ELSEVIER

Statistics & Probability Letters 39 (1998) 1–10

STATISTICS & PROBABILITY LETTERS

The weighted average information criterion for order selection in time series and regression models

Tiee-Jian Wu [a,*], Alfred Sepulveda [b]

[a] *Department of Applied Mathematics, National Donghwa University, Hualien 974, Taiwan*
[b] *Department of Mathematics, University of Houston, Houston, TX 77204, USA*

Received June 1997; received in revised form October 1997

Abstract

We propose a consistent criterion for model order selection in the model identification phase of time series and regression, based on a weighted average of an asymptotically efficient selection criterion, AICC (bias-corrected Akaike information criterion) and a consistent selection criterion, BIC (Akaike's Bayesian modification of AIC). The weights attached to AICC and BIC are optimal choices from a natural class of possible weights, and are proportional to the model-complexity penalty term of AICC and BIC, respectively. Thus, the AICC part of the criterion receives most of the weight for small sample size n, and the BIC part receives the most weight for large n. It is shown that this weighted average criterion, WIC, is essentially equivalent to AICC for small n and to BIC for large n. An extensive simulation study comparing the performance of WIC with several popular criteria has been done. It clearly shows that WIC is a very reliable and practical criterion. In particular, for small n, WIC performs as well as AICC and outperforms other criteria, and for large n, WIC performs as well as BIC and outperforms other criteria. This demonstrates the overall strength of WIC.

Keywords: AIC; AICC; Asymptotic efficiency; BIC; Consistency; Order selection; SIC

1. Introduction

Model identification involving order-selection criteria are usually based on the minimization of a loss function of the following form:

$$G(\hat{\sigma}_\varepsilon^2) + P(n,m). \quad (1)$$

Here, $P(n,m)$ is a nonnegative random variable depending directly on sample size n and the number of fitted parameters m of the candidate (or approximating) model. $P(n,m)$ measures the complexity of the candidate model and serves as a penalty term for overfitting. $G(\hat{\sigma}_\varepsilon^2)$ is a measure of goodness-of-fit of the candidate model to the data and is dependent on a sample estimator of the residual variance, $\hat{\sigma}_\varepsilon^2$. Criterion may be categorized, under certain conditions on the true and candidate models, to be asymptotically efficient, or consistent. See,

* Corresponding author.

0167-7152/98/$19.00
PII S0167-7152(98)00003-0

e.g., Shibata (1980) or Brockwell and Davis (1995, pp. 304–305) for definitions and conditions. Criteria that are asymptotically efficient are AIC (Akaike, 1974), AICC (Hurvich and Tsai, 1989), CP (Mallows, 1973), CAT (Parzen, 1977) and others. Those which have been shown to be consistent are BIC (Akaike, 1978), SIC (Schwarz, 1978), HQ (Hannan and Quinn, 1979), PLS (Rissanen, 1986), FIC (Wei, 1992) and others. The more recent ODQ (Zhang and Wang, 1994) is an example of a "separation point" method that does not minimize a given loss function. It has been shown to be consistent under mild conditions.

All these criteria have relative advantages depending upon the situation in which they are used. Recent studies have shown that the asymptotically efficient criterion AICC tends to give the best estimate of the true model order when m/n is large, and the consistent criterion BIC tends to outperform other criteria when m/n is small (Hurvich and Tsai, 1989; Lütkepohl, 1985). Therefore, large sample sizes favor BIC, while small sample sizes favor AICC. Nonetheless, it may be difficult to pick a criterion based on intermediate sample sizes. In addition, at what sample size does one switch criteria for model order selection? Factors involving one's decision depend on unknown information inherent to the true model. Hence, a criterion which eliminates this seemingly arbitrary decision should prove valuable in model building.

To achieve the above goal, we propose a weighted average of AICC and BIC, as a new order selection criterion, for both regression and time-series models. The weights attached to AICC and BIC are proportional to the model-complexity penalty terms of AICC and BIC, respectively. Thus, the AICC part of the criterion receives most of the weight for small n, and the BIC part receives the most weight for large n. These weights are optimal choices from a natural class of possible weights. This criterion, called weighted-average information criterion (WIC), is consistent and is asymptotically equivalent to BIC as $n \to \infty$. On the other hand, WIC retains characteristics of AICC for small n since it is asymptotically equivalent to AICC as $n \downarrow m+2$. We compare the performances of WIC with several popular criteria by means of simulations in which the true model is finite dimensional. It clearly shows that WIC is a very reliable and practical criterion over a wide range of models and sample sizes. In particular, for small n, WIC performs as well as AICC and outperforms BIC and other criteria, and for large n, WIC performs as well as BIC and outperforms AICC and other criteria. In addition, a maximum cut off on the allowable dimension of the candidate models is not needed for WIC. The theoretical and simulation results demonstrate the overall strength of WIC, as a whole.

Section 2 contains the definition of *WIC* and theoretical results. Section 3 presents simulation results for regression, AR, MA, and ARMA model selection.

2. The weighted average information criterion

Both AICC and BIC attempt to correct the overfitting nature of AIC. To that end, AICC is most successful at small n, whereas BIC is most successful at large n. The WIC, as proposed below, combines the strengths of AICC and BIC in some optimal way. Consequently, WIC dramatically improves AIC with any n.

The AICC, up to a constant, can be expressed as

$$\mathrm{AICC} = n \log \hat{\sigma}_{\varepsilon}^2 + A,$$

where

$$A = 2n(m+1)/(n-m-2)$$

(Brockwell and Davis, 1995). The BIC, up to a constant, can be written as

$$\mathrm{BIC} = n \log \hat{\sigma}_{\varepsilon}^2 + B,$$

where

$$B = (m-n)\log(1-m/n) + m\log n + m\log\{m^{-1}(\hat{\sigma}_X^2/\hat{\sigma}_{\varepsilon}^2 - 1)\}$$

with $\hat{\sigma}_X^2$ denoting the sample variance of the observations. Furthermore, in typical situations

$$\mathrm{BIC} = \mathrm{SIC} + \mathrm{O}(m),$$

where

$$\mathrm{SIC} = n \log \hat{\sigma}_\varepsilon^2 + m \log n \tag{2}$$

and $\mathrm{O}(m)$ denotes a term which is functionally independent of n (Priestley, 1982, p. 376; Schwarz, 1978). The proposed criterion, WIC, is expressed as

$$\mathrm{WIC} = \{A/(A+B)\}\mathrm{AICC} + \{B/(A+B)\}\mathrm{BIC} = n \log \hat{\sigma}_\varepsilon^2 + W,$$

where

$$W = (A^2 + B^2)/(A + B). \tag{3}$$

Note that W is the penalty term, which is always between A and B. We now explain why WIC should work, as given by (i)–(iv) below. Results (i) and (ii) measure the asymptotic closeness among the penalty terms and among the criteria, (iii) gives the consistency of WIC, and (iv) establishes the optimality of WIC (or equivalently, of its weights). In what follows, n is used as if it were real in the convergence statement "$n \downarrow m+2$". The purpose of this usage is to help us to get a real feeling of the closeness between A and W, and between AICC and WIC for small sample size. Also, unless otherwise specified, all convergence involving random variables are in the strong sense:

(i) As $n \to \infty$, the penalty terms $A \to 2(m+1)$ and $B \to \infty$, so that the weights $B/(A+B) \to 1$ and $A/(A+B) \to 0$. Hence WIC behaves like BIC for large n. On the other hand, as $n \downarrow m+2$, the penalty terms $A \to \infty$ and $B \to m \log\,(m+2) + \mathrm{O}(m)$, so that $A/(A+B) \to 1$ and $B/(A+B) \to 0$. Hence *WIC* behaves like AICC for small n.

(ii) The difference between the penalty terms, B and W, is given by $B - W = A(B-A)/(A+B)$, so that as $n \to \infty$,

$$(B - W)/B = \mathrm{O}(A/B) = \mathrm{O}(\log^{-1} n),$$
$$(\mathrm{BIC} - \mathrm{WIC})/\mathrm{BIC} = (B - W)/\mathrm{BIC} = \mathrm{O}(A/n) = \mathrm{O}(n^{-1}) \tag{4}$$

and, consequently, $W \sim B$ and WIC $\sim$ BIC (here "$a \sim b$" means that a and b are asymptotically equivalent, i.e., $a/b \to 1$). On the other hand, the difference between the penalty terms, A and W, is given by $A - W = B(A-B)/(A+B)$, so that as $n \downarrow m+2$,

$$(A - W)/A = \mathrm{O}(B/A) = \mathrm{o}(1),$$
$$(\mathrm{AICC} - \mathrm{WIC})/\mathrm{AICC} = (A - W)/\mathrm{AICC} = \mathrm{O}(B/A) = \mathrm{o}(1)$$

and, consequently, $W \sim A$ and WIC $\sim$ AICC.

(iii) The criterion WIC is weakly consistent under general conditions. This can be seen from equations (2) and (4), and from Hannan (1980) (see his paper for conditions).

(iv) Among possible weighted averages of AICC and BIC, it is natural to consider the class of criteria $\mathscr{F} = \{\mathrm{WIC}(r) : r \geqslant 1\}$, where

$$\mathrm{WIC}(r) = \{A^r/(A^r + B^r)\}\mathrm{AICC} + \{B^r/(A^r + B^r)\}\mathrm{BIC} = n \log \hat{\sigma}_\varepsilon^2 + W_r,$$

and

$$W_r = (A^{r+1} + B^{r+1})/(A^r + B^r)$$

(see Remark 1 below for a more rigorous derivation of the class $\mathscr{F}$). Evidently, W_r is between A and B. Further, it is easy to see that for $r \geqslant 1$, the WIC(r), along with its weights and penalty term, possess asymptotic properties which are similar to those stated in (i)–(iii) above. Thus, every member of $\mathscr{F}$ appears to be a feasible criterion. Let us consider a measure of the overall discrepancy of W_r from A and B (or equivalently, of WIC(r) from AICC and BIC), namely,

$$d(r) = (A - W_r)^2 + (B - W_r)^2 = (\mathrm{AICC} - \mathrm{WIC}(r))^2 + (\mathrm{BIC} - \mathrm{WIC}(r))^2, r \geqslant 1.$$

For any sample size this measure puts equal weight on $(A - W_r)^2$ and $(B - W_r)^2$, which is conservative but sensible due to the uncertainty of the circumstances. The quantity $d(r)$ can be thought of as a quadratic loss function. The goal is to minimize $d(r)$ over $r \geqslant 1$. The derivatives of W_r and $d(r)$ are, respectively, equal to

$$W_r' = (A^r + B^r)^{-2} A^r B^r (A - B) \log(A/B)$$

and

$$d'(r) = 2W_r'(2W_r - A - B) = 2W_r'(A^r - B^r)(A - B)/(A^r + B^r).$$

Evidently, $d'(r)$ is positive for all $r \geqslant 1$. Therefore, the minimum of $d(r)$ occurs at the left endpoint $r = 1$. This shows that for any sample size, WIC (the special case $r = 1$ here) is the optimal choice from the class $\mathscr{F}$ in the sense of minimizing $d(r)$.

It can be easily shown that for any $r < 1$, $(B - W_r) = \mathrm{O}(B^{1-r}) \to \infty$ as $n \to \infty$, and $(A - W_r) = \mathrm{O}(A^{1-r}) \to \infty$ as $n \downarrow m+2$, so that W_r does not have the desired asymptotic closeness to A or B. Therefore, the case $r < 1$ is excluded from the above consideration.

It is worthwhile to consider another measure of the overall discrepancy of W_r from A and B, namely,

$$d_s(r) = c_s(A - W_r)^2 + (1 - c_s)(B - W_r)^2,$$

where

$$c_s = A^s/(A^s + B^s), \quad s \geqslant 1.$$

Evidently, $c_s \to 1$ as $n \downarrow m+2$, and $c_s \to 0$ as $n \to \infty$. Thus, $d_s(r)$ applies the most weight on $(A - W_r)^2$ for small n, and on $(B - W_r)^2$ for large n, and so $d_s(r)$ is not so conservative as is $d(r)$. Upon differentiating $d_s(r)$ with respect to r, we get

$$d_s'(r) = 2W_r'(W_r - c_s A - (1 - c_s)B) = 2W_r'(A^r B^s - A^s B^r)(A - B)/\{(A^r + B^r)(A^s + B^s)\}.$$

Consequently, $d_s'(r) >, =$ or < 0 according as $r >, =$ or $< s$, and the minimum of $d_s(r)$ occurs at $r = s$. Therefore, the WIC(s) is the optimal choice from the class $\mathscr{F}$ in the sense of minimizing $d_s(r)$. This provides a different point of view for the optimality of WIC (which corresponds to the special case $s = 1$ here).

Remark 1. Instead of choosing the class $\mathscr{F}$ in advance, one may ask the following question: What are the conditions on the weight $0 < x < 1$, such that the linear combination

$$L(x) = x\,\mathrm{AICC} + (1 - x)\mathrm{BIC} = n \log \hat{\sigma}_\varepsilon^2 + xA + (1 - x)B$$

is essentially equivalent to AICC as $n \downarrow m+2$ and to BIC as $n \to \infty$? To achieve this, a necessary condition is that $x \to 1$ as $n \downarrow m+2$ and $x \to 0$ as $n \to \infty$. Furthermore, by the arguments in (i) above, we see that for any $t > 1$,

$$(\mathrm{AICC} - L(x))^t = (1 - x)^t (A - B)^t = (1 - x)^t (A - \mathrm{O}(m))^t < (1 - x)^t A^t \text{ as } \quad n \downarrow m+2$$

and

$$(\mathrm{BIC} - L(x))^t = x^t(B - A)^t = x^t(B - \mathrm{O}(m))^t < x^t B^t \text{ as } \quad n \to \infty.$$

It follows that $(\mathrm{AICC} - L(x))^t = \mathrm{o}(1)$ if and only if $(1-x)^t A^t = \mathrm{o}(1)$ as $n \downarrow m+2$, and $(\mathrm{BIC} - L(x))^t = \mathrm{o}(1)$ if and only if $x^t B^t = \mathrm{o}(1)$ as $n \to \infty$. Therefore, $(1-x)^t A^t$ should be as small as possible at small n and $x^t B^t$ should be as small as possible at large n. Now, for any sample size n, one may not be able to ascertain whether n is small or large, since these are relative terms that depend on the unknown complexity of the true model. Thus, at any n, a reasonable step to take, is to choose x such that

$$D_t(x) = x^t B^t + (1-x)^t A^t$$

is minimized. This scheme obtains a balance between minimizing $x^t B^t$ and minimizing $(1-x)^t A^t$. Now, a simple calculation shows that for any $t > 1$, $D_t(x)$ is minimized at $x_r = A^r/(A^r + B^r)$ where $r = t/(t-1)$. This explains the rationale in using $\mathscr{F}$ as the candidate family for the optimality problem discussed in (iv) above, since $\mathscr{F}$ is the closure of $\{L(x_r) : r = t/(t-1), t > 1\}$.

We conclude this section by noting that in (3) the SIC and its penalty term $m\log(n)$ may be substituted for BIC and B, respectively, to give an asymptotically equivalent version of WIC. Denote this version by $\mathrm{WIC_S}$. The $\mathrm{WIC_S}$ is computationally simple and should perform nearly as well as WIC for small or large n. However, its performance is worse than that of WIC for intermediate or moderately large n.

3. Simulations

We conducted extensive simulations to investigate the performance of WIC with respect to other criteria including AICC, $\mathrm{WIC_S}$, BIC, ODQ, SIC, HQ, PLS, and FIC. The FIC and PLS criteria were used only in the regression case, while the ODQ and HQ criteria were used only in the time series case. We generated 100 realizations of samples at several different sizes from time series or regression models with various structures. For each realization, parameters and residual variance of the candidate models were estimated by the maximum likelihood method (the PEST program from the ITSM Package, Brockwell et al., 1995, was used), and the above- mentioned criteria were used to select from among the candidate models. Out of the 100 realizations the frequencies of the model orders selected were tabulated for each criterion, sample size, and model.

Specifically, we generated 100 realizations from the following time series models:

AR(2): $x_t = 0.99x_{t-1} - 0.8x_{t-2} + \varepsilon_t,$
AR(3): $x_t = -0.95x_{t-1} + x_{t-2} + 0.95x_{t-3} + \varepsilon_t,$
AR(7): $x_t = 0.1x_{t-1} + 0.5x_{t-7} + \varepsilon_t,$
MA(1): $x_t = \varepsilon_t + 0.95\varepsilon_{t-1},$
ARMA(2,2): $x_t = 0.99x_{t-1} - 0.8x_{t-2} + \varepsilon_t + 0.95\varepsilon_{t-1} + \varepsilon_{t-2},$

where $\varepsilon_t \sim \mathrm{N}(0,1)$, $t = 0,1,\dots,n-1$, are uncorrelated Gaussian random variables. The AR(2), AR(3), and MA(1) models were used as in Hurvich and Tsai (1989), and the AR(7) model was used as in Zhang and Wang (1994). Here, the AR(2) model is stationary. The AR(3) model is nonstationary. The MA(1) model is invertible, but its characteristic root is close to the unit circle. The ARMA(2,2) model is stationary but noninvertible. Tables 1–5 show the frequencies of model orders selected by the various criteria. Seven different sample sizes were generated, $n = 15,20,25,30,35,50,100$. To save space, only the four cases $n = 15,25,35,100$ were included. Maximum model order cut -off of 10 was also utilized. Table 5 demonstrates the applicability of model selection criteria when considering mixed ARMA models. The alternative candidates in Table 5 include a variety of pure AR, pure MA, and mixed ARMA models. Fig. 1 plots the relative frequency curves of correct model order selections of the criteria, as functions of n, from 1000 realizations of samples of sizes n = 15, 25, 35, 50, 100, respectively, from the AR(3) model. Fig. 2 plots the average values of the

Table 1
Number of time-series model order selections out of 100 realization under an AR(2) model (n = 15, 25, 35, 100). Maximum order = 10

	Model order			
Criterion	1	2	3–5	6–10
AICC	8, 5, 3, 0	69, 78, 84, 95	18, 14, 13, 5	5, 3, 0, 0
WIC	7, 5, 1, 0	68, 77, 90, 99	20, 14, 9, 1	5, 4, 0, 0
WIC_S	8, 5, 2, 0	65, 75, 84, 95	21, 16, 11, 5	6, 4, 3, 0
BIC	7, 4, 0, 0	65, 77, 91, 99	23, 16, 9, 1	5, 3, 0, 0
SIC	9, 6, 2, 0	61, 72, 85, 95	24, 17, 11, 5	6, 5, 2, 0
HQ	3, 2, 0, 0	51, 61, 82, 93	22, 18, 14, 7	24, 19, 4, 0
ODQ	5, 1, 0, 0	58, 69, 81, 93	32, 25, 15, 7	5, 5, 4, 0

Table 2
Number of time-series model order selections out of 100 realizations under an AR(3) model. Maximum order = 10 (n = 15, 25, 35, 100).

	Model order			
Criterion	1–2	3	4–5	6–10
AICC	19, 3, 3, 0	45, 70, 82, 95	18, 16, 14, 6	18, 11, 1, 0
WIC	8, 5, 1, 0	43, 66, 91, 99	24, 15, 8, 1	25, 14, 0, 0
WIC_S	9, 6, 2, 0	31, 60, 83, 95	44, 29, 10, 5	16, 5, 5, 0
BIC	8, 4, 0, 0	30, 58, 91, 99	25, 21, 9, 1	37, 17, 0, 0
SIC	9, 3, 2, 0	19, 44, 83, 95	21, 12, 10, 5	51, 41, 5, 0
HQ	5, 4, 2, 0	10, 39, 80, 92	21, 14, 11, 8	64, 43, 7, 0
ODQ	8, 1, 0, 0	29, 54, 81, 93	33, 31, 14, 7	30, 14, 5, 0

Table 3
Number of time-series model order selections out of 100 realizations under an AR(7) model. Maximum order = 10 (n = 15, 25, 35, 100).

	Model order			
Criterion	1–3	4–6	7	8–10
AICC	4, 3, 1, 0	12, 10, 9, 6	48, 60, 68, 82	36, 27, 22, 12
WIC	4, 3, 1, 0	14, 12, 8, 0	47, 56, 76, 95	35, 29, 15, 5
WIC_S	7, 6, 10, 0	20, 23, 21, 12	44, 49, 51, 73	29, 22, 18, 15
BIC	4, 2, 0, 0	25, 23, 6, 0	39, 53, 82, 95	32, 22, 12, 5
SIC	4, 2, 6, 0	12, 17, 15, 11	25, 34, 45, 73	59, 47, 34, 16
HQ	2, 1, 0, 0	9, 7, 5, 0	24, 33, 42, 72	65, 59, 53, 28
ODQ	12, 10, 0, 0	31, 27, 11, 4	30, 40, 71, 94	27, 23, 18, 2

criteria, as functions of m, from the 100 realizations for the AR(2) case, while the maximum model order cut-off was 20.

Some conclusions may be warranted from Tables 1–5, and from Fig. 1. Firstly, the WIC exhibited robustness and stability in comparison to the aforementioned criteria in this study. The WIC was either as good or came in a strong second throughout every sample size and model order scenario considered. Secondly, for any particular model included, the relative frequency of successfully selecting the correct model order from the

Table 4
Number of time-series model order selections out of 100 realizations under an MA(1) model. Maximum order = 10 (n = 15, 25, 35, 100).

Criterion	Model order: 1	3–5	6–10
AICC	58, 66, 71, 89	30, 25, 23, 11	12, 9, 6, 0
WIC	55, 64, 75, 99	32, 23, 18, 1	13, 13, 7, 0
WIC_S	52, 60, 72, 96	32, 26, 20, 4	16, 14, 8, 0
BIC	18, 51, 79, 99	53, 36, 18, 1	29, 13, 3, 0
SIC	16, 42, 75, 97	50, 48, 17, 3	34, 10, 8, 0
HQ	17, 40, 71, 95	25, 21, 12, 4	58, 39, 17, 1
ODQ	34, 50, 72, 97	42, 39, 24, 3	24, 11, 4, 0

Table 5
Number of time-series model order selections out of 100 realizations under an ARMA(2,2) model. Maximum order = 10 (n = 15, 25, 35, 100). Totally, there are 20 candidate models.

Criterion	Model order: AR(1)–AR(6)	MA(1)–MA(10)	ARMA(1,1)	ARMA(1,2)& ARMA(2,1)	ARMA(2,2)
AICC	2, 1, 0, 0	7, 7, 0, 0	18, 7, 7, 0	31,26,24,19	42,59,69,81
WIC	7, 0, 0, 0	13, 6, 0, 0	10, 8, 1, 0	29,32,19,12	41,54,80,88
WIC_S	2, 0, 0, 0	10, 5, 2, 0	14,10, 3, 0	39,31,24,14	35,54,71,86
BIC	2, 1, 0, 0	13,13, 0, 0	16,10, 1, 0	46,28,20,12	23,48,81,88
SIC	3, 2, 0, 0	17, 9, 6, 0	12, 9, 6, 0	46,35,16,15	22,45,72,85
HQ	31,20,10, 6	18,15, 5, 2	7, 5, 3, 1	21,16,11, 8	23,44,71,83
ODQ	0, 0, 0, 0	10, 2, 0, 0	15,12, 3, 0	49,39,22,18	26,47,75,82

WIC, was very favorably comparable or nearly as high as that of the best criterion for the situation. In particular, again as regards to the relative frequency of correct model order selections (see again Fig. 1), for small sample sizes, the WIC was nearly as good as the AICC; for large sample sizes, the WIC asymptotically approached the behavior of the BIC quickly; and for intermediate sample sizes, the WIC performed much better than the worst of the two criteria, AICC or BIC, and nearly as well as the best of these two.

In regards to performance at specific sample sizes, the WIC outperformed other criteria and was comparable to AICC at the small sizes of $n = 15, 20, 25$. Moreover, it outperformed other criteria, and was comparable to BIC at the sample sizes of $n = 35, 50$. In addition, it was equal to BIC at the large sample size of $n = 100$. Finally, at the intermediate transitional zone between what may be considered small and large samples (approximately between $n = 20$ and 40 for the time series models considered here), the WIC dominates other criteria due to its stable performance (WIC was either the best or a very strong second , whereas other criteria varied more in ranking).

Fig. 2 demonstrates that the average value of WIC attains a global minimum at the correct value. Moreover, it implies that the maximum cut-off has virtually no effect on the model selected by WIC. Indeed, WIC preserves such desirable properties from AICC and BIC (Hurvich and Tsai, 1989).

The regression simulations were done as in Wei (1992) and Hurvich and Tsai (1989). Specifically, we generated 100 realizations of samples at three different sizes $n = 15, 20, 50$, from each of the following three models: the linear trend model M_1 and the random walk (with drift) model M_2, both of Wei (1992), and the third order regression model of Hurvich and Tsai (1989). The readers are referred to their respective

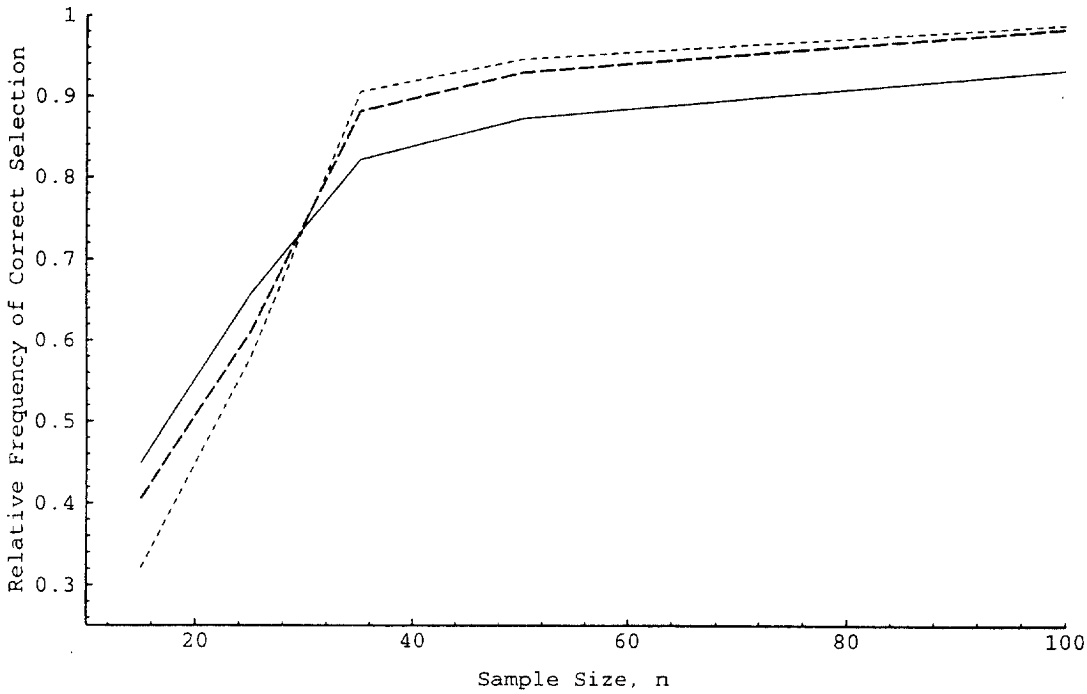

Fig. 1. Relative frequency of correct model order selections for AICC (solid line), WIC (long-dashed line), BIC (dashed line). The relative frequency of correct model order selections are taken from 1000 replication of samples of size n from the AR(3) model.

Table 6
Number of regression order selections out of 100 realizations under the second-order model M_1 (a liner trend model). Maximum order = 10 ($n = 15, 20, 50$)

	Model order			
Criterion	1	2	3–6	7–10
AICC	0, 0, 0	89, 93, 94	11, 7, 6	0, 0, 0
WIC	0, 0, 0	87, 92, 98	13, 8, 2	0, 0, 0
WIC_S	0, 0, 0	85, 86, 88	15, 14, 12	0, 0, 0
BIC	0, 0, 0	79, 90, 98	21, 10, 2	0, 0, 0
SIC	0, 0, 0	71, 79, 86	29, 21, 14	0, 0, 0
FIC	0, 0, 0	85, 89, 93	15, 11, 7	0, 0, 0
PLS	0, 0, 0	87, 91, 98	13, 9, 2	0, 0, 0

papers for detailed descriptions of the models. For each of the above three models, linear regression models ranging from 1 to 10 independent variables were used as candidates. The candidate family always includes the true model. In addition, the design matrices of the candidate models are sequentially nested. That is, columns $1, \ldots, m-1$ of the design matrix of the candidate model of order m are identical to the design matrix of the candidate model of order $m-1$. Tables 6–8 show the frequencies of model orders selected by the various criteria. Again, the WIC is the most stable, and therefore most practical criterion among those criteria considered. The WIC also showed characteristics of AICC and BIC at small and large samples, respectively.

It is interesting to note that for $n = 100$, the asymptotics became effective for most of the criteria, except when considering more complex models, such as the ARMA(2,2) or AR(7) models. We also saw that WIC outperformed the WIC_S criteria, although both displayed asymptotically similar behavior at $n = 100$. We

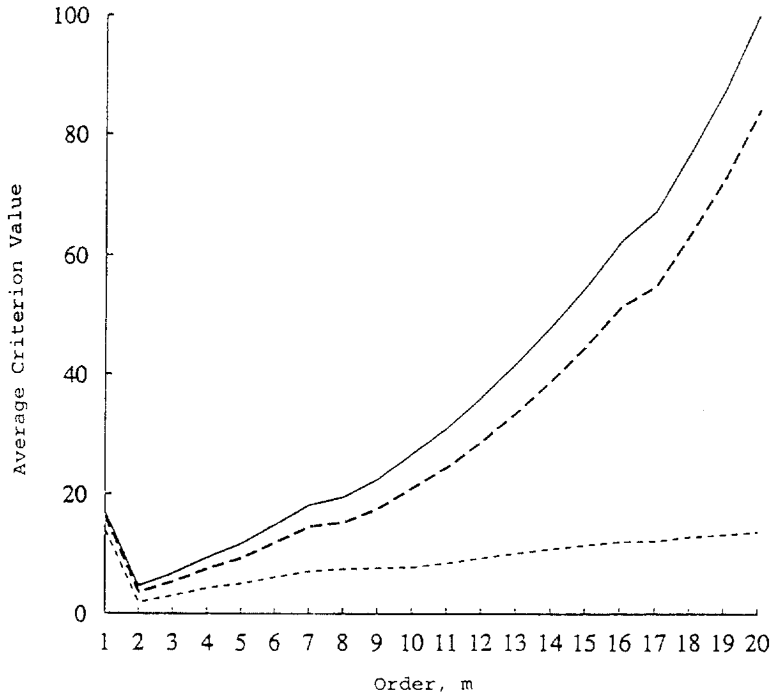

Fig. 2. Average criterion values for AICC (solid line), WIC (long-dashed line), BIC (dashed line). The average criterion values are taken from 100 replication of samples of size n=35 from the AR(2) model.

Table 7
Number of regression order selections out of 100 realizations under the second-order model M_2 (a random walk with drift). Maximum order = 10 (n = 15, 20, 50)

	Model order			
Criterion	1	2	3–6	7–10
AICC	0, 0, 0	88, 90, 92	10, 10, 8	2, 0, 0
WIC	0, 0, 0	87, 90, 97	11, 10, 3	2, 0, 0
WIC_S	0, 0, 0	85, 88, 90	12, 9, 10	3, 3, 0
BIC	0, 0, 0	78, 89, 97	20, 9, 3	2, 2, 0
SIC	0, 0, 0	75, 85, 87	19, 11, 13	6, 4, 0
FIC	0, 0, 0	79, 90, 92	20, 10, 8	1, 0, 0
PLS	0, 0, 0	80, 90, 91	19, 10, 9	1, 0, 0

mention that in both simulations of Wei (1992) and Zhang and Wang (1994), the form of the BIC taken was that of SIC. Finally, our simulation results confirmed the consistency of WIC. Indeed, as n increases, the model order selected by WIC converges to the true order, at the same fast speed as does BIC.

In conclusion, we recommend the use of WIC for intermediate sample sizes. In addition, at moderately large to large sample sizes, WIC performed very nearly as did BIC. At the practically small sample sizes, one may not lose much by using WIC instead of AICC. In the case where one is not certain of the relative sample size (more often than not this is the case since the terms "small", "moderate", and "large" are relative terms, and depend on the unknown complexity of the true model), the WIC may be a practical and safe alternative to any criterion.

Table 8
Number of regression order selections out of 100 realizations under a third-order model. Maximum order = 10 (n = 15, 20, 50)

Criterion	Model order 1–2	3	4–6	7–10
AICC	0, 0, 0	75, 88, 91	20, 11, 9	5, 1, 0
WIC	0, 0, 0	74, 88, 96	21, 12, 4	5, 0, 0
WIC_S	0, 0, 0	73, 85, 86	18, 14, 14	9, 1, 0
BIC	0, 0, 0	71, 89, 96	22, 11, 4	7, 0, 0
SIC	0, 0, 0	70, 83, 86	20, 15, 14	10, 2, 0
FIC	0, 0, 0	70, 88, 90	17, 10, 10	13, 2, 0
PLS	8, 12, 0	74, 80, 89	10, 8, 11	8, 0, 0

Acknowledgements

The first author thanks C.-L. Tsai for valuable discussions on time series model identification and order determination during his visit at Academia Sinica, Taiwan. The authors thank M. Kinter for his substantial computing assistance at the early stage of the development of the work. Furthermore, the authors thank the referees for valuable comments and suggestions which improved the presentation of the paper.

References

Akaike, H., 1974 A new look at the statistical model identification IEEE Trans. Automat. Control 19, 716–723.
Akaike, H., 1978 Time series analysis and control through parameteric models in: Findley, D.F. (Eds.), Applied Time Series Analysis. Academic Press, New York, pp. 1–24.
Brockwell, P.J., Davis, R.A., 1995 Time Series: Theory and Methods, 2nd Ed. Springer, New York.
Brockwell, P.J., Davis, R.A., Mandarino, J., 1995 PEST program ITSM: An Interactive Time Series Modeling Package for the PC. Springer, New York.
Hannan, E.J., 1980 The estimation of the order of an ARMA process Ann. Statist. 8, 1071–1081.
Hannan, E.J., Quinn, B.G., 1979 The determination of the order of autoregression, J. Roy. Statist. Soc. Ser. B 41, 190–195.
Hurvich, C.M., Tsai, C.-L., 1989 Regression and time series model selection in small samples Biometrika 76, 297–307.
Lütkepohl, H., 1985 Comparison of criteria for estimating the order of a vector autoregressive process J. Time Ser. Anal. 6, 35–52.
Mallows, C.L., 1973 Some comments on Cp Technometrics 15, 661–675.
Parzen, E., 1977 Multiple time series modeling: determining the order of approximating autoregressive schemes In: Krishnaiah, P. (Eds.), Multivariate Analysis IV. North-Holland, Amsterdam pp. 283–295.
Priestley, M.B., 1982 Spectral Analysis and Time Series, vol. 1. Wiley, New York.
Rissanen, J., 1986 Stochastic complexity and modeling Ann. Statist. 14, 1080–1100.
Schwarz, G., 1978 Estimating the dimension of a model Ann. Statist. 6, 461–464.
Shibata, R., 1980 Asymptotically efficient selection of the order of the model for estimating parameters of a linear process Ann. Statist. 8, 147–164.
Wei, C.Z., 1992 On predictive least squares principles Ann. Statist. 20, 1–42.
Zhang, H.-M., Wang, P., 1994 A new way to estimate orders in time series J. Time Ser. Anal. 15, 545–559.