Thailand Statistician
July 2014; 12(2): 161-178
http://statassoc.or.th
Contributed paper

Study on the penalty functions of model selection criteria

Warangkhana Keerativibool
Department of Mathematics and Statistics, Faculty of Science, Thaksin University, Phatthalung 93110 Thailand.
E-mail: warang27@gmail.com

Received: 19 July 2013
Accepted: 2 December 2014

Abstract

The aim of this paper is to study the penalty functions of the well-known model selection criteria, AIC, BIC, and KIC, which can unify their formulas as

$$APIC\alpha = \log\left(\hat{\sigma}^2\right) + \alpha\left(p+1\right)/n,$$

called Adjusted Penalty Information Criterion. The appropriate value of α for $APIC\alpha$ has been found to reduce the probabilities of over- and underfitting and also to overcome the weak signal-to-noise ratio. The value of α is selected based on four measurements: the probability of over- and underfitting, the signal-to-noise ratio, the probability of order selected, and the observed L_2 efficiency. Performance of $APIC\alpha$ is examined by theoretical and extensive simulation study. The theoretical results show that, the probability of overfitting tends to zero and the signal-to-noise ratio tends to strong if the value of α tends to infinity. However, the simulation results show that, when the true model is weakly identifiable, the small value of α gives a high probability of correct order being selected. But, if the true model is very difficult to detect, the observed L_2 efficiency is a meaningful measurement than the probability of order selected. The observed L_2 efficiency suggests the large value of α causes the high efficiency of $APIC\alpha$ which indicates that the correct model is also the closet model, except when the true model can be specified more easily and sample sizes are moderate to large, then the small value of α is preferable. For the strongly identifiable true model, the large value of α performs well, whereas if the regression coefficients are not large enough and the sample sizes are small to moderate, the value of α should be moderate.

Keywords: model selection, penalty function, probability of overfitting, signal-to-noise ratio, observed L_2 efficiency.

1. Introduction

In regression analysis, the choice of an appropriate model from a class of candidate models to characterize the study data is a key issue. In real life, we may not know what the true model is, but we hope to find a model that is a reasonably accurate representation. A model selection criterion represents a useful tool to judge the propriety of a fitted model by assessing whether it offers an optimal balance between goodness of fit and parsimony. The first model selection criterion to gain widespread acceptance was the Akaike information criterion, AIC [1-3]. This serves as an asymptotically unbiased estimator of a variant of Kullback's directed divergence between the true model and a fitted approximating model. The directed divergence, also known as the Kullback-Leibler information, the I-divergence, or the relative entropy, assesses the dissimilarity between two statistical models. Other well-known criteria were subsequently introduced and studied such as Bayesian information criterion, BIC [4-6], and Kullback information criterion, KIC [7-14]. BIC is an asymptotic approximation to a transformation of the Bayesian posterior probability of a candidate model [5]. KIC is a symmetric measure, meaning that an alternate directed divergence may be obtained by reversing the roles of the two models in the definition of the measure. The sum of the two directed divergences is Kullback's symmetric divergence, also known as the J-divergence [7-8]. Although AIC remains arguably the most widely used model selection criterion, BIC and KIC, are popular competitors. In fact, BIC is often preferred over AIC by practitioners who find appeal in either its Bayesian justification or its tendency to choose more parsimonious models than AIC [5]. Likewise, KIC is a symmetric measure which combines the information in two related, though distinct measures; it functions as a gauge of model disparity that is arguably more sensitive than AIC that corresponds to only individual components [7-8]. However, AIC, BIC, and KIC, still have the problems of overfitting and weak signal-to-noise ratio due to the weak penalty functions.

With this motivation, the aim of this paper is to study the penalty functions based on these criteria for the case of univariate regression model in order to find the appropriate value of penalty to reduce the probabilities of over- and underfitting and also to overcome the weak signal-to-noise ratio. The remainder of this paper is organized as follows. In Section 2, we unify AIC, BIC, and KIC, in one form, called Adjusted Penalty Information Criterion $(APIC\alpha)$. The studies on the probability of overfitting and signal-to-noise ratio are also considered in this section. In Section 3, we simulate 1,000 realizations of multiple regression models in order to study the probability of the order selected and the observed L_2 efficiency of $APIC\alpha$ where the values of α range from 1 to 14. Finally, Section 4 is the conclusions, discussion, and further study.

2. Model selection criteria, probability of overfitting, and signal-to-noise ratio

Suppose data are generated by the operating model, i.e., true model [15]

$$\mathbf{y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}_0,\ \boldsymbol{\varepsilon}_0 \sim \mathrm{N}_n\left(\mathbf{0},\ \sigma_0^2\mathbf{I}_n\right), \tag{1}$$

and the candidate or approximating model is in the form [15]

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim \mathrm{N}_n\left(\mathbf{0},\ \sigma^2\mathbf{I}_n\right), \tag{2}$$

where $\mathbf{y}$ is an $n\times 1$ dependent random vector of observations, $\mathbf{X}_0$ and $\mathbf{X}$ are $n\times p_0$ and $n\times p$ matrices of independent variables with full-column rank, respectively, $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}$ are $p_0\times 1$ and $p\times 1$ parameter vectors of regression coefficients, respectively, $\boldsymbol{\varepsilon}_0$ and $\boldsymbol{\varepsilon}$ are $n\times 1$ noise vectors. The $(p+1)\times 1$ vector of parameters is $\boldsymbol{\theta}_0 = \left[\boldsymbol{\beta}_0' \quad \sigma_0^2\right]'$ and the maximum likelihood estimator of $\boldsymbol{\theta}_0$ is $\hat{\boldsymbol{\theta}} = \left[\hat{\boldsymbol{\beta}}' \quad \hat{\sigma}^2\right]'$ where

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \text{ and } \hat{\sigma}^2 = \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)/n.$$

For each data set, we can construct many fitted candidate models. Nevertheless, we cannot know which model is the best. Criterion for model selection is a way to solve this problem. AIC, BIC, and KIC, are three well-known criteria to consider in this study. We scale these criteria by $1/n$ in order to express them as a rate per observation. The formulas for them are based on the form of the log of the likelihood function of the maximum likelihood estimator of σ^2 plus a penalty function, called Adjusted Penalty Information Criterion,

$$APIC\alpha = \log\left(\hat{\sigma}^2\right) + \alpha\left(p+1\right)/n. \quad (3)$$

When the values of α in (3) are equal to 2, $\log(n)$, 3; $APIC\alpha$ becomes AIC, BIC, and KIC, respectively. The appropriate value of α has been found to reduce the probabilities of over- and underfitting and also to overcome the weak signal-to-noise ratio. The value of α is selected by four measurements: the probability of over- and underfitting, the signal-to-noise ratio, the probability of order selected, and the observed L_2 efficiency. Theoretical and empirical methods are used to examine the performance of $APIC\alpha$.

The terms over- and underfitting can be defined in two ways. Under consistency, when a true model is itself a candidate model, overfitting is a situation when the model has extra variables with more parameters than the optimal model and underfitting is defined as choosing a model that either has too few variables or is incomplete. In view of efficiency, overfitting can be defined as choosing a model that has more variables than the model identified as closest to the true model, thereby reducing efficiency. Underfitting is defined as choosing a model with too few variables compared to the closest model, also reducing efficiency. Both over- and underfitting can lead to problems with the predictive abilities of a model. An underfitted model may have poor predictive ability due to a lack of detail in the model, while an overfitted model may be unstable in the sense that repeated samples from the same process can lead to widely differing predictions due to variability in the extraneous variables. A criterion that can balance the tendencies of over- and underfitted is preferable [16-17].

The probability of model selection criterion preferring the overfitted model is analyzed here by comparing the true model of order p_0 to a more complex model or overfitted model of order p_0+l, $l>0$. Hence for finite n, the probability that $APIC\alpha$ prefers the overfitted model is defined by

$$P\left\{APIC\alpha_{p_0+l} < APIC\alpha_{p_0}\right\} = P\left\{\log\left(\hat{\sigma}^2_{p_0+l}\right) + \frac{\alpha\left(p_0+l+1\right)}{n} < \log\left(\hat{\sigma}^2_{p_0}\right) + \frac{\alpha\left(p_0+1\right)}{n}\right\}$$

$$= P\left\{\log\left(\frac{\hat{\sigma}^2_{p_0}}{\hat{\sigma}^2_{p_0+l}}\right) > \frac{\alpha l}{n}\right\} = P\left\{\frac{\hat{\sigma}^2_{p_0}}{\hat{\sigma}^2_{p_0+l}} > \exp\left(\frac{\alpha l}{n}\right)\right\} = P\left\{\frac{\hat{\sigma}^2_{p_0} - \hat{\sigma}^2_{p_0+l}}{\hat{\sigma}^2_{p_0+l}} > \exp\left(\frac{\alpha l}{n}\right) - 1\right\}. \quad (4)$$

Under the assumption of nested models; $p \geq p_0$ and $l > 0$, we have $n\left(\hat{\sigma}_p^2 - \hat{\sigma}_{p+l}^2\right) \sim \sigma_0^2 \chi_l^2$, $n\hat{\sigma}_p^2 \sim \sigma_0^2 \chi_{n-p}^2$, where χ_k^2 represents the chi-square distribution with k degrees of freedom, and $\hat{\sigma}_p^2 - \hat{\sigma}_{p+l}^2$ is independent of $\hat{\sigma}_{p+l}^2$ [16]. (5)

Then the probability of overfitting by l extra variables of $APIC\alpha$ in (4) becomes

$$P\left\{APIC\alpha_{p_0+l} < APIC\alpha_{p_0}\right\} = P\left\{F_{l,\, n-p_0-l} > \frac{n-p_0-l}{l}\left[\exp\left(\frac{\alpha l}{n}\right) - 1\right]\right\}. \quad (6)$$

In (6), we found that $APIC\alpha$'s probability of overfitting depends on the value of α in (3). If the value of α tends to infinity under the same values of the sample size (n), the order of true model (p_0), and the additional variable (l), $APIC\alpha$ tends to less overfitting. When we replace the values of α in (6) by 2, $\log(n)$, 3, we get the probabilities of overfitting of AIC, BIC, and KIC, respectively. The proof of the probability of overfitting can be confirmed numerically in Table 1. The examples of the calculation for the probability of overfitting by l extra variables of $APIC\alpha$ in (6) are as follows: for $n = 15$, $p_0 = 3$

$$P\left\{APIC1_{p_0+1} < APIC1_{p_0}\right\} = P\left\{F_{1,\,11} > 0.7583\right\} = 0.4025$$

$$P\left\{APIC1_{p_0+2} < APIC1_{p_0}\right\} = P\left\{F_{2,\,10} > 0.7132\right\} = 0.5134$$

$$P\left\{APIC3_{p_0+1} < APIC3_{p_0}\right\} = P\left\{F_{1,\,11} > 2.4354\right\} = 0.1469$$

$$P\left\{APIC3_{p_0+2} < APIC3_{p_0}\right\} = P\left\{F_{2,\,10} > 2.4591\right\} = 0.1353.$$

The explanation of the result in Table 1 is that, e.g. for $n = 15$, $p_0 = 3$, and $l = 1$, the probability of overfitting of $APIC1$ is 0.4025, this means that this criterion would select the model whose order is higher by one order than true model with a probability of 0.4025. Although the large value of α resulted in $APIC\alpha$ having the low probability of overfitting, sometimes it will be prone to underfitting. This result will be shown in the simulation study.

Table 1. Probability of overfitting by l extra variables of *APIC*α for different values of n, p_0, and l.

n	p_0	*l*	Criteria *APIC*1	*APIC*2	*APIC*3	*APIC*4	*APIC*5	*APIC*6	*APIC*7
15	3	1	0.4025	0.2363	0.1469	0.0939	0.0611	0.0402	0.0266
15	3	2	0.5134	0.2636	0.1353	0.0695	0.0357	0.0183	0.0094
15	3	3	0.5947	0.2857	0.1287	0.0561	0.0240	0.0101	0.0042
15	3	4	0.6664	0.3143	0.1305	0.0508	0.0190	0.0070	0.0025
15	4	1	0.4257	0.2599	0.1676	0.1110	0.0747	0.0509	0.0349
15	4	2	0.5488	0.3012	0.1653	0.0907	0.0498	0.0273	0.0150
15	4	3	0.6384	0.3362	0.1667	0.0802	0.0378	0.0176	0.0082
15	4	4	0.7154	0.3784	0.1780	0.0788	0.0336	0.0140	0.0058
30	3	1	0.3565	0.1922	0.1102	0.0651	0.0392	0.0239	0.0147
30	3	2	0.4346	0.1889	0.0821	0.0357	0.0155	0.0067	0.0029
30	3	3	0.4846	0.1795	0.0617	0.0204	0.0066	0.0021	0.0007
30	3	4	0.5256	0.1720	0.0482	0.0125	0.0031	0.0007	0.0002
30	4	1	0.3661	0.2012	0.1175	0.0706	0.0433	0.0268	0.0168
30	4	2	0.4493	0.2019	0.0907	0.0408	0.0183	0.0082	0.0037
30	4	3	0.5033	0.1954	0.0705	0.0245	0.0083	0.0028	0.0009
30	4	4	0.5475	0.1902	0.0568	0.0157	0.0042	0.0011	0.0003
100	3	1	0.3284	0.1670	0.0905	0.0506	0.0289	0.0167	0.0097
100	3	2	0.3867	0.1496	0.0578	0.0224	0.0087	0.0033	0.0013
100	3	3	0.4178	0.1288	0.0367	0.0100	0.0027	0.0007	0.0002
100	3	4	0.4395	0.1109	0.0236	0.0046	0.0009	0.0002	0.0000
100	4	1	0.3310	0.1692	0.0922	0.0519	0.0297	0.0173	0.0101
100	4	2	0.3906	0.1526	0.0596	0.0233	0.0091	0.0036	0.0014
100	4	3	0.4227	0.1322	0.0382	0.0106	0.0029	0.0008	0.0002
100	4	4	0.4453	0.1144	0.0248	0.0050	0.0009	0.0002	0.0000

n	p_0	*l*	Criteria *APIC*8	*APIC*9	*APIC*10	*APIC*11	*APIC*12	*APIC*13	*APIC*14
15	3	1	0.0178	0.0119	0.0080	0.0054	0.0037	0.0025	0.0017
15	3	2	0.0048	0.0025	0.0013	0.0007	0.0003	0.0002	0.0001
15	3	3	0.0018	0.0007	0.0003	0.0001	0.0001	0.0000	0.0000
15	3	4	0.0009	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000
15	4	1	0.0241	0.0167	0.0117	0.0082	0.0057	0.0040	0.0028
15	4	2	0.0082	0.0045	0.0025	0.0014	0.0007	0.0004	0.0002
15	4	3	0.0037	0.0017	0.0008	0.0004	0.0002	0.0001	0.0000
15	4	4	0.0023	0.0009	0.0004	0.0002	0.0001	0.0000	0.0000
30	3	1	0.0091	0.0057	0.0035	0.0022	0.0014	0.0009	0.0006
30	3	2	0.0013	0.0006	0.0002	0.0001	0.0000	0.0000	0.0000
30	3	3	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
30	3	4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
30	4	1	0.0106	0.0067	0.0043	0.0027	0.0017	0.0011	0.0007
30	4	2	0.0017	0.0007	0.0003	0.0002	0.0001	0.0000	0.0000
30	4	3	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
30	4	4	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	3	1	0.0057	0.0034	0.0020	0.0012	0.0007	0.0004	0.0003
100	3	2	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
100	3	3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	3	4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	4	1	0.0060	0.0035	0.0021	0.0013	0.0008	0.0005	0.0003
100	4	2	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
100	4	3	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	4	4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

The signal-to-noise ratio is the second measure used to study the property of $APIC\alpha$. McQuarrie and Tsai [16] defined the signal-to-noise ratio as a measurement that is basically a ratio of the expectation to the standard deviation of the difference in criterion values for two models. The ratio tends to assess whether the penalty function is sufficiently strong in relation to the goodness of fit term. From the true model order p_0 and a candidate model order p_0+l where $l>0$, the true model is considered better than a candidate model if the criterion value of a model of order p_0 is less than that of order $p_0+l,\ APIC\alpha_{p_0} < APIC\alpha_{p_0+l}$. Then, the signal-to-noise ratio that the true model has selected compared to a candidate model is defined by

$$\frac{signal}{noise} = \frac{E\left[APIC\alpha_{p_0+l} - APIC\alpha_{p_0}\right]}{sd\left[APIC\alpha_{p_0+l} - APIC\alpha_{p_0}\right]}, \tag{7}$$

where $E\left[APIC\alpha_{p_0+l} - APIC\alpha_{p_0}\right]$

$$= E\left[\log\left(\hat{\sigma}^2_{p_0+l}\right) + \frac{\alpha\left(p_0+l+1\right)}{n} - \log\left(\hat{\sigma}^2_{p_0}\right) - \frac{\alpha\left(p_0+1\right)}{n}\right]$$

$$= E\left[\log\left(\frac{\hat{\sigma}^2_{p_0+l}}{\hat{\sigma}^2_{p_0}}\right) + \frac{\alpha l}{n}\right],$$

and $sd\left[APIC\alpha_{p_0+l} - APIC\alpha_{p_0}\right]$

$$= sd\left[\log\left(\hat{\sigma}^2_{p_0+l}\right) + \frac{\alpha\left(p_0+l+1\right)}{n} - \log\left(\hat{\sigma}^2_{p_0}\right) - \frac{\alpha\left(p_0+1\right)}{n}\right]$$

$$= sd\left[\log\left(\frac{\hat{\sigma}^2_{p_0+l}}{\hat{\sigma}^2_{p_0}}\right) + \frac{\alpha l}{n}\right].$$

Applying the second-order of Taylor's series expansions in order to find the signal in (7) is as follows: suppose $X \sim \chi^2_p$, expanding $\log(X)$ about $E(X)=p$, we have

$$\log(X) \doteq \log(p) + \frac{1}{p}(X-p) - \frac{1}{2p^2}(X-p)^2 \text{ and } E\left[\log(X)\right] \doteq \log(p) - \frac{1}{p}. \tag{8}$$

Using the results in (8) and the assumption in (5), the approximate signal in (7) is

$$E\left[APIC\alpha_{\mathrm{U}_{p_0+l}} - APIC\alpha_{\mathrm{U}_{p_0}}\right] = E\left[\log\left(n\hat{\sigma}^2_{p_0+l}\right)\right] - E\left[\log\left(n\hat{\sigma}^2_{p_0}\right)\right] + \frac{\alpha l}{n},$$

where $E\left[\log\left(n\hat{\sigma}^2_{p_0+l}\right)\right] = \log\left(\sigma_0^2\right) + \log\left(n-p_0-l\right) - \dfrac{1}{n-p_0-l},$

and $E\left[\log\left(n\hat{\sigma}^2_{p_0}\right)\right] = \log\left(\sigma_0^2\right) + \log\left(n-p_0\right) - \dfrac{1}{n-p_0}.$

Therefore,

$$E\left[APIC\alpha_{\mathrm{U}_{p_0+l}} - APIC\alpha_{\mathrm{U}_{p_0}}\right] = \log\left(\frac{n-p_0-l}{n-p_0}\right) - \frac{l}{\left(n-p_0-l\right)\left(n-p_0\right)} + \frac{\alpha l}{n}. \quad (9)$$

Using the assumption in (5) to find the noise in (7) by the *Q*-statistic which has the Beta distribution as follows:

$$Q = \frac{n\hat{\sigma}^2_{p_0+l}}{n\hat{\sigma}^2_{p_0}} \sim Beta\left(\frac{n-p_0-l}{2}, \frac{l}{2}\right), \quad (10)$$

and the log-distribution of *Q*-statistic is

$$\log\left(Q\right) = \log\left(\frac{n\hat{\sigma}^2_{p_0+l}}{n\hat{\sigma}^2_{p_0}}\right) \sim \log\text{-}Beta\left(\frac{n-p_0-l}{2}, \frac{l}{2}\right). \quad (11)$$

Applying the first-order of Taylor's series expansions when $X \sim \chi^2_p$, we expand $\log\left(X\right)$ about $E\left(X\right) = p$ as follows:

$$\log\left(X\right) \doteq \log\left(p\right) + \frac{1}{p}\left(X-p\right). \quad (12)$$

Using (12) to expand $\log\left(Q\right)$ in (11) about

$$E\left(Q\right) = \frac{\left(n-p_0-l\right)/2}{\left(n-p_0-l\right)/2 + l/2} = \frac{n-p_0-l}{n-p_0},$$

we have

$$\log\left(Q\right) \doteq \log\left(\frac{n-p_0-l}{n-p_0}\right) + \frac{n-p_0}{n-p_0-l}\left(Q - \frac{n-p_0-l}{n-p_0}\right). \quad (13)$$

The variance of $\log(Q)$ in (11) is approximated by the variance of $\log(Q)$ in (13) as follows:

$$\operatorname{var}\left[\log(Q)\right] = \operatorname{var}\left[\log\left(\frac{n\hat{\sigma}^2_{p_0+l}}{n\hat{\sigma}^2_{p_0}}\right)\right] \doteq \operatorname{var}\left[\log\left(\frac{n-p_0-l}{n-p_0}\right)+\frac{n-p_0}{n-p_0-l}\left(Q-\frac{n-p_0-l}{n-p_0}\right)\right]$$

$$=\left(\frac{n-p_0}{n-p_0-l}\right)^2 \operatorname{var}(Q),$$

where $\operatorname{var}(Q)=\dfrac{\left[(n-p_0-l)/2\right](l/2)}{\left((n-p_0-l)/2+l/2\right)^2\left((n-p_0-l)/2+l/2+1\right)}$,

Therefore,

$$\operatorname{var}\left[\log(Q)\right]=\frac{2l}{(n-p_0-l)(n-p_0+2)}, \tag{14}$$

and the standard deviation of $\log(Q)$ in (14) or the approximate noise in (7) is

$$sd\left[\log(Q)\right]=sd\left[\log\left(\frac{\hat{\sigma}^2_{p_0+l}}{\hat{\sigma}^2_{p_0}}\right)+\frac{\alpha l}{n}\right]\doteq\frac{\sqrt{2l}}{\sqrt{(n-p_0-l)(n-p_0+2)}}. \tag{15}$$

Combined, the approximations of signal in (9) and noise in (15) to be the approximate signal-to-noise ratio in (7) is as follows:

$$\frac{signal}{noise}\doteq\frac{\sqrt{(n-p_0-l)(n-p_0+2)}}{\sqrt{2l}}\left[\log\left(\frac{n-p_0-l}{n-p_0}\right)-\frac{l}{(n-p_0-l)(n-p_0)}+\frac{\alpha l}{n}\right]. \tag{16}$$

In (16), we found that the signal-to-noise ratio of $APIC\alpha$ depends on the value of α in (3). This conclusion is similar to the probability of overfitting, that is if the value of α tends to infinity under the same values of n, p_0, and l, $APIC\alpha$ has a strong signal-to-noise ratio. When we replace the values of α in (16) by 2, $\log(n)$, and 3, we have the approximate signal-to-noise ratios for AIC, BIC, and KIC, respectively. The proof of the signal-to-noise ratio can be confirmed numerically in Table 2.

Table 2. Signal-to-noise ratio of *APIC* α for different values of n, p_0, and l.

n	p_0	l	Criteria						
			***APIC*1**	***APIC*2**	***APIC*3**	***APIC*4**	***APIC*5**	***APIC*6**	***APIC*7**
15	3	1	-0.2450	0.3400	0.9250	1.5100	2.0950	2.6800	3.2650
15	3	2	-0.3884	0.4004	1.1892	1.9780	2.7668	3.5556	4.3444
15	3	3	-0.5291	0.3874	1.3039	2.2204	3.1370	4.0535	4.9700
15	3	4	-0.6752	0.3225	1.3203	2.3181	3.3159	4.3136	5.3114
15	4	1	-0.3042	0.2333	0.7707	1.3082	1.8457	2.3832	2.9207
15	4	2	-0.4734	0.2477	0.9688	1.6899	2.4110	3.1321	3.8532
15	4	3	-0.6351	0.1976	1.0302	1.8629	2.6956	3.5282	4.3609
15	4	4	-0.8002	0.0992	0.9985	1.8979	2.7973	3.6967	4.5961
30	3	1	-0.1132	0.5340	1.1812	1.8284	2.4756	3.1229	3.7701
30	3	2	-0.1785	0.7190	1.6166	2.5141	3.4116	4.3092	5.2067
30	3	3	-0.2414	0.8356	1.9127	2.9897	4.0667	5.1438	6.2208
30	3	4	-0.3054	0.9120	2.1295	3.3470	4.5644	5.7819	6.9994
30	4	1	-0.1389	0.4847	1.1083	1.7319	2.3555	2.9791	3.6027
30	4	2	-0.2149	0.6492	1.5133	2.3774	3.2415	4.1056	4.9697
30	4	3	-0.2861	0.7499	1.7859	2.8219	3.8579	4.8940	5.9300
30	4	4	-0.3573	0.8127	1.9827	3.1527	4.3227	5.4927	6.6627
100	3	1	-0.0324	0.6569	1.3463	2.0356	2.7250	3.4143	4.1037
100	3	2	-0.0510	0.9188	1.8886	2.8584	3.8282	4.7980	5.7678
100	3	3	-0.0687	1.1128	2.2942	3.4757	4.6572	5.8387	7.0202
100	3	4	-0.0867	1.2703	2.6273	3.9843	5.3413	6.6982	8.0552
100	4	1	-0.0396	0.6426	1.3249	2.0072	2.6895	3.3717	4.0540
100	4	2	-0.0612	0.8986	1.8584	2.8182	3.7780	4.7378	5.6976
100	4	3	-0.0813	1.0880	2.2572	3.4264	4.5957	5.7649	6.9341
100	4	4	-0.1011	1.2417	2.5845	3.9274	5.2702	6.6130	7.9559

n	p_0	l	Criteria						
			***APIC*8**	***APIC*9**	***APIC*10**	***APIC*11**	***APIC*12**	***APIC*13**	***APIC*14**
15	3	1	3.8500	4.4350	5.0200	5.6050	6.1900	6.7750	7.3600
15	3	2	5.1333	5.9221	6.7109	7.4997	8.2885	9.0773	9.8661
15	3	3	5.8865	6.8030	7.7195	8.6360	9.5526	10.4691	11.3856
15	3	4	6.3092	7.3070	8.3047	9.3025	10.3003	11.2981	12.2958
15	4	1	3.4582	3.9956	4.5331	5.0706	5.6081	6.1456	6.6831
15	4	2	4.5743	5.2954	6.0166	6.7377	7.4588	8.1799	8.9010
15	4	3	5.1936	6.0262	6.8589	7.6916	8.5242	9.3569	10.1896
15	4	4	5.4955	6.3948	7.2942	8.1936	9.0930	9.9924	10.8917
30	3	1	4.4173	5.0645	5.7117	6.3589	7.0062	7.6534	8.3006
30	3	2	6.1042	7.0017	7.8993	8.7968	9.6943	10.5918	11.4894
30	3	3	7.2978	8.3749	9.4519	10.5289	11.6060	12.6830	13.7600
30	3	4	8.2168	9.4343	10.6518	11.8692	13.0867	14.3041	15.5216
30	4	1	4.2263	4.8500	5.4736	6.0972	6.7208	7.3444	7.9680
30	4	2	5.8338	6.6979	7.5620	8.4261	9.2902	10.1543	11.0184
30	4	3	6.9660	8.0020	9.0380	10.0740	11.1101	12.1461	13.1821
30	4	4	7.8327	9.0027	10.1727	11.3427	12.5127	13.6827	14.8527
100	3	1	4.7930	5.4824	6.1717	6.8611	7.5504	8.2398	8.9291
100	3	2	6.7376	7.7074	8.6772	9.6470	10.6168	11.5866	12.5564
100	3	3	8.2016	9.3831	10.5646	11.7461	12.9276	14.1091	15.2905
100	3	4	9.4122	10.7692	12.1262	13.4831	14.8401	16.1971	17.5541
100	4	1	4.7363	5.4186	6.1008	6.7831	7.4654	8.1477	8.8299
100	4	2	6.6574	7.6171	8.5769	9.5367	10.4965	11.4563	12.4161
100	4	3	8.1034	9.2726	10.4418	11.6111	12.7803	13.9495	15.1187
100	4	4	9.2987	10.6415	11.9844	13.3272	14.6700	16.0129	17.3557

The examples of the calculation for the signal-to-noise ratio of $APIC\alpha$ in (16) are as follows: for $n=15$, $p_0=3$

$$l=1,\ \alpha=1;\ \frac{signal}{noise} \doteq \frac{\sqrt{(11)(14)}}{\sqrt{2}}\left[\log\left(\frac{11}{12}\right)-\frac{1}{(11)(12)}+\frac{1}{15}\right]=-0.2450$$

$$l=2,\ \alpha=1;\ \frac{signal}{noise} \doteq \frac{\sqrt{(10)(14)}}{\sqrt{4}}\left[\log\left(\frac{10}{12}\right)-\frac{2}{(10)(12)}+\frac{2}{15}\right]=-0.3884$$

$$l=1,\ \alpha=3;\ \frac{signal}{noise} \doteq \frac{\sqrt{(11)(14)}}{\sqrt{2}}\left[\log\left(\frac{11}{12}\right)-\frac{1}{(11)(12)}+\frac{3}{15}\right]=0.9250$$

$$l=2,\ \alpha=3;\ \frac{signal}{noise} \doteq \frac{\sqrt{(10)(14)}}{\sqrt{4}}\left[\log\left(\frac{10}{12}\right)-\frac{2}{(10)(12)}+\frac{6}{15}\right]=1.1892\,.$$

McQuarrie and Tsai [16] concluded that the signal-to-noise ratios are strong or weak as follows. A strong signal-to-noise ratio refers to a large positive value (often greater than 2) and leads to small probability of overfitting. A weak signal-to-noise ratio usually refers to one that is small (less than 0.5) or negative and results in high probability of overfitting. The model selection criterion that has strong signal-to-noise ratio and lowest probability of overfitting is preferable.

3. Simulation study

In addition to the proofs of probability of overfitting in (6) and the approximate signal-to-noise ratio in (16), we use the simulation data to find the appropriate value of α for $APIC\alpha$ in (3). Four cases of the true multiple regression models in (1) are constructed as follows.

Model 1 (very weakly identifiable true model with the true order $p_0=7$):

$$y_1 = X_1 + 0.5X_2 + 0.1X_3 + 0.05X_4 + 0.01X_5 + 0.005X_6 + 0.001X_7 + \varepsilon_1,$$

Model 2 (weakly identifiable true model with the true order $p_0=3$):

$$y_2 = X_1 + 0.5X_2 + 0.25X_3 + \varepsilon_2,$$

Model 3 (very strongly identifiable true model with the true order $p_0=4$):

$$y_3 = X_1 + 2X_2 + 2X_3 + 2X_4 + \varepsilon_3,$$

Model 4 (strongly identifiable true model with the true order $p_0=8$):

$$y_4 = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + \varepsilon_4.$$

Model 1 and Model 2 represent the weakly identifiable true models which mean they are not easily identified compared to the strongly identifiable true models such as Model 3 and Model 4. In this study, the true variance σ_0^2 in (1) is assumed equal to one. For each model, we consider 1,000 realizations for three levels of the sample sizes which are $n=15$ (small), $n=30$ (moderate), and $n=100$ (large). Ten candidate variables, X_1 to X_{10}, are stored in an $n\times 10$ matrix $\mathbf{X}$ of the candidate model in (2), where X_1 is given as a constant which equals one, followed by nine independent identically distributed normal random variables with zero mean and equal variance-covariance matrix to identity matrix $\mathbf{I}_{10}$. The candidate models include the columns of $\mathbf{X}$ in a sequentially nested fashion; i.e., columns 1 to p define the design matrix for the candidate model with dimension p. Over 1,000 realizations, we apply $APIC\alpha$ in (3) with the values of α ranging from 1 to 14 on the datasets y of four models constructed. The probability of order selected by $APIC\alpha$ is a measure used to examine the effects of weak or strong penalty function in the proposed criterion. In addition to above measure, many authors [18-19] use the observed L_2 efficiency to assess model selection criterion performance, especially when the true model is very difficult to detect. The observed L_2 distance, scaled by $1/n$, between the true model in (1) and the fitted candidate model in (2) is defined as

$$L_2(p)=\left(\mathbf{X}_0\boldsymbol{\beta}_0-\mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{X}_0\boldsymbol{\beta}_0-\mathbf{X}\hat{\boldsymbol{\beta}}\right)/n.$$

Observed L_2 efficiency is defined by the ratio

$$\text{Observed } L_2 \text{ efficiency} = \frac{\min_{1\le p\le P} L_2(p)}{L_2(p_s)},$$

where P is the class of all possible candidate models, p is the rank of fitted candidate model, and p_s is the model selected by specific model selection criterion. The closer the selected model is to the true model, the higher the efficiency. Therefore, the best model selection criterion will select a model which yields high efficiency even in small samples or the true model is weakly identifiable. In order to summarize the results in this study, the average observed L_2 efficiencies over the 1,000 realizations are ranked for $APIC\alpha$ where the values of α range from 1 to 14. The first rank of average observed L_2 efficiencies goes to the highest efficiency criterion and denotes better relative performance. Results of comparing the probability of order selected by $APIC\alpha$ and average observed L_2 efficiencies are shown in Table 3.

From the results of simulation in Table 3 we found that, for Model 1 and Model 2 which are the situations where the true model cannot be easily identified, $APIC\alpha$ with the small value of α (about 1 to 3) gives the greater probability of correct order being selected than the case of large value and also prevents the probability of underfitting. While, the observed L_2 efficiency suggests the large value of α causes the high efficiency of $APIC\alpha$, except when the true model can be specified more easily, such as Model 2, and sample sizes are moderate to large, the small value of α (about 3 to 4) is preferable. For Model 3 and Model 4 which are the situations where the true model is strongly identifiable, the value of α should be large (at least 8), except when the regression coefficients are not large enough, such as Model 4, and the sample sizes are small to moderate, the value of α should be moderate (about 4 to 6).

For all models, if the value of α tends to infinity, the probability of overfitted tends to decrease whereas the probability of underfitting tends to increase. The point that has the optimal probability of over- and underfitting always presents the maximum probability of correct order being selected.

4. Conclusions, discussion, and further study

In this paper, we study the penalty functions based on the well-known model selection criteria, AIC, BIC, and KIC which can be unified in the form of the log likelihood function of the maximum likelihood estimator of σ^2 plus a penalty function, called Adjusted Penalty Information Criterion, i.e.,

$$APIC\alpha = \log\left(\hat{\sigma}^2\right) + \alpha\left(p+1\right)/n,$$

when the values of α are equal to 2, $\log\left(n\right)$, 3; $APIC\alpha$ becomes AIC, BIC, and KIC respectively. Each criterion has a different value due to its penalty function, the differences in strong or weak penalty affecting the probabilities of over- and underfitting, including the problem of signal-to-noise ratio being weak.

Table 3. Probability of the order selected by *APIC* α and average observed L_2 efficiencies over 1,000 realizations.

Model	*n*	Order	Criteria						
			APIC1	***APIC2***	***APIC3***	***APIC4***	***APIC5***	***APIC6***	***APIC7***
1 very weakly identifiable (true order $p_0 = 7$)	15	Underfitted	0.191	0.560	0.809	0.931	0.980	0.997	0.997
		Correct	**0.055**	0.044	0.018	0.006	0.001	0.000	0.000
		Overfitted	0.754	0.396	0.173	0.063	0.019	0.003	0.003
		Ave. L_2 eff.	0.266	0.483	0.687	0.811	0.890	0.922	0.937
		Rank	14	13	12	11	10	9	8
	30	Underfitted	0.441	0.853	0.982	0.998	0.999	1.000	1.000
		Correct	**0.067**	0.029	0.006	0.001	0.001	0.000	0.000
		Overfitted	0.492	0.118	0.012	0.001	0.000	0.000	0.000
		Ave. L_2 eff.	0.386	0.646	0.795	0.858	0.885	0.913	0.923
		Rank	14	13	12	11	10	9	8
	100	Underfitted	0.588	0.927	0.996	0.999	1.000	1.000	1.000
		Correct	**0.079**	0.022	0.001	0.000	0.000	0.000	0.000
		Overfitted	0.333	0.051	0.003	0.001	0.000	0.000	0.000
		Ave. L_2 eff.	0.470	0.642	0.703	0.723	0.735	0.748	0.756
		Rank	14	13	12	11	10	9	8
2 weakly identifiable (true order $p_0 = 3$)	15	Underfitted	0.058	0.288	0.545	0.721	0.826	0.890	0.930
		Correct	0.038	0.136	**0.167**	0.158	0.123	0.090	0.061
		Overfitted	0.904	0.576	0.288	0.121	0.051	0.020	0.009
		Ave. L_2 eff.	0.301	0.469	0.615	0.703	0.746	0.771	0.786
		Rank	14	13	12	11	10	9	8
	30	Underfitted	0.102	0.376	0.584	0.712	0.799	0.857	0.900
		Correct	0.124	**0.282**	0.271	0.234	0.183	0.135	0.096
		Overfitted	0.774	0.342	0.145	0.054	0.018	0.008	0.004
		Ave. L_2 eff.	0.402	0.602	0.663	**0.670**	0.659	0.648	0.642
		Rank	14	13	2	**1**	3	8	12
	100	Underfitted	0.029	0.118	0.223	0.333	0.417	0.499	0.582
		Correct	0.271	0.575	**0.663**	0.628	0.565	0.496	0.415
		Overfitted	0.700	0.307	0.114	0.039	0.018	0.005	0.003
		Ave. L_2 eff.	0.515	0.748	**0.806**	0.782	0.732	0.679	0.616
		Rank	9	3	**1**	2	4	5	6
Model	***n***	**Order**	**Criteria**						
			APIC8	***APIC9***	***APIC10***	***APIC11***	***APIC12***	***APIC13***	***APIC14***
1 very weakly identifiable (true order $p_0 = 7$)	15	Underfitted	0.999	1.000	1.000	1.000	1.000	1.000	1.000
		Correct	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Overfitted	0.001	0.000	0.000	0.000	0.000	0.000	0.000
		Ave. L_2 eff.	0.952	0.960	0.961	0.962	0.964	0.965	**0.966**
		Rank	7	6	5	4	3	2	**1**
	30	Underfitted	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Correct	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Overfitted	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Ave. L_2 eff.	0.929	0.934	0.935	0.939	0.941	**0.942**	**0.942**
		Rank	7	6	5	4	3	**1.5**	**1.5**
	100	Underfitted	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Correct	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Overfitted	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Ave. L_2 eff.	0.765	0.772	0.778	0.782	0.784	0.784	**0.786**
		Rank	7	6	5	4	3	2	**1**
2 weakly identifiable (true order $p_0 = 3$)	15	Underfitted	0.955	0.970	0.978	0.986	0.990	0.990	0.992
		Correct	0.042	0.030	0.022	0.014	0.010	0.010	0.008
		Overfitted	0.003	0.000	0.000	0.000	0.000	0.000	0.000
		Ave. L_2 eff.	0.797	0.802	0.805	0.808	0.810	0.810	**0.811**
		Rank	7	6	5	4	2.5	2.5	**1**
	30	Underfitted	0.927	0.941	0.959	0.972	0.978	0.982	0.990
		Correct	0.069	0.057	0.039	0.028	0.022	0.018	0.010
		Overfitted	0.004	0.002	0.002	0.000	0.000	0.000	0.000
		Ave. L_2 eff.	0.643	0.643	0.646	0.650	0.650	0.652	0.656
		Rank	11	10	9	7	6	5	4
	100	Underfitted	0.652	0.704	0.768	0.814	0.847	0.876	0.892
		Correct	0.346	0.295	0.231	0.186	0.153	0.124	0.108
		Overfitted	0.002	0.001	0.001	0.000	0.000	0.000	0.000
		Ave. L_2 eff.	0.562	0.524	0.479	0.449	0.427	0.407	0.397
		Rank	7	8	10	11	12	13	14

Note: Boldface type indicates the maximum value.

Table 3. (Continued).

Model	*n*	Order	Criteria *APIC1*	*APIC2*	*APIC3*	*APIC4*	*APIC5*	*APIC6*	*APIC7*
3 very strongly identifiable (true order $p_0 = 4$)	15	Underfitted	0.000	0.000	0.000	0.002	0.005	0.008	0.010
		Correct	0.091	0.312	0.558	0.728	0.851	0.909	0.944
		Overfitted	0.909	0.688	0.442	0.270	0.144	0.083	0.046
		Ave. L_2 eff.	0.435	0.568	0.719	0.828	0.906	0.942	0.964
		Rank	14	13	12	10	8	5	2
	30	Underfitted	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Correct	0.223	0.602	0.789	0.890	0.937	0.961	0.978
		Overfitted	0.777	0.398	0.211	0.110	0.063	0.039	0.022
		Ave. L_2 eff.	0.525	0.753	0.868	0.928	0.958	0.973	0.984
		Rank	14	13	12	11	10	9	8
	100	Underfitted	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Correct	0.307	0.684	0.855	0.932	0.961	0.982	0.989
		Overfitted	0.693	0.316	0.145	0.068	0.039	0.018	0.011
		Ave. L_2 eff.	0.577	0.805	0.910	0.955	0.974	0.988	0.992
		Rank	14	13	12	11	10	9	8
4 strongly identifiable (true order $p_0 = 8$)	15	Underfitted	0.011	0.036	0.094	0.171	0.300	0.503	0.680
		Correct	0.253	0.444	0.532	**0.555**	0.517	0.384	0.251
		Overfitted	0.736	0.520	0.374	0.274	0.183	0.113	0.069
		Ave. L_2 eff.	0.788	0.815	**0.830**	0.812	0.746	0.602	0.449
		Rank	4	2	**1**	3	5	6	7
	30	Underfitted	0.001	0.001	0.003	0.006	0.011	0.019	0.047
		Correct	0.489	0.759	0.875	0.932	0.964	**0.967**	0.944
		Overfitted	0.510	0.240	0.122	0.062	0.025	0.014	0.009
		Ave. L_2 eff.	0.848	0.912	0.950	0.969	**0.982**	0.981	0.962
		Rank	8	7	5	3	**1**	2	4
	100	Underfitted	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Correct	0.593	0.815	0.925	0.966	0.985	0.995	0.999
		Overfitted	0.407	0.185	0.075	0.034	0.015	0.005	0.001
		Ave. L_2 eff.	0.857	0.919	0.960	0.980	0.991	0.997	0.999
		Rank	14	13	12	11	10	9	7.5

Model	*n*	Order	Criteria *APIC8*	*APIC9*	*APIC10*	*APIC11*	*APIC12*	*APIC13*	*APIC14*
3 very strongly identifiable (true order $p_0 = 4$)	15	Underfitted	0.020	0.030	0.042	0.062	0.099	0.144	0.192
		Correct	**0.948**	0.946	0.942	0.929	0.895	0.851	0.805
		Overfitted	0.032	0.024	0.016	0.009	0.006	0.005	0.003
		Ave. L_2 eff.	**0.964**	0.961	0.955	0.941	0.909	0.867	0.823
		Rank	**1**	3	4	6	7	9	11
	30	Underfitted	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Correct	0.984	0.989	0.991	0.996	**1.000**	**1.000**	**1.000**
		Overfitted	0.016	0.011	0.009	0.004	0.000	0.000	0.000
		Ave. L_2 eff.	0.988	0.993	0.994	0.997	**1.000**	**1.000**	**1.000**
		Rank	7	6	5	4	**2**	**2**	**2**
	100	Underfitted	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Correct	0.993	0.997	0.997	0.999	0.999	**1.000**	**1.000**
		Overfitted	0.007	0.003	0.003	0.001	0.001	0.000	0.000
		Ave. L_2 eff.	0.995	0.998	0.998	0.999	0.999	**1.000**	**1.000**
		Rank	7	5.5	5.5	3.5	3.5	**1.5**	**1.5**
4 strongly identifiable (true order $p_0 = 8$)	15	Underfitted	0.834	0.922	0.968	0.995	0.997	0.998	0.999
		Correct	0.140	0.069	0.028	0.003	0.002	0.001	0.000
		Overfitted	0.026	0.009	0.004	0.002	0.001	0.001	0.001
		Ave. L_2 eff.	0.311	0.224	0.171	0.134	0.129	0.124	0.121
		Rank	8	9	10	11	12	13	14
	30	Underfitted	0.104	0.209	0.350	0.560	0.736	0.871	0.947
		Correct	0.895	0.790	0.649	0.440	0.264	0.129	0.053
		Overfitted	0.001	0.001	0.001	0.000	0.000	0.000	0.000
		Ave. L_2 eff.	0.917	0.820	0.688	0.485	0.317	0.185	0.109
		Rank	6	9	10	11	12	13	14
	100	Underfitted	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		Correct	0.999	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
		Overfitted	0.001	0.000	0.000	0.000	0.000	0.000	0.000
		Ave. L_2 eff.	0.999	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
		Rank	7.5	**3.5**	**3.5**	**3.5**	**3.5**	**3.5**	**3.5**

Note: Boldface type indicates the maximum value.

The theoretical results show that, when the value of α tends to infinity, the probability of overfitting tends to zero and the signal-to-noise ratio tends to strong. At the same time, the results of simulation based on values of α for $APIC\alpha$ ranging from 1 to 14 suggest that, when the true model is weakly identifiable, the value of α should be small to give a high probability of correct order being selected and to prevent the probability of underfitting. However in the case of the true model is very difficult to detect, such as Model 1; none of the criteria correctly identify the true model more than 8% of the time. As a result, the probability of correct order being selected may not be meaningful. For this reason, we used the observed L_2 efficiency to assess the appropriate value of α . This measure suggests the large value of α causes the high efficiency of $APIC\alpha$ which indicates that the correct model is also the closet model, except when the true model can be specified more easily, such as Model 2, and sample sizes are moderate to large, then the small value of α is preferable. For the strongly identifiable true model, the large value of α performs well. Because the problem of underfitting does not occur in this situation, the underfitted order often gives the maximum value of the estimated mean squared error and hence, under the model selection criterion, it is not possible to select the underfitted model. In the situation where the regression coefficients are not large enough, such as Model 4, and the sample sizes are small to moderate, the value of α should be moderate.

In further work, we attempt to construct the model selection criteria to overcome the probability of over- and underfitting in the multivariate regression and simultaneous equations models.

Acknowledgements

This project is financial supported by The Thailand Research Fund, Office of The Higher Education Commission, and Thaksin University under grant No. MRG5480044. I would like to give special thanks to Prof. Dr. Joseph E. Cavanaugh, Department of Biostatistics, College of Public Health, University of Iowa, USA, Assoc. Prof. Dr. Jirawan Jitthavech, and Assoc. Prof. Dr. Vichit Lorchirachoonkul, National Institute of Development Administration (NIDA), THAILAND for all comments and suggestions. The author wishes to express his gratitude to an anonymous referee for carefully reading and constructively critiquing this manuscript.

References

[1] Akaike, H., Information theory and an extension of the maximum likelihood principle, In *2nd International Symposium on Information Theory*, B.N. Petrov and F. Csaki, eds. Akademiai Kiado, Budapest, 1973: 267-281.

[2] Akaike, H., A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 1974; 19: 716-723.

[3] Bedrick, E.J., and Tsai, C.L., Model selection for multivariate regression in small samples, *Biometrics*, 1994; 50: 226-231.

[4] Schwarz, G., Estimating the dimension of a model, *The Annals of Statistics*, 1978; 6: 461-464.

[5] Neath, A., and Cavanaugh, J.E., Regression and time series model selection using variants of the schwarz information criterion, *Communications in Statistics-Theory and Methods*, 1997; 26: 559-580.

[6] McQuarrie, A.D., A Small-Sample correction for the Schwarz SIC model selection criterion, *Statistics & Probability Letters*, 1999; 44: 79-86.

[7] Cavanaugh, J.E., A large-sample model selection criterion based on Kullback's symmetric divergence, *Statistics & Probability Letters*, 1999; 42: 333-343.

[8] Cavanaugh, J.E., Criteria for linear model selection based on Kullback's symmetric divergence, *Australian & New Zealand Journal of Statistics,* 2004; 46: 257-274.

[9] Seghouane, A.K., and Bekara, M., A small sample model selection criterion based on Kullback's symmetric divergence, *IEEE Transactions on Signal Processing*, 2004; 52: 3314-3323.

[10] Kim, H.J., and Cavanaugh, J.E., Model selection criteria based on kullback information measures for nonlinear regression, *Journal of Statistical Planning and Inference*, 2005; 134: 332-349.

[11] Seghouane, A.K., Bekara, M., and Fleury, G., A criterion for model selection in the presence of incomplete data based on Kullback's symmetric divergence, *Signal Processing*, 2005; 85: 1405-1417.

[12] Hafidi, B., A Small-Sample criterion based on Kullback's symmetric divergence for vector autoregressive modeling, *Statistics & Probability Letters*, 2006; 76: 1647-1654.

[13] Hafidi, B., and Mkhadri, A., A corrected akaike criterion based on Kullback's symmetric divergence: Applications in Time Series, Multiple and Multivariate Regression, *Computational Statistics & Data Analysis*, 2006; 50: 1524-1550.

[14] Seghouane, A.K., Multivariate regression model selection from small samples using Kullback's symmetric divergence, *Signal Processing*, 2006; 86: 2074-2084.

[15] Montgomery, D.C., Peck, E.A., and Vining, G.G., *Introduction to Linear Regression Analysis,* 4th ed., New York, Wiley, 2006.

[16] McQuarrie, A.D., and Tsai, C.L., *Regression and Time Series Model Selection*, Singapore, World Scientific, 1998.

[17] Seghouane, A.K., A note on overfitting properties of KIC and KICc. *Signal Process* 2006; 86: 3055-3060.

[18] McQuarrie, A.D., Shumway, R., and Tsai, C.L., The model selection criterion AICu. *Statistics & Probability Letters*, 1997; 34: 285-292.

[19] Seghouane, A.K., Multivariate regression model selection from small samples using Kullback's symmetric divergence, *Signal Processing*, 2006; 86: 2074-2084.