A Note On the Unification of the Akaike Information Criterion

Peide Shi
Department of Probability and Statistics
Peking University, Beijing 100871
P. R. China

Chih-Ling Tsai
Graduate School of Management
University of California
Davis, California 95616-8608
U. S. A.

September, 1997

SUMMARY

In order to measure the distance between a robust function evaluated under the true regression model and under a fitted model, we propose a generalized Kullback-Leibler information. Using this generalization we have developed three robust model selection criteria, $AICR^*$, $AICCR^*$ and $AICCR$, that allow the selection of candidate models that not only fit the majority of the data, but also take into account non-normally distributed errors. These two criteria, $AICR^*$ and $AICCR$, can unify most existing Akaike information criteria; three examples of such unification are given. Simulation studies are presented to illustrate the relative performance of each criterion.

Some key words: AIC; $AICC$; $AICR^*$; $AICCR$; Kullback-Leibler information; Robust model selection.

JRSS(b) in press

1. Introduction

For the last three decades the Akaike Information Criterion (AIC) as well as a number of variants, have been widely used in the area of model selection. Because AIC can be a drastically biased estimate of the expected Kullback-Leibler information of the fitted model, as Hurvich and Tsai (1989) have shown, they developed one such variant for regression models based on the normal distribution, $AICC$, a corrected version of AIC which is more nearly unbiased and which provides better model choices in small samples. For non-normal regression models, Ronchetti (1985) has proposed a robust model selection criterion $AICR$, a generalized Akaike information criterion that takes into account the presence of outliers.

The aim of this paper is first to propose a generalized Kullback-Leibler information that can measure the discrepancy between a robust function evaluated under both the true model and fitted models. Next, we use this generalized Kullback-Leibler information to obtain three generalized Akaike information criteria ($AICR^*$, $AICCR^*$, and $AICCR$) that will allow us to choose the best model with two considerations: good model fit for the majority of the data, and accommodation of non-normally distributed errors. The relationships between the two unified criteria, $AICR^*$ and $AICCR$, and each of the AIC and $AICC$ criteria are illustrated with three analytical examples using location-scale regression models. Simulation results are also presented which show that $AICCR^*$ outperforms the others.

2. The Unification of the Akaike Information Criterion Family

Suppose that the data y_i $(i = 1, ..., n)$ are independently generated from a true model with the probability density function $f(y_i; x_{0i}'\beta_0, \sigma_0)$, where x_{0i} is a $p_0 \times 1$ vector of explanatory variables, β_0 is a $p_0 \times 1$ vector of parameters, and σ_0 is a scale parameter. Because in practice the true model is unknown, we would first consider fitting the data with a family of candidate models with probability density functions $f(y_i; x_i'\beta, \sigma)$, where x_i is a $p \times 1$ vector, β is a $p \times 1$ vector of unknown parameters, and σ is a scale parameter. Hence, the true model and the candidate models will differ both in their location (regression) and scale parameters.

After fitting the candidate models to the data, we next assume that the unknown regression parameter β and the scale parameter σ can be estimated by minimizing the function

$$D(\beta, \sigma) = \sum_{i=1}^{n} \rho\{(y_i - x_i'\beta)/\sigma\}, \tag{1}$$

where ρ is a suitable function chosen with respect to the data and the candidate model structure. The resulting parameter estimators, $\hat{\beta}$ and $\hat{\sigma}$, are usually called M-estimators. Note that by choosing particular ρ functions, a few well known parameter estimators can be obtained (*e.g.*, the least squares estimators, L_p-norm estimators, and maximum likelihood estimators).

To assess the discrepancy between the function in (1) evaluated under the true model and candidate models, we propose the following generalized Kullback-Leibler information:

$$\Delta(\beta, \sigma) = E_0\{D(\beta, \sigma)\}, \tag{2}$$

where E_0 denotes the expectation evaluated under the true model. If $\rho = -2log(f)$, then equation (2) gives the classical Kullback-Leibler information. Next we make the following three assumptions:

Assumption 1. The ρ is a second-order differentiable function;
Assumption 2. There is a $p \times p_0$ matrix M such that $M\beta_0 = \beta^*$; and
Assumption 3. For any given $\sigma > 0$, $E_0[\psi(\varepsilon_1/\sigma)] = 0$, where $\varepsilon_1 = (y_1 - x'_{01}\beta_0)$ and $\psi(t) = \frac{d\,\rho(t)}{d\,t}$.

Note that Assumption 3 is directly adopted from Huber (1981, p. 165). Under Assumptions 1 and 2, the second-order Taylor expansion of $\rho\{(y_i - x'_i\beta)/\sigma\}$ around β^* leads to the following approximation of the generalized Kullback-Leibler information in (2):

$$\Delta(\beta, \sigma) \simeq \Delta(\beta^*, \sigma) + \dot{\Delta}(\beta^*, \sigma)(\beta - \beta^*) + (\beta - \beta^*)'\ddot{\Delta}(\beta^*, \sigma)(\beta - \beta^*)/2, \tag{3}$$

where $\dot{\Delta}(\beta^*, \sigma) = E_0\left\{\partial D(\beta, \sigma)/\partial\beta\right\}_{\beta=\beta^*}$, and $\ddot{\Delta}(\beta^*, \sigma) = E_0\left\{\partial^2 D(\beta, \sigma)/\partial\beta\partial\beta'\right\}_{\beta=\beta^*}$. Under Assumption 3, Equation (3) becomes

$$\Delta(\beta, \sigma) \simeq \Delta(\beta^*, \sigma) + (\beta - \beta^*)'\ddot{\Delta}(\beta^*, \sigma)(\beta - \beta^*)/2. \tag{4}$$

A reasonable criterion for judging the quality of candidate models with respect to the data is $E_0\{\Delta(\hat{\beta}, \hat{\sigma})\}$, where $\hat{\beta}$ and $\hat{\sigma}$ are the M-estimators of β and σ, respectively (Huber, 1981, p. 176). Since, in practice, $E_0\{\Delta(\hat{\beta}, \hat{\sigma})\}$ cannot be computed, we obtain two versions of its approximation, which are given in the following theorem.

Theorem. Under above three assumptions, we have

$$E_0\{\Delta(\hat{\beta}, \hat{\sigma})\} \simeq E_0\{\Delta(\beta^*, \hat{\sigma})\} + \text{trace}(A)/2 \tag{5}$$

and

$$E_0\{\Delta(\hat{\beta}, \hat{\sigma})\} \simeq E_0\left(\sum_{i=1}^{n} E_0[\rho\{(y_i - x'_i\hat{\beta})/\sigma\}]_{\hat{\sigma}}\right) + \text{trace}(A), \tag{6}$$

where $A = c_1 E_0\{\ddot{\Delta}(\beta^*, \hat{\sigma})\}(X'X)^{-1}$, $c_1 = \text{Var}\{\psi(\varepsilon^*)\}/[E_0\{\dot{\psi}(\varepsilon^*)\}]^2$ with $\varepsilon^* = \varepsilon_1/\sigma_0$, $\dot{\psi}(t) = \frac{d\,\psi(t)}{d\,t}$, X is an $n \times p$ matrix with the ith row equal to x'_i, and Var$\{.\}$ indicates the variance evaluated under the true model.

Proof. Applying Silvapulle's (1985, p. 1494) results stating that $\sqrt{n}(\hat{\beta} - \beta^*)$ and $\ddot{\Delta}(\beta^*, \hat{\sigma})$ are asymptotically independent, we can replace Var$\{\sqrt{n}(\hat{\beta} - \beta^*)\}$ by its asymptotic covariance. Hence, $E_0\{\Delta(\hat{\beta}, \hat{\sigma})\}$ can be approximated as (5). Furthermore, applying Silvapulle's (1985) Theorem 1 and given the asymptotic covariance of Var$\{\sqrt{n}(\hat{\beta} - \beta^*)\}$, we have

$$E_0\{\Delta(\beta^*, \hat{\sigma})\} \simeq E_0\left(\sum_{i=1}^{n} E_0[\rho\{(y_i - x'_i\hat{\beta})/\sigma\}]_{\hat{\sigma}}\right) + \text{trace}(A)/2.$$

Then, substituting the above equation into (5) leads to (6). This completes the proof of Theorem 1.

Equations (5) and (6) unify most existing Akaike information criteria, and also allow the selected model to be more robust against outliers and other departures from the distributional assumptions of the model. In order to show how equations (5) and (6) can be useful, in the next section we will illustrate with some examples using location-scale regression models.

3. Model Selection Criteria for Location-Scale Regression Models

3.1 Model Selection Criteria $AICR^*$ and $AICCR$

Assume that the data y_i are generated from the following true model:

$$y_i = x_{0i}'\beta_0 + \epsilon_i, \qquad (i = 1, ..., n), \tag{7}$$

and then fitted with the family of candidate models described by

$$y_i = x_i'\beta + e_i, \qquad (i = 1, ..., n),$$

where x_{0i}, x_i, β_0 and β are as defined in Section 2, ϵ_i and e_i are independent and identically distributed, and have location-scale probability density functions of $f(\epsilon_i, \sigma_0) = \sigma_0^{-1} g(\epsilon_i/\sigma_0)$ and $f(e_i, \sigma) = \sigma^{-1} g(e_i/\sigma)$, respectively. Also, $g(.)$ is a positive function, and σ_0 and σ are the scale parameters of the true model and the candidate model, respectively.

After a simple calculation, we can see that

$$\ddot{\Delta}(\beta^*, \sigma) = \{\int \dot{\psi}(t\sigma_0/\sigma) g(t) d\, t\} X'X\sigma_0^2/\sigma^2.$$

Hence, trace(A) from equations (5) and (6) becomes

$$\text{trace}(A) = E_0 \left(\frac{\sigma_0^2}{\hat{\sigma}^2} \int \dot{\psi}(\frac{t\sigma_0}{\hat{\sigma}}) g(t) d\, t \right) \frac{p \int \psi^2(t) g(t) d\, t}{(\int \dot{\psi}(t) g(t) d\, t)^2}. \tag{8}$$

Substituting (8) into equation (5), we obtain the generalized $AICC$ model selection criterion:

$$AICCR = E_0 \{\Delta(\beta^*, \hat{\sigma})\} + E_0 \left(\frac{\sigma_0^2}{\hat{\sigma}^2} \int \dot{\psi}(\frac{t\sigma_0}{\hat{\sigma}}) g(t) d\, t \right) \frac{p \int \psi^2(t) g(t) d\, t}{2(\int \dot{\psi}(t) g(t) d\, t)^2}. \tag{9}$$

If the first term of equation (9) can be simplified so that it does not depend on the true unknown parameter β^*, then we may consider replacing the non-pivotal quantity by its unbiased estimator (see Examples 2 and 3 in Section 3.2). If the first term cannot be simplified and it depends on the true parameter β^*, then we may consider replacing it with $\sum_{i=1}^n \rho\{(y_i - x_i'\hat{\beta})/\hat{\sigma})\}$. However, this quantity equals $n/2$ if $\rho(t) = t^2/2$. To avoid this situation, we suggest replacing $\hat{\sigma}$ given above by $\hat{\sigma}(full)$. The $\hat{\sigma}(full)$ is the scaled parameter estimate of σ which is calculated through the full model (i.e., the model

includes all explanatory variables). The same suggestion is also applied to the first term of equation (10) given below. Finally, if the first component of the second term (the penalty term) cannot be simplified so that it only depends on the sample size and the number of unknown parameters, then $AICCR$ may not perform well (see simulation results given in Section 4). This is primarily because $\hat{\sigma}$ may not be a good estimate of σ_0.

If we next substitute (8) into equation (6) by replacing σ_0 with its consistent estimator $\hat{\sigma}$, and then use $\sum_{i=1}^{n} \rho\{(y_i - x_i'\hat{\beta})/\hat{\sigma}\}$ to estimate the following expectation,

$$E_0\left(\sum_{i=1}^{n} E_0[\rho\{(y_i - x_i'\hat{\beta})/\sigma\}]_{\hat{\sigma}}\right),$$

we can also obtain the generalized AIC selection criterion:

$$AICR^* = \sum_{i=1}^{n} \rho\{(y_i - x_i'\hat{\beta})/\hat{\sigma}\} + p\frac{\int \psi^2(t)g(t)dt}{\int \dot{\psi}(t)g(t)dt}. \tag{10}$$

The $AICR^*$ given in equation (10) is in fact the same as Ronchetti's (1985) $AICR$ when his α is allowed to be the ratio appearing in the last term of equation (10), $\frac{\int \psi^2(t)g(t)dt}{\int \dot{\psi}(t)g(t)dt}$. Note that Ronchetti's penalty term arises from applying an argument of Stone (1977).

In order to reduce the bias of AIC, Hurvich and Tsai (1989 and 1995) obtained a corrected version of AIC, $AICC$, where the difference between the two lies in the penalty term. In fact, the penalty term of $AICC$ is equal to the penalty term of AIC multiplied by $n/(n-p-2)$ (a detailed explanation is given at the end of Example 2). Hence, we can directly apply this approach to obtain the following corrected version of $AICR^*$:

$$AICCR^* = \sum_{i=1}^{n} \rho\{(y_i - x_i'\hat{\beta})/\hat{\sigma}\} + \{n/(n-p-2)\}p\frac{\int \psi^2(t)g(t)dt}{\int \dot{\psi}(t)g(t)dt}.$$

It can be seen from the simulation studies presented in Section 4 that $AICCR^*$ performs better than $AICR^*$. We next turn to some examples in order to illustrate how the generalized $AICR^*$ and $AICCR$ criteria unify the relationships between Akaike information criteria.

3.2 Examples

Three analytical examples are given in this subsection: the first one considers $AICR^*$ and its relationship to AIC. The second one studies the relationship between $AICCR$ and $AICC$. The last one builds up a relationship between $AICCR$ and the Monte Carlo approach to $AICC$ proposed by Hurvich and Tsai (1990) and then used by Hurvich, Shumway and Tsai (1990).

Example 1: $AICR^*$ **and** AIC

Let $\rho\{(y_i - x_i'\beta)/\sigma\} = -2\ln[f\{(y_i - x_i'\beta), \sigma\}]$, where $f(t,\sigma) = \sigma^{-1}g(t/\sigma)$ and $g(\cdot)$ is the density function of the standard normal distribution. After straightforward computations,

the second term of equation (10) can be reduced to $2p$. Hence, if the constant term $n\ln(2\pi)+n-2$ is ignored, $AICR^*$ reduces to the Akaike Information Criterion:

$$AIC = n\ln(\hat{\sigma}^2) + 2(p+1).$$

Example 2: $AICCR$ **and** $AICC$

Here let $\rho\{(y_i - x_i'\beta)/\sigma\} = -2\ln[f\{(y_i - x_i'\beta), \sigma\}]$, where $f(t,\sigma) = \sigma^{-1}g(t/\sigma)$ and $g(\cdot)$ is the density function of the standard normal distribution. Also let $\hat{\beta}$ and $\hat{\sigma}^2$ be the maximum likelihood estimators of β and σ^2, respectively. Given these conditions,

$$E_0\{\Delta(\beta^*, \hat{\sigma})\} = E_0\{n\ln(\hat{\sigma}^2)\} + E_0(n\sigma_0^2/\hat{\sigma}^2) + n + n\ln(2\pi)$$

and

$$E_0(\sigma_0^2/\hat{\sigma}^2) = n/(n-p-2).$$

Furthermore, the second term of equation (9) can be simplified to $E_0(\sigma_0^2/\hat{\sigma}^2)p$. Then, replacing $E_0\{n\ln(\hat{\sigma}^2)\}$ by its unbiased estimator $n\ln(\hat{\sigma}^2)$, $AICCR$ reduces to the $AICC$ proposed by Hurvich and Tsai (1989 and 1995):

$$AICC = n\ln(\hat{\sigma}^2) + \frac{n}{n-p-2}2(p+1),$$

ignoring the constant $2n + n\ln(2\pi)$. Since modifying the penalty term of AIC to $2(p+1)$ does not affect AIC's performance, the difference between AIC and $AICC$ is the factor of $n/(n-p-2)$ that appears in their penalty terms.

Example 3: $AICCR$ **and a Monte Carlo approach to** $AICC$

Let $\rho\{(y_i - x_i'\beta)/\sigma\} = -2\ln[\sigma^{-1}g\{(y_i - x_i'\beta)/\sigma\}]$, where $g(t) = \exp(-t)\exp(-e^{-t})$ and is the density of the extreme value distribution (see Lawless, 1982, p. 298). After simple calculations we find that

$$\frac{\int \psi^2(t)g(t)dt}{(\int \dot{\psi}(t)g(t)dt)^2} = 1,$$

and

$$\int \dot{\psi}(\frac{t\sigma_0}{\hat{\sigma}})g(t)dt = 2\int \exp(-\frac{t\sigma_0}{\hat{\sigma}})g(t)dt = 2\Gamma(1+\frac{\sigma_0}{\hat{\sigma}}).$$

In addition,

$$\Delta(\beta^*, \hat{\sigma}) = n\ln(\hat{\sigma}^2) + 2nE_0(\varepsilon_1)/\hat{\sigma} + 2nE_0\left\{\exp(-\frac{\varepsilon_1}{\sigma})\right\}|_{\hat{\sigma}} = n\ln(\hat{\sigma}^2) + 2n\gamma\frac{\sigma_0}{\hat{\sigma}} + 2n\Gamma(1+\frac{\sigma_0}{\hat{\sigma}}),$$

where γ is Euler's constant. Using the above results and replacing $E_0\{n\ln(\hat{\sigma}^2)\}$ by its unbiased estimator $n\ln(\hat{\sigma}^2)$, we can reduce $AICCR$ to

$$n\ln(\hat{\sigma}^2) + 2nE_0\left\{\gamma\frac{\sigma_0}{\hat{\sigma}} + \Gamma(1+\frac{\sigma_0}{\hat{\sigma}})\right\} + pE_0\left\{\frac{\sigma_0^2}{\hat{\sigma}^2}\Gamma(1+\frac{\sigma_0}{\hat{\sigma}})\right\}. \qquad (11)$$

According to Theorem G4 in Lawless (1982, p. 538), $\frac{\sigma_0}{\hat{\sigma}}$ is pivotal. In other words, its distribution is independent of the parameters β_0 and σ_0. This means that Hurvich and Tsai's (1990) Monte Carlo approach of $AICC$ is a suitable method for computing the last two terms of equation (11).

4. Simulation Results

To compare the performance of AIC, $AICC$, $AICR^*$, $AICCR$, and $AICCR^*$, 1000 realizations were generated from the true model (7) with $p_0 = 5$, $\beta_0 = (1, 1, 1, 1, 1)'$, and $x_{0i} \sim N(0, I_5)$ for $i = 1, ..., n$. In addition, ε_i are iid random errors chosen from the distributions of $N(0, 0.7^2)$, t_4, and $0.9N(0, 1) + 0.1N(0, 7^2)$, respectively. Three sample sizes were used, n=20, 50 and 80. There are nine candidate variables, stored in an $n \times 9$ matrix of independent identically distributed normal random variables. The candidate models are linear, and are listed in columns in a sequential nested fashion. Hence, the set of candidate models includes the true model. In order to assess the effect of robust functions on model selection criteria, we used Huber's (1964) ψ-function with the bending point $c = 1.345$, as proposed by Ronchetti and Staudte (1994), to conduct our simulation studies. We also apply Huber's (1981, Section 7.8) algorithm to compute the M-estimates of β and σ.

Table 1 gives the proportions of the order selected by the various criteria. In this case, $AICCR$ has the worst performance, mainly because the first component of its second term does not depend on n and p, but does depend on $\hat{\sigma}$. We also note that, in this case, $\hat{\sigma}$ usually does not provide a good estimate of σ_0. Hence, we conclude that while $AICCR$ is of theoretical interest, we do not recommend its use in practice.

It is obvious from Table 1 that $AICC$, $AICR^*$, and $AICCR^*$ outperform AIC for all error distributions and sample sizes. Although $AICC$ selects the correct model order more often than $AICR^*$ when the error distribution is $N(0, 0.7^2)$, $AICR^*$ outperforms $AICC$ when the error distributions are t_4 (except $n = 20$) and contaminated normal. It is also not surprising that robust model selection criterion $AICCR^*$ performs better than $AICC$ when the error distributions are apart from normality. In summary, the results of the limited scope simulation study in Table 1 show that $AICCR^*$ provides the best model selections across all error distributions and sample sizes.

We next study the sensitivities of the heavy-tailed and contaminated normal densities g against the standard normal density, as we compute the penalty term component $\frac{\int \psi^2(t)g(t)dt}{\int \dot{\psi}(t)g(t)dt}$ for $AICR^*$ and $AICCR^*$. After replacing t_4 and $0.9N(0, 1) + 0.1N(0, 7^2)$ by $N(0, 1)$ when computing $AICR^*$ and $AICCR^*$, we see from Table 2 that the two criteria perform somewhat differently than under the conditions in Table 1. However, their performances are still reasonably well. Especially, $AICCR^*$ performs better than $AICC$ given in Table 1. Hence we conclude that $AICR^*$ and $AICCR^*$ can be used for model selection even when g is not known or well-specified.

INSERT TABLES 1 and 2 HERE

5. Discussion

In this paper we have proposed three criteria which unify many versions of the Akaike information criterion family. These unified criteria can be extended to various model structures with appropriate robust functions, such as autoregressive models (Hurvich and Tsai, 1989), multivariate regression models (Bedrick and Tsai, 1994), extended quasi-likelihood models (Hurvich and Tsai, 1995), and others. In addition, the expectations in (5) and (6) remain applicable to the location-scale family as long as the explanatory variables $x_1, ..., x_n$ (Section 2) are identically distributed and independent random variables, and they are also independent of the random errors. The derivations of these variations are virtually identical to those given in Section 2, except that Silvapulle's (1985) results, used for the fixed carrier case, are replaced by the corresponding Maronna and Yohai (1981) results in the random carrier case. Detailed proofs can be obtained from the authors by request.

Also, in the particular case where the true model is included among the candidate models and ($\hat{\beta}$ and $\hat{\sigma}$) are the M-estimators of (β and σ) as defined in Huber (1981, p. 176), we can then follow Hurvich andTsai's (1990) reasoning to prove that the distributions of $(\hat{\beta}-\beta^*)/\sigma_0$ and $\hat{\sigma}/\sigma_0$ are independent of β_0 and σ_0. This allows us to apply their Monte Carlo method to calculate the term

$$E_0\{\frac{\sigma_0^2}{\hat{\sigma}^2}\int \dot{\psi}(\frac{t\sigma_0}{\hat{\sigma}})g(t)d\,t\}$$

from equations (8) and (9).

ACKNOWLEDGEMENTS

The research of Peide Shi was supported by NSF of China. The research of Chih-Ling Tsai was supported by National Science Foundation grant DMS 95-10511. The helpful comments of the Associate Editor and the referee are gratefully acknowledged.

REFERENCES

Bedrick, E. J. and Tsai, C.-L. (1994). Model selection for multivariate regression in small samples. *Biometrics* **50**, 226-231.

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Stat.* **35**, 73-101.

Huber, P. J. (1981). *Robust Statistics*, New York: John Wiley & Sons.

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

Hurvich, C. M. and Tsai, C.-L. (1990). Model selection for least absolute deviations regression in small samples. *Statist. and Probab. Letters* **9**, 259-265.

Hurvich, C. M., Shumway R. and Tsai, C.-L. (1990). Improved estimators of Kullback-Leibler information for autoregressive model selection in small samples. *Biometrika* **88**, 709-719.

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

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*, New York: John Wiley and Sons.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics* **15**, 661-675.

Maronna, R. A. and Yohai, V. J. (1981). Asymptotic behavior of general M-estimates for regression and scale with random carriers. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **58**, 7-20.

Ronchetti, E. (1985). Robust Model Selection in Regression. *Statistics and Probability Letters* **3**, 21-23.

Ronchetti, E. and Staudte, R. G. (1994). A robust version of Mallows' C_p. *J. Am. Statist. Assoc.* **89**, 550-9.

Silvapulle, M. J. (1985). Asymptotic behavior of robust estimators of regression and scale parameters with fixed carriers. *Ann. Statist.* **13**, 1490-1497.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *J. R. Statist. Soc.* **B, 39**, 44-47.

Table 1. Proportions of the correct model order selected by AIC, $AICC$, $AICR^*$, $AICCR^*$ and $AICCR$ criteria in 1000 realizations.

Error Distribution	Criteria	n=20	n=50	n=80
$N(0, 0.7^2)$	AIC	0.531	0.640	0.711
	$AICC$	0.909	0.778	0.785
	$AICR^*$	0.757	0.721	0.750
	$AICCR^*$	0.942	0.827	0.814
	$AICCR$	0.399	0.430	0.413
t_4	AIC	0.450	0.683	0.670
	$AICC$	0.709	0.834	0.770
	$AICR^*$	0.526	0.850	0.796
	$AICCR^*$	0.740	0.919	0.857
	$AICCR$	0.216	0.321	0.344
$0.9N(0,1) + 0.1N(0, 7^2)$	AIC	0.361	0.609	0.694
	$AICC$	0.445	0.699	0.756
	$AICR^*$	0.555	0.819	0.825
	$AICCR^*$	0.658	0.897	0.874
	$AICCR$	0.127	0.275	0.314

Table 2. Proportions of the correct model order selected by $AICR^*$ and $AICCR^*$ criteria in 1000 realizations, where the penalties are computed with the standard normal density function.

Error Distribution	Criteria	n=20	n=50	n=80
t_4	$AICR^*$	0.517	0.775	0.723
	$AICCR^*$	0.751	0.881	0.792
$0.9N(0,1) + 0.1N(0, 7^2)$	$AICR^*$	0.505	0.742	0.749
	$AICCR^*$	0.669	0.854	0.812