THE APPLICABILITY OF BIASED ESTIMATION IN MODEL AND MODEL ORDER SELECTION

Weaam Alkhaldi[1], *D. Robert Iskander*[2] *and Abdelhak M. Zoubir*[1]

[1]Signal Processing Group, Institute for Communications, Technische Universität Darmstadt
Merckstr. 25, 64283 Darmstadt, Germany
{alkhaldi,zoubir}@spg.tu-darmstadt.de
[2]Contact Lens & Visual Optics Lab, School of Optometry, Queensland University of Technology
Kelvin Grove Q4059, Brisbane, Australia
d.iskander@qut.edu.au

ABSTRACT

Biased estimation has the advantage of reducing the Mean Squared Error (MSE) of an estimator. The question of interest is how biased estimation affects model selection. In this paper, we introduce biased estimation to a range of model selection criteria. Specifically, we analyze the performance of the Minimum Description Length (MDL) criterion based on biased and unbiased estimation and compare it against modern model selection criteria such as Kay's conditional model order estimator (CME), the Bootstrap and the more recently proposed Hook-and-Loop resampling based model selection. The advantages and limitations of the considered techniques are discussed. The results indicate that, in some cases, biased estimators can slightly improve the selection of the correct model. We also give an example for which the CME with an unbiased estimator fails, but could regain its power when a biased estimator is used.

Index Terms— biased estimation, model selection, model order estimation, bootstrap

1. INTRODUCTION

There are many engineering applications in which there is an interest in determining an optimal set of parameters in a model. Techniques for performing such a task are termed model selection. In many applications the considered model can be expressed in terms of a linear (in the parameters) matrix equation. Consider the linear model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{w},$$

where $\boldsymbol{X}$ is a known full rank matrix of size $N \times p$, $\boldsymbol{\theta}$ is a p vector-valued unknown paramter and $\boldsymbol{w}$ is the noise vector of an unknown distribution with mean zero and covariance $\sigma_w^2\boldsymbol{I}$. In such a model, two fundamental problems arise. Firstly, the question to ask is how to estimate the number of the parameter p, and then whether there is a subset of $\{1, \ldots, p\}$ that will best (in some sense) represent the data. Although in most practical cases, it is sufficient to choose only the model order, p, we also consider selecting a particular model being a subset of $\{1, \ldots, p\}$.

Several model selection procedures can be applied to find the optimal model. Some of the more commonly used model selection procedures include Akaike's information criterion (AIC) [1], Rissanen's minimum description length (MDL) criterion [2], and their variants. Recently, Kay [3], proposed a conditional model order estimator (CME) that is based on the theory of sufficient statistics. Also, bootstrap based model selection techniques have been widely used in many practical problems [4, 5]. They proved to be asymptotically more accurate than other techniques [6]. One drawback of the bootstrap for model selection is the necessity of selecting a suitable scaling parameter for the detrended residuals in the resampling procedure. A resampling scheme, called the " Hook-and-Loop" (HL), has been recently proposed to avoid scaling problems [7]. The HL resampling scheme has shown superior performance to its bootstrap predecessor in the context of model selection and model order selection [8].

Biased estimators have recently shown superior perfromance as compared to minimum variance unbiased (MVU) estimators in view of minimizing the MSE of an estimator [9]. This is accomplished by scaling (i.e., shrinking) the unbiased estimator in order to get a biased estimator which has a smaller MSE than the unbiased one. It has been proved in [9] that for a wide range of theoretical models, such a scaling factor can be found using convex optimization techniques (solving a minimax problem).

In this work, we investigate the applicability of biased estimation in the context of model and model order selection. We compare the performance of several model selection techniques (MDL, CME, Bootstrap and HL) derived from unbiased and biased estimators. We show by simulation that biased estimation based model selection techniques can outperform unbiased estimation based model selection.

2. METHODOLOGIES

Let β, which we call from now on the model, be a subset of $\{1, \ldots, p\}$. Subsequently, let $\boldsymbol{X}_\beta$ be a sub-matrix of $\boldsymbol{X}$ and $\boldsymbol{\theta}_\beta$ be a sub-vector of $\boldsymbol{\theta}$, both containing the components indexed by the integers in β. Then, a model corresponding to β is [5]

$$\boldsymbol{y} = \boldsymbol{X}_\beta \boldsymbol{\theta}_\beta + \boldsymbol{w}.$$

The aim of model selction is to choose, or more precisely estimate, model β so that it best represents the data.

The MDL and the CME [3] criteria find the model that minimizes

$$\text{MDL}(\beta) = \frac{N}{2} \ln \hat{\sigma}_\beta^2 + \frac{p_\beta + 1}{2} \ln N,$$

and

$$\begin{aligned} \text{CME}(\beta) \quad &= \quad \frac{N - p_\beta - 2}{2} \ln \hat{\sigma}_\beta^2 + \frac{1}{2} \ln |\boldsymbol{X}'_\beta \boldsymbol{X}_\beta| \\ &+ \quad \ln \frac{[\pi(N-p_\beta)]^{(N-p_\beta)/2}}{\Gamma\left(\frac{N-p_\beta}{2}\right)}, \end{aligned}$$

respectively. Herein, p_β is the number of elements in the model β, $\Gamma(\cdot)$ is the Euler's Gamma function, and

$$\hat{\sigma}_\beta^2 = \frac{1}{N - p_\beta} (\boldsymbol{y} - \boldsymbol{X}_\beta \hat{\boldsymbol{\theta}}_\beta)' (\boldsymbol{y} - \boldsymbol{X}_\beta \hat{\boldsymbol{\theta}}_\beta).$$

In the above, $\hat{\boldsymbol{\theta}}_\beta$ is simply the least-squares estimator of the parameter vector $\boldsymbol{\theta}$ conditioned on the model β, given by

$$\hat{\boldsymbol{\theta}}_\beta = \left(\boldsymbol{X}'_\beta \boldsymbol{X}_\beta\right)^{-1} \boldsymbol{X}'_\beta \boldsymbol{y}.$$

It should be noted that CME has been designed in the context of a full model and as such can only be used for model order estimation.

For the Bootstrap approach, given observations

$$y_1, y_2, \ldots, y_N,$$

we first calculate the least-squares estimate $\hat{\boldsymbol{\theta}}_\alpha$ and the resulting residuals

$$\hat{w}_t = y_t - \boldsymbol{X}'_{\alpha t} \hat{\boldsymbol{\theta}}_\alpha, \quad t = 1, 2, \ldots, N,$$

where $\alpha = \{1, 2, \ldots, p\}$ is the full model and $\boldsymbol{X}'_{\alpha t}$ is the tth row of $\boldsymbol{X}_\alpha$. Then, we resample with replacement from

$$\sqrt{N/m}(\hat{w}_t - \hat{w}.)/\sqrt{1 - p/N},$$

$t = 1, 2, \ldots, N$ to obtain $\hat{w}_t^*$, where $\hat{w}. = 1/N \sum_{t=1}^{N} \hat{w}$. A scaling parameter m is introduced such that $m/N \to 0$ and

$$\frac{N}{m} \max_{t \leq N} \boldsymbol{X}'_{\beta t} (\boldsymbol{X}'_\beta \boldsymbol{X}_\beta)^{-1} \boldsymbol{X}_{\beta t} \to 0$$

for all β.

In the HL resampling scheme, we sort the detrended residuals

$$\hat{w}_1 - \hat{w}., \hat{w}_2 - \hat{w}., \ldots, \hat{w}_n - \hat{w}.$$

in an increasing order to obtain a set of residuals

$$\hat{w}_{(1)} < \hat{w}_{(2)} < \ldots < \hat{w}_{(N)}$$

and generate a new HL sample using, for example,

$$\hat{w}_{(i)}^* = \frac{1}{2}\left(w_{(i)} + w_{(i+1)}\right) + \varepsilon_i$$

where

$$\varepsilon_i \sim \mathcal{N}\left(0, \left[\frac{1}{6}\left(w_{(i+1)} - w_{(i)}\right)\right]^2\right).$$

The HL residuals are then further ordered according to the strength of the signal y_t [7]. Next, we compute

$$y_t^* = \boldsymbol{X}'_{\beta t} \hat{\boldsymbol{\theta}}_\beta + \hat{v}_t^*, \quad t = 1, 2, \ldots, N$$

where $\hat{v}_t^*$, $t = 1, 2, \ldots, N$, denote either the Bootstrap or the HL residuals, and the least-squares estimate $\hat{\boldsymbol{\theta}}_{\beta,m}^*$ is calculated from $(y_t^*, \boldsymbol{X}_{\beta t})$. Resampling is repeated B times to obtain $\hat{\boldsymbol{\theta}}_{\beta,m}^{*(i)}$ and the bootstrap estimate of the residual squared error

$$\hat{\Gamma}_{N,m}^{*(i)}(\beta) = \frac{\|\boldsymbol{y} - \boldsymbol{X}_\beta \hat{\boldsymbol{\theta}}_{\beta,m}^{*(i)}\|^2}{N}, \quad i = 1, \ldots, B.$$

Finally, we average $\hat{\Gamma}_{N,m}^{*(i)}(\beta)$ over $i = 1, \ldots, B$ to obtain $\bar{\Gamma}_{N,m}^*$ and minimize over β to obtain $\hat{\beta}_0$.

To implement the biased estimator ($\hat{\theta}_b$), we assume it is a scaled version of the unbiased one ($\hat{\theta}_u$) as given by [9]

$$\hat{\theta}_b = (1 + k)\hat{\theta}_u. \tag{1}$$

where k is the scaling factor. The MSE of the biased estimator can be expressed as

$$\text{MSE}(\hat{\theta}_b) = (1 + k)^2 \text{var}(\hat{\theta}_u) + k^2 \theta^2. \tag{2}$$

We assume that our true parameter, θ, equals $\hat{\theta}_u$ and we simply search for the value k which minimizes Eq. (2). Therefore, we can calculate the biased estimator by substituting in Eq. (1). Obtaining the model estimator or the model order estimator is hence straightforward as we have to repeat calculations for the MDL, the CME, the Bootstrap and the HL techniques using the biased estimator, $\hat{\theta}_b$.

3. SIMULATION RESULTS WITH DISCUSSION

We present the simulation results for two examples in the context of model and model order selection. We compare the performance of the different approaches described above across

Table 1. Percentages of selecting the correct model evaluated over 1000 independent Monte Carlo runs for the polynomial of Eq. (3) with Gaussian noise. The highest probabilities achieved for each of the cases are indicated in bold fonts.

N	SNR [dB]	MDL U	MDL B	Bootstrap U	Bootstrap B	HL U	HL B
32	0	0.5	0.5	0.0	0.0	**1.0**	**1.0**
	10	14.7	11.5	0.5	0.4	**52.1**	45.0
	20	87.1	88.5	74.9	75.1	**95.0**	28.0
	30	95.5	97.1	99.5	**99.7**	97.4	93.5
	40	95.9	97.7	99.7	**99.8**	97.8	94.6
64	0	39.4	39.1	0.5	0.3	**47.4**	38.7
	10	96.2	96.2	74.3	73.9	**97.2**	96.1
	20	96.7	97.9	**99.7**	**99.7**	**99.7**	98.7
	30	98.1	99.4	99.9	**100.0**	**100.0**	98.8
	40	98.1	99.2	99.7	**99.9**	99.8	98.8
128	0	46.5	48.0	0.1	0.3	**50.5**	18.4
	10	90.6	**91.2**	63.1	61.9	82.3	77.2
	20	98.8	99.1	99.5	99.4	**100.0**	**100.0**
	30	98.9	99.3	99.9	**100.0**	**100.0**	**100.0**
	40	98.3	99.0	99.6	99.6	**100.0**	**100.0**

different Signal-to-Noise Ratios (SNR) for both biased and unbiased estimators. In all simulations, the Bootstrap and the HL repetitions are set to $B = 100$ and the scaling paramter, m, is set to 2 for the Bootstrap procedure [5], [7] and [8]. Only Gaussian noise models with identically and independently distributed (i.i.d.) random variables are considered here. It was shown in [6] that bootstrap techniques perform better than information criteria for model selection in the non-Gaussian case. It is not the objective to demonstrate here that the bootstrap is superior also when using biased estimators. In this work, we are primarily concerned with the effect of biased estimation on some model selectors.

3.1. Model Selection

We choose a polynomial from [5] given by

$$y_t = 0 + 0 + 0.035t^2 - 0.0005t^3 + w_t, \quad t = 1, \dots, N, \quad (3)$$

to evaluate the performance of biased (B) and unbiased (U) estimators for model selection. The model is then normalized over t. Table 1 shows the probabilities of selecting the correct model using the MDL, the Bootstrap and the HL. Different signal lengths ($N = 32$, $N = 64$ and $N = 128$) are used with an SNR ranging between 0 dB and 40 dB.

We observe that for an SNR larger than 20 dB, the biased estimators yield higher probabilities of selecting the correct model than the unbiased ones for the MDL criterion, but not necessarily for the Bootstrap and HL model selction schemes. For the Bootstrap, the application of a biased estimator increases sometimes the probability of selecting the correct model, but it does not for the HL method when the probability of correct detection decreases, except when both the SNR and the sample size are high. It is worth noting that a biased estimator for Bootstrap model selection leads to higher probabilities as compared to the biased ones for the MDL criterion. In Table 1, the highest probabilities achieved for each of the cases are indicated in bold fonts, showing that the overall performance achieved with the HL method based on unbiased estimators is superior to the MDL and the bootstrap based technique counterparts.

Table 2. Percentages of correct selection of the model order evaluated over 1000 independent Monte Carlo runs for the model in Eq. (4) with $w(t)$ Gaussian. The highest probabilities achieved for each of the cases are indicated in bold fonts.

σ^2	MDL U	MDL B	CME U	CME B
10	94.0	97.7	**100.0**	**100.0**
15	92.5	94.4	**100.0**	**100.0**
20	93.2	92.6	**100.0**	99.9
25	94.3	90.9	**100.0**	99.9
30	92.3	89.1	**99.9**	99.6
35	92.9	86.8	**99.8**	98.4
40	93.8	86.3	**99.7**	96.2
45	93.6	84.1	99.3	**93.5**
50	92.4	78.1	**99.3**	87.9

3.2. Model Order Selection: Example 1

For evaluating the performance of the biased estimators in the context of non-iterative model order selection (i.e., the MDL and CME criteria), we choose the polynomial trend from [3] in noise, given by

$$y_t = 0 + 0.4t + 0.1t^2 + w_t, \quad (4)$$

where $t = 1, \dots, N$. In Table 2, We compare the probability of selecting the correct model order (second order in this case) for a signal length of $N = 30$ and a noise variance, ranging from $\sigma^2 = 10$ to $\sigma^2 = 50$. The highest possible model order was set to 10.

We observe in Table 2 that the unbiased estimators lead to a CME which performs better than the MDL (as proved in [3]). Moreover, biased estimators for the CME yield higher accuracies than their counterparts of the MDL for all noise variances. However, biased estimators give lower probabilities than unbiased ones for both the CME and the MDL criteria.

3.3. Model Order Selection: Example 2

We use the polynomial of Eq. (3) to evaluate the performance of biased and unbiased estimators for model order selection

Table 3. The percentages of selecting the correct model order evaluated over 1000 independent Monte Carlo runs for the polynomial of Eq. (3) with Gaussian noise. The highest probabilities achieved for each of the cases are indicated in bold fonts.

	SNR	MDL		CME		Bootstrap		HL	
N	[dB]	U	B	U	B	U	B	U	B
32	0	**0.9**	0.8	0.0	0.0	0.0	0.0	0.0	0.0
	10	**3.3**	2.1	0.0	0.0	0.2	0.0	0.0	0.5
	20	**35.9**	24.2	0.0	3.5	13.5	6.6	29.0	32.0
	30	**96.4**	79.1	0.0	47.8	95.2	38.1	84.5	44.4
	40	93.8	97.5	0.0	91.1	**97.8**	96.1	91.9	82.2
64	0	**3.9**	3.2	0.0	0.0	0.0	0.0	0.0	0.0
	10	**88.9**	41.7	0.0	1.6	12.2	4.7	0.0	0.0
	20	96.6	79.7	0.0	29.4	**98.3**	73.2	11.7	12.5
	30	98.5	97.0	0.0	94.0	**99.2**	95.4	91.6	70.5
	40	97.2	**99.5**	0.0	98.7	98.7	99.3	89.9	60.2
128	0	**18.7**	12.5	0.0	0.1	0.0	0.1	0.0	0.0
	10	**98.0**	41.0	0.0	2.2	8.7	4.8	0.0	0.0
	20	98.0	85.6	0.0	46.5	**98.6**	73.9	60.0	10.7
	30	98.3	99.2	0.0	89.4	99.3	99.7	**100.0**	66.1
	40	98.6	99.2	0.0	97.2	98.5	99.6	**100.0**	77.0

(third order in this case) using the MDL, the CME, the Bootstrap and the HL. Different signal lengths ($N = 32$, $N = 64$ and $N = 128$) are used with an SNR, ranging between 0 dB and 40 dB, as shown in Table 3. The highest considered model order was set to 10.

The example chosen here shows a breakdown of the CME, which always selects the highest possible model order (i.e., 10 rather than 3). Interestingly, the CME regains some of its power when a biased estimator is used, but its performance is still lower than that of the corresponding MDL. This indicates that biased estimates for the CME lead to better results than unbiased estimates. The results also indicate that biased estimators can lead to a slightly higher probability of selecting the correct model order at high SNR, but not for the HL method. Overall the performance of the unbiased estimator based MDL is best for a low SNR while the Bootstrap and HL excel at a higher SNR.

4. CONCLUSIONS

We have investigated the applicability of biased estimation in model and model order selection. A model selection is often based on the minimization of a MSE based criterion. Hence, further minimization of the MSE via biased estimation could lead to better results. This study was undertaken to assess whether biased estimation would affect the probability of selecting the correct model or model order. No theoretical results for the consistency for model selection with biased estimators have been reported here. We primarily were interested in evaluating the behavior of a few model selection techniques in view of biased estimation. Techniques such as AIC, the corrected AIC, and Hannan and Quinn's criterion yielded very similar results to those of the MDL and hence we restricted our choice here to the MDL criterion. Emphasis was given to modern techniques such as the CME, the Bootstrap and the HL model selectors. We have shown that model selection with biased estimators can sometimes slightly increase the probability of selecting the correct model but this is not true for model order selection, except for the CME. The results of this paper highlight the continued difficulty in selecting an appropriate model order selection technique in practice.

5. REFERENCES

[1] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, vol. 19, pp. 716–723, 1974.

[2] J. Rissanen. A universal prior for integers and estimating by minimum description length. *The Annals of Statistics*, vol. 11, pp. 416–431, 1983.

[3] S. Kay. Conditional model order estimation. *IEEE Transactions on Signal Processing*, vol. 49, no. 9, pp. 1910–1917, 2001.

[4] A. M. Zoubir and D. R. Iskander. The bootstrap: A tutorial for the signal processing practitioner. *IEEE Signal Processing Magazine*, 24(4):10-19, July 2007.

[5] A. M. Zoubir and D. R. Iskander, *Bootstrap techniques for signal processing*. Cambridge University Press, Cambridge, UK, 2004.

[6] A. M. Zoubir. Bootstrap Methods for Model Selection. *AEÜ International Journal of Electronics and Communications*, 53(6):386–392, December 1999.

[7] D. R. Iskander and W. Alkhaldi. The "hook and loop" resampling plane. *In The Proceedings IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 65–68, St. Thomas, U.S. Virgin Islands, December 2007.

[8] D. R. Iskander, W. Alkhaldi and A. M. Zoubir. On the computer intensive methods in model selection. In *The Proceedings of the 33rd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3461–3464, Las Vegas NV, USA, March 2008.

[9] S. Kay and Y. Eldar. Rethinking biased estimation. *IEEE Signal Processing Magazine*, vol. 25, no. 3, pp. 133–136, May 2008.