UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Model Selection : Information Criteria Comparison

Marie Clara Fancello, Eva Germini and Eloise Gutfreund

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

University of Panthéon Sorbonne

Model Selection
Information Criteria Comparison

Master's Thesis

To fulfill the requirements for the degree of
Master in Econometrics, Statistics
at University of Panthéon Sorbonne under the supervision of
Prof. Philippe De Perretti

Marie Clara Fancello, Eva Germini and Eloise Gutfreund

Contents

Abstract

In this article, we analyze model selection by Information Criteria. Comparing two estimation methods, different processes, different coefficients for each series and four sample sizes, we analyse Information Criteria performances resulting from these factors. On the one hand, in terms of Estimation Methods, the Maximum Likelihood Estimation does not converge and therefore has no interesting and robust elements to conclude on. On the other hand, the estimation by Generalized Autocorrelation allows us to obtain interesting results. Some Information Criteria give the same performances even if, mathematically, they are different. For the sample size, we see that the more the sample size is large, the more Information Criteria will find the right process and right orders. Underfitting and overfitting percentages vary from one Information Criteria to another depending on the sample size and their penalty function. Another interesting conclusion comes from corrected Information Criteria. They tend to have a lot of underfitting, especially when the sample size increases. As our coefficients vary, their penalty function may be too restrictive. Finally, some processes and even some orders are easily selected by Information Criteria while some others are much more complicated to cast.

1 Introduction

One of the most important questions in temporal econometrics is the selection of the right model. As a matter of fact, the model you will use to describe and analyze your data will define the results you will get. One could think that model selection is easy because it depends on the question the study is trying to answer. Having the right model for your data and problematic is in fact crucial. In temporal econometrics, data are time-based ones and series are studied across time. Therefore, the well estimation of determinist and stochastic terms and orders of the series is very important. In order to describe best the reality, and bring significant conclusions, model selection have to be done and this is the subject of this study.

In order to face this problematic, there are different types of methods that help choosing the right model. Theoretically, the determination of these orders can be done thanks to the AutoCorrelation Function (ACF) which makes it possible to determine the order of the Moving Average process and by the Partial AutoCorrelation Function (PACF) which makes it possible to give the order of the Autoregressive process. This method, which is sometimes too approximate, leads to the use of another method: the method of minimizing the selection criteria. In this paper we will concentrate on the Information Criteria (IC). Information Criteria are estimators of the quality of a model. They estimate the goodness of fit taking into account the number of parameters used in the model. Thus, it allows to know if a model could estimate more efficiently a process depending on the number of parameters considered. Different Information Criteria exist, some are more or less convenient depending on the process observed and the sample size. This study will take eight different Information Criteria : the AIC [Akaike, 1973], the BIC [Akaike, 1979], the FPE [Akaike, 1969], the HQ [Hannan and Quinn, 1978], the SIC [Schwarz, 1978], the SICc [McQuarrie, 1998], the AICc [Hurvich and Tsai, 1989], and the AICu [McQuarrie and Tsai, 1998]. Thus, we will use some commonly used Information Criteria that are well known for their strong performances in large samples, and other ones, lesser-known, that have been made to be consistent even in small samples. In the way they operate, Information Criteria will compare several models. Examining each goodness's of fit and number of parameters, it will select the best model by giving the smallest Information Criteria result.

A problem commonly encountered is due to small sample. When the number of observations is large, Information Criteria tend to select automatically the right model. But when the number of observations is limited, Information Criteria are more likely to select a wrong model. Several studies have already been made on this subject, and we will present them in the next part. In order to account for this problem, we will use sample sizes ranging from 15 (corresponding to a small sample with limited data) to 150 (a threshold that can be considered asymptotic). This will also allow us to study the possible Information Criteria overfitting and underfitting.

Also, the process followed by the times series observed adds complications. Genuinely, the more the series depend on their past stochastic and determinist terms, the more the model selection gets difficult. In fact, it is harder for an Information Criteria to select autoregressive moving average (ARMA) processes than autoregressive (AR) or moving average (MA) ones. Even if Information Criteria have penalty functions, if the sample size is small while the number of lags is high, this penalty function will laboriously overcome these restrictions and will potentially bias the Information Criteria estimation. As real data are never perfect autoregressive or moving average models, we will analyze twelve different ARMA, with orders ranging from 1 to 4 for respectively AR and MA processes, and orders ranging from 1 to 2 for ARMA processes.

Finally, in order to analyze the performances of these Information Criteria in model selection, depending on the processes and sample sizes used, we will proceed using two estimations methods. This will not only allow to study the Information Criteria, but also the differences and performances of estimation methods. One method commonly used is the Maximum Likelihood Estimation, that is quite efficient but also has some limits, the other one is the Generalized Autocorrelation, a lesser-known estimation method. Using a Monte Carlo Simulation, we will vary the ARMA coefficients and repeat our estimations with both methods 10 000 times. This will enable this study to help in model selection not only by knowing what Information Criteria performs best depending on sample size and processes, but also to know what type of method is more efficient when it comes to estimation.

We find interesting conclusions for the three parts of this study. In terms of Estimation Methods, the Maximum Likelihood Estimation does not converge and therefore has no interesting and robust elements to conclude on. For the sample size, we see that the more the sample size is large, the more Information Criteria will find the right process and right orders. Underfitting and overfitting percentages vary from one IC to another depending on the sample size and their penalty function. Finally, some processes and even some orders are easily selected by IC while some others are much more complicated to cast.

This study aims to clarify the selection Criteria problem by comparing estimation methods, processes, sample sizes and the performance resulting from these factors. Indeed, we do not select a single Information Criterion, but rather to help in the choice of the estimation methods, and the criteria, depending on the process studied and the size of the sample. We will begin by presenting the existing literature on the subject. Articles have been made on Information Criteria Selection using different criteria, methodologies and therefore disparate results. Subsequently, we will introduce the Information Criteria selected, processes considered and methodology. Ultimately, we will submit our Monte Carlo study and the obtained results of Information Criteria performances depending on the process, the sample size and their overfitting and underfitting appearance.

2 Overview of the existing literature

In time series analysis, the use of autoregressive processes (AutoRegressive (AR), Moving Average (MA), and ARMA (AutoRegressive Moving Average processes) is essential, however, it is important to determine the right specification of these models. This specification goes through the definition of the orders of the ARMA: the order of the AR part which will be denoted by p and the order of the MA part which will be denoted by q. We will denote k as the sum of p and q.

If selection criteria are relatively equivalent in asymptotic, they are not on finite samples. When the sample size varies, the performances of the criteria will also vary. Through different studies, we will discuss the performance of the criteria in finite sample. In a second part, we will analyze their performance in an infinite sample. Finally, we will see the problem of overfitting and underfitting.

First, we will analyze the performance of the criteria for samples with finite size. The article On Autoregressive Order Selection Criteria [Liew, 2004] analyzes an AR process of order 3 generated on 1000 series of varying sample sizes (from 25 to 1600 observations). For a sample size equal to 25, the results show a better performance of the HQ followed by the BIC (Bayesian Information Criterion), which find respectively 603 times and 596 times the correct model. The performances of the AIC (Akaike Information Criterion), the FPE (Final Prediction Error),and the SIC (Schwarz Information Criterion) is lower, but the gap remains low.

The article "Regression and time series model selection in small samples" [Hurvich and Tsai, 1989] adds a new information criterion: the AICc (Akaike Information Criterion corrected),. The main objective of this study is to observed the performance of the AICc compared to the AIC and the BIC. The AICc adds a penalty for additional parameters. Using AIC when the number of observations is not much greater than the order increases the probability of overfitting. We will discuss overfitting in a next paragraph. The selection criteria used by the author are therefore AIC, AICc, BIC, HQ, FPE and CAT (Criterion Autoregressive Transfer Function) [Parzen, 1977] on different ARMA models (AR(1), AR(2), ARMA(3,5) etc.). When the number of observations is equal to 23, the AICc is more efficient than the AIC and the BIC. For example, for an AR(2), it finds the correct specification of the order in about 81% of all simulations against only 8% for the AIC and 78% for the BIC. For an ARMA (3,5), the AICc still performs better than AIC, however, it is equivalent to BIC.

The article "Information criteria: How do they behave in different models?" [Emiliano et al., 2014] confirms this analysis and add a new element. The authors analyze the criteria AIC, AICc and BIC on eight different ARMA (AR(1), AR(2), ARMA(1,2) ...). When the sample contains 23 observations then the AICc performs better than the BIC and the AIC. However, when the sample size is 100, the BIC performs better than the AIC and the AICc. With an ARMA(1.1), an ARMA(1.2) and an ARMA(2.1) none of the three criteria select the correct model. Inspired by Hurvich and Tsai's 1989 theory of correcting AIC to AICc, the author of the article A small-sample correction for the Schwarz SIC model selection criterion [McQuarrie, 1998] creates a new information criterion: the SICc (Schwarz Information Criterion corrected). The SICc is a criterion which corrects the poor performance of the SIC with a small sample by adding a new penalty criterion (the number of additional parameters). When this new information is added and the number of observations is large, then the SICc has a low impact. However, when the number of observations is low then the penalty has an impact on the criterion performance. In order to analyze the performance of this new criterion, the author simulates an ARMA having a $k = 5$ out of 10 000 series. The criteria are SICc, SIC, HQ, HQc (Hannan-Quinn corrected, which adds the same penalty as the SICc and the AICc), GM [Geweke and Messe, 1981] and AICc. For a sample with 15 observations, the best performing criteria are the corrected criteria (SICc, HQc and AICc) which respectively find the correct order in 86.6%, 82.4% and 84.4% of all simulations. The SIC performance and the HQ performance are much poor. They are about 55.3% for the SIC and 51.1% for the HQ.

Then, we will show through various studies that the performance in an asymptotic sample is relatively equivalent for all the information criteria. Venus Khim-Sen Liew [Liew, 2004] analyzes the criteria discussed previously for sample sizes 50, 100, 200, 400, 800 and 1600. Overall, for all the criteria, the more the sample size increases, the more the criteria perform. These performances are all the more increasing when the sample goes from a size of 50 to 100 (which can be considered as the transition to asymptotic). When the sample size exceeds 200 observations then the BIC and the HQ perform better than the other criteria, however, the differences remain insignificant.

The performance of AICc in asymptotic is analyzed by the authors Liew and Shitan in 2000 [Liew, 2000]. The article The Performance of AICC as an Order Selection Criterion in ARMA Time Series Models use ten ARMA (simulated 100 times) of order p and q ranging from 1 to 4 on a sample size of 555 observations. First, it is important to note that the AICc for this sample performs poorly. It finds the correct model in only 60% of simulations. Also, we observe a performance of 72% for an AR(1) against only 39% for an ARMA(2,2). That means that the higher the orders are, the less efficient this criterion is in asymptotic. The article Information criteria: How do they behave in different models? confirms this analysis. The information criteria generally perform better in asymptotic. In addition, when the sample size is 5000 then the BIC becomes more efficient than the AIC and the AICc. The AIC finds the correct ARMA in 85% versus 96% of all simulations for the BIC.

McQuarrie [McQuarrie, 1998] presents interesting information when the sample size is 100. SICc, HQc and AICc always outperform other information criteria. However, they never find the correct model for an order ranging from 1 to 4. The other criteria do not work for these ARMA either. Also, there is a sharp drop in the performance gaps between these three criteria and the SIC and the HQ. With a sample consisting of 15 observations the difference between the performance of the SICc and the SIC is more than 30 points. While with a sample size of 100 the difference is only 0.02 points.

So far, we have analyzed the performance of the criteria when the orders of the real ARMA were known. However, in empirical analyzes these orders are not known and must be estimated. The article "Bridging AIC and BIC: a new criterion for autoregression" [Jie et al., 2016] adds additional information into the comparison between the BIC criterion and the AIC criterion. They compare the performance of these criteria in a well-specified model and in a mis-specified model. According to them, a pattern is well-specified when the AR part orders and the MA part orders are correct. Conversely, if the AR and/or the MA orders are wrong then the model is mis-specified. This distinction is important because when we analyze time series on real data, we do not know the orders of the process. If the sample size is large with a finite order model and the model is well-specified then the BIC is more efficient than the AIC. The AIC underperforms because its probability of overfitting is fixed when the sample size increases. When the model is mis-specified, the results are reversed. The BIC has a poorer performance. Also, when the orders of autoregressive processes are infinite, then AIC is better.

We analyzed the performance of criteria such as the ability of the model to find the right orders in an autoregressive process. When the criterion is not efficient, then it leads to overfitting or underfitting. When the sum of the AR orders and the MA orders is greater than the known order then there is overfitting. Conversely, when this sum is less than the known order then there is underfitting. In the article by Liew [Liew, 2004], in which the criteria AIC, SIC, FPE, HQ and BIC are used, we find a decreasing relationship between sample size and the probability of underfitting. When the sample size is larger, the probability of underfitting is naturally lower. Conversely, the relationship between sample size and overfitting is positive. It therefore increases with increasing sample size. Also, when we study the probability of underfitting, we do not observe any noticeable difference between the different criteria. When we analyze the probability of overfitting, a difference is observable for AIC and FPE. When the sample size is 1600, the probability is equal to 12.1% for the AIC and the FPE against only 0.8% for the SIC and 1% for the BIC.

In the previous paragraph, we analyzed the probabilities of underfitting and overfitting for criteria that do not have additional penalties. As we studied previously, the AICc, the SICc and the HQc are information criteria that perform better in a finite sample because they have an additional parameter penalty. According to the McQuarrie article [McQuarrie, 1998], AIC and SIC tend to overfit when the sample size is small. The main objective of the corrected criteria is to avoid this overfitting. When the sample size is equal to 15 then the AIC, the SIC and the HQ perform poorly, contrary to the corrected criteria.

Tsai at the origin of the AICc which corrects the overfitting of the AIC and MCQuarrie, author of the SICc correcting the SIC overfittig for finite samples, together create a new criterion: the AICu (Akaike Information Criterion unbaised) [McQuarrie and Tsai, 1998]. The AICu corrects overfitting problems of the AIC and the AICc for asymptotic samples. Their studies focus on AR(2) and AR(5) with sample sizes ranging from 15 to 100. When the sample size is 15 and 25, the probability of AIC overfitting is greater than the probability of AICc overfitting. For a sample size of 25, the probability of overfitting the AIC is 49% versus 0% for the AICc. AICu criterion is also very efficient in a small sample with a probability of 0%. The larger the sample size, the less efficient the AICc is. The AICc has an overfitting probability equal to 0% for a sample size of 35 against 3.7% for a sample size of 100. The AICu, whose main function is to correct the outperformance in asymptotic sample performs perfectly well in asymptotic with zero probability of overfitting.

To conclude, if in asymptotic samples, selection criteria are relatively equivalent (with a better performance of the AICu), in a finite sample, the analysis is more contrasted. Criteria corrected by penalties correct overfitting and obtain better performance in small samples.

3 Methodology

3.1 AR, MA and ARMA processes

The goal of this essay is to identify the best model selection criteria depending on the process studied and the number of observations available. Therefore, it is crucial to identify the process times series are experiencing. There are three major types of processes that we will study: the Autoregressive model so-called AR, the Moving Average model so-called MA, and the Autoregressive Moving Average model so-called ARMA.

The stationarity of autoregressive models (AR, MA, ARMA) is essential to obtain good estimates and forecasts. We define X_t as our autoregressive process. For the process to be stationary, three conditions are required:

- Its expectation must be constant: $E(X_t) = m$ with m a constant.
- Its variance should not depend on time (t): $Var(X_t) = \gamma(0)$ with $\gamma(0)$ a function which does not depend on time.
- Its covariance should not depend on time (t) too: $Cov(X_t, X_{t+h} = \gamma(h)$ with $\gamma(h)$ a function which does not depend on time and h the lag.

We will, in this section, present these models, the hypothesis associated and their stationary.

3.1.1 Autoregressive models

An autoregressive process of order p, noted AR(p) is given by:

$$X_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + ... + \varphi_p X_{t-p} + \varepsilon_t$$

With φ_i the coefficient of variables with a lag i, where $i = 1, .., p$, c being the constant, ε_t a white noise and X_t the serie with X_{t-i} the with a lag i, where $i = 1, .., p$.

An autoregressive model is defined by the impact of the past of a variable on its actual value.This type of process has a white noise and p orders, p being the number of lags impacting X_t. This means that the past noise of this process does not define the actual value of the interest variable. Autoregressive models are not always stationary.

We can also write the AR process with a backshift operator defined as B. It allows to analyse the backshift operator independently of the other variables to define if the process is stationary or not.

$$X_t = c + \sum_{i=1}^{p} \varphi_i B^i X_t + \varepsilon_t$$
$$\Longleftrightarrow \varphi[B] X_t = c + \varepsilon_t$$

Its characteristic polynomial is:

$$\varphi(x) = (1 - \varphi_1 x - \varphi_1 x^2 - ... - \varphi_p x^p$$

If the roots of this polynomial are different from 1, then the process is stationary. If the roots of this polynomial are greater than 1, then the residuals are the process of innovations.

3.1.2 Moving average models

A Moving Average process is defined by the impact of the past stochastic terms on the actual value of the variable analyzed.

$$X_t = c + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_p \varepsilon_{t-p} + \varepsilon_t$$

With c being the constant, θ_i the coefficient of variables with a lag i and ε_{t-i} the white noise with a lag i, where $i = 1,..,q$.

In this equation, the order is q, q being the number of lags of the stochastic variable having an impact on X_t. It is also possible to write X_t with a backshift operator.

$$X_t = c + (1 + \theta_1 B + ... + \theta_q B^q)\varepsilon_t$$

According to Wold's theorem, the MA is always stationary: For any stationary process there is an MA infinite which defines it. For a stationary process (X_t) centered, there exists a white noise (ε_t) and a series of infinite order and of coefficients (θ) such as:

$$\forall t, X_t = \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_h \varepsilon_{t-h} + ...$$
$$\iff \forall t, X_t = \sum_{h \geq 0} \theta_h \varepsilon_{t-h}$$

Where $\lim\limits_{H \to +\infty} E[(\sum_{h \geq H} \theta_h \varepsilon_{t-h})^2] \underset{H \to +\infty}{\longrightarrow} 0$
$\iff \lim\limits_{H \to +\infty} E[\sum_{h \geq H} \theta_h \varepsilon_{t-h}] \underset{H \to +\infty}{\longrightarrow} 0$ and $\lim\limits_{H \to +\infty} Var[\sum_{h \geq H} \theta_h \varepsilon_{t-h}] \underset{H \to +\infty}{\longrightarrow} 0$

h represents the number of lag.The mean and the variance tend towards a constant equal to 0 and do not depend on time, so the process is stationary.

3.1.3 Autoregressive moving average models

Finally, an ARMA process is a combination of AR and MA processes. An ARMA time series is therefore defined by having both lags on p and q orders.

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

We can also write the ARMA model using a backshift operator as follows:

$$\varphi(B) X_t = \theta(B) \varepsilon_t$$

According to the Wold theorem the MA part of the ARMA is stationary. The stationary conditions therefore depend on the AR part. Its characteristic polynomials are:

$$\varphi(B) = (1 - \varphi_1 B - \ldots - \varphi_p B^p)$$

and

$$\theta(B) = (\theta_0 + \theta_1 B - \ldots - \theta_q B^q)$$

If the roots of the polynomial $\varphi(B)$ are different from 1 then the process is stationary. If the roots of this polynomial $\varphi(B)$ are greater than 1, then the residuals are a process of innovations.

3.2 Maximum likelihood and generalized autocorrelation estimations

In order to estimate the Information Criteria, there are two possible estimations : the Maximum Likelihood Estimation (MLE) and the Generalized Autocorrelation Estimation (GAE). We will define both estimations methods in this part.

3.2.1 Likelihood estimation

The Maximum Likelihood Estimation is also a selection model criteria. We will in this section begin by introducing the Likelihood estimation, and then present its formula. In Bayesian inference, likelihood can be interpreted as the probability density of the data conditionally on a value of the stochastic parameters. The Maximum Likelihood Method is used to estimate the parameters of a regression model, under the assumption that the true distribution of the parameters is known. This can be a limit to its use because we do not always know this distribution.

The likelihood function describes the goodness of fit of the model considered. Shumway and Stoffer [Shumway and Stoffer, 2010] define this method as follows. We can describe it as follow with x the estimated variables, μ its mean, σ^2 its variance, $f(.)$ the density function and φ the AR coefficient.

$$L(\mu, \varphi, \sigma^2) = f(x_1)f(x_2|x_1)...f(x_n|x_{n-1})$$

For the sake of clarity, we will be presenting the likelihood function for an AR(1) that is defined as:

$$L(\mu, \varphi, \sigma^2) = (2\pi\sigma^2)^{-n/2}(1-\varphi^2)^{1/2}\exp[-\frac{S(\mu, \varphi)}{2\sigma^2}$$

$S(\mu, \varphi)$ also called the unconditional sum of squares and it is defined as:

$$S(\mu, \varphi) = (1-\varphi^2)(x_1-\mu)^2 + \sum_{t=2}^{n}[(x_t-\mu)-\varphi(x_{t-1}-\mu)]^2$$

It is used in the LE in order to find the estimated variance of the series considered:

$$\widehat{\sigma}^2 = n^{-1}S(\widehat{\mu}, \widehat{\varphi})$$

This estimation method has the goal to find the parameter with the higher probability of reproducing the true observations of the time series analyzed. One problem generally encountered with this estimate is the lack of convergence. This explains why we will be using both MLE and GAE estimations methods. Under the assumption that the stochastic terms are normally distributed, the GAE and MLE should have the same results. It therefore is interesting to test this assumption in our study.

3.2.2 Generalized autocorrelation method

The theory of the specification of ARMA models by the generalized autocorrelation method, using the Ordinary Least Squared estimation was developed by the econometrics research center of the University of Montreal [Dufour, 2002].

The first step is the estimation of the X_t process by the OLS:

$$X_t = \sum_{i=1}^{p}\varphi_{i(p)}X_{t-i} + \varepsilon_{p,t}, t = p+1, ..., n$$

The coefficients φ ranging from order 1 to order p are estimated. The estimated coefficients are $\widehat{\varphi}$. ε_t correspond to the residual of the regression.

If the ARMA process is non-stationary then the OLS estimators are convergent:

$$\widehat{\varphi_{i(p)}} \underset{n\to+\infty}{\longrightarrow} \varphi_i, i = 1,..,p$$

If the ARMA process is stationary then the OLS the estimators are not convergent:

$$\widehat{\varepsilon_{p,t}} = X_t - \sum_{i=1}^{p} \widehat{\varphi_{i(p)}} X_{t-i}, t = p+1,...,n$$

$\widehat{\varepsilon}$ corresponds to the estimated residual. This residual is not a white noise and their past values are used to predict X_t.

Secondly, we estimate the regression taking the past value of the residue $\widehat{\varepsilon_{t-1}}$:

$$X_t = \sum_{i=1}^{p} \widehat{\varphi_{i(p)}} X_{t-i} + \beta_{i(p)} \widehat{\varepsilon_{p,t-1}} + \varepsilon_{p,t}, t = p+2,...,n$$

If the process X_t is non-stationary then the OLS estimators are convergent:

$$\widehat{\varphi_{i(p)}} \underset{n\to+\infty}{\varphi}_i, i = 1,...,p$$

If the process X_t is stationary then the OLS the estimators are not convergent:

$$\widehat{\varepsilon_{p,t}} = X_t - \sum_{i=1}^{p} \widehat{\varphi_{i(p)}} X_{t-i} + \widehat{\beta_{i(p)}} \widehat{\varepsilon_{p,t-1}} + \varepsilon_{p,t}, t = p+2,...,n$$

$\widehat{\varepsilon}$ corresponds to the estimated residual. These residuals are not white noise and their past values are used to predict X_t.

These steps are reiterated until we obtain convergent estimators of: $\varphi_1,...,\varphi_p$.

3.3 Definition of the study

In this study we are interested in ARMA model selection by information criteria. It will be based on 10 000 realizations of 12 different ARMA with different sample sizes. This allows to not only study when the information criteria have selected the right model, but also to see how many times it overfits (possible for $k < 4$), and underfits (possible for $k > 1$). We will, in this part, be presenting the Information Criteria selected, the Monte Carlo study we will be doing, and the performance analysis method we will be using.

We note that k represent the total number of orders combined both in the AR(p) process, MA(q) process and ARMA(p,q). The number n is the size of the time series.

3.3.1 Information criteria selected

The sample size having a great impact on overfitting and underfitting probabilities, we will be having both small and large sample sizes. This allows to study not only the accuracy of the information criteria, but also the risk associated with each criterion, whether it tends to overfit or underfit.

We will be looking at 8 information criteria in this study. In these criteria some are expected to do very well in large samples but the perform poorly in small samples because they tend to overfit. This is the case of AIC, BIC, SIC, HQ and FPE. Some, on the contrary are supposed to compensate this lack of adaptation to small sample sizes with a penalty function more restrictive. This is the case of AICc and SICc, but they also tend to underfit when the sample size increases. Finally, we will be looking at the AICu, created to balance underfitting and overfitting while improving the accuracy.

Let σ_k^2 be the variance, k the number of orders, n the sample size and s_k^2 the unbiased variance. The information criteria we will used are defined as:

$$AIC = log(\widehat{\sigma_k^2}) + \frac{2(k+1)}{n}$$

$$BIC = nlog(\widehat{\sigma_k^2}) + klog(n)$$

$$SIC = log(\widehat{\sigma_k^2} + \frac{klog(n)}{n}$$

$$HQ = log(\widehat{\sigma_k^2} + \frac{2klog(log(n))}{n}$$

$$FPE = \widehat{\sigma_k^2} + \frac{n+k}{n-k}$$

$$AICc = log(\widehat{\sigma_k^2} + \frac{n+k}{n-k-2}$$

$$SICc = log(\widehat{\sigma_k^2} + \frac{klog(n)}{n-k-2}$$

$$AICu = log(\widehat{s_k^2} + \frac{n+k}{n-k-2}$$

3.3.2 Monte Carlo simulation

The Monte Carlo simulation procedure involves four sub-routines. We perform these steps for sample sizes 15, 50, 100, and 150.

The first sub-routine consists to generate the coefficients of different ARMA. We generate the coefficients φ_i ranging from 0 to p and the coefficients θ_i ranging from 0 to q from a random distribution in the range $[0;1]$. This makes it possible to satisfy the following condition:

$$\sum_{i=1}^{p} \varphi_i < 1$$

This condition makes it possible to obtain stationary ARMA.

The second sub-routine estimates the number of order p and q. The ARMA used are: AR(1), AR(2), AR(3), AR(4), MA(1), MA(2), MA(3), MA(4), ARMA(1,2) ARMA(2,1) and ARMA(2,2). We estimate these ARMA both using GAE and MLE methods as described before.

The third subroutine is to use each selection criterion to determine the estimated order $\widehat{p}$ and $\widehat{q}$. We do this for the well-specified model and mis-specified models created on our simulated series.

The last subroutine repeats step one to three 10 000 times, in order to count the number of times Information Criteria finds the right process, when it underfits and when it overfits.

3.3.3 The probability of well performance

The performance calculates the number of times the minimized criterion finds the correct order of autoregressive processes. It finds the order p if it is an AR, the order q if it is an MA and the order k $(p+q)$ if it is a ARMA. If the correct order is found, we count 1. If the order is not found, we count 0.

We denote by P_c the percentage that the criterion finds the right order such that:

$$P_c = \frac{\text{number of time ''picks up'' occured}}{10000} * 100$$

When $Pc = 0\%$ then the criterion is never efficient. When $Pc = 100\%$, the criterion performs perfectly.

3.3.4 Right orders, Overfitting and Underfitting

An interesting thing to look at is not only how many times Information Criteria did select the right process. We also want to know how many times they found the right order but not the right process. If the sum of p and q selected by the criterion is the right number, but is a different process than the one simulated, we count 1. The percentage that the event “right k but wrong process” happened was calculated as:

$$P_k = \frac{\text{number of times the right number of k but wrong process had been found}}{10000} * 100$$

When the selection criterion is not efficient two cases are possible: either it overestimates the number of orders (overfitting) or it underestimates the number of orders (underfitting). If the sum of p and q selected by the criterion is greater than the sum p and q known then there is overfitting. The percentage that the event ”overffiting” happened was calculated as:

$$P_o = \frac{\text{number of time overfitting occured}}{10000} * 100$$

If the sum of p and q selected by the criterion is less than the sum p and q known then there is underfitting. The percentage that the event ”underffitng” happened was calculated as:

$$P_u = \frac{numberoftimeunderfittingoccured}{10000} * 100$$

4 Results

4.1 The Maximum Likelihood Estimation results

Now that we have our results, an important conclusion can be drawn. The Maximum Likelihood Estimation did not converge, whatever the information criteria, the different sample sizes and the different processes. This result is surprising because MLE is extremely common, and we have never found such a result in the literature. However, the articles on the subject never vary that much parameters. Indeed, we simulate our series 10 000 times, with different coefficients at each new simulation. We also analyzed the behavior of different processes. Finally, as our study also treats with different sizes, we also simulated our series on different sample sizes. All these variations can explain the convergence limit of the maximum likelihood. Thus, we cannot conclude on the MLE, except the fact that this method has convergence problems. Therefore, we will analyze the results obtained by the Generalized Autocorrelation Estimation method, which always converged.

4.2 The Generalized Autocorrelation Estimation results

The analysis of the results by the GAE method is presented in this part by tables and graphs. First, the tables, which are in the appendix (p. 31), give the percentage of success (Tables 1 to 13), the percentage of good k found (Tables 14 to 26), the percentage of underfitting (Tables 27 to 37) and the percentage of overfitting (Tables 38 to 47) for each criterion and for each sample size. Second, the graphs represent the correct estimation of the orders for each criterion and for each sample. One the one hand, when the color on the graph is red then it means that the criterion is performing well. On the other hand, when the color is blue it means that the criterion is not performing. The more performing (non-performing) the criteria, the darker the red (blue) color is. White color is the transition between performance and non-performance. The black part presents the none estimated ARMA.

4.2.1 The well-performances

The well-performance of the criteria is described in this part for each process according to the different sample sizes. At the beginning, we will evaluate it within finite sample sizes (15 and 50) then we will study the asymptotic sizes (100 and 150). The ultimate goal is to obtain a comparison between the performance of the different criteria.

First, we observe that for a sample size equal to 15 the criteria all have the same performance for an AR(1) at 58.31%. Thus, the color for this process is red for all the criteria (Figure 1). Then, none of the criteria finds the right process for the AR(2), the AR(3), the AR(4), the MA(1), the MA(4), the ARMA(1,1), the ARMA(1,2) and the ARMA(2,2). For the Moving Average process of orders 2 and 3, the AICc, the SICc, the AICu and the FPE do not perform. For the MA(3), the HQ and the SICc never find the right orders. The BIC and the SIC are the best performers for this process but their percentages remain low (0.79%). The AIC, the BIC, the SIC and the HQ perform at over 58% for the ARMA(2,1).The criteria corrected by penalty (SICc, AICc and AICu), built to perform better in small samples, do not perform.

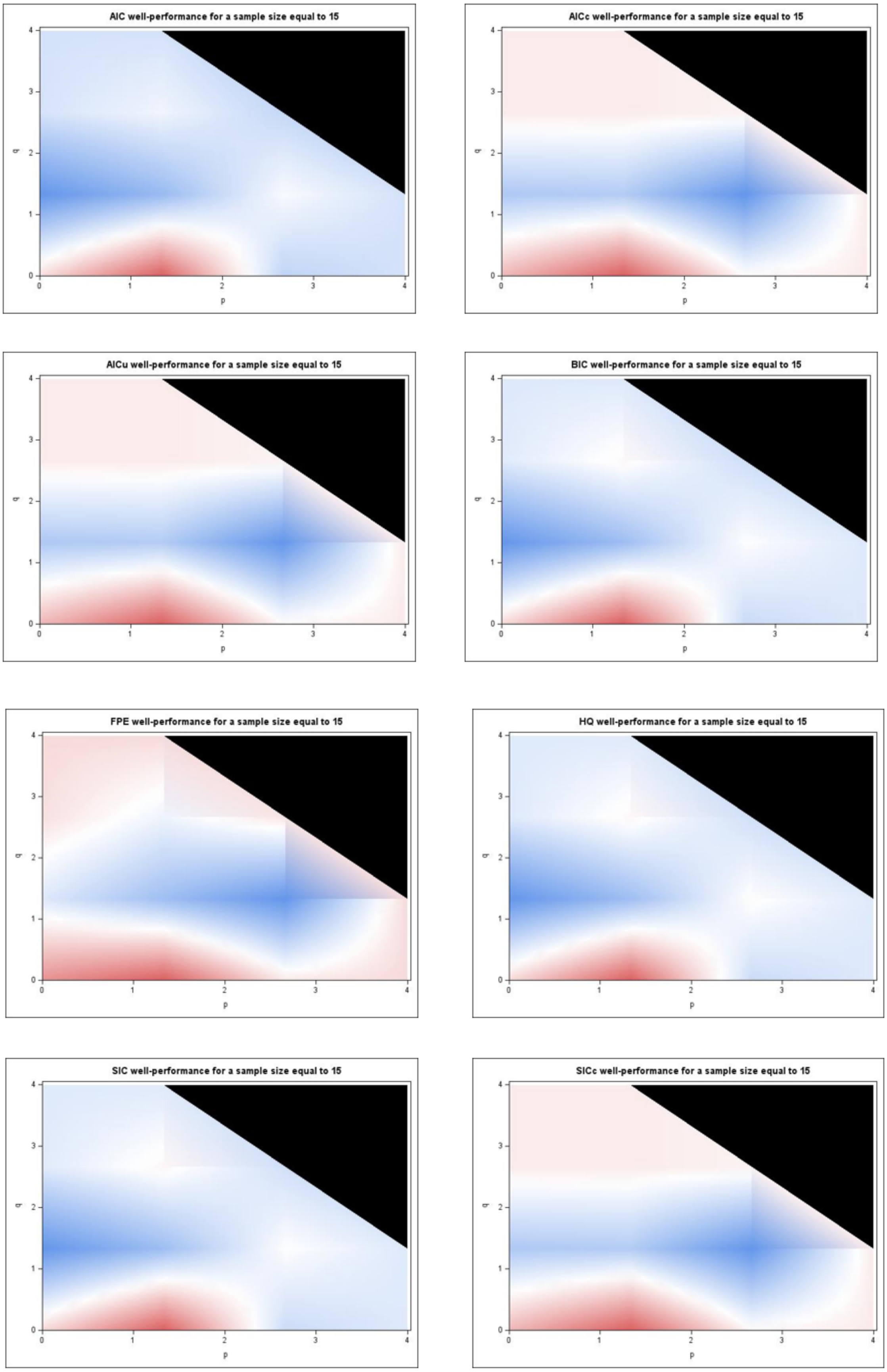

Figure 1: The well-performance of Information Criteria for a sample size equal to 15

Next, we analyze the correct specification of the orders for a sample size $n = 50$. For the AR(1), the criteria are still effective. The BIC, the SIC, the AICc, the SICc, the AICu and the FPE performed at 60.38%. The performance of HQ and AIC is lower. Also, for the AR(2), the AR(3), the ARMA(1,1) and the ARMA(1,2), no criterion is effective. For these processes, the color is either blue (meaning that the criterion is inefficient) or white (marking the transition between efficiency and inefficiency) (Figure 2). For the AR(4), the AIC performed very well at 98.95% followed by the HQ at 83.81%, the BIC and the SIC at 43.65% and the AICc and the SICc at almost 20%. The AIC and the HQ do not perform for the MA(1). For this process, the criteria that most often find the right process are the AICu and the FPE at 77.21%. They are followed by the BIC, the SIC, the AICc and the SICc which have a performance between 40% and 55%. For the MA(2), the best performing criteria are the BIC, the SIC, the AICc and the SICc. They find the correct model in 20% to 30% of all simulations. The HQ and the AIC have poorer performances. For the MA(3), all the criteria are either inefficient or have a weak efficency. All criteria (except the AICu and the FPE) have a performance between 33% and 43% for the process ARMA(2,1). Finally, for the ARMA(2,2) only the HQ and the AIC finds the correct model in approximately 25% of cases, the BIC and the SIC in 14% of cases and the AICc in 8% of cases. Overall, we observe that AICu and the FPE are efficient only when k=1. The BIC, the SIC, the SICc and the AICc behave similarly regardless of the process. Increasing the sample size generally improves the quality of the criteria estimation. Also, increasing the number of orders has no impact on the efficiency of the criteria.

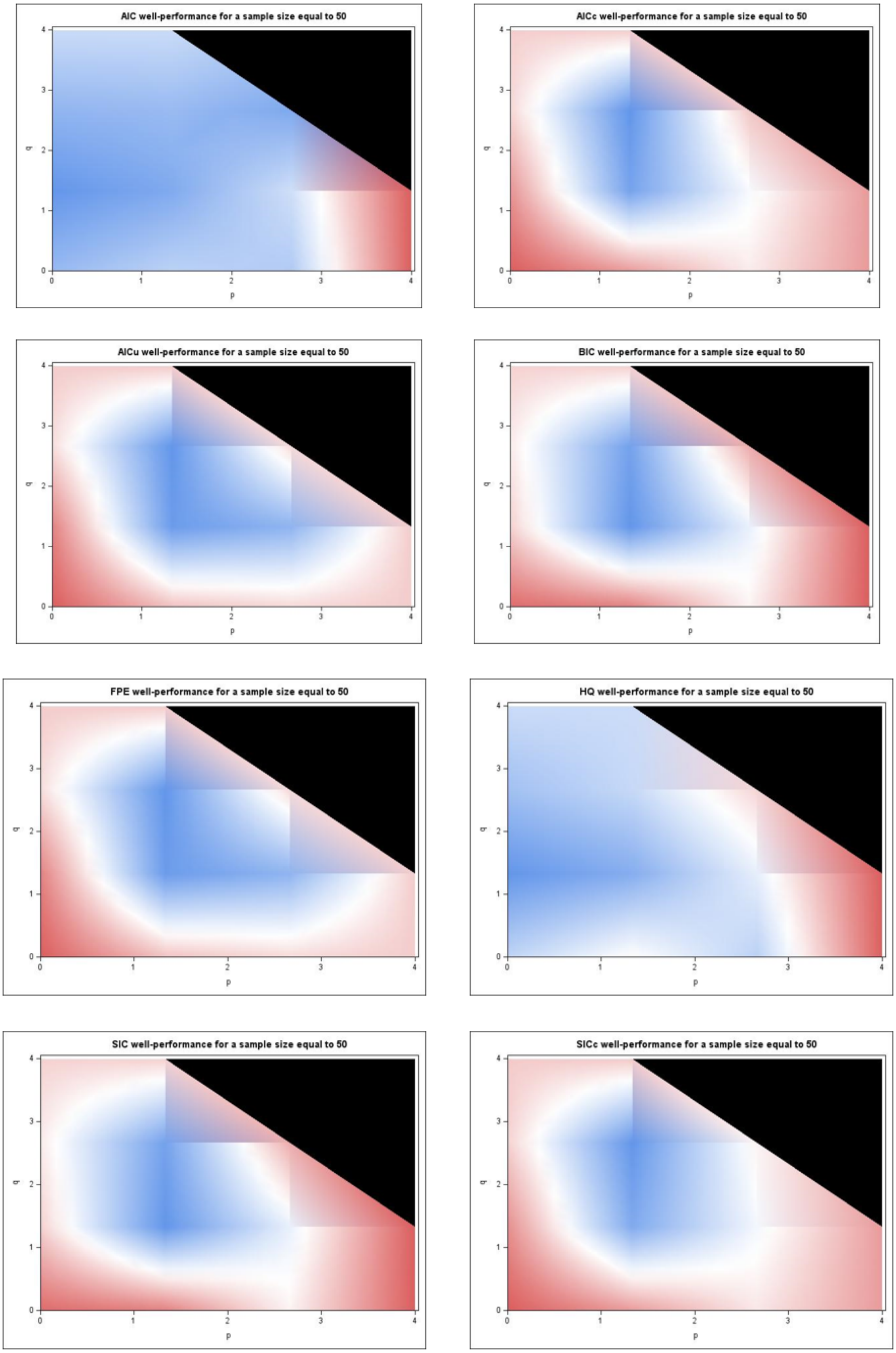

Figure 2: The well-performance of Information Criteria for a sample size equal to 50

In this paragraph, we consider the performance of the criteria when the sample size is asymptotic ($n = 100$). All of the criteria with the exception of the HQ and the AIC perform perfectly in the analysis of an AR(1). In this case, we observe dark red color for all these criteria (Figure 3). The SICc finds the correct AR(2) process in 7% of all simulations followed by the SIC and the BIC in 1% of cases. The performance of the criteria for this process is therefore very low. Also, the criteria never find the right process for the AR(3), the ARMA(1,1) and the ARMA(1,2). The AIC is the criterion that performs best for an AR(4). Its performance is 91.63% against 55.11% for the HQ and the AICc and between 13% and 19% for the BIC, the SIC and the SICc. For the MA(1), only the FPE and the AICu are efficient. The performance of the BIC, the SIC, the AICc and the SICc is approximately 30% and the performance of the HQ is approximately 10% for an MA(2). The MA(3) process is found in about 30% of cases by the BIC, the SIC, the AIC, the HQ, the AICc and the SICc. For the MA(4), the AIC is the one who performs the best with a success rate equal to 80.22%. It is followed by the HQ (74%), the AICc (62%), the BIC and the SIC (59%), and the SICc (55%). For an ARMA(2,1), performances are between 20% and 28% for all criteria (except AICu and FPE). The performance of the criteria when the ARMA has a $p = 2$ and a $q = 2$ is lower: it is 15% for the HQ, 19% for the AIC less than 10% for the BIC, the SIC, the AICc and SICc and zero for the FPE and the AICu. In summary, the AICu and the FPE perform only for an order equal to 1. The BIC and the SIC always behave similarly. Going from a finite sample to an asymptotic sample improves the efficiency of the criteria.

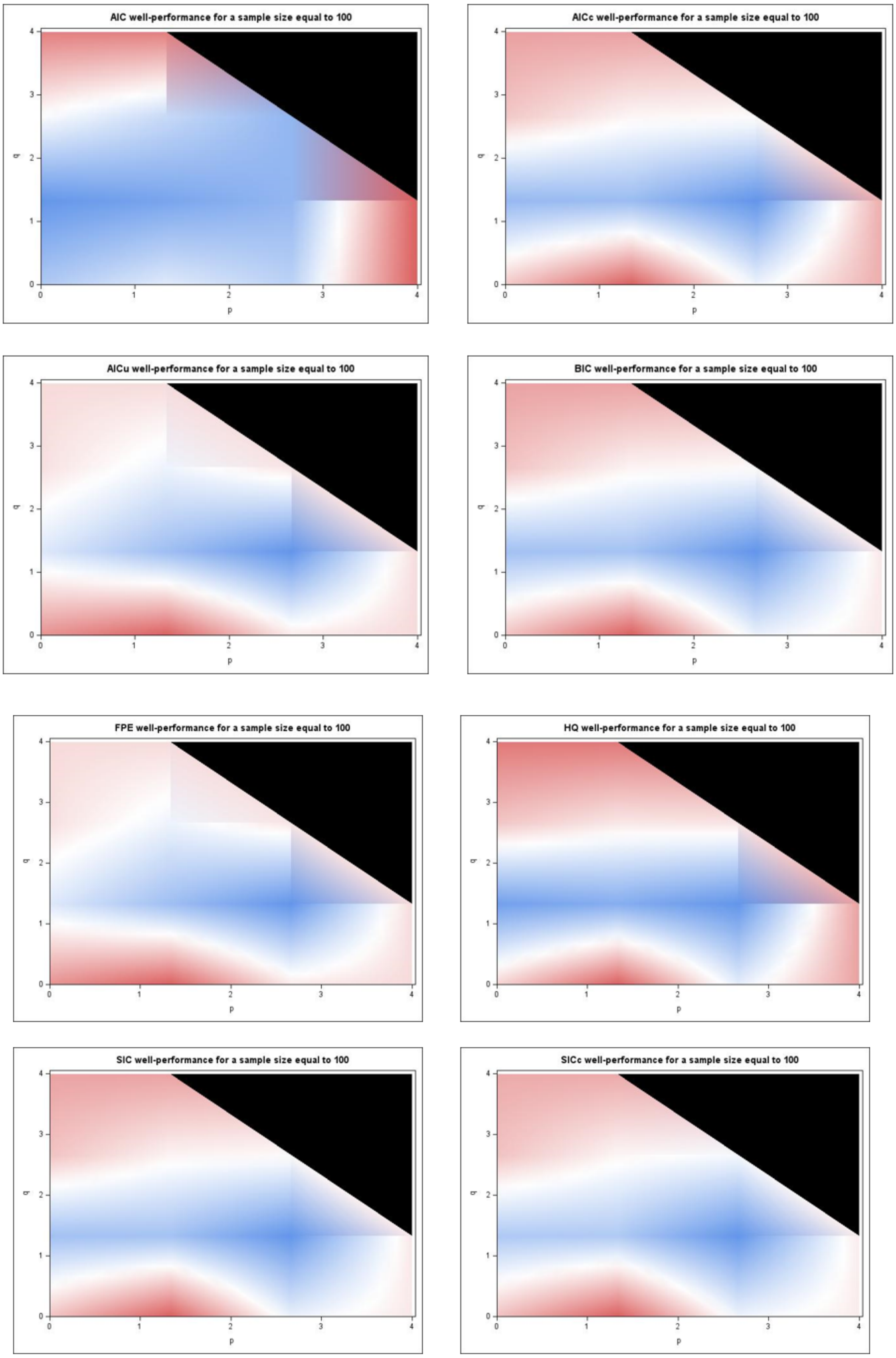

Figure 3: The well-performance of Information Criteria for a sample size equal to 100

For an even larger sample size ($n = 150$), right performances are all the more important. Indeed, when we analyze the AR(1), we find that all the criteria perform at 100% with the exception of the AIC which performs at 51.17%. For an AR(2), the AIC, the AICu and the FPE are not efficient, however, the other IC have a performance ranging from 32% to 54%. All the performance are zero for an AR(3), an ARMA(1,1) and an ARMA(1,2). For these processes, the color is either blue or white (Figure 4). The AIC performs very well for an AR(4) at 80.99%, followed by the HQ (57.88%) and the AICc (43.39%). For an MA(1), it is still the AICu and the FPE criteria that perform best, followed by the SICc, the BIC and the SIC. The BIC, the SIC, the SICc and the AICc have a performance of about 50% followed by the HQ and the AIC for an MA(2). The number of times the criteria find the correct MA(3) is very low for all the criteria. The MA(4) and the ARMA(2,1) are respectively found by all the criteria (with the exception of AICu and FPE) in about 60% and 20% of the simulations. The AIC is the criterion that performs best for a known ARMA(2,2), however, the overall percentages are low.

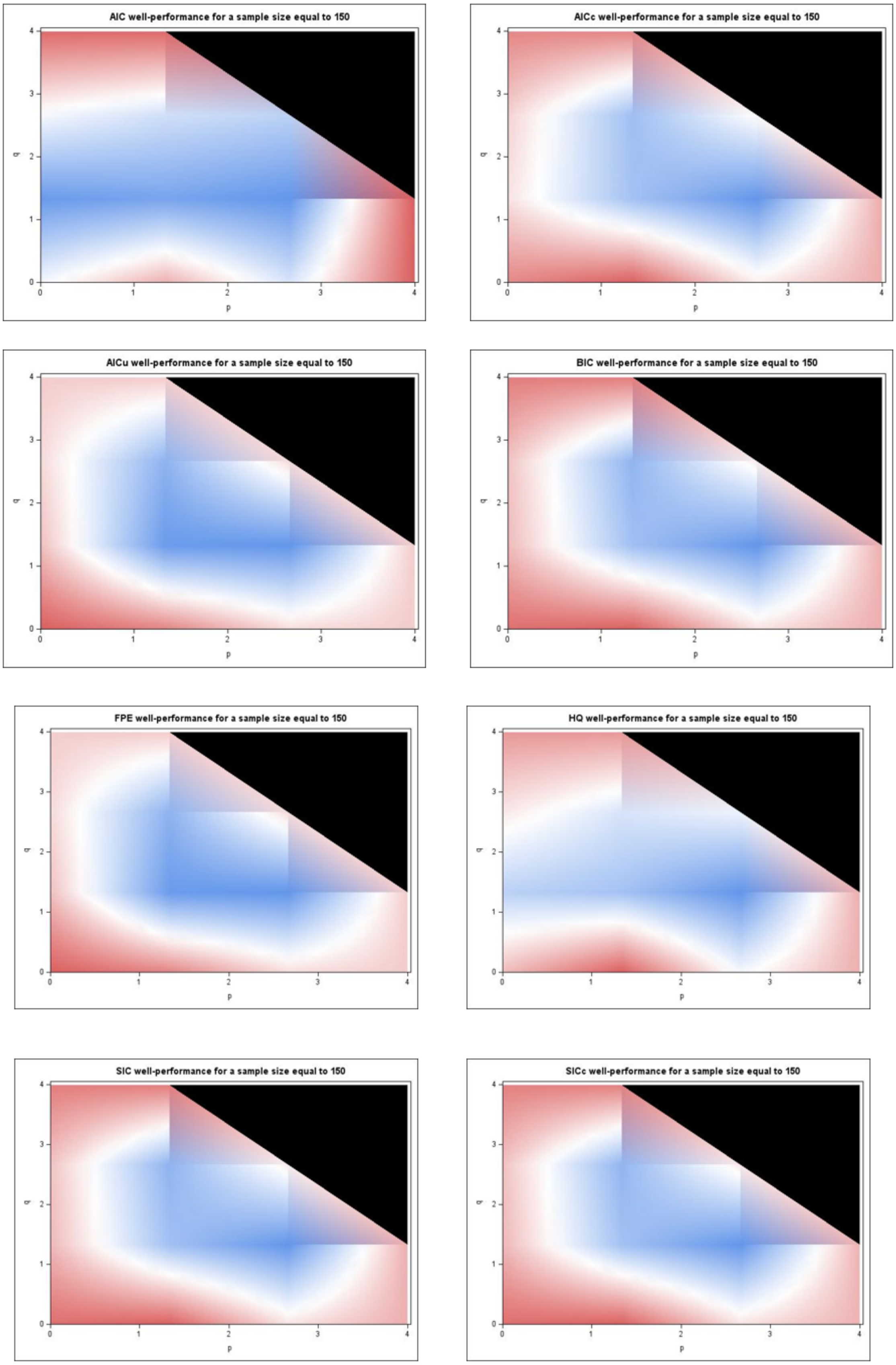

Figure 4: The well-performance of Information Criteria for a sample size equal to 150

In conclusion, we find that the performance of the criteria is dependent on the size of the sample: the larger it is, the better the precision is. Second, performances of IC depend on the orders. By way of example, whatever the size of the sample is, the FPE and the AICu are the criteria which perform best for an order equal to 1. The AIC is the best criterion for evaluating the MA(4) and the ARMA(2,2). Overall, the more important the order are, the more difficult it is to obtain a right precision: the performances are more important for an AR(1) than for an AR(3), an ARMA(1,1) and an ARMA (1,2) which are never found. Third, we find that certain criteria have exactly the same performances : the BIC and the SIC, and, the FPE and the AICu work in pairs. Finally, the corrected criteria (AICc, SICc and AICu) created to obtain better performances in small sample does not excel when n is low.

4.2.2 The percentage of times when the criterion finds the right order but not the right process

In this part, we have calculated the average number where the criterion finds the right order per process (Table 26) and for all the processes (Tables 14 to 25). Overall, we observe that the average number of estimation of the right order rather than the right process is greater when $n = 15$ than when $n = 150$. This finding is not surprising because as the sample size is increased the number of right process estimates is greater. Then, we find that for an AR(1) regardless of the size of the sample, the right process or the right order is estimated. Also, the AIC is the criterion that finds the most right orders for all samples for an ARMA(2,2). We thus find an efficient AIC when orders are high. The corrected IC (AICc, SICc, AICu) do not detect the right k in small samples. For $n = 15$, the AIC, the BIC, the HQ find respectively the right k in 19.13%, 21.73% and 17.22% against 13.39%, 16.54% and 11.82% for the AICc, the SICc and the AICu on average.

4.2.3 Underfitting : The order determined by the criterion is lower than the known order

In this part, we have calculated the average number where the criterion finds an order lower than the k known by process (Table 37) and for all the processes (Tables 27 to 36). These calculations cannot be operated on the AR(1) and MA(1) because underfitting is not observable. We therefore consider $k > 1$.

First, the percentage of underfitting is greater when the number of observations is 15 than for a sample size 150. For sample $n = 15$, the HQ is the criterion that underfits the least. The AIC is the criterion with the least underfitting for samples 50, 100 and 150. In addition, the more the sample increases, the less this criterion presents underfitting. The BIC and the SIC criteria do not perform because they often underfit. As an example, the average percentage of underfitting of these two criteria is equal to 67.27% for a sample size equal to 15. The corrected criteria known to be efficient in a small sample are those that underfit the most on average. For $n = 15$, the averages for the AICc, the SICc and the AICu are 98.10%, 94.32% and 100%, respectively. Underfitting could therefore be at the origin of the inefficiency of the adjusted criteria when the sample is small. The FPE, having a similar behavior to the AICu, underfits in all cases when $k > 1$. We can therefore conclude that the AICu and the FPE only perform for an order equal to 1. Previously, we spotted that the performances of the criteria for an AR(3), an ARMA(1,1) and an ARMA(1,2) was zero. Indeed, the criteria present underfitting for these three processes. However, overfitting can also be the cause of this poor performances.

4.2.4 Overfitting : The order determined by the criterion is greater than the known order

In this part, we have calculated the average number where the criterion finds an order greater than the k known by process (Table 47) and for all the processes (Table 38 to 46). These calculations cannot be operated on the AR(4), the MA(4) and the ARMA(2,2) because overfitting is not noticeable. We therefore consider $k < 4$.

First of all, unlike underffiting, overfitting is not necessarily lower as the sample size increases. The criteria that present little underfitting (AIC and HQ) demonstrate overfitting. The HQ has an average overfitting of 44.24% for a sample size equal to 15. The AIC has an average of overfitting of 87.45%, 76.63% and 79.40% for sample sizes 50, 100 and 150. The BIC and the SIC also present overfitting, but to a lesser extent compared to underfitting. For $n = 15$, the corrected criteria never overfit. These criteria have been considered to eliminate overfitting in small samples. As mentioned previously, the inefficiency of these criteria is therefore mainly due to underfitting. However, we note that AICc and SICc overfit for sample sizes 50, 100 and 150. Finally, the overffiting of the criteria for the estimation of an ARMA(1,1) and an ARMA(1,2) is the source of the non-efficiency of the criteria in the same way as the underfitting. The conclusion is different if the known process is an AR(3): the cause of the ineffectiveness of the criteria is essentially the underfitting (with the exception of the AIC).

5 Conclusion

In this study, we analyzed the model selection by Information Criteria. Using two estimation methods, 12 different types of process, 4 sample sizes and 10 000 simulations for each process and sample size, we obtain interesting results. For our results, we used the Maximum Likelihood and Generalized Autocorrelation Estimations.

The literature helped realizing that all Information Criteria depend on process and sample sizes. Some are created to perform in small samples while other are created to avoid overfitting. The Generalized Autocorrelation Method was not used in these papers, while the Maximum Likelihood method seems to be the reference method. In order to compare these methods, we will be using both of them in this study.

First of all, the Maximum Likelihood Method does not bring any interpretable results. This result is surprising because Maximum Likelihood Estimation is extremely common. However, articles on the subject never vary so many parameters. Also, we analyzed more processes than the articles of interest. Finally, we also simulated our series on different sample sizes. All these variations can explain the limit of convergence of the maximum likelihood. Thus, we cannot conclude on the maximum likelihood estimation, except that this method presents convergence problems.

The use of the Generalized Autocorrelation Method allowed this study to obtain robust results because it allowed convergence even when varying the coefficients. It is also difficult to do a conclusion comparing the literature, that used the Maximum Likelihood Estimation, and our study, because of this lack of adaptation of the estimation method to our study. It will therefore not be possible to draw a conclusion on this point, by comparing the results obtained with the main results expected from the literature review.

When it comes to the sample size, we see that the more the sample size increases, the more the Information Criteria find the right process. We thought that it would have been more complicated to find ARMAs than ARs or MAs because of the difference in terms. However, their performance depends on the number of orders considered. For example, all Information Criteria almost systematically find processes for AR(1) or MA(1). However for k=3, we notice that Information Criteria almost never find the right process. Whether it is for an AR(3), MA(3) or even ARMA(1,2) and ARMA(2,1).

Regarding the results of the Information Criteria, we will start by presenting the results of the uncorrected criteria, then those of the corrected criteria.

A first thing to notice is that the uncorrected criteria (AIC, BIC, SIC, HQ and FPE) are the ones performing the best in this study. The AIC, BIC, SIC, and HQ have great performances in large samples, and perform also quite effectively in small samples, a conclusion not expected from the literature. Those great performances are in comparison to the corrected criteria, and for processes where the right order is found, because as said before, some orders are never found.

When the AIC and HQ do not perform, this is due to their overfitting probabilities. When it comes to SIC and BIC, it is most likely due to their underfitting probabilities, and their probability of overfitting to a lesser extent. Also, the BIC and SIC have the exact same results for all the simulations, processes and samples sizes even if their formula differ. Looking at the FPE and AICu a strong result appear :

they select the models in a completely identical way. Yet the AICu is supposed to be made to avoid overfitting, and the FPE being at the origin of the AIC, we could have expected it to have similar results to those of the AIC.

When it comes to the corrected criteria (AICu, AICc, SICc), it is expected that they perform better in small samples, as their penalty functions are created to this purpose. In this study, this is not the case. In fact, their performances grow with the sample size. Also, they tend to find the right number of orders more often than good processes in small samples. The SICc results are also very close to SIC's and BIC's results when the sample size increases. Finally, when corrected criteria do not find the right order, they underfit. This conclusion is even more common in small samples. This could mean that their penalty function is too strong and restrictive.

The results obtained by this study are interesting when it comes to Information Criteria selection. However, it could be interesting to modify the type of simulation. Indeed, in order to have results comparable to the literature, not varying the coefficients of the simulations could allow to have a more important convergence. Thus, we could see if the penalty functions of the corrected criteria remain as important and limiting as in the presented study. By not varying the coefficients, we could vary the sample size. This would tell us what the asymptotic threshold really is, and if it changes from one criterion to another.

Bibliography

[Akaike, 1969] Akaike, H. (1969). Fitting autoregressive models for prediction. (21):243 – 247.

[Akaike, 1973] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. pages 267 – 281.

[Akaike, 1979] Akaike, H. (1979). A bayesian extension of the minimum AIC procedure of autoregressive model fitting. 66:237 – 242.

[Dufour, 2002] Dufour, J. (2002). Spécification de modèles ARIMA par la méthode des autocorrélations généralisées.

[Emiliano et al., 2014] Emiliano, P., Vivanco, M., and Menezes, F. (2014). Information criteria: How do they behave in different models? Volume 69:Pages 141–153.

[Geweke and Messe, 1981] Geweke, J. and Messe, R. (1981). Estimating regression models of finite but unknown order. 16(1):162.

[Hannan and Quinn, 1978] Hannan, E. J. and Quinn, B. G. (1978). The determination of the order of an autoregression. (41):190 – 195.

[Hurvich and Tsai, 1989] Hurvich, M. and Tsai, C. (1989). Regression and time series model selection in small samples. 76-2:292–307.

[Jie et al., 2016] Jie, D., Vahid, T., and Yuhong, Y. (2016). Bridging AIC and BIC: a new criterion for autoregression.

[Liew, 2000] Liew, V. K.-S. (2000). The performance of AICC as order determination criterion in the selection of ARMA time series models.

[Liew, 2004] Liew, V. K.-S. (2004). On autoregressive order selection criteria.

[McQuarrie, 1998] McQuarrie, A. (1998). A small-sample correction for the schwarz SIC model selection criterion.

[McQuarrie and Tsai, 1998] McQuarrie, A. and Tsai, C. (1998). *Regression and Time Series Model Selection - McQuarrie & Tsai, 1998*. World Economic.

[Parzen, 1977] Parzen, E. (1977). Multiple time series modeling: determining the order of approximating autoregressive schemes. pages 283–925.

[Schwarz, 1978] Schwarz, G. (1978). Estimating the dimension of a model. 6(2):461–464.

[Shumway and Stoffer, 2010] Shumway, R. and Stoffer, D. (2010). Time series analysis and its applications : With r examples.

Appendices

Table 1 : Percentage of well-performance for AR(1)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	58.31%	58.31%	58.31%	58.31%	58.31%	58.31%	58.31%	58.31%
50	17.63%	60.38%	32.36%	60.38%	60.38%	60.38%	60.38%	60.38%
100	25.13%	100%	86.00%	100%	100%	100%	100%	100%
150	51.17%	100%	100%	100%	100%	100%	100%	100%

Table 2 : Percentage of well-performance for AR(2)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0%	0%	0%	0%	0%	0%	0%	0%
50	0%	0%	0%	0%	0%	0%	0%	0%
100	0%	1%	0%	1%	0%	7.19%	0%	0%
150	0%	49%	33%	49%	53%	48%	0%	0%

Table 3 : Percentage of well-performance for AR(3)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0%	0%	0%	0%	0%	0%	0%	0%
50	0%	0%	0%	0%	0%	0%	0%	0%
100	0%	0%	0%	0%	0%	0%	0%	0%
150	0%	0%	0%	0%	0%	0%	0%	0%

Table 4 : Percentage of well-performance for AR(4)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0%	0%	0%	0%	0%	0%	0%	0%
50	98.95%	43.65%	83.81%	43.65%	25.19%	24.15%	0%	0%
100	91.63%	18.54%	55.11%	18.54%	55.11%	13.34%	0%	0%
150	80.99%	31.52%	57.88%	31.52%	43.39%	28.09%	0%	0%

Table 5 : Percentage of well-performance for MA(1)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0%	0%	0%	0%	0%	0%	0%	0%
50	0%	40.18%	0%	40.18%	53.59%	54.18%	77.21%	77.21%
100	0%	0%	0%	0%	0%	0%	26.95%	26.95%
150	0%	65.38%	15.02%	65.38%	53.09%	66.15%	78.79%	78.79%

Table 6 : Percentage of well-performance for MA(2)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	9.35%	0.95%	0.17%	0.95%	0%	0%	0%	0%
50	0.94%	23.54%	10.51%	23.54%	28.66%	29.96%	0%	0%
100	0.16%	30.56%	10.41%	30.56%	28.77%	32.50%	0%	0%
150	8.90%	49.93%	32.15%	49.93%	45.80%	54.43%	0%	0%

Table 7 : Percentage of well-performance for MA(3)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0.09%	0.79%	0%	0.79%	0%	0%	0%	0%
50	0%	5.07%	0%	5.07%	8.27%	8.68%	0%	0%
100	24.64%	31.12%	27.70%	31.12%	30.52%	32.03%	0%	0%
150	7.46%	11.95%	9.32%	11.95%	10.67%	13.12%	0%	0%

Table 8 : Percentage of well-performance for MA(4)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0%	0%	0%	0%	0%	0%	0%	0%
50	25.03%	10.99%	17.25%	10.99%	7.93%	7.65%	0%	0%
100	80.22%	59.66%	73.99%	59.66%	62.65%	55.63%	0%	0%
150	76.89%	62.24%	72.38%	62.24%	67.51%	60.08%	0%	0%

Table 9 : Percentage of well-performance for ARMA(1,1)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0%	0%	0%	0%	0%	0%	0%	0%
50	0%	0%	0%	0%	0%	0%	0%	0%
100	0%	0%	0%	0%	0%	0%	0%	0%
150	0%	0%	0%	0%	0%	0%	0%	0%

Table 10 : Percentage of well-performance for ARMA(1,2)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0%	0%	0%	0%	0%	0%	0%	0%
50	0%	0%	0%	0%	0%	0%	0%	0%
100	0%	0%	0%	0%	0%	0%	0%	0%
150	0%	0%	0%	0%	0%	0%	0%	0%

Table 11 : Percentage of well-performance for ARMA(2,1)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	59.76%	58.87%	59.76%	58.87%	0%	0%	0%	0%
50	33.07%	42.34%	37.65%	42.34%	42.32%	42.53%	0%	0%
100	21.07%	22.28%	27.46%	22.28%	26.38%	19.44%	0%	0%
150	20.51%	16.82%	29.91%	16.82%	23.33%	15.46%	0%	0%

Table 12 : Percentage of well-performance for ARMA(2,2)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0%	0%	0%	0%	0%	0%	0%	0%
50	24.68%	14.10%	25.45%	14.10%	8.49%	0%	0%	0%
100	19.04%	6.40%	14.79%	6.40%	8.49%	0%	0%	0%
150	19.04%	0.04%	1.32%	0.04%	0.15%	1.10%	0%	0%

Table 13 : Percentage of well-performance for all processes by sample size								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	10.63%	9.91%	9.85%	9.91%	4.86%	4.86%	4.86%	4.86%
50	16.69%	20.02%	17.25%	20.02%	19.57%	18.96%	11.47%	11.47%
100	21.82%	22.44%	24.62%	22.44%	25.93%	21.69%	10.58%	10.58%
150	21.13%	32.24%	29.22%	32.43%	33.08%	32.23%	14.90%	14.90%

Table 14 : Percentage of times the right k is found for an AR(1)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	41.69%	41.69%	41.69%	41.69%	41.69%	41.69%	41.69%	41.69%
50	0%	39.62%	0%	39.62%	39.62%	39.62%	39.62%	39.62%
100	0%	0%	0%	0%	0%	0%	0%	0%
150	0%	0%	0%	0%	0%	0%	0%	0%

Table 15 : Percentage of times the right k is found for an AR(2)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	27.35%	21.15%	21.52%	21.15%	0%	0.65%	0%	0%
50	0%	0%	0%	0%	0.17%	0.18%	0%	0%
100	0%	0.65%	0%	0.65%	0.05%	1.27%	0%	0%
150	0%	1.40%	0.64%	1.40%	1.60%	1.36%	0%	0%

Table 16 : Percentage of times the right k is found for an AR(3)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0%	0%	0%	0%	0%	0%	0%	0%
50	0%	100%	5.65%	100%	0.71%	0%	0%	0%
100	0%	0%	1.62%	0%	0.1%	0%	0%	0%
150	0%	100%	4.12%	100%	3.06%	1.58%	0%	0%

Table 17 : Percentage of times the right k is found for an AR(4)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0%	0%	0%	0%	0%	0%	0%	0%
50	0.59%	0%	0.08%	0%	0%	0%	0%	0%
100	0.05%	0%	0%	0%	0%	0%	0%	0%
150	17.58%	0.38%	7.38%	0.38%	2.02%	0.17%	0%	0%

Table 18 : Percentage of times the right k is found for an MA(1)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	9.43%	49.16%	0%	49.16%	100%	100%	100%	100%
50	0%	17.40%	0%	17.40%	18.00%	17.77%	22.79%	22.79%
100	0%	85.82%	0%	85.82%	73.69%	83.06%	73.05%	73.05%
150	0%	34.60%	17.44%	34.60%	34.72%	33.85%	21.21%	21.21%

Table 19 : Percentage of times the right k is found for an MA(2)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	11.64%	16.25%	0.41%	16.25%	18.97%	34.24%	0%	0%
50	4.14%	21.58%	12.36%	21.58%	22.09%	22.24%	0%	0%
100	0.45%	16.76%	11.28%	16.76%	17.43%	16.61%	0%	0%
150	4.31%	17.03%	11.72%	17.83%	15.22%	18.35%	0%	0%

Table 20 : Percentage of times the right k is found for an MA(3)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	52.16%	52.02%	47.41%	52.03%	0%	16.81%	0%	0%
50	1.95%	4.56%	3.16%	4.56%	5.41%	5.54%	0%	0%
100	0.98%	0.86%	0.92%	0.86%	0.88%	0.88%	0%	0%
150	4.84%	6.32%	5.84%	6.32%	6.27%	6.68%	0%	0%

Table 21 : Percentage of times the right k is found for an MA(4)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	44.17%	39.47%	50.59%	39.47%	0%	0.41%	0%	0%
50	17.53%	12.99%	15.78%	12.99%	11.13%	10.99%	0%	0%
100	3.36%	1.21%	15.78%	1.21%	1.35%	1.06%	0%	0%
150	2.73%	0.84%	1.71%	0.84%	1.21%	0.72%	0%	0%

Table 22 : Percentage of times the right k is found for an ARMA(1,1)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0%	0%	0%	0%	0%	0%	0%	0%
50	0%	0%	0%	0%	0%	0%	0%	0%
100	0%	0%	0%	0%	0%	0%	0%	0%
150	0%	0%	0%	0%	0%	0%	0%	0%

Table 23 : Percentage of times the right k is found for an ARMA(1,2)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	2.87%	1.44%	3.58%	1.44%	0%	0%	0%	0%
50	8.17%	3.45%	7.56%	3.45%	2.14%	2.11%	0%	0%
100	16.83%	5.97%	11.73%	5.97%	6.88%	5.28%	0%	0%
150	22.69%	12.68%	19.66%	12.68%	15.12%	12.09%	0%	0%

Table 24 : Percentage of times the right k is found for an ARMA(2,1)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	40.24%	39.53%	40.24%	39.53%	0%	4.70%	0%	0%
50	2.50%	10.31%	5.41%	10.31%	13.10%	13.46%	0%	0%
100	11.22%	28.36%	19.32%	28.36%	27.03%	30.82%	0%	0%
150	0%	18.51%	12.33%	18.51%	16.52%	19.74%	0%	0%

Table 25 : Percentage of times the right k is found for an ARMA(2,2)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0.1%	0%	1.21%	0%	0%	0%	0%	0%
50	4.86%	0.20%	1.26%	0.20%	0%	0%	0%	0%
100	36.78%	1.21%	14.63%	1.21%	2.45%	4.98%	0%	0%
150	39.18%	2.30%	22.13%	2.30%	8.84%	0.03%	0%	0%

Table 26 : Percentage of times the right k but wrong process for all processes by sample size								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	19.13%	21.73%	17.22%	21.73%	13.39%	16.54%	11.82%	11.82%
50	5.81%	17.51%	4.27%	17.51%	9.47%	10.03%	5.20%	5.20%
100	5.81%	11.74%	5.16%	11.74%	10.82%	12.00%	6.09%	6.09%
150	7.61%	16.17%	8.58%	16.17%	8.72%	7.88%	1.77%	1.77%

Table 27 : Percentage of underfitting for an AR(2)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	44.00%	62.00%	38.00%	62.00%	100%	99%	100%	100%
50	0%	13.00%	0%	13.00%	37.00%	24.00%	100%	100%
100	0%	58.00%	23.29%	58.00%	51.74%	62.37%	100%	100%
150	1.91%	50.00%	39.00%	50.00%	45.00%	50.00%	100%	100%

Table 28 : Percentage of underfitting for an AR(3)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	99.93%	100%	95.88%	100%	100%	100%	100%	100%
50	22.11%	0%	72.23%	0%	99.29%	100%	100%	100%
100	73.73%	100%	98.34%	100%	99.90%	100%	100%	100%
150	39.79%	0%	91.34%	0%	96.63%	98.42%	100%	100%

Table 29 : Percentage of underfitting for an AR(4)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	100%	100%	100%	100%	100%	100%	100%	100%
50	0.46%	56.35%	16.11%	56.35%	74.81%	75.85%	100%	100%
100	8.32%	81.46%	44.89%	81.46%	44.89%	86.66%	100%	100%
150	1.43%	68.10%	34.74%	68.10%	54.59%	71.74%	100%	100%

Table 30 : Percentage of underfitting for an MA(2)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	1.99%	36.68%	0%	36.68%	81.03%	65.76%	100%	100%
50	0%	19.28%	0%	19.28%	26.19%	25.77%	100%	100%
100	0%	11.64%	0.06%	11.64%	9.24%	12.73%	100%	100%
150	0%	10.57%	0.29%	10.57%	2.31%	6.98%	100%	100%

Table 31 : Percentage of underfitting for an MA(3)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	6.71%	13.84%	2.72%	13.84%	100%	83.19%	100%	100%
50	1.82%	11.21%	3.77%	11.21%	16.89%	17.06%	100%	100%
100	0.66%	3.77%	1.49%	3.77%	3.12%	4.45%	100%	100%
150	0.66%	7.05%	2.20%	7.05%	4.19%	4.69%	100%	100%

Table 32 : Percentage of underfitting for an MA(4)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	55.83%	60.53%	49.41%	60.53%	100%	99.59%	100%	100%
50	57.44%	76.02%	66.97%	76.02%	80.94%	81.36%	100%	100%
100	16.42%	39.13%	23.56%	39.13%	36.00%	43.31%	100%	100%
150	20.38%	36.92%	25.91%	36.92%	31.28%	39.20%	100%	100%

Table 33 : Percentage of underfitting for an ARMA(1,1)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0.38%	99.03%	0%	99.03%	100%	100%	100%	100%
50	0%	39.92%	0%	30.92%	71.67%	71.70%	100%	100%
100	0%	99.75%	6.60%	99.75%	96.15%	100%	100%	100%
150	0%	100%	74.45%	100%	98.89%	100%	100%	100%

Table 34 : Percentage of underfitting for an ARMA(1,2)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	97.13%	98.56%	96.42%	98.56%	100%	100%	100%	100%
50	20.57%	85.11%	48.57%	85.11%	94.88%	95.41%	100%	100%
100	16.83%	5.97%	11.73%	5.97%	6.88%	5.28%	0%	0%
150	22.69%	12.68%	19.66%	12.68%	15.12%	12.09%	0%	0%

Table 35 : Percentage of underfitting for an ARMA(2,1)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0%	1.60%	0%	1.60%	100%	95.30%	100%	100%
50	0%	1.56%	0%	1.56%	6.12%	6.12%	100%	100%
100	0%	14.34%	0%	14.34%	9.24%	18.31%	100%	100%
150	0%	37.71%	3.76%	37.71%	21.03%	39.08%	100%	100%

Table 36 : Percentage of underfitting for an ARMA(2,2)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	98.59%	100%	98.79%	100%	100%	100%	100%	100%
50	49.56%	85.69%	73.29%	85.69%	90.30%	91.49%	100%	100%
100	39.30%	92.39%	70.58%	92.39%	89.87%	94.90%	100%	100%
150	42.56%	97.66%	76.54%	97.66%	91.01%	98.87%	100%	100%

Table 37 : Percentage of times the Information Criteria underfitted (for k>1) for all processes by sample size								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	50.41%	67.27%	48.10%	67.27%	98.10%	94.32%	100%	100%
50	15.20%	38.81%	28.09%	38.81%	58.82%	58.88%	100%	100%
100	17.39%	59.27%	34.54%	59.27%	53.12%	61.69%	100%	100%
150	12.99%	49.72%	42.35%	49.72%	53.00%	59.71%	100%	100%

Table 38 : Percentage of overfitting for an AR(1)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0%	0%	0%	0%	0%	0%	0%	0%
50	82.37%	0%	67.64%	0%	0%	0%	0%	0%
100	74.87%	0%	14.00%	0%	0%	0%	0%	0%
150	48.83%	0%	0%	0%	0%	0%	0%	0%

Table 39 : Percentage of overfitting for an AR(2)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	29.00%	16.00%	45.00%	0%	0%	0%	0%	0%
50	100%	59.00%	100%	59.00%	63.00%	63.00%	0%	0%
100	100%	39.00%	76.71%	39.00%	46.18%	29.17%	0%	0%
150	98.09%	0%	28.00%	0%	0%	0%	0%	0%

Table 40 : Percentage of overfitting for an AR(3)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0.07%	0%	4.12%	0%	0%	0%	0%	0%
50	77.89%	0%	22.12%	0%	0%	0%	0%	0%
100	26.27%	0%	0.04%	0%	0%	0%	0%	0%
150	60.21%	0%	3.95%	0%	0.31%	0%	0%	0%

Table 41 : Percentage of overfitting for an MA(1)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	90.57%	50.84%	100%	50.84%	0%	0%	0%	0%
50	100%	42.41%	100%	42.41%	28.41%	28.05%	0%	0%
100	100%	14.18%	100%	14.18%	26.31%	16.94%	0%	0%
150	100%	0.02%	67.54%	0.02%	12.18%	0%	0%	0%

Table 42 : Percentage of overfitting for an MA(2)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	77.02%	46.12%	99.42%	46.12%	0%	0%	0%	0%
50	94.92%	35.60%	77.13%	35.50%	23.06%	22.03%	0%	0%
100	99.40%	41.04%	78.25%	41.04%	44.35%	38.17%	0%	0%
150	86.79%	22.47%	55.84%	22.47%	36.67%	20.23%	0%	0%

Table 43 : Percentage of overfitting for an MA(3)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	41.04%	33.35%	49.87%	33.34%	0%	0%	0%	0%
50	96.23%	79.16%	93.07%	79.16%	69.43%	68.72%	0%	0%
100	73.72%	64.25%	69.89%	64.25%	65.50%	62.64%	0%	0%
150	87.04%	74.68%	82.64%	74.68%	78.87%	75.51%	0%	0%

Table 44 : Percentage of overfitting for an ARMA(1,1)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	99.62%	0.97%	100%	0.97%	0%	0%	0%	0%
50	100%	60.08%	100%	60.08%	28.33%	28.30%	0%	0%
100	100%	0.25%	93.40%	0.25%	3.85%	0%	0%	0%
150	100%	0%	25.55%	0%	1.11%	0%	0%	0%

Table 45 : Percentage of overfitting for an ARMA(1,2)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0%	0%	0%	0%	0%	0%	0%	0%
50	71.25%	11.44%	43.86%	11.44%	2.97%	2.48%	0%	0%
100	47.73%	1.45%	11.69%	1.44%	2.26%	0.59%	0%	0%
150	54.16%	0.17%	5.03%	0.17%	0.22%	0.13%	0%	0%

Table 46 : Percentage of overfitting for an ARMA(2,1)								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	0%	0%	0%	0%	0%	0%	0%	0%
50	64.43%	45.79%	56.94%	45.79%	38.58%	37.89%	0%	0%
100	67.71%	35.02%	53.22%	35.02%	37.35%	31.43%	0%	0%
150	79.49%	28.96%	54.00%	28.96%	39.12%	25.72%	0%	0%

Table 47 : Percentage of times the Information Criteria overfitted (for k<4) for all processes by sample size								
Sample size	AIC	BIC	HQ	SIC	AICc	SICc	AICu	FPE
15	37.50%	16.41%	44.24%	16.41%	0%	0%	0%	0%
50	87.45%	37.10%	73.42%	37.10%	28.16%	27.78%	0%	0%
100	76.63%	21.73%	55.24%	21.73%	25.09%	19.88%	0%	0%
150	79.40%	14.03%	35.88%	14.03%	18.72%	13.51%	0%	0%