^2) distributions. Thus, we were able to calculate the p.d.f.'s (and their threshold values) using scenario-free Monte Carlo simulations.

Two different AR models have been simulated and tested. The first originates from a simple HF OTHR sea-clutter model that uses a stationary AR(2) process. We added a periodic FM that emulates ionospheric phase-path variations during the coherent integration time to obtain a TVAR(2) process. For this model, we demonstrated that both the existing ITC method and the new GLRT thresholding method give reliable order estimation, regardless of the model stationarity. The new test that discriminates between stationary and time-varying AR models demonstrated remarkable sensitivity, and makes it possible to detect a quite insignificant FM. We also checked the robustness of our discrimination test with respect to AR order overestimation.

Our second AR model was a more challenging scenario: it originates from a single plane-wave source (tone) mixed in noise, with an $(m+1)$-sensor uniform linear antenna array. While the GLRT approach again gave high order-estimation accuracy, the ITC methods demonstrated scenario-dependent performance. More specifically, MAP and MDL grossly underestimated the AR order $m_0 = 4$, while AIC (which is known for its overestimation properties) in this case gave correct estimates; for $m_0 = 3$, only MAP failed.

To summarize, both analytic and simulation results have demonstrated that the problem of $\mathrm{AR}(m)$ or $\mathrm{TVAR}(m)$ order estimation, and stationary versus time-varying model discrimination under limited training sample support now has an adequate and efficient solution.

Appendix I
Proof of Theorem 1

Let us simplify the notation in (24)

$$\mathrm{LR}_0(m) = \prod_{q=1}^{N-m-1} \frac{\hat{v}_{qq}^{(m)}}{\hat{v}_{qq}^{(\mu)}} \equiv \prod_{q=1}^{N-m-1} \frac{\alpha_q}{\gamma_q}. \qquad (74)$$

We first investigate the sequence

$$\alpha_1 \equiv \left[S_{11.2}^{(1)}\right]^{-1}, \ldots, \alpha_{N-m-1} \equiv \left[S_{11.2}^{(N-m-1)}\right]^{-1} \qquad (75)$$

where

$$S_{11.2}^{(q)} \equiv \left[\boldsymbol{e}_{m+1}^T \left(\hat{R}_q^{(m)}\right)^{-1} \boldsymbol{e}_{m+1}\right]^{-1}. \tag{76}$$

To prove that $S_{11.2}^{(q)}$ $(q = 1, \ldots, N-m-1)$ is a sequence of independent variables provided that $m \geq m_0$, consider the normalized variables

$$\psi_1 \equiv \Sigma_{11.2}^{(1)} \left[S_{11.2}^{(1)}\right]^{-1}, \ldots, \psi_{N-m-1} \equiv \Sigma_{11.2}^{(N-m-1)} \left[S_{11.2}^{(N-m-1)}\right]^{-1}. \tag{77}$$

It is clear that the statistical independence of ψ_q leads to the statistical independence of $[S_{11.2}^q]^{-1}$.

First, we investigate the properties of the matrix $\hat{C}_N^{(m)}$, where

$$\hat{C}_N^{(m)} \equiv V_N^{(m_0)H} \hat{R}_N^{(m)} V_N^{(m_0)}$$
$$\left[R_N^{(m)}\right]^{-1} \equiv V_N^{(m_0)} V_N^{(m_0)H} \tag{78}$$

which is the Dym–Gohberg factorization of the true $\text{TVAR}(m_0)$ or $\text{AR}(m_0)$ covariance matrix of the input data, with

$$\left\{V_N^{(m_0)}\right\}_{ij} \begin{cases} \neq 0, & \text{for } j \leq i \leq L(j) \\ = 0, & \text{otherwise} \end{cases} \tag{79}$$

[see (9) and [2], [9]]. Therefore, $V_N^{(m_0)H}$ is an upper triangular matrix with bandwidth (m_0+1), and means that

$$\text{diag}\left[V_N^{(m_0)H} \hat{R}_N^{(m)} V_N^{(m_0)}\right] = \text{diag}\left[V_N^{(m_0)H} \hat{R} V_N^{(m_0)}\right] \tag{80}$$

if $m \geq m_0$. Indeed

$$\left\{\hat{C}_N^{(m)}\right\}_{qq} = \boldsymbol{e}_N^{(q)T} V_N^{(m_0)H} \hat{R}_N^{(m)} V_N^{(m_0)} \boldsymbol{e}_N^{(q)} = \left\{V_N^{(m_0)H} \hat{R}_N^{(m)} V_N^{(m_0)}\right\}_{qq} \tag{81}$$

where $\boldsymbol{e}_N^{(q)}$ is the N-variate column vector with a single unit element in the qth position. Note that the qth column of the N-variate matrix $V_N^{(m_0)}$, specified by $V_N^{(m_0)} \boldsymbol{e}_N^{(q)}$, has only $L(q) - q + 1 \leq m_0 + 1$ non-zero elements, and so

$$\left\{\hat{C}_N^{(m)}\right\}_{qq} = \boldsymbol{e}_{L(q)-q+1}^T V_q^{(m_0)H} \hat{R}_q^{(L)} V_q^{(m_0)} \boldsymbol{e}_{L(q)-q+1} \equiv \boldsymbol{e}_{L(q)-q+1}^T \hat{C}_q \boldsymbol{e}_{L(q)-q+1} \tag{82}$$

where $\hat{R}_q^{(L)}$ is specified in (10) and $V_q^{(m_0)}$ is the $(L(q) - q + 1)$-variate lower-triangular matrix in the Dym–Gohberg factorization of the matrix $R_q^{(L)}$

$$\left[R_q^{(L)}\right]^{-1} = V_q^{(m_0)} V_q^{(m_0)H}. \tag{83}$$

Thus, (m_0+1) non-zero elements of the $(m+1)$-variate vector $V_q^{(m_0)} \boldsymbol{e}_{L(q)-q+1}$ are exactly the same as the (m_0+1) non-zero elements of the vector $V_N^{(m_0)} \boldsymbol{e}_N^{(q)}$, and

$$V_q^{(m_0)} \boldsymbol{e}_{L(q)-q+1} \equiv \left[\boldsymbol{e}_{L(q)-q+1}^{(q)T} V_q^{(m_0)} \boldsymbol{e}_{L(q)-q+1}\right] \left[R_q^{(L)}\right]^{-1} \boldsymbol{e}_{L(q)-q+1}. \tag{84}$$

From the latter, we conclude that

$$\hat{C}_q = V_q^{(m_0)H} \hat{R}_q^{(L)} V_q^{(m_0)} \sim \mathcal{CW}(T, L(q) - q + 1, I_{L(q)-q+1}) \tag{85}$$

hence, for any $m \geq m_0$, the diagonal elements of the original N-variate prewhitened sample matrix $V_q^{(m_0)H} \hat{R}_q^{(L)} V_q^{(m_0)}$ are identical to those of the matrix $\hat{C}_N^{(m)}$, which is the prewhitened Dym–Gohberg matrix $\hat{R}_N^{(m)} = \text{DG}_m(\hat{R})$, and most importantly, equal to the upper-left corner of the (small) $(L(q) - q + 1)$-variate matrix $\hat{C}_q$, i.e.,

$$\hat{C}_q = \begin{bmatrix} \hat{C}_{qq} & \hat{C}_{12}^{(q)} \\ \hat{C}_{21}^{(q)} & \hat{C}_{22}^{(q)} \end{bmatrix} \tag{86}$$

where

$$\hat{C}_{qq} = \left\{V^{(m_0)H} \hat{R}_N^{(m)} V^{(m_0)}\right\}_{qq} = \left\{V^{(m_0)H} \hat{R} V^{(m_0)}\right\}_{qq}. \tag{87}$$

The latter means that $\hat{C}_{qq}$ $(q = 1, \ldots, N)$ are the diagonal elements of an N-variate matrix with the complex anti-Wishart distribution $\mathcal{ACW}(T < N, N, I_N)$ for $T < N$, hence, are mutually independent, and have a chi-squared p.d.f. [31]

$$\hat{C}_{qq} \sim \frac{1}{\Gamma(T)} \hat{C}_{qq}^{T-1} \exp(-\hat{C}_{qq}). \tag{88}$$

Naturally, the same p.d.f. follows from (84) and (85), but the *mutual independence* of the $\hat{C}_{qq}$ can be proven only for $\text{TVAR}(m \geq m_0)$ models.

Now, let us consider the qth member of the sequence (77)

$$\psi_q \equiv \Sigma_{11.2}^{(q)} \left[S_{11.2}^{(q)}\right]^{-1} = \frac{\boldsymbol{e}_{m+1}^T \left[\hat{R}_q^{(m)}\right]^{-1} \boldsymbol{e}_{m+1}}{\boldsymbol{e}_{m+1}^T \left[R_q^{(m_0)}\right]^{-1} \boldsymbol{e}_{m+1}}, \quad \text{for } q = 1, \ldots, N-m-1 \tag{89}$$

where

$$\hat{R}_q^{(m)} \sim \mathcal{CW}(T, m+1, R_q^{(m_0)})$$
$$T > m+1$$
$$m \geq m_0. \tag{90}$$

From (84) and (85), we have

$$\psi_q = \frac{\boldsymbol{e}_{m+1}^T V_q^{(m_0)} \hat{C}_q^{-1} V_q^{(m_0)H} \boldsymbol{e}_{m+1}}{\boldsymbol{e}_{m+1}^T V_q^{(m_0)} V_q^{(m_0)H} \boldsymbol{e}_{m+1}} = \boldsymbol{e}_{m+1}^T \hat{C}_q^{-1} \boldsymbol{e}_{m+1} \tag{91}$$

and so by (85)

$$\psi_q = \left(\hat{C}_{qq} - \hat{C}_{12}^{(q)} \left[\hat{C}_{22}^{(q)}\right]^{-1} \hat{C}_{21}^{(q)}\right)^{-1} = \left[\hat{C}_{11.2}^{(q)}\right]^{-1} \tag{92}$$

then according to [32, Th. 3.3.9, iii], $\hat{C}_{11.2}^{(q)}$ depends only on $\hat{C}_{qq}$, and *does not depend* on $\hat{C}_{12}^{(q)}$ nor $\hat{C}_{22}^{(q)}$. Since the $\hat{C}_{qq}$ $(q = 1, \ldots, N)$ are independent, we conclude that the ψ_q $(q = 1, \ldots, N-m-1)$ are also mutually independent. Since ψ_q is just a deterministically normalized sequence α_q, we conclude that for a $\mathrm{TVAR}(m_0)$ (and $\mathrm{AR}(m_0)$) model for any $m \geq m_0$, the sequence α_q in (75) is a sequence with independent random variables.

Similarly, we can establish that the sequence $\gamma_1, \gamma_2, \ldots, \gamma_{N-m-1}$ is also independent, as well as the α_k and γ_ℓ for $k \neq \ell$. The only dependent entries in (74) are α_q and γ_q with the same index q. Therefore, we have to specify the p.d.f. of the ratio

$$\frac{\alpha_q}{\gamma_q} = \frac{\boldsymbol{e}^T \left[\hat{R}_q^{(m)}\right]^{-1} \boldsymbol{e}}{\boldsymbol{e}^T \left[\hat{R}_q^{(\mu)}\right]^{-1} \boldsymbol{e}}, \quad \text{for } m_{\max} \geq \mu \geq m+1. \tag{93}$$

Here, $\hat{R}_q^{(m)}$ is a $(m+1)$-variate Hermitian matrix, and $\hat{R}_q^{(\mu)}$ is the $(\mu+1)$-variate Hermitian matrix that can be partitioned as

$$\hat{R}_q^{(\mu)} = \left[\begin{array}{c|c} \hat{R}_q^{(m)} & \hat{R}_q^{12} \\ \hline \hat{R}_q^{21} & \hat{R}_q^{22} \end{array}\right] \begin{array}{c} (m+1) \\ (\mu-m) \end{array} \quad \begin{array}{cc} (m+1) & (\mu-m). \end{array} \tag{94}$$

Note that instead of $\hat{R}_q^{(m)}$, we can analyze $T\hat{R}_q^{(m)}$ which for $T > m_{\max}+1$ is described by the complex Wishart distributions

$$\begin{aligned} T\hat{R}_q^{(\mu)} &\sim \mathcal{CW}\left(T, \mu+1, R_q^{(\mu)}\right) \\ T\hat{R}_q^{(m)} &\sim \mathcal{CW}\left(T, m+1, R_q^{(m)}\right). \end{aligned} \tag{95}$$

When $m \geq m_0$, $R_q^{(\mu)}$ has the $(2m_0+1)$-wide band inverse [6], i.e.,

$$\left[R_q^{(\mu)}\right]_{k\ell}^{-1} = 0, \quad \text{for } |k-\ell| > m_0. \tag{96}$$

Moreover, according to the Dym–Gohberg formula [9]

$$\left[R_q^{(\mu)}\right]^{-1} = V_q^{(\mu)H} V_q^{(\mu)} \tag{97}$$

where the $(m+1)$-variate lower triangular matrix, $V_q^{(\mu)}$ is a (m_0+1)-wide band matrix

$$\left[V_q^{(\mu)}\right]_{k\ell} = 0, \quad \text{for } k-\ell > m_0 \quad (\text{and } \ell > k). \tag{98}$$

Thus, the last $(\mu - m_0)$ elements in the $(\mu+1)$-variate vector $a_q^{(\mu)} \equiv V_q^{(\mu)} \boldsymbol{e}_{m+1}$ are equal to zero, and more importantly, for any $m_0 \leq m < \mu$, this vector can be presented as

$$a_q^{(\mu)} = \left[\begin{array}{c} a_q^{(m)} \\ \hline 0 \end{array}\right] \begin{array}{c} (m+1) \\ (\mu-m). \end{array} \tag{99}$$

Naturally, property (99) holds both for stationary $\mathrm{AR}(m)$ and time-varying $\mathrm{TVAR}(m)$ models.

Now, let us introduce an $(\mu+1)$-variate unitary matrix $U_q^{(\mu)}$ in a partitioned form

$$U_q^{(\mu)} = \left[\begin{array}{c|c} U_q^{(m)} & 0 \\ \hline 0 & I_{\mu-m} \end{array}\right] \begin{array}{c} (m+1) \\ (\mu-m) \end{array} \quad \begin{array}{cc} (m+1) & (\mu-m) \end{array} \tag{100}$$

Here, $U_q^{(\mu)}$ is a $(m+1)$-variate unitary matrix that is specified by the condition

$$U_q^{(m)} \boldsymbol{e}_{m+1} = \frac{a_q^{(m)}}{\sqrt{a_q^{(m)H} a_q^{(m)}}}. \tag{101}$$

At the same time

$$U_q^{(\mu)} \boldsymbol{e}_{\mu+1} = \frac{a_q^{(\mu)}}{\sqrt{a_q^{(\mu)H} a_q^{(\mu)}}} \tag{102}$$

since due to (99)

$$a_q^{(\mu)H} a_q^{(\mu)} = a_q^{(m)H} a_q^{(m)} \tag{103}$$

with respect to (100)–(102), we get

$$\frac{\alpha_q}{\gamma_q} = \frac{\boldsymbol{e}_{m+1}^T \left[\hat{C}_q^{(m)}\right]^{-1} \boldsymbol{e}_{m+1}}{\boldsymbol{e}_{\mu+1}^T \left[\hat{C}_q^{(\mu)}\right]^{-1} \boldsymbol{e}_{\mu+1}} \tag{104}$$

since $\hat{R}_q^{(\mu)} \sim [V_q^{(\mu)}]^H \hat{C}_q^{(\mu)} [V_q^{(\mu)}]$, where

$$\begin{aligned} \hat{C}_q^{(\mu)} &\equiv \left[\begin{array}{c|c} \hat{C}_q^{(m)} & \hat{C}_q^{12} \\ \hline \hat{C}_q^{21} & \hat{C}_q^{22} \end{array}\right] \\ \left[\hat{C}_q^{(\mu)}\right]^{-1} &\equiv \left[\begin{array}{c|c} \hat{C}_q^{11} & \hat{C}_{12}^{q} \\ \hline \hat{C}_{21}^{q} & \hat{C}_{22}^{q} \end{array}\right] \end{aligned} \tag{105}$$

and

$$\begin{aligned} \hat{C}_q^{(\mu)} &\sim \mathcal{CW}(T, \mu+1, I_{\mu+1}) \\ \hat{C}_q^{(m)} &\sim \mathcal{CW}(T, m+1, I_{m+1}). \end{aligned} \tag{106}$$

[Note that (104) and (105) also follow from (85).] Since [33]

$$\begin{aligned} \hat{C}_q^{11} = \left[\hat{C}_q^{(m)}\right]^{-1} + \left[\hat{C}_q^{(m)}\right]^{-1} \hat{C}_q^{12} \\ \times \left[\hat{C}_q^{22} - \hat{C}_q^{21} \left(\hat{C}_q^{(m)}\right)^{-1} \hat{C}_q^{12}\right]^{-1} \hat{C}_q^{21} \left[\hat{C}_q^{(m)}\right]^{-1} \end{aligned} \tag{107}$$

we get

$$\frac{\alpha_q}{\gamma_q} = \frac{1}{1+\nu_q} \tag{108}$$

where ν_q is as shown in (109) and (110) at the bottom of the page. According to [32, Th. 3.3.9, ii], $\hat{W}_q^{(22)}$ and $\{\hat{C}_q^{12}, \hat{C}_q^{(\mu)}\}$ are always independent. Moreover, since $\mathcal{E}\{\hat{C}_q^{(\mu)}\} = I_{\mu+1}$, then by the theorem, we have

$$\hat{C}_q^{21}\left[\hat{C}_q^{(m)}\right]^{-\frac{1}{2}}\Big|\hat{C}_q^{(m)} \sim \mathcal{CN}_{(\mu-m);m+1}(0, I_{m+1}) \quad (111)$$

which means that for the matrix $\hat{C}_q^{\mu}$, the vector $\hat{C}_q^{21}[\hat{C}_q^{(m)}]^{-\frac{1}{2}}$ and matrices $\hat{C}_q^{(m)}$ and $\hat{W}_q^{(22)}$ are all mutually independent. Let us, therefore, find

$$\nu_q(b) = \frac{\boldsymbol{b}^H\left[X^H\left(\hat{W}_q^{(22)}\right)^{-1}X\right]\boldsymbol{b}}{\boldsymbol{b}^H\boldsymbol{b}} \quad (112)$$

where

$$\boldsymbol{b} = \left[\hat{C}_q^{(m)}\right]^{-\frac{1}{2}}\boldsymbol{e}_{m+1},$$
$$X = \hat{C}_q^{21}\left[\hat{C}_q^{(m)}\right]^{-\frac{1}{2}} \sim \mathcal{CN}_{(\mu-m);m+1}(0, I_{m+1}) \quad (113)$$

with X and $\hat{W}_q^{(22)}$ being mutually independent. By another unitary $(m+1)$-variate transform

$$L_q^{(m)}\boldsymbol{e}_{m+1} = \frac{\boldsymbol{b}}{\sqrt{\boldsymbol{b}^H\boldsymbol{b}}} \quad (114)$$

we now convert $\nu_q(\boldsymbol{b})$ into

$$\nu_q(\boldsymbol{b}) = \nu_q = \boldsymbol{x}^H\left(\hat{W}_q^{(22)}\right)^{-1}\boldsymbol{x} \quad (115)$$

which means that despite $\boldsymbol{b}$ being a random vector, its independence from X and $W_q^{(22)}$ makes ν_q independent of its statistics. According to (111), $\boldsymbol{x} \sim \mathcal{CN}(0, I_{\mu-m})$, and

$$\hat{W}_q^{(22)} \sim \mathcal{CN}(T-m-1, \mu-m, I_{\mu-m}) \quad (116)$$

according to (ii).

The p.d.f. for the Hermitian form ν_q is described by the well-known F-distribution [34], [35]

$$f(\nu_q) = \frac{1}{B[(\mu-m), T-\mu]}\frac{\nu_q^{(\mu-m-1)}}{[1+\nu_q]^{T-m}} \quad (117)$$

where $B(\nu_1, \nu_2) = \Gamma(\nu_1)\Gamma(\nu_2)/\Gamma(\nu_1+\nu_2)$ is the β-function, and the direct transformation $\beta_q = 1/(1+v_q)$ leads to the also well-known β-distribution [31], [34]

$$f(\beta_q) = \frac{1}{B[(\mu-m), T-\mu]}\beta_q^{(T-\mu-1)}(1-\beta_q)^{(\mu-m-1)}, \quad 0 \le \beta_q < 1 \quad (118)$$

(in [31] notations, $N = \mu - m + 1$ and $K = T - m - 1$). Now, according to (74)

$$\mathrm{LR}_0(m > m_0) = \prod_{q=1}^{N-m-1}\beta_q. \quad (119)$$

Since β_q are independent, we get

$$\mathcal{E}\left\{\mathrm{LR}_0^S(m)\right\} = \prod_{q=1}^{N-m-1}\mathcal{E}\left\{\beta_q^S\right\}. \quad (120)$$

For β-distribution (118), we get

$$\mathcal{E}\left\{\beta_q^S\right\} = \frac{\Gamma[T-m]\Gamma[T+S-\mu]}{\Gamma[T-\mu]\Gamma[T+S-m]}. \quad (121)$$

The p.d.f. $f(\boldsymbol{x})$ for $\mathrm{LR}_0(m > m_0)$ can now be specified using the inverse Mellin transform [15]

$$f(\boldsymbol{x}) = \prod_{q=1}^{N-m-1}\frac{\Gamma[T-m]}{\Gamma[T-\mu]}\frac{1}{2\pi i} \times \oint \boldsymbol{x}^{-(S+1)}\prod_{q=1}^{N-m-1}\frac{\Gamma[T+S-\mu]}{\Gamma[T+S-m]}\,dS \quad (122)$$

with the particular integration path specified in [15]. Note that

$$\mu = m_{\max}, \quad \text{for } q \le N - m_{\max} \quad (123)$$
$$\mu = m_{\max}-1, \ldots, (m+1), \quad \text{for } q_{\max}, \ldots, N-m-1. \quad (124)$$

Therefore, the integral in (121) can be calculated as [15]

$$\oint \boldsymbol{x}^{-(S+1)}\prod_{q=1}^{N-m-1}\frac{\Gamma[T+S-\mu]}{\Gamma[T+S-m]}\,dS = G^{(N-m-1),0}_{(N-m-1),(N-m-1)}\left(\boldsymbol{x}\,\Big|\,\begin{matrix}T-m-1,\ldots,T-m-1\\ T-m-2,\ldots,T-m_{\max}-1\end{matrix}\right). \quad (125)$$

$$\nu_q = \frac{\boldsymbol{e}_{m+1}^T\left[\hat{C}_q^{(m)}\right]^{-\frac{1}{2}}\left\{\left[\hat{C}_q^{(m)}\right]^{-\frac{1}{2}}\hat{C}_q^{12}\left(\hat{W}_q^{(22)}\right)^{-1}\hat{C}_q^{21}\left[\hat{C}_q^{(m)}\right]^{-\frac{1}{2}}\right\}\left[\hat{C}_q^{(m)}\right]^{-\frac{1}{2}}\boldsymbol{e}_{m+1}}{\boldsymbol{e}_{m+1}^T\left[\hat{C}_q^{(m)}\right]^{-\frac{1}{2}}\left[\hat{C}_q^{(m)}\right]^{-\frac{1}{2}}\boldsymbol{e}_{m+1}} \quad (109)$$

$$\hat{W}_q^{(22)} = \hat{C}_q^{22} - \hat{C}_q^{21}\left[\hat{C}_q^{(m)}\right]^{-1}\hat{C}_q^{12} = \hat{C}_{22.1}. \quad (110)$$

According to [15, (9.31.5*)], this integral could be also presented as

$$F = \boldsymbol{x}^{(T-m_{\max}-1)} G^{(N-m-1),0}_{(N-m-1),(N-m-1)} \times \left(\boldsymbol{x}\left|^{m_{\max}-m,\ldots,m_{\max}-m}_{m_{\max}-m-1,\ldots,1,0,\ldots,0}\right.\right). \quad (126)$$

Finally we get the following expression for $f(\boldsymbol{x})$:

$$f(\boldsymbol{x}) = C(T,N,m_{\max},m)\boldsymbol{x}^{(T-m_{\max}-1)} \times G^{(N-m-1),0}_{(N-m-1),(N-m-1)}\left(\boldsymbol{x}\left|^{m_{\max}-m,\ldots,m_{\max}-m}_{m_{\max}-m-1,\ldots,1,0,\ldots,0}\right.\right) \quad (127)$$

where

$$C(T,N,m_{\max},m) = \prod_{q=1}^{N-m-1} \frac{\Gamma(T-m)}{\Gamma(T-\mu(q))}. \quad (128)$$

Using [15, 7.811(2) and 9.303], we can demonstrate that $\int_0^1 f(\boldsymbol{x})d\boldsymbol{x} = 1$. ■

APPENDIX II
PROOF OF THEOREM 2

According to (34)

$$\Lambda(T_{m_0}) = \frac{\det\left[(T_{m_0})^{-1}\hat{R}^{(m)}_{\rm TVAR}\right]\exp N}{\exp\left(\operatorname{tr}\left[(T_{m_0})^{-1}\hat{R}^{(m)}_{\rm TVAR})\right]\right)} \quad (129)$$

where

$$T_{m_0} = \mathcal{E}\left\{\frac{1}{T}\sum_{j=1}^{T} x_j x_j^H\right\} \quad (\equiv \mathcal{E}\{\hat{R}\}). \quad (130)$$

Since T_{m_0} is an N-variate Toeplitz Hermitian matrix with a $(2m_0+1)$-wide band inverse, and using property (11) of the ML estimate, we have for $\hat{R}^{(m)}_{\rm TVAR}$ with $m \geq m_0$

$$\operatorname{tr}\left(T_{m_0}^{-1}\hat{R}^{(m)}_{\rm TVAR}\right) = \operatorname{tr}\left(T_{m_0}^{-1}\hat{R}\right) = \operatorname{tr}\left(T_{m_0}^{-\frac{1}{2}}\hat{R}T_{m_0}^{-\frac{1}{2}}\right) = \operatorname{tr}(\hat{C}/T) \quad (131)$$

where

$$\hat{C} \equiv T_{m_0}^{-\frac{1}{2}}\sum_{j=1}^{T} x_j x_j^H T_{m_0}^{-\frac{1}{2}} \sim \mathcal{ACW}(T,N,I_N). \quad (132)$$

Despite $\hat{C}$ being a degenerate matrix, all its $(m+1)$-variate central block matrices are nondegenerate with probability one, since $m+1 \leq T$. In particular, the trace of $\hat{C}$ is $\operatorname{tr}\hat{C} = \sum_{q=1}^{N}\hat{C}_{qq}$ with each $\hat{C}_{qq}$ being mutually independent, and having a chi-squared p.d.f. (88) [31]

$$\hat{C}_{qq} \sim \frac{1}{\Gamma(T)}\hat{C}_{qq}^{T-1}\exp[-\hat{C}_{qq}]. \quad (133)$$

Now, similarly to (19)

$$\det\left[T_{m_0}^{-1}\hat{R}^{(m)}_{\rm TVAR}\right] = \prod_{q=1}^{N} \frac{\boldsymbol{e}^T_{L(q)-q+1}\left[T_q^{(m_0)}\right]^{-1}\boldsymbol{e}^T_{L(q)-q+1}}{\boldsymbol{e}^T_{L(q)-q+1}\left[R_q^{(m)}\right]^{-1}\boldsymbol{e}^T_{L(q)-q+1}} \quad (134)$$

where

$$T_q^{(m_0)} \equiv \begin{bmatrix} t_{qq} & \cdots & t_{q,L(q)} \\ \vdots & & \vdots \\ t_{L(q),q} & \cdots & t_{L(q),L(q)} \end{bmatrix} \quad (135)$$

is a Toeplitz matrix of dimension $L(q)-q+1$. For simplicity, let us introduce the notation

$$0 \leq m_1(q) \equiv L(q) - q \leq m. \quad (136)$$

For $\hat{R}_q^{(m)}$ introduced in (10), we then have $T\hat{R}_q^{(m)} \sim \mathcal{CW}(T, m_1+1, T_q^{(m_0)})$, hence

$$\det\left[T_{m_0}^{-1}\hat{R}^{(m)}_{\rm TVAR}\right] = \prod_{q=1}^{N}\left[\boldsymbol{e}^T_{m_1+1}\left(T\hat{C}_q^{-1}\right)\boldsymbol{e}_{m_1+1}\right]^{-1} \quad (137)$$

where

$$\hat{C}_q = [\hat{C}_{\ell k}], \quad \text{for } \ell, k = q, \ldots, L(q) \quad (138)$$

and $\hat{C}$ is specified in (132). Since (again see [32, Th. 3.3.9])

$$\boldsymbol{e}^T_{m_1+1}\hat{C}_q^{-1}\boldsymbol{e}_{m_1+1} \equiv \left[\hat{C}^{(q)}_{11.2}\right]^{-1} = \left[\hat{C}_{qq} - \hat{C}^{(q)}_{12}\left(\hat{C}^{(q)}_{22}\right)^{-1}\hat{C}^{(q)}_{12}\right]^{-1} \quad (139)$$

where

$$\hat{C}_q \equiv \left[\begin{array}{c|c} \hat{C}_{qq} & \hat{C}^{(q)}_{12} \\ \hline \hat{C}^{(q)}_{21} & \hat{C}^{(q)}_{22} \end{array}\right] \begin{array}{c} (1) \\ (m_1) \end{array} \quad (140)$$
$$\qquad (1) \qquad (m_1)$$

[see (85)], we may apply the same argument as in Appendix I to declare that $\hat{C}^{(q)}_{11.2}$ $(q=1,\ldots,N)$ is a sequence of mutually independent random numbers.

Similarly to (107), we may present $[\hat{C}^{(q)}_{11.2}]^{-1}$ as

$$\hat{C}^{(q)}_{11.2} = \hat{C}_{qq}^{-1} + \hat{C}_{qq}^{-1}\hat{C}^{(q)}_{12} \times \left[\hat{C}^{(q)}_{22} - \hat{C}^{(q)}_{21}\hat{C}_{qq}^{-1}\hat{C}^{(q)}_{12}\right]^{-1}\hat{C}^{(q)}_{21}\hat{C}_{qq}^{-1} \quad (141)$$

where, according to [32, Th. 3.3.9]

$$\hat{C}_{qq}^{-\frac{1}{2}}\hat{C}^{(q)}_{21} \sim \mathcal{CN}(0, I_{m_1}) \quad (142)$$

$$\hat{W}^{(q)}_{22} \equiv \hat{C}^{(q)}_{22.1} = \hat{C}^{(q)}_{22} - \hat{C}^{(q)}_{21}\hat{C}_{qq}^{-1}\hat{C}^{(q)}_{12} \sim \mathcal{CW}(T-1, m_1, I_{m_1}) \quad (143)$$

$$\hat{\Omega}^{(q)} \equiv \frac{\hat{C}_{qq}}{T} \exp\left[-\frac{\hat{C}_{qq}}{T}\right] \tag{146}$$

$$\hat{\beta}^{(q)} \equiv \begin{cases} 1, & m_1 = 0 \\ \frac{1}{1+\boldsymbol{x}^H\left(\hat{W}_{22}^{(q)}\right)^{-1}\boldsymbol{x}}, \quad f(\hat{\beta}^{(q)}) = \frac{\left(\hat{\beta}^{(q)}\right)^{(T-m_1-1)}}{B[m_1, T-m_1]}\left(1-\hat{\beta}^{(q)}\right)^{(m_1-1)}, & m_1 > 0. \end{cases} \tag{147}$$

with $\hat{C}_{qq}^{-\frac{1}{2}}\hat{C}_{21}^{(q)}, \hat{W}_{22}^{(q)}$ and $\hat{C}_{qq}$ being mutually independent. Therefore

$$\left[\hat{C}_{11.2}^{(q)}\right]^{-1} = \hat{C}_{qq}^{-1}\left[1+\boldsymbol{x}^H\left(\hat{W}_{22}^{(q)}\right)^{-1}\boldsymbol{x}\right], \quad \boldsymbol{x} \sim \mathcal{CN}(0, I_{m_1}) \tag{144}$$

and

$$\Lambda(T_{m_0}) = \exp N \prod_{q=1}^{N} \frac{\hat{C}_{qq}}{T\left[1+\boldsymbol{x}^H\left(\hat{W}_{22}^{(q)}\right)^{-1}\boldsymbol{x}\right]\exp\left[\frac{\hat{C}_{qq}}{T}\right]} = \exp N \prod_{q=1}^{N} \hat{\Omega}^{(q)}\hat{\beta}^{(q)} \tag{145}$$

where (146) and (147) hold, shown at the top of the page. The p.d.f. for $\hat{\Omega}^{(q)}$ could be found similarly to [20] in a form of an infinite series by applying an inverse Mellin transform to the moment function $f(p) \equiv \mathcal{E}\{\Lambda^p(T_{m_0})\}$, where $\mathcal{E}\{\hat{\beta}^{(q)})^p\}$ is given by (121) and

$$\mathcal{E}\left\{\left(\hat{\Omega}^{(q)}\right)^p\right\} = \frac{T^T\Gamma(T+p)}{(T+p)^{(T+p)}\Gamma(T)}. \tag{148}$$

Since $\hat{\Omega}^{(q)}$ and $\hat{\beta}^{(q)}$ are independent, we get

$$\mathcal{E}\left\{\left(\hat{\Omega}^{(q)}\hat{\beta}^{(q)}\right)^p\right\} = \frac{T^T\Gamma(T+p-m_1)}{(T+p)^{(T+p)}\Gamma(T-m_1)}. \tag{149}$$

Finally

$$\mathcal{E}\{\Lambda^p(T_{m_0})\} = \exp pN \prod_{q=1}^{N} \frac{\Gamma(T+p-m_1(q))T^T}{(T+p)^{(T+p)}\Gamma(T-m_1(q))} \tag{150}$$

or

$$\mathcal{E}\{\Lambda^p(T_{m_0})\} = \frac{T^{TN}\exp pN}{(T+p)^{N(T+p)}} \frac{\prod_{q=1}^{N}\Gamma(T+p-m_1(q))}{\prod_{q=1}^{N}\Gamma(T-m_1(q))}. \tag{151}$$

A comparison of (151) with the derivations in [20] suggests that, with minor modifications, we can apply the same transformations to the inverse Mellin transform of (151) to get a serial representation for the p.d.f. of $\Lambda(T_{m_0})$, similar to the expression introduced in [36] for $\mathrm{LR}(R_0)$

$$\mathrm{LR}(R_0) = \frac{\det\left[R_0^{-1}\hat{R}\right]\exp M}{\exp\left(\mathrm{tr}\left[R_0^{-1}\hat{R}\right]\right)} \tag{152}$$

with $\hat{R} \sim \mathcal{CW}(N > M, M, R_0)$.

These derivations suggest that computationally it is more convenient to deal with

$$\Lambda_0(T_m) \equiv [\Lambda(T_m)]^{\frac{1}{TN}}. \tag{153}$$

REFERENCES

[1] E. Hannan, *Multiple Time Series*. New York: Wiley, 1970.

[2] B. Porat, *Digital Processing of Random Signals*, 5th ed. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[3] A. McQuarrie and C. Tsai, *Regression and Time Series Model Selection*. Singapore: World Scientific, 1998.

[4] E. Polak, "Algorithms for maximum likelihood constrained covariance estimation," presented at the CSSIP Short Courses Program, Adelaide, Australia, 2001.

[5] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 33, no. 2, pp. 387–392, Apr. 1985.

[6] Y. Abramovich, N. Spencer, and M. Turley, "Time-varying autoregressive (TVAR) models for multiple radar observations," *IEEE Trans. Signal Process.*, vol. 55, no. 4, Apr. 2007, to be published.

[7] R. Janik and M. Nowak, "Wishart and anti-Wishart random matrices," *J. Phys. A: Math. Gen.*, vol. 36, pp. 3629–3637, 2003.

[8] T. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1958.

[9] H. Dym and I. Gohberg, "Extensions of band matrices with band inverses," *Linear Algebra Appl.*, vol. 36, pp. 1–24, 1981.

[10] R. Grone, C. Johnson, E. M. de Sá, and H. Wolkowicz, "Positive definite completions of partial Hermitian matrices," *Linear Algebra Appl.*, vol. 58, pp. 109–124, 1984.

[11] H. Dym and I. Gohberg, "A new class of contractive interpolants and maximum entropy principles," in *Topics in Operator Theory and Interpolation*. Cambridge, MA: Birkhäuser Verlag, 1988, pp. 117–150.

[12] H. Woerdeman, "Matrix and operator extensions," Ph.D. dissertation, Vrije Universiteit, Amsterdam, The Netherlands, 1989.

[13] P. Forster and G. Vezzosi, "Optimal toeplitzification with a given rank constraint," in *Proc. ICASSP*, 1989, pp. 2783–2786.

[14] K. Grigoriadis, A. Frazho, and R. Skelton, "Application of alternating convex projection methods for computation of positive definite Toeplitz matrices," *IEEE Trans. Signal Process.*, vol. 42, no. 7, pp. 1873–1975, Jul. 1994.

[15] I. Gradshteyn and I. Ryzhik, *Tables of Integrals, Series, and Products*, 6th ed. New York: Academic, 2000.

[16] P. Djurić, "A model selection rule for sinusoids in white Gaussian noise," *IEEE Trans. Signal Process.*, vol. 44, no. 7, pp. 1744–1757, Jul. 1996.

[17] R. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.

[18] Y. Abramovich, N. Spencer, and A. Gorokhov, "Bounds on maximum likelihood ratio—Part I: Application to antenna array detection-estimation with perfect wavefront coherence," *IEEE Trans. Signal Process.*, vol. 52, no. 6, pp. 1524–1536, Jun. 2004.

[19] Y. Abramovich, N. Spencer, and A. Gorokhov, "Bounds on maximum likelihood ratio—Part II: Application to antenna array detection-estimation with imperfect wavefront coherence," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2046–2058, Jun. 2005.

[20] B. Nagarsenker and K. Pillai, "Distribution of the likelihood ratio criterion for testing a hypothesis specifying a covariance matrix," *Biometrica*, vol. 60, no. 2, pp. 359–361, 1973.

[21] Y. Abramovich, D. Arov, and V. Kachur, "Adaptive cancellation filters for stationary interference with a Toeplitz correlation matrix," *Soviet J. Commun. Technol. Electron.*, vol. 33, no. 4, pp. 54–61, 1988, English translation of *Radioteknika i Elektronika*.

[22] Y. Abramovich, D. Gray, A. Gorokhov, and N. Spencer, "Positive-definite Toeplitz completion in DOA estimation for nonuniform linear antenna arrays—Part I: Fully augmentable arrays," *IEEE Trans. Signal Process.*, vol. 46, no. 9, pp. 2458–2471, Sep. 1998.

[23] R. Nitzberg, "Application of maximum likelihood estimation of persymmetric covariance matrices to adaptive processing," *IEEE Trans. Aerosp. Electr. Syst.*, vol. 16, no. 1, pp. 124–127, Jan. 1980.

[24] I. Gohberg and A. Semencul, "On the inversion of finite Toeplitz matrices and their continuous analogues," (in Russian) *Mat. Issled.*, vol. 7, no. 2, pp. 201–223, 1972.

[25] J. Parent and A. Bourdillon, "A method to correct HF skywave backscattered signals for ionospheric frequency modulation," *IEEE Trans. Ant. Propag.*, vol. 36, no. 1, pp. 127–135, Jan. 1988.

[26] Y. Abramovich, N. Spencer, S. Anderson, and A. Gorokhov, "Stochastic-constraints method in nonstationary hot-clutter cancellation—Part I: Fundamentals and supervised training applications," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 4, pp. 1271–1292, Oct. 1998.

[27] Y. Abramovich, N. Spencer, and S. Anderson, "Stochastic-constraints method in nonstationary hot-clutter cancellation—Part II: Unsupervised training applications," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 1, pp. 132–150, Jan. 2000.

[28] E. Hannan and B. Quinn, "The determination of the order of an autoregression," *J. Royal Soc. Series B*, vol. 41, no. 2, pp. 190–195, 1979.

[29] Y. Abramovich, N. Spencer, and A. Gorokhov, "Detection-estimation of more uncorrelated Gaussian sources than sensors in nonuniform linear antenna arrays—Part I: Fully augmentable arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 5, pp. 959–971, May 2001.

[30] Y. Abramovich, N. Spencer, and A. Gorokhov, "Detection-estimation of more uncorrelated Gaussian sources than sensors in nonuniform linear antenna arrays—Part II: Partially augmentable arrays," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1492–1507, Jun. 2003.

[31] I. Reed, J. Mallett, and L. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 10, no. 6, pp. 853–863, Nov. 1974.

[32] A. Gupta and D. Nagar, *Matrix Variate Distributions*. London, U.K.: Chapman & Hall/CRC, 2000.

[33] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[34] M. Siotani, T. Hayakawa, and Y. Fujikoshi, *Modern Multivariate Statistical Analysis*. OH: Amer. Sci. Press, 1985.

[35] Y. Abramovich, "Analysis of a direct adaptive tuning method for interference compensation systems with auxiliary linear constraints," *Soviet J. Commun. Technol. Electron.*, vol. 35, no. 1, pp. 30–37, 1990, English translation of *Radioteknika i Elektronika*.

[36] Y. Abramovich, N. Spencer, and A. Gorokhov, "A modified GLRT and AMF framework for diagonally loaded and fast maximum-likelihood adaptive detectors," *IEEE Trans. Aerosp. Electron. Syst.*, Jan. 2005, to be published.

Yuri I. Abramovich (M'96–SM'06) received the Dipl.Eng. (honors) degree in radio electronics and the Cand.Sci. degree (Ph.D. equivalent) in theoretical radio techniques, both from the Odessa State Polytechnic University, Odessa, Ukraine, U.S.S.R., in 1967 and in 1971, respectively, and the D.Sc. degree in radar and navigation from the Leningrad Institute for Avionics, Leningrad, Russia, U.S.S.R., in 1981.

Since 2000, he has been with the Australian Defence Science and Technology Organization (DSTO), Adelaide, Australia, as a Principal Research Scientist, seconded to CSSIP until its closure. From 1968 to 1994, he was with the Odessa State Polytechnic University as a Research Fellow, a Professor, and ultimately a Vice-Chancellor of Science and Research. From 1994 to 2006, he was at the Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP), Adelaide, Australia. His research interests include signal processing (particularly spatio-temporal adaptive processing, beamforming, signal detection, and estimation), its application to radar (particularly over-the-horizon radar), electronic warfare, and communication. He served as an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2002 to 2005.

Nicholas K. Spencer received the B.Sc. (honors) degree in applied mathematics and the M.Sc. degree in computational mathematics from the Australian National University, Canberra, Australia, in 1985 and 1992, respectively.

He is currently a Senior Researcher at Adelaide Research and Innovation Pty. Ltd. (ARI), Adelaide, Australia. He has been with the Australian Department of Defense, Canberra, Australia; the Flinders University of South Australia, Adelaide, Australia; the University of Adelaide; the Australian Centre for Remote Sensing, Canberra, Australia; and the Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP), Adelaide, Australia, in the areas of computational and mathematical sciences. His research interests include array signal processing, parallel and supercomputing, software best-practice, human-machine interfaces, multilevel numerical methods, modelling, and simulation of physical systems, theoretical astrophysics, and cellular automata.

Michael D. E. Turley received the Ph.D. degree in plasma physics from The Flinders University of South Australia, Adelaide, Australia, in 1986.

Since 1986, he has been with HF Skywave and Surface Wave Radar Projects, with the Defense Science and Technology Organization, Edinburgh, South Australia. As a Principal Research Scientist, he presently heads the Signal Processing and Propagation Group, the Intelligence, Surveillance, and Reconnaissance Division. His research interests include the area of ionospheric physics and signal processing techniques, and clutter mitigation, signal detection, CFAR, EW, adaptive beamforming, and spectral analysis.