**R programming task**

**Teremy Zsombor, Till Gábor**

There is a growing body of economic and computational literature which compare the two most applied Akaike (AIC) and Bayesian (BIC) information criteria (e.g. Kuha (2004), Burnham (2004), Vrieze (2012)). The philosophical context of what is assumed about the data generating process, the model approximations, and the aim of the empirical study should determine whether AIC or BIC should be used as an evaluation principle.

The aim of our study is to investigate the hit ratio of the AIC and BIC information criteria on various ARIMA simulations. We generated 1000 ARIMA (1,1) processes with different length of time series and with different noise levels. Logically, the hit ratios must be evaluated for each simulation separately, because different information criterions might perform well under different circumstances.

Time series were simulated using 100, 250, 500, 1000 time series length because we wanted to show that the increasing time series length must be grow the probability that criterions will find the right data generating process. But too long time series need significantly higher computational capacity and it is rare that financial analysts fit models for longer period than 1000 observations without rolling or extended window techniques. To pick standard deviation values we motivated that simulate a process with irrationally low sigma (0.25) and our maximum deviation is 2 since it is an empirically high value.

The hit of the criterion-matched model worth only if both parameters (lag, order) of the data generating process are predicted right. Based on these, finally we compare which information criteria-based method could hit the parameters more often of the thousand simulation.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **AIC** | | | | |
| **t/sig** | **100** | **250** | **500** | **1000** |
| **0.25** | 12% | 17% | 21% | 32% |
| **0.5** | 14% | 13% | 17% | 27% |
| **1** | 8% | 8% | 17% | 38% |
| **2** | 13% | 12% | 32% | 31% |
| **BIC** | | | | |
| **t/sig** | **100** | **250** | **500** | **1000** |
| **0.25** | 3% | 9% | 10% | 9% |
| **0.5** | 6% | 6% | 4% | 7% |
| **1** | 3% | 8% | 3% | 11% |
| **2** | 4% | 5% | 9% | 14% |

Table 1 Performance of BIC and AIC from 1000 cases in percentage in case of different time series length and variance

Table 1. represents our empirical results. Clearly, AIC outperforms BIC independently from the time series length or level of standard deviation. In case of ARMA (1,1), the use of AIC is suggested based on our results. The longer time periods enhance the precision of AIC and BIC- based decision. But information criterions can deal with the higher values of sigma. For example, the highest sigma value in case of BIC cannot cause significantly lower hit rates compared to the others with the same length of time periods.

To draw the conclusion the AIC information criteria could hit the original model parameters more frequently. However, the hit rate could be increased by a greater sample and by a lower noise level.

The analysis of the performance of AIC and BIC would be interesting in case of the different number of ARs and MAs. Our intuition is that the more complex the DSG is the lower performance information criterions have.

Beside the simulations a further comparison of the AIC and BIC criteria could also be implemented on empirical data set from different science fields. Evaluation of the aforementioned models would be interesting in order to support our findings. Different noise levels could be modeled on FX datasets (high noise), macroeconomic data (moderate noise) and average temperate on earth (low noise data). Comparing these results with our conclusion from the simulation could really add to the reliability and credibility of the outcome.

**Biblography**

Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological methods & research*, *33*(2), 188-229.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, *33*(2), 261-304.

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological methods*, *17*(2), 228.