เกณฑ์ใหม่เพื่อแก้ไขปัญหาอัตราส่วน Signal-to-Noise ที่น้อยเกินไปและความน่าจะเป็นที่จะ Over/Underfitting สำหรับตัวแบบถดถอย

New Criterion to Correct the Problems of Weak Signal-to-Noise Ratio and the Probability of Over/Underfitting for Regression Model

วรางคณา เรียนสุทธิ์[1*]

Warangkhana Riansut[1*]

Received: 26 January 2017 ; Accepted: 19 April 2017

บทคัดย่อ

บทความนี้เสนอเกณฑ์การคัดเลือกตัวแบบเพื่อแก้ไขปัญหาอัตราส่วน signal-to-noise ที่น้อยเกินไป และความน่าจะเป็นที่จะ over/underfitting สำหรับตัวแบบถดถอย โดยการปรับ penalty term ของเกณฑ์การคัดเลือกตัวแบบที่เป็นที่รู้จัก *(AIC*, *BIC*, *KIC*) เรียกชื่อเกณฑ์ใหม่นี้ว่า adjusted penalty information criterion (*APIC*) เกณฑ์ที่จัดว่าเป็นเกณฑ์ที่ดี เมื่อมีอัตราส่วน signal-to-noise ที่มาก มีความน่าจะเป็นที่จะ over/underfitting ต่ำ และมีความน่าจะเป็นที่จะคัดเลือกตัวแบบได้ถูกต้องสูง การพิสูจน์ทางทฤษฎี พบว่า ถ้าค่าของ α เข้าใกล้อนันต์ ความน่าจะเป็นที่จะ overfitting จะเข้าใกล้ 0 และอัตราส่วน signal-to-noise จะมีแนวโน้มมาก แต่ความน่าจะเป็นที่จะ underfitting จะเข้าใกล้ 1 ผลการจำลองข้อมูล พบว่า เมื่อตัวแบบสามารถระบุได้ยาก การแจกแจงของตัวแปรอิสระ คือ การแจกแจงปกติหรือการแจกแจงเอกรูป ค่าที่เหมาะสมของ σ^2 ควรมีค่าน้อย แต่สำหรับการแจกแจงของตัวแปรอิสระ คือ การแจกแจงปกติ ขนาดตัวอย่างเพิ่มขึ้น และความแปรปรวนของความคลาดเคลื่อนมีค่าน้อยถึงปานกลาง α ควรมีค่าปานกลาง ถ้าตัวแบบสามารถระบุได้ง่าย การแจกแจงของตัวแปรอิสระ คือ การแจกแจงปกติ และความแปรปรวนของความคลาดเคลื่อนมีค่าน้อยถึงปานกลาง α ควรมีค่ามาก เมื่อความแปรปรวนของความคลาดเคลื่อนเพิ่มขึ้น α ควรมีค่าปานกลาง ถ้าการแจกแจงของตัวแปรอิสระเปลี่ยนเป็นการแจกแจงเอกรูป และความแปรปรวนของความคลาดเคลื่อนมีค่าน้อยถึงปานกลาง α ควรมีค่าปานกลาง นอกเหนือจากนี้ α ควรมีค่าน้อย ถ้าความแปรปรวนของความคลาดเคลื่อนเพิ่มขึ้นจะส่งผลต่อความถูกต้องของ *APIC* ลดลง แต่เมื่อขนาดตัวอย่างเพิ่มขึ้น ความถูกต้องของ *APIC* จะเพิ่มขึ้น

คำสำคัญ: Kullback's Directed Divergence Kullback's Symmetric Divergence การคัดเลือกตัวแบบ ตัวแบบถดถอย

Abstract

This article proposed a model selection criterion in order to correct the weak signal-to-noise ratio and to reduce the probability of over/underfitting for regression model by adjusting the penalty term of the well-known model selection criteria *(AIC*, *BIC*, *KIC*), called adjusted penalty information criterion (*APIC*). Criterion is classified to be the best when it has the strong signal-to-noise ratio, lowest probability of over/underfitting and maximum probability of correct order being selected. The theoretical results show that, if the value of α tends to infinity, the probability of overfitting tends to zero and the signal-to-noise ratio tends to strong, but the probability of underfitting tends to one. The simulation results show that, when the true model is difficult to identify, distributions of independent variables are normal or uniform, the appropriate α is small. But for the independent variables are normal distributed, sample size increases and variances of error terms are small to moderate, α should be moderate. If the true model is easily to identify,

[1] *ผู้ช่วยศาสตราจารย์, สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยทักษิณ พัทลุง 93210*

[1] *Assistant Professor, Department of Mathematics and Statistics, Thaksin University, Phatthalung, 93210.*

** Corresponding author: Tel 074-609-600ext2573, E-mail: warang27@gmail.com*

distribution of independent variables is normal and variances of error terms are small to moderate, the appropriate α is large. When the variance of error terms increases, α should be moderate. If the distribution of independent variables changes to be uniform and variances of error terms are small to moderate, α should be moderate, otherwise α should be small. If the variance of error terms increases, the validity of *APIC* decreases, but when the sample size increases, the validity of *APIC* also increases.

Keywords: Kullback's directed divergence, Kullback's symmetric divergence, model selection, regression model

Introduction

In the application of statistics, the statistical modeling is considered as a major task of study. Three statistical processes to guide a model, which has the parsimony, goodness-of-fit and generalizability properties, are the hypothesis testing of parameters, variable selection algorithms and model selection criterion. The model selection criterion is a popular tool for selecting the best model. The first model selection criterion to gain widespread acceptance was 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. Other well-known criteria were subsequently introduced and studied such as, Bayesian information criterion, *BIC*[4] and Kullback information criterion, *KIC*[5-6]. *BIC* is an asymptotic approximation to a transformation of Bayesian posterior probability of a candidate model[7]. *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[5,8]. 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*[7]. Likewise, *KIC* is a symmetric measure which combines the information in two related, though distinct measures; its functions as a gauge of model disparity that is arguably more sensitive than *AIC* that corresponds to only individual component[5,8]. However, *AIC*, *BIC* and *KIC* still have the problems of weak signal-to-noise ratios and high probabilities of overfitting when the sample size is not large enough which both problems have an effect on the frequency of selection the correct model. With this motivation, the aim of this paper is to propose a model selection criterion to correct the weak signal-to-noise ratio and to reduce the probability of over/underfitting by adjusting the penalty term of the model selection criterion, called adjusted penalty information criterion, denoted by *APIC*. The proposed criterion performance is examined by the extensive simulation study relative to the well-known criteria, *AIC*, *BIC* and *KIC*, under the difference circumstances: sample sizes, orders of true model, regression coefficients, variances of error terms and distributions of independent variables[9-12]. The criterion is classified to be the best when it has the strong signal-to-noise ratio, has the lowest probability of over/underfitting and has the maximum probability of correct order being selected. The remainder of this paper is organized as follows. Adjusted Penalty Information Criterion (*APIC*) in order to correct the weak signal-to-noise ratio and to reduce the probability of over/underfitting is proposed in Section 2. In Section 3, we simulate 1,000 realizations of multiple regression models in order to examine the performance of *APIC* relative to *AIC*, *BIC* and *KIC*. Finally, Section 4 is the conclusions, discussion and further study.

Materials and Methods

The true regression model to consider in this paper is in the form[13]

$$\mathbf{y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}_0 , \quad (1)$$

and the candidate or approximating regression model is in the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} , \quad (2)$$

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 $n\times 1$ error vectors with zero means and variance $\sigma_0^2\mathbf{I}_n$ and $\sigma^2\mathbf{I}_n$, respectively. The maximum likelihood estimators of $\boldsymbol{\beta}$ and σ^2 are, respectively, $\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and $\hat{\sigma}^2=\frac{1}{n}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})$. 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 paper. Many authors usually scale these criteria by 1/n in order to express them as a rate per observation. The formulae for them are based on the following form,

$$APIC=\log(\hat{\sigma}^2)+\frac{\alpha(p+1)}{n}. \tag{3}$$

When the values of α in (3) are equal to 2, $\log(n)$ and 3, *APIC* becomes *AIC*[1-2], *BIC*[4] and *KIC*[5], respectively. In this paper, the methods used to compare which criterion is the best are the ratio of signal-to-noise, the probability of over/underfitting and the probability of correct order being selected. McQuarrie and Tsai[14] 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 term 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 $APIC_{p_0}<APIC_{p_0+}$. Then the signal-to-noise ratio that the true model is selected compared to a candidate model is

$$\begin{aligned}\frac{signal}{noise}&=\frac{E[APIC_{p_0+l}-APIC_{p_0}]}{sd[APIC_{p_0+l}-APIC_{p_0}]}\\&=\frac{E\left[\log\left(\hat{\sigma}^2_{p_0+l}/\hat{\sigma}^2_{p_0}\right)+\frac{\alpha l}{n}\right]}{sd\left[\log\left(\hat{\sigma}^2_{p_0+l}/\hat{\sigma}^2_{p_0}\right)+\frac{\alpha l}{n}\right]}\end{aligned} \tag{4}$$

In order to find the signal in (4), we apply the second-order of Taylor's series expansions as follows. Suppose $X\sim\chi^2$, expanding $\log(X)$ about $E(X)=p$, we have

$$\log(X)\doteq \quad p \; +(X-p)/p-(X-p)^2/2p^2$$

and $$E[\log(X)]\doteq \quad p \;\; -1/p. \tag{5}$$

Under the assumption of nested models; $p\geq p_0$ and $l>0$, we have

$$n(\hat{\sigma}^2_p-\hat{\sigma}^2_{p+l})\sim\sigma\ \chi_l\ ,\ n\hat{\sigma}^2_p\sim\sigma\ \chi_{n\ \ p}$$ and $\hat{\sigma}^2_p-\hat{\sigma}^2_{p+l}$ is independent of $\hat{\sigma}^2_{p+l}$, (6)

where χ^2_k represents the chi-square distribution with k degrees of freedom.

Using the result of Taylor's series expansions in (5) and the assumptions in (6), we have

$$E[\log(n\hat{\sigma}^2_p)]\doteq \quad \sigma \quad +\log(n-p)-1/(n-p). \tag{7}$$

From (7), the signal in (4) is approximated by

$$\begin{aligned}&E[APIC_{p_0+l}-APIC_{p_0}]\\&\doteq \left(n \; \frac{p_0-l}{n-p_0}\right)-\frac{l}{(n-p_0-l)(n-p_0)}+\frac{\alpha l}{n}.\end{aligned} \tag{8}$$

In order to find the noise in (4), we use the assumptions in (6), then we have

$$Q=\frac{n\hat{\sigma}^2_{p_0+l}}{n\hat{\sigma}^2_{p_0}}\sim\frac{\chi^2_{n\ \ p\ \ -l}}{\chi_{n-p_0-l}\ +\chi^2_l}, \tag{9}$$

the *Q*-statistic in (9) has the Beta distribution $Q\sim Beta\quad n-p_0-l)/2,\, l/2)$,

and the log-distribution is

$$\begin{aligned}&\log(Q)=\log\left(n\hat{\sigma}^2_{p_0+l}/n\hat{\sigma}^2_{p_0}\right)\\&\sim\quad Beta((n-p_0-l)/2,\, l/2).\end{aligned} \tag{10}$$

Applying the first-order of Taylor's series expansions to log(*Q*) in (10) about

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

we have

$$\log(Q) \doteq \log[E(Q)] + [Q - E(Q)]/E(Q)$$
$$= \log\left(\frac{n-p_0-l}{n-p_0}\right) + \frac{n-p_0}{n-p_0-l}\left(Q - \frac{n-p_0-l}{n-p_0}\right).$$

Hence

$$\operatorname{var}[\log(Q)] \doteq \left(\frac{n-p_0}{n-p_0-l}\right)^2 \times \left[\frac{(n-p_0-l)/2 \cdot l/2}{\left((n-p_0-l)/2 + l/2\right)^2 \left((n-p_0-l)/2 + l/2 + 1\right)}\right]$$
$$= \frac{2l}{(n-p_0-l)(n-p_0+2)}. \tag{11}$$

Combined the results in (8) and (11) to be the approximate signal-to-noise ratio in (4) as follows:

$$\frac{signal}{noise} \doteq \frac{\sqrt{(n-p_0-l)(n-p_0+2)}}{\sqrt{2l}} \times \left[\log\left(\frac{n-p_0-l}{n-p_0}\right) - \frac{l}{(n-p_0-l)(n-p_0)} + \frac{\alpha l}{n}\right]. \tag{12}$$

In (12), the signal-to-noise ratio of *APIC* depends on the value of α as mention earlier. When we replace the values of α by 2, $\log(n)$ and 3, we have the signal-to-noise ratios of *AIC*, *BIC* and *KIC*, respectively. 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* has a strong signal-to-noise ratio. The proof of the signal-to-noise ratio can be confirmed numerically in Table 1. The example of the calculation for the signal-to-noise ratio of *APIC*, for $n = 15$, $p_0 = 3$, $l = 1$ and $\alpha = 1$, is as follows:

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

From Table 1 we found that when the sample size is small (*n* = 15), *KIC* has a strong signal-to-noise ratio than *BIC* and *AIC*, respectively, because the value of α in (3) from *KIC* is larger than *BIC* and *AIC*, respectively (3 > log(15) > 2). Whereas the sample size are moderate to large (*n* = 30, 100), *BIC* has a strong signal-to-noise ratio than *KIC* and *AIC*, respectively, because the value of α in (3) from *BIC* is larger than *KIC* and *AIC*, respectively (log(30) or log(100) > 3 > 2). Therefore, we can conclude that, *APIC* with a much more value of α, make its signal-to-noise to be strong.

The probability of over/underfitting is the second method to compare which criterion is the best. Both overfitting 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. The probability of overfitting is defined based on a model that has extra variables with more parameters than the optimal model[15]. The probability of *APIC* preferring the overfitted model by l extra variables 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* prefers the overfitted model is defined by

$$P\{APIC_{p_0+l} < APIC_{p_0}\}$$
$$= P\left\{\log\left(\hat{\sigma}^2_{p_0+l}\right) + \frac{\alpha(p_0+l+1)}{n} < \log\left(\hat{\sigma}^2_{p_0}\right) + \frac{\alpha(p_0+1)}{n}\right\}$$
$$= P\left\{\log\left(\frac{\hat{\sigma}^2_{p_0}}{\hat{\sigma}^2_{p_0+l}}\right) > \frac{\alpha l}{n}\right\} = P\left\{\frac{\hat{\sigma}^2_{p_0}}{\hat{\sigma}^2_{p_0+l}} > \exp\left(\frac{\alpha l}{n}\right)\right\} \tag{13}$$
$$= P\left\{\frac{\hat{\sigma}^2_{p_0} - \hat{\sigma}^2_{p_0+l}}{\hat{\sigma}^2_{p_0+l}} > \exp\left(\frac{\alpha l}{n}\right) - 1\right\}.$$

Using the assumptions in (6), the probability of overfitting by l extra variables for *APIC* in (13) becomes

$$P\{APIC_{p_0+l} < APIC_{p_0}\}$$
$$= P\left\{\frac{\chi^2_l}{\chi^2_{n-p_0-l}} > \exp\left(\frac{\alpha l}{n}\right) - 1\right\} \tag{14}$$
$$= P\left\{F_{l,\, n-p_0-l} > \frac{n-p_0-l}{l}\left[\exp\left(\frac{\alpha l}{n}\right) - 1\right]\right\}.$$

In the opposite, the probability of underfitting is defined based on a model with too few variables compared to the optimal model[15]. It is defined by

Table 1 Signal-to-noise ratio of *APIC* α for different values of n, p_0 and l.

n	p_0	l	Criteria							
			APIC1	***APIC2 (AIC)***	***APIClog(n) (BIC)***	***APIC3 (KIC)***	***APIC4***	***APIC5***	***APIC6***	***APIC7***
15	3	1	-0.2450	0.3400	0.7542	0.9250	1.5100	2.0950	2.6800	3.2650
15	3	2	-0.3884	0.4004	0.9589	1.1892	1.9780	2.7668	3.5556	4.3444
15	3	3	-0.5291	0.3874	1.0364	1.3039	2.2204	3.1370	4.0535	4.9700
15	3	4	-0.6752	0.3225	1.0290	1.3203	2.3181	3.3159	4.3136	5.3114
15	5	1	-0.3660	0.1239	0.4708	0.6138	1.1037	1.5936	2.0835	2.5734
15	5	2	-0.5625	0.0907	0.5532	0.7439	1.3971	2.0503	2.7035	3.3567
n	***p_0***	***l***	***APIC8***	***APIC9***	***APIC10***	***APIC11***	***APIC12***	***APIC13***	***APIC14***	
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	5	1	3.0633	3.5532	4.0431	4.5330	5.0229	5.5128	6.0027	
15	5	2	4.0099	4.6631	5.3163	5.9695	6.6227	7.2759	7.9291	
n	***p_0***	***l***	***APIC1***	***APIC2 (AIC)***	***APIC3 (KIC)***	***APIClog(n) (BIC)***	***APIC4***	***APIC5***	***APIC6***	***APIC7***
30	3	1	-0.1132	0.5340	1.1812	1.4409	1.8284	2.4756	3.1229	3.7701
30	3	2	-0.1785	0.7190	1.6166	1.9767	2.5141	3.4116	4.3092	5.2067
30	3	3	-0.2414	0.8356	1.9127	2.3448	2.9897	4.0667	5.1438	6.2208
30	3	4	-0.3054	0.9120	2.1295	2.6179	3.3470	4.5644	5.7819	6.9994
30	5	1	-0.1648	0.4352	1.0352	1.2759	1.6352	2.2352	2.8352	3.4352
30	5	2	-0.2516	0.5791	1.4097	1.7430	2.2404	3.0710	3.9017	4.7324
n	***p_0***	***l***	***APIC8***	***APIC9***	***APIC10***	***APIC11***	***APIC12***	***APIC13***	***APIC14***	
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	5	1	4.0352	4.6352	5.2352	5.8352	6.4352	7.0352	7.6352	
30	5	2	5.5630	6.3937	7.2244	8.0550	8.8857	9.7163	10.5470	
n	***p_0***	***l***	***APIC1***	***APIC2 (AIC)***	***APIC3 (KIC)***	***APIC4***	***APIClog(n) (BIC)***	***APIC5***	***APIC6***	***APIC7***
100	3	1	-0.0324	0.6569	1.3463	2.0356	2.4528	2.7250	3.4143	4.1037
100	3	2	-0.0510	0.9188	1.8886	2.8584	3.4453	3.8282	4.7980	5.7678
100	3	3	-0.0687	1.1128	2.2942	3.4757	4.1907	4.6572	5.8387	7.0202
100	3	4	-0.0867	1.2703	2.6273	3.9843	4.8055	5.3413	6.6982	8.0552
100	5	1	-0.0469	0.6283	1.3035	1.9787	2.3874	2.6539	3.3292	4.0044
100	5	2	-0.0714	0.8784	1.8282	2.7780	3.3527	3.7277	4.6775	5.6273
n	***p_0***	***l***	***APIC8***	***APIC9***	***APIC10***	***APIC11***	***APIC12***	***APIC13***	***APIC14***	
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	5	1	4.6796	5.3548	6.0300	6.7052	7.3804	8.0556	8.7308	
100	5	2	6.5771	7.5269	8.4767	9.4265	10.3763	11.3261	12.2758	

$$
\begin{aligned}
&P\left\{APIC_{p_0-l} < APIC_{p_0}\right\} \\
&= P\left\{\log\left(\hat{\sigma}^2_{p_0-l}\right) + \frac{\alpha\left(p_0-l+1\right)}{n} < \log\left(\hat{\sigma}^2_{p_0}\right) + \frac{\alpha\left(p_0+1\right)}{n}\right\} \\
&= P\left\{\log\left(\frac{\hat{\sigma}^2_{p_0-l}}{\hat{\sigma}^2_{p_0}}\right) < \frac{\alpha l}{n}\right\} = P\left\{\frac{\hat{\sigma}^2_{p_0-l}}{\hat{\sigma}^2_{p_0}} < \exp\left(\frac{\alpha l}{n}\right)\right\} \\
&= P\left\{\frac{\hat{\sigma}^2_{p_0-l} - \hat{\sigma}^2_{p_0}}{\hat{\sigma}^2_{p_0}} < \exp\left(\frac{\alpha l}{n}\right) - 1\right\} \\
&= P\left\{\frac{\chi^2_l}{\chi^2_{n-p_0}} < \exp\left(\frac{\alpha l}{n}\right) - 1\right\} \\
&= P\left\{F_{l,\,n-p_0} < \frac{n-p_0}{l}\left[\exp\left(\frac{\alpha l}{n}\right) - 1\right]\right\}.
\end{aligned} \tag{15}
$$

In (14) and (15), *APIC*'s probability of over/underfitting depends on the value of α same as the signal-to-noise ratio. When we replace the values of α by 2, $\log(n)$ and 3, we have the probabilities of over/underfitting of *AIC*, *BIC* and *KIC*, respectively. If the value of α tends to infinity under the same values of *n*, p_0 and *l*, *APIC* having the low probability of overfitting but it will be prone to underfitting. The proof of the probability of over/underfitting can be confirmed numerically in Table 2 and 3. The example of the calculation for the probability of overfitting by l extra variables of *APIC*, for $n = 15$, $p_0 = 3$, $l = 1$ and $\alpha = 1$, is as follows:

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

It means that *APIC* for $\alpha = 1$ would select the model whose order is higher by one order than true model with a probability of 0.4025. In the same manner, the probability of underfitting by l variables of *APIC* for this case is

$$P\left\{APIC_{p_0-1} < APIC_{p_0}\right\} = P\left\{F_{1,\,12} < 0.8273\right\} = 0.6190.$$

It means that *APIC* for $\alpha = 1$ would select the model whose order is lower by one order than true model with a probability of 0.6190. The model selection criterion that has strong signal-to-noise ratio and lowest probability of over/underfitting is preferable. As a result, the main objective of this paper is to find the appropriate value of α, by proving and verifying the result of study with simulation data, in order to make the strength of penalty function in the model selection criterion. Then, the performance of *APIC* is examined relative to the well-known criteria, *AIC*, *BIC* and *KIC*, under various circumstances.

From Table 2 and 3 we found that when the sample size is small (*n* = 15), *KIC* has probability of overfitted less than *BIC* and *AIC*, respectively, in the opposite it has more probability of underfitted because the value of α in (3) from *KIC* is larger than *BIC* and *AIC*, respectively (3 > log(15) > 2). Whereas the sample size are moderate to large (*n* = 30, 100), *BIC* has probability of overfitted less than *KIC* and *AIC*, respectively, in the opposite it has more probability of underfitted because the value of α in (3) from *BIC* is larger than *KIC* and *AIC*, respectively (log(30) or log(100) > 3 > 2). Therefore, the conclusion is that, *APIC* with a much more value of α, make its probability of overfitting to be smaller but make more probability of underfitting.

Table 2 Probability of overfitting by l extra variables of *APIC* α for different values of n, p_0 and l .

n	p_0	l	Criteria							
			APIC1	***APIC2 (AIC)***	***APIClog(n) (BIC)***	***APIC3 (KIC)***	***APIC4***	***APIC5***	***APIC6***	***APIC7***
15	3	1	0.4025	0.2363	0.1682	0.1469	0.0939	0.0611	0.0402	0.0266
15	3	2	0.5134	0.2636	0.1644	0.1353	0.0695	0.0357	0.0183	0.0094
15	3	3	0.5947	0.2857	0.1631	0.1287	0.0561	0.0240	0.0101	0.0042
15	3	4	0.6664	0.3143	0.1701	0.1305	0.0508	0.0190	0.0070	0.0025
15	5	1	0.4511	0.2865	0.2148	0.1917	0.1316	0.0918	0.0647	0.0460
15	5	2	0.5866	0.3442	0.2359	0.2019	0.1184	0.0695	0.0408	0.0239
n	***p_0***	***l***	***APIC8***	***APIC9***	***APIC10***	APIC11	APIC12	APIC13	APIC14	
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	5	1	0.0329	0.0236	0.0170	0.0123	0.0089	0.0065	0.0047	
15	5	2	0.0140	0.0082	0.0048	0.0028	0.0017	0.0010	0.0006	
n	p_0	l	APIC1	APIC2 (AIC)	APIC3 (KIC)	APIClog(n) (BIC)	APIC4	APIC5	***APIC6***	***APIC7***
30	3	1	0.3565	0.1922	0.1102	0.0890	0.0651	0.0392	0.0239	0.0147
30	3	2	0.4346	0.1889	0.0821	0.0588	0.0357	0.0155	0.0067	0.0029
30	3	3	0.4846	0.1795	0.0617	0.0397	0.0204	0.0066	0.0021	0.0007
30	3	4	0.5256	0.1720	0.0482	0.0282	0.0125	0.0031	0.0007	0.0002
30	5	1	0.3761	0.2106	0.1252	0.1026	0.0766	0.0478	0.0301	0.0192
30	5	2	0.4646	0.2158	0.1003	0.0737	0.0466	0.0216	0.0101	0.0047
n	***p_0***	***l***	***APIC8***	***APIC9***	***APIC10***	***APIC11***	***APIC12***	***APIC13***	***APIC14***	
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	5	1	0.0123	0.0079	0.0051	0.0033	0.0022	0.0014	0.0009	
30	5	2	0.0022	0.0010	0.0005	0.0002	0.0001	0.0000	0.0000	
n	***p_0***	***l***	***APIC1***	***APIC2 (AIC)***	***APIC3 (KIC)***	***APIC4***	***APIClog(n) (BIC)***	***APIC5***	***APIC6***	***APIC7***
100	3	1	0.3284	0.1670	0.0905	0.0506	0.0360	0.0289	0.0167	0.0097
100	3	2	0.3867	0.1496	0.0578	0.0224	0.0126	0.0087	0.0033	0.0013
100	3	3	0.4178	0.1288	0.0367	0.0100	0.0045	0.0027	0.0007	0.0002
100	3	4	0.4395	0.1109	0.0236	0.0046	0.0017	0.0009	0.0002	0.0000
100	5	1	0.3336	0.1715	0.0940	0.0531	0.0380	0.0306	0.0179	0.0105
100	5	2	0.3946	0.1557	0.0614	0.0242	0.0138	0.0096	0.0038	0.0015
n	***p_0***	***l***	***APIC8***	***APIC9***	***APIC10***	***APIC11***	***APIC12***	***APIC13***	***APIC14***	
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	5	1	0.0062	0.0037	0.0022	0.0013	0.0008	0.0005	0.0003	
100	5	2	0.0006	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	

Table 3 Probability of underfitting by l variables of *APIC* α for different values of n, p_0 and l.

n	p_0	l	Criteria							
			APIC1	***APIC2 (AIC)***	***APIClog(n) (BIC)***	***APIC3 (KIC)***	***APIC4***	***APIC5***	***APIC6***	***APIC7***
15	3	1	0.6190	0.7847	0.8507	0.8709	0.9204	0.9500	0.9682	0.9796
15	3	2	0.5507	0.7981	0.8854	0.9093	0.9592	0.9817	0.9918	0.9963
15	3	3	0.5238	0.8272	0.9197	0.9418	0.9811	0.9940	0.9981	0.9994
15	3	4	0.5146	0.8581	0.9464	0.9646	0.9918	0.9982	0.9996	0.9999
15	5	1	0.5743	0.7401	0.8102	0.8324	0.8890	0.9253	0.9491	0.9651
15	5	2	0.4866	0.7364	0.8356	0.8647	0.9305	0.9643	0.9817	0.9906
n	***p_0***	***l***	***APIC8***	***APIC9***	***APIC10***	***APIC11***	***APIC12***	***APIC13***	***APIC14***	
15	3	1	0.9869	0.9915	0.9945	0.9964	0.9976	0.9984	0.9990	
15	3	2	0.9983	0.9993	0.9997	0.9998	0.9999	1.0000	1.0000	
15	3	3	0.9998	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	
15	3	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
15	5	1	0.9759	0.9833	0.9883	0.9918	0.9943	0.9960	0.9972	
15	5	2	0.9952	0.9975	0.9987	0.9993	0.9997	0.9998	0.9999	
n	***p_0***	***l***	***APIC1***	***APIC2 (AIC)***	***APIC3 (KIC)***	***APIClog(n) (BIC)***	***APIC4***	***APIC5***	***APIC6***	***APIC7***
30	3	1	0.6528	0.8163	0.8965	0.9170	0.9399	0.9645	0.9787	0.9871
30	3	2	0.5934	0.8347	0.9328	0.9532	0.9727	0.9889	0.9955	0.9982
30	3	3	0.5680	0.8612	0.9588	0.9750	0.9882	0.9967	0.9991	0.9998
30	3	4	0.5561	0.8863	0.9754	0.9870	0.9951	0.9991	0.9998	1.0000
30	5	1	0.6339	0.7988	0.8825	0.9045	0.9294	0.9567	0.9732	0.9832
30	5	2	0.5654	0.8111	0.9179	0.9412	0.9643	0.9845	0.9933	0.9971
n	***p_0***	l	APIC8	APIC9	APIC10	APIC11	APIC12	APIC13	APIC14	
30	3	1	0.9922	0.9952	0.9971	0.9982	0.9989	0.9993	0.9996	
30	3	2	0.9993	0.9997	0.9999	0.9999	1.0000	1.0000	1.0000	
30	3	3	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
30	3	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
30	5	1	0.9894	0.9933	0.9957	0.9973	0.9983	0.9989	0.9993	
30	5	2	0.9987	0.9994	0.9998	0.9999	1.0000	1.0000	1.0000	
n	p_0	l	APIC1	APIC2 (AIC)	***APIC3 (KIC)***	***APIC4***	***APIClog(n) (BIC)***	***APIC5***	***APIC6***	***APIC7***
100	3	1	0.6741	0.8352	0.9111	0.9506	0.9650	0.9720	0.9839	0.9907
100	3	2	0.6209	0.8563	0.9455	0.9793	0.9885	0.9922	0.9970	0.9989
100	3	3	0.5967	0.8808	0.9676	0.9915	0.9963	0.9978	0.9995	0.9999
100	3	4	0.5830	0.9023	0.9808	0.9965	0.9988	0.9994	0.9999	1.0000
100	5	1	0.6690	0.8308	0.9078	0.9481	0.9630	0.9703	0.9827	0.9899
100	5	2	0.6133	0.8504	0.9422	0.9776	0.9874	0.9913	0.9967	0.9987
n	***p_0***	***l***	***APIC8***	***APIC9***	***APIC10***	***APIC11***	***APIC12***	APIC13	APIC14	
100	3	1	0.9945	0.9968	0.9981	0.9989	0.9993	0.9996	0.9998	
100	3	2	0.9996	0.9998	0.9999	1.0000	1.0000	1.0000	1.0000	
100	3	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
100	3	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
100	5	1	0.9940	0.9965	0.9979	0.9987	0.9992	0.9995	0.9997	
100	5	2	0.9995	0.9998	0.9999	1.0000	1.0000	1.0000	1.0000	

Simulation Study and Results

In addition to the proofs of signal-to-noise ratio in (12) and the probability of over/underfitting in (14) and (15), we use the simulation data to find the appropriate value of α for *APIC* in (3). True multiple regression models in (1) are constructed as follows.

Model 1 (very weakly identifiable true model due to the small values of regression coefficients):

$y_1 = 1 + 0.5X_2 + 0.4X_3 + 0.3X_4 + 0.2X_5 + \varepsilon_1$, the true order $p_0 = 5$

Model 2 (weakly identifiable true model due to the small values of regression coefficients):

$y_2 = 1 + 0.5X_2 + 0.4X_3 + \varepsilon_2$, the true order $p_0 = 3$

Model 3 (strongly identifiable true model due to the large values of regression coefficients):

$y_3 = 1 + 2X_2 + 2X_3 + \varepsilon_3$, the true order $p_0 = 3$

Model 4 (very strongly identifiable true model due to the large values of regression coefficients):

$y_4 = 1 + 2X_2 + 2X_3 + 2X_4 + 2X_5 + \varepsilon_4$. the true order $p_0 = 5$

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). The error terms for all models are assumed to be $N(0, \sigma_0^2)$ where σ_0^2 in (1) is assumed equal to three levels: 0.25, 1, 9. Seven candidate variables, X_1 to X_7, are stored in an $n \times 7$ matrix $\mathbf{X}$ of the candidate model in (2). X_1 is given as a constant which equals 1, followed by six independent variables which have two distributions: $N(0,1)$ and $U(a,b)$. For the uniform distribution, we given

$$X_2 \sim U(5,10),\ X_3 \sim U(10,20),\ X_4 \sim U(7,9),$$
$$X_5 \sim U(6,11),\ X_6 \sim U(9,19),\ X_7 \sim U(4,8).$$

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* 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* is measure and used for examining the effects of weak or strong penalty function in the proposed criterion. Findings are the following.

For the very weakly identifiable situation of true models with the true orders p_0 = 5, Model 1, the sample size is small (*n* = 15) and the distribution of independent variable is normal, the appropriate values of α when the true variances σ_0^2 = 0.25, 1, 9, are 2, 1 and 1, respectively with the probabilities of correct order being selected are 29.7%, 15.5% and 11.9%. While, the distribution of independent variable is changed to be uniform, the appropriate values of α for all three levels of true variances are 1 with the probabilities are reduced to be 13.2%, 11.3% and 10.6%.

For the weakly identifiable situation of true models with the true orders p_0 = 3, Model 2, the sample size is small (*n* = 15) and the distribution of independent variable is normal, the appropriate values of α when the true variances σ_0^2 = 0.25, 1, 9, are 4, log n and 2, respectively with the probabilities of correct order being selected are 65.8%, 33.3% and 11.9%. While, the distribution of independent variable is changed to be uniform, the appropriate values of α for all three levels of true variances are 2 with the probabilities are reduced to be 17.8%, 12.6% and 13.6%.

For the strongly identifiable situation of true models with the true orders p_0 = 3, Model 3, the sample size is small (*n* = 15) and the distribution of independent variable is normal, the appropriate values of α when the true variances σ_0^2 = 0.25, 1, 9, are 14, 9 and 4, respectively with the probabilities of correct order being selected are 99.8%, 97.7% and 55.4%. While, the distribution of independent variable is changed to be uniform, the appropriate values of α for three levels of true variances are 6, 4 and log n with the probabilities are reduced to be 85.8%, 48.5% and 15.7%.

For the very strongly identifiable situation of true models with the true orders p_0 = 5, Model 4, the sample size is small (*n* = 15) and the distribution of independent variable is normal, the appropriate values of α when the true variances σ_0^2 = 0.25, 1, 9, are 13, 7 and log n, respectively with the probabilities of correct order being selected are 98.5%, 91.6% and 46.6%. While, the distribution of independent variable is changed to be uniform, the appropriate values of α for three levels of true variances are 5, log n and 1 with the probabilities are reduced to be 78.2%, 42.3% and 14.8%.

For very weakly identifiable situation of true models with the true orders α = 5, Model 1, the sample size is moderate (*n* = 30) and the distribution of independent variable is normal, the appropriate values of α when the true variances σ_0^2 = 0.25, 1, 9, are 3, 1 and 1, respectively with the probabilities of correct order being selected are 55%, 24.6% and 13.5%. While, the distribution of independent variable is changed to be uniform, the appropriate values of α for all three levels of true variances are 1 with the probabilities are reduced to be 17.5%, 13% and 13.3%.

For the weakly identifiable situation of true models with the true orders p_0 = 3, Model 2, the sample size is moderate (*n* = 30) and the distribution of independent variable is normal, the appropriate values of α when the true variances σ_0^2 = 0.25, 1, 9, are 5, 3 and 2, respectively with the probabilities of correct order being selected are 90.8%, 55.5% and 18.5%. While, the distribution of independent variable is changed to be uniform, the appropriate values of α for all three levels of true variances are 2 with the probabilities are reduced to be 29.2%, 16.6% and 11.8%.

For strongly identifiable situation of true models with the true orders p_0 = 3, Model 3, the sample size is moderate (*n* = 30) and the distribution of independent variable is normal, the appropriate values of α when the true variances σ_0^2 = 0.25, 1, 9, are 14, 11 and 5, respectively with the probabilities of correct order being selected are 100%, 99.9% and 85.5%. While, the distribution of independent variable is changed to be uniform, the appropriate values of α for three levels of true variances are 10, 4 and 2 with the probabilities are reduced to be 98.8%, 75.9% and 23.2%.

For very strongly identifiable situation of true models with the true orders p_0 = 5, Model 4, the sample size is moderate (*n* = 30) and the distribution of independent variable is normal, the appropriate values of α when the true variances σ_0^2 = 0.25, 1, 9, are 14, 14 and 4, respectively with the probabilities of correct order being selected are 100%, 100% and 79.7%. While, the distribution of independent variable is changed to be uniform, the appropriate values of α for three levels of true variances are 8, 3 and 1 with the probabilities are reduced to be 98.6%, 72.3% and 22.3%.

For very weakly identifiable situation of true models with the true orders p_0 = 5, Model 1, the sample size is large (*n* = 100) and the distribution of independent variable is normal, the appropriate values of α when the true variances σ_0^2 = 0.25, 1, 9, are 5, 2 and 1, respectively with the probabilities of correct order being selected are 91.4%, 53.5% and 17.4%. While, the distribution of independent variable is changed to be uniform, the appropriate values of α for all three levels of true variances are 1 with the probabilities are reduced to be 31.3%, 18% and 11.7%.

For weakly identifiable situation of true models with the true orders p_0 = 3, Model 2, the sample size is large (*n* = 100) and the distribution of independent variable is normal, the appropriate values of α when the true variances σ_0^2 = 0.25, 1, 9, are 12, 5 and 2, respectively with the probabilities of correct order being selected are 100%, 92.9% and 33.9%. While, the distribution of independent variable is changed to be uniform, the appropriate values of α for three levels of true variances are 3, 2 and 1 with the probabilities are reduced to be 63.5%, 28.7% and 12.5%.

For strongly identifiable situation of true models with the true orders p_0 = 3, Model 3, the sample size is large (*n* = 100) and the distribution of independent variable is normal, the appropriate values of α when the true variances σ_0^2 = 0.25, 1, 9, are 12, 13 and 9, respectively with the probabilities of correct order being selected are 100%, 100% and 99.3%. While, the distribution of independent variable is changed to be uniform, the appropriate values of α for three levels of true variances are 14, 8 and 2 with the probabilities are reduced to be 100%, 99.3% and 50.6%.

For very strongly identifiable situation of true models with the true orders p_0 = 5, Model 4, the sample size is large (*n* = 100) and the distribution of independent variable is normal, the appropriate values of α when the true variances σ_0^2 = 0.25, 1, 9, are 14, 10 and 10, respectively with the probabilities of correct order being selected are 99.9%, 99.9% and 99.5%. While, the distribution of independent variable is changed to be uniform, the appropriate values of α for three levels of true variances are 13, 9 and 2 with the probabilities are reduced to be 100%, 99.3% and 49.2%.

From the results in **1.** to **12.** the conclusion is that, the weakly or very weakly identifiable situations of true models, Model 1 and Model 2, the true orders p_0 = 3, 5 and the distribution of independent variable is normal, the appropriate α is small. If sample size increases and variances of error terms are small (σ_0^2 = 0.25) to moderate (σ_0^2 = 1), α should be moderate. For the distribution of independent variable is changed to be uniform, the appropriate α is small, regardless the sample size or the variances of error terms. When the true model is very weakly identifiable, the appropriate α should be smaller than the weakly identifiable situation. The strongly or very strongly identifiable situations of true models, Model 3 and Model 4, the true orders p_0 = 3, 5 and the distribution of independent variable is normal, the appropriate α is large. If the variance of error terms increases, α should be moderate. For the distribution of independent variable is changed to be uniform, the appropriate α is moderate. If the variance of error terms increases, α should be small.

Conclusions, Discussion and Future Works

In this paper, we propose the model selection criterion, called Adjusted Penalty Information Criterion,

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

when the values of α are equal to 2, $\log(n)$ and 3; *APIC* 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 probability of over/underfitting, including the problem of signal-to-noise ratio being weak. 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. However, the probability of underfitting tends to one. At the same time, the results of simulation suggest that, the appropriate α is small when true models are weakly or very weakly identifiable and distributions of independent variables are normal or uniform. But α should be moderate, if distribution of independent variables is normal, sample size increases and variances of error terms are small to moderate. The appropriate α is large, if the true model is strongly identifiable, distribution of independent variables is normal, and variance of error terms is small to moderate. But α should be moderate, if the variance of error terms increases. When the distribution of independent variables changes to be uniform, the appropriate α is moderate for the case of variance of error terms is small to moderate. But α should be small, if the variance of error terms increases. All of these conclusions can be summarized in Table 4. The variance of error terms and sample size affects the validity of *APIC*. The variance of error terms increases, the validity of *APIC* decreases. Whereas the sample size increases, the validity of *APIC* also increases. In further work, we attempt to construct the model selection criteria to correct the weak signal-to-noise ratio and to reduce the probability of over/underfitting in the multivariate regression and simultaneous equations models.

Table 4 Appropriate value of α in *APIC*.

<table>
<tr><th rowspan="2">Model</th><th rowspan="2">n</th><th colspan="3">X~ Normal</th><th colspan="3">X~ Uniform</th></tr>
<tr><th>σ^2= 0.25</th><th>σ^2= 1</th><th>σ^2= 9</th><th>σ^2 = 0.25</th><th>σ^2= 1</th><th>σ^2 = 9</th></tr>
<tr><td rowspan="3">Model 1, 2 Weakly</td><td>15</td><td colspan="3">small</td><td colspan="3" rowspan="3">small</td></tr>
<tr><td>30</td><td colspan="2" rowspan="2">moderate</td><td rowspan="2">small</td></tr>
<tr><td>100</td></tr>
<tr><td rowspan="3">Model 3, 4 Strongly</td><td>15</td><td colspan="2" rowspan="3">large</td><td rowspan="3">moderate</td><td colspan="2" rowspan="3">moderate</td><td rowspan="3">small</td></tr>
<tr><td>30</td></tr>
<tr><td>100</td></tr>
</table>

Acknowledgements

This project is financial supported by the Thailand Research Fund and Thaksin University under grant No. TRG5780219.

References

1. Akaike H. Information theory and an extension of the maximum likelihood principle. Proceedings of the 2nd International Symposium on Information Theory.

Akademia Kiado. Budapest; 1973. P. 267-281.

2. Akaike H. A new look at the statistical model identification. IEEE T. Automat. Contr. 1974; 19: 716-723.
3. Bedrick EJ, Tsai CL. Model selection for multivariate regression in small samples. Biometrics 1994; 50: 226-231.
4. Schwarz G. Estimating the dimension of a model. Ann. Stat. 1978; 6: 461-464.
5. Cavanaugh JE. A large-sample model selection criterion based on Kullback's symmetric divergence. Stat. Probabil. Lett. 1999; 42: 333-343.
6. Hafidi B, Mkhadri A. A corrected akaike criterion based on Kullback's symmetric divergence: applications in time series: multiple and multivariate regression. Comput. Stat. Data. An. 2006; 50: 1524-1550.
7. Neath A, Cavanaugh JE. Regression and time series model selection using variants of the Schwarz information criterion. Commun. Stat. Theory. 1997; 26: 559-580.
8. Cavanaugh JE. Criteria for LINEAR MODEL SELECTION BASED on Kullback's symmetric divergence. Aust. NZ. J. Stat. 2004; 46: 257-274.
9. McQuarrie AD, Shumway R, Tsai CL. The model selection criterion AIC_u. Stat. Probabil. Lett. 1997; 34: 285-292.
10. McQuarrie AD. A small-sample correction for the Schwarz SIC model selection criterion. Stat. Probabil. Lett. 1999; 44: 79-86.
11. Mills JA, Prasad K. A comparison of model selection criteria. Econometric Reviews. 2007; 11: 201-234.
12. Rahman MS, King ML. Improved model selection criterion. Commun. Stat. Simulat. 2007; 28: 51-71.
13. Montgomery DC, Peck EA, Vining GG. Introduction to linear regression analysis. 4th ed. New York: Wiley; 2006.
14. McQuarrie AD, Tsai CL. Regression and time series model selection. Singapore: World Scientific; 1998.
15. Seghouane AK. Multivariate regression model selection from small samples using Kullback's symmetric divergence. Signal Processing 2006; 86: 2074-2084.