The Study of the Probability of Overfitting and the Signal-to-Noise Ratio of the $\mathbf{KIC_U}$ Criterion

Rujirek Boosarawongse

Department of Applied Statistics, Faculty of Science,

King Mongkut Institute of Technology Ladkrabang

Chalongkrung Road, Ladkrabang District, Bangkok 10520, THAILAND

(rujirek@mozart.inet.co.th)

Abstract

This paper shows the derivation of the probability of overfitting and the signal-to-noise ratio of the $\mathbf{KIC_U}$ criterion. Comparing them with **AIC**, $\mathbf{AIC_C}$, $\mathbf{AIC_U}$, **SBC, KIC** and $\mathbf{KIC_C}$ on AR(1), AR(2), and AR(3) models were examined. The results show that, for small to medium sample sizes, the $\mathbf{KIC_U}$ criterion had the lowest probability of overfitting and the highest signal-to-noise ratio. However, the **SBC** criterion is the best for large sample.

1. Introduction

Model selection criteria are often compared using results from simulation studies. Often, a count of time that a selection criterion identifies the correct model is a useful measure of model selection performance. However, the more variety in models, the more unreliable can become [7]. They proposed the two theoretical properties to measure a model selection performance that is the probability of overfitting and the signal to noise ratio. The aim of this paper is to derive this two theoretical properties of the $\mathbf{KIC_U}$ criterion [3] in autoregressive model. Omitting the constant $n\ln 2\pi$, the $\mathbf{KIC_U}$ criterion for AR model at order p is given by

$$\mathbf{KIC_U}(p) = n\ln \hat{s}_p^2 + n\ln\left(\frac{n}{n-p}\right) + \frac{n[(n+p)(n-p)+(n-p-2)]}{(n-p-2)(n-p)} \tag{1}$$

where $\hat{s}_p^2$ is an unbiased estimator of error variance of the process with sample of size n.

2. Probability of Overfitting

When choosing the best model from the candidate models for each model selection criterion, the best model is assumed to have the lowest selection criterion value. In the case of overfitting, we assume that the true model order is k and a candidate model order is p where $p=k+L$ and $L>0$, L being the amount of overfitting. If the criterion value of a model of order $k+L$ is less than that of order k, $\mathbf{KIC_U}(k+L) < \mathbf{KIC_U}(k)$, the candidate model with an order $k+L$ is selected instead of the true model of order k. The model with order $k+L$ is said to be *overfitted*. Then the probability of overfitting by L for the $\mathbf{KIC_U}$ criterion is given by

$$P\{\mathbf{KIC_U}(k+L) < \mathbf{KIC_U}(k)\} \; . \tag{2}$$

The criterion that gives the lower probability of overfitting is the better one. Then we have

$$\begin{aligned} &P\{\mathbf{KIC_U}(k+L) < \mathbf{KIC_U}(k)\} \\ &= P\left\{\left[\ln \hat{s}_{k+L}^2 + \ln\left(\frac{n}{n-k-L}\right) + \frac{[(n+k+L)(n-k-L)+(n-k-L-2)]}{(n-k-L-2)(n-k-L)}\right] \right. \\ &\qquad \left. < \left[\ln \hat{s}_k^2 + \ln\left(\frac{n}{n-k}\right) + \frac{[(n+k)(n-k)+(n-k-2)]}{(n-k-2)(n-k)}\right]\right\} \end{aligned}$$

Since $SSE_k = (n-k)\hat{s}_k^2$ and therefore $\ln SSE_k = \ln(n-k) + \ln \hat{s}_k^2$, then

$$\begin{aligned} &P\{\mathbf{KIC_U}(k+L) < \mathbf{KIC_U}(k)\} \\ &= P\left\{\left[\ln SSE_{k+L} - \ln(n-k-L) - \ln(n-k-L) + \frac{[(n+k+L)(n-k-L)+(n-k-L-2)]}{(n-k-L-2)(n-k-L)}\right] \right. \\ &\qquad \left. < \left[\ln SSE_k - \ln(n-k) - \ln(n-k) + \frac{[(n+k)(n-k)+(n-k-2)]}{(n-k-2)(n-k)}\right]\right\} \end{aligned}$$

$$P\{\mathbf{KIC_U}(k+L) < \mathbf{KIC_U}(k)\}$$
$$= P\left\{ \ln\frac{SSE_k}{SSE_{k+L}} > \left[2\ln\frac{(n-k)}{(n-k-L)} + \frac{[(n+k+L)(n-k-L)+(n-k-L-2)]}{(n-k-L-2)(n-k-L)} - \frac{[(n+k)(n-k)+(n-k-2)]}{(n-k-2)(n-k)} \right] \right\}. \quad (3)$$

We know that

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

then, let

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

Substituting in (3), the following is obtained;

$$P\{\mathbf{KIC_U}(k+L) < \mathbf{KIC_U}(k)\}$$
$$= P\left\{ \ln\frac{SSE_k}{SSE_{k+L}} > 2\ln\frac{(n-k)}{(n-k-L)} + \mathbf{A} \right\}$$
$$= P\left\{ \frac{SSE_k}{SSE_{k+L}} > \left\{\frac{(n-k)}{(n-k-L)}\right\}^2 \exp(\mathbf{A}) \right\}$$
$$= P\left\{ \frac{SSE_k}{SSE_{k+L}} - 1 > \left\{\frac{(n-k)}{(n-k-L)}\right\}^2 \exp(\mathbf{A}) - 1 \right\}$$
$$= P\left\{ \frac{SSE_k - SSE_{k+L}}{SSE_{k+L}} > \left\{\frac{(n-k)}{(n-k-L)}\right\}^2 \exp(\mathbf{A}) - 1 \right\}.$$

From [7] , $SSE_k - SSE_{k+L} \sim \sigma_k^2\chi_L^2$, $SSE_{k+L} \sim \sigma_k^2\chi_{n-k-L}^2$ and $SSE_k - SSE_{k+L}$ are independent of SSE_{k+L}. Then

$$P\{\mathbf{KIC_U}(k+L) < \mathbf{KIC_U}(k)\}$$
$$= P\left\{ \frac{\chi_L^2}{\chi_{n-k-L}^2} > \left\{\frac{(n-k)}{(n-k-L)}\right\}^2 \exp(\mathbf{A}) - 1 \right\}$$
$$= P\left\{ F_{L,n-k-L} > \left(\frac{n-k-L}{L}\right) \times \left(\left\{\frac{(n-k)}{(n-k-L)}\right\}^2 \exp(\mathbf{A}) - 1 \right) \right\}. \quad (4)$$

From [7], for a fixed k and L, and where n approaches infinity, $\frac{\chi_{n-k-L}^2}{(n-k-L)} \to 1$, $F_{L,n-k-L} \to \frac{\chi_L^2}{L}$, and if $\lim_{n\to\infty} f_n \to f$ then $\lim_{n\to\infty} P\{F_{L,n-k-L} > f_n\} \to P\{\chi_L^2 > fL\}$, and $\exp(x) = 1 + x + \sum_{i=2}^{\infty}\frac{x^i}{i!}$. is replaced by $\exp(x) = 1 + x + o(x^2)$ for $0 \le x \le 1$. From (4), if

$$f_n = \left(\frac{n-k-L}{L}\right) \times \left(\left\{\frac{(n-k)}{(n-k-L)}\right\}^2 \exp(\mathbf{A}) - 1 \right)$$

then $\lim_{n\to\infty} f_n \to 4$ and

$$\lim_{n\to\infty} P\{F_{L,n-k-L} > f_n\} \to P\{\chi_L^2 > 4L\}.$$

Therefore, the asymptotic probability of overfitting of the $\mathbf{KIC_U}$ criterion by L is $P\{\chi_L^2 > 4L\}$. Notice that $P\{\chi_L^2 > 4L\}$ decreases as the amount of overfitting L increases.

3. Signal-to-Noise Ratio

The signal-to-noise ratio is a measurement which is basically a ratio of the expectation to the standard deviation of the difference in criterion values for two models. The ratio tends to assess whether the penalty term is sufficiently strong in relation to the goodness-of-fit term. From the true model order k and a candidate model order p where $p=k+L$ and $L>0$, the true model is considered better than a candidate model if the criterion value of a model of order k is less than that of order $k+L$, $\mathbf{KIC_U}(k) < \mathbf{KIC_U}(k+L)$. For the $\mathbf{KIC_U}$ criterion, McQuarrie and Tsai [7] defined the signal as $E[\mathbf{KIC_U}(k+L) - \mathbf{KIC_U}(k)]$, and the noise as the standard deviation of the difference, $sd[\mathbf{KIC_U}(k+L) - \mathbf{KIC_U}(k)]$. Then the signal-to-noise ratio that the true model is selected compared with a candidate model is defined as

$$\frac{E[\mathbf{KIC_U}(k+L) - \mathbf{KIC_U}(k)]}{sd[\mathbf{KIC_U}(k+L) - \mathbf{KIC_U}(k)]} \quad (5)$$

The criterion that gives a higher signal-to-noise ratio is the better one. Notice that when the amount of overfitting, L, increases, the signal-to-noise ratio will increase to indicate a higher overfit.

Since $E(\ln \hat{s}_p^2) \simeq \ln\sigma_k^2 - \frac{1}{n-p}$, the following is obtained;

$$E(\mathbf{KIC_U}(k+L))-E(\mathbf{KIC_U}(k))$$
$$\simeq\left\{\left[n\ln\sigma_k^2-\frac{n}{n-k-L}+n\ln\left(\frac{n}{n-k-L}\right)+\frac{n[(n+k+L)(n-k-L)+(n-k-L-2)]}{(n-k-L-2)(n-k-L)}\right]\right\}$$
$$-\left\{\left[n\ln\sigma_k^2-\frac{n}{n-k}+n\ln\left(\frac{n}{n-k}\right)+\frac{n[(n+k)(n-k)+(n-k-2)]}{(n-k-2)(n-k)}\right]\right\}$$

$$E(\mathbf{KIC_U}(k+L))-E(\mathbf{KIC_U}(k))$$
$$\simeq\frac{-nL}{(n-k-L)(n-k)}+n\ln\left(\frac{n-k}{n-k-L}\right)+\frac{nL\left[2(n-1)(n-k)(n-k-L)+(n-k-2)(n-k-L-2)\right]}{(n-k-L-2)(n-k-L)(n-k-2)(n-k)}.$$

To measure the noise;

$$sd\left[\mathbf{KIC_U}(k+L)-\mathbf{KIC_U}(k)\right]$$
$$=sd\left[n\ln\hat{s}_{k+L}^2-n\ln\hat{s}_k^2+\text{constant}\right]$$
$$=sd\left[n\ln\frac{SSE_{k+L}}{SSE_k}\right].$$

McQuarrie and Tsai [7] suggested an approximation of $Var\left(n\ln\frac{SSE_{k+L}}{SSE_k}\right)$ by $\frac{2n^2L}{(n-k-L)(n-k+2)}$. Then the noise becomes

$$\text{noise}\simeq\frac{n\sqrt{2L}}{\sqrt{(n-k-L)(n-k+2)}}. \tag{6}$$

Therefore, the signal-to-noise ratio of the $\mathbf{KIC_U}$ criterion for the AR(k) model is given by

$$\frac{E(\mathbf{KIC_U}(k+L))-E(\mathbf{KIC_U}(k))}{sd\left[\mathbf{KIC_U}(k+L)-\mathbf{KIC_U}(k)\right]}$$
$$\simeq\frac{\sqrt{(n-k-L)(n-k+2)}}{n\sqrt{2L}}\times\left[\frac{-nL}{(n-k-L)(n-k)}+n\ln\left(\frac{n-k}{n-k-L}\right)+\frac{nL\left[2(n-1)(n-k)(n-k-L)+(n-k-2)(n-k-L-2)\right]}{(n-k-L-2)(n-k-L)(n-k-2)(n-k)}\right]. \tag{7}$$

For a fixed k and L,

$$\lim_{n\to\infty}\frac{E\left[\mathbf{KIC_U}(k+L)-\mathbf{KIC_U}(k)\right]}{sd\left[\mathbf{KIC_U}(k+L)-\mathbf{KIC_U}(k)\right]}=\frac{3L}{\sqrt{2L}}$$

Therefore, the asymptotic signal-to-noise ratio of the $\mathbf{KIC_U}$ criterion for the AR model is $\frac{3L}{\sqrt{2L}}$. Notice that $\frac{3L}{\sqrt{2L}}$ increases as the amount of overfitting L increases.

4. Simulation Study

In this section, the probability of overfitting and the signal-to-noise ratio of the $\mathbf{KIC_U}$ criterion were tested by comparing them with a number of considered criteria using simulation. The considered criteria are **AIC** [1], $\mathbf{AIC_C}$ [5], $\mathbf{AIC_U}$ [6], **SBC** [8], **KIC** [4] and $\mathbf{KIC_C}$ [2] which their probability of overfitting and signal-to-noise ratio are given in McQuarrie and Tsai [7].

AR(1), AR(2) and AR(3) models with sample sizes 25, 40, 60 and 100, ranging from small to large, were used. All results are shown in tables 1-6.

The data shown in table 1 are the probability of overfitting for the AR(1) model by each criteria for different amounts of overfitting L and different sample sizes, e.g. for $\mathbf{KIC_U}$ where n = 25 and L = 1 the probability of overfitting was 0.0382. This means that this criterion would select the model whose order is higher by one order than true model, AR(1), with a probability of 0.0382. It can be seen that, where n = 25, $\mathbf{KIC_U}$ is the best among the selection criteria as it had the lowest probability of overfitting; 0.0382 and 0.0128 for L = 1 and 2 respectively. Where n = 40, $\mathbf{KIC_U}$ is the best with the lowest probability of overfitting; 0.0411, 0.0150 and 0.0053 for L = 1, 2 and 3 respectively. Where n = 60, $\mathbf{KIC_U}$ performs the best with the lowest probability of overfitting; 0.0427, 0.0161 and 0.0060 for L = 1, 2 and 3 respectively. Where n = 100, the best criterion is **SBC** with the lowest probability of overfitting; 0.0341, 0.0115 and 0.0040 for L = 1, 2 and 3 respectively.

The data shown in table 4 are the signal-to-noise ratios for the AR(1) model for each criteria using different amounts of overfitting L for different sample sizes, e.g. for $\mathbf{KIC_U}$, where n = 25 and L = 1, the signal-to-noise ratio was 2.5325. This means that this criterion will select the model whose order is higher by one order than the true model, AR(1), with a signal-to-noise ratio of 2.5325. It was found that, where n = 25, $\mathbf{KIC_U}$ has the highest signal-to-noise ratio; 2.5325 and 3.6494 for L = 1 and 2 respectively. Where n = 40, the highest signal-to-noise ratio was generated by $\mathbf{KIC_U}$; 2.3594, 3.3711 and 4.1729 for L = 1, 2 and 3 respectively. Where n = 60, $\mathbf{KIC_U}$ generated the highest signal-to-noise ratio; 2.2738, 3.2362 and 3.9894 for L = 1, 2 and 3 respectively. Where n = 100, the highest signal-to-noise ratio was generated by **SBC**; 2.5182, 3.5378 and 4.3041 for L = 1, 2 and 3 respectively.

Table 1 The probability of overfitting for AR(1) model

n	L	probability of overfitting						
		AIC	**AIC_C**	**AIC_U**	**SBC**	**KIC**	**KIC_C**	**KIC_U**
25	1	0.1796	0.1262	0.0887	0.0887	0.1003	0.0677	0.0382
	2	0.1720	0.0907	0.0589	0.0589	0.0714	0.0334	0.0128
	3	0.1608	0.0602	0.0403	0.0403	0.0519	0.0155	0.0041
	4	0.1524	0.0380	0.0291	0.0291	0.0396	0.0069	0.0013
40	1	0.1709	0.1383	0.0629	0.0629	0.0935	0.0738	0.0411
	2	0.1572	0.1077	0.0330	0.0330	0.0623	0.0396	0.0150
	3	0.1402	0.0790	0.0178	0.0178	0.0419	0.0205	0.0053
	4	0.1260	0.0565	0.0101	0.0101	0.0290	0.0104	0.0019
60	1	0.1662	0.1448	0.0476	0.0476	0.0900	0.0771	0.0427
	2	0.1496	0.1170	0.0205	0.0205	0.0578	0.0431	0.0161
	3	0.1300	0.0898	0.0090	0.0090	0.0372	0.0234	0.0060
	4	0.1133	0.0678	0.0042	0.0042	0.0245	0.0126	0.0022
100	1	0.1626	0.1499	0.0341	0.0341	0.0872	0.0796	0.0438
	2	0.1437	0.1244	0.0115	0.0115	0.0545	0.0458	0.0170
	3	0.1223	0.0985	0.0040	0.0040	0.0338	0.0257	0.0066
	4	0.1041	0.0771	0.0014	0.0014	0.0213	0.0144	0.0025

Table 3 The probability of overfitting for AR(3) model

n	L	probability of overfitting						
		AIC	**AIC_C**	**AIC_U**	**SBC**	**KIC**	**KIC_C**	**KIC_U**
25	1	0.2002	0.1075	0.1041	0.1041	0.1167	0.0581	0.0330
	2	0.2019	0.0695	0.0761	0.0761	0.0907	0.0256	0.0099
	3	0.1972	0.0412	0.0570	0.0570	0.0716	0.0105	0.0028
	4	0.1945	0.0229	0.0449	0.0449	0.0589	0.0041	0.0007
40	1	0.1827	0.1271	0.0704	0.0704	0.1027	0.0682	0.0381
	2	0.1738	0.0941	0.0396	0.0396	0.0724	0.0346	0.0131
	3	0.1595	0.0656	0.0229	0.0229	0.0513	0.0169	0.0044
	4	0.1472	0.0446	0.0138	0.0138	0.0374	0.0081	0.0015
60	1	0.1738	0.1375	0.0516	0.0516	0.0957	0.0734	0.0407
	2	0.1599	0.1079	0.0234	0.0234	0.0639	0.0397	0.0149
	3	0.1417	0.0805	0.0109	0.0109	0.0426	0.0209	0.0054
	4	0.1258	0.0591	0.0053	0.0053	0.0290	0.0109	0.0019
100	1	0.1670	0.1456	0.0360	0.0360	0.0905	0.0775	0.0427
	2	0.1496	0.1190	0.0126	0.0126	0.0578	0.0438	0.0163
	3	0.1288	0.0928	0.0045	0.0045	0.0367	0.0242	0.0062
	4	0.1109	0.0716	0.0017	0.0017	0.0236	0.0133	0.0023

Table 2 The probability of overfitting for AR(2) model

n	L	probability of overfitting						
		AIC	**AIC_C**	**AIC_U**	**SBC**	**KIC**	**KIC_C**	**KIC_U**
25	1	0.1896	0.1169	0.0961	0.0961	0.1081	0.0630	0.0356
	2	0.1864	0.0800	0.0669	0.0669	0.0805	0.0295	0.0113
	3	0.1781	0.0503	0.0479	0.0479	0.0610	0.0129	0.0034
	4	0.1723	0.0299	0.0362	0.0362	0.0484	0.0054	0.0010
40	1	0.1767	0.1328	0.0665	0.0665	0.0980	0.0710	0.0396
	2	0.1653	0.1009	0.0362	0.0362	0.0672	0.0371	0.0140
	3	0.1496	0.0722	0.0202	0.0202	0.0463	0.0187	0.0048
	4	0.1362	0.0504	0.0118	0.0118	0.0330	0.0092	0.0017
60	1	0.1699	0.1412	0.0496	0.0496	0.0928	0.0753	0.0417
	2	0.1546	0.1125	0.0219	0.0219	0.0608	0.0414	0.0155
	3	0.1357	0.0851	0.0099	0.0099	0.0398	0.0221	0.0057
	4	0.1194	0.0634	0.0047	0.0047	0.0266	0.0117	0.0021
100	1	0.1648	0.1478	0.0350	0.0350	0.0889	0.0785	0.0433
	2	0.1466	0.1217	0.0120	0.0120	0.0561	0.0448	0.0166
	3	0.1255	0.0956	0.0042	0.0042	0.0352	0.0249	0.0064
	4	0.1074	0.0743	0.0015	0.0015	0.0224	0.0139	0.0024

Table 4 Signal to noise ratio for AR(1) model

n	L	signal to noise ratio						
		AIC	**AIC_C**	**AIC_U**	**SBC**	**KIC**	**KIC_C**	**KIC_U**
25	1	0.6161	1.0293	1.7652	1.4591	1.3077	1.7616	2.5325
	2	0.8275	1.5233	2.5638	1.9936	1.7842	2.5067	3.6496
	3	0.9589	1.9557	3.2295	2.3541	2.1036	3.0909	4.5601
	4	1.0428	2.3719	3.8418	2.6151	2.3328	3.5957	5.3789
40	1	0.6517	0.8906	1.6156	1.8301	1.3495	1.6143	2.3594
	2	0.8952	1.2938	2.3190	2.5397	1.8690	2.2880	3.3711
	3	1.0637	1.6285	2.8839	3.0504	2.2401	2.8087	4.1729
	4	1.1901	1.9338	3.3831	3.4520	2.5294	3.2511	4.8716
60	1	0.6707	0.8236	1.5425	2.1388	1.3717	1.5420	2.2738
	2	0.9312	1.1852	2.2019	2.9895	1.9140	2.1828	3.2362
	3	1.1192	1.4774	2.7225	3.6178	2.3123	2.6759	3.9894
	4	1.2676	1.7366	3.1743	4.1268	2.6328	3.0929	4.6373
100	1	0.6855	0.7744	1.4886	2.5182	1.3890	1.4884	2.2100
	2	0.9593	1.1064	2.1164	3.5378	1.9491	2.1056	3.1368
	3	1.1623	1.3691	2.6061	4.3041	2.3683	2.5796	3.8558
	4	1.3275	1.5975	3.0258	4.9364	2.7128	2.9797	4.4688

Table 5 Signal to noise ratio for AR(2) model

n	L	signal to noise ratio						
		AIC	AIC$_C$	AIC$_U$	SBC	KIC	KIC$_C$	KIC$_U$
25	1	0.5567	1.1253	1.8624	1.3652	1.2201	1.8569	2.5792
	2	0.7434	1.6668	2.7090	1.8605	1.6599	2.6439	3.6435
	3	0.8555	2.1422	3.4181	2.1907	1.9510	3.2623	4.4591
	4	0.9231	2.6014	4.0736	2.4258	2.1559	3.7982	5.1474
40	1	0.6153	0.9392	1.6646	1.7639	1.2954	1.6628	2.3791
	2	0.8438	1.3646	2.3905	2.4460	1.7925	2.3570	3.3569
	3	1.0006	1.7181	2.9743	2.9355	2.1463	2.8937	4.1019
	4	1.1171	2.0407	3.4909	3.3191	2.4210	3.3500	4.7258
60	1	0.6467	0.8526	1.5718	2.0901	1.3359	1.5710	2.2842
	2	0.8973	1.2270	2.2440	2.9206	1.8634	2.2239	3.2238
	3	1.0776	1.5296	2.7751	3.5335	2.2502	2.7264	3.9404
	4	1.2195	1.7982	3.2363	4.0294	2.5611	3.1513	4.5407
100	1	0.6712	0.7904	1.5046	2.4855	1.3676	1.5044	2.2151
	2	0.9390	1.1293	2.1394	3.4916	1.9188	2.1282	3.1282
	3	1.1375	1.3974	2.6346	4.2474	2.3312	2.6074	3.8258
	4	1.2989	1.6305	3.0590	4.8709	2.6700	3.0118	4.4114

Table 6 Signal to noise ratio for AR(3) model

n	L	signal to noise ratio						
		AIC	AIC$_C$	AIC$_U$	SBC	KIC	KIC$_C$	KIC$_U$
25	1	0.4971	1.2324	1.9708	1.2711	1.1321	1.9630	2.6859
	2	0.6588	1.8273	2.8714	1.7270	1.5352	2.7968	3.7964
	3	0.7516	2.3516	3.6297	2.0268	1.7978	3.4538	4.6493
	4	0.8027	2.8602	4.3349	2.2358	1.9784	4.0249	5.3711
40	1	0.5789	0.9908	1.7168	1.6976	1.2413	1.7143	2.4308
	2	0.7922	1.4400	2.4665	2.3522	1.7159	2.4302	3.4301
	3	0.9374	1.8135	3.0705	2.8204	2.0524	2.9841	4.1918
	4	1.0440	2.1546	3.6057	3.1862	2.3124	3.4552	4.8299
60	1	0.6227	0.8827	1.6021	2.0414	1.3001	1.6011	2.3144
	2	0.8633	1.2704	2.2878	2.8517	1.8127	2.2665	3.2665
	3	1.0360	1.5839	2.8298	3.4490	2.1882	2.7788	3.9926
	4	1.1714	1.8621	3.3006	3.9318	2.4894	3.2121	4.6009
100	1	0.6569	0.8067	1.5210	2.4528	1.3463	1.5207	2.2315
	2	0.9188	1.1526	2.1629	3.4453	1.8886	2.1513	3.1513
	3	1.1128	1.4264	2.6636	4.1907	2.2942	2.6357	3.8541
	4	1.2703	1.6643	3.0929	4.8055	2.6273	3.0445	4.4440

5. Conclusions

For small to medium sample sizes, the **KIC$_U$** criterion had the lowest probability of overfitting and the highest signal-to-noise-ratio. However, as the sample size increases, **SBC** had the lowest probability of overfitting and the highest signal-to-noise-ratio.

Therefore, for small to medium sample sizes, the **KIC$_U$** performed the best. However, the **SBC** criterion is the best for large sample.

References

[1] Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F. Csaki, eds. *2nd International Symposium on Information Theory* (Akademia Kiado, Budapest).

[2] Boosarawongse R. and Chongcharoen S. 2003. Model Selection Criteria for Autoregressive Models. *The Proceeding of the 8th APDSI*, 4th– 8th July, China.

[3] Boosarawongse R., 2004. The **KIC$_U$** Model Selection Criteria for Autoregressive Models. *The Proceeding of the 9th APDS***I**, 1th– 4th July, Korea.

[4] Cavanaugh, J.E. 1999. A large-sample model selection criterion based on Kullback's symmetric divergence. *Statistics & Probability letter.* 42: 333-343.

[5] Hurvich, C.M. and Tsai, C.L. 1989. Regression and time series model selection in small samples. *Biometrika.* 76: 297-307.

[6] Hurvich, C.M., Shumway, R.H. and Tsai, C.L. 1990. Improved estimators of Kullback-Leibler Information for autoregressive model selection in small samples. *Biometrika.* 77: 709-719.

[7] McQuarrie, A.D.R., and Tsai, C.L. 1998. *Regression and Time Series Model Selection.* Singapore: World Scientific.

[8] Schwarz,G. 1978. Estimating the dimension of a model. *Annals of Statistics*. 6: 461-464.