



ELSEVIER

Economics Letters 59 (1998) 17–22

**economics
letters**

Effects of skewness and kurtosis on model selection criteria

Sıdıka Başçı*, Asad Zaman

Department of Economics, Bilkent University, 06533, Bilkent, Ankara, Turkey

Received 29 July 1997; accepted 4 February 1998

Abstract

We consider the behavior of model selection criteria in AR models where the error terms are not normal by varying skewness and kurtosis. The probability of estimating the true lag order for varying degrees of freedom (k) is the interest. For both small and large samples skewness does not effect the performance of criteria under consideration. On the other hand, kurtosis does effect some of the criteria considerably. In large samples and for large values of k the usual asymptotic theory results for normal models are confirmed. Moreover, we showed that for small sample sizes performance of some newly introduced criteria which were not considered in Monte Carlo studies before are better. © 1998 Elsevier Science S.A.

Keywords: Model Selection Criteria; AR lag order determination; Robustness

JEL classification: C22; C51

1. Introduction

In a Monte-Carlo study Lütkepohl (1985) compares the finite-sample performance of 12 different identification approaches for AR models. In his study, Schwarz Criterion (SC) (Schwarz, 1978; Rissanen, 1978) and Hannan-Quinn Criterion (HQC) (Hannan and Quinn, 1979; Quinn, 1980) emerge as being the best among the compared criteria. Akaike Information Criterion (AIC) (Akaike, 1973, 1974) also performs well. Koreisha and Pukkila (1993) augment Lütkepohl (1985) and show that the performance of the above mentioned criteria depends on the number of nonzero elements of the matrices of the AR parameters and the maximum possible lag order that is used. These two studies and some other prior theoretical and empirical studies assume that the distribution of the error terms are normal. Our object in this paper is to relax this assumption and compare the performance of above mentioned criteria, as well as a few others, when the AR process may have nonnormal errors. We do this by varying skewness and kurtosis. Since it is known that some of the financial data posses skewness and kurtosis it is important for applied econometricians to know the behaviour of various criteria under such circumstances.

We find that although skewness does not have much effect on the performance of criteria while estimating the lag order of the AR model, kurtosis effects the performance of some of the criteria. We reached same conclusion as Lütkepohl (1985) that SC and HQC are best criteria under normal error

*Corresponding author. Tel.: +90 312 2664769; fax: +90 312 2664960/2665140; e-mail: sbasci@bilkent.edu.tr

distributions and large samples. More than that we showed that they are also best for nonnormal error distributions if the sample size is large. For small sample sizes performance of other criteria which are not included in Lütkepohl (1985) are better than SC and HQC under the assumptions of both normal and nonnormal error distributions.

2. Criteria for AR lag order selection

Assume that a given set of data $\mathbf{X} = (X_0, X_1, \dots, X_T)'$ is generated by a stationary AR(p) process (allowing for nonzero mean a_0):

$$X_t = a_0 + a_1 X_{t-1} + \dots + a_p X_{t-p} + u_t$$

where the a_i , $i=0, \dots, p$, are unknown parameters and u_t is an i.i.d error term with common distribution F . Our object is to find p , the true unknown lag order. It is assumed that there is an a prior maximum order M , so that our estimate for p can be any integer between $0, 1, \dots, M$. All prior studies of the issue, both theoretical and empirical, assume the errors to be normally distributed. In this paper, our goal is to assess the performance of these criteria when errors may be nonnormal. From the theory of Edgeworth expansions, we know that the first two terms in asymptotic approximations arising from lack of normality correspond to skewness and kurtosis respectively. Thus a good approach to assess robustness is to vary skewness and kurtosis and study the behavior of the model selection criteria. It appears that skewness does not have much effect on the relative performance of the criteria we studied, while kurtosis does have a great effect.

We first briefly review the model selection criteria included in this study. To describe the criteria, let $\hat{a}_0, \dots, \hat{a}_p$ be the OLS estimates of the parameters of the AR(p) model. Let $\hat{X}_t = \hat{a}_0 + \hat{a}_1 X_{t-1} + \dots + \hat{a}_p X_{t-p}$ be the OLS estimate of X_t . Define the usual variance estimate $\hat{\sigma}_p^2 = (T - (p + 1))^{-1} \sum_{t=p}^T (X_t - \hat{X}_t)^2$. The first criteria AIC was introduced by Akaike (1973), and has been a very popular criterion for model selection. Define $AIC(p) = \ln(\hat{\sigma}_p^2) + \frac{2p}{T}$, and estimate p to be that integer between $1, \dots, M$ which minimizes the criterion AIC. Shibata (1976) shows that AIC is not consistent but overestimate p asymptotically with positive probability if $M \geq p$. Zaman (1984) shows how to calculate the probability distribution of the estimate \hat{p} produced by AIC. Shibata (1980) discusses an asymptotic efficiency property of AIC.

Hurvich and Tsai (1989) makes a bias correction to the AIC, defined by $AIC_C(p) = T \ln(\hat{\sigma}_p^2) + T[1 + (p/T)]/[1 - (p+2)T]$. They show that it is asymptotically efficient if the true model is infinite dimensional. When the true model is finite dimensional AIC_C chooses the true lag order most often compared to other asymptotically efficient criteria.

The Schwarz Criterion, $SC(p) = \ln(\hat{\sigma}_p^2) + p \ln(T)/T$ was introduced by Schwarz (1978) based on Bayesian reasoning. It has the advantage of being consistent over the AIC. This and the Hannan-Quinn criterion $HQC(p) = \ln(\hat{\sigma}_p^2) + 2p \ln(\ln(T))/T$ both perform well in the Lütkepohl (1985) study, and hence are included in our study.

In addition to the AIC_C which was not part of the Lütkepohl (1985) study, we introduce two other criterion which have not been studied in this context. The criterion PRESS, closely related to Quenouille (1949) jackknife, was first introduced by Allen (1974). This is defined as $PRESS(p) =$

$\sum_{t=1}^T (x_t - \tilde{x}_t)^2$ where \tilde{x}_t , $t=1, \dots, T$, is the predicted value of x_t from an AR(p) model after omitting the t -th observation in the process of estimation.¹ The value of p minimizing PRESS is selected.

The last criterion included in our study is the often-used sequential F test. We start with the largest model AR(M), and examine the t statistic of \hat{a}_M . If this is insignificant, we drop it and re-estimate an AR($M-1$) model. We keep dropping the last coefficient until we get a significant one. In the simulation study we took the significance level as 5%.

3. Simulation results

To compare the small sample performance of the criteria discussed in the previous section we have two types of processes where one assumes three lags and the other assumes two lags as the true lag order.² a_0 is one. In order to have a stationary process we generate a_i , $i=1,2,3$, from a uniform distribution in the region $(-1/2, 1/2)$. We set $M=6$. For our robustness studies, we use two classes of error distributions. Let $X \sim \mathcal{G}(k, 2) \equiv \chi_{2k}^2$. To study the effect of skewness, we considered $u_t \stackrel{iid}{\sim} F$, where F is the distribution of $(X/(2\sqrt{k})) - \sqrt{k}$. This has mean 0, variance 1, and $\mathbf{E}u_t^3 = 2\sqrt{k}$ so that skewness increases with k .

Fig. 1a,b below give the probability of correctly estimating p using the various criteria for varying k . Fig. 1a has sample size 30, representing the small sample situation, and Fig. 1b has sample size 200, representing the large sample situation. From the figures we can see that skewness does not have much effect on the performance of criteria while estimating the true lag order. Although AIC_C is the best criterion for sample size 30, it becomes one of the worst criteria for sample size 200. Performance of PRESS is similar to AIC_C . Asymptotically consistent criteria, SC and HQC, are not doing that well for sample size 30 but they estimate the true lag most often when we increase sample size to 200. Performance of AIC and sequential F test are close to each other.

To study the effect of kurtosis, let $u_t \stackrel{iid}{\sim} F_k$ where F is a Student's t distribution scaled to have variance 1. As k increases, the kurtosis decrease, converging to that of the normal asymptotically. Note that skewness is 0 for this class of distributions. When k is one we have the Slash distribution defined in Morgenthaler and Tukey (1991) as a normal divided by a Uniform (0,1).

Fig. 2a,b below give the probabilities for correctly estimating p . We see from the figures that kurtosis has a considerable effect on the criteria under consideration, especially on AIC_C and PRESS. When k is small, that is for heavy-tailed error distributions, performance of PRESS and sequential F test are good but as k increase, that is when we get closer to normal distribution, we see a decline in their performance. Converse is true for AIC_C . The result for AIC_C is what is to be expected because Hurvich and Tsai (1989) makes the bias correction under the assumption of normality. Results can be seen more clearly for the case of sample size 200. Performance of other criteria slightly increase as we move towards normal distribution when sample size is 30. For sample size 200, it seems that kurtosis

¹There are potential problems which arise from omitting middle observations in time series. We made adjustments to account for such problems but found that such adjustments made no difference in the outcome, and hence have chosen to report results without any adjustments.

²Since the results for three lags are similar to the ones for two lags, we only present in this paper the results for two lags.

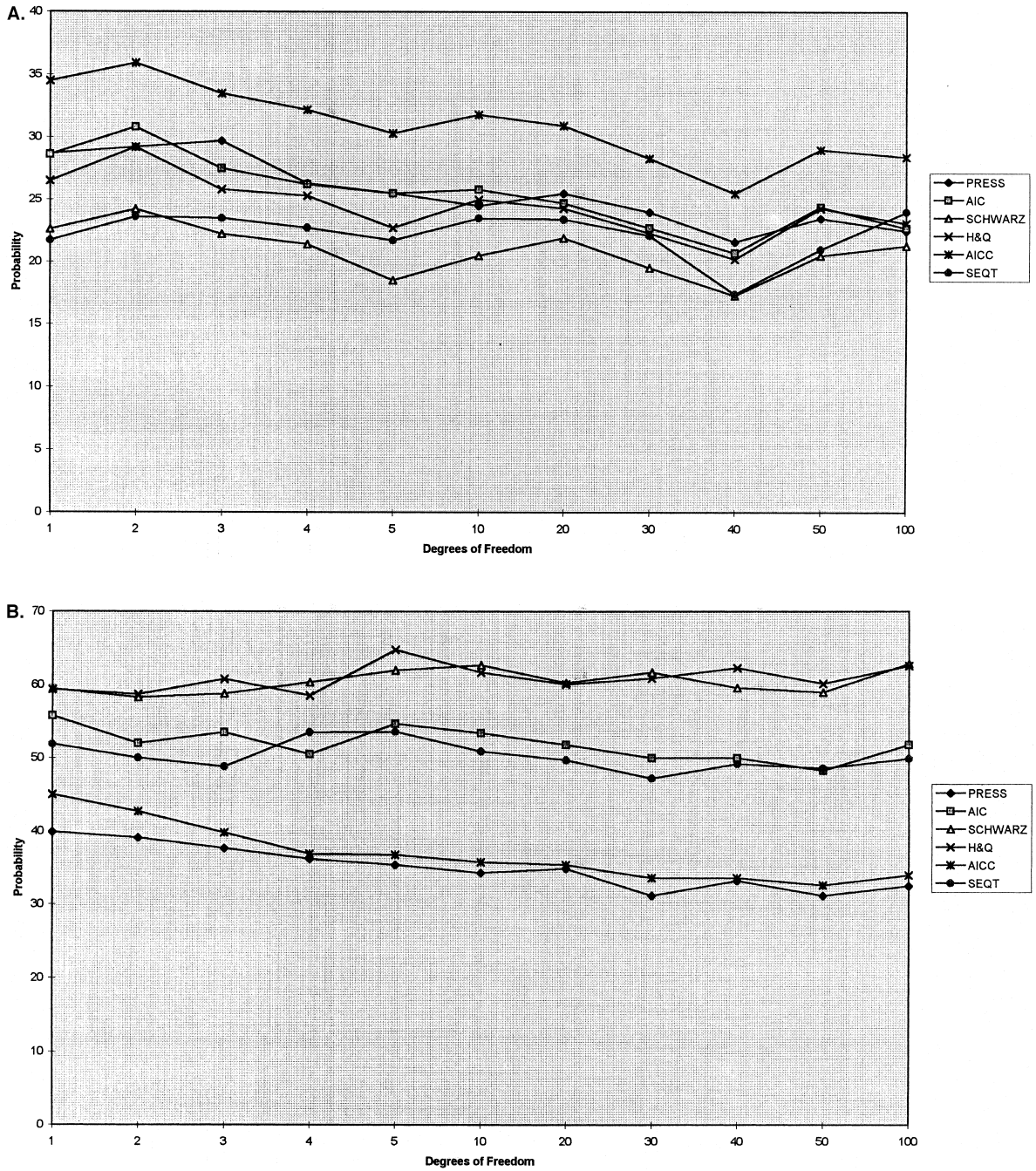


Fig. 1. (a) Probability of estimating the true lag order (true lag order=2, sample size=30, a skewed distribution). (b) Probability of estimating the true lag order (true lag order=2, sample size=200, a skewed distribution).

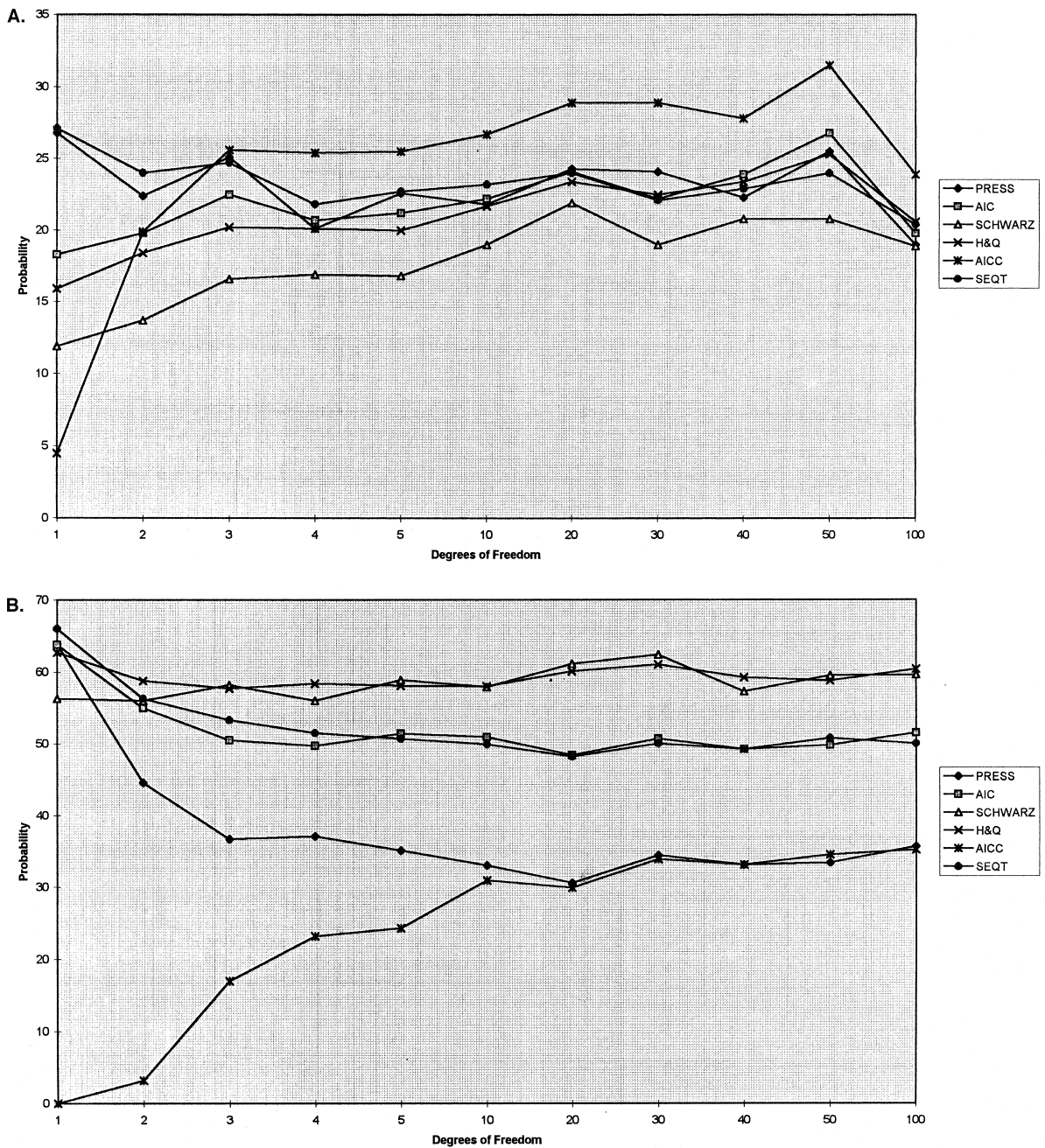


Fig. 2. (a) Probability of estimating the true lag order (true lag order=2, sample size=30, t distribution). (b) Probability of estimating the true lag order (true lag order=2, sample size=200, t distribution).

does not effect these criteria. Once again SC and HQC are performing badly when sample size is 30 but they are the best ones when we increase sample size to 200. PRESS and AIC_C are the worst criteria for sample size 200.

4. Conclusions

To conclude we can say that skewness does not effect the probability of estimating the true lag order but kurtosis effects it especially for the criteria PRESS and AIC_C . When the sample size is small and error distributions are not heavy-tailed (corresponding to large values of degrees of freedom, k) we suggest the use of AIC_C for lag order selection. On the other hand, for heavy-tailed error distributions PRESS and sequential F test perform very well. For large samples SC and HQC are definitely the best criteria to be used.

We also studied forecasting performance based on models selected by the various criteria. This does not correlate very well with the ability of the model to pick out the true p . In addition, we also studied the probability distribution of the estimate of p . This reveals that PRESS has higher probabilities for models size m satisfying $m \geq p$ relative to the other criteria. If the maximum size M is reduced the performance of PRESS improves substantially. Results for the normal distribution are very similar to those for t_{100} .

References

- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csàki, F., 2nd International Symposium on Information Theory. Akadèmiai Kiadó, Budapest, pp. 267–281.
- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* AC 19, 716–723.
- Allen, D.M., 1974. The relationship between variable selection and data augmentation and a method for prediction. *Technometrics* 16, 125–127.
- Hannan, E.J., Quinn, B.G., 1979. The determination of the order of an autoregression. *Journal of the Royal Statistical Society* B41, 190–195.
- Hurvich, C.M., Tsai, C.L., 1989. Regression and time series model selection in small samples. *Biometrika* 76 (2), 297–307.
- Koreisha, S.G., Pukkila, T., 1993. Determining the order of a vector autoregression when the number of component series is large. *Journal of Time Series Analysis* 14 (1), 47–69.
- Lütkepohl, H., 1985. Comparison of criteria for estimating the order of a vector autoregressive process. *Journal of Time Series Analysis* 6 (1), 35–52.
- Quenouille, M., 1949. Approximate tests of correlation in time series. *J. Royal Statistical Society Series B*11, 18–84.
- Morgenthaler, S., Tukey, J.W. (Eds.), 1991. *Configural Polysampling: A Route to Practicle Robustness*. Wiley, New York.
- Quinn, B.G., 1980. Order determination for a multivariate autoregression. *Journal of the Royal Statistical Society* B42, 182–185.
- Rissanen, J., 1978. Modeling by shortest data description. *Automatica* 14, 465–471.
- Schwarz, G., 1978. Estimating the dimension of a model. *The Annals of Statistics* 6, 461–464.
- Shibata, R., 1976. Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* 63, 117–126.
- Shibata, R., 1980. Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *The Annals of Statistics* 8, 147–164.
- Zaman, A., 1984. Avoiding model selection by the use of shrinkage techniques. *Journal of Econometrics* 25, 239–346.