IIT Kanpur

MS Project(2016-17 2^{nd} sem)

On consistency of Exponentially Embedded Family (EEF) rule for 1-D and 2-D cisoid model and Interval estimation of model order using residual bootstrapping

Anupreet Porwal
(12817143)

Supervisors:
Prof. Amit Mitra
Prof. Sharmishtha Mitra

April 12, 2017

Certificate

It is certified that the work embodied in this project entitled " On consistency of Exponentially Embedded Family (EEF) rule for 1-D and 2-D cisoid model and Interval estimation of model order using residual bootstrapping" by Anupreet Porwal has been carried out under our supervision.

(Amit Mitra)
Professor
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

(Sharmishtha Mitra)
Associate Professor
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Declaration

I hereby declare that the work presented in this project entitled " On consistency of Exponentially Embedded Family (EEF) rule for 1-D and 2-D cisoid model and Interval estimation of model order using residual bootstrapping" contains my own ideas in my own words. At places, where ideas and words are borrowed from other sources, proper references, as applicable, have been cited. To the best of my knowledge this work does not emanate or resemble to other work created by person(s) other than mentioned herein.

This work was created on April 12, 2017.

-Anupreet Porwal

Acknowledgement

I want to express my sincere gratitude towards my supervisors Prof. Amit Mitra and Prof. Sharmishtha Mitra whose aspiring guidance, unfailing support and continuous efforts and encouragement helped me to learn through out the semester and overcome the difficulties faced throughout the project. I am gratefully indebted to them for helping me acquire a decent knowledge in the concerned topic.

-Anupreet Porwal
April 12, 2017

Chapter 1

Large sample consistency properties of Exponential Embedded Family (EEF) rule for model order selection of 1-D cisoid model in white noise

Large sample consistency properties of Exponential Embedded Family (EEF) rule for model order selection of 1-D cisoid model in white noise

***Abstract*— The problem of estimating the model order of a 1-D cisoid model in presence of white noise using a recently proposed EEF rule (see [1], [13]), is being considered. We prove that the EEF rule estimator is consistent in large sample scenario. We further demonstrate the finite sample performance of EEF rule based estimator for order selection with other popular order selection rules using simulation examples.**

I. INTRODUCTION

Model order selection is a pivotal problem in parametric approaches to signal processing. Model order is an important specification for a data model as it determines the complexity of the model and thus lays the foundation for estimation of real valued model parameters pertinent to the signal and noise component. A considerable amount of literature is available on various order estimation techniques (see [2]-[3] and the references therein), with two of the most widely used techniques being Akaike Information criteria (AIC) and Bayesian Information Criteria (BIC).

Recently [1] formulated a new rule to estimate model order called the Exponential Embedded Family (EEF) rule and discussed a Generalised Likelihood Ratio(GLR)- based derivation of the same. In this paper, we discuss EEF rule formulation for 1-D cisoid model in presence of white noise to estimate the model order and study its asymptotic properties.

We start our discussion with model description assumptions, followed by formulation of EEF rule for the concerned model. We then present lemmas that will help us to show consistency of order selection rule, followed by proof of consistency. We analyze the results of numerical simulation performed for various sample sizes and model parameters followed by conclusion.

II. Model Definition, Notations and Assumptions

A. Model Specifications

Cisoid Models are considered as building blocks of digital signal processing and estimation of its signal components have received considerable attention in the past few years. Formally, the model can be defined as

$\forall t = 1, 2, ..., n$

$$y_t = f(t, \theta_m) + \epsilon_t \tag{1}$$

$$f(t, \theta_m) = \sum_{k=1}^{m} \alpha_k e^{it\omega_k} \tag{2}$$

Therefore, the model becomes

$$y_t = \sum_{k=1}^{m} \alpha_k e^{it\omega_k} + \epsilon_t \tag{3}$$

where $\theta_m = (\alpha_{1_R}, \alpha_{1_C}, \omega_1, ..., \alpha_{m_R}, \alpha_{m_C}, \omega_m)'$ is a $3m$x1 vector of unknown signal parameters; α_{j_R} and α_{j_C} denotes the real and the imaginary part of α_j for j=1,..., m.

Let m_o be the true number of components in the the observed signal. Given a sample of size n, $y = (y_1, y_2, ..., y_n)'$, the model order selection problem will be to estimate m_o.

B. Model Assumptions

- ϵ_t are i.i.d complex valued gaussian with zero mean s.t.

$$\epsilon_t = \epsilon_{t_R} + i\epsilon_{t_C} \tag{4}$$

$$\epsilon_{t_R} \sim \mathcal{N}(0, \sigma^2/2) \tag{5}$$

$$\epsilon_{t_C} \sim \mathcal{N}(0, \sigma^2/2) \tag{6}$$

  and ϵ_{t_R} and ϵ_{t_C} are independent.
- $\forall k = 1, 2, ..., m_o : \omega_k \in (0, 2\pi)$; $\omega_j \neq \omega_k, \forall j \neq k$. Furthermore, $\forall k = 1, 2..., m_o : \alpha_k$ are bounded.
- The true model parameter vector θ_{m_o} is an interior point in the parameter space $\Theta \subset \mathbb{R}^{3m_o}$.

III. EEF rule formulation for cisoid model

[1] gives a general framework for formulating EEF rule. We devise the same for cisoid model using motivation from the above. In order to define the rules below, define

$$y = (y_1, y_2, ..., y_n)'$$
$$\epsilon = (\epsilon_1, \epsilon_2, ..., \epsilon_n)'$$
$$f(\theta_m) = (f(1, \theta_m), f(2, \theta_m), ..., f(n, \theta_m))';$$

For a m- component model, let $\theta_m^* = (\theta_m', \sigma_m^2)'$ denotes the vector containing the underlying signal and noise parameters, then the likelihood function of y under these assumptions can be written as

$$f_m(y, \theta_m^*) = \frac{1}{\left(\frac{2\pi\sigma_m^2}{2}\right)^n} e^{\frac{-(y-f(\theta_m))^H (y-f(\theta_m))}{\sigma_m^2}} \tag{7}$$

We consider the set of $\tilde{m}$ nested models given by $\{M_m\}_{m=1}^{\tilde{m}}$ where M_m is the m component cisoid model with parameter vector θ_m^*. We assume that the true model M_{m_o} is contained in this set .i.e., $m_o \leq \tilde{m}$ and y is not completely a white noise process i.e., $M_0 \neq M_{m_o}$.

Consider the Generalized likelihood ratio,

$$\hat{r}_{m+1} = 2 \ln \left[\frac{f_m(y, \hat{\theta}_m^*)}{f_0(y, \hat{\theta}_0^*)} \right] \tag{8}$$

where $\hat{\theta}_k^*$ is maximum likelihood estimate of underlying signal and noise parameter vector θ_k^* and $f_0(y, \hat{\theta}_0^*)$ denotes the p.d.f of y when M_0 is the model i.e., $f(\theta_0) = 0$.

The EEF rule for cisoid model mentioned above can be given by maximising the following statistic for $m = 1, 2, \dots, \tilde{m}$

$$EEF(m) = \begin{cases} \hat{r}_{m+1} - (3m+1)\left[1 + \ln\left(\frac{\hat{r}_{m+1}}{3m+1}\right)\right] & \text{if } \hat{r}_{m+1} > 3m+1 \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{m} = \underset{m \in \{1,2,\dots,\tilde{m}\}}{\arg\max} \, [EEF(m)] \tag{9}$$

where $\hat{m}$ is the estimated model order for the model.

IV. Consistency of EEF Rule

In order to prove the consistency of the rule, we prove the following lemma whose proofs are attached in the Appendix.

Lemma 1: Under the assumptions B1-B3, $\forall m \leq m_o$,

$$\hat{\sigma}^2_m = \sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j - \sum_{j=1}^{m} \hat{\alpha}_j^H \hat{\alpha}_j + o(1) \tag{10}$$

a.s. as $n \to \infty$.
Proof: See Appendix A.

Lemma 2: Under B1-B3, for any integer $k \geq 1$

$$\hat{\sigma}^2_{m_o+k} = \hat{\sigma}^2_{m_o} - \sum_{j=1}^{k} \frac{I_\epsilon(\hat{\omega}_{m_o+j})}{n} + o\left(\frac{\ln n}{n}\right) \text{ a.s. as } n \to \infty \tag{11}$$

or

$$\hat{\sigma}^2_{m_o+k} = \hat{\sigma}^2_{m_o} - \frac{G_k}{n} + o\left(\frac{\ln n}{n}\right) \text{ a.s. as } n \to \infty \tag{12}$$

where $G_k = \sum_{j=1}^{k} I_\epsilon(\hat{\omega}_{m_o+j})$
and

$$I_\epsilon(\omega) = \frac{1}{n}\left|\sum_{t=1}^{n} \epsilon_t e^{-it\omega}\right|^2 \tag{13}$$

$I_\epsilon(\omega)$ corresponds to periodogram of underlying white noise process and $\hat{\omega}_{m_o+1}, \hat{\omega}_{m_o+2}, ..., \hat{\omega}_{m_o+k}$ are the k largest frequencies corresponding to $I_\epsilon(\omega)$. Thus, G_k is the sum of k largest elements of the periodogram of noise.
Proof: See Appendix A.

Lemma 3: Under assumptions B1-B3, r_m as defined before in (8) satisfies

$$r_m = \begin{cases} 0 & m = 1 \\ O(n) & 2 \leq m \leq \tilde{m} \end{cases} \tag{14}$$

Proof: See Appendix A.

A. Proof of consistency

Theorem: Under the assumptions B1-B3, If m_o is the true model order and $\hat{m}$ is the estimated model order using EEF rule, then

$$P(\hat{m} \neq m_o) \to 0 \text{ as } \; n \to \infty \tag{15}$$

Proof:
Case I: $m \leq m_o$ **(Underestimation)**
Subcase I: $\hat{r}_{m+1} > 3m+1$;$\hat{r}_{m_o+1} > 3m_o+1$

$$\begin{aligned} &EEF(m) - EEF(m_o) = \hat{r}_{m+1} - \hat{r}_{m_o+1} \\ &- (3m+1)\ln(\hat{r}_{m+1}) + (3m_o+1)\ln(\hat{r}_{m_o+1}) + k \end{aligned} \tag{16}$$

where,

$$k = 3(m_o - m) + (3m+1)\ln(3m+1) - (3m_o+1)\ln(3m_o+1) \tag{17}$$

Note that, k is independent of sample size n. Also since

$$\hat{r}_{m+1} = 2n\ln\left(\frac{\hat{\sigma^2_o}}{\hat{\sigma^2_m}}\right) \tag{18}$$

then,

$$\begin{aligned} &EEF(m) - EEF(m_o) = -2n\ln\left(\frac{\hat{\sigma^2_m}}{\hat{\sigma^2_{m_o}}}\right) \\ &- (3m+1)\ln(\hat{r}_{m+1}) + (3m_o+1)\ln(\hat{r}_{m_o+1}) + k \end{aligned} \tag{19}$$

Also, using result of lemma 3, we get $\hat{r}_{m+1} = O(n)$ $\forall m \leq m_o$. Using lemma 1, we have $\forall m \leq m_o$

$$\hat{\sigma^2_m} \to \sigma^2 + \sum_{k=m+1}^{m_o} \alpha_k^H \alpha_k \text{ a.s.} \tag{20}$$

$$\hat{\sigma^2_{m_o}} \to \sigma^2 \tag{21}$$

as $n \to \infty$.Substituting values in the first part of the above expression we get $\forall m \leq m_o$

$$\ln\left(\frac{\hat{\sigma^2_m}}{\hat{\sigma^2_{m_o}}}\right) \to \ln\left(1 + \frac{\sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2}\right) \text{a.s. as } n \to \infty \tag{22}$$

which is a strictly positive and bounded quantity, substituting back in equation (19), we get

$$\begin{aligned} EEF(m) - EEF(m_o) = O(n) - \\ (3m+1)O(\ln n) + (3m_o+1)O(\ln n) + k \text{ a.s.} \end{aligned} \tag{23}$$

$$\text{or, } \frac{EEF(m) - EEF(m_o)}{n} \to -2\ln\left(1 + \frac{\sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2}\right) \tag{24}$$

a.s. as $n \to \infty$.

Thus, $EEF(m) < EEF(m_o)$ with probability 1.

Subcase 2 : $\hat{r}_{m+1} < 3m+1$;$\hat{r}_{m_o+1} > 3m_o+1$

$$EEF(m) = 0 \tag{25}$$

$$EEF(m_o) = \hat{r}_{m_o+1} - (3m_o+1)\left[1 + \ln\left(\frac{\hat{r}_{m_o+1}}{3m_o+1}\right)\right] \tag{26}$$

as the above expression is of form $g(x) = x - \ln x - 1$ which has a unique minimum value of 0 at x=1, $EEF(m_o) > 0$ for $\hat{r}_{m_o+1} \neq 3m_o+1$. Thus $EEF(m_o) > EEF(m) \forall m \leq m_o$ and hence underestimation is not possible in this case.

Subcase 3 : $\hat{r}_{m+1} > 3m+1$;$\hat{r}_{m_o+1} < 3m_o+1$

We know from [1], for large samples, EEF rule looks as the following

$$EEF(m) \cong \hat{r}_{m+1} - (3m+1)\ln n \tag{27}$$

Since $\hat{r}_{m_o+1} < 3m_o+1$, $EEF(m_o) = 0$.Thus,

$$\frac{EEF(m) - EEF(m_o)}{n\ln n} = \frac{\hat{r}_{m+1}}{n\ln n} - \frac{3m+1}{n} \text{ as } n \to \infty \tag{28}$$

Since we know from Lemma 3, $\hat{r}_{m+1} = O(n)$

$$\frac{EEF(m) - EEF(m_o)}{n \ln n} \to 0 \text{ as } n \to \infty \tag{29}$$

$$\text{or, } P\Big(\frac{EEF(m) - EEF(m_o)}{n \ln n} > 0\Big) \to 0 \text{ as } n \to \infty. \tag{30}$$

Thus for all the subcases discussed above,

$$\begin{aligned} &P(\hat{m} < m_o) \\ &= P(EEF(m) > EEF(m_o) \text{ for some } m < m_o) \\ &\to 0 \text{ as } n \to \infty \end{aligned} \tag{31}$$

Case 2: $m > m_o$ **(Overestimation)**
Subcase 1: $\hat{r}_{m+1} > 3m + 1$;$\hat{r}_{m_o+1} > 3m_o + 1$

$$\begin{aligned} &EEF(m) - EEF(m_o) = -2n \ln\Big(\frac{\hat{\sigma}^2_m}{\hat{\sigma}^2_{m_o}}\Big) \\ &- (3m + 1) \ln(\hat{r}_{m+1}) + (3m_o + 1) \ln(\hat{r}_{m_o+1}) + k \end{aligned} \tag{32}$$

Since $m > m_o$, $\hat{r}_{m+1} = 2 \ln\Big(\frac{\hat{f}_m}{\hat{f}_o}\Big) > 2 \ln\Big(\frac{\hat{f}_{m_o}}{\hat{f}_o}\Big) = \hat{r}_{m_o+1}$, using which we have

$$\begin{aligned} EEF(m) - EEF(m_o) \le -2n \ln\Big(\frac{\hat{\sigma}^2_m}{\hat{\sigma}^2_{m_o}}\Big) \\ + k + 3(m_o - m) \ln(\hat{r}_{m+1}) \end{aligned} \tag{33}$$

It follows from the asymptotic theory of likelihood ratios (see [5], [4]) that,

$$2n \ln\Bigg(\frac{\hat{\sigma}^2_m}{\hat{\sigma}^2_{m_o}}\Bigg) \sim \chi^2_{3(m-m_o)} \tag{34}$$

Thus, $\frac{1}{\ln(n)}\Bigg(2n \ln\Big(\frac{\hat{\sigma}^2_m}{\hat{\sigma}^2_{m_o}}\Big)\Bigg) = o_p(1)$.Since k is independent of n and $m > m_o$, and from lemma 3 we have $\hat{r}_{m+1} = O(n)$, we get

$$\begin{aligned} &P\Bigg(\frac{1}{\ln(n)}\Bigg(-2n \ln\Big(\frac{\hat{\sigma}^2_m}{\hat{\sigma}^2_{m_o}}\Big) + k+ \\ &3(m_o - m) \ln(r_{\hat{m+1}})\Bigg) < 0\Bigg) \to 1 \text{ as } n \to \infty \end{aligned} \tag{35}$$

that implies from (33)

$$P\Bigg(\frac{EEF(m) - EEF(m_o)}{\ln n} < 0\Bigg) \to 1 \text{ as } n \to \infty \tag{36}$$

Therefore,

$$\begin{aligned} &P(\hat{m} > m_o) \\ &= P(EEF(m) - EEF(m_o) > 0 \text{ for some } m > m_o) \\ &= P\Bigg(\frac{EEF(m) - EEF(m_o)}{\ln n} > 0\Bigg) \to 0 \text{ as } n \to \infty \end{aligned} \tag{37}$$

Subcase 2 : $\hat{r}_{m+1} < 3m + 1$;$\hat{r}_{m_o+1} > 3m_o + 1$

$$EEF(m) = 0 \tag{38}$$

$$EEF(m_o) = \hat{r}_{m_o+1} - (3m_o + 1)\Bigg[1 + \ln\Big(\frac{\hat{r}_{m_o+1}}{3m_o + 1}\Big)\Bigg] \tag{39}$$

as the above expression is of form $g(x) = x - \ln x - 1$ which has a unique minimum value of 0 at x=1, $EEF(m_o) > 0$ for $\hat{r}_{m_o+1} \neq 3m_o + 1$. Thus $EEF(m_o) > EEF(m) \forall m \ge m_o$ and hence overestimation is not possible in this case.

Subcase 3 : $\hat{r}_{m+1} > 3m + 1$;$\hat{r}_{m_o+1} < 3m_o + 1$

Similar to the underestimation approach, We know from [1], for large samples, EEF rule looks as the following

$$EEF(m) \cong \hat{r}_{m+1} - (3m + 1) \ln n \tag{40}$$

Since $\hat{r}_{m_o+1} < 3m_o + 1$, $EEF(m_o) = 0$.Thus,

$$\frac{EEF(m) - EEF(m_o)}{n \ln n} = \frac{\hat{r}_{m+1}}{n \ln n} - \frac{3m + 1}{n} \text{ as } n \to \infty \tag{41}$$

Since we know from Lemma 3, $\hat{r}_{m+1} = O(n)$

$$\frac{EEF(m) - EEF(m_o)}{n \ln n} \to 0 \text{ as } n \to \infty \tag{42}$$

$$\text{or, } P\Big(\frac{EEF(m) - EEF(m_o)}{n \ln n} < 0\Big) \to 1 \text{ as } n \to \infty. \tag{43}$$

Thus for all the subcases discussed above,

$$\begin{aligned} P(\hat{m} < m_o) = P(EEF(m) < EEF(m_o) \text{ for some } m < m_o) \\ \to 0 \text{ as } n \to \infty \end{aligned} \tag{44}$$

Hence, from the two cases of over and under estimation we have that

$$P(\hat{m} \neq m_o) \to 0 \text{ as } \ n \to \infty \tag{45}$$

V. Numerical Simulation

In this section, we present results of numerical simulations that we run to analyze the performance of model order estimate based on EEF rule. We consider the following model:

$$y_t = \sum_{k=1}^{3} \alpha_k e^{it\omega_k} + \epsilon_t \tag{46}$$

$$\begin{aligned} \alpha_1 = 3 + 2i \qquad \alpha_2 &= 2 + 1.66i \qquad \alpha_3 = 1.75 + i \\ \omega_1 = 0.8\pi \qquad \omega_2 &= 1.2\pi \qquad \omega_3 = 1.4\pi \end{aligned}$$

where, ϵ_t are i.i.d complex valued gaussian with zero mean s.t.

$$\begin{aligned} \epsilon_t &= \epsilon_{t_R} + i\epsilon_{t_C} \\ \epsilon_{t_R} &\sim \mathcal{N}(0, \sigma^2/2) \\ \epsilon_{t_C} &\sim \mathcal{N}(0, \sigma^2/2) \end{aligned}$$

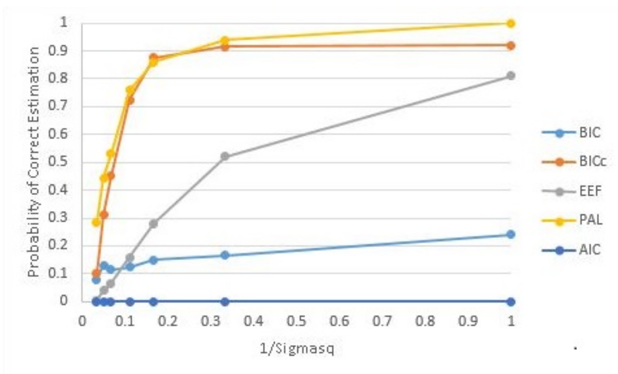

Fig. 1. Probability of correct order estimation; n=25

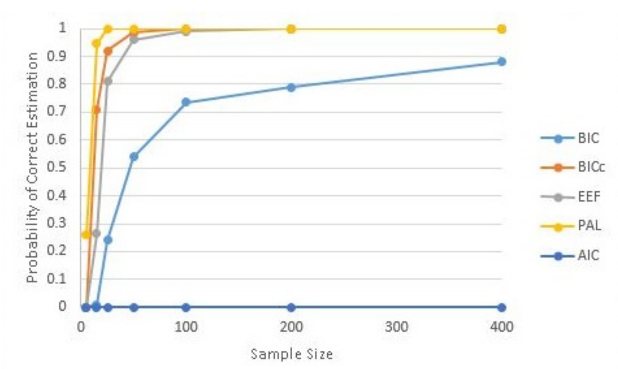

Fig. 2. Probability of correct order estimation;$\sigma^2 = 1$

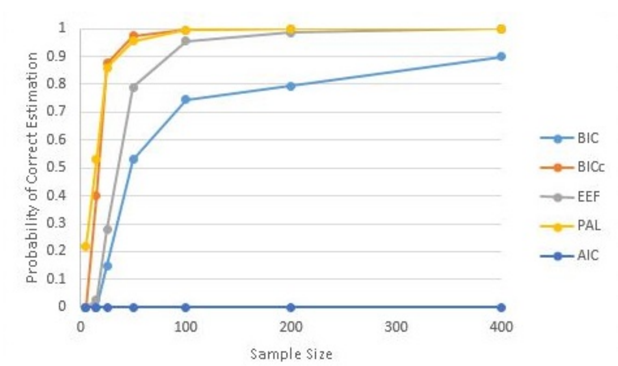

Fig. 3. Probability of correct order estimation; $\sigma^2 = 6$

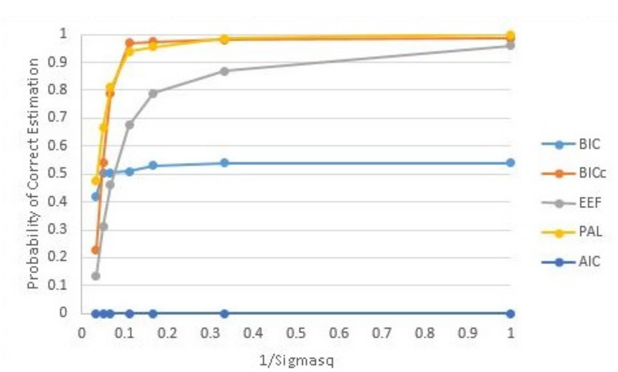

Fig. 4. Probability of correct order estimation; n=50

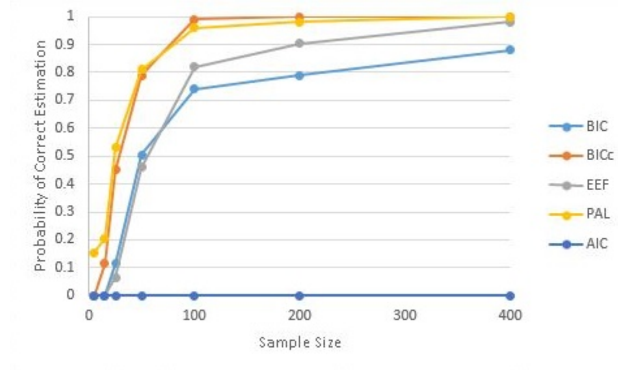

Fig. 5. Probability of correct order estimation;$\sigma^2 = 15$

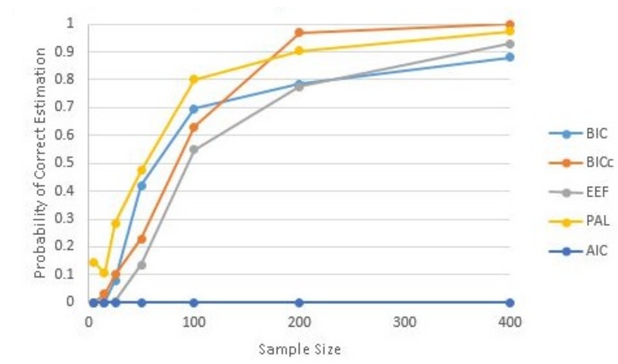

Fig. 6. Probability of correct order estimation; $\sigma^2 = 30$

and ϵ_{t_R} and ϵ_{t_C} are independent $\forall t = 1 \ldots n$. For the above model, we vary σ^2 from 0.5 to 30 and vary the sample size n from 5 to 400. We keep the maximum model order fixed at $\tilde{m} = 10$. We calculate the Probability of correct estimation for the various rules, for 200 simulation run. We evaluate the performance of EEF rule in comparison to other popular model selection rule for the above 1-D cisoid model, namely, Bayesian Information Criterion(see, [2], [14]), Akaike Information crtierion ([2], [14]), corrected BIC (see, [2]) and model order selection based on Penalizing adaptively the likelihood (see [16]). The forms of the following rules given the model of order m is given as:

$$PAL(m) = -2\ln(f_m(y, \hat{\theta}_m^*)) + (3m+1)\ln(3\tilde{m}+1) * \frac{\ln(r_m + 1)}{\ln(\rho_m + 1)} \tag{47}$$

$$BIC(m) = -2\ln(f_m(y, \hat{\theta}_m^*)) + (3m+1)\ln n \tag{48}$$

$$BICc(m) = -2\ln(f_m(y, \hat{\theta}_m^*)) + (5m+1)\ln n \tag{49}$$

$$AIC(m) = -2\ln(f_m(y, \hat{\theta}_m^*)) + 2 * (3m+1) \tag{50}$$

where, usual notations defined above are followed and ρ_m is defined as

$$\rho_m = 2\ln\left[\frac{f_{\tilde{m}}(y, \hat{\theta}_{\tilde{m}}^*)}{f_{m-1}(y, \hat{\theta}_{m-1}^*)}\right] \tag{51}$$

and the usual notations for the rest of the terms as defined above in the paper. We understand from the above figures that EEF rule performs well for high signal to noise (SNR) ratio and low sample sizes. The rule particularly outperforms BIC and AIC and its performance is comparable to corrected form of BIC and PAL rule. It is interesting to note that corrected BIC is asymptotically equivalent to Maximum-a posteriori (MAP) rule and PAL rule, due to its "oracle like" properties as discussed in [16], is one of the most preferable rules in comparison to latter rules in simulation tests. We understand that probability of correct order estimation using the EEF rule approaches limiting value of one for as n=50 for $\sigma^2 = 0.5$ and reaches one at a comparable rate when compared to other rules taken in the simulation even for a high σ^2 value of 15. Further, we infer that all rules except AIC have increasing probability of correct estimation with sample size and SNR values.

VI. Conclusions

In this paper, we discussed the EEF rule formulation for 1-D cisoid model in presence of white noise and proved the large sample consistency properties for the same. We also, illustrated the EEF rule which verifies the excellent performance of the rule when compared with other standard rules like AIC, BIC, corrected form of BIC and a novel rule with data adaptive penalty called PAL rule, even for small small sample sizes and low SNR values. We thus believe that EEF rule can be casted in main stream of order selection particular for signal processing models like this one. It would thus be interesting to look at extending the consistency of EEF rule various other popular nested non linear models in the time series and signal processing literature with more liberal assumptions. However, extra results may be needed to be derived depending on the model under consideration.

Appendix

In order to show the consistency of the rule, we look at the behaviour of $\hat{\sigma}_m^2$ as sample size tends to infinity for both the under estimation and overestimation cases which helps us in dealing with understanding the behaviour of likelihood terms and ratios involved with the rule. All the proofs of various lemmas fall on the same line as done for real 1-D sinusoidal models in [3]. Before we start proving the lemmas mentioned in the paper, we define a few notations. Let us write the model in (3) in matrix form using the following definitions:
$\forall j = 1, 2, ..., m$

$$e_j = (e^{i\omega_j}, e^{i2\omega_j}, ..., e^{in\omega_j})' \tag{52}$$

$$\omega = (\omega_1, \omega_2, ..., \omega_m)' \tag{53}$$

$$D_m(\omega) = (e_1, e_2, ..., e_m)_{nxm} \tag{54}$$

$$a_m = (\alpha_1, \alpha_2, ..., \alpha_m)'_{mx1} \tag{55}$$

Under the above notation model in (3) can be written as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} e^{i\omega_1} & e^{i\omega_2} & \dots & e^{i\omega_m} \\ e^{i2\omega_1} & e^{i2\omega_2} & \dots & e^{i2\omega_m} \\ \vdots & \vdots & \ddots & \vdots \\ e^{in\omega_1} & e^{in\omega_2} & \dots & e^{in\omega_m} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{56}$$

or, with the notation defined above

$$y = D_m(\omega)a_m + \epsilon \tag{57}$$

For notational simplicity, we write D_m for $D_m(\omega)$ and $\hat{D_m}$ for $D_m(\hat{\omega})$.Note that in the following proofs, $O(.)$ and $o(.)$ denote either the order in probability or deterministic order depending on the context

The proofs of the three theorems can be seen below:

Proof: (Lemma 1) The estimated variance of noise from the model in (57) can be written as:

$$\hat{\sigma}_m^2 = \frac{y^H(I_n - P_m(\hat{\omega}))y}{n} \tag{58}$$

$$\hat{\sigma}_m^2 = \frac{y^H y}{n} - \frac{y^H P_m(\hat{\omega}) y}{n} \tag{59}$$

where $P_m(\hat{\omega}) = \hat{D_m}(\hat{D_m}^H\hat{D_m})^{-1}\hat{D_m}^H$ is the projection matrix. Substituting value of y from true model $y = D_{m_o}a_{m_o} + \epsilon$ and expanding gives:

$$\hat{\sigma}_m^2 = \frac{1}{n}\Bigg[\epsilon^H\epsilon + 2\epsilon^H D_{m_o}a_{m_o} + a_{m_o}^H D_{m_o}^H D_{m_o}a_{m_o} \\ - \hat{a}_{m_o}^H \hat{D}_{m_o}^H \hat{D}_{m_o}\hat{a}_{m_o}\Bigg] \tag{60}$$

where $\hat{a}_m = (\hat{D_m}^H\hat{D_m})^{-1}\hat{D_m}^H y$. Let us analyze each of the terms one by one

$$\frac{\epsilon^H\epsilon}{n} = \frac{1}{n}\sum_{t=1}^{n}(\epsilon_{t_R}^2 + \epsilon_{t_C}^2) \tag{61}$$

$$= \frac{\sum_{t=1}^n \epsilon_{t_R}^2}{n} + \frac{\sum_{t=1}^n \epsilon_{t_C}^2}{n} \tag{62}$$

$$\tag{63}$$

Since both ϵ_{t_R} and ϵ_{t_C} are independent, by Kolmogorov's Strong law of Large Numbers(see e.g.[6])

$$\frac{\epsilon^H\epsilon}{n} = \frac{\sum_{t=1}^n \epsilon_{t_R}^2}{n} + \frac{\sum_{t=1}^n \epsilon_{t_C}^2}{n} \to \frac{\sigma^2}{2} + \frac{\sigma^2}{2} = \sigma^2 \tag{64}$$

a.s. as $n \to \infty$.

$$\frac{\epsilon^H D_{m_o}a_{m_o}}{n} = \frac{1}{n}\sum_{k=1}^{n}\left(\bar{\epsilon}_k\left(\sum_{j=1}^{m_o}\alpha_j e^{ik\omega_j}\right)\right) \tag{65}$$

$$= \frac{1}{n}\left(\sum_{j=1}^{m_o}\alpha_j\left(\sum_{k=1}^{n}\bar{\epsilon}_k e^{ik\omega_j}\right)\right) \tag{66}$$

Using lemma 1 of Kundu and Mitra [7], [8],

$$\frac{\epsilon^H D_{m_o}a_{m_o}}{n} = o(1) \text{ a.s.} \tag{67}$$

The third term of (60) can be shown the following behaviour

$$\frac{1}{n} a_{m_o}^H D_{m_o}^H D_{m_o} a_{m_o} = \sum_{j=1}^{m_o} \alpha_j^H \alpha_j + o(1) \text{ a.s. as } n \to \infty \tag{68}$$

using the fact that $\lim_{n\to\infty} \frac{\sum_{t=1}^{n} e^{it\tilde{\omega}}}{n} = o(1)$ for some fixed $\tilde{\omega} \in (0, 2\pi)$. Similarly, $\hat{\theta}_m$ is M.L.E. of θ_m, we have

$$\hat{a}_{m_o}^H \hat{D}_{m_o}^H \hat{D}_{m_o} \hat{a}_{m_o} = \sum_{j=1}^{m} \hat{\alpha}_j^H \hat{\alpha}_j + o(1) \text{ a.s. as } n \to \infty \tag{69}$$

Thus using (64), (67), (68) and (69), we hae $\forall m \leq m_o$

$$\hat{\sigma}_m^2 = \sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j - \sum_{j=1}^{m} \hat{\alpha}_j^H \hat{\alpha}_j + o(1) \text{ a.s. as } n \to \infty \tag{70}$$

The above lemma thus tells us about the way $\hat{\sigma}_m^2$ varies asymptotically when m is underestimated. We now look at the behaviour of $\hat{\sigma}_m^2$ when the estimated model order m is greater than the true model order m_o.

Proof:(Lemma 2) For l=1,2,..., k let

$$\hat{\alpha}_{m_o+l} = \frac{1}{n} \sum_{t=1}^{n} \epsilon_t e^{-it\hat{\omega}_{m_o+l}} \tag{71}$$

$$\hat{D}_{m_o+k} = (\hat{D}_{m_o}, \hat{e}_{m_o+1}, \hat{e}_{m_o+2}, \ldots, \hat{e}_{m_o+k}) \tag{72}$$

$$\hat{a}_{m_o+k} = (\hat{a}_{m_o}', \hat{\alpha}_{m_o+1}, \hat{\alpha}_{m_o+2}, \ldots, \hat{\alpha}_{m_o+k})' \tag{73}$$

Using result presented in theorem 2 of [9] (see also [10], [11]), we note that in the overestimated scenario, M.L.E of the parameter vector contains a sub vector equal in dimension to the true model order, that converges almost surely to the true parameter vector. Also, the frequencies of the over estimated components are those that maximise the noise periodogram. Thus for the variance of the over estimated model can be given as,

$$\hat{\sigma}_{m_o+k}^2 = \frac{y^H (I_n - P_{m_o+k}(\hat{\omega}_{m_o+k})) y}{n} \tag{74}$$

$$\tag{75}$$

Substituting the values of $\hat{a}_{m_o+k}$ and $\hat{D}_{m_o+k}$ in the above expressions and simplifying we get

$$\hat{\sigma}_{m_o+k}^2 = \hat{\sigma}_{m_o}^2 - T_1 - T_2 \tag{76}$$

where,

$$\hat{\sigma}_{m_o}^2 = \frac{1}{n} (y^H y - \hat{a}_{m_o}^H \hat{D}_{m_o}^H \hat{D}_{m_o} \hat{a}_{m_o}) \tag{77}$$

$$T_1 = 2 \sum_{k=1}^{k} T_{1_j} \text{ and } T_2 = \sum_{j=1}^{k} T_{2_j} + 2 \sum_{j=1}^{k} \sum_{\substack{\tilde{j}=1 \\ j \neq \tilde{j}}}^{k} T_{2(j,\tilde{j})} \tag{78}$$

$$T_{1_j} = \frac{1}{n} \delta_j \Delta_j^H \hat{D}_{m_o} \hat{a}_{m_o} \tag{79}$$

$$T_{2_j} = \frac{1}{n} \delta_j \Delta_j^H \Delta_j \delta_j^H \tag{80}$$

$$T_{2(j,\tilde{j})} = \frac{1}{n} \delta_j \Delta_j^H \Delta_{\tilde{j}} \delta_{\tilde{j}}^H \tag{81}$$

and $\delta_j = \hat{\alpha}_{m_o+j}$ and $\Delta_j = \hat{e}_{m_o+j}$.

If ϵ_t is circularly symmetric gaussian with zero mean and finite variance then, $E(\epsilon_{1_R}^2 log|\epsilon_{1_R}|) < \infty$ and $E(\epsilon_{1_C}^2 log|\epsilon_{1_C}|) < \infty$, it thus follows from Theorem 2.2 of [12]:

$$\limsup_{n\to\infty} \frac{\sup_\omega I_\epsilon(\omega)}{\sigma^2 \ln(n)} \leq 8 \text{ a.s.} \tag{82}$$

Also, note that $\forall \omega \in (0, 2\pi)$

$$\frac{1}{(n \ln n)^{1/2}} \sum_{t=1}^{n} e^{it\omega} = o(1) \implies \frac{1}{n} \sum_{t=1}^{n} e^{it\omega} = o\left(\frac{\ln n}{n}\right)^{1/2} \tag{83}$$

Using (82), Observe that $\forall j = 1, 2, \ldots, k$

$$|\hat{\alpha}_{m_o+j}|^2 = \frac{1}{n} I_\epsilon(\omega_{m_o+j}\hat{}) \implies |\hat{\alpha}_{m_o+j}| = o\left(\frac{\ln n}{n}\right)^{1/2} \text{ a.s.} \tag{84}$$

Using (82), (83) and (84), $\forall j = 1, 2, \ldots, m_o$ as $n \to \infty$

$$\frac{1}{n} \hat{\alpha}_{m_o+1} \hat{e}_{m_o+1}^H \hat{e}_j \hat{\alpha}_j^H = o\left(\frac{\ln n}{n}\right) \tag{85}$$

Thus,

$$T_{1_1} = \frac{1}{n} \delta_1 \Delta_1^H \hat{D}_{m_o} \hat{a}_{m_o} = o\left(\frac{\ln n}{n}\right) \text{ a.s.} \tag{86}$$

Similarly, $\forall j = 1, 2, \ldots, k$

$$T_{1_j} = o\left(\frac{\ln n}{n}\right) \text{ a.s.} \tag{87}$$

$$T_1 = 2 \sum_{k=1}^{k} T_{1_j} = o\left(\frac{\ln n}{n}\right) \text{a.s. as } n \to \infty \tag{88}$$

Similarly, $\forall j \neq \tilde{j}$

$$T_{2(j,\tilde{j})} = \delta_j \Delta_j^H \Delta_{\tilde{j}} \delta_{\tilde{j}}^H = o\left(\frac{\ln n}{n}\right)^{(3/2)} \tag{89}$$

$$\implies T_{2(j,\tilde{j})} = o\left(\frac{\ln n}{n}\right) \text{a.s. as } n \to \infty \tag{90}$$

$$T_{2_j} = \frac{1}{n} \hat{\alpha}_{m_o+j} \hat{e}_{m_o+j}^H \hat{e}_{m_o+j} \hat{\alpha}_{m_o+j}^H = |\hat{\alpha}_{m_o+j}|^2 = \frac{1}{n} I_\epsilon(\hat{\omega}_{m_o+j}) \tag{91}$$

Finally substituting value of T_1 and T_2 in (76) we have as $n \to \infty$:

$$\hat{\sigma}_{m_o+k}^2 = \hat{\sigma}_{m_o}^2 - \sum_{j=1}^{k} \frac{I_\epsilon(\hat{\omega}_{m_o+j})}{n} + o\left(\frac{\ln n}{n}\right) \text{ a.s.} \tag{92}$$

In the next lemma, we look at the likelihood ratio 8 and investigate its asymptotic behaviour.

Proof: (Lemma 3) We understand the substituting the value of likelihoods in 8, r_m reduces to

$$r_m = 2n \ln\left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_{m-1}^2}\right) \tag{93}$$

where $\hat{\sigma}_0^2 = \frac{1}{n}\sum_{t=1}^{n} y_t^H y_t$ as $n \to \infty$. When $m = 1$, clearly $r_1 = 0$. Under normal assumptions of noise, M.L.E $\hat{\theta}_k$ of θ_k is same as the non- linear least sqaure estimate, we conclude using results from [10], [11] as $n \to \infty$,

$$\hat{\alpha}_j \to \alpha_j \text{ for j=1,2,\ldots,m; } m \leq m_o \tag{94}$$

For the underestimation case i.e $\forall m < m_o$, $\hat{\sigma}_m^2$ can thus be written using lemma 1 as follows:

$$\hat{\sigma}_m^2 \to \sigma^2 + \sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j \text{a.s. and } \hat{\sigma}_{m_o}^2 \to \sigma^2 \quad \text{a.s. as } n \to \infty \tag{95}$$

Thus, for $2 \leq m \leq m_o$

$$\frac{r_m}{n} = 2\ln\left(\frac{\sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2 + \sum_{j=m}^{m_o} \alpha_j^H \alpha_j}\right) \text{ a.s. as } n \to \infty \tag{96}$$

$$= 2\ln\left(1 + \frac{\sum_{j=1}^{m-1} \alpha_j^H \alpha_j}{\sigma^2 + \sum_{j=m}^{m_o} \alpha_j^H \alpha_j}\right) \text{ a.s. as } n \to \infty \tag{97}$$

since the R.H.S is bounded by definition of model and is strictly positive. Hence, we have $r_m = O(n)$ a.s. for all $m \leq m_o$.

Similarly, for overestimation i.e $m > m_o$, from Lemma 1 and lemma 2 we have

$$\frac{r_m}{n} = 2\ln\left(\frac{\sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2 + \frac{G_{m-m_o}}{n} + o\left(\frac{\ln n}{n}\right)}\right) \text{ a.s. as } n \to \infty \tag{98}$$

where $G_k = \sum_{j=1}^{k} I_\epsilon(\hat{\omega}_{m_o+j}) \, \forall k = 1, 2, \ldots, \tilde{m} - m_o$. Using (82), $G_k = O(\ln n)$

$$\frac{G_{m-m_o}}{n} \to 0 \text{ a.s and } o\left(\frac{\ln n}{n}\right) \to 0 \text{ as } n \to \infty \tag{99}$$

Combining this we have for all $m > m_o$

$$\frac{r_m}{n} = 2\ln\left(1 + \frac{\sum_{j=1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2}\right) = O(1) \text{ a.s. as } n \to \infty \tag{100}$$

Hence, we get the desired result.

Chapter 2

On the consistency of Exponential Embedded Family (EEF) rule for 2-D cisoid model in white noise

On the consistency of Exponential Embedded Family (EEF) rule for 2-D cisoid model in white noise

***Abstract*— Model order selection is a critical problem in signal processing models. Recently, [1] presented Exponential Embedded Family (EEF) approach for model order selection using Generalised Likelihood Ratios(GLR) framework. In this paper, we aim to extend the consistency of EEF rule for order selection to 2-D cisoid model in presence of white noise. Simulation examples are provided to validate the performance of EEF rule on different parameter specifications and compare it with other popular model order selection rules.**

I. INTRODUCTION

The need for model order selection is canonical to parametric signal processing and time series methods. Abundant literature on model order selection techniques exists, including some of the most popular techniques, Bayesian Information Criteria (BIC), see [17]-[20] and Akaike Information Criteria (AIC), see [21]-[22]. There has been some recent advancements too, where novel rules have been designed with favourable properties like Penalising adaptively the likelihood approach, see [16].

In [1], Petre and Babu (2012) presented a GLR based derivation of a recently proposed EEF rule and thus bridged a connection between EEF and other conventional model order selection techniques. Large sample consistency properties of EEF rule for 1-D cisoid models in presence of white noise has been discussed in the last chapter. Extending the results to 2-D cisoid model seems to be a natural extension. In this paper, we begin by establishing discussing the formulation of EEF rule for the model and prove its large sample consistency for the same. We also illustrate numerical simulation results and compare its performance with other popular model order selection techniques.

The resr of the paper is organised as follows: In next section, we start with discussion of the model, various assumptions needed for the proof of consistency of the rule followed by formulation of EEF rule. We then establish consistency of the rule followed by numerical simulation example and conclusions.

II. MODEL DEFINITION, NOTATIONS AND ASSUMPTIONS

A. Model Specifications

Detection of the signal component in presence of noise is an important problem in Statistical Signal Processing. 2-D cisoid model have gained widespread attention due to its applicability in texture analysis. Formally, the model is defined as

$\forall s, t$ such that$1 \leq s \leq S; 1 \leq t \leq T$

$$y(s,t) = f(s,t,\theta_m) + \epsilon(s,t) \tag{1}$$

$$f(s,t,\theta_m) = \sum_{k=1}^{m} \alpha_k e^{i(s\beta_k + t\omega_k)} \tag{2}$$

Therefore, the model becomes

$$y(s,t) = \sum_{k=1}^{m} \alpha_k e^{i(s\beta_k + t\omega_k)} + \epsilon(s,t) \tag{3}$$

where $\theta_m = (\alpha_{1_R}, \alpha_{1_C}, \beta_1, \omega_1, ..., \alpha_{m_R}, \alpha_{m_C}, \beta_m, \omega_m)'$ is a $4m$x1 vector of unknown signal parameters; α_{j_R} and α_{j_C} denotes the real and the imaginary part of α_j for j=1,..., m. Let m_o be the true number of components in the the observed signal. Given a sample of size ST, $y = (y(1,1), \ldots, y(S,1), y(1,2), \ldots, y(S,2), \ldots, y(S,T))'$ the model order selection problem will be to estimate m_o.

B. Model Assumptions

- $\epsilon(s,t)$ are i.i.d complex valued gaussian with zero mean s.t.

$$\epsilon(s,t) = \epsilon_R(s,t) + i\epsilon_C(s,t) \tag{4}$$
$$\epsilon_R(s,t) \sim \mathcal{N}(0, \sigma^2/2) \tag{5}$$
$$\epsilon_C(s,t) \sim \mathcal{N}(0, \sigma^2/2) \tag{6}$$

  and $\epsilon_R(s,t)$ and $\epsilon_C(s,t)$ are independent.
- $\forall k = 1, 2, ..., m_o : (\beta_k, \omega_k) \in (0, 2\pi)x(0, 2\pi)$; where (β_k, ω_k) are pairwise different i.e. $\omega_j \neq \omega_k$ or $\beta_j \neq \beta_k, \forall j \neq k$. Furthermore, $\forall k = 1, 2..., m_o : \alpha_k$ are bounded.
- The true model parameter vector θ_{m_o} is an interior point in the parameter space $\Theta \subset \mathbb{R}^{4m_o}$.

III. EEF RULE FORMULATION FOR 2-D CISOID MODEL

For the model in (3), define

$$y = (y(1,1), \ldots, y(S,1), y(1,2), \ldots, y(S,T))'$$
$$\epsilon = (\epsilon(1,1), \ldots, \epsilon(S,1), \epsilon(1,2), \ldots, \epsilon(S,T))'$$
$$f(\theta_m) = (f(1,1,\theta_m), \ldots, f(S,1,\theta_m), \ldots, f(S,T,\theta_m))'$$

For a m- component model, let $\theta_m^* = (\theta_m', \sigma_m^2)'$ denotes the vector containing the underlying signal and noise parameters, then the likelihood function of y under these assumptions can be written as

$$f_m(y, \theta_m^*) = \frac{1}{\left(\frac{2\pi\sigma^2}{2}\right)^{ST}} e^{\frac{-(y-f(\theta_m))^H (y-f(\theta_m))}{\sigma_m^2}} \tag{7}$$

We consider the set of $\tilde{m}$ nested models given by $\{M_m\}_{m=1}^{\tilde{m}}$ where M_m is the m component 2-D cisoid model with parameter vector θ_m^*. We assume that the true model M_{m_o} is contained in this set .i.e., $m_o \leq \tilde{m}$ and y is not completely a white noise process i.e., $M_0 \neq M_{m_o}$.

Consider the Generalized likelihood ratio,

$$\hat{r}_{m+1} = 2\ln\left[\frac{f_m(y,\hat{\theta}_m^*)}{f_0(y,\hat{\theta}_0^*)}\right] \tag{8}$$

where $\hat{\theta}_k^*$ is maximum likelihood estimate of underlying signal and noise parameter vector θ_k^* and $f_0(y,\hat{\theta}_0^*)$ denotes the p.d.f of y when M_0 is the model i.e., $f(\theta_0) = 0$.

The EEF rule for 2-D cisoid model mentioned above can be given by maximising the following statistic for $m = 1, 2, \ldots, \tilde{m}$

$$EEF(m) = \begin{cases} \hat{r}_{m+1} - (4m+1)\left[1 + \ln\Big(\frac{\hat{r}_{m+1}}{4m+1}\Big)\right] & \text{if } \hat{r}_{m+1} > 4m+1 \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{m} = \underset{m\in\{1,2,\ldots,\tilde{m}\}}{\arg\max}\,[EEF(m)] \tag{9}$$

where $\hat{m}$ is the estimated model order for the model.

IV. Consistency of EEF Rule

To prove the consistency of the model order selection rule, we need some results which discuss and analyze the behaviour of model variance $\hat{\sigma}_m^2$ for different choices of m and hence the behaviour of likelihood ratio r_m as defined above. In order to discuss the large sample consistency properties of EEF rule for 2-D cisoid models, we first define the notion of large sample asymptotics using the following remark.

Remark: Let $\{\Psi_i\}$ be a sequence of rectangles s.t.

$$\Psi = \{(s,t) \in \mathbb{Z}^2 | 1 \leq s \leq S_i, 1 \leq t \leq T_i\} \tag{10}$$

Then, sequence of subsets $\{\Psi_i\}$ is said to tend to infinity as $i \to \infty$ if

$$\lim_{i\to\infty} min(S_i,T_i) \to \infty \text{ and } 0 < \lim_{i\to\infty}\left(\frac{S_i}{T_i}\right) < \infty \tag{11}$$

To simplify notation, omit subscript i. Thus, $\Psi(S,T) \to \infty$ implies both S and T tend to infinity as a function of i, and roughly at the same rate.

We now present these results in the following lemmas and the proofs of the same are attached in the appendix.

Lemma 1: Under the assumptions B1-B3, $\forall m \leq m_o$,

$$\hat{\sigma}_m^2 = \sigma^2 + \sum_{j=1}^{m_o}\alpha_j^H\alpha_j - \sum_{j=1}^{m}\hat{\alpha}_j^H\hat{\alpha}_j + o(1) \tag{12}$$

a.s. as $\Psi(S,T) \to \infty$.
Proof: See Appendix A.

Lemma 2: Under B1-B3, for any integer $k \geq 1$

$$\hat{\sigma}_{m_o+k}^2 = \hat{\sigma}_{m_o}^2 - \frac{G_k}{ST} + o\left(\frac{(\ln(ST)\ln S\ln T)^{1/2}}{ST}\right) \tag{13}$$

a.s. as $\Psi(S,T) \to \infty$, where $G_k = \sum_{j=1}^{k} I_\epsilon(\hat{\beta}_{m_o+j}, \hat{\omega}_{m_o+j})$ and

$$I_\epsilon(\beta,\omega) = \frac{1}{ST}\left|\sum_{s=1}^{S}\sum_{t=1}^{T}\epsilon(s,t)e^{-i(s\beta+t\omega)}\right|^2 \tag{14}$$

$I_\epsilon(\beta,\omega)$ corresponds to periodogram of underlying white noise process and $(\hat{\beta}_{m_o+1},\hat{\omega}_{m_o+1}), (\hat{\beta}_{m_o+2},\hat{\omega}_{m_o+2}), \ldots, (\hat{\beta}_{m_o+k},\hat{\omega}_{m_o+k})$ are the k largest frequencies corresponding to $I_\epsilon(\beta,\omega)$. Thus, G_k is the sum of k largest elements of the periodogram of noise.
Proof: See Appendix A.

Lemma 3: Under assumptions B1-B3, r_m as defined before in (8) satisfies

$$r_m = \begin{cases} 0 & m = 1 \\ O(ST) & 2 \leq m \leq \tilde{m} \end{cases} \tag{15}$$

Proof: See Appendix A.

In the next subsection, we present the proof of consistency for EEF rule. We prove the result for various sub cases based on under and over estimation, and form of EEF rule based on relationship between r_m and model order m. We show that for each of the possible cases, probability of wrong estimation of model order goes to zero as $\Psi(S,T) \to \infty$.

A. *Proof of Consistency*

Theorem: Under the assumptions B1-B3, If m_o is the true model order and $\hat{m}$ is the estimated model order using EEF rule, then

$$P(\hat{m} \neq m_o) \to 0 \text{ as } \Psi(S,T) \to \infty \tag{16}$$

Proof:
Case I: $m \leq m_o$ **(Underestimation)**
Subcase I: $\hat{r}_{m+1} > 4m+1$;$\hat{r}_{m_o+1} > 4m_o+1$

$$\begin{aligned} &EEF(m) - EEF(m_o) = \hat{r}_{m+1} - \hat{r}_{m_o+1} \\ &- (4m+1)\ln(\hat{r}_{m+1}) + (4m_o+1)\ln(\hat{r}_{m_o+1}) + k \end{aligned} \tag{17}$$

where,

$$k = 4(m_o-m)+(4m+1)\ln(4m+1)-(4m_o+1)\ln(4m_o+1) \tag{18}$$

Note that, k is independent of sample size n. Also since

$$\hat{r}_{m+1} = 2ST\ln\Big(\frac{\hat{\sigma_o^2}}{\hat{\sigma_m^2}}\Big) \tag{19}$$

then,

$$\begin{aligned} &EEF(m) - EEF(m_o) = -2ST\ln\Big(\frac{\hat{\sigma_m^2}}{\hat{\sigma_{m_o}^2}}\Big) \\ &- (4m+1)\ln(\hat{r}_{m+1}) + (4m_o+1)\ln(\hat{r}_{m_o+1}) + k \end{aligned} \tag{20}$$

Also, using result of lemma 3, we get $\hat{r}_{m+1} = O(ST)\ \forall m \leq m_o$. Using lemma 1, we have $\forall m \leq m_o$

$$\hat{\sigma_m^2} \to \sigma^2 + \sum_{k=m+1}^{m_o}\alpha_k^H\alpha_k \text{ a.s.} \tag{21}$$

$$\hat{\sigma_{m_o}^2} \to \sigma^2 \tag{22}$$

as $\Psi(S,T) \to \infty$.Substituting values in the first part of the above expression we get $\forall m \le m_o$

$$\ln\Big(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}\Big) \to \ln\Big(1 + \frac{\sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2}\Big)\text{a.s. as } \Psi(S,T) \to \infty \tag{23}$$

which is a strictly positive and bounded quantity, substituting back in equation (20), we get

$$\begin{aligned} EEF(m) - EEF(m_o) &= O(ST) - \\ &(4m+1)O(\ln(ST)) + (4m_o+1)O(\ln(ST)) + k \text{ a.s.} \end{aligned} \tag{24}$$

$$\text{or, } \frac{EEF(m) - EEF(m_o)}{ST} \to -2\ln\Big(1 + \frac{\sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2}\Big) \tag{25}$$

a.s. as $\Psi(S,T) \to \infty$.

Thus, $EEF(m) < EEF(m_o)$ with probability 1.

Subcase 2 : $\hat{r}_{m+1} < 4m+1$;$\hat{r}_{m_o+1} > 4m_o+1$

$$EEF(m) = 0 \tag{26}$$

$$EEF(m_o) = \hat{r}_{m_o+1} - (4m_o+1)\Bigg[1 + \ln\Big(\frac{\hat{r}_{m_o+1}}{4m_o+1}\Big)\Bigg] \tag{27}$$

as the above expression is of form $g(x) = x - \ln x - 1$ which has a unique minimum value of 0 at x=1, $EEF(m_o) > 0$ for $\hat{r}_{m_o+1} \neq 3m_o+1$. Thus $EEF(m_o) > EEF(m) \forall m \le m_o$ and hence underestimation is not possible in this case.

Subcase 3 : $\hat{r}_{m+1} > 4m+1$;$\hat{r}_{m_o+1} < 4m_o+1$

We know from [1], for large samples, EEF rule looks as the following

$$EEF(m) \cong \hat{r}_{m+1} - (4m+1)\ln n \tag{28}$$

Since $\hat{r}_{m_o+1} < 4m_o+1$, $EEF(m_o) = 0$.Thus,

$$\frac{EEF(m) - EEF(m_o)}{ST\ln(ST)} = \frac{\hat{r}_{m+1}}{ST\ln(ST)} - \frac{3m+1}{ST} \tag{29}$$

as $\Psi(S,T) \to \infty$ Since we know from Lemma 3, $\hat{r}_{m+1} = O(ST)$

$$\frac{EEF(m) - EEF(m_o)}{ST\ln(ST)} \to 0 \text{ as } \Psi(S,T) \to \infty \tag{30}$$

$$\text{or, } P\Big(\frac{EEF(m) - EEF(m_o)}{ST\ln(ST)} > 0\Big) \to 0 \text{ as } \Psi(S,T) \to \infty. \tag{31}$$

Thus for all the subcases discussed above,

$$\begin{aligned} &P(\hat{m} < m_o) \\ &= P(EEF(m) > EEF(m_o) \text{ for some } m < m_o) \\ &\to 0 \text{ as } \Psi(S,T) \to \infty \end{aligned} \tag{32}$$

Case 2: $m > m_o$ **(Overestimation)**

Subcase 1: $\hat{r}_{m+1} > 4m+1$;$\hat{r}_{m_o+1} > 4m_o+1$

$$\begin{aligned} EEF(m) - EEF(m_o) &= -2ST\ln\Big(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}\Big) \\ &- (4m+1)\ln(\hat{r}_{m+1}) + (4m_o+1)\ln(\hat{r}_{m_o+1}) + k \end{aligned} \tag{33}$$

Since $m > m_o$, $\hat{r}_{m+1} = 2\ln\Big(\frac{\hat{f}_m}{\hat{f}_o}\Big) > 2\ln\Big(\frac{\hat{f}_{m_o}}{\hat{f}_o}\Big) = \hat{r}_{m_o+1}$, using which we have

$$\begin{aligned} EEF(m) - EEF(m_o) \le -2ST\ln\Big(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}\Big) \\ + k + 4(m_o - m)\ln(\hat{r}_{m+1}) \end{aligned} \tag{34}$$

It follows from the asymptotic theory of likelihood ratios (see [20], [19]) that,

$$2ST\ln\Big(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}\Big) \sim \chi^2_{4(m-m_o)} \tag{35}$$

Thus, $\frac{1}{\ln(ST)}\Bigg(2ST\ln\Big(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}\Big)\Bigg) = o_p(1)$.Since k is independent of n and $m > m_o$, and from lemma 3 we have $\hat{r}_{m+1} = O(ST)$, we get

$$\begin{aligned} &P\Bigg(\frac{1}{\ln(ST)}\Bigg(-2ST\ln\Big(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}\Big) + k + \\ &4(m_o - m)\ln(\hat{r_{m+1}})\Bigg) < 0\Bigg) \to 1 \text{ as } \Psi(S,T) \to \infty \end{aligned} \tag{36}$$

that implies from (34)

$$P\Bigg(\frac{EEF(m) - EEF(m_o)}{\ln(ST)} < 0\Bigg) \to 1 \text{ as } \Psi(S,T) \to \infty \tag{37}$$

Therefore,

$$\begin{aligned} &P(\hat{m} > m_o) \\ &= P(EEF(m) - EEF(m_o) > 0 \text{ for some } m > m_o) \\ &= P\Bigg(\frac{EEF(m) - EEF(m_o)}{\ln(ST)} > 0\Bigg) \to 0 \text{ as } \Psi(S,T) \to \infty \end{aligned} \tag{38}$$

Subcase 2 : $\hat{r}_{m+1} < 4m+1$;$\hat{r}_{m_o+1} > 4m_o+1$

$$EEF(m) = 0 \tag{39}$$

$$EEF(m_o) = \hat{r}_{m_o+1} - (4m_o+1)\Bigg[1 + \ln\Big(\frac{\hat{r}_{m_o+1}}{4m_o+1}\Big)\Bigg] \tag{40}$$

as the above expression is of form $g(x) = x - \ln x - 1$ which has a unique minimum value of 0 at x=1, $EEF(m_o) > 0$ for $\hat{r}_{m_o+1} \neq 4m_o+1$. Thus $EEF(m_o) > EEF(m) \forall m \ge m_o$ and hence overestimation is not possible in this case.

Subcase 3 : $\hat{r}_{m+1} > 4m+1$; $\hat{r}_{m_o+1} < 4m_o+1$

Similar to the underestimation approach, We know from [1], for large samples, EEF rule looks as the following

$$EEF(m) \cong \hat{r}_{m+1} - (4m+1)\ln(ST) \tag{41}$$

Since $\hat{r}_{m_o+1} < 4m_o + 1$, $EEF(m_o) = 0$.Thus,

$$\frac{EEF(m) - EEF(m_o)}{ST\ln(ST)} = \frac{\hat{r}_{m+1}}{ST\ln(ST)} - \frac{3m+1}{ST} \tag{42}$$

as $\Psi(S,T) \to \infty$ Since we know from Lemma 3, $\hat{r}_{m+1} = O(ST)$

$$\frac{EEF(m) - EEF(m_o)}{ST\ln(ST)} \to 0 \text{ as } \Psi(S,T) \to \infty \tag{43}$$

$$\text{or, } P\Big(\frac{EEF(m) - EEF(m_o)}{ST\ln(ST)} < 0\Big) \to 1 \text{ as } \Psi(S,T) \to \infty. \tag{44}$$

Thus for all the subcases discussed above,

$$\begin{aligned} P(\hat{m} < m_o) &= P(EEF(m) < EEF(m_o) \text{ for some } m < m_o) \\ &\to 0 \text{ as } \Psi(S,T) \to \infty \end{aligned} \tag{45}$$

Hence, from the two cases of over and under estimation we have that

$$P(\hat{m} \neq m_o) \to 0 \text{ as } \Psi(S,T) \to \infty \tag{46}$$

V. Numerical Simulations

We conducted numerical experiments to investigate the performance of EEF rule of model order selection for 2-D cisoid models. We compared the performance of the same with other popular order selection rules like AIC, BIC and a more recent Penalizing the likelihood (PAL) based approach for model order selection. We consider the following model for simulation purposes:

$$y(s,t) = \sum_{k=1}^{2} \alpha_k e^{i(s\beta_k + t\omega_k)} + \epsilon(s,t) \tag{47}$$

$$\begin{aligned} \alpha_1 &= 1 + \sqrt{2}i & \beta_1 &= 0.26\pi & \omega_1 &= 0.26\pi \\ \alpha_2 &= 2 + 2i & \beta_2 &= 0.62\pi & \omega_2 &= 0.62\pi \end{aligned}$$

where $\epsilon(s,t)$ is complex valued gaussian as defined in assumption B1. We vary the value of σ^2 from 0.5 to 15, vary S and T from 2 to 40 and note the probability of correct estimation of model order for 200 simulation runs and compare the performances with forms of rules given below:

$$PAL(m) = -2\ln(f_m(y, \hat{\theta}^*_m)) + (4m+1)\ln(4\tilde{m}+1) * \frac{\ln(r_m+1)}{\ln(\rho_m+1)} \tag{48}$$

$$BIC(m) = -2\ln(f_m(y, \hat{\theta}^*_m)) + (4m+1)\ln(ST) \tag{49}$$

$$AIC(m) = -2\ln(f_m(y, \hat{\theta}^*_m)) + 2 * (4m+1) \tag{50}$$

where, usual notations defined above are followed and ρ_m is defined as

$$\rho_m = 2\ln\left[\frac{f_{\tilde{m}}(y, \hat{\theta}^*_{\tilde{m}})}{f_{m-1}(y, \hat{\theta}^*_{m-1})}\right] \tag{51}$$

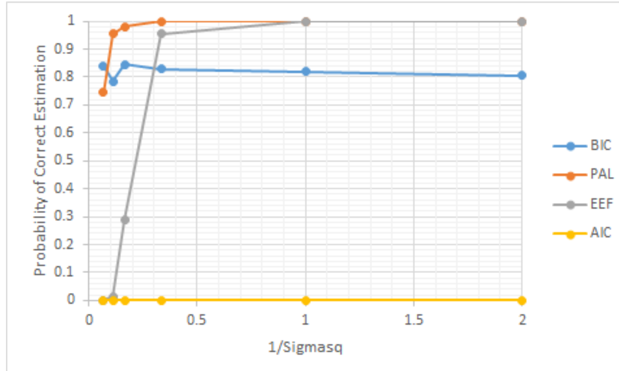

Fig. 1. Probability of correct order estimation; S,T=9

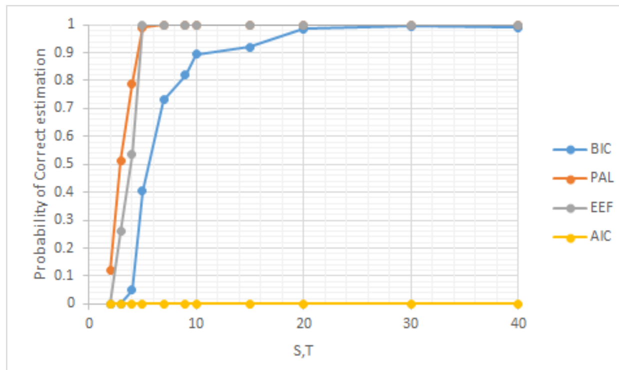

Fig. 2. Probability of correct order estimation;$\sigma^2 = 1$

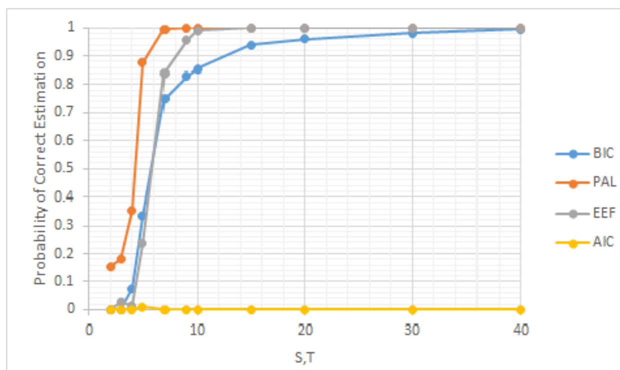

Fig. 3. Probability of correct order estimation; $\sigma^2 = 3$

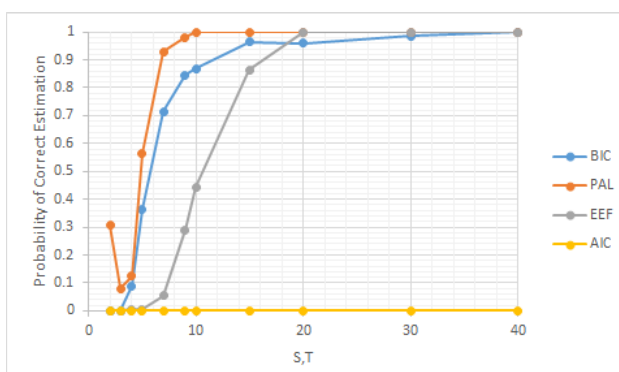

Fig. 4. Probability of correct order estimation; $\sigma^2 = 6$

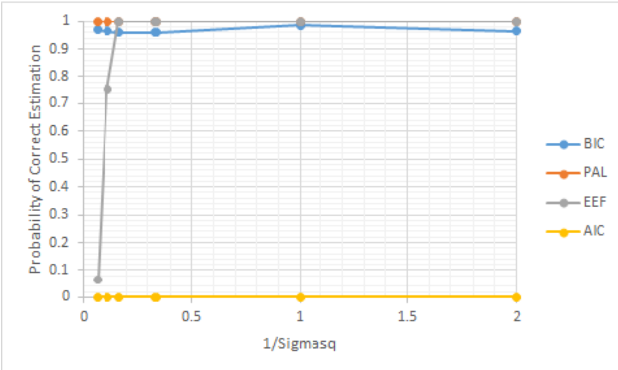

Fig. 5. Probability of correct order estimation; S,T=20

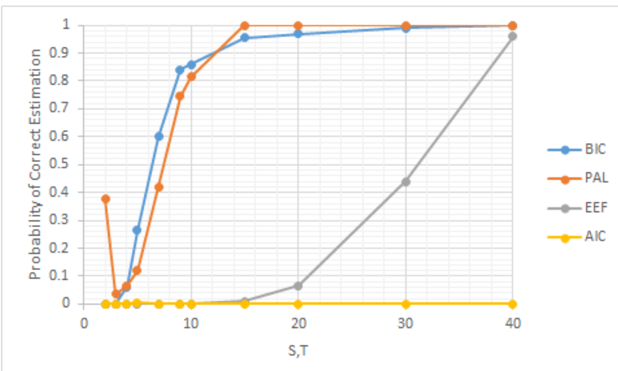

Fig. 6. Probability of correct order estimation;$\sigma^2 = 15$

and the usual notations for the rest of the terms as defined above in the paper. We understand that EEF rules performs well for low noise variance and large sample sizes of S and T. Probability of correct estimation of model order at S,T=9 reaches 1 even for high sigma square values of 3. Its performance is better than AIC and BIC for small values of variance $sigma^2$ and its comparable to PAL rule. However for large values of noise variance, the probability of correct estimation takes large number of samples for attaining correct model order with probability 1 is taken over by BIC and PAL in doing the same.

VI. CONCLUSIONS

In this paper, we looked at the asymptotic statistical properties of EEF rule for 2-D complex exponential model as a natural extension to 1-D case. We derived the EEF rule formulation for the same proved that this novel rule is consistent. We supported the proof with numerical simulations which demonstrated the good performance of EEF rule in comparison to other popular model order selection rules like AIC, BIC and PAL, particularly at large values of S, T and at large values of Signal-to-noise(SNR) ratio.

APPENDIX

In the following section, we present the proofs of the lemma 1-3 that were used in proof of consistency of EEF for 2-D cisoid models. The proof follow along the same lines as the analogous 1-D cisoid case discussed before. The model (3) can be written as:

$$y = D_m(\beta,\omega)a_m + \epsilon \tag{52}$$

where, $\forall j = 1, 2, ..., m$

$$e_j = (e^{i(\beta_k+\omega_k)}, \ldots, e^{i(S\beta_k+\omega_k)}, e^{i(S\beta_k+T\omega_k)})' \tag{53}$$

$$\beta = (\beta_1, \beta_2, ..., \beta_m)' \tag{54}$$

$$\omega = (\omega_1, \omega_2, ..., \omega_m)' \tag{55}$$

$$D_m(\beta,\omega) = (e_1, e_2, ..., e_m)_{ST\text{x}m} \tag{56}$$

$$a_m = (\alpha_1, \alpha_2, ..., \alpha_m)'_{m\text{x}1} \tag{57}$$

We write D_m for $D_m(\beta,\omega)$ and $\hat{D_m}$ for $D_m(\hat{\beta},\hat{\omega})$ to simplify the notations. $O(.)$ and $o(.)$ denote either the order in probability or deterministic order depending on the context. The proofs of the three lemmas are presented below:
Proof: (Lemma 1) The estimated variance of noise from the model in 52 can be written as:

$$\hat{\sigma}^2_m = \frac{y^H(I_{ST} - P_m(\hat{\beta},\hat{\omega}))y}{ST} \tag{58}$$

where $P_m(\hat{\beta},\hat{\omega}) = \hat{D_m}(\hat{D_m}^H\hat{D_m})^{-1}\hat{D_m}^H$ is the projection matrix. Substituting value of y from true model $y = D_{m_o}a_{m_o} + \epsilon$ and expanding gives:

$$\hat{\sigma}^2_m = \frac{1}{ST}\Bigg[\epsilon^H\epsilon + 2\epsilon^H D_{m_o}a_{m_o} + a^H_{m_o}D^H_{m_o}D_{m_o}a_{m_o} - \hat{a}^H_{m_o}\hat{D}^H_{m_o}\hat{D}_{m_o}\hat{a}_{m_o}\Bigg] \tag{59}$$

where $\hat{a}_m = (\hat{D_m}^H\hat{D_m})^{-1}\hat{D_m}^H y$. Let us analyze each of the terms one by one Since both $\epsilon_R(s,t)$ and $\epsilon_C(s,t)$ are independent, by Kolmogorov's Strong law of Large Numbers(see e.g. [[6])

$$\frac{\epsilon^H\epsilon}{ST} = \frac{\sum_{s=1}^S\sum_{t=1}^T\epsilon_R^2(s,t)}{ST} + \frac{\sum_{s=1}^S\sum_{t=1}^T\epsilon_C^2(s,t)}{ST} \to \frac{\sigma^2}{2} + \frac{\sigma^2}{2} = \sigma^2 \tag{60}$$

a.s. as $\Psi(S,T) \to \infty$. Now let us have a look at

$$\frac{\epsilon^H D_{m_o}a_{m_o}}{ST} = \frac{1}{ST}\left(\sum_{k=1}^{m_o}\alpha_k\left(\sum_{l=1}^{S}\sum_{j=1}^{T}\bar{\epsilon}(l,j)e^{i(l\beta_k+j\omega_k)}\right)\right) \tag{61}$$

Using lemma 2 of Kundu and Mitra ([7], [8]),

$$\frac{\epsilon^H D_{m_o}a_{m_o}}{ST} = o(1) \text{ a.s.} \tag{62}$$

as $\Psi(S,T) \to \infty$. Using the fact that

$$\lim_{n\to\infty}\frac{\sum_{t=1}^n e^{it\tilde{w}}}{n} = o(1) \text{ for some fixed } \tilde{\omega} \in (0,2\pi) \tag{63}$$

it can be shown that, the third and fourth terms of 59 reduces to

$$\frac{1}{ST} a_{m_o}^H D_{m_o}^H D_{m_o} a_{m_o} = \sum_{j=1}^{m_o} \alpha_j^H \alpha_j + o(1) \text{ a.s.} \quad (64)$$

$$\frac{1}{ST} \hat{a}_{m_o}^H \hat{D}_{m_o}^H \hat{D}_{m_o} \hat{a}_{m_o} = \sum_{j=1}^{m} \hat{\alpha}_j^H \hat{\alpha}_j + o(1) \text{ a.s.} \quad (65)$$

as $\Psi(S,T) \to \infty$. Thus using (60), (62), (64) and (65), we have $\forall m \leq m_o$

$$\hat{\sigma}_m^2 = \sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j - \sum_{j=1}^{m} \hat{\alpha}_j^H \hat{\alpha}_j + o(1) \text{ a.s. as } \Psi(S,T) \to \infty \quad (66)$$

The above lemma shows the dependence of $\hat{\sigma}_m^2$ on the model parameters asymptotically and will be used in tracking the nature of EEF rule for the underestimation cases. The next lemma discusses the behaviour of model variance for the overestimation case and investigates the dependence of the same on model parameters and noise periodogram.

Proof:(Lemma 2) For l=1,2,..., k let

$$\hat{\alpha}_{m_o+l} = \frac{1}{ST} \sum_{s=1}^{S} \sum_{t=1}^{T} \epsilon(s,t) e^{-i(s\hat{\beta}_{m_o+l} + t\hat{\omega}_{m_o+l})} \quad (67)$$

$$\hat{D}_{m_o+k} = (\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_{m_o+k}) \quad (68)$$

$$= (\hat{D}_{m_o}, \hat{e}_{m_o+1}, \hat{e}_{m_o+2}, \ldots, \hat{e}_{m_o+k}) \quad (69)$$

Similarly, (70)

$$\hat{a}_{m_o+k} = (\hat{a}_{m_o}', \hat{\alpha}_{m_o+1}, \hat{\alpha}_{m_o+2}, \ldots, \hat{\alpha}_{m_o+k})' \quad (71)$$

Using results presented in theorem 2 of [9] (see also [10], [11]) we note that in the overestimated scenario, M.L.E of the parameter vector contains a sub vector equal in dimension to the true model order, that converges almost surely to the true parameter vector. Also, the frequencies of the over estimated components are those that maximise the noise periodogram. Thus for the variance of the over estimated model can be given as,

$$\hat{\sigma}_{m_o+k}^2 = \frac{y^H (I_{ST} - P_{m_o+k}(\hat{\beta}_{m_o+k}, \hat{\omega}_{m_o+k})) y}{ST} \quad (72)$$

Substituting the values of $\hat{a}_{m_o+k}$ and $\hat{D}_{m_o+k}$ in the above expressions and simplifying we get

$$\hat{\sigma}_{m_o+k}^2 = \hat{\sigma}_{m_o}^2 - T_1 - T_2 \quad (73)$$

where,

$$\hat{\sigma}_{m_o}^2 = \frac{1}{ST}(y^H y - \hat{a}_{m_o}^H \hat{D}_{m_o}^H \hat{D}_{m_o} \hat{a}_{m_o}) \quad (74)$$

$$T_1 = 2\sum_{k=1}^{k} T_{1_j} \text{ and } T_2 = \sum_{j=1}^{k} T_{2_j} + 2\sum_{j=1}^{k} \sum_{\substack{\tilde{j}=1 \\ j \neq \tilde{j}}}^{k} T_{2(j,\tilde{j})} \quad (75)$$

$$T_{1_j} = \frac{1}{ST} \delta_j \Delta_j^H \hat{D}_{m_o} \hat{a}_{m_o} \quad (76)$$

$$T_{2_j} = \frac{1}{ST} \delta_j \Delta_j^H \Delta_j \delta_j^H \quad (77)$$

$$T_{2(j,\tilde{j})} = \frac{1}{ST} \delta_j \Delta_j^H \Delta_{\tilde{j}} \delta_{\tilde{j}}^H \quad (78)$$

and $\delta_j = \hat{\alpha}_{m_o+j}$ and $\Delta_j = \hat{e}_{m_o+j}$. If $\epsilon(s,t)$ is circularly symmetric gaussian with zero mean and finite variance then, $E(\epsilon_R(1,1)^2 log|\epsilon_R(1,1)|) < \infty$ and $E(\epsilon_C(1,1)^2 log|\epsilon_C(1,1)|) < \infty$, it thus follows from Theorem 2.2 of [12]:

$$\limsup_{\Psi(S,T)\to\infty} \frac{\sup_\omega I_\epsilon(\omega)}{\sigma^2 \ln(ST)} \leq 8 \text{ a.s.} \quad (79)$$

Also, note that $\forall \omega \in (0, 2\pi)$

$$\frac{1}{(n \ln n)^{1/2}} \sum_{t=1}^{n} e^{it\omega} = o(1) \implies \frac{1}{n} \sum_{t=1}^{n} e^{it\omega} = o\left(\frac{\ln n}{n}\right)^{1/2} \quad (80)$$

Observe that $\forall j = 1, 2, \ldots, k$

$$|\hat{\alpha}_{m_o+j}|^2 = \frac{1}{ST} I_\epsilon(\hat{\beta_{m_o+j}}, \hat{\omega_{m_o+j}}) \quad (81)$$

$$|\hat{\alpha}_{m_o+j}| = o\left(\frac{\ln(ST)}{ST}\right)^{1/2} \text{ a.s.} \quad (82)$$

Using (79),(80) and (81), $\forall j = 1, 2, \ldots, m_o$ as $\Psi(S,T) \to \infty$

$$\frac{1}{ST} \hat{\alpha}_{m_o+1} \hat{e}_{m_o+1}^H \hat{e}_j \hat{\alpha}_j^H = o\left(\frac{(\ln(ST) \ln S \ln T)^{1/2}}{ST}\right) \quad (83)$$

Thus,

$$T_{1_1} = o\left(\frac{(\ln(ST) \ln S \ln T)^{1/2}}{ST}\right) \text{ a.s.} \quad (84)$$

Similarly, $\forall j = 1, 2, \ldots, k$

$$T_{1_j} = o\left(\frac{(\ln(ST) \ln S \ln T)^{1/2}}{ST}\right) \text{ a.s.} \quad (85)$$

$$T_1 = 2\sum_{k=1}^{k} T_{1_j} = o\left(\frac{(\ln(ST) \ln S \ln T)^{1/2}}{ST}\right) \text{a.s.} \quad (86)$$

as $\Psi(S,T) \to \infty$. Similarly, $\forall j \neq \tilde{j}$

$$T_{2(j,\tilde{j})} = o\left(\frac{(\ln(ST) \ln S \ln T)^{1/2}}{ST}\right) \text{a.s. as } \Psi(S,T) \to \infty \quad (87)$$

Also, $\forall 1 \leq j \leq k$,

$$T_{2_j} = \frac{1}{ST} \hat{\alpha}_{m_o+j} \hat{e}_{m_o+j}^H \hat{e}_{m_o+j} \hat{\alpha}_{m_o+j}^H \quad (88)$$

$$= |\hat{\alpha}_{m_o+j}|^2 = \frac{1}{ST} I_\epsilon(\hat{\beta_{m_o+j}}, \hat{\omega_{m_o+j}}) \quad (89)$$

Finally substituting value of T_1 and T_2 in (73) we have as $\Psi(S,T) \to \infty$:

$$\hat{\sigma}_{m_o+k}^2 = \hat{\sigma}_{m_o}^2 - \frac{G_k}{ST} + o\left(\frac{(\ln(ST) \ln S \ln T)^{1/2}}{ST}\right) \quad (90)$$

almost surely, where $G_k = \sum_{j=1}^{k} I_\epsilon(\hat{\beta}_{m_o+j}, \hat{\omega}_{m_o+j})$ We now look at the proof of lemma which discuss the asymptotic properties of the likelihood ratio r_m defined above in the paper.

Proof:(Lemma 3) Substituting the value of likelihood in 8, we get

$$r_m = 2ST \ln\bigg(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_{m-1}^2}\bigg) \tag{91}$$

where $\hat{\sigma}_0^2 = \frac{1}{ST}\sum_{s=1}^{S}\sum_{t=1}^{T} y(s,t)^H y(s,t)$ as $\Psi(S,T) \to \infty$.

When $m = 1$, clearly $r_1 = 0$. Under normal assumptions of noise, M.L.E $\hat{\theta}_k$ of θ_k is same as the non- linear least square estimate, we conclude using results from [10], [11] as $\Psi(S,T) \to \infty$,

$$\hat{\alpha}_j \to \alpha_j \text{ for j=1,2,}\ldots\text{,m; } m \leq m_o \tag{92}$$

Case 1:$m \leq m_o$(Underestimation) $\forall m \leq m_o$,

$$\hat{\sigma}_m^2 \to \sigma^2 + \sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j \text{a.s. and } \hat{\sigma}_{m_o}^2 \to \sigma^2 \text{ a.s.} \tag{93}$$

as $\Psi(S,T) \to \infty$. For $2 \leq m \leq m_o$

$$\frac{r_m}{ST} = 2\ln\bigg(1 + \frac{\sum_{j=1}^{m-1} \alpha_j^H \alpha_j}{\sigma^2 + \sum_{j=m}^{m_o} \alpha_j^H \alpha_j}\bigg) \text{ a.s.} \tag{94}$$

as $\Psi(S,T) \to \infty$. Since the R.H.S is bounded by definition of model and is strictly positive. Hence, we have $r_m = O(n)$ a.s. for all $m \leq m_o$.

Case 2: $m > m_o$(Overestimation)Similar, to the underestimation case For $m > m_o$, from Lemma 1 and lemma 2 we have

$$\frac{r_m}{ST} = 2\ln\Bigg(\frac{\sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2 + \frac{G_{m-m_o}}{ST} + o\big(\frac{(\ln(ST)\ln S \ln T)^{1/2}}{ST}\big)}\Bigg) \text{ a.s.} \tag{95}$$

as $\Psi(S,T) \to \infty$. Since $\forall k = 1,2,\ldots,\tilde{m} - m_o$, $G_k = \sum_{j=1}^{k} I_\epsilon(\beta_{m_o+j}, \omega_{m_o+j})$. Using (79), $G_k = O(\ln(ST))$

$$\frac{G_{m-m_o}}{ST} \to 0 \text{ a.s and } \; o\Bigg(\frac{(\ln(ST)\ln S \ln T)^{1/2}}{ST}\Bigg) \to 0 \tag{96}$$

as $\Psi(S,T) \to \infty$. Combining this we have for all $m > m_o$

$$\frac{r_m}{ST} = 2\ln\bigg(1 + \frac{\sum_{j=1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2}\bigg) = O(1) \text{ a.s. as } \Psi(S,T) \to \infty \tag{97}$$

Hence, we get the desired result.

Chapter 3

On confidence interval estimation of model order using Penalizing adaptively the likelihood (PAL) approach for 1-D cisoid model in white noise using residual bootstrapping

On confidence interval estimation of model order using Penalizing adaptively the likelihood (PAL) approach for 1-D cisoid model in white noise using residual bootstrapping

***Abstract*— We discussed the consistency properties of moder order selection using Penalizing adaptively the likelihood (PAL) approach for model order selection. It was suggested that a major drawback of criterion rule based approaches is the face that they only provide with the point estimate and not the confidence interval for model order. Algorithm and numerical simulations for confidence interval estimation using residual bootstrap for model order selection using PAL rule for 1-D cisoid models is discussed.**

I. Algorithm

Algorithm 1 Confidence interval estimation of model order

Define the number of components m_o and model parameter values $\alpha_k, \beta_k \forall k = 1 \ldots m_o$
$n \leftarrow$ Sample size
$\sigma^2 \leftarrow$ noise variance
maxcomp $\leftarrow$ maximum number of components
nsim $\leftarrow$ number of simulations
bootsim $\leftarrow$ number of bootstrap
for `isim=1 to nsim` **do**
    $y \leftarrow$ generate-data(n, m_o, α, β)
    Estimate model order for y using PAL rule
    Estimate $\alpha_k, \beta_k \ \forall k = 1 \cdots$
    Calculate $\hat{y}_i \forall i = 1 \ldots n$
    Calculate $\hat{\epsilon}_i = y_i - \hat{y}_i \ \forall i = 1, \ldots, n$
    for `b=1 to bootsim` **do**
        Draw a bootstrap sample $\hat{\epsilon}_b$ of size n from $\hat{\epsilon}$
        Calculate $y_b = \hat{y} + \hat{\epsilon}_b$
        Estimate model order for y_b using PAL rule
    end for
    Find the empirical interval with α level of confidence using Bootsim number of estimated model orders
end for
Calculate average over confidence interval for nsim simulations

II. Numerical Simulation

In this section, we present results of numerical simulations that we run to analyze the confidence interval of model order estimate based on PAL rule. We consider the following model:

$$y_t = \sum_{k=1}^{3} \alpha_k e^{it\omega_k} + \epsilon_t \tag{1}$$

$$\alpha_1 = 3 + 2i \qquad \alpha_2 = 2 + 1.66i \qquad \alpha_3 = 1.75 + i$$
$$\omega_1 = 0.8\pi \qquad \omega_2 = 1.2\pi \qquad \omega_3 = 1.4\pi$$

where, ϵ_t are i.i.d complex valued gaussian with zero mean s.t.

$$\epsilon_t = \epsilon_{t_R} + i\epsilon_{t_C}$$
$$\epsilon_{t_R} \sim \mathcal{N}(0, \sigma^2/2)$$
$$\epsilon_{t_C} \sim \mathcal{N}(0, \sigma^2/2)$$

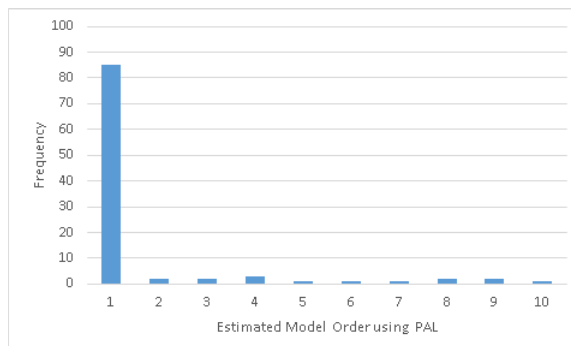

Fig. 1. Histogram for estimated model order over 100 bootstrap samples for a simulation; n=15

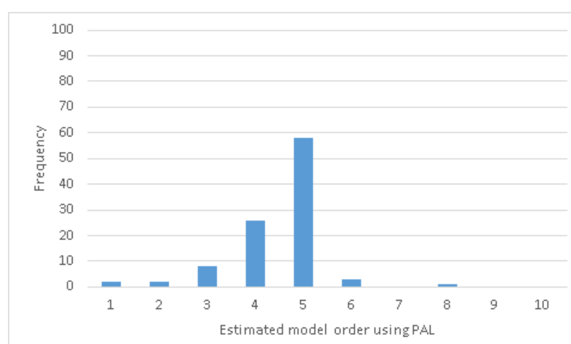

Fig. 2. Histogram for estimated model order over 100 bootstrap samples for a simulation; n=25

and ϵ_{t_R} and ϵ_{t_C} are independent $\forall t = 1 \ldots n$. For the above model, we fix σ^2 at a high value of 30 and vary the sample size n from 15 to 200. We keep the maximum model order fixed at $\tilde{m} = 10$. For 100 simulations, we generate the data, estimate the model order using PAL rule and then

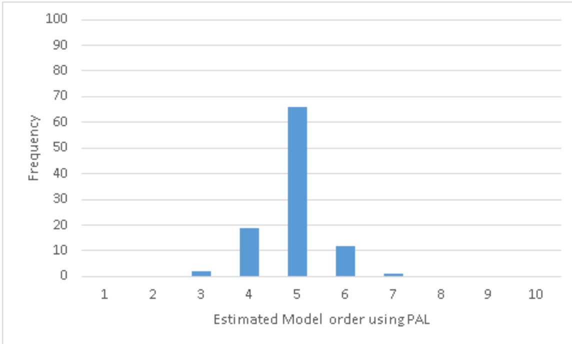

Fig. 3. Histogram for estimated model order over 100 bootstrap samples for a simulation; n=50

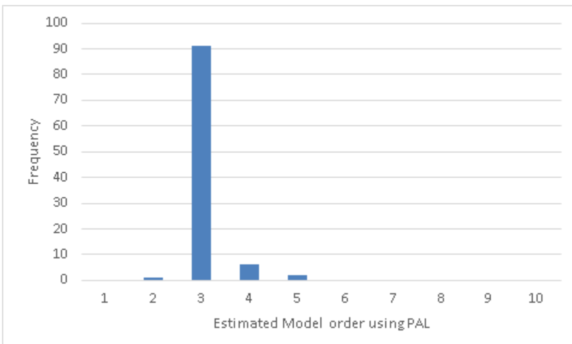

Fig. 4. Histogram for estimated model order over 100 bootstrap samples for a simulation; n=100

do residual bootstrapping for 100 times to get an empirical distribution over the model order which is then further used to define 90% and 95% confidence intervals for the model order. The form of the PAL rule given the model of order m is given as:

$$PAL(m) = -2\ln(f_m(y, \hat{\theta}^*_m)) + (3m+1)\ln(3\tilde{m}+1) * \frac{\ln(r_m+1)}{\ln(\rho_m+1)} \tag{2}$$

where, usual notations defined above are followed and ρ_m

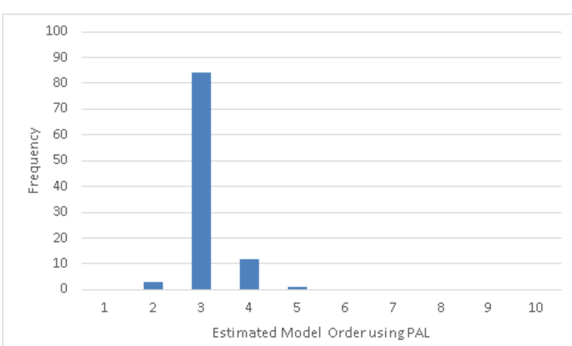

Fig. 5. Histogram for estimated model order over 100 bootstrap samples for a simulation; n=200

is defined as

$$\rho_m = 2\ln\left[\frac{f_{\tilde{m}}(y, \hat{\theta}^*_{\tilde{m}})}{f_{m-1}(y, \hat{\theta}^*_{m-1})}\right] \tag{3}$$

and the usual notations for the rest of the terms as defined above in the paper. We understand that the probability of correct estimation of model order increases as the sample size increase, which validates the consistency results we proved last semester. The confidence interval results for different samples are mentioned below:

TABLE I
AVERAGE QUANTILES FROM 100 BOOTSTRAP SAMPLES OF DIFFERENT SAMPLE SIZES

	Quantile values				
Sample Size	0.025	0.05	0.5	0.95	0.975
15	1.51	1.725	2.2	7.45	8.5
25	1.18	1.265	1.905	3.62	4.34
50	1.56	1.79	2.775	3.995	4.35
100	2.39	2.6	3.06	4.125	4.35
200	2.95	3	3.05	3.915	4.15

We understand the interval gets more precise as the amount of data increase and for a sample size as high as 200, even for a high noise variance of 30, we can accurately predict the model order upto order one with 95% confidence.

III. CONCLUSIONS

In this part, we present residual bootsrapping as a techniques for interval estimation for model order. We understand that the confidence interval improves as the sample size increases.Due to time constraints, we were not able to analyze the effect that noise variance and amplitudes as well as frequency parameters have on the confidence interval of model order. However, they can be looked at as a future work to understand the dependence of distribution of model order on these parameters. We can also look at devising bayesian techniques for model order estimation using a appropriate prior on the same as a future work in interval estimation problem.

References

[1] Stoica, Petre, and Prabhu Babu. "On the exponentially embedded family (EEF) rule for model order selection." IEEE Signal Processing Letters 19.9 (2012): 551-554.

[2] Stoica, Petre, and Yngve Selen. "Model-order selection: a review of information criterion rules." IEEE Signal Processing Magazine 21.4 (2004): 36-47.

[3] Surana, Khushboo, et al. "Estimating the order of sinusoidal models using the adaptively penalized likelihood approach: Large sample consistency properties." Signal Processing 128 (2016): 204-211.

[4] Wilks, Samuel S. "Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution." The Annals of Mathematical Statistics (1946): 257-281.

[5] Lehmann, Erich L., and Joseph P. Romano. "Testing Statistical Hypotheses (Springer Texts in Statistics)." (2005).

[6] Chung, Kai Lai. A course in probability theory. Academic press, 2001.

[7] D. Kundu and A. Mitra, "On Asymptotic Properties and Confidence Intervals for Exponential Signals ", Signal Processing , Vol. 72, No. 2, 129-139, 1999.

[8] Kundu, Debasis, and Amit Mitra. "Asymptotic properties of the least squares estimates of 2-D exponential signals." Multidimensional Systems and Signal Processing 7.2 (1996): 135-150.

[9] Francos, Joseph M., and Mark Kliger. "Strong consistency of the over- and under-determined LSE of 2-D exponentials in white noise." Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.. Vol. 4. IEEE, 2005.

[10] Kliger, Mark, and Joseph M. Francos. "Strong consistency of a family of model order selection rules for estimating 2D sinusoids in noise." Statistics & Probability Letters 78.17 (2008): 3075-3081.

[11] Prasad, Anurag, Debasis Kundu, and Amit Mitra. "Sequential estimation of the sum of sinusoidal model parameters." Journal of Statistical Planning and Inference 138.5 (2008): 1297-1313.

[12] S.He, uniform convergence of weighted periodogram of stationary linear random fields, Chinese annals of mathematics 16B (1995) 331-340

[13] Kay, Steven. "Exponentially embedded families-new approaches to model order estimation." IEEE Transactions on Aerospace and Electronic Systems 41.1 (2005): 333-345.

[14] Schwarz, Gideon. "Estimating the dimension of a model." The annals of statistics 6.2 (1978): 461-464.

[15] Akaike, Hirotugu. "A new look at the statistical model identification." IEEE transactions on automatic control 19.6 (1974): 716-723.

[16] Stoica, Petre, and Prabhu Babu. "Model order estimation via penalizing adaptively the likelihood (PAL)." Signal processing 93.11 (2013): 2865-2871.

[17] Schwarz, Gideon. "Estimating the dimension of a model." The annals of statistics 6.2 (1978): 461-464.

[18] Kashyap, Rangasami L. "Optimal choice of AR and MA parts in autoregressive moving average models." IEEE Transactions on Pattern Analysis and Machine Intelligence 2 (1982): 99-104.

[19] Rissanen, Jorma. "Modeling by shortest data description." Automatica 14.5 (1978): 465-471.

[20] Rissanen, Jorma. "Estimation of structure by minimum description length." Rational Approximation in Systems Engineering. Birkhuser Boston, 1982. 395-406.

[21] Sakamoto, Yosiyuki, Makio Ishiguro, and Genshiro Kitagawa. "Akaike information criterion statistics." Dordrecht, The Netherlands: D. Reidel (1986).

[22] Akaike, Hirotugu. "A new look at the statistical model identification." IEEE transactions on automatic control 19.6 (1974): 716-723.