Information criteria

Tamás Jászai & Jenő Fáró

October 2020

1 Introduction

Facing with a modeling task our aim is to find the underlying data generating process (hereinafter DGP). This task might be quite difficult. Although we might have some ideas and we can form several possible models based on theory, professional expertise, etc finally we have to select a model from the set of possible models to proceed with.

One potential tool to accomplish this is the use of model selection criteria. During this analysis we investigate the ability of 2 information criteria (Akaike Information Criterion (hereinafter AIC) and Bayesian Information Criterion (hereinafter BIC)) to identify the true model from among several ARMA-models with different complexity. Whether or not we can find the true DGP during the model-building is of utmost importance which is underpinned by the fact that simulation studies distinguish between the cases of (non)-correctly specified models to the DGP (see Huang & Lee, 2007).

As DGP we consider altogether 3 cases (all of which being ARMA processes with different complexity) starting with ARMA(1,1). We simulate 1000 series from each DGP with 4-4 different sample size and noise level forming a 4 * 4 grid. As possible sample sizes we consider 25; 50; 100 and 1000 the former 3 of which reflects to data recorded with low frequency (with yearly or quarterly frequency we can easily have 25 observation but we rarely have more than 100 data points), while the involvement of 1000 corresponds to a more frequent data recording. The noise levels are determined by the -1st, 0th, 1st and 2nd powers of 2 - in this aspect we follow the methodology of Huang & Lee (2007) but omit the lowest and the highest - almost irrealistically high - noise levels.

We proceed as follows: we firstly simulate time series form different data generating processes then fit different ARMA-models with the (p;q) orders being within the range of (0:2) to the simulated series and let the information criteria to select the most adequate one. Then we calculate in what percentage of the total cases the considered information criteria was able to find the DGP from the set of possible models. We present and briefly discuss the results and give some suggestions about the possible directions of the analysis.

2 Results

2.1 ARMA(1;1)

2.1.1 AIC

As for the AIC it can clearly be seen from the results that its ability for finding the true model does not really depend on the noise level of the data. On the contrary it is strongly dependent of the sample size since besides short time-series (i.e 25 data points) it was only able to find the true model in app. 3 % of the total cases which is lower than 1/9 (the maximum number of AR- and MA-lags considered was 2 thus a model space of 3x3 was explored by the built-in fuction of R (auto.arima) for finding the best model. This indicates a rather poor performance and the phenomenon still prevails with sample sizes equal to 50 and 100. But after increasing the number of data points simulated to 1000 the ratio of correct selections by AIC jumped over 20 %. Besides this sample size the performance of AIC seems still independent from noise level.

AIC	noise = 0.5	noise = 1	noise = 2	noise = 4
n = 25	0.030	0.031	0.032	0.029
n = 50	0.058	0.056	0.052	0.062
n = 100	0.062	0.084	0.078	0.066
n = 1000	0.218	0.209	0.245	0.239

Table 1: Identification performance of AIC - ARMA(1,1)

2.1.2 BIC

As for the BIC the ratio of correctly selected models is even lower - exceeding 5% only in the case of large samples (n = 1000). Again there are no clear patterns regarding the development of the correctly classified cases as the noise level increases. However the sample size again seems to be an important factor: as it increases the performance of the BIC as a model selection tool is getting better but is still not satisfying.

BIC	noise = 0.5	noise = 1	noise = 2	noise = 4
n = 25	0.009	0.014	0.013	0.012
n = 50	0.015	0.019	0.015	0.017
n = 100	0.022	0.020	0.020	0.014
n = 1000	0.051	0.051	0.047	0.052

Table 2: Identification performance of BIC - ARMA(1,1)

2.2 AR(1) and ARMA(2,1)

The results of the ARMA(1,1) case gave us some suggestions about which direction to proceed. Our MA-coefficient was quite low with 0.1 making the process close to an AR(1) process which was underpinned by the highest frequency of AR(1) models selected. Thus we set the MA-coefficient vector to zero and investigated the AR(1) process. We also considered a DGP which is more complex than ARMA(1,1) and finally choose the ARMA(2,1).

The result of these simulations are presented in Tables 3 and 4 and the main differences in comparison to the ARMA(1,1) case highlighted here.

In the AR(1) case BIC proved to be more suitable for selecting the true model than AIC. Another important difference compared to the previous case is that the performance of both information criteria increased and exceeded 50% in almost all of the cases except from the ones with low sample sizes and high level of noise in the series. The ratios are increasing as for both indicators if the sample size increases and in the case of BIC it is very close to 100% if n = 1000. The huge difference might be attributable to the magnitude of the AR- and MA-coefficients: the MA-coefficient was set to 0.1 in the ARMA(1,1) case with the AR-coefficient being -0.5. Thus the DGP was very close to an AR(1) process which explains the high ratio of AR(1) choices when ARMA(1,1) was the DGP. But after the MA-coefficient was set to zero the suggestions of the criteria for AR(1) become correct and thus the ratio of correct identifications jumped from below 30 % to well above 50 %. See Table 3 for the details.

	Criterion	noise = 0.5	noise = 1	noise = 2	noise = 4
n = 25	AIC	0.525	0.525	0.503	0.467
	BIC	0.556	0.521	0.463	0.435
n = 50	AIC	0.570	0.611	0.575	0.565
	BIC	0.657	0.678	0.664	0.631
n = 100	AIC	0.661	0.663	0.673	0.681
	BIC	0.816	0.823	0.820	0.839
n = 1000	AIC	0.659	0.645	0.652	0.670
	BIC	0.964	0.975	0.976	0.976

Table 3: AIC vs BIC - AR(1) case

As for the ARMA(2,1) case the AR-coefficient was set to be quite high with (-1.5 and -0.9) while the MA-coefficient was not changed. And this again resulted in quite high ratios (above 80% in almost all sample-size - noise combination) and the prevalence of BIC the performance of which was not really influenced by the noise level but become almost perfect besides large samples (above 98.5%). For further details please refer to Table 4.

	Criterion	noise = 0.5	noise = 1	noise = 2	noise = 4
n = 25	AIC	0.816	0.820	0.808	0.814
	BIC	0.863	0.845	0.838	0.853
n = 50	AIC	0.784	0.786	0.787	0.810
	BIC	0.911	0.918	0.917	0.920
n = 100	AIC	0.804	0.807	0.835	0.834
	BIC	0.0.958	0.953	0.965	0.960
n = 1000	AIC	0.834	0.853	0.840	0.841
	BIC	0.991	0.992	0.986	0.989

Table 4: AIC vs BIC - ARMA(2,1) case

3 Concluding remarks

Summarizing the results of the analysis we can conclude that the sample size is an important factor as for the performance of the information criteria considered during the analysis and they perform better if we make use of them with large sample analyses.

The noise level of the data does not really play a role in determining how well the criteria performs in selecting the optimal model.

Consequently if we have high frequency data (for example stock prices or returns) we can rely on the information criteria more confidently to select the final model from a set of possible models independently from the noisyness of the data. But if we have low frequency data (financials of companies reported with at most quarterly frequency) we can not be that sure whether we have found the true model based on some information criteria.

Another important aspect which has to be emphasized is the role of the magnitude of the true coefficients. Beside a low MA-coefficient in the ARMA(1,1) DGP neither the performance AIC nor that of BIC was satisfying because they kept voting for (mainly) AR(1) and other models (the RW was also quite frequent which is in line with the fact the neither the AR nor the MA-coefficient was high in absolute terms. Setting the MA-coefficient to zero resulted in a higher rate of correct identifications. And if we applied high AR-coefficient and a low MA-coefficient the identification of ARMA(2,1) was succesfully realized in most of the cases. Thus the magnitude of the AR-coefficients and their relative size compared to the MA-coefficients seem to play an important role whether the true models can be succesfully identified by some information criteria.

This finding indicates the possible future direction of the research: it would be worth to consider several AR(1,1) DGP-s with a fine grid within which the AR- and the MA-coefficient varies from high negative to high positive values (but within the bounds of stationarity) and follow how the relative magnitude of the coefficients influence the performance of the information criteria.

4 References

Huang, H., Lee, T. H. (2010). To combine forecasts or to combine information?. Econometric Reviews, 29(5-6), 534-570.