New Criteria of Model selection and model averaging in linear regression models

Magda Mohamed Mohamed Haggag

Department of Statistics, Mathematics, and Insurance, Faculty of Commerce, Damanhour University, Egypt

Email address:

magda.hagag@damanhour.edu.eg, magmhag@yahoo.com

Abstract: Model selection is an important part of any statistical analysis. Many tools are suggested for selecting the best model including frequentist and Bayesian perspectives. There is often a considerable uncertainty in the selection of a particular model to be the best approximating model. Model selection uncertainty arises when the data are used for both model selection and parameter estimation. Bias in estimators of model parameters often arise when data based selection has been done. Therefore, model averaging of the parameter estimators will be done to alleviate the bias in model selection in a set of candidate models, by combining the information from a set of candidate models. This paper is two-fold, new criteria of model selection are proposed based on different averages of AIC, BIC, AICc, and HQC. Also, model averaging is introduced to compare the parameter estimators in model averaging with the ones in model selection. Two Simulation studies are considered, the first is for model selection and showed that the new proposed criteria are lies between some of the known criteria such as AIC, BIC, AICc, and HQC, and so they can be used as new criteria of model selection. The second simulation study is for model averaging and showed that the parameter estimators have less bias and less predicted mean square error (PMSE) compared with the parameter estimators in model selection.

Keywords: AIC, BIC, AICc, HQC, Kullback-Leibler (K-L) Distance, Model Averaging, Model Selection

1. Introduction

Model selection is the task of choosing from a candidate set of models, the one that fits the input data. Kullback and Leibler (1951) derived information measure referred as the Kullback-Leibler (K-L) distance. The K-L distance can be defined as a directed distance between two models. (Kullback (1959)). It is the most fundamental of all information measures and it is the logical basis for model selection. Estimation of Kullback-Leibler information is a key to derive the information criterion which is widely used for selecting a statistical model as defined by Akaike. (AKaike, 1973).

Based on a Akaike's (1973) information criterion (AIC), many model-selection criteria have been proposed. Takeuchi (1976) introduced a very general derivation of an information criterion called Takeuchi's information criterion (TIC). The AIC and TIC are designed for the likelihood or quasi-likelihood context, and they select the fitted model whose densities are close to the true density, which is a broad and useful feature. Sugiura (1978) proposed a corrected version

of AIC denoted by AIC_C. The advantage of using AIC_C relies on its superior performance as a selection criterion in small-sample applications. Lebreton et al. (1992) suggested simple modifications to AIC and AIC_C for overdispersed count data denoted by QAIC and $QAIC_C$, respectively.

Schwarz (1978) introduced the Bayesian information criterion (BIC) as a competitor to the AIC. He derived BIC to serve as an asymptotic approximation to a transformation of the Bayesian posterior probability of a candidate model. The computation of BIC is based on the empirical log-likelihood and does not require the specification of priors. AIC and BIC share the goodness-of-fit term, but the penalty term of BIC is potentially much more stringent than the penalty term of AIC. Thus, BIC tends to choose fitted models that are more parsimonious than those favored by AIC.

Hannan and Quinn (1979) suggested a criterion for identifying an autoregressive model denoted by HQC(p), and the adjusted version of it can be applied to regression models.(See Al-Saubaihi (2007)).

Cavanaugh (1999), proposed a new class of criterion for linear model selection denoted by KIC, KIC_C, and MKIC as

analogue to AIC, AIC$_C$, MAIC respectively. He illustrated its performance in a simulation study for choosing an order of autoregression.

When using model selection criteria, the model is fitted under a specific parametric probability distribution. The fitted model (or the approximated model) considers two types of risks (errors), the risk of modeling and the risk of estimation. A risk of modeling is considered in terms of the incorrect specification of the probability model. A risk of estimation is considered when estimating the true parameter vector in the restricted parameter space of the model. The last risk contains two components, the variance and the bias. The variance can be interpreted as a penalty of the size of the parameter space of the model, and the bias is the penalty for the distance between the restricted parameter space and the true parameter vector the penalty for the distance between the restricted parameter space and the true parameter vector of the model. The overall risk (which includes both the risks of modeling and estimation) is aimed to be minimized. Model selection criteria are the estimators of the overall risk of a model under the maximum-likelihood estimation.

The aim of this paper is to provide a new class of criteria based on the averages of the most popular information criteria such that AIC, BIC, HQC, and AICc. The proposed criteria are applied to regression models from both theoretical and empirical point of view. Also, model averaging is introduced as a preferred method to best model inference in the regression setting.

This paper is organized as follows. Section (2) reviews the Kullback-Leibler (K-L) information and some information criteria as estimators of the Kullback-Leibler information. Section (3) introduces the model weights of model selection criteria. Section (4) presents the statistical inference based on model averaging. Section (5) considers the proposed class of information criteria. Section (6) introduces simulation studies. Section (7) is devoted to the conclusion of this paper.

2. The Cullback-Leibler Information

Let $Y_1, \ldots\ldots, Y_n$ are iid observations of a (px1) random vector Y. Let $f(y)$ be a real probability distribution function which has an infinite number of parameters, and let $g(y;\theta)$ be an approximating model (a probability distribution), and θ are the parameters in the approximating model g that must be estimated from the data. A set of M approximated models (candidate models) for the representation of the data will be considered and denoted by,

$$\{g_i(y;\theta): i = 1,...., M\}. \qquad (1)$$

Information lost when approximating model is used to approximate the full real model $f(\cdot)$. The aim is to find an approximating model that loses as little information as possible, which means minimizing the distance between a candidate model $g(y;\theta)$ and the real model $f(y)$. A candidate model is a preferred one if it is the closest model to

the true one.

The kullback-Leibler information (K-L) (or distance) is considered a popular measure of closeness between the two models f and g. The K-L information is defined for continuous functions as follows,

$$I(f,g) = \int f(y)\ln\left[\frac{f(y)}{g(y;\theta)}\right]dy , \qquad (2)$$

where ln denotes by natural logarithm. It is always nonnegative and it is only equal to zero iff $f = g$. Also, $I(f,g) \neq I(g,f)$ which implies that the K-L information is not the real distance.

Kullback and Leibler (1951) developed the quantity $I(f,g)$ in (2) from the information theory, and so they used the notation $I(f,g)$. It is the information lost when using the model g as an approximate model to f.

The K-L distance in the case of discrete distributions such as Poisson, binomial, or multinomial is defined as,

$$I(f,g) = \sum_{i=1}^{K} f_i(y)ln\left[\frac{f_i(y)}{g_i(y;\theta)}\right], \qquad (3)$$

where K are possible outcomes of the random variable Y.

The expression for K-L distance in (2) can be written equivalently as,

$$I(f,g) = \int f(y)\ln f(y)dy - \int f(y)\ln g(y;\theta)dy . \qquad (4)$$

The two terms on the right hand in (4) are the statistical expectations with respect to the real function $f(y)$. Therefore, the K-L information in (4) can be written as follows,

$$I(f,g) = E[\ln f(y)] - E[\ln g(y;\theta)]. \qquad (5)$$

The first expectation on the right hand side of (5) is a constant that depends on the unknown true function f, which is clearly not known. Therefore, computing the second expectation, $E[\ln g(y;\theta)]$ will give a relative measure of the distance between f and g. (See Bozdogan (1987); and Kapur and Kesavan (1992)). Therefore, the quantity $E[\ln g(y;\theta)]$ is the quantity of interest. Minimizing $I(f,g)$ is equivalent to maximizing of $E[\ln g(y;\theta)]$ with respect to θ. Since θ is unknown, then model selection criterion will be changed to minimizing the expected estimated K-L information based on θ. So, the concept of selecting a model based on minimizing the estimated Kullback-Leibler information will be as follows,

$$\hat{I}(f,g) = E[\ln f(y)] - E[\ln g(y;\hat{\theta})] = \text{constant} - E[\ln g(y;\hat{\theta})], \qquad (6)$$

where constant= $E[\ln f(y)]$, since $E[\ln f(y)]$ depends only on the unknown true distribution. So, this unknown term is treated as a constant. The term $\hat{I}(f,g) - \text{constant}$, is a relative directed distance between the two models f and g

when estimating $E[\ln g(y;\theta)]$. Thus $E[\ln g(y;\theta)]$ becomes the quantity of interest.

In the following section, some of the most common information crtiteria are presented as estimators of the K-L information criterion. It is assumed that the model parameters are unknown and will be estimated using Fisher's maximum likelihood method, and a log-likelihood function, $ln\left[L(\theta\backslash y)\right] = ln(g(y;\theta))$, is associated with each probability model in the set of M candidate models. (See Anderson et al. (1998)).

2.1. Akaike's Information Criterion (AIC)

Akaike information criterion (AIC), developed by Akaike (1973, 1974) is devoted to estimate the expected Kullback-Leibler information in (6) between the model generating the data and a fitted candidate model.

Akaike (1973, 1974) showed that the critical term in model selection is the second term of (6), i.e., $-E\left[ln g(y;\hat{\theta})\right]$. He found a relationship between this term and the maximized log-likelihood, which has allowed practical and theoretical advances in model selection and analysis of complex data. (See Stone (1982); Shibata (1983); and Deleauw (1992)).

Akaike (1973) showed that the maximum log-likelihood is a biased upword estimator of the model selection criterion. He found that under certain conditions, this bias is approximately equal to k, the number of estimable parameters in the approximating model. Therefore, an unbiased estimator of the expected K-L information will be,

$$\hat{E}\left[\hat{I}(f,g)\right] = ln\left[L(\hat{\theta}\backslash y)\right] - k, \qquad (7)$$

where $L(\cdot)$ is the likelihood function which is equivalent to,

$$L(\hat{\theta}\backslash y) - k = \text{constant} - \hat{E}\left[\hat{I}(f,g)\right],$$

or

$$-L(\hat{\theta}\backslash y) + k = -\text{constant} + \hat{E}\left[\hat{I}(f,g)\right],$$

or

$$-L(\hat{\theta}\backslash y) + k = \text{the estimated relative expected K-L distance. (8)}$$

The author Akaike (1973) introduced an information criterion denoted AIC by multiplying (8) by -2 to get,

$$AIC = -2ln\left[L(\hat{\theta}\backslash y)\right] + 2k, \qquad (9)$$

where the first term of the right hand side represents the bias but the second term corresponds to the for estimated model.

Thus, as shown in (9), AIC is defined without specific reference to a " true model " since the expectation of the logarithm of f (x) drops out as a constant independent of the data.

Among candidate models, the model will be selected which yields the smallest value of AIC. This means that one should select the fitted approximated model, which on the average, to be closest to the unknown f.

If the data are iid with normally distributed errors, then the AIC in (9) using the least squares estimator will be:

$$AIC = n ln\left[\frac{RSS}{n}\right] + 2k, \qquad (10)$$

where RSS is the residual sum of squares of the fitted model.

The greatest advantage of AIC is its potential in model selection (i.e., variable selection), since AIC is independent of the order in which models are computed. Also, the ordered models according to AIC can be used to incorporate model uncertainty to obtain robust estimates. On the other hand, according to AIC, the model is good if it is in the set of candidate models and using the same data which have generated it. (See Anderson et al. (1998)).

2.2. The Corrected Akaike's Information Criterion (AIC)

Sugiura (1978); and Sakamoto eta al. (1986) found that AIC may not perform well if there are too many parameters with respect to the sample size. As a result, AIC exhibits a potentially high degree of negative bias. Sugiura (1978) derived a second version of AIC called the corrected AIC. Hurvich and Tsai(1989), studied the corrected AIC for small sample bias adjustment. The corrected AIC is denoted by AICc and defined as:

$$AICc = -2ln\left[L(\hat{\theta}\backslash y)\right] + 2k\left[\frac{n}{n-k-1}\right], \qquad (11)$$

where $\left[\dfrac{n}{n-k-1}\right]$ is the correction factor. Eq.(11) can be written as:

$$AICc = -2ln\left[L(\hat{\theta}\backslash y)\right] + 2k + \frac{2k(k+1)}{n-k-1},$$

or

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}. \qquad (12)$$

The AICc is an unbiased criterion for linear regression if the candidate models include the true one. AICc has an additional bias correction term, $\dfrac{2k(k+1)}{n-k-1}$ and must be used when the ratio (n/k) is small. If the ratio (n/k) is sufficiently large, then AIC and AICc are similar. (See Hurvich and Tsai (1989); (1991); and (1995)).

While the AICc estimates the expected discrepancy with less bias than the AIC, the AIC is more universally applicable than AICc because the derivation of AIC is quite general but the derivation of AICc relies on the form of candidate model.

(See Cavanaugh, 1997). AIC is asymptotically efficient and it is not consistent, but AICc is an efficient criterion and it is asymptotically consistent. (See Bozdogan 1987; and 1994).

2.3. Takeuchi's Information Criterion (TIC)

Takeuchi (1976) proposed a general derivation of an information criterion, without taking expectation with respect to g. this criterion is called Takeuchi's information criterion (TIC) and it is useful when the candidate model is not close approximation to f. TIC is defined as follows:

$$TIC = -2\ln\left[L\left(\hat{\theta}\right)\right] + 2tr\left(\hat{J}^{-1}\hat{I}\right), \qquad (13)$$

where,

$$\hat{J} = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta^T}\ln g\left(y_i;\hat{\theta}\right), \qquad (14)$$

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta}\ln g\left(y_i;\hat{\theta}\right)\frac{\partial}{\partial\theta}\ln g\left(y_i;\hat{\theta}\right)^T\right). \qquad (15)$$

This does not require that g is correctly specified. If $g \equiv f$, then $\hat{J} = \hat{I}$. Hence $tr\left[\hat{J}^{-1}\hat{I}\right] = k$. Also, if g is close to f, then $tr\left[\hat{J}^{-1}\hat{I}\right] \approx k$. TIC is rarely used in practice because it needs a very large sample size to obtain the two estimated matrices $\hat{J}$ and $\hat{I}$.

The AIC and TIC are designed for the likelihood (or quasi-likelihood) models. The model needs to be a conditional density, not just a conditional mean or a set of moment conditions. Both AIC and TIC select models whose densities are close to the true one. (See Takeuchi (1987)).

2.4. Mallows' C_p Criterion

Mallows (1973) proposed the following predictive statistics, for (p-1) variable model (M) to MSE of full model and penalizes for the number of variables,

$$C_p\left(M\right) = \frac{SSE\left(M\right)}{MSE\left(full\right)} - n + 2p\left(M\right), \qquad (16)$$

Where $SSE\left(M\right) = \left\|Y - \hat{Y}_M\right\|^2$, is the SSE of the model M, $MSE\left(full\right) = \frac{SSE\left(full\right)}{df\left(full\right)}$, is the estimate of σ^2, and $p(M)$ is the number of predictors in model M.

The basic idea of Mallows' C_p criterion is to compare the predictive ability of subset models to that of full model. Full model, generally, is best for prediction, but if multicollinearity exist, then the parameter estimates will not be useful. Subset of full model that doesn't have as much collinearity will be better as long as there is no substantial bias in the predicted values to the full model.

A model is considered good if $C_p\left(M\right) \leq p$. The benefit of C_p is that it can be used to select the model size and then get

a good model which contains as few variables as possible. Mallows' C_p is used primarily for variable selection in linear regression, and it is the most popular criteria for this purpose. C_p is almost a special case of AIC as shown in (9).

2.5. The Bayesian Information Criterion (BIC)

Schwarz (1978) proposed the Baysian information criterion as follows:

$$BIC = -2\ln\left[L\left(\hat{\theta} \backslash y\right)\right] + K\ln\left(n\right), \qquad (17)$$

As usually used, BIC is computed for each model and the model with the smallest criterion value is selected.

Eq.(17) is the same as AIC but the penalty term is greater. BIC tends to choose simpler models, but behaves quite differently than AIC. It is also based on different assumptions. BIC assumes that the candidate models contain the true model and the model most likely to be true is found in the Bayesian sense. (See Burnham and Anderson (2002)).

2.6. Hannan and Quinn Criterion (HQC)

Hannan and Quinn (1979) introduced a criterion for identifying an autoregressive model. The adjusted version for regression model can be shown as follows:

$$HQC = -2\ln\left[L\left(\hat{\theta} \backslash y\right)\right] + 2K\ln\left(\ln\left(n\right)\right). \qquad (18)$$

The best model is the model which corresponds to minimum HQC. It is shown that HQC like BIC but unlike AIC in which it is not an estimator of Kullback-Leibler distance and also it is not an asymptotically efficient criterion.

3. Selecting The Model Weights

Selecting the best model depends on a set of candidate models which are fitted and then finding the corresponding information criteria such as AIC, AICc, TIC, BIC, HQC, and Cp. There will always be information lost due to choosing one of the candidate models to represent the true model. Let AIC values of the candidate models denoted by AIC$_i$, i=1,.......,R, and let AIC$_{min}$ be the minimum of those values, then the AIC differences will be denoted by Δ_i, i=1,.......,R, and defined as,

$$\Delta_i = AIC_i - AIC_{min}, \text{ i=1,........,R,} \qquad (19)$$

over all the candidate R models. (See Burnham and Anderson (2002)).

The Δ_i's values allow a quick comparison and ranking of the candidate models. The fitted mode i will be best if it has $\Delta_i \equiv \Delta_{min} \equiv 0$. Anderson et al. (1998) introduced some rough rules of thumb as follows:

- A model with a Δ_i within 0-2 units of the best model, i.e., $0 \leq \Delta_i \leq 2$ will have substantial support.

- A model with a Δ_i within 4-7 units of the best model, i.e., $4 \leq \Delta_i \leq 7$ will have less support.

- A model with a Δ_i greater than 10 units of the best model, i.e., indicates that the model is worse, will have no support, and can be omitted from further consideration.

Akaike weights are the weights of evidence in favor of model **i** being the actual best model for the set of R models. Akaike weight is denoted by w_i and defined as follows,

$$w_i = \frac{exp\left(-\frac{1}{2}\Delta_i\right)}{\sum\limits_{r=1}^{R} exp\left(-\frac{1}{2}\Delta_r\right)}, \qquad (20)$$

where Δ_i as defined in (19). Akaike weights for all models combined should add to 1.

Akaike weights are used for finding the following:
- The probability that the candidate model is the best model.
- The relative strength of evidence (evidenc ratio).
- The variable selection, i.e., finding the variables which have the greatest influence.
- The model averaging, as will be discussed in the following section.

The weight W_i depend on the entire set of R models, so that if a model is added or dropped during analysis, the W_i's must be recomputed for all the models in the new defined set of models. (See Burnham and Anderson (2002); and (2004)).

Burnham and Anderson (2002), showed that given any set of prior probability τ_i, generalized Akaike weights are defined as follows,

$$w_i = \frac{L(g_i \setminus y)\tau_i}{\sum\limits_{r=1}^{R} L(g_r \setminus y)\tau_r}, \qquad (21)$$

where $L(g_i \setminus y)$ is the likelihood function of model g_i given the data y.

4. Inference Based on Model Averaging

4.1. Averaging of Model Parameters

A large number of closely related models can be obtained in linear regression through all subsets selection. In the case that no single model is superior to some other models in a set of models, then inference based on a single model will cause risks. That is estimated model may vary from data set to another, causes more highly variable. In this case model averaging gives a relatively much more stabilized inference. (See Burnham and Anderson (2002); and 2004)).

Burnham and Anderson (2002) showed that a model average estimator of the regression parameters will have reduced bias and sometimes have better precision compared to the parameter estimators from the selected best model.

If there are R models in a set of regression models, then each model have the parameter β to be estimated. Each model i (i=1,.....,R) allows an estimate of the parameter β_i. If the estimators $\hat{\beta}_i$'s differ across the R models, then there will be a risk in depending on a single selected model. In this case, a weighted estimator of these estimators may be computed as follows,

$$\hat{\bar{\beta}}_j = \frac{\sum\limits_{i=1}^{R} w_i I_j(g_i)\hat{\beta}_{j,i}}{w_+(j)}, \qquad (22)$$

$$w_+(j) = \sum\limits_{i=1}^{R} w_i I_j(g_i), \qquad (23)$$

and,

$$I_j(g_i) = \begin{cases} 1 & \text{if the regressor } x_j \text{is in model } g_i. \\ 0 & \text{otherwise.} \end{cases}, \qquad (24)$$

Where $\hat{\beta}_{j,i}$ refers to the estimator of β_j based on model g_i, w_i is the Akaike weight of model g_i, and $w_+(j)$ is the sum of the Akaike weights over all models in the set where regressor j exists in the model.

The second model averaging is to consider that the variable x_j is in every model. In this case, in some models β_j is set to zero rather than being unknown. Conditional model selection of g_i gives a biased estimator of $\hat{\beta}_{j,i}$. Therefore, a second model-averaged estimator is proposed as follows,

$$\hat{\bar{\beta}} = w_+(j)\hat{\bar{\beta}}_j, \qquad (25)$$

where $\hat{\bar{\beta}}$ denotes a second-order average estimator derived from model averaging over all R models. In this case, x_j is not in some models in the set, so $\hat{\beta}_{j,i} \equiv 0$ is used instead of the estimator $\hat{\beta}_{j,i}$. It is clear that $w_+(j)$ shrinks the conditional $\hat{\bar{\beta}}_j$ to zero and thus ameliorates the model selection bias of $\hat{\bar{\beta}}_j$.(See Burnham and Anderson (2002); and (2004)).

In this work, conditional inference which is inference based on a best model is considered and unconditional inference, which is based on all models are applied for comparison. A model averaged estimator will have a reduced bias and a better precision compared to the estimator

computed from the selected best model. (See Burnham and Anderson (2002); and (2004) for regression models; and Leamer (1978); and Hoeting et al. (1999) for Bayesian viewpoint).

4.2. Unconditional Confidence Intervals

A *(1-α)100%* unconditional confidence interval will be set. There are two general approaches:
- The first approach is based on the bootstrap used by Buckland et al. (1997).
- The second approach is based on the analysis results of the one data set.
- The second analytical approach will be used and applied in this work.

The conditional confidence interval is given by:

$$\hat{\beta}_i \pm Z_{1-\frac{\alpha}{2}} \, \hat{se}(\hat{\beta}_i), \qquad (26)$$

where α is the level of significance, $\hat{se}(\hat{\beta}_i) = \sqrt{\hat{var}(\hat{\beta}_i)}$, and $\hat{var}(\hat{\beta}_i)$ is the estimated conditional variance of the selected model. Also, the following formula can be used for the conditional confidence interval,

$$\hat{\beta}_i \pm t_{df,1-\frac{\alpha}{2}} \, \hat{se}(\hat{\beta}_i \setminus g_i), \qquad (27)$$

where *df* denote the degrees of freedom. The Akaike (w_i) weights that are used to rank and scale models, can also be used in estimating the unconditional precision where the interest in the parameter β over all R models (for i=1,......,R), is defined as,

$$\hat{Var}(\hat{\bar{\beta}}_i) = \left[\sum_{i=1}^{R} w_i \sqrt{\hat{Var}(\hat{\beta}_i \setminus g_i) + (\hat{\beta}_i - \hat{\bar{\beta}})^2} \right]^2, \quad (28)$$

Where $\hat{\bar{\beta}}_i = \sum_{i=1}^{R} w_i \hat{\beta}_i$, and $\hat{\bar{\beta}}$ represents a form of model averaging. $\hat{\beta}_i$ means that the parameter β is estimated based on model g_i, but β is a parameter in common to all R models (even if its value is 0 in model i so that $\hat{\beta}_i = 0$). The estimator in (28) include two components of variance, the first one is the conditional sampling variance given model g_i $\left(Var(\hat{\beta}_i \setminus g_i) \right)$, and the second part is a variance component for model selection uncertainty $(\hat{\beta}_i - \hat{\bar{\beta}})^2$. The sum of the two components is multiplied by the Akaike weight, which reflects the relative support of model i. (See Burnham and Anderson (2002)).

The unconditional confidence interval will be as follows:

$$\hat{\bar{\beta}}_i \pm Z_{1-\frac{\alpha}{2}} \, \hat{se}(\hat{\bar{\beta}}_i), \qquad (29)$$

where

$$\hat{se}(\hat{\bar{\beta}}_i) = \sqrt{\hat{var}(\hat{\bar{\beta}}_i)}, \qquad (30)$$

and

$\hat{var}(\hat{\bar{\beta}}_i)$ is as defined in (28). In this work, the conditional and unconditional confidence intervals for β_i's are considered for the purpose of comparison.

5. The Proposed Information Criteria

Information criteria are used to select the best approximated model to the true model in a set of candidate models. These criteria such as AIC, BIC, HQC, and AICc, include two components, the logarithm of the sum of squared residuals which decreases with the increasing number of the estimated parameters. These criteria differ in the punishing terms. Therefore by taking the averages of these criteria, new criteria can be obtained which differ also in the punishing terms. The following are the proposed criteria based on AIC:

- The average of AIC and BIC which is denoted by AVGAB, and defined as follows:

$$\begin{aligned} AVGAB &= \frac{1}{2}(AIC + BIC) \\ &= -2\ln[L(\hat{\theta} \setminus y)] + K(1 + 0.5\ln(n)) \end{aligned} \qquad (31)$$

- The average of AIC and HQC which is denoted by AVGAH, and defined as follows:

$$\begin{aligned} AVGAH &= \frac{1}{2}(AIC + HQC) \\ &= -2\ln[L(\hat{\theta} \setminus y)] + 2K(1 + \ln(\ln(n))) \end{aligned} \qquad (32)$$

- The average of AIC and AICc which is denoted by AVGACc, and defined as follows:

$$\begin{aligned} AVGACc &= \frac{1}{2}(AIC + AICc) \\ &= -2ln[L(\hat{\theta} \setminus y)] + K\left(\left(\frac{n}{n-k-1}\right)+1\right) \end{aligned} \qquad (33)$$

Lemma 5.1

The average of AIC and BIC criteria, denoted by AVGAB in (31), is always greater than BIC and less than AIC when the sample size n<8, but AVGAB is always greater than AIC and less than BIC when the sample size n>8. The equality, between AIC, BIC, and AVGAB, holds when n ≈ 8. Therefore AVGAB lies between AIC and BIC and can be used as a new criterion.

Proof

It is found that: AIC-BIC= 2k- kln(n) > 0, when n < 8, and BIC-AIC=kln(n) – 2k >0, when n > 8. Also, AVGAB-AIC = k + 0.5kln(n) – 2k=0.5kln(n) –k >0, for n >8, and BIC-AVGAB = 0.5kln(n)-k > 0, for n > 8.

Thus, AVGAB is between AIC and BIC when n < 8 as follows:

$$(BIC < AVGAB < AIC), n < 8, \qquad (34)$$

and, AVGAB is between AIC and BIC when n > 8 as follows:

$$(AIC < AVGAB < BIC), n > 8, \qquad (35)$$

From (34) and (35), AVGAB can be used as a new criterion for selecting a model.

Lemma 5.2

The average of AIC and HQC criteria, denoted by AVGAH in (32), is always greater than HQC and less than AIC when the sample size n<16, but AVGAH is always greater than HQC and less than AIC when the sample size n>16. The equality, between AIC, HQC, and AVGAH, holds when n≈16. Therefore AVGAH lies between AIC and HQC and can be used as a new criterion for selecting a model.

Proof

It is found that: AIC-HQC= 2k- 2kln(ln(n)) > 0, when n < 16, and HQC-AIC=2kln(ln(n)) – 2k >0, when n >16. Also, AIC - AVGAH = k(1- ln(ln(n)) >0, for n <16, and AVGAH - AIC=k(ln(ln(n))-1) > 0, n >16.

Thus, AVGAH is between AIC and HQC when n < 16, as follows:

$$(HQC < AVGAH < AIC), n <16, \qquad (36)$$

and, AVGAH is between AIC and HQC when n > 16 as follows:

$$(AIC < AVGAH < HQC), n > 16, \qquad (37)$$

From (36) and (37), AVGAH can be used as a new criterion for selecting a model.

Lemma 5.3

The average of AIC and AICc criteria, denoted by AVGACc in (33), is always greater than AIC , and less than AICc when k >0, n>k+1, but AVGACc is always greater than AICc and less than AIC when k >0, n<k+1 (overfitting). The equality, between AIC, AICc, and AVGACc, holds for large values of n. Therefore AVGACc lies between AIC and AICc and can be used as a new criterion for selecting a model.

Proof

It   is   found   that:

$$AIC - AICc = 2k - \frac{2kn}{n-k-1} = 2k\left[1 - \frac{n}{n-(k+1)}\right] > 0, \text{ when}$$

k   >0, $n < k+1$ (overfitting),   and

$$AICc - AIC = \frac{2kn}{n-k-1} - 2k = 2k\left[\frac{n}{n-(k+1)} - 1\right] > 0, \text{ when}$$

k >0, $n > k+1$.   Also,   AIC   -   AVGACc   =

$$= 2k - k - \frac{kn}{n-k-1} = k - k\frac{n}{n-(k+1)} = k\left[1 - \frac{n}{n-(k+1)}\right] >$$

0,   for   k>0,   n<k+1,   and   AVGACc   -   AICc

$$= k + \frac{kn}{n-k-1} - \frac{2kn}{n-(k+1)} = k - \frac{kn}{n-k-1} = k\left[1 - \frac{n}{n-(k+1)}\right] >$$

0, for k >0, n<k+1 (overfitting). But AVGACc- AIC =

$$= k + \frac{kn}{n-k-1} - 2k = \frac{kn}{n-(k+1)} - k = k\left[\frac{n}{n-(k+1)} - 1\right] >0, \text{ for}$$

k>0, n>k+1,and

AICc   -   AVGACc

$$= \frac{2kn}{n-k-1} - k - \frac{kn}{n-(k+1)} = \frac{kn}{n-k-1} - k = k\left[\frac{n}{n-(k+1)} - 1\right]$$

> 0, for k >0, n<k+1 (over fitting). AIC=AICc=AVGACc for large values of n.

Thus, AVGACc is between AIC and AICc when n < k+1 (over fitting), as follows:

$$(AICc < AVGACc < AIC), n <k+1, \text{ (over fitting)} \qquad (38)$$

and, AVGACc is between AIC and AICc when n > k+1 as follows:

$$(AIC < AVGACc < AICc), n > k+1, \qquad (39)$$

From (38) and (39), AVGACc can be used as a new criterion for selecting a model.

6. Simulation Studies

Simulation is a very useful method to gain insight into model selection and model averaging. The following simulation studies consider some simulation studies in both model selection and model averaging.

6.1. Simulations of Model Selection

Suppose that the generating linear regression model for the data is defined as,

$$Y = X_\circ \beta_\circ + \varepsilon , \; \varepsilon \sim N\left(0, \sigma_\circ^2 I\right), \qquad (40)$$

where Y is an (nx1) observation vector, ε is an (nx1) error vector, $X_\circ$ is an $(n \times p_\circ)$ design matrix of rank $p_\circ$, and $\beta_\circ$ is a $(p_\circ \times 1)$ vector of unknown parameters. The objective is to find a suitable approximate model to (40). Define $\theta_\circ = \left(\beta_\circ, \sigma_\circ^2\right)$. The candidate linear regression model for selection takes the form,

$$Y = X\beta + \varepsilon , \; \varepsilon \sim N\left(0, \sigma^2 I\right), \qquad (41)$$

where X is an $(n \times p)$ design matrix of rank p, and β is a $(p \times 1)$ vector of unknown parameters. Define $\theta = \left(\beta, \sigma^2\right)$,

and $\hat{\theta} = \left(\hat{\beta}, \hat{\sigma}^2 \right)$ is a vector of the maximum likelihood estimators of θ obtained by maximizing the likelihood function $f(y \setminus \theta)$. The candidate models correspond to design matrices of ranks 2, 3, ..., k, where k=(p+1), and the matrix of rank 2 contains only one regressor and a column vector of ones. One of the candidate models will be correctly specified by containing the same regressors as the true one. Different sizes of samples 6, 8, 16, 20, 30, 50, 100, and 500 are used. For every sample size, the candidate models are fitted and evaluated by the AIC, AICc, BIC, HQC, AVGAB, AVGAH, and AVGACc criteria. This part of simulation consists of two practical applications. The first application considers the nested models, where the candidate models will be a sequence of models with design matrices of ranks, 2, 3, ..., p+1. The models for $2 \leq p < p_\circ$ will be underfitted and those for $p_\circ \leq p \leq k$, will be overfitted. Here, p is the true order and k=p+1, which contains the regressors and the constant term (column of ones). The second application considers all possible regressions which is more realistic than the nested models. (See Cavanaugh and Neath (1999); and Burnham and Anderson (2002); and 2004)).

6.1.1. Model Selection Using Nested Models

In all sets k=11(10 regressors + 1) with different sample sizes n=6, 8, 16, 20, 30, 50, 100, 200, and 500 with 1000 replications. The true model is parameterized by $\beta_\circ = (1,1,2,3,4)$, and $p_\circ = 5$. Thus there are 10 candidate models, from which the smallest 3 models are under fitted and the largest 6 models are over fitted as follows:

$M(1) = 1 + x_1 + \varepsilon$
$M(2) = 1 + x_1 + 2x_2 + \varepsilon$
$M(3) = 1 + x_1 + 2x_2 + 3x_3 + \varepsilon$
$M(4) = 1 + x_1 + 2x_2 + 3x_3 + 4x_4 + \varepsilon$
$M(5) = 1 + x_1 + 2x_2 + 3x_3 + 4x_4 + x_5 + \varepsilon$
$M(6) = 1 + x_1 + 2x_2 + 3x_3 + 4x_4 + x_5 + x_6 + \varepsilon$
$M(7) = 1 + x_1 + 2x_2 + 3x_3 + 4x_4 + x_5 + x_6 + x_7 + \varepsilon$
$M(8) = 1 + x_1 + 2x_2 + 3x_3 + 4x_4 + x_5 + x_6 + x_7 + x_8 + \varepsilon$
$M(9) = 1 + x_1 + 2x_2 + 3x_3 + 4x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + \varepsilon$
$M(10) = 1 + x_1 + 2x_2 + 3x_3 + 4x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + \varepsilon$

The regressors are generated from a N(0,1) distribution. The true model variance is set at $\sigma_\circ^2 = 1$, $\sigma_\circ^2 = 15$, and $\sigma_\circ^2 = 30$ such that the signal-to-noise ratio (SNR) will be (1:0.03), (1:0.5), and (1:1), respectively. Therefore the results are shown in three groups, the first group is for 1:0.03 SNR, the second group is for 1:0.5, the third group is for 1:1 SNR. The performance of each criterion is affected by the form of the candidate model, the sample size n, and the signal-to-noise ratio (SNR) for the true model (SNR). (See Cavavough and Neath (1999)).

Table (1) in Appendix (A) shows the results of the first group (1:0.03 SNR) for the four criteria AIC, BIC, AICc, HQC, in addition to the three proposed criteria AVGAB, AVGAH, and AVGACc. Table (2) in Appendix (A) shows the

results of the second group (1:0.5 SNR) for the four criteria AIC, BIC, AICc, HQC, in addition to the three proposed criteria AVGAB, AVGAH, and AVGACc. Also, Table (3) in Appendix (A) shows the results of the third group (1:1 SNR) for the four criteria AIC, BIC, AICc, HQC, in addition to the three proposed criteria AVGAB, AVGAH, and AVGACc. All the criteria are showed difficulty in identifying the correctly specified model. Often underfitted models are chosen in all cases of SNR 1:0.03, 1:0.5 and 1:1, except BIC, HQC, AICc, AVGAB, and AVGAH for n<8.(See Table(1), Table(2) and Table(3)), where these criteria identified the correct model M(4). In Table (3), all the criteria, except AVGACc, identified the correct model (M4) for small sample sizes (less than 16). The values of all criteria are increasing more rapidly in the second and third groups than in the first one as a result of the large variance of the noise term in the second and third groups. Comparing the proposed criteria AVGAB, AVGAH, and AVGACc, for the model M(4) it is found that AVGAH outperforms for smaller sample sizes (n=6 and n=8), and AVGACc outperforms for larger sample sizes (greater than n=20). The results of our proposed criteria will be shown for M(4). These results support the three lemmas in section (5) as follows:

• For AVGAB

It is found that AVGAB, when n<8, i.e., when n=6, is greater than BIC and less than AIC. On the other side, when n>8, i.e., it is found that AVGAB is less than BIC and greater than AIC. Also, when n=8, it is shown that AIC, BIC, and AVGAB are approximately equal. These results support our lemma (1) as shown in Tables (1), (2), and (3) for the three groups. Therefore, AVGAB can be used as a new criteria of model selection for small and large sample sizes.

• For AVGAH

It is found when n<16, i.e., when n=6, and 8, that AVGAH is greater than HQC and less than AIC. But when n>16, i.e., it is found that AVGAH is less than HQC and greater than AIC. Also, when n=16, it is shown that AIC, HQC, and AVGAH are approximately equal. These results support our lemma (2) as shown in Tables (1), (2), and (3) for the three groups. Therefore, AVGAH can be used as a new criteria of model selection for small and large sample sizes.

• For AVGACc

It is found that AVGACc is always greater than AIC and less than AICc in the three groups for small sample sizes. For example, AIC< AVGACc<AICc for n=20, 50 and SNR is 1:0.03. (See Table (1)). Also, AIC<AVGACc<AICc for n=8, 20 and SNR is 1:0.5, but AIC<AVGACc when n=8, 20, 50 and SNR is 1:1. These results mean that AVGACc is always greater than AIC and less than AICc for smaller sample size. These results also, support lemma (3) when n>k+1. Therefore, AVGACc can be used as a new criteria of model selection for small and medium sample sizes. It is found for large sample sizes in all groups that AIC and AICc and AVGACc are approximately equal. (See Tables (1), (2), and (3) for n=500).

6.1.2. Model Selection Using All Possible Regression

All possible regression are more realistic than the nested models. A study is considered with 10 regressors in addition to the intercept and different sample sizes n=6, 8, 16, 20, 30, 50, 100, 200, and 500. As aresult, 91 models are considered: one model with all 10 regressors, 10 models with 9 regressors, 10 models with 8 regressors, 10 models with 7 regressors, 10 models with 6 regressors, 10 models with 5 regressors, 10 models with 4 regressors, 10 models with 3 regressors, 10 models with 2 regressors, and 10 with one regressor. The regressors are generated from a N(0,1). The true model is parameterized by $\beta_\circ = (1,1,2,3,4)$, $\sigma_\circ^2 = 15$, and $p_\circ = 5$, such that the SNR is 1:0.5.

Figure (1) in Appendix (B) show all possible regressions. Each row in each graph represents a model, and the shaded rectangles in the columns indicate the variables included in the given model. The numbers on the left margin are the values of BIC, and the darkness of the shading represents the ordering of the BIC values. In Figure (1-1) show all possible regressions in the first group when sample size=20, the first best model has the intercept, x1, x3, and x7 with lowest value of BIC(=-1.1); and the second best model has the intercept, x3, and x7 with a value of BIC=-0.61; etc. The variables in the various models differ according to sample size, for example, in Figure (1-6) when n=500, the first best model has the intercept, x3, and x7 with lowest value of BIC(=-1.4); and the second best model has the intercept, x3, x4, and x7 with a value of BIC=0.51; etc. It is found that the regressor variables in the various models differ according to sample size and the SNR as shown in Figures (1) in Appendix(B).

It is found that most of the best models contain the regressors, x_1, x_2, x_3, x_4, and x_7.

6.2. Simulation of Model Averaging

Model averaging is used to base point inferences on the entire set of models rather than a single best model. For comparison, the Akaike (AIC), corrected Akaike (AICc), and Bayesian information criterion (BIC) weights are used as different point inference methodology. The inference based on all models can be used to reduce the bias effects of regression coefficients. (See Burnham and Anderson (2002)).

The entire set of models which will be used here is the second group simulation study considered in section 6-1-1 with SNR 1:0.5. Table (4) in Appendix (A) shows the AIC weights by model, sample size n=100, and regressors used in each model. Table (5) in Appendix (A) shows the AICc weights by model, sample size n=100, and regressors used in each model. Table (6) in Appendix (A) shows the BIC weights by model, sample size n=100, and regressors used in each model. From these Tables, it is shown that the best four models chosen by AIC are M(3), M(1), M(4), and M(7), respectively. But the best four models chosen by AICc and BIC are M(1), M(3), M(2), and M(4), respectively.

Table (7) in Appendix (A) shows the squared bias, the predicted mean squared error (PMSE), and the 95% confidence intervals (CI's) of regression coefficients, in all models using model selection and model averaging. It is clear that model averaged has less bias and less PMSE compared to model selection. (See Figure (2) in Appendix (B)). Also, CI's for model averaging are greater than those of model selection for regression coefficients. These results assure that we need methods to reduce model selection bias and PMSE of the alternative models. Model averaging is the one which based on all models.

7. Conclusions

The values of all the criteria are affected by the different SNR used in this work; these values are increased by increasing the noise. In selecting the true model, it is found that the proposed criterion AVGAH gave best results, compared to the other two proposed criteria (AVGAB and AVGACc), for sample sizes (n=6,8,and 20) and for all signal-to-noise ratio (SNR) used in this work. On the other hand, the proposed criterion AVGACc gave best results, compared to the other two proposed criteria (AVGAB and AVGAH), for large sample sizes (n=30, 50,100,200, and 500) and for all signal-to-noise ratio (SNR) used in this work

When comparing the four criteria AIC, BIC, AICc, and HQC, in selecting the true model, it is found that these criteria are also not affected by the different SNR. For example when n=6, it is found that the criterion AICc gave the best results for all SNR. when n=8, it is found that the criterion HQC gave the best results for all SNR. It is found that AIC selected the true model when n=200 for all SNR.

When comparing all the criteria, AIC, BIC, AICc, HQC, AVGAB, AVGAH, and AVGACc in selecting the true model, it is found that all the criteria are affected a little by the different SNR used in this work as follows:

- AICc selected the true model when n=6 in all SNR.
- HQC selected the true model when n=8 in all SNR.
- AVGACc selected the true model when n=16,30,100, 500 in all SNR, and when n=50 and SNR 1:0.03, and 1:0.5.
- AIC selected the true model when n=20,200 in all SNR, and n=50 when SNR is 1:1 only.

Model selection is just a method of selecting the best model, and the statistical inference is based on that model. Model averaging is based on the entire set of models (the candidate models) of the problem at hand. Our results showed smaller values of bias, variance, and PMSE for regression coefficient estimates of model averaging than that of model selection. It is found that statistical inference based on all models can be used to reduce model selection bias and variance of the regression coefficients estimates. Model averaging is a new field which need more research, advances, and updating.

Appendix (A)

Table (1). Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.03. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models. Negative values are the lowest values, which refer to best criteria.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
6	M(1)	21.31	20.69	33.31	18.81	21.00	20.06	22.81
	M(2)	21.31	20.48	61.31	17.98	20.90	19.64	24.31
	M(3)	22.23	21.19	INF	18.06	21.71	20.14	27.31
	M(4)	21.52	20.27	-62.48	16.52	20.89	19.02	33.31
	M(5)							
	M(6)							
	M(7)							
	M(8)							
	M(9)							
	M(10)							

Table (1). Continued : Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.03. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models. Negative values are the lowest values, which refer to best criteria.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
8	M(1)	23.11	23.35	29.11	21.51	23.23	22.31	24.11
	M(2)	24.56	24.88	37.90	22.42	24.72	23.49	24.91
	M(3)	26.49	26.89	56.49	23.81	26.69	25.15	26.11
	M(4)	26.64	27.12	110.64	23.43	26.89	25.04	25.11
	M(5)	15.29	15.84		11.54	15.57	13.41	32.11
	M(6)	-15.27	-14.63		-19.56	-14.95	-17.41	44.11
	M(7)							
	M(8)							
	M(9)							
	M(10)							

Table (1). Continued: Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.03. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
16	M(1)	49.18	51.50	51.18	49.30	50.34	49.24	49.87
	M(2)	50.18	53.28	53.82	50.34	51.73	50.26	50.18
	M(3)	52.03	55.90	58.03	52.23	53.96	52.13	47.54
	M(4)	52.91	56.54	61.24	52.15	54.23	52.02	50.98
	M(5)	53.88	59.28	67.88	54.15	56.58	54.01	51.51
	M(6)	55.47	61.65	76.04	55.79	58.56	55.63	52.18
	M(7)	57.26	64.22	87.27	57.62	60.74	57.45	53.03
	M(8)	58.55	66.27	102.55	58.94	62.41	58.75	54.18
	M(9)	58.24	66.74	124.24	58.68	62.49	58.46	55.78
	M(10)	60.22	69.49	164.22	60.69	64.85	60.45	58.18

Table (1). Continued: Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.03. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
20	M(1)	66.54	69.53	68.04	67.13	68.04	66.83	66.88
	M(2)	68.41	72.39	71.07	69.18	70.40	68.79	67.07
	M(3)	64.72	69.69	69.00	65.69	67.21	65.20	64.48
	M(4)	66.67	72.64	73.13	67.84	69.66	67.25	67.54
	M(5)	68.62	75.59	77.95	69.98	72.11	69.30	67.83
	M(6)	68.97	76.94	82.07	70.53	72.96	69.75	68.16
	M(7)	67.07	76.03	85.07	68.82	71.55	67.94	68.54
	M(8)	69.03	78.99	93.47	70.97	74.01	70.00	69.00
	M(9)	68.72	79.67	101.72	70.85	74.19	69.78	69.54
	M(10)	68.70	80.64	113.27	71.03	74.67	69.86	70.21

Table (1). Continued: Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.03. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
30	M(1)	85.70	89.90	86.62	87.05	87.80	86.37	85.92
	M(2)	87.61	93.22	89.21	89.41	90.42	88.51	86.03
	M(3)	86.40	93.41	88.90	88.64	89.90	87.52	83.28
	M(4)	87.10	95.51	90.75	89.79	91.30	88.44	86.30
	M(5)	88.61	98.41	93.70	91.74	93.51	90.17	86.45
	M(6)	90.34	101.55	97.19	93.92	95.94	92.13	86.61
	M(7)	91.48	104.09	100.48	95.52	97.79	93.50	86.79
	M(8)	92.38	106.39	103.95	96.86	99.38	94.62	87.00
	M(9)	94.33	109.74	109.00	99.26	102.04	96.79	87.02
	M(10)	95.88	112.70	114.24	101.26	104.29	98.57	87.44

Table (1). Continued: Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.03. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
50	M(1)	136.76	142.50	137.28	138.95	139.63	137.86	136.89
	M(2)	135.85	143.50	136.74	138.76	139.67	137.31	136.95
	M(3)	137.85	147.41	139.21	141.49	142.63	139.67	134.09
	M(4)	136.35	147.82	138.30	140.72	142.08	138.53	137.10
	M(5)	136.87	150.25	139.53	141.96	143.56	139.41	137.17
	M(6)	138.77	154.06	142.28	144.59	146.41	141.68	137.25
	M(7)	139.78	156.99	144.28	146.33	148.38	143.05	137.33
	M(8)	137.43	156.55	143.07	144.71	146.99	141.07	137.42
	M(9)	139.19	160.22	146.14	147.20	149.71	143.20	137.51
	M(10)	139.81	162.76	148.24	148.55	151.29	144.18	137.61

Table (1). Continued: Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.03. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
100	M(1)	292.73	300.55	292.98	295.90	296.64	294.31	292.79
	M(2)	293.52	303.94	293.94	297.74	298.73	295.63	292.83
	M(3)	295.39	308.41	296.02	300.66	301.90	298.02	289.89
	M(4)	297.38	313.01	298.28	303.71	305.20	300.54	292.89
	M(5)	299.22	317.45	300.43	306.60	308.34	302.91	292.92
	M(6)	300.71	321.55	302.29	309.14	311.13	304.92	292.96
	M(7)	302.67	326.12	304.67	312.16	314.39	307.42	293.99
	M(8)	304.24	330.29	306.71	314.79	317.27	309.51	293.03
	M(9)	304.98	333.64	307.98	316.58	319.31	310.78	293.07
	M(10)	306.62	337.89	310.21	319.28	322.26	312.95	293.10

Table (1). Continued: Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.03. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
200	M(1)	602.82	612.72	602.94	606.83	607.77	604.82	602.85
	M(2)	598.81	612.01	599.02	604.15	605.41	601.48	602.87
	M(3)	600.40	616.90	600.71	607.08	608.65	603.74	599.90
	M(4)	602.15	621.94	602.59	610.16	612.05	606.16	602.90
	M(5)	602.78	625.87	603.37	612.13	614.33	607.46	602.92
	M(6)	604.64	631.03	605.40	615.32	617.84	609.99	602.93
	M(7)	605.24	634.93	606.19	617.26	620.09	611.25	602.95
	M(8)	606.93	639.91	608.09	620.28	623.42	613.60	602.96
	M(9)	608.78	645.06	610.18	623.46	626.92	616.12	602.98
	M(10)	609.79	645.37	611.46	625.80	629.58	617.80	603.00

Table (1). Continued: Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.03. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
500	M(1)	1373.22	1385.87	1373.27	1378.18	1379.55	1375.70	1373.23
	M(2)	1375.14	1392.00	1375.22	1381.76	1383.27	1378.45	1373.24
	M(3)	1376.81	1397.88	1376.93	1385.08	1387.35	1380.95	1370.25
	M(4)	1376.39	1401.68	1376.56	1386.32	1389.04	1381.35	1373.25
	M(5)	1378.19	1407.69	1378.42	1389.77	1392.94	1383.98	1373.26
	M(6)	1379.11	1412.83	1379.41	1392.34	1395.97	1385.73	1373.27
	M(7)	1367.40	1405.33	1367.78	1382.29	1386.37	1374.85	1373.27
	M(8)	1368.90	1411.05	1369.35	1385.44	1389.98	1377.17	1373.28
	M(9)	1370.86	1417.22	1371.40	1389.05	1394.04	1379.96	1373.28
	M(10)	1371.67	1422.24	1372.31	1391.51	1396.96	1381.59	1373.29

Table (2). Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.5. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models. Negative values are the lowest values, which refer to best criteria.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
6	M(1)	53.81	53.18	65.81	51.31	53.50	52.56	55.31
	M(2)	53.81	52.97	93.81	50.47	53.39	52.14	56.81
	M(3)	54.72	53.68	INF	50.56	54.20	52.64	58.31
	M(4)	54.02	52.77	-29.98	49.02	53.39	51.52	68.81
	M(5)							
	M(6)							
	M(7)							
	M(8)							
	M(9)							
	M(10)							

Table (2). Continued: Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.5. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models. Negative values are the lowest values, which refer to best criteria.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
8	M(1)	66.44	66.68	72.44	64.83	66.56	65.64	67.44
	M(2)	67.89	68.21	81.22	65.75	68.05	66.82	68.24
	M(3)	69.82	70.21	99.82	67.14	70.01	68.48	67.19
	M(4)	69.97	70.45	153.97	66.76	70.21	68.37	71.44
	M(5)	58.62	59.17	INF	54.87	58.89	56.74	75.44
	M(6)	28.06	28.69	-115.94	23.77	28.38	25.92	87.44
	M(7)							
	M(8)							
	M(9)							
	M(10)							

Table (2). Continued: Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.5. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
16	M(1)	135.84	138.15	137.84	135.96	136.99	135.90	136.24
	M(2)	136.84	139.93	140.48	137.00	138.39	136.92	136.53
	M(3)	138.69	142.55	144.69	138.89	140.62	138.79	134.09
	M(4)	138.57	143.20	147.90	138.81	140.89	138.69	137.20
	M(5)	140.53	145.94	154.53	140.81	143.23	140.67	137.64
	M(6)	142.13	148.31	162.70	142.44	145.22	142.29	138.17
	M(7)	143.92	150.88	173.92	144.28	147.40	144.10	138.84
	M(8)	145.21	152.93	189.21	145.60	149.07	145.40	139.69
	M(9)	144.90	153.40	210.90	145.33	149.15	145.12	140.84
	M(10)	146.87	156.14	250.87	147.35	151.51	147.11	142.44

Table (2). Continued: *Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.5. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.*

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
20	M(1)	174.87	177.85	176.37	175.45	176.36	175.16	175.20
	M(2)	176.73	180.71	179.40	177.51	178.72	177.12	175.40
	M(3)	173.04	178.02	177.32	174.01	175.53	173.52	172.80
	M(4)	174.99	180.97	181.46	176.16	177.98	175.58	175.87
	M(5)	176.94	183.91	186.28	178.30	180.43	177.62	176.15
	M(6)	177.30	185.26	190.39	178.85	181.28	178.07	176.48
	M(7)	175.39	184.35	193.39	177.14	179.87	176.27	176.87
	M(8)	177.35	187.31	201.80	179.29	182.33	178.32	177.32
	M(9)	177.04	187.99	210.04	179.18	182.51	178.11	177.87
	M(10)	177.02	188.97	221.59	179.35	182.99	178.18	178.53

Table (2). Continued: *Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.5. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.*

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
30	M(1)	248.18	252.39	249.11	249.53	250.29	248.86	248.40
	M(2)	250.10	255.70	251.70	251.89	252.90	250.99	248.52
	M(3)	248.88	255.89	251.38	251.12	252.39	250.00	245.76
	M(4)	249.58	257.99	253.23	252.27	253.78	250.93	248.78
	M(5)	251.09	260.90	256.18	254.23	255.99	252.66	248.93
	M(6)	252.82	264.03	259.68	256.41	258.42	254.61	249.10
	M(7)	253.96	266.57	262.96	257.99	260.27	255.98	249.28
	M(8)	254.86	268.87	266.44	259.34	261.87	257.10	249.47
	M(9)	256.81	272.23	271.48	261.74	264.52	259.28	249.68
	M(10)	258.37	275.18	276.72	263.75	266.77	261.06	249.92

Table (2). Continued: *Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.5. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.*

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
50	M(1)	407.57	413.30	408.09	409.75	410.44	408.66	407.69
	M(2)	406.65	414.30	407.54	409.57	410.48	408.11	407.76
	M(3)	408.65	418.21	410.02	412.29	413.43	410.47	404.89
	M(4)	407.15	418.62	409.11	411.52	412.89	409.34	404.90
	M(5)	407.67	421.06	410.34	412.77	414.36	410.22	407.98
	M(6)	409.57	424.87	413.08	415.40	417.22	412.48	408.06
	M(7)	410.58	427.79	415.08	417.13	419.19	413.86	408.14
	M(8)	408.23	427.35	413.87	415.51	417.79	411.87	408.23
	M(9)	409.99	431.03	416.94	418.01	420.51	414.00	408.32
	M(10)	410.62	433.56	419.05	419.36	422.09	414.99	408.41

Table (2). Continued: *Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.5. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.*

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
100	M(1)	834.34	842.16	834.59	837.51	838.25	835.92	834.40
	M(2)	835.13	845.55	835.56	839.35	840.34	837.24	834.44
	M(3)	836.99	850.02	837.63	842.27	843.51	839.63	831.50
	M(4)	838.99	854.62	839.89	845.32	846.81	842.15	834.50
	M(5)	840.83	859.06	842.04	848.21	849.95	844.52	834.53
	M(6)	842.32	863.16	843.90	850.75	852.74	846.54	834.57
	M(7)	844.28	867.73	846.28	853.77	856.00	849.03	834.60
	M(8)	845.85	871.90	848.32	856.40	858.88	851.12	834.64
	M(9)	846.60	875.25	849.60	858.19	860.92	852.39	834.68
	M(10)	848.23	879.50	851.82	860.89	863.87	854.56	834.71

Table (2). Continued: Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.5. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
200	M(1)	1686.04	1695.94	1686.17	1690.05	1690.99	1688.05	1686.07
	M(2)	1682.03	1695.23	1682.24	1687.37	1688.63	1684.70	1686.09
	M(3)	1683.62	1700.12	1683.93	1690.30	1691.87	1686.96	1683.10
	M(4)	1685.37	1705.17	1685.81	1693.38	1695.27	1689.38	1686.12
	M(5)	1686.00	1709.09	1686.59	1695.35	1697.55	1690.68	1686.14
	M(6)	1687.87	1714.25	1688.62	1698.54	1701.06	1693.21	1686.15
	M(7)	1688.46	1718.15	1689.41	1700.48	1703.31	1694.47	1686.17
	M(8)	1690.15	1723.13	1691.31	1703.50	1706.64	1696.82	1686.18
	M(9)	1695.00	1728.28	1693.40	1706.68	1710.14	1699.34	1686.20
	M(10)	1693.01	1732.59	1694.68	1709.03	1412.80	1701.02	1686.22

Table (2). Continued: Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:0.5. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
500	M(1)	4081.27	4093.92	4081.32	4086.23	4087.59	4083.75	4081.28
	M(2)	4083.19	4100.05	4083.27	4089.81	4091.62	4086.50	4081.29
	M(3)	4084.86	4105.93	4084.98	4093.13	4095.40	4088.99	4078.30
	M(4)	4084.44	4109.73	4084.61	4094.37	4097.09	4089.40	4081.30
	M(5)	4086.24	4115.74	4086.47	4097.82	4100.99	4092.03	4081.31
	M(6)	4087.16	4120.88	4087.46	4100.39	4104.02	4093.78	4081.32
	M(7)	4075.45	4113.39	4075.82	4090.34	4094.42	4082.90	4081.32
	M(8)	4076.95	4119.10	4077.40	4093.49	4098.03	4085.22	4081.33
	M(9)	4078.91	4125.27	4079.45	4097.10	4102.09	4088.01	4081.33
	M(10)	4079.72	4130.29	4080.36	4099.56	4105.00	4089.64	4081.34

Table (3). Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:1. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models. Negative values are the lowest values, which refer to best criteria.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
6	M(1)	62.13	61.50	74.13	59.62	61.81	60.88	63.63
	M(2)	62.13	61.29	102.13	58.79	61.71	60.46	65.13
	M(3)	63.04	62.00	INF	58.87	62.52	60.96	66.63
	M(4)	62.034	61.09	-21.66	57.34	61.71	59.84	77.13
	M(5)							
	M(6)							
	M(7)							
	M(8)							
	M(9)							
	M(10)							

Table (3). Continued: Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:1. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models. Negative values are the lowest values, which refer to best criteria.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
8	M(1)	77.53	77.77	83.53	75.92	77.65	76.73	78.53
	M(2)	78.98	80.00	92.32	76.84	79.14	77.91	79.33
	M(3)	80.91	81.30	110.91	78.23	81.11	79.57	78.28
	M(4)	81.06	81.54	165.06	77.85	81.30	79.46	82.53
	M(5)	69.71	70.26	INF	65.96	69.99	67.83	86.53
	M(6)	39.15	39.78	-104.85	34.86	39.47	37.01	98.53
	M(7)							
	M(8)							
	M(9)							
	M(10)							

Table (3). Continued: *Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:1. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.*

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
16	M(1)	158.02	160.33	160.02	158.14	159.18	158.08	158.45
	M(2)	159.02	162.11	162.66	159.18	160.57	159.10	158.71
	M(3)	160.87	164.73	166.87	161.07	162.80	160.97	156.27
	M(4)	160.75	165.38	170.08	160.98	163.07	160.87	159.38
	M(5)	162.71	168.12	176.71	162.99	165.42	162.85	159.82
	M(6)	164.31	170.49	184.88	164.63	167.40	164.47	160.35
	M(7)	166.11	173.06	196.11	166.46	169.58	166.28	161.02
	M(8)	167.39	175.11	211.39	167.78	171.25	167.58	161.87
	M(9)	167.08	175.58	233.08	167.52	171.33	167.30	163.02
	M(10)	169.05	178.32	273.05	169.53	173.69	169.29	164.62

Table (3). Continued: *Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:1. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.*

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
20	M(1)	202.59	205.58	204.09	203.18	204.09	202.88	202.93
	M(2)	204.46	208.44	207.12	205.23	206.45	204.84	203.12
	M(3)	200.76	205.74	205.05	201.74	203.25	201.25	200.53
	M(4)	202.72	208.69	209.18	203.89	205.71	203.30	203.59
	M(5)	204.67	211.64	214.00	206.03	208.15	205.35	203.88
	M(6)	205.02	212.99	218.11	206.58	209.00	205.80	204.21
	M(7)	203.12	212.08	221.12	204.87	207.60	203.99	204.59
	M(8)	205.08	215.03	229.52	207.02	210.06	206.05	205.05
	M(9)	204.76	215.72	237.76	206.90	210.24	205.83	205.59
	M(10)	204.74	216.69	249.32	207.08	210.72	205.91	206.26

Table (3). Continued: *Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:1. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.*

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
30	M(1)	289.77	293.98	290.70	291.11	291.87	290.45	289.99
	M(2)	291.68	297.29	293.28	293.48	294.49	292.58	290.11
	M(3)	290.47	297.48	292.97	292.71	293.97	291.59	287.35
	M(4)	291.17	299.58	294.82	293.86	295.37	292.51	290.37
	M(5)	292.68	302.49	297.77	295.81	297.58	294.25	290.52
	M(6)	294.41	305.62	301.27	297.99	300.01	296.20	290.69
	M(7)	295.55	308.16	304.55	299.59	301.86	297.57	290.86
	M(8)	296.45	310.46	308.03	300.93	303.45	298.69	291.06
	M(9)	298.40	313.81	313.07	303.33	306.11	300.87	291.27
	M(10)	299.96	316.77	318.31	305.33	308.36	302.64	291.51

Table (3). Continued: *Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:1. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.*

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
50	M(1)	476.88	482.62	477.40	479.07	479.75	477.97	477.01
	M(2)	475.97	483.62	476.86	478.88	479.79	477.43	477.07
	M(3)	477.97	487.53	479.33	481.61	482.75	479.79	474.21
	M(4)	476.47	487.94	478.42	480.83	482.20	478.65	477.22
	M(5)	476.99	490.37	479.65	482.08	483.68	479.54	477.29
	M(6)	478.89	494.18	482.40	484.71	486.54	481.80	477.37
	M(7)	479.90	497.11	484.40	486.45	488.50	483.17	477.45
	M(8)	477.55	496.67	483.19	484.83	487.11	481.19	477.54
	M(9)	479.31	500.34	486.26	487.32	489.83	483.32	477.63
	M(10)	479.93	502.88	488.37	488.67	491.40	484.30	477.73

Table (3). continued: Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:1. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
100	M(1)	972.97	980.79	973.22	976.14	976.88	974.55	973.03
	M(2)	973.76	984.18	974.18	977.98	978.97	975.87	973.07
	M(3)	975.63	988.65	976.26	980.90	982.14	978.26	970.13
	M(4)	977.62	993.25	978.52	983.95	985.44	980.78	973.13
	M(5)	979.46	997.69	980.67	986.84	988.58	983.15	973.16
	M(6)	980.95	1001.79	982.53	989.38	991.37	985.16	973.20
	M(7)	982.91	1006.36	984.91	992.40	994.63	987.66	973.23
	M(8)	984.48	1010.53	986.95	995.03	997.51	989.75	973.27
	M(9)	985.22	1013.88	988.22	996.82	999.55	991.02	973.31
	M(10)	986.86	1018.13	990.45	999.52	1002.50	993.19	973.34

Table (3). Continued: Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:1. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
200	M(1)	1963.30	1973.20	1963.42	1967.31	1968.25	1965.30	1963.33
	M(2)	1959.29	1972.49	1959.50	1964.63	1965.89	1961.96	1963.35
	M(3)	1960.88	1977.37	1961.19	1967.56	1969.13	1964.22	1960.38
	M(4)	1962.63	1982.42	1963.07	1970.64	1972.53	1966.64	1963.38
	M(5)	1963.26	1986.35	1963.85	1972.61	1974.81	1967.93	1963.39
	M(6)	1965.13	1991.51	1965.88	1975.80	1978.32	1970.47	1963.41
	M(7)	1965.72	1995.41	1966.67	1977.74	1980.56	1971.73	1963.42
	M(8)	1967.41	2000.39	1968.57	1980.76	1983.90	1974.08	1963.44
	M(9)	1969.26	2005.54	1970.66	1983.94	1987.40	1976.60	1963.46
	M(10)	1970.27	2009.85	1971.94	1986.28	1990.06	1978.28	1963.48

Table (3). continued: Criteria of Model Selection when signal-to-noise ratio (SNR) is 1:1. A sequence of sample sizes are used (n=6, 8, 16, 20, 30, 50, 100, 200, 500) for 10 nested models.

Sample size (n)	Model	AIC	BIC	AICc	HQC	AVGAB	AVGAH	AVGACc
500	M(1)	4774.42	4787.06	4774.47	4779.38	4780.74	4776.90	4774.43
	M(2)	4776.34	4793.20	4776.42	4782.95	4784.77	4779.65	4774.44
	M(3)	4778.01	4799.08	4778.13	4786.28	4788.54	4782.14	4771.45
	M(4)	4777.59	4802.88	4777.76	4787.51	4790.23	4782.55	4774.45
	M(5)	4779.39	4808.89	4779.62	4790.96	4794.14	4785.18	4774.46
	M(6)	4780.31	4814.03	4780.60	4793.54	4797.17	4786.93	4774.46
	M(7)	4768.60	4806.53	4768.97	4783.49	4787.57	4776.04	4774.47
	M(8)	4770.10	4812.25	4770.55	4786.64	4791.17	4778.37	4774.48
	M(9)	4772.06	4818.42	4772.60	4790.25	4795.24	4781.16	4774.48
	M(10)	4772.87	4823.44	4773.51	4792.71	4798.15	4782.79	4774.49

Table (4). The AIC weights by model, sample size n=100, and different regressors used in each model according to the second simulation study (second group) with signal-to-noise ratio (SNR) is 1:0.5.

Model	Regressors used	AIC	Delta Δ_i	Weight w_i
M(3)	x_1, x_2, x_3	64.72	0.00	0.34
M(1)	x_1	66.54	1.83	0.14
M(4)	x_1, x_2, x_3, x_4	66.67	1.96	0.13
M(7)	$x_1, x_2, x_3, x_4, x_5, x_6, x_7$	67.07	2.35	0.11
M(2)	x_1, x_2	68.41	3.69	0.05
M(5)	x_1, x_2, x_3, x_4, x_5	68.62	3.90	0.05
M(10)	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$	68.70	3.98	0.05
M(9)	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9$	68.72	4.00	0.05
M(6)	$x_1, x_2, x_3, x_4, x_5, x_6$	68.97	4.26	0.04
M(8)	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	69.03	4.31	0.04

Table (5). *The AICc weights by model, sample size n=100, and different regressors used in each model according to the second simulation study (second group) with signal-to-noise ratio (SNR) is 1:0.5.*

Model	Regressors used	AICc	Delta Δ_i	Weight w_i
M(1)	x_1	68.04	0.00	0.52
M(3)	x_1,x_2,x_3	69.00	0.96	0.32
M(2)	x_1,x_2	71.07	3.03	0.11
M(4)	x_1,x_2,x_3,x_4	73.13	5.09	0.04
M(5)	x_1,x_2,x_3,x_4,x_5	77.95	9.91	0.00
M(6)	x_1,x_2,x_3,x_4,x_5,x_6	82.07	14.02	0.00
M(7)	$x_1,x_2,x_3,x_4,x_5,x_6,x_7$	85.07	17.02	0.00
M(8)	$x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8$	93.47	25.43	0.00
M(9)	$x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9$	101.72	33.67	0.00
M(10)	$x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10}$	113.27	45.22	0.00

Table (6). *The BIC weights by model, sample size n=100, and different regressors used in each model according to the second simulation study (second group) with signal-to-noise ratio (SNR) is 1:0.5.*

Model	Regressors used	BIC	Delta Δ_i	Weight w_i
M(1)	x_1	69.53	0.00	0.40
M(3)	x_1,x_2,x_3	69.69	0.16	0.37
M(2)	x_1,x_2	72.39	2.86	0.10
M(4)	x_1,x_2,x_3,x_4	72.65	3.11	0.08
M(5)	x_1,x_2,x_3,x_4,x_5	75.59	6.06	0.02
M(7)	$x_1,x_2,x_3,x_4,x_5,x_6,x_7$	76.03	6.50	0.01
M(6)	x_1,x_2,x_3,x_4,x_5,x_6	76.94	7.41	0.00
M(8)	$x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8$	78.99	9.45	0.00
M(9)	$x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9$	79.67	10.14	0.00
M(10)	$x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10}$	80.64	11.11	0.00

Table (7). *Bias square , predicted mean square error(PMSE),and 95% confidence intervals (CI's) of the coefficient in model selection compared to model averaging for in the second group simulation study, sample size n=100, and with signal-to-noise ratio (SNR) is 1:0.5.*

Coefficient / Model	$\hat{\beta}_o$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\beta}_9$	$\hat{\beta}_{10}$
M(1)	1.102	-1.362									
M(2)	1.037	-1.364	0.249								
M(3)	1.050	-1.353	0.220	2.453							
M(4)	1.056	-1.359	0.214	2.464	3.889						
M(5)	1.087	-1.463	0.123	2.447	3.866	1.536					
M(6)	1.053	-1.590	0.102	2.417	3.757	1.471	-0.137				
M(7)	1.109	-1.613	0.136	2.425	3.795	1.452	-0.138	1.276			
M(8)	1.090	-1.618	0.074	2.574	3.756	1.386	0.019	1.182	1.934		
M(9)	1.096	-1.144	0.092	2.530	4.196	1.509	0.366	1.219	1.888	-0.632	
M(10)	1.127	-1.546	0.110	2.707	4.116	1.612	0.451	1.296	1.837	-0.810	1.911
Model -Averaged	1.248	0.724	1.99	3.767	4.052	1.017	1.127	1.658	1.060	1.408	1.493
Bias^2 of model averaged	0.062	0.076	0.0001	0.588	0.003	0.0003	0.016	0.433	0.0036	0.166	0.243
Bias^2 of model selection	0.007	5.958	3.433	2.248	0.008	0.235	0.760	0.060	0.785	2.962	0.830
PMSE of model selection	2.526	9.084	6.185	4.629	2.678	2.167	3.600	2.314	3.025	5.360	3.415
PMSE of model averaged	0.142	0.195	0.111	0.722	0.119	0.158	0.154	0.655	0.096	0.275	0.497
CI of model selection	0.569	0.278	0.638	0.959	0.951	0.905	0.859	0.869	0.869	0.854	0.844
CI of model averaging	0.999	0.945	0.999	0.999	0.999	0.978	0.993	0.998	0.998	0.999	0.989

Appendix (B)
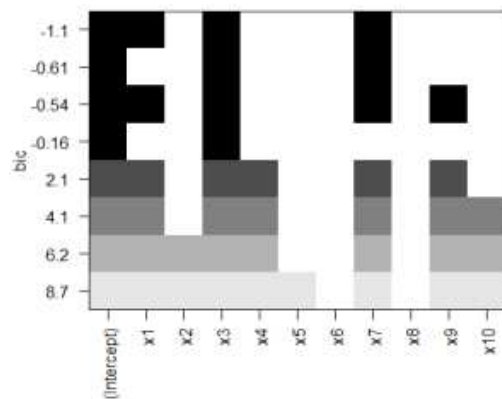
Figure 1.1. *n=20.*
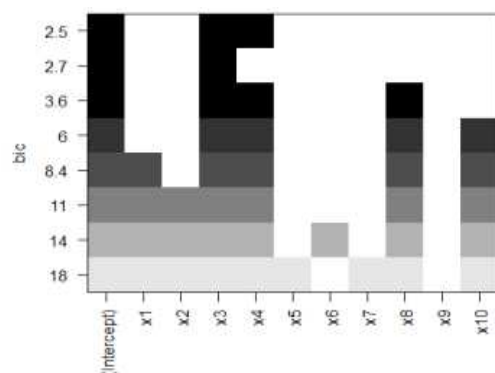
Figure 1.2. *n=30.*
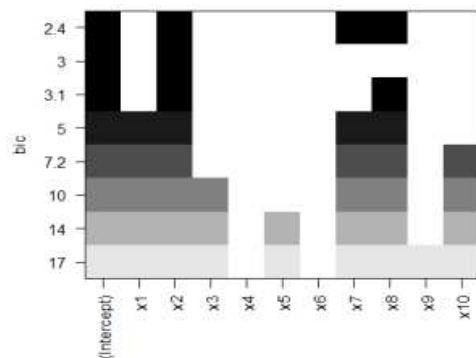
Figure 1.3. *n=50.*
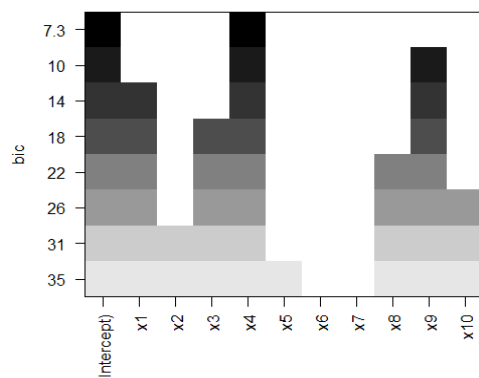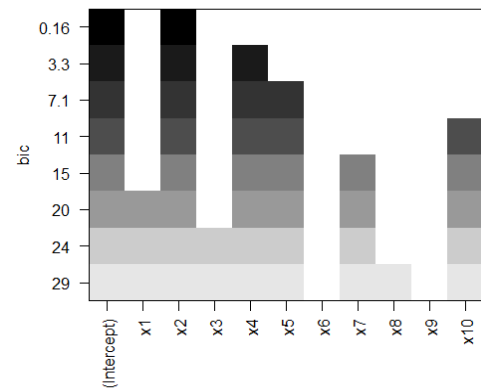

Figure 1.4. *n=100.*
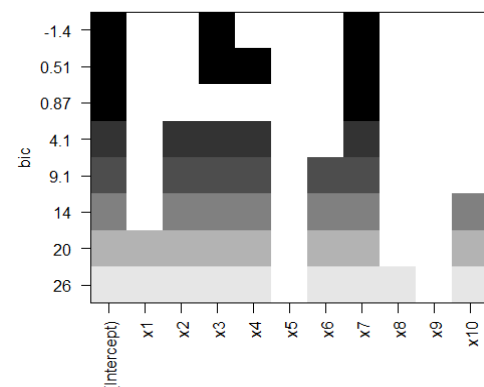

Figure 1.5. *n=200.*

Figure 1.6. *n=500.*

Figure (1). *All Possible Regression with different sample sizes, n={20,30,50,100,200,500) when SNR is 1:0.5.*
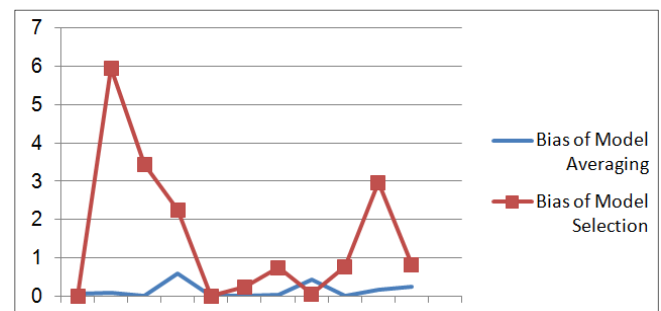

Figure (2). *Bias of Model Selection and Model Averaging of the Regression Coefficient estimates for the multivariate linear regression model(42).*

References

[1] Akaike, H., 1973. "Information theoryas an extension of the maximum likelihood principle", P.267-281 in B.N P Petrov, and FCsaki, (eds.) *Second International Symposium on Information Theory. Akademiai Kiado, Budapest.*

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

[3] Al-Subaihi, Ali A. (2007). "Variable selection in multivariate regression using SAS/IML, Saudi Arabia.

[4]   Anderson, D. R., Burnham,K.P., and White, G.C.(1998). "Comparison of AIC and CAIC *for model selection and statistical inference from capture-recapture studies". Journal of Applied Statistics 25, 263-282.*

[5]   Bozdogan, H. (1987). "Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika 52, 345-370.*

[6]   Bozdogan, H. (1994). "Editor's general preface". *Pages ix-xii in H. Bozdogan (ed.) Engineering and Scientific applicatioins. Vol., 3, Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach. Kluwer Academic Publishers, Dordrecht, the Netherlands.*

[7]   Buckland, S.T., Burnham, K.P., and Augustin, N.H. (1997). "Model selection: an integral part of inference", *Biometrics 53, 603-618.*

[8]   Burnham, K.P.; Anderson, D.R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.), Springer-Verlage ISBN 0-387-9536-7.

[9]   Burnham, K.P.; Anderson, D.R. (2004). "Multimodel inference: understanding AIC and BIC in Model Selection", Sociological Methods and research 33:201-208.

[10]   Cavanaugh, J.E., and Neath, A.A. (1999). "Generlizing the derivation of the Schwarz information criterion", *Communication in Statistics-Theory and Methods 28, 49-66.*

[11]   Cavanaugh, J.E., and Shumway, R.H. (1997). "A bootsrap variant of AIC for state-space model selection", *Statistica Sinica 7, 473-496.*

[12]   Deleauw, j. (1992). "Introduction to Akaike (1973) information theory and an extension of the maximum likelihood principle". *Pages 599-609 in S. Kotz, and N.L. Johnson (eds) Breakthroughs in statistics. Vol.1. Springer Verlag, London.*

[13]   Hannan, E.J., and Quinn, B.G. (1979). "The determination of the order of an autoregression", *Journal of the Royal Statistical Society, Series B 41, 190-195.*

[14]   Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T. (1999). "Bayesian model averaging: a tutorial (with discussion)", *Statistical Science 14, 382-417.*

[15]   Hurvich, C.M., and Tsai, (1989). "Regression and time series model selection in small samples", *Biometrika 76, 297-307.*

[16]   Hurvich, C.M., and Tsai, (1991). "Bias of the corrected AIC criterion for underfitted regression and time series models", *Biometrika 78, 499-509.*

[17]   Hurvich, C.M., and Tsai, (1995). "Model selection for extended quasi-likelihood models in small samples", *Biometrics 51, 1077-1084.*

[18]   Kapur, J.N., and Kesavan, H.K. (1992). *Entropy optimization principles with applications.* Academic press, London.

[19]   Kullback, S. (1959). *Information theory and statistics.* John Wiley and Sons, New York, NY.

[20]   Kullback, S., and Leibler, R.A. (1951). On information and sufficiency, *Annals of Mathematical Statistics 22, 79-86.*

[21]   Leamer, E.E. (1978). *Specification searches: ad hoc inference with nonexperimental data*, John Wiley and Sons, New York, NY.

[22]   Lebreton, J-D., Burnham, K.P., Clobert, J., and Anderson, D.R. (1992). "Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies", *Ecological Monograph 62, 67-118.*

[23]   Mallows, C.L. (1973). "Some comments on C_p", *Technometrics 12, 591-612.*

[24]   Mallows, C.L. (1995). "More comments on C_p", *Technometrics 37, 362-372.*

[25]   Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986). Akaike information criterion statistics, KTK Scientific Publishers, Tokyo, Japan.

[26]   Schwarz, Gideon E. (1978). "Estimating the dimension of a model". Annals of statistics 6 (2):461-464.

[27]   Shibata, R. (1983). A theoretical view of the use of AIC. Pages 237-244 in O.D. Anderson (ed.) Time series analysis: theory and practice. Elsevier Science Publication, North-Holland, the Netherlands.

[28]   Stone, C.J. (1982). "Local asymptotic admissibility of a generalization of Akaike's model selection rule", *Annals of the Institute of Statistical Mathematics Part A 34, 123-133.*

[29]   Sugiura, N. (1978). "Further analysis of the data by Akaike's information criterion and the finite corrections", *Communications in Statistics, Theory and Methods A7, 13-26.*

[30]   Takeuchi, K. (1976). "Distribution of informational statistics and a criterion of model fitting", *Suri-Kagaku (Mathematic Sciences) 153, 12-18. (In Japanese).*