

A New Nonstationarity Detector

Steven Kay*

Dept. of Electrical, Computer, and Biomedical Engineering

University of Rhode Island

Kingston, RI 02881

401-874-5804 (voice) 401-782-6422 (fax)

kay@ele.uri.edu

Keywords: signal detection, spectral analysis

EDICS: SSP-SPEC, SSP-SNMD

September 8, 2007

Abstract

A new test to determine the stationarity length of a locally wide sense stationary Gaussian random process is proposed. Based on the modeling of the process as a time-varying autoregressive process, the time-varying model parameters are tested using a Rao test. The use of a Rao test avoids the necessity of obtaining the maximum likelihood estimator of the model parameters under the alternative hypothesis, which is intractable. Computer simulation results are given to demonstrate its effectiveness and to verify the asymptotic theoretical performance of the test. Applications are to spectral analysis, noise estimation, and time series modeling.

1 Introduction

There are many statistical signal processing approaches that are based on the assumption of a wide sense stationary (WSS) Gaussian random process. Some of these are spectral analysis [1,2], signal detection [3], and general time series modeling [1,4]. For example, in spectral analysis we wish to base our estimate on the largest data record that retains the stationarity of the process, while in signal detection, it is imperative that an accurate estimate of the stationary noise floor be available. In time series modeling such as for autoregressive, moving average, and autoregressive moving average models the primary assumption is that

*This work was supported by the Naval Air Warfare Center, Patuxent, MD under the Office of Naval Research contract N0001404-M-0331.

of a WSS random process. In practice, however, a test for stationarity is seldom invoked before choosing a data record length. Generally, the choice of a stationarity interval is based on physical arguments, which may not always be valid, or even if valid, may become violated as time evolves. Performing, for example, a spectral analysis on a data record that exhibits a nonstationarity will result in a severely biased estimate.

The difficulty in designing an efficient test for stationarity is in having to assume an alternative hypothesis and to estimate some set of parameters under the alternative hypothesis. In this paper we will consider a Gaussian random process that exhibits a “slowly varying” type of nonstationarity. That is to say, the power spectral density (PSD) of the process is slowly varying as opposed to an abrupt change for which many efficient tests exist [3,5]. In order to design an efficient test, i.e., one that is able to quickly determine when the PSD has changed significantly, we will require a model for the alternative hypothesis that is accurately estimated using only a short data record. Such a model for a WSS random process is the AR model [1] and its extension to the locally stationary case is the time-varying AR (TVAR) model [6,7,8]. The main advantages of this model is that it is capable of representing any PSD and the AR *filter* parameters may be accurately estimated using a linear model type of estimate. Some areas in which the TVAR model has been used successfully are in speech processing [7,8], in estimation of the time-varying center frequency of a narrowband process [18], and in classification of EEGs [19]. Note that in the previous work cited, it had to be assumed that the excitation noise variance was constant and known. This restriction was placed on this parameter in order to retain the linearity since otherwise the estimation problem become highly nonlinear. The approach that we will describe shortly will be able to accommodate a time-varying excitation noise by circumventing the estimation problem under the alternative hypothesis. It is critical that this time variation be allowed since in practice it is quite common for the spectral shape to remain nearly constant but to have an overall power that is time-varying.

Some previous tests for general nonstationarity can be found in [9–12], as well as many other papers that treat only special cases of nonstationarity, for example in [13]. The tests of [9–12] are based on the statistics of the Fourier transform, which are only true asymptotically. Therefore, it is not clear that the approaches are viable for the shorter data records employed here. For hypotheses that only prescribe that the process be WSS, there do not appear to be many approaches.

In summary, we propose the use of the TVAR model for the alternative hypothesis. This is because under the null hypothesis, i.e., the stationary case, the AR model has been shown to be easily estimated using an approximate maximum likelihood estimator (MLE). The approximate MLE is linear and yields the asymptotic properties of the MLE for relatively short data records, less than 100 data samples. Also, the model is capable of representing any PSD [1]. The TVAR model retains most of the properties of the time invariant AR model, except that the estimation of the excitation noise variance makes the estimation procedure nonlinear. To circumvent this problem we propose the Rao test [3], which only requires the

MLE under the null hypothesis.

The paper is organized as follows. Section 2 summarizes the modeling used and the resultant nonstationarity detector. In Section 3 some examples are given to illustrate the evaluation of the test, as well as some computer simulation results. An application to a practical problem of interest is described in Section 4 while Section 5 discusses the proposed test and possible desired extensions.

2 Modeling and Summary of Test

The TVAR model is given as [6]

$$x[n] = - \sum_{i=1}^p a_i[n-i]x[n-i] + b[n]w[n] \quad (1)$$

where $w[n]$ is white Gaussian noise (WGN) with unit variance and the time-varying AR parameters are

$$\begin{aligned} a_i[n] &= \sum_{j=0}^m a_{ij}f_j[n] & i = 1, 2, \dots, p \\ b[n] &= \sum_{j=0}^m b_jf_j[n] \end{aligned}$$

for some suitable set of basis functions $\{f_0[n], f_1[n], \dots, f_m[n]\}$. We select $f_0[n] = c$, for c a constant, so that if $f_j[n] = 0$ for $j = 1, 2, \dots, m$, then $x[n]$ corresponds to a stationary AR process. (Note that with the Gaussian assumption wide sense stationarity implies the stronger condition of stationarity.) In order for the model to be identifiable we assume that $b[n] > 0$ for all n . A nonstationary process will result whenever any of the parameters $\{a_{1j}, a_{2j}, \dots, a_{pj}, b_j\}$ for $j = 1, 2, \dots, m$ are nonzero. Hence, the Rao test will be testing whether or not $a_{ij} = 0$ for $i = 1, 2, \dots, p; j = 1, 2, \dots, m$ and $b_j = 0$ for $j = 1, 2, \dots, m$. Under \mathcal{H}_0 the AR process is stationary so that we have the usual representation (letting $f_j[n] = 0$ for $j = 1, 2, \dots, m$)

$$x[n] = - \sum_{i=1}^p a_{i0}cx[n-i] + b_0cw[n] \quad (2)$$

with AR filter parameters $\{a_{10}c, a_{20}c, \dots, a_{p0}c\}$ and excitation noise variance $b_0^2c^2$.

The Rao test is derived in Appendix A. It is important to note that in implementing the test we only require the MLE of the TVAR parameters under \mathcal{H}_0 , which is just the MLE of the stationary AR process parameters. This greatly simplifies the implementation and amounts to a simple standard AR parameter estimation where only the parameters in (2) need be estimated. In Appendix B we give a simple explanation as to how the Rao test is able to avoid computing the MLE under \mathcal{H}_1 , as this is a crucial property. Also, note that the Rao test is referred to in the statistical literature as the Lagrange multiplier

test [20]. We assume that we wish to decide whether a segment of the realization composed of the data samples $\{x[0], x[1], \dots, x[N-1]\}$ is stationary. To do so we reject the stationarity hypothesis if

$$T_N(\mathbf{x}) = \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{a}} \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}^T \left[\mathbf{I}_{\mathbf{a}'\mathbf{a}'}^{-1}(\tilde{\boldsymbol{\theta}}) \right]_{\mathbf{aa}} \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{a}} \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} + \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{b}} \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}^T \left[\mathbf{I}_{\mathbf{b}'\mathbf{b}'}^{-1}(\tilde{\boldsymbol{\theta}}) \right]_{\mathbf{bb}} \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{b}} \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} > \gamma \quad (3)$$

where the threshold γ is chosen to maintain a constant false alarm probability (a false alarm occurs if we say it is nonstationary when it is actually stationary). *All quantities are evaluated at $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$, where $\tilde{\boldsymbol{\theta}}$ is the MLE under \mathcal{H}_0 .* The MLE required is that of the AR parameters under the assumption of stationarity or for the process given by (2). The various gradients and matrices in (3) are defined as follows.

$$\frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{a}} = \begin{bmatrix} \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{11}} \\ \vdots \\ \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{p1}} \\ \text{---} \\ \vdots \\ \text{---} \\ \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{1m}} \\ \vdots \\ \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{pm}} \end{bmatrix} \quad (mp \times 1) \quad (4)$$

where for $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$

$$\frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{rs}} \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = - \sum_{n=p}^{N-1} \frac{\hat{u}[n] f_s[n-r] x[n-r]}{\hat{b}_0^2 c^2} \quad (5)$$

for $r = 1, 2, \dots, p; s = 1, 2, \dots, m$, and where $\hat{u}[n] = x[n] + \sum_{i=1}^p \hat{a}_{i0} c x[n-i]$. Also,

$$\left[\frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{b}} \right]_r \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial b_r} \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = \sum_{n=p}^{N-1} \frac{f_r[n]}{\hat{b}_0^3 c^3} (\hat{u}^2[n] - \hat{b}_0^2 c^2) \quad (6)$$

for $r = 1, 2, \dots, m$. The estimates indicated, which are $\{\hat{a}_{10}c, \hat{a}_{20}c, \dots, \hat{a}_{p0}c, \hat{b}_0^2 c^2\}$, are just the usual covariance method estimates for the parameters of an AR process based on the data $x[n]$ for $n = 0, 1, \dots, N-1$ [1].

The matrices are next defined. For the AR filter parameters we have

$$\left[\mathbf{I}_{\mathbf{a}'\mathbf{a}'}^{-1}(\boldsymbol{\theta}) \right]_{\mathbf{aa}} = \left(\mathbf{I}_{\mathbf{aa}} - \mathbf{I}_{\mathbf{aa}0} \mathbf{I}_{\mathbf{a}_0\mathbf{a}_0}^{-1} \mathbf{I}_{\mathbf{a}_0\mathbf{a}} \right)^{-1}$$

where the matrices are partitions of the Fisher information matrix (FIM) given by

$$\begin{aligned} \mathbf{I}_{\mathbf{a}'\mathbf{a}'}(\boldsymbol{\theta}) &= \begin{bmatrix} \mathbf{I}_{\mathbf{a}'}(0,0) & | & \mathbf{I}_{\mathbf{a}'}(0,1) & \dots & \mathbf{I}_{\mathbf{a}'}(0,m) \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ \mathbf{I}_{\mathbf{a}'}(1,0) & | & \mathbf{I}_{\mathbf{a}'}(1,1) & \dots & \mathbf{I}_{\mathbf{a}'}(1,m) \\ \vdots & | & \vdots & \ddots & \vdots \\ \mathbf{I}_{\mathbf{a}'}(m,0) & | & \mathbf{I}_{\mathbf{a}'}(m,1) & \dots & \mathbf{I}_{\mathbf{a}'}(m,m) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_{\mathbf{a}_0\mathbf{a}_0} & \mathbf{I}_{\mathbf{a}_0\mathbf{a}} \\ \mathbf{I}_{\mathbf{a}\mathbf{a}_0} & \mathbf{I}_{\mathbf{a}\mathbf{a}} \end{bmatrix} = \begin{bmatrix} p \times p & p \times mp \\ mp \times p & mp \times mp \end{bmatrix} \end{aligned} \quad (7)$$

with the dimensions of the partitions indicated. Note that each submatrix of $\mathbf{I}_{\mathbf{a}'\mathbf{a}'}(\boldsymbol{\theta})$ is $p \times p$. The submatrices when evaluated at $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$ are given by

$$[\mathbf{I}_{\mathbf{a}'}(r,s)]_{kl} = \frac{\hat{r}_x[k-l]}{\hat{b}_0^2 c^2} \sum_{n=p}^{N-1} f_r[n-k] f_s[n-l] \quad (8)$$

for $k, l = 1, 2, \dots, p; r, s = 0, 1, \dots, m$. Since the FIM is evaluated under \mathcal{H}_0 the estimates $\hat{r}_x[k-l]$ and $\hat{b}_0^2 c^2$ are obtained by first using the covariance method, which immediately gives the latter estimate, and then constructing the estimated autocorrelation sequence estimates $\{\hat{r}_x[0], \hat{r}_x[1], \dots, \hat{r}_x[p-1]\}$ using the step-down procedure followed by a recursive difference equation [1]. The excitation noise variance matrix is defined as

$$[\mathbf{I}_{\mathbf{b}'\mathbf{b}'}^{-1}(\boldsymbol{\theta})]_{\mathbf{b}\mathbf{b}} = \left(\mathbf{I}_{\mathbf{b}\mathbf{b}} - \mathbf{I}_{\mathbf{b}b_0} \mathbf{I}_{b_0 b_0}^{-1} \mathbf{I}_{b_0 \mathbf{b}} \right)^{-1}$$

where the partitioned FIM matrix is defined as

$$\begin{aligned} \mathbf{I}_{\mathbf{b}'\mathbf{b}'}(\boldsymbol{\theta}) &= \begin{bmatrix} I_{b'}(0,0) & | & I_{b'}(0,1) & \dots & I_{b'}(0,m) \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ I_{b'}(1,0) & | & I_{b'}(1,1) & \dots & I_{b'}(1,m) \\ \vdots & | & \vdots & \ddots & \vdots \\ I_{b'}(m,0) & | & I_{b'}(m,1) & \dots & I_{b'}(m,m) \end{bmatrix} \\ &= \begin{bmatrix} I_{b_0 b_0} & \mathbf{I}_{b_0 \mathbf{b}} \\ \mathbf{I}_{\mathbf{b} b_0} & \mathbf{I}_{\mathbf{b}\mathbf{b}} \end{bmatrix} = \begin{bmatrix} 1 \times 1 & 1 \times m \\ m \times 1 & m \times m \end{bmatrix} \end{aligned} \quad (9)$$

with the dimensions of the partitions indicated. The elements of $\mathbf{I}_{\mathbf{b}'\mathbf{b}'}(\boldsymbol{\theta})$ when evaluated at $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$ are

$$I_{b'}(r,s) = \frac{2}{\hat{b}_0^2 c^2} \sum_{n=p}^{N-1} f_r[n] f_s[n]. \quad (11)$$

The performance of the Rao test can be found asymptotically or as $N \rightarrow \infty$. In practice, because of our choice of an AR model the asymptotic performance will usually be attained for relatively short

data records. Depending on the sharpness of the PSD the necessary data record length can be as short as $N = 100$ samples. Hence, under \mathcal{H}_0 it can be shown that [3] the Rao test has a central chi-squared distribution or

$$T_N(\mathbf{x}) \sim \chi_{m(p+1)}^2 \quad (12)$$

and under \mathcal{H}_1 , it has a noncentral chi-squared distribution or

$$T_N(\mathbf{x}) \sim \chi_{m(p+1)}^2(\lambda) \quad (13)$$

where the noncentrality parameter is given by

$$\lambda = \mathbf{a}^T \left(\mathbf{I}_{\mathbf{a}\mathbf{a}} - \mathbf{I}_{\mathbf{a}\mathbf{a}_0} \mathbf{I}_{\mathbf{a}_0\mathbf{a}_0}^{-1} \mathbf{I}_{\mathbf{a}_0\mathbf{a}} \right) \mathbf{a} + \mathbf{b}^T \left(\mathbf{I}_{\mathbf{b}\mathbf{b}} - \mathbf{I}_{\mathbf{b}b_0} \mathbf{I}_{b_0b_0}^{-1} \mathbf{I}_{b_0\mathbf{b}} \right) \mathbf{b}. \quad (14)$$

The vector \mathbf{a} , which is $mp \times 1$, and \mathbf{b} , which is $m \times 1$, are the AR filter parameter and excitation noise parameter vectors defined as

$$\mathbf{a} = \begin{bmatrix} a_{11} \\ \vdots \\ a_{p1} \\ \text{---} \\ \vdots \\ \text{---} \\ a_{1m} \\ \vdots \\ a_{pm} \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

and are evaluated at the true values of the parameters under \mathcal{H}_1 , i.e., for the nonstationary AR process. The matrices, on the other hand, are all evaluated under \mathcal{H}_0 . Hence, all matrices are defined as before except that we evaluate them for the true parameters under \mathcal{H}_0 and not estimates. As a result, we have that

$$[\mathbf{I}_{\mathbf{a}'}(r, s)]_{kl} = \frac{r_x[k-l]}{b_0^2 c^2} \sum_{n=p}^{N-1} f_r[n-k] f_s[n-l]$$

and

$$I_{b'}(r, s) = \frac{2}{b_0^2 c^2} \sum_{n=p}^{N-1} f_r[n] f_s[n]. \quad (15)$$

3 Some Examples

In this section we explicitly evaluate the Rao test and illustrate its performance for two simple cases. The first is that of a white Gaussian noise (WGN) process whose power is changing in time and the second is an AR process of order one whose filter parameter is changing in time.

3.1 WGN process with time-varying power

Assume that $x[n] = w[n]$, where $w[n]$ is nominally WGN but whose power, which is $b^2[n]$, may be time-varying. Since $b[n] = \sum_{j=0}^m b_j f_j[n]$, we will test if $b_1 = b_2 = \dots = b_m = 0$, i.e., our hypothesis under \mathcal{H}_0 . The Rao test is then given from (3) as

$$T_N(\mathbf{x}) = \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{b}} \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}^T \left[\mathbf{I}_{\mathbf{b}'\mathbf{b}'}^{-1}(\tilde{\boldsymbol{\theta}}) \right]_{\mathbf{bb}} \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{b}} \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} > \gamma$$

where the elements of the gradient vector are from (6) with $\hat{u}[n]$ replaced by $x[n]$ since for this example, $x[n] = u[n]$,

$$\frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial b_r} \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = \sum_{n=0}^{N-1} \frac{f_r[n]}{\hat{b}_0^3 c^3} (x^2[n] - \hat{b}_0^2 c^2) \quad (16)$$

for $r = 1, 2, \dots, m$. Also, the FIM is given by (9) and (10) with elements defined in (11). The required MLE of b_0 under \mathcal{H}_0 can be obtained from the known result for the MLE of the variance of WGN [16]

$$\hat{\sigma}_x^2 = \frac{1}{N} \sum_{n=0}^{N-1} x^2[n]$$

and noting that since $u[n] = x[n]$, $\hat{b}_0^2 c^2 = \hat{\sigma}_x^2$.

For basis functions we will choose those corresponding to a second-order polynomial or the set $\{1, n, n^2\}$. The choice of the basis functions is dictated by the need to represent a slowly varying function since the nonstationarity is slowly varying. Other possible basis functions could be a set of low frequency sinusoids. We have found satisfactory performance with a low-order polynomial and thus have not pursued this matter further. The number of basis functions should be kept as small as possible since ultimately we will have to estimate the parameters. Too many basis functions will result in having to raise the detection threshold to maintain a given probability of false alarm. It is even possible to estimate the number of basis functions and incorporate this estimate into the nonstationarity detector. Such techniques as the minimum description length (MDL) [14] and the exponentially embedded family (EEF) model order estimator [15] could be used. This could be the topic of a future paper.

To simplify the computations and implementation we use the orthogonal polynomials given for $n = 0, 1, \dots, N-1$ by [17]

$$\begin{aligned} f_0[n] &= \frac{1}{\sqrt{N}} \\ f_1[n] &= \frac{n - \mu_1}{\|\mathbf{n} - \mu_1 \mathbf{1}\|} \\ f_2[n] &= \frac{(n - \mu_1)^2 - (\mu_3/\mu_2)(n - \mu_1) - \mu_2}{\|\tilde{\mathbf{n}}_2 - (\mu_3/\mu_2)(\mathbf{n} - \mu_1 \mathbf{1}) - \mu_2 \mathbf{1}\|} \end{aligned}$$

where $\mathbf{n} = [0 \ 1 \ \dots \ N - 1]^T$, $\bar{\mathbf{n}}_2 = [(0 - \mu_1)^2 \ (1 - \mu_1)^2 \ \dots \ ((N - 1) - \mu_1)^2]^T$, $\mathbf{1} = [1 \ 1 \ \dots \ 1]^T$, $\|\cdot\|$ denotes Euclidean norm, and

$$\begin{aligned}\mu_1 &= \frac{1}{N} \sum_{n=0}^{N-1} n \\ \mu_i &= \frac{1}{N} \sum_{n=0}^{N-1} (n - \mu_1)^i \quad i = 2, 3.\end{aligned}$$

Note that a linear combination of these polynomials produces a second-order polynomial and also that the orthogonality property is

$$\sum_{n=0}^{N-1} f_r[n] f_s[n] = \delta_{rs}.$$

The orthogonalized polynomials are the result of a Gram-Schmidt orthogonalization of the set $\{1, n, n^2\}$. Finally, we have that $f_0[n] = c = 1/\sqrt{N}$. An example of the polynomials is shown in Figure 1.

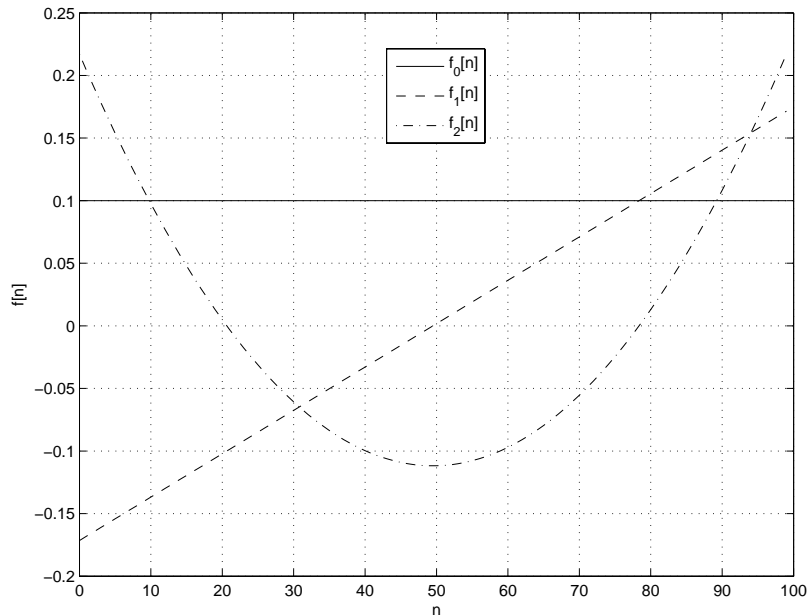


Figure 1: Orthogonal polynomials for $N = 100$.

The elements of the Fisher information matrix evaluated at $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$ are now from (15)

$$I'_b(r, s) = \frac{2}{\hat{\sigma}_0^2 c^2} \sum_{n=0}^{N-1} f_r[n] f_s[n] = \frac{2}{\sigma_x^2} \delta_{rs}$$

so that

$$\mathbf{I}_{\mathbf{b}'\mathbf{b}'}(\tilde{\boldsymbol{\theta}}) = \frac{2}{\sigma_x^2} \mathbf{I}_3$$

where \mathbf{I}_n denotes an $n \times n$ identity matrix. As a result, we have that

$$\left[\mathbf{I}_{\mathbf{b}'\mathbf{b}'}^{-1}(\tilde{\boldsymbol{\theta}})\right]_{\mathbf{b}\mathbf{b}} = \frac{\hat{\sigma}_x^2}{2}\mathbf{I}_2. \quad (17)$$

Finally, we have from (16) and (17) that we should reject the hypothesis of stationarity if

$$T_N(\mathbf{x}) = \frac{1}{2(\hat{\sigma}_x^2)^2} \sum_{r=1}^2 \left[\sum_{n=0}^{N-1} f_r[n] \left(x^2[n] - \hat{\sigma}_x^2 \right) \right]^2 > \gamma.$$

To determine the asymptotic detection performance we use (12)–(14) with $p = 0$ and $m = 2$ to yield $T_N(\mathbf{x}) \sim \chi_2^2$ under \mathcal{H}_0 , $T_N(\mathbf{x}) \sim \chi_2^2(\lambda)$ under \mathcal{H}_1 , where

$$\lambda = \mathbf{b}^T (\mathbf{I}_{\mathbf{b}\mathbf{b}} - \mathbf{I}_{\mathbf{b}\mathbf{b}_0} \mathbf{I}_{\mathbf{b}_0\mathbf{b}_0}^{-1} \mathbf{I}_{\mathbf{b}_0\mathbf{b}}) \mathbf{b} = \mathbf{b}^T \left(\left[\mathbf{I}_{\mathbf{b}'\mathbf{b}'}^{-1}(\boldsymbol{\theta}_0)\right]_{\mathbf{b}\mathbf{b}} \right)^{-1} \mathbf{b}.$$

Here, we have that $\mathbf{b} = [b_1 \ b_2]^T$ and from (17) (with the true value of σ_x^2 under \mathcal{H}_0 used)

$$\left[\mathbf{I}_{\mathbf{b}'\mathbf{b}'}^{-1}(\boldsymbol{\theta}_0)\right]_{\mathbf{b}\mathbf{b}} = \frac{\sigma_x^2}{2}\mathbf{I}_2 = \frac{b_0^2}{2N}\mathbf{I}_2$$

so that

$$\lambda = \frac{2N(b_1^2 + b_2^2)}{b_0^2}.$$

As an example, first consider a stationary WGN process for $N = 100$. The parameters are chosen as $b_0 = 0.3$, $b_1 = b_2 = 0$. The estimated PDF (shown as a bar plot) and the theoretical asymptotic χ_2^2 PDF are shown in Figure 2. Next, we simulate a nonstationary WGN process using $b_0 = 0.3$, $b_1 = 0.04$, and $b_2 = 0.01$. A typical realization is shown in Figure 3 along with the square-root of the time-varying variance, i.e., the standard deviation (shown dashed). Finally, in Figure 4 is shown the estimated PDF (shown as a bar plot) and the theoretical asymptotic $\chi_2^2(\lambda)$ PDF. It is seen that the performance is described quite accurately using the asymptotic results, even for the relatively short data record of 100 samples. Another example follows.

3.2 AR process with time-varying filter parameter

For this example we assume an TVAR process for $p = 1$ with a time-varying filter parameter and a constant excitation noise variance so that $x[n] = -a_1[n-1] + b_0 c w[n]$. As before, we let $m = 2$ and hence the time-varying filter parameter is given by $a_1[n] = a_{10}f_0[n] + a_{11}f_1[n] + a_{12}f_2[n]$. To determine the Rao test statistic we first note that since $b[n]$ is constant, we have from (3)

$$T_N(\mathbf{x}) = \left. \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{a}} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}^T \left[\mathbf{I}_{\mathbf{a}'\mathbf{a}'}^{-1}(\tilde{\boldsymbol{\theta}})\right]_{\mathbf{a}\mathbf{a}} \left. \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{a}} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}$$

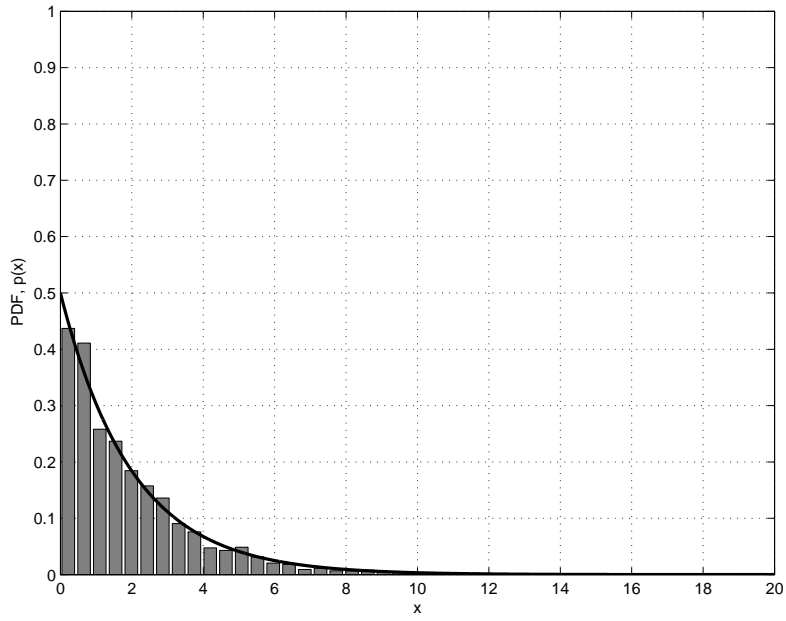


Figure 2: Estimated and theoretical PDF for $N = 100$ - stationary WGN.

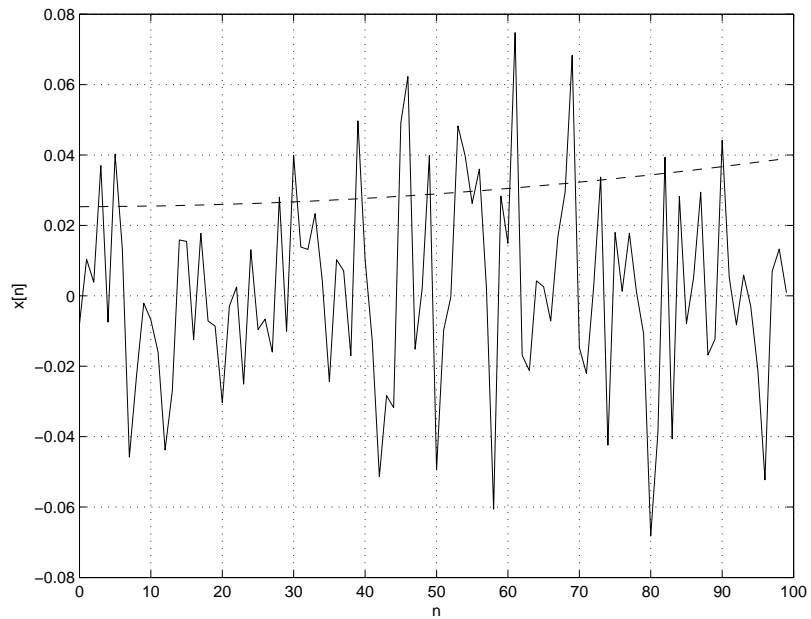


Figure 3: Typical realization for nonstationary WGN and time-varying standard deviation.

where from (4)

$$\frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{a}} = \begin{bmatrix} \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{11}} \\ \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{12}} \end{bmatrix}$$

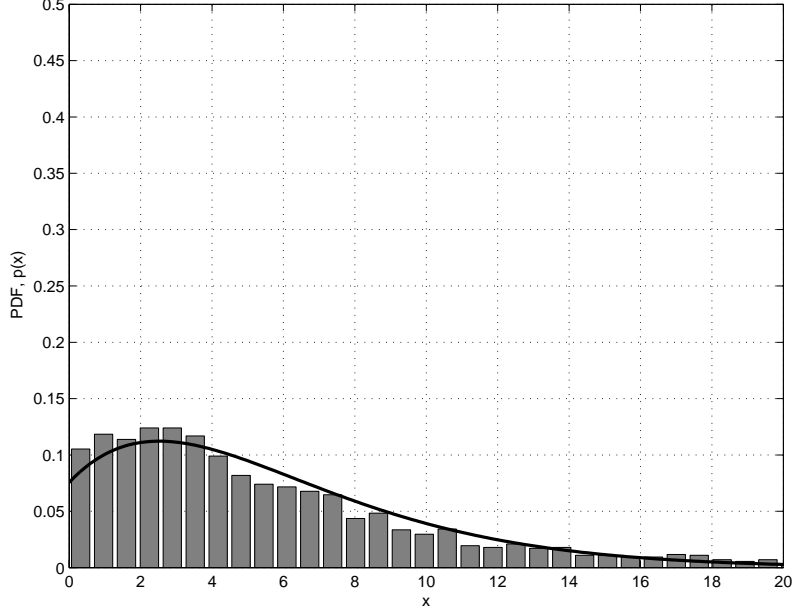


Figure 4: Estimated and theoretical PDF for $N = 100$ - nonstationary WGN.

and from (5)

$$\left. \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{rs}} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = - \sum_{n=1}^{N-1} \frac{\hat{u}[n] f_s[n-r] x[n-r]}{\hat{b}_0^2 c^2} \quad (18)$$

where $\hat{u}[n] = x[n] + \hat{a}_{10} c x[n-1]$. Since $p = 1$ the FIM is a 3×3 matrix of scalars. The FIM is from (7)

$$\mathbf{I}_{\mathbf{a}'\mathbf{a}'}(\boldsymbol{\theta}) = \begin{bmatrix} I_{\mathbf{a}'}(0,0) & I_{\mathbf{a}'}(0,1) & I_{\mathbf{a}'}(0,2) \\ I_{\mathbf{a}'}(1,0) & I_{\mathbf{a}'}(1,1) & I_{\mathbf{a}'}(1,2) \\ I_{\mathbf{a}'}(2,0) & I_{\mathbf{a}'}(2,1) & I_{\mathbf{a}'}(2,2) \end{bmatrix}$$

where the elements are from (8) with $k = l = 1$ and evaluated at $\boldsymbol{\theta}$

$$I_{\mathbf{a}'}(r,s) = [\mathbf{I}_{\mathbf{a}'}(r,s)]_{11} = \frac{r_x[0]}{b_0^2 c^2} \sum_{n=1}^{N-1} f_r[n-1] f_s[n-1].$$

Once more, by using the orthogonal basis functions we can assert that approximately (for $N \gg 1$) $\sum_{n=1}^{N-1} f_r[n-1] f_s[n-1] = \delta_{rs}$ and therefore

$$I_{\mathbf{a}'}(r,s) = \frac{r_x[0]}{b_0^2 c^2} \delta_{rs}$$

and as a result we have

$$\mathbf{I}_{\mathbf{a}'\mathbf{a}'}(\boldsymbol{\theta}) = \frac{r_x[0]}{b_0^2 c^2} \mathbf{I}_3.$$

Thus, it follows that

$$\left[\mathbf{I}_{\mathbf{a}'\mathbf{a}'}^{-1}(\boldsymbol{\theta}) \right]_{\mathbf{a}\mathbf{a}} = \frac{b_0^2 c^2}{r_x[0]} \mathbf{I}_2.$$

Since we are evaluating the latter matrix under \mathcal{H}_0 for which $x[n]$ is a stationary AR process of order one, we can use the result that $r_x[0] = b_0^2 c^2 / (1 - (a_{10}c)^2)$ [1]. This produces

$$\left[\mathbf{I}_{\mathbf{a}'\mathbf{a}'}^{-1}(\boldsymbol{\theta}) \right]_{\mathbf{a}\mathbf{a}} = \left(1 - (a_{10}c)^2 \right) \mathbf{I}_2 \quad (19)$$

which when evaluated for $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$ yields

$$\left[\mathbf{I}_{\mathbf{a}'\mathbf{a}'}^{-1}(\tilde{\boldsymbol{\theta}}) \right]_{\mathbf{a}\mathbf{a}} = \left(1 - (\hat{a}_{10}c)^2 \right) \mathbf{I}_2. \quad (20)$$

Finally, from (18) with $r = 1$ and $s = 1, 2$ we decide that the process is nonstationary if

$$T_N(\mathbf{x}) = \left(1 - (\hat{a}_{10}c)^2 \right) \sum_{j=1}^2 \left[\sum_{n=1}^{N-1} \frac{\hat{u}[n] f_j[n-1] x[n-1]}{\hat{b}_0^2 c^2} \right]^2 > \gamma$$

where $\hat{u}[n] = x[n] + \hat{a}_{10}c x[n-1]$. The estimates that are needed are for $a_{10}c$ and b_0c under \mathcal{H}_0 . Using the covariance method estimate [1] we have

$$\begin{aligned} \hat{a}_{10}c &= - \frac{\sum_{n=1}^{N-1} x[n] x[n-1]}{\sum_{n=1}^{N-1} x^2[n-1]} \\ \hat{b}_0^2 c^2 &= \frac{1}{N-1} \sum_{n=1}^{N-1} \hat{u}^2[n]. \end{aligned}$$

As before the asymptotic PDF of the Rao test statistic is χ_2^2 under \mathcal{H}_0 and $\chi_2^2(\lambda)$ under \mathcal{H}_1 . The noncentrality parameter is from (14) and (19)

$$\begin{aligned} \lambda &= \begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix}^T \left(\left[\mathbf{I}_{\mathbf{a}'\mathbf{a}'}^{-1}(\boldsymbol{\theta}_0) \right]_{\mathbf{a}\mathbf{a}} \right)^{-1} \begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix} \\ &= \frac{a_{11}^2 + a_{12}^2}{1 - (a_{10}c)^2}. \end{aligned}$$

As an illustration, under \mathcal{H}_0 we have that $a_{11} = a_{12} = 0$ and we choose $a_{10} = -0.8\sqrt{N}$. The estimated PDF and theoretical asymptotic PDF are shown in Figure 5 for $N = 1000$. Under \mathcal{H}_1 we let $a_{10} = -0.8\sqrt{N}$, $a_{11} = -0.0015N$, and $a_{12} = 0.000005N^{3/2}$ to produce the time-varying AR filter parameter shown in Figure 6, along with a typical realization of the process. The estimated PDF and the theoretical asymptotic PDF are shown in Figure 7. It can be seen that under either hypothesis the agreement is good. Note, however, that in order for the asymptotic PDF to hold we have had to increase the data record to $N = 1000$. Hence, unlike the previous case it has been found that the time variation of the AR filter parameter must be sufficiently slow for the asymptotic PDF to be valid. The Rao test can still be used for shorter data records although the exact performance under \mathcal{H}_1 will be difficult to quantify analytically. However, the PDF under \mathcal{H}_0 , which is needed to set the threshold and hence implement the test is quite accurate even for short data records of $N = 100$ samples.

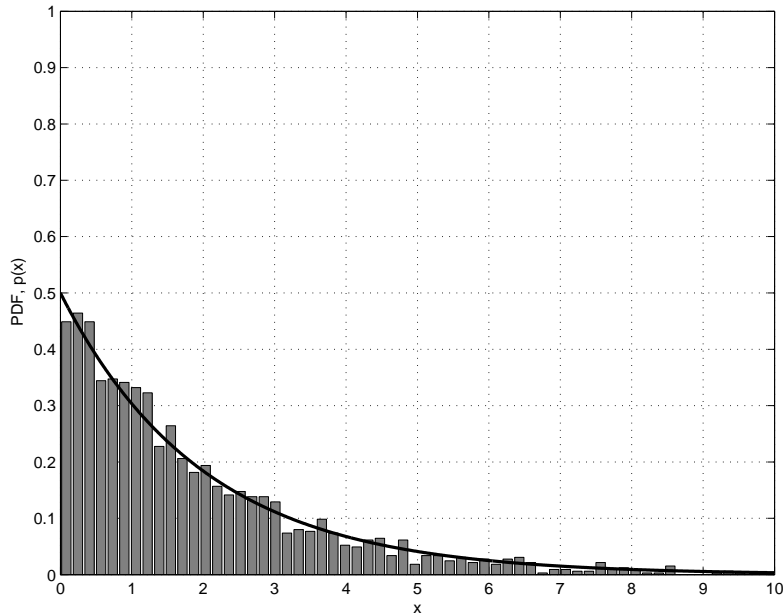


Figure 5: Estimated and theoretical PDF for $N = 1000$ - stationary AR process.

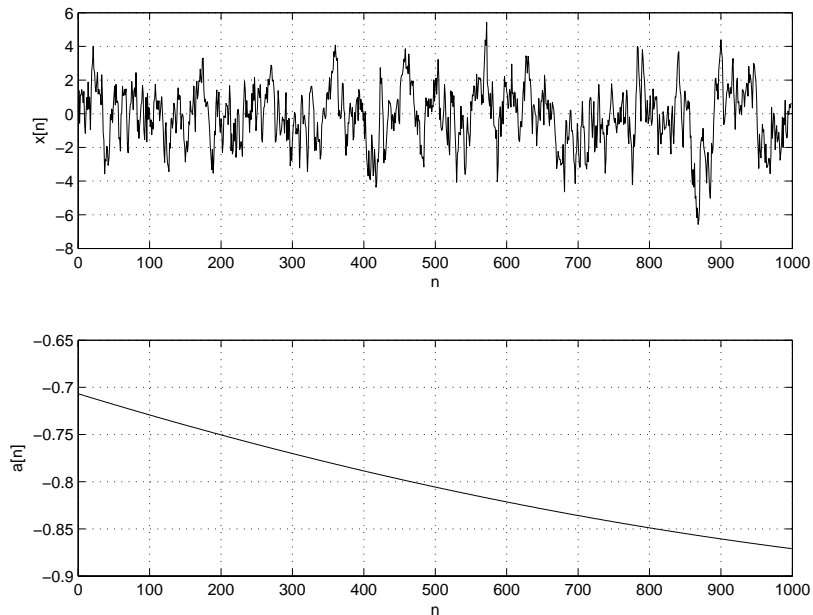


Figure 6: Typical realization of the nonstationary AR process and the time-varying AR filter parameter.

4 An Application to Real-time Detection - WGN process with time-varying power

In a practical implementation of the nonstationarity detector just described one might want to monitor the process in real-time. Hence, the Rao test statistic $T_N(\mathbf{x})$ would be computed for each value of N as

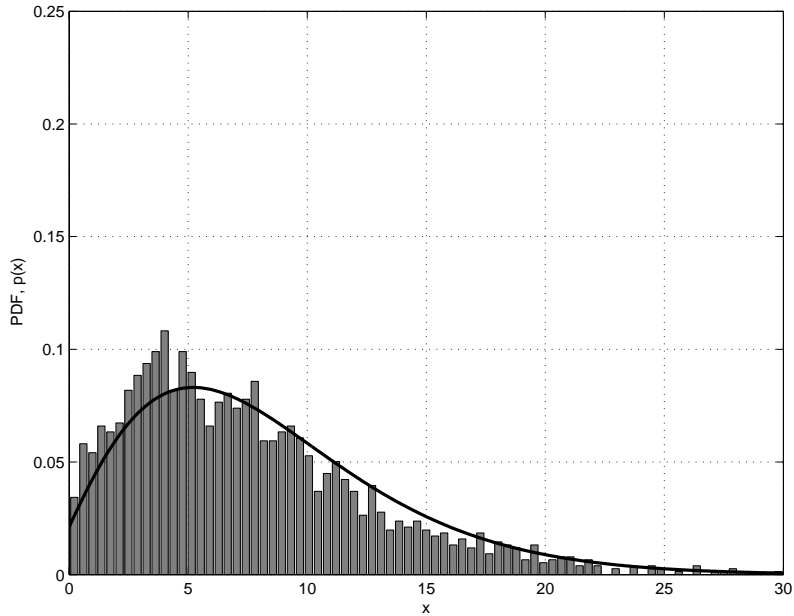


Figure 7: Estimated and theoretical PDF for $N = 1000$ - nonstationary AR process.

N increased to assess the maximum data record length possible for stationarity to hold. To illustrate this application assume that the Rao test statistic is computed for the first case examined, that of the WGN process with time-varying power. A realization of this process is shown in Figure 8 along with the standard deviation of the noise (shown as the dashed line). A gradual increase in the noise power is seen. Note that it is not until about $n = 150$ to $n = 200$ that the noise power appears to increase substantially. This is due to the influence of the linear and quadratic terms in the basis functions relative to the constant term, which is dominant for $n < 150$. It would be expected that for a reasonable probability of false alarm the threshold would be exceeded in this region. The Rao test statistic is shown in Figure 9 as a function of the data record length N . Also, is shown the threshold which produces a false alarm probability of $P_{FA} = 0.01$. It is given by $\gamma = 2\ln(1/P_{FA})$, which is found by noting that the test statistic under \mathcal{H}_0 is χ_2^2 . It should be observed that the false alarm probability is the probability that the threshold will be exceeded at a single value of N , based on the data record up to and including that data sample at time N , not the probability that there will be at least one false alarm up to that time. It is seen that by comparing Figures 8 and 9 the data record length during which the process is stationary appears to have been found with reasonable accuracy.

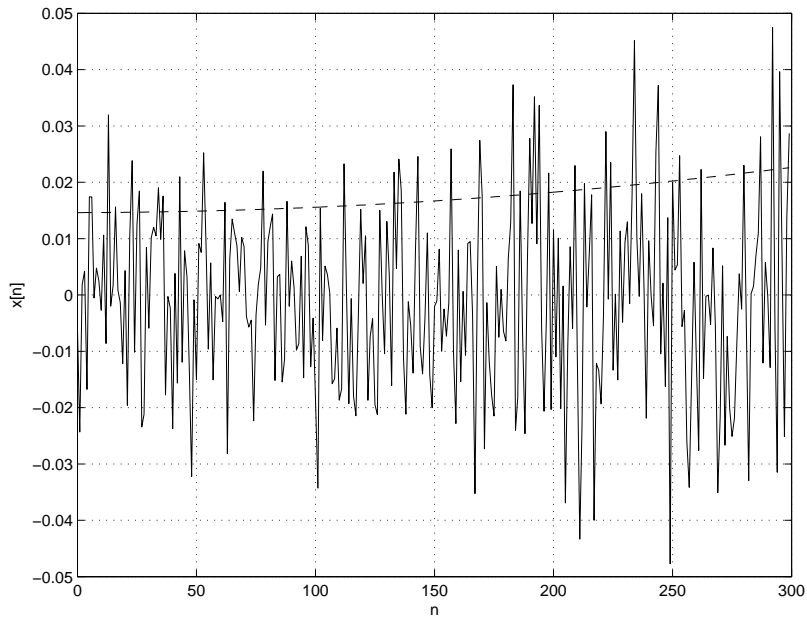


Figure 8: Typical realization for nonstationary WGN and time-varying standard deviation.

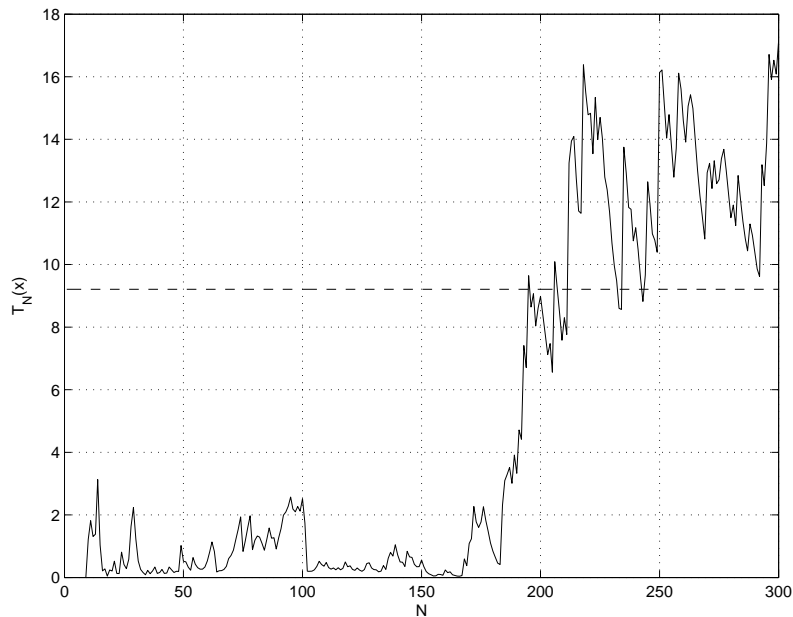


Figure 9: Rao test statistic as a function of data record length and threshold for $P_{FA} = 0.01$.

5 Discussion and Conclusions

A new test for stationarity of a WSS Gaussian random process has been introduced. It allows the user to determine how long a data record should be employed for a statistical analysis before a slowly varying nonstationarity will cause the results to be biased. To do so the user will be required to set a threshold,

as is customary in hypothesis testing, in order to assure a required probability of false alarm. Since this is problem dependent, we have omitted any discussion of this necessary step. The test is based in the time domain so that the usual asymptotic assumptions that must be made in the frequency domain are avoided. It is believed that this is responsible for the test achieving its predicted performance for relatively short data records. A future paper will explore this in more detail. Also, the use of a Rao test allows us to circumvent the nonlinear parameter estimation problem. As a byproduct of the approach, the test allows one to assess the stationarity interval of an AR random process. In speech or economics, for example, this is of great concern. One issue that has not been addressed and that warrants further study is the question of what model order to use for the TVAR model. Such techniques as the MDL [14] and the EEF [15] would appear to be possible solutions to the problem of automating the procedure and will be addressed later. Finally, it would appear that a sequential technique to be able to estimate the maximum data record length for stationarity would be important in practice. The current algorithm requires the user to recompute the statistic for each new data sample or block of data, if the test is applied to successive blocks.

References

1. Kay, S., *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs, NJ, 1988.
2. Priestley, M.B., *Spectral Analysis and Time Series*, Academic Press, New York, 1981.
3. Kay, S., *Fundamentals of Statistical Signal Processing: Detection Theory*, Prentice-Hall, Upper Saddle River, NJ, 1998.
4. Box, G.E.P., G.M. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, 1970.
5. Basseville, M., I.V. Nikiforov, *Detection of Abrupt Changes*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
6. Rao, T.S., "The Fitting of Non-Stationary Time Series Models with Time-Dependent Parameters", *J. Royal Stat. Soc.*, Vol. 32, pp. 312–322, 1970.
7. Liporace, L.A., "Linear Estimation of Nonstationary Signals", *J. Acoust. Soc. Am.*, Vol. 58, pp. 1288–1295, 1975.
8. Grenier, Y., "Time-Dependent ARMA Modeling of Nonstationary Signals", *IEEE Trans. on Signal Processing*, Vol. 31, pp. 899–911, Aug. 1983.

9. Priestley, M.B., T.S. Rao, “A Test for Non-Stationarity of Time-Series”, *J. Royal Stat. Soc.*, Vol. 31, pp. 1401–1449, 1969.
10. Brcich, R.F., D.R. Iskander, “Testing for Stationarity in the Frequency Domain using a Sphericity Statistic”, *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 464–467, 2006.
11. Epps, T.W., “Testing that a Gaussian Process is Stationary”, *Annals of Statistics*, Vol. 16, pp. 1667–1683, 1988.
12. Dalhaus, R., L. Giraitis, “On the Optimal Segment Length for Parameter Estimates for Locally Stationary Time Series”, *J. Time Series Analysis*, Vol. 19, pp. 629 –656, 1998.
13. Kwiatkowski,D., P.C.B. Phillips, P. Schmidt, Y. Shin, “Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root”, *J. of Econometrics*, Vol. 54, pp. 159–178, 1992.
14. Rissanen, J., “Modeling by Shortest Data Description”, *Automatica*, Vol. 14, pp. 465–471, 1978.
15. Kay, S., “Embedded Exponential Families: A New Approach to Model Order Selection”, *IEEE Trans. on Aerospace and Electronics*, Jan. 2005
16. Kay, S., *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Upper Saddle River, NJ, 1993.
17. Kendall, Sir M., A. Stuart, *The Advanced Theory of Statistics, Vol. 2*, Macmillan, New York, 1979.
18. Sharman, K.C., B. Friedlander, “Time-Varying Autoregressive Modeling of a Class of Nonstationary Signals”, *Proc. ICASSP*, pp. 22.2.1–22.2.4, New York, 1984.
19. Rajan, J.J., P.J.W. Rayner, “Generalized Feature Extraction for Time-Varying Autoregressive Models”, *IEEE Trans. on Signal Processing*, Vol. 44, pp. 2498–2507, 1996.
20. Harvey, A.C., *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge Univ. Press, New York, 1989.

A Derivation of the Rao Test and its Performance

It is assumed that the TVAR process is given by

$$x[n] = - \sum_{i=1}^p a_i[n-i]x[n-i] + b[n]w[n] \quad (21)$$

where $w[n]$ is WGN noise with unit variance and the time-varying AR parameters are

$$\begin{aligned} a_i[n] &= \sum_{j=0}^m a_{ij} f_j[n] \quad i = 1, 2, \dots, p \\ b[n] &= \sum_{j=0}^m b_j f_j[n]. \end{aligned}$$

Furthermore, $f_0[n] = c$, for c a constant. Hence, the Rao test will be testing whether $a_{ij} = 0$ for $i = 1, 2, \dots, p; j = 1, 2, \dots, m$ and $b_j = 0$ for $j = 1, 2, \dots, m$. This corresponds to \mathcal{H}_0 for which the AR process is stationary so that

$$x[n] = - \sum_{i=1}^p a_{i0} c x[n-i] + b_0 c w[n].$$

To set up the Rao test we define the total set of AR filter parameters by the $p \times (m+1)$ matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_0 & \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$$

where each \mathbf{a}_i is a column vector of dimension $p \times 1$ with $[\mathbf{A}]_{ij} = a_{ij} = [\mathbf{a}_j]_i$ and the excitation noise variance vector

$$\mathbf{b}' = \begin{bmatrix} b_0 & b_1 & \dots & b_m \end{bmatrix}^T = \begin{bmatrix} b_0 & \mathbf{b}^T \end{bmatrix}^T.$$

The entire parameter vector can now be written as

$$\boldsymbol{\theta} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \\ \mathbf{b} \\ \text{---} \\ \mathbf{a}_0 \\ b_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}_r \\ \text{---} \\ \boldsymbol{\theta}_s \end{bmatrix}$$

and thus we have the hypothesis test

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{\theta}_r = \boldsymbol{\theta}_{r_0} = \mathbf{0}, \boldsymbol{\theta}_s \\ \mathcal{H}_1 &: \boldsymbol{\theta}_r \neq \boldsymbol{\theta}_{r_0} = \mathbf{0}, \boldsymbol{\theta}_s \end{aligned}$$

where $\boldsymbol{\theta}_s$ can take on any values. For this hypothesis test the Rao test statistic is

$$T_R(\mathbf{x}) = \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r} \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}^T \left[\mathbf{I}^{-1}(\tilde{\boldsymbol{\theta}}) \right]_{\boldsymbol{\theta}_r, \boldsymbol{\theta}_r} \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r} \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}$$

where $\tilde{\boldsymbol{\theta}} = [\boldsymbol{\theta}_{r_0}^T = \mathbf{0}^T \quad \hat{\boldsymbol{\theta}}_{s_0}^T]^T$ is the MLE under \mathcal{H}_0 or assuming that $\boldsymbol{\theta}_r = \mathbf{0}$. We therefore need to determine the probability density function (PDF) $p(\mathbf{x}; \boldsymbol{\theta})$, its gradient, and the Fisher information matrix. We begin with the PDF.

A.1 PDF

Assuming observed samples of $x[n]$ for $n = 0, 1, \dots, N - 1$ we let $\mathbf{x} = [x[0] x[1] \dots x[N - 1]]^T$. Since from (21) we have $x[n] + \sum_{i=1}^p a_i[n - i]x[n - i] = b[n]w[n]$, we can transform from the $w[n]$'s to the $x[n]$'s for $n = p, p + 1, \dots, N - 1$ using

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_1[p] & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_2[p] & a_1[p + 1] & 1 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_p[p] & a_{p-1}[p + 1] & \dots & a_1[2p - 1] & 1 & \dots & 0 & 0 \\ 0 & a_p[p + 1] & a_{p-1}[p + 2] & \dots & a_1[2p] & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & a_p[N - 1 - p] & \dots & a_1[N - 2] & 1 \end{bmatrix}}_{\mathbf{T}} \underbrace{\begin{bmatrix} x[p] \\ x[p + 1] \\ \vdots \\ x[N - 1] \end{bmatrix}}_{\mathbf{x}'}$$

$$\approx \underbrace{\begin{bmatrix} b[p]w[p] \\ b[p + 1]w[p + 1] \\ \vdots \\ b[N - 1]w[N - 1] \end{bmatrix}}_{\mathbf{u}'}$$

It is an approximate transformation since we have set the initial conditions $\{x[0], x[1], \dots, x[p - 1]\}$ equal to zero to effect a one-to-one transformation from \mathbf{u}' to \mathbf{x}' . For $N \gg p$ this will be a good approximation. Since $\mathbf{u}' \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{u}'})$, where the covariance matrix is $\mathbf{C}_{\mathbf{u}'} = \text{diag}(b^2[p], b^2[p + 1], \dots, b^2[N - 1])$, we have

$$p_{\mathbf{U}'}(\mathbf{u}') = \frac{1}{(2\pi)^{(N-p)/2} \sqrt{\prod_{n=p}^{N-1} b^2[n]}} \exp \left[-\frac{1}{2} \sum_{n=p}^{N-1} \frac{u^2[n]}{b^2[n]} \right]. \quad (22)$$

Noting now that $\det(\mathbf{T}) = 1$, the PDF of \mathbf{x}' is just

$$p(\mathbf{x}'; \boldsymbol{\theta}) = \frac{1}{(2\pi)^{(N-p)/2} \sqrt{\prod_{n=p}^{N-1} b^2[n]}} \exp \left[-\frac{1}{2} \sum_{n=p}^{N-1} \frac{(x[n] + \sum_{i=1}^p a_i[n - i]x[n - i])^2}{b^2[n]} \right] \quad (23)$$

The log-likelihood function is therefore approximately

$$\ln p(\mathbf{x}'; \boldsymbol{\theta}) = c_1 - \frac{1}{2} \sum_{n=p}^{N-1} \ln b^2[n] - \frac{1}{2} \sum_{n=p}^{N-1} \frac{(x[n] + \sum_{i=1}^p a_i[n - i]x[n - i])^2}{b^2[n]} \quad (24)$$

where c_1 is a constant not dependent on $\boldsymbol{\theta}$.

A.2 Gradients and Expected Value of Second-order Partial Derivatives

First consider the partials with respect to b_r . Then, since $b[n] = \sum_{j=0}^m b_j f_j[n]$ we have from (24)

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial b_r} &= - \sum_{n=p}^{N-1} \frac{f_r[n]}{b[n]} - \frac{1}{2} \sum_{n=p}^{N-1} u^2[n] \left(\frac{-2}{b^3[n]} \right) f_r[n] \\ &= \sum_{n=p}^{N-1} \left[\frac{u^2[n] f_r[n]}{b^3[n]} - \frac{f_r[n]}{b[n]} \right] \quad r = 0, 1, \dots, m \end{aligned} \quad (25)$$

where we have let $u[n] = x[n] + \sum_{i=1}^p a_i[n-i]x[n-i]$. Next the partials with respect to a_{rs} are using $\partial u[n]/\partial a_{rs} = f_s[n-r]x[n-r]$

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{rs}} &= - \frac{1}{2} \sum_{n=p}^{N-1} \frac{2u[n] \partial u[n]/\partial a_{rs}}{b^2[n]} \\ &= - \sum_{n=p}^{N-1} \frac{u[n] f_s[n-r] x[n-r]}{b^2[n]} \quad r = 1, 2, \dots, p; s = 0, 1, \dots, m. \end{aligned} \quad (26)$$

The second-order partials are from (25)

$$\begin{aligned} \frac{\partial^2 \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial b_r \partial b_s} &= \sum_{n=p}^{N-1} \left[u^2[n] f_r[n] \left(-\frac{3}{b^4[n]} f_s[n] \right) + \frac{f_r[n]}{b^2[n]} f_s[n] \right] \\ &= \sum_{n=p}^{N-1} \left[\frac{1}{b^2[n]} - \frac{3u^2[n]}{b^4[n]} \right] f_r[n] f_s[n] \quad r = 0, 1, \dots, m; s = 0, 1, \dots, m \end{aligned} \quad (27)$$

and from (26)

$$\frac{\partial^2 \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{rs} \partial b_k} = 2 \sum_{n=p}^{N-1} \frac{u[n] f_s[n-r] x[n-r]}{b^3[n]} f_k[n] \quad r = 1, \dots, p; s = 0, \dots, m; k = 0, 1, \dots, m \quad (28)$$

and also from (26)

$$\frac{\partial^2 \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{rs} \partial a_{kl}} = - \sum_{n=p}^{N-1} \frac{f_s[n-r] x[n-r]}{b^2[n]} \underbrace{\frac{\partial u[n]}{\partial a_{kl}}}_{f_l[n-k] x[n-k]} \quad (29)$$

To compute the expected values for the Fisher information matrix we have from (27) with $E[u^2[n]] = b^2[n]$

$$-E \left[\frac{\partial^2 \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial b_r \partial b_s} \right] = 2 \sum_{n=p}^{N-1} \frac{f_r[n] f_s[n]}{b^2[n]}$$

which becomes upon evaluation at $\tilde{\boldsymbol{\theta}}$ since under \mathcal{H}_0 , $b[n] = b_0 c$

$$-E \left[\frac{\partial^2 \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial b_r \partial b_s} \right] \Bigg|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = 2 \sum_{n=p}^{N-1} \frac{f_r[n] f_s[n]}{b^2[n]} \Bigg|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = 2 \sum_{n=p}^{N-1} \frac{f_r[n] f_s[n]}{\hat{b}_0^2 c^2} \quad r = 0, \dots, m; s = 0, \dots, m \quad (30)$$

where \hat{b}_0 is the MLE of b_0 under \mathcal{H}_0 . Also, under \mathcal{H}_0 we have $E[u[n]x[n-r]] = 0$ for $r = 1, 2, \dots, p$. As a result, from (28)

$$E \left[\frac{\partial^2 \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{rs} \partial b_k} \right] = 0 \quad (31)$$

for $r = 1, 2, \dots, p$, $s = 0, 1, \dots, m$, and $k = 0, 1, \dots, m$. Clearly, evaluating this at $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$ also produces zero. Finally, from (29) under \mathcal{H}_0 we have

$$-E \left[\frac{\partial^2 \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{rs} \partial a_{kl}} \right] = \sum_{n=p}^{N-1} f_s[n-r] f_l[n-k] \frac{r_x[k-r]}{b_0^2 c^2}$$

where $r_x[k]$ is the autocorrelation sequence of $x[n]$ under \mathcal{H}_0 . Evaluating this term at $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$ yields

$$-E \left[\frac{\partial^2 \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{rs} \partial a_{kl}} \right] \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = \sum_{n=p}^{N-1} f_s[n-r] f_l[n-k] \frac{\hat{r}_x[k-r]}{\hat{b}_0^2 c^2} \quad r, k = 1, \dots, p; s, l = 0, \dots, m \quad (32)$$

A.3 Fisher Information Matrix

We reorder the parameters to take advantage of the block diagonal nature of the FIM. To do so define

$$\boldsymbol{\theta} = \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \\ \text{---} \\ b_0 \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} (m+1) \times p \\ \text{---} \\ (m+1) \times 1 \end{bmatrix}.$$

Recall that \mathbf{a}_i is just the i th column of \mathbf{A} . Because of (31) the FIM will be block diagonal with respect to the partitions of $\boldsymbol{\theta}$ given above. We need only determine the FIM for each of the partitions of $\boldsymbol{\theta}$. First consider $\mathbf{b}' = [b_0 \mathbf{b}^T]^T$, which is $(m+1) \times 1$. Define the partitioned FIM as

$$\begin{aligned} \mathbf{I}_{\mathbf{b}'\mathbf{b}'}(\boldsymbol{\theta}) &= \begin{bmatrix} I_{b'}(0,0) & | & I_{b'}(0,1) & \dots & I_{b'}(0,m) \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ I_{b'}(1,0) & | & I_{b'}(1,1) & \dots & I_{b'}(1,m) \\ \vdots & | & \vdots & \ddots & \vdots \\ I_{b'}(m,0) & | & I_{b'}(m,1) & \dots & I_{b'}(m,m) \end{bmatrix} \\ &= \begin{bmatrix} I_{b_0 b_0} & \mathbf{I}_{b_0 \mathbf{b}} \\ \mathbf{I}_{\mathbf{b} b_0} & \mathbf{I}_{\mathbf{b} \mathbf{b}} \end{bmatrix} = \begin{bmatrix} 1 \times 1 & 1 \times m \\ m \times 1 & m \times m \end{bmatrix}. \end{aligned}$$

For the Rao test we will need the (\mathbf{b}, \mathbf{b}) partition of the inverse of $\mathbf{I}_{\mathbf{b}'\mathbf{b}'}(\boldsymbol{\theta})$. This is given by

$$\left[\mathbf{I}_{\mathbf{b}'\mathbf{b}'}^{-1}(\boldsymbol{\theta})\right]_{\mathbf{bb}} = \left(\mathbf{I}_{\mathbf{bb}} - \mathbf{I}_{\mathbf{bb}_0} I_{b_0 b_0}^{-1} \mathbf{I}_{b_0 \mathbf{b}}\right)^{-1}.$$

When evaluated under \mathcal{H}_0 the elements of $\mathbf{I}_{\mathbf{b}'\mathbf{b}'}(\tilde{\boldsymbol{\theta}})$ become from (30)

$$I_{b'}(r, s) = \frac{2}{\hat{b}_0^2 c^2} \sum_{n=p}^{N-1} f_r[n] f_s[n]. \quad (33)$$

Similarly, we will partition

$$\mathbf{a}' = \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \end{bmatrix} = \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} p \times 1 \\ mp \times 1 \end{bmatrix}.$$

Each \mathbf{a}_i is a $p \times 1$ vector so that \mathbf{a}' is $(m+1)p \times 1$. The FIM is written in partitioned form as

$$\begin{aligned} \mathbf{I}_{\mathbf{a}'\mathbf{a}'}(\boldsymbol{\theta}) &= \begin{bmatrix} \mathbf{I}_{\mathbf{a}'}(0,0) & | & \mathbf{I}_{\mathbf{a}'}(0,1) & \dots & \mathbf{I}_{\mathbf{a}'}(0,m) \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ \mathbf{I}_{\mathbf{a}'}(1,0) & | & \mathbf{I}_{\mathbf{a}'}(1,1) & \dots & \mathbf{I}_{\mathbf{a}'}(1,m) \\ \vdots & | & \vdots & \ddots & \vdots \\ \mathbf{I}_{\mathbf{a}'}(m,0) & | & \mathbf{I}_{\mathbf{a}'}(m,1) & \dots & \mathbf{I}_{\mathbf{a}'}(m,m) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_{\mathbf{a}_0 \mathbf{a}_0} & \mathbf{I}_{\mathbf{a}_0 \mathbf{a}} \\ \mathbf{I}_{\mathbf{a} \mathbf{a}_0} & \mathbf{I}_{\mathbf{a} \mathbf{a}} \end{bmatrix} = \begin{bmatrix} p \times p & p \times mp \\ mp \times p & mp \times mp \end{bmatrix} \end{aligned} \quad (34)$$

where each submatrix in (34) is $p \times p$. For the Rao test we require the (\mathbf{a}, \mathbf{a}) partition of the inverse of $\mathbf{I}_{\mathbf{a}'\mathbf{a}'}(\boldsymbol{\theta})$, which is

$$\left[\mathbf{I}_{\mathbf{a}'\mathbf{a}'}^{-1}(\boldsymbol{\theta})\right]_{\mathbf{aa}} = \left(\mathbf{I}_{\mathbf{aa}} - \mathbf{I}_{\mathbf{aa}_0} \mathbf{I}_{\mathbf{a}_0 \mathbf{a}_0}^{-1} \mathbf{I}_{\mathbf{a}_0 \mathbf{a}}\right)^{-1}.$$

Note that the elements of $\mathbf{I}_{\mathbf{a}'\mathbf{a}'}(\boldsymbol{\theta})$ are given from (32) when evaluated at $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$

$$\begin{aligned} [\mathbf{I}_{\mathbf{a}'}(r, s)]_{kl} &= -E \left[\frac{\partial^2 \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{kr} \partial a_{ls}} \right] \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} \\ &= \frac{\hat{r}_x[k-l]}{\hat{b}_0^2 c^2} \sum_{n=p}^{N-1} f_r[n-k] f_s[n-l] \end{aligned}$$

for $k, l = 1, 2, \dots, p; r, s = 0, 1, \dots, m$.

A.4 Evaluated Gradients

To evaluate the gradient under \mathcal{H}_0 we use (25) with $b[n] = b_0 c$ and replace $\boldsymbol{\theta}$ by $\tilde{\boldsymbol{\theta}}$ to yield

$$\frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial b_r} \Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = \sum_{n=p}^{N-1} \frac{f_r[n]}{\hat{b}_0^3 c^3} (\hat{u}^2[n] - \hat{b}_0^2 c^2) \quad (35)$$

for $r = 1, 2, \dots, m$ and since under \mathcal{H}_0

$$u[n] = x[n] + \sum_{i=1}^p a_{i0} f_0[n-i] x[n-i]$$

where $f_0[n] = c$, we have at $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$

$$\hat{u}[n] = x[n] + \sum_{i=1}^p \hat{a}_{i0} c x[n-i] \quad (36)$$

for use in (35). Next we have under \mathcal{H}_0 using (26) with $b[n] = b_0 c$

$$\frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{rs}} = - \sum_{n=p}^{N-1} \frac{u[n] f_s[n-r] x[n-r]}{b_0^2 c^2}$$

which when evaluated at $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$ becomes

$$\left. \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{rs}} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = - \sum_{n=p}^{N-1} \frac{\hat{u}[n] f_s[n-r] x[n-r]}{\hat{b}_0^2 c^2}$$

where $\hat{u}[n]$ is given by (36) and $r = 1, 2, \dots, p; s = 1, 2, \dots, m$. To construct the gradient vector for use in the Rao test, we note the partitioned form

$$\frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{a}} = \begin{bmatrix} \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{a}_1} \\ \vdots \\ \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{a}_m} \end{bmatrix} = \begin{bmatrix} \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{11}} \\ \vdots \\ \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{p1}} \\ \text{---} \\ \vdots \\ \text{---} \\ \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{1m}} \\ \vdots \\ \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial a_{pm}} \end{bmatrix}.$$

A.5 Final Rao Statistic and Performance

Because of the block-diagonal nature of the FIM we have that

$$\begin{aligned} T_N(\mathbf{x}) &= \left. \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{a}} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}^T \left[\mathbf{I}_{\mathbf{a}'\mathbf{a}'}^{-1}(\tilde{\boldsymbol{\theta}}) \right]_{\mathbf{a}\mathbf{a}} \left. \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{a}} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} \\ &+ \left. \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{b}} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}^T \left[\mathbf{I}_{\mathbf{b}'\mathbf{b}'}^{-1}(\tilde{\boldsymbol{\theta}}) \right]_{\mathbf{b}\mathbf{b}} \left. \frac{\partial \ln p(\mathbf{x}'; \boldsymbol{\theta})}{\partial \mathbf{b}} \right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} \end{aligned}$$

where $\mathbf{a} = [\mathbf{a}_1^T \mathbf{a}_2^T \dots \mathbf{a}_m^T]^T$ and $\mathbf{b} = [b_1 \ b_2 \ \dots \ b_m]^T$. The distribution of $T_N(\mathbf{x})$ is $\chi_{m(p+1)}^2$ under \mathcal{H}_0 and $\chi_{m(p+1)}'^2(\lambda)$ under \mathcal{H}_1 , where λ is the noncentrality parameter given in general for $\boldsymbol{\theta}_{r_0} = \mathbf{0}$ by

$$\lambda = \boldsymbol{\theta}_{r_1}^T \left[\mathbf{I}_{\boldsymbol{\theta}_r \boldsymbol{\theta}_r}(\mathbf{0}, \boldsymbol{\theta}_s) - \mathbf{I}_{\boldsymbol{\theta}_r \boldsymbol{\theta}_s}(\mathbf{0}, \boldsymbol{\theta}_s) \mathbf{I}_{\boldsymbol{\theta}_s \boldsymbol{\theta}_s}^{-1}(\mathbf{0}, \boldsymbol{\theta}_s) \mathbf{I}_{\boldsymbol{\theta}_s \boldsymbol{\theta}_r}(\mathbf{0}, \boldsymbol{\theta}_s) \right] \boldsymbol{\theta}_{r_1}$$

where $\boldsymbol{\theta}_{r_1}$ is the value of $\boldsymbol{\theta}_r$ under \mathcal{H}_1 , and $\boldsymbol{\theta}_s$ is the true value, which is the same under either hypothesis. The matrix in brackets is actually a partition of the inverse FIM and is given as

$$\left(\left[\begin{array}{c|c} \mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} & \\ \hline \boldsymbol{\theta}_r, \boldsymbol{\theta}_r \end{array} \right] \right)^{-1} \Big|_{\boldsymbol{\theta}_r=0, \boldsymbol{\theta}_s}$$

To evaluate it we need to revert back to the original parameter ordering as

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_r \\ \boldsymbol{\theta}_s \end{bmatrix} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \text{---} \\ \mathbf{a}_0 \\ b_0 \end{bmatrix}.$$

With this partitioning the FIM becomes under \mathcal{H}_0

$$\mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \begin{bmatrix} \mathbf{I}_{\mathbf{a}\mathbf{a}} & \underline{\mathbf{I}_{\mathbf{a}\mathbf{b}}} & | & \mathbf{I}_{\mathbf{a}\mathbf{a}_0} & \underline{\mathbf{I}_{\mathbf{a}b_0}} \\ \underline{\mathbf{I}_{\mathbf{b}\mathbf{a}}} & \mathbf{I}_{\mathbf{b}\mathbf{b}} & | & \underline{\mathbf{I}_{\mathbf{b}\mathbf{a}_0}} & \mathbf{I}_{\mathbf{b}b_0} \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ \mathbf{I}_{\mathbf{a}_0\mathbf{a}} & \underline{\mathbf{I}_{\mathbf{a}_0\mathbf{b}}} & | & \mathbf{I}_{\mathbf{a}_0\mathbf{a}_0} & \underline{\mathbf{I}_{\mathbf{a}_0b_0}} \\ \underline{\mathbf{I}_{b_0\mathbf{a}}} & \mathbf{I}_{b_0\mathbf{b}} & | & \underline{\mathbf{I}_{b_0\mathbf{a}_0}} & \mathbf{I}_{b_0b_0} \end{bmatrix}$$

where the partitions that are underlined are zero due to (31). We require the $(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) = (1, 1)$ partition of the inverse of $\mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}$. To do so note that if

$$\mathbf{D} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{bmatrix}$$

and furthermore

$$\mathbf{D} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} & | & \mathbf{B}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} & | & \mathbf{0} & \mathbf{B}_{22} \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ \mathbf{B}_{11}^T & \mathbf{0} & | & \mathbf{C}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{22}^T & | & \mathbf{0} & \mathbf{C}_{22} \end{bmatrix}$$

then

$$\left[\mathbf{D}^{-1} \right]_{11} = (\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T)^{-1}$$

and therefore

$$\left(\left[\mathbf{D}^{-1} \right]_{11} \right)^{-1} = \mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T.$$

Since $\mathbf{A}, \mathbf{B}, \mathbf{B}^T, \mathbf{C}$ are all block-diagonal, we have that

$$\left([\mathbf{D}^{-1}]_{11} \right)^{-1} = \begin{bmatrix} \mathbf{A}_{11} - \mathbf{B}_{11} \mathbf{C}_{11}^{-1} \mathbf{B}_{11}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} - \mathbf{B}_{22} \mathbf{C}_{22}^{-1} \mathbf{B}_{22}^T \end{bmatrix}$$

and finally because $\mathbf{I}_{\mathbf{a}_0 \mathbf{a}} = \mathbf{I}_{\mathbf{a} \mathbf{a}_0}^T$ we have

$$\left([\mathbf{I}_{\theta\theta}^{-1}]_{\theta_r \theta_r} \right)^{-1} = \begin{bmatrix} \mathbf{I}_{\mathbf{a} \mathbf{a}} - \mathbf{I}_{\mathbf{a} \mathbf{a}_0} \mathbf{I}_{\mathbf{a}_0 \mathbf{a}_0}^{-1} \mathbf{I}_{\mathbf{a}_0 \mathbf{a}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\mathbf{b} \mathbf{b}} - \mathbf{I}_{\mathbf{b} b_0} \mathbf{I}_{b_0 b_0}^{-1} \mathbf{I}_{b_0 \mathbf{b}} \end{bmatrix}.$$

As a result, the noncentrality parameter is

$$\begin{aligned} \lambda &= \mathbf{a}^T \left(\mathbf{I}_{\mathbf{a} \mathbf{a}} - \mathbf{I}_{\mathbf{a} \mathbf{a}_0} \mathbf{I}_{\mathbf{a}_0 \mathbf{a}_0}^{-1} \mathbf{I}_{\mathbf{a}_0 \mathbf{a}} \right) \mathbf{a} + \mathbf{b}^T \left(\mathbf{I}_{\mathbf{b} \mathbf{b}} - \mathbf{I}_{\mathbf{b} b_0} \mathbf{I}_{b_0 b_0}^{-1} \mathbf{I}_{b_0 \mathbf{b}} \right) \mathbf{b} \\ &= \mathbf{a}^T \left([\mathbf{I}_{\mathbf{a}' \mathbf{a}'}^{-1}(\theta_0)]_{\mathbf{a} \mathbf{a}} \right)^{-1} \mathbf{a} + \mathbf{b}^T \left([\mathbf{I}_{\mathbf{b}' \mathbf{b}'}^{-1}(\theta_0)]_{\mathbf{b} \mathbf{b}} \right)^{-1} \mathbf{b} \end{aligned}$$

where \mathbf{a}, \mathbf{b} are the true values under \mathcal{H}_1 , and the FIMs are evaluated under \mathcal{H}_0 using $\mathbf{a} = \mathbf{0}$ and $\mathbf{b} = \mathbf{0}$, and also the true values of \mathbf{a}_0 and b_0 .

B The Rao Test - A Simplified Description

In this appendix we describe the essence of the Rao test and how it is possible to carry out a hypothesis test without estimating unknown parameters under the alternative hypothesis. We consider only the simplest case of an unknown scalar parameter θ and examine the test of whether $\theta = \theta_0$ under \mathcal{H}_0 or $\theta \neq \theta_0$ under \mathcal{H}_1 . The MLE $\hat{\theta}$ is the value of θ that maximizes $g(\theta) = \ln p(\mathbf{x}; \theta)$ (this is the log-likelihood function as shown in Figure 10) for a given data set $\mathbf{x} = [x[0] x[1] \dots x[N-1]]^T$. To test whether the parameter θ is a *known value* $\theta = \theta_0$ or not, a good hypothesis test relies on computing the normalized squared difference or

$$\frac{(\hat{\theta} - \theta_0)^2}{\text{var}(\hat{\theta})}$$

where the denominator normalizes the squared deviation of the estimated θ , i.e., the MLE $\hat{\theta}$, from the known value of θ , i.e., θ_0 . It can be shown that

$$\text{var}(\hat{\theta}) \approx \frac{1}{-\left. \frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \right|_{\theta=\theta_0}} = \frac{1}{-g''(\theta_0)}$$

if $\hat{\theta}$ is close to θ_0 . (The larger the curvature, as measured by the second derivative, the sharper $g(\theta)$ is and thus, the smaller will be the variance of the MLE). Hence, the desired test is based on

$$(\hat{\theta} - \theta_0)^2 |g''(\theta_0)| \tag{37}$$

since the second derivative of the log-likelihood function will be negative at the MLE $\theta = \hat{\theta}$ as shown in Figure 10. *To compute this, we usually require the MLE. However, the Rao test avoids this sometimes impossible task.* The Rao test approximates $g'(\theta)$ by a straight line as is usually done in a Newton-Raphson approach to find the zero of a function. Then, the zero of the linearized $g'(\theta)$ will be approximately at the MLE as shown in Figure 11. Hence, the value of the derivative of the log-likelihood function at $\theta = \hat{\theta}$ approximately satisfies

$$0 = g'(\hat{\theta}) \approx g'(\theta_0) + g''(\theta_0)(\hat{\theta} - \theta_0)$$

which yields

$$\hat{\theta} - \theta_0 = \frac{-g'(\theta_0)}{g''(\theta_0)}$$

and upon squaring yields

$$(\hat{\theta} - \theta_0)^2 = \frac{(g'(\theta_0))^2}{(g''(\theta_0))^2}$$

so that finally

$$(\hat{\theta} - \theta_0)^2 |g''(\theta_0)| = \frac{(g'(\theta_0))^2}{|g''(\theta_0)|} \quad (38)$$

which is just (37), the desired result. Note that the left-hand-side of (38) contains the MLE while the right-hand-side *does not*. In fact the right-hand-side only depends on θ_0 , *which is assumed known*. This is the essence of the Rao test. To show that (38) is indeed the Rao test (although a special case in which

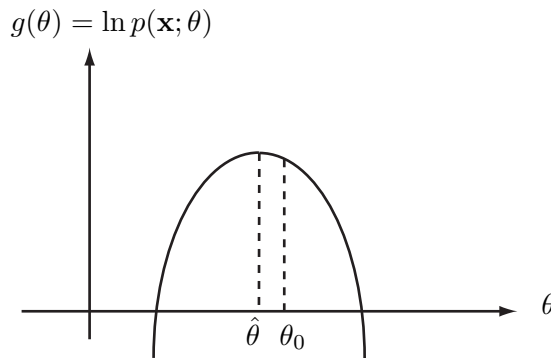


Figure 10: Log-likelihood function.

there are no nuisance paramters) we have upon rewriting it

$$\begin{aligned} \frac{(g'(\theta_0))^2}{|g''(\theta_0)|} &= \frac{\left(\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \Big|_{\theta=\theta_0} \right)^2}{\left| - \frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \Big|_{\theta=\theta_0} \right|} \\ &\approx \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \Big|_{\theta=\theta_0} I^{-1}(\theta_0) \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \Big|_{\theta=\theta_0} \end{aligned}$$

which should be compared with (3).

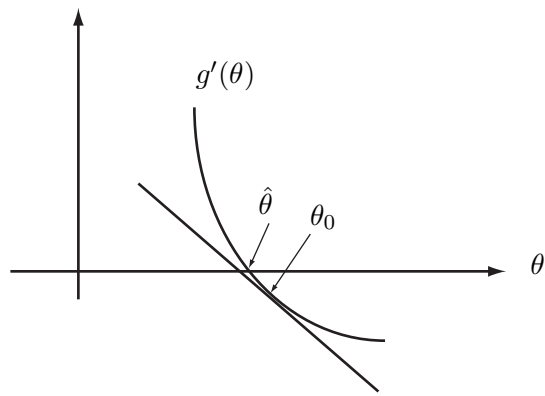


Figure 11: Derivative of log-likelihood function and tangent approximation near the MLE.