The Korean Journal of Applied Statistics (2009)
22(1), 197–204

On Information Criteria in Linear Regression Model

Man Sik Park[1]

[1]Dept. of Biostatistics, Korea University

(Received July 2008; accepted October 2008)

Abstract

In the model selection problem, the main objective is to choose the true model from a manageable set of candidate models. An information criterion gauges the validity of a statistical model and judges the balance between goodness-of-fit and parsimony; "how well observed values can approximate to the true values" and "how much information can be explained by the lower dimensional model". In this study, we introduce some information criteria modified from the Akaike Information Criterion(AIC) and the Bayesian Information Criterion(BIC). The information criteria considered in this study are compared via simulation studies and real application.

Keywords: Linear regression, information criterion, model selection.

1. Introduction

In the model selection problem confronted by statisticians, the main objective is to choose the true model from a manageable set of competing models. Since statisticians devoted their attention to this problem, they have addressed several approaches.

One way is to use an Information Criterion(IC) or model selection criterion for choosing an appropriate model. It gauges the validity of a statistical model and judges the balance between goodness-of-fit and parsimony; "how well observed values can approximate to the true values" and "how much information can be explained by the lower dimensional model". In general, an information criterion consists of a log likelihood function(L) and a complexity penalty parameter(λ) given by, for $\lambda \in \mathbb{R}$,

$$\mathrm{IC} = -2L + \lambda.$$

During the period(1960s through 2000s), a number of information criteria for the model selection have been proposed. The first information criterion accepted widespread is the Akaike Information Criterion(AIC) (Akaike, 1973) defined by

$$\mathrm{AIC} = -2L + 2p, \tag{1.1}$$

[1]Research professor, Dept. of Biostatistics & Dept. of Preventive Medicine, Medical Research Center for Environmental Toxico-Genomics and Proteomics, College of Medicine, Korea University, Seoul 136-705, Korea. E-mail: bayesia@korea.ac.kr

where p is the number of parameters considered in a model. Schwarz (1978) contrived the Bayesian Information Criterion(BIC) defined by

$$\mathrm{BIC} = -2L + p\log(n). \tag{1.2}$$

In the case that the true model has an infinite dimension, the AIC efficiently selects an approximating model with finite dimension. However, when the true model has a finite dimension, the AIC does not select an approximating model consistently. In contrast, in the same situation, the BIC is a consistent estimator but tends to select an approximating model that is too simple due to heavy complexity penalty.

To overcome the disadvantages of the AIC and the BIC, many researchers have proposed new information criteria satisfying the consistency and asymptotic efficiency. The purpose of this paper is to introduce some modified versions of the AIC and the BIC. This paper is organized as follows. In Section 2, we introduce various improved versions of information criteria. Sections 3 and 4 focus on the comparison of the criteria in linear regression models from simulation studies and real data analysis. We make a conclusion in Section 5.

2. Modified Versions of the AIC and the BIC

Let us consider the linear regression model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.1}$$

where $\mathbf{y}$ is an $n \times 1$ column vector, $\mathbf{X}$ is an $n \times p$ matrix, $\boldsymbol{\beta}$ is a $p \times 1$ parameter vector, and $\boldsymbol{\epsilon}$ is an $n \times 1$ random vector from the multivariate normal distribution, $N_n(0, \sigma^2\mathbf{I})$, where $\mathbf{I}$ is an identity matrix with order n. Under the model (2.1), the main objective for model selection is to choose an approximating model, M_α, denoted by

$$\mathbf{y} = \mathbf{X}_\alpha\boldsymbol{\beta}_\alpha + \boldsymbol{\epsilon},$$

where the dimension index set, $\alpha \in \mathscr{A}$, a class of all possible index subsets of $\{1, \ldots, p\}$, $\mathbf{X}_\alpha$ and $\boldsymbol{\beta}_\alpha$ are the matrix and the parameter vector, based on M_α, respectively. We define the number of elements of α as p_α, which is the dimension of M_α.

Now we present some modified versions of the AIC and the BIC. To overcome the inconsistency of the AIC, Hurvich and Tsai (1989) proposed the corrected AIC(AICC) defined by

$$\mathrm{AICC}_\alpha = \mathrm{AIC}_\alpha + \frac{2(p_\alpha + 1)(p_\alpha + 2)}{n - p_\alpha - 2}$$

for linear and non-linear regression and autoregressive models. They showed that the AICC is appropriate for choosing the true model as p/n increases, while the ABIC(Akaike's Bayesian Information Criterion; Akaike, 1980) performs the best as p/n decreases, where the ABIC is defined by

$$\mathrm{ABIC}_\alpha = n\log\widehat{\sigma}^2_\alpha + (p_\alpha - n)\log\left(1 - \frac{p_\alpha}{n}\right) + p_\alpha\log n + p_\alpha\log\left\{\frac{1}{p_\alpha}\left(\frac{\widehat{\sigma}^2_y}{\widehat{\sigma}^2} - 1\right)\right\},$$

where $\widehat{\sigma}^2_y$ is a sample variance of dependent variable, and $\widehat{\sigma}^2_\alpha$ and $\widehat{\sigma}^2$ are mean error variance estimators based on a candidate model, M_α and the full model, respectively.

Wu and Sepulveda (1998) proposed the Weighted-Average Information Criterion(WAIC) modified from the AICC and the ABIC. The WAIC is defined as

$$\mathrm{WAIC}_\alpha = \frac{A}{A+B}\mathrm{AICC}_\alpha + \frac{B}{A+B}\mathrm{ABIC}_\alpha = n\log\widehat{\sigma}^2_\alpha + \frac{A^2+B^2}{A+B},$$

where $A = 2n(p_\alpha+1)/(n-p_\alpha-2)$ and $B = (p_\alpha - n)\log(1-p_\alpha/n)+p_\alpha\log n+p_\alpha\log\{(\widehat{\sigma}^2_y/\widehat{\sigma}^2-1)/p_\alpha\}$. They showed that, for small sample sizes, the WAIC performs as well as the AICC and for large sample sizes, performs as well as the BIC.

By using the symmetric measure different from asymmetric Kullback-Leibler information (Kullback and Liebler, 1951), Cavanaugh (1999) proposed a new information criterion providing an asymptotically unbiased estimator of asymmetric measure between the true model and an approximating model. Kullbacks Information Criterion(KIC) is expressed as

$$\mathrm{KIC}_\alpha = \frac{\mathrm{RSS}_\alpha}{\widehat{\sigma}^2} + 3p_\alpha,$$

where RSS denotes residual sum of squares.

Rahman and King (1997) compared some popular criteria in terms of the probability of correct selection of the true model among a lot of candidate models. Rahman and King (1999) showed that, when the sample size is small and the true model has a finite dimension, the BIC has better performance than the AIC and the AIC performs better than Theil's R^2 (Theil, 1961), but the R^2 has better performance than the AIC and the AIC performs better than the BIC when the true model has an infinite dimension. Investigating this interesting pattern, they proposed the Joint Information Criterion(JIC) defined by

$$\mathrm{JIC}_\alpha = \frac{\mathrm{RSS}_\alpha}{\widehat{\sigma}^2} - \frac{1}{4}\left\{p_\alpha - \log n - n\log\left(1-\frac{p_\alpha}{n}\right)\right\}$$

with a simple average of the BIC penalty and the implicit R^2 penalty. They showed that the JIC is strongly consistent and usually selects more parsimonious models than the AIC does, but less parsimonious models than the BIC does.

Wei (1992) proposed a new information criterion, Fisher Information Criterion(FIC) defined by

$$\mathrm{FIC}_\alpha = \frac{\mathrm{RSS}_\alpha}{\widehat{\sigma}^2} + \log\left\{\det\left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right)\right\}$$

based on the Fisher information matrix. This criterion replaces the original complexity penalty proportional to the dimension of an approximating model with a new one proportional to the Fisher information contained in an approximating model.

Foster and George (1994) proposed the Risk Inflation Criterion(RIC) defined by

$$\mathrm{RIC}_\alpha = \frac{\mathrm{RSS}_\alpha}{\widehat{\sigma}^2} + 2p_\alpha\log p$$

when $n-p$ is reasonably large. The main idea behind this criterion is to use the risk function in order to eliminate the meaningless variables correctly since good performance with respect to $\widehat{\boldsymbol{\beta}}_\alpha$ has small risk inflation. In addition, George and Foster (2000) proposed the Modified Risk Inflation Criterion(MRIC) derived from the fact that the expected size of the k_{th} largest squared t-statistic is approximately $2\log(p/p_k)$.

Tibshirani and Knight (1999) proposed a new information criterion called the Covariance Inflation Criterion(CIC), a nonparametric model selection method based on adjusting the bias of in-sample estimates. The CIC is defined by

$$\mathrm{CIC}_\alpha = \frac{\mathrm{RSS}_\alpha}{\widehat{\sigma}^2} + \frac{2}{\widehat{\sigma}_y^2}\sum_{i=1}^{n}\mathrm{Cov}_0\left(y_i^*, \mathbf{x}_{\alpha,i}\widehat{\boldsymbol{\beta}}_\alpha^*\right),$$

where $\widehat{\boldsymbol{\beta}}_\alpha^*$ is the least square estimator based on permutation data, $(\mathbf{X}_\alpha, \mathbf{y}^*)$ under the model, M_α and Cov_0 is the covariance under the permutation distribution of $\mathbf{X}$ and $\mathbf{y}$. The main idea of the CIC is that the harder we fit the data, the more $\mathbf{X}\widehat{\boldsymbol{\beta}}_\alpha$ affects its own prediction, and hence the greater prediction error.

3. Simulation Studies

In this simulation study, we compare various information criteria that we have presented in the main paper in terms of their performance for the model selection. We fix the number of independent variables to 4 and the sample size to $n = 100$. The matrix $\mathbf{X}$ is the 100×5 matrix of which columns consist of the first column vector with all 1's and the other 4 independent variable column vectors. These independent variables are generated from multivariate normal distribution with mean 10 and variance 1. That is,

$$\mathbf{x}_i \sim N_{100}(\boldsymbol{\mu}_{10}, \mathbf{I}),$$

where, for $i = 1, \ldots, 4$, $\mathbf{x}_i$ is $(i+1)^{th}$ column vector in $\mathbf{X}$ and $\boldsymbol{\mu}_{10}$ is a column vector with all 10's. We also generate the errors, $\boldsymbol{\epsilon} \sim N_{100}(\mathbf{0}, \mathbf{I})$. We consider three cases of true regression coefficient parameter vectors: $\boldsymbol{\beta} = (1, 0, 0, 0.2, 0)'$, $(1, 0, 0, 0.2, 0.3)'$ and $(1, 0.4, 0, 0.2, 0.3)'$. Here, we consider all possible combinations of variables as candidate models where the intercept is included in every candidate model. So the number of all candidate models, $\{M_\alpha\}$, is $2^{(p-1)} - 1 = 15$ and the dimension index set, α, corresponding to M_α is $\alpha_1 = \{1, 5\}$, $\alpha_2 = \{1, 4\}$, $\alpha_3 = \{1, 4, 5\}$, $\alpha_4 = \{1, 3\}$, $\alpha_5 = \{1, 3, 5\}$, $\alpha_6 = \{1, 3, 4\}$, $\alpha_7 = \{1, 3, 4, 5\}$, $\alpha_8 = \{1, 2\}$, $\alpha_9 = \{1, 2, 5\}$, $\alpha_{10} = \{1, 2, 4\}$, $\alpha_{11} = \{1, 2, 4, 5\}$, $\alpha_{12} = \{1, 2, 3\}$, $\alpha_{13} = \{1, 2, 3, 5\}$, $\alpha_{14} = \{1, 2, 3, 4\}$ and $\alpha_{15} = \{1, 2, 3, 4, 5\}$. We set the number of replication to $T = 1000$. Most of the information criteria which we presented in Section 2 are easy to compute. In order to compute the GIC, we consider the preassigned level of significance, $\delta = 0.05$. We fix the number of permutation to 10 for obtaining the CIC.

In order to compare the information criteria used in this simulation study, we need the standard measures. We consider the true model error(ME) and empirical probabilities of correct selection, over-selection, under-selection and biased selection as the standard measures. The ME is defined by

$$\mathrm{ME} = \left\|\mathbf{X}_{\alpha_0}\boldsymbol{\beta}_{\alpha_0} - \mathbf{X}_{\widehat{\alpha}}\widehat{\boldsymbol{\beta}}_{\widehat{\alpha}}\right\|^2, \tag{3.1}$$

where α_0 is is the dimension index set of the true model, M_{α_0} and $\widehat{\alpha}$ is the dimension index set of the estimated model, $M_{\widehat{\alpha}}$, by the particular information criterion. The empirical probabilities are defined by

$$\Pr(\widehat{\alpha} = \alpha_0) = \frac{1}{T}\sum_{t=1}^{T} I(\widehat{\alpha}_t = \alpha_0), \tag{3.2}$$

Table 3.1. True model error(ME) and empirical probability of selecting models under $\boldsymbol{\beta} = (1, 0, 0, 0.2, 0)'$

	ME	Over	Under	Biased	Correct	Incorrectly selected models		
						$\widehat{\alpha}_a$	$\widehat{\alpha}_b$	$\widehat{\alpha}_c$
AIC	.042	.335	.000	.148	.517	10(.096)	3(.092)	6(.077)
BIC	.036	.064	.000	.199	.737	4(.075)	8(.063)	1(.061)
AICC	.038	.160	.000	.182	.658	4(.069)	1(.056)	8(.054)
ABIC	.038	.179	.000	.167	.654	4(.061)	1(.054)	6(.052)
WAIC	.038	.159	.000	.178	.663	4(.068)	1(.055)	8(.052)
KIC	.038	.180	.000	.177	.643	4(.065)	10(.058)	1(.054)
JIC	.039	.199	.000	.173	.628	10(.065)	4(.063)	3(.058)
FIC	.037	.060	.000	.222	.718	4(.080)	1(.071)	8(.070)
RIC	.038	.148	.000	.186	.666	4(.069)	1(.059)	8(.056)
MRIC	.034	.013	.000	.209	.778	4(.078)	8(.066)	1(.065)
CIC	.043	.507	.000	.140	.353	10(.117)	3(.115)	6(.110)
GIC	.034	.007	.000	.210	.783	4(.078)	8(.067)	1(.065)

$$\Pr(\widehat{\alpha} \subsetneq \alpha_0) = \frac{1}{T}\sum_{t=1}^{T} I(\widehat{\alpha}_t \subset \alpha_0,\ \widehat{\alpha}_t \neq \alpha_0), \tag{3.3}$$

$$\Pr(\widehat{\alpha} \supsetneq \alpha_0) = \frac{1}{T}\sum_{t=1}^{T} I(\widehat{\alpha}_t \supset \alpha_0,\ \widehat{\alpha}_t \neq \alpha_0), \tag{3.4}$$

$$\Pr(\widehat{\alpha} \neq \alpha_0) = \frac{1}{T}\sum_{t=1}^{T} I(\widehat{\alpha}_t \not\subseteq \alpha_0,\ \widehat{\alpha}_t \not\supseteq \alpha_0), \tag{3.5}$$

where $I(\cdot)$ is an indicator function. The Equations (3.2) through (3.5) are regarded as the probabilities of correct selection, under selection, over selection and biased selection, respectively. Now we interpret the simulation results based on the linear regression model. Through the simulation studies, we focus on the true model error(ME) (3.1) of an approximating model selected by information criteria, and the probabilities of selecting the true model (3.2) and over-fitted models (3.4). Tables 3.1 through 3.3 show the simulation results. $\widehat{\alpha}_a$, $\widehat{\alpha}_b$ and $\widehat{\alpha}_c$ mean the dimension index sets of the incorrectly estimated models including over-dimensional models, under-dimensional models and biased models, where $\widehat{\alpha}_a$ is the set with the highest empirical probability, $\widehat{\alpha}_b$ is the second and $\widehat{\alpha}_a$ is the third. In addition, the probabilities of selecting the incorrect models with these sets are in the parentheses.

Table 3.1 shows the simulation results for $\boldsymbol{\beta} = (1, 0, 0, 0.2, 0)'$. The MRIC and the GIC have the smallest ME and the highest empirical probabilities of selecting the true model with $\alpha_2 = \{1, 4\}$. However, the CIC has the lowest empirical probability of selecting a biased model although its probability of selecting an over-dimensional model is the largest. The reason is that the models with the first (.117) and third (.115) highest probabilities of selecting incorrect models are both over-dimensional models, that is, models with $\alpha_{10} = \{1, 2, 4\}$ and with $\alpha_6 = \{1, 3, 4\}$. The ABIC, the AICC, the WAIC, the KIC and the RIC have the same ME and have very similar performance in terms of the probability of selecting an over-dimensional model and the pattern of the incorrect selections.

Tables 3.2 and 3.3 show the simulation results for $\boldsymbol{\beta} = (1, 0, 0, 0.2, 0.3)'$ and $\boldsymbol{\beta} = (1, 0.4, 0, 0.2, 0.3)'$, where the true dimension index sets are $\alpha_3 = \{1, 4, 5\}$ and $\alpha_{11} = \{1, 2, 4, 5\}$, respectively. The AIC

Table 3.2. True model error(ME) and empirical probability of selecting models under $\beta = (1, 0, 0, 0.2, 0.3)'$

	ME	Over	Under	Biased	Correct	Incorrectly selected models		
						$\widehat{\alpha}_a$	$\widehat{\alpha}_b$	$\widehat{\alpha}_c$
AIC	.047	.256	.153	.064	.527	1(.140)	11(.124)	7(.109)
BIC	.057	.040	.499	.035	.426	1(.392)	2(.107)	7(.021)
AICC	.050	.095	.310	.056	.539	1(.261)	7(.052)	2(.049)
ABIC	.051	.096	.344	.044	.516	1(.292)	2(.052)	7(.047)
WAIC	.050	.093	.327	.052	.528	1(.279)	7(.049)	2(.048)
KIC	.049	.117	.275	.063	.545	1(.235)	7(.058)	11(.052)
JIC	.049	.136	.254	.063	.547	1(.217)	7(.064)	11(.063)
FIC	.058	.038	.515	.035	.412	1(.405)	2(.110)	7(.022)
RIC	.051	.091	.332	.056	.521	1(.279)	2(.053)	7(.050)
MRIC	.068	.010	.727	.015	.248	1(.557)	2(.170)	7(.006)
CIC	.048	.383	.113	.111	.393	11(.173)	7(.140)	1(.098)
GIC	.072	.002	.799	.015	.184	1(.607)	2(.192)	5(.005)

Table 3.3. True model error(ME) and empirical probability of selecting models under $\beta = (1, 0.4, 0, 0.2, 0.3)'$

	ME	Over	Under	Biased	Correct	Incorrectly selected models		
						$\widehat{\alpha}_a$	$\widehat{\alpha}_b$	$\widehat{\alpha}_c$
AIC	.054	.132	.192	.051	.625	9(.153)	15(.132)	13(.036)
BIC	.073	.020	.510	.032	.438	9(.338)	10(.098)	8(.055)
AICC	.062	.046	.361	.042	.551	9(.257)	10(.070)	15(.046)
ABIC	.064	.034	.403	.037	.526	9(.287)	10(.074)	15(.034)
WAIC	.063	.037	.386	.040	.537	9(.277)	10(.072)	15(.037)
KIC	.060	.068	.319	.042	.571	9(.231)	15(.068)	10(.063)
JIC	.059	.075	.301	.044	.580	9(.219)	15(.075)	10(.060)
FIC	.074	.018	.526	.031	.425	9(.350)	10(.096)	8(.059)
RIC	.063	.039	.373	.042	.546	9(.265)	10(.072)	15(.039)
MRIC	.094	.005	.676	.015	.304	9(.360)	8(.155)	10(.110)
CIC	.057	.245	.173	.082	.500	15(.245)	9(.120)	13(.060)
GIC	.108	.001	.835	.009	.155	9(.427)	8(.198)	10(.142)

has the smallest ME and the second lowest probability of selecting an under-dimensional model for both of the cases. The AIC also has the second highest probability of selecting an over-dimensional model. However, the BIC has the reverse result to the AIC. From these results, we can conclude that the AIC tends to select an over-dimensional model, but the BIC tends to select an under-dimensional model due to a heavy complexity penalty. The modified versions have almost the same ME even though their empirical probabilities are slightly different. As can be seen, their probabilities of selecting an over-dimensional model are less than the AIC's, but larger than the BIC's, and those of selecting an under-dimensional model are less than the BIC's, but larger than the AIC's. Especially, the ABIC, the AICC and the WAIC have an effect on decreasing the inconsistency of the AIC and the KIC and the JIC are effective to treat inefficiency of the BIC.

In the simulation studies, we have discussed the performance of various information criteria. Since the earliest information criteria was used in the model selection problem, a lot of adjusted criteria have been motivated and proposed in order to overcome their disadvantages.

Table 4.1. Variable selection with Boston Housing data

	L	P	C	R	D	N	B	R	T	C	Z	I	A
	S	T	R	M	S	X	L	A	X	H	N	D	G
AIC	×	×	×	×	×	×	×	×	×	×	×	·	·
BIC	×	×	×	×	×	×	×	×	×	×	·	·	·
AICC	×	×	×	×	×	×	×	×	×	×	×	·	·
ABIC	×	×	×	×	×	×	×	×	×	×	·	·	·
WAIC	×	×	×	×	×	×	×	×	×	×	·	·	·
KIC	×	×	×	×	×	×	×	×	×	×	×	·	·
JIC	×	×	×	×	×	×	×	×	×	×	×	·	·
FIC	×	×	×	×	×	×	×	×	×	×	·	·	·
RIC	×	×	×	×	×	×	×	×	×	×	·	·	·
MRIC	×	×	×	×	×	×	×	×	×	×	·	·	·
CIC	×	×	×	×	×	×	×	×	×	×	×	·	·
GIC	×	×	×	×	×	×	×	×	×	·	·	·	·
FWD	×	×	×	×	×	×	×	×	×	×	×	×	·
BWD	×	×	×	×	×	×	×	×	×	×	×	·	·
STP	×	×	×	×	×	×	×	×	×	×	×	·	·

Notes. LS: percentage lower status of the population; PT: pupil-teacher ratio by town; CR: per capita crime rate by town; RM: average number of rooms per dwelling; DS: weighted distances to five Boston employment centers; NX: nitric oxides concentration (parts per 10 million); BL: $1000 \times (B - 0.63)^2$ where B is the population of blacks by town; RA: index of accessibility to radial highways; TX: full-value property-tax rate per \$10,000; CH: Carles River dummy variable (1 bounds river; 0 otherwise); ZN: proportion of residential land zoned for lots over 25,000 sq.ft.; ID: proportion of non-retail business acres per town; AG: proportion of owner-occupied units built prior to 1940; FWD: Forward selection; BWD: Backward elimination; STP: Stepwise selection

4. Real Application

We now represent a real data analysis using the Boston Housing data. This data contain 14 variables (including a dependent variable, median value of owner-occupied homes in \$1000's) that describe the 506 census tracks in the Boston SMSA in 1970, excluding tracks with no housing units (Belsley *et al.*, 1980; Breiman *et al.*, 1984). Since there are 13 independent variables, it is impossible to consider all possible combinations of these variables, $2^{13} - 1$, as candidate models. So, we first set up the order of the 13 independent variables by Stepwise selection method before doing variable selection by information criteria. Traditional variable selection methods(Forward selection, Backward elimination and Stepwise selection) in the multiple linear regression model are also considered.

Table 4.1 illustrates which variables among the 13 independent variables are selected by each of the criteria. Here, "×" means that the corresponding variable is selected by an information criterion. The AIC, the AICC, the KIC, the JIC and the CIC select the model with 11 variables except the variables ID and AG, but the other criteria except the GIC select the model with 10 variables except ZN, ID and AG. As mentioned above in Section 3, the JIC has an effect on treating the over-fitting problem that the BIC usually has while the ABIC and the WAIC is effective to the under-fitting problem of the AIC. In addition, Backward elimination and Stepwise selection choose the model with 11 variables, but Forward selection chooses the model with 12 variables. From Table 3.2 through Table 4.1, most of the criteria that select the model with 10 variables have the under-fitting probability at least 0.33 and most of the criteria choosing the model with 11 variables have the under-fitting probability at most 0.36.

5. Conclusion

We have discussed a class of information criteria for model selection. In general, an approximating model is selected by minimizing an information criterion, which consists of log-likelihood function and complexity penalty parameter. Hence, the researchers have tried to select a model with the greater information about observed data but the lower dimensional structure. Each criterion has the specific complexity penalty, which leads to different performances in different situations. The results from simulation study and real application show that most of the information criteria considered in this paper are superior to the AIC and the BIC, but the AIC and the BIC still have advantage due to the ease of computation.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *Second International Symposium on Information Theory*, 267–281.

Akaike, H. (1980). *Likelihood and the Bayes procedure, Bayesian Statistics*, University Press, Valencia Spain.

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, New York.

Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. (1984). *Classification and Regression Trees*, Chapman & Hall/CRC, New York.

Cavanaugh, J. E. (1999). A large-sample model selection criterion based on Kullback's symmetric divergence, *Statistics and Probability Letters*, **42**, 333–343.

Foster, D. P. and George, E. I. (1994). The risk information criterion for multiple regression, *The Annals of Statistics*, **22**, 1947–1975.

George, E. I. and Foster, D. P. (2000). Calibration and empirical Bayes variable selection, *Biometrika*, **87**, 731–747.

Hurvich, C. M. and Tsai, C. L. (1989). Regression and time series model selection in small samples, *Biometrika*, **76**, 297–307.

Kullback, S. and Liebler, R. A. (1951). On information and sufficiency, *The Annals of Mathematical Statistics*, **22**, 79–86.

Rahman, M. S. and King, M. L. (1997). Probability of correct identification in model selection, In *Proceedings of the Econometric Society Australasian Meeting*, 631–640, University of Melbourne, Melbourne.

Rahman, M. S. and King, M. L. (1999). 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.

Theil, H. (1961). *Economic Forecasts and Policy*, (Second Ed.), North-Holland, Amsterdam.

Tibshirani, R. and Knight, K. (1999). The covariance inflation criterion for adaptive model selection, *Journal of the Royal Statistical Society, Series B*, **61**, 529–546.

Wei, C. Z. (1992). On predictive least squares principles, *The Annals of Statistics*, **20**, 1–42.

Wu, T. J. and Sepulveda, A. (1998). The weighted average information criterion for order selection in time series and regression models, *Statistics and Probability Letters*, **39**, 1–10.