E C O N O M I C S B U L L E T I N

Which Lag Length Selection Criteria Should We Employy?

Venus Khim−Sen Liew
Universiti Putra Malaysia

Abstract

Estimating the lag length of autoregressive process for a time series is a crucial econometric exercise in most economic studies. This study attempts to provide helpfully guidelines regarding the use of lag length selection criteria in determining the autoregressive lag length. The most interesting finding of this study is that Akaike's information criterion (AIC) and final prediction error (FPE) are superior than the other criteria under study in the case of small sample (60 observations and below), in the manners that they minimize the chance of under estimation while maximizing the chance of recovering the true lag length. One immediate econometric implication of this study is that as most economic sample data can seldom be considered "large" in size, AIC and FPE are recommended for the estimation the autoregressive lag length.

Citation: Liew, Venus Khim−Sen, (2004) "Which Lag Length Selection Criteria Should We Employ?." *Economics Bulletin*, Vol. 3, No. 33 pp. 1−9
Submitted: May 16, 2004. **Accepted:** September 17, 2004.
URL: http://www.economicsbulletin.com/2004/volume3/EB−04C20021A.pdf

1. Introduction

It is well known that most economic data are time series in nature and that a popular kind of time series model known as autoregressive (AR) model has been directly or indirectly applied in most economic researches. Note that the foremost exercise in the application of AR model is none other than the determination of autoregressive lag length. In this respect, many lag length selection criteria have been employed in economic study to determine the Autoregressive (AR) lag length of time series variables. Briefly, an AR process of lag length p refers to a time series in which its current value is dependent on its first p lagged values and is normally denoted by AR (p). Note that the AR lag length p is always unknown and therefore has to be estimated via various lag length selection criteria such as the Aikaike's information criterion (AIC) (Akaike 1973), Schwarz information criterion (SIC) (Schwarz 1978) Hannan-Quinn criterion (HQC) (Hannan and Quinn 1979), final prediction error (FPE) (Akaike 1969), and Bayesian information criterion (BIC) (Akaike 1979); see Liew (2000) for an overview of these criteria. These criteria especially the AIC have been popularly adopted in economic studies, see for examples the works of Sarantis (1999, 2001) and Baum *et al*. (2001), Baharumshah *et al*. (2002), Ng (2002) and Tang (2003) who employed the AIC, Sarno and Taylor (1998) who employed the AIC and SIC, Ahmed (2000) who used the AIC and BIC, Yamada (2000) who used AIC and HQC, Tan and Baharumshah (1999) and Ibrahim (2001) who deployed the FPE, Dropsy (1996), Azali *et al*. (2001) and Xu (2003) who utilized the SIC in their empirical research. However, no special study has been allocated to contrast the performances of these lag length selection criteria, although few empirical studies (Taylor and Peel 2000, Baum *et al*. 2001, Guerra 2001) do notify the inconsistency of these criterion and their tendency to under estimate the autoregressive lag length[1]. This simulation study is specially conducted to compare the empirical performances of various lag length selected criteria, with the principle objective of discovering the best choice of lag length criteria, an issue which has substantial econometric impact on most empirical economic studies.

The major findings in the current simulation study are previewed as follows. First, these criteria managed to pick up the correct lag length at least half of the time in small sample. Second, this performance increases substantially as sample size grows. Third, with relatively large sample (120 or more observations), HQC is found to outdo the rest in correctly identifying the true lag length. In contrast, AIC and FPE should be a better choice for smaller sample. Fourth, AIC and FPE are found to produce the least probability of under estimation among all criteria under study. Finally, the problem of over estimation, however, is negligible in all cases. The findings in this simulation study, besides providing formal groundwork supportive of the popular choice of AIC in previous empirical researches, may as well serve as useful guiding principles for future economic researches in the determination of autoregressive lag length.

The rest of this paper is organized as follows. Section 2 briefly describes the AR process, the lag length selection criteria and simulation procedure. Section 3 presents and discusses the results of this simulation study. Section 4 offers a summary of this study.

[1] A related work by Liew (2000) studies the performance of an individual criteria, namely the Aikaike's biased corrected information criterion, AICC. The current study is more comprehensive than Liew (2000) in the sense that more criteria are involved for the purpose of comparative study.

2. Methodology of Study

2.1 Autoregressive process

Mathematically, an AR(p) process of a series y_t may be represented by

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_p y_{t-p} + \varepsilon_t \tag{1}$$

where $a_1, a_2, \ldots, a_p$ are autoregressive parameters and ε_t are normally distributed random error terms with a zero mean and a finite variance σ^2.

The estimation of AR (p) process involves 2 stages: First, identify the AR lag length p based on certain rules such as lag length selection criteria. Second, estimate the numerical values for intercept and parameters using regression analysis. This study is confined to the study of the performances of various commonly used lag length selection criteria in identifying the true lag length p. In particular, this study generates AR processes with p arbitrary fixed at a value of 4 and uses these criteria to determine the lag length of each generated series as if the lag length is unknown. The autoregressive parameters are independently generated from uniform distribution with values ranging from -1 to 1 exclusively. Measures are taken to ensure that the sum of these simulated autoregressive parameters is less than unity in magnitude ($|a_1 + a_2 + a_3 + a_4| < 1$) so as to avoid non-stationary AR process. The error term is generated from standard normal distribution. We simulate data sets for various usable sample sizes, S: 30, 60, 120, 240, 480 and 960. For each combination of processes and sample sizes, we simulated 1000 independent series for the purpose of lag length estimation. In every case, the initial value, y_0 is arbitrary set to zero. In an effort to minimize the initial effect, we simulate 3S observations and discard the first 2S observations, leaving the last S observations for lag length estimation. The estimated lag length $\hat{p}$ is allowed to be determined from any integer ranging from 1 to 20 inclusively. In this respect, we compute the values for all 20 lag lengths for each specific criterion and $\hat{p}$ is taken from the one that minimizes that criterion. Note that each criterion independently selects one $\hat{p}$ for the same simulated series.

2.2 Lag length selection criteria

The lag length selection criteria to be evaluated include[2]:

(a) Akaike information criterion, $\text{AIC}_p = -2T[\ln(\hat{\sigma}_p^2)] + 2p$; (2)

(b) Schwarz information criterion, $\text{SIC}_p = \ln(\hat{\sigma}_p^2) + [p\ln(T)]/T$; (3)

(c) Hannan-Quinn criterion, $\text{HQC}_p = \ln(\hat{\sigma}_p^2) + 2T^{-1}p\ln[\ln(T)]$; (4)

(d) the final prediction error, $\text{FPE}_p = \hat{\sigma}_p^2 (T-p)^{-1}(T+p)$ and (5)

[2] Among other criteria not taken up in this study include: First, the Schwert (1987, 1989) criteria, which are defined as $[(4S/100)^{0.25}]$ and $[(12S/100)^{0.25}]$ respectively, with S denoting the sample size and $[A]$ stands for the integer part of the real number A, see for instance Habibullah (2001) and Habibullah and Baharumshah (2001) for their applications. Second, the Akaike's corrected information criterion, $\text{AICC}_p = -2T[\ln(\hat{\sigma}_p^2)] + 2Tp/(T-p)$, see Liew (2000) for a simulation study on its performance as well as its application. Last but not least, the partial autocorrelation function as applied in among others, Taylor and Peel (2000), Guerrra (2001) and Liew *et al*. (2003).

(e) Bayesian information criterion,

$$\mathrm{BIC}_p = (T-p)\ln[(T-p)^{-1}T\sigma_p^2] + T[1+\ln(\sqrt{2\pi})] + p\ln[p^{-1}(\sum_{t=1}^{T} y_t^2 - T\hat{\sigma}_p^2)], \quad (6)$$

where $\hat{\sigma}_p^2 = (T-p-1)^{-1}\sum_{t=p}^{T}\hat{\varepsilon}_t^2$, ε_t is the model's residuals and T is the sample size.

Note that the cap sign (^) indicates an estimated value. Liew (2000) provides an overview on these criteria, whereas details are given in, for instance, Brockwell and Davis (1996) and the references therein.

The main task of this study is to compute the probability of each of these criteria in correctly estimated the true autoregressive lag length. Note that this probability takes a value between zero and one inclusively, with a probability of zero means that the criterion fails to pick up any true lag length and thereby is a poor criterion. On the other hand, a probability of one implies that the criterion manages to correctly select the true lag length in all cases and hence is an excellent criterion.

Besides, we also inspect the selected lag lengths of the estimated lag length for 1000 simulated series of known lag length (that is, p = 4), so as to gain deeper understanding on the performance of various criteria. We will refer to the situation whereby a criterion selected lower lag lengths than the true ones as under estimate, whereas over estimate would mean the selection of higher lag lengths than the true ones.

2.3 Simulation procedure

Briefly, the simulation procedure involves three sub-routines: with the first sub-routine generates a series of from the AR process, whereas the second sub-routine selects the autoregressive lag length of the simulated series and the third sub-routine evaluates the performance of the lag length selection criteria. The algorithm for the simulation procedure for each combination of sample size S and AR lag length p is outlined as follows:

1. Independently generate a_1, a_2 and a_3 from a uniform distribution in the range (-1, 1), conditioned on $|\sum_{i=1}^{4} a_i| < 1$.
2. Generate a series of size 3S from the AR process as represented in Equation (1) of lag length $p = 4$ with a_1, a_2 and a_3 obtained from Step 1. Initialize the starting value, y_0 = 0. Discard the first 2S observations to minimize the effect of initial value.
3. Use each selection criterion to determine the autoregressive lag length ($\hat{p}$) for the last S observations of the series simulated in Step 2. Five selection criteria are involved.
4. Repeat Step 1 to Step 3 for B times, where B is fixed at 1000 in this study.
5. Compute the probabilities of (i) correct estimate, which is computed as $\#(\hat{p} = p)/B$; (ii) under estimate, which is computed as $\#(\hat{p} < p)/B$; and (iii) over estimate, which is computed as $\#(\hat{p} > p)/B$, where #($\bullet$) denotes numbers of time event ($\bullet$) happens.

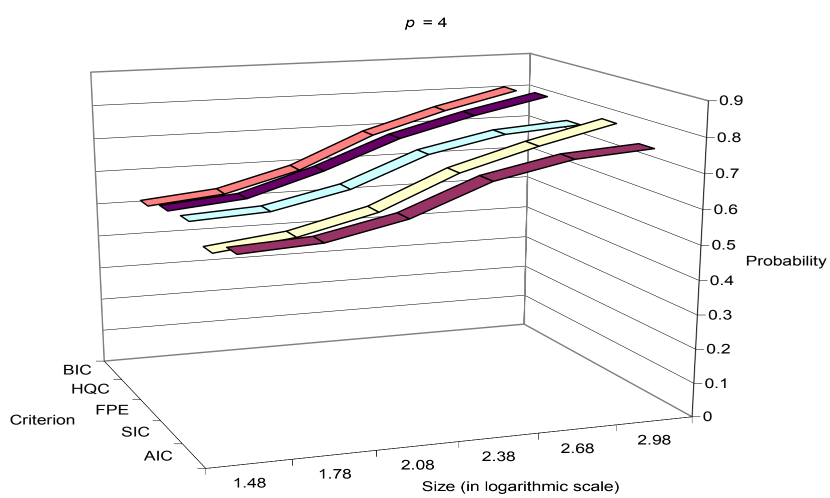
3. Results and discussions

The probability of various criteria in correctly estimated the true lag length of the AR, process is tabulated in Table 1. Generally, Table 1 shows more than half of the time, AIC, SIC, FPE, HQC and BIC correctly estimated the true autoregressive lag length, in all cases. For example, for the case of sample size equals 30, the probability in correctly recovering the true lag length for each of the above criterion is, in that lag length, 0.554, 0.510, 0.554, 0.542 and 0.515. This means that out of 1000 simulated series of known lag length, AIC, SIC, FPE, HQC and BIC respectively have correctly identified the true lag length 554, 510, 554, 642 and 515 times. Table 1 also shows that these criteria perform better and better as the sample size grows. With a sample size of 960, the probability concerned for each of the same the five criteria has reached a value of 0.765, 0.802, 0.765, 0.818 and 0.807 respectively. This conclusion of improvement in performance for each of these five criteria as the sample size grows is clearly depicted in Figure 1. Thus, around 80% of the true lag length has been correctly detected by these five criteria under study. Summing up these two findings, we may conclude that these criteria perform fairly well in picking up the true lag length especially when one has large enough sample size.

Table 1: Probability of correctly estimated the true lag length of AR process, ($\hat{p} = 4$).

Sample Size	Lag length Selection Criteria				
(Logarithmic Scale)	AIC	SIC	FPE	HQC	BIC
30 (1.48)	0.554	0.510	0.554	0.542	0.515
60 (1.78)	0.567	0.537	0.567	0.563	0.537
120 (2.08)	0.616	0.592	0.616	0.631	0.596
240 (2.38)	0.703	0.687	0.703	0.715	0.691
480 (2.68)	0.749	0.750	0.749	0.772	0.755
960 (2.98)	0.765	0.802	0.765	0.818	0.807

Figure 1: Performances of various criteria in correctly selected the true lag length.



The third finding revealed by Table 1 is that AIC and FPE (both constructed by Akaike) seems to have identical performance in terms of their ability to correctly locating the true lag length. In fact, a closer inspection on the selected lag length for each simulated series (results not shown) discovered that they consistently choose the same lag length at all times[3]. One would expect AIC to improve over FPE as it was proposed by Akaike to overcome the inconsistency of the latter (Akaike 1973). However, such improvement is not observed in this study.

An interesting question in mind is whether we can identify the best criterion in selecting the AR lag length. However, it is difficult to just from Table 1 regarding this matter, as no criterion is found to consistently perform better than the rest in all cases. Nonetheless, it is observed that HQC performs substantially better than others, in when the sample size is equal to or larger than 120. However, for sample size smaller than this figure, AIC and FPE turns out to be the better choice.

Further analysis of the distribution of the selected lag lengths is conducted and the results are summarized in Tables 2 and 3. Table 2 reveals that for a sample data containing up to 120 observations, AIC, SIC, FPE, HQC and BIC have under-estimated the true lag length with a probability falling in the range of 0.289 and 0.473 inclusively. On the other hand, the probability of under estimation reduces as sample size grows, to an acceptable extent for a sample size as large as 960, with a respective probability of 0.128, 0.192, 0.128, 0.151 and 0.182. This finding is may be clearly seen from Figure 2. However, as researchers hardly have large sample, identifying the criterion that minimizes the probability of under estimation may be a more practically effort. In this regards, it is observed from Table 2 that AIC and FPE consistently out-do the rest across all sample sizes. Thus, if our objective is to avoid too low the lag length being selected, it is advisable to adopt AIC and/or FPE. The gain in choosing of these two criteria is even significant in sample size of not more than 60 observations. In such ideal case, apart from minimizing the chance of under estimation, one can simultaneously maximize the chance of getting the correct lag length. This conclusion may be taken as formal statistical support for the well-liked use of AIC criterion in previous empirical studies.

Table 2: Probability of under estimated the true lag length of AR process, ($\hat{p} < 4$).

Sample Size (Logarithmic Scale)	Lag length Selection Criteria				
	AIC	SIC	FPE	HQC	BIC
30 (1.48)	0.362	0.473	0.362	0.418	0.463
60 (1.78)	0.353	0.453	0.353	0.402	0.451
120 (2.08)	0.289	0.399	0.289	0.336	0.387
240 (2.38)	0.216	0.307	0.216	0.258	0.299
480 (2.68)	0.168	0.247	0.168	0.201	0.234
960 (2.98)	0.128	0.192	0.128	0.151	0.182

Regarding over estimation, Table 3 shows that AIC, SIC, FPE, HQC and BIC is negligible in all cases regardless of small sample size. In fact, the probability of over estimation is well less than 10% for all criteria across most sample sizes. This empirical finding is in line with the built-in property of these criteria, which are designed in such a way that larger lag length is less preferable, in the spirit of parsimony (that is the simpler the better).

[3] Hence, these two criteria also have the same level of under estimation and over estimation as will be shown in Tables 2 and 3 later.

Figure 2: Performances of various criteria in under estimated the true lag length.

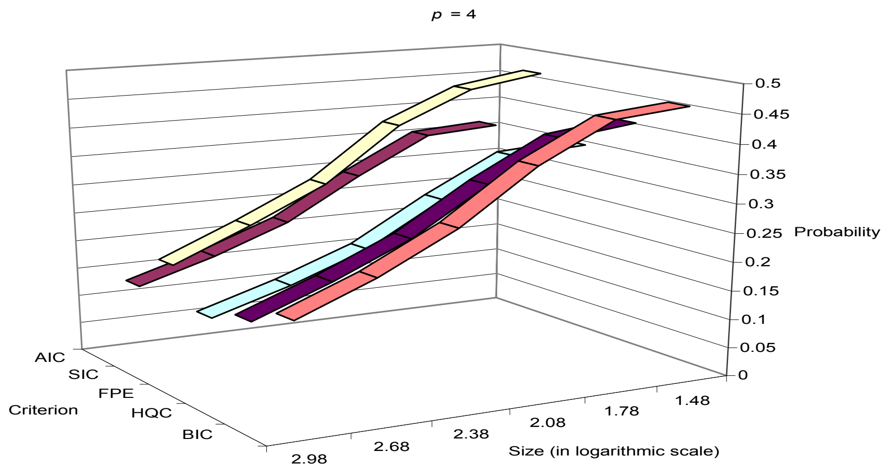


Table 3: Probability of over estimated the true lag length of AR process, ($\hat{p} > 4$).

Sample Size (Logarithmic Scale)	Lag length Selection Criteria				
	AIC	SIC	FPE	HQC	BIC
30 (1.48)	0.084	0.017	0.084	0.040	0.022
60 (1.78)	0.080	0.010	0.080	0.035	0.012
120 (2.08)	0.095	0.009	0.095	0.033	0.017
240 (2.38)	0.081	0.006	0.081	0.027	0.010
480 (2.68)	0.083	0.003	0.083	0.027	0.011
960 (2.98)	0.107	0.006	0.107	0.031	0.011

4. Summary

The determination of autoregressive lag length for a time series is especially important in economics studies. Various lag length selection criteria such as the Aikaike's information criterion (AIC), Schwarz information criterion (SIC), Hannan-Quinn criterion (HQC), final prediction error (FPE) and Bayesian information criterion (BIC) have been employed for this while by researchers in this respect. As the outcomes of these criteria may influence the ultimate findings of a study, a throughout understanding on the empirical performance of these criteria is warranted. This simulation study is specially conducted to shed light on this matter.

The current study independently simulate 1000 series from autoregressive process of known lag length (p = 4) each of the various sample sizes ranging from 30 to 960 observations in each series. Each lag length selection criterion is then allowed to independently estimate the autoregressive lag length for each simulated series, yielding some 1000 selected lag lengths for each criterion. Based on these selected lag lengths, we compute the probabilities in which the true lag length is correctly identified, under estimate and over estimate. The results, which provide useful insights for empirical researchers are summarized as follows.

First, these criteria managed to pick up the correct lag length at least half of the time in small sample. Second, this performance increases substantially as sample size grows. Third, with relatively large sample (120 or more observations), HQC is found to outdo the rest in correctly identifying the true lag length. In contrast, AIC and FPE should be a better choice for smaller sample. Fourth, AIC and FPE are found to produce the least probability of under estimation among all criteria under study. Finally, the problem of over estimation, however, is negligible in all cases. As many econometric testing procedures such as unit root tests, causality tests, cointegration tests and linearity tests involved the determination of autoregressive lag lengths, the findings in this simulation study may be taken as useful guidelines for future economic researches.

References

Ahmed, M. (2000) "Money–income and money–price–causality in selected SAARC countries: some econometric exercises" *Indian Economic Journal* **48**, 55 – 62.

Akaike, H. (1969) "Fitting autoregressive models for prediction" *Annals of the Institute of Statistical Mathematics* **21**, 243 – 247.

Akaike, H. (1973) "Information theory and an extension of the maximum likelihood principle" in *2nd International Symposium on Information Theory* by B. N. Petrov and F. Csaki, eds., Akademiai Kiado: Budapest.

Akaike, H. (1979) "A Bayesian extension of the minimum AIC procedure of Autoregressive model fitting" *Biometrika* **66**, 237 – 242.

Azali, M., A. Z. Baharumshah, and M.S. Habibullah (2001) "Cointegration test for demand for money in Malaysia: Does exchange rate matter?" in *Readings on Malaysia Economy: Issues in Applied Macroeconomics* by M. Azali, ed. Imagepac Print: Kuala Lumpur.

Baharumshah, A. Z., A.M.M. Masih and M. Azali (2002) "The stock market and the ringgit exchange rate: A note" *Japan and the World Economy* **14**, 471 – 486.

Brockwell, P. J., R.A. Davis (1996) *Introduction to Time Series and Forecasting*. Springer: New York.

Dropsy, V. (1996) "Macroeconomics determinants of exchange rates: A frequency-specific analysis" *Applied Economics* **28**, 55 – 63.

Guerra, R. (2001) "Nonlinear adjustment towards Purchasing Power Parity: the Swiss France - German Mark case" Working Paper, Department of Economics, University of Geneva.

Habibullah, M. S. and A. Z. Baharumshah (2001) "Money, output and stock prices in Malaysia" in *Readings on Malaysia Economy: Issues in Applied Macroeconomics* by M. Azali, ed. Imagepac Print: Kuala Lumpur.

Habibullah, M. S. (2001) "Rational expections, survey data and cointegration" in *Readings on Malaysia Economy: Issues in Applied Macroeconomics* by M. Azali, ed. Imagepac Print: Kuala Lumpur.

Ibrahim, M. H.(2001) "Financial factors and the empirical behaviour of money demand: A case study of Malaysia" *International Economic Journal* **15**, 55 – 72.

Hannan, E. J. and B.G. Quinn (1978) "The determination of the lag length of an autoregression" *Journal of Royal Statistical Society* **41**, 190 – 195.

Liew, K. S. (2000) "The performance of AICC as lag length determination criterion in the selection of ARMA time series models" Unpublished Thesis, Department of Mathematics, Universiti Putra Malaysia.

Liew, V. K. S., T. T. L. Chong, K.P. and Lim (2003) ‘The inadequacy of linear autoregressive model for real exchange rates: Empirical evidences from Asian economies” *Applied Economics* **35**, 1387 – 1392.

Ng, T. H. (2002) “Stock Market Linkages in South East Asia” ASEAN *Economic Journal* **16**, 353 – 377.

Sarantis, N. (1999) “Modelling non-linearities in real effective exchange rates” *Journal of International Money and Finance* **18**, 27 – 45.

Sarantis, N. (2001) “Nonlinearities, cyclical behaviour and predictability in stock markets: international evidence” *International Journal of Forecasting* **17**, 439 – 482.

Sarno, L. and M.P. Taylor (1998) “Real exchange rates under the recent float: unequivocal evidence of mean reversion” *Economics Letters* **60**, 131 – 137.

Schwarz, G. (1978) “Estimating the dimension of a model” *Annals of Statistics* **6**, 461 – 464.

Schwert, G. W. (1987) “Effects of model specification on tests for unit roots in macroeconomic data” *Journal of Monetary Economics* **20**, 73 – 103.

Schwert, G. W. (1989) “Tests for unit roots: A Monte Carlo investigation” *Journal of Business and Economics Statistics* **7**, 147 – 159.

Tan, H. B. and A. Z. Baharumshah (1999) “Dynamic causal chain of money, output, interest rate and prices in Malaysia: evidence based on vector error-correction modeling analysis” *International Economic Journal* **13**, 103 – 120.

Tang. T. C. (2003) “Singapore’s aggregrate import demand function: Southeast Asian economies compared” *Labuan Buletin of International Business and Finance* **1**, 13 – 28.

Taylor, M. P. and Peel, D. (2000) “Nonlinear adjustment, long-run equilibrium and exchange rate fundamentals” *Journal of International Money and Finance* **19**, 33 – 53.

Xu, Z. (2003) “Purchasing power parity, price indices, and exchange rate forecasts” *Journal of International Money and Finance* **22**, 105 – 130.

Yamada, H. (2000) “M2 demand relation and effective exchange rate in Japan: a cointegration analysis” *Applied Economics Letters* **7**, 229 – 232.