



รายงานวิจัยฉบับสมบูรณ์

โครงการ

Model Selection Criterion to Overcome the Weak Signal-to-Noise Ratio
and to Reduce the Probability of Over/Underfitting

โดย

ผู้ช่วยศาสตราจารย์ ดร.วรางคณา กীরติวิบูลย์
มหาวิทยาลัยทักษิณ วิทยาเขตพัทลุง

มิถุนายน 2559

รายงานวิจัยฉบับสมบูรณ์

โครงการ

Model Selection Criterion to Overcome the Weak Signal-to-Noise Ratio
and to Reduce the Probability of Over/Underfitting

โดย

ผู้ช่วยศาสตราจารย์ ดร.วรางคณา กิริติวิบูลย์
มหาวิทยาลัยทักษิณ วิทยาเขตพัทลุง

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย และมหาวิทยาลัยทักษิณ

(ความเห็นในรายงานนี้เป็นของผู้วิจัย สกว. และมหาวิทยาลัยทักษิณ ไม่จำเป็นต้องเห็นด้วยเสมอไป)

EXECUTIVE SUMMARY

Background

The problem of choosing the appropriate regression models 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 Akaike information criterion, *AIC* (Akaike, 1973, 1974; Bedrick and Tsai, 1994). 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* (Schwarz, 1978; Gorobets, 2005), and Kullback information criterion, *KIC* (Cavanaugh, 1999; Hafidi and Mkhadri, 2006). *BIC* is an asymptotic approximation to a transformation of Bayesian posterior probability of a candidate model (Neath and Cavanaugh, 1997). *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 two directed divergences is Kullback's symmetric divergence, also known as the J-divergence (Cavanaugh, 1999; Cavanaugh, 2004). 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* (Neath and Cavanaugh, 1997). 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 (Cavanaugh, 1999; Cavanaugh, 2004). 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 research is to propose a model selection criterion to overcome 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 various circumstances as follows (McQuarrie, Shumway, and Tsai, 1997; McQuarrie, 1999; Mills and Prasad, 2007; Rahman and King, 2007).

- Sample sizes (n) are difference,
- Orders of true model (p_0) are difference,
- Regression coefficients (β) are difference,
- Variances of error terms (σ^2) are difference,
- Distributions of independent variables (Dist. of $\mathbf{X}$) are difference.

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 frequency of correct order being selected.

Objectives of the Research

The objectives of this research are as follows:

1) To derive the model selection criterion in order to overcome the weak signal-to-noise ratio and to reduce the probability of over/underfitting in univariate regression model, called adjusted penalty information criterion (*APIC*).

2) To examine the performance of *APIC*, the proposed model selection criterion, relative to the well-known criteria, *AIC*, *BIC*, and *KIC*, under various circumstances as follows:

- Sample sizes (n) are difference,
- Orders of true model (p_0) are difference,
- Regression coefficients (β) are difference,

- Variances of error terms (σ^2) are difference,
- Distributions of independent variables (Dist. of $\mathbf{X}$) are difference.

Methodologies

The methodologies of this research are as follows:

1) Derive the model selection criterion in order to overcome the weak signal-to-noise ratio and to reduce the probability of over/underfitting in univariate regression model, called adjusted penalty information criterion, denoted by *APIC*.

2) Use the SAS programming to simulate the univariate regression model under various circumstances as follows:

- Sample sizes (n) are difference,
- Orders of true model (p_0) are difference,
- Regression coefficients (β) are difference,
- Variances of error terms (σ^2) are difference,
- Distributions of independent variables (Dist. of $\mathbf{X}$) are difference.

3) Examine the properties of the model must be consistent with step 2.

4) Calculate the values of *APIC*, the proposed model selection criterion, and those of well-known criteria, *AIC*, *BIC*, and *KIC*, under various circumstances in step 2.

5) Examine the performance of *APIC*, the proposed model selection criterion, relative to the well-known criteria, *AIC*, *BIC*, and *KIC*, under various circumstances in step 2. 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 frequency of correct order being selected.

Plans of the Research

The plans of the research are as follows:

Activities	Jun 15, 2014	Dec 15, 2014	Jun 15, 2015	Dec 15, 2015
	–	–	–	–
	Dec 14, 2014	Jun 14, 2015	Dec 14, 2015	Jun 14, 2016
1) Derive the model selection criterion in order to overcome the weak signal-to-noise ratio and to reduce the probability of over/underfitting in univariate regression model, called adjusted penalty information criterion, denoted by <i>APIC</i> .				
2) Use the SAS programming to simulate the univariate regression model under various circumstances as follows: • Sample sizes (n) are difference, • Orders of true model (p_0) are difference, • Regression coefficients (β) are difference, • Variances of error terms (σ^2) are difference, • Distributions of independent variables (Dist. of X) are difference.				
3) Examine the properties of the model must be consistent with step 2.				
4) Calculate the values of <i>APIC</i> , the proposed model selection criterion, and those of well-known criteria, <i>AIC</i> , <i>BIC</i> , and <i>KIC</i> , under various circumstances in step 2.				
5) Examine the performance of <i>APIC</i> , the proposed model selection criterion, relative to the well-known criteria, <i>AIC</i> , <i>BIC</i> , and <i>KIC</i> , under various circumstances in step 2. 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 frequency of correct order being selected.				

ABSTRACT

PROJECT CODE : TRG5780219
PROJECT TITLE : Model Selection Criterion to Overcome the Weak Signal-to-Noise Ratio and to Reduce the Probability of Over/Underfitting
INVESTIGATOR : Asst.Prof.Dr. Warangkhan Keerativibool
Thaksin University, Phatthalung Campus
E-MAIL ADDRESS : warang27@gmail.com
PROJECT PERIOD : June 15, 2014 – June 14, 2016
ABSTRACT :

This research proposed a model selection criterion in order to overcome the weak signal-to-noise ratio and to reduce the probability of over/underfitting by adjusting the penalty term of the well-known model selection criteria (*AIC*, *BIC*, *KIC*), called adjusted penalty information criterion, $APIC = \log(\hat{\sigma}^2) + \alpha(p+1)/n$. 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, 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.

ACKNOWLEDGEMENTS

This research would not have been completed without the help and support of several people and my organization, Thaksin University. Thank is also dedicated to the research fund by The Thailand Research Fund and Thaksin University under grant No. TRG5780219.

I would like to express my sincere gratitude and appreciation to Assoc. Prof. Dr. Pachitjanut Siripanich from Dhurakij Pundit University for their valuable and continual advice, encouragement, and constructive criticisms throughout this research, which enabled me to complete this research successfully. Appreciation is also tendered to all teachers in the School of Applied Statistics, National Institute of Development Administration for the knowledge provided.

Finally, I would particularly like to thank my family for their patience, understanding, and encouragement throughout the duration of this work. They had more faith in me than could ever be justified by logical argument.

Asst.Prof.Dr. Warangkhana Keerativibool

June 2016

TABLE OF CONTENTS

	Page
EXECUTIVE SUMMARY	
ABSTRACT	
ACKNOWLEDGEMENTS	
TABLE OF CONTENTS	
CHAPTER 1 INTRODUCTION	1
1.1 Background	1
1.2 Objectives of the Research	3
1.3 Scope of the Research	3
CHAPTER 2 LITERATURE REVIEW	4
CHAPTER 3 METHODOLOGY	11
CHAPTER 4 SIMULATION STUDY	18
CHAPTER 5 CONCLUSIONS AND FUTURE WORKS	30
5.1 Conclusions	30
5.2 Future Works	31
BIBLIOGRAPHY	32
APPENDIX	34
Outputs of this Research	35

CHAPTER 1

INTRODUCTION

1.1 Background

The problem of choosing the appropriate regression models 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 Akaike information criterion, *AIC* (Akaike, 1973, 1974; Bedrick and Tsai, 1994). 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* (Schwarz, 1978; Gorobets, 2005), and Kullback information criterion, *KIC* (Cavanaugh, 1999; Hafidi and Mkhadri, 2006). *BIC* is an asymptotic approximation to a transformation of Bayesian posterior probability of a candidate model (Neath and Cavanaugh, 1997). *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 two directed divergences is Kullback's symmetric divergence, also known as the J-divergence (Cavanaugh, 1999; Cavanaugh, 2004). 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* (Neath and Cavanaugh, 1997). 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 (Cavanaugh, 1999; Cavanaugh, 2004). 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 research is to propose a model selection criterion to overcome 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 various circumstances as follows (McQuarrie, Shumway, and Tsai, 1997; McQuarrie, 1999; Mills and Prasad, 2007; Rahman and King, 2007).

- Sample sizes (n) are difference,
- Orders of true model (p_0) are difference,
- Regression coefficients (β) are difference,
- Variances of error terms (σ^2) are difference,
- Distributions of independent variables (Dist. of $\mathbf{X}$) are difference.

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 frequency of correct order being selected.

1.2 Objectives of the Research

The objectives of this research are as follows:

1) To derive the model selection criterion in order to overcome the weak signal-to-noise ratio and to reduce the probability of over/underfitting in univariate regression model, called adjusted penalty information criterion (*APIC*).

2) To examine the performance of *APIC*, the proposed model selection criterion, relative to the well-known criteria, *AIC*, *BIC*, and *KIC*, under various circumstances as follows:

- Sample sizes (n) are difference,
- Orders of true model (p_0) are difference,
- Regression coefficients (β) are difference,
- Variances of error terms (σ^2) are difference,
- Distributions of independent variables (Dist. of $\mathbf{X}$) are difference.

1.3 Scope of the Research

In this research, the model selection criterion focuses on the univariate regression model (Montgomery, Peck, and Vining, 2006),

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (1.1)$$

where $\mathbf{y}$ is an $n \times 1$ dependent random vector of observations, $\mathbf{X}$ is a $n \times p$ matrix of independent variables with full-column rank, $\boldsymbol{\beta}$ is a $p \times 1$ parameter vector of regression coefficients, $\boldsymbol{\varepsilon}$ is an $n \times 1$ error vector with zero mean and variance $\sigma^2 \mathbf{I}_n$.

The maximum likelihood estimators of $\boldsymbol{\beta}$ and σ^2 are, respectively,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \text{ and } \hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \quad (1.2)$$

CHAPTER 2

LITERATURE REVIEW

The generating or true univariate regression model to consider in this research is in the form (Montgomery, Peck, and Vining, 2006)

$$\mathbf{y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}_0, \quad (2.1)$$

and the candidate or approximating univariate regression model is in the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.2)$$

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 variances $\sigma_0^2 \mathbf{I}_n$ and $\sigma^2 \mathbf{I}_n$, respectively. The maximum likelihood estimators of $\boldsymbol{\beta}$ and σ^2 are, respectively,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \text{ and } \hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \quad (2.3)$$

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 research. 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,

$$\text{Model Selection Criterion} = MSC = \log(\hat{\sigma}^2) + \frac{\alpha(p+1)}{n}. \quad (2.4)$$

When the values of α in (2.4) are equal to 2, $\log(n)$, and 3, *MSC* becomes *AIC* (Akaike, 1973, 1974), *BIC* (Schwarz, 1978), and *KIC* (Cavanaugh, 1999), respectively, i.e.,

$$AIC = \log(\hat{\sigma}^2) + \frac{2(p+1)}{n}, \quad (2.5)$$

$$BIC = \log(\hat{\sigma}^2) + \frac{(p+1)\log(n)}{n}, \quad (2.6)$$

$$KIC = \log(\hat{\sigma}^2) + \frac{3(p+1)}{n}. \quad (2.7)$$

In this research, the methods used to compare which criterion is the best are the ratio of signal-to-noise and the probability of over/underfitting. McQuarrie and Tsai (1998) 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 the criterion value of a model of order p_0 is less than that of order $p_0 + l$, $MSC_{p_0} < MSC_{p_0+l}$. Then the signal-to-noise ratio that the true model is selected compared to a candidate model is

$$\begin{aligned} \frac{\text{signal}}{\text{noise}} &= \frac{E[MSC_{p_0+l} - MSC_{p_0}]}{sd[MSC_{p_0+l} - MSC_{p_0}]} \\ &= \frac{E\left[\log(\hat{\sigma}_{p_0+l}^2) + \frac{\alpha(p_0+l+1)}{n} - \log(\hat{\sigma}_{p_0}^2) - \frac{\alpha(p_0+1)}{n}\right]}{sd\left[\log(\hat{\sigma}_{p_0+l}^2) + \frac{\alpha(p_0+l+1)}{n} - \log(\hat{\sigma}_{p_0}^2) - \frac{\alpha(p_0+1)}{n}\right]} \\ &= \frac{E\left[\log\left(\frac{\hat{\sigma}_{p_0+l}^2}{\hat{\sigma}_{p_0}^2}\right) + \frac{\alpha l}{n}\right]}{sd\left[\log\left(\frac{\hat{\sigma}_{p_0+l}^2}{\hat{\sigma}_{p_0}^2}\right) + \frac{\alpha l}{n}\right]}. \end{aligned} \quad (2.8)$$

In order to find the signal in (2.8), we apply the second-order of Taylor's series expansions as follows. Suppose $X \sim \chi_p^2$, expanding $\log(X)$ about $E(X) = p$, we have

$$\log(X) \doteq \log(p) + \frac{1}{p}(X - p) - \frac{1}{2p^2}(X - p)^2 \text{ and } E[\log(X)] \doteq \log(p) - \frac{1}{p}. \quad (2.9)$$

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

$$n(\hat{\sigma}_p^2 - \hat{\sigma}_{p+l}^2) \sim \sigma_0^2 \chi_l^2, \quad n\hat{\sigma}_p^2 \sim \sigma_0^2 \chi_{n-p}^2, \quad \text{and } \hat{\sigma}_p^2 - \hat{\sigma}_{p+l}^2 \text{ is independent of } \hat{\sigma}_{p+l}^2, \quad (2.10)$$

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

Using the result of Taylor's series expansions in (2.9) and the assumptions in (2.10), we have

$$E \left[\log \left(\frac{n\hat{\sigma}_p^2}{\sigma_0^2} \right) \right] = E \left[\log(n\hat{\sigma}_p^2) \right] - E \left[\log(\sigma_0^2) \right] \doteq \log(n-p) - \frac{1}{n-p},$$

then

$$E \left[\log(n\hat{\sigma}_p^2) \right] \doteq \log(\sigma_0^2) + \log(n-p) - \frac{1}{n-p}. \quad (2.11)$$

From (2.11), the signal in (2.8) is approximated by

$$\begin{aligned} E[MSC_{p_0+l} - MSC_{p_0}] &= E \left[\log \left(\frac{\hat{\sigma}_{p_0+l}^2}{\hat{\sigma}_{p_0}^2} \right) + \frac{\alpha l}{n} \right] \\ &= E \left[\log(n\hat{\sigma}_{p_0+l}^2) \right] - E \left[\log(n\hat{\sigma}_{p_0}^2) \right] + \frac{\alpha l}{n} \\ &\doteq \left\{ \log(\sigma_0^2) + \log(n-p_0-l) - \frac{1}{n-p_0-l} \right\} - \left\{ \log(\sigma_0^2) + \log(n-p_0) - \frac{1}{n-p_0} \right\} + \frac{\alpha l}{n} \\ &= \log \left(\frac{n-p_0-l}{n-p_0} \right) - \frac{l}{(n-p_0-l)(n-p_0)} + \frac{\alpha l}{n}. \end{aligned} \quad (2.12)$$

In order to find the noise in (2.8), we use the assumptions in (2.10), then we have

$$Q = \frac{n\hat{\sigma}_{p_0+l}^2}{n\hat{\sigma}_{p_0}^2} \sim \frac{\chi_{n-p_0-l}^2}{\chi_{n-p_0-l}^2 + \chi_l^2}, \quad (2.13)$$

the Q -statistic in (2.13) has the Beta distribution

$$Q \sim \text{Beta} \left(\frac{n-p_0-l}{2}, \frac{l}{2} \right),$$

and the log-distribution is

$$\log(Q) = \log\left(\frac{n\hat{\sigma}_{p_0+l}^2}{n\hat{\sigma}_{p_0}^2}\right) \sim \log\text{-Beta}\left(\frac{n-p_0-l}{2}, \frac{l}{2}\right). \quad (2.14)$$

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

$$\log(X) \doteq \log(p) + \frac{1}{p}(X - p). \quad (2.15)$$

Applying (2.15) to expand $\log(Q)$ in (2.14) about

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

we have

$$\log(Q) \doteq \log\left(\frac{n-p_0-l}{n-p_0}\right) + \frac{n-p_0}{n-p_0-l} \left(Q - \frac{n-p_0-l}{n-p_0}\right). \quad (2.16)$$

The variance of $\log(Q)$ in (2.14) is approximated by the variance $\log(Q)$ in (2.16) as follows:

$$\begin{aligned} \text{var}[\log(Q)] &= \text{var}\left[\log\left(\frac{n\hat{\sigma}_{p_0+l}^2}{n\hat{\sigma}_{p_0}^2}\right)\right] \\ &\doteq \text{var}\left[\log\left(\frac{n-p_0-l}{n-p_0}\right) + \frac{n-p_0}{n-p_0-l} \left(Q - \frac{n-p_0-l}{n-p_0}\right)\right] \\ &= \left(\frac{n-p_0}{n-p_0-l}\right)^2 \text{var}(Q) \\ &= \frac{(n-p_0)^2}{(n-p_0-l)^2} \left[\frac{\frac{n-p_0-l}{2} \cdot \frac{l}{2}}{\left(\frac{n-p_0-l}{2} + \frac{l}{2}\right)^2 \left(\frac{n-p_0-l}{2} + \frac{l}{2} + 1\right)} \right] \\ &= \frac{2l}{(n-p_0-l)(n-p_0+2)}. \end{aligned} \quad (2.17)$$

Therefore, the standard deviation of $\log(Q)$ in (2.17) or the approximate noise in (2.8) is

$$\begin{aligned} sd[MSC_{p_0+l} - MSC_{p_0}] &= sd \left[\log \left(\frac{\hat{\sigma}_{p_0+l}^2}{\hat{\sigma}_{p_0}^2} \right) + \frac{\alpha l}{n} \right] = sd \left[\log \left(\frac{n \hat{\sigma}_{p_0+l}^2}{n \hat{\sigma}_{p_0}^2} \right) \right] \\ &= sd[\log(Q)] \doteq \frac{\sqrt{2l}}{\sqrt{(n-p_0-l)(n-p_0+2)}}. \end{aligned} \quad (2.18)$$

Combined the approximations of signal in (2.12) and noise in (2.18) to be the approximate signal-to-noise ratio in (2.8) as follows:

$$\begin{aligned} \frac{signal}{noise} &= \frac{E[MSC_{p_0+l} - MSC_{p_0}]}{sd[MSC_{p_0+l} - MSC_{p_0}]} \\ &\doteq \frac{\sqrt{(n-p_0-l)(n-p_0+2)}}{\sqrt{2l}} \left[\log \left(\frac{n-p_0-l}{n-p_0} \right) - \frac{l}{(n-p_0-l)(n-p_0)} + \frac{\alpha l}{n} \right]. \end{aligned} \quad (2.19)$$

Replacing the values of α in (2.19) by 2, $\log(n)$, and 3, we have the approximate signal-to-noise ratios for *AIC*, *BIC*, and *KIC*, respectively, i.e.,

$$\begin{aligned} \frac{E[AIC_{p_0+l} - AIC_{p_0}]}{sd[AIC_{p_0+l} - AIC_{p_0}]} &\doteq \frac{\sqrt{(n-p_0-l)(n-p_0+2)}}{\sqrt{2l}} \left[\log \left(\frac{n-p_0-l}{n-p_0} \right) - \frac{l}{(n-p_0-l)(n-p_0)} + \frac{2l}{n} \right], \end{aligned} \quad (2.20)$$

$$\begin{aligned} \frac{E[BIC_{p_0+l} - BIC_{p_0}]}{sd[BIC_{p_0+l} - BIC_{p_0}]} &\doteq \frac{\sqrt{(n-p_0-l)(n-p_0+2)}}{\sqrt{2l}} \left[\log \left(\frac{n-p_0-l}{n-p_0} \right) - \frac{l}{(n-p_0-l)(n-p_0)} + \frac{l \log(n)}{n} \right], \end{aligned} \quad (2.21)$$

$$\begin{aligned} & \frac{E[KIC_{p_0+l} - KIC_{p_0}]}{sd[KIC_{p_0+l} - KIC_{p_0}]} \\ & \doteq \frac{\sqrt{(n-p_0-l)(n-p_0+2)}}{\sqrt{2l}} \left[\log\left(\frac{n-p_0-l}{n-p_0}\right) - \frac{l}{(n-p_0-l)(n-p_0)} + \frac{3l}{n} \right]. \end{aligned} \quad (2.22)$$

The probability of overfitting is the second method used to compare which criterion is the best. It is defined based on a model that has more parameters than the optimal model (Seghouane, 2006). The probabilities of the criteria *AIC*, *BIC*, and *KIC* preferring the overfitted model by l extra variables are 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 *MSC* prefers the overfitted model is defined by

$$\begin{aligned} P\{MSC_{p_0+l} < MSC_{p_0}\} &= P\left\{\log(\hat{\sigma}_{p_0+l}^2) + \frac{\alpha(p_0+l+1)}{n} < \log(\hat{\sigma}_{p_0}^2) + \frac{\alpha(p_0+1)}{n}\right\} \\ &= P\left\{\log\left(\frac{\hat{\sigma}_{p_0}^2}{\hat{\sigma}_{p_0+l}^2}\right) > \frac{\alpha l}{n}\right\} = P\left\{\frac{\hat{\sigma}_{p_0}^2}{\hat{\sigma}_{p_0+l}^2} > \exp\left(\frac{\alpha l}{n}\right)\right\} = P\left\{\frac{\hat{\sigma}_{p_0}^2 - \hat{\sigma}_{p_0+l}^2}{\hat{\sigma}_{p_0+l}^2} > \exp\left(\frac{\alpha l}{n}\right) - 1\right\}. \end{aligned} \quad (2.23)$$

Using the assumptions in (2.10), the probability of overfitting by l extra variables for *MSC* in (2.23) becomes

$$\begin{aligned} P\{MSC_{p_0+l} < MSC_{p_0}\} &= P\left\{\frac{\chi_l^2}{\chi_{n-p_0-l}^2} > \exp\left(\frac{\alpha l}{n}\right) - 1\right\} \\ &= P\left\{F_{l, n-p_0-l} > \frac{n-p_0-l}{l} \left[\exp\left(\frac{\alpha l}{n}\right) - 1\right]\right\}. \end{aligned} \quad (2.24)$$

Replacing the values of α in (2.24) by 2, $\log(n)$, and 3, we have the probabilities of overfitting by l extra variables for *AIC*, *BIC*, and *KIC*, respectively, i.e.,

$$P\{AIC_{p_0+l} < AIC_{p_0}\} = P\left\{F_{l, n-p_0-l} > \frac{n-p_0-l}{l} \left[\exp\left(\frac{2l}{n}\right) - 1\right]\right\}, \quad (2.25)$$

$$P\{BIC_{p_0+l} < BIC_{p_0}\} = P\left\{F_{l, n-p_0-l} > \frac{n-p_0-l}{l} \left[\exp\left(\frac{l \log(n)}{n}\right) - 1 \right] \right\}, \quad (2.26)$$

$$P\{KIC_{p_0+l} < KIC_{p_0}\} = P\left\{F_{l, n-p_0-l} > \frac{n-p_0-l}{l} \left[\exp\left(\frac{3l}{n}\right) - 1 \right] \right\}. \quad (2.27)$$

From (2.20) until (2.22) and (2.25) until (2.27), we found that the signal-to-noise ratios and probabilities of overfitting from *AIC*, *BIC*, and *KIC* depend on the values of α in (2.4). Therefore, any criterion has the value of α is greater than 2 (from *AIC*), or greater than $\log(n)$ (from *BIC*), or greater than 3 (from *KIC*), means that the signal-to-noise ratio tend to strong and the model tends to less overfitting. However, if the values of α is too large the signal-to-noise ratio becomes weak in the underfitting case, and the model selection criterion will be prone to underfitting. McQuarrie and Tsai (1998) concluded that, a strong signal-to-noise ratio refers to a large positive value (often greater than 2) and then 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 then results in high probability of overfitting. The model selection criterion that has strong signal-to-noise ratio and lowest probability of overfitting is preferable.

CHAPTER 3

METHODOLOGY

This research attempted to derive the model selection criterion in order to overcome the weak signal-to-noise ratio and to reduce the probability of over/underfitting in univariate regression model, called adjusted penalty information criterion (*APIC*) and to examine the performance of *APIC* relative to the well-known criteria, *AIC*, *BIC*, and *KIC*, under various circumstances as follows:

- Sample sizes (n) are difference,
- Orders of true model (p_0) are difference,
- Regression coefficients (β) are difference,
- Variances of error terms (σ^2) are difference,
- Distributions of independent variables (Dist. of $\mathbf{X}$) are difference.

Recalled the equation (2.4) in Chapter 2 as

$$APIC = \log(\hat{\sigma}^2) + \frac{\alpha(p+1)}{n}. \quad (3.1)$$

APIC in (3.1) has the signal-to-noise ratio as shown in the equation (2.19),

$$\frac{\text{signal}}{\text{noise}} \doteq \frac{\sqrt{(n-p_0-l)(n-p_0+2)}}{\sqrt{2l}} \left[\log\left(\frac{n-p_0-l}{n-p_0}\right) - \frac{l}{(n-p_0-l)(n-p_0)} + \frac{\alpha l}{n} \right]. \quad (3.2)$$

In (3.2), we found that 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 3.1. 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:

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

From Table 3.1 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.

Table 3.1 Signal-to-noise ratio of $APIC\alpha$ for different values of n , p_0 and l .

n	p_0	l	<i>Criteria</i>														
			$APIC1$	$APIC2$ (AIC)	$APIC$ $\log(n)$ (BIC)	$APIC3$ (KIC)	$APIC4$	$APIC5$	$APIC6$	$APIC7$	$APIC8$	$APIC9$	$APIC10$	$APIC11$	$APIC12$	$APIC13$	$APIC14$
15	3	1	-0.2450	0.3400	0.7542	0.9250	1.5100	2.0950	2.6800	3.2650	3.8500	4.4350	5.0200	5.6050	6.1900	6.7750	7.3600
15	3	2	-0.3884	0.4004	0.9589	1.1892	1.9780	2.7668	3.5556	4.3444	5.1333	5.9221	6.7109	7.4997	8.2885	9.0773	9.8661
15	3	3	-0.5291	0.3874	1.0364	1.3039	2.2204	3.1370	4.0535	4.9700	5.8865	6.8030	7.7195	8.6360	9.5526	10.4691	11.3856
15	3	4	-0.6752	0.3225	1.0290	1.3203	2.3181	3.3159	4.3136	5.3114	6.3092	7.3070	8.3047	9.3025	10.3003	11.2981	12.2958
15	5	1	-0.3660	0.1239	0.4708	0.6138	1.1037	1.5936	2.0835	2.5734	3.0633	3.5532	4.0431	4.5330	5.0229	5.5128	6.0027
15	5	2	-0.5625	0.0907	0.5532	0.7439	1.3971	2.0503	2.7035	3.3567	4.0099	4.6631	5.3163	5.9695	6.6227	7.2759	7.9291
n	p_0	l	$APIC1$	$APIC2$ (AIC)	$APIC3$ (KIC)	$APIC$ $\log(n)$ (BIC)	$APIC4$	$APIC5$	$APIC6$	$APIC7$	$APIC8$	$APIC9$	$APIC10$	$APIC11$	$APIC12$	$APIC13$	$APIC14$
30	3	1	-0.1132	0.5340	1.1812	1.4409	1.8284	2.4756	3.1229	3.7701	4.4173	5.0645	5.7117	6.3589	7.0062	7.6534	8.3006
30	3	2	-0.1785	0.7190	1.6166	1.9767	2.5141	3.4116	4.3092	5.2067	6.1042	7.0017	7.8993	8.7968	9.6943	10.5918	11.4894
30	3	3	-0.2414	0.8356	1.9127	2.3448	2.9897	4.0667	5.1438	6.2208	7.2978	8.3749	9.4519	10.5289	11.6060	12.6830	13.7600
30	3	4	-0.3054	0.9120	2.1295	2.6179	3.3470	4.5644	5.7819	6.9994	8.2168	9.4343	10.6518	11.8692	13.0867	14.3041	15.5216
30	5	1	-0.1648	0.4352	1.0352	1.2759	1.6352	2.2352	2.8352	3.4352	4.0352	4.6352	5.2352	5.8352	6.4352	7.0352	7.6352
30	5	2	-0.2516	0.5791	1.4097	1.7430	2.2404	3.0710	3.9017	4.7324	5.5630	6.3937	7.2244	8.0550	8.8857	9.7163	10.5470
n	p_0	l	$APIC1$	$APIC2$ (AIC)	$APIC3$ (KIC)	$APIC4$	$APIC$ $\log(n)$ (BIC)	$APIC5$	$APIC6$	$APIC7$	$APIC8$	$APIC9$	$APIC10$	$APIC11$	$APIC12$	$APIC13$	$APIC14$
100	3	1	-0.0324	0.6569	1.3463	2.0356	2.4528	2.7250	3.4143	4.1037	4.7930	5.4824	6.1717	6.8611	7.5504	8.2398	8.9291
100	3	2	-0.0510	0.9188	1.8886	2.8584	3.4453	3.8282	4.7980	5.7678	6.7376	7.7074	8.6772	9.6470	10.6168	11.5866	12.5564
100	3	3	-0.0687	1.1128	2.2942	3.4757	4.1907	4.6572	5.8387	7.0202	8.2016	9.3831	10.5646	11.7461	12.9276	14.1091	15.2905
100	3	4	-0.0867	1.2703	2.6273	3.9843	4.8055	5.3413	6.6982	8.0552	9.4122	10.7692	12.1262	13.4831	14.8401	16.1971	17.5541
100	5	1	-0.0469	0.6283	1.3035	1.9787	2.3874	2.6539	3.3292	4.0044	4.6796	5.3548	6.0300	6.7052	7.3804	8.0556	8.7308
100	5	2	-0.0714	0.8784	1.8282	2.7780	3.3527	3.7277	4.6775	5.6273	6.5771	7.5269	8.4767	9.4265	10.3763	11.3261	12.2758

Similarly, $APIC$ in (3.1) has the probability of overfitting as shown in the equation (2.24),

$$P\{APIC_{p_0+l} < APIC_{p_0}\} = P\left\{F_{l, n-p_0-l} > \frac{n-p_0-l}{l} \left[\exp\left(\frac{\alpha l}{n}\right) - 1 \right] \right\}. \quad (3.3)$$

In the opposite, the probability of underfitting is defined based on a model with too few variables compared to the optimal model (Seghouane, 2006). It is defined by

$$\begin{aligned} P\{APIC_{p_0-l} < APIC_{p_0}\} &= P\left\{\log(\hat{\sigma}_{p_0-l}^2) + \frac{\alpha(p_0-l+1)}{n} < \log(\hat{\sigma}_{p_0}^2) + \frac{\alpha(p_0+1)}{n}\right\} \\ &= P\left\{\log\left(\frac{\hat{\sigma}_{p_0-l}^2}{\hat{\sigma}_{p_0}^2}\right) < \frac{\alpha l}{n}\right\} = P\left\{\frac{\hat{\sigma}_{p_0-l}^2}{\hat{\sigma}_{p_0}^2} < \exp\left(\frac{\alpha l}{n}\right)\right\} = P\left\{\frac{\hat{\sigma}_{p_0-l}^2 - \hat{\sigma}_{p_0}^2}{\hat{\sigma}_{p_0}^2} < \exp\left(\frac{\alpha l}{n}\right) - 1\right\} \\ &= P\left\{\frac{\chi_l^2}{\chi_{n-p_0}^2} < \exp\left(\frac{\alpha l}{n}\right) - 1\right\} = P\left\{F_{l, n-p_0} < \frac{n-p_0}{l} \left[\exp\left(\frac{\alpha l}{n}\right) - 1 \right] \right\}. \end{aligned} \quad (3.4)$$

In (3.3) and (3.4), we found that $APIC$'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 3.2 and 3.3. 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:

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

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

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

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.

Table 3.2 Probability of overfitting by l extra variables of $APIC\alpha$ for different values of n , p_0 and l .

n	p_0	l	<i>Criteria</i>														
			$APIC1$	$APIC2$ (AIC)	$APIC$ $\log(n)$ (BIC)	$APIC3$ (KIC)	$APIC4$	$APIC5$	$APIC6$	$APIC7$	$APIC8$	$APIC9$	$APIC10$	$APIC11$	$APIC12$	$APIC13$	$APIC14$
15	3	1	0.4025	0.2363	0.1682	0.1469	0.0939	0.0611	0.0402	0.0266	0.0178	0.0119	0.0080	0.0054	0.0037	0.0025	0.0017
15	3	2	0.5134	0.2636	0.1644	0.1353	0.0695	0.0357	0.0183	0.0094	0.0048	0.0025	0.0013	0.0007	0.0003	0.0002	0.0001
15	3	3	0.5947	0.2857	0.1631	0.1287	0.0561	0.0240	0.0101	0.0042	0.0018	0.0007	0.0003	0.0001	0.0001	0.0000	0.0000
15	3	4	0.6664	0.3143	0.1701	0.1305	0.0508	0.0190	0.0070	0.0025	0.0009	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000
15	5	1	0.4511	0.2865	0.2148	0.1917	0.1316	0.0918	0.0647	0.0460	0.0329	0.0236	0.0170	0.0123	0.0089	0.0065	0.0047
15	5	2	0.5866	0.3442	0.2359	0.2019	0.1184	0.0695	0.0408	0.0239	0.0140	0.0082	0.0048	0.0028	0.0017	0.0010	0.0006
n	p_0	l	$APIC1$	$APIC2$ (AIC)	$APIC3$ (KIC)	$APIC$ $\log(n)$ (BIC)	$APIC4$	$APIC5$	$APIC6$	$APIC7$	$APIC8$	$APIC9$	$APIC10$	$APIC11$	$APIC12$	$APIC13$	$APIC14$
30	3	1	0.3565	0.1922	0.1102	0.0890	0.0651	0.0392	0.0239	0.0147	0.0091	0.0057	0.0035	0.0022	0.0014	0.0009	0.0006
30	3	2	0.4346	0.1889	0.0821	0.0588	0.0357	0.0155	0.0067	0.0029	0.0013	0.0006	0.0002	0.0001	0.0000	0.0000	0.0000
30	3	3	0.4846	0.1795	0.0617	0.0397	0.0204	0.0066	0.0021	0.0007	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
30	3	4	0.5256	0.1720	0.0482	0.0282	0.0125	0.0031	0.0007	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
30	5	1	0.3761	0.2106	0.1252	0.1026	0.0766	0.0478	0.0301	0.0192	0.0123	0.0079	0.0051	0.0033	0.0022	0.0014	0.0009
30	5	2	0.4646	0.2158	0.1003	0.0737	0.0466	0.0216	0.0101	0.0047	0.0022	0.0010	0.0005	0.0002	0.0001	0.0000	0.0000
n	p_0	l	$APIC1$	$APIC2$ (AIC)	$APIC3$ (KIC)	$APIC4$	$APIC$ $\log(n)$ (BIC)	$APIC5$	$APIC6$	$APIC7$	$APIC8$	$APIC9$	$APIC10$	$APIC11$	$APIC12$	$APIC13$	$APIC14$
100	3	1	0.3284	0.1670	0.0905	0.0506	0.0360	0.0289	0.0167	0.0097	0.0057	0.0034	0.0020	0.0012	0.0007	0.0004	0.0003
100	3	2	0.3867	0.1496	0.0578	0.0224	0.0126	0.0087	0.0033	0.0013	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
100	3	3	0.4178	0.1288	0.0367	0.0100	0.0045	0.0027	0.0007	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	3	4	0.4395	0.1109	0.0236	0.0046	0.0017	0.0009	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	5	1	0.3336	0.1715	0.0940	0.0531	0.0380	0.0306	0.0179	0.0105	0.0062	0.0037	0.0022	0.0013	0.0008	0.0005	0.0003
100	5	2	0.3946	0.1557	0.0614	0.0242	0.0138	0.0096	0.0038	0.0015	0.0006	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000

Table 3.3 Probability of underfitting by l variables of $APIC\alpha$ for different values of n , p_0 and l .

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

From Table 3.2 and 3.3 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 the equation (3.1) 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, we can conclude that, *APIC* with a much more value of α , make its probability of overfitting to be smaller but make more probability of underfitting. As a result, the main objective of this research 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. The proposed criterion, *APIC*, tries to overcome the weak signal-to-noise ratio and to reduce the probability of over/underfitting in order to select the most correct model for univariate regression. Then, the performance of *APIC* is examined relative to the well-known criteria, *AIC*, *BIC*, and *KIC*, under various circumstances such as differences in sample sizes, the orders of true model, the regression coefficients, the variances of error terms, and the distributions of independent variables (McQuarrie, Shumway, and Tsai, 1997; McQuarrie, 1999; Mills and Prasad, 2007; Rahman and King, 2007).

CHAPTER 4

SIMULATION STUDY

In addition to the proofs of signal-to-noise ratio in (3.2) and the probability of over/underfitting in (3.3) and (3.4), we use the simulation data to find the appropriate value of α for *APIC* in (3.1). True multiple regression models in (2.1) are constructed as follows.

Model 1 (very weakly identifiable true model with the true order $p_0 = 5$):

$$y_1 = 1 + 0.5X_2 + 0.4X_3 + 0.3X_4 + 0.2X_5 + \varepsilon_1,$$

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

$$y_2 = 1 + 0.5X_2 + 0.4X_3 + \varepsilon_2,$$

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

$$y_3 = 1 + 2X_2 + 2X_3 + \varepsilon_3,$$

Model 4 (very strongly identifiable true model with the true order $p_0 = 5$):

$$y_4 = 1 + 2X_2 + 2X_3 + 2X_4 + 2X_5 + \varepsilon_4.$$

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 (2.1) 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 (2.2). 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 (3.1) 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. Results are shown in Table 4.1 to Table 4.3. Findings are the following.

[1] In Table 4.1, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[2] In Table 4.1, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[3] In Table 4.1, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[4] In Table 4.1, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

Table 4.1 Probability of the order selected by *APIC* for $n = 15$.

Model	Dist. of X	σ_0^2	Order	Criteria														
				<i>APIC1</i>	<i>APIC2</i> (AIC)	<i>APIClog(n)</i> (BIC)	<i>APIC3</i> (KIC)	<i>APIC4</i>	<i>APIC5</i>	<i>APIC6</i>	<i>APIC7</i>	<i>APIC8</i>	<i>APIC9</i>	<i>APIC10</i>	<i>APIC11</i>	<i>APIC12</i>	<i>APIC13</i>	<i>APIC14</i>
1 very weakly $p_0 = 5$	Normal	0.25	Underfitted	0.157	0.353	0.472	0.53	0.669	0.785	0.868	0.93	0.958	0.974	0.983	0.99	0.994	0.998	1
			Correct	0.229	0.297	0.293	0.281	0.232	0.165	0.107	0.061	0.038	0.025	0.016	0.009	0.005	0.002	0
			Overfitted	0.614	0.35	0.235	0.189	0.099	0.05	0.025	0.009	0.004	0.001	0.001	0.001	0.001	0	0
		1	Underfitted	0.269	0.562	0.731	0.79	0.9	0.951	0.978	0.99	0.998	0.999	0.999	0.999	1	1	1
			Correct	0.155	0.135	0.104	0.085	0.05	0.025	0.011	0.006	0.001	0	0	0	0	0	0
			Overfitted	0.576	0.303	0.165	0.125	0.05	0.024	0.011	0.004	0.001	0.001	0.001	0.001	0	0	0
		9	Underfitted	0.367	0.69	0.826	0.863	0.946	0.98	0.993	0.998	1	1	1	1	1	1	1
			Correct	0.119	0.097	0.064	0.054	0.03	0.011	0.005	0.001	0	0	0	0	0	0	0
			Overfitted	0.514	0.213	0.11	0.083	0.024	0.009	0.002	0.001	0	0	0	0	0	0	0
	Uniform	0.25	Underfitted	0.346	0.679	0.82	0.855	0.937	0.971	0.99	0.998	1	1	1	1	1	1	1
			Correct	0.132	0.093	0.066	0.053	0.027	0.014	0.004	0.001	0	0	0	0	0	0	0
			Overfitted	0.522	0.228	0.114	0.092	0.036	0.015	0.006	0.001	0	0	0	0	0	0	0
		1	Underfitted	0.365	0.691	0.838	0.878	0.947	0.98	0.997	0.998	0.999	0.999	1	1	1	1	1
			Correct	0.113	0.079	0.052	0.039	0.015	0.007	0.001	0.001	0.001	0	0	0	0	0	0
			Overfitted	0.522	0.23	0.11	0.083	0.038	0.013	0.002	0.001	0	0	0	0	0	0	0
		9	Underfitted	0.359	0.695	0.836	0.868	0.946	0.985	0.995	0.997	0.999	1	1	1	1	1	1
			Correct	0.106	0.074	0.041	0.034	0.018	0.007	0.004	0.003	0.001	0	0	0	0	0	0
			Overfitted	0.535	0.231	0.123	0.098	0.036	0.008	0.001	0	0	0	0	0	0	0	0
2 weakly $p_0 = 3$	Normal	0.25	Underfitted	0.02	0.057	0.112	0.124	0.195	0.274	0.366	0.44	0.522	0.59	0.654	0.722	0.762	0.804	0.844
			Correct	0.215	0.465	0.587	0.614	0.658	0.639	0.578	0.528	0.458	0.4	0.34	0.276	0.236	0.194	0.156
			Overfitted	0.765	0.478	0.301	0.262	0.147	0.087	0.056	0.032	0.02	0.01	0.006	0.002	0.002	0.002	0
		1	Underfitted	0.091	0.282	0.416	0.464	0.604	0.704	0.779	0.842	0.882	0.915	0.942	0.965	0.97	0.981	0.987
			Correct	0.147	0.301	0.333	0.33	0.286	0.237	0.191	0.143	0.11	0.082	0.057	0.034	0.03	0.019	0.013
			Overfitted	0.762	0.417	0.251	0.206	0.11	0.059	0.03	0.015	0.008	0.003	0.001	0.001	0	0	0
		9	Underfitted	0.181	0.52	0.693	0.739	0.85	0.899	0.943	0.961	0.974	0.98	0.984	0.99	0.992	0.993	0.998
			Correct	0.095	0.119	0.104	0.101	0.081	0.065	0.046	0.032	0.023	0.018	0.016	0.01	0.008	0.007	0.002
			Overfitted	0.724	0.361	0.203	0.16	0.069	0.036	0.011	0.007	0.003	0.002	0	0	0	0	0
	Uniform	0.25	Underfitted	0.152	0.425	0.583	0.644	0.768	0.855	0.896	0.93	0.96	0.972	0.982	0.991	0.995	0.996	0.997
			Correct	0.111	0.178	0.173	0.17	0.149	0.11	0.084	0.059	0.036	0.026	0.018	0.009	0.005	0.004	0.003
			Overfitted	0.737	0.397	0.244	0.186	0.083	0.035	0.02	0.011	0.004	0.002	0	0	0	0	0
		1	Underfitted	0.179	0.478	0.635	0.695	0.841	0.915	0.941	0.961	0.976	0.985	0.991	0.995	0.997	0.999	1
			Correct	0.103	0.126	0.119	0.108	0.079	0.054	0.04	0.031	0.02	0.013	0.009	0.005	0.003	0.001	0
			Overfitted	0.718	0.396	0.246	0.197	0.08	0.031	0.019	0.008	0.004	0.002	0	0	0	0	0
		9	Underfitted	0.189	0.491	0.658	0.717	0.844	0.925	0.957	0.977	0.985	0.993	0.996	0.997	0.998	0.998	0.998
			Correct	0.094	0.136	0.132	0.115	0.075	0.04	0.028	0.014	0.009	0.006	0.003	0.003	0.002	0.002	0.002
			Overfitted	0.717	0.373	0.21	0.168	0.081	0.035	0.015	0.009	0.006	0.001	0.001	0	0	0	0

Note: Boldface type indicates the maximum probability of correct order being selected.

Table 4.1 (Continued).

Model	Dist. of X	σ_0^2	Order	Criteria														
				APIC1	APIC2 (AIC)	APIClog(n) (BIC)	APIC3 (KIC)	APIC4	APIC5	APIC6	APIC7	APIC8	APIC9	APIC10	APIC11	APIC12	APIC13	APIC14
3 strongly $p_0 = 3$	Normal	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.22	0.515	0.654	0.707	0.814	0.891	0.942	0.962	0.971	0.983	0.991	0.992	0.995	0.997	0.998
			Overfitted	0.78	0.485	0.346	0.293	0.186	0.109	0.058	0.038	0.029	0.017	0.009	0.008	0.005	0.003	0.002
		1	Underfitted	0	0	0	0	0	0.001	0.001	0.004	0.005	0.01	0.014	0.019	0.029	0.055	0.074
			Correct	0.238	0.542	0.683	0.725	0.835	0.902	0.946	0.961	0.975	0.977	0.977	0.977	0.968	0.943	0.924
			Overfitted	0.762	0.458	0.317	0.275	0.165	0.097	0.053	0.035	0.02	0.013	0.009	0.004	0.003	0.002	0.002
		9	Underfitted	0.025	0.118	0.189	0.213	0.313	0.412	0.516	0.578	0.661	0.714	0.776	0.826	0.864	0.902	0.921
			Correct	0.197	0.427	0.518	0.541	0.554	0.514	0.447	0.399	0.323	0.275	0.22	0.171	0.133	0.097	0.078
			Overfitted	0.778	0.455	0.293	0.246	0.133	0.074	0.037	0.023	0.016	0.011	0.004	0.003	0.003	0.001	0.001
	Uniform	0.25	Underfitted	0.004	0.009	0.016	0.019	0.039	0.063	0.095	0.121	0.158	0.206	0.25	0.316	0.382	0.446	0.511
			Correct	0.215	0.52	0.654	0.703	0.799	0.846	0.858	0.854	0.823	0.784	0.745	0.68	0.616	0.552	0.487
			Overfitted	0.781	0.471	0.33	0.278	0.162	0.091	0.047	0.025	0.019	0.01	0.005	0.004	0.002	0.002	0.002
		1	Underfitted	0.041	0.154	0.237	0.272	0.377	0.473	0.587	0.661	0.713	0.779	0.828	0.87	0.903	0.917	0.942
			Correct	0.198	0.389	0.461	0.476	0.485	0.467	0.387	0.322	0.276	0.215	0.169	0.128	0.097	0.083	0.058
			Overfitted	0.761	0.457	0.302	0.252	0.138	0.06	0.026	0.017	0.011	0.006	0.003	0.002	0	0	0
		9	Underfitted	0.153	0.45	0.611	0.671	0.797	0.876	0.924	0.953	0.972	0.979	0.986	0.993	0.999	0.999	1
			Correct	0.112	0.154	0.157	0.149	0.121	0.083	0.054	0.035	0.022	0.02	0.014	0.007	0.001	0.001	0
			Overfitted	0.735	0.396	0.232	0.18	0.082	0.041	0.022	0.012	0.006	0.001	0	0	0	0	0
4 very strongly $p_0 = 5$	Normal	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0.001	0.004	0.007	0.016
			Correct	0.342	0.55	0.659	0.686	0.804	0.874	0.914	0.94	0.958	0.97	0.978	0.983	0.984	0.985	0.977
			Overfitted	0.658	0.45	0.341	0.314	0.196	0.126	0.086	0.06	0.042	0.03	0.022	0.016	0.012	0.008	0.007
		1	Underfitted	0	0.001	0.001	0.001	0.003	0.005	0.011	0.019	0.036	0.061	0.108	0.169	0.248	0.348	0.475
			Correct	0.309	0.546	0.669	0.698	0.797	0.851	0.894	0.916	0.913	0.902	0.866	0.813	0.738	0.642	0.518
			Overfitted	0.691	0.453	0.33	0.301	0.2	0.144	0.095	0.065	0.051	0.037	0.026	0.018	0.014	0.01	0.007
		9	Underfitted	0.052	0.167	0.252	0.292	0.437	0.589	0.728	0.819	0.885	0.934	0.955	0.976	0.989	0.995	0.997
			Correct	0.304	0.448	0.466	0.462	0.418	0.328	0.234	0.158	0.103	0.061	0.043	0.023	0.011	0.005	0.003
			Overfitted	0.644	0.385	0.282	0.246	0.145	0.083	0.038	0.023	0.012	0.005	0.002	0.001	0	0	0
	Uniform	0.25	Underfitted	0.004	0.016	0.026	0.032	0.057	0.109	0.175	0.274	0.414	0.545	0.669	0.761	0.853	0.923	0.959
			Correct	0.344	0.569	0.656	0.69	0.764	0.782	0.76	0.684	0.556	0.438	0.323	0.234	0.145	0.077	0.041
			Overfitted	0.652	0.415	0.318	0.278	0.179	0.109	0.065	0.042	0.03	0.017	0.008	0.005	0.002	0	0
		1	Underfitted	0.078	0.209	0.314	0.363	0.535	0.676	0.798	0.878	0.938	0.966	0.983	0.993	0.998	1	1
			Correct	0.298	0.399	0.423	0.415	0.345	0.252	0.162	0.1	0.056	0.031	0.015	0.006	0.002	0	0
			Overfitted	0.624	0.392	0.263	0.222	0.12	0.072	0.04	0.022	0.006	0.003	0.002	0.001	0	0	0
		9	Underfitted	0.278	0.593	0.755	0.801	0.911	0.965	0.987	0.995	0.998	0.999	0.999	0.999	1	1	1
			Correct	0.148	0.127	0.089	0.074	0.043	0.017	0.007	0.004	0.002	0.001	0.001	0.001	0	0	0
			Overfitted	0.574	0.28	0.156	0.125	0.046	0.018	0.006	0.001	0	0	0	0	0	0	0

Note: Boldface type indicates the maximum probability of correct order being selected.

[5] In Table 4.2, for very weakly identifiable situation of true models with the true orders $p_0 = 5$, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[6] In Table 4.2, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[7] In Table 4.2, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[8] In Table 4.2, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

Table 4.2 Probability of the order selected by *APIC* for $n = 30$.

Model	Dist. of X	σ_0^2	Order	Criteria														
				<i>APIC1</i>	<i>APIC2</i> (<i>AIC</i>)	<i>APIC3</i> (<i>KIC</i>)	<i>APIClog(n)</i> (<i>BIC</i>)	<i>APIC4</i>	<i>APIC5</i>	<i>APIC6</i>	<i>APIC7</i>	<i>APIC8</i>	<i>APIC9</i>	<i>APIC10</i>	<i>APIC11</i>	<i>APIC12</i>	<i>APIC13</i>	<i>APIC14</i>
1 very weakly $p_0 = 5$	Normal	0.25	Underfitted	0.076	0.199	0.312	0.366	0.446	0.547	0.656	0.745	0.827	0.877	0.927	0.958	0.973	0.985	0.99
			Correct	0.406	0.532	0.55	0.522	0.481	0.41	0.328	0.249	0.171	0.122	0.073	0.042	0.027	0.015	0.01
			Overfitted	0.518	0.269	0.138	0.112	0.073	0.043	0.016	0.006	0.002	0.001	0	0	0	0	0
		1	Underfitted	0.275	0.55	0.761	0.799	0.855	0.937	0.976	0.985	0.995	0.999	1	1	1	1	1
			Correct	0.246	0.237	0.161	0.146	0.108	0.054	0.023	0.015	0.005	0.001	0	0	0	0	0
			Overfitted	0.479	0.213	0.078	0.055	0.037	0.009	0.001	0	0	0	0	0	0	0	0
		9	Underfitted	0.48	0.804	0.915	0.94	0.974	0.993	0.999	1	1	1	1	1	1	1	1
			Correct	0.135	0.075	0.045	0.032	0.017	0.005	0.001	0	0	0	0	0	0	0	0
			Overfitted	0.385	0.121	0.04	0.028	0.009	0.002	0	0	0	0	0	0	0	0	0
	Uniform	0.25	Underfitted	0.392	0.696	0.871	0.905	0.952	0.982	0.998	0.999	1	1	1	1	1	1	1
			Correct	0.175	0.137	0.071	0.056	0.031	0.012	0.002	0.001	0	0	0	0	0	0	0
			Overfitted	0.433	0.167	0.058	0.039	0.017	0.006	0	0	0	0	0	0	0	0	0
		1	Underfitted	0.48	0.828	0.96	0.976	0.985	0.997	0.999	1	1	1	1	1	1	1	1
			Correct	0.13	0.074	0.023	0.016	0.011	0.003	0.001	0	0	0	0	0	0	0	0
			Overfitted	0.39	0.098	0.017	0.008	0.004	0	0	0	0	0	0	0	0	0	0
		9	Underfitted	0.48	0.818	0.95	0.971	0.988	0.996	0.999	1	1	1	1	1	1	1	1
			Correct	0.133	0.063	0.024	0.016	0.007	0.002	0	0	0	0	0	0	0	0	0
			Overfitted	0.387	0.119	0.026	0.013	0.005	0.002	0.001	0	0	0	0	0	0	0	0
2 weakly $p_0 = 3$	Normal	0.25	Underfitted	0.001	0.003	0.009	0.018	0.025	0.04	0.066	0.093	0.135	0.171	0.224	0.28	0.338	0.392	0.45
			Correct	0.321	0.632	0.802	0.837	0.884	0.908	0.906	0.892	0.857	0.824	0.772	0.719	0.661	0.607	0.55
			Overfitted	0.678	0.365	0.189	0.145	0.091	0.052	0.028	0.015	0.008	0.005	0.004	0.001	0.001	0.001	0
		1	Underfitted	0.057	0.173	0.305	0.346	0.413	0.514	0.606	0.661	0.742	0.804	0.85	0.875	0.896	0.919	0.935
			Correct	0.296	0.514	0.555	0.55	0.522	0.452	0.378	0.328	0.252	0.194	0.15	0.125	0.104	0.081	0.065
			Overfitted	0.647	0.313	0.14	0.104	0.065	0.034	0.016	0.011	0.006	0.002	0	0	0	0	0
		9	Underfitted	0.236	0.577	0.756	0.792	0.84	0.907	0.938	0.959	0.974	0.982	0.986	0.993	0.995	0.998	0.999
			Correct	0.147	0.185	0.154	0.145	0.118	0.077	0.055	0.04	0.026	0.018	0.014	0.007	0.005	0.002	0.001
			Overfitted	0.617	0.238	0.09	0.063	0.042	0.016	0.007	0.001	0	0	0	0	0	0	0
	Uniform	0.25	Underfitted	0.174	0.437	0.616	0.667	0.734	0.811	0.871	0.904	0.937	0.953	0.972	0.98	0.985	0.991	0.996
			Correct	0.21	0.292	0.276	0.254	0.217	0.166	0.115	0.088	0.06	0.045	0.027	0.02	0.015	0.009	0.004
			Overfitted	0.616	0.271	0.108	0.079	0.049	0.023	0.014	0.008	0.003	0.002	0.001	0	0	0	0
		1	Underfitted	0.257	0.599	0.776	0.826	0.867	0.925	0.952	0.972	0.982	0.988	0.991	0.996	0.999	0.999	1
			Correct	0.127	0.166	0.14	0.121	0.103	0.063	0.046	0.028	0.018	0.012	0.009	0.004	0.001	0.001	0
			Overfitted	0.616	0.235	0.084	0.053	0.03	0.012	0.002	0	0	0	0	0	0	0	0
		9	Underfitted	0.317	0.655	0.83	0.875	0.913	0.953	0.978	0.989	0.994	0.996	0.997	0.999	0.999	0.999	1
			Correct	0.107	0.118	0.079	0.069	0.058	0.036	0.019	0.009	0.004	0.003	0.003	0.001	0.001	0.001	0
			Overfitted	0.576	0.227	0.091	0.056	0.029	0.011	0.003	0.002	0.002	0.001	0	0	0	0	0

Note: Boldface type indicates the maximum probability of correct order being selected.

Table 4.2 (Continued).

Model	Dist. of X	σ_0^2	Order	Criteria														
				APIC1	APIC2 (AIC)	APIC3 (KIC)	APIClog(n) (BIC)	APIC4	APIC5	APIC6	APIC7	APIC8	APIC9	APIC10	APIC11	APIC12	APIC13	APIC14
3 strongly $p_0 = 3$	Normal	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.317	0.634	0.824	0.861	0.906	0.942	0.972	0.977	0.986	0.99	0.996	0.998	0.998	0.998	1
			Overfitted	0.683	0.366	0.176	0.139	0.094	0.058	0.028	0.023	0.014	0.01	0.004	0.002	0.002	0.002	0
		1	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.348	0.664	0.832	0.874	0.91	0.96	0.979	0.99	0.995	0.998	0.998	0.999	0.999	0.999	0.999
			Overfitted	0.652	0.336	0.168	0.126	0.09	0.04	0.021	0.01	0.005	0.002	0.002	0.001	0.001	0.001	0.001
		9	Underfitted	0.003	0.021	0.038	0.044	0.06	0.091	0.137	0.185	0.256	0.304	0.37	0.437	0.485	0.544	0.593
			Correct	0.316	0.62	0.765	0.807	0.842	0.855	0.832	0.796	0.735	0.691	0.627	0.56	0.513	0.454	0.406
			Overfitted	0.681	0.359	0.197	0.149	0.098	0.054	0.031	0.019	0.009	0.005	0.003	0.003	0.002	0.002	0.001
	Uniform	0.25	Underfitted	0	0	0	0	0	0	0	0.001	0.003	0.006	0.007	0.012	0.021	0.03	0.038
			Correct	0.308	0.625	0.813	0.862	0.908	0.948	0.968	0.976	0.984	0.985	0.988	0.986	0.978	0.969	0.961
			Overfitted	0.692	0.375	0.187	0.138	0.092	0.052	0.032	0.023	0.013	0.009	0.005	0.002	0.001	0.001	0.001
		1	Underfitted	0.009	0.044	0.095	0.122	0.155	0.23	0.298	0.374	0.44	0.497	0.566	0.627	0.686	0.734	0.782
			Correct	0.331	0.62	0.739	0.756	0.759	0.735	0.68	0.611	0.549	0.496	0.43	0.37	0.312	0.264	0.216
			Overfitted	0.66	0.336	0.166	0.122	0.086	0.035	0.022	0.015	0.011	0.007	0.004	0.003	0.002	0.002	0.002
		9	Underfitted	0.189	0.504	0.688	0.74	0.797	0.864	0.905	0.934	0.957	0.974	0.982	0.987	0.99	0.992	0.994
			Correct	0.196	0.232	0.209	0.198	0.167	0.119	0.087	0.063	0.042	0.026	0.018	0.013	0.01	0.008	0.006
			Overfitted	0.615	0.264	0.103	0.062	0.036	0.017	0.008	0.003	0.001	0	0	0	0	0	0
4 very strongly $p_0 = 5$	Normal	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.481	0.713	0.831	0.867	0.903	0.945	0.969	0.979	0.985	0.99	0.993	0.994	0.996	0.998	1
			Overfitted	0.519	0.287	0.169	0.133	0.097	0.055	0.031	0.021	0.015	0.01	0.007	0.006	0.004	0.002	0
		1	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.454	0.705	0.845	0.881	0.908	0.95	0.971	0.985	0.992	0.996	0.998	0.998	0.998	0.998	1
			Overfitted	0.546	0.295	0.155	0.119	0.092	0.05	0.029	0.015	0.008	0.004	0.002	0.002	0.002	0.002	0
		9	Underfitted	0.009	0.029	0.064	0.074	0.102	0.172	0.26	0.354	0.457	0.578	0.708	0.801	0.86	0.907	0.947
			Correct	0.447	0.675	0.765	0.793	0.797	0.767	0.703	0.623	0.527	0.413	0.287	0.198	0.139	0.093	0.053
			Overfitted	0.544	0.296	0.171	0.133	0.101	0.061	0.037	0.023	0.016	0.009	0.005	0.001	0.001	0	0
	Uniform	0.25	Underfitted	0	0	0	0	0	0	0	0.001	0.002	0.011	0.024	0.04	0.066	0.12	0.204
			Correct	0.419	0.699	0.829	0.869	0.901	0.946	0.971	0.982	0.986	0.981	0.971	0.958	0.934	0.88	0.796
			Overfitted	0.581	0.301	0.171	0.131	0.099	0.054	0.029	0.017	0.012	0.008	0.005	0.002	0	0	0
		1	Underfitted	0.018	0.053	0.116	0.154	0.214	0.317	0.427	0.544	0.664	0.763	0.844	0.894	0.947	0.967	0.983
			Correct	0.445	0.658	0.723	0.719	0.703	0.635	0.554	0.448	0.329	0.234	0.154	0.106	0.053	0.033	0.017
			Overfitted	0.537	0.289	0.161	0.127	0.083	0.048	0.019	0.008	0.007	0.003	0.002	0	0	0	0
		9	Underfitted	0.323	0.646	0.826	0.87	0.925	0.969	0.992	0.998	0.998	0.999	0.999	1	1	1	1
			Correct	0.223	0.175	0.115	0.092	0.055	0.024	0.006	0.001	0.001	0.001	0.001	0	0	0	0
			Overfitted	0.454	0.179	0.059	0.038	0.02	0.007	0.002	0.001	0.001	0	0	0	0	0	0

Note: Boldface type indicates the maximum probability of correct order being selected.

[9] In Table 4.3, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[10] In Table 4.3, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[11] In Table 4.3, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[12] In Table 4.3, for very strongly identifiable situation of true models with the true orders $p_0 = 5$, Model 4, the sample size is moderate ($n = 100$) and the distribution of independent variable is normal, the appropriate values of α when the true variances $\sigma_0^2 = 0.25, 1, 9$, 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%.

Table 4.3 Probability of the order selected by *APIC* for $n = 100$.

Model	Dist. of X	σ_0^2	Order	Criteria														
				<i>APIC1</i>	<i>APIC2</i> (<i>AIC</i>)	<i>APIC3</i> (<i>KIC</i>)	<i>APIC4</i>	<i>APIClog(n)</i> (<i>BIC</i>)	<i>APIC5</i>	<i>APIC6</i>	<i>APIC7</i>	<i>APIC8</i>	<i>APIC9</i>	<i>APIC10</i>	<i>APIC11</i>	<i>APIC12</i>	<i>APIC13</i>	<i>APIC14</i>
1 very weakly $p_0 = 5$	Normal	0.25	Underfitted	0	0.005	0.016	0.031	0.045	0.05	0.078	0.107	0.144	0.199	0.246	0.293	0.341	0.394	0.458
			Correct	0.537	0.766	0.856	0.904	0.912	0.914	0.898	0.88	0.849	0.798	0.752	0.706	0.658	0.605	0.542
			Overfitted	0.463	0.229	0.128	0.065	0.043	0.036	0.024	0.013	0.007	0.003	0.002	0.001	0.001	0.001	0
		1	Underfitted	0.115	0.259	0.401	0.505	0.566	0.621	0.704	0.783	0.847	0.893	0.932	0.961	0.977	0.985	0.991
			Correct	0.434	0.535	0.502	0.444	0.394	0.351	0.289	0.216	0.153	0.107	0.068	0.039	0.023	0.015	0.009
			Overfitted	0.451	0.206	0.097	0.051	0.04	0.028	0.007	0.001	0	0	0	0	0	0	0
		9	Underfitted	0.489	0.783	0.916	0.97	0.986	0.988	0.996	0.997	0.999	0.999	0.999	1	1	1	1
			Correct	0.174	0.128	0.061	0.027	0.012	0.01	0.004	0.003	0.001	0.001	0	0	0	0	0
			Overfitted	0.337	0.089	0.023	0.003	0.002	0.002	0	0	0	0	0	0	0	0	0
	Uniform	0.25	Underfitted	0.276	0.575	0.763	0.87	0.905	0.922	0.954	0.972	0.985	0.996	0.998	0.999	0.999	1	1
			Correct	0.313	0.273	0.201	0.121	0.089	0.074	0.046	0.028	0.015	0.004	0.002	0.001	0.001	0	0
			Overfitted	0.411	0.152	0.036	0.009	0.006	0.004	0	0	0	0	0	0	0	0	0
		1	Underfitted	0.485	0.808	0.933	0.979	0.986	0.991	0.996	0.999	1	1	1	1	1	1	1
			Correct	0.18	0.109	0.055	0.017	0.011	0.008	0.004	0.001	0	0	0	0	0	0	0
			Overfitted	0.335	0.083	0.012	0.004	0.003	0.001	0	0	0	0	0	0	0	0	0
		9	Underfitted	0.576	0.883	0.958	0.988	0.993	0.998	0.999	1	1	1	1	1	1	1	1
			Correct	0.117	0.053	0.024	0.009	0.006	0.002	0.001	0	0	0	0	0	0	0	0
			Overfitted	0.307	0.064	0.018	0.003	0.001	0	0	0	0	0	0	0	0	0	0
2 weakly $p_0 = 3$	Normal	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.393	0.712	0.873	0.939	0.958	0.966	0.981	0.99	0.996	0.998	0.999	0.999	1	1	1
			Overfitted	0.607	0.288	0.127	0.061	0.042	0.034	0.019	0.01	0.004	0.002	0.001	0.001	0	0	0
		1	Underfitted	0.001	0.002	0.011	0.029	0.036	0.041	0.063	0.092	0.125	0.162	0.212	0.271	0.338	0.381	0.445
			Correct	0.377	0.719	0.856	0.905	0.92	0.929	0.918	0.894	0.866	0.832	0.783	0.727	0.661	0.618	0.554
			Overfitted	0.622	0.279	0.133	0.066	0.044	0.03	0.019	0.014	0.009	0.006	0.005	0.002	0.001	0.001	0.001
		9	Underfitted	0.193	0.438	0.608	0.726	0.773	0.799	0.852	0.887	0.924	0.95	0.967	0.982	0.988	0.991	0.993
			Correct	0.24	0.339	0.304	0.236	0.205	0.183	0.139	0.109	0.075	0.049	0.033	0.018	0.012	0.009	0.007
			Overfitted	0.567	0.223	0.088	0.038	0.022	0.018	0.009	0.004	0.001	0.001	0	0	0	0	0
	Uniform	0.25	Underfitted	0.048	0.161	0.252	0.352	0.404	0.442	0.528	0.625	0.683	0.756	0.805	0.839	0.868	0.893	0.915
			Correct	0.351	0.586	0.635	0.599	0.564	0.532	0.456	0.37	0.314	0.244	0.195	0.161	0.132	0.107	0.085
			Overfitted	0.601	0.253	0.113	0.049	0.032	0.026	0.016	0.005	0.003	0	0	0	0	0	0
		1	Underfitted	0.21	0.511	0.668	0.786	0.83	0.841	0.892	0.926	0.947	0.967	0.98	0.986	0.993	0.997	0.999
			Correct	0.243	0.287	0.259	0.186	0.155	0.148	0.103	0.07	0.051	0.033	0.02	0.014	0.007	0.003	0.001
			Overfitted	0.547	0.202	0.073	0.028	0.015	0.011	0.005	0.004	0.002	0	0	0	0	0	0
		9	Underfitted	0.363	0.723	0.864	0.92	0.936	0.953	0.974	0.99	0.994	0.995	0.997	0.998	0.998	0.999	0.999
			Correct	0.125	0.111	0.079	0.063	0.058	0.042	0.025	0.01	0.006	0.005	0.003	0.002	0.002	0.001	0.001
			Overfitted	0.512	0.166	0.057	0.017	0.006	0.005	0.001	0	0	0	0	0	0	0	0

Note: Boldface type indicates the maximum probability of correct order being selected.

Table 4.3 (Continued).

Model	Dist. of X	σ_0^2	Order	Criteria														
				APIC1	APIC2 (AIC)	APIC3 (KIC)	APIC4	APIClog(n) (BIC)	APIC5	APIC6	APIC7	APIC8	APIC9	APIC10	APIC11	APIC12	APIC13	APIC14
3 strongly $p_0 = 3$	Normal	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.394	0.717	0.88	0.934	0.954	0.962	0.976	0.991	0.996	0.996	0.998	0.998	1	1	1
			Overfitted	0.606	0.283	0.12	0.066	0.046	0.038	0.024	0.009	0.004	0.004	0.002	0.002	0	0	0
		1	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.376	0.71	0.869	0.939	0.955	0.969	0.98	0.988	0.993	0.996	0.996	0.998	0.999	1	1
			Overfitted	0.624	0.29	0.131	0.061	0.045	0.031	0.02	0.012	0.007	0.004	0.004	0.002	0.001	0	0
		9	Underfitted	0	0	0	0	0	0	0	0	0	0	0.002	0.003	0.007	0.007	0.008
			Correct	0.365	0.704	0.855	0.933	0.95	0.959	0.98	0.985	0.989	0.993	0.993	0.993	0.991	0.992	0.991
			Overfitted	0.635	0.296	0.145	0.067	0.05	0.041	0.02	0.015	0.011	0.007	0.005	0.004	0.002	0.001	0.001
	Uniform	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.404	0.73	0.873	0.942	0.962	0.975	0.985	0.989	0.995	0.998	0.998	0.999	0.999	0.999	1
			Overfitted	0.596	0.27	0.127	0.058	0.038	0.025	0.015	0.011	0.005	0.002	0.002	0.001	0.001	0.001	0
		1	Underfitted	0	0	0	0	0	0.001	0.001	0.001	0.002	0.005	0.008	0.014	0.021	0.032	0.042
			Correct	0.391	0.721	0.867	0.929	0.949	0.961	0.98	0.99	0.993	0.992	0.989	0.985	0.978	0.967	0.958
			Overfitted	0.609	0.279	0.133	0.071	0.051	0.038	0.019	0.009	0.005	0.003	0.003	0.001	0.001	0.001	0
		9	Underfitted	0.09	0.237	0.386	0.515	0.567	0.604	0.699	0.757	0.813	0.853	0.882	0.916	0.936	0.961	0.969
			Correct	0.324	0.506	0.501	0.442	0.405	0.376	0.294	0.237	0.184	0.146	0.117	0.083	0.063	0.038	0.031
			Overfitted	0.586	0.257	0.113	0.043	0.028	0.02	0.007	0.006	0.003	0.001	0.001	0.001	0.001	0.001	0
4 very strongly $p_0 = 5$	Normal	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.516	0.753	0.858	0.922	0.938	0.951	0.971	0.981	0.991	0.992	0.996	0.998	0.998	0.998	0.999
			Overfitted	0.484	0.247	0.142	0.078	0.062	0.049	0.029	0.019	0.009	0.008	0.004	0.002	0.002	0.002	0.001
		1	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.52	0.776	0.886	0.94	0.965	0.971	0.981	0.986	0.995	0.996	0.999	0.999	0.999	0.999	0.999
			Overfitted	0.48	0.224	0.114	0.06	0.035	0.029	0.019	0.014	0.005	0.004	0.001	0.001	0.001	0.001	0.001
		9	Underfitted	0	0	0	0	0	0	0	0	0	0.001	0.002	0.004	0.005	0.008	0.012
			Correct	0.527	0.762	0.882	0.93	0.959	0.965	0.978	0.987	0.99	0.992	0.995	0.994	0.993	0.991	0.988
			Overfitted	0.473	0.238	0.118	0.07	0.041	0.035	0.022	0.013	0.01	0.007	0.003	0.002	0.002	0.001	0
	Uniform	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.546	0.797	0.902	0.944	0.957	0.965	0.981	0.986	0.99	0.993	0.998	0.999	0.999	1	1
			Overfitted	0.454	0.203	0.098	0.056	0.043	0.035	0.019	0.014	0.01	0.007	0.002	0.001	0.001	0	0
		1	Underfitted	0	0.001	0.001	0.001	0.002	0.002	0.002	0.005	0.006	0.006	0.01	0.017	0.024	0.037	0.048
			Correct	0.531	0.781	0.886	0.939	0.957	0.962	0.981	0.982	0.988	0.993	0.989	0.983	0.976	0.963	0.952
			Overfitted	0.469	0.218	0.113	0.06	0.041	0.036	0.017	0.013	0.006	0.001	0.001	0	0	0	0
		9	Underfitted	0.138	0.314	0.504	0.674	0.741	0.776	0.868	0.929	0.965	0.983	0.99	0.996	0.998	0.998	1
			Correct	0.406	0.492	0.417	0.294	0.239	0.212	0.127	0.07	0.035	0.017	0.01	0.004	0.002	0.002	0
			Overfitted	0.456	0.194	0.079	0.032	0.02	0.012	0.005	0.001	0	0	0	0	0	0	0

Note: Boldface type indicates the maximum probability of correct order being selected.

From the results in [1] to [12] we can conclude 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 ($\sigma_0^2 = 0.25$) to moderate ($\sigma_0^2 = 1$), 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. All of these conclusions can be summarized in Table 4.4.

Table 4.4 Appropriate value of α in $APIC = \log(\hat{\sigma}^2) + \frac{\alpha(p+1)}{n}$.

Model	n	X ~ Normal			X ~ Uniform		
		$\sigma^2 = 0.25$	$\sigma^2 = 1$	$\sigma^2 = 9$	$\sigma^2 = 0.25$	$\sigma^2 = 1$	$\sigma^2 = 9$
Model 1, 2 Weakly	15	small			small		
	30	moderate		small			
	100						
Model 3, 4 Strongly	15	large		moderate	moderate		small
	30						
	100						

CHAPTER 5

CONCLUSIONS AND FUTURE WORKS

5.1 Conclusions

In this research, we propose the model selection criteria, called Adjusted Penalty Information Criterion,

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

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. 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.

5.2 Future Works

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

BIBLIOGRAPHY

- Akaike, H. 1973. 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. pp. 267 – 281.
- Akaike, H. 1974. A New Look at the Statistical Model Identification. **IEEE Transactions on Automatic Control**. 19: 716 – 723.
- Bedrick, E. J. and Tsai, C. L. 1994. Model Selection for Multivariate Regression in Small Samples. **Biometrics**. 50: 226 – 231.
- Cavanaugh, J. E. 1999. A Large-Sample Model Selection Criterion Based on Kullback's Symmetric Divergence. **Statistics & Probability Letters**. 42: 333 – 343.
- Cavanaugh, J. E. 2004. Criteria for Linear Model Selection Based on Kullback's Symmetric Divergence. **Australian & New Zealand Journal of Statistics**. 46: 257 – 274.
- Gorobets, A. 2005. The Optimal Prediction Simultaneous Equations Selection. **Econometrics Bulletin**. 36: 1 – 8.
- Hafidi, B. and Mkhadri, A. 2006. A Corrected Akaike Criterion Based on Kullback's Symmetric Divergence: Applications in Time Series, Multiple and Multivariate Regression. **Computational Statistics & Data Analysis**. 50: 1524 – 1550.
- McQuarrie, A. D.; Shumway, R. and Tsai, C. L. 1997. The Model Selection Criterion AICu. **Statistics & Probability Letters**. 34: 285 – 292.
- McQuarrie, A. D. and Tsai, C. L. 1998. **Regression and Time Series Model Selection**. Singapore: World Scientific.
- McQuarrie, A. D. 1999. A Small-Sample Correction for the Schwarz SIC Model Selection Criterion. **Statistics & Probability Letters**. 44: 79 – 86.
- Mills, J. A. and Prasad, K. 2007. A Comparison of Model Selection Criteria. **Econometric Reviews**. 11: 201 – 234.
- Montgomery, D. C.; Peck, E. A. and Vining, G. G. 2006. **Introduction to Linear Regression Analysis**. 4th ed. New York: Wiley.

- Neath, A. and Cavanaugh, J. E. 1997. Regression and Time Series Model Selection Using Variants of the Schwarz Information Criterion. **Communication in Statistic-Theory and Method**. 26: 559 – 580.
- Rahman, M. S. and King, M. L. 2007. Improved Model Selection Criterion. **Communications in Statistics – Simulation and Computation**. 28: 51 – 71.
- Schwarz, G. 1978. Estimating the Dimension of a Model. **The Annals of Statistics**. 6: 461 – 464.
- Seghouane, A. K. 2006. A Note on Overfitting Properties of KIC and KICc. **Signal Processing**. 86: 3055 – 3060.

APPENDIX

OUTPUTS OF THIS RESEARCH

Submitted paper

Author(s)	Title	Journal	Vol	Page	Year	Data Base	Impact factor/year
Warangkhan Keerativibool and Pachitjanut Siripanich	Comparison of the Model Selection Criteria Based on Kullback- Leibler's Information	Chiang Mai Journal of Science				ISI	0.317/2014
Warangkhan Keerativibool and Pachitjanut Siripanich	Model Selection Criterion to Overcome the Weak Signal-to- Noise Ratio and to Reduce the Probability of Over/Underfitting	Chiang Mai Journal of Science				ISI	0.317/2014

Comparison of the Model Selection Criteria Based on Kullback-Leibler's Information

Warangkhan Keerativibool^{1,*} and Pachitjanut Siripanich²

¹ Department of Mathematics and Statistics, Faculty of Science, Thaksin University, Phatthalung, Thailand

² Faculty of Business Administration, Dhurakij Pundit University, Bangkok, Thailand

* Author for correspondence; e-mail: warang27@gmail.com

Abstract

This paper presents the derivations to unify the justifications of the criteria based on Kullback's divergence; AIC, AIC_c, KIC, KIC_{cc}, KIC_{csb}, and KIC_{chm}. The results show that KIC_{cc} has the strongest penalty function under some condition, followed, respectively, by KIC_{csb}, KIC_{chm}, KIC and AIC. Also, KIC is greater than AIC_c under some condition, but AIC_c always greater than AIC. The performances of all model selection criteria are examined by the extensive simulation study. It can be concluded that, for the small to moderate sample sizes and the true model is somewhat difficult to identify, the performances of AIC and AIC_c are better than others. However, they can identify the true model actually less accurate. When the sample size is large and the model is still weakly identifiable, the performances of all model selection criteria are insignificant difference, but all criteria can identify the true model still less accurate. As a result, we used the observed L_2 efficiency to assess model selection criteria performances. On the average, this measure suggests that in a weakly identifiable true model, whether the sample size is small or large, KIC_{cc} is the best criterion. For the small sample size and the true model can be specified more easily with small error variance, every model selection criteria still have the ability to select the correct model. If the error variance increase, the performances of all model selection criteria are bad. When the sample sizes are moderate to large and the true model can be specified more easily, KIC_c performs the best, it can identify a lot of true model for small error variance. But, if the error variance increases and the sample size is not large enough, all model selection criteria can identify a little true model.

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

1. INTRODUCTION

The problem of choosing the appropriate regression models 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 Akaike information criterion, AIC [1-2]. This serves as an asymptotically unbiased estimator of a variant of Kullback's directed divergence between the true and the candidate models. The directed divergence, also known as the I-divergence or the relative entropy, assesses the dissimilarity between two statistical models. Other well-known criterion was subsequently introduced and studied such as, Kullback information criterion, KIC [3]. It 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 two directed divergences is Kullback's symmetric divergence, also known as the J-divergence. Although AIC remains arguably the most widely used model selection criterion, KIC is a popular competitor. In fact, 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. However, when the sample size is small or the

dimension of candidate model is large relative to the sample size, both estimators suffer from a large negative bias. As a result, they have the problem of high probability of overfitting. In this setting, Hurvich and Tsai [4] proposed a corrected Akaike information criterion (AIC_c), for linear and non-linear regression and for autoregressive modeling. The AIC_c has been extended in a number of directions, including autoregressive moving average modeling [5], vector autoregressive modeling [6], and multivariate regression modeling [7]. Further, the KIC tends to underestimate the Kullback's symmetric divergence in small-sample applications, as indicated by Cavanaugh [8], Seghouane and Bekara [9], and Hafidi and Mkhadri [10]. Therefore, they [8-10] proposed KIC_c in order to correct this problem. Although AIC [1-2], AIC_c [4], KIC [3], KIC_{cc} [8], KIC_{csb} [9], and KIC_{chm} [10] share the same fundamental objective, the justifications of the criteria proceed along different directions, making it difficult to reconcile how the different model selection criteria refine the approximations used to establish Kullback's divergence. With this motivation, the aims of this paper are to unify the derivations of all criteria based on Kullback's directed divergence, AIC , AIC_c , and Kullback's symmetric divergence, KIC , KIC_{cc} , KIC_{csb} , KIC_{chm} , in order to link the justifications of these criteria and the performances of them are then examined by the extensive simulation study, under the difference various circumstances: sample sizes (n), regression coefficients (β), and variances of error terms (σ^2) [11-14]. Model selection criteria performances are examined by a consistent measure which is a measure of counting the frequency of order being selected. Particularly for the case of true model being weakly identifiable, we use an efficient measure which is the observed L_2 efficiency. This is a useful measure when the criteria do not select the correct model.

The true and the candidate models to consider in this study are, respectively, given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}_0, \boldsymbol{\varepsilon}_0 \sim N_n(\mathbf{0}, \sigma_0^2 \mathbf{I}_n), \quad (1)$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n), \quad (2)$$

where $\mathbf{y}$ is an $n \times 1$ dependent random vector of observations, $\mathbf{X}$ is an $n \times p$ matrix of independent variables with full-column rank, $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}$ are $p \times 1$ parameter vectors of regression coefficients, $\boldsymbol{\varepsilon}_0$ and $\boldsymbol{\varepsilon}$ are $n \times 1$ noise vectors. The true model is assumed to be correctly specified or overfitted by all the candidate models. This means that $\boldsymbol{\beta}_0$ has p_0 nonzero entries with $0 < p_0 \leq p$ and the rest of the $(p - p_0)$ entries are equal to zero. The $(p+1) \times 1$ vector of parameters is $\boldsymbol{\theta}_0 = [\boldsymbol{\beta}_0' \ \sigma_0^2]'$ and the maximum likelihood estimator of $\boldsymbol{\theta}_0$ is $\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\beta}}' \ \hat{\sigma}^2]'$ where

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \text{ and } \hat{\sigma}^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})/n. \quad (3)$$

The observed L_2 distance or squared error distance, scaled by $1/n$, between the true and the candidate models is defined as [13-14]

$$L_2(p) = (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})' \mathbf{X}'\mathbf{X}(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})/n. \quad (4)$$

Observed L_2 efficiency is defined by the ratio

$$\text{Observed } L_2 \text{ efficiency} = \frac{\min_{1 \leq p \leq P} L_2(p)}{L_2(p_s)}, \quad (5)$$

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 if the true model is weakly identifiable. The remainder of this paper is organized as follows. In Section 2, we show the

unifications for the derivations of all criteria based on Kullback's directed divergence, AIC, AIC_c , and Kullback's symmetric divergence, KIC, KIC_{cc} , KIC_{csb} , KIC_{chm} . Simulation study and results for 1,000 realizations of multiple regression models to examine the performances of all model selection criteria are shown in Section 3. Finally, Section 4 is the conclusions, discussion, and further works.

2. MATERIALS AND METHODS

The minus twice log likelihood of the candidate model in (2) when replacing the dependent vector $\mathbf{y}$ in (1) is defined by

$$-2\log L(\boldsymbol{\theta}|\mathbf{y}) = n\log 2\pi + n\log \sigma^2 + \frac{1}{\sigma^2} \boldsymbol{\varepsilon}'_0 \boldsymbol{\varepsilon}_0 + \frac{1}{\sigma^2} (\boldsymbol{\beta}_0 - \boldsymbol{\beta})' \mathbf{X}'\mathbf{X}(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + \frac{2}{\sigma^2} \boldsymbol{\varepsilon}'_0 \mathbf{X}(\boldsymbol{\beta}_0 - \boldsymbol{\beta}). \quad (6)$$

A well-known measure to separate the discrepancy between two models in (1) and (2) is given by Kullback's directed divergence or I-divergence [15],

$$2I(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = E_{\boldsymbol{\theta}_0} \left\{ 2\log \frac{L(\boldsymbol{\theta}_0|\mathbf{y})}{L(\boldsymbol{\theta}|\mathbf{y})} \right\} = d(\boldsymbol{\theta}_0, \boldsymbol{\theta}) - d(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0), \quad (7)$$

$$\text{where } d(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = E_{\boldsymbol{\theta}_0} \{ -2\log L(\boldsymbol{\theta}|\mathbf{y}) \}, \quad d(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0) = E_{\boldsymbol{\theta}_0} \{ -2\log L(\boldsymbol{\theta}_0|\mathbf{y}) \}, \quad (8)$$

and the expectation $E_{\boldsymbol{\theta}_0}$ is taken with respect to the true model in (1). Because $d(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0)$ does not depend on $\boldsymbol{\theta}$, any ranking of the candidate models according to (7) would be identical to ranking them according to $d(\boldsymbol{\theta}_0, \boldsymbol{\theta})$. Given a set of maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ in (3), the estimated directed measure $d(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ is

$$d(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}) = E_{\boldsymbol{\theta}_0} \{ -2\log L(\boldsymbol{\theta}|\mathbf{y}) \} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = n\log 2\pi + n\log \hat{\sigma}^2 + \frac{n\sigma_0^2}{\hat{\sigma}^2} + \frac{1}{\hat{\sigma}^2} (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})' \mathbf{X}'\mathbf{X}(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}). \quad (9)$$

However, the evaluation in (9) is not possible because it requires the knowledge of $\boldsymbol{\theta}_0$, Akaike [1-2] proposed an asymptotically unbiased estimator of

$$\Delta(\boldsymbol{\theta}_0, p) = E_{\boldsymbol{\theta}_0} \{ d(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}) \} \quad (10)$$

as

$$AIC = n\log \hat{\sigma}^2 + 2(p+1), \quad (11)$$

$$\text{i.e., } E_{\boldsymbol{\theta}_0} \{ AIC \} + o(1) = \Delta(\boldsymbol{\theta}_0, p).$$

Because of a large negative bias of AIC when the sample size is small or the dimension of candidate model is large relative to the sample size, Hurvich and Tsai [4] proposed an exactly unbiased estimator of (10) as follows:

$$AIC_c = n\log \hat{\sigma}^2 + \frac{2n(p+1)}{n-p-2}, \quad (12)$$

$$\text{i.e., } E_{\boldsymbol{\theta}_0} \{ AIC_c \} = \Delta(\boldsymbol{\theta}_0, p).$$

Cavanaugh [3], Seghouane and Bekara [9], Seghouane [12] summarized that the Kullback's directed divergence produced too underfitted value of model selection, and then it tended to be large for overparameterized models. An alternate measure to prevent both overfitting and underfitting problems is obtained by reversing the roles of two models in the definition of the measure, called Kullback's symmetric divergence or J-divergence,

$$2J(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = 2I(\boldsymbol{\theta}_0, \boldsymbol{\theta}) + 2I(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = [d(\boldsymbol{\theta}_0, \boldsymbol{\theta}) - d(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0)] + [d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) - d(\boldsymbol{\theta}, \boldsymbol{\theta})], \quad (13)$$

where $d(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ and $d(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$ are exhibited in (8), $d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = E_{\boldsymbol{\theta}} \{ -2\log L(\boldsymbol{\theta}_0|\mathbf{y}) \}$, and $d(\boldsymbol{\theta}, \boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \{ -2\log L(\boldsymbol{\theta}|\mathbf{y}) \}$.

Dropping $d(\theta_0, \theta_0)$, the ranking of the candidate models according to (13) is identical to ranking them according to

$$K(\theta_0, \theta) = d(\theta_0, \theta) + d(\theta, \theta_0) - d(\theta, \theta).$$

Given a set of maximum likelihood estimator $\hat{\theta}$ in (3), the estimated symmetric measure $K(\theta_0, \theta)$ is

$$K(\theta_0, \hat{\theta}) = d(\theta_0, \hat{\theta}) + d(\hat{\theta}, \theta_0) - d(\hat{\theta}, \hat{\theta}), \quad (14)$$

where $d(\theta_0, \hat{\theta})$ is exhibited in (9),

$$d(\hat{\theta}, \theta_0) = E_{\theta} \left\{ -2 \log L(\theta_0 | \mathbf{y}) \right\} \Big|_{\theta=\hat{\theta}} = n \log 2\pi + n \log \sigma_0^2 + \frac{n\hat{\sigma}^2}{\sigma_0^2} + \frac{1}{\sigma_0^2} (\hat{\beta} - \beta_0)' \mathbf{X}'\mathbf{X} (\hat{\beta} - \beta_0), \quad (15)$$

$$\text{and } d(\hat{\theta}, \hat{\theta}) = E_{\theta} \left\{ -2 \log L(\theta | \mathbf{y}) \right\} \Big|_{\theta=\hat{\theta}} = n \log 2\pi + n \log \hat{\sigma}^2 + n. \quad (16)$$

From (9), (15), and (16), the estimated symmetric measure in (14) can be written as

$$\begin{aligned} K(\theta_0, \hat{\theta}) &= n \log 2\pi + n \log \hat{\sigma}^2 + \frac{n\sigma_0^2}{\hat{\sigma}^2} + \frac{n\hat{\sigma}^2}{\sigma_0^2} + \frac{1}{\hat{\sigma}^2} (\beta_0 - \hat{\beta})' \mathbf{X}'\mathbf{X} (\beta_0 - \hat{\beta}) \\ &\quad + \frac{1}{\sigma_0^2} (\hat{\beta} - \beta_0)' \mathbf{X}'\mathbf{X} (\hat{\beta} - \beta_0) - n \log \left(\frac{\hat{\sigma}^2}{\sigma_0^2} \right) - n \end{aligned} \quad (17)$$

Similarly, the evaluation in (17) requires θ_0 , Cavanaugh [3] proposed an asymptotically unbiased estimator of

$$\Omega(\theta_0, p) = E_{\theta_0} \left\{ K(\theta_0, \hat{\theta}) \right\} \quad (18)$$

as

$$\text{KIC} = n \log \hat{\sigma}^2 + 3(p+1), \quad (19)$$

i.e., $E_{\theta_0} \{ \text{KIC} \} + o(1) = \Omega(\theta_0, p)$.

Seghouane and Bekara [9] proposed an exactly unbiased estimator of (18) in order to correct a large negative bias of KIC in (19) as follows:

$$\text{KIC}_c = n \log \hat{\sigma}^2 + \frac{2n(p+1)}{n-p-2} - n \psi \left(\frac{n-p}{2} \right) + n \log \left(\frac{n}{2} \right),$$

i.e., $E_{\theta_0} \{ \text{KIC}_c \} = \Omega(\theta_0, p)$.

Because the phi (ψ) or digamma function in KIC_c has no closed-form solution, Cavanaugh [8], Seghouane and Bekara [9], Hafidi and Mkhadri [10] gave the asymptotically unbiased estimators of (18) called, respectively, in this paper KIC_{cc} , KIC_{csb} , and KIC_{chm} ,

$$\text{KIC}_{cc} = n \log \hat{\sigma}^2 + n \log \left(\frac{n}{n-p} \right) + \frac{n[(n-p)(2p+3)-2]}{(n-p-2)(n-p)}, \quad (20)$$

$$\text{KIC}_{csb} = n \log \hat{\sigma}^2 + \frac{(p+1)(3n-p-2)}{n-p-2} + \frac{p}{n-p}, \quad (21)$$

$$\text{KIC}_{chm} = n \log \hat{\sigma}^2 + \frac{(p+1)(3n-p-2)}{n-p-2}. \quad (22)$$

To begin the unification of the derivations AIC in (11), AIC_c in (12), KIC in (19), KIC_{cc} in (20), KIC_{csb} in (21), and KIC_{chm} in (22), we give the following lemmas.

Lemma 1. $E_{\theta_0} \left\{ \frac{n\hat{\sigma}^2}{\sigma_0^2} \right\} = n-p$ (23)

$$E_{\theta_0} \left\{ \frac{1}{\sigma_0^2} (\hat{\beta} - \beta_0)' \mathbf{X}' \mathbf{X} (\hat{\beta} - \beta_0) \right\} = p \quad (24)$$

$$E_{\theta_0} \left\{ \frac{n\sigma_0^2}{\hat{\sigma}^2} \right\} = \frac{n^2}{n-p-2} \quad (25)$$

$$E_{\theta_0} \left\{ \frac{1}{\hat{\sigma}^2} (\hat{\beta} - \beta_0)' \mathbf{X}' \mathbf{X} (\hat{\beta} - \beta_0) \right\} = \frac{np}{n-p-2}. \quad (26)$$

Proof. From the fact that [16], $E_{\theta_0} \{ \chi_{d.f.}^2 \} = d.f.$ and the terms

$$\frac{n\hat{\sigma}^2}{\sigma_0^2} \text{ and } \frac{1}{\sigma_0^2} (\hat{\beta} - \beta_0)' \mathbf{X}' \mathbf{X} (\hat{\beta} - \beta_0)$$

are the independent χ^2 distributions with the degrees of freedom (d.f.) which are, respectively, $n-p$ and p . Therefore, we obtain (23) and (24). Since, we can write

$$E_{\theta_0} \left\{ \frac{n\sigma_0^2}{\hat{\sigma}^2} \right\} = n^2 E_{\theta_0} \left\{ \frac{1}{n\hat{\sigma}^2 / \sigma_0^2} \right\}. \quad (27)$$

Let $Y = \frac{n\hat{\sigma}^2}{\sigma_0^2}$ and we know that Y is the χ^2 distribution with $n-p$ degrees of freedom,

$$E_{\theta_0} \left\{ \frac{1}{n\hat{\sigma}^2 / \sigma_0^2} \right\} = E_{\theta_0} \left\{ \frac{1}{Y} \right\} = \int_0^\infty \frac{y^{\frac{n-p}{2}-2} \exp(-y/2)}{\Gamma\left(\frac{n-p}{2}\right) 2^{\frac{n-p}{2}}} dy = \frac{\Gamma\left(\frac{n-p}{2}-1\right) 2^{\frac{n-p}{2}-1}}{\Gamma\left(\frac{n-p}{2}\right) 2^{\frac{n-p}{2}}} = \frac{1}{n-p-2}. \quad (28)$$

Substituting the results in (28) into (27), we obtain (25). Since, we can write

$$E_{\theta_0} \left\{ \frac{1}{\hat{\sigma}^2} (\hat{\beta} - \beta_0)' \mathbf{X}' \mathbf{X} (\hat{\beta} - \beta_0) \right\} = \frac{1}{n} E_{\theta_0} \left\{ \frac{n\sigma_0^2}{\hat{\sigma}^2} \right\} E_{\theta_0} \left\{ \frac{1}{\sigma_0^2} (\hat{\beta} - \beta_0)' \mathbf{X}' \mathbf{X} (\hat{\beta} - \beta_0) \right\}.$$

Using the results in (24) and (25), we obtain (26).

Applying Lemma 1 into $\Delta(\theta_0, p)$ in (10) or the expected of the estimated directed measure in (9), we obtain

$$\begin{aligned} \Delta(\theta_0, p) &= E_{\theta_0} \left\{ d(\theta_0, \hat{\theta}) \right\} = n \log 2\pi + E_{\theta_0} \left\{ n \log \hat{\sigma}^2 \right\} + \frac{n^2}{n-p-2} + \frac{np}{n-p-2} \\ &= n(\log 2\pi + 1) + E_{\theta_0} \{ \text{AIC}_c \}, \end{aligned} \quad (29)$$

where AIC_c is the corrected version of AIC that was exhibited in (12).

Applying Lemma 1 into $\Omega(\theta_0, p)$ in (18) or the expected of the estimated symmetric measure in (17), we obtain

$$\begin{aligned} \Omega(\theta_0, p) &= E_{\theta_0} \left\{ K(\theta_0, \hat{\theta}) \right\} = n \log 2\pi + E_{\theta_0} \left\{ n \log \hat{\sigma}^2 \right\} + \frac{n^2}{n-p-2} + \frac{np}{n-p-2} - E_{\theta_0} \left\{ n \log \frac{\hat{\sigma}^2}{\sigma_0^2} \right\} \\ &= n(\log 2\pi + 1) + E_{\theta_0} \{ \text{AIC}_c \} - E_{\theta_0} \left\{ n \log \frac{\hat{\sigma}^2}{\sigma_0^2} \right\}, \end{aligned} \quad (30)$$

where AIC_c is the corrected version of AIC that was exhibited in (12).

It is noteworthy that, in KIC and KIC_c derived from $K(\theta_0, \hat{\theta})$ in (17), the differences in all formulas in (20) to (22) come from the last term of the right-hand side in (30). Therefore, in order to show the connections of KIC, KIC_{cc} , KIC_{csb} , and KIC_{chm} , we give the following lemmas.

Lemma 2. $-E_{\theta_0} \left\{ n \log \frac{\hat{\sigma}^2}{\sigma_0^2} \right\} = -n \log \left(\frac{n-p}{2} \right) + \frac{n}{n-p} + n \log \left(\frac{n}{2} \right) + o \left(\frac{n}{(n-p)^2} \right).$ (31)

Proof. From [16-17] we have, respectively,

$$E_{\theta_0} \left\{ \log \chi_{d.f.}^2 \right\} = \psi \left(\frac{d.f.}{2} \right) + \log 2 \text{ and } \psi(x) = \log x - \frac{1}{2x} + o \left(\frac{1}{x^2} \right) \text{ as } x \rightarrow \infty. \quad (32)$$

Applying (23) in Lemma 1 and the facts in (32), we have

$$\begin{aligned} -E_{\theta_0} \left\{ n \log \frac{\hat{\sigma}^2}{\sigma_0^2} \right\} &= -E_{\theta_0} \left\{ n \log \frac{n \hat{\sigma}^2}{\sigma_0^2} \right\} + n \log n = -n \left[\psi \left(\frac{n-p}{2} \right) + \log 2 \right] + n \log n \\ &= -n \left[\log \left(\frac{n-p}{2} \right) - \frac{1}{n-p} + o \left(\frac{1}{(n-p)^2} \right) \right] - n \log 2 + n \log n \\ &= -n \log \left(\frac{n-p}{2} \right) + \frac{n}{n-p} + n \log \left(\frac{n}{2} \right) + o \left(\frac{n}{(n-p)^2} \right). \end{aligned}$$

Lemma 3. $-n \log \left(\frac{n-p}{2} \right) + \frac{n}{n-p} + n \log \left(\frac{n}{2} \right) + o \left(\frac{n}{(n-p)^2} \right) = p + \frac{n}{n-p} + o \left(\frac{p^2}{n} \right) + o \left(\frac{n}{(n-p)^2} \right).$ (33)

Proof. Applying the first-order Taylor's series expansion to expand the term $\log((n-p)/2)$

about $n/2$, i.e., $\log \left(\frac{n-p}{2} \right) = \log \left(\frac{n}{2} \right) - \frac{p}{n} + o \left(\left(\frac{p}{n} \right)^2 \right)$, to obtain the approximation in (33).

Lemma 4. $p + \frac{n}{n-p} + o \left(\frac{p^2}{n} \right) + o \left(\frac{n}{(n-p)^2} \right) = (p+1) + o(1).$ (34)

Proof. Rearrange $p + n/(n-p)$ to be $(p+1) + p/(n-p)$. As $n \rightarrow \infty$ and p is held constant, the

term $\frac{p}{n-p} + o \left(\frac{p^2}{n} \right) + o \left(\frac{n}{(n-p)^2} \right)$ is $o(1)$ which yields the approximation in (34).

Applying Lemma 2 into $\Omega(\theta_0, p)$ in (30), we obtain

$$\begin{aligned} \Omega(\theta_0, p) &= n(\log 2\pi + 1) + E_{\theta_0} \{AIC_c\} - n \log \left(\frac{n-p}{2} \right) + \frac{n}{n-p} + n \log \left(\frac{n}{2} \right) + o \left(\frac{n}{(n-p)^2} \right) \\ &= n(\log 2\pi + 1) + E_{\theta_0} \left\{ KIC_{cc} + o \left(\frac{n}{(n-p)^2} \right) \right\}, \end{aligned}$$

where KIC_{cc} is the corrected version of KIC from Cavanaugh [8] that was exhibited in (20).

Applying Lemmas 2 and 3 into $\Omega(\theta_0, p)$ in (30), we obtain

$$\begin{aligned} \Omega(\theta_0, p) &= n(\log 2\pi + 1) + E_{\theta_0} \{AIC_c\} + p + \frac{n}{n-p} + o \left(\frac{p^2}{n} \right) + o \left(\frac{n}{(n-p)^2} \right) \\ &= n(\log 2\pi + 1) + E_{\theta_0} \left\{ KIC_{csb} + o \left(\frac{p^2}{n} \right) + o \left(\frac{n}{(n-p)^2} \right) \right\}, \end{aligned}$$

where KIC_{csb} is the corrected version of KIC from Seghouane and Bekara [9] that was exhibited in (21).

Applying Lemmas 2, 3, and 4 into $\Omega(\theta_0, p)$ in (30), we obtain

$$\Omega(\theta_0, p) = n(\log 2\pi + 1) + E_{\theta_0} \{AIC_c\} + (p+1) + o(1)$$

$$= n(\log 2\pi + 1) + E_{\theta_0} \{KIC_{\text{cHM}} + o(1)\},$$

where KIC_{cHM} is the corrected version of KIC from Hafidi and Mkhadri [10] that was exhibited in (22).

The connections of AIC , AIC_c , KIC , KIC_{cHM} , KIC_{cSB} , and KIC_{cC} are given in Lemma 5.

$$\textbf{Lemma 5. } AIC_c = AIC + \frac{2(p+1)(p+2)}{n-p-2}, \quad (35)$$

$$KIC = AIC + (p+1), \quad (36)$$

$$KIC = AIC_c + \frac{(p+1)(n-3p-6)}{n-p-2}, \quad (37)$$

$$KIC_{\text{cHM}} = KIC + \frac{2(p+1)(p+2)}{n-p-2}, \quad (38)$$

$$KIC_{\text{cSB}} = KIC_{\text{cHM}} + \frac{p}{n-p}, \quad (39)$$

$$KIC_{\text{cC}} = KIC_{\text{cSB}} + n \log \left(\frac{n}{n-p} \right) - p. \quad (40)$$

Proof. From AIC in (11) and AIC_c in (12),

$$\begin{aligned} AIC_c &= AIC - 2(p+1) + \frac{2n(p+1)}{n-p-2} = AIC + \frac{2(p+1)}{n-p-2}(-n+p+2+n) \\ &= AIC + \frac{2(p+1)(p+2)}{n-p-2}. \end{aligned}$$

From AIC in (11) and KIC in (19),

$$KIC = AIC - 2(p+1) + 3(p+1) = AIC + (p+1).$$

From AIC_c in (12) and KIC in (19),

$$\begin{aligned} KIC &= AIC_c - \frac{2n(p+1)}{n-p-2} + 3(p+1) = AIC_c + \frac{(p+1)}{n-p-2}(-2n+3n-3p-6) \\ &= AIC_c + \frac{(p+1)(n-3p-6)}{n-p-2}. \end{aligned}$$

From KIC in (19) and KIC_{cHM} in (22),

$$\begin{aligned} KIC_{\text{cHM}} &= KIC - 3(p+1) + \frac{(p+1)(3n-p-2)}{n-p-2} \\ &= KIC + \frac{(p+1)}{n-p-2}(-3n+3p+6+3n-p-2) = KIC + \frac{2(p+1)(p+2)}{n-p-2}. \end{aligned}$$

From KIC_{cSB} in (21) and KIC_{cHM} in (22), we already have (39). From KIC_{cC} in (20) and KIC_{cSB} in (21),

$$\begin{aligned} KIC_{\text{cC}} &= KIC_{\text{cSB}} - \frac{(p+1)(3n-p-2)}{n-p-2} - \frac{p}{n-p} + n \log \left(\frac{n}{n-p} \right) + \frac{n[(n-p)(2p+3)-2]}{(n-p-2)(n-p)} \\ &= KIC_{\text{cSB}} + n \log \left(\frac{n}{n-p} \right) \\ &\quad + \frac{-(p+1)(3n-p-2)(n-p) - p(n-p-2) + n(n-p)(2p+3) - 2n}{(n-p-2)(n-p)} \end{aligned}$$

$$\begin{aligned}
&= \text{KIC}_{\text{CSB}} + n \log \left(\frac{n}{n-p} \right) + \frac{-n^2p + 2np^2 - p^3 + 2np - 2p^2}{(n-p-2)(n-p)} \\
&= \text{KIC}_{\text{CSB}} + n \log \left(\frac{n}{n-p} \right) - \frac{p(n^2 - 2np + p^2 - 2n + 2p)}{(n-p-2)(n-p)} \\
&= \text{KIC}_{\text{CSB}} + n \log \left(\frac{n}{n-p} \right) - p.
\end{aligned}$$

Consider the connection in (35) and (38), we found that the term

$$\frac{2(p+1)(p+2)}{n-p-2} \quad (41)$$

is at most zero if and only if $n-p > 2$ and p belongs to the set of $[-2, -1]$. The connection in (36),

$$\frac{p+1}{n-p-2} \quad (42)$$

is at most zero if and only if $p \leq -1$. The connection in (37),

$$\frac{(p+1)(n-3p-6)}{n-p-2} \quad (43)$$

is at least zero if and only if

$$n-p > 2, n > 3 \text{ and } p \text{ belongs to the set of } [-1, n/3-2]. \quad (44)$$

The connection in (39),

$$\frac{p}{n-p} \quad (45)$$

is at most zero if and only if $n-p > 0$ and p belongs to the set of $(-\infty, 0]$.

Therefore, we can argue that the terms in (41), (42), and (45) have values of at least zero because p represents the number of regression coefficients which is an integer that has the value of at least one and all these terms are very close to zero if the ratio of p/n tends to zero. This conclusion links to $\text{AIC} \leq \text{AIC}_c$, $\text{AIC} \leq \text{KIC} \leq \text{KIC}_{\text{CHM}} \leq \text{KIC}_{\text{CSB}}$ whereas KIC is greater than AIC_c when the condition in (44) is true, such as when n equals to 15 and p equals to 2. While the term

$$n \log \left(\frac{n}{n-p} \right) - p \quad (46)$$

has the value in the range $[-p, \infty)$ where it is close to the lower bound $-p$ if the ratio of p/n tends to zero. If the value of p is fixed, this term is the decreasing function of n , whereas when the value of n is fixed, it is the increasing function of p . Whenever $n-p > 0$ and the condition

$$(1-p/n) \exp(p/n) < 1 \quad (47)$$

is true, we have the term in (46) being greater than zero. This means that the penalty function of KIC_{CC} is stronger than other criteria, KIC_{CSB} , KIC_{CHM} , KIC , AIC_c , and AIC under the condition in (47). The strong penalty may cause model selection criteria to have the maximum frequency of the correct order being selected. However, occasionally it causes them to select underparameterized models [16]. This confusion is studied by the extensive simulation in the next section.

3. SIMULATION STUDY AND RESULTS

To examine the model selection criteria performances, we generated the true multiple regression models in (1) as follows.

Model I represents a weakly identifiable true model ($p_0 = 3$):

$$y = 1 + 0.5X_2 + 0.1X_3 + \varepsilon_0. \quad (48)$$

Model II represents a strongly identifiable true model ($p_0 = 5$):

$$y = 1 + X_2 + X_3 + X_4 + X_5 + \varepsilon_0. \quad (49)$$

A weakly identifiable true model, Model I, means it is not easily identified compared to the strongly identifiable true models such as Model II. The error terms ε_0 in (48) and (49) are assumed to be normally distributed with zero mean and variances σ_0^2 equal to three levels: 0.25, 1, and 9. For both model in (48) and (49), 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 steps for simulation and all results are as follows.

3.1 Use the RAND function of SAS programming to generate the error terms ε_0 in (48) and (49) about 100,000 observations to be normal random variables with zero mean and variances equal to 0.25, 1, and 9.

3.2 Split the series of error terms in step 1 into 1,000 samples, each of which consists of three levels of sample sizes, $n = 15, 30, 100$ observations.

3.3 Use the RAND function of SAS programming to generate the independent variables X_2 until X_7 about 100,000 observations to be the normal random variables with zero mean and variance equals to one where the relevant independent variables of Model I are X_2, X_3 and irrelevant independent variables are X_4, X_5 , whereas the relevant independent variables of Model II are X_2 until X_5 and irrelevant independent variables are X_6, X_7 .

3.4 Split the series of independent variables in step 3 into 1,000 samples, each of which consists of 15, 30, 100 observations. For this study, X_1 is given as a constant which equals one.

3.5 Use the corresponding relevant independent variables obtained in Step 4 and the error terms obtained in Step 2 to construct the dependent variables described in (48) and (49).

3.6 Use the concept of all possible subsets as potential candidate models. For Model I, we consider $2^4 - 1 = 15$ subsets, while Model II, we consider $2^6 - 1 = 63$ subsets. For each subset, calculate AIC in (11), AIC_c in (12), KIC in (19), KIC_{cc} in (20), KIC_{csb} in (21), and KIC_{chm} in (22). The subset with the minimum value of model selection criterion can be classified to be the best model. Due to the large number of subsets, it is impractical to summarize the individual models chosen. Hence, Tables 1 to 3 summarize $p = \text{rank}(X)$ of the selected subset to be three groups: the selected order less than p is called underfitted order, the selected order equals to p is called correct order, and the selected order greater than p is called overfitted order. Although $p = 3, 5$ are, respectively, the correct orders of Model I and Model II, $p = 3, 5$ may include models with the correct number of variables but are not the correct model. The “True Order” row summarizes counts for correctly selecting the true model. Tables 4 display the candidate models that are closest to the true model in the L_2 sense. The ave. and S.D. L_2 rows denote, respectively, the average and standard deviation of observed L_2 efficiency in (5) over 1,000 realizations. For 1,000 realizations, the results of comparing the model selection criteria performances can be concluded as follows.

(1) For the small to moderate sample sizes and the true model is somewhat difficult to identify (weakly identifiable) as Model I for $n = 15, 30$, AIC and AIC_c perform the best, but these criteria can identify the true model about 15 – 25% of the time for $\sigma_0^2 = 0.25$, about 10 – 15% of the time for $\sigma_0^2 = 1$, and about 3 – 5% of the time for $\sigma_0^2 = 9$.

(2) For the sample size is large and the model is still weakly identifiable as Model I for $n = 100$, the performances of all model selection criteria are insignificant difference, but all criteria can identify the true model about 47 – 50% of the time for $\sigma_0^2 = 0.25$, about 21 – 24% of the time for $\sigma_0^2 = 1$, and about 3 – 7% of the time for $\sigma_0^2 = 9$.

(3) For the small sample size and the true model can be specified more easily (strongly identifiable) with small error variance as Model II for $n = 15$, every model selection criteria still have the ability to select the correct model about 51 – 90% of the time for $\sigma_0^2 = 0.25$ and about 29 – 45% of the time for $\sigma_0^2 = 1$. Whereas, when the error variance increase as Model II for $n = 15$ and $\sigma_0^2 = 9$, the performances of all model selection criteria are bad.

(4) For the sample sizes are moderate to large and the true model can be specified more easily (strongly identifiable) as Model II for $n = 30, 100$, KIC_c performs the best, it can identify the

true model about 84 – 87% of the time for small error variance $\sigma_0^2 = 0.25, 1$. If the error variance increases ($\sigma_0^2 = 9$) and the sample size is moderate as Model II for $n = 30$, all model selection criteria can identify the true model about 2 – 10% of the time. However, when the sample size increases ($n = 100$), the performances of all model selection criteria are insignificant difference. All criteria can identify the true model about 58 – 61% of the time.

(5) Since the strongly penalty term of KIC family, KIC and KIC_c are underfitted more than AIC and AIC_c in AIC family, especially when the true model is very difficult to detect, sample size is small, and error variance is large.

(6) On the average, the observed L_2 efficiency in Table 4 suggests that KIC_{cc} in KIC family is the best criterion for all sample sizes of a weakly identifiable true model.

(7) This simulation also found that; when the sample size increases including the regression coefficient increases or the model can be easily to identify, the frequency of order being selected and the observed L_2 efficiency also increase. While, the error variance is effect to the frequency of order being selected and the observed L_2 efficiency decrease.

Place Table 1. about here

Place Table 2. about here

Place Table 3. about here

Place Table 4. about here

4. CONCLUSIONS, DISCUSSION, AND FUTURE WORKS

This paper presents the derivations to unify the justifications of the criteria based on Kullback's directed and symmetric divergence; Akaike information criterion (AIC) by Akaike [1-2] and the corrected version; AIC_c by Hurvich and Tsai [4], Kulback information criterion (KIC) by Cavanaugh [3] and the corrected versions; KIC_{cc} by Cavanaugh [8], KIC_{csb} by Seghouane and Bekara [9], and KIC_{chm} by Hafidi and Mkhadri [10]. The results show that KIC_{cc} has the strongest penalty function under the condition in (47), followed, respectively, by KIC_{csb} , KIC_{chm} , KIC, and AIC. If the condition in (44) is true, KIC is greater than AIC_c . However, AIC_c always greater than AIC. The strong penalty may cause model selection criteria to have the maximum frequency of the correct order being selected. However, sometimes it may cause the underfitted problem. The performances of all model selection criteria, AIC, AIC_c , KIC, KIC_{cc} , KIC_{csb} , and KIC_{chm} , are examined by the extensive simulation study. Our simulation study indicates that, for the small to moderate sample sizes and the true model is somewhat difficult to identify, the performances of AIC and AIC_c are better than others. However, they can identify the true model about 10 – 25% of the time for small error variance and can identify the true model about 3 – 5% of the time for large error variance. When the sample size is large and the model is still weakly identifiable, the performances of all model selection criteria are insignificant difference, but all criteria can identify the true model about 21 – 50% of the time for small error variance and can identify the true model about 3 – 7% of the time for large error variance. As a result, the frequency of correct order being selected may not be meaningful. For this reason, we have also used the observed L_2 efficiency to assess model selection criteria performances. On the average, this measure suggests that in a weakly identifiable true model, whether the sample size is small or large, KIC_{cc} is the best criterion because it has highest average value of the observed L_2 efficiency and lowest standard deviation. The better performance of KIC_{cc} may be because its formula is closer to the expected estimated symmetric discrepancy in (18) than other. But, KIC_{cc} is more likely to select an underfitted model than other criteria which is because its penalty function is strong. Nevertheless, even if KIC_{cc} tends to select underfitted models, these selected models are close to the true model. For the small sample size and the true model can be specified more easily with small error variance, every model

selection criteria have the ability to select the correct model about 29 – 90% of the time. Whereas, the error variance increase, the performances of all model selection criteria are bad. When the sample sizes are moderate to large and the true model can be specified more easily, KIC_c performs the best, it can identify the true model about 84 – 87% of the time for small error variance. If the error variance increases and the sample size is moderate, all model selection criteria can identify the true model about 2 – 10% of the time. However, when the sample size increases, the performances of all model selection criteria are insignificant difference. All criteria can identify the true model about 58 – 61% of the time. Since the strongly penalty term of KIC family, they are underfitted more than the criteria in AIC family, especially when the true model is very difficult to detect, sample size is small, and error variance is large. This study also found that; when the sample size increases including the regression coefficient increases or the model can be easily to identify, the frequency of order being selected and the observed L_2 efficiency also increase. While, the error variance is effect to the frequency of order being selected and the observed L_2 efficiency decrease, this result is opposite to Kundu and Murali [18] which concluded that the criteria performances did not change much when the error variance was changed. In future work, we hope to find a model selection criterion to overcome 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 under various circumstances as follows: sample sizes, orders of true model, regression coefficients, variances of error terms, distributions of independent variables, and distributions of error terms.

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*, 1973, Akademia Kiado, Budapest, pp. 267-281. DOI 10.1007/978-1-4612-1694-0_15.
- [2] Akaike, H., A New Look at the Statistical Model Identification, *IEEE T. Automat. Contr.*, 1974; **19**: 716-723. DOI 10.1109/TAC.1974.1100705.
- [3] Cavanaugh, J.E., A Large-Sample Model Selection Criterion Based on Kullback's Symmetric Divergence, *Stat. Probabil. Lett.*, 1999; **42**: 333-343. DOI 10.1016/S0167-7152(98)00200-4.
- [4] Hurvich, C.M. and Tsai, C.L., Regression and Time Series Model Selection in Small Samples, *Biometrika*, 1989; **76**: 297-307. DOI 10.1093/biomet/76.2.297.
- [5] Hurvich, C.M., Shumway, R.H. and Tsai, C.L., Improved Estimators of Kullback-Leibler Information for Autoregressive Model Selection in Small Samples. *Biometrika*, 1990; **77**: 709-719. DOI 10.1093/biomet/77.4.709.
- [6] Hurvich, C.M. and Tsai, C.L., A Corrected Akaike Information Criterion for Vector Autoregressive Model Selection. *J. Time Ser.* 1993; **14**: 271-279. DOI 10.1111/j.1467-9892.1993.tb00144.x.
- [7] Bedrick, E.J. and Tsai, C.L., Model Selection for Multivariate Regression in Small Samples, *Biometrics*, 1994; **50**: 226-231. DOI 10.2307/2533213.
- [8] Cavanaugh, J.E., Criteria for Linear Model Selection Based on Kullback's Symmetric Divergence, *Aust. NZ. J. Stat.*, 2004; **46**: 257-274. DOI 10.1111/j.1467-842X.2004.00328.x.
- [9] Seghouane, A.K. and Bekara, M., A Small Sample Model Selection Criterion Based on Kullback's Symmetric Divergence, *IEEE T. Signal Proces.*, 2004; **52**: 3314-3323. DOI 10.1109/TSP.2004.837416.
- [10] Hafidi, B. and 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. DOI 10.1016/j.csda.2005.01.007.

- [11] Rahman, M.S. and King, M.L. Improved model selection criterion. *Commun. Stat. Simulat.*, 2007; **28**: 51-71. DOI 10.1080/03610919908813535.
- [12] Seghouane, A.K., A Note on Overfitting Properties of KIC and KICc, *Signal Process.*, 2006; **86**: 3055-3060. DOI 10.1016/j.sigpro.2006.01.002.
- [13] McQuarrie, A.D., Shumway, R. and Tsai, C.L. The Model Selection Criterion AIC_u, *Stat. Probabil. Lett.*, 1997; **34**: 285-292. DOI 10.1016/S0167-7152(96)00192-7.
- [14] McQuarrie, A.D., A Small-Sample Correction for the Schwarz SIC Model Selection Criterion, *Stat. Probabil. Lett.*, 1999; **44**: 79-86. DOI 10.1016/S0167-7152(98)00294-6.
- [15] Kullback, S., *Information Theory and Statistics*, New York, Dover, 1968.
- [16] McQuarrie, A.D. and Tsai, C.L., *Regression and Time Series Model Selection*, World Scientific, Singapore, 1998.
- [17] Bernardo, J.M., Psi (Digamma) Function, *J. Roy. Stat. Soc. C-App.*, 1976; **25**: 315-317. DOI 10.2307/2347257.
- [18] Kundu, D. and Murali G., Model Selection in Linear Regression, *Comput. Stat. Data An.*, 1996; **22**: 461-469. DOI 10.1016/0167-9473(96)00008-4.

Table 1. Frequency of the model order being selected by each criterion for 1,000 realizations: $\sigma_0^2 = 0.25$.

Model	n	Counts for order p	Criteria					
			AIC	AIC _c	KIC	KIC _{CHM}	KIC _{CSB}	KIC _{CC}
I	15	order < p	393	686	565	773	786	804
		p = 3	398	277	326	210	199	185
		order > p	209	37	109	17	15	11
		True Order	187	139	156	108	102	95
I	30	order < p	409	538	591	681	688	695
		p = 3	430	384	343	282	278	276
		order > p	161	78	66	37	34	29
		True Order	246	219	198	172	169	169
I	100	order < p	201	227	334	362	362	364
		p = 3	579	585	549	543	545	546
		order > p	220	188	117	95	93	90
		True Order	493	504	469	468	468	469
II	15	order < p	1	28	5	42	43	53
		p = 5	516	900	652	910	910	906
		order > p	483	72	343	48	47	41
		True Order	512	895	646	903	903	899
II	30	order < p	0	0	0	0	0	0
		p = 5	609	802	762	859	863	874
		order > p	391	198	238	141	137	126
		True Order	609	802	762	859	863	874
II	100	order < p	0	0	0	0	0	0
		p = 5	688	722	815	846	847	851
		order > p	312	278	185	154	153	149
		True Order	688	722	815	846	847	851

Note: Boldface type indicates the maximum frequency of correct order being selected; more than one boldface type is insignificant difference.

Table 2. Frequency of the model order being selected by each criterion for 1,000 realizations: $\sigma_0^2 = 1$.

Model	n	Counts for order p	Criteria					
			AIC	AIC _c	KIC	KIC _{CHM}	KIC _{CSB}	KIC _{CC}
I	15	order < p	534	829	735	902	905	915
		p = 3	320	156	202	90	87	78
		order > p	146	15	63	8	8	7
		True Order	104	49	64	36	36	32
I	30	order < p	524	644	703	805	806	814
		p = 3	342	291	245	175	175	168
		order > p	134	65	52	20	19	18
		True Order	151	130	110	81	81	78
I	100	order < p	446	465	584	608	609	613
		p0 = 3	410	413	351	334	333	330
		order > p	144	122	65	58	58	57
		True Order	233	241	219	216	215	214
II	15	order < p	100	530	208	639	649	681
		p = 5	451	435	510	342	334	307
		order > p	450	35	283	19	17	12
		True Order	391	395	448	320	314	292
II	30	order < p	3	12	11	22	23	25
		p = 5	640	789	758	854	856	864
		order > p	357	199	231	124	121	111
		True Order	638	788	757	852	854	861
II	100	order < p	0	0	0	0	0	0
		p = 5	674	722	812	842	843	846
		order > p	326	278	188	158	157	154
		True Order	674	722	812	842	843	846

Note: Boldface type indicates the maximum frequency of correct order being selected; more than one boldface type is insignificant difference.

Table 3. Frequency of the model order being selected by each criterion for 1,000 realizations: $\sigma_0^2 = 9$.

Model	n	Counts for order p	Criteria					
			AIC	AIC _c	KIC	KIC _{CHM}	KIC _{CSB}	KIC _{CC}
I	15	order < p	705	920	846	960	965	968
		p = 3	215	74	126	37	32	30
		order > p	80	6	28	3	3	2
		True Order	47	20	24	11	10	10
I	30	order < p	738	856	887	941	942	945
		p = 3	202	128	102	56	55	52
		order > p	60	16	11	3	3	3
		True Order	44	28	24	12	11	9
I	100	order < p	708	749	858	874	876	877
		p0 = 3	239	214	126	118	116	115
		order > p	53	37	16	8	8	8
		True Order	70	65	38	36	35	35
II	15	order < p	999	1000	1000	1000	1000	1000
		p = 5	1	0	0	0	0	0
		order > p	0	0	0	0	0	0
		True Order	0	0	0	0	0	0
II	30	order < p	665	877	857	946	951	957
		p = 5	244	107	122	50	45	41
		order > p	91	16	21	4	4	2
		True Order	97	48	52	28	25	24
II	100	order < p	97	117	212	245	246	249
		p = 5	615	647	645	635	634	637
		order > p	288	236	143	120	120	114
		True Order	575	606	611	602	602	604

Note: Boldface type indicates the maximum frequency of correct order being selected; more than one boldface type is insignificant difference.

Table 4. Average and standard deviation of the observed L_2 efficiency over 1,000 realizations

Circumstance	n	Stat.	Criteria					
			AIC	AICc	KIC	KIC _{CHM}	KIC _{CSB}	KIC _{CC}
Model I $\sigma_0^2 = 0.25$	15	Ave. L_2 eff.	0.6013	0.7004	0.6548	0.7278	0.7319	0.7394
		Rank	6	4	5	3	2	1
		S.D. L_2 eff.	0.3108	0.2985	0.3084	0.2916	0.2907	0.2891
		Rank	6	4	5	3	2	1
Model I $\sigma_0^2 = 0.25$	30	Ave. L_2 eff.	0.6301	0.6388	0.6404	0.6553	0.6563	0.6578
		Rank	6	5	4	3	2	1
		S.D. L_2 eff.	0.3076	0.2950	0.2917	0.2827	0.2822	0.2821
		Rank	6	5	4	3	2	1
Model I $\sigma_0^2 = 0.25$	100	Ave. L_2 eff.	0.7210	0.7249	0.6955	0.6927	0.6926	0.6930
		Rank	2	1	3	5	6	4
		S.D. L_2 eff.	0.3182	0.3195	0.3313	0.3330	0.3331	0.3332
		Rank	2	1	3	5	6	4
Model I $\sigma_0^2 = 1$	15	Ave. L_2 eff.	0.5483	0.6517	0.6195	0.6808	0.6820	0.6863
		Rank	6	4	5	3	2	1
		S.D. L_2 eff.	0.3421	0.3603	0.3610	0.3563	0.3562	0.3550
		Rank	6	4	5	3	2	1
Model I $\sigma_0^2 = 1$	30	Ave. L_2 eff.	0.7103	0.7372	0.7458	0.7703	0.7705	0.7716
		Rank	6	5	4	3	2	1
		S.D. L_2 eff.	0.3148	0.3131	0.3150	0.3098	0.3095	0.3094
		Rank	6	5	4	3	2	1
Model I $\sigma_0^2 = 1$	100	Ave. L_2 eff.	0.6291	0.6384	0.6606	0.6674	0.6673	0.6681
		Rank	6	5	4	2	3	1
		S.D. L_2 eff.	0.2993	0.2986	0.2891	0.2855	0.2853	0.2847
		Rank	6	5	4	2	3	1
Model I $\sigma_0^2 = 9$	15	Ave. L_2 eff.	0.3634	0.3951	0.3838	0.4010	0.4018	0.4026
		Rank	6	4	5	3	2	1
		S.D. L_2 eff.	0.2806	0.2913	0.2879	0.2913	0.2914	0.2918
		Rank	6	4	5	3	2	1
Model I $\sigma_0^2 = 9$	30	Ave. L_2 eff.	0.4534	0.4768	0.4829	0.4946	0.4946	0.4954
		Rank	6	5	4	2	3	1
		S.D. L_2 eff.	0.3225	0.3280	0.3297	0.3314	0.3315	0.3317
		Rank	6	5	4	2	3	1
Model I $\sigma_0^2 = 9$	100	Ave. L_2 eff.	0.6115	0.6251	0.6622	0.6717	0.6721	0.6730
		Rank	6	5	4	3	2	1
		S.D. L_2 eff.	0.3545	0.3555	0.3543	0.3542	0.3543	0.3540
		Rank	6	5	4	3	2	1

Note: Boldface type indicates the best performance.

Table 4. (Continued.)

Circumstance	n	Stat.	Criteria					
			AIC	AICc	KIC	KIC _{c_{HM}}	KIC _{c_{SB}}	KIC _{c_C}
Model I	15	Ave. L ₂ eff.	0.5043	0.5824	0.5527	0.6032	0.6052	0.6094
Ave. all σ_0^2		Rank	6	4	5	3	2	1
		S.D. L ₂ eff.	0.3112	0.3167	0.3191	0.3131	0.3128	0.3120
		Rank	6	4	5	3	2	1
Model I	30	Ave. L ₂ eff.	0.5979	0.6176	0.6230	0.6401	0.6405	0.6416
Ave. all σ_0^2		Rank	6	5	4	3	2	1
		S.D. L ₂ eff.	0.3150	0.3120	0.3122	0.3080	0.3077	0.3077
		Rank	6	5	4	3	2	1
Model I	100	Ave. L ₂ eff.	0.6539	0.6628	0.6728	0.6773	0.6773	0.6781
Ave. all σ_0^2		Rank	6	5	4	3	2	1
		S.D. L ₂ eff.	0.3240	0.3246	0.3249	0.3242	0.3242	0.3240
		Rank	6	5	4	3	2	1

Note: Boldface type indicates the best performance.

Model Selection Criterion to Overcome the Weak Signal-to-Noise Ratio and to Reduce the Probability of Over/Underfitting

Warangkhan Keerativibool^{1,*} and Pachitjanut Siripanich²

¹ Department of Mathematics and Statistics, Faculty of Science, Thaksin University, Phatthalung, Thailand

² Faculty of Business Administration, Dhurakij Pundit University, Bangkok, Thailand

* Author for correspondence; e-mail: warang27@gmail.com

Abstract

This paper proposed a model selection criterion in order to overcome the weak signal-to-noise ratio and to reduce the probability of over/underfitting by adjusting the penalty term of the well-known model selection criteria (*AIC*, *BIC*, *KIC*), called adjusted penalty information criterion, $APIC = \log(\hat{\sigma}^2) + \alpha(p+1)/n$. 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, 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.

1. INTRODUCTION

In the application of statistics, the statistical modeling is considered 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 overcome 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 [9-12]: sample sizes, orders of true model, regression coefficients, variances of error terms and distributions of independent variables. 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. In Section 2, we propose Adjusted Penalty Information Criterion (*APIC*) in order to overcome the weak signal-to-noise ratio and to reduce the probability of over/underfitting. 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.

2. MATERIALS AND METHODS

The true univariate 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 univariate 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,

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

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 $APIC_{p_0} < APIC_{p_0+l}$. Then the signal-to-noise ratio that the true model is selected compared to a candidate model is

$$\frac{\text{signal}}{\text{noise}} = \frac{E[APIC_{p_0+l} - APIC_{p_0}]}{sd[APIC_{p_0+l} - APIC_{p_0}]} = \frac{E\left[\log\left(\hat{\sigma}_{p_0+l}^2 / \hat{\sigma}_{p_0}^2\right) + \frac{\alpha l}{n}\right]}{sd\left[\log\left(\hat{\sigma}_{p_0+l}^2 / \hat{\sigma}_{p_0}^2\right) + \frac{\alpha l}{n}\right]}. \quad (4)$$

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

$$\log(X) \doteq \log(p) + (X-p)/p - (X-p)^2/2p^2 \text{ and } E[\log(X)] \doteq \log(p) - 1/p. \quad (5)$$

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

$$n(\hat{\sigma}_p^2 - \hat{\sigma}_{p+l}^2) \sim \sigma_0^2 \chi_l^2, \quad n\hat{\sigma}_p^2 \sim \sigma_0^2 \chi_{n-p}^2 \text{ and } \hat{\sigma}_p^2 - \hat{\sigma}_{p+l}^2 \text{ is independent of } \hat{\sigma}_{p+l}^2, \quad (6)$$

where χ_k^2 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

$$E[\log(n\hat{\sigma}_p^2)] \doteq \log(\sigma_0^2) + \log(n-p) - 1/(n-p). \quad (7)$$

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

$$E[APIC_{p_0+l} - APIC_{p_0}] \doteq \log\left(\frac{n-p_0-l}{n-p_0}\right) - \frac{l}{(n-p_0-l)(n-p_0)} + \frac{\alpha l}{n}. \quad (8)$$

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

$$Q = \frac{n\hat{\sigma}_{p_0+l}^2}{n\hat{\sigma}_{p_0}^2} \sim \frac{\chi_{n-p_0-l}^2}{\chi_{n-p_0-l}^2 + \chi_l^2}, \quad (9)$$

the Q -statistic in (9) has the Beta distribution

$$Q \sim \text{Beta}((n-p_0-l)/2, l/2),$$

and the log-distribution is

$$\log(Q) = \log(n\hat{\sigma}_{p_0+l}^2/n\hat{\sigma}_{p_0}^2) \sim \log\text{-Beta}((n-p_0-l)/2, l/2). \quad (10)$$

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

$$\begin{aligned} \text{var}[\log(Q)] &\doteq \left(\frac{n-p_0}{n-p_0-l}\right)^2 \left[\frac{(n-p_0-l)/2 \cdot l/2}{((n-p_0-l)/2 + l/2)^2 ((n-p_0-l)/2 + l/2 + 1)} \right] \\ &= \frac{2l}{(n-p_0-l)(n-p_0+2)}. \end{aligned} \quad (11)$$

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

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

In (12), we found that 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:

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

Place Table 1. about here

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 used 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

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

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

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

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

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

In (14) and (15), we found that *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:

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

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

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

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.

Place Table 2. about here

Place Table 3. about here

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, we can conclude that, $APIC$ with a much more value of α , make its probability of overfitting to be smaller but make more probability of underfitting.

3. 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 with the true order $p_0 = 5$):

$$y_1 = 1 + 0.5X_2 + 0.4X_3 + 0.3X_4 + 0.2X_5 + \varepsilon_1,$$

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

$$y_2 = 1 + 0.5X_2 + 0.4X_3 + \varepsilon_2,$$

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

$$y_3 = 1 + 2X_2 + 2X_3 + \varepsilon_3,$$

Model 4 (very strongly identifiable true model with the true order $p_0 = 5$):

$$y_4 = 1 + 2X_2 + 2X_3 + 2X_4 + 2X_5 + \varepsilon_4.$$

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. Results are shown in Table 4 to Table 6. Findings are the following.

[1] In Table 4, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[2] In Table 4, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[3] In Table 4, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[4] In Table 4, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[5] In Table 5, for very weakly identifiable situation of true models with the true orders $p_0 = 5$, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[6] In Table 5, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[7] In Table 5, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[8] In Table 5, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[9] In Table 6, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[10] In Table 6, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[11] In Table 6, 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 $\sigma_0^2 = 0.25, 1, 9$, 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%.

[12] In Table 6, for very strongly identifiable situation of true models with the true orders $p_0 = 5$, Model 4, the sample size is moderate ($n = 100$) and the distribution of independent variable is normal, the appropriate values of α when the true variances $\sigma_0^2 = 0.25, 1, 9$, 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%.

Place Table 4. about here

Place Table 5. about here

Place Table 6. about here

From the results in [1] to [12] we can conclude 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 ($\sigma_0^2 = 0.25$) to moderate ($\sigma_0^2 = 1$), 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.

4. CONCLUSIONS, DISCUSSION and FUTURE WORKS

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

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

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. 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 overcome the weak signal-to-noise ratio and to reduce the probability of over/underfitting in the multivariate regression and simultaneous equations models.

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*, 1973, Akademia Kiado, Budapest, pp. 267-281. DOI 10.1007/978-1-4612-1694-0_15.
- [2] Akaike, H., A New Look at the Statistical Model Identification, *IEEE T. Automat. Contr.*, 1974; **19**: 716-723. DOI 10.1109/TAC.1974.1100705.
- [3] Bedrick, E.J. and Tsai, C.L., Model Selection for Multivariate Regression in Small Samples, *Biometrics*, 1994; **50**: 226-231. DOI 10.2307/2533213.
- [4] Schwarz, G., Estimating the Dimension of a Model. *Ann. Stat.*, 1978; **6**: 461-464. DOI. 10.2307/2958889.
- [5] Cavanaugh, J.E., A Large-Sample Model Selection Criterion Based on Kullback's Symmetric Divergence, *Stat. Probabil. Lett.*, 1999; **42**: 333-343. DOI 10.1016/S0167-7152(88)00200-4.
- [6] Hafidi, B. and 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. DOI 10.1016/j.csda.2005.01.007.
- [7] Neath, A. and Cavanaugh, J.E., Regression and Time Series Model Selection using Variants of the Schwarz Information Criterion, *Commun. Stat. Theory.*, 1997; **26**: 559-580. DOI. 10.1080/03610929708831934.

- [8] Cavanaugh, J.E., Criteria for Linear Model Selection Based on Kullback's Symmetric Divergence, *Aust. NZ. J. Stat.*, 2004; **46**: 257-274. DOI 10.1111/j.1467-842X.2004.00328.x.
- [9] McQuarrie, A.D., Shumway, R. and Tsai, C.L., The Model Selection Criterion AIC_u , *Stat. Probabil. Lett.*, 1997; **34**: 285-292. DOI 10.1016/S0167-7152(86)00192-7.
- [10] McQuarrie, A.D., A Small-Sample Correction for the Schwarz SIC Model Selection Criterion, *Stat. Probabil. Lett.*, 1999; **44**: 79-86. DOI 10.1016/S0167-7152(88)00294-6.
- [11] Mills, J.A. and Prasad, K., A Comparison of Model Selection Criteria. *Econometric Reviews*. 2007; **11**: 201- 234. DOI. 10.1080/07474939208800232.
- [12] Rahman, M.S. and King, M.L., Improved model selection criterion. *Commun. Stat. Simulat.*, 2007; **28**: 51-71. DOI. 10.1080/03610919908813535.
- [13] Montgomery, D.C., Peck, E.A. and Vining, G.G., *Introduction to Linear Regression Analysis*, 4th Edn., New York, Wiley, 2006.
- [14] McQuarrie, A.D. and Tsai, C.L., *Regression and Time Series Model Selection*, World Scientific, Singapore, 1998.
- [15] Seghouane, A.K., Multivariate Regression Model Selection from Small Samples using Kullback's Symmetric Divergence, *Signal Processing*, 2006; **86**: 2074-2084. DOI. 10.1016/j.sigpro.2005.10.009.

Table 1. Signal-to-noise ratio of $APIC\alpha$ for different values of n , p_0 and l .

n	p_0	l	<i>Criteria</i>														
			$APIC1$	$APIC2$ (AIC)	$APIClog(n)$ (BIC)	$APIC3$ (KIC)	$APIC4$	$APIC5$	$APIC6$	$APIC7$	$APIC8$	$APIC9$	$APIC10$	$APIC11$	$APIC12$	$APIC13$	$APIC14$
15	3	1	-0.2450	0.3400	0.7542	0.9250	1.5100	2.0950	2.6800	3.2650	3.8500	4.4350	5.0200	5.6050	6.1900	6.7750	7.3600
15	3	2	-0.3884	0.4004	0.9589	1.1892	1.9780	2.7668	3.5556	4.3444	5.1333	5.9221	6.7109	7.4997	8.2885	9.0773	9.8661
15	3	3	-0.5291	0.3874	1.0364	1.3039	2.2204	3.1370	4.0535	4.9700	5.8865	6.8030	7.7195	8.6360	9.5526	10.4691	11.3856
15	3	4	-0.6752	0.3225	1.0290	1.3203	2.3181	3.3159	4.3136	5.3114	6.3092	7.3070	8.3047	9.3025	10.3003	11.2981	12.2958
15	5	1	-0.3660	0.1239	0.4708	0.6138	1.1037	1.5936	2.0835	2.5734	3.0633	3.5532	4.0431	4.5330	5.0229	5.5128	6.0027
15	5	2	-0.5625	0.0907	0.5532	0.7439	1.3971	2.0503	2.7035	3.3567	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)	$APIC4$	$APIC5$	$APIC6$	$APIC7$	$APIC8$	$APIC9$	$APIC10$	$APIC11$	$APIC12$	$APIC13$	$APIC14$
30	3	1	-0.1132	0.5340	1.1812	1.4409	1.8284	2.4756	3.1229	3.7701	4.4173	5.0645	5.7117	6.3589	7.0062	7.6534	8.3006
30	3	2	-0.1785	0.7190	1.6166	1.9767	2.5141	3.4116	4.3092	5.2067	6.1042	7.0017	7.8993	8.7968	9.6943	10.5918	11.4894
30	3	3	-0.2414	0.8356	1.9127	2.3448	2.9897	4.0667	5.1438	6.2208	7.2978	8.3749	9.4519	10.5289	11.6060	12.6830	13.7600
30	3	4	-0.3054	0.9120	2.1295	2.6179	3.3470	4.5644	5.7819	6.9994	8.2168	9.4343	10.6518	11.8692	13.0867	14.3041	15.5216
30	5	1	-0.1648	0.4352	1.0352	1.2759	1.6352	2.2352	2.8352	3.4352	4.0352	4.6352	5.2352	5.8352	6.4352	7.0352	7.6352
30	5	2	-0.2516	0.5791	1.4097	1.7430	2.2404	3.0710	3.9017	4.7324	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)	$APIC5$	$APIC6$	$APIC7$	$APIC8$	$APIC9$	$APIC10$	$APIC11$	$APIC12$	$APIC13$	$APIC14$
100	3	1	-0.0324	0.6569	1.3463	2.0356	2.4528	2.7250	3.4143	4.1037	4.7930	5.4824	6.1717	6.8611	7.5504	8.2398	8.9291
100	3	2	-0.0510	0.9188	1.8886	2.8584	3.4453	3.8282	4.7980	5.7678	6.7376	7.7074	8.6772	9.6470	10.6168	11.5866	12.5564
100	3	3	-0.0687	1.1128	2.2942	3.4757	4.1907	4.6572	5.8387	7.0202	8.2016	9.3831	10.5646	11.7461	12.9276	14.1091	15.2905
100	3	4	-0.0867	1.2703	2.6273	3.9843	4.8055	5.3413	6.6982	8.0552	9.4122	10.7692	12.1262	13.4831	14.8401	16.1971	17.5541
100	5	1	-0.0469	0.6283	1.3035	1.9787	2.3874	2.6539	3.3292	4.0044	4.6796	5.3548	6.0300	6.7052	7.3804	8.0556	8.7308
100	5	2	-0.0714	0.8784	1.8282	2.7780	3.3527	3.7277	4.6775	5.6273	6.5771	7.5269	8.4767	9.4265	10.3763	11.3261	12.2758

Table 2. Probability of overfitting by l extra variables of $APIC\alpha$ for different values of n , p_0 and l .

n	p_0	l	<i>Criteria</i>														
			$APIC1$	$APIC2$ (AIC)	$APIClog(n)$ (BIC)	$APIC3$ (KIC)	$APIC4$	$APIC5$	$APIC6$	$APIC7$	$APIC8$	$APIC9$	$APIC10$	$APIC11$	$APIC12$	$APIC13$	$APIC14$
15	3	1	0.4025	0.2363	0.1682	0.1469	0.0939	0.0611	0.0402	0.0266	0.0178	0.0119	0.0080	0.0054	0.0037	0.0025	0.0017
15	3	2	0.5134	0.2636	0.1644	0.1353	0.0695	0.0357	0.0183	0.0094	0.0048	0.0025	0.0013	0.0007	0.0003	0.0002	0.0001
15	3	3	0.5947	0.2857	0.1631	0.1287	0.0561	0.0240	0.0101	0.0042	0.0018	0.0007	0.0003	0.0001	0.0001	0.0000	0.0000
15	3	4	0.6664	0.3143	0.1701	0.1305	0.0508	0.0190	0.0070	0.0025	0.0009	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000
15	5	1	0.4511	0.2865	0.2148	0.1917	0.1316	0.0918	0.0647	0.0460	0.0329	0.0236	0.0170	0.0123	0.0089	0.0065	0.0047
15	5	2	0.5866	0.3442	0.2359	0.2019	0.1184	0.0695	0.0408	0.0239	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)	$APIC4$	$APIC5$	$APIC6$	$APIC7$	$APIC8$	$APIC9$	$APIC10$	$APIC11$	$APIC12$	$APIC13$	$APIC14$
30	3	1	0.3565	0.1922	0.1102	0.0890	0.0651	0.0392	0.0239	0.0147	0.0091	0.0057	0.0035	0.0022	0.0014	0.0009	0.0006
30	3	2	0.4346	0.1889	0.0821	0.0588	0.0357	0.0155	0.0067	0.0029	0.0013	0.0006	0.0002	0.0001	0.0000	0.0000	0.0000
30	3	3	0.4846	0.1795	0.0617	0.0397	0.0204	0.0066	0.0021	0.0007	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
30	3	4	0.5256	0.1720	0.0482	0.0282	0.0125	0.0031	0.0007	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
30	5	1	0.3761	0.2106	0.1252	0.1026	0.0766	0.0478	0.0301	0.0192	0.0123	0.0079	0.0051	0.0033	0.0022	0.0014	0.0009
30	5	2	0.4646	0.2158	0.1003	0.0737	0.0466	0.0216	0.0101	0.0047	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)	$APIC5$	$APIC6$	$APIC7$	$APIC8$	$APIC9$	$APIC10$	$APIC11$	$APIC12$	$APIC13$	$APIC14$
100	3	1	0.3284	0.1670	0.0905	0.0506	0.0360	0.0289	0.0167	0.0097	0.0057	0.0034	0.0020	0.0012	0.0007	0.0004	0.0003
100	3	2	0.3867	0.1496	0.0578	0.0224	0.0126	0.0087	0.0033	0.0013	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
100	3	3	0.4178	0.1288	0.0367	0.0100	0.0045	0.0027	0.0007	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	3	4	0.4395	0.1109	0.0236	0.0046	0.0017	0.0009	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
100	5	1	0.3336	0.1715	0.0940	0.0531	0.0380	0.0306	0.0179	0.0105	0.0062	0.0037	0.0022	0.0013	0.0008	0.0005	0.0003
100	5	2	0.3946	0.1557	0.0614	0.0242	0.0138	0.0096	0.0038	0.0015	0.0006	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000

Table 3. Probability of underfitting by l variables of $APIC\alpha$ for different values of n , p_0 and l .

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

Table 4. Probability of the order selected by *APIC* for $n = 15$.

Model	Dist. of X	σ_0^2	Order	Criteria														
				<i>APIC1</i>	<i>APIC2</i> (AIC)	<i>APIClog(n)</i> (BIC)	<i>APIC3</i> (KIC)	<i>APIC4</i>	<i>APIC5</i>	<i>APIC6</i>	<i>APIC7</i>	<i>APIC8</i>	<i>APIC9</i>	<i>APIC10</i>	<i>APIC11</i>	<i>APIC12</i>	<i>APIC13</i>	<i>APIC14</i>
1 very weakly $p_0 = 5$	Normal	0.25	Underfitted	0.157	0.353	0.472	0.53	0.669	0.785	0.868	0.93	0.958	0.974	0.983	0.99	0.994	0.998	1
			Correct	0.229	0.297	0.293	0.281	0.232	0.165	0.107	0.061	0.038	0.025	0.016	0.009	0.005	0.002	0
			Overfitted	0.614	0.35	0.235	0.189	0.099	0.05	0.025	0.009	0.004	0.001	0.001	0.001	0.001	0	0
		1	Underfitted	0.269	0.562	0.731	0.79	0.9	0.951	0.978	0.99	0.998	0.999	0.999	0.999	1	1	1
			Correct	0.155	0.135	0.104	0.085	0.05	0.025	0.011	0.006	0.001	0	0	0	0	0	0
			Overfitted	0.576	0.303	0.165	0.125	0.05	0.024	0.011	0.004	0.001	0.001	0.001	0.001	0	0	0
		9	Underfitted	0.367	0.69	0.826	0.863	0.946	0.98	0.993	0.998	1	1	1	1	1	1	1
			Correct	0.119	0.097	0.064	0.054	0.03	0.011	0.005	0.001	0	0	0	0	0	0	0
			Overfitted	0.514	0.213	0.11	0.083	0.024	0.009	0.002	0.001	0	0	0	0	0	0	0
	Uniform	0.25	Underfitted	0.346	0.679	0.82	0.855	0.937	0.971	0.99	0.998	1	1	1	1	1	1	1
			Correct	0.132	0.093	0.066	0.053	0.027	0.014	0.004	0.001	0	0	0	0	0	0	0
			Overfitted	0.522	0.228	0.114	0.092	0.036	0.015	0.006	0.001	0	0	0	0	0	0	0
		1	Underfitted	0.365	0.691	0.838	0.878	0.947	0.98	0.997	0.998	0.999	0.999	1	1	1	1	1
			Correct	0.113	0.079	0.052	0.039	0.015	0.007	0.001	0.001	0.001	0	0	0	0	0	0
			Overfitted	0.522	0.23	0.11	0.083	0.038	0.013	0.002	0.001	0	0	0	0	0	0	0
		9	Underfitted	0.359	0.695	0.836	0.868	0.946	0.985	0.995	0.997	0.999	1	1	1	1	1	1
			Correct	0.106	0.074	0.041	0.034	0.018	0.007	0.004	0.003	0.001	0	0	0	0	0	0
			Overfitted	0.535	0.231	0.123	0.098	0.036	0.008	0.001	0	0	0	0	0	0	0	0
2 weakly $p_0 = 3$	Normal	0.25	Underfitted	0.02	0.057	0.112	0.124	0.195	0.274	0.366	0.44	0.522	0.59	0.654	0.722	0.762	0.804	0.844
			Correct	0.215	0.465	0.587	0.614	0.658	0.639	0.578	0.528	0.458	0.4	0.34	0.276	0.236	0.194	0.156
			Overfitted	0.765	0.478	0.301	0.262	0.147	0.087	0.056	0.032	0.02	0.01	0.006	0.002	0.002	0.002	0
		1	Underfitted	0.091	0.282	0.416	0.464	0.604	0.704	0.779	0.842	0.882	0.915	0.942	0.965	0.97	0.981	0.987
			Correct	0.147	0.301	0.333	0.33	0.286	0.237	0.191	0.143	0.11	0.082	0.057	0.034	0.03	0.019	0.013
			Overfitted	0.762	0.417	0.251	0.206	0.11	0.059	0.03	0.015	0.008	0.003	0.001	0.001	0	0	0
		9	Underfitted	0.181	0.52	0.693	0.739	0.85	0.899	0.943	0.961	0.974	0.98	0.984	0.99	0.992	0.993	0.998
			Correct	0.095	0.119	0.104	0.101	0.081	0.065	0.046	0.032	0.023	0.018	0.016	0.01	0.008	0.007	0.002
			Overfitted	0.724	0.361	0.203	0.16	0.069	0.036	0.011	0.007	0.003	0.002	0	0	0	0	0
	Uniform	0.25	Underfitted	0.152	0.425	0.583	0.644	0.768	0.855	0.896	0.93	0.96	0.972	0.982	0.991	0.995	0.996	0.997
			Correct	0.111	0.178	0.173	0.17	0.149	0.11	0.084	0.059	0.036	0.026	0.018	0.009	0.005	0.004	0.003
			Overfitted	0.737	0.397	0.244	0.186	0.083	0.035	0.02	0.011	0.004	0.002	0	0	0	0	0
		1	Underfitted	0.179	0.478	0.635	0.695	0.841	0.915	0.941	0.961	0.976	0.985	0.991	0.995	0.997	0.999	1
			Correct	0.103	0.126	0.119	0.108	0.079	0.054	0.04	0.031	0.02	0.013	0.009	0.005	0.003	0.001	0
			Overfitted	0.718	0.396	0.246	0.197	0.08	0.031	0.019	0.008	0.004	0.002	0	0	0	0	0
		9	Underfitted	0.189	0.491	0.658	0.717	0.844	0.925	0.957	0.977	0.985	0.993	0.996	0.997	0.998	0.998	0.998
			Correct	0.094	0.136	0.132	0.115	0.075	0.04	0.028	0.014	0.009	0.006	0.003	0.003	0.002	0.002	0.002
			Overfitted	0.717	0.373	0.21	0.168	0.081	0.035	0.015	0.009	0.006	0.001	0.001	0	0	0	0

Note: Boldface type indicates the maximum probability of correct order being selected.

Table 4. (Continued).

Model	Dist. of X	σ_0^2	Order	Criteria														
				APIC1	APIC2 (AIC)	APIClog(n) (BIC)	APIC3 (KIC)	APIC4	APIC5	APIC6	APIC7	APIC8	APIC9	APIC10	APIC11	APIC12	APIC13	APIC14
3 strongly $p_0 = 3$	Normal	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.22	0.515	0.654	0.707	0.814	0.891	0.942	0.962	0.971	0.983	0.991	0.992	0.995	0.997	0.998
			Overfitted	0.78	0.485	0.346	0.293	0.186	0.109	0.058	0.038	0.029	0.017	0.009	0.008	0.005	0.003	0.002
		1	Underfitted	0	0	0	0	0	0.001	0.001	0.004	0.005	0.01	0.014	0.019	0.029	0.055	0.074
			Correct	0.238	0.542	0.683	0.725	0.835	0.902	0.946	0.961	0.975	0.977	0.977	0.977	0.968	0.943	0.924
			Overfitted	0.762	0.458	0.317	0.275	0.165	0.097	0.053	0.035	0.02	0.013	0.009	0.004	0.003	0.002	0.002
		9	Underfitted	0.025	0.118	0.189	0.213	0.313	0.412	0.516	0.578	0.661	0.714	0.776	0.826	0.864	0.902	0.921
			Correct	0.197	0.427	0.518	0.541	0.554	0.514	0.447	0.399	0.323	0.275	0.22	0.171	0.133	0.097	0.078
			Overfitted	0.778	0.455	0.293	0.246	0.133	0.074	0.037	0.023	0.016	0.011	0.004	0.003	0.003	0.001	0.001
	Uniform	0.25	Underfitted	0.004	0.009	0.016	0.019	0.039	0.063	0.095	0.121	0.158	0.206	0.25	0.316	0.382	0.446	0.511
			Correct	0.215	0.52	0.654	0.703	0.799	0.846	0.858	0.854	0.823	0.784	0.745	0.68	0.616	0.552	0.487
			Overfitted	0.781	0.471	0.33	0.278	0.162	0.091	0.047	0.025	0.019	0.01	0.005	0.004	0.002	0.002	0.002
		1	Underfitted	0.041	0.154	0.237	0.272	0.377	0.473	0.587	0.661	0.713	0.779	0.828	0.87	0.903	0.917	0.942
			Correct	0.198	0.389	0.461	0.476	0.485	0.467	0.387	0.322	0.276	0.215	0.169	0.128	0.097	0.083	0.058
			Overfitted	0.761	0.457	0.302	0.252	0.138	0.06	0.026	0.017	0.011	0.006	0.003	0.002	0	0	0
		9	Underfitted	0.153	0.45	0.611	0.671	0.797	0.876	0.924	0.953	0.972	0.979	0.986	0.993	0.999	0.999	1
			Correct	0.112	0.154	0.157	0.149	0.121	0.083	0.054	0.035	0.022	0.02	0.014	0.007	0.001	0.001	0
			Overfitted	0.735	0.396	0.232	0.18	0.082	0.041	0.022	0.012	0.006	0.001	0	0	0	0	0
4 very strongly $p_0 = 5$	Normal	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0.001	0.004	0.007	0.016
			Correct	0.342	0.55	0.659	0.686	0.804	0.874	0.914	0.94	0.958	0.97	0.978	0.983	0.984	0.985	0.977
			Overfitted	0.658	0.45	0.341	0.314	0.196	0.126	0.086	0.06	0.042	0.03	0.022	0.016	0.012	0.008	0.007
		1	Underfitted	0	0.001	0.001	0.001	0.003	0.005	0.011	0.019	0.036	0.061	0.108	0.169	0.248	0.348	0.475
			Correct	0.309	0.546	0.669	0.698	0.797	0.851	0.894	0.916	0.913	0.902	0.866	0.813	0.738	0.642	0.518
			Overfitted	0.691	0.453	0.33	0.301	0.2	0.144	0.095	0.065	0.051	0.037	0.026	0.018	0.014	0.01	0.007
		9	Underfitted	0.052	0.167	0.252	0.292	0.437	0.589	0.728	0.819	0.885	0.934	0.955	0.976	0.989	0.995	0.997
			Correct	0.304	0.448	0.466	0.462	0.418	0.328	0.234	0.158	0.103	0.061	0.043	0.023	0.011	0.005	0.003
			Overfitted	0.644	0.385	0.282	0.246	0.145	0.083	0.038	0.023	0.012	0.005	0.002	0.001	0	0	0
	Uniform	0.25	Underfitted	0.004	0.016	0.026	0.032	0.057	0.109	0.175	0.274	0.414	0.545	0.669	0.761	0.853	0.923	0.959
			Correct	0.344	0.569	0.656	0.69	0.764	0.782	0.76	0.684	0.556	0.438	0.323	0.234	0.145	0.077	0.041
			Overfitted	0.652	0.415	0.318	0.278	0.179	0.109	0.065	0.042	0.03	0.017	0.008	0.005	0.002	0	0
		1	Underfitted	0.078	0.209	0.314	0.363	0.535	0.676	0.798	0.878	0.938	0.966	0.983	0.993	0.998	1	1
			Correct	0.298	0.399	0.423	0.415	0.345	0.252	0.162	0.1	0.056	0.031	0.015	0.006	0.002	0	0
			Overfitted	0.624	0.392	0.263	0.222	0.12	0.072	0.04	0.022	0.006	0.003	0.002	0.001	0	0	0
		9	Underfitted	0.278	0.593	0.755	0.801	0.911	0.965	0.987	0.995	0.998	0.999	0.999	0.999	1	1	1
			Correct	0.148	0.127	0.089	0.074	0.043	0.017	0.007	0.004	0.002	0.001	0.001	0.001	0	0	0
			Overfitted	0.574	0.28	0.156	0.125	0.046	0.018	0.006	0.001	0	0	0	0	0	0	0

Note: Boldface type indicates the maximum probability of correct order being selected.

Table 5. Probability of the order selected by *APIC* for $n = 30$.

Model	Dist. of X	σ_0^2	Order	Criteria														
				<i>APIC1</i>	<i>APIC2</i> (<i>AIC</i>)	<i>APIC3</i> (<i>KIC</i>)	<i>APIClog(n)</i> (<i>BIC</i>)	<i>APIC4</i>	<i>APIC5</i>	<i>APIC6</i>	<i>APIC7</i>	<i>APIC8</i>	<i>APIC9</i>	<i>APIC10</i>	<i>APIC11</i>	<i>APIC12</i>	<i>APIC13</i>	<i>APIC14</i>
1 very weakly $p_0 = 5$	Normal	0.25	Underfitted	0.076	0.199	0.312	0.366	0.446	0.547	0.656	0.745	0.827	0.877	0.927	0.958	0.973	0.985	0.99
			Correct	0.406	0.532	0.55	0.522	0.481	0.41	0.328	0.249	0.171	0.122	0.073	0.042	0.027	0.015	0.01
			Overfitted	0.518	0.269	0.138	0.112	0.073	0.043	0.016	0.006	0.002	0.001	0	0	0	0	0
		1	Underfitted	0.275	0.55	0.761	0.799	0.855	0.937	0.976	0.985	0.995	0.999	1	1	1	1	1
			Correct	0.246	0.237	0.161	0.146	0.108	0.054	0.023	0.015	0.005	0.001	0	0	0	0	0
			Overfitted	0.479	0.213	0.078	0.055	0.037	0.009	0.001	0	0	0	0	0	0	0	0
		9	Underfitted	0.48	0.804	0.915	0.94	0.974	0.993	0.999	1	1	1	1	1	1	1	1
			Correct	0.135	0.075	0.045	0.032	0.017	0.005	0.001	0	0	0	0	0	0	0	0
			Overfitted	0.385	0.121	0.04	0.028	0.009	0.002	0	0	0	0	0	0	0	0	0
	Uniform	0.25	Underfitted	0.392	0.696	0.871	0.905	0.952	0.982	0.998	0.999	1	1	1	1	1	1	1
			Correct	0.175	0.137	0.071	0.056	0.031	0.012	0.002	0.001	0	0	0	0	0	0	0
			Overfitted	0.433	0.167	0.058	0.039	0.017	0.006	0	0	0	0	0	0	0	0	0
		1	Underfitted	0.48	0.828	0.96	0.976	0.985	0.997	0.999	1	1	1	1	1	1	1	1
			Correct	0.13	0.074	0.023	0.016	0.011	0.003	0.001	0	0	0	0	0	0	0	0
			Overfitted	0.39	0.098	0.017	0.008	0.004	0	0	0	0	0	0	0	0	0	0
		9	Underfitted	0.48	0.818	0.95	0.971	0.988	0.996	0.999	1	1	1	1	1	1	1	1
			Correct	0.133	0.063	0.024	0.016	0.007	0.002	0	0	0	0	0	0	0	0	0
			Overfitted	0.387	0.119	0.026	0.013	0.005	0.002	0.001	0	0	0	0	0	0	0	0
2 weakly $p_0 = 3$	Normal	0.25	Underfitted	0.001	0.003	0.009	0.018	0.025	0.04	0.066	0.093	0.135	0.171	0.224	0.28	0.338	0.392	0.45
			Correct	0.321	0.632	0.802	0.837	0.884	0.908	0.906	0.892	0.857	0.824	0.772	0.719	0.661	0.607	0.55
			Overfitted	0.678	0.365	0.189	0.145	0.091	0.052	0.028	0.015	0.008	0.005	0.004	0.001	0.001	0.001	0
		1	Underfitted	0.057	0.173	0.305	0.346	0.413	0.514	0.606	0.661	0.742	0.804	0.85	0.875	0.896	0.919	0.935
			Correct	0.296	0.514	0.555	0.55	0.522	0.452	0.378	0.328	0.252	0.194	0.15	0.125	0.104	0.081	0.065
			Overfitted	0.647	0.313	0.14	0.104	0.065	0.034	0.016	0.011	0.006	0.002	0	0	0	0	0
		9	Underfitted	0.236	0.577	0.756	0.792	0.84	0.907	0.938	0.959	0.974	0.982	0.986	0.993	0.995	0.998	0.999
			Correct	0.147	0.185	0.154	0.145	0.118	0.077	0.055	0.04	0.026	0.018	0.014	0.007	0.005	0.002	0.001
			Overfitted	0.617	0.238	0.09	0.063	0.042	0.016	0.007	0.001	0	0	0	0	0	0	0
	Uniform	0.25	Underfitted	0.174	0.437	0.616	0.667	0.734	0.811	0.871	0.904	0.937	0.953	0.972	0.98	0.985	0.991	0.996
			Correct	0.21	0.292	0.276	0.254	0.217	0.166	0.115	0.088	0.06	0.045	0.027	0.02	0.015	0.009	0.004
			Overfitted	0.616	0.271	0.108	0.079	0.049	0.023	0.014	0.008	0.003	0.002	0.001	0	0	0	0
		1	Underfitted	0.257	0.599	0.776	0.826	0.867	0.925	0.952	0.972	0.982	0.988	0.991	0.996	0.999	0.999	1
			Correct	0.127	0.166	0.14	0.121	0.103	0.063	0.046	0.028	0.018	0.012	0.009	0.004	0.001	0.001	0
			Overfitted	0.616	0.235	0.084	0.053	0.03	0.012	0.002	0	0	0	0	0	0	0	0
		9	Underfitted	0.317	0.655	0.83	0.875	0.913	0.953	0.978	0.989	0.994	0.996	0.997	0.999	0.999	0.999	1
			Correct	0.107	0.118	0.079	0.069	0.058	0.036	0.019	0.009	0.004	0.003	0.003	0.001	0.001	0.001	0
			Overfitted	0.576	0.227	0.091	0.056	0.029	0.011	0.003	0.002	0.002	0.001	0	0	0	0	0

Note: Boldface type indicates the maximum probability of correct order being selected.

Table 5. (Continued).

Model	Dist. of X	σ_0^2	Order	Criteria														
				APIC1	APIC2 (AIC)	APIC3 (KIC)	APIClog(n) (BIC)	APIC4	APIC5	APIC6	APIC7	APIC8	APIC9	APIC10	APIC11	APIC12	APIC13	APIC14
3 strongly $p_0 = 3$	Normal	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.317	0.634	0.824	0.861	0.906	0.942	0.972	0.977	0.986	0.99	0.996	0.998	0.998	0.998	1
			Overfitted	0.683	0.366	0.176	0.139	0.094	0.058	0.028	0.023	0.014	0.01	0.004	0.002	0.002	0.002	0
		1	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.348	0.664	0.832	0.874	0.91	0.96	0.979	0.99	0.995	0.998	0.998	0.999	0.999	0.999	0.999
			Overfitted	0.652	0.336	0.168	0.126	0.09	0.04	0.021	0.01	0.005	0.002	0.002	0.001	0.001	0.001	0.001
		9	Underfitted	0.003	0.021	0.038	0.044	0.06	0.091	0.137	0.185	0.256	0.304	0.37	0.437	0.485	0.544	0.593
			Correct	0.316	0.62	0.765	0.807	0.842	0.855	0.832	0.796	0.735	0.691	0.627	0.56	0.513	0.454	0.406
			Overfitted	0.681	0.359	0.197	0.149	0.098	0.054	0.031	0.019	0.009	0.005	0.003	0.003	0.002	0.002	0.001
	Uniform	0.25	Underfitted	0	0	0	0	0	0	0	0.001	0.003	0.006	0.007	0.012	0.021	0.03	0.038
			Correct	0.308	0.625	0.813	0.862	0.908	0.948	0.968	0.976	0.984	0.985	0.988	0.986	0.978	0.969	0.961
			Overfitted	0.692	0.375	0.187	0.138	0.092	0.052	0.032	0.023	0.013	0.009	0.005	0.002	0.001	0.001	0.001
		1	Underfitted	0.009	0.044	0.095	0.122	0.155	0.23	0.298	0.374	0.44	0.497	0.566	0.627	0.686	0.734	0.782
			Correct	0.331	0.62	0.739	0.756	0.759	0.735	0.68	0.611	0.549	0.496	0.43	0.37	0.312	0.264	0.216
			Overfitted	0.66	0.336	0.166	0.122	0.086	0.035	0.022	0.015	0.011	0.007	0.004	0.003	0.002	0.002	0.002
		9	Underfitted	0.189	0.504	0.688	0.74	0.797	0.864	0.905	0.934	0.957	0.974	0.982	0.987	0.99	0.992	0.994
			Correct	0.196	0.232	0.209	0.198	0.167	0.119	0.087	0.063	0.042	0.026	0.018	0.013	0.01	0.008	0.006
			Overfitted	0.615	0.264	0.103	0.062	0.036	0.017	0.008	0.003	0.001	0	0	0	0	0	0
4 very strongly $p_0 = 5$	Normal	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.481	0.713	0.831	0.867	0.903	0.945	0.969	0.979	0.985	0.99	0.993	0.994	0.996	0.998	1
			Overfitted	0.519	0.287	0.169	0.133	0.097	0.055	0.031	0.021	0.015	0.01	0.007	0.006	0.004	0.002	0
		1	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.454	0.705	0.845	0.881	0.908	0.95	0.971	0.985	0.992	0.996	0.998	0.998	0.998	0.998	1
			Overfitted	0.546	0.295	0.155	0.119	0.092	0.05	0.029	0.015	0.008	0.004	0.002	0.002	0.002	0.002	0
		9	Underfitted	0.009	0.029	0.064	0.074	0.102	0.172	0.26	0.354	0.457	0.578	0.708	0.801	0.86	0.907	0.947
			Correct	0.447	0.675	0.765	0.793	0.797	0.767	0.703	0.623	0.527	0.413	0.287	0.198	0.139	0.093	0.053
			Overfitted	0.544	0.296	0.171	0.133	0.101	0.061	0.037	0.023	0.016	0.009	0.005	0.001	0.001	0	0
	Uniform	0.25	Underfitted	0	0	0	0	0	0	0	0.001	0.002	0.011	0.024	0.04	0.066	0.12	0.204
			Correct	0.419	0.699	0.829	0.869	0.901	0.946	0.971	0.982	0.986	0.981	0.971	0.958	0.934	0.88	0.796
			Overfitted	0.581	0.301	0.171	0.131	0.099	0.054	0.029	0.017	0.012	0.008	0.005	0.002	0	0	0
		1	Underfitted	0.018	0.053	0.116	0.154	0.214	0.317	0.427	0.544	0.664	0.763	0.844	0.894	0.947	0.967	0.983
			Correct	0.445	0.658	0.723	0.719	0.703	0.635	0.554	0.448	0.329	0.234	0.154	0.106	0.053	0.033	0.017
			Overfitted	0.537	0.289	0.161	0.127	0.083	0.048	0.019	0.008	0.007	0.003	0.002	0	0	0	0
		9	Underfitted	0.323	0.646	0.826	0.87	0.925	0.969	0.992	0.998	0.998	0.999	0.999	1	1	1	1
			Correct	0.223	0.175	0.115	0.092	0.055	0.024	0.006	0.001	0.001	0.001	0.001	0	0	0	0
			Overfitted	0.454	0.179	0.059	0.038	0.02	0.007	0.002	0.001	0.001	0	0	0	0	0	0

Note: Boldface type indicates the maximum probability of correct order being selected.

Table 6. Probability of the order selected by *APIC* for $n = 100$.

Model	Dist. of X	σ_0^2	Order	Criteria														
				APIC1	APIC2 (AIC)	APIC3 (KIC)	APIC4	APIClog(n) (BIC)	APIC5	APIC6	APIC7	APIC8	APIC9	APIC10	APIC11	APIC12	APIC13	APIC14
1 very weakly $p_0 = 5$	Normal	0.25	Underfitted	0	0.005	0.016	0.031	0.045	0.05	0.078	0.107	0.144	0.199	0.246	0.293	0.341	0.394	0.458
			Correct	0.537	0.766	0.856	0.904	0.912	0.914	0.898	0.88	0.849	0.798	0.752	0.706	0.658	0.605	0.542
			Overfitted	0.463	0.229	0.128	0.065	0.043	0.036	0.024	0.013	0.007	0.003	0.002	0.001	0.001	0.001	0
		1	Underfitted	0.115	0.259	0.401	0.505	0.566	0.621	0.704	0.783	0.847	0.893	0.932	0.961	0.977	0.985	0.991
			Correct	0.434	0.535	0.502	0.444	0.394	0.351	0.289	0.216	0.153	0.107	0.068	0.039	0.023	0.015	0.009
			Overfitted	0.451	0.206	0.097	0.051	0.04	0.028	0.007	0.001	0	0	0	0	0	0	0
		9	Underfitted	0.489	0.783	0.916	0.97	0.986	0.988	0.996	0.997	0.999	0.999	0.999	1	1	1	1
			Correct	0.174	0.128	0.061	0.027	0.012	0.01	0.004	0.003	0.001	0.001	0.001	0	0	0	0
			Overfitted	0.337	0.089	0.023	0.003	0.002	0.002	0	0	0	0	0	0	0	0	0
	Uniform	0.25	Underfitted	0.276	0.575	0.763	0.87	0.905	0.922	0.954	0.972	0.985	0.996	0.998	0.999	0.999	1	1
			Correct	0.313	0.273	0.201	0.121	0.089	0.074	0.046	0.028	0.015	0.004	0.002	0.001	0.001	0	0
			Overfitted	0.411	0.152	0.036	0.009	0.006	0.004	0	0	0	0	0	0	0	0	0
		1	Underfitted	0.485	0.808	0.933	0.979	0.986	0.991	0.996	0.999	1	1	1	1	1	1	1
			Correct	0.18	0.109	0.055	0.017	0.011	0.008	0.004	0.001	0	0	0	0	0	0	0
			Overfitted	0.335	0.083	0.012	0.004	0.003	0.001	0	0	0	0	0	0	0	0	0
		9	Underfitted	0.576	0.883	0.958	0.988	0.993	0.998	0.999	1	1	1	1	1	1	1	1
			Correct	0.117	0.053	0.024	0.009	0.006	0.002	0.001	0	0	0	0	0	0	0	0
			Overfitted	0.307	0.064	0.018	0.003	0.001	0	0	0	0	0	0	0	0	0	0
2 weakly $p_0 = 3$	Normal	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.393	0.712	0.873	0.939	0.958	0.966	0.981	0.99	0.996	0.998	0.999	0.999	1	1	1
			Overfitted	0.607	0.288	0.127	0.061	0.042	0.034	0.019	0.01	0.004	0.002	0.001	0.001	0	0	0
		1	Underfitted	0.001	0.002	0.011	0.029	0.036	0.041	0.063	0.092	0.125	0.162	0.212	0.271	0.338	0.381	0.445
			Correct	0.377	0.719	0.856	0.905	0.92	0.929	0.918	0.894	0.866	0.832	0.783	0.727	0.661	0.618	0.554
			Overfitted	0.622	0.279	0.133	0.066	0.044	0.03	0.019	0.014	0.009	0.006	0.005	0.002	0.001	0.001	0.001
		9	Underfitted	0.193	0.438	0.608	0.726	0.773	0.799	0.852	0.887	0.924	0.95	0.967	0.982	0.988	0.991	0.993
			Correct	0.24	0.339	0.304	0.236	0.205	0.183	0.139	0.109	0.075	0.049	0.033	0.018	0.012	0.009	0.007
			Overfitted	0.567	0.223	0.088	0.038	0.022	0.018	0.009	0.004	0.001	0.001	0	0	0	0	0
	Uniform	0.25	Underfitted	0.048	0.161	0.252	0.352	0.404	0.442	0.528	0.625	0.683	0.756	0.805	0.839	0.868	0.893	0.915
			Correct	0.351	0.586	0.635	0.599	0.564	0.532	0.456	0.37	0.314	0.244	0.195	0.161	0.132	0.107	0.085
			Overfitted	0.601	0.253	0.113	0.049	0.032	0.026	0.016	0.005	0.003	0	0	0	0	0	0
		1	Underfitted	0.21	0.511	0.668	0.786	0.83	0.841	0.892	0.926	0.947	0.967	0.98	0.986	0.993	0.997	0.999
			Correct	0.243	0.287	0.259	0.186	0.155	0.148	0.103	0.07	0.051	0.033	0.02	0.014	0.007	0.003	0.001
			Overfitted	0.547	0.202	0.073	0.028	0.015	0.011	0.005	0.004	0.002	0	0	0	0	0	0
		9	Underfitted	0.363	0.723	0.864	0.92	0.936	0.953	0.974	0.99	0.994	0.995	0.997	0.998	0.998	0.999	0.999
			Correct	0.125	0.111	0.079	0.063	0.058	0.042	0.025	0.01	0.006	0.005	0.003	0.002	0.002	0.001	0.001
			Overfitted	0.512	0.166	0.057	0.017	0.006	0.005	0.001	0	0	0	0	0	0	0	0

Note: Boldface type indicates the maximum probability of correct order being selected.

Table 6. (Continued).

Model	Dist. of X	σ_0^2	Order	Criteria														
				APIC1	APIC2 (AIC)	APIC3 (KIC)	APIC4	APIClog(n) (BIC)	APIC5	APIC6	APIC7	APIC8	APIC9	APIC10	APIC11	APIC12	APIC13	APIC14
3 strongly $p_0 = 3$	Normal	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.394	0.717	0.88	0.934	0.954	0.962	0.976	0.991	0.996	0.996	0.998	0.998	1	1	1
			Overfitted	0.606	0.283	0.12	0.066	0.046	0.038	0.024	0.009	0.004	0.004	0.002	0.002	0	0	0
		1	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.376	0.71	0.869	0.939	0.955	0.969	0.98	0.988	0.993	0.996	0.996	0.998	0.999	1	1
			Overfitted	0.624	0.29	0.131	0.061	0.045	0.031	0.02	0.012	0.007	0.004	0.004	0.002	0.001	0	0
		9	Underfitted	0	0	0	0	0	0	0	0	0	0	0.002	0.003	0.007	0.007	0.008
			Correct	0.365	0.704	0.855	0.933	0.95	0.959	0.98	0.985	0.989	0.993	0.993	0.993	0.991	0.992	0.991
			Overfitted	0.635	0.296	0.145	0.067	0.05	0.041	0.02	0.015	0.011	0.007	0.005	0.004	0.002	0.001	0.001
	Uniform	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.404	0.73	0.873	0.942	0.962	0.975	0.985	0.989	0.995	0.998	0.998	0.999	0.999	0.999	1
			Overfitted	0.596	0.27	0.127	0.058	0.038	0.025	0.015	0.011	0.005	0.002	0.002	0.001	0.001	0.001	0
		1	Underfitted	0	0	0	0	0	0.001	0.001	0.001	0.002	0.005	0.008	0.014	0.021	0.032	0.042
			Correct	0.391	0.721	0.867	0.929	0.949	0.961	0.98	0.99	0.993	0.992	0.989	0.985	0.978	0.967	0.958
			Overfitted	0.609	0.279	0.133	0.071	0.051	0.038	0.019	0.009	0.005	0.003	0.003	0.001	0.001	0.001	0
		9	Underfitted	0.09	0.237	0.386	0.515	0.567	0.604	0.699	0.757	0.813	0.853	0.882	0.916	0.936	0.961	0.969
			Correct	0.324	0.506	0.501	0.442	0.405	0.376	0.294	0.237	0.184	0.146	0.117	0.083	0.063	0.038	0.031
			Overfitted	0.586	0.257	0.113	0.043	0.028	0.02	0.007	0.006	0.003	0.001	0.001	0.001	0.001	0.001	0
4 very strongly $p_0 = 5$	Normal	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.516	0.753	0.858	0.922	0.938	0.951	0.971	0.981	0.991	0.992	0.996	0.998	0.998	0.998	0.999
			Overfitted	0.484	0.247	0.142	0.078	0.062	0.049	0.029	0.019	0.009	0.008	0.004	0.002	0.002	0.002	0.001
		1	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.52	0.776	0.886	0.94	0.965	0.971	0.981	0.986	0.995	0.996	0.999	0.999	0.999	0.999	0.999
			Overfitted	0.48	0.224	0.114	0.06	0.035	0.029	0.019	0.014	0.005	0.004	0.001	0.001	0.001	0.001	0.001
		9	Underfitted	0	0	0	0	0	0	0	0	0	0.001	0.002	0.004	0.005	0.008	0.012
			Correct	0.527	0.762	0.882	0.93	0.959	0.965	0.978	0.987	0.99	0.992	0.995	0.994	0.993	0.991	0.988
			Overfitted	0.473	0.238	0.118	0.07	0.041	0.035	0.022	0.013	0.01	0.007	0.003	0.002	0.002	0.001	0
	Uniform	0.25	Underfitted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			Correct	0.546	0.797	0.902	0.944	0.957	0.965	0.981	0.986	0.99	0.993	0.998	0.999	0.999	1	1
			Overfitted	0.454	0.203	0.098	0.056	0.043	0.035	0.019	0.014	0.01	0.007	0.002	0.001	0.001	0	0
		1	Underfitted	0	0.001	0.001	0.001	0.002	0.002	0.002	0.005	0.006	0.006	0.01	0.017	0.024	0.037	0.048
			Correct	0.531	0.781	0.886	0.939	0.957	0.962	0.981	0.982	0.988	0.993	0.989	0.983	0.976	0.963	0.952
			Overfitted	0.469	0.218	0.113	0.06	0.041	0.036	0.017	0.013	0.006	0.001	0.001	0	0	0	0
		9	Underfitted	0.138	0.314	0.504	0.674	0.741	0.776	0.868	0.929	0.965	0.983	0.99	0.996	0.998	0.998	1
			Correct	0.406	0.492	0.417	0.294	0.239	0.212	0.127	0.07	0.035	0.017	0.01	0.004	0.002	0.002	0
			Overfitted	0.456	0.194	0.079	0.032	0.02	0.012	0.005	0.001	0	0	0	0	0	0	0

Note: Boldface type indicates the maximum probability of correct order being selected.