Parameter Magnitude-Based Information Criterion in Identification of Discrete-Time Dynamic System

Md Fahmi Abd Samad, Abdul Rahman Mohd Nasir*
Faculty of Mechanical Engineering, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Malacca, Malaysia
**mdfahmi@utem.edu.my*

ABSTRACT

Information criterion is an important factor for model structure selection in system identification. It is used to determine the optimality of a particular model structure with the aim of selecting an adequate model. A good information criterion not only evaluate predictive accuracy but also the parsimony of model. There are many information criterions those are widely used such as Akaike information criterion (AIC), corrected Akaike information criterion (AICc) and Bayesian information criterion (BIC). This paper introduces a new parameter-magnitude based information criterion (PMIC2) for identification of linear and non-linear discrete time model. It presents a study on comparison between AIC, AICc, BIC and PMIC2 in selecting the correct model structure for simulated models. This shall be tested using computational software on a number of simulated systems in the form of discrete-time models of various lag orders and number of terms/variables. It is shown that PMIC2 performed in optimum model structure selection better than AIC, AICc and BIC.

Keywords: *Akaike Information Criterion, Bayesian Information Criterion, Model Structure Selection, Parameter Magnitude-Based Information Criterion*

Introduction

System identification can be defined as approximating dynamic system models using experimental data [1]. Its basic idea is to compare the time dependent responses of the actual system and identified model based on a

ISSN 1823- 5514, eISSN 2550-164X

Received for review: 2017-04-13
Accepted for publication: 2017-05-18
Published: 2017-08-15

performance function. Referring to information criterion, it measures how well the model response fits the system response [2]. Data acquisition, model structure selection, parameter estimation and model validity tests is the procedure of identification steps [3]. Normally, the identification procedure consists of estimating parameters of different models, then selecting the optimum model complexity within the set. The systematic errors will decrease when increasing the model complexity and at the same time the model variability also increases [4].

Model accuracy and model parsimony are also known as variance and bias: $f(J) = Var(J) + Bias(J)$. These two considerations need to be evaluated when selecting a model structure [5]. Therefore, selecting the model with the smallest variance within the set will not be good because when more parameters are added, it will continue to decrease. At a certain complexity, the additional parameters no longer reduce the systematic errors but are used to follow the actual noise realization on the data. To overcome this problem, such loss function or information criterion is extended with a bias term in order to compensate the model complexly. In short, the information criterion should be able to detect undermodelling (too simple model) as well as overmodelling (too complex model). Undermodelling occurs when the considered model does not include the true model and the number of parameters is less than the true model e.g. when there is unmodelled dynamics and/or nonlinear distortions in linear system identification, or too small a number of sine waves and/or nonperiodic deterministic disturbances in signal modelling. Overmodelling is described by too many parameters and it occurs when the considered model includes the true model [4]. Here, a study on the effectiveness of information criterion is warranted.

This paper studies the effectiveness of Akaike information criterion (AIC) [6], corrected Akaike information criterion (AICc) [7], Bayesian information criterion (BIC) [8] and parameter magnitude-based information criterion (PMIC2), [9]. All these information criterions are compared by testing on four simulated dynamic models in the form of difference equations model. Linear and nonlinear autoregressive models with exogenous input (ARX and NARX) are used in this simulation [5]. The benefit of using simulated models is the presence of an opportunity to compare the final model directly with the true model.

The next sections are as follows: Section 2 explains system identification; Section 3 lays out the information criterions; Section 4 explains the simulated models; Section 5 provides results and discussion and lastly Section 6 concludes the paper along with recommendation of future works.

System Identification

System identification is a pre-requisite to analysis of a dynamic system and design of an appropriate controller for improving its performance. The more accurate the mathematical model identified for a system, the more effective will be the controller designed for it [12]. Often, in order to deal with the bias-variance trade-off, the loss function or information criterion is augmented with a penalty term intended to guide the search for the "optimal" relationship penalizing undesired regressors, where regressors refer to possible terms and variables identified from model order and linearity specifications. Regularized estimation has been widely applied also in the context of system identification [13]. Several strategies have been proposed to avoid over-parameterization while utilizing all the data for training the model. The most popular strategy is to minimize a theoretically derived formula or criterion, which includes a goodness-of-fit index and a penalty factor for model complexity [14]. System identification can be framed as an optimization problem, as in Equation (1):

$$\hat{\theta} = \arg \underset{\theta}{min} J_F\,(\theta, D_N) \tag{1}$$

where $J_F\,(\theta, D_N)$ measure how well the model described by parameter θ describes the measured data. A widely used variation of the estimation criterion includes a so-called 'regularization term' in the loss function to be minimized, as in Equation (2):

$$\hat{\theta} = \arg \underset{\theta}{min} J_F\,(\theta, D_N) + J_R\,(\theta, \mathrm{n}) \tag{2}$$

In this case, $\hat{\theta}$ is estimated by trading-off the data fitting term $J_F\,(\theta, D_N)$and the regularization term $J_R\,(\theta, \mathrm{n})$ which act as a penalty to penalize certain parameters vectors θ which describe 'unlikely' systems [11].

In today's literature, various types of models are proposed for system modelling such as linear autoregressive with exogenous input (ARX) model and nonlinear autoregressive with exogenous input (NARX) model [5].

Information Criterion

Model complexity selection is the sub-problem of model selection [15]. Parsimony, working hypotheses, and strength of evidence are three principles that regulate the ability to make inferences [16]. To overcome this, many information criterions were developed such as AIC, AICc, BIC and PMIC2. Estimation of the Kullback-Leibler information is the key to deriving the

AIC, which was the first model selection criterion to gain widespread acceptance [17]. AIC is written as in Equation (3):

$$AIC = n\, ln\, \frac{RSS}{n} + 2p \tag{3}$$

where n is the number of observations, RSS is the residual sum of square, while RSS/n is the maximized value of the likelihood function for the estimated model and p is the number of parameters in the statistical model [6]. RSS can be defined as in Equation (4):

$$RSS = \sum_{t=k}^{N} \varepsilon^2(t) = \sum_{t=k}^{N} (y(t) - \hat{y}(t))^2 \tag{4}$$

where $\varepsilon(t)$ is the residual; $\hat{y}(t)$ and $y(t)$ are the k-step-ahead predicted output and actual output value at time t, respectively; and N is the number of data. The k-step-ahead prediction is used when the value of k depends on the output's smallest lag order in the selected model structure, which in turn depends on the variables selected by the search method.

Although AIC has been proven to be widely applicable, it can have serious deficiencies. Indeed, AIC was designed as an approximately unbiased estimator of the expected Kullback-Leibler divergence between the generating model and the fitted approximating model under the assumption that the true model is correctly specified or overfitted [18]. When the sample size is small or when the number of fitted parameters is a moderate to large fraction of the sample size, AIC becomes a strongly negatively biased estimate of the Kullback-Leibler divergence and leads to the choice of overparameterized models [15].

From [7], improvement has been made to AIC called corrected Akaike information criterion (AICc). Different approaches have been made to improve AIC by correcting its penalty term. One such approach is to asymptotically evaluate the penalty term as precisely as possible to provide better estimates of the model order [17]. AICc can be written as in Equation (5):

$$AICc = n\, ln\, \frac{RSS}{n} + \frac{2k(k+1)}{n-k-1} \tag{5}$$

Example of another widely used criterion is BIC which is based, respectively, on Bayesian and coding theory. The BIC is a likelihood criterion penalized by the model complexity. The penalty in BIC for additional parameters is known to be stronger than that of the AIC. The BIC is an asymptotic result derived under the assumptions that the data distribution is in the exponential family [8]. BIC is defined as in Equation (6):

$$BIC = n \ln \frac{RSS}{n} + \mathrm{k}\ \ln(n) \tag{6}$$

The PMIC2 is developed from the approach of using parameter magnitude information in information criterion [10, 11]. The bias term or known as penalty function are based on the magnitude of the parameters themselves, considering models with terms/variables with small parameter values to be discarded in the selection. It is written as in Equation (7):

$$PMIC2 = \sum_n (y(t) - \hat{y}(t))^2 + \sum_j \frac{1}{\theta_j} \tag{7}$$

where, θ_j is the magnitude of parameter in the model and j is the number of parameter.

Simulation Setup

In this simulation, three ARX models and a NARX model are simulated using computer simulation software MATLAB. All models are denoted as Model 1, Model 2, Model 3 and Model 4 and each model is further classified as having d.c. level and not having d.c. level. The difference between the two are one having the output and input average subtracted (hence has no d.c. level) and the other not subtracted. The following are the models written as linear regression models, its specifications, number of correct regressors and number of possible regressors:

Model 1:

$y(t) = 0.1y(t-2) + 0.3u(t-1) + 0.8u(t-3) + e(t)$

Specification: l=1, assumed maximum output order, n_y=3, maximum input order, n_u=3

Number of correct regressor = 3 out of 7 (if d.c. level is assumed present) or 6 (if d.c. level is assumed absent)

Number of possible model =127 (with d.c. level) or 63 (without d.c. level)

Model 2:

$y(t) = 0.1y(t-1) + 0.4y(t-5) - 0.3u(t-3) + 0.5u(t-4) + e(t)$

Specification: l=1, n_y=5, n_u=5

Number of correct regressor = 4 out of 11 or 10

Number of possible model =2047 or 1023

Model 3:

$y(t) = 0.1y(t-2) + 0.3y(t-4) - 0.3y(t-7) + 0.2u(t-3)$
$+0.3u(t-5) + e(t)$

Specification: l=1,n_y=7, n_u=7
Number of correct regressor = 5 out of 15 or 14
Number of possible model = 32767 or 16383

Model 4:

$$y(t) = 0.3y(t-2) - 0.4u(t-2) + 0.1y(t-1)y(t-2)$$
$$+0.1y(t-1)u(t-2) - 0.3u(t-2)u(t-2) + e(t)$$

Specification: l =2, n_y=2, n_u=2
Number of correct regressor = 5 out of 15 or 14
Number of possible model = 32767 or 16383

The input $u(t)$ is generated from a random uniform distribution in the interval [-1, 1] to represent white signal, while noise $e(t)$is generated from a random uniform distribution [-0.01, 0.01] to represent white noise. Five hundred data points are generated for each model. All models are evaluated by AIC, AICc, BIC and PMIC2, respectively, in order to identify the assumed optimum model.

Result and Discussion

In this section, comparisons are made between AIC, AICc, BIC and PMIC2 for all models based on their selected model. The selected models are based on the minimization of respective criterions. Tables 1 to 4 show the results. Models with d.c. level are denoted as Model 1a, Model 2a, Model 3a and Model 4a while models without d.c. level are denoted as Model 1b, Model 2b, Model 3b and Model 4b. The simulated model is denoted as S.M. For brevity, the variables and terms are not included but numbered.

Table 1: Results on Model 1

Model 1a	**Regressor Number**						
	1	**2**	**3**	**4**	**5**	**6**	**7**
S.M			0.1		0.3		0.8
AIC, AICc, BIC	-0.02			0.1			
PMIC2			0.1		0.3		0.8

Model 1b	**Regressor Number**					
	1	**2**	**3**	**4**	**5**	**6**
S.M		0.1		0.3		0.8
AIC, AICc, BIC	0.02				-0.02	
PMIC2		0.1		0.3		0.8

Table 2: Results on Model 2

Model 2a	**Regressor Number**										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
S.M		0.1				0.4			-0.3	0.5	
AIC, AICc, BIC	-0.6		-0.06			0.3	0.01	0.01	-0.3		
PMIC2						0.4			-0.3	0.5	

Model 2b	**Regressor Number**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
S.M	0.1				0.4			-0.3	0.5	
AIC, AICc, BIC			-0.02			0.02		-0.3		
PMIC2					0.4			-0.3	0.5	

Table 3: Results on Model 3

Model 3a	**Regressor Number**														
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
S.M			0.1		0.3			-0.3			0.2		0.3		
AIC, AICc, BIC	-0.2		0.7									0.02			
PMIC2					0.37			-0.31			0.2		0.32		

Model 3b	**Regressor Number**													
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
S.M		0.1		0.3			-0.3			0.2		0.3		
AIC, AICc, BIC	-0.01	1.3	-0.01	0.1	-0.1	-0.4	-0.04		0.01	0.2			0.01	-0.4
PMIC2				0.37			-0.31			0.2		0.32		

Table 4: Results on Model 4

Model 4a	**Regressor Number**														
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
S.M			0.3		-0.4		0.1			0.1					-0.3
AIC, AICc, BIC	-1.0		0.14	0.1		0.01	0.1	-0.01				-0.1		0.01	
PMIC2			0.3		-0.4		0.1		0.1						-0.3

Model 4b	**Regressor Number**													
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
S.M		0.3		-0.4		0.1			0.1					-0.3
AIC, AICc, BIC		0.2		0.2	0.01			0.14		0.06		-0.02		
PMIC2		0.39		0.16	0.04	0.1		0.1	0.04			0.07		-0.24

Figure 1 shows the comparison of number of regressors selected between AIC, AICc, BIC and PMIC2. From the tables and figure, AIC, AICc and BIC showed the same outcome but PMIC2 performed differently. As can be seen, PMIC2 selected the same model as the simulated model in Model 1a and Model 1b but AIC, AICc and BIC selected too simple model. However, for Model 2 and Model 3, PMIC2 almost selected a model as the simulated model where only one out of all regressors for all models has not been selected. It means that PMIC2 selected a more parsimonious model than simulated model for model 2 and model 3 while AIC, AICc and BIC cannot select the correct model where either under-modelling or overmodelling occurred in the case of the ARX models.

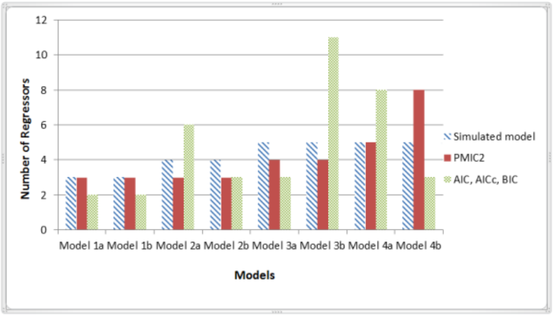

Figure 1: A comparison of regressor number selected between AIC, AICc, BIC and PMIC2 for all models

Besides, at Model 4b, AIC, AICc and BIC selected a wrong model, however, the number of regressor was not too far different from the right number. This occurred when AIC, AICc and BIC selected a model with 6 regressors out of 14 while simulated model has 5 regressors. Overmodelling occurred for these criterions in the case of Model 4a. PMIC2 almost able to select the right model on Model 4a with only one regressor different which is $0.1y(t-1)u(t-1)$ instead of $0.1y(t-1)u(t-2)$. PMIC2 choose too complex model with 8 regressors out of 14 in the case of Model 4b. Among these 8, the true regressors were selected. Overall, PMIC2 capability in selecting the correct model can be considered much better than AIC, AICc and BIC.

Conclusion

From this simulation, PMIC2 proved that it can select better models than AIC, AICc and BIC in order to choose the correct linear model (ARX model). However, in simulation of nonlinear model (NARX model), PMIC2 almost picked the right model on Model 4a but overmodelling occurred when selecting Model 4b, a NARX model assuming no d.c. level. It is noted that although these two models are the same, they constitute different data (due to the step at which the averages were subtracted) causing information criterions to behave differently. More rigorous analysis could be made to identify such weakness and therefore improve the capability of information criterions.

Acknowledgement

The authors would like to acknowledge all the support from Universiti Teknikal Malaysia Melaka and Ministry of Higher Education Malaysia for research grant FRGS/1/2015/TK03/UTEM/02/16.

References

[1] Saleem, B. Taha, T. Tutunji, and A. AI-Qaisia, "Identification and cascade control of servo-pneumatic system using particle swarm optimization", *Simulation Modelling Practice and Theory*, vol. 52, pp. 164-179, March 2015.

[2] Alfi, and M. M. Fateh, "Parameter identification based on a modified PSO applied to suspension system", *Journal of Software Engineering & Applications*, 3, 221-229 (2010).

[3] K. J. Keesman, *System Identification: An Introduction*, (Springer Science & Business Media, 2011).

[4] F. D. Riddef', R. Pintelon, J.Schoukens, and D. P. Gillikinb, "Modified AIC and MDL model selection criteria for short data records", *IEEE Transactions on Instrumentation and Measurement,* 54, 144-150 (2004).

[5] L. Ljung, *System Identification: Theory for the User,* 2nd ed., (Prentice Hall, Upper Saddle River, 1999).

[6] H. Akaike, "A new look at the statistical model identification", *IEEE Transactions on Automatic Control*, 19(6) (1974).

[7] D. R. Anderson, K. P. Burnham and G. C. White, "AIC model selection in overdispersed capture–recapture data", *Ecology*, 75, 1780-1793 (1994).

[8] G. Schwartz, "Estimating the dimension of a model", *The annals of statistics*, 6(2), 461-464 (1978).

[9] M.F. Abd Samad and A.R. Mohd Nasir, "Parameter-Magnitude based information criterion for identification of linear discrete-time model", *Proceedings of Innovative Research and Industrial Dialogue'16,* pp. 129-130, May 2017.

[10] M.F. Abd Samad, H. Jamaluddin, R. Ahmad, M.S. Yaacob and A.K.M. Azad, "Effect of penalty function parameter in objective function of system identification", *Int. J. of Automotive and Mechanical Engineering*, 7, 940-954 (2013).

[11] M.F. Abd Samad and A.R. Mohd Nasir, "Comparison of information criterion on identification of discrete-time dynamic

system", *Int. J. of Engineering and Technology*. To be published.

[12] Z. Zibo and F. Naghdy, "Application of genetic algorithms to system identification", *IEEE International Conference*, 2, 777-782 (1995).

[13] G. Prando, G. Pillonetto and A. Chiuso, "The role of rank penalties in linear system identification", *IFAC-Papers online*, 48(28), 1293-1300 (2015).

[14] X. Xinshu, L. Ying, and R. Mukkamala, "A model order selection criterion with applications to cardio-respiratory-renal systems", *IEEE Transactions on Biomedical Engineering*, 52(3), 445-453 (2005).

[15] K. Kristinsson and G. A. Dumont, "System identification and control using genetic algorithms", *IEEE Transactions on Systems, Man, and Cybernetics*, 22(5), 1033-1046 (1992).

[16] L. Xiaoyong, F. Huajing and C. Zhaoxu, "A novel cost function based on decomposing least-square support vector machine for Takagi-Sugeno fuzzy system identification", *IET Control Theory & Applications*, 8(5), 338-347 (2013).

[17] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed., (Springer-Verlag, New York, 2002).

[18] *A.* K. Seghouane and S. I. Amari, "The AIC Criterion and Symmetrizing the Kullback– Leibler Divergence", *IEEE Transactions on Neural Networks*, 18(1), 97-106 (2007).