



**HAL**  
open science

# An Improved Divergence Information Criterion for the Determination of the Order of an AR Process

Panagiotis Mantalos, Kyriacos Mattheou, Alex Karagrignoriou

► **To cite this version:**

Panagiotis Mantalos, Kyriacos Mattheou, Alex Karagrignoriou. An Improved Divergence Information Criterion for the Determination of the Order of an AR Process. *Communications in Statistics - Simulation and Computation*, 2010, 39 (05), pp.865-879. 10.1080/03610911003650391 . hal-00591710

**HAL Id: hal-00591710**

**<https://hal.science/hal-00591710>**

Submitted on 10 May 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**An Improved Divergence Information Criterion for the Determination of the Order of an AR Process**

Journal:	<i>Communications in Statistics - Simulation and Computation</i>
Manuscript ID:	LSSP-2009-0219.R2
Manuscript Type:	Original Paper
Date Submitted by the Author:	13-Jan-2010
Complete List of Authors:	MANTALOS, PANAGIOTIS; UNIV OF LUND MATTHEOU, KYRIACOS; UNIV OF CYPRUS KARAGRIGORIOU, ALEX; UNIV OFR CYPRUS
Keywords:	model selection, measure of divergence, AR process, information criterion
Abstract:	In this paper we propose a modification of the Divergence Information Criterion (DIC) for the determination of the order of an autoregressive process and show that it is an asymptotically unbiased estimator of the expected overall discrepancy. Further, we use Monte Carlo methods and various Data Generating Processes for small, medium and large sample sizes in order to explore the capabilities of the new criterion in selecting the optimal order in autoregressive processes and in general in a time series context. The new criterion shows remarkably good results by choosing the correct model more frequently than traditional Information Criteria.



**RESPONSE TO THE REVIEWER'S COMMENTS**

We would like to express our appreciation to the reviewer because his/hers careful reading resulted in identifying a typo in our GAUSS program that was oversighted. When in the original report, the reviewer raised concerns about the simulations, we looked into the programming of HQ, assuming that the error was there. The reviewer's persistence made us look more closely and we have now identified a typo in the programming not of HQ but of SIC. The error, which has partially affected the SIC powers, has been corrected and the correct rates of success for SIC are reported in the present revision (Tables 1-6). Note that all other values stay intact. Due to the new values for SIC, some minor changes were made into the text on pages 11-13 (all the changes are highlighted in yellow) which though do not effect in any way our original conclusions. On the contrary, a new special feature of MDIC has been revealed and decided to be included in the present revision:

First note that we included in this revision a brief discussion about MDIC with  $a$  equal to that value in  $[0, 0.50]$  which maximizes the power (see p. 11, lines 6-9). However, due to insignificant differences between MDIC and MDIC with max power, the powers of the latter for Models 3-6 are suppressed from the Tables (p. 11, lines -5 to -4). Our simulations revealed that especially for small and medium sample sizes, MDIC with  $a = 0.25$  provides an ideal model identification technique for both small and large lag models. As a result, we have chosen to include in the discussion about the choice of  $a$  (p. 13, middle paragraph), a statement about the small and medium sample performance of MDIC (p.13, lines -8 to -6). A similar statement appears also on p. 14, lines 9-13. In these statements, we explicitly state that for small and medium sample sizes, the choice  $a = 0.25$  is *highly recommended* since it clearly serves as a fair balance between small and large lag models.

Finally note that throughout the manuscript, the index  $a$  has been properly referred to.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# An Improved **Divergence Information Criterion** for the Determination of the Order of an AR Process

P. Mantalos<sup>a</sup>,

K. Mattheou<sup>b</sup>,

A. Karagrigoriou<sup>b,1</sup>,

<sup>a</sup>Department of Statistics

<sup>b</sup>Department of Mathematics and Statistics

University of Lund

University of Cyprus

Sweden

Cyprus

## Abstract

In this paper we propose a modification of the recently introduced Divergence Information Criterion (DIC, Mattheou et. al, 2009) for the determination of the order of an autoregressive process and show that it is an asymptotically unbiased estimator of the expected overall discrepancy, a nonnegative quantity which measures the distance between the true unknown model and a fitted approximating model. Further, we use Monte Carlo methods and various Data Generating Processes for small, medium and large sample sizes in order to explore the capabilities of the new criterion in selecting the optimal order in autoregressive processes and in general in a time series context. The new criterion shows remarkable good results by choosing the correct model more frequently than traditional Information Criteria.

*Key words:* AR process, Information Criterion, Model selection, Measure of divergence

---

<sup>1</sup> Corresponding author: Dept. of Math. and Stat., University of Cyprus, Cyprus, e-mail: alex@ucy.ac.cy

## 1. Introduction

Consider a discrete-time stationary AR(p) process of the form

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \cdots + a_p X_{t-p} + e_t \quad (1)$$

where  $\{e_t\}$  is a sequence of uncorrelated  $N(0, \sigma^2)$  random variables and all the roots of the

autoregressive polynomial are outside the unit circle, that is,  $A(z) = 1 - a_1 z - \cdots - a_p z^p \neq 0$

for all  $z \in \mathbb{C}$  such that  $|z| < 1$ .

One of the crucial problems in time series modeling is selecting the optimal model order and specifying a parsimonious model for the data generating process (DGP). Many techniques have been developed for the order selection of linear models, namely a) graphical methods with representative the Box and Jenkins methods (1970) and b) hypothesis test procedures with the pioneering work of Whittle (1952) as representative.

Akaike (1974) viewed the problem of model fitting in the context of time series analysis as a multiple decision procedure rather than a hypothesis testing problem. However the same author (Akaike, 1973) initiated the research on automatic model selection techniques by developing the popular Akaike Information Criterion (AIC) which is considered as a significant contribution to statistical modelling where the focus is on the selection of a satisfactory model for a given set of observations. It might appear at first sight that the higher the order of the selected process (model) chosen, the better the fitted model will be. Such a thought may be true for fitting purposes but not for predicting ones. Indeed, the fit may be perfect for the given data but the use of the selected model for predictive purposes may result in gross errors. Numerous selection criteria have been developed over the last 30 years which attempt to prevent overfitting by assigning a penalty to the introduction of an unnecessary

1  
2  
3 large order and consequently to an unnecessary number of parameters. Schwarz (1978),  
4  
5 Hannan and Quinn (1979), Seghouane et. al (2005), Bengtsson and Cavanaugh (2006) and  
6  
7 others followed the pioneering work of Akaike. Most of the criteria proposed in the literature  
8  
9 consist in minimizing, with respect to the model order, a function of the given observations  
10  
11 plus a penalty term for the introduction of additional parameters, which generally depends on  
12  
13 the model order and sample size. Thus, according to automatic criteria, the parameters of a  
14  
15 variety of competing models are estimated and the model chosen is the one with the smallest  
16  
17 criterion value.  
18  
19  
20  
21

22  
23 In this work we propose a modification of the recently introduced Divergence Information  
24  
25 Criterion (DIC) (Mattheou (2007); Mattheou, et. al (2009)) and show that it is an  
26  
27 asymptotically unbiased estimator of the expected overall discrepancy, a nonnegative quantity  
28  
29 which measures the distance between the true unknown model and a fitted approximating  
30  
31 model. Furthermore, we explore the capabilities of the new criterion in selecting the optimal  
32  
33 order in autoregressive processes and in general in a time series context. The paper is  
34  
35 organised as follows. In Section 2, we first briefly discuss various Information Criteria and  
36  
37 their association with measures of divergence and then introduce the Modified Divergence  
38  
39 Information Criterion (MDIC) and provide for practical purposes, an optimum choice for the  
40  
41 index  $\alpha$  which the new criterion depends on. In Section 3 in order to explore the capabilities  
42  
43 of the new MDIC criterion, we perform a comparative study of various Information Criteria  
44  
45 for different time series examples by using a Monte Carlo method. Finally in Section 4 we  
46  
47 summarize our results. For the simulations we make use of the Gauss 8 program.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 2. The New Modified Divergence Information Criterion

### 2.1. Measures of Divergence and Selection Criteria

The measures of divergence are used as indices of similarity or dissimilarity between populations and for measuring mutual information concerning two variables and as such they can be used for the construction of model selection criteria. The distance between a candidate model and the true but unknown model could be measured by a measure of divergence. The candidate model for which the measure of divergence is minimized will be selected. The well known Kullback-Leibler measure of divergence (Kullback and Leibler, 1951) was the one used by Akaike (1973) to develop the Akaike Information Criterion (AIC). Akaike proposed the evaluation of the fit of the candidate model using minus twice the mean expected loglikelihood (also known as expected overall discrepancy). Furthermore, he provided an unbiased estimator of the expected loglikelihood so that the resulting criterion is given by

$$AIC(p) = n \log(\hat{\sigma}_p^2) + 2p,$$

where  $n$  the sample size,  $p$  the order of the candidate model and  $\hat{\sigma}_p^2$  the estimator of the variance of the  $p^{\text{th}}$ -order candidate model.

For fitting autoregressive processes, Jones (1975) suggested that AIC has a tendency to overestimate the order of the process and Shibata (1976) showed that the probability of overestimation for a large sample size is nonzero. To correct this tendency, the SIC criterion was proposed independently by Akaike (1978) and Schwarz (1978):

$$SIC(p) = n \log(\hat{\sigma}_p^2) + p \log n.$$

The above two criteria can be considered as members of the General Information Criterion given by  $GIC(p) = n \log(\hat{\sigma}_p^2) + c_n p$  where  $c_n$  a quantity that may depend on  $n$ . For  $c_n = 2$  the

1  
2  
3 criterion reduces to AIC, while for  $c_n = \log n$  to SIC. The case  $c_n = c \log \log n$  with  $c > 2$   
4  
5  
6 corresponds to the Hannan and Quinn criterion (HQ, Hannan and Quinn, 1979) which has  
7  
8 been found to be equivalent to SIC in the sense that they are both consistent (Hannan, 1980).  
9

10  
11 The problem of avoiding overfitting especially for purely autoregressive processes can also be  
12  
13 dealt with the minimization of the final prediction error criterion (FPE) of Akaike (1969). The  
14  
15 FPE is an estimate of the one-step ahead prediction mean squared error for a realization of the  
16  
17 process independent of the one observed. If we fit AR processes of steadily increasing order  $p$   
18  
19 the maximum likelihood estimator (MLE) of the white noise variance will usually decrease  
20  
21 with  $p$ . However, FPE will decrease to a minimum value and then increase as  $p$  will increase.  
22  
23 According to FPE, we choose the order of the candidate process to be the value of  $p$  for which  
24  
25 FPE is minimized. The FPE is given by  
26  
27  
28  
29

$$30 \quad FPE(p) = \hat{\sigma}_p^2 \frac{n+p}{n-p}.$$

31  
32  
33 In all the above cases, any parameter estimation is handled through the maximum likelihood  
34  
35 method. A general class of criteria has been introduced by Konishi and Kitagawa (1996)  
36  
37 which also estimates the Kullback-Leibler measure where the estimation is not necessarily  
38  
39 based on maximum likelihood.  
40  
41  
42  
43  
44

## 45 2.2. The Development of the MDIC Criterion

46  
47  
48 One of the most recently proposed measures of divergence is the Basu-Harris-Hjort-Jones  
49  
50 power divergence between the candidate model  $f_\theta(\cdot)$  and the true model  $g$  (Basu et. al, 1998)

51  
52 which is denoted by BHHJ, indexed by a positive parameter  $a$ , and defined as:

$$53 \quad I^a(g, f_\theta) = \int \left\{ f_\theta^{1+a}(z) - \left(1 + \frac{1}{a}\right) g(z) f_\theta^a(z) + \left(\frac{1}{a}\right) g^{1+a}(z) \right\} dz.$$



This family of measures was proposed by Basu et al. (1998) for the development of a minimum divergence estimating method for robust parameter estimation. The index  $\alpha$  controls the trade-off between robustness and asymptotic efficiency of the parameter estimators which are the values of  $\theta$  that minimize the measure over a parametric space  $\Theta$ . It should be also noted that the BHHJ family reduces to the Kullback-Leibler measure of divergence for  $\alpha \downarrow 0$  (see Mattheou, 2007) and as it can be easily seen, to the square of the standard  $L_2$  distance between the candidate and the true model for  $\alpha=1$ . Mattheou et al. (2008) applying the same methodology used for AIC to the BHHJ divergence developed a new criterion, the Divergence Information Criterion (DIC) which for a set of observations  $x_1, \dots, x_n$  is given by

$$DIC(p) = nQ_{\hat{\theta}} + (2\pi)^{-a/2} (1+a)^{2+p/2} p,$$

where  $Q_{\hat{\theta}} = \int f_{\hat{\theta}}^{1+a}(z) dz - \left(1 + \frac{1}{a}\right) \frac{1}{n} \sum_{i=1}^n f_{\hat{\theta}}^a(x_i)$  and  $\hat{\theta}$  a consistent and asymptotically normal estimator of  $\theta$ . Preliminary simulation studies for regression models (Mattheou, 2007) showed a very good medium sample size performance for DIC for values of  $\alpha$  close to zero.

Although the DIC criterion was constructed so that it will be an asymptotically unbiased estimator of the BHHJ divergence measure between the candidate and the true model, the calculation of the first part of  $Q_{\hat{\theta}}$ , namely the integral  $\int f_{\hat{\theta}}^{1+a}(z) dz$  is not computationally attractive for practitioners. Furthermore, the simulation study shows that the difference in the calculations of the above integrals, for the different candidate models is negligible compared with the difference in the calculation for the entire quantity  $Q_{\hat{\theta}}$ . In other words the integral term does not affect the selection of the correct model and therefore the criterion can be properly revised. In fact the criterion can be modified without affecting neither its small

sample nor its asymptotic performance. As a result, we propose now a modified new criterion, the Modified Divergence Information Criterion (MDIC) which is given by

$$MDIC(p) = n * MQ_{\hat{\theta}} + (2\pi)^{-a/2} (1+a)^{2+p/2} p$$

$$\text{where } MQ_{\hat{\theta}} = - \left[ (1 + \alpha^{-1}) \frac{1}{n} \sum_{i=1}^n f_{\hat{\theta}}^{\alpha}(x_i) \right].$$

Note that a model selection criterion can be considered as an approximately unbiased estimator of the expected overall discrepancy, a nonnegative quantity which measures the distance between the true unknown model and a fitted approximating model. Observe also that, as it was mentioned in the introduction, a criterion consists of two terms, the first of which is a biased estimator of the expected overall discrepancy. As a result, if we choose the model with the smallest estimator of the expected overall discrepancy we may end up with a selection with an unnecessarily large order. The estimator becomes asymptotically unbiased by introducing the appropriate correction term. **The expected overall discrepancy in the present setting is defined by**

$$E(W_{\hat{\theta}}) = E(W_{\theta} / \theta = \hat{\theta}) \quad \text{with } W_{\theta} = - \int \left\{ \left( 1 + \frac{1}{a} \right) g(z) f_{\theta}^{\alpha}(z) \right\} dz$$

**and  $\hat{\theta}$  a consistent and asymptotically normal estimator of  $\theta$ . Observe that  $W_{\theta}$  is the targeted divergence  $I^{\alpha}(g, f_{\theta})$  without the first and the last terms. Notice that the last term remains constant irrespectively of the candidate model  $f_{\theta}$  while the first term, as stated earlier, is approximately constant for the various candidate models. In that sense, the expected overall discrepancy  $E(W_{\hat{\theta}})$  represents the average distance between the true model  $g$  and the candidate model  $f_{\theta}$  up to a constant.** The relevant theorem for MDIC is stated below. The proof is omitted since it follows the lines of the proof of Theorem 2.4 (Mattheou et.al, 2009) where the unbiasedness of the original DIC criterion was established.

**Theorem 1.** *An asymptotically unbiased estimator of  $n$ -times the expected overall discrepancy evaluated at  $\hat{\theta}$  is given by*

$$MDIC(p) = n * MQ_{\hat{\theta}} + (2\pi)^{-a/2} (1+a)^{2+p/2} p$$

where  $\hat{\theta}$  is a consistent and asymptotically normal estimator of  $\theta$ .

Note that although the discussion in Mattheou et. al (2009) is limited to iid random variables, by following the approach by Akaike (1973 and 1974) and Billingsley (1961), the same line of discussion can be extended to cover autoregressive models. Recall that in time series models, even under the Gaussian assumption, the evaluation of the log-likelihood cannot be expressed as the sum of the logarithms of the density function of each observation. However, the log-likelihood can generally be expressed by *conditional* distributions, so that in relatively simple models, like the autoregressive models, we use as an approximation the mean *conditional* log-likelihood which multiplied by the number of observations is used in place of the log-likelihood in model selection.

In addition to the above, numerous other techniques have been introduced over the years some of which are associated with the order determination and the forecasting performance in time series. Such techniques include approaches based on bootstrapping (Shang and Cavanaugh, 2008), approaches for high dimensional medical data (Koukouvinos et al., 2008), supersaturated designs (Koukouvinos and Mylona, 2008), variations of AIC (Cavanaugh and Johnson, 1999; Bengtsson and Cavanaugh, 2006; Seghouane, 2006; Seghouane et. al, 2005) and time series models (Lee, 1998).

### 2.3 Optimal Choice for the Index $a$

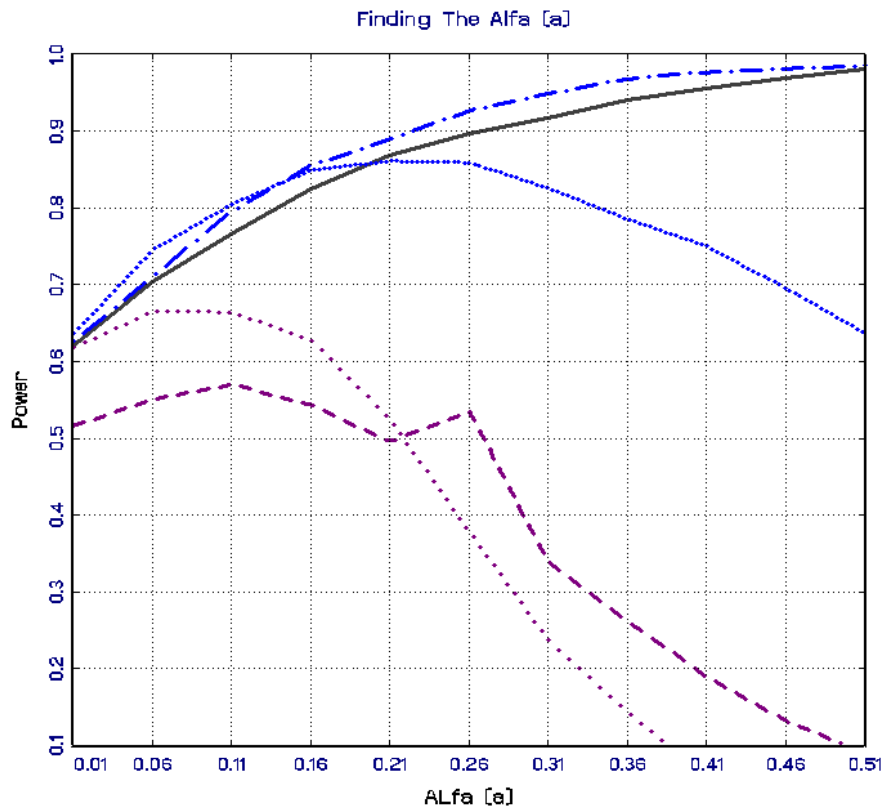
For practical purposes one has to decide the optimal choice of the positive index  $a$ . For this purpose, we simulate a 100 observation series for 5 different models with  $a$  between 0.01 and

1  
2  
3 0.5 and we provide in Figure 1, the power of the selection, namely the proportion of times the  
4 correct model is selected as a function of  $a$  .  
5  
6  
7

8  
9 In Figure 1, [Dot dash] line is the AR(1) model defined as Model 1 in section 3.2; [Whole]  
10 line shows the AR(2) appearing as Model 5 in section 3.2; [Dot close] line refers to the AR(3)  
11 process:  $x_t = 1 + 0.2x_{t-1} + 0.5x_{t-2} - 0.35x_{t-3} + e_t$ ; [Dash] line is the AR(4) process:  
12  $x_t = 1 + 0.2x_{t-1} + 0.5x_{t-2} - 0.35x_{t-3} - 0.22x_{t-4} + e_t$ ; finally [Dot] is the AR(5) process defined as  
13 Model 6 in section 3.2.  
14  
15  
16  
17

18  
19 Figure 1 shows that for small lags (models AR(1) and AR(2)) the power increases as the value  
20 of  $\alpha$  increases. In all the other cases the power increases up to a value of  $\alpha$  and then  
21 decreases. More specifically, for the complete AR(3) and AR(4) processes the power stays  
22 high approximately up to the value  $a = 0.25$  . For the AR(5) process the power decreases  
23 after the value  $a = 0.10$  .  
24  
25  
26  
27  
28

29  
30 Although an optimum value of  $a$  for all types of autoregressive models may be considered to  
31 be the value of 0.10, in the Monte Carlo study of section 3.2 we choose the value of 0.25 since  
32 it appears to serve as a fair balance between small and large lag models.  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Figure 1: Optimal choice of the index  $a$  (AR(1) - - - AR(2) — AR(3) ..... AR(4)---- AR(5) ····)

### 3. Simulations

#### 3.1 The Monte Carlo Experiment

In this section we provide the characteristics of the Monte Carlo experiment undertaken. We calculate the *estimated "percent"* by simply observing how many times the correct AR(p) model is selected in repeated samples. By varying factors such as the number of observations 50 (small sample) 75, 100 (medium sample) and 200, 500 (large sample); and the order of AR(p) model we obtain a succession of estimated percent of the correct selection model under different conditions.

The Monte Carlo experiment has been performed by generating data according to the following Data generating processes:

$$\text{Model 1: } x_t = 1 + 0.65x_{t-1} + e_t$$

$$\text{Model 2: } x_t = 1 + 0.65x_{t-2} + e_t$$

$$\text{Model 3: } x_t = 1 + 0.5x_{t-1} - 0.85x_{t-2} + e_t$$

$$\text{Model 4: } x_t = 1 + 0.5x_{t-1} - 0.35x_{t-2} + e_t$$

$$\text{Model 5: } x_t = 1 + 1.5x_{t-1} - 0.5x_{t-2} + e_t$$

$$\text{Model 6: } x_t = 1 + 0.23x_{t-1} - 0.22x_{t-3} - 0.45x_{t-5} + e_t$$

where  $\{e_t\}$  is a sequence of uncorrelated  $N(0, \sigma^2)$  random variables.

The criteria used in this experiment are the ones defined in Section 2, namely AIC, SIC, HQ with  $c = 2$ , FPE, and MDIC with  $a = 0.25$ . We also obtain the results of MDIC with the index  $a$  taken to be equal to that value in  $[0, 0.50]$  which maximizes the power. According to the analysis in Section 2.3, the maximum powers are obtained for  $a = 0.50$  (Models 1-5) and  $a = 0.10$  (Model 6). For the implementation of the criteria we use

$$f_{\hat{\theta}}(x_i) = \frac{1}{\sqrt{(2\pi\hat{\sigma}^2)}} \exp\left\{-\frac{(x_i - \hat{x}_i)^2}{2\hat{\sigma}^2}\right\},$$

where  $\hat{x}_i$  the predicted value of  $x_i$  and  $\hat{\sigma}^2$  the estimated variance.

For each time series 20 presample values are generated with zero initial conditions, taking net sample sizes of  $n = 50, 75, 100, 200, 500$  in order to cover small, medium and large samples.

The number of replications per model used is 10000. The calculations were performed using GAUSS 8.

### 3.2 Simulation Results

In this section we present the results of the Monte Carlo experiment concerning the percent of correct selected model. Due to insignificant differences between MDIC and MDIC with max power, the powers of the latter for Models 3-6 are suppressed from the Tables.

Table 1 shows the results for Model 1. It is not difficult to see the good performance of the MDIC criterion for all sample sizes. The rate of success of MDIC is almost 98% for the small sample with 50 observations, while the other criteria have a rate of success between 76.9%

1  
2  
3 and 90.9%. We also observe the sample effect. Indeed, by increasing the number of  
4 observations, the only criterion for which the percent of correct model selection increases  
5 significantly is the Hannan and Quinn criterion (HQ). A smaller increase is observed for the  
6 SIC criterion. For large sample sizes SIC and MDIC come first with a rate of success of  
7 approximately 98% and HQ close second with a success rate almost equal to 92%. At the  
8 same time, as the sample size increases, the difference between the MDIC and the remaining  
9 Information Criteria decreases, although it is still in favor of MDIC. The good performance of  
10 MDIC is evident from the high percent of selecting the correct AR(1) model in repeated  
11 samples.  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24

25 While in Table 1 MDIC performs quite well with though some competition from the SIC  
26 criterion, in the case of the second model, as seen in Table 2, the MDIC is superior to all other  
27 criteria for all sample sizes. Observe that MDIC is the best criterion among the competing  
28 criteria with the SIC coming second. The worst performance was observed by AIC. The  
29 sample effect shows the tendency of the best 2 criteria, namely MDIC and SIC, to approach  
30 each other in terms of their rate of success as  $n$  increases. Observe the impressive success rate  
31 of MDIC even for small sample sizes where for  $n=75$  reaches a remarkable success rate of at  
32 least 95%. Our attempt to see if the incomplete AR(2) process misleads the Information  
33 criteria does not seem to work since the rate of success is not significantly affected by the  
34 particular form of the underlying process.  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

50 In case of Models 3 and 5 as Tables 3 and 5 show the results are not different from the  
51 previous cases. Indeed, MDIC is superior to all other criteria with success rate as high as 95%  
52 for small and medium sample sizes and stays slightly behind SIC for large sample sizes.  
53  
54  
55  
56  
57

58 In case of Model 4 as Table 4 shows all criteria for small sample sizes ( $n=50$ ) have a high  
59 tendency of selecting the simplest model (AR(1)) which is due to the fact that the coefficient  
60

1  
2  
3 of lag2 is relatively close to zero (0.35). As the sample size increases the rate of success  
4 improves. MDIC is the best criterion in conjunction with HQ, for small sample sizes, and with  
5  
6 SIC, for large sample sizes.  
7  
8

9  
10  
11 Finally for Model 6, Table 6 shows that for small sample sizes the rate of success is less than  
12 50% with AIC being the best, MDIC being the worse and SIC coming second to last. As the  
13 sample size increases, MDIC and SIC have the higher improvement so that for  $n \geq 100$  they  
14  
15 become again the best models with MDIC reaching an impressive success rate of 98.5% for  
16  $n=200$  and 99.1% for  $n=500$ .  
17  
18

19  
20  
21 It should be pointed out that the choice of the index  $a$  plays an important role in model  
22 selection and controls the rate of success of MDIC. As indicated in Section 2.3 for small lag  
23 models, values of  $a$  in the interval  $[0.21, 0.50]$  appear to maximize the rate. On the other  
24 hand, values of  $a$  in the interval  $[0.05, 0.25]$  produce the best results for large lag models.  
25 Note though that the maximum powers are obtained for values of  $a$  in the upper end of the  
26 former interval, for small lag models and the lower end of the latter, for large lag models. The  
27 value of 0.25 proposed in this work, appears to offer a good balance irrespectively of the type  
28 of the underlying AR model. If the candidate set consists exclusively of AR models of one  
29 type, we suggest a larger value of  $a$ , around 0.35, for small lag models and a smaller one,  
30 around 0.10, for large lag models. In reference to the sample size, our simulations clearly  
31 show that especially for small and medium sample sizes, the MDIC criterion with  $a = 0.25$   
32 provides an ideal model identification technique for both small and large lag models.  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46

47  
48 To summarize the findings we could safely conclude that the MDIC criterion performs much  
49 better than the other selection criteria for AR processes and for various sample sizes. In  
50 particular, the magnitude of the superiority of MDIC is extremely high for AIC and FPE and  
51 relatively high for HQ, for all sample sizes. In reference to the SIC criterion, the superiority of  
52 MDIC is of much lesser magnitude as compared to the other criteria.  
53  
54  
55  
56  
57  
58  
59  
60



#### 4. Concluding Remarks

In this paper we have proposed the Modified Divergence Information Criterion (MDIC) for the determination of the order of an autoregressive process and proved its asymptotic unbiasedness. Furthermore, we performed a comparative study of model selection criteria for autoregressive processes for small, medium, and large sample sizes in order to investigate the practical implications of the new criterion. Based on the results of our simulation study we conclude that for regular AR models, the performance of MDIC is excellent with very high rate of success, for all sample sizes. The rate of success of MDIC increases with sample size but usually not as much as the HQ and SIC criteria which is expected though since they are both consistent. It is important to point out that in almost all cases MDIC with  $a = 0.25$  performs extremely well with a high rate of success (approx. equal to 90%) for both small and medium sample sizes and for all types of models. As a result and especially for small and medium sample sizes, the choice  $a = 0.25$  is highly recommended since it serves as a fair balance between small and large lag models.

For irregular models (like Model 6), MDIC needs a sufficient number of observations for performing well. More specifically, a medium sample size of order 75 or 100 seems to be enough in order to pick up high rates of success.

Finally observe that MDIC is the only criterion that never selects too large models. In fact, in all cases the probability of overestimation is at most 8% and for lags at most 2 or 3 higher than the true lag while other criteria may select lags as large as 7 lags higher than the true one.

A similar behaviour but to a lesser extend is observed by SIC.

Based on the above observations we can conclude that MDIC chooses, at least for regular models, the correct model quite easily even with sample sizes as small as  $n=50$ . At the same time, the rate of success appears to plateau as the sample size increases. As it is known, consistency is an attractive *asymptotic* property. In particular, as  $n$  tends to infinity we have that  $P[\text{correct selection}] \rightarrow 1$ . For MDIC, our results indicate that the probability of correct

1  
2  
3 selection is *significantly close* to 1 even for very small values of  $n$  but stays *below* 1 even for  
4 very large sample sizes. This behaviour may not be the best one from the asymptotic  
5 (theoretical) point of view but it is ideal from the finite one due to desirable practical  
6 implications. In that sense, MDIC may not possess the characteristics of a consistent criterion  
7 but the fact that it is superior to all other criteria for finite sample sizes, makes it an excellent  
8 selection criterion for practical purposes. Furthermore, another attractive feature of MDIC is  
9 the one related to overfitting, where the criterion never selects very large models. Indeed, if  $k^*$   
10 is the true order of the model and  $\hat{k}$  the order selected by MDIC, we observe that  
11 for  $n \geq 50$ ,  $P[\hat{k} = k^* + L, L \geq 2] \leq 0.005$ . In most cases, the probability drops to 0 even for  $L=3$ .  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24

25 **Acknowledgements:** The authors would like to express their appreciation to the Editor and an  
26 anonymous referee for valuable comments and suggestions that greatly improved both the  
27 quality and the presentation of the manuscript. We especially thank the referee whose careful  
28 and thorough reading helped in identifying and correcting a typo in our GAUSS program that  
29 was overlooked.  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table nr 1: Model 1

		Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8
50 obs									
<b>AIC</b>		0,769	0,108	0,049	0,022	0,018	0,012	0,013	0,009
<b>SIC</b>		0,909	0,068	0,015	0,003	0,004	0,001	0	0
<b>HQ</b>		0,836	0,097	0,028	0,015	0,009	0,008	0,005	0,002
<b>FPE</b>		0,769	0,11	0,049	0,021	0,017	0,012	0,013	0,009
<b>MDIC</b>	<b>0.25</b>	0,889	0,095	0,014	0,002	0	0	0	0
	<b>max</b>	0,974	0,026	0	0				
75 obs									
<b>AIC</b>		0,775	0,124	0,04	0,025	0,016	0,009	0,005	0,006
<b>SIC</b>		0,936	0,050	0,011	0,001	0,001	0,001	0	0
<b>HQ</b>		0,873	0,085	0,024	0,008	0,004	0,002	0,001	0,003
<b>FPE</b>		0,777	0,124	0,04	0,025	0,017	0,008	0,003	0,006
<b>MDIC</b>	<b>0.25</b>	0,894	0,086	0,018	0,001	0,001	0	0	0
	<b>max</b>	0,976	0,024	0	0	0			
100 obs									
<b>AIC</b>		0,774	0,107	0,057	0,023	0,019	0,01	0,004	0,006
<b>SIC</b>		0,954	0,034	0,008	0,003	0,001	0	0	0
<b>HQ</b>		0,881	0,071	0,027	0,009	0,009	0,002	0,001	0
<b>FPE</b>		0,774	0,107	0,057	0,023	0,019	0,01	0,005	0,005
<b>MDIC</b>	<b>0.25</b>	0,896	0,082	0,017	0,004	0,001	0	0	0
	<b>max</b>	0,982	0,018	0	0	0			
200 obs									
<b>AIC</b>		0,797	0,116	0,038	0,023	0,012	0,007	0,005	0,002
<b>SIC</b>		0,966	0,032	0,001	0,001	0	0	0	0
<b>HQ</b>		0,916	0,062	0,013	0,005	0,001	0,003	0	0
<b>FPE</b>		0,797	0,116	0,038	0,023	0,012	0,007	0,005	0,002
<b>MDIC</b>	<b>0.25</b>	0,896	0,088	0,012	0,002	0,002	0	0	0
	<b>max</b>	0,981	0,019	0	0	0			
500 obs									
<b>AIC</b>		0,776	0,117	0,049	0,019	0,016	0,013	0,008	0,002
<b>SIC</b>		0,977	0,022	0,001	0	0	0	0	0
<b>HQ</b>		0,918	0,07	0,01	0,001	0,001	0	0	0
<b>FPE</b>		0,776	0,117	0,049	0,019	0,016	0,013	0,008	0,002
<b>MDIC</b>	<b>0.25</b>	0,893	0,098	0,009	0	0	0	0	0
	<b>max</b>	0,977	0,022	0,010					

The shading indicates best performance Information Criterion.

Table nr 2: Model 2

	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8	
50 obs									
AIC	0,018	0,748	0,103	0,05	0,026	0,02	0,015	0,02	
SIC	0,047	0,865	0,063	0,014	0,006	0,003	0,001	0,001	
HQ	0,034	0,805	0,08	0,034	0,02	0,014	0,006	0,007	
FPE	0,018	0,751	0,103	0,047	0,026	0,02	0,015	0,02	
MDIC	0.25	0,037	0,894	0,06	0,008	0,001	0	0	0
	max	0,095	0,902	0,003	0	0			
75 obs									
AIC	0,001	0,777	0,104	0,051	0,028	0,017	0,009	0,013	
SIC	0,005	0,947	0,039	0,008	0,001	0	0	0	
HQ	0,002	0,878	0,076	0,024	0,01	0,006	0,001	0,003	
FPE	0,001	0,777	0,104	0,051	0,028	0,017	0,009	0,013	
MDIC	0.25	0,003	0,948	0,041	0,007	0,001	0	0	0
	max	0,013	0,985	0,002	0	0			
100 obs									
AIC	0,001	0,775	0,12	0,042	0,031	0,011	0,009	0,011	
SIC	0,001	0,944	0,046	0,004	0,005	0	0	0	
HQ	0,001	0,874	0,086	0,019	0,013	0,004	0,003	0	
FPE	0,001	0,776	0,12	0,042	0,031	0,011	0,008	0,011	
MDIC	0.25	0,001	0,932	0,063	0,004	0	0	0	0
	max	0,001	0,993	0,006	0				
200 obs									
AIC	0,001	0,814	0,104	0,027	0,032	0,007	0,004	0,011	
SIC	0,001	0,974	0,022	0,003	0	0	0	0	
HQ	0,001	0,915	0,061	0,013	0,007	0,002	0,001	0	
FPE	0,001	0,814	0,104	0,027	0,032	0,007	0,004	0,011	
MDIC	0.25	0,001	0,945	0,051	0,003	0	0	0	0
	max	0,001	0,996	0,003	0				
500 obs									
AIC	0,001	0,793	0,114	0,042	0,024	0,008	0,011	0,007	
SIC	0,001	0,987	0,012	0	0	0	0	0	
HQ	0,001	0,915	0,066	0,012	0,005	0,001	0	0	
FPE	0,001	0,793	0,114	0,042	0,024	0,008	0,011	0,007	
MDIC	0.25	0,001	0,939	0,057	0,003	0	0	0	0
	max	0,001	0,996	0,003	0				

The shading indicates best performance Information Criterion.

Table nr 3: Model 3

	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8
50 obs								
<b>AIC</b>	0,001	0,749	0,121	0,052	0,024	0,027	0,015	0,011
<b>SIC</b>	0,001	0,903	0,072	0,017	0,003	0,004	0,001	0,001
<b>HQ</b>	0,001	0,825	0,106	0,031	0,011	0,02	0,004	0,002
<b>FPE</b>	0,001	0,749	0,122	0,052	0,024	0,027	0,015	0,01
<b>MDIC</b>	0,001	0,920	0,072	0,007	0	0	0	0
75 obs								
<b>AIC</b>	0,001	0,78	0,094	0,059	0,034	0,012	0,01	0,01
<b>SIC</b>	0,001	0,944	0,039	0,008	0,001	0	0	0
<b>HQ</b>	0,001	0,882	0,068	0,024	0,016	0,003	0,003	0,003
<b>FPE</b>	0,001	0,781	0,094	0,058	0,034	0,012	0,01	0,01
<b>MDIC</b>	0,001	0,944	0,048	0,007	0	0	0	0
100 obs								
<b>AIC</b>	0,001	0,764	0,121	0,046	0,027	0,022	0,013	0,006
<b>SIC</b>	0,001	0,931	0,049	0,007	0,002	0	0	0
<b>HQ</b>	0,001	0,879	0,08	0,019	0,014	0,005	0,002	0
<b>FPE</b>	0,001	0,765	0,121	0,046	0,027	0,022	0,012	0,006
<b>MDIC</b>	0,001	0,933	0,061	0,005	0	0	0	0
200 obs								
<b>AIC</b>	0,001	0,776	0,116	0,042	0,031	0,018	0,009	0,007
<b>SIC</b>	0,001	0,968	0,029	0,002	0	0	0	0
<b>HQ</b>	0,001	0,913	0,066	0,009	0,009	0,002	0	0
<b>FPE</b>	0,001	0,777	0,116	0,042	0,03	0,018	0,009	0,007
<b>MDIC</b>	0,001	0,949	0,047	0,003	0	0	0	0
500 obs								
<b>AIC</b>	0,001	0,795	0,095	0,058	0,025	0,012	0,011	0,003
<b>SIC</b>	0,001	0,983	0,014	0,002	0	0	0	0
<b>HQ</b>	0,001	0,931	0,046	0,018	0,004	0	0	0
<b>FPE</b>	0,001	0,795	0,095	0,058	0,025	0,012	0,011	0,003
<b>MDIC</b>	0,001	0,945	0,048	0,006	0	0	0	0

The shading indicates best performance Information Criterion.

Table nr 4: Model 4

	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8
50 obs								
<b>AIC</b>	0,159	0,607	0,111	0,046	0,026	0,026	0,013	0,012
<b>SIC</b>	0,306	0,622	0,045	0,019	0,005	0,002	0,001	0,001
<b>HQ</b>	0,214	0,643	0,078	0,036	0,009	0,013	0,002	0,005
<b>FPE</b>	0,161	0,61	0,109	0,045	0,026	0,025	0,012	0,012
<b>MDIC</b>	0,268	0,673	0,052	0,007	0	0	0	0
75 obs								
<b>AIC</b>	0,075	0,71	0,112	0,045	0,03	0,015	0,007	0,006
<b>SIC</b>	0,200	0,751	0,037	0,009	0,001	0	0	0
<b>HQ</b>	0,123	0,756	0,074	0,024	0,014	0,003	0,002	0,004
<b>FPE</b>	0,077	0,71	0,112	0,045	0,029	0,014	0,008	0,005
<b>MDIC</b>	0,144	0,800	0,049	0,007	0	0	0	0
100 obs								
<b>AIC</b>	0,033	0,739	0,11	0,054	0,021	0,024	0,01	0,009
<b>SIC</b>	0,102	0,846	0,038	0,010	0,002	0	0	0
<b>HQ</b>	0,059	0,818	0,072	0,033	0,009	0,006	0,002	0,001
<b>FPE</b>	0,033	0,739	0,11	0,054	0,021	0,024	0,01	0,009
<b>MDIC</b>	0,072	0,863	0,051	0,013	0,001	0	0	0
200 obs								
<b>AIC</b>	0,001	0,779	0,114	0,047	0,033	0,014	0,005	0,007
<b>SIC</b>	0,007	0,965	0,019	0,007	0,002	0	0	0
<b>HQ</b>	0,002	0,909	0,06	0,016	0,007	0,005	0,001	0
<b>FPE</b>	0,001	0,779	0,114	0,047	0,033	0,014	0,005	0,007
<b>MDIC</b>	0,003	0,938	0,048	0,009	0,002	0	0	0
500 obs								
<b>AIC</b>	0,001	0,795	0,106	0,053	0,022	0,013	0,008	0,002
<b>SIC</b>	0,001	0,980	0,016	0,003	0	0	0	0
<b>HQ</b>	0,001	0,917	0,061	0,018	0,003	0	0	0
<b>FPE</b>	0,001	0,795	0,106	0,053	0,022	0,013	0,008	0,002
<b>MDIC</b>	0,001	0,933	0,059	0,007	0	0	0	0

The shading indicates best performance Information Criterion.

Table nr 5: Model 5

	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8
50 obs								
<b>AIC</b>	0,025	0,712	0,137	0,045	0,027	0,027	0,013	0,014
<b>SIC</b>	0,203	0,699	0,069	0,014	0,010	0,003	0,001	0,001
<b>HQ</b>	0,042	0,782	0,106	0,031	0,012	0,016	0,004	0,007
<b>FPE</b>	0,026	0,716	0,136	0,045	0,026	0,027	0,012	0,012
<b>MDIC</b>	0,058	0,861	0,073	0,007	0,001	0	0	0
75 obs								
<b>AIC</b>	0,005	0,763	0,127	0,046	0,028	0,014	0,008	0,009
<b>SIC</b>	0,058	0,869	0,062	0,010	0,001	0	0	0
<b>HQ</b>	0,009	0,861	0,086	0,027	0,011	0,003	0,001	0,002
<b>FPE</b>	0,005	0,764	0,127	0,046	0,028	0,014	0,007	0,009
<b>MDIC</b>	0,011	0,927	0,053	0,008	0,001	0	0	0
100 obs								
<b>AIC</b>	0,001	0,759	0,119	0,055	0,024	0,025	0,009	0,008
<b>SIC</b>	0,019	0,918	0,050	0,006	0,006	0,001	0	0
<b>HQ</b>	0,002	0,859	0,089	0,027	0,013	0,009	0,001	0
<b>FPE</b>	0,001	0,76	0,118	0,055	0,024	0,025	0,009	0,008
<b>MDIC</b>	0,002	0,927	0,062	0,007	0,002	0	0	0
200 obs								
<b>AIC</b>	0,001	0,775	0,12	0,044	0,033	0,013	0,009	0,005
<b>SIC</b>	0,001	0,966	0,029	0,003	0,001	0	0	0
<b>HQ</b>	0,001	0,898	0,071	0,019	0,004	0,004	0,003	0
<b>FPE</b>	0,001	0,775	0,12	0,044	0,033	0,013	0,009	0,005
<b>MDIC</b>	0,001	0,941	0,05	0,008	0	0	0	0
500 obs								
<b>AIC</b>	0,001	0,789	0,107	0,052	0,026	0,011	0,01	0,004
<b>SIC</b>	0,001	0,979	0,019	0,001	0	0	0	0
<b>HQ</b>	0,001	0,917	0,055	0,018	0,007	0,002	0	0
<b>FPE</b>	0,001	0,789	0,107	0,052	0,026	0,011	0,01	0,004
<b>MDIC</b>	0,001	0,936	0,056	0,007	0	0	0	0

The shading indicates best performance Information Criterion.

Table nr 6: Model 6

	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8
50 obs								
<b>AIC</b>	0,149	0,02	0,096	0,033	0,498	0,114	0,049	0,041
<b>SIC</b>	0,423	0,025	0,113	0,026	0,358	0,040	0,011	0,004
<b>HQ</b>	0,237	0,022	0,113	0,033	0,466	0,08	0,029	0,02
<b>FPE</b>	0,15	0,02	0,097	0,033	0,497	0,114	0,048	0,041
<b>MDIC</b>	0,495	0,051	0,205	0,026	0,22	0,003	0	0
75 obs								
<b>AIC</b>	0,059	0,001	0,031	0,014	0,702	0,12	0,047	0,026
<b>SIC</b>	0,238	0,007	0,075	0,015	0,627	0,024	0,011	0,003
<b>HQ</b>	0,121	0,008	0,051	0,013	0,701	0,068	0,026	0,012
<b>FPE</b>	0,059	0,001	0,031	0,014	0,704	0,119	0,046	0,026
<b>MDIC</b>	0,284	0,026	0,146	0,016	0,524	0,004	0	0
100 obs								
<b>AIC</b>	0,012	0,001	0,008	0,004	0,751	0,139	0,05	0,035
<b>SIC</b>	0,121	0,006	0,035	0,004	0,787	0,036	0,004	0,007
<b>HQ</b>	0,038	0,004	0,013	0,005	0,813	0,087	0,025	0,015
<b>FPE</b>	0,012	0,001	0,008	0,004	0,752	0,139	0,049	0,035
<b>MDIC</b>	0,128	0,015	0,08	0,011	0,758	0,008	0	0
200 obs								
<b>AIC</b>	0,001	0,001	0,001	0,001	0,801	0,111	0,047	0,037
<b>SIC</b>	0,001	0,001	0,001	0,001	0,961	0,032	0,003	0,001
<b>HQ</b>	0,001	0,001	0,001	0,001	0,901	0,078	0,015	0,002
<b>FPE</b>	0,001	0,001	0,001	0,001	0,801	0,111	0,047	0,037
<b>MDIC</b>	0,002	0,001	0,001	0,001	0,985	0,01	0	0
500 obs								
<b>AIC</b>	0,001	0,001	0,001	0,001	0,813	0,117	0,034	0,032
<b>SIC</b>	0,001	0,001	0,001	0,001	0,973	0,020	0,002	0,001
<b>HQ</b>	0,001	0,001	0,001	0,001	0,922	0,067	0,005	0,002
<b>FPE</b>	0,001	0,001	0,001	0,001	0,813	0,117	0,035	0,031
<b>MDIC</b>	0,001	0,001	0,001	0,001	0,991	0,005	0	0

The shading indicates best performance Information Criterion.



## REFERENCES

Akaike, H. (1969). Fitting autoregression for predictions, *Ann. of the Instit. of Statist. Math.*, 21, 243-247.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, Proc. of the 2<sup>nd</sup> Intern. Symposium on Information Theory, Petrov B. N. and Csáki F., eds., 267-281, Akademiai Kiado, Budapest.

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, AC-19, 716–723.

Akaike, H. (1978). Time series analysis and control through parametric models, in *Applied Time Series Analysis*, D.F. Findley (ed.), Academic Press, New York.

Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence, *Biometrika*, 85, 549–559.

Bengtsson, T. and Cavanaugh. J.E. (2006). An improved Akaike information criterion for state-space model selection, *Computational Statistics and Data Analysis*, 50, 2635-2654.

Billingsley, P. (1961). *Statistical Inference for Markov Processes*. Univ. Chicago Press, Chicago.

Box, G.E.P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Fransisco.

Cavanaugh, J.E. and Johnson, W.O. (1999). Assessing the predictive influence of cases in a stat-space process, *Biometrika*, 86, 183-190.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression, *Journal of Royal Statistical Society B*, 41, 190–195.

Hannan, E. J. (1980). The estimation of the order of an ARMA process, *Ann. Statist.*, 8, 1071-1081.

Jones, R. H. (1975). Fitting autoregressions, *J. Amer. Statist. Assoc.*, 70, 590-592.

Konishi, S. and Kitagawa, G. (1996). Generalised information criteria in model selection, *Biometrika*, 83, 875-890.

Koukouvinos, C. and Mylona, K. (2008). A method for analyzing supersaturated designs with block orthogonal structure, *Commun. in Statist.- Simul. and Comput.*, 37, 290-300.

Koukouvinos, C., Mylona, K. and Vonta, F. (2008). A comparative study of variable selection procedures applied in high dimensional medical problems, *J. of Appl. Prob. and Statist.*, 3 (2), 195-209.

Kullback, S. and Leibler, R. (1951). On information and sufficiency, *Annals of Math. Statist.*, 22, 79-86.

1  
2  
3  
4 Lee, S. (1998). On the quantile process based on the autoregressive residuals. *J. Statist. Plann. Inference*, **67**, 17--28.

7  
8 Mattheou, K. (2007). On new developments in statistical inference for measures of divergence, *PhD Thesis*, Department of Mathematics and Statistics, University of Cyprus.

11  
12 Mattheou, K., Lee, S., and Karagrigoriou, A. (2009). A model selection criterion based on the BHHJ measure of divergence, *J. of Statist. Plann. Inf.* **139**, pp. 128-135.

15  
16 Schwarz, G. (1978). Estimating the dimension of a model, *Ann. Statist.*, **6**, 461–464.

18  
19 Seghouane, A.-K. (2006). Vector autoregressive model order selection from finite samples using Kullback's symmetric divergence, *IEEE Transc. on Circuits and Systems*, **53(10)**, 2327-2335.

22  
23 Seghouane, A.-K., Bekara, M., and Fleury, G. (2005). A criterion for model selection in the presence of incomplete data based on Kullback's symmetric divergence, *Signal Process.*, **85**, 1405-1417.

26  
27 Shang, J. and Cavanaugh, J. E. (2008). Bootstrap variants of the Akaike information criterion for mixed model selection, *Computational Statistics and Data Analysis*, **52**, 2004-2021.

30  
31 Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion, *Biometrika*, **63**, 117-126.

34  
35 Whittle, P. (1952). Test of Fit in Time Series, *Biometrika*, **39**, 309 - 318.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# An Improved **Divergence Information Criterion** for the Determination of the Order of an AR Process

P. Mantalos<sup>a</sup>,

K. Mattheou<sup>b</sup>,

A. Karagrigoriou<sup>b,1</sup>,

<sup>a</sup>Department of Statistics

<sup>b</sup>Department of Mathematics and Statistics

University of Lund

University of Cyprus

Sweden

Cyprus

## Abstract

In this paper we propose a modification of the recently introduced Divergence Information Criterion (DIC, Mattheou et. al, 2009) for the determination of the order of an autoregressive process and show that it is an asymptotically unbiased estimator of the expected overall discrepancy, a nonnegative quantity which measures the distance between the true unknown model and a fitted approximating model. Further, we use Monte Carlo methods and various Data Generating Processes for small, medium and large sample sizes in order to explore the capabilities of the new criterion in selecting the optimal order in autoregressive processes and in general in a time series context. The new criterion shows remarkable good results by choosing the correct model more frequently than traditional Information Criteria.

*Key words:* AR process, Information Criterion, Model selection, Measure of divergence

---

<sup>1</sup> Corresponding author: Dept. of Math. and Stat., University of Cyprus, Cyprus, e-mail: alex@ucy.ac.cy

## 1. Introduction

Consider a discrete-time stationary AR(p) process of the form

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \cdots + a_p X_{t-p} + e_t \quad (1)$$

where  $\{e_t\}$  is a sequence of uncorrelated  $N(0, \sigma^2)$  random variables and all the roots of the

autoregressive polynomial are outside the unit circle, that is,  $A(z) = 1 - a_1 z - \cdots - a_p z^p \neq 0$

for all  $z \in \mathbb{C}$  such that  $|z| < 1$ .

One of the crucial problems in time series modeling is selecting the optimal model order and specifying a parsimonious model for the data generating process (DGP). Many techniques have been developed for the order selection of linear models, namely a) graphical methods with representative the Box and Jenkins methods (1970) and b) hypothesis test procedures with the pioneering work of Whittle (1952) as representative.

Akaike (1974) viewed the problem of model fitting in the context of time series analysis as a multiple decision procedure rather than a hypothesis testing problem. However the same author (Akaike, 1973) initiated the research on automatic model selection techniques by developing the popular Akaike Information Criterion (AIC) which is considered as a significant contribution to statistical modelling where the focus is on the selection of a satisfactory model for a given set of observations. It might appear at first sight that the higher the order of the selected process (model) chosen, the better the fitted model will be. Such a thought may be true for fitting purposes but not for predicting ones. Indeed, the fit may be perfect for the given data but the use of the selected model for predictive purposes may result in gross errors. Numerous selection criteria have been developed over the last 30 years which attempt to prevent overfitting by assigning a penalty to the introduction of an unnecessary

1  
2  
3 large order and consequently to an unnecessary number of parameters. Schwarz (1978),  
4  
5 Hannan and Quinn (1979), Seghouane et. al (2005), Bengtsson and Cavanaugh (2006) and  
6  
7 others followed the pioneering work of Akaike. Most of the criteria proposed in the literature  
8  
9 consist in minimizing, with respect to the model order, a function of the given observations  
10  
11 plus a penalty term for the introduction of additional parameters, which generally depends on  
12  
13 the model order and sample size. Thus, according to automatic criteria, the parameters of a  
14  
15 variety of competing models are estimated and the model chosen is the one with the smallest  
16  
17 criterion value.  
18  
19  
20  
21

22  
23 In this work we propose a modification of the recently introduced Divergence Information  
24  
25 Criterion (DIC) (Mattheou (2007); Mattheou, et. al (2009)) and show that it is an  
26  
27 asymptotically unbiased estimator of the expected overall discrepancy, a nonnegative quantity  
28  
29 which measures the distance between the true unknown model and a fitted approximating  
30  
31 model. Furthermore, we explore the capabilities of the new criterion in selecting the optimal  
32  
33 order in autoregressive processes and in general in a time series context. The paper is  
34  
35 organised as follows. In Section 2, we first briefly discuss various Information Criteria and  
36  
37 their association with measures of divergence and then introduce the Modified Divergence  
38  
39 Information Criterion (MDIC) and provide for practical purposes, an optimum choice for the  
40  
41 index  $\alpha$  which the new criterion depends on. In Section 3 in order to explore the capabilities  
42  
43 of the new MDIC criterion, we perform a comparative study of various Information Criteria  
44  
45 for different time series examples by using a Monte Carlo method. Finally in Section 4 we  
46  
47 summarize our results. For the simulations we make use of the Gauss 8 program.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 2. The New Modified Divergence Information Criterion

### 2.1. Measures of Divergence and Selection Criteria

The measures of divergence are used as indices of similarity or dissimilarity between populations and for measuring mutual information concerning two variables and as such they can be used for the construction of model selection criteria. The distance between a candidate model and the true but unknown model could be measured by a measure of divergence. The candidate model for which the measure of divergence is minimized will be selected. The well known Kullback-Leibler measure of divergence (Kullback and Leibler, 1951) was the one used by Akaike (1973) to develop the Akaike Information Criterion (AIC). Akaike proposed the evaluation of the fit of the candidate model using minus twice the mean expected loglikelihood (also known as expected overall discrepancy). Furthermore, he provided an unbiased estimator of the expected loglikelihood so that the resulting criterion is given by

$$AIC(p) = n \log(\hat{\sigma}_p^2) + 2p,$$

where  $n$  the sample size,  $p$  the order of the candidate model and  $\hat{\sigma}_p^2$  the estimator of the variance of the  $p^{\text{th}}$ -order candidate model.

For fitting autoregressive processes, Jones (1975) suggested that AIC has a tendency to overestimate the order of the process and Shibata (1976) showed that the probability of overestimation for a large sample size is nonzero. To correct this tendency, the SIC criterion was proposed independently by Akaike (1978) and Schwarz (1978):

$$SIC(p) = n \log(\hat{\sigma}_p^2) + p \log n.$$

The above two criteria can be considered as members of the General Information Criterion given by  $GIC(p) = n \log(\hat{\sigma}_p^2) + c_n p$  where  $c_n$  a quantity that may depend on  $n$ . For  $c_n = 2$  the

1  
2  
3 criterion reduces to AIC, while for  $c_n = \log n$  to SIC. The case  $c_n = c \log \log n$  with  $c > 2$   
4  
5  
6 corresponds to the Hannan and Quinn criterion (HQ, Hannan and Quinn, 1979) which has  
7  
8 been found to be equivalent to SIC in the sense that they are both consistent (Hannan, 1980).  
9

10  
11 The problem of avoiding overfitting especially for purely autoregressive processes can also be  
12  
13 dealt with the minimization of the final prediction error criterion (FPE) of Akaike (1969). The  
14  
15 FPE is an estimate of the one-step ahead prediction mean squared error for a realization of the  
16  
17 process independent of the one observed. If we fit AR processes of steadily increasing order  $p$   
18  
19 the maximum likelihood estimator (MLE) of the white noise variance will usually decrease  
20  
21 with  $p$ . However, FPE will decrease to a minimum value and then increase as  $p$  will increase.  
22  
23 According to FPE, we choose the order of the candidate process to be the value of  $p$  for which  
24  
25 FPE is minimized. The FPE is given by  
26  
27  
28  
29

$$30 \quad FPE(p) = \hat{\sigma}_p^2 \frac{n+p}{n-p}.$$

31  
32  
33 In all the above cases, any parameter estimation is handled through the maximum likelihood  
34  
35 method. A general class of criteria has been introduced by Konishi and Kitagawa (1996)  
36  
37 which also estimates the Kullback-Leibler measure where the estimation is not necessarily  
38  
39 based on maximum likelihood.  
40  
41  
42  
43  
44

## 45 2.2. The Development of the MDIC Criterion

46  
47  
48 One of the most recently proposed measures of divergence is the Basu-Harris-Hjort-Jones  
49  
50 power divergence between the candidate model  $f_\theta(\cdot)$  and the true model  $g$  (Basu et. al, 1998)

51  
52 which is denoted by BHHJ, indexed by a positive parameter  $a$ , and defined as:

$$53 \quad I^a(g, f_\theta) = \int \left\{ f_\theta^{1+a}(z) - \left(1 + \frac{1}{a}\right) g(z) f_\theta^a(z) + \left(\frac{1}{a}\right) g^{1+a}(z) \right\} dz.$$

This family of measures was proposed by Basu et al. (1998) for the development of a minimum divergence estimating method for robust parameter estimation. The index  $\alpha$  controls the trade-off between robustness and asymptotic efficiency of the parameter estimators which are the values of  $\theta$  that minimize the measure over a parametric space  $\Theta$ . It should be also noted that the BHHJ family reduces to the Kullback-Leibler measure of divergence for  $\alpha \downarrow 0$  (see Mattheou, 2007) and as it can be easily seen, to the square of the standard  $L_2$  distance between the candidate and the true model for  $\alpha=1$ . Mattheou et al. (2008) applying the same methodology used for AIC to the BHHJ divergence developed a new criterion, the Divergence Information Criterion (DIC) which for a set of observations  $x_1, \dots, x_n$  is given by

$$DIC(p) = nQ_{\hat{\theta}} + (2\pi)^{-a/2} (1+a)^{2+p/2} p,$$

where  $Q_{\hat{\theta}} = \int f_{\hat{\theta}}^{1+a}(z) dz - \left(1 + \frac{1}{a}\right) \frac{1}{n} \sum_{i=1}^n f_{\hat{\theta}}^a(x_i)$  and  $\hat{\theta}$  a consistent and asymptotically normal estimator of  $\theta$ . Preliminary simulation studies for regression models (Mattheou, 2007) showed a very good medium sample size performance for DIC for values of  $\alpha$  close to zero.

Although the DIC criterion was constructed so that it will be an asymptotically unbiased estimator of the BHHJ divergence measure between the candidate and the true model, the calculation of the first part of  $Q_{\hat{\theta}}$ , namely the integral  $\int f_{\hat{\theta}}^{1+a}(z) dz$  is not computationally attractive for practitioners. Furthermore, the simulation study shows that the difference in the calculations of the above integrals, for the different candidate models is negligible compared with the difference in the calculation for the entire quantity  $Q_{\hat{\theta}}$ . In other words the integral term does not affect the selection of the correct model and therefore the criterion can be properly revised. In fact the criterion can be modified without affecting neither its small



sample nor its asymptotic performance. As a result, we propose now a modified new criterion, the Modified Divergence Information Criterion (MDIC) which is given by

$$MDIC(p) = n * MQ_{\hat{\theta}} + (2\pi)^{-a/2} (1+a)^{2+p/2} p$$

$$\text{where } MQ_{\hat{\theta}} = - \left[ (1 + \alpha^{-1}) \frac{1}{n} \sum_{i=1}^n f_{\hat{\theta}}^{\alpha}(x_i) \right].$$

Note that a model selection criterion can be considered as an approximately unbiased estimator of the expected overall discrepancy, a nonnegative quantity which measures the distance between the true unknown model and a fitted approximating model. Observe also that, as it was mentioned in the introduction, a criterion consists of two terms, the first of which is a biased estimator of the expected overall discrepancy. As a result, if we choose the model with the smallest estimator of the expected overall discrepancy we may end up with a selection with an unnecessarily large order. The estimator becomes asymptotically unbiased by introducing the appropriate correction term. **The expected overall discrepancy in the present setting is defined by**

$$E(W_{\hat{\theta}}) = E(W_{\theta} / \theta = \hat{\theta}) \quad \text{with } W_{\theta} = - \int \left\{ \left( 1 + \frac{1}{a} \right) g(z) f_{\theta}^{\alpha}(z) \right\} dz$$

**and  $\hat{\theta}$  a consistent and asymptotically normal estimator of  $\theta$ . Observe that  $W_{\theta}$  is the targeted divergence  $I^{\alpha}(g, f_{\theta})$  without the first and the last terms. Notice that the last term remains constant irrespectively of the candidate model  $f_{\theta}$  while the first term, as stated earlier, is approximately constant for the various candidate models. In that sense, the expected overall discrepancy  $E(W_{\hat{\theta}})$  represents the average distance between the true model  $g$  and the candidate model  $f_{\theta}$  up to a constant.** The relevant theorem for MDIC is stated below. The proof is omitted since it follows the lines of the proof of Theorem 2.4 (Mattheou et.al, 2009) where the unbiasedness of the original DIC criterion was established.

**Theorem 1.** An asymptotically unbiased estimator of  $n$ -times the expected overall discrepancy evaluated at  $\hat{\theta}$  is given by

$$MDIC(p) = n * MQ_{\hat{\theta}} + (2\pi)^{-a/2} (1+a)^{2+p/2} p$$

where  $\hat{\theta}$  is a consistent and asymptotically normal estimator of  $\theta$ .

Note that although the discussion in Mattheou et. al (2009) is limited to iid random variables, by following the approach by Akaike (1973 and 1974) and Billingsley (1961), the same line of discussion can be extended to cover autoregressive models. Recall that in time series models, even under the Gaussian assumption, the evaluation of the log-likelihood cannot be expressed as the sum of the logarithms of the density function of each observation. However, the log-likelihood can generally be expressed by *conditional* distributions, so that in relatively simple models, like the autoregressive models, we use as an approximation the mean *conditional* log-likelihood which multiplied by the number of observations is used in place of the log-likelihood in model selection.

In addition to the above, numerous other techniques have been introduced over the years some of which are associated with the order determination and the forecasting performance in time series. Such techniques include approaches based on bootstrapping (Shang and Cavanaugh, 2008), approaches for high dimensional medical data (Koukouvinos et al., 2008), supersaturated designs (Koukouvinos and Mylona, 2008), variations of AIC (Cavanaugh and Johnson, 1999; Bengtsson and Cavanaugh, 2006; Seghouane, 2006; Seghouane et. al, 2005) and time series models (Lee, 1998).

### 2.3 Optimal Choice for the Index $a$

For practical purposes one has to decide the optimal choice of the positive index  $a$ . For this purpose, we simulate a 100 observation series for 5 different models with  $a$  between 0.01 and

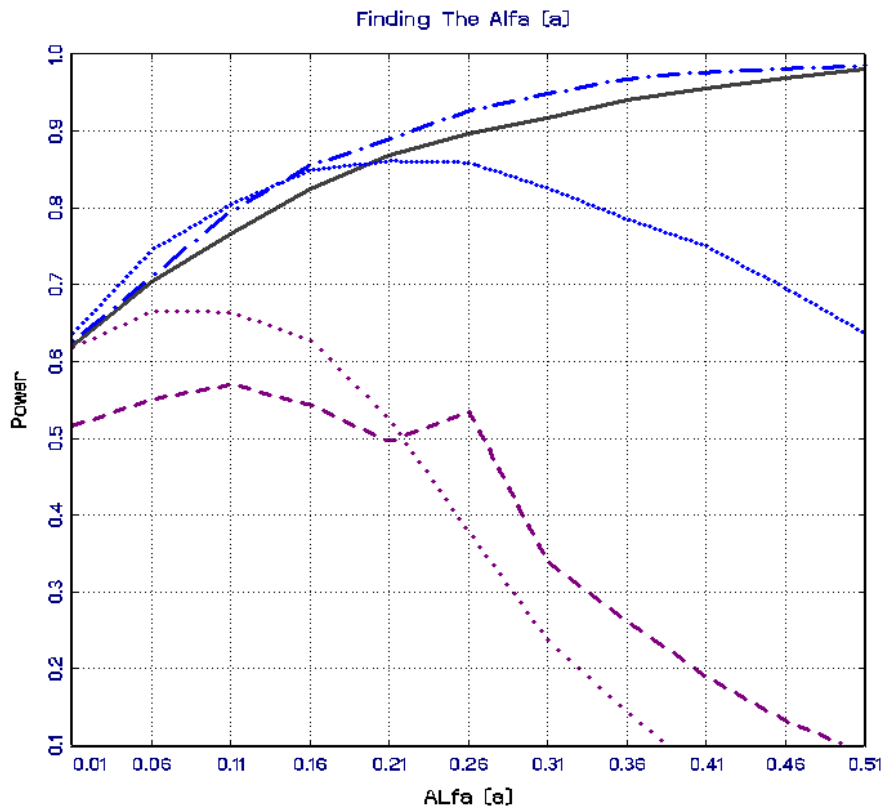
1  
2  
3 0.5 and we provide in Figure 1, the power of the selection, namely the proportion of times the  
4 correct model is selected as a function of  $a$  .  
5  
6  
7

8 In Figure 1, [Dot dash] line is the AR(1) model defined as Model 1 in section 3.2; [Whole]  
9 line shows the AR(2) appearing as Model 5 in section 3.2; [Dot close] line refers to the AR(3)  
10 process:  $x_t = 1 + 0.2x_{t-1} + 0.5x_{t-2} - 0.35x_{t-3} + e_t$ ; [Dash] line is the AR(4) process:  
11  
12  $x_t = 1 + 0.2x_{t-1} + 0.5x_{t-2} - 0.35x_{t-3} - 0.22x_{t-4} + e_t$ ; finally [Dot] is the AR(5) process defined as  
13 Model 6 in section 3.2.  
14  
15  
16  
17  
18

19 Figure 1 shows that for small lags (models AR(1) and AR(2)) the power increases as the value  
20 of  $\alpha$  increases. In all the other cases the power increases up to a value of  $\alpha$  and then  
21 decreases. More specifically, for the complete AR(3) and AR(4) processes the power stays  
22 high approximately up to the value  $a = 0.25$  . For the AR(5) process the power decreases  
23 after the value  $a = 0.10$  .  
24  
25  
26  
27  
28

29 Although an optimum value of  $a$  for all types of autoregressive models may be considered to  
30 be the value of 0.10, in the Monte Carlo study of section 3.2 we choose the value of 0.25 since  
31 it appears to serve as a fair balance between small and large lag models.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Figure 1: Optimal choice of the index  $a$  (AR(1) -•- AR(2) — AR(3) •••• AR(4)---- AR(5)••••)



### 3. Simulations

#### 3.1 The Monte Carlo Experiment

In this section we provide the characteristics of the Monte Carlo experiment undertaken. We calculate the *estimated "percent"* by simply observing how many times the correct AR(p) model is selected in repeated samples. By varying factors such as the number of observations 50 (small sample) 75, 100 (medium sample) and 200, 500 (large sample); and the order of AR(p) model we obtain a succession of estimated percent of the correct selection model under different conditions.

The Monte Carlo experiment has been performed by generating data according to the following Data generating processes:

Model 1:  $x_t = 1 + 0.65x_{t-1} + e_t$

Model 2:  $x_t = 1 + 0.65x_{t-2} + e_t$

Model 3:  $x_t = 1 + 0.5x_{t-1} - 0.85x_{t-2} + e_t$

$$\text{Model 4: } x_t = 1 + 0.5x_{t-1} - 0.35x_{t-2} + e_t$$

$$\text{Model 5: } x_t = 1 + 1.5x_{t-1} - 0.5x_{t-2} + e_t$$

$$\text{Model 6: } x_t = 1 + 0.23x_{t-1} - 0.22x_{t-3} - 0.45x_{t-5} + e_t$$

where  $\{e_t\}$  is a sequence of uncorrelated  $N(0, \sigma^2)$  random variables.

The criteria used in this experiment are the ones defined in Section 2, namely AIC, SIC, HQ with  $c = 2$ , FPE, and MDIC with  $a = 0.25$ . We also obtain the results of MDIC with the index  $a$  taken to be equal to that value in  $[0, 0.50]$  which maximizes the power. According to the analysis in Section 2.3, the maximum powers are obtained for  $a = 0.50$  (Models 1-5) and  $a = 0.10$  (Model 6). For the implementation of the criteria we use

$$f_{\hat{\theta}}(x_i) = \frac{1}{\sqrt{(2\pi\hat{\sigma}^2)}} \exp\left\{-\frac{(x_i - \hat{x}_i)^2}{2\hat{\sigma}^2}\right\},$$

where  $\hat{x}_i$  the predicted value of  $x_i$  and  $\hat{\sigma}^2$  the estimated variance.

For each time series 20 presample values are generated with zero initial conditions, taking net sample sizes of  $n = 50, 75, 100, 200, 500$  in order to cover small, medium and large samples.

The number of replications per model used is 10000. The calculations were performed using GAUSS 8.

### 3.2 Simulation Results

In this section we present the results of the Monte Carlo experiment concerning the percent of correct selected model. Due to insignificant differences between MDIC and MDIC with max power, the powers of the latter for Models 3-6 are suppressed from the Tables.

Table 1 shows the results for Model 1. It is not difficult to see the good performance of the MDIC criterion for all sample sizes. The rate of success of MDIC is almost 98% for the small sample with 50 observations, while the other criteria have a rate of success between 76.9%

1  
2  
3 and 90.9%. We also observe the sample effect. Indeed, by increasing the number of  
4 observations, the only criterion for which the percent of correct model selection increases  
5 significantly is the Hannan and Quinn criterion (HQ). A smaller increase is observed for the  
6 SIC criterion. For large sample sizes SIC and MDIC come first with a rate of success of  
7 approximately 98% and HQ close second with a success rate almost equal to 92%. At the  
8 same time, as the sample size increases, the difference between the MDIC and the remaining  
9 Information Criteria decreases, although it is still in favor of MDIC. The good performance of  
10 MDIC is evident from the high percent of selecting the correct AR(1) model in repeated  
11 samples.  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24

25 While in Table 1 MDIC performs quite well with though some competition from the SIC  
26 criterion, in the case of the second model, as seen in Table 2, the MDIC is superior to all other  
27 criteria for all sample sizes. Observe that MDIC is the best criterion among the competing  
28 criteria with the SIC coming second. The worst performance was observed by AIC. The  
29 sample effect shows the tendency of the best 2 criteria, namely MDIC and SIC, to approach  
30 each other in terms of their rate of success as  $n$  increases. Observe the impressive success rate  
31 of MDIC even for small sample sizes where for  $n=75$  reaches a remarkable success rate of at  
32 least 95%. Our attempt to see if the incomplete AR(2) process misleads the Information  
33 criteria does not seem to work since the rate of success is not significantly affected by the  
34 particular form of the underlying process.  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

50 In case of Models 3 and 5 as Tables 3 and 5 show the results are not different from the  
51 previous cases. Indeed, MDIC is superior to all other criteria with success rate as high as 95%  
52 for small and medium sample sizes and stays slightly behind SIC for large sample sizes.  
53  
54  
55  
56  
57

58 In case of Model 4 as Table 4 shows all criteria for small sample sizes ( $n=50$ ) have a high  
59 tendency of selecting the simplest model (AR(1)) which is due to the fact that the coefficient  
60

1  
2  
3 of lag2 is relatively close to zero (0.35). As the sample size increases the rate of success  
4 improves. MDIC is the best criterion in conjunction with HQ, for small sample sizes, and with  
5  
6 SIC, for large sample sizes.  
7  
8

9  
10  
11 Finally for Model 6, Table 6 shows that for small sample sizes the rate of success is less than  
12 50% with AIC being the best, MDIC being the worse and SIC coming second to last. As the  
13 sample size increases, MDIC and SIC have the higher improvement so that for  $n \geq 100$  they  
14  
15 become again the best models with MDIC reaching an impressive success rate of 98.5% for  
16  
17  $n=200$  and 99.1% for  $n=500$ .  
18  
19  
20  
21  
22  
23

24 It should be pointed out that the choice of the index  $a$  plays an important role in model  
25 selection and controls the rate of success of MDIC. As indicated in Section 2.3 for small lag  
26 models, values of  $a$  in the interval  $[0.21, 0.50]$  appear to maximize the rate. On the other  
27 hand, values of  $a$  in the interval  $[0.05, 0.25]$  produce the best results for large lag models.  
28 Note though that the maximum powers are obtained for values of  $a$  in the upper end of the  
29 former interval, for small lag models and the lower end of the latter, for large lag models. The  
30 value of 0.25 proposed in this work, appears to offer a good balance irrespectively of the type  
31 of the underlying AR model. If the candidate set consists exclusively of AR models of one  
32 type, we suggest a larger value of  $a$ , around 0.35, for small lag models and a smaller one,  
33 around 0.10, for large lag models. In reference to the sample size, our simulations clearly  
34 show that especially for small and medium sample sizes, the MDIC criterion with  $a = 0.25$   
35 provides an ideal model identification technique for both small and large lag models.  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46

47 To summarize the findings we could safely conclude that the MDIC criterion performs much  
48 better than the other selection criteria for AR processes and for various sample sizes. In  
49 particular, the magnitude of the superiority of MDIC is extremely high for AIC and FPE and  
50 relatively high for HQ, for all sample sizes. In reference to the SIC criterion, the superiority of  
51 MDIC is of much lesser magnitude as compared to the other criteria.  
52  
53  
54  
55  
56  
57  
58  
59  
60

#### 4. Concluding Remarks

In this paper we have proposed the Modified Divergence Information Criterion (MDIC) for the determination of the order of an autoregressive process and proved its asymptotic unbiasedness. Furthermore, we performed a comparative study of model selection criteria for autoregressive processes for small, medium, and large sample sizes in order to investigate the practical implications of the new criterion. Based on the results of our simulation study we conclude that for regular AR models, the performance of MDIC is excellent with very high rate of success, for all sample sizes. The rate of success of MDIC increases with sample size but usually not as much as the HQ and SIC criteria which is expected though since they are both consistent. It is important to point out that in almost all cases MDIC with  $a = 0.25$  performs extremely well with a high rate of success (approx. equal to 90%) for both small and medium sample sizes and for all types of models. As a result and especially for small and medium sample sizes, the choice  $a = 0.25$  is highly recommended since it serves as a fair balance between small and large lag models.

For irregular models (like Model 6), MDIC needs a sufficient number of observations for performing well. More specifically, a medium sample size of order 75 or 100 seems to be enough in order to pick up high rates of success.

Finally observe that MDIC is the only criterion that never selects too large models. In fact, in all cases the probability of overestimation is at most 8% and for lags at most 2 or 3 higher than the true lag while other criteria may select lags as large as 7 lags higher than the true one.

A similar behaviour but to a lesser extend is observed by SIC.

Based on the above observations we can conclude that MDIC chooses, at least for regular models, the correct model quite easily even with sample sizes as small as  $n=50$ . At the same time, the rate of success appears to plateau as the sample size increases. As it is known, consistency is an attractive *asymptotic* property. In particular, as  $n$  tends to infinity we have that  $P[\text{correct selection}] \rightarrow 1$ . For MDIC, our results indicate that the probability of correct



1  
2  
3 selection is *significantly close* to 1 even for very small values of  $n$  but stays *below* 1 even for  
4 very large sample sizes. This behaviour may not be the best one from the asymptotic  
5 (theoretical) point of view but it is ideal from the finite one due to desirable practical  
6 implications. It that sense, MDIC may not possess the characteristics of a consistent criterion  
7 but the fact that it is superior to all other criteria for finite sample sizes, makes it an excellent  
8 selection criterion for practical purposes. Furthermore, another attractive feature of MDIC is  
9 the one related to overfitting, where the criterion never selects very large models. Indeed, if  $k^*$   
10 is the true order of the model and  $\hat{k}$  the order selected by MDIC, we observe that  
11 for  $n \geq 50$ ,  $P[\hat{k} = k^* + L, L \geq 2] \leq 0.005$ . In most cases, the probability drops to 0 even for  $L=3$ .  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24

25 **Acknowledgements:** The authors would like to express their appreciation to the Editor and an  
26 anonymous referee for valuable comments and suggestions that greatly improved both the  
27 quality and the presentation of the manuscript. We especially thank the referee whose careful  
28 and thorough reading helped in identifying and correcting a typo in our GAUSS program that  
29 was overlooked.  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table nr 1: Model 1

	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8
50 obs								
AIC	<b>0,769</b>	<b>0,108</b>	<b>0,049</b>	<b>0,022</b>	<b>0,018</b>	<b>0,012</b>	<b>0,013</b>	<b>0,009</b>
SIC	<b>0,909</b>	<b>0,068</b>	<b>0,015</b>	<b>0,003</b>	<b>0,004</b>	<b>0,001</b>	<b>0</b>	<b>0</b>
HQ	<b>0,836</b>	<b>0,097</b>	<b>0,028</b>	<b>0,015</b>	<b>0,009</b>	<b>0,008</b>	<b>0,005</b>	<b>0,002</b>
FPE	<b>0,769</b>	<b>0,11</b>	<b>0,049</b>	<b>0,021</b>	<b>0,017</b>	<b>0,012</b>	<b>0,013</b>	<b>0,009</b>
MDIC	0.25	<b>0,889</b>	<b>0,095</b>	<b>0,014</b>	<b>0,002</b>	0	0	0
	max	<b>0,974</b>	<b>0,026</b>	<b>0</b>	<b>0</b>			
75 obs								
AIC	<b>0,775</b>	<b>0,124</b>	<b>0,04</b>	<b>0,025</b>	<b>0,016</b>	<b>0,009</b>	<b>0,005</b>	<b>0,006</b>
SIC	<b>0,936</b>	<b>0,050</b>	<b>0,011</b>	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0</b>	<b>0</b>
HQ	<b>0,873</b>	<b>0,085</b>	<b>0,024</b>	<b>0,008</b>	<b>0,004</b>	<b>0,002</b>	<b>0,001</b>	<b>0,003</b>
FPE	<b>0,777</b>	<b>0,124</b>	<b>0,04</b>	<b>0,025</b>	<b>0,017</b>	<b>0,008</b>	<b>0,003</b>	<b>0,006</b>
MDIC	0.25	<b>0,894</b>	<b>0,086</b>	<b>0,018</b>	<b>0,001</b>	0	0	0
	max	<b>0,976</b>	<b>0,024</b>	<b>0</b>	<b>0</b>			
100 obs								
AIC	<b>0,774</b>	<b>0,107</b>	<b>0,057</b>	<b>0,023</b>	<b>0,019</b>	<b>0,01</b>	<b>0,004</b>	<b>0,006</b>
SIC	<b>0,954</b>	<b>0,034</b>	<b>0,008</b>	<b>0,003</b>	<b>0,001</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,881</b>	<b>0,071</b>	<b>0,027</b>	<b>0,009</b>	<b>0,009</b>	<b>0,002</b>	<b>0,001</b>	<b>0</b>
FPE	<b>0,774</b>	<b>0,107</b>	<b>0,057</b>	<b>0,023</b>	<b>0,019</b>	<b>0,01</b>	<b>0,005</b>	<b>0,005</b>
MDIC	0.25	<b>0,896</b>	<b>0,082</b>	<b>0,017</b>	<b>0,004</b>	0	0	0
	max	<b>0,982</b>	<b>0,018</b>	<b>0</b>	<b>0</b>			
200 obs								
AIC	<b>0,797</b>	<b>0,116</b>	<b>0,038</b>	<b>0,023</b>	<b>0,012</b>	<b>0,007</b>	<b>0,005</b>	<b>0,002</b>
SIC	<b>0,966</b>	<b>0,032</b>	<b>0,001</b>	<b>0,001</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,916</b>	<b>0,062</b>	<b>0,013</b>	<b>0,005</b>	<b>0,001</b>	<b>0,003</b>	<b>0</b>	<b>0</b>
FPE	<b>0,797</b>	<b>0,116</b>	<b>0,038</b>	<b>0,023</b>	<b>0,012</b>	<b>0,007</b>	<b>0,005</b>	<b>0,002</b>
MDIC	0.25	<b>0,896</b>	<b>0,088</b>	<b>0,012</b>	<b>0,002</b>	0	0	0
	max	<b>0,981</b>	<b>0,019</b>	<b>0</b>	<b>0</b>			
500 obs								
AIC	<b>0,776</b>	<b>0,117</b>	<b>0,049</b>	<b>0,019</b>	<b>0,016</b>	<b>0,013</b>	<b>0,008</b>	<b>0,002</b>
SIC	<b>0,977</b>	<b>0,022</b>	<b>0,001</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,918</b>	<b>0,07</b>	<b>0,01</b>	<b>0,001</b>	<b>0,001</b>	<b>0</b>	<b>0</b>	<b>0</b>
FPE	<b>0,776</b>	<b>0,117</b>	<b>0,049</b>	<b>0,019</b>	<b>0,016</b>	<b>0,013</b>	<b>0,008</b>	<b>0,002</b>
MDIC	0.25	<b>0,893</b>	<b>0,098</b>	<b>0,009</b>	0	0	0	0
	max	<b>0,977</b>	<b>0,022</b>	<b>0,010</b>				

The shading indicates best performance Information Criterion.

Table nr 2: Model 2

	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8
50 obs								
AIC	<b>0,018</b>	<b>0,748</b>	<b>0,103</b>	<b>0,05</b>	<b>0,026</b>	<b>0,02</b>	<b>0,015</b>	<b>0,02</b>
SIC	<b>0,047</b>	<b>0,865</b>	<b>0,063</b>	<b>0,014</b>	<b>0,006</b>	<b>0,003</b>	<b>0,001</b>	<b>0,001</b>
HQ	<b>0,034</b>	<b>0,805</b>	<b>0,08</b>	<b>0,034</b>	<b>0,02</b>	<b>0,014</b>	<b>0,006</b>	<b>0,007</b>
FPE	<b>0,018</b>	<b>0,751</b>	<b>0,103</b>	<b>0,047</b>	<b>0,026</b>	<b>0,02</b>	<b>0,015</b>	<b>0,02</b>
MDIC	0.25	<b>0,037</b>	<b>0,894</b>	<b>0,06</b>	<b>0,008</b>	<b>0,001</b>	<b>0</b>	<b>0</b>
	max	<b>0,095</b>	<b>0,902</b>	<b>0,003</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
75 obs								
AIC	<b>0,001</b>	<b>0,777</b>	<b>0,104</b>	<b>0,051</b>	<b>0,028</b>	<b>0,017</b>	<b>0,009</b>	<b>0,013</b>
SIC	<b>0,005</b>	<b>0,947</b>	<b>0,039</b>	<b>0,008</b>	<b>0,001</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,002</b>	<b>0,878</b>	<b>0,076</b>	<b>0,024</b>	<b>0,01</b>	<b>0,006</b>	<b>0,001</b>	<b>0,003</b>
FPE	<b>0,001</b>	<b>0,777</b>	<b>0,104</b>	<b>0,051</b>	<b>0,028</b>	<b>0,017</b>	<b>0,009</b>	<b>0,013</b>
MDIC	0.25	<b>0,003</b>	<b>0,948</b>	<b>0,041</b>	<b>0,007</b>	<b>0,001</b>	<b>0</b>	<b>0</b>
	max	<b>0,013</b>	<b>0,985</b>	<b>0,002</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
100 obs								
AIC	<b>0,001</b>	<b>0,775</b>	<b>0,12</b>	<b>0,042</b>	<b>0,031</b>	<b>0,011</b>	<b>0,009</b>	<b>0,011</b>
SIC	<b>0,001</b>	<b>0,944</b>	<b>0,046</b>	<b>0,004</b>	<b>0,005</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,001</b>	<b>0,874</b>	<b>0,086</b>	<b>0,019</b>	<b>0,013</b>	<b>0,004</b>	<b>0,003</b>	<b>0</b>
FPE	<b>0,001</b>	<b>0,776</b>	<b>0,12</b>	<b>0,042</b>	<b>0,031</b>	<b>0,011</b>	<b>0,008</b>	<b>0,011</b>
MDIC	0.25	<b>0,001</b>	<b>0,932</b>	<b>0,063</b>	<b>0,004</b>	<b>0</b>	<b>0</b>	<b>0</b>
	max	<b>0,001</b>	<b>0,993</b>	<b>0,006</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
200 obs								
AIC	<b>0,001</b>	<b>0,814</b>	<b>0,104</b>	<b>0,027</b>	<b>0,032</b>	<b>0,007</b>	<b>0,004</b>	<b>0,011</b>
SIC	<b>0,001</b>	<b>0,974</b>	<b>0,022</b>	<b>0,003</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,001</b>	<b>0,915</b>	<b>0,061</b>	<b>0,013</b>	<b>0,007</b>	<b>0,002</b>	<b>0,001</b>	<b>0</b>
FPE	<b>0,001</b>	<b>0,814</b>	<b>0,104</b>	<b>0,027</b>	<b>0,032</b>	<b>0,007</b>	<b>0,004</b>	<b>0,011</b>
MDIC	0.25	<b>0,001</b>	<b>0,945</b>	<b>0,051</b>	<b>0,003</b>	<b>0</b>	<b>0</b>	<b>0</b>
	max	<b>0,001</b>	<b>0,996</b>	<b>0,003</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
500 obs								
AIC	<b>0,001</b>	<b>0,793</b>	<b>0,114</b>	<b>0,042</b>	<b>0,024</b>	<b>0,008</b>	<b>0,011</b>	<b>0,007</b>
SIC	<b>0,001</b>	<b>0,987</b>	<b>0,012</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,001</b>	<b>0,915</b>	<b>0,066</b>	<b>0,012</b>	<b>0,005</b>	<b>0,001</b>	<b>0</b>	<b>0</b>
FPE	<b>0,001</b>	<b>0,793</b>	<b>0,114</b>	<b>0,042</b>	<b>0,024</b>	<b>0,008</b>	<b>0,011</b>	<b>0,007</b>
MDIC	0.25	<b>0,001</b>	<b>0,939</b>	<b>0,057</b>	<b>0,003</b>	<b>0</b>	<b>0</b>	<b>0</b>
	max	<b>0,001</b>	<b>0,996</b>	<b>0,003</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>

The shading indicates best performance Information Criterion.

Table nr 3: Model 3

	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8
50 obs								
AIC	<b>0,001</b>	<b>0,749</b>	<b>0,121</b>	<b>0,052</b>	<b>0,024</b>	<b>0,027</b>	<b>0,015</b>	<b>0,011</b>
SIC	<b>0,001</b>	<b>0,903</b>	<b>0,072</b>	<b>0,017</b>	<b>0,003</b>	<b>0,004</b>	<b>0,001</b>	<b>0,001</b>
HQ	<b>0,001</b>	<b>0,825</b>	<b>0,106</b>	<b>0,031</b>	<b>0,011</b>	<b>0,02</b>	<b>0,004</b>	<b>0,002</b>
FPE	<b>0,001</b>	<b>0,749</b>	<b>0,122</b>	<b>0,052</b>	<b>0,024</b>	<b>0,027</b>	<b>0,015</b>	<b>0,01</b>
MDIC	<b>0,001</b>	<b>0,920</b>	<b>0,072</b>	<b>0,007</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
75 obs								
AIC	<b>0,001</b>	<b>0,78</b>	<b>0,094</b>	<b>0,059</b>	<b>0,034</b>	<b>0,012</b>	<b>0,01</b>	<b>0,01</b>
SIC	<b>0,001</b>	<b>0,944</b>	<b>0,039</b>	<b>0,008</b>	<b>0,001</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,001</b>	<b>0,882</b>	<b>0,068</b>	<b>0,024</b>	<b>0,016</b>	<b>0,003</b>	<b>0,003</b>	<b>0,003</b>
FPE	<b>0,001</b>	<b>0,781</b>	<b>0,094</b>	<b>0,058</b>	<b>0,034</b>	<b>0,012</b>	<b>0,01</b>	<b>0,01</b>
MDIC	<b>0,001</b>	<b>0,944</b>	<b>0,048</b>	<b>0,007</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
100 obs								
AIC	<b>0,001</b>	<b>0,764</b>	<b>0,121</b>	<b>0,046</b>	<b>0,027</b>	<b>0,022</b>	<b>0,013</b>	<b>0,006</b>
SIC	<b>0,001</b>	<b>0,931</b>	<b>0,049</b>	<b>0,007</b>	<b>0,002</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,001</b>	<b>0,879</b>	<b>0,08</b>	<b>0,019</b>	<b>0,014</b>	<b>0,005</b>	<b>0,002</b>	<b>0</b>
FPE	<b>0,001</b>	<b>0,765</b>	<b>0,121</b>	<b>0,046</b>	<b>0,027</b>	<b>0,022</b>	<b>0,012</b>	<b>0,006</b>
MDIC	<b>0,001</b>	<b>0,933</b>	<b>0,061</b>	<b>0,005</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
200 obs								
AIC	<b>0,001</b>	<b>0,776</b>	<b>0,116</b>	<b>0,042</b>	<b>0,031</b>	<b>0,018</b>	<b>0,009</b>	<b>0,007</b>
SIC	<b>0,001</b>	<b>0,968</b>	<b>0,029</b>	<b>0,002</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,001</b>	<b>0,913</b>	<b>0,066</b>	<b>0,009</b>	<b>0,009</b>	<b>0,002</b>	<b>0</b>	<b>0</b>
FPE	<b>0,001</b>	<b>0,777</b>	<b>0,116</b>	<b>0,042</b>	<b>0,03</b>	<b>0,018</b>	<b>0,009</b>	<b>0,007</b>
MDIC	<b>0,001</b>	<b>0,949</b>	<b>0,047</b>	<b>0,003</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
500 obs								
AIC	<b>0,001</b>	<b>0,795</b>	<b>0,095</b>	<b>0,058</b>	<b>0,025</b>	<b>0,012</b>	<b>0,011</b>	<b>0,003</b>
SIC	<b>0,001</b>	<b>0,983</b>	<b>0,014</b>	<b>0,002</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,001</b>	<b>0,931</b>	<b>0,046</b>	<b>0,018</b>	<b>0,004</b>	<b>0</b>	<b>0</b>	<b>0</b>
FPE	<b>0,001</b>	<b>0,795</b>	<b>0,095</b>	<b>0,058</b>	<b>0,025</b>	<b>0,012</b>	<b>0,011</b>	<b>0,003</b>
MDIC	<b>0,001</b>	<b>0,945</b>	<b>0,048</b>	<b>0,006</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>

The shading indicates best performance Information Criterion.

Table nr 4: Model 4

	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8
50 obs								
AIC	<b>0,159</b>	<b>0,607</b>	<b>0,111</b>	<b>0,046</b>	<b>0,026</b>	<b>0,026</b>	<b>0,013</b>	<b>0,012</b>
SIC	<b>0,306</b>	<b>0,622</b>	<b>0,045</b>	<b>0,019</b>	<b>0,005</b>	<b>0,002</b>	<b>0,001</b>	<b>0,001</b>
HQ	<b>0,214</b>	<b>0,643</b>	<b>0,078</b>	<b>0,036</b>	<b>0,009</b>	<b>0,013</b>	<b>0,002</b>	<b>0,005</b>
FPE	<b>0,161</b>	<b>0,61</b>	<b>0,109</b>	<b>0,045</b>	<b>0,026</b>	<b>0,025</b>	<b>0,012</b>	<b>0,012</b>
MDIC	<b>0,268</b>	<b>0,673</b>	<b>0,052</b>	<b>0,007</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
75 obs								
AIC	<b>0,075</b>	<b>0,71</b>	<b>0,112</b>	<b>0,045</b>	<b>0,03</b>	<b>0,015</b>	<b>0,007</b>	<b>0,006</b>
SIC	<b>0,200</b>	<b>0,751</b>	<b>0,037</b>	<b>0,009</b>	<b>0,001</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,123</b>	<b>0,756</b>	<b>0,074</b>	<b>0,024</b>	<b>0,014</b>	<b>0,003</b>	<b>0,002</b>	<b>0,004</b>
FPE	<b>0,077</b>	<b>0,71</b>	<b>0,112</b>	<b>0,045</b>	<b>0,029</b>	<b>0,014</b>	<b>0,008</b>	<b>0,005</b>
MDIC	<b>0,144</b>	<b>0,800</b>	<b>0,049</b>	<b>0,007</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
100 obs								
AIC	<b>0,033</b>	<b>0,739</b>	<b>0,11</b>	<b>0,054</b>	<b>0,021</b>	<b>0,024</b>	<b>0,01</b>	<b>0,009</b>
SIC	<b>0,102</b>	<b>0,846</b>	<b>0,038</b>	<b>0,010</b>	<b>0,002</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,059</b>	<b>0,818</b>	<b>0,072</b>	<b>0,033</b>	<b>0,009</b>	<b>0,006</b>	<b>0,002</b>	<b>0,001</b>
FPE	<b>0,033</b>	<b>0,739</b>	<b>0,11</b>	<b>0,054</b>	<b>0,021</b>	<b>0,024</b>	<b>0,01</b>	<b>0,009</b>
MDIC	<b>0,072</b>	<b>0,863</b>	<b>0,051</b>	<b>0,013</b>	<b>0,001</b>	<b>0</b>	<b>0</b>	<b>0</b>
200 obs								
AIC	<b>0,001</b>	<b>0,779</b>	<b>0,114</b>	<b>0,047</b>	<b>0,033</b>	<b>0,014</b>	<b>0,005</b>	<b>0,007</b>
SIC	<b>0,007</b>	<b>0,965</b>	<b>0,019</b>	<b>0,007</b>	<b>0,002</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,002</b>	<b>0,909</b>	<b>0,06</b>	<b>0,016</b>	<b>0,007</b>	<b>0,005</b>	<b>0,001</b>	<b>0</b>
FPE	<b>0,001</b>	<b>0,779</b>	<b>0,114</b>	<b>0,047</b>	<b>0,033</b>	<b>0,014</b>	<b>0,005</b>	<b>0,007</b>
MDIC	<b>0,003</b>	<b>0,938</b>	<b>0,048</b>	<b>0,009</b>	<b>0,002</b>	<b>0</b>	<b>0</b>	<b>0</b>
500 obs								
AIC	<b>0,001</b>	<b>0,795</b>	<b>0,106</b>	<b>0,053</b>	<b>0,022</b>	<b>0,013</b>	<b>0,008</b>	<b>0,002</b>
SIC	<b>0,001</b>	<b>0,980</b>	<b>0,016</b>	<b>0,003</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,001</b>	<b>0,917</b>	<b>0,061</b>	<b>0,018</b>	<b>0,003</b>	<b>0</b>	<b>0</b>	<b>0</b>
FPE	<b>0,001</b>	<b>0,795</b>	<b>0,106</b>	<b>0,053</b>	<b>0,022</b>	<b>0,013</b>	<b>0,008</b>	<b>0,002</b>
MDIC	<b>0,001</b>	<b>0,933</b>	<b>0,059</b>	<b>0,007</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>

The shading indicates best performance Information Criterion.

Table nr 5: Model 5

	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8
50 obs								
AIC	<b>0,025</b>	<b>0,712</b>	<b>0,137</b>	<b>0,045</b>	<b>0,027</b>	<b>0,027</b>	<b>0,013</b>	<b>0,014</b>
SIC	<b>0,203</b>	<b>0,699</b>	<b>0,069</b>	<b>0,014</b>	<b>0,010</b>	<b>0,003</b>	<b>0,001</b>	<b>0,001</b>
HQ	<b>0,042</b>	<b>0,782</b>	<b>0,106</b>	<b>0,031</b>	<b>0,012</b>	<b>0,016</b>	<b>0,004</b>	<b>0,007</b>
FPE	<b>0,026</b>	<b>0,716</b>	<b>0,136</b>	<b>0,045</b>	<b>0,026</b>	<b>0,027</b>	<b>0,012</b>	<b>0,012</b>
MDIC	<b>0,058</b>	<b>0,861</b>	<b>0,073</b>	<b>0,007</b>	<b>0,001</b>	<b>0</b>	<b>0</b>	<b>0</b>
75 obs								
AIC	<b>0,005</b>	<b>0,763</b>	<b>0,127</b>	<b>0,046</b>	<b>0,028</b>	<b>0,014</b>	<b>0,008</b>	<b>0,009</b>
SIC	<b>0,058</b>	<b>0,869</b>	<b>0,062</b>	<b>0,010</b>	<b>0,001</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,009</b>	<b>0,861</b>	<b>0,086</b>	<b>0,027</b>	<b>0,011</b>	<b>0,003</b>	<b>0,001</b>	<b>0,002</b>
FPE	<b>0,005</b>	<b>0,764</b>	<b>0,127</b>	<b>0,046</b>	<b>0,028</b>	<b>0,014</b>	<b>0,007</b>	<b>0,009</b>
MDIC	<b>0,011</b>	<b>0,927</b>	<b>0,053</b>	<b>0,008</b>	<b>0,001</b>	<b>0</b>	<b>0</b>	<b>0</b>
100 obs								
AIC	<b>0,001</b>	<b>0,759</b>	<b>0,119</b>	<b>0,055</b>	<b>0,024</b>	<b>0,025</b>	<b>0,009</b>	<b>0,008</b>
SIC	<b>0,019</b>	<b>0,918</b>	<b>0,050</b>	<b>0,006</b>	<b>0,006</b>	<b>0,001</b>	<b>0</b>	<b>0</b>
HQ	<b>0,002</b>	<b>0,859</b>	<b>0,089</b>	<b>0,027</b>	<b>0,013</b>	<b>0,009</b>	<b>0,001</b>	<b>0</b>
FPE	<b>0,001</b>	<b>0,76</b>	<b>0,118</b>	<b>0,055</b>	<b>0,024</b>	<b>0,025</b>	<b>0,009</b>	<b>0,008</b>
MDIC	<b>0,002</b>	<b>0,927</b>	<b>0,062</b>	<b>0,007</b>	<b>0,002</b>	<b>0</b>	<b>0</b>	<b>0</b>
200 obs								
AIC	<b>0,001</b>	<b>0,775</b>	<b>0,12</b>	<b>0,044</b>	<b>0,033</b>	<b>0,013</b>	<b>0,009</b>	<b>0,005</b>
SIC	<b>0,001</b>	<b>0,966</b>	<b>0,029</b>	<b>0,003</b>	<b>0,001</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,001</b>	<b>0,898</b>	<b>0,071</b>	<b>0,019</b>	<b>0,004</b>	<b>0,004</b>	<b>0,003</b>	<b>0</b>
FPE	<b>0,001</b>	<b>0,775</b>	<b>0,12</b>	<b>0,044</b>	<b>0,033</b>	<b>0,013</b>	<b>0,009</b>	<b>0,005</b>
MDIC	<b>0,001</b>	<b>0,941</b>	<b>0,05</b>	<b>0,008</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
500 obs								
AIC	<b>0,001</b>	<b>0,789</b>	<b>0,107</b>	<b>0,052</b>	<b>0,026</b>	<b>0,011</b>	<b>0,01</b>	<b>0,004</b>
SIC	<b>0,001</b>	<b>0,979</b>	<b>0,019</b>	<b>0,001</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
HQ	<b>0,001</b>	<b>0,917</b>	<b>0,055</b>	<b>0,018</b>	<b>0,007</b>	<b>0,002</b>	<b>0</b>	<b>0</b>
FPE	<b>0,001</b>	<b>0,789</b>	<b>0,107</b>	<b>0,052</b>	<b>0,026</b>	<b>0,011</b>	<b>0,01</b>	<b>0,004</b>
MDIC	<b>0,001</b>	<b>0,936</b>	<b>0,056</b>	<b>0,007</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>

The shading indicates best performance Information Criterion.

Table nr 6: Model 6

	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6	Lag7	Lag8
50 obs								
AIC	<b>0,149</b>	<b>0,02</b>	<b>0,096</b>	<b>0,033</b>	<b>0,498</b>	<b>0,114</b>	<b>0,049</b>	<b>0,041</b>
SIC	<b>0,423</b>	<b>0,025</b>	<b>0,113</b>	<b>0,026</b>	<b>0,358</b>	<b>0,040</b>	<b>0,011</b>	<b>0,004</b>
HQ	<b>0,237</b>	<b>0,022</b>	<b>0,113</b>	<b>0,033</b>	<b>0,466</b>	<b>0,08</b>	<b>0,029</b>	<b>0,02</b>
FPE	<b>0,15</b>	<b>0,02</b>	<b>0,097</b>	<b>0,033</b>	<b>0,497</b>	<b>0,114</b>	<b>0,048</b>	<b>0,041</b>
MDIC	<b>0,495</b>	<b>0,051</b>	<b>0,205</b>	<b>0,026</b>	<b>0,22</b>	<b>0,003</b>	<b>0</b>	<b>0</b>
75 obs								
AIC	<b>0,059</b>	<b>0,001</b>	<b>0,031</b>	<b>0,014</b>	<b>0,702</b>	<b>0,12</b>	<b>0,047</b>	<b>0,026</b>
SIC	<b>0,238</b>	<b>0,007</b>	<b>0,075</b>	<b>0,015</b>	<b>0,627</b>	<b>0,024</b>	<b>0,011</b>	<b>0,003</b>
HQ	<b>0,121</b>	<b>0,008</b>	<b>0,051</b>	<b>0,013</b>	<b>0,701</b>	<b>0,068</b>	<b>0,026</b>	<b>0,012</b>
FPE	<b>0,059</b>	<b>0,001</b>	<b>0,031</b>	<b>0,014</b>	<b>0,704</b>	<b>0,119</b>	<b>0,046</b>	<b>0,026</b>
MDIC	<b>0,284</b>	<b>0,026</b>	<b>0,146</b>	<b>0,016</b>	<b>0,524</b>	<b>0,004</b>	<b>0</b>	<b>0</b>
100 obs								
AIC	<b>0,012</b>	<b>0,001</b>	<b>0,008</b>	<b>0,004</b>	<b>0,751</b>	<b>0,139</b>	<b>0,05</b>	<b>0,035</b>
SIC	<b>0,121</b>	<b>0,006</b>	<b>0,035</b>	<b>0,004</b>	<b>0,787</b>	<b>0,036</b>	<b>0,004</b>	<b>0,007</b>
HQ	<b>0,038</b>	<b>0,004</b>	<b>0,013</b>	<b>0,005</b>	<b>0,813</b>	<b>0,087</b>	<b>0,025</b>	<b>0,015</b>
FPE	<b>0,012</b>	<b>0,001</b>	<b>0,008</b>	<b>0,004</b>	<b>0,752</b>	<b>0,139</b>	<b>0,049</b>	<b>0,035</b>
MDIC	<b>0,128</b>	<b>0,015</b>	<b>0,08</b>	<b>0,011</b>	<b>0,758</b>	<b>0,008</b>	<b>0</b>	<b>0</b>
200 obs								
AIC	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,801</b>	<b>0,111</b>	<b>0,047</b>	<b>0,037</b>
SIC	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,961</b>	<b>0,032</b>	<b>0,003</b>	<b>0,001</b>
HQ	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,901</b>	<b>0,078</b>	<b>0,015</b>	<b>0,002</b>
FPE	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,801</b>	<b>0,111</b>	<b>0,047</b>	<b>0,037</b>
MDIC	<b>0,002</b>	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,985</b>	<b>0,01</b>	<b>0</b>	<b>0</b>
500 obs								
AIC	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,813</b>	<b>0,117</b>	<b>0,034</b>	<b>0,032</b>
SIC	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,973</b>	<b>0,020</b>	<b>0,002</b>	<b>0,001</b>
HQ	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,922</b>	<b>0,067</b>	<b>0,005</b>	<b>0,002</b>
FPE	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,813</b>	<b>0,117</b>	<b>0,035</b>	<b>0,031</b>
MDIC	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,001</b>	<b>0,991</b>	<b>0,005</b>	<b>0</b>	<b>0</b>

The shading indicates best performance Information Criterion.

## REFERENCES

Akaike, H. (1969). Fitting autoregression for predictions, *Ann. of the Instit. of Statist. Math.*, 21, 243-247.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, Proc. of the 2<sup>nd</sup> Intern. Symposium on Information Theory, Petrov B. N. and Csáki F., eds., 267-281, Akademiai Kiado, Budapest.

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, AC-19, 716–723.

Akaike, H. (1978). Time series analysis and control through parametric models, in *Applied Time Series Analysis*, D.F. Findley (ed.), Academic Press, New York.

Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence, *Biometrika*, 85, 549–559.

Bengtsson, T. and Cavanaugh. J.E. (2006). An improved Akaike information criterion for state-space model selection, *Computational Statistics and Data Analysis*, 50, 2635-2654.

Billingsley, P. (1961). *Statistical Inference for Markov Processes*. Univ. Chicago Press, Chicago.

Box, G.E.P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Fransisco.

Cavanaugh, J.E. and Johnson, W.O. (1999). Assessing the predictive influence of cases in a stat-space process, *Biometrika*, 86, 183-190.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression, *Journal of Royal Statistical Society B*, 41, 190–195.

Hannan, E. J. (1980). The estimation of the order of an ARMA process, *Ann. Statist.*, 8, 1071-1081.

Jones, R. H. (1975). Fitting autoregressions, *J. Amer. Statist. Assoc.*, 70, 590-592.

Konishi, S. and Kitagawa, G. (1996). Generalised information criteria in model selection, *Biometrika*, 83, 875-890.

Koukouvinos, C. and Mylona, K. (2008). A method for analyzing supersaturated designs with block orthogonal structure, *Commun. in Statist.- Simul. and Comput.*, 37, 290-300.

Koukouvinos, C., Mylona, K. and Vonta, F. (2008). A comparative study of variable selection procedures applied in high dimensional medical problems, *J. of Appl. Prob. and Statist.*, 3 (2), 195-209.

Kullback, S. and Leibler, R. (1951). On information and sufficiency, *Annals of Math. Statist.*, 22, 79-86.



1  
2  
3  
4 Lee, S. (1998). On the quantile process based on the autoregressive residuals. *J. Statist. Plann. Inference*, **67**, 17--28.

7  
8 Mattheou, K. (2007). On new developments in statistical inference for measures of divergence, *PhD Thesis*, Department of Mathematics and Statistics, University of Cyprus.

10  
11 Mattheou, K., Lee, S., and Karagrigoriou, A. (2009). A model selection criterion based on the BHHJ measure of divergence, *J. of Statist. Plann. Inf.* **139**, pp. 128-135.

14  
15 Schwarz, G. (1978). Estimating the dimension of a model, *Ann. Statist.*, **6**, 461–464.

17  
18 Seghouane, A.-K. (2006). Vector autoregressive model order selection from finite samples using Kullback's symmetric divergence, *IEEE Transc. on Circuits and Systems*, **53(10)**, 2327-2335.

20  
21 Seghouane, A.-K., Bekara, M., and Fleury, G. (2005). A criterion for model selection in the presence of incomplete data based on Kullback's symmetric divergence, *Signal Process.*, **85**, 1405-1417.

23  
24 Shang, J. and Cavanaugh, J. E. (2008). Bootstrap variants of the Akaike information criterion for mixed model selection, *Computational Statistics and Data Analysis*, **52**, 2004-2021.

26  
27 Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion, *Biometrika*, **63**, 117-126.

29  
30 Whittle, P. (1952). Test of Fit in Time Series, *Biometrika*, **39**, 309 - 318.