A NOVEL ALGORITHM FOR TIME SERIES MODEL SELECTION: MUTUAL INFORMATION MODEL SELECTION ALGORITHM (MIMSA)

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ELİF AKÇA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
STATISTICS

JULY 2017

Approval of the thesis:

A NOVEL ALGORITHM FOR TIME SERIES MODEL SELECTION: MUTUAL INFORMATION MODEL SELECTION ALGORITHM (MIMSA)

submitted by **ELİF AKÇA** in partial fulfillment of the requirements for the degree of **Master of Science in Statistics Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences** ________________

Prof. Dr. Ayşen Akkaya
Head of Department, **Statistics** ________________

Assoc. Prof. Dr. Ceylan Yozgatlıgil
Supervisor, **Statistics Department, METU** ________________

Examining Committee Members:

Assoc. Prof. Dr. Özlem İlk Dağ
Statistics Department, METU ________________

Assoc. Prof. Dr. Ceylan Yozgatlıgil
Statistics Department, METU ________________

Assoc. Prof. Dr. Berna Burçak Başbuğ Erkan
Statistics Department, METU ________________

Assoc. Prof. Dr. Seher Nur Sülkü
Econometrics Department, Gazi University ________________

Prof. Dr. Mehmet Yılmaz
Statistics Department, Ankara University ________________

Date: ________________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: ELİF AKÇA

Signature :

ABSTRACT

A NOVEL ALGORITHM FOR TIME SERIES MODEL SELECTION: MUTUAL INFORMATION MODEL SELECTION ALGORITHM (MIMSA)

AKÇA, ELİF
M.S., Department of Statistics
Supervisor : Assoc. Prof. Dr. Ceylan Yozgatlıgil

July 2017, 165 pages

Time series model selection has gained a significant popularity, and a variety of methods has been introduced in the recent years. It is crucial for a method to propose a candidate model as the final model that explains the procedure underlying a series best and provides accurate forecasts among many candidates. In this study, it is aimed to create an algorithm for order selection in Box-Jenkins models that combines penalized natural logarithm of mutual information among the original series and predictions coming from each candidate. The penalization is achieved by subtracting the number of parameters in each candidate and empirical information the data provide. Simulation studies under various scenarios and application on a real data set imply that our algorithm offers a promising and satisfactory alternative to its counterparts.

Keywords: Time Series, Box-Jenkins Models, Order Selection, Mutual Information.

ÖZ

ZAMAN SERİLERİ İÇİN YENİ BİR MODEL SEÇME ALGORİTMASININ OLUŞTURULMASI: KARŞILIKLI BİLGİ MODEL SEÇME ALGORİTMASI

AKÇA, ELİF
Yüksek Lisans, İstatistik Bölümü
Tez Yöneticisi : Doç. Dr. Ceylan Yozgatlıgil

Temmuz 2017 , 165 sayfa

Zaman serilerinde model seçimi, son yıllarda birçok araştırmacı tarafından ilgi gören bir alandır. Bu kapsamda birçok yöntem literatüre kazandırılmıştır. Herhangi bir seri için aday modeller arasından en iyi modeli seçmek, veriyi oluşturan prosedürün açıklanması ve verinin bir sonraki zaman diliminde nasıl bir değer alacağını öngörmek açısından oldukça önemlidir. Bu tezde, orijinal veri ve her bir aday modelin sağladığı tahminler arasındaki ortak bilgiyi, veriden gelen ampirik bilgi ve aday modellerdeki parametre sayısı ile düzenleyen yeni bir algoritmanın geliştirilmesi amaçlanmaktadır. Farklı senaryolar için gerçekleştirilen simülasyon çalışmaları ve gerçek veri üzerine uygulama, Box-Jenkins modellerinde sıra seçimi için geliştirilen bu algoritmanın var olan diğer yöntemler için iyi bir alternatif olabileceğini doğrulamaktadır.

Anahtar Kelimeler: Zaman Serileri, Box-Jenkins Modelleri, Model Seçme Yöntemleri, Karşılıklı Bilgi.

To my mom

ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to my advisor Assoc. Prof. Dr. Ceylan Yozgatlıgil for her precious guidance and endless patience during this thesis study. I would like to state that she is more than an advisor to me. She was very supportive and understanding in each stage of my study. Whenever I need, she was always with me as a mentor. Because of her mastery in our study, kindness and helpfulness, I have seen her as a role model both in my academic career and personal life. I feel very fortune to be guided by her as her student.

I would also like to thank to my dearest friends Ebru Öztürk and Özgecan Yıldırım for their valuable friendship and tactful support from the beginning of my study. We achieved our theses like our other accomplishments since our B.S. studies together. I feel very lucky to have been accompanied by them. Furthermore, I would like to thank to all members of METU Statistics Department, especially, Neşe Bayram, Gamze Musluoğlu, Berna Tuncer, Melih Ağraz and my officemate, Onur Çamlı, for listening to my complaints about my thesis study. Especially Onur was very patient and supportive.

My special thanks go to my lovely mom, Mesudiye Akça, for her endless and loving support through every stage of my life. She loved me unconditionally and supported me for being a successful woman.

Finally, I would like to express my deepest gratitude to the other members of my family for their lovely support. Without their encouragement, I would not succeed in being here. I am especially appreciative to my lovely sister, Betül Akça, for her valuable assistance and considerate tolerance. Now, I promise to make my family proud of my further achievements in the future.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
ANN	Artificial Neural Networks
ARIMA	AutoRegressive Integrated Moving Average Model
AIC_c	Corrected Akaike Information Criterion
BIC	Bayesian Information Criterion
EIC	Empirical Information Criterion
FPE	Final Prediction Error
GA	Genetic Algorithms
HQIC	Hannan-Quinn Information Criterion
KPSS	Kwiatkowski–Phillips–Schmidt–Shin Test
KL	Kullback- Leibler Distance
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
ME	Mean Error
MSE	Mean Square Error
MI	Mutual Information
MIMSA	Mutual Information Model Selection Algorithm
RMSE	Root Mean Square Error
RWM4	Weighted Model Selection with MIMSA, MAPE and BIC by Rolling Windows
RWMS	Weighted Model Selection by Rolling Windows
W3	Blowfly B data set
WM1	Weighted Model Selection with MIMSA and MAPE
WM2	Weighted Model Selection with MIMSA and BIC
WM3	Weighted Model Selection with MAPE and BIC
WM4	Weighted Model Selection with MIMSA, MAPE and BIC
WMS	Weighted Model Selection

CHAPTER 1

INTRODUCTION

There is a famous quotation appreciated a lot in model selection environment, "All models are wrong, but some are useful." (George Box, 1976). From this expression, it could be inferred that it is not possible to find the exact process that governs a time series. However, it is of great importance to approximate it with the highest possible accuracy.

Time series model selection has been studied a lot in the recent century. People from a variety of disciplines improved many criteria and methods to select the best model among a set of possible candidate models. The ultimate goal with this aim is improving a method that evaluates the candidate models according to some criteria and proposes the most satisfactory one in terms of fitting and forecasting time series.

Since time series has been significantly used in a diversity of branches from Econometrics to Signal Processing, it is crucial to achieve at selecting the best candidate model in terms of forecasting. For instance, governmental authorities or stock holders might take precautions against a financial crisis if it could be forecast from the econometric indicators. If the relevant indicators are modeled and forecast appropriately, this could be achieved. To cite another example, one might refer to meteorological variables. The risk of rain is in the concern of many like farmers or individuals who wish to go out. If the model selection part is conducted with an appropriate method, precise predictions and forecasts could be obtained.

Model selection in time series could be classified into two steps: i. determination of model class and ii.order selection within the specified class(es). Although the latter one is more complex and requires comprehensive work, many methods have been proposed to this end. Model class of a series are generally determined by some visual inspection tools like the sample Autocorrelation- Partial Autocorrelation Function (ACF-PACF) plots or pattern identification methods such as Extended Sample Autocorrelation Function (ESACF) or the Smallest Canonical Correlation (SCAN) tables.

Following the selection of model class, current model order selection methods for time series might be classified into two main groups: i. information criteria and ii. error based forecast measures. The most popular criteria of type (i) is Akaike Information Criterion (AIC). This criterion is based on penalized model likelihood. In order to improve the theoretical properties of this criterion, numerous derivations has been presented via different criteria like Corrected Akaike Information Criterion (AIC_c), Bayesian Information Criterian (BIC) and Hannan-Quinn Information Criterion (HQIC). Later on, Billah et al. (2003) proposed Empirical Information Criterion (EIC) as a new method of order selection in time series. Numerous measures and their comparison for type (ii) are provided by Hyndman and Koehler (2006). The most widely used ones might be Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). On the other hand, there are other procedures that may not belong to this classification. For example, Ong et al. (2005) suggested to use Genetic Algorithms (GA) to determine the lag order for Box-Jenkins models. When all mentioned model selection methods are investigated, one could notice that model selection is still a demanding issue and should immediately be improved for time series.

When we examine the literature for model selection in time series, we find a criterion named LIC and its derivations that calculate the mutual information among the past and future observations of a series and penalize this by the information both the sample size and candidate models provide. Since mutual information is a comprehensive measure of dependencies, we plan to improve an algorithm that benefits from this measure and concerns about fitting and forecasting time series.

This thesis focuses on developing an algorithm that takes the mutual information among the series of interest to be modeled and predictions of each candidate model proposed. As in our knowledge, there is no such algorithm that considers lag order selection problem in time series along this approach. Here, we concentrate on order selection for non-seasonal Box-Jenkins models, which are the most globally applied type of time series models (Box and Jenkins, 1976).

This thesis is organized as follows. Chapter 2 reviews some literature on Box-Jenkins models and currently available methods to select lag order for that type of time series models. In Chapter 3, we provide background information on mutual information and introduce our algorithm, Mutual Information Model Selection Algorithm (MIMSA) in detail. Chapter 4 summarizes how we benefit from weighting model selection and Chapter 5 presents combining weighting model selection with rolling windows approach to increase the forecast performance of MIMSA. While Chapter 6 provides our experience with application on real data, in Chapter 7, we finally mention our main findings and noteworthy inferences for this thesis.

CHAPTER 2

LITERATURE REVIEW

2.1 Non-Seasonal Box- Jenkins Models

Time series refers to data that were collected over regular or irregular time points. Because of this, a series enables us to monitor how a characteristic of interest varies in both space and time by incorporating the time effect. What distinguishes time series data is the dependency structure it has. Since the observations in a series are interrelated to each other sequentially in time, the methods and models that generally assume independency are not usable. For this reason, there have been many advances in time series analysis as a distinct discipline.

Although a variety of methods and approaches have been ameliorated such as exponential smoothing or Artificial Neural Networks (ANN), Box-Jenkins modeling is still a universally accepted and classical procedure that comes in mind by default to analyze a univariate time series. This type of models are appropriate to accomplish for any type of series that is discrete or continuous (Pankratz, 2009).

A series, Y_t has the following form if it could be modeled by a standard Box-Jenkins AutoRegressive Integrated Moving Average, ARIMA(p,d,q) model with no seasonality:

$$(1-\sum_{k=1}^{p}\Phi_k B^k)(1-B)^d Y_t = \mu + (1+\sum_{k=1}^{q}\Theta_k B^k)\varepsilon_t \tag{2.1}$$

where B is the difference operator that is mentioned in 2.1.2, p,d and q are to be estimated in order to determine the lag order of Y_t and $\widetilde{\Phi}$ and $\widetilde{\Theta}$ are vectors of parameters to characterize the series and μ is the process mean (Box&Jenkins, 1976).

In Equation 2.1, $\widetilde{\Phi}$ and $\widetilde{\Theta}$ are used to represent the autoregressive and moving average parameters of dimension p and q, respectively. If k represents the number of parameters to estimate the dimension of a series, it is defined by $k = p + q$. In order to construct a model which has the above form, the required assumptions relating to the error term are clarified in the following part, 2.1.1.

2.1.1 Assumptions for Constructing Box-Jenkins Models

- **Stationarity**
  One should validate that a series is stationary in order to implement Box-Jenkins models. That favors the stability of model parameters over time so that the constancy of the probability structure dictating the series is guaranteed.
- **White Noise**
  This assumption requires the residuals of ARIMA models to be independent and identically distributed with mean 0 and a constant variance. Gaussian distributed residuals are generally desired in order to benefit from the theoretical properties of this distribution.
- **Temperate Sample Size**
  In order to attain satisfactory approximations through ARIMA modeling, a time series should have at least 50 observations (Pankratz, 2009).

2.1.2 Elements of Box-Jenkins Models

An ARIMA(p,d,q) model is composed of three main segments: autoregressive, integration and moving average.

- **AutoRegressive Segment**
  This element represents the self reliance of time series on its values on the preceding lags. In other words,it is hired to reflect the dependency structure of a series on the lagged observations.
  If a series is identified by an autoregressive part, it is called an AR(p) model:

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + ... + \Phi_p Y_{t-p} + \varepsilon_t \tag{2.2}$$

where ε_t stands for error term or noise in model.

- **Integrated Segment**

  If the stationarity assumption for the series of interest to be modeled is not satisfied, differencing should be taken to construct a Box-Jenkins model. Differencing transforms the series by using the difference of the lagged observations, ∇Y_t rather than the original series, where

  $$\nabla Y_t = Y_t - Y_{t-1} = (1 - B)Y_t \tag{2.3}$$

  Differencing might be needed to apply a couple of times, say d, to achieve a stationary series. Hence, d represents how many times differencing has been taken in an ARIMA model.

- **Moving Average Segment**

  This part particularizes how dominant the noise is in specifying the present value of Y_t.

  If a series is identified by a moving average part, it is called an MA(q) model:

  $$Y_t = \Theta_1 \varepsilon_{t-1} + \Theta_2 \varepsilon_{t-2} + ... + \Theta_q \varepsilon_{t-q} \tag{2.4}$$

  As a result, Box-Jenkins models establish three separate processes to investigate various aspects of time series (Wagenmakers et al., 2006).

2.2 Order Selection Methods

Following the specification of model class for a series, the corresponding lag order should be determined. Any method to this end should discriminate the difference between candidate orders. For instance, a method should select the most appropriate one through making a comparison among candidates AR(1) and AR(3). Although each method considers its own criteria to determine the most attractive model, simplicity should also be taken into account while comparing. In the literature, there exists many methods to determine lag order for time series. These methods might be categorized into three groups as information criteria, error-based forecast measures and model selection algorithms. The most widely used methods for each category is discussed in the following sub-chapters.

2.2.1 Information Criteria

For the purpose of estimation of the dimensionality of a time series model, information criteria are commonly applied. Such criteria takes the likelihoods of the set of candidates models into consideration and penalize them by using the information the series of interest and each candidate provide in terms of number of observations and parameters.

- **Akaike Information Criterion (AIC)**

  AIC was proposed by Akaike (1973) and a universally preferred method in order selection problems for time series. This criterion is an estimate for the Kullback- Leibler distance (KL) (Kullback & Leibler, 1951) between the underlying procedure that governs series and candidate models.

$$KL(f,g) = \int f(x)log(\frac{f(x|\theta)}{g(x|\widehat{\theta})})dx = E(log(f(x|\theta))) - E(log(g(x|\widehat{\theta})) \quad (2.5)$$

  where $f(x|\theta)$ and $g(x|\widehat{\theta})$ represent the probability function of the true model and approximating model, respectively. This points out that a candidate which provides the closest distance to the true model should be chosen by AIC.

  Since KL has a close relation to the likelihood function, AIC has the following form:

$$AIC = -2log(L(\widehat{\theta})) + 2k \quad (2.6)$$

  where, k is the number of parameters to be estimated in a model. The smaller the value of AIC, the better the model. Therefore, one should select a model which offers a good approximation to the generating process and increases the likelihood of the observed series.

- **Bayesian Information Criterion (BIC)**

  BIC is the second commonly employed method in modeling environment.

$$BIC = -2log(L(\widehat{\theta})) + klog(n) \quad (2.7)$$

where n symbolizes the length of series of interest (Schwarz, 1978). BIC stands for the implementation of Bayesian approach which assigns equal prior probabilities for each candidate model. It is classified as a dimension consistent criterion due to searching for an optimal estimation of the true model dimension that is unbiased and has variance approximating 0, that are the required qualifications through consistency (Anderson & Burnham, 1999). Small values of BIC signifies preferable models.

- **Hannan- Quinn Information Criterion (HQIC)**

  Like AIC and BIC, HQIC relies on the computation of model likelihood, but its penalization term is different.

$$HQIC = -2log(L(\widehat{\theta})) + 2k * log(log(n)) \tag{2.8}$$

  It is known to be highly consistent in dimension estimation (Hannan & Quinn, 1979).

- **Corrected Akaike Information Criterion (AIC_c)**

  To enhance the overparametrization problem of AIC in small samples, its penalization term was adjusted by Hurvish and Tsai (1989). Sen and Shitan (2002) reveals that AIC_c tolerately outperforms AIC.

$$AIC_c = -2log(L(\widehat{\theta})) + \frac{2k(k+1)}{n-k-2} \tag{2.9}$$

  For sufficiently large sample sizes, AIC_c converges to AIC.

- **Final Prediction Error (FPE)**

  This criterion as a method for model selection was proffered by Akaike (1970).

$$FPE = -2log(L(\widehat{\theta})) + n(log(n+k) - log(n-k)) \tag{2.10}$$

  For selecting a better model, one should prefer smaller values of FPE.

2.2.2 Error Based Forecast Measures

If the main purpose of modeling is forecasting a series, out-of-sample performance based on forecast measures is evaluated. Myung (2000) highlights that a model selection method should consider the competing models' capability of picking up the behaviour of future observations that are not seen yet. If a series of length n is of concern to be modeled, each candidate is fitted to the first $n-h$ observations and the mean forecast accuracy measures for the remaining h observations are calculated. Then, any candidate model providing the minimum forecast error is proposed as the best. Applying such a procedure in model selection is known cross validation.
Inoue and Kilian (2006) suggests using error based forecast measures in model selection, especially, for small samples to obtain good forecasts. There are many measures of forecast accuracy. Hyndman and Koehler (2006) reviews the mostly used ones as follows:

- **Mean Error (ME)**

  This measure requires the computation of mean deviation between the original series and predictions coming from any candidate.

  $$ME = \frac{\sum_{i=1}^{n} e_{ti}}{n} \tag{2.11}$$

  where n denotes the length of the series in a model and

  $$e_{ti} = Y_{ti} - \widehat{Y_{ti}} \tag{2.12}$$

  For selecting the best, one should choose a candidate model with minimum ME.

- **Mean Absolute Error (MAE)**

  To compute this measure, one needs to obtain the mean of the absolute differences between the original series and predictions.

  $$MAE = \frac{\sum_{i=1}^{n} |e_{ti}|}{n} \tag{2.13}$$

Like ME, small values of MAE are preferable. It provides a practical way for interpreting compared to ME.

- **Mean Square Error (MSE)**

  The calculation of the mean of squared difference between the observations of the original series and the corresponding predictions is necessary for this measure. It was firstly used by Thompson (1990).

$$MSE = \frac{\sum_{i=1}^{n} e_{ti}^2}{n} \tag{2.14}$$

  A small value of MSE is a reason for preference through selecting the best candidate model.

- **Root Mean Square Error (RMSE)**

  This forecast accuracy measure is nothing but the squared root of MSE.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} e_{ti}^2}{n}} \tag{2.15}$$

- **Mean Absolute Percentage Error (MAPE)**

  This measure is claimed to perform better than RMSE (Armstrong, 2001).

$$MAPE = \frac{\sum_{i=1}^{n} |p_{ti}|}{n} \tag{2.16}$$

  where

$$p_{ti} = \frac{100 * e_{ti}}{Y_{ti}} \tag{2.17}$$

  As the previously mentioned measures, a smaller value of $MAPE$ indicates good forecasts.

- **Mean Absolute Scaled Error (MASE)**

  This was introduced by Hyndman and Koehler (2006) as an alternative measure for forecast accuracy. It makes a relative comparison between in-sample one-step forecasts and the forecast performance measured by taking the difference of original observations and predictions.

$$MASE = \frac{\sum_{i=1}^{n} |q_{ti}|}{n} \tag{2.18}$$

where

$$q_{ti} = \frac{e_{ti} * (n-1)}{\sum_{i=2}^{n} |Y_{ti} - Y_{ti-1}|}. \tag{2.19}$$

It is implied that MASE values less than 1 point out that a model provides more accurate forecasts than in-sample one-step forecasts.

2.2.3 Other Model Selection Methods and Algorithms

In the literature, there exists other procedures that may not be grouped into the previous classes to accomplish order selection in time series.

- **Empirical Information Criterion (EIC)**
  Billah et al. (2003) emphasizes that information criteria are composed of two terms regarding likelihood and penalization term. They assert only penalization term differs for each criteria but the term for (log)likelihood is constant.

$$IC = log(L(\theta)) - f(n,k) \tag{2.20}$$

  where penalization component,$f(n,k)$, is a function of data-driven and model based terms. Hence, they constructed an algorithm that estimates the penalization term according to the information series provides instead of using a fixed functional form. This method is offered for large series and a bootstrapped version is recommended for small samples.

- **Model Identification of ARIMA Family Using Genetic Algorithms**
  Ong et al. (2005) presents a novel algorithm for order of lag identification in ARIMA modeling. This algorithm implements Genetic Algorithms (GA), in which the order of candidate models are treated as chromosomes. As the principle of natural selection of evolutionary theory, the fittest chromosomes, orders, survive at the end and selected as the best order approximating to the

true model order. They state that this algorithm outperforms its counterparts in terms of forecasts.

CHAPTER 3

MUTUAL INFORMATION MODEL SELECTION ALGORITHM (MIMSA)

In this chapter, we introduce our proposed algorithm after providing some background information on the quantity our algorithm benefits from. Then, we present the results of simulation studies to evaluate its performance with the made up data, in which the procedure governing time series is exactly known.

3.1 Mutual Information Concept

Mutual information (MI) was pioneered by Tzannes and Noonan (1973) as an alternative method of allocating priors in Bayesian framework. Later on, it was examined deeply and generalized as a means of functional association among variables.

MI could be prescribed as a metric to express the amount of dependency or cooperation among variables. By definition, it might be presumed as correlation. Unlike correlation, it does not require any assumption related to the nature of dependency. In fact, the information it provides covers any type of linear or nonlinear relationship. For this reason, Lange and Grubmüller (2006) mentions MI generalized correlation as an extensive measure of relation. On the other hand, if variables of interest follow bivariate normal density, there exists a direct relationship between correlation and MI.

To clarify MI in other words, one could define it as an indication for how much information is shared or interrelated by the variables of interest. Thus, it can be used to represent the quantity of the reduced uncertainty about a random variable resulting from having introduced the other.

The MI between two random variables, X and Y, is computed by:

$$I(X,Y) = \int\int f_{XY}(x,y)log(\frac{f_{XY}(x,y)}{f_X(x)f_Y(y)})dxdy \tag{3.1}$$

where $I(X,Y)$ denotes the MI between X and Y, and $f_{XY}(x,y)$, $f_X(x)$ and $f_Y(y)$ symbolize joint and marginal probability density functions of X and Y, respectively. The ratio term, $\frac{f_{XY}(x,y)}{f_X(x)f_Y(y)}$ in the above equation is known as density ratio and represented by $w(x,y)$ in some sources.

MI is also computable when one or both of the variables are discrete. The integration operator in Equation 3.1 is substituted by a single or double summation in that case. Besides, it is also possible to benefit from MI to search for an evidence of linkage for more than two variables. Joint MI by using the conditional MI's serves to that end. The joint MI between three random variables, X, Y and C is calculated by:

$$I(X,Y,C) = I(X,C|Y) + I(Y,C) \tag{3.2}$$

where $I(X,C|Y)$ quantifies the measure of information in common for two variables provided that the other has been put on stage.

Many methods are employed such as maximum likelihood density estimation (Suzuki et al.,2008(b)), least squares approach (Suzuki et al., 2008(a)) and k-nearest neighbors method (Kraskov et al., 2004) to estimate MI. Kernel density estimation (KDE) is the most commonly used method to approximate MI. In this method, marginal and joint densities are required to be separately estimated from the information coming from the corresponding samples and used in calculation (Silverman, 1986).

MI plays a crucial role in a diverse set of machine learning problems like independent component analysis (Comon, 1994), feature selection (Peng et al., 2005& Bennasar et al., 2015), blind source separation (El Rhabi et al., 2004) and input selection (Božić et al., 2013) due to its attractive properties listed below:

- MI is nonnegative, $I(X,Y) \geq 0$.
- MI has symmetricity property, $I(X,Y) = I(Y,X)$.

- MI can handle detecting any type of dependencies even with no correlation.
- MI has invariance property for linear transformations of X and Y.
- MI has an association with entropy which is a quantity of uncertainty a random variable carries, $I(X,Y) = H(X) + H(Y) - H(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$, where $H(.)$ represents entropy.
- $I(X,Y) = 0$, if and only if the components, X and Y, are statistically independent.

Moreover, Simon and Verleysen (2006) and Gao and Tian (2010) utilize MI to detect lags in regression and nonlinear modeling, respectively. Use of MI is also common in time series analysis. Božić et al.(2013) use MI for forecasting while Sorjamaa et al. (2005) use it for prediction.

3.2 Mutual Information Model Selection Algorithm (MIMSA)

Li and Xie (1996) created a new selection criterion named LIC which requires minimizing the MI between past and future observations of a series with a penalization term. Afterwards, they built another criterion, MIC, on LIC by modifying the penalization term in order to overcome inconsistency. Their primary goal is to select the precise order in modeling a Gaussian stationary series via the employment of autocovariances. Thereafter, Proietti and Luati (2015) improved the procedure of implementing MI between past and future observations of a series in order selection by using partial autocorrelation coefficients.

However, we prefer to benefit from MI among the original series and the correspondent predictions. Pitt and Myung (2002) apprises that a model selection method should favor a candidate which approximate the underlying true model by predictions that do not deviate too much from the observed data. By the default evocation of MI, a candidate model which maximizes the MI between observations and predictions should be selected as the final model so that redundant models are eliminated. Hence, our algorithm, Mutual Information Model Selection Algorithm (MIMSA), is constructed on the maximization of the following criterion:

$$MIMSA_i = log(I(Y_t, \widehat{Y_t^i})) - \frac{2k(k+1)}{n-k-1} \tag{3.3}$$

where Y_t : series of interest to be modeled, $\widehat{Y_t^i}$: predicted values of candidate model i, n : length of Y_t and k : number of parameters in candidate model i

In this quantity provided in 3.3, we hire the natural logarithm of MI by considering the link between MI and log-likelihood suggested by Galka et al. (2005). Indeed, we penalize this by using a penalization term very similar to the one AIC_c employs. Bozdoğan (1987) states that the penalization term in AIC_c makes a criterion model independent and prevents overparametrization. Therefore, we take the amount of relevance between the predictions and original series into account by implementing the following algorithm:

MIMSA Algorithm: For a time series, the following steps could be achieved in determining the order of the model underlying that series:

Step 1- Determine the candidate models,

Step 2- Calculate $MIMSA_i$ for each candidate,

Step 3- Propose the candidate with the highest $MIMSA_i$ as the final model.

3.3 Simulation Studies

In order to evaluate the performance of MIMSA, a set of simulation studies are made under various scenario. To determine the scenarios, we try parameter value combinations from -0.9 to 0.9 with an incremental increase of size 0.1 in a preliminary study. Then, we observe that significant differences might be observed when the parameter of interest determining the lag order is close to 1 and far away from 1 since the absolute sum of the parameters must be lower than 1 in order to satisfy the stationarity assumption. Besides, a relative comparison of MIMSA with the most preferred criteria, AIC and BIC, is achieved. These scenarios cover different combinations of non-seasonal Box-Jenkins models up to lag length 3 in a stepwise manner. In any scenario, we generate a series of size n from which train and test sets were determined by using the 90% and 10% of the observations. While the first set is hired to measure the ability of detecting the true lag order, the latter serves to quantify how capable of providing satisfactory forecasts a model selection method is. The number of simulations are conducted for 1,000 and 10,000 times for sample sizes 100 and 250 and 500, respectively, since for small sized samples, non-convergence problems in generating series might occur. Although increasing the number of iterations in the generation mechanism of a series from an ARIMA model is applied, which is usually suggested for such type of problems, based on our simulation experience, we note that even 550,000 iterations may not be enough but computational memory overflow burdens prevent one to increase more. Therefore, we consider such an approach to overcome this issue.

The procedure for a simulation could be explained clearly by the following algorithm:

1. Generate a series of size n from normal distribution and assign the first 90% and last 10% of the observations to two sets: train and test, respectively,
2. Determine candidate models for train series and calculate $MIMSA_i$, AIC_i and BIC_i for each candidate,
3. Propose the candidate with the highest $MIMSA_i$ and lowest AIC_i and BIC_i as the final model,
4. Based on the proposed model, obtain the predicted values for the test set and calculate the mostly used forecast measures (ME, RMSE, MAPE and MASE)
5. Repeat steps (1)-(4) N times, where N takes the values of 1,000 for a series of size 100 and 10,000 for the other sample sizes,
6. Summarize N outputs.

Table 3.1 summarizes how MIMSA performs when the true model is AR(1) relative to AIC and BIC. This table indicates that the highest frequency of finding the correct lag order is achieved by BIC. On the other hand, MIMSA outperforms AIC regardless of the sample size and parameter value that determines the amount of auto-regression

Table 3.1: Frequency distribution of model identification for simulation of AR(1)

Parameter	Sample Size	Criterion	Candidate Model					
			AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
$\Phi = 0.85$	n=100	MIMSA	764	164	49	18	5	0
		AIC	580	161	78	54	57	70
		BIC	901	63	26	6	2	2
	n=250	MIMSA	7404	1670	524	262	111	29
		AIC	6037	1482	858	633	476	514
		BIC	9503	399	76	21	1	0
	n=500	MIMSA	6277	1704	834	544	376	265
		AIC	6211	1443	801	599	459	487
		BIC	9682	282	30	4	1	1
$\Phi = 0.5$	n=100	MIMSA	750	159	47	22	19	3
		AIC	584	144	76	75	54	67
		BIC	911	61	17	7	4	0
	n=250	MIMSA	7782	1251	474	287	148	58
		AIC	6092	1449	882	604	470	503
		BIC	9503	411	65	12	9	0
	n=500	MIMSA	7621	1441	518	231	133	56
		AIC	6242	1428	818	554	480	478
		BIC	9634	315	39	8	4	0
$\Phi = -0.9$	n=100	MIMSA	7629	1716	524	106	22	3
		AIC	5792	1419	875	664	580	670
		BIC	9017	686	188	60	30	19
	n=250	MIMSA	7782	1457	486	208	54	13
		AIC	6094	1436	847	631	501	491
		BIC	9524	389	62	17	4	4
	n=500	MIMSA	7498	1515	618	239	105	25
		AIC	6194	1427	831	620	469	459
		BIC	9663	297	32	8	0	0

Note: $Y_t = \Phi Y_{t-1} + \varepsilon_t$ *is generated from AR(1).*

for AR(1). For instance, when $\Phi = 0.5$ and 10000 samples of size 500 are generated, 9634, 315, 39, 8 and 4 of them are specified as AR(1), AR(2), AR(3), AR(4) and AR(5), respectively, by BIC. Conversely, the frequency of correctly specifying the true lag order declines to 6242 and 7621 when AIC and MIMSA are preferred as a way of order selection for the same situation.

Table 3.2: Forecast measures for the simulation of AR(1) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Phi = 0.85$	n=100	MIMSA	0	0.7289	160.6954	0.6750
		AIC	0	0.6127	131.1023	0.5626
		BIC	0	0.7588	167.3982	0.7098
	n=250	MIMSA	0	0.8950	739.9936	0.8496
		AIC	0	0.8691	727.7268	0.8170
		BIC	0	0.9107	744.6871	0.8703
	n=500	MIMSA	0	0.9452	426.5954	0.8981
		AIC	0	0.9423	416.4980	0.8942
		BIC	0	0.9609	425.7974	0.9199
$\Phi = 0.5$	n=100	MIMSA	0	0.7427	265.4917	0.6313
		AIC	0	0.6284	237.7435	0.5283
		BIC	0	0.7889	276.3243	0.6762
	n=250	MIMSA	0	0.9090	439.0654	0.7775
		AIC	0	0.8843	608.5371	0.7486
		BIC	0	0.9240	437.0474	0.7953
	n=500	MIMSA	0	0.9545	401.1285	0.8217
		AIC	0	0.9440	399.8254	0.8087
		BIC	0	0.9620	409.6043	0.8312
$\Phi = -0.9$	n=100	MIMSA	0	0.7741	321.4988	0.2250
		AIC	0	0.6410	157.5623	0.1849
		BIC	0	0.7961	328.2495	0.2335
	n=250	MIMSA	0	0.9151	3634.676	0.2355
		AIC	0	0.8869	3631.615	0.2256
		BIC	0	0.9270	3645.079	0.2397
	n=500	MIMSA	0	0.9573	345.0190	0.2321
		AIC	0	0.9465	342.7461	0.2284
		BIC	0	0.9650	349.0548	0.2348

The above table represents the forecast performance of MIMSA, AIC and BIC. When this table is investigated, one may infer that AIC provides more correct forecasts than MIMSA and BIC since its ME, RMSE, MAPE and MASE values for each Φ value and sample size combination are the smallest.

Table 3.3: Frequency distribution of model identification for simulation of AR(2)

Parameter	Sample Size	Criterion	Candidate Model					
			AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
$\Phi_1 = 0.01$, $\Phi_2 = 0.85$	n=100	MIMSA	101	668	177	34	12	8
		AIC	0	586	146	98	79	91
		BIC	0	889	75	28	7	1
	n=250	MIMSA	61	7929	1317	430	183	80
		AIC	0	6212	1509	874	701	704
		BIC	0	9426	453	93	19	9
	n=500	MIMSA	40	7350	1682	594	229	105
		AIC	0	6419	1440	865	652	624
		BIC	0	9663	285	41	8	3
$\Phi_1 = 0.45$, $\Phi_2 = 0.5$	n=100	MIMSA	665	270	48	12	2	3
		AIC	1	591	147	109	73	79
		BIC	10	904	62	15	4	5
	n=250	MIMSA	5036	4964	0	0	0	0
		AIC	0	6188	1509	929	703	671
		BIC	0	9477	434	66	16	7
	n=500	MIMSA	2762	7238	0	0	0	0
		AIC	0	6356	1445	909	668	622
		BIC	0	9674	279	40	6	1
$\Phi_1 = -0.02$, $\Phi_2 = 0.9$	n=100	MIMSA	124	876	0	0	0	0
		AIC	0	603	140	98	72	87
		BIC	0	909	60	19	9	3
	n=250	MIMSA	21	9979	0	0	0	0
		AIC	0	6265	1486	844	753	652
		BIC	0	9513	404	56	20	7
	n=500	MIMSA	2	9998	0	0	0	0
		AIC	0	6266	1564	919	658	593
		BIC	0	9638	305	50	6	1

Note: $Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \varepsilon_t$ *is generated from AR(2).*

Similar to the AR(1) case, BIC has the highest frequency of correctly determining the true model order when all simulation scenarios are considered. In converse, MIMSA outperforms AIC and BIC when one of the parameters is negative and sample size is moderate or large.

Table 3.4: Forecast measures for the simulation of AR(2) process

Parameter	Sample Size	Criterion	Forecast Measure ME	RMSE	MAPE	MASE
$\Phi_1 = 0.01$, $\Phi_2 = 0.85$	n=100	MIMSA	0	0.6112	160.2081	0.2816
		AIC	0	0.4758	114.3022	0.2169
		BIC	0	0.6135	156.8120	0.2821
	n=250	MIMSA	0	0.8557	283.7554	0.3694
		AIC	0	0.8291	278.4128	0.3545
		BIC	0	0.8642	292.2900	0.3749
	n=500	MIMSA	0	0.9303	1023.764	0.3815
		AIC	0	0.9200	1014.497	0.3758
		BIC	0	0.9356	1027.880	0.3850
$\Phi_1 = 0.45$, $\Phi_2 = 0.5$	n=100	MIMSA	0	0.7005	156.2913	0.5879
		AIC	0	0.4722	102.8269	0.3740
		BIC	0	0.6083	133.4671	0.4895
	n=250	MIMSA	0	0.9108	250.4962	0.7632
		AIC	0	0.8192	232.5865	0.6700
		BIC	0	0.9647	261.4427	0.8139
	n=500	MIMSA	0	0.9651	371.5828	0.8111
		AIC	0	0.9146	353.6347	0.7609
		BIC	0	0.9304	370.9577	0.7798
$\Phi_1 = -0.02$, $\Phi_2 = 0.9$	n=100	MIMSA	0	0.6661	124.4016	0.2505
		AIC	0	0.4788	90.8517	0.1754
		BIC	0	0.6095	114.6603	0.2299
	n=250	MIMSA	0	0.8645	267.4457	0.2988
		AIC	0	0.8258	251.9903	0.2810
		BIC	0	0.8608	266.6286	0.2974
	n=500	MIMSA	0	0.9359	300.0023	0.2991
		AIC	0	0.9191	294.7662	0.2915
		BIC	0	0.9348	329.9794	0.2986

Table 3.4 outlines the forecast performance of the corresponding selection methods. The last 10% of the observations are predicted and the required forecast measures are computed to evaluate the forecast performance. One might understand from this table that MIMSA provides more accurate forecasts than BIC when the parameters of AR(2) process are close to each other and one for AR(1) part is negative.

Table 3.5: Frequency distribution of model identification for simulation of AR(3)

Parameter	Sample Size	Criterion	Candidate Model					
			AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
$\Phi_1 = 0.01$, $\Phi_2 = 0.01$, $\Phi_3 = 0.9$	n=100	MIMSA	54	60	772	103	8	3
		AIC	0	0	641	155	98	106
		BIC	0	0	901	64	21	14
	n=250	MIMSA	257	223	6330	914	804	442
		AIC	378	441	6546	839	796	563
		BIC	2	0	9447	436	85	24
	n=500	MIMSA	816	592	4488	1963	805	616
		AIC	352	358	6057	1847	822	564
		BIC	0	3	9689	269	34	5
$\Phi_1 = 0.01$, $\Phi_2 = 0.01$, $\Phi_3 = 0.5$	n=100	MIMSA	485	101	339	48	20	7
		AIC	4	0	628	165	100	103
		BIC	37	1	869	63	23	7
	n=250	MIMSA	648	242	7143	1630	267	70
		AIC	0	0	6427	1552	1035	986
		BIC	0	0	9496	410	71	23
	n=500	MIMSA	147	60	7550	1443	570	230
		AIC	0	0	6411	1659	1018	912
		BIC	0	0	9647	305	42	6
$\Phi_1 = 0.01$, $\Phi_2 = 0.01$, $\Phi_3 = -0.875$	n=100	MIMSA	95	42	764	94	3	2
		AIC	0	0	594	162	118	126
		BIC	0	0	893	60	36	11
	n=250	MIMSA	183	118	6577	2012	742	368
		AIC	0	0	6433	1625	1029	913
		BIC	0	0	9507	390	85	18
	n=500	MIMSA	352	221	5931	1902	1016	578
		AIC	0	0	6612	1580	961	847
		BIC	0	0	9645	314	38	3

Note: $Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \Phi_3 Y_{t-3} + \varepsilon_t$ *is generated from AR(3).*

When the true model is simulated from an AR(3) process, BIC performs better than its counterparts in terms of identifying the underlying lag order. Moreover, its forecasts are the most acceptable.

One may refer to Table 3.6 to examine this issue. Although ME of BIC is not the smallest when, especially, moderate and small samples are generated, its related

MASE, MAPE and RMSE quantities would compensate this.

Table 3.6: Forecast measures for the simulation of AR(3) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Phi_1 = 0.01$, $\Phi_2 = 0.01$, $\Phi_3 = 0.9$	n=100	MIMSA	0	0.5189	99.9739	0.1811
		AIC	0	0.3556	63.0813	0.1227
		BIC	0	0. 4540	76.6621	0.1575
	n=250	MIMSA	0	0.8073	294.0234	0.2787
		AIC	0	0.7598	349.6460	0.2606
		BIC	0	0.8140	356.4897	0.2844
	n=500	MIMSA	0	0.9278	284.5884	0.3149
		AIC	0	0.8889	292.4207	0.3027
		BIC	0	0.9117	299.8160	0.3136
$\Phi_1 = 0.01$, $\Phi_2 = 0.01$, $\Phi_3 = 0.5$	n=100	MIMSA	0	0.6760	244.4539	0.4240
		AIC	0	0.3795	134.3695	0.2218
		BIC	0	0.5059	172.1733	0.2992
	n=250	MIMSA	0	0.8357	402.6798	0.5025
		AIC	0	0.8004	405.3907	0.4757
		BIC	0	0.8298	404.1375	0.4990
	n=500	MIMSA	0	0.9147	473.2951	0.5600
		AIC	0	0.9047	496.3330	0.5520
		BIC	0	0.9180	489.8604	0.5635
$\Phi_1 = 0.01$, $\Phi_2 = 0.01$, $\Phi_3 = -0.875$	n=100	MIMSA	0	0.6413	145.0410	0.2299
		AIC	0	0.3617	109.4613	0.1334
		BIC	0	0.4880	130.0853	0.1784
	n=250	MIMSA	0	0.8495	230.5143	0.3078
		AIC	0	0.8127	230.9287	0.2938
		BIC	0	0.8413	237.5636	0.3078
	n=500	MIMSA	0	0.9676	389.3895	0.3463
		AIC	0	0.9118	398.0751	0.3268
		BIC	0	0.9241	399.2046	0.3332

In order to observe MIMSA's order detection achievement in MA processes, a collection of simulation studies for diverse parameter values and sample sizes is conducted. For each scenario, orders from 1 to 6 are evaluated in the candidate set.

Table 3.7 might be referred to interpret on model identification of AIC, BIC and

Table 3.7: Frequency distribution of model identification for simulation of MA(1)

Parameter	Sample Size	Criterion	Candidate Model					
			MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
$\Theta = 0.85$	n=100	MIMSA	725	213	35	17	8	2
		AIC	640	128	62	57	64	70
		BIC	918	43	14	8	9	8
	n=250	MIMSA	6001	2611	704	342	209	133
		AIC	7139	1199	619	393	328	322
		BIC	9742	226	1	1	1	0
	n=500	MIMSA	4433	2714	1162	771	518	402
		AIC	7185	1236	622	416	303	238
		BIC	9860	119	19	2	0	0
$\Theta = 0.5$	n=100	MIMSA	709	187	59	23	13	9
		AIC	666	99	62	46	62	65
		BIC	935	26	10	5	15	9
	n=250	MIMSA	6531	2132	682	334	204	117
		AIC	7153	1207	414	315	294	503
		BIC	9758	205	31	2	4	0
	n=500	MIMSA	5036	2431	1035	657	482	359
		AIC	7290	1174	614	400	292	230
		BIC	9841	147	11	1	0	0
$\Theta = -0.9$	n=100	MIMSA	667	213	68	32	11	9
		AIC	635	130	76	51	53	55
		BIC	911	61	15	7	4	2
	n=250	MIMSA	6164	2471	681	369	206	109
		AIC	6784	1209	727	495	380	405
		BIC	9584	323	68	21	2	2
	n=500	MIMSA	4823	2707	1051	664	431	324
		AIC	7251	1127	648	390	278	306
		BIC	9839	145	13	3	0	0

Note: $Y_t = \Theta\varepsilon_{t-1}$ *is generated from MA(1).*

MIMSA when the generated series follows an MA(1) process. This table points out that BIC has the best performance.

Table 3.8: Forecast measures for the simulation of MA(1) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Phi = 0.85$	n=100	MIMSA	-0.0079	0.7567	482.4884	0.6107
		AIC	-0.0033	0.6980	462.1222	0.5610
		BIC	-0.0036	0.7937	491.0681	0.6405
	n=250	MIMSA	0.0008	0.9247	575.1472	0.7145
		AIC	-0.0003	0.9221	557.4347	0.7122
		BIC	0.0008	0.9514	600.7783	0.7354
	n=500	MIMSA	-0.0008	0.9702	440.3121	0.7426
		AIC	-0.0009	0.9772	441.8922	0.7479
		BIC	-0.0010	0.9856	444.3805	0.7544
$\Phi = 0.5$	n=100	MIMSA	-0.0013	0.7269	271.4019	0.6050
		AIC	0	0.6892	261.7742	0.5687
		BIC	0.0018	0.7867	294.7708	0.6529
	n=250	MIMSA	0	0.9216	365.4768	0.7647
		AIC	0.0006	0.9177	367.5233	0.7618
		BIC	0.0004	0.9471	358.6052	0.7861
	n=500	MIMSA	0	0.9615	778.5448	0.7930
		AIC	0.0003	0.9676	827.224	0.7978
		BIC	0.00013	0.9761	811.3184	0.8048
$\Phi = -0.9$	n=100	MIMSA	-0.0082	0.7120	249.6382	0.3236
		AIC	0.0017	0.6890	215.948	0.3129
		BIC	-0.0037	0.7406	248.471	0.3394
	n=250	MIMSA	0.0021	0.9158	339.0515	0.4053
		AIC	0.0029	0.9115	333.146	0.4031
		BIC	0.0030	0.9309	351.1506	0.4127
	n=500	MIMSA	-0.0009	0.9669	433.0144	0.4214
		AIC	-0.0003	0.9717	442.6404	0.4236
		BIC	-0.0008	0.9789	448.7697	0.4268

Even though AIC finds the true lag order more frequently than MIMSA, it is clear in Table 3.8 that the latter results in more accurate forecasts.

Table 3.9: Frequency distribution of model identification for simulation of MA(2)

Parameter	Sample Size	Criterion	Candidate Model					
			MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
$\Theta_1 = 0.01$, $\Theta_2 = 0.85$	n=100	MIMSA	545	7411	1647	254	94	49
		AIC	0	6746	1223	725	621	685
		BIC	0	9256	430	114	88	112
	n=250	MIMSA	303	6742	1651	734	385	185
		AIC	0	7167	1302	712	440	379
		BIC	0	9739	220	35	6	0
	n=500	MIMSA	687	5197	1739	1110	699	568
		AIC	0	7405	1218	638	395	344
		BIC	0	9833	149	15	3	0
$\Theta_1 = 0.45$, $\Theta_2 = 0.5$	n=100	MIMSA	1669	7088	872	260	78	33
		AIC	10	6672	1179	714	624	801
		BIC	94	9184	394	116	83	129
	n=250	MIMSA	452	6737	1732	668	255	156
		AIC	0	7207	1240	710	471	372
		BIC	0	9712	251	33	2	2
	n=500	MIMSA	414	5069	2095	1161	748	513
		AIC	0	7333	1197	667	416	387
		BIC	0	9827	150	18	5	0
$\Theta_1 = -0.2$, $\Theta_2 = 0.9$	n=100	MIMSA	53	737	165	35	8	2
		AIC	0	658	141	74	64	63
		BIC	0	908	62	13	5	12
	n=250	MIMSA	243	6701	1693	807	372	184
		AIC	0	7323	1198	704	414	361
		BIC	0	9743	207	40	9	1
	n=500	MIMSA	574	5117	1769	1226	755	559
		AIC	0	7345	1237	622	417	379
		BIC	0	9835	150	13	2	0

Note: $Y_t = \Theta_1\varepsilon_{t-1} + \Theta_2\varepsilon_{t-2}$ *is generated from MA(2).*

Like MA(1) scenario, simulation results for an MA(2) process signify that MIMSA underperforms AIC and BIC in terms of identifying the underlying model order.

Table 3.10: Forecast measures for the simulation of MA(2) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Theta_1 = 0.01$, $\Theta_2 = 0.85$	n=100	MIMSA	-0.0013	0.7483	295.069	0.4431
		AIC	-0.0007	0.6909	278.9755	0.4083
		BIC	-0.0020	0.7617	299.5676	0.4516
	n=250	MIMSA	-0.0020	0.9515	366.790	0.5386
		AIC	-0.0008	0.9412	373.0537	-0.5325
		BIC	-0.0014	0.9674	363.0238	0.54799
	n=500	MIMSA	0.0010	1.0002	469.296	0.5543
		AIC	-0.00004	0.9873	481.4512	0.5471
		BIC	0.0003	0.9948	464.5949	0.5514
$\Theta_1 = 0.45$, $\Theta_2 = 0.5$	n=100	MIMSA	-0.0015	0.6908	310.9572	0.5781
		AIC	-0.0007	0.6015	277.0303	0.5000
		BIC	-0.0008	0.6701	328.7044	0.5609
	n=250	MIMSA	-0.0008	0.9058	723.5341	0.7419
		AIC	-0.0013	0.8943	1983.516	0.7324
		BIC	-0.0003	0.9231	1975.733	0.7565
	n=500	MIMSA	0.00002	0.9605	443.3993	0.7792
		AIC	-0.00009	0.9611	424.7086	0.7799
		BIC	0.0001	-0.0004	443.2716	0.7867
$\Theta_1 = -0.2$, $\Theta_2 = 0.9$	n=100	MIMSA	-0.0028	0.7826	255.2015	0.4057
		AIC	-0.0011	0.7272	285.9753	0.3750
		BIC	-0.0020	0.7927	278.5696	0.4097
	n=250	MIMSA	0.0007	0.9673	1085.01	0.4814
		AIC	0.0007	0.9599	1098.470	0.4779
		BIC	0.0016	0.9848	1089.584	0.4906
	n=500	MIMSA	-0.0006	1.0126	580.3695	0.4893
		AIC	-0.0004	1.0015	533.032	0.4842
		BIC	-0.0004	1.0089	537.3983	0.4878

Unlike MA(1) simulation scenario, weak forecast performance of MIMSA when compared to AIC and BIC might be concluded from Table 3.10.

Table 3.11: Frequency distribution of model identification for simulation of MA(3)

Parameter	Sample Size	Criterion	Candidate Model					
			MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
$\Theta_1 = 0.01$, $\Theta_2 = 0.01$, $\Theta_3 = 0.9$	n=100	MIMSA	123	453	382	34	7	1
		AIC	0	0	714	128	84	74
		BIC	0	0	920	48	20	12
	n=250	MIMSA	280	2863	4946	1220	499	192
		AIC	0	0	7437	1300	759	594
		BIC	0	0	9657	266	52	25
	n=500	MIMSA	1503	4523	1024	556	1576	818
		AIC	0	0	10000	0	0	0
		BIC	0	0	10000	0	0	0
$\Theta_1 = 0.01$, $\Theta_2 = 0.01$, $\Theta_3 = 0.5$	n=100	MIMSA	519	272	182	16	9	2
		AIC	4	665	144	85	102	0
		BIC	65	3	856	42	16	18
	n=250	MIMSA	879	1221	6420	1122	223	135
		AIC	0	0	7317	1299	791	593
		BIC	0	0	9729	231	28	12
	n=500	MIMSA	713	717	5361	1733	936	540
		AIC	0	0	7497	1243	728	532
		BIC	0	0	9835	148	15	2
$\Theta_1 = 0.01$, $\Theta_2 = 0.01$, $\Theta_3 = -0.875$	n=100	MIMSA	133	369	394	63	24	17
		AIC	0	0	679	137	107	77
		BIC	0	0	907	62	25	6
	n=250	MIMSA	0	3320	6671	9	0	0
		AIC	0	0	7096	1305	824	775
		BIC	0	0	9649	373	111	47
	n=500	MIMSA	749	2051	4216	1454	880	650
		AIC	0	0	732	1283	728	617
		BIC	0	0	9756	196	37	11

Note: $Y_t = \Theta_1\varepsilon_{t-1} + \Theta_2\varepsilon_{t-2} + \Theta_3\varepsilon_{t-3}$ *is generated from MA(3).*

If one makes a comparison between the results of Table 3.11 and the previous scenarios, it is apparent that its model identifiability worsens as the order of the MA process increases.

Table 3.12: Forecast measures for the simulation of MA(3) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Theta_1 = 0.01$, $\Theta_2 = 0.01$, $\Theta_3 = 0.9$	n=100	MIMSA	0.0076	0.7027	375.3105	0.3967
		AIC	0.0064	0.6107	380.944	0.3361
		BIC	0.0013	0.6416	388.493	0.3553
	n=250	MIMSA	-0.0020	1.0014	423.404	0.5473
		AIC	-0.0029	0.9382	412.080	0.5116
		BIC	-0.0033	0.9598	410.6569	0.5239
	n=500	MIMSA	0.0019	1.4465	385.6184	0.4932
		AIC	0.0020	0.9727	287.1509	0.3318
		BIC	0.0020	0.9727	287.1509	0.3318
$\Theta_1 = 0.01$, $\Theta_2 = 0.01$, $\Theta_3 = 0.5$	n=100	MIMSA	-0.0033	0.6876	254.461	0.4672
		AIC	0.0017	0.5102	218.3293	0.3344
		BIC	-0.0016	0.5581	234.588	0.3686
	n=250	MIMSA	-0.0001	0.8940	1333.280	0.5861
		AIC	-0.0007	0.8571	1342.256	0.5618
		BIC	-0.0002	0.8803	1337.994	0.5774
	n=500	MIMSA	-0.0004	0.9614	2075.991	0.6213
		AIC	-0.0006	0.9479	819.3995	0.6126
		BIC	-0.0004	0.9551	818.356	0.6173
$\Theta_1 = 0.01$, $\Theta_2 = 0.01$, $\Theta_3 = -0.875$	n=100	MIMSA	-0.0035	0.7440	305.291	0.4129
		AIC	-0.0013	0.5587	223.872	0.3047
		BIC	0	0.58025	233.4159	0.3199
	n=250	MIMSA	-0.0008	0.9615	466.518	0.5258
		AIC	0.0003	0.8616	432.334	0.4704
		BIC	0.0007	0.873	440.392	0.4774
	n=500	MIMSA	0.0004	1.0352	382.183	0.5599
		AIC	-0.0007	0.9568	617.553	0. 5177
		BIC	-0.0002	0.9623	621.395	0.5209

The same conclusion might be drawn that MIMSA produces less accurate forecasts as the generated series is governed by an MA process with a greater order.

In order to monitor how MIMSA performs from a various perspective, a diverse set of simulation scenarios for non-stationary ARIMA processes have been generated. By doing so, we aim to observe the effect of integrated segment on MIMSA's performance.

Table 3.13: Frequency distribution of model identification for simulation of ARIMA(0,1,1)

Parameter	Sample Size	Criterion	Candidate Model								
			ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (0,1,3)	ARIMA (0,2,1)	ARIMA (0,2,2)	ARIMA (0,2,3)	ARIMA (0,3,1)	ARIMA (0,3,2)	ARIMA (0,3,3)
$\Theta = 0.9$	n=100	MIMSA	57	0	0	379	0	0	559	5	0
		AIC	513	106	91	0	214	66	0	0	10
		BIC	815	36	10	0	120	15	0	0	4
	n=250	MIMSA	664	16	0	4215	16	0	4998	90	1
		AIC	6638	1077	802	0	1202	278	0	0	3
		BIC	9359	190	39	0	385	25	0	0	2
	n=500	MIMSA	665	8	2	4285	10	0	4952	75	5
		AIC	7132	1196	783	0	714	175	0	0	0
		BIC	9638	145	15	0	198	4	0	0	0
$\Theta = 0.5$	n=100	MIMSA	56	0	0	315	1	0	616	12	0
		AIC	640	104	77	2	147	26	0	0	4
		BIC	878	27	10	5	74	5	0	0	1
	n=250	MIMSA	683	9	0	3652	20	0	5510	119	7
		AIC	6967	1199	809	0	820	203	0	0	2
		BIC	9490	207	22	0	265	16	0	0	0
	n=500	MIMSA	647	6	0	3804	15	1	5392	133	2
		AIC	7325	1177	793	0	583	121	0	0	1
		BIC	9728	118	13	0	138	3	0	0	0

Table 3.13 emphasizes that 6 scenarios for different parameter values and sample sizes are generated from ARIMA (0,1,1) and 9 candidates are evaluated to mimic this process. In this scenario, one needs to differentiate the series of interest once to achieve order selection. For instance, when $\Theta = 0.9$ and n=100, MIMSA misidentifies the process as ARIMA(0,3,1) with the highest frequency in the above table. Although the frequency for identifying the true model is lower than AR or MA cases for AIC and BIC, they still have the ability to specify correctly.

Table 3.14: Forecast measures for the simulation of ARIMA(0,1,1) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Theta = 0.9$	n=100	MIMSA	-0.0129	0.8006	7.3356	0.39230
		AIC	0.0011	0.7422	8.1304	0.3628
		BIC	0.0010	0.7937	9.8609	0.3948
	n=250	MIMSA	-0.0007	0.9348	4.7208	0.4448
		AIC	0.00003	0.9098	4.7164	0.4327
		BIC	0.0005	0.9275	4.9129	0.4448
	n=500	MIMSA	-0.0001	1.3081	32.1214	0.9592
		AIC	0.0003	0.9999	24.1035	0.74158
		BIC	0.0002	1.0024	24.0525	0.7441
$\Theta = 0.5$	n=100	MIMSA	0.0021	1.0570	33.0634	0.8518
		AIC	0.0047	0.8420	32.9582	0.7273
		BIC	0.0051	0.8613	33.5609	0.7504
	n=250	MIMSA	0.0019	1.1605	34.8439	1.0040
		AIC	0.0012	0.9439	30.3047	0.8401
		BIC	0.0003	0.9505	30.4947	0.8474
	n=500	MIMSA	-0.0015	1.1879	78.5779	1.0430
		AIC	0.0017	0.9763	48.2332	0.87028
		BIC	0.0015	0.9783	48.1955	0.8727

The above table, 3.14, might be useful to compare the forecast accuracy of the corresponding order selection methods. This table means that forecasts of MIMSA are the least faultless. However, they are better than the one for BIC when MAPE and MASE are taken into account.

Table 3.15: Frequency distribution of model identification for simulation of ARIMA(0,1,2)

Parameter	Sample Size	Criterion	Candidate Model								
			ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (0,1,3)	ARIMA (0,2,1)	ARIMA (0,2,2)	ARIMA (0,2,3)	ARIMA (0,3,1)	ARIMA (0,3,2)	ARIMA (0,3,3)
$\Theta_1 = 0.01$, $\Theta_2 = 0.9$	n=100	MIMSA	56	2	0	387	0	0	552	3	0
		AIC	0	606	144	0	0	250	0	0	0
		BIC	0	849	39	0	0	112	0	0	0
	n=250	MIMSA	741	15	0	3892	61	0	5164	123	4
		AIC	0	6941	1396	0	0	1663	0	0	0
		BIC	0	9291	201	0	0	508	0	0	0
	n=500	MIMSA	449	5	0	4141	25	0	5290	90	0
		AIC	0	7548	1447	0	0	1005	0	0	0
		BIC	0	9612	139	0	0	249	0	0	0
$\Theta_1 = 0.01$, $\Theta_2 = 0.5$	n=100	MIMSA	59	0	0	483	1	0	454	3	0
		AIC	0	676	145	4	2	173	0	0	0
		BIC	6	876	43	20	2	53	0	0	0
	n=250	MIMSA	717	6	0	4712	32	0	4452	80	1
		AIC	0	7433	1460	0	0	1107	0	0	0
		BIC	0	9481	204	0	0	315	0	0	0
	n=500	MIMSA	535	0	0	4880	16	0	4519	50	0
		AIC	0	7717	1531	0	0	752	0	0	0
		BIC	0	9713	142	0	0	145	0	0	0

A very similar conclusion might be reasonable to be drawn that MIMSA underperforms its counterparts in order identification. Although it is not reasonable to produce a non-stationary series that requires double-differentiating, we include those simulations in our study in order to examine MIMSA from various aspects. Also, candidate models that have the order of up to 3 for the integrated segment are included in the candidate set to ensure a systematic order selection procedure. In dealing with the real data applications, order selection is achieved after implementing stationarity tests like Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test, so the value of d parameter is already determined at the beginning.

Table 3.16: Forecast measures for the simulation of ARIMA(0,1,2) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Theta_1 = 0.01$, $\Theta_2 = 0.9$	n=100	MIMSA	-0.0161	1.5104	42.6072	1.0032
		AIC	-0.0062	0.9050	33.3386	0.6362
		BIC	-0.0107	0.9192	36.2102	0.6565
	n=250	MIMSA	0.0028	1.6215	45.1332	1.1618
		AIC	-0.0015	1.0035	25.8382	0.7402
		BIC	-0.0016	1.0049	27.1889	0.7440
	n=500	MIMSA	-0.0009	1.6636	46.0241	1.2142
		AIC	0.0008	1.0147	25.0644	0.7523
		BIC	0.0011	1.0150	24.2498	0.7532
$\Theta_1 = 0.01$, $\Theta_2 = 0.5$	n=100	MIMSA	-0.0232	1.0641	44.1885	0.8221
		AIC	-0.0056	0.8050	44.7182	0.6518
		BIC	-0.0082	0.8207	44.1994	0.6714
	n=250	MIMSA	-0.0008	1.1670	28.0307	0.9619
		AIC	-0.0007	0.9288	22.4847	0.7829
		BIC	-0.0003	0.9325	23.0729	0.7872
	n=500	MIMSA	0.0022	1.1969	28.5652	0.9970
		AIC	-0.0002	0.9673	24.2529	0.8162
		BIC	-0.0007	0.9683	24.4511	0.8175

Table 3.16 points out that MIMSA has the least precise forecasts for the future observations when compared to AIC and BIC.

Table 3.17: Frequency distribution of model identification for simulation of ARIMA(0,2,1)

Parameter	Sample Size	Criterion	Candidate Model								
			ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (0,1,3)	ARIMA (0,2,1)	ARIMA (0,2,2)	ARIMA (0,2,3)	ARIMA (0,3,1)	ARIMA (0,3,2)	ARIMA (0,3,3)
$\Theta = 0.9$	n=100	MIMSA	153	0	0	552	0	0	295	0	0
		AIC	0	0	1	521	93	74	0	254	57
		BIC	0	0	1	810	36	7	0	138	8
	n=250	MIMSA	503	0	0	965	0	0	8530	2	0
		AIC	0	0	0	6647	1073	756	0	1255	269
		BIC	0	0	0	9347	196	22	0	415	20
	n=500	MIMSA	1387	1	0	4275	0	0	4337	0	0
		AIC	0	0	0	7085	1237	787	0	724	167
		BIC	0	0	0	9658	146	9	0	183	4
$\Theta = 0.5$	n=100	MIMSA	144	0	0	509	0	0	347	0	0
		AIC	0	0	1	680	107	74	3	111	24
		BIC	0	0	0	896	34	13	7	48	2
	n=250	MIMSA	544	0	0	950	1	0	8503	2	0
		AIC	0	0	0	7092	1195	749	0	767	197
		BIC	0	0	0	9528	195	36	0	228	13
	n=500	MIMSA	1382	0	0	4355	0	0	4263	0	0
		AIC	0	0	0	7353	1210	787	0	534	116
		BIC	0	0	0	9760	116	6	0	116	2

When the order of integration inclines to 2, MIMSA's performance to identify the true lag order improves. Although it accomplishes it with the smallest frequency among BIC and AIC, MIMSA is able to determine the true model.

Table 3.18: Forecast measures for the simulation of ARIMA(0,2,1) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Theta = 0.9$	n=100	MIMSA	-0.0129	0.8006	7.3356	0.3923
		AIC	0.0011	0.7422	8.1304	0.3628
		BIC	0.0010	0.7937	9.8609	0.3948
	n=250	MIMSA	-0.0007	0.9348	4.7208	0.4448
		AIC	0.00005	0.9098	4.7164	0.4327
		BIC	0.0005	0.9275	4.9129	0.4448
	n=500	MIMSA	0.0420	2.1420	0.5706	0.1137
		AIC	-0.0012	0.9890	0.1059	0.0509
		BIC	-0.0017	0.9917	0.1063	0.0511
$\Theta = 0.5$	n=100	MIMSA	0.0443	1.1979	1.1200	0.1600
		AIC	-0.0083	0.7714	0.4466	0.1058
		BIC	-0.0181	0.7950	0.4664	0.1102
	n=250	MIMSA	-0.0028	1.3013	0.5922	0.1114
		AIC	-0.0003	0.9252	0.2066	0.0830
		BIC	-0.0002	0.9311	0.2076	0.0837
	n=500	MIMSA	-0.0339	1.7639	0.42001	0.1211
		AIC	-0.0037	0.9637	0.2262	0.0619
		BIC	-0.0042	0.9659	0.2267	0.0621

When Table 3.18 is analyzed, it might be inferred that forecast performance of MIMSA is more acceptable when $\Theta = 0.9$.

Table 3.19: Frequency distribution of model identification for simulation of ARIMA(0,2,2)

Parameter	Sample Size	Criterion	Candidate Model								
			ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (0,1,3)	ARIMA (0,2,1)	ARIMA (0,2,2)	ARIMA (0,2,3)	ARIMA (0,3,1)	ARIMA (0,3,2)	ARIMA (0,3,3)
$\Theta_1 = 0.01$, $\Theta_2 = 0.9$	n=100	MIMSA	140	0	0	564	0	0	296	0	0
		AIC	0	0	0	0	560	133	0	0	307
		BIC	0	0	0	0	806	36	0	0	158
	n=250	MIMSA	498	1	0	1045	0	0	8456	0	0
		AIC	0	0	0	0	6922	1400	0	0	1678
		BIC	0	0	0	0	9237	174	0	0	589
	n=500	MIMSA	1422	0	0	4322	1	0	4254	1	0
		AIC	0	0	0	0	7545	1416	0	0	1039
		BIC	0	0	0	0	9642	134	0	0	224
$\Theta_1 = 0.01$, $\Theta_2 = 0.5$	n=100	MIMSA	166	0	0	590	0	0	244	0	0
		AIC	0	0	0	1	643	180	3	2	171
		BIC	0	0	0	11	862	46	15	1	65
	n=250	MIMSA	574	0	0	1078	0	0	8347	1	0
		AIC	0	0	0	0	7301	1528	0	0	1171
		BIC	0	0	0	0	9471	216	0	0	313
	n=500	MIMSA	1334	0	0	4007	0	0	4658	1	0
		AIC	0	0	0	0	7713	1476	0	0	811
		BIC	0	0	0	0	9713	137	0	0	150

One might discover that MIMSA's frequency of finding the true model gets increased when the sample size goes up. This conclusion is valid for both values of Θ_2.

Table 3.20: Forecast measures for the simulation of ARIMA(0,2,2) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Theta_1 = 0.01$, $\Theta_2 = 0.9$	n=100	MIMSA	0.0648	1.6123	1.0807	0.1668
		AIC	0.0146	0.8378	0.4270	0.0880
		BIC	0.0129	0.8498	0.4381	0.0903
	n=250	MIMSA	0.0216	1.6008	0.6824	0.1100
		AIC	-0.0024	0.9819	0.1814	0.0708
		BIC	-0.0011	0.9834	0.1828	0.1828
	n=500	MIMSA	0.0138	2.4707	1.3649	0.1267
		AIC	-0.0025	1.0026	0.0962	0.0513
		BIC	-0.0028	1.0023	0.0976	0.0514
$\Theta_1 = 0.01$, $\Theta_2 = 0.5$	n=100	MIMSA	-0.0135	1.3988	3.5001	0.1602
		AIC	-0.0099	0.7136	0.3140	0.0791
		BIC	-0.0104	0.7321	0.3187	0.0820
	n=250	MIMSA	-0.0142	1.4181	0.8650	0.0969
		AIC	-0.0026	0.9083	0.1632	0.0650
		BIC	-0.0028	0.9128	0.1637	0.0654
	n=500	MIMSA	0.0376	2.1047	0.3097	0.1112
		AIC	-0.0007	0.9593	0.0645	0.0478
		BIC	$5.9x10^{-6}$	0.9604	0.0657	0.0479

Table 3.20 might be informative when examining the forecast performance of MIMSA, AIC and BIC when the underlying series is from MA(2) process and needs to be differenced twice. It might be derived that MIMSA's forecasts improve towards the ones ones for AIC and BIC.

Table 3.21: Frequency distribution of model identification for simulation of ARIMA(1,1,0)

Parameter	Sample Size	Criterion	Candidate Model								
			ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (3,1,0)	ARIMA (1,2,0)	ARIMA (2,2,0)	ARIMA (3,2,0)	ARIMA (1,3,0)	ARIMA (2,3,0)	ARIMA (3,3,0)
$\Phi = 0.9$	n=100	MIMSA	312	0	0	623	0	0	65	0	0
		AIC	499	95	61	178	99	68	0	0	0
		BIC	639	25	6	263	52	15	0	0	0
	n=250	MIMSA	2649	0	0	7196	0	0	155	0	0
		AIC	6884	1625	1296	14	36	145	0	0	0
		BIC	9447	285	56	159	47	6	0	0	0
	n=500	MIMSA	2676	0	0	7311	0	0	13	0	0
		AIC	6930	1801	1268	0	0	1	0	0	0
		BIC	9651	307	42	0	0	0	0	0	0
$\Phi = 0.5$	n=100	MIMSA	55	0	0	242	0	0	703	0	0
		AIC	698	175	123	0	0	4	0	0	0
		BIC	909	72	14	0	3	2	0	0	0
	n=250	MIMSA	319	0	0	2271	0	0	7410	0	0
		AIC	7003	1741	1256	0	0	0	0	0	0
		BIC	9495	444	61	0	0	0	0	0	0
	n=500	MIMSA	373	0	0	1574	0	0	0	0	0
		AIC	6945	1772	1283	0	0	0	8053	0	0
		BIC	9676	285	39	0	0	0	0	0	0

One might make an inference that MIMSA is not able to determine the exact process that governs the simulated series unlike BIC and AIC.

Table 3.22: Forecast measures for the simulation of ARIMA(1,1,0) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Phi = 0.9$	n=100	MIMSA	-0.0129	0.8006	7.3356	0.3923
		AIC	0.0011	0.7422	8.1304	0.3628
		BIC	0.0010	0.7937	9.8609	0.3948
	n=250	MIMSA	-0.0007	0.9348	4.7208	0.4448
		AIC	0.00003	0.9098	4.7164	0.4327
		BIC	0.0005	0.9275	4.9129	0.4448
	n=500	MIMSA	-0.0003	0.9753	3.4173	0.4511
		AIC	0.0012	0.9560	3.3189	0.4420
		BIC	0.0011	0.9641	3.3662	0.4476
$\Phi = 0.5$	n=100	MIMSA	-0.0026	0.9986	18.9037	0.7531
		AIC	-0.0084	0.7623	16.3353	0.6108
		BIC	-0.0087	0.8042	20.7482	0.6553
	n=250	MIMSA	0.0006	1.2623	42.3275	1.0412
		AIC	-0.0022	0.9111	32.6757	0.7654
		BIC	-0.0024	0.9270	33.0842	0.7848
	n=500	MIMSA	0.00007	1.3652	22.5446	1.1554
		AIC	-0.0004	0.9557	15.8619	0.8163
		BIC	-0.0003	0.9640	16.1711	0.8268

It might be inferrible from Table 3.22 that MIMSA has closer forecasts to the ones for AIC and BIC when $\Phi = 0.9$.

Table 3.23: Frequency distribution of model identification for simulation of ARIMA(1,2,0)

Parameter	Sample Size	Criterion	Candidate Model								
			ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (3,1,0)	ARIMA (1,2,0)	ARIMA (2,2,0)	ARIMA (3,2,0)	ARIMA (1,3,0)	ARIMA (2,3,0)	ARIMA (3,3,0)
$\Phi = 0.9$	n=100	MIMSA	688	0	0	12	0	0	300	0	0
		AIC	0	24	6	489	77	62	166	84	92
		BIC	0	9	0	640	15	6	284	36	10
	n=250	MIMSA	194	3	0	4682	1	0	5118	2	0
		AIC	0	260	84	6621	1654	1170	16	41	154
		BIC	0	49	4	9338	360	54	147	32	16
	n=500	MIMSA	3850	1	0	1092	0	0	5057	0	0
		AIC	0	276	78	6720	1680	1244	0	0	0
		BIC	0	44	2	9627	283	44	0	0	2
$\Phi = 0.5$	n=100	MIMSA	907	0	0	29	0	0	64	0	0
		AIC	0	26	5	639	179	138	0	0	13
		BIC	0	10	1	904	67	18	0	0	0
	n=250	MIMSA	418	1	0	5711	0	0	3870	0	0
		AIC	0	218	85	6736	1727	1234	0	0	0
		BIC	0	41	4	9477	402	76	0	0	0
	n=500	MIMSA	4525	0	0	1116	1	0	4357	1	0
		AIC	0	239	62	6764	1685	1250	0	0	0
		BIC	0	32	2	9657	266	43	0	0	0

Table 3.23 leads one to conclude that MIMSA might not be capable of finding the true lag order for the integration segment in a Box-Jenkins model as opposed to AIC and BIC.

Table 3.24: Forecast measures for the simulation of ARIMA(1,2,0) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Phi = 0.9$	n=100	MIMSA	-0.0129	0.8006	7.3356	0.3923
		AIC	0.0011	0.7422	8.1304	0.3628
		BIC	0.0010	0.7937	9.8609	0.3948
	n=250	MIMSA	-0.0007	0.9348	4.7208	0.4448
		AIC	0.00003	0.9098	4.7164	0.4327
		BIC	0.0005	0.9275	4.9129	0.4448
	n=500	MIMSA	0.0028	1.3081	0.0234	0.0145
		AIC	-0.0011	0.9448	0.0160	0.0099
		BIC	-0.0010	0.9532	0.0165	0.0100
$\Phi = 0.5$	n=100	MIMSA	0.0047	0.8372	0.4552	0.0903
		AIC	-0.0005	0.6906	0.3204	0.0673
		BIC	0.0005	0.7480	0.3673	0.0741
	n=250	MIMSA	-0.0034	0.9458	0.2238	0.0632
		AIC	-0.0055	0.8881	0.1690	0.0594
		BIC	-0.0054	0.9050	0.18256	0.0611
	n=500	MIMSA	-0.0006	1.0635	0.0899	0.0514
		AIC	-0.0020	0.9456	0.0806	0.0452
		BIC	-0.0019	0.9540	0.0794	0.0458

Unlike the previous scenarios, AIC might not outperform BIC and MIMSA in terms of forecast accuracy.

Table 3.25: Frequency distribution of model identification for simulation of ARIMA(2,1,0)

Parameter	Sample Size	Criterion	Candidate Model								
			ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (3,1,0)	ARIMA (1,2,0)	ARIMA (2,2,0)	ARIMA (3,2,0)	ARIMA (1,3,0)	ARIMA (2,3,0)	ARIMA (3,3,0)
$\Phi_1 = 0.01$, $\Phi_2 = 0.9$	n=100	MIMSA	416	0	0	476	0	0	108	0	0
		AIC	0	124	40	582	143	111	0	0	0
		BIC	0	47	3	875	61	14	0	0	0
	n=250	MIMSA	4040	3	0	4985	3	0	967	2	0
		AIC	0	4679	1247	2518	708	848	0	0	0
		BIC	0	1896	89	7545	386	84	0	0	0
	n=500	MIMSA	4150	0	0	5260	2	0	588	0	0
		AIC	0	7599	2137	49	26	189	0	0	0
		BIC	0	7226	202	2432	105	35	0	0	0
$\Phi_1 = 0.01$, $\Phi_2 = 0.5$	n=100	MIMSA	316	0	0	602	0	0	82	0	0
		AIC	1	191	49	505	144	110	0	0	0
		BIC	6	82	10	815	68	19	0	0	0
	n=250	MIMSA	3032	5	0	6301	3	0	657	2	0
		AIC	0	6543	1848	741	298	570	0	0	0
		BIC	0	4186	183	5249	300	82	0	0	0
	n=500	MIMSA	3044	1	0	6589	4	0	362	0	0
		AIC	0	7704	2238	0	1	57	0	0	0
		BIC	0	9446	254	0	22	18	0	0	0

The above table, 3.25, might be consulted to summarize that MIMSA may not accomplish determining the true lag order when the simulated process is dominated by ARIMA(2,1,0).

Table 3.26: Forecast measures for the simulation of ARIMA(2,1,0) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Phi_1 = 0.01$, $\Phi_2 = 0.9$	n=100	MIMSA	0.0071	1.0421	4.4799	0.4661
		AIC	0.0097	0.6776	3.0219	0.3106
		BIC	0.0071	0.7298	3.3366	0.3402
	n=250	MIMSA	0.0015	1.2795	5.5758	0.5789
		AIC	-0.0004	0.8807	3.9090	0.4052
		BIC	0.0003	0.9063	4.0219	0.4208
	n=500	MIMSA	-0.0029	1.3840	4.4399	0.6218
		AIC	-0.0011	0.9396	3.2497	0.4288
		BIC	-0.0011	0.9495	3.3506	0.4344
$\Phi_1 = 0.01$, $\Phi_2 = 0.5$	n=100	MIMSA	-0.0044	0.8365	4.0367	0.4756
		AIC	-0.0136	0.6911	3.3400	0.3820
		BIC	-0.0112	0.7500	3.9073	0.4236
	n=250	MIMSA	0.0008	0.9804	4.4208	0.5615
		AIC	0.0002	0.8797	3.8745	0.4954
		BIC	-0.0004	0.9065	3.9993	0.5159
	n=500	MIMSA	-0.0008	1.0181	3.6934	0.5653
		AIC	0.0010	0.9380	3.4431	0.5167
		BIC	0.0010	0.9430	3.4713	0.5205

Table 3.26 might highlight that MIMSA has the weakest forecast performance when compared to AIC and BIC.

Table 3.27: Frequency distribution of model identification for simulation of ARIMA(2,2,0)

Parameter	Sample Size	Criterion	Candidate Model								
			ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (3,1,0)	ARIMA (1,2,0)	ARIMA (2,2,0)	ARIMA (3,2,0)	ARIMA (1,3,0)	ARIMA (2,3,0)	ARIMA (3,3,0)
$\Phi_1 = 0.01$, $\Phi_2 = 0.9$	n=100	MIMSA	546	0	0	140	0	0	314	0	0
		AIC	0	0	12	0	115	43	561	147	122
		BIC	0	0	3	0	50	4	862	61	20
	n=250	MIMSA	217	0	0	4635	0	0	5148	0	0
		AIC	0	0	274	0	4482	1258	2380	791	815
		BIC	0	0	30	0	1840	93	7582	370	85
	n=500	MIMSA	3298	0	0	1905	0	0	4795	2	0
		AIC	0	0	339	0	7268	2167	35	28	163
		BIC	0	0	37	0	7147	191	2479	125	21
$\Phi_1 = 0.01$, $\Phi_2 = 0.5$	n=100	MIMSA	696	0	0	86	0	0	218	0	0
		AIC	0	0	18	0	201	67	457	138	119
		BIC	0	0	2	3	76	9	834	60	16
	n=250	MIMSA	197	1	0	4613	0	0	5188	1	0
		AIC	0	0	290	0	6327	1785	750	303	545
		BIC	0	0	40	0	4212	171	5166	330	81
	n=500	MIMSA	3928	1	0	1710	0	0	4361	0	0
		AIC	0	0	318	0	7456	2177	0	1	48
		BIC	0	0	35	0	9437	246	243	27	12

If one investigates Table 3.27, it might be summed up that MIMSA misdetermines the exact lag order when a series is a reflection from an ARIMA (2,2,0) process.

Table 3.28: Forecast measures for the simulation of ARIMA(2,2,0) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Phi_1 = 0.01$, $\Phi_2 = 0.9$	n=100	MIMSA	-0.0072	1.1120	0.1410	0.0347
		AIC	-0.0112	0.5958	0.04915	0.0156
		BIC	-0.0125	0.6592	0.0631	0.0177
	n=250	MIMSA	0.0064	1.2817	0.0622	0.0203
		AIC	0.0004	0.8586	0.0260	0.0129
		BIC	0.0005	0.8870	0.0268	0.0135
	n=500	MIMSA	-0.0075	1.4388	0.0193	0.0152
		AIC	-0.0020	0.9269	0.0100	0.0093
		BIC	-0.0025	0.9375	0.0101	0.0095
$\Phi_1 = 0.01$, $\Phi_2 = 0.5$	n=100	MIMSA	-0.0156	0.8352	0.0943	0.0300
		AIC	-0.0064	0.6034	0.0682	0.0151
		BIC	-0.0071	0.6760	0.0745	0.0177
	n=250	MIMSA	-0.0004	0.9601	0.1009	0.0161
		AIC	0.0012	0.8549	0.1234	0.0138
		BIC	0.0007	0.8817	0.0949	0.0143
	n=500	MIMSA	0.0020	1.1837	0.0190	0.0133
		AIC	-0.0017	0.9283	0.0119	0.0098
		BIC	-0.0016	0.9338	0.0120	0.0099

It might be observable from Table 3.28 that MIMSA's performance for providing accurate forecasts for unobserved future values is not well when a series is simulated from an ARIMA(2,2,0) process.

CHAPTER 4

WEIGHTED MODEL SELECTION WITH MIMSA

As it is concluded in Chapter 3, MIMSA has a weak forecast performance even though its capability of determining the true lag order is satisfactory. Its both forecast and order identification performances should be improved, especially, for ARIMA models. In the model selection framework, to increase the performance of a single model selection procedure, utilizing a combination of that method with its counterparts is a common approach. This chapter covers our manner to implement weighted model selection with MIMSA.

4.1 Weighted Model Selection

In the literature, there are some researches stating that using the weighted sum of the existing model selection criteria as a new tool of selecting the best among candidate models might be used. Wu and Sepulveda (1998) proposed Weighted Average Information Criterion for order selection in time series. They claim that the weighted combination of AIC_C and BIC, where weights are determined by the length of series of interest, has a better performance than the individual criterion. Besides, Ayalew et al. (2012) summarized that taking median of any two criteria outperforms individual performance in model selection. Wagenmakers and Farrell (2004) suggests Akaike weights for the sake of interpretability over raw AIC. In fact, pooling in model selection is offered to increase forecast performance (Kuzin & Schumacher, 2013). Building on Weighted Information Criterion (Eğrioğlu et al., 2008), Aladağ et al. (2010) have developed Adaptive Weighted Information Criterion for ANN. While the former requires to take a weighted average of several model selection criteria with fixed weights, the latter figures the corresponding weights by optimization.

4.2 Weighted Model Selection With MIMSA

With the aim of enriching the forecast performance of MIMSA, our motivation stands for the idea of combining it with an information criterion and error based forecast measure. The corresponding methods to combine with MIMSA are decided to be MAPE and BIC. To make this decision, attractive properties of BIC such as consistency and propensity of evading overfitting are influential.We also considered that other forecast measures might cover a smaller scale than MIMSA and preferred MAPE. In this way, we take dual and triple combinations of each method into consideration and name them WM1, WM2, WM3 and WM4 provided below:

$$WM1 = w_{11} * MIMSA + w_{12} * MAPE \tag{4.1}$$

$$WM2 = w_{21} * MIMSA + w_{22} * BIC \tag{4.2}$$

$$WM3 = w_{31} * MAPE + w_{32} * BIC \tag{4.3}$$

$$WM4 = w_{41} * MIMSA + w_{42} * MAPE + w_{43} * BIC \tag{4.4}$$

Here, we should emphasize that we transform MIMSA by computing its negative version and setting the rule of selection through minimizing in order to make it conformable with MAPE and BIC. Therefore, a candidate model with the minimal WM value should be selected as the best possible model.

We assign the weights in each combination by Genetic Algorithms (GA). GA were proposed by Holland in the early 1960's and are inspired from evolutionary theory. In these algorithms, a starting value is determined and it is expected to evolve limited to a pre-determined search space. Evolution is achieved according to crossover and mutation probabilities as in the natural selection. Therefore, the optimum solution in that search space is found. We prefer GA because of some attractive properties like efficiency, being easily applicable and requiring no more assumptions as compared to the other optimization methods (Sivanandam & Deepa, 2008 and Scrucca, 2013).
As a result, Weighted Model Selection with MIMSA is implemented by following the below steps:

Step 1- Determine candidate models,

Step 2- Calculate BIC, $MAPE$ and $MIMSA$ for each candidate,

Step 3- Assign the weights by GA and calculate $WM1_i$-$WM4_i$ for each candidate,

Step 4- Propose the candidate with the lowest $WM-_i$ as the final model.

4.3 Simulation Studies

To monitor the performance of weighted model selection with MIMSA, similar simulation studies as in the previous chapter that mimic various scenarios for nonseasonal Box-Jenkins models up to lag order 3 are created.
Any scenario is implemented on the below algorithm:

1. Generate a series of size n from normal distribution and assign the first 90% and last 10% of the observations to two sets: train and test, respectively,
2. Determine candidate models for train series and calculate BIC, $MAPE$ and $MIMSA$ for each candidate,
3. Assign the weights by GA and calculate $WM1_i$-$WM4_i$ for each candidate,
4. Propose the candidate with the lowest $WM-_i$ as the final model,

5. Based on the proposed model for each combination, obtain the predicted values for the test set and calculate forecast measures (ME, RMSE, MAPE, etc.)

6. Repeat steps (1)-(5) $N = 1,000$ times,

7. Summarize N outputs to make a comparison between the combinations of model selection methods as well as the individual ones.

For each simulation scenario, a similar approach with the previous chapter has been applied when determining the candidate models. Since the computational time required to apply for GA in any procedure takes a long time, the number of simulation studies for this algorithm is fixed as $N = 1,000$.

Table 4.1: Frequency distribution of model identification for simulation of AR(1); $\Phi = 0.85$

Sample Size	Criterion	Candidate Model					
		AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
n=100	MIMSA	1000	0	0	0	0	0
	BIC	899	70	17	8	3	3
	MAPE	97	69	102	123	146	463
	WM1	176	177	174	164	166	143
	WM2	175	195	153	169	157	151
	WM3	170	130	146	175	176	203
	WM4	172	190	170	156	161	151
n=250	MIMSA	1000	0	0	0	0	0
	BIC	941	51	8	0	0	0
	MAPE	144	102	91	111	163	389
	WM1	198	162	167	150	169	154
	WM2	171	172	164	147	182	164
	WM3	153	144	164	173	190	176
	WM4	177	162	170	162	154	175
n=500	MIMSA	999	1	0	0	0	0
	BIC	970	25	5	0	0	0
	MAPE	150	100	107	124	152	367
	WM1	177	164	172	166	166	155
	WM2	169	189	149	154	164	175
	WM3	160	176	163	152	169	180
	WM4	160	158	169	180	162	171

Note: WM1-MIMSA&MAPE, WM2-MIMSA&BIC, WM3-MAPE&BIC and WM4-MIMSA&MAPE&BIC.

Table 4.1 compares the relevant order selection methods when the AR parameter is $\Phi = 0.85$. If one observes this table to derive a conclusion, it might be clear that MAPE underperforms MIMSA and BIC in terms of identifying the true lag order. In addition, weighted combinations of the corresponding order selection methods have a significantly weaker performance when compared to the individual methods.

If Table 4.2 is taken as a reference to make a forecast accuracy comparison, one might summarize that MAPE outperforms the rest of the order selection methods. On the other hand, it may be inferred that there is a meaningful difference between the forecast measures of the single and weighted methods. It might be valid to state that using weighted combinations might increase forecast performance.

Table 4.2: Forecast measures for the simulation of AR(1) process; $\Phi = 0.85$

Sample Size	Selection Criterion	Forecast Measure ME	RMSE	MAPE	MASE
	MIMSA	0	0.7642	243.2161	0.7239
	BIC	0	0.7402	228.3848	0.6981
	MAPE	0	0.2026	47.2993	0.1799
n=100	WM1	0	0.3740	118.762	0.3331
	WM2	0	0.3745	105.0836	0.3332
	WM3	0	0.3314	70.4031	0.2965
	WM4	0	0.3755	124.5702	0.3335
	MIMSA	0	0.9101	286.6067	0.8718
	BIC	0	0.9073	285.0416	0.8681
	MAPE	0	0.7606	233.2534	0.6812
n=250	WM1	0	0.8015	253.0716	0.7310
	WM2	0	0.7963	265.1447	0.7248
	WM3	0	0.7874	242.4143	0.7157
	WM4	0	0.7954	253.6655	0.7240
	MIMSA	0	0.9622	329.9861	0.9216
	BIC	0	0.9616	330.3678	0.9208
	MAPE	0	0.8965	309.5652	0.8315
n=500	WM1	0	0.9105	304.7059	0.8506
	WM2	0	0.9096	330.0806	0.8497
	WM3	0	0.9090	311.8513	0.8489
	WM4	0	0.9088	306.8942	0.8487

Note: WM1-MIMSA&MAPE, WM2-MIMSA&BIC, WM3-MAPE&BIC and WM4-MIMSA&MAPE&BIC.

Table 4.3: Frequency distribution of model identification for simulation of AR(1); $\Phi = 0.5$

Sample Size	Criterion	Candidate Model					
		AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
n=100	MIMSA	999	1	0	0	0	0
	BIC	890	82	17	8	2	1
	MAPE	155	126	109	113	147	350
	WM1	164	170	170	167	177	152
	WM2	165	163	183	175	164	150
	WM3	182	169	160	164	165	160
	WM4	179	173	164	177	162	145
n=250	MIMSA	982	18	0	0	0	0
	BIC	954	35	7	2	2	0
	MAPE	176	116	112	126	151	319
	WM1	177	161	160	189	151	162
	WM2	168	158	172	163	180	159
	WM3	165	172	175	163	163	162
	WM4	154	180	197	153	146	170
n=500	MIMSA	999	1	0	0	0	0
	BIC	961	34	5	0	0	0
	MAPE	226	134	124	104	153	259
	WM1	157	187	169	165	148	174
	WM2	159	183	182	165	168	143
	WM3	159	182	183	148	161	167
	WM4	185	157	153	180	154	171

When $\Phi = 0.5$, it might be concluded that MAPE is not able to specify the true model order correctly.

Table 4.4: Forecast measures for the simulation of AR(1) process; $\Phi = 0.5$

Sample Size	Selection Criterion	ME	Forecast Measure RMSE	MAPE	MASE
n=100	MIMSA	0	0.8027	292.6825	0.6872
	BIC	0	0.7774	285.031	0.6606
	MAPE	0	0.3015	110.0363	0.2441
	WM1	0	0.3770	157.5036	0.3079
	WM2	0	0.3820	147.1109	0.3054
	WM3	0	0.3952	177.8965	0.3152
	WM4	0	0.3961	170.3807	0.3188
n=250	MIMSA	0	0.9214	261.3059	0.7972
	BIC	0	0.9190	262.588	0.7946
	MAPE	0	0.7865	232.8813	0.6400
	WM1	0	0.8099	245.4801	0.6666
	WM2	0	0.8053	257.6113	0.6628
	WM3	0	0.8076	265.437	0.6644
	WM4	0	0.8074	238.8543	0.6633
n=500	MIMSA	0	0.9609	289.7241	0.8311
	BIC	0	0.9598	285.2334	0.8299
	MAPE	0	0.9055	296.6502	0.7629
	WM1	0	0.9084	284.1525	0.7665
	WM2	0	0.9098	288.2789	0.7681
	WM3	0	0.9085	303.1532	0.7664
	WM4	0	0.9091	286.408	0.7673

Although its order identification performance is the weakest among the other methods, MAPE has the best forecast accuracy.

Table 4.5: Frequency distribution of model identification for simulation of AR(1); $\Phi = -0.9$

Sample Size	Criterion	Candidate Model					
		AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
n=100	MIMSA	1000	0	0	0	0	0
	BIC	908	56	21	5	8	2
	MAPE	97	75	83	114	178	453
	WM1	184	165	170	163	165	153
	WM2	172	178	181	167	147	155
	WM3	150	166	170	166	149	199
	WM4	196	170	167	172	151	144
n=250	MIMSA	1000	0	0	0	0	0
	BIC	941	44	13	2	0	0
	MAPE	125	113	101	108	144	409
	WM1	181	174	166	149	164	166
	WM2	188	154	189	157	158	154
	WM3	185	173	161	152	171	158
	WM4	191	172	160	157	157	163
n=500	MIMSA	1000	0	0	0	0	0
	BIC	977	21	1	1	0	0
	MAPE	156	111	107	116	168	342
	WM1	163	180	168	160	183	146
	WM2	177	161	178	145	170	169
	WM3	171	157	168	179	180	145
	WM4	148	179	179	176	162	156

When the parameter of the AR(1) process is negative, WM4 and WM1 combinations might be able to determine the exact lag order for this scenario.

Table 4.6: Forecast measures for the simulation of AR(1) process; $\Phi = -0.9$

Sample Size	Selection Criterion	Forecast Measure ME	RMSE	MAPE	MASE
	MIMSA	0	0. 8196	157.5201	0.2328
	BIC	0	0.7958	153.8846	0.2244
	MAPE	0	0.2133	45.4598	0.0567
n=100	WM1	0	0.4067	73.1972	0.1083
	WM2	0	0.4129	56.9627	0.1099
	WM3	0	0.3856	66.9736	0.1032
	WM4	0	0.4281	70.4232	0.1163
	MIMSA	0	0.9214	261.3059	0.7972
	BIC	0	0.9190	207.5155	0.2380
	MAPE	0	0.7742	158.9905	0.1873
n=250	WM1	0	0.8122	165.6436	0.2011
	WM2	0	0.8165	167.9598	0.2017
	WM3	0	0.8135	172.9889	0. 2009
	WM4	0	0.8150	166.3760	0.2017
	MIMSA	0	0.9660	614.2113	0.2361
	BIC	0	0.9655	618.7154	0.2360
	MAPE	0	0.9029	597.8768	0.2138
n=500	WM1	0	0.9150	1181.464	0.2182
	WM2	0	0.9150	1021.269	0.2180
	WM3	0	0.9144	1019.674	0.2180
	WM4	0	0.9133	1213.568	0.2176

Table 4.6 might be referred to claim that preferring weighted combinations might increase forecast performance.

Table 4.7: Frequency distribution of model identification for simulation of AR(2); $\Phi_1 = 0.05$, $\Phi_2 = 0.85$

Sample Size	Criterion	Candidate Model					
		AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
n=100	MIMSA	28	972	0	0	0	0
	BIC	904	63	21	9	3	0
	MAPE	264	82	75	72	137	370
	WM1	171	181	160	165	166	157
	WM2	152	181	165	159	171	172
	WM3	115	176	159	157	182	211
	WM4	132	189	188	175	186	130
n=250	MIMSA	3	997	0	0	0	0
	BIC	0	946	47	6	1	0
	MAPE	356	95	62	76	125	286
	WM1	153	172	171	169	172	163
	WM2	159	179	142	180	178	162
	WM3	129	155	193	168	181	174
	WM4	156	197	143	174	175	155
n=500	MIMSA	0	997	3	0	0	0
	BIC	0	960	38	2	0	0
	MAPE	487	84	63	71	94	201
	WM1	140	176	165	172	187	160
	WM2	150	204	169	162	151	164
	WM3	183	160	142	178	187	150
	WM4	136	178	185	169	168	164

As the order of AR process increases, WM4 improves its performance in identifying the precise lag order.

Table 4.8: Forecast measures for the simulation of AR(2) process; $\Phi_1 = 0.05$, $\Phi_2 = 0.85$

Sample Size	Selection Criterion	ME	RMSE	MAPE	MASE
	MIMSA	0	0.6454	134.1855	0.3246
	BIC	0	0.6148	129.8518	0.3083
	MAPE	0	0.3428	79.8036	0.1703
n=100	WM1	0	0.3906	86.3348	0.1918
	WM2	0	0.3733	71.8766	0.1815
	WM3	0	0.3390	87.7833	0.1649
	WM4	0	0.3750	72.0162	0.1845
	MIMSA	0	0.8611	294.2521	0.4005
	BIC	0	0.8565	293.3337	0.3980
	MAPE	0	0.9124	301.6930	0.411
n=250	WM1	0	0.8440	278.5554	0.3788
	WM2	0	0.8393	234.8882	0.3764
	WM3	0	0.83041	272.7126	0.3708
	WM4	0	0.8437	281.4889	0.3787
	MIMSA	0	0.9370	418.7445	0.4268
	BIC	0	0.9361	418.4337	0.4263
	MAPE	0	1.1725	518.5830	0.5239
n=500	WM1	0	0.9836	402.3539	0.4384
	WM2	0	0.9150	410.1774	0.4405
	WM3	0	1.0070	438.2350	0.4483
	WM4	0	0.9740	395.4071	0.4357

(Forecast Measure spans ME, RMSE, MAPE, MASE.)

Even though its order determination performance gets better, WM1 has the biggest forecast accuracy measures among the other weighted combinations.

Table 4.9: Frequency distribution of model identification for simulation of AR(2); $\Phi_1 = 0.45$, $\Phi_2 = 0.5$

Sample Size	Criterion	Candidate Model					
		AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
n=100	MIMSA	813	187	0	0	0	0
	BIC	8	896	68	21	5	2
	MAPE	128	83	65	63	150	511
	WM1	182	167	179	161	165	146
	WM2	179	182	178	162	156	143
	WM3	105	167	161	151	208	208
	WM4	172	155	175	168	172	158
n=250	MIMSA	479	521	0	0	0	0
	BIC	0	954	37	7	2	0
	MAPE	187	101	95	109	155	353
	WM1	159	169	164	163	167	178
	WM2	168	154	170	190	153	165
	WM3	141	174	168	176	164	177
	WM4	142	183	162	193	167	153
n=500	MIMSA	296	704	3	0	0	0
	BIC	0	969	27	3	1	0
	MAPE	170	128	96	128	129	349
	WM1	158	173	177	173	150	169
	WM2	161	165	179	167	164	164
	WM3	137	167	147	189	177	183
	WM4	147	177	157	173	165	181

A comparable conclusion might be drawn that MAPE misclassifies the lag order while WM4 achieves.

Table 4.10: Forecast measures for the simulation of AR(2) process; $\Phi_1 = 0.45$, $\Phi_2 = 0.5$

Sample Size	Selection Criterion	Forecast Measure ME	RMSE	MAPE	MASE
n=100	MIMSA	0	0.7344	192.3547	0.6164
	BIC	0	0.6126	145.8352	0.4896
	MAPE	0	0.2065	58.1468	0.1670
	WM1	0	0.3817	89.8026	0.3036
	WM2	0	0.3871	95.0368	0.3033
	WM3	0	0.3124	70.4231	0.2433
	WM4	0	0.3648	110.0883	0.2899
n=250	MIMSA	0	0.9072	212.6724	0.75813
	BIC	0	0.8536	188.3676	0.7065
	MAPE	0	0.7736	169.6910	0.6155
	WM1	0	0.7944	176.6181	0.6392
	WM2	0	0.7974	170.9482	0.6410
	WM3	0	0.7932	181.0795	0.6364
	WM4	0	0.7946	169.9244	0.6380
n=500	MIMSA	0	0.9740	224.1140	0.8127
	BIC	0	0.9365	219.7203	0.7802
	MAPE	0	0.9113	211.9821	0.7435
	WM1	0	0.9227	214.3793	0.7566
	WM2	0	0.9235	226.9574	0.7568
	WM3	0	0.9165	212.6430	0.7495
	WM4	0	0.9184	216.3142	0.7525

Table 4.10 might point out that forecast accuracy of WM4 is better than WM2 for this simulation scenario.

Table 4.11: Frequency distribution of model identification for simulation of AR(2); $\Phi_1 = 0.02$, $\Phi_2 = -0.9$

Sample Size	Criterion	Candidate Model					
		AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
n=100	MIMSA	0	997	3	0	0	0
	BIC	0	915	69	10	5	1
	MAPE	652	20	23	28	76	201
	WM1	185	187	164	170	150	144
	WM2	141	190	188	152	155	174
	WM3	286	147	136	157	138	136
	WM4	145	187	193	169	158	148
n=250	MIMSA	0	1000	0	0	0	0
	BIC	0	952	35	9	4	0
	MAPE	909	9	8	7	9	58
	WM1	174	161	167	171	158	169
	WM2	154	182	172	180	154	158
	WM3	410	125	106	119	113	127
	WM4	165	174	187	163	141	170
n=500	MIMSA	0	1000	0	0	0	0
	BIC	0	975	19	5	1	0
	MAPE	982	2	2	2	4	8
	WM1	174	154	152	187	181	152
	WM2	142	167	181	148	178	184
	WM3	427	126	102	116	121	108
	WM4	186	138	177	155	171	173

One may refer to Table 4.11 to summarize that MIMSA has a considerably well performance of identifying the true lag order over the other individual and weighted combinations of methods of order selection.

Table 4.12: Forecast measures for the simulation of AR(2) process; $\Phi_1 = 0.02$, $\Phi_2 = -0.9$

Sample Size	Selection Criterion	ME	RMSE	MAPE	MASE
		Forecast Measure			
n=100	MIMSA	0	0.6898	182.7092	0.2309
	BIC	0	0.6694	178.2768	0.2233
	MAPE	0	1.3205	157.1102	0.4242
	WM1	0	0.6227	306.6192	0.2023
	WM2	0	0.5697	86.8715	0.1779
	WM3	0	0.7796	147.1984	0.2500
	WM4	0	0.5824	149.7786	0.1857
n=250	MIMSA	0	0.8825	268.6169	0.3005
	BIC	0	0.8795	265.5125	0.2992
	MAPE	0	1.9639	136.7313	0.6387
	WM1	0	0.1024	170.6058	0.3319
	WM2	0	0.9822	209.6494	0.3218
	WM3	0	1.3262	188.4780	0.4334
	WM4	0	1.0215	216.0892	0.3315
n=500	MIMSA	0	0.9465	242.4629	0.3086
	BIC	0	0.9459	242.4289	0.3083
	MAPE	0	2.1810	118.6156	0.6900
	WM1	0	1.1218	201.6832	0.3582
	WM2	0	1.0763	210.6269	0.3450
	WM3	0	1.4670	161.1644	0.4639
	WM4	0	1.1359	194.1874	0.3626

Unlike the other scenarios, MAPE provides the weakest forecasts for this simulation scenario when MASE and RMSE are considered as a ways of evaluation.

Table 4.13: Frequency distribution of model identification for simulation of AR(3); $\Phi_1 = 0.01$, $\Phi_2 = 0.01$, $\Phi_3 = 0.9$

Sample Size	Criterion	Candidate Model					
		AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
n=100	MIMSA	2	22	976	0	0	0
	BIC	0	0	911	64	17	8
	MAPE	138	98	102	92	143	427
	WM1	148	174	181	159	178	160
	WM2	177	152	171	186	166	148
	WM3	67	59	201	218	210	245
	WM4	112	128	199	203	192	166
n=250	MIMSA	0	0	1000	0	0	0
	BIC	0	0	960	30	9	1
	MAPE	245	132	109	77	130	307
	WM1	140	163	180	183	161	173
	WM2	150	164	194	170	164	158
	WM3	108	110	186	208	196	192
	WM4	117	132	200	184	201	166
n=500	MIMSA	0	0	1000	0	0	0
	BIC	0	0	968	27	4	1
	MAPE	354	186	89	65	96	210
	WM1	167	155	195	182	149	152
	WM2	168	144	188	152	172	176
	WM3	166	179	168	178	139	170
	WM4	144	143	172	194	149	198

If the order of the AR process increases, one might observe that WM4 misspecifies the order that governs the simulated series.

Table 4.14: Forecast measures for the simulation of AR(3) process; $\Phi_1 = 0.01$, $\Phi_2 = 0.01$, $\Phi_3 = 0.9$

Sample Size	Selection Criterion	ME	RMSE	MAPE	MASE
n=100	MIMSA	0	0.5014	162.1973	0.1737
	BIC	0	0.4705	158.2756	0.1618
	MAPE	0	0.3606	190.1461	0.1199
	WM1	0	0.4754	139.9311	0.1670
	WM2	0	0.5034	90.7561	0.1753
	WM3	0	0.3070	113.2984	0.1018
	WM4	0	0.4296	164.4061	0.1453
n=250	MIMSA	0	0.8173	203.8609	0.2842
	BIC	0	0.8143	203.8691	0.2832
	MAPE	0	1.0254	225.3764	0.3445
	WM1	0	0.9585	210.1101	0.3243
	WM2	0	0.9669	208.0808	0.3288
	WM3	0	0.9071	205.0091	0.3038
	WM4	0	0.9209	210.4728	0.3113
n=500	MIMSA	0	0.9070	298.8994	0.3133
	BIC	0	0.9065	298.8724	0.3131
	MAPE	0	1.3329	350.1203	0.4516
	WM1	0	1.1377	325.1420	0.3879
	WM2	0	1.1459	315.2575	0.3878
	WM3	0	1.1634	322.1737	0.3940
	WM4	0	1.1103	282.2918	0.3772

As converse to the previous scenarios, it might be inferred that forecast accuracy of the weighted combinations is weaker than the individual order selection methods.

Table 4.15: Frequency distribution of model identification for simulation of AR(3); $\Phi_1 = 0.01$, $\Phi_2 = 0.01$, $\Phi_3 = 0.5$

Sample Size	Criterion	Candidate Model					
		AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
n=100	MIMSA	594	12	394	0	0	0
	BIC	33	3	860	82	13	9
	MAPE	468	335	36	34	41	86
	WM1	203	171	173	159	174	120
	WM2	179	165	187	165	157	147
	WM3	333	269	99	89	110	100
	WM4	218	200	152	147	146	137
n=250	MIMSA	1	0	999	0	0	0
	BIC	0	0	955	36	8	1
	MAPE	604	347	9	7	15	18
	WM1	177	179	163	149	165	167
	WM2	185	167	166	172	161	149
	WM3	322	323	74	104	91	86
	WM4	188	202	162	148	148	152
n=500	MIMSA	0	0	1000	0	0	0
	BIC	0	0	966	26	7	1
	MAPE	656	331	3	0	2	8
	WM1	178	185	179	145	159	154
	WM2	177	167	164	164	168	160
	WM3	355	337	79	85	69	75
	WM4	210	194	142	139	147	168

The above table might lead one to conclude that weighted methods fail identifying the exact model order for this simulation scenario.

Table 4.16: Forecast measures for the simulation of AR(3) process; $\Phi_1 = 0.01$, $\Phi_2 = 0.01$, $\Phi_3 = 0.5$

Sample Size	Selection Criterion	Forecast Measure ME	RMSE	MAPE	MASE
n=100	MIMSA	0	0.7300	1410.210	0.4478
	BIC	0	0.4967	1298.376	0.2864
	MAPE	0	0.6629	223.3846	0.4078
	WM1	0	0.4337	1032.711	0.2588
	WM2	0	0.4038	1053.195	0.2444
	WM3	0	0.5448	183.0075	0.3337
	WM4	0	0.4393	1767.688	0.2676
n=250	MIMSA	0	0.8333	218.2141	0.5004
	BIC	0	0.8300	218.2662	0.2832
	MAPE	0	0.9822	207.1429	0.6077
	WM1	0	0.8355	206.8600	0.4994
	WM2	0	0.8418	206.1758	0.5023
	WM3	0	0.8998	214.4714	0.5478
	WM4	0	0.8494	208.2393	0.5091
n=500	MIMSA	0	0.9177	355.8488	0.5601
	BIC	0	0.9169	355.7123	0.5594
	MAPE	0	1.0699	241.8662	0.6608
	WM1	0	0.9514	273.7329	0.5777
	WM2	0	0.9478	276.5137	0.5746
	WM3	0	1.0139	305.4841	0.6206
	WM4	0	0.9625	322.1859	0.5844

Table 4.16 might be an indication to assert that WM4 might provide the most considerable forecast accuracy among the other weighted combinations.

Table 4.17: Frequency distribution of model identification for simulation of AR(3); $\Phi_1 = 0.01$, $\Phi_2 = 0.01$, $\Phi_3 = -0.875$

Sample Size	Criterion	Candidate Model					
		AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
n=100	MIMSA	34	4	962	0	0	0
	BIC	0	0	894	76	22	8
	MAPE	380	164	52	65	87	252
	WM1	184	163	194	156	153	150
	WM2	186	143	176	180	154	161
	WM3	190	132	158	173	168	179
	WM4	152	148	188	191	166	155
n=250	MIMSA	0	0	1000	0	0	0
	BIC	0	0	949	40	9	2
	MAPE	591	202	26	34	49	98
	WM1	166	156	185	164	174	155
	WM2	170	141	173	174	164	178
	WM3	263	206	135	134	145	117
	WM4	167	165	174	171	153	170
n=500	MIMSA	0	0	1000	0	0	0
	BIC	0	0	959	36	5	0
	MAPE	759	177	15	10	12	27
	WM1	155	178	180	164	173	150
	WM2	155	169	161	168	169	178
	WM3	297	269	112	113	97	112
	WM4	174	176	151	165	156	178

When the governor parameter for this scenario, Φ_3, is negative, individual methods outclass the combined methods.

Table 4.18: Forecast measures for the simulation of AR(3) process; $\Phi_1 = 0.01$, $\Phi_2 = 0.01$, $\Phi_3 = -0.875$

Sample Size	Selection Criterion	Forecast Measure ME	RMSE	MAPE	MASE
n=100	MIMSA	0	0.5667	110.4978	0.2042
	BIC	0	0.4917	101.4926	0.1769
	MAPE	0	0.8785	123.5599	0.3150
	WM1	0	0.6748	100.3429	0.2393
	WM2	0	0.6456	87.7967	0.2296
	WM3	0	0.5973	92.5174	0.2168
	WM4	0	0.4393	93.1554	0.2139
n=250	MIMSA	0	0.8361	189.1338	0.3025
	BIC	0	0.8318	190.1689	0.3005
	MAPE	0	1.5547	202.0378	0.5583
	WM1	0	1.0592	180.2076	0.3774
	WM2	0	1.0636	189.2199	0.3779
	WM3	0	1.2158	198.5531	0.4336
	WM4	0	1.0764	187.9049	0.3805
n=500	MIMSA	0	0.9223	339.2595	0.3388
	BIC	0	0.9213	339.2037	0.3383
	MAPE	0	1.8030	256.4973	0.6534
	WM1	0	1.2089	324.3600	0.4350
	WM2	0	1.1934	282.3835	0.4320
	WM3	0	1.4367	297.3095	0.5174
	WM4	0	1.2158	282.9949	0.4398

A similar summary might be made with the previous scenarios that WM4 outperforms WM3 in providing accurate forecasts.

Table 4.19: Frequency distribution of model identification for simulation of MA(1); $\Theta = 0.85$

Sample Size	Criterion	Candidate Model					
		MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
n=100	MIMSA	950	49	1	0	0	0
	BIC	920	45	8	5	9	13
	MAPE	222	157	126	120	150	225
	WM1	198	181	164	146	161	150
	WM2	161	190	152	178	157	162
	WM3	191	188	173	149	134	165
	WM4	185	171	177	185	150	132
n=250	MIMSA	827	173	0	0	0	0
	BIC	968	27	3	2	0	0
	MAPE	293	137	100	118	125	227
	WM1	155	185	157	165	168	170
	WM2	171	180	171	137	183	158
	WM3	187	172	159	160	150	172
	WM4	180	151	173	163	156	177
n=500	MIMSA	747	253	0	0	0	0
	BIC	968	14	0	0	0	0
	MAPE	319	137	114	80	124	226
	WM1	162	155	170	188	182	143
	WM2	169	167	178	157	176	153
	WM3	171	168	146	178	171	166
	WM4	182	150	182	169	152	165

Unlike AR scenarios, weighted combinations are capable of distinguishing the precise lag order.

Table 4.20: Forecast measures for the simulation of MA(1) process; $\Theta = 0.85$

Sample Size	Selection Criterion	ME	Forecast Measure RMSE	MAPE	MASE
	MIMSA	-0.0002	0.8074	274.6745	0.6439
	BIC	0.0025	0.7899	269.2438	0.6300
	MAPE	0.0002	0.5360	196.7679	0.4179
n=100	WM1	-0.0067	0.5434	192.0104	0.4184
	WM2	0.0068	0.5276	168.9089	0.4064
	WM3	-0.0029	0.5313	172.3978	0.4134
	WM4	-0.0003	0.5304	170.5939	0.4104
	MIMSA	0.0007	0.9412	279.5984	0.7294
	BIC	0.0009	0.9458	280.5419	0.7321
	MAPE	-0.0005	0.8398	274.3391	0.6491
n=250	WM1	0.0022	0.8347	300.3304	0.6443
	WM2	-0.0009	0.8300	272.0004	0.6408
	WM3	-0.0023	0.8324	284.2089	0.6432
	WM4	-0.0047	0.8296	269.6052	0.6397
	MIMSA	0.00006	0.9810	950.7894	0.7503
	BIC	-0.0003	0.9839	950.8085	0.7521
	MAPE	-0.0032	0.9491	796.2034	0.7256
n=500	WM1	-0.0011	0.9477	921.8829	0.7244
	WM2	-0.0012	0.9469	811.1142	0.7235
	WM3	-0.0023	0.9463	867.7786	0.7231
	WM4	-0.0014	0.9486	833.4172	0.7245

Similar to the previous cases, it might be notified that weighted combinations result in more accurate forecasts than the individual methods.

Table 4.21: Frequency distribution of model identification for simulation of MA(1); $\Theta = 0.5$

Sample Size	Criterion	Candidate Model					
		MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
n=100	MIMSA	966	31	3	0	0	0
	BIC	927	41	5	5	9	13
	MAPE	282	152	153	115	137	161
	WM1	166	187	164	174	152	157
	WM2	183	165	170	171	156	155
	WM3	257	183	157	134	142	127
	WM4	166	195	181	132	171	155
n=250	MIMSA	931	69	0	0	0	0
	BIC	969	26	5	0	0	0
	MAPE	294	193	102	106	110	195
	WM1	153	193	157	167	153	177
	WM2	194	182	166	162	135	161
	WM3	193	185	160	165	145	152
	WM4	169	175	159	154	186	157
n=500	MIMSA	864	136	0	0	0	0
	BIC	979	19	2	0	0	0
	MAPE	338	153	96	104	116	193
	WM1	155	161	165	181	179	159
	WM2	169	169	162	177	155	168
	WM3	174	167	167	150	175	167
	WM4	160	169	183	161	172	155

In converse of the other weighted combinations, WM1 and WM4 might not identify the correct order.

Table 4.22: Forecast measures for the simulation of MA(1) process; $\Theta = 0.5$

Sample Size	Selection Criterion	ME	Forecast Measure RMSE	MAPE	MASE
	MIMSA	-0.0005	0.7879	258.8042	0.6766
	BIC	-0.0013	0.7753	256.4379	0.6656
	MAPE	-0.0060	0.5429	211.8633	0.4525
n=100	WM1	-0.0031	0.5016	198.9587	0.4171
	WM2	0.0004	0.5147	190.6201	0.4271
	WM3	-0.0053	0.5435	198.9575	0.4540
	WM4	-0.0045	0.5098	181.5428	0.4213
	MIMSA	0.0025	0.9438	641.2203	0.7823
	BIC	0.0024	0.9461	641.0980	0.7837
	MAPE	-0.0003	0.8484	633.8094	0.7028
n=250	WM1	0.0068	0.8278	866.6662	0.6846
	WM2	-0.0024	0.8325	632.4741	0.6868
	WM3	-0.00005	0.8345	677.8495	0.6902
	WM4	0.0044	0.8303	868.7453	0.6868
	MIMSA	-0.0003	0.9725	381.5733	0.8021
	BIC	-0.0002	0.9741	379.3626	0.8032
	MAPE	0.0004	0.9383	391.1911	0.7729
n=500	WM1	-0.00008	0.9323	456.1130	0.7678
	WM2	0.0023	0.9335	454.6211	0.7696
	WM3	0.0014	0.9339	409.9072	0.7699
	WM4	-0.0009	0.9331	397.8950	0.7692

When MASE and RMSE values are considered, one might notice that WM4 has better forecasts than WM3.

Table 4.23: Frequency distribution of model identification for simulation of MA(1); $\Theta = -0.9$

Sample Size	Criterion	Candidate Model					
		MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
n=100	MIMSA	944	56	0	0	0	0
	BIC	913	57	17	7	6	0
	MAPE	246	147	122	134	145	206
	WM1	182	187	178	163	145	145
	WM2	160	173	174	194	167	132
	WM3	213	177	177	145	141	147
	WM4	176	182	176	146	163	157
n=250	MIMSA	848	152	0	0	0	0
	BIC	956	31	7	5	0	1
	MAPE	281	141	108	107	132	231
	WM1	173	162	159	175	186	145
	WM2	176	181	158	169	157	159
	WM3	191	170	179	168	148	144
	WM4	165	180	160	170	174	151
n=500	MIMSA	756	243	1	0	0	0
	BIC	983	17	0	0	0	0
	MAPE	302	146	101	106	146	199
	WM1	172	167	159	173	159	170
	WM2	164	175	167	154	169	171
	WM3	147	194	162	161	162	174
	WM4	161	182	180	153	139	185

Like the other scenarios, it might be inferred that weighted methods could not find the model order.

Table 4.24: Forecast measures for the simulation of MA(1) process; $\Theta = -0.9$

Sample Size	Selection Criterion	ME	Forecast Measure RMSE	MAPE	MASE
n=100	MIMSA	0.0019	0.7530	322.7600	0.3552
	BIC	0.0031	0.7485	321.2752	0.3527
	MAPE	0.0067	0.6159	186.6086	0.2793
	WM1	0.0076	0.6187	320.0117	0.2813
	WM2	0.0043	0.6110	296.1750	0.2756
	WM3	0.0086	0.6157	215.7954	0.2797
	WM4	0.0038	0.6083	194.7896	0.2744
n=250	MIMSA	0.0041	0.9275	926.0274	0.4135
	BIC	0.0037	0.9297	896.6298	0.4145
	MAPE	5.9545	0.8657	539.3037	0.3826
	WM1	0.0052	0.8615	873.3900	0.3814
	WM2	0.0060	0.8635	575.6860	0.3821
	WM3	-0.0012	0.8646	970.3020	0.3818
	WM4	0.0064	0.8642	907.6450	0.3825
n=500	MIMSA	-0.0011	0.9741	316.0057	0.4242
	BIC	-0.0006	0.9772	318.2126	0.4256
	MAPE	0.0020	0.9502	319.2493	0.4135
	WM1	0.0006	0.9474	306.6564	0.4126
	WM2	0.0031	0.9462	297.4019	0.4124
	WM3	0.0003	0.9475	305.6033	0.4123
	WM4	0.0012	0.9482	293.3933	0.4128

In this simulation scenario, MAPE leads to the most accurate forecasts for the future values of a series.

Table 4.25: Frequency distribution of model identification for simulation of MA(2); $\Theta_1 = 0.05$, $\Theta_2 = 0.85$

Sample Size	Criterion	Candidate Model					
		MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
n=100	MIMSA	998	2	0	0	0	0
	BIC	0	924	37	21	8	10
	MAPE	858	33	30	23	22	34
	WM1	171	181	166	180	143	159
	WM2	160	188	179	183	158	132
	WM3	491	111	100	108	97	93
	WM4	203	173	164	159	156	145
n=250	MIMSA	0	996	0	4	0	0
	BIC	0	972	24	3	1	0
	MAPE	978	5	3	5	3	6
	WM1	197	163	153	163	164	160
	WM2	145	186	172	174	157	166
	WM3	491	118	90	101	88	112
	WM4	191	163	180	160	154	152
n=500	MIMSA	0	987	1	12	0	0
	BIC	0	982	18	0	0	0
	MAPE	998	2	0	0	0	0
	WM1	208	160	173	157	147	155
	WM2	161	187	160	169	159	164
	WM3	501	95	119	104	94	87
	WM4	209	168	147	149	167	160

For this scenario, except for WM2, other weighted combinations are not able to determine the true model order.

Table 4.26: Forecast measures for the simulation of MA(2) process; $\Theta_1 = 0.05$, $\Theta_2 = 0.85$

Sample Size	Selection Criterion	ME	Forecast Measure RMSE	MAPE	MASE
n=100	MIMSA	-0.0065	0.7868	1211.760	0.4749
	BIC	-0.0063	0.7697	1205.861	0.4648
	MAPE	0.0002	0.8958	1272.131	0.5463
	WM1	-0.0082	0.6238	452.5015	0.3729
	WM2	0.0002	0.6262	456.9985	0.3736
	WM3	0.0046	0.7574	1254.171	0.4570
	WM4	-0.0102	0.6430	1168.962	0.3846
n=250	MIMSA	0.0044	0.9651	389.7006	0.5539
	BIC	0.0041	0.9642	388.6498	0.5533
	MAPE	-0.0003	1.1960	306.5430	0.6880
	WM1	0.0033	0.9324	358.0206	0.5336
	WM2	-0.0030	0.9144	296.2560	0.5237
	WM3	0.0021	1.0310	260.5951	0.5910
	WM4	0.0008	0.9339	281.6703	0.5355
n=500	MIMSA	-0.0038	0.9962	438.1180	0.5615
	BIC	-0.0037	0.9963	439.6835	0.5615
	MAPE	-0.00002	1.2620	200.3159	0.7119
	WM1	-0.0023	1.0296	379.1181	0.5803
	WM2	-0.0018	1.0151	374.6852	0.5725
	WM3	-0.0005	1.1132	253.2646	0.6274
	WM4	-0.0027	1.0266	288.9770	0.5797

Table 4.26 might indicate that WM4 provides the worst forecasts in terms of MAPE when series of size 100 are generated from an MA(2) process.

Table 4.27: Frequency distribution of model identification for simulation of MA(2); $\Theta_1 = 0.45$, $\Theta_2 = 0.5$

Sample Size	Criterion	Candidate Model					
		MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
n=100	MIMSA	41	951	8	0	0	0
	BIC	5	909	46	16	7	17
	MAPE	770	54	41	40	45	50
	WM1	177	174	166	160	162	161
	WM2	182	155	171	177	152	163
	WM3	431	134	130	113	105	87
	WM4	199	173	178	150	159	141
n=250	MIMSA	0	982	18	0	0	0
	BIC	0	977	20	1	2	0
	MAPE	889	26	19	21	22	23
	WM1	165	151	166	179	166	173
	WM2	169	175	158	176	168	154
	WM3	361	127	147	136	124	105
	WM4	201	170	150	166	164	149
n=500	MIMSA	0	938	60	2	0	0
	BIC	0	984	13	1	1	1
	MAPE	925	21	13	9	14	18
	WM1	180	156	178	171	146	169
	WM2	174	155	198	162	140	171
	WM3	333	132	156	152	114	113
	WM4	190	164	161	135	172	178

Similar to the previous scenario, weighted combinations misdetermine the lag order.

Table 4.28: Forecast measures for the simulation of MA(2) process; $\Theta_1 = 0.45$, $\Theta_2 = 0.5$

Sample Size	Selection Criterion	ME	Forecast Measure RMSE	MAPE	MASE
n=100	MIMSA	-0.0096	0.6963	250.8661	0.5793
	BIC	-0.0035	0.6689	238.8434	0.5569
	MAPE	-0.0031	0.7788	240.2211	0.6494
	WM1	-0.0022	0.5466	183.8134	0.4466
	WM2	0.0022	0.5417	182.4134	0.4455
	WM3	-0.0075	0.6517	233.9836	0.5391
	WM4	-0.0097	0.5584	187.8690	0.4554
n=250	MIMSA	0.0046	0.9303	383.0314	0.7589
	BIC	0.0050	0.9298	383.0326	0.7585
	MAPE	0.0002	1.0224	283.9826	0.8329
	WM1	-0.0020	0.8560	365.2674	0.6963
	WM2	-0.0012	0.8619	368.0204	0.7001
	WM3	0.0022	0.9029	388.2366	0.7377
	WM4	-0.0012	0.8638	342.3948	0.7038
n=500	MIMSA	0.0010	0.9712	386.8675	0.7813
	BIC	0.0009	0.9716	386.1791	0.7818
	MAPE	0.0001	1.0702	290.0324	0.8599
	WM1	-0.0011	0.9635	373.8074	0.7745
	WM2	-0.0009	0.9638	413.3963	0.7750
	WM3	-0.0014	0.9881	349.9860	0.7935
	WM4	-0.0019	0.9644	397.4590	0.7739

If Table 4.28 is investigated, it is likely to summarize that forecast values of weighted model selection methods are more acceptable than the individual methods.

Table 4.29: Frequency distribution of model identification for simulation of MA(2); $\Theta_1 = 0.02$, $\Theta_2 = -0.9$

Sample Size	Criterion	Candidate Model					
		MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
n=100	MIMSA	554	449	0	0	0	0
	BIC	0	934	41	12	11	2
	MAPE	329	201	129	76	104	161
	WM1	203	180	174	128	161	154
	WM2	165	173	175	168	167	152
	WM3	175	202	159	159	151	154
	WM4	167	182	159	158	176	158
n=250	MIMSA	478	520	2	0	0	0
	BIC	0	952	35	8	5	0
	MAPE	370	205	146	67	93	119
	WM1	151	185	174	179	141	170
	WM2	173	182	168	153	164	160
	WM3	178	185	174	161	157	145
	WM4	154	167	168	173	158	180
n=500	MIMSA	417	583	0	0	0	0
	BIC	0	983	16	1	0	0
	MAPE	401	220	138	79	53	109
	WM1	176	146	159	173	173	173
	WM2	151	170	179	186	155	159
	WM3	191	171	155	180	148	155
	WM4	129	173	178	175	182	163

WM4 correctly identifies the true lag order when a small sample is of interest in this scenario.

Table 4.30: Forecast measures for the simulation of MA(2) process; $\Theta_1 = 0.02$, $\Theta_2 = -0.9$

Sample Size	Selection Criterion	ME	Forecast Measure RMSE	MAPE	MASE
n=100	MIMSA	0.0076	0.8005	257.4791	0.4537
	BIC	-0.0079	0.6393	225.2643	0.3582
	MAPE	-0.0090	0.6830	237.4451	0.3761
	WM1	-0.0018	0.624	214.4952	0.3426
	WM2	-0.0016	0.6036	227.6729	0.3288
	WM3	-0.0105	0.6223	209.1597	0.3391
	WM4	-0.0045	0.6089	200.8214	0.3324
n=250	MIMSA	-0.0014	1.0416	537.0416	0.5569
	BIC	0.0025	0.904	432.7791	0.4838
	MAPE	0.0035	0.9764	329.9171	0.5199
	WM1	0.0030	0.9046	488.2632	0.4825
	WM2	-0.0025	0.9123	515.5570	0.4850
	WM3	0.0002	0.9138	416.8472	0.4858
	WM4	0.0007	0.9007	404.1132	0.4797
n=500	MIMSA	0.0031	0.9738	568.9926	0.5785
	BIC	-0.0006	0.9716	497.6802	0.5162
	MAPE	-0.0002	1.0771	470.1893	0.5710
	WM1	0.0018	1.0023	559.5736	0.5316
	WM2	-0.0035	0.9989	562.3612	0.5292
	WM3	-0.0006	1.0105	514.9331	0.5355
	WM4	-0.0031	0.9908	452.0749	0.5250

The above table, 4.30, might lead one to conclude that WM4 provides a higher value of RMSE than MIMSA and BIC when the sample size is 500.

Table 4.31: Frequency distribution of model identification for simulation of MA(3); $\Theta_1 = 0.01, \Theta_2 = 0.01, \Theta_3 = 0.9$

Sample Size	Criterion	Candidate Model					
		MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
n=100	MIMSA	35	543	422	0	0	0
	BIC	0	0	911	63	15	11
	MAPE	721	106	51	37	39	46
	WM1	183	192	168	147	149	161
	WM2	168	184	179	162	159	148
	WM3	432	127	113	123	102	103
	WM4	214	142	163	187	144	150
n=250	MIMSA	0	365	634	1	0	0
	BIC	0	0	978	20	1	1
	MAPE	910	63	16	0	3	8
	WM1	201	187	155	159	157	141
	WM2	180	163	177	169	169	142
	WM3	460	123	96	108	117	96
	WM4	184	155	174	173	166	148
n=500	MIMSA	0	299	701	0	0	0
	BIC	0	0	986	12	2	0
	MAPE	942	41	11	1	1	4
	WM1	186	139	174	163	163	175
	WM2	154	181	173	190	139	163
	WM3	491	106	116	87	100	100
	WM4	233	143	154	155	135	180

A similar conclusion might be derived from Table 4.31 that weighted methods could not handle determination of the model order.

When all forecast accuracy measures are considered, one might observe that WM4 has a better forecast performance than WM3.

Table 4.32: Forecast measures for the simulation of MA(3) process; $\Theta_1 = 0.01$, $\Theta_2 = 0.01, \Theta_3 = 0.9$

Sample Size	Selection Criterion	ME	RMSE	MAPE	MASE
			Forecast Measure		
n=100	MIMSA	-0.0068	0.6973	212.9689	0.3945
	BIC	0.0027	0.63597	210.8277	0.3500
	MAPE	0.0028	0.8171	210.0700	0.4640
	WM1	-0.0025	0.6244	173.9054	0.3458
	WM2	-0.0058	0.6185	176.0088	0.3438
	WM3	0.0030	0.7118	193.8768	0.3987
	WM4	0.0029	0.6209	174.8279	0.3452
n=250	MIMSA	-0.0097	1.0272	356.1259	0.5623
	BIC	0.0108	0.9619	340.4056	0.5256
	MAPE	-0.0026	1.2131	374.5522	0.6622
	WM1	0.0017	1.0025	418.7463	0.5468
	WM2	-0.0002	0.9900	319.1146	0.5396
	WM3	-0.0011	1.0715	453.2245	0.5847
	WM4	-0.0001	0.9905	369.1509	0.5379
n=500	MIMSA	-0.0020	1.0697	378.6225	0.5756
	BIC	-0.0014	1.0016	381.7737	0.5394
	MAPE	-0.0001	1.2732	263.6151	0.6851
	WM1	-0.0030	1.0720	350.3562	0.5773
	WM2	0.0017	1.0698	391.8931	0.5762
	WM3	0.00076	1.1511	310.0384	0.6200
	WM4	-0.0017	1.0847	391.5009	0.5840

Table 4.33: Frequency distribution of model identification for simulation of MA(3); $\Theta_1 = 0.01, \Theta_2 = 0.01, \Theta_3 = 0.5$

Sample Size	Criterion	Candidate Model					
		MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
n=100	MIMSA	578	256	166	0	0	0
	BIC	65	3	847	44	26	15
	MAPE	678	195	38	31	27	31
	WM1	184	186	171	155	156	148
	WM2	183	160	185	171	157	144
	WM3	444	201	98	80	95	82
	WM4	171	199	166	163	161	140
n=250	MIMSA	4	77	919	0	0	0
	BIC	0	0	972	23	5	0
	MAPE	794	180	10	4	8	4
	WM1	190	198	161	151	151	149
	WM2	174	183	171	157	166	149
	WM3	411	244	96	80	80	89
	WM4	176	187	163	167	171	136
n=500	MIMSA	0	5	995	0	0	0
	BIC	0	0	989	11	0	0
	MAPE	815	178	4	1	0	2
	WM1	186	164	175	156	169	150
	WM2	156	147	171	193	168	165
	WM3	397	272	79	74	88	90
	WM4	243	175	158	160	129	135

Table 4.33 might lead to the same conclusion that weighted methods could not specify the exact model order.

Table 4.34: Forecast measures for the simulation of MA(3) process; $\Theta_1 = 0.01$, $\Theta_2 = 0.01$, $\Theta_3 = 0.5$

Sample Size	Selection Criterion	ME	Forecast Measure RMSE	MAPE	MASE
n=100	MIMSA	0.0049	0.7080	209.3257	0.4859
	BIC	0.0084	0.5645	221.6663	0.3776
	MAPE	0.0032	0.7121	221.4711	0.4902
	WM1	0.0023	0.5503	214.8556	0.3695
	WM2	-0.0015	0.5459	197.7914	0.3679
	WM3	0.0052	0.6332	241.8709	0.4329
	WM4	0.0026	0.5471	186.9393	0.3651
n=250	MIMSA	-0.0004	0.8939	296.0424	0.5833
	BIC	0.0001	0.8858	293.4758	0.5783
	MAPE	-0.0018	1.0175	243.2090	0.6613
	WM1	0.0044	0.8822	272.8310	0.5740
	WM2	0.0039	0.8821	277.0079	0.5736
	WM3	0.0002	0.9427	273.3017	0.6134
	WM4	0.0033	0.8831	266.8267	0.5744
n=500	MIMSA	0.0006	0.9632	702.5779	0.6184
	BIC	0.0006	0.9623	702.6665	0.6178
	MAPE	-0.00003	1.0797	274.9212	0.6925
	WM1	-0.0006	0.9839	611.1652	0.6314
	WM2	0.0002	0.9777	644.6577	0.6275
	WM3	0.0007	1.0301	352.6094	0.6608
	WM4	-0.0003	0.9955	603.2498	0.6393

When a large sample is generated from this scenario, WM4 might not produce faultless forecasts according to RMSE.

Table 4.35: Frequency distribution of model identification for simulation of MA(3); $\Theta_1 = 0.01, \Theta_2 = 0.01, \Theta_3 = -0.875$

Sample Size	Criterion	Candidate Model					
		MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
n=100	MIMSA	34	520	446	0	0	0
	BIC	0	0	902	51	30	17
	MAPE	847	68	24	18	17	26
	WM1	192	168	188	159	139	154
	WM2	171	193	167	188	154	127
	WM3	473	118	108	87	114	100
	WM4	182	164	169	176	157	152
n=250	MIMSA	1	350	642	1	0	0
	BIC	0	0	949	42	7	2
	MAPE	937	47	8	3	1	4
	WM1	222	145	152	162	164	155
	WM2	167	181	167	164	159	162
	WM3	507	118	95	91	97	92
	WM4	212	161	170	159	155	143
n=500	MIMSA	0	272	726	2	0	0
	BIC	0	0	977	19	4	0
	MAPE	968	28	3	0	0	1
	WM1	173	168	171	165	176	147
	WM2	162	165	168	171	167	167
	WM3	482	133	77	104	110	94
	WM4	273	133	165	124	152	153

As the other scenarios, it might be summarized that individual order selection methods outperform the weighted combinations in identifying the exact lag of the generated process.

Table 4.36: Forecast measures for the simulation of MA(3) process; $\Theta_1 = 0.01$, $\Theta_2 = 0.01$, $\Theta_3 = -0.875$

Sample	Selection		Forecast Measure		
Size	Criterion	ME	RMSE	MAPE	MASE
	MIMSA	-0.0105	0.6994	7903.427	0.3984
	BIC	-0.0029	0.5620	6952.222	0.3171
	MAPE	0.0152	0.9693	7864.838	0.5632
n=100	WM1	-0.0002	0.6626	983.2591	0.3708
	WM2	-0.0071	0.6572	936.9747	0.3709
	WM3	0.0002	0.7864	7845.425	0.4493
	WM4	-0.0036	0.6504	5580.231	0.3659
	MIMSA	-0.0012	0.9689	309.1632	0.5315
	BIC	-0.0037	0.8738	239.6369	0.4805
	MAPE	-0.0016	1.2435	197.7873	0.6822
n=250	WM1	0.0020	0.9698	232.3001	0.5315
	WM2	-0.0058	0.9579	250.3653	0.5265
	WM3	-0.0021	1.0854	219.1860	0.5950
	WM4	-0.0021	0.9746	245.378	0.5348
	MIMSA	-0.0048	1.0345	364.7831	0.5580
	BIC	-0.0063	0.9631	342.7605	0.5190
	MAPE	0.0007	1.2968	199.7068	0.6987
n=500	WM1	-0.0023	1.0545	337.5167	0.5678
	WM2	-0.0051	1.0502	314.2927	0.5649
	WM3	-0.0053	1.1528	320.1238	0.6200
	WM4	-0.0069	1.0813	284.7756	0.5826

Table 4.36 might indicate that weighted combinations are preferable over the individual methods in order to provide more accurate forecasts for a series.

Table 4.37: Frequency distribution of model identification for simulation of ARIMA(1,1,0); $\Phi = 0.9$

Sample Size	Criterion	Candidate Model								
		ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (3,1,0)	ARIMA (1,2,0)	ARIMA (2,2,0)	ARIMA (3,2,0)	ARIMA (1,3,0)	ARIMA (2,3,0)	ARIMA (3,3,0)
n=100	MIMSA	292	0	0	663	0	0	45	0	0
	BIC	672	18	1	271	26	12	0	0	0
	MAPE	28	32	243	14	26	408	27	37	185
	WM1	73	105	121	92	140	180	64	95	130
	WM2	137	130	112	118	149	118	122	114	0
	WM3	103	101	148	109	108	167	60	81	123
	WM4	100	111	130	118	116	134	77	95	119
n=250	MIMSA	259	0	0	728	0	0	13	0	0
	BIC	943	31	5	19	2	0	0	0	0
	MAPE	32	64	291	24	43	356	27	40	123
	WM1	101	118	127	117	113	155	73	78	118
	WM2	148	111	118	130	124	120	127	122	0
	WM3	89	110	140	91	126	139	77	106	122
	WM4	119	108	119	111	124	120	99	97	103
n=500	MIMSA	234	0	0	764	0	0	2	0	0
	BIC	970	27	3	0	0	0	0	0	0
	MAPE	31	55	294	32	50	358	20	53	107
	WM1	110	111	117	122	115	174	59	77	115
	WM2	148	135	114	118	126	121	125	113	0
	WM3	107	113	135	122	137	128	78	94	86
	WM4	125	106	117	104	132	129	86	88	113

An identical conclusion might be derived from Table 4.37 that BIC outperforms the rest of the order selection methods since it finds the true lag order with the highest frequency.

Table 4.38: Forecast measures for the simulation of ARIMA(1,1,0) process; $\Phi = 0.9$

Sample Size	Selection Criterion	ME	Forecast Measure RMSE	MAPE	MASE
n=100	MIMSA	-0.0032	0.7874	8.1270	0.3807
	BIC	0.0009	0.7888	9.8229	0.3881
	MAPE	0.0025	0.5188	2.0075	0.2179
	WM1	0.0110	0.6101	3.5754	0.2672
	WM2	-0.0008	0.6860	4.5534	0.3110
	WM3	-0.0038	0.6253	7.7413	0.2789
	WM4	0.0074	0.6293	3.1530	0.2803
n=250	MIMSA	0.0002	0.9371	4.8004	0.4488
	BIC	-0.0027	0.92793	4.8392	0.4473
	MAPE	-0.0015	0.8773	4.6697	0.4011
	WM1	-0.0003	0.9143	4.8333	0.4223
	WM2	-0.00006	0.9394	4.7563	0.4411
	WM3	-0.0034	0.9230	4.5905	0.4281
	WM4	-0.0043	0.9312	4.7890	0.4320
n=500	MIMSA	0.0036	0.9795	3.9273	0.4453
	BIC	0.0020	0.9678	4.2288	0.4414
	MAPE	-0.0016	0.9681	3.3410	0.4316
	WM1	0.0023	0.9929	3.8805	0.4452
	WM2	0.0023	1.0057	4.1122	0.4522
	WM3	0.0027	0.9989	3.9844	0.4484
	WM4	0.0006	1.0044	3.2859	0.4509

Like the previous scenarios, WM4 might lead to more precise forecasts than individual methods.

Table 4.39: Frequency distribution of model identification for simulation of ARIMA(1,1,0); $\Phi = 0.5$

Sample Size	Criterion	Candidate Model								
		ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (3,1,0)	ARIMA (1,2,0)	ARIMA (2,2,0)	ARIMA (3,2,0)	ARIMA (1,3,0)	ARIMA (2,3,0)	ARIMA (3,3,0)
n=100	MIMSA	53	0	0	219	0	0	728	0	0
	BIC	920	66	13	0	0	1	0	0	0
	MAPE	59	64	413	34	49	240	23	37	81
	WM1	109	121	138	110	124	127	75	95	101
	WM2	132	132	116	125	132	114	133	116	0
	WM3	125	142	151	93	124	155	34	75	101
	WM4	112	117	136	107	111	129	89	97	102
n=250	MIMSA	42	0	0	212	0	0	746	0	0
	BIC	961	38	1	0	0	0	0	0	0
	MAPE	54	98	459	34	64	192	18	20	61
	WM1	124	138	150	91	107	140	75	87	88
	WM2	127	132	132	127	127	109	132	114	0
	WM3	126	154	146	110	118	146	46	60	94
	WM4	121	155	164	107	126	132	44	62	89
n=500	MIMSA	37	0	0	147	0	0	816	0	0
	BIC	956	39	5	0	0	0	0	0	0
	MAPE	81	98	517	36	47	142	20	19	40
	WM1	120	138	158	103	112	131	74	76	88
	WM2	127	127	114	129	125	118	117	143	0
	WM3	126	154	146	110	118	146	46	60	94
	WM4	130	132	137	96	109	130	84	82	100

In this scenario, it might be observable that order selection methods could not overcome the problem of specifying the true model order except BIC.

Table 4.40: Forecast measures for the simulation of ARIMA(1,1,0) process; $\Phi = 0.5$

Sample Size	Selection Criterion	ME	Forecast Measure RMSE	MAPE	MASE
	MIMSA	0.0006	1.0137	60.0172	0.7612
	BIC	-0.0027	0.8113	61.1764	0.6614
	MAPE	0.0027	0.5853	49.0969	0.4188
n=100	WM1	-0.0091	0.6858	55.1847	0.5035
	WM2	-0.0028	0.7457	55.5968	0.5510
	WM3	0.0008	0.6621	51.0207	0.4890
	WM4	-0.0093	0.6893	50.4791	0.5038
	MIMSA	-0.0001	1.2783	38.1790	1.0534
	BIC	-0.0091	0.9345	29.5918	0.7914
	MAPE	-0.0057	0.9124	26.0149	0.7376
n=250	WM1	0.0039	0.9897	25.0781	0.8054
	WM2	-0.0009	1.0164	38.8238	0.83508
	WM3	-0.0092	0.9685	26.2366	0.7875
	WM4	-0.0034	0.9971	26.1977	0.8156
	MIMSA	-0.0008	1.3706	22.9202	1.1618
	BIC	-0.0019	0.9642	21.6553	0.8274
	MAPE	-0.0039	0.9771	20.2554	0.8215
n=500	WM1	0.0013	1.0549	22.2418	0.8870
	WM2	-0.0014	1.0790	19.4946	0.9116
	WM3	0.0003	1.0397	22.0964	0.8764
	WM4	-0.0011	1.0652	21.7002	0.8993

Table 4.40 might be mentioned to draw the same conclusion with the previous one. All other scenarios lead one to notice that individual order selection methods have a tendency to determine the lag order underlying a series with a higher frequency. On the other hand, weighted combinations of the related methods provide more precise forecasts. This might result from the fact that when more complex models with higher numbers of parameters are selected, smaller forecast measures could be computed since such models would lead to higher R^2 values. When weighted combinations and individual methods are compared, one might observe that the former selects higher order models more frequently than the latter. Therefore, the mentioned situation might cause weighted combinations to obtain smaller forecast accuracy measures. The corresponding outputs for the conducted simulation studies are available on Appendix

A.

Once the performances of MIMSA alone for Chapters 3 and 4 are reviewed, one would find out that MIMSA detects the true lag order for almost 1000 times in 1000 simulations for the latter. On the other hand, its performance is not so satisfactory in the former chapter. This might result from the fact that taking the negative of $MIMSA_i$ and adjusting its selection rule for the minimization of that quantity may make MIMSA theoretically appealing. The same issue is also valid for AIC for which two variations are available in the literature. Bozdoğan (1987) proves that the version explained in Chapter 2 is unbiased. Hu (2007) also clarifies that those variations depend on using KL or likelihood function based estimates in computing AIC. KL should be minimized while log-likelihood function should be maximized. Therefore, it would be better to implement the negative version of MIMSA in a model selection problem.

CHAPTER 5

WEIGHTED MODEL SELECTION WITH MIMSA BY ROLLING WINDOWS

Since it has been explicated in the previous chapter that combining MIMSA with MAPE and BIC does not provide decent results we look for. Therefore, we offer to implement rolling window approach as a proposal which requires investigating the data in pieces.

5.1 Rolling Window Analysis

Rolling window analysis is a method to assess the stability of model parameters, in which the data are divided into overlapping or non-overlapping several windows and analyses are conducted to examine if the information coming from each window is stable (Zivot & Wang, 2006). This method is significant for time series in order to satisfy stationarity condition.

5.2 Weighted Model Selection with MIMSA by Rolling Windows

In order to upgrade the forecast performance of MIMSA without exacerbating its lag detection capability, we consider developing an algorithm that implements the combination of MIMSA, MAPE and BIC along with rolling window approach.
By this approach, we recommend to handle an order selection problem by examining the series of interest step-by-step. We expect this algorithm to perform well,

especially, in the presence of outliers since any outlier would have dominance only, around the window it belongs to, instead of the whole series.

This algorithm divides the series of interest into a specified number of non-overlapping windows and evaluates each window separately. Then, candidate models are examined with respect to the weighted sum of the correspondent model selection criteria, for which the corresponding weights are assigned by optimization via GA. After updating the selected model information according to the information provided in the previous window as the windows roll up, the final model concerning the whole series is proposed.
The succeeding steps elucidate the algorithm for the application of Weighted Model Selection with MIMSA by rolling windows:

Step 1- Generate a series of size n and divide it into k windows,

Step 2- Determine the candidate models for the first window, calculate the model selection criteria, $BIC1_i$, $MIMSA1_i$ and $MAPE1_i$, required by weighted model selection for each model and determine the weights by GA : $W1_i$,

$$RWM1_i = \begin{bmatrix} MIMSA1_i & MAPE1_i & BIC1_i \end{bmatrix} * W1_i^T \tag{5.1}$$

Step 3- Compute the model selection criteria in the second window for each candidate, determined in window1 : $BIC2_i$, $MIMSA2_i$ and $MAPE2_i$,

(i) Calculate $RWM21$ by using the weights computed in the first window : $W1_i$,

$$RWM21_i = \begin{bmatrix} MIMSA2_i & MAPE2_i & BIC2_i \end{bmatrix} * W1_i^T \tag{5.2}$$

(ii) Calculate $RWM22$ by using the weights determined by applying GA : $W2_i$,

$$RWM22_i = \begin{bmatrix} MIMSA2_i & MAPE2_i & BIC2_i \end{bmatrix} * W2_i^T \tag{5.3}$$

(iii) Compute $\rho_1 = Corr(RWM1, RWM21)$ and $\rho_2 = Corr(RWM1, RWM22)$. If $\rho_1 \geq \rho_2$, set $W2 = W1$. Else, set $W2 = W2$. Then, calculate $RWM2_i$,

$$RWM2_i = \begin{bmatrix} MIMSA2_i & MAPE2_i & BIC2_i \end{bmatrix} * W2_i^T \tag{5.4}$$

Step 4- Calculate $RWM32$ by using $W2$ and $RWM33$ by using $W3$. Compute $\rho_1 = Corr(RWM2, RWM32)$ and $\rho_2 = Corr(RWM2, RWM33)$. If $\rho_1 \geq \rho_2$, set $W3 = W2$. Else, set $W3 = W3$.Then, calculate $RWM3_i$,

$$RWM3_i = \begin{bmatrix} MIMSA3_i & MAPE3_i & BIC3_i \end{bmatrix} * W3_i^T \tag{5.5}$$

Step 5- Repeat these steps until the k_{th} window and determine Wk_i,

$$RWMk,(k-1)_i = \begin{bmatrix} MIMSAk_i & MAPEk_i & BICk_i \end{bmatrix} * W(k-1)_i^T \tag{5.6}$$

$$RWMkk_i = \begin{bmatrix} MIMSAk_i & MAPEk_i & BICk_i \end{bmatrix} * Wk_i^T \tag{5.7}$$

Compute $\rho_1 = Corr(RWM(k-1), RWMk(k-1))$ and $\rho_2 = Corr(RWM(k-1), RWMkk)$. If $\rho_1 \geq \rho_2$, set $Wk = W(k-1)$. Else, set $Wk = Wk$.

Step 6- Calculate BIC, MIMSA and MAPE for each candidate by using the whole series and weights determined in the k^{th} window and compute $WM4R_i$,

$$WM4R_i = \begin{bmatrix} MIMSA_i & MAPE_i & BIC_i \end{bmatrix} * Wk_i^T \tag{5.8}$$

Step 7- Propose the candidate with the smallest $WM4R_i$ as the final model.

5.3 Simulation Studies

In order to observe how weighted combination of MIMSA by rolling windows manages, a variety of series for many scenario is simulated. Those scenarios cover different types of nonseasonal Box-Jenkins models for lags 1 to 3.

A single simulation scenario is processed on the below algorithm:

1. Generate a series of size n from normal distribution and assign 5 windows by using 20% of the observations,
2. Allocate the first four windows to train set, and appoint the last one to test set,

3. Determine candidate models for the first window on train series and calculate weights for each candidate,
4. Update the corresponding weights until the last window,
5. Calculate $WM4R_i$ for each candidate by using the last updated value of weights,
6. Propose the candidate with the lowest $WM4R_i$ as the final model,
7. Based on the proposed model for each combination, obtain the predicted values for the test set and calculate forecast measures (ME, RMSE, MAPE, and MASE),
8. Repeat steps (1)-(6) for $N = 1,000$ times,
9. Summarize N outputs.

Table 5.1: Frequency distribution of model identification for simulation of AR(1)

Parameter	Sample Size	Criterion	Candidate Model					
			AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
$\Phi = 0.85$	n=100	MIMSA	1000	0	0	0	0	0
		BIC	901	59	23	7	6	4
		MAPE	91	77	78	103	197	454
		WM4	180	165	169	187	164	135
		WM4R	228	200	188	157	119	108
	n=250	MIMSA	974	26	0	0	0	0
		BIC	931	48	12	8	0	1
		MAPE	143	85	104	103	168	397
		WM4	176	179	191	140	153	161
		WM4R	164	196	177	149	159	155
	n=500	MIMSA	1000	0	0	0	0	0
		BIC	965	30	3	1	1	0
		MAPE	150	107	98	114	170	361
		WM4	146	168	196	159	178	153
		WM4R	182	179	181	147	155	156
$\Phi = 0.5$	n=100	MIMSA	997	3	0	0	0	0
		BIC	903	59	27	5	4	2
		MAPE	165	128	107	106	172	322
		WM4	182	188	193	145	156	136
		WM4R	239	180	190	168	127	96
	n=250	MIMSA	998	1	1	0	0	0
		BIC	936	41	16	6	0	1
		MAPE	190	115	108	145	149	293
		WM4	165	178	155	176	162	164
		WM4R	178	164	179	165	161	153
	n=500	MIMSA	1000	0	0	0	0	
		BIC	962	32	5	0	0	1
		MAPE	227	136	109	120	149	259
		WM4	180	167	171	159	151	172
		WM4R	169	174	157	149	163	188

Note: WM4-MIMSA&MAPE&BIC and WM4R-MIMSA&MAPE&BIC by rolling windows.

When one investigates Table 5.1, it might be noticed that approaching an order selection problem by rolling windows has increased the performance of weighted combination of MIMSA with BIC and MAPE.

Table 5.2: Forecast measures for the simulation of AR(1) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Phi = 0.85$	n=100	MIMSA	0	0.8958	205.6662	0.8548
		BIC	0	0.8851	204.0180	0.8424
		MAPE	0	0.6773	157.6072	0.5848
		WM4	0	0.7538	166.1468	0.6762
		WM4R	0	0.7751	183.3684	0.7016
	n=250	MIMSA	0	0.9562	439.2022	0.9199
		BIC	0	0.9545	446.4103	0.9173
		MAPE	0	0.8861	317.7900	0.8244
		WM4	0	0.9045	423.4389	0.8487
		WM4R	0	0.9051	409.6371	0.8493
	n=500	MIMSA	0	0.9795	304.6353	0.9431
		BIC	0	0.9792	304.8121	0.9425
		MAPE	0	0.9468	276.5901	0.8972
		WM4	0	0.9529	290.5555	0.9061
		WM4R	0	0.9548	286.7459	0.9084
$\Phi = 0.5$	n=100	MIMSA	0	0.91658	390.6398	0.7798
		BIC	0	0.9081	285.5492	0.7698
		MAPE	0	0.7371	302.4658	0.5804
		WM4	0	0.7774	314.9821	0.6218
		WM4R	0	0.7932	305.9807	0.6390
	n=250	MIMSA	0	0.9575	369.8152	0.8343
		BIC	0	0.95527	369.3186	0.8315
		MAPE	0	0.8988	350.1058	0.7604
		WM4	0	0.9073	354.4312	0.7701
		WM4R	0	0.9069	360.4131	0.7702
	n=500	MIMSA	0	0.9852	1629.691	0.8491
		BIC	0	0.9849	1629.309	0.8486
		MAPE	0	0.9584	1346.489	0.8151
		WM4	0	0.9597	1568.940	0.8170
		WM4R	0	0.9588	1470.554	0.8154

Table 5.2 summarizes the forecast performance of BIC, MAPE and MIMSA and their weighted combination by rolling windows. It might be inferred that there exist an improvement in this performance after rolling windows approach.

A very similar inference may be made from the other simulation scenarios for which the corresponding tables summarizing the results of the simulation studies could be found at Appendix B. In short, one might analyze the series of interest in nonoverlapping windows in order to increase both forecast and model identification performances.

CHAPTER 6

APPLICATION

This chapter demonstrates how our algorithm, MIMSA, performs with a real data set. There are 364 observations in the data that represent the number of adult blowfly in a population. This series was collected by Nicholson (1950) in a laboratory work. In that work, a fixed number of adult blowflies was kept in a cage and the population size were observed on alternate days while providing with food in a stable amount. This procedure was applied for almost two years. This data were used by many researchers such as Brillinger et al. (1980) and Tong (1977).
Since this data set has been analyzed in many textbooks on time series, we decided to use it in our application study. A detailed description of the data set is provided in Section 6.1 while 6.2 summarizes how a set of various candidate models fits this data set.

6.1 Data Description

The original data set is comprised of 364 observations. However, Tong (1983) claimed that it could be clustered into two subgroups that should be separately analyzed. The corresponding subgroups are named Blowfly A and Blowfly B. While the former requires to be modeled nonlinearly, the latter is stated to follow an AR(1) model (Wei, 2006).
This chapter summarizes the experience of MIMSA while handling Blowfly B data which have 82 observations. This series is also called W3 series in Wei (2006).
Figure 6.1 represents the time series plot of Blowfly B data. It seems stationarity in

mean is satisfied while variance instability is a problem to be solved by log transformation. After this transformation, KPSS test validates that the series is stationary.

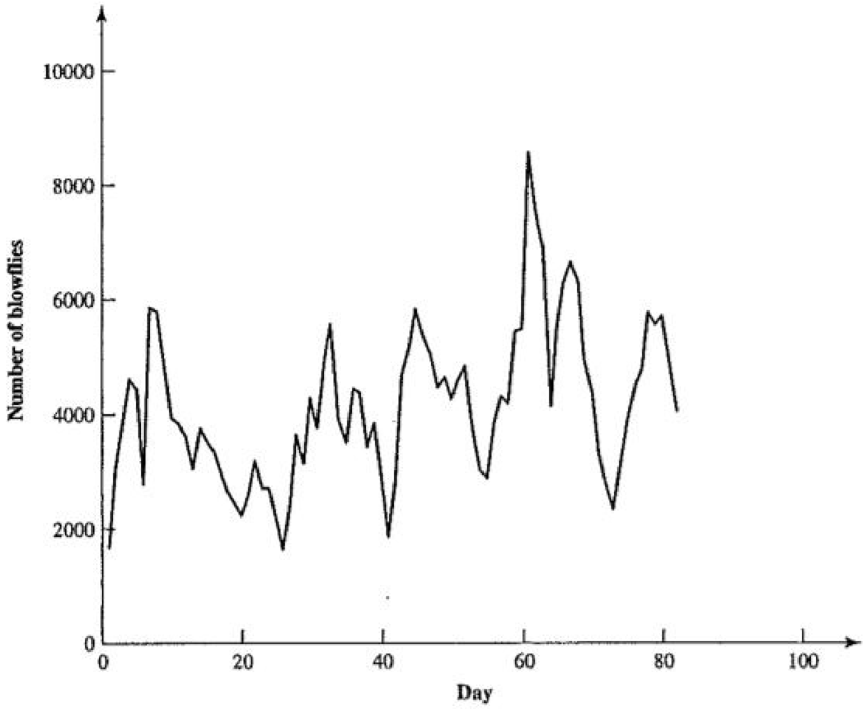

Figure 6.1: Time series plot of the number of blowflies

In order to determine the candidate models, the most commonly preferred visual inspection tools, ACF- PACF plots are illustrated in Figure 6.2.

This figure indicates that Blowfly B series must be governed by an AR(1) process since there is a sinusoidal behavior in the ACF plot and all lags rather than the first one are within white noise bands in the PACF plot.

6.2 Model Selection

For the sake of illustration, a diverse set of candidate models to explore the underlying process for Blowfly B series is evaluated. At the same time, order detectability of various methods is compared. Since the length of series under interest is not very much and $RWM4$ does seem to underperform, it is not applied in this part. In fact, dividing the series into much smaller pieces would cause one not to obtain reliable results.
In addition, for the concern of quantifying out-of-sample performance of each method,

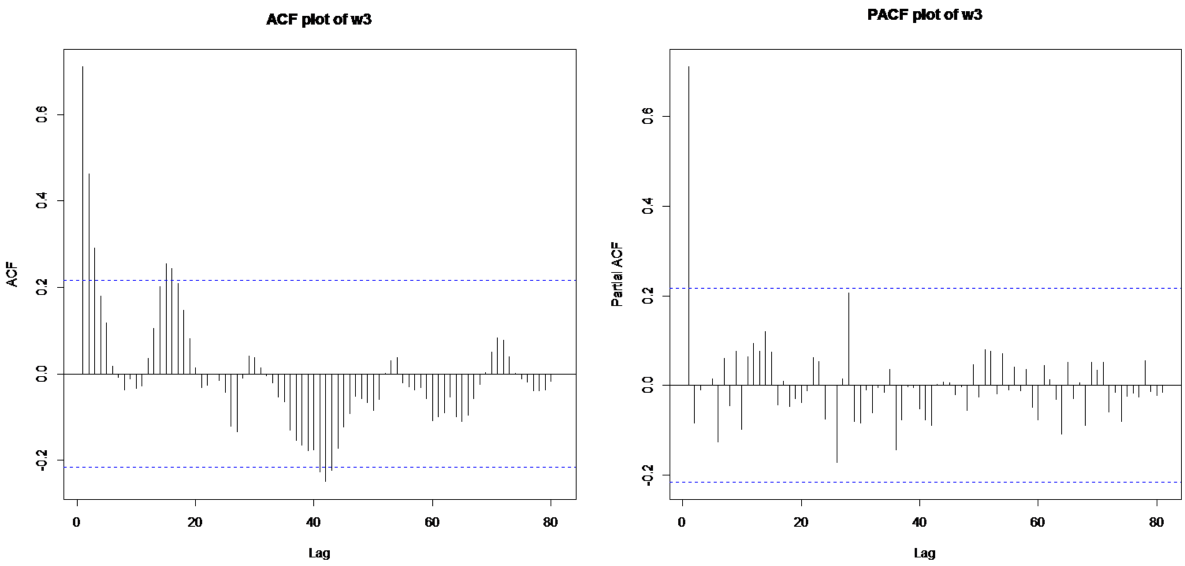

Figure 6.2: ACF-PACF plot for Blowfly B series

the last 5 observations in this series are kept in a test set. Based on the proposed orders, forecast values for each observation for this set is computed. The corresponding measures to provide an insight for this concern could be found in Table 6.2. To this end, RMSE which is the most widely used measure is calculated.

Table 6.1: Respective Outputs of Selection Criteria for Each Candidate

Candidate	Selection Method						
	AIC	BIC	MIMSA	MAPE	HQIC	FPE	WM4
AR(1)	-10.5311	**-8.1244**	**-14.0795**	0.0218	-9.5648	-10.5310	**-7.7897**
AR(2)	**-12.1591**	-7.3457	-14.1778	**0.0203**	**-10.2266**	**-12.1583**	-6.1190
AR(3)	-9.1986	-1.9785	-14.3448	0.0203	-6.2999	-9.1960	-1.6837
MA(1)	15.1290	17.5357	-14.2214	0.0253	16.0952	15.1291	1.0329
MA(2)	7.7716	12.5851	-14.2808	0.0234	9.7041	7.7724	0.2654
MA(3)	5.0303	12.2504	-14.3464	0.0228	7.9290	5.0329	0.3435
ARMA(1,1)	-11.3608	-6.5474	-14.1731	0.0208	-9.4283	-11.3600	-6.2173
ARMA(1,2)	-9.7859	-2.5657	-14.3529	0.0206	-6.8871	-9.7832	-2.4011

Table 6.1 evaluates 8 candidates to fit the series of interest according to 7 model selection methods. The proposed model for each criterion is provided in bold. For instance, one conclude that the underlying process for Blowfly B data is governed by

an AR(2) process if AIC is taken into account as a measure of evaluation of candidate models. When this table is examined, it is observable that only BIC, MIMSA and WM4 manage to correctly specify the true model as AR(1) for this data set. Indeed, AIC, MAPE, HQIC and FPE misdetermine.

Table 6.2: Forecast Performance of Selection Criteria for Each Candidate

Forecast Measure	Selection Method						
	AIC	BIC	MIMSA	MAPE	HQIC	FPE	WM4
RMSE	72.9931	72.9910	72.9910	72.9931	72.9931	72.9931	72.9910

From Table 6.2, it is mentionable that the methods that are able to detect the true order produce better forecasts since the corresponding RMSE values are smaller.

CHAPTER 7

CONCLUSION

In this study, our main focus is based on developing an order determination algorithm to provide a novel and applicable insight into model selection literature of time series. To this end, we aim to benefit from mutual information quantity to improve a fresh method. Box-Jenkins models are the most commonly preferred ways to mimic the process underlying a series. Non-seasonal type of such models are deeply studied in this thesis.

In the first chapter of this thesis, Box-Jenkins models and their properties are introduced. Then, an extensive review of literature on achieving order selection for those models is presented. This review covers information criteria, error based forecast measures and other methods. For each method, the formula and selection rule are reviewed.

In the following chapter, the proposed algorithm is orientated. Prior to this, a detailed description of mutual information concept and its distinguishable characteristics are provided. MIMSA is constructed on the most vital property of mutual information that zero mutual information indicates statistical independence. This algorithm computes the mutual information among the predicted values and observations for the series of interest to evaluate any candidate model. In order to make MIMSA more practical, the natural logarithm of this quantity is penalized by using the information coming from the number of parameters in a candidate model and observations in the series to be modeled. Then, the results of simulation studies conducted to observe how MIMSA performs relative to AIC and BIC are demonstrated. In order to make a more detailed comparison, the frequency of model identifiability and forecast performance of each order selection method have been monitored. In terms of ME, almost all methods in each scenario perform satisfactorily since it is very well known that ME has the tendency to take the value of 0, especially, in AR processes.

Since the forecast performance of MIMSA does not outperform its mostly used counterparts, a weighted combination of it with an information criteria, BIC, and an error based forecast measure, MAPE, is suggested. Since these methods are well known for performing well, they are believed to be improve MIMSA if they take part in order selection process in a cooperation. Weights in this newly offered method are determined by GA's. Simulation studies summarize that model identification frequency of MIMSA dropped when compared to its individual performance. Although this weighted combination yields more acceptable forecasts, its outputs are not satisfactory.

In the next part, the previously built approach on MIMSA that requires to employ a weighted combination is proposed to be enhanced by focusing on an order selection problem in non-overlapping windows. To accomplish this, a method that divides the series of interest into several pieces and investigates each window separately is offered. As the windows roll up, weight information for each piece is updated. Afterwards, a quantity to evaluate the members in the set of candidate models is computed regarding to the whole series. For the aim of examining lag detection and forecast abilities of the proposed method, vibrant simulation scenarios are generated. As a result, we observe that this method by rolling windows manages to identify the true model order better than the weighted combination of order selection methods. However, the corresponding frequencies are very below the ones for MIMSA.

The following part mentions our experience on a real-life data set. We make an application to model a well-known data, Blowfly B, by using widely preferred order selection methods, namely, AIC, BIC, MAPE, HQIC, FPE, MIMSA and WM4. To this end, a diverse collection of candidate models are investigated. The results of this application study point out that only MIMSA and BIC are capable of detecting the true lag length.

In conclusion, the following inferences has been made from this study:

- BIC is able to discover the underlying lag order with the highest accuracy,
- Even though AIC and BIC might select higher order lag lengths more frequently, MIMSA has a tendency to select less complex models with smaller

orders,

- Although MAPE results in the smallest forecast accuracy measures, AIC might be preferred since its order identification ability is also acceptable,

- MIMSA performs better when the process governing a series is AR when compared to MA and ARIMA.

Although both a good forecast and order identification performance are desirable for an order selection to be ideal, detecting the precise lag order with the highest frequency should outweigh the forecast performance. Besides, it might be significant to mention the computational time required to implement each algorithm. Applying for MIMSA takes 10 minutes, on the average. However, this time inclines to 6 hours and 30 hours, on the average, for WMS and WMS by rolling windows, respectively. When these issues are taken into consideration, one may conclude that MIMSA should be preferred over weighted combination and weighted combination via rolling windows approaches. However, our findings reminds that MIMSA requires further research for ARIMA models. In the future, we aim to investigate statistical properties of this algorithm such as consistency and unbiasedness due to the reasons mentioned in Chapter 4. We also look for improving it for those type of Box-Jenkins models and modify MIMSA for seasonal type of time series. Then, all developed methods are planned to be offered as an R package as a free and open source so that many researchers may apply MIMSA comfortably. As a result, we recommend the negative version of MIMSA to handle order selection problems for non-seasonal Box-Jenkins models as a promising alternative to AIC and BIC.

REFERENCES

[1] Akaike, H. (1970).Statistical predictor identification. *Annals of Institute of Statistical Mathematics*, 22, 203-217.

[2] Akaike, H. (1973). *Information theory and the maximum likelihood principle.* In: Petrov, B. N. & Csaki, F. (ed.), Second International Symposium on Information Theory, Akademiai Kiado, Budapest.

[3] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.

[4] Akaike, H. (1978). On the likelihood of a time series model. *The Statistician*, 27, 217-235.

[5] Aladağ,Ç. H., Günay, S., Eğrioğlu, E., & Başaran, M. A. (2010). Improving weighted information criterion by using optimization. *Journal of Computational and Applied Mathematics*, 233(10), 2683–2687.

[6] Anderson, D. R. & Burnham, K. P. (1999). Understanding information criteria for selection among capture-recapture or ring recovery models. *Bird Study*, 46, 14-21.

[7] Armstrong, J. S. (2001). Evaluating forecasting methods. Retrieved on June 23, 2017 from $http://repository.upenn.edu/marketing\_papers/146$.

[8] Ayalew, S., Babu, M. C., & Rao, L.K.M. (2012). Comparison of new approach criteria for estimating the order of autoregressive process. *IOSR Journal of Mathematics (IOSRJM)*, 1(3), 10-20.

[9] Bennasar, M., Hicks, Y. & Setchi, R. Feature selection using joint mutual information maximization. *Expert Systems with Applications*, 42, 8520-8532.

[10] Billah, M. B., Hyndman, R., J., & Koehler, A. B. (2003). Empirical information criteria for time series forecasting model selection. *Journal of Statistical Computation and Simulation*, 75(10), 831-840.

[11] , G. E., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

[12] Brillinger, D. R., Guckenheimer, J., Guttorp, P. E. & Oster, G. (1980). Empirical modeling of population time series data: The case of age and density dependent vital rates. *Lectures on Mathematics in the Life Science (Amer. Math. Soc.)*, 13, 65-90.

[13] Bozdoğan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345-370.

[14] Božić, M., Stojanović, M., Stajić, Z., & Floranović, N. (2013). Mutual information based inputs selection for electric load time series forecasting. *Entropy*, 15, 926-942.

[15] Comon, P. (1994). Independent component analysis, a new concept?. *Signal Processing*, 36(3), 287-314.

[16] Cryer, J.D. & Chan, K.-S. (2008). *Time series analysis with applications in R*, Springer Science+ Business Media, LLC.

[17] Eğrioğlu, E., Aladağ,Ç., H., & Günay, S. (2008). A new model selection strategy in artificial neural networks. *Applied Mathematics and Computation*, 195, 591–597.

[18] El Rhabi, M., Gelle, G., Fenniri, H. & Delaunay, G. (2004). A penalized mutual information criterion for blind separation of convolutive mixtures. *Signal Processing*, 84, 1979-1984.

[19] Gao, W. & Tian, Z. (2010). Detecting lags in nonlinear models using general mutual information. *Journal of Mathematical Research & Exposition*, 30(1), 87-98.

[20] Galka, A., Ozaki, T. & Yamashita, O. (2005). A new approach to mutual information between pairs of time series. Retrieved on June 25, from http://www.ieice.org/proceedings/NOLTA2005/HTMLS/paper/7111.pdf.

[21] Hannan, E. J., & Quinn, B.G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society B.*, 41(2), 190-195.

[22] Hannan, E. J., & Rissanen, J. (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika*, 69(1), 81-94.

[23] Holland, J. H. (1992). Genetic Algorithms. Retrieved on June 25, 2017 from $http://www.cc.gatech.edu/\ turk/bio\_sim/articles/genetic\_algorithm.$

[24] Hu, S. (2007). Akaike Information Criterion. Lecture notes for Center for Research in Scientific Computation, North Carolina State University. Retrieved on July 18, 2017 from *http://www4.ncsu.edu/ shu3/Presentation/AIC.pdf.*

[25] Hurvich, C. M. & Tsai, C.-L. (1995). Model selection for extended quasi-likelihood models in small samples. *Biometrics*, 51(3), 1077-1084.

[26] Hyndman, R. J. (2001). ARIMA processes. Retrieved on June 24, 2017 from https://robjhyndman.com/papers/ARIMA.pdf.

[27] Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679-688.

[28] Inoue, A. & Kilian, L. (2006). On the selection of forecasting models. *Journal of Econometrics*, 130, 273-306.

[29] Kraskov, A., Stögbauer, H. & Grassberger, P. (2004). Estimating mutual information. *Physical Review E*, 69, 066138.

[30] Kullback, S. & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematicl Statistics*, 22, 76-86.

[31] Kuzin, V., Marcellino, M., & Schumacher, C. (2013). Pooling versus model selection for nowcasting with many predictors:an application to German GDP. *Journal of Applied Econometrics*, 28(3), 392–411.

[32] Lange, O. F. & Grubmüller, H. (2006). Generalized correlation for biomolecular dynamics. *PROTEINS: Structure, Function, and Bioinformatics*, 62, 1053-1061.

[33] Li, L. M. (2005). Some notes on mutual information between past and future. *Journal of Time Series Analysis*, 27(2), 309-322. doi:10.1111/j.1467-9892.2005.00469.x.

[34] Li, L. M., & Xie, Z. (1996). Model selection and order determination for time series by information between the past and the future. *Journal of Time Series Analysis*, 17(1), 65-84. doi:10.1111/j.1467-9892.1996.tb00265.x.

[35] Myung, I.J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, 44, 190-204.

[36] Nicholson, A. J. (1950). Population oscillations caused by competition for food. *Nature*, 165, 476-477.

[37] Ong, C.–S., Huang, J.–J., & Tzeng, G. –H. (2005). Model identification of ARIMA family using genetic algorithms. *Applied Mathematics and Computation*, 164, 885-912.

[38] Pankratz, A. (2009). *Forecasting with Univariate Box-Jenkins models: Concepts and cases*, John Wiley & Sons.

[39] Peng, H., Long, F.,& Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8), 1226-1238.

[40] Pham, D. T. (n.d.). Fast algorithm for estimating mutual information, entropies and score functions.Retrieved on June 25, 2017 from
http://www-ljk.imag.fr/membres/Dinh-Tuan.Pham/BSS/mutinf-score.pdf

[41] Pitt, M. A. & Myung, I. J. (2002). When a good fit can be bad. *TRENDS in Cognitive Sciences*, 6 (10), 421-425.

[42] Proietti, T., & Luati, A. (2015). Generalized partial autocorrelations and the mutual information between past and future, presented at Aarhus Conference on Probability, Statistics and Their Applications: Celebrating the Scientific Achievements of Ole E. Barndoff-Nielsen, Aarhus, 15-19 June 2015.

[43] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.

[44] Scrucca, L. (2013). GA: A package for genetic algorithms in R. *Journal of Statistical Software*. 53(4), 1-37.

[45] Sen, L. K. & Shitan, M. (2002). The performance of AICC as an order selection criterion in ARMA time series models. *Pertanika Journal of Science and Technology*, 10(1), 25-33.

[46] Simon, G. & Verleysen, M. (2006). Lag selection for regression models using high-dimensional mutual information. 14th European Symposium on Artificial Neural Networks (ESANN 2006) (Bruges (Belgium), du 26/04/2006 au 28/04/2006). In: Proceedings of the 14th European Symposium on Artificial Neural Networks (ESANN 2006), D-side2006, p.395-400.

[47] Sivanandam, S.N., Deepa, S.N. (2008). Introduction to genetic algorithms. Springer-Verlag Berlin Heidelberg.

[48] Sorjamaa, A., Hao, J. & Lendasse, A. (2005). Mutual information and k-nearest neighbors approximator for time series prediction.In: Duch W., Kacprzyk J., Oja E., Zadrouzny S. (eds) Artificial Neural Networks: Formal Models and Their Applications. ICANN 2005. Lecture Notes in Computer Science, vol 3697. Springer, Berlin, Heidelberg.

[49] Suzuki, T., Sugiyama, M., Sese, J. & Kanamori, T. (2008). A least-squares approach to mutual information estimation. Retrieved on June 25, 2017 from http://www.cs.titech.ac.jp/ tr/reports/2008/TR08-0001.pdf.

[50] Suzuki, T., Sugiyama, M., Sese, J. & Kanamori, T. (2008). Approximating mutual information by maximum likelihood density ratio estimation. Paper published at JMLR: Workshop and Conference Proceedings 4: 5-20.

[51] Tasy, R. S., & Tiao, G.C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA model. *Journal of the American Statistical Association*, 79(1), 84-96.

[52] Tasy, R. S., & Tiao, G.C. (1985). Use of canonical analysis in time series model identification. *Biometrika*, 72(2), 299-315.

[53] Thompson, P.A. (1990). An MSE statistic for comparing forecast accuracy across series. *International Journal of Forecasting*, 6, 219-227.

[54] Tong, H. (1977). Some comments on the Canadian Iynx data with discussion. *Journal of Royal Statistical Society*, Ser. A, 140, 448-468.

[55] Tzannes, N. S., & Noonan, J. P. (1973). The mutual information principle and applications. *Information and Control*, 22, 1-12.

[56] Wagenmakers, E.-J. & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1), 192–196.

[57] Wagenmakers, E-J., Grünwald, P. & Stevyers, M. (2006). Accumulative prediction error and the selection of time series models. *Journal of Mathematical Psychology*, 50, 149-166.

[58] Wei, W. S. W. (2006). *Time Series Analysis: Univariate and Multivariate Methods*, Pearson Education, Inc.

[59] Wu, T.-J., & Sepulveda, A. (1998). The weighted average information criterion for order selection in time series and regression models. *Statistics & Probability Letters*, 39, 1-10.

[60] Zivot, E. & Wang, J. (2006). *Modeling Financial Time Series with S- PLUS*, Springer New York,313-360.

APPENDIX A

SIMULATION RESULTS FOR WEIGHTED MODEL SELECTION (WMS) WITH MIMSA

Table A.1: Frequency distribution of model identification for simulation of ARIMA(2,1,0); $\Phi_1 = 0.01$, $\Phi_2 = 0.9$, N=1000

Sample Size	Criterion	Candidate Model								
		ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (3,1,0)	ARIMA (1,2,0)	ARIMA (2,2,0)	ARIMA (3,2,0)	ARIMA (1,3,0)	ARIMA (2,3,0)	ARIMA (3,3,0)
n=100	MIMSA	426	0	0	450	0	0	124	0	0
	BIC	0	48	5	850	80	17	0	0	0
	MAPE	16	29	175	14	22	538	15	40	151
	WM1	30	100	135	114	151	203	50	81	136
	WM2	115	147	125	122	128	139	120	104	0
	WM3	18	116	144	130	150	158	49	94	141
	WM4	32	127	128	123	139	152	61	108	130
n=250	MIMSA	370	0	0	570	0	0	60	0	0
	BIC	0	187	14	755	38	6	0	0	0
	MAPE	10	52	209	13	35	528	19	34	100
	WM1	35	123	142	127	149	175	46	73	130
	WM2	103	130	140	121	130	127	131	118	0
	WM3	15	126	128	122	145	166	57	99	142
	WM4	28	132	126	133	144	136	84	110	107
n=500	MIMSA	458	0	0	526	0	0	16	0	0
	BIC	0	740	16	231	11	2	0	0	0
	MAPE	12	43	263	16	34	481	23	36	92
	WM1	18	135	160	133	153	164	53	70	114
	WM2	108	117	116	149	137	118	133	122	0
	WM3	18	129	151	134	130	164	61	96	117
	WM4	35	125	116	128	144	138	86	105	123

Table A.2: Forecast measures for the simulation of ARIMA(2,1,0) process; $\Phi_1 = 0.01, \Phi_2 = 0.9$

Sample Size	Selection Criterion	Forecast Measure ME	RMSE	MAPE	MASE
	MIMSA	-0.0041	1.0521	6.0523	0.4677
	BIC	0.0088	0.7334	4.1607	0.3396
	MAPE	0.0024	0.4744	2.3437	0.1947
n=100	WM1	-0.0052	0.5901	3.0200	0.2529
	WM2	0.0065	0.7357	4.1707	0.3212
	WM3	0.0011	0.5784	3.3318	0.2515
	WM4	0.0039	0.6005	3.7436	0.2596
	MIMSA	0.0125	1.2380	4.8063	0.5612
	BIC	0.0043	0.9089	3.7450	0.4208
	MAPE	0.0064	0.8728	3.5461	0.3895
n=250	WM1	-0.0077	0.9391	3.9853	0.4220
	WM2	0.0032	1.0206	3.9014	0.4624
	WM3	0.0025	0.9299	3.7612	0.4185
	WM4	-0.0035	0.9542	3.6502	0.4308
	MIMSA	-0.0097	1.4302	4.8648	0.6362
	BIC	-0.0061	0.9484	3.0290	0.4319
	MAPE	-0.0039	0.9698	3.2170	0.4351
n=500	WM1	0.0029	1.0043	3.0910	0.4524
	WM2	-0.0188	1.1260	3.5580	0.5083
	WM3	-0.0020	1.0161	3.6310	0.4573
	WM4	-0.0100	1.0505	3.2759	0.4718

Table A.3: Frequency distribution of model identification for simulation of ARIMA(2,1,0); $\Phi_1 = 0.4$, $\Phi_2 = 0.5$, N=1000

Sample Size	Criterion	Candidate Model								
		ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (3,1,0)	ARIMA (1,2,0)	ARIMA (2,2,0)	ARIMA (3,2,0)	ARIMA (1,3,0)	ARIMA (2,3,0)	ARIMA (3,3,0)
n=100	MIMSA	279	0	0	617	0	0	104	0	0
	BIC	5	76	5	844	53	17	0	0	0
	MAPE	34	33	193	13	28	485	20	30	164
	WM1	47	108	115	116	134	182	59	104	135
	WM2	158	99	117	117	122	123	120	144	0
	WM3	60	101	148	112	151	156	49	92	131
	WM4	62	103	135	117	131	131	75	103	143
n=250	MIMSA	287	0	0	691	0	0	22	0	0
	BIC	0	425	14	535	20	6	0	0	0
	MAPE	52	51	211	18	45	461	14	24	124
	WM1	54	125	128	129	152	170	53	74	115
	WM2	129	126	104	143	125	114	135	124	0
	WM3	78	141	130	104	144	159	56	83	105
	WM4	96	108	120	116	144	126	71	106	113
n=500	MIMSA	299	0	0	697	0	0	4	0	0
	BIC	0	925	40	31	4	0	0	0	0
	MAPE	61	46	261	26	39	421	16	37	93
	WM1	73	139	155	104	142	153	46	81	107
	WM2	124	115	130	125	134	130	120	122	0
	WM3	95	107	132	108	137	165	67	84	105
	WM4	109	104	117	131	116	123	77	99	124

Table A.4: Forecast measures for the simulation of ARIMA(2,1,0) process; $\Phi_1 = 0.4$, $\Phi_2 = 0.5$

Sample Size	Selection Criterion	Forecast Measure ME	RMSE	MAPE	MASE
n=100	MIMSA	-0.0130	0.8173	7.0008	0.4597
	BIC	-0.0072	0.7450	7.1714	0.4167
	MAPE	0.0008	0.5109	3.0382	0.2534
	WM1	-0.0081	0.6056	6.0912	0.3172
	WM2	-0.0104	0.7020	8.0819	0.3747
	WM3	-0.0014	0.6110	6.7156	0.3192
	WM4	-0.0032	0.6182	6.1587	0.32437
n=250	MIMSA	-0.0046	0.9598	8.9884	0.5446
	BIC	-0.0026	0.9019	7.3989	0.5073
	MAPE	0.0031	0.8731	8.0520	0.4737
	WM1	-0.0004	0.9141	7.1818	0.5016
	WM2	0.0011	0.9576	8.1869	0.5324
	WM3	-0.0004	0.9207	8.0219	0.5066
	WM4	0.0052	0.9329	8.9424	0.5147
n=500	MIMSA	0.0030	1.0115	3.7324	0.5587
	BIC	0.0032	0.9455	3.4688	0.5192
	MAPE	0.0036	0.9683	3.6170	0.52457
	WM1	0.0047	0.9996	3.6517	0.5437
	WM2	-0.0033	1.0295	3.9922	0.5628
	WM3	0.0016	1.0100	3.8047	0.5517
	WM4	0.0043	1.0235	4.0599	0.5582

Table A.5: Frequency distribution of model identification for simulation of ARIMA(1,2,0); $\Phi = 0.9$, N=1000

Sample Size	Criterion	Candidate Model								
		ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (3,1,0)	ARIMA (1,2,0)	ARIMA (2,2,0)	ARIMA (3,2,0)	ARIMA (1,3,0)	ARIMA (2,3,0)	ARIMA (3,3,0)
n=100	MIMSA	670	0	0	9	0	0	321	0	0
	BIC	0	9	0	630	20	3	279	46	13
	MAPE	4	4	22	3	5	129	7	35	791
	WM1	139	8	13	355	21	16	379	45	24
	WM2	245	124	27	353	101	24	105	21	0
	WM3	4	62	126	48	111	160	109	155	225
	WM4	44	96	25	179	128	61	251	143	73
n=250	MIMSA	10	0	0	462	0	0	528	0	0
	BIC	0	9	1	943	29	0	13	5	0
	MAPE	5	2	17	2	2	161	8	29	774
	WM1	260	2	0	334	5	0	396	3	0
	WM2	147	147	86	208	132	71	148	67	0
	WM3	12	97	103	97	119	156	110	143	163
	WM4	58	108	84	140	109	115	149	132	105
n=500	MIMSA	423	0	0	16	0	0	561	0	0
	BIC	0	4	0	956	35	4	0	0	1
	MAPE	9	2	12	0	5	165	8	33	766
	WM1	390	25	1	271	17	1	283	11	1
	WM2	115	131	116	143	134	113	134	114	0
	WM3	34	107	119	100	116	120	117	135	152
	WM4	67	118	137	120	122	110	114	100	112

Table A.6: Forecast measures for the simulation of ARIMA(1,2,0) process; $\Phi = 0.9$

Sample Size	Selection Criterion	ME	RMSE	MAPE	MASE
			Forecast Measure		
n=100	MIMSA	-0.0032	0.9390	0.1050	0.0295
	BIC	0.0033	0.7219	0.0605	0.0178
	MAPE	0.0038	0.3606	0.0434	0.0074
	WM1	-0.0049	0.7534	0.0627	0.0189
	WM2	-0.0103	0.7663	0.0672	0.0199
	WM3	0.0073	0.5210	0.0470	0.0118
	WM4	0.0035	0.6466	0.0497	0.0154
n=250	MIMSA	0.0004	0.9137	0.0352	0.0147
	BIC	-0.0027	0.9044	0.0298	0.0144
	MAPE	0.0008	0.8178	0.0274	0.0123
	WM1	-0.0019	1.0530	0.0354	0.0170
	WM2	-0.0015	0.9665	0.0287	0.0153
	WM3	-0.0035	0.8574	0.0288	0.0133
	WM4	0.0018	0.8996	80.0292	0.0142
n=500	MIMSA	0.0068	1.3429	0.0245	0.0156
	BIC	0.0020	0.9537	0.0110	0.0105
	MAPE	-0.00073	0.9305	0.0010	0.0100
	WM1	0.0023	1.3064	0.0159	0.0147
	WM2	0.0082	1.0497	0.0108	0.0115
	WM3	0.0030	0.9681	0.0115	0.0104
	WM4	0.0043	0.9983	0.0108	0.0109

Table A.7: Frequency distribution of model identification for simulation of ARIMA(1,2,0); $\Phi = 0.5$, N=1000

Sample Size	Criterion	Candidate Model								
		ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (3,1,0)	ARIMA (1,2,0)	ARIMA (2,2,0)	ARIMA (3,2,0)	ARIMA (1,3,0)	ARIMA (2,3,0)	ARIMA (3,3,0)
n=100	MIMSA	903	0	0	23	0	0	74	0	0
	BIC	0	7	0	919	58	14	0	0	2
	MAPE	27	12	33	3	9	286	25	48	557
	WM1	109	52	57	135	103	102	187	100	155
	WM2	178	124	82	196	125	84	112	99	0
	WM3	26	61	108	64	117	169	86	148	221
	WM4	60	72	100	119	123	126	138	143	119
n=250	MIMSA	31	0	0	612	0	0	357	0	0
	BIC	0	9	0	933	47	11	0	0	0
	MAPE	29	3	26	3	14	324	29	50	522
	WM1	153	32	13	312	37	21	368	33	31
	WM2	159	139	89	184	147	79	130	73	0
	WM3	33	69	125	90	118	158	108	141	158
	WM4	64	107	90	135	94	131	136	144	99
n=500	MIMSA	512	0	0	25	0	0	463	0	0
	BIC	0	3	0	960	32	5	0	0	0
	MAPE	26	1	24	4	7	380	33	46	479
	WM1	289	13	1	338	10	3	322	21	3
	WM2	155	127	108	144	137	108	130	91	0
	WM3	57	111	122	106	110	129	94	125	146
	WM4	86	118	125	113	113	103	101	130	111

Table A.8: Forecast measures for the simulation of ARIMA(1,2,0) process; $\Phi = 0.5$

Sample Size	Selection Criterion	ME	Forecast Measure RMSE	MAPE	MASE
	MIMSA	-0.0010	0.8252	0.3503	0.0914
	BIC	0.0015	0.7333	0.3084	0.0747
	MAPE	-0.0018	0.4202	0.1483	0.0378
n=100	WM1	0.0035	0.6237	0.2143	0.0553
	WM2	-0.0010	0.6823	0.2514	0.0663
	WM3	-0.0019	0.5433	0.2073	0.0496
	WM4	0.0014	0.6080	0.2284	0.0561
	MIMSA	-0.0070	0.9444	0.1502	0.0632
	BIC	-0.0035	0.9070	0.1321	0.0614
	MAPE	-0.0018	0.8605	0.1213	0.0551
n=250	WM1	-0.0037	0.9542	0.1114	0.0629
	WM2	-0.0033	0.9148	0.1284	0.0611
	WM3	-0.0048	0.8895	0.1181	0.0584
	WM4	-0.0022	0.9001	0.1177	0.0595
	MIMSA	-0.0008	1.0837	0.0902	0.0513
	BIC	0.0020	0.9587	0.0767	0.0448
	MAPE	-0.0011	0.9625	0.0752	0.0442
n=500	WM1	-0.0010	1.0347	0.0748	0.0485
	WM2	0.0015	0.9875	0.0767	0.0460
	WM3	-0.0008	0.9731	0.0760	0.0449
	WM4	0.0035	0.9790	0.0754	0.0452

Table A.9: Frequency distribution of model identification for simulation of ARIMA(2,2,0); $\Phi_1 = 0.01$, $\Phi_2 = 0.9$, N=1000

Sample Size	Criterion	Candidate Model								
		ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (3,1,0)	ARIMA (1,2,0)	ARIMA (2,2,0)	ARIMA (3,2,0)	ARIMA (1,3,0)	ARIMA (2,3,0)	ARIMA (3,3,0)
n=100	MIMSA	571	0	0	127	0	0	302	0	0
	BIC	0	0	3	0	48	7	850	68	24
	MAPE	0	8	21	2	6	132	5 805		
	WM1	162	11	11	239	29	18	442	56	32
	WM2	277	110	29	260	146	32	119	27	0
	WM3	6	17	113	13	152	145	126	166	262
	WM4	52	45	55	73	153	74	283	181	84
n=250	MIMSA	1	0	0	454	0	0	545	0	0
	BIC	52	45	55	73	153	74	283	181	84
	MAPE	4	5	15	4	5	120	4	19	824
	WM1	289	1	0	304	1	1	400	3	1
	WM2	156	106	85	152	154	79	162	106	0
	WM3	12	38	128	39	127	182	130	143	201
	WM4	73	54	117	82	127	128	158	149	112
n=500	MIMSA	337	0	0	122	0	0	541	0	0
	BIC	0	0	4	0	713	20	246	14	3
	MAPE	1	3	8	6	2	120	7	21	832
	WM1	311	30	0	325	22	1	284	27	0
	WM2	118	98	129	110	160	112	144	129	0
	WM3	941	55	124	59	141	157	124	146	153
	WM4	80	72	125	81	123	128	131	142	118

Table A.10: Forecast measures for the simulation of ARIMA(2,2,0) process; $\Phi_1 = 0.01$, $\Phi_2 = 0.9$

Sample Size	Selection Criterion	Forecast Measure ME	RMSE	MAPE	MASE
n=100	MIMSA	-0.0190	1.1420	4.5964	0.0333
	BIC	-0.0099	0.6581	5.2968	0.0155
	MAPE	-0.0069	0.3301	0.0432	0.0068
	WM1	-0.0137	0.9084	0.0861	0.0208
	WM2	0.0091	1.0139	3.9287	0.0269
	WM3	0.0003	0.4918	0.0546	0.0115
	WM4	-0.0004	0.6588	0.0764	0.0149
n=250	MIMSA	-0.0138	1.2426	0.0406	0.0193
	BIC	0.0001	0.8857	0.0232	0.0132
	MAPE	-0.0003	0.8156	0.0202	0.0117
	WM1	0.0160	1.3569	0.0321	0.0209
	WM2	-0.0128	1.1515	0.0339	0.0183
	WM3	-0.0115	0.8952	0.0225	0.0128
	WM4	-0.0067	0.9961	0.0268	0.0147
n=500	MIMSA	0.0108	1.3715	0.0114	0.0141
	BIC	0.0026	0.9318	0.0066	0.0091
	MAPE	0.0055	0.9145	0.0062	0.0088
	WM1	-0.0025	1.5473	0.0112	0.0161
	WM2	0.0122	1.2060	0.0078	0.0121
	WM3	0.0042	1.0434	0.0068	0.0103
	WM4	0.0090	1.1234	0.0077	0.0109

Table A.11: Frequency distribution of model identification for simulation of ARIMA(0,1,1); $\Theta = 0.9$, N=1000

Sample Size	Criterion	Candidate Model								
		ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (0,1,3)	ARIMA (0,2,1)	ARIMA (0,2,2)	ARIMA (0,2,3)	ARIMA (0,3,1)	ARIMA (0,3,2)	ARIMA (0,3,3)
n=100	MIMSA	40	0	0	410	0	0	548	2	0
	BIC	823	34	9	0	120	12	0	0	2
	MAPE	86	76	167	55	167	250	24	41	134
	WM1	105	126	91	97	138	140	102	85	116
	WM2	139	118	114	128	133	115	140	113	0
	WM3	121	141	133	83	134	133	60	68	127
	WM4	122	120	122	96	128	118	87	98	109
n=250	MIMSA	25	0	0	455	0	0	520	0	0
	BIC	938	15	4	0	40	3	0	0	0
	MAPE	102	89	169	54	185	227	37	24	113
	WM1	135	110	118	99	128	134	84	83	109
	WM2	109	126	117	148	143	124	116	117	0
	WM3	144	149	136	77	132	125	67	75	95
	WM4	125	132	92	92	132	133	85	99	110
n=500	MIMSA	37	0	0	430	0	0	533	0	0
	BIC	971	16	0	0	11	2	0	0	0
	MAPE	119	77	187	50	172	214	24	24	133
	WM1	141	117	126	93	106	148	79	81	109
	WM2	129	129	130	112	128	115	120	137	0
	WM3	132	127	111	80	141	157	61	68	123
	WM4	115	118	144	100	122	116	84	89	112

Table A.12: Forecast measures for the simulation of ARIMA(0,1,1) process; $\Theta = 0.9$

Sample	Selection		Forecast Measure		
Size	Criterion	ME	RMSE	MAPE	MASE
	MIMSA	-0.0270	1.1368	16651.79	0.7764
	BIC	0.0148	0.9423	15967.89	0.6867
	MAPE	-0.0021	0.8878	3869.660	0.6115
n=100	WM1	-0.00775	0.9466	13959.57	0.6526
	WM2	0.0136	0.9514	16661.45	0.6616
	WM3	0.0019	0.9175	3867.146	0.6342
	WM4	0.0075	0.9369	15962.54	0.6439
	MIMSA	-0.0101	1.2796	33.7023	0.9285
	BIC	0.0015	0.9901	27.8747	0.7389
	MAPE	0.0002	1.0217	25.7396	0.7486
n=250	WM1	-0.0208	1.0693	30.8039	0.7839
	WM2	-0.0183	1.0911	28.3964	0.7977
	WM3	0.0003	1.0558	25.3008	0.7734
	WM4	0.0103	1.0804	26.1557	0.7901
	MIMSA	-0.0002	1.3247	37.4924	0.9675
	BIC	-0.0017	1.0038	32.7370	0.7451
	MAPE	0.0037	1.0513	36.5928	0.7719
n=500	WM1	0.0015	1.1001	36.1753	0.8071
	WM2	-0.0012	1.1231	36.7911	0.8250
	WM3	-0.0055	1.0863	32.5704	0.7984
	WM4	0.0031	1.1093	36.1460	0.8131

Table A.13: Frequency distribution of model identification for simulation of ARIMA(0,1,1); $\Theta = 0.5$, N=1000

Sample Size	Criterion	Candidate Model								
		ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (0,1,3)	ARIMA (0,2,1)	ARIMA (0,2,2)	ARIMA (0,2,3)	ARIMA (0,3,1)	ARIMA (0,3,2)	ARIMA (0,3,3)
n=100	MIMSA	14	0	0	328	0	0	658	0	0
	BIC	902	32	5	2	49	10	0	0	0
	MAPE	116	102	185	88	125	183	52	50	99
	WM1	122	105	112	97	121	126	99	112	106
	WM2	140	128	111	146	123	112	121	119	0
	WM3	138	141	157	96	133	131	54	77	73
	WM4	111	117	115	114	116	110	109	107	101
n=250	MIMSA	28	0	0	369	0	0	603	0	0
	BIC	967	13	2	0	18	0	0	0	0
	MAPE	155	121	213	84	113	152	47	40	75
	WM1	134	128	121	85	116	129	114	85	88
	WM2	110	116	136	128	126	108	154	122	0
	WM3	125	148	138	103	113	116	86	67	104
	WM4	115	121	129	110	126	108	96	91	104
n=500	MIMSA	27	0	0	390	0	0	583	0	0
	BIC	971	10	3	0	16	0	0	0	0
	MAPE	145	118	227	88	131	110	49	46	86
	WM1	120	127	129	101	118	109	99	86	111
	WM2	131	108	133	148	108	119	147	106	0
	WM3	119	128	142	105	127	145	77	76	81
	WM4	113	124	130	109	103	118	91	105	107

Table A.14: Forecast measures for the simulation of ARIMA(0,1,1) process; $\Theta = 0.5$

Sample Size	Selection Criterion	ME	RMSE	MAPE	MASE
			Forecast Measure		
n=100	MIMSA	0.0030	1.0580	33.1732	0.8593
	BIC	0.0073	0.8544	31.0548	0.7563
	MAPE	-0.0080	0.8284	26.7071	0.6888
	WM1	0.0146	0.8619	27.4270	0.7176
	WM2	0.0106	0.8823	29.2633	0.7395
	WM3	-0.0104	0.8350	22.3105	0.7026
	WM4	0.0007	0.8704	25.2569	0.7210
n=250	MIMSA	-0.0052	1.1802	30.3412	1.0193
	BIC	0.0012	0.9512	21.6627	0.8449
	MAPE	-0.0029	0.9870	24.6769	0.8622
	WM1	-0.0017	1.0258	22.8191	0.8948
	WM2	-0.0029	1.0353	22.6598	0.9022
	WM3	-0.0019	1.0172	29.0753	0.8868
	WM4	-0.0025	1.0283	28.6777	0.8929
n=500	MIMSA	0.0061	1.1996	20.1053	1.0550
	BIC	-0.0068	0.9780	17.1485	0.8763
	MAPE	-0.0029	1.0280	18.2768	0.9136
	WM1	-0.0024	1.0623	18.1335	0.9420
	WM2	-0.0023	1.0656	18.3605	0.9446
	WM3	0.0013	1.0485	17.9721	0.9301
	WM4	0.0001	1.0681	17.7991	0.9454

Table A.15: Frequency distribution of model identification for simulation of ARIMA(0,1,2); $\Theta_1 = 0.01$, $\Theta_2 = 0.9$, N=1000

Sample Size	Criterion	Candidate Model								
		ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (0,1,3)	ARIMA (0,2,1)	ARIMA (0,2,2)	ARIMA (0,2,3)	ARIMA (0,3,1)	ARIMA (0,3,2)	ARIMA (0,3,3)
n=100	MIMSA	26	0	0	355	0	0	619	0	0
	BIC	0	817	49	0	0	134	0	0	0
	MAPE	71	143	216	22	50	381	13	21	83
	WM1	109	133	105	102	103	171	95	86	96
	WM2	105	134	143	121	113	142	121	121	0
	WM3	93	185	188	93	100	180	27	52	82
	WM4	111	129	119	108	118	151	73	80	111
n=250	MIMSA	29	0	0	397	1	0	575	1	0
	BIC	0	919	17	0	0	64	0	0	0
	MAPE	67	193	237	31	36	358	7	20	51
	WM1	98	156	150	85	107	147	75	101	81
	WM2	129	123	110	136	111	122	127	142	0
	WM3	90	187	184	101	108	157	31	53	89
	WM4	111	139	150	102	119	129	57	89	104
n=500	MIMSA	24	0	0	442	1	0	0	531	2
	BIC	0	959	15	0	0	26	0	0	0
	MAPE	48	225	281	18	32	324	5	19	48
	WM1	87	176	172	81	93	153	76	69	93
	WM2	136	120	119	104	116	136	127	142	0
	WM3	103	180	192	81	79	179	38	60	88
	WM4	116	137	145	106	98	161	73	78	86

Table A.16: Forecast measures for the simulation of ARIMA(0,1,2) process; $\Theta_1 = 0.01$, $\Theta_2 = 0.9$

Sample Size	Selection Criterion	Forecast Measure ME	RMSE	MAPE	MASE
n=100	MIMSA	0.0054	1.5759	55.8827	1.0327
	BIC	-0.0117	0.9372	27.4258	0.6651
	MAPE	-0.0084	0.9845	26.2481	0.6632
	WM1	0.0124	1.1450	33.9995	0.7611
	WM2	-0.0160	1.1535	37.8331	0.7791
	WM3	0.0103	1.0562	31.9550	0.7127
	WM4	-0.0058	1.1294	30.1854	0.7520
n=250	MIMSA	0.0041	1.6544	40.9816	1.1865
	BIC	-0.0041	1.0040	21.5025	0.7436
	MAPE	-0.0067	1.0668	26.9478	0.7768
	WM1	-0.0082	1.2300	24.2795	0.8878
	WM2	-0.0073	1.2893	27.5416	0.9350
	WM3	-0.0153	1.1542	37.4944	0.8404
	WM4	-0.0096	1.2165	33.0219	0.8790
n=500	MIMSA	0.0072	1.6626	29.0777	1.2054
	BIC	0.0044	1.0150	20.6952	0.7472
	MAPE	-0.0048	1.0787	20.3675	0.7882
	WM1	0.0038	1.2361	21.8456	0.8999
	WM2	0.0068	1.3184	26.6597	0.9598
	WM3	-0.0046	1.1943	22.3202	0.8697
	WM4	0.0059	1.2519	22.6475	0.9109

Table A.17: Frequency distribution of model identification for simulation of ARIMA(0,1,2); $\Theta_1 = 0.4$, $\Theta_2 = 0.5$, N=1000

Sample Size	Criterion	Candidate Model								
		ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (0,1,3)	ARIMA (0,2,1)	ARIMA (0,2,2)	ARIMA (0,2,3)	ARIMA (0,3,1)	ARIMA (0,3,2)	ARIMA (0,3,3)
n=100	MIMSA	21	0	0	452	0	0	527	0	0
	BIC	9	874	31	11	1	73	0	1	0
	MAPE	108	132	173	75	66	305	45	28	68
	WM1	107	127	95	124	78	148	111	103	107
	WM2	138	121	108	136	124	116	126	131	0
	WM3	102	154	160	98	115	136	64	72	99
	WM4	118	134	105	120	98	121	98	102	104
n=250	MIMSA	26	0	0	502	0	0	472	0	0
	BIC	0	951	16	0	0	33	0	0	0
	MAPE	89	163	199	64	54	311	45	35	40
	WM1	113	120	122	121	91	139	98	88	108
	WM2	132	125	103	114	130	136	128	132	0
	WM3	112	160	135	109	113	134	72	65	100
	WM4	99	131	134	111	123	124	83	95	100
n=500	MIMSA	32	0	0	497	0	0	471	0	0
	BIC	0	973	11	0	0	16	0	0	0
	MAPE	90	180	208	40	47	327	37	27	44
	WM1	111	156	135	97	111	152	77	78	83
	WM2	121	131	113	122	135	131	122	125	0
	WM3	96	145	126	112	100	149	82	95	95
	WM4	113	112	132	111	113	119	104	106	90

Table A.18: Forecast measures for the simulation of ARIMA(0,1,2) process; $\Theta_1 = 0.4$, $\Theta_2 = 0.5$

Sample Size	Selection Criterion	ME	RMSE	MAPE	MASE
		Forecast Measure			
n=100	MIMSA	0.0114	1.0757	28.3023	0.8402
	BIC	-0.0029	0.8151	26.2313	0.6741
	MAPE	0.0025	0.8412	27.5683	0.6673
	WM1	0.0063	0.9144	25.6489	0.7206
	WM2	-0.0060	0.9180	23.3675	0.7255
	WM3	-0.0207	0.8709	22.9149	0.6896
	WM4	-0.0202	0.9008	24.4699	0.7048
n=250	MIMSA	-0.00391	1.1702	20.1144	0.9727
	BIC	-0.0077	0.9274	16.3035	0.7893
	MAPE	-0.0150	0.9833	17.5362	0.8254
	WM1	-0.0049	1.0416	18.4354	0.8687
	WM2	-0.0064	1.0435	19.5056	0.8720
	WM3	-0.0109	1.0254	18.7079	0.8584
	WM4	-0.0141	1.0391	17.3029	0.8660
n=500	MIMSA	0.0022	1.2012	18.5538	0.9929
	BIC	-0.0078	0.9686	14.2650	0.7472
	MAPE	0.0003	1.0190	14.6719	0.8500
	WM1	-0.0031	1.0725	15.5357	0.8926
	WM2	-0.0083	1.0897	17.7213	0.9048
	WM3	-0.0008	1.0787	17.5702	0.8979
	WM4	-0.0056	1.0925	16.1753	0.9084

Table A.19: Frequency distribution of model identification for simulation of ARIMA(0,2,1); $\Theta = 0.9$, N=1000

Sample Size	Criterion	Candidate Model								
		ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (0,1,3)	ARIMA (0,2,1)	ARIMA (0,2,2)	ARIMA (0,2,3)	ARIMA (0,3,1)	ARIMA (0,3,2)	ARIMA (0,3,3)
n=100	MIMSA	162	0	0	552	0	0	286	0	0
	BIC	0	0	0	819	35	14	0	117	15
	MAPE	29	28	33	20	18	30	144	332	366
	WM1	107	67	33	167	81	152	226	170	97
	WM2	135	115	90	200	142	95	141	84	0
	WM3	29	36	55	108	109	103	157	193	210
	WM4	44	55	47	142	120	84	172	188	148
n=250	MIMSA	27	0	0	18	0	0	955	0	0
	BIC	0	0	0	949	16	5	0	28	2
	MAPE	20	14	18	26	19	33	158	360	352
	WM1	162	28	13	270	34	14	386	60	33
	WM2	93	74	65	227	149	105	175	112	0
	WM3	28	35	34	128	137	126	142	189	181
	WM4	31	44	37	156	158	109	150	165	150
n=500	MIMSA	150	0	0	453	0	0	397	0	0
	BIC	0	0	0	979	7	1	0	12	1
	MAPE	14	17	11	39	18	31	146	379	345
	WM1	199	19	3	370	9	5	362	28	5
	WM2	64	86	80	181	178	123	136	152	0
	WM3	20	27	41	153	154	133	129	155	188
	WM4	29	44	56	155	117	145	135	158	161

Table A.20: Forecast measures for the simulation of ARIMA(0,2,1) process; $\Theta = 0.9$

Sample	Selection	Forecast Measure			
Size	Criterion	ME	RMSE	MAPE	MASE
	MIMSA	-0.0167	1.5059	5.0945	0.1581
	BIC	-0.00005	0.8806	0.4292	0.0933
	MAPE	-0.0409	1.3267	0.4764	0.1087
n=100	WM1	0.0042	2.4440	0.6606	0.1633
	WM2	0.2101	3.0680	4.7787	0.2004
	WM3	-0.0692	1.4235	0.5577	0.1167
	WM4	-0.0053	1.7155	0.6110	0.1304
	MIMSA	0.0214	1.3250	0.4671	0.0908
	BIC	0.0075	0.9763	0.1357	0.0689
	MAPE	0.0195	1.3531	0.1514	0.0839
n=250	WM1	0.0602	3.2941	0.3448	0.1571
	WM2	-0.2407	3.0578	0.3387	0.14778
	WM3	-0.0038	1.7665	0.2187	0.0980
	WM4	-0.1199	1.9446	0.2448	0.1038
	MIMSA	0.0708	2.0311	0.7794	0.1193
	BIC	0.0007	0.9963	0.1137	0.05199
	MAPE	0.0235	1.5514	0.1246	0.0671
n=500	WM1	-0.0189	4.4834	0.2120	0.1556
	WM2	-0.1884	3.6534	0.4105	0.1248
	WM3	-0.0788	1.9039	0.1387	0.0784
	WM4	0.0510	2.2791	0.1661	0.0904

Table A.21: Frequency distribution of model identification for simulation of ARIMA(0,2,1); $\Theta = 0.5$, N=1000

Sample Size	Criterion	Candidate Model								
		ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (0,1,3)	ARIMA (0,2,1)	ARIMA (0,2,2)	ARIMA (0,2,3)	ARIMA (0,3,1)	ARIMA (0,3,2)	ARIMA (0,3,3)
n=100	MIMSA	162	0	0	537	0	0	301	0	0
	BIC	0	0	0	898	32	7	1	48	14
	MAPE	42	24	44	33	20	42	193	272	330
	WM1	95	59	51	126	74	45	260	194	96
	WM2	138	122	100	162	144	98	146	90	0
	WM3	20	47	64	112	90	92	177	214	184
	WM4	47	62	61	126	103	98	192	164	147
n=250	MIMSA	35	0	0	17	0	0	948	0	0
	BIC	0	0	0	938	20	2	0	37	3
	MAPE	29	24	34	58	36	56	167	312	284
	WM1	158	32	25	237	46	18	384	64	36
	WM2	106	93	69	215	162	113	140	102	0
	WM3	28	32	44	124	147	110	166	165	184
	WM4	31	47	41	158	150	112	187	166	108
n=500	MIMSA	144	0	0	428	0	0	428	0	0
	BIC	0	0	0	970	16	0	0	14	0
	MAPE	21	21	17	73	44	61	166	284	313
	WM1	198	13	3	389	13	12	332	32	8
	WM2	75	68	75	186	161	125	181	129	0
	WM3	33	35	66	121	168	125	146	147	159
	WM4	29	44	51	146	154	138	134	161	143

Table A.22: Forecast measures for the simulation of ARIMA(0,2,1) process; $\Theta = 0.5$

Sample Size	Selection Criterion	ME	RMSE	MAPE	MASE
		Forecast Measure			
n=100	MIMSA	0.1127	1.2472	1.3303	0.1680
	BIC	0.0127	0.8027	0.5719	0.1094
	MAPE	0.0246	1.1994	1.0071	0.1268
	WM1	0.0076	1.8823	1.0460	0.1716
	WM2	0.0380	2.4284	1.2222	0.2172
	WM3	0.0459	1.3112	0.7277	0.1314
	WM4	-0.0758	1.4320	1.0344	0.1492
n=250	MIMSA	0.0272	1.2167	0.6764	0.0997
	BIC	0.0087	0.9318	0.2525	0.0789
	MAPE	-0.0636	1.5864	0.3405	0.1048
	WM1	0.1045	3.0325	0.4978	0.1665
	WM2	-0.1023	2.8888	0.4787	0.1663
	WM3	0.0281	1.5942	0.3241	0.1086
	WM4	-0.0548	1.5598	0.3807	0.1122
n=500	MIMSA	-0.0238	1.7192	0.3534	0.1244
	BIC	-0.0002	0.9675	0.0950	0.0627
	MAPE	0.0304	1.5192	0.1065	0.0825
	WM1	-0.0783	3.7337	0.2168	0.1595
	WM2	0.2752	2.7687	0.1908	0.1307
	WM3	0.1685	2.1708	0.1282	0.1016
	WM4	0.2353	2.1015	0.1405	0.0987

Table A.23: Frequency distribution of model identification for simulation of ARIMA(0,2,2); $\Theta_1 = 0.01$, $\Theta_2 = 0.9$, N=1000

Sample Size	Criterion	Candidate Model								
		ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (0,1,3)	ARIMA (0,2,1)	ARIMA (0,2,2)	ARIMA (0,2,3)	ARIMA (0,3,1)	ARIMA (0,3,2)	ARIMA (0,3,3)
n=100	MIMSA	170	0	0	550	0	0	280	0	0
	BIC	0	0	0	0	808	44	0	0	148
	MAPE	10	13	31	42	63	71	60	159	551
	WM1	110	64	41	152	83	46	250	118	136
	WM2	151	110	94	198	128	92	143	84	0
	WM3	14	19	59	82	133	125	147	196	225
	WM4	46	50	40	125	143	126	171	149	150
n=250	MIMSA	24	0	0	19	0	0	957	0	0
	BIC	0	0	0	0	920	23	0	0	57
	MAPE	12	11	21	62	68	90	79	108	549
	WM1	139	26	8	268	32	19	422	55	31
	WM2	96	89	68	215	179	104	153	96	0
	WM3	16	36	34	98	163	154	145	168	186
	WM4	40	35	39	163	163	127	155	131	147
n=500	MIMSA	127	0	0	422	0	0	451	0	0
	BIC	0	0	0	957	17	0	0	0	26
	MAPE	5	3	5	67	91	80	81	123	545
	WM1	178	8	1	385	25	1	378	19	5
	WM2	62	76	65	166	176	155	145	155	0
	WM3	22	29	52	133	164	156	130	152	162
	WM4	28	33	46	142	171	154	142	135	149

Table A.24: Forecast measures for the simulation of ARIMA(0,2,2) process; $\Theta_1 = 0.01$, $\Theta_2 = 0.9$

Sample Size	Selection Criterion	Forecast Measure ME	RMSE	MAPE	MASE
n=100	MIMSA	0.0370	1.6364	1.8035	0.1842
	BIC	-0.0054	0.8651	0.5129	0.1007
	MAPE	-0.0530	1.1340	0.7783	0.1091
	WM1	0.0759	2.5739	0.8941	0.1869
	WM2	-0.1646	3.0805	1.9592	0.2293
	WM3	0.0064	1.3767	0.7849	0.1293
	WM4	0.0576	1.7406	0.9910	0.1491
n=250	MIMSA	0.0323	1.3655	0.3568	0.10433
	BIC	0.0181	0.9836	0.1209	0.0749
	MAPE	0.0060	1.3524	0.1467	0.0913
	WM1	0.2055	2.9159	0.2616	0.1605
	WM2	0.0871	3.2181	0.3515	0.1665
	WM3	0.0494	1.5942	1.7223	0.1058
	WM4	-0.0222	1.8778	0.2214	0.1189
n=500	MIMSA	-0.1206	2.3393	0.4096	0.1166
	BIC	-0.0045	1.0062	0.1034	0.0474
	MAPE	0.0677	1.2081	0.1057	0.0548
	WM1	-0.2626	4.4657	0.2270	0.1467
	WM2	0.3090	3.4616	0.2603	0.1188
	WM3	0.0370	2.3057	0.2985	0.0830
	WM4	0.0172	2.3322	0.1383	0.0857

Table A.25: Frequency distribution of model identification for simulation of ARIMA(0,2,2); $\Theta_1 = 0.4$, $\Theta_2 = 0.5$, N=1000

Sample Size	Criterion	Candidate Model								
		ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (0,1,3)	ARIMA (0,2,1)	ARIMA (0,2,2)	ARIMA (0,2,3)	ARIMA (0,3,1)	ARIMA (0,3,2)	ARIMA (0,3,3)
n=100	MIMSA	153	0	0	632	0	0	215	0	0
	BIC	0	0	0	12	863	35	20	0	70
	MAPE	21	24	30	63	37	32	141	119	533
	WM1	119	67	39	146	76	41	249	145	118
	WM2	135	114	79	230	127	94	140	81	0
	WM3	13	24	69	108	132	94	182	166	212
	WM4	46	50	39	147	116	91	199	172	140
n=250	MIMSA	28	0	0	32	0	0	940	0	0
	BIC	0	0	0	0	939	28	1	0	32
	MAPE	7	6	16	62	49	38	149	141	532
	WM1	116	20	22	280	27	22	424	50	39
	WM2	111	81	58	215	166	109	161	99	0
	WM3	16	25	46	123	132	144	163	170	181
	WM4	48	32	46	155	135	114	170	156	144
n=500	MIMSA	132	0	0	419	0	0	449	0	0
	BIC	0	0	0	0	978	12	0	0	10
	MAPE	7	9	15	62	42	34	181	114	536
	WM1	178	9	2	361	16	5	404	19	6
	WM2	56	86	86	181	182	124	148	137	0
	WM3	24	23	49	136	150	146	153	152	167
	WM4	21	40	40	129	163	146	162	159	140

Table A.26: Forecast measures for the simulation of ARIMA(0,2,2) process; $\Theta_1 = 0.4$, $\Theta_2 = 0.5$

Sample Size	Selection Criterion	Forecast Measure ME	RMSE	MAPE	MASE
	MIMSA	0.0337	1.3777	2.0619	0.1519
	BIC	0.0107	0.7363	1.0357	0.0797
	MAPE	0.0171	1.1188	1.2831	0.0953
n=100	WM1	-0.0006	2.4263	1.5210	0.1648
	WM2	-0.0688	2.8209	1.6929	0.1958
	WM3	-0.0148	1.2722	1.3339	0.1062
	WM4	0.0189	1.5563	0.6924	0.1250
	MIMSA	-0.0484	1.1641	0.3557	0.0869
	BIC	-0.0067	0.9064	0.1522	0.0681
	MAPE	0.0117	1.1451	0.1607	0.0776
n=250	WM1	-0.3409	2.5365	0.2967	0.1363
	WM2	-0.2038	3.1712	0.3011	0.1577
	WM3	0.0105	1.5493	0.1949	0.0944
	MIMSA	-0.0035	2.2781	0.5515	0.1129
	BIC	-0.0037	0.9583	0.2614	0.0500
	MAPE	0.1388	1.3902	0.2765	0.0606
n=500	WM1	-0.6236	4.2001	0.2966	0.1423
	WM2	-0.4941	3.5964	0.3756	0.1214
	WM3	0.08131	1.8900	0.2230	0.0801
	WM4	-0.0995	2.0989	0.5722	0.0814

APPENDIX B

SIMULATION RESULTS FOR RWMS

Table B.1: Frequency distribution of model identification for simulation of AR(2), N=1000

Parameter	Sample Size	Criterion	Candidate Model					
			AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
$\Phi_1 = 0.01$, $\Phi_2 = 0.9$	n=100	MIMSA	95	905	0	0	0	0
		BIC	0	890	78	16	11	5
		MAPE	151	89	76	89	163	432
		WM4	147	180	180	188	169	136
		WM4R	160	217	210	169	143	101
	n=250	MIMSA	0	1000	0	0	0	0
		BIC	0	947	35	13	3	2
		MAPE	302	110	72	83	125	308
		WM4	134	183	176	173	162	172
		WM4R	132	220	177	164	162	145
	n=500	MIMSA	0	1000	0	0	0	0
		BIC	0	974	23	3	0	0
		MAPE	460	76	77	98	91	198
		WM4	129	198	184	176	144	169
		WM4R	154	179	167	180	170	150

Note: WM4-MIMSA&MAPE&BIC and WM4R-MIMSA&MAPE&BIC by rolling windows.

Table B.2: Frequency distribution of model identification for simulation of AR(2), N=1000

Parameter	Sample Size	Criterion	Candidate Model					
			AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
$\Phi_1 = 0.4$, $\Phi_2 = 0.5$	n=100	MIMSA	751	249	0	0	0	0
		BIC	15	867	79	23	12	4
		MAPE	169	77	77	85	160	432
		WM4	176	175	174	157	158	160
		WM4R	224	214	193	154	121	94
	n=250	MIMSA	300	700	0	0	0	0
		BIC	0	939	43	14	4	0
		MAPE	244	96	84	96	146	334
		WM4	150	188	162	175	166	159
		WM4R	190	179	166	151	154	160
	n=500	MIMSA	78	922	0	0	0	
		BIC	0	976	19	4	1	0
		MAPE	263	124	95	116	111	291
		WM4	138	169	180	170	165	178
		WM4R	158	191	161	177	162	151

Table B.3: Forecast measures for the simulation of AR(2) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Phi_1 = 0.01$, $\Phi_2 = 0.9$	n=100	MIMSA	0	0.8644	285.1367	0.3452
		BIC	0	0.8210	286.1744	0.3268
		MAPE	0	0.7278	324.1587	0.2729
		WM4	0	0.7865	236.1453	0.3010
		WM4R	0	0.8009	219.4950	0.3089
	n=250	MIMSA	0	0.9429	237.4356	0.3294
		BIC	0	0.9412	237.1695	0.3286
		MAPE	0	1.1145	237.7818	0.3813
		WM4	0	0.9948	234.8695	0.3417
		WM4R	0	1.0077	227.0138	0.3455
	n=500	MIMSA	0	0.9725	494.8115	0.3384
		BIC	0	0.9721	492.7796	0.3382
		MAPE	0	1.3766	493.0073	0.4715
		WM4	0	1.0722	494.2597	0.3699
		WM4R	0	1.0980	470.6681	0.3768
$\Phi_1 = 0.4$, $\Phi_2 = 0.5$	n=100	MIMSA	0	0.9082	356.9583	0.7501
		BIC	0	0.8172	817.2043	0.6623
		MAPE	0	0.6873	154.8705	0.5272
		WM4	0	0.7444	292.9195	0.5841
		WM4R	0	0.7726	321.5321	0.6137
	n=250	MIMSA	0	0.9734	329.6362	0.7980
		BIC	0	0.9345	289.1377	0.7629
		MAPE	0	0.9234	287.3283	0.7411
		WM4	0	0.9209	282.8567	0.7413
		WM4R	0	0.9259	282.3579	0.7461
	n=500	MIMSA	0	0.9806	366.7546	0.8035
		BIC	0	0.9694	359.6427	0.7940
		MAPE	0	0.9870	337.9526	0.8023
		WM4	0	0.9707	342.9341	0.7889
		WM4R	0	0.9766	363.3709	0.7934

Table B.4: Frequency distribution of model identification for simulation of AR(3), N=1000

Parameter	Sample Size	Criterion	Candidate Model					
			AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
$\Phi_1 = 0.01$, $\Phi_2 = 0.01$, $\Phi_3 = 0.9$	n=100	MIMSA	6	36	958	0	0	0
		BIC	0	0	886	79	25	10
		MAPE	121	78	110	104	137	450
		WM4	149	128	183	184	180	176
		WM4R	170	163	204	185	158	120
	n=250	MIMSA	0	0	1000	0	0	0
		BIC	0	0	946	43	8	3
		MAPE	227	104	124	117	140	288
		WM4	124	144	205	171	173	183
		WM4R	135	158	191	189	181	146
	n=500	MIMSA	0	0	1000	0	0	0
		BIC	0	0	975	25	0	0
		MAPE	320	159	105	74	97	245
		WM4	119	143	184	165	198	191
		WM4R	149	140	185	187	171	168
$\Phi_1 = 0.01$, $\Phi_2 = 0.01$, $\Phi_3 = 0.5$	n=100	MIMSA	749	10	241	0	0	0
		BIC	61	4	833	65	24	13
		MAPE	468	324	38	48	44	78
		WM4	198	198	160	161	147	136
		WM4R	257	228	160	150	116	89
	n=250	MIMSA	10	0	990	0	0	0
		BIC	0	0	947	44	6	3
		MAPE	587	360	12	10	8	23
		WM4	199	187	171	128	152	163
		WM4R	193	196	164	167	150	130
	n=500	MIMSA	0	0	1000	0	0	0
		BIC	0	0	973	18	7	2
		MAPE	656	332	4	1	2	5
		WM4	189	203	147	154	157	150
		WM4R	231	173	159	144	136	157

Note: WM4-MIMSA&MAPE&BIC and WM4R-MIMSA&MAPE&BIC by rolling windows.

Table B.5: Forecast measures for the simulation of AR(3) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Phi_1 = 0.01$, $\Phi_2 = 0.01$, $\Phi_3 = 0.9$	n=100	MIMSA	0	0.7819	550.7273	2.7743
		BIC	0	0.7514	527.0401	0.2662
		MAPE	0	0.7724	292.0430	0.2638
		WM4	0	0.8504	318.6244	0.2948
		WM4R	0	0.9147	611.9534	0.3145
	n=250	MIMSA	0	0.9062	205.6930	0.3234
		BIC	0	0.9048	205.5961	0.3226
		MAPE	0	1.1448	219.8594	0.4001
		WM4	0	1.0837	211.9682	0.3779
		WM4R	0	1.1091	221.2660	0.3876
	n=500	MIMSA	0	0.9552	377.5534	0.3214
		BIC	0	0.9550	377.4895	0.3213
		MAPE	0	1.4179	373.7476	0.4696
		WM4	0	1.1945	382.6723	0.3978
		WM4R	0	1.2190	324.1684	0.4077
$\Phi_1 = 0.01$, $\Phi_2 = 0.01$, $\Phi_3 = 0.5$	n=100	MIMSA	0	0.9359	286.6013	0.5806
		BIC	0	0.7913	273.0440	0.4733
		MAPE	0	0.8946	297.9196	0.5461
		WM4	0	0.7971	264.0045	0.4753
		WM4R	0	0.8185	293.2474	0.4915
	n=250	MIMSA	0	0.9194	243.6363	0.5654
		BIC	0	0.9167	242.0286	0.5632
		MAPE	0	1.0591	209.7844	0.6577
		WM4	0	0.9565	229.4137	0.5845
		WM4R	0	0.9571	231.4726	0.5854
	n=500	MIMSA	0	0.9617	391.9856	0.5917
		BIC	0	0.9614	391.9649	0.5913
		MAPE	0	1.1098	220.9843	0.6871
		WM4	0	1.0121	360.1635	0.6210
		WM4R	0	1.0131	337.4989	0.6218

Table B.6: Frequency distribution of model identification for simulation of MA(1)

Parameter	Sample Size	Criterion	Candidate Model					
			MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
$\Theta = 0.85$	n=100	MIMSA	957	42	31	0	0	0
		BIC	895	48	13	13	16	15
		MAPE	219	171	138	130	131	211
		WM4	192	179	164	166	140	159
		WM4R	229	198	194	151	125	103
	n=250	MIMSA	845	155	0	0	0	0
		BIC	967	24	9	0	0	0
		MAPE	270	121	117	116	139	237
		WM4	170	159	168	169	161	173
		WM4R	182	164	174	178	162	140
	n=500	MIMSA	758	239	3	0	0	0
		BIC	981	16	3	0	0	0
		MAPE	308	115	117	106	124	230
		WM4	158	187	180	174	152	149
		WM4R	163	169	179	167	160	162
$\Theta = 0.5$	n=100	MIMSA	967	29	4	0	0	0
		BIC	897	49	13	12	13	16
		MAPE	255	166	131	138	147	163
		WM4	187	187	157	174	171	124
		WM4R	229	210	189	143	129	100
	n=250	MIMSA	912	86	2	0	0	0
		BIC	979	16	4	1	0	0
		MAPE	303	155	115	107	131	189
		WM4	167	168	166	159	172	168
		WM4R	161	186	174	157	160	162
	n=500	MIMSA	899	101	0	0	0	0
		BIC	988	12	0	0	0	0
		MAPE	268	174	124	140	116	178
		WM4	160	161	171	173	179	156
		WM4R	166	161	166	170	171	166

Table B.7: Forecast measures for the simulation of MA(1) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Theta = 0.85$	n=100	MIMSA	-0.0016	0.9484	211.4117	0.7161
		BIC	-0.0113	0.9258	196.8803	0.6995
		MAPE	0.0223	0.7908	187.4695	0.5926
		WM4	0.02110	0.7699	189.3570	0.5795
		WM4R	-0.0029	0.7911	195.5754	0.5952
	n=250	MIMSA	0.0001.	0.9794	572.8028	0.7543
		BIC	0.0003	0.9804	559.1144	0.7550
		MAPE	-0.0013	0.9445	561.8923	0.7268
		WM4	-0.0024	0.9444	589.6925	0.7273
		WM4R	-0.0011	0.9451	584.4724	0.7274
	n=500	MIMSA	0.00003	0.9958	411.7141	0.7594
		BIC	0.0001	0.9972	412.3460	0.7604
		MAPE	-0.0001	0.9842	422.2120	0.7503
		WM4	-0.0007	0.9838	407.7660	0.7504
		WM4R	-0.0003	0.9833	408.9636	0.7499
$\Theta = 0.5$	n=100	MIMSA	-0.0059	0.9147	273.3082	0.7512
		BIC	-0.0108	0.9080	272.3083	0.7446
		MAPE	-0.0166	0.7706	251.2846	0.6258
		WM4	-0.0217	0.7913	284.8977	0.6434
		WM4R	-0.0218	0.7870	272.4639	0.6413
	n=250	MIMSA	0.0004	0.9732	409.0689	0.8017
		BIC	0.0005	0.9738	407.8465	0.8021
		MAPE	0.0030	0.9395	570.5029	0.7737
		WM4	0	0.9354	454.7373	0.7704
		WM4R	0	0.9361	578.9828	0.7714
	n=500	MIMSA	0.0001	0.9866	433.0066	0.8115
		BIC	0.0001	0.9870	430.4207	0.8119
		MAPE	-0.0001	0.9747	438.9482	0.8016
		WM4	0	0.9727	626.8861	0.8001
		WM4R	-0.0001	0.9728	426.5862	0.8000

Table B.8: Frequency distribution of model identification for simulation of MA(2)

Parameter	Sample Size	Criterion	Candidate Model					
			MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
$\Theta_1 = 0.01$, $\Theta_2 = 0.9$	n=100	MIMSA	0	997	3	0	0	0
		BIC	0	902	48	22	13	15
		MAPE	800	55	34	24	34	53
		WM4	171	181	182	143	172	151
		WM4R	250	210	176	154	117	93
	n=250	MIMSA	0	997	1	2	0	0
		BIC	0	967	22	9	1	1
		MAPE	969	12	5	3	2	9
		WM4	181	172	170	166	154	157
		WM4R	202	187	171	159	138	143
	n=500	MIMSA	0	963	24	13	0	0
		BIC	0	974	24	2	0	0
		MAPE	989	2	1	2	2	4
		WM4	196	168	162	140	162	172
		WM4R	226	140	148	160	172	154
$\Theta_1 = 0.4$, $\Theta_2 = 0.5$	n=100	MIMSA	99	898	3	0	0	0
		BIC	29	874	53	13	6	25
		MAPE	728	59	52	40	51	70
		WM4	203	175	189	161	126	146
		WM4R	244	190	190	152	125	99
	n=250	MIMSA	0	0	973	26	1	0
		BIC	0	966	28	4	2	0
		MAPE	883	36	15	22	6	38
		WM4	184	167	149	176	168	156
		WM4R	198	173	159	162	159	149
	n=500	MIMSA	0	957	41	2	0	0
		BIC	0	987	13	0	0	0
		MAPE	934	19	7	10	6	24
		WM4	193	169	172	153	163	150
		WM4R	187	190	166	157	149	151

Table B.9: Forecast measures for the simulation of MA(2) process

Parameter	Sample Size	Criterion	Forecast Measure ME	RMSE	MAPE	MASE
$\Theta_1 = 0.01$, $\Theta_2 = 0.9$	n=100	MIMSA	-0.0097	0.9767	463.1308	0.5362
		BIC	-0.0166	0.9574	444.4672	0.5253
		MAPE	-0.0027	1.1553	444.6172	0.6273
		WM4	-0.0065	0.8875	452.8269	0.4805
		WM4R	-0.0041	0.9421	509.9172	0.5123
	n=250	MIMSA	-0.0042	1.0128	353.2220	0.5507
		BIC	-0.0038	1.0122	354.4149	0.5504
		MAPE	-0.0008	1.2842	210.7962	0.6963
		WM4	0.0002	1.0402	295.2720	0.5646
		WM4R	-0.0003	1.0491	305.4201	0.5693
	n=500	MIMSA	-0.0017	1.0122	459.6984	0.5406
		BIC	-0.0018	1.0123	458.2981	0.5407
		MAPE	-0.0003	1.3159	215.3293	0.7034
		WM4	-0.0022	1.0625	405.3197	0.5677
		WM4R	-0.0033	1.0730	362.8454	0.5740
$\Theta_1 = 0.4$, $\Theta_2 = 0.5$	n=100	MIMSA	0.0170	0.8577	285.7435	0.7281
		BIC	0.0218	0.8381	290.3448	0.7336
		MAPE	0.0106	0.8996	270.3296	0.7479
		WM4	0.0324	0.7754	274.7784	0.6517
		WM4R	0.0330	0.7942	287.8871	0.6727
	n=250	MIMSA	0.00031	0.9618	402.0791	0.7738
		BIC	0.0005	0.9616	403.9318	0.7737
		MAPE	0.00007	1.0552	308.1390	0.8494
		WM4	0.00004	0.9527	380.5750	0.7670
		WM4R	0.0002	0.9568	364.5576	0.76975
	n=500	MIMSA	-0.0008	0.9873	378.8124	0.7795
		BIC	-0.0008	0.9876	378.6452	0.7797
		MAPE	-0.0003	1.0939	447.0469	0.8635
		WM4	-0.0009	1.0019	315.3252	0.7906
		WM4R	-0.00104	1.0014	525.3202	0.7902

Table B.10: Frequency distribution of model identification for simulation of MA(3)

Parameter	Sample Size	Criterion	Candidate Model					
			MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)
		MIMSA	78	530	392	0	0	0
		BIC	0	0	910	53	21	16
	n=100	MAPE	681	92	79	42	35	71
		WM4	203	182	151	177	144	143
		WM4R	253	175	177	187	123	85
		MIMSA	0	410	589	1	0	0
		BIC	0	0	972	18	8	2
$\Theta_1 = 0.01$, $\Theta_2 = 0.01$, $\Theta_3 = 0.9$	n=250	MAPE	931	46	4	3	9	7
		WM4	184	150	176	152	158	180
		WM4R	206	175	169	158	168	124
		MIMSA	0	299	700	1	0	0
		BIC	0	0	982	17	1	0
	n=500	MAPE	957	35	3	2	1	2
		WM4	218	156	156	176	144	150
		WM4R	209	139	166	175	135	176
	MIMSA	640	246	114	0	0	0	
		BIC	106	3	792	46	29	24
		MAPE	642	189	58	30	38	43
		WM4	212	171	176	131	156	154
		WM4R	242	192	195	163	109	99
		MIMSA	0	0	974	22	2	2
		BIC	42	105	850	3	0	0
$\Theta_1 = 0.01$, $\Theta_2 = 0.01$, $\Theta_3 = 0.5$	n=250	MAPE	769	196	7	6	4	18
		WM4	199	164	158	179	162	138
		WM4R	217	180	154	138	165	146
		MIMSA	0	10	988	2	0	0
		BIC	0	0	979	16	5	0
	n=500	MAPE	787	204	3	2	1	3
		WM4	219	185	146	151	147	152
		WM4R	195	191	155	143	146	170

Table B.11: Forecast measures for the simulation of MA(3) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Theta_1 = 0.01$, $\Theta_2 = 0.01$, $\Theta_3 = 0.9$	n=100	MIMSA	-0.0032	0.9908	960.1285	0.5507
		BIC	0.0263	0.9071	686.7539	0.4989
		MAPE	0.0105	1.0932	949.3890	0.6144
		WM4	0.0275	0.9169	657.3624	0.5072
		WM4R	0.0281	0.9665	672.1545	0.5369
	n=250	MIMSA	-0.0006	1.0995	460.0245	0.5941
		BIC	0.0004	1.0057	458.3745	0.5428
		MAPE	0.0002	1.2763	252.2144	0.6877
		WM4	-0.0003	1.0739	410.5018	0.5783
		WM4R	-0.0042	1.0897	409.8296	0.5879
	n=500	MIMSA	-0.0001	1.0914	346.6737	0.5799
		BIC	-0.0003	1.0108	373.6090	0.5371
		MAPE	-0.0004	1.3148	201.3579	0.6993
		WM4	0.001	1.1149	315.0312	0.5933
		WM4R	-0.0007	1.1077	358.7120	0.5890
$\Theta_1 = 0.01$, $\Theta_2 = 0.01$, $\Theta_3 = 0.5$	n=100	MIMSA	-0.0098	0.9562	246.3810	0.6430
		BIC	-0.0180	0.8418	314.4942	0.5639
		MAPE	-0.0198	0.9425	279.4435	0.6305
		WM4	-0.0263	0.8115	281.6955	0.5455
		WM4R	-0.0244	0.8320	333.6028	0.5582
	n=250	MIMSA	0.0010	0.9702	307.6128	0.6265
		BIC	0.0027	0.9550	309.6855	0.6165
		MAPE	0.0006	1.0635	224.6155	0.6868
		WM4	0.0015	0.9801	289.9742	0.6324
		WM4R	0.0022	0.9810	301.2317	0.6328
	n=500	MIMSA	-0.0004	0.9849	280.6843	0.6287
		BIC	-0.0004	0.9838	279.900	0.6281
		MAPE	-0.00002	1.0984	189.6761	0.7004
		WM4	-0.0009	1.0235	256.9938	0.6526
		WM4R	-0.00006	1.0205	260.3194	0.6510

Table B.12: Frequency distribution of model identification for simulation of ARIMA(1,1,0)

Parameter	Sample Size	Criterion	Candidate Model ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (3,1,0)	ARIMA (1,2,0)	ARIMA (2,2,0)	ARIMA (3,2,0)	ARIMA (1,3,0)	ARIMA (2,3,0)	ARIMA (3,3,0)
$\Phi = 0.9$	n=100	MIMSA	303	0	0	628	0	0	69	0	0
		BIC	594	20	5	327	43	11	0	0	0
		MAPE	27	31	210	17	27	448	20	39	181
		WM4	107	106	106	119	117	145	68	102	130
		WM4R	113	131	133	117	137	126	90	126	27
	n=250	MIMSA	264	0	0	725	0	0	11	0	0
		BIC	929	30	4	30	4	3	0	0	0
		MAPE	40	52	262	23	36	405	31	39	112
		WM4	104	108	113	121	117	118	99	101	119
		WM4R	107	126	114	153	129	126	110	119	16
	n=500	MIMSA	256	0	0	741	0	0	3	0	0
		BIC	962	28	8	9	0	0	0	0	0
		MAPE	26	48	308	29	41	380	15	49	104
		WM4	110	126	120	122	128	107	66	109	112
		WM4R	119	126	128	119	123	120	114	142	9
$\Phi = 0.5$	n=100	MIMSA	50	0	0	233	0	0	717	0	0
		BIC	908	63	23	0	5	1	0	0	0
		MAPE	48	68	425	29	42	244	14	32	98
		WM4	109	111	128	122	132	124	68	104	102
		WM4R	139	150	117	134	122	137	78	97	26
	n=250	MIMSA	36	0	0	226	0	0	738	0	0
		BIC	954	40	6	0	0	0	0	0	0
		MAPE	54	92	474	35	57	186	18	29	55
		WM4	120	123	139	117	122	118	68	91	102
		WM4R	145	138	123	112	139	113	101	112	17
	n=500	MIMSA	42	0	0	163	0	0	795	0	0
		BIC	961	32	7	0	0	0	0	0	0
		MAPE	68	106	514	42	51	136	18	27	38
		WM4	111	117	141	97	132	137	72	100	93
		WM4R	126	139	134	111	119	141	97	125	8

Table B.13: Forecast measures for the simulation of ARIMA(1,1,0) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Theta = 0.85$	n=100	MIMSA	-0.0050	0.9229	5.9222	0.4544
		BIC	-0.0054	0.9035	6.0077	0.4440
		MAPE	-0.0023	0.8296	4.5069	0.3855
		WM4	-0.0077	0.8803	4.7404	0.4181
		WM4R	-0.0068	0.8873	4.7002	0.4223
	n=250	MIMSA	0.0012	0.9763	5.4391	0.4536
		BIC	0.0042	0.9625	6.4391	0.4482
		MAPE	0.0036	0.9642	4.9881	0.4383
		WM4	0.0025	1.0049	5.1490	0.4603
		WM4R	0.0017	0.9988	4.9761	0.4579
	n=500	MIMSA	0.0013	0.9987	4.0671	0.4511
		BIC	0.0028	0.9837	4.0453	0.4448
		MAPE	0	1.0031	4.1514	0.4483
		WM4	0.0010	1.0356	4.1035	0.4647
		WM4R	0.0029	1.0406	4.0947	0.4671
$\Theta = 0.5$	n=100	MIMSA	-0.0028	1.2148	41.5620	0.9926
		BIC	0.0050	0.9059	23.8374	0.7609
		MAPE	-0.0011	0.8644	23.9092	0.6844
		WM4	-0.0026	0.9432	35.5799	0.7572
		WM4R	-0.0024	0.9444	24.2516	0.7620
	n=250	MIMSA	-0.0012	1.3441	44.7395	1.1429
		BIC	0.0057	0.9640	25.8181	0.8312
		MAPE	0.0046	0.9869	27.4607	0.8323
		WM4	0.0040	1.0664	38.8766	0.9030
		WM4R	0.0014	1.0700	31.9618	0.9080
	n=500	MIMSA	-0.0001	1.3998	24.3548	1.2020
		BIC	-0.0037	0.9810	19.2181	0.8481
		MAPE	-0.0031	1.0147	20.2526	0.8693
		WM4	-0.0005	1.1040	19.2274	0.9458
		WM4R	-0.0027	1.1027	20.4406	0.9457

Table B.14: Frequency distribution of model identification for simulation of ARIMA(1,2,0)

Parameter	Sample Size	Criterion	Candidate Model ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (3,1,0)	ARIMA (1,2,0)	ARIMA (2,2,0)	ARIMA (3,2,0)	ARIMA (1,3,0)	ARIMA (2,3,0)	ARIMA (3,3,0)
$\Phi = 0.9$	n=100	MIMSA	709	0	0	8	0	0	283	0	0
		BIC	0	7	1	583	16	4	329	42	18
		MAPE	6	10	20	3	9	148	7	35	762
		WM4	29	89	40	209	110	58	273	147	45
		WM4R	49	83	141	113	152	45	145	166	106
	n=250	MIMSA	515	0	0	10	0	0	475	0	0
		BIC	0	9	0	910	27	5	39	7	3
		MAPE	6	0	21	1	4	150	14	23	781
		WM4	51	105	73	141	126	92	163	150	99
		WM4R	65	115	80	168	126	49	185	156	56
	n=500	MIMSA	423	0	0	17	0	0	560	0	0
		BIC	0	4	1	948	41	5	1	0	0
		MAPE	9	1	14	1	4	153	14	27	777
		WM4	71	120	106	112	111	126	122	125	107
		WM4R	59	126	124	119	133	100	148	151	40
$\Phi = 0.5$	n=100	MIMSA	913	0	0	24	0	0	63	0	0
		BIC	0	6	1	901	67	20	0	2	3
		MAPE	22	7	46	2	16	259	30	34	584
		WM4	42	92	110	102	108	122	133	149	142
		WM4R	84	98	134	116	134	102	125	151	56
	n=250	MIMSA	691	0	0	34	0	0	275	0	0
		BIC	0	8	0	937	43	12	0	0	0
		MAPE	31	6	23	1	5	318	31	52	533
		WM4	69	93	98	130	120	130	136	131	121
		WM4R	104	106	144	121	113	114	116	160	22
	n=500	MIMSA	511	0	0	37	0	0	452	0	0
		BIC	0	6	1	966	23	4	0	0	0
		MAPE	30	5	18	3	9	352	28	55	500
		WM4	90	123	119	112	119	106	117	95	119
		WM4R	114	123	115	124	127	96	149	136	16

Table B.15: Forecast measures for the simulation of ARIMA(1,2,0) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Phi = 0.85$	n=100	MIMSA	0.0018	1.2885	0.2352	0.0354
		BIC	0.00003	0.8805	0.2922	0.0212
		MAPE	0.0056	0.7718	0.2619	0.0173
		WM4	-0.0041	0.8610	0.2653	0.0203
		WM4R	0.0041	0.8480	0.2005	0.0201
	n=250	MIMSA	0.0063	1.4439	0.0766	0.0220
		BIC	0.0023	0.9574	0.0427	0.0139
		MAPE	0.0003	0.9334	0.0431	0.0142
		WM4	0.0002	0.9902	5.1490	0.4603
		WM4R	0.0047	1.0137	0.0473	0.0146
	n=500	MIMSA	0.0033	1.4533	0.0171	0.0145
		BIC	0.0012	0.9782	0.0099	0.0093
		MAPE	0.0037	0.9823	0.0097	0.0092
		WM4	0.0022	1.0531	0.0108	0.0099
		WM4R	-0.0019	1.0371	0.0106	0.0098
$\Phi = 0.5$	n=100	MIMSA	0.0037	1.0034	0.7769	0.1041
		BIC	0.0030	0.8821	0.5921	0.0890
		MAPE	0.0042	0.8045	0.4600	0.0762
		WM4	0.0063	0.8592	0.5124	0.0832
		WM4R	0.0119	0.8673	0.6044	0.0856
	n=250	MIMSA	-0.0017	1.0751	0.1451	0.0686
		BIC	0.0028	0.9493	0.1343	0.0601
		MAPE	-0.0002	0.9569	0.1366	0.0594
		WM4	0.0040	0.9735	0.1267	0.0610
		WM4R	0.0023	0.9723	0.1374	0.0608
	n=500	MIMSA	-0.0019	1.1061	0.2769	0.0524
		BIC	0.0010	0.9786	0.0797	0.0462
		MAPE	0.0020	1.0096	0.1984	0.0472
		WM4	0.0006	1.0134	0.2366	0.0474
		WM4R	-0.0017	1.0184	0.1011	0.0479

Table B.16: Frequency distribution of model identification for simulation of ARIMA(2,1,0)

Parameter	Sample Size	Criterion	Candidate Model ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (3,1,0)	ARIMA (1,2,0)	ARIMA (2,2,0)	ARIMA (3,2,0)	ARIMA (1,3,0)	ARIMA (2,3,0)	ARIMA (3,3,0)
$\Phi_1 = 0.01$, $\Phi_2 = 0.9$	n=100	MIMSA	387	0	0	466	0	0	147	0	0
		BIC	0	34	3	885	63	15	0	0	0
		MAPE	10	38	181	7	20	531	12	31	170
		WM4	30	125	125	103	143	143	76	123	132
		WM4R	49	130	151	136	149	137	101	135	12
	n=250	MIMSA	418	0	0	532	0	0	50	0	0
		BIC	0	153	6	803	31	7	0	0	0
		MAPE	9	38	225	16	38	499	20	36	119
		WM4	29	123	144	137	144	143	72	105	103
		WM4R	82	143	126	125	125	170	103	117	9
	n=500	MIMSA	444	0	0	505	0	0	51	0	0
		BIC	0	600	19	358	19	4	0	0	0
		MAPE	12	54	263	19	27	474	12	36	103
		WM4	26	139	122	140	136	149	68	114	106
		WM4R	77	136	128	144	139	128	114	120	14
$\Phi_1 = 0.4$, $\Phi_2 = 0.5$	n=100	MIMSA	280	0	0	616	0	0	104	0	0
		BIC	4	75	3	819	75	24	0	0	0
		MAPE	34	30	198	16	28	474	11	26	183
		WM4	90	129	113	122	116	124	76	123	107
		WM4R	122	117	124	134	136	152	70	117	28
	n=250	MIMSA	292	0	0	682	0	0	26	0	0
		BIC	0	326	18	8603	45	8	0	0	0
		MAPE	52	45	220	20	30	454	25	30	124
		WM4	81	116	130	114	146	116	77	93	127
		WM4R	96	140	147	126	140	123	96	119	13
	n=500	MIMSA	296	0	0	698	0	0	6	0	0
		BIC	0	897	18	72	10	3	0	0	0
		MAPE	56	47	250	23	41	427	26	26	104
		WM4	118	114	109	106	146	127	69	95	116
		WM4R	112	118	143	123	132	137	103	124	8

Table B.17: Forecast measures for the simulation of ARIMA(2,1,0) process

Parameter	Sample Size	Criterion	Forecast Measure ME	RMSE	MAPE	MASE
$\Phi_1 = 0.01$, $\Phi_2 = 0.9$	n=100	MIMSA	-0.0138	1.2463	11.9589	0.5598
		BIC	0.0017	0.8927	6.4238	0.4134
		MAPE	0.0003	0.8302	4.1938	0.3630
		WM4	-0.0088	0.9161	8.8976	0.4083
		WM4R	0.0027	0.9391	6.5450	0.4208
	n=250	MIMSA	0.0059	1.3827	5.7916	0.6172
		BIC	-0.0024	0.9601	4.3858	0.4347
		MAPE	-0.0037	0.9657	4.4521	0.4290
		WM4	0.0019	1.0398	4.5393	0.4628
		WM4R	0.0062	1.0852	4.9019	0.4867
	n=500	MIMSA	-0.0203	1.4857	7.9290	0.6602
		BIC	-0.0033	0.9769	5.0653	0.4396
		MAPE	-0.0034	1.0086	4.9481	0.4504
		WM4	-0.0041	1.0652	5.0029	0.4781
		WM4R	-0.0051	1.1267	5.5945	0.5056
$\Phi_1 = 0.4$, $\Phi_2 = 0.5$	n=100	MIMSA	0.0019	0.9545	12.0497	0.5455
		BIC	0.0075	0.8875	11.9374	0.5062
		MAPE	0.0155	0.8230	10.8784	0.4460
		WM4	0.0030	0.9010	11.1616	0.4999
		WM4R	0.0047	0.8974	9.9273	0.4994
	n=250	MIMSA	0.0024	1.0228	7.4495	0.5662
		BIC	0.0016	0.9654	7.3581	0.5340
		MAPE	-0.0008	0.9743	6.1770	0.5318
		WM4	0.0004	1.0194	7.4641	0.5579
		WM4R	0.0033	1.0229	21.5562	0.5618
	n=500	MIMSA	-0.0044	1.0312	6.3146	0.5550
		BIC	-0.0020	0.9731	6.4458	0.5224
		MAPE	-0.0003	1.0082	6.5995	0.5381
		WM4	-0.0005	1.0561	6.7819	0.5656
		WM4R	0.0012	1.0551	6.4785	0.5649

Table B.18: Frequency distribution of model identification for simulation of ARIMA(0,1,1)

Parameter	Sample Size	Criterion	Candidate Model ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (0,1,3)	ARIMA (0,2,1)	ARIMA (0,2,2)	ARIMA (0,2,3)	ARIMA (0,3,1)	ARIMA (0,3,2)	ARIMA (0,3,3)
$\Theta = 0.9$	n=100	MIMSA	41	0	0	378	0	0	581	0	0
		BIC	807	28	12	0	131	22	0	0	0
		MAPE	98	78	137	73	165	238	26	46	139
		WM4	116	114	107	110	128	135	87	92	111
		WM4R	134	160	110	114	146	107	95	102	32
	n=250	MIMSA	40	0	0	432	0	0	528	0	0
		BIC	916	21	6	0	56	1	0	0	0
		MAPE	100	84	181	45	175	219	24	46	126
		WM4	147	126	121	93	122	127	96	77	91
		WM4R	122	139	145	113	130	118	110	113	10
	n=500	MIMSA	32	0	0	481	0	0	487	0	0
		BIC	957	19	2	0	22	0	0	0	0
		MAPE	128	94	183	40	169	202	24	33	127
		WM4	125	137	128	81	114	117	96	92	110
		WM4R	149	143	132	104	121	122	104	112	13
$\Theta = 0.5$	n=100	MIMSA	16	0	0	377	0	0	607	0	0
		BIC	879	41	12	5	55	8	0	0	0
		MAPE	115	97	216	89	140	171	44	31	97
		WM4	136	107	111	110	108	121	93	117	97
		WM4R	132	139	125	112	122	144	85	111	27
	n=250	MIMSA	28	0	0	331	0	0	641	0	0
		BIC	957	14	1	0	1	27	0	0	0
		MAPE	147	111	192	98	120	153	42	47	90
		WM4	134	129	135	99	107	118	85	96	97
		WM4R	137	100	142	136	124	108	128	114	11
	n=500	MIMSA	25	0	0	370	0	0	605	0	0
		BIC	969	13	2	0	16	0	0	0	0
		MAPE	135	118	220	92	118	145	53	42	77
		WM4	137	111	95	103	134	126	94	102	98
		WM4R	126	128	126	140	122	116	105	16	8

Table B.19: Forecast measures for the simulation of ARIMA(0,1,1) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Theta = 0.9$	n=100	MIMSA	0.0059	1.2753	143.6477	0.9142
		BIC	-0.0031	0.9852	72.9470	0.7287
		MAPE	-0.0036	1.0163	69.8463	0.7370
		WM4	0.0061	1.0590	122.1878	0.7680
		WM4R	-0.0048	1.0508	82.0429	0.7652
	n=250	MIMSA	-0.0060	1.3194	31.7053	0.9662
		BIC	0.0066	0.9996	33.5235	0.7410
		MAPE	0.0041	1.0505	32.3139	0.7724
		WM4	0.0026	1.1006	34.9875	0.8083
		WM4R	0.0026	1.105607	41.1176	0.8149
	n=500	MIMSA	0.0007	1.3303	26.4581	0.9823
		BIC	0.0072	1.0056	20.4936	0.7472
		MAPE	0.0045	1.0576	21.1005	0.7817
		WM4	0.0051	1.1164	21.6755	0.8247
		WM4R	-0.0024	1.1183	24.0933	0.8270
$\Theta = 0.5$	n=100	MIMSA	0.0035	1.1675	51.6193	0.9983
		BIC	-0.0034	0.9410	40.6655	0.8341
		MAPE	-0.0011	0.9634	46.2207	0.8389
		WM4	-0.0005	1.0129	43.6138	0.8773
		WM4R	0.0119	0.8673	0.6044	0.0856
	n=250	MIMSA	-0.0128	1.2122	35.1316	1.0659
		BIC	-0.0005	0.9804	29.2350	0.8748
		MAPE	0.0013	1.0318	30.9190	0.9139
		WM4	0.0050	1.0619	29.0463	0.9386
		WM4R	-0.0023	1.0672	30.9161	0.9430
	n=500	MIMSA	0.0015	1.2036	24.0989	1.0685
		BIC	0.0068	0.9890	20.3200	0.8853
		MAPE	0.0030	1.0402	20.8294	0.9258
		WM4	0.0050	1.0722	22.1405	0.9537
		WM4R	-0.0005	1.0684	22.0215	0.9507

Table B.20: Frequency distribution of model identification for simulation of ARIMA(0,2,1)

Parameter	Sample Size	Criterion	Candidate Model ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (0,1,3)	ARIMA (0,2,1)	ARIMA (0,2,2)	ARIMA (0,2,3)	ARIMA (0,3,1)	ARIMA (0,3,2)	ARIMA (0,3,3)
$\Theta = 0.9$	n=100	MIMSA	145	0	0	587	0	0	268	0	0
		BIC	0	0	0	797	42	8	0	140	13
		MAPE	40	26	26	20	12	25	144	328	379
		WM4	49	59	65	126	123	88	181	173	136
		WM4R	66	68	80	168	133	31	159	188	107
	n=250	MIMSA	162	0	0	476	0	0	362	0	0
		BIC	0	0	0	930	26	1	0	38	5
		MAPE	19	9	27	29	8	19	158	361	370
		WM4	30	38	45	172	148	92	155	181	139
		WM4R	75	69	66	158	159	39	167	167	10
	n=500	MIMSA	139	0	0	444	0	0	417	0	0
		BIC	0	0	0	972	12	3	0	12	1
		MAPE	18	15	13	32	20	29	149	353	371
		WM4	30	36	37	154	154	138	139	178	134
		WM4R	61	71	65	159	155	44	189	163	93
$\Theta = 0.5$	n=100	MIMSA	144	0	0	516	0	0	340	0	0
		BIC	0	0	2	886	28	10	4	62	8
		MAPE	33	30	46	39	33	39	197	269	314
		WM4	53	48	58	138	128	92	188	171	124
		WM4R	81	83	92	140	149	52	161	163	79
	n=250	MIMSA	178	0	0	464	0	0	358	0	0
		BIC	0	0	0	948	21	3	0	25	3
		MAPE	21	20	38	43	39	57	186	292	304
		WM4	34	47	44	168	155	99	156	173	124
		WM4R	76	75	83	141	159	61	181	158	66
	n=500	MIMSA	159	0	0	454	0	0	387	0	0
		BIC	0	0	0	974	15	0	0	11	0
		MAPE	24	16	31	56	34	65	199	277	298
		WM4	43	45	44	171	133	143	141	148	132
		WM4R	63	74	60	157	136	54	176	197	83

Table B.21: Forecast measures for the simulation of ARIMA(0,2,1) process

Parameter	Sample Size	Criterion	Forecast Measure			
			ME	RMSE	MAPE	MASE
$\Theta = 0.9$	n=100	MIMSA	0.0614	1.4769	2.1534	0.1512
		BIC	-0.0055	0.9546	0.7098	0.0939
		MAPE	0.0337	1.4660	0.8328	0.1214
		WM4	-0.0483	1.8019	0.9736	0.1429
		WM4R	0.0156	2.0643	1.0282	0.1576
	n=250	MIMSA	-0.0331	2.0069	0.7991	0.1356
		BIC	-0.0074	0.9925	0.1786	0.0643
		MAPE	0.0655	1.4534	0.2156	0.0822
		WM4	0.0080	1.7958	0.2777	0.0982
		WM4R	0.1717	2.6184	0.3527	0.1298
	n=500	MIMSA	0.1509	2.1686	0.2737	0.1105
		BIC	0.0005	0.9933	0.07684	0.0460
		MAPE	0.0331	1.4984	0.0889	0.0626
		WM4	0.1704	2.0258	0.1453	0.0779
		WM4R	0.0618	3.1483	0.1978	0.1090
$\Theta = 0.5$	n=100	MIMSA	-0.0628	1.3247	1.6062	0.1698
		BIC	-0.0052	0.9095	0.6303	0.1159
		MAPE	-0.0077	1.3592	1.2407	0.1445
		WM4	0.0630	1.5306	0.8888	0.1587
		WM4R	-0.0005	1.8455	1.0308	0.1859
	n=250	MIMSA	-0.0791	1.7759	0.8053	0.1518
		BIC	-0.0025	0.9761	0.2704	0.0795
		MAPE	0.0468	1.4301	0.3240	0.1028
		WM4	-0.0328	1.6917	0.3611	0.1149
		WM4R	-0.0337	2.3440	0.4594	0.1478
	n=500	MIMSA	-0.1258	1.9023	0.5665	0.1279
		BIC	0.0029	0.9792	0.2420	0.0595
		MAPE	-0.0641	1.6294	0.2614	0.0820
		WM4	-0.1472	2.0910	0.2835	0.0100
		WM4R	0.0605	2.6073	0.3110	0.1198

Table B.22: Frequency distribution of model identification for simulation of ARIMA(0,1,2)

Parameter	Sample Size	Criterion	Candidate Model ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (0,1,3)	ARIMA (0,2,1)	ARIMA (0,2,2)	ARIMA (0,2,3)	ARIMA (0,3,1)	ARIMA (0,3,2)	ARIMA (0,3,3)
$\Theta_1 = 0.01$, $\Theta_2 = 0.9$	n=100	MIMSA	34	0	0	354	1	0	611	0	0
		BIC	0	821	35	0	0	144	0	0	0
		MAPE	69	154	222	28	62	356	8	22	79
		WM4	113	142	139	108	97	148	84	89	80
		WM4R	141	135	158	108	100	137	87	114	20
	n=250	MIMSA	20	0	0	402	0	0	577	1	0
		BIC	0	927	22	0	0	51	0	0	0
		MAPE	65	194	245	30	44	336	11	26	49
		WM4	106	158	134	95	111	132	75	92	97
		WM4R	127	161	124	119	118	149	77	115	10
	n=500	MIMSA	22	0	0	423	2	0	553	0	0
		BIC	0	960	11	0	0	29	0	0	0
		MAPE	66	204	250	15	39	340	10	17	59
		WM4	87	140	136	94	118	167	67	100	91
		WM4R	134	164	164	109	105	127	89	104	4
$\Theta_1 = 0.4$, $\Theta_2 = 0.5$	n=100	MIMSA	27	0	0	468	0	0	505	0	0
		BIC	11	842	41	28	2	76	0	0	0
		MAPE	95	110	204	67	42	315	36	38	93
		WM4	116	110	119	111	113	127	91	115	98
		WM4R	126	134	135	114	140	107	119	109	16
	n=250	MIMSA	21	0	0	494	0	0	485	0	0
		BIC	0	939	23	0	0	38	0	0	0
		MAPE	99	140	199	43	56	346	37	27	53
		WM4	103	133	113	123	101	141	92	91	103
		WM4R	138	136	127	97	123	134	105	125	15
	n=500	MIMSA	28	0	0	523	0	0	449	0	0
		BIC	0	973	12	0	0	15	0	0	0
		MAPE	96	154	228	40	71	299	35	32	45
		WM4	110	122	118	129	118	119	91	95	98
		WM4R	104	144	150	118	118	116	119	116	15

Table B.23: Forecast measures for the simulation of ARIMA(0,1,2) process

Parameter	Sample Size	Criterion	Forecast Measure ME	RMSE	MAPE	MASE
$\Theta_1 = 0.01$, $\Theta_2 = 0.9$	n=100	MIMSA	0.0041	1.6861	65.4978	1.1786
		BIC	-0.0088	1.0031	43.1572	0.7315
		MAPE	-0.0092	1.0843	39.6289	0.7724
		WM4	0.0015	1.2444	45.1377	0.8790
		WM4R	-0.0056	1.2328	50.2057	0.8750
	n=250	MIMSA	-0.0098	1.6996	58.2084	1.2399
		BIC	0.0068	1.0112	30.8619	0.7496
		MAPE	0.0035	1.1016	32.9045	0.8109
		WM4	0.0042	1.2619	38.9547	0.9219
		WM4R	0.0034	1.2593	38.6928	0.9242
	n=500	MIMSA	-0.0101	1.6930	38.2177	1.2379
		BIC	0.0052	1.0154	21.3464	0.7489
		MAPE	0.0058	1.0995	21.3439	0.8071
		WM4	0.0045	1.2683	26.5176	0.9263
		WM4R	0.0021	1.2748	23.9177	0.9343
$\Theta_1 = 0.4$, $\Theta_2 = 0.5$	n=100	MIMSA	-0.0002	1.1738	39.6533	0.9592
		BIC	0.0007	0.9214	31.9415	0.7783
		MAPE	-0.0028	0.9720	33.6845	0.8071
		WM4	-0.0079	1.0321	33.2925	0.8495
		WM4R	-0.0033	1.0309	33.1807	0.8482
	n=250	MIMSA	-0.0042	1.2103	32.8402	1.0030
		BIC	0.0010	0.9732	30.1902	0.8179
		MAPE	-0.0008	1.0304	29.8615	0.8586
		WM4	-0.0016	1.0930	31.5650	0.9067
		WM4R	-0.0068	1.0885	27.6619	0.9054
	n=500	MIMSA	0.0009	1.2078	30.2881	1.0156
		BIC	0.0002	0.9837	27.1703	0.8325
		MAPE	-0.0011	1.0420	26.6146	0.8775
		WM4	0.0034	1.1033	26.5475	0.9290
		WM4R	-0.0012	1.0964	28.5364	0.9233