KWAME NKRUMAH UNIVERSITY

OF

SCIENCE AND TECHNOLOGY, KUMASI

DILEMMAS IN MODEL SELECTION IN TIME SERIES ANALYSIS

By

Doris Arthur

B.Ed Mathematics

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF M.PHIL MATHEMATICAL STATISTICS

June, 2018

Declaration

I hereby declare that this submission is my own work towards the award of the M.Phil degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university, except where due acknowledgement had been made in the text.

Doris Arthur	....	....
Student	Signature	Date

Certified by:

Dr. Sampson Twumasi-Ankrah	....	....
Supervisor	Signature	Date

Certified by:

Rev. Dr. William Obeng – Denteh	....	....
Head of Department	Signature	Date

Dedication

I gratefully dedicate this thesis to God Almighty who by His grace and mercies granted me this strength and knowledge. I again dedicate it to my mum and all my siblings

Abstract

This study seeks to resolve two important dilemmas in model selection in time series analysis. These are to compare the performance of the graphical and the information criterion methods in selecting the true model. In addition, Yu et al. (2005) relative precision performance stability was modified. For the graphical and information criterion comparison, dataset from ARIMA models were simulated. Also the cocoa production and rainfall datasets in Ghana were used to validate the modified relative precision performance stability of Yu et al. (2005). It was observed from the study that, in comparison to the performance of the graphical method and the Akaike information criterion (AIC) in selecting the ARIMA models, the information criterion performs better than the graphical method. Also, in verifying for the size of the evaluation sets in forecasting, whether to select a single model or combine the models of different models, our findings showed that the size of the evaluation sets may not influence the decision of selecting or combining since 97% of the decisions were to combine the models. In addition to that, though there was a modification on the computations of the relative prediction performance stability formerly utilized by [Yu et al., 2005], the decision rule still remains the same. Hence, whether the use of the mean or median on different the size of evaluation sets and interval, the combining strategy still outperforms.

Acknowledgements

My greatest thanks and appreciation goes to the Almighty God for his grace and mercy that has kept me through out my life.

I would like to express my gratitude to my affable supervisor Dr. Sampson Twumasi-Ankrah for the magnificent support and encouragement given me throughout the success of this work. Also, I would like to thank Professor A.O Adebanji for her advice given me throughout my entire programme and to all lecturers who taught me. My sincerest gratitude also goes to my mother and the entire family members who supported me both physically and spiritually. This work would not have come this far without the support given to me by my Headmaster Dr. Johnson Yeboah Gyekye for his immense contributions and encouragement and to my able friend Wilhemina Adoma Pels for given me a great support.

Contents

List of Abbreviation

SQL .... Structured Query Language

WHO ....World Health Organisation

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1 Background

The increase of advanced computational science and technology has offered several statistical models (both linear and nonlinear) which have been broadly utilized for time series modeling and forecasting. For instance, the autoregressive integrated moving average (ARIMA) which is a typical linear model, has been very effective in numerous time series forecasting. Moreover , the artificial neural system (ANN) which is a nonlinear model has likewise appeared to be an extremely encouraging methodology for time series modeling and forecasting. This because of the fact that true models are unknown, the selected model based on a selection criterion (e.g., Akaike information criterion (AIC) , hypothesis testing and graphical inspections, is often mis-specified. This may cause an unexpectedly high variability in the final prediction and affect forecasting accuracy and reliability.

In time series forecasting applications, the standard practice is to choose one suitable model from the various competing models regarding some selection strategies; before other inference on the selected model can be made. According [Burnham and Anderson, 2002], there are three techniques for model selection, and these techniques are graphical method with examination of, autocorrelations functions (ACF) and Partial autocorrelations functions (PACF);the hypothesis testing, a formal method for model selection in statistics,and information criteria method. Notwithstanding, these techniques show a few limitations. For instance, the graphical strategy is excessively subjective and too rough by and large for model selection. Moreover, there are difficulties with the second approach because of the testing issue of various testing (e.g., there is no target rule for the decision of the measure of every individual test, and it is totally hazy how such a decision affect the forecasting accuracy). Additionally, the initial two techniques does not measure the model sufficiency since pertinent data might be lost. [Breiman, 1996] called attention to that the estimators based on model selection are not stable. [Zou and Yang, 2001] additionally thought that as a noteworthy downside with model selection is its instability. For instance, with a little or moderate number of observations, of course, it is normally difficult to recognize models that are near each other (the model choice basis esteems are for the most part very close). For this situation, the decision of the model with the smallest criterion value is unstable. That is, a slight change in the information may prompt a different model being picked. Accordingly, the forecast based on the selected model are variable [Zou and Yang, 2001]. Moreover [Zou and Yang, 2001] called attention to distinguishing the true model which is not really ideal for forecasting. To reduce variability in model selection, researchers have turned to combining or mixing the same or different competing models for prediction purpose. So many literature has been in the past decades for model selection or combining competing models for forecasting. In all these, the decision is, combined models seem to outperformed a single selected model as the best or true model whether the models were combined from same or different models [Yu et al., 2005].

1.2 Problem Statement

A variety of modeling approaches for example regression models, neutral networks and others has helped researchers to solve real life problems. This enabled researchers to build models in a wide range of fields. As a result,statistical modeling has made tremendous contributions to many fields.

In model building, it is well noted that there is a information in an observed data and we need to express this information in a compact form called a model [Burnham and Anderson, 2002]. Each data originates from a model and in reality, all models are not right, however, some are useful

[Burnham and Anderson, 2002]. Since the true model that generates the data is not known,the standard practice has been to fit competing models and use a criterion to select an appropriate one. Thus, model selection is best observed as approximating instead of recognizing the "true model" . There are three ways to deal with model selection in time series literature. The general practice is to utilize any of these methods to choose a proper model from series of competing models for final estimation and inference to be made.

According to [Akaike, 1974] the first is the graphical method together with simple summary statistics, for example, autocorrelations function (ACF) and Partial autocorrelations function (PACF) which is helpful for preliminary model analysis. The second is hypothesis testing, a formal method for model selection in statistics and the third is a well-defined formal model selection criteria, for example, AIC or Bayesian Information Criteria (BIC). These techniques show a few defects. The first approach is too subjective and too rough, making it impossible to be used for model selection. Model selection based on hypothesis testing is frequently poor since most researches perceive that experiments are not conducted simply to reject null hypothesis or claim statistical significance since they need further understanding [Burnham and Anderson, 2002]. Likewise there is no statistical theory that supports the idea that hypothesis testing with fixed α-level is the premise of model selection.

Moreover, [Breiman, 1996] called attention to estimators in view of model selection criteria are unstable. For instance, with a little or moderate number of observations, of course it is generally difficult to separate models that are near each other with a model choice criterion values. For this situation a decision with the smallest criterion value is not steady. That is, a slight change in the data may bring about a different model being chosen. The graphical and the information criteria approaches are the commonly used methods in identifying the best model for a data. As indicated, these methods have limitations, therefore it is appropriate to compare their performance in selecting the true model for a particular dataset.

[Zou and Yang, 2001] pointed out that, identifying the true model is not really ideal for forecasting. To reduce fluctuation in model selection, researchers have turned to joining different candidate models for prediction purposes with various approaches which have been used in literature. The literature reports that joining models outperforms better in forecasting than using a single model [Yu et al., 2005]. At the point when a single model is chosen among candidate models for forecasting, a model that has not been chosen may predict better than "the best model that has been chosen" . In this regard the researcher becomes confused in whether to choose a single best model or combine models and is faced with two situations. In view of that, [Yu et al., 2005]proposed a decision rule (called relative precision performance stability) in order for researchers to overcome the troubles of making a decision to whether select or combine models. [Yu et al., 2005] further suggested that their relative prediction performance stability may be sensitive to parameters such as size and interval of evaluation sets. This research extends [Yu et al., 2005] decision rule by varying the size and interval of the evaluation sets to check if influence on the decision to select a single model or combine the models. Therefore, the study compares two different approaches (Graphical Method and Information Criterion) in choosing the best ARIMA models. The study also modifies the performance stability used in [Yu et al., 2005].

1.3 Objectives of the Study

The general aim of the study is to resolve dilemmas in model selection in univariate time series Analysis.

1.3.1 Specific Objectives

The specific objectives of the study are;

- To compare the performance of the graphical and the information criteria approaches in correctly selecting or specifying the true models.
- To analyze the effect of varying evaluation sets and the intervals on the decision of selecting or combining proposed by [Yu et al., 2005].
- To modify the computations of the relative prediction performance stability proposed by [Yu et al., 2005].

1.4 Methodology

The study will look at generating time series data by stimulating ARIMA models by using the R Software package for different sample sizes (with n=15, 25, 35, 50, 100 and 1000) of different parameters. Each sample size will be replicated five times and the data generated will be used to determine the order for autoregressive models (AR), moving average (MA) models ,autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models by using the graphical method and their respective ACFs and PACFs will be recorded for the preliminary data analysis for the selection of the true model. Further analysis for the estimation of the model selection is done by using the Akaike information criteria to select the true model from each sample of the simulated data to compare the performance of the two methods. In order to compare the prediction performance, it is necessary to introduce the forecasting evaluation criteria on the various sizes of evaluation sets to test for the performance stability of the model by using cocoa production data from 1948-2015 and that of rainfall from 1900- 2016. The study will employ two measures of accuracy, that is root mean square error (RMSE) and mean absolute percentage error (MAPE) to test for the performance stability of the model either to select a single model or combine the models.

1.5 Motivation of the Study

In order for a mathematical model in time series to be used to predict future unseen data successfully, researchers need to test the said true model on the three ways of selecting models for the final decisions to be taken in order not to make wrong statistical inferences. Hypothesis testing, graphical method and information criteria (AIC) are the three methods that are normally used. What motivated the choice of the graphical method is that the ACF and the PACF define distinct pattern for the autocorrelations and the partial autocorrelations of the models that are been selected. The information criteria have been used predominantly for model selection since it minimizes errors. These two methods are combined to compare performance since there is not a perfect method for the model that has been selected. The model that has been selected by a single method might even perform worse than the one that has not been selected and in certain instances both procedures might be working well so the need to consider combining methods. Also the evaluation of the performance stability of the selected model is done without considering the size of the evaluation sets for the verification of the performance stability at which forecast is done. However, the selected model or combined models is evaluated since stability is a key tool when it comes to predictability? The study will add to knowledge of time series analysts to consider this aspect critically in order to get a better prediction of a model selected to be the best by the use of two performance measures of accuracy for forecasting, the root mean square error (RMSE) and mean absolute percentage error (MAPE). These will be used to test the performance indicating values with the evaluation sets as to whether to select a single model or combine the models for prediction. The next issue of concern is how to select an appropriate model or combine models that can produce a better forecast.

1.6 Organization of Work

The whole thesis is structured into five main chapters as follows; Chapter one comprises the introduction, background of the study, statement of the problem, aims or objectives ,methodology, justification or significance of the study ,limitations and delimitations and organization of the study. Chapter two will deal with the relevant literature in the area of model selection (publications by other researchers), in the field of time series, the theoretical reviews of autoregressive integrated moving averages(ARIMA), Exponential smoothing (ES) Artificial Neural Network(ANN) model and measures of performance accuracy in forecasting. Chapter three will look at the detail description of the methodology, chapter four will look at the results and discussions and chapter five discusses the findings, conclusion and recommendations of the study.

CHAPTER 2

Literature Review

2.1 Introduction

This chapter presents a review and discussion on the problem under study. The chapter seeks to examine some research work done in the area of model selection (publications by other researchers), in the field of time series, the theoretical reviews of autoregressive integrated moving averages(ARIMA) , Exponential smoothing (ES) Artificial Neural Network(ANN) models .

2.2 Theoretical Concepts on Time Series

Time series modeling is a research area which has attracted attentions of researchers over last few decades. The primary goal of time series modeling is to carefully collect and study the past observations of a time series to develop an appropriate model which describes the nature and the structure of the series. This model is then used to generate future values for the series, i.e. to make projections or prediction. Time series forecasting thus can be termed as the act of predicting the future by understanding the past [Raicharoen et al., 2003]. Due to the uncountable importance of time series forecasting in various practical fields such as business, economics, finance, science and engineering, etc. Zhang (2007,2008) and Tong (2003) , measures must be put in place to fit an appropriate model to the observed time series. It is apparent that an accurate forecast of a time series is dependent on fitting an adequate model. Researchers have in time past invested resources into the development of efficient models to make better the correctness of forecasts. Thus, different time series forecasting models have been introduced and modified in literature and measures of performance accuracy in forecasting.

2.2.1 Time Series Analysis

A Time Series is a sequential set of data points, measured typically over equal time intervals ([Hipel and McLeod, 1994] [Raicharoen et al., 2003]). The measurements recorded during an event in a time series are organized in an appropriate sequential order. A univariate time series is one that has only one variable measured against time, whereas a multivariate time series has more than one variable considered. A time series can be a continuous or a discrete series. In a continuous time series the measurements are taken for all time, whereas in a discrete time series, the observations are made at specific time intervals. Examples of continuous time series include temperature readings, flow of a river, concentration of a chemical process etc. Then again the populace of a particular city, production of a company, trade rates between two unique monetary forms will be examples of discrete time series.

2.2.2 Components Of Time Series

Time series in general has four main basic components, which can be separated from the observed data. These components are: Trend, Cyclical, Seasonal and Irregular or random components. The propensity of a time series to rise, fall or stagnate over a long time period is simply known as Trend. Thus, it can be said that trend is a long term path of a time series. For instance, series relating to population growth, number of houses in a city etc. show a positive trend, whereas a negative trend can be observed in series relating to mortality rates, epidemics, employments, etc. Seasonal variations in a time series are fluctuations within a year during the season. The vital components influencing regular deviations are: atmospheric and climatic conditions, customs, traditional habits, etc. For example sales of air conditioners and fans increase during hot weather conditions. Variations that are dependent on the season are essential factors for business practitioners and producers for planning into the future. The variations dependent on cycles in time series depicts the medium-term changes in the series caused by conditions which recur in cycles. The span of a cycle stretches over longer time frames, typically at least two years. The greater part of the economic and financial time series demonstrates some type of cycle-based variation.

Sporadic or random deviations in a time series are influenced by uncertain causes which are irregular and as well, do not recur in a specific pattern. These variations are caused by occurrences such as war, strike, earthquake, flood, revolution, fire outbreaks and so forth. No defined statistical method exists for estimating random fluctuations in a time series. Factoring the effects of the four components, two distinct types of models are often utilized for a time series and they are the Multiplicative and Additive models. Multiplicative model is based on the assumption that the four components of a time series may or may not be independent and that all four can affect each other; whereas in the additive model it is assumed that the four components are independent of each other.

$$Z_t = T_t + S_t + C_t + I_t \tag{2.1}$$

Equation 2.1 presents the additive models

$$Z_t = T_t \times S_t \times C_t \times I_t \tag{2.2}$$

Equation 2.2 presents the multiplicative models. A stationary time series is independent on time and a non-stationary time series is dependent on time.

2.2.3 Autoregressive Integrated Moving Average Models

Models for time series data can have many forms and represent different stochastic processes. There are two widely used linear time series models in literature. Autoregressive (AR) [Box and Jenkins, 1970] and Moving Average (MA) models. Combining these two, the Autoregressive Moving Average (ARMA) [Hipel and McLeod, 1994] and Autoregressive Integrated Moving Average (ARIMA) models have been proposed in literature. For seasonal time series forecasting, a variation of ARIMA the Seasonal Autoregressive Integrated Moving Average (SARIMA) [Hamzaçebi, 2008] model is used. ARIMA model and its different variations are based on the Box-Jenkins principle and so these are also broadly known as the Box-Jenkins models. In ARIMA models a non-stationary time series is made stationary by using finite differencing of the data points. The mathematical formulation of the ARIMA (p,d,q) model using lag polynomial is given as

$$\Phi(L)(1-L)^d y_t = \theta(L)\varepsilon_t$$

$$\Big(1-\sum_{i=1}^{p}\Phi_i L^i\Big)(1-L)^d y_t = \Big(1+\sum_{j=1}^{q}\theta_j L^j\Big)\varepsilon_t \tag{2.3}$$

where p is the order of the autoregressive, d which controls the differencing is integrated, and q is moving average parts of the model.

The Autoregressive Model (AR)

In an AR(p) model the predicted value of a variable is assumed to be a linear combination of p past observations and a random error together with a constant term. Normally when measuring the parameters of an AR process for the given time series, the Yule Walker are used.

Assume $\{\epsilon_t\}$ is a sequence of independent and identically distributed random variables of mean 0 and variance σ_t^2. Then a process $\{Y_t\}$ is an Autoregressive process of order p (abbreviated to an AR(p) process) if

$$\Phi(B)Y_t = (1-\phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p)Y_t = \epsilon_t \text{ or}$$

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t \tag{2.4}$$

where ϕ's are the estimated Autoregressive parameters, Y is the original series, p is the order of the AR model and the ϵ_t , white noise is the source of randomness which has the following characteristics:

- $\mathbb{E}[\epsilon_t] = 0,$
- $\mathbb{E}[\epsilon_t^2] = 0 \ \forall \ t \neq s$,
- $\mathbb{E}[\epsilon_t, \epsilon_s] = 0 \ \forall \ t \neq s$.

Moving Average (MA) Model

This model is a linear regression of the current value of the series against current and the previous (unobserved) white noise error. That is, the lagged error terms are in this process. The order for an MA model is q and is formulate as MA(q):

$$Y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q} \tag{2.5}$$

where the θ is the moving average parameters of the model and the ϵ_t's are the series of unknown random errors which are assumed to follow a normal probability distribution.

The Autoregressive Moving Average (ARMA) Model

An ARMA(p, q) model is a combination of AR(p) and MA(q) models and is suitable for univariate time series modeling. Some assumption of ARMA processes is

- $\mathbb{E}[\epsilon_t] = 0,$
- $\mathbb{V}[\epsilon_t] = \sigma^2$,
- $Cov[\epsilon_t\epsilon_{t-k}] = 0$ for all t

It can be formulated as

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q} \quad (2.6)$$

2.2.4 Stationarity Analysis

The point at which an AR(p) process is presented as $\varepsilon_t = \Phi(L) y_t$, the $\Phi(L) = 0$ is understood to be the characteristic equation for the procedure. It is demonstrated by [Box and Jenkins, 1970] that an essential and adequate dependency for the AR(p) procedure to be stationary is that every one of the underlying foundations of the characteristic equation must fall outside the unit circle. [Hipel and McLeod, 1994]] said another straightforward algorithm (by Schur and Pagano) for deciding stationarity of an AR procedure. For instance, for the AR(1) model $Y_t = \mu + \phi_1 Y_{t-1} + \varepsilon_t$ is stationary when $|\phi| < 1$.

A MA(q) process is constantly stationary, regardless of the values the MA parameters. The conditions with respect to stationarity and invertibility of AR and MA processes additionally hold for an ARMA process. An ARMA(p, q) process is stationary if every one of the root of the characteristic $\phi(L) = 0$ lie outside the unit circle. Additionally, if every one of the roots of the lag equation $\theta(L) = 0$ lie outside the unit circle, at that point the ARMA(p, q) process is invertible and can be expressed as an pure AR process.

2.2.5 Exponential Smoothing Technique

Typically, time series could be observed as an amalgamation of varying parts which include the trend (τ), cycle (c), seasonal (s), and irregular or error (ε) parts. All these could be added together in a diverse number of ways. A purely additive model can be expressed as

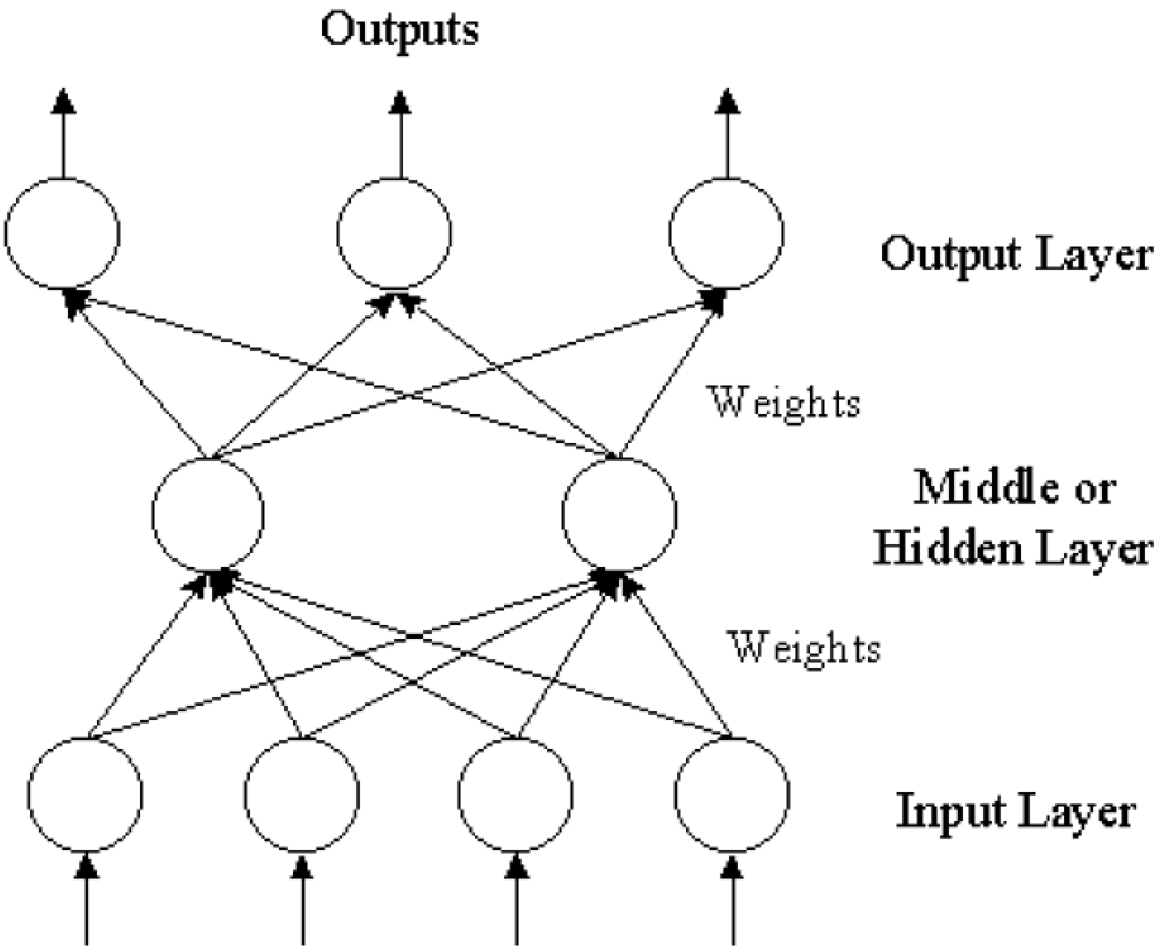
$$y = \tau + s + \varepsilon \quad (2.7)$$

where the three parts are combined to form the series that is seen. The trend component, an amalgamation of a level term (τ_l) and a growth term (τ_g), at all times, is the beginning component in exponential smoothing. If the error part is overlooked, there are precisely fifteen exponential smoothing methods as shown in the table below. To be clear, the simple exponential smoothing method is defined by cell (N, N), Holt's linear method by cell (A, N), the damped trend method by cell (Ad, N), Holt-Winters' additive method by cell (A, A), and Holt-Winters' multiplicative method is given by cell (A, M).

Trend Component	Seasonal Component		
	N (None)	A (Additive)	M (Multiplicative)
N (None)	N, N	N, A	N, M
A (Additive)	A, N	A, A	A, M
Ad (Additive Damped)	Ad, N	Ad, A	Ad, M
M (Multiplicative)	M, N	M, A	M, M
Md(Multiplicative Damped)	Md, N	Md, A	Md, M

2.2.6 Artificial Neural Network

The most broadly utilized ANNs in gauging issues are multi-layer perceptrons (MLPs), which utilize a solitary hidden layer feed forward network (FNN). The model is portrayed by a system of three layers, viz. input, hidden and output layer, connected by acyclic links. There might be in excess of one hidden layer. The nodes in various layers are generally called processing components. The three-layer feed forward architecture of ANN models diagrammatically portrayed as



2.3 Forecast Accuracy Measure

2.3.1 The Mean Absolute Percentage Error (MAPE)

The Mean Absolute Percentage Error (MAPE) is a forecasting accuracy measure which represents the percentage of average absolute error occurred. It is independent of the size of estimation, however influenced by a change of the information. MAPE does not punish extraordinary deviations.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100 \tag{2.8}$$

2.3.2 The Root Mean Squared Error (RMSE)

The Root Mean Squared Error (RMSE) is nothing but the square root of the mean square error. The mean square error a measure of the mean squared deviation of predicted values. This is formulated as

$$RMSE = \sqrt{MSE}$$

$$\sqrt{\frac{1}{n}\sum_{t=1}^{n} y_t - \hat{y}_t^2} \tag{2.9}$$

2.4 Akaike Information Criteria (AIC)

Akaike Information Criteria (AIC) proposed by [Akaike, 1974] uses log-likelihood and adds a penalizing term which is associated with the number variables. It uses the goodness of fit and then balances it by taking into account the inclusion of variables in the model. It is formulated as

$$AIC = -2log[L(\hat{\theta})] + 2K \tag{2.10}$$

where $-2log[L(\hat{\theta})]$ is the deviance term with $[L(\hat{\theta})]$ as the maximum log-likelihood of the parameter θ and K is the number of unknown parameters.

2.5 Empirical Studies on Model Selection

Statistical models for example autoregressive integrated models have been used for some decades for time series data analysis and has received much attention in prediction into future unseen data. Best models are selected by the use of information criteria (AIC, AIC_c, BIC), hypothesis testing and or a graphical method though they have their disadvantages. The selected model or models is or are used for predicting future values. According to earlier research done by researchers in time series analysis, the model preferred by a test or information criterion does not perform better than other competing models in terms of forecasting risk. In addition, one major drawback of model selection is its instability.

[Zou and Yang, 2004] argued that with a small or moderate number of observations, models closed to each other are usually hard to distinguish and the model selection criteria are usually quite close to each other. Thus a slight change in data may result in choice of different model. The unstable nature of the model selection criteria often may inflate variability in estimation or prediction.

[Chatfield, , Hoeting et al., 1999]; have used the term 'model uncertainty' to capture the difficulty in identifying the correct model. To address this problem, the combining forecast was introduced for some decades now ([Bates and Granger, 1969]; [Clemen, 1989]) and different methods have been proposed by time series researchers. Thus, when there is difficulty in selecting a single model, an alternative method, such as the combined method could be considered. Because of the way different time series have diverse properties and features, the chosen model based on a selection criterion (e.g., Akaike information criterion (AIC) [Akaike, 1974], information theory and an extension of maximum likelihood principle)), hypothesis testing and graphical procedures, is regularly mis-specified and thus may cause an unexpectedly high variations in the final prediction and influence forecasting accuracy and reliability. In practical application however, scientists and business professionals looked with two vital difficulties. One is whether to choose a proper modeling approach for prediction purposes or to join these diverse individual methods into a single forecast for same or different modeling approaches. The other is whether to choose the best competing model to forecast or to combine the various candidates with different parameters into a new forecast for the similar modeling approaches. At the end of the day, for time series determining the emphasis is on whether to choose a model from contending models or to consolidate these distinctive models into a single forecast. As a rule, the two issues are exceedingly non-trivial and have gotten significant consideration with various methodologies being examined .

In applications involving time series forecasting, the convention is to choose one fitting model from the different candidate models based on some selection strategies; thus final estimation, interpretation, and forecast are then based on the selected model. In general, there are three techniques for multiple candidate model selection. The first is graphical investigation, together with examination of basic summary statistics (such as autocorrelations (AC) and partial autocorrelations (PAC)), which is extremely useful for pre-modeling analysis. The second is hypothesis testing, a formal system for model selection in statistics. The last strategy is to utilize a well-defined and formal model selection criterion, such as AIC [Akaike, 1974] or Bayesian information criterion BIC [Schwarz et al., 1978]. [Breiman, 1996] pointed out that the estimators in view of model choice are instable.

[Yang, 2003] also considered that a significant setback with model choice is its inability to remain stable. For example, except when a large number of observations are utilized, it is typically difficult to differentiate models that are similar to each other (the model choice criterion values are similar). In this sense, the choice of the model with the smallest criterion value is unstable. That is, a slight change in the data may lead to a different model being selected. As a consequence, the forecasts based on the selected model are highly variable, [Zou and Yang, 2001].

[Zou and Yang, 2001] further additionally brought up that recognizing the true model is not really ideal for forecasting. To decrease variability in model selection, scientists have opted to consolidate the distinctive contending models for forecast purposes. A variety of literature on consolidated forecasting is accounted for. In the blend of a similar technique, [Draper, 1995, George and McCulloch, 1997] proposed an "averaging" strategy to combine a steady estimator. [Raftery, 1995] suggested using a BIC approximation for Bayesian model averaging. From their work, they identify that model uncertainty in linear regression on a single selected model made the model to underestimate the prediction and that leads to an incorrect inference and so their procedure looked at averaging over all possible combinations of predictors of a model. [Madigan et al., 1995] used a Markov chain Monte Carlo approximation to gain a steady forecast.

Additionally, [Breiman, 1996]) proposed a "bagging" strategy to produce numerous adaptations of the estimator and afterward average them into a steady estimator. In a comparative study, [Buckland et al., 1997] proposed a conceivable model weighting strategy as indicated by the estimation of a model selection rule (e.g., AIC). Cross-approval and bootstrapping have additionally been utilized to straightly consolidate distinctive models with the expectation of enhancing precision by finding the best linear combination ([Breiman, 1996, Wolpert, 1992].

Moreover, [Juditsky and Nemirovski, 2000] likewise proposed a stochastic estimation strategy to join k forecasts in the best linear mix. [Yang, 2003] adaptive regression method (ARM) by blending distinctive candidates from a similar model for regression. Likewise, [Zou and Yang, 2004] utilized the aggregated forecast through exponential re-weighting strategy (AFTER) to consolidate the forecasts from the distinctive individual autoregressive moving average(ARMA) models. Nonetheless, their works are just restricted to linear combinations of similar techniques with various parameters. What's more, the mix of various models, literature reporting their works indicates it to be very diverse.

Combining forecasts with different methods has been studied for over three decades, starting with the pioneering work of [Reid, 1968, Reid, 1969] and [Bates and Granger, 1969]. Various methods have been involved. With the rising improvements and application of new computational sciences, the growing interest in time series analysis is becoming alarming for statisticians, economists and other practitioners for example businessmen for the development of different models (linear and non-linear) for prediction in future unseen data values.

[Ankrah et al., 2015] in their paper propose a standard approach of combining forecast by proposing weights which were based on ranking the performance of forecast accuracy measures of models. Their study was essential due to the problems that were associated with the Akaike weights, equal weights and forecast from the 'best' model selected by the minimum AIC_c value. Their study pointed out that mean squared forecast error (MSFE) from the combined forecast of the proposed weights (weighted ranking procedure) outperformed all other approaches that were investigated. The findings from their research also revealed that the three combined forecast approaches consistently outperformed the forecast from the best model selected by the minimum AIC_c. The researchers recommended the use of the weighted ranking procedure in combining models in time series analysis. One ultimate concern is to be able to forecast the future values of a series from a 'best' model. This is to say that before forecasting, one is faced with a challenge of choosing the 'best' model among a variety of competing models. The selection procedures of the 'best' model have several difficulties since there is no special guideline for evaluating to final outcome of making a decision [Zou and Yang, 2004].

[Terui and Van Dijk, 2002] used the combined forecasts from a linear and a nonlinear model for investigating for time series with possibly nonlinear characteristics. The forecast combined a constant coefficient regression method as well as a time varying method. From the researchers, the time varying method allows for a locally (non)linear modeling. Three methods were applied to their data and the outcome showed that the combined forecasts perform well, especially with time varying coefficient.

[Tseng et al., 2002] in their paper combining neural network show with seasonal ARIMA time series demonstrate likewise proposes a crossover forecasting model, which consolidates the seasonal time series ARIMA (SARIMA) and the neural system back propagation(BP) models, known as SARIMABP. The model was utilized to foresee two seasonal time series of aggregate production value for Taiwan machinery industry and the soda time series. The forecasting performance was looked at among four models,(i.e the SARIMABP and SARIMA) models and the two neural network models with differenced and deseasonalized data, respectively. Among these strategies, the mean square error(MSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE) of the SARIMABP model were the most minimal. The SARIMABP model was likewise ready to forecast certain critical defining moments of the test time series.

[Van Der Voort et al., 1996] also used a hybrid method of short-term traffic forecasting the (KARIMA) method. The technique uses a Kohonen self-organizing map as an initial classifier; each class has an individually tuned ARIMA model associated with it. Using a Kohonen map which is hexagonal in layout eases the problem of defining the classes. The explicit separation of the tasks of classification and functional approximation greatly improves forecasting performance compared to either a single ARIMA model or a back propagation neural network. The model is demonstrated by producing forecasts of traffic flow Performance is similar to other layered models.

[Adhikari and Agrawal, 2012] utilized simple linear ensemble strategies frequently the conceivable connections between at least two partaking models. In their paper ,they argue that the simple linear combination most at times do have problems so they decided to use non linear models since there is a considerable account for uncertainties. They suggested a vigorous weighted nonlinear ensemble method which considers the individual predictions from various models (ARIMA,ANN ,SVM) and additionally the relationships among them while combining. Comparison is made among the suggested plan and three other broadly utilized linear combination techniques, as far as the forecast errors are concerned. The comparison demonstrated that their technique gives essentially lower prediction errors than every individual model and in addition every one of the four linear combination strategies. Empirical discoveries from their paper exhibit that the proposed method outperforms every individual model and in addition three other mainstream linear combination strategies, as far as acquired forecast correctness and their method is very productive when the contributing forecasts are strongly associated.

Diverse literature proves that most of the research done in this area of time series analysis regularly noticed that an appropriate mix of different forecasts generously enhances the general precision and in addition outperforms every individual model [Armstrong, 2001] and [Zhang, 2003] additionally add their voice to the issue. Combination strategies are the best instinctive options to utilize, when there is a lot of uncertainty related with the determination of the one vital model for forecasting. However, bringing together or mixing multiple forecasts models minimizes the errors arising from faulty assumptions, bias, or mistakes in the data to a great extent.

[Yu et al., 2005] also propose combining different models (ARIMA, ES, SMA, ANN) for linear and non-linear models by using two judgmental criteria to solve the problems of whether to choose a single model or consolidate numerous contending models. They inferred that, in the individual forecasting model, the ANN model is the best as far as RMSE is concerned; the combined forecasting models by and large perform superior to the individual prediction models ; the forecasting exactness of nonlinear combination is superior to that of thelinear; combination of different techniques outperforms a combination of comparable strategies ; the PCA-based combination models perform better than the other mix models and other individual models; and the PANC model performs the best of all the linear (nonlinear) mix models and individual models. They finally concluded that the combine models of different methods perform better.

The study would be made successful by using the following three proposed time series models for forecasting. These models are Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing (ES),and Artificial

Neural Network (ANN).

CHAPTER 3

Methodology

3.1 Introduction

The study is in two folds: 1. evaluating the performance of graphical method and information criteria in selecting the order of univariate time series; and 2. extending the work of [Yu et al., 2005] on deciding whether a single or combining model is appropriate for forecasting by varying the number of evaluation sets and interval. Thus, in this chapter the theoretical background in achieving these two objectives or directions are discussed.

3.2 Graphical Method and Information Criteria Approach

In other to validate the performance of the graphical and the information criteria methods of selecting the true order of a time series model, a simulation study was conducted from ARIMA models. The rule of thumb in both approaches are discussed.

3.2.1 The rules of thumb for graphically selecting the true order of a Model

In the graphical method, a best fitted model was identified by the partial autocorrelation function (PACF) and the autocorrelation function (ACF).

- ACF and PACF are plotted individually to the generated dataset.

- If the PACF decays exponentially, the model is identified as an AR model and the number of significant spikes or lags of the ACF plot is used as the order of the AR model.

- If the ACF decays exponentially it gives an indication that the model is an MA model and the number of significant spikes or lags of the ACF is used as the order of the MA model.

- If the PACF and ACF do not decay exponentially it gives an indication that the model is an ARMA model and the number of significant spikes or lags of the PACF and ACF is used as the order of the AR and MA model respectively.

3.2.2 The rules of thumb for selecting the correct order of a model using Information Criteria

- For each generated data a certain number of competing models or order are fitted to the data.

- Since the true order of the model is known, the order or the model with the minimum information criterion is chosen.

Generally, if a specific method or approach selects the true order, it is then recorded and the frequency of the method selecting the correct model or true model is tallied. This is then record in percentage, thus the frequency (f) of a particular approach being able to pick the true model across the replicates (R) is recorded and divided by the chosen number of replicate times 100. This should be formulated as:

$$\frac{f}{R} \times 100$$

3.2.3 Simulation Study

The nine datasets were generated from AR(p), where p=1,2,3; MA(q) with q=1,2,3; and ARMA (p,q) p,q=1,2,3 with six replications and six different sample sizes (15,25,35,50,100.1000) for both approaches. Hence the datasets generated from each case are fitted to three competing models including the true model which generated the datasets. That is, for AR(p), AR(1), AR(2), and AR(3) were fitted, for MA(q), MA(1), MA(2), and MA(3) and for ARMA (p, q); ARMA(1,1), ARMA(2,2), and ARMA(3,3) were also fitted.

For graphical method, the ACF and PACF are plotted on each dataset generated and the rule of thumb is applied to select the appropriate order or model. Pertaining to the purpose of these sample sizes is to test the influence of sample size on the ability of the graphical or the information criterion to correctly select the true model, underestimate and overestimate the model as sample size increases. The underline models or the true models are known since the datasets were derived from the true models.

3.3 Decisions on single or combining methods

Here, several forecasting procedures and series of experiments are conducted in order to compute the prediction performance stability of time series methods. This is to verify whether to select a single model or combine the models. This was done by using two real time series data on yearly cocoa production from 1948-2015 with 68 observations and the second was monthly rainfall patterns from 1900-2016 with 116 observations. In this subsection of the study, three forecasting methods namely autoregressive integrated (ARIMA), Exponential smoothing (ES), Artificial Neural Networks (ANN) models are used.

3.3.1 Algorithm on Prediction Performance Stability

The following algorithms are provided in order to achieve a decision rule.

- Subset the data into a training and test set (we called the test set evaluation sets).

- Fit competing models to the training sets and select the best model.

- Forecast from the best model up to the length of the test set or evaluation set.

- Compute the forecast accuracy measure for each evaluation set.

$$RootMeanSquareError(RMSE) = \sqrt{\frac{1}{n}(Y_t - \hat{Y}_t)^2} \tag{3.1}$$

where Y_t $(t = 1, 2, 3, \cdots, n)$ represent the actual values and $\hat{Y}_t$ $(t = 1, 2, 3, \cdots, n)$ represent the predicted values and n is the number of observations in the evaluation sets.

$$MeanAbsolutePercentageError(MAPE) = \frac{1}{n}|\frac{e_t}{Y_t}| \times 100 \tag{3.2}$$

where Y_t $(t = 1, 2, 3, \cdots, n)$ represent the actual values and e_t $(t = 1, 2, 3, \cdots, n)$ represent the random error and n is the number of observations in the evaluation sets.

- The mean and standard deviation are calculated from the number of evaluation set sets.

- According to [Yu et al., 2005] the performance indicators is computed as

$$PI_i = \sum_{j=1}^{m} \frac{PI_{ij}}{m} \tag{3.3}$$

where m is the number of evaluation set used and PI_i is the prediction performance indicator which is the RMSE or the MAPE, j $= 1 \cdots, m$ is the number of evaluation sets, $i = 1, \cdots, k$ is the number of forecasting methods .

- Compute the relative prediction performance stability (S_i). This is formulated as

$$S_i = \frac{(\frac{1}{m}\sum_{j=1}^{m} PI_{ij})/\sigma_i}{Min\{PI_i\}/\sigma PI_i} \qquad i = 1, 2, 3, \cdots, k \tag{3.4}$$

  where σ is the standard deviation of the PI_{ij}.

- Make a decision by comparing S_i and S_o. Where S_o is the assumed threshold value which according to literature is 4.0.

$$\text{Decision} = \begin{cases} selecting, \text{ if } S_i \geq S_o & i = 1, 2, 3, \cdots, k \\ combining \text{ if } S_i < S_o \end{cases}$$

The yearly data of cocoa production is taken from 1948-2005 as training periods with 58 observations and the rest as testing periods with 10 observations. The rainfall data was taken from 1900-2006 as training with 106 observations and 10 as testing and are used to evaluate the size of the evaluation periods in the performance of the measures of accuracy on the root mean square error (RMSE) and mean absolute percentage error (MAPE). In order to compare the prediction performance, the necessary forecasting criteria are introduced and its computational forms are defined in equations 3.1 and 3.2.

3.4 Modified Relative Prediction Performance Stability

According to [Yu et al., 2005] the performance indicators are computed using standard deviation and mean of the forecast accuracy measures. The limitation of this method is that, the mean and standard deviations are sensitive to extreme values. Therefore, we propose the use of median and standard deviation (which uses median in place of mean in its computation) instead of mean and standard deviation respectively. Thus, PI_i which is the same as the PI_{mean} is replaced by

PI_{median} which is the median of the PI_{ij}

$$PI_{mean} = PI_{median}(PI_{ij}), \sigma = \sigma_{median}$$

Hence

$$S_i = \frac{PI_{median}/\sigma_i median}{min(PI_{median})/\sigma_i median} \tag{3.5}$$

where $i = 1, \cdots, n$

CHAPTER 4

Results and Discussions

4.1 Introduction

The Chapter looks at the two main aspects of the study. The results on the Graphical and the Information Criterion (AIC) model selection methods and the size of the performance of the evaluation sets on the dissimilar models.

4.2 Graphical Method and Information Criteria

This section presents the results from the simulation study on the selection of different models by the ACFs and PACFS across different sample sizes. These tabulated results show the percentage or probability of correctly selecting the true model by the Graphical Method. The procedures are based on the methodology explained in section 3.2.

4.2.1 Performance of individual methods in selecting the true model

Table 4.1 gives the percentage of selecting the true model using the Graphical method. In the autoregressive processes (AR(1) and AR(3), performance of the graphical method was poor as sample size increased. Nonetheless, regarding AR (2) the performance was better and consistent as sample size increased. Pertaining to the MA process, as the model order increased performance became poor. However, in MA (2) at large sample size performance was good. In the mixed processes the graphical approach was inconsistent and poor with regards to sample size and model order. In general, the graphical approach performed poorly.

Table 4.1: Summary of percentage of selecting the true model using Graphical method varying across sample sizes

	Sample Sizes					
Models	15	25	35	50	100	1000
AR (1)	60	40	20	0	0	0
AR (2)	40	60	80	100	100	100
AR (3)	0	0	0	0	0	0
MA (1)	40	60	80	40	40	0
MA (2)	0	0	0	60	60	100
MA (3)	0	0	0	0	0	0
ARMA (1,1)	40	0	0	0	0	0
ARMA (2,2)	20	20	20	40	20	0
ARMA (3,3)	0	0	0	0	0	0

Table 4.2 presents the percentage of selecting the true model using the AIC. The performance of the AIC was very poor in AR(1) since it could not select the true model as the sample size increased. However, its performance in AR(2) for all sample sizes was good and consistent. Also, as sample size was small, performance in selecting the true model in MA (1) was good but very poor in MA (2) and MA (3). The AIC performed poorly in ARMA(3,3). It also performed well in ARMA(2,2) as sample size increases but performed poorly in AR (1).

Table 4.2: Summary of percentage of selecting the true model using AIC varying across sample sizes

	Sample Sizes					
Models	15	25	35	50	100	1000
AR (1)	60	0	0	0	0	0
AR (2)	40	80	80	80	80	100
AR (3)	0	20	40	20	40	0
MA (1)	80	100	40	40	20	0
MA (2)	0	0	60	60	20	0
MA (3)	0	0	0	0	60	100
ARMA (1,1)	60	20	0	20	20	0
ARMA (2,2)	40	80	60	80	80	100
ARMA (3,3)	0	0	0	0	20	0

Generally, in comparing the two methods in 4.1 and 4.2 by the percentage of selecting the true model as order and sample size increases, the performance of AR(2) as a single model by Graphical and AIC methods is very good and that of the mixed models ARMA(2,2). However, the performance of other ones is unacceptable in both methods. The percentage of selecting the true models from both methods, AR(2) and ARMA(2,2) performs better than all other models.

4.2.2 The overall performance of graphical and information criteria to select the true order

A weighted ranking scale based on the performance categories were used in order to compare the overall performance of the methods. The overall performance rating summary of the Graphical and the AIC correctly selecting the true models is given in table 4.2. A five performance capability categories have been defined. That is, a decreasing weight has been assigned to a decreasing performance category. At each method, the number of times a method occurred (frequencies) with respect to the five categories are multiplied by its corresponding weight assigned. The results are summed across categories as the overall performance score of the method. The scores are then ranked in order of the highest score given rank 1 and the other 2. The score can be calculated as;

$$Score = f \times w_i \tag{4.1}$$

where f is the frequency of a method selecting the true model and w_i is the weight assigned to the performance of the model.

Hence, the performance rating was defined as “Very good” and assigned percentage is [90 - 100] and a weight "4" of the number of times a procedure or method selects the true or correct model. The next assigned percentage as [70-80] and is given “good” and weight “3” followed by [50-60] as “acceptable” and weight “2”, the fourth category is rated “poor” with an assigned interval [30-40] and weight "1" and the last but not least with an assigned percentage rating [0 -20] as unacceptable is given weight “0” as per the decreasing weight.

From Table 4.3, it can be deduced that the overall performance of the AIC method is ranked as the best performing methods/procedures for selecting the true models with the highest score of 58.

Table 4.3: Overall performance rating summary of graphical method and information criterion selecting the true models

Weight (w_i)	4	3	2	1	0		
Performance Capabilities category						Score	Rank
Methods	Very Good	Good	Acceptable	Poor	Unacceptable		
Graphical	4	2	5	7	36	39	2
AIC	4	8	6	6	30	58	1

4.2.3 Percentage of selecting under-fitted order when the true order is not selected

We look at the rate of the graphical and information criterion to select a model that has less parameters contrasted with the true model. Our aim here is to recognize which method has the likelihood to under-fit if the true model or order is not selected.

Table 4.4 gives the percentage of the Graphical method underestimating (under-fitting) the models as order and sample size increases. It is clear from the table that, AR (3) seems to be performing poorly (i.e severely under-fits) . This is followed by MA (3) then ARMA (3,3) and MA (2). In all AR (1) and ARMA (1,1) truly selected the best model as compared to the others. Also with the exception of sample sizes in the case of ARMA (2,2), its performance was also better and in the single methods too, AR (2) does quite well.

Table 4.4: Summary of percentage of underestimating the true model using the Graphical approach varying across sample sizes

	Sample Sizes					
Models	15	25	35	50	100	1000
AR (1)	0	0	0	0	0	0
AR (2)	40	40	20	0	0	0
AR (3)	100	100	100	100	100	100
MA (1)	0	0	0	0	100	0
MA (2)	40	60	80	40	40	0
MA (3)	40	60	80	100	100	100
ARMA (1,1)	0	0	0	0	0	0
ARMA (2,2)	80	20	0	0	0	0
ARMA (3,3)	100	60	80	80	40	0

Table 4.5 also gives the percentage performance of the AIC under-fitting the model. From the above table, it is obvious that, AR (1), MA(1), ARMA(1,1) performs better as sample size increases. ARMA (2,2) in a single case when sample size is 15 is acceptable but performs well from sample sizes 25, 35, 50,100, 1000. AR (3) and ARMA (3, 3) performs poorly as sample size increases with exception of sample size 35 where it is acceptable.

Table 4.5: Summary of percentage of underestimating the true model using AIC varying across sample sizes

	Sample Sizes					
Models	15	25	35	50	100	1000
AR (1)	0	0	0	0	0	0
AR (2)	60	20	0	0	0	0
AR (3)	100	80	60	80	80	100
MA (1)	0	0	0	0	0	0
MA (2)	100	100	40	40	20	0
MA (3)	100	100	100	100	40	0
ARMA (1,1)	0	0	0	0	0	0
ARMA (2,2)	60	20	20	20	0	0
ARMA (3,3)	100	100	60	100	80	100

4.2.4 Overall Performance of the Graphical and Information Criterion Underestimating the true order

We tend to look at the probability of the two approaches in choosing a model that has a lesser parameter in comparison to the true model. The focus here is to examine; the overall performance of these two methods that is Graphical and Akaike Information Criterion underestimating the true model with respect to the weight assigned to the performance capabilities categories.

The last performance rating is defined with a weight of 0 to the category or class [90% – 100%] which is "unacceptable". The next lowest rating is with a weight of 1 and the alloted rate of category is [70% - 80%] which is "poor". The third performance rating is also with a weight of 2 and the category on the alloted rate is [50% – 60%] and is "acceptable". The second rate is again defined with a weight of 3 with the doled out rate as [30% - 40%] and is termed "good". The first rate, which is "very good", is defined with a weight of 4 and is therefore assigned percentage category as [0% – 20%]. That is, a decreasing weight is alloted to an increasing performance class.

From table 4.6, it is obvious that the highest ranked method is AIC with a score of 335 and underestimating (under-fitting) the model, AIC is ahead of that of the Graphical method. This shows that the Graphical method seems to under-fit models more than that of the Akaike Information Criterion.

Table 4.6: Overall performance rating summary of graphical method and information criterion underestimating the true models

Weight (w_i)	0	1	2	3	4		
Performance Capabilities category						Score	Rank
Methods	Unacceptable	Poor	Acceptable	Good	Very Good		
Graphical	11	4	4	7	27	141	2
AIC	12	4	4	3	31	335	1

4.2.5 Overfitting

We also look at the rate of the graphical and information criterion being able to pick a model that has more parameters contrasted with the true model. Our goal is to recognize which approach has the likelihood to over-fit, if the true order is not selected.

Table 4.7 also gives a clear presentation of the percentage performance of the Graphical method overestimating (over-fitting) the true model. The performance of AR (2), AR (3), MA (2), MA (3) and ARMA (3, 3) is better compared to the rest as over-fitting is concerned. The performance of ARMA (2,2) is not consistent and that MA (1)..

Table 4.7: Summary of percentage of overestimating the true model using the Graphical approach varying across sample sizes

	Sample Sizes					
Models	15	25	35	50	100	1000
AR (1)	20	60	80	100	100	100
AR (2)	0	0	0	0	100	0
AR (3)	0	0	0	0	100	0
MA (1)	0	20	20	40	60	100
MA (2)	0	0	0	0	100	0
MA (3)	0	0	0	0	100	0
ARMA (1,1)	60	60	80	60	60	60
ARMA (2,2)	0	20	60	20	40	60
ARMA (3,3)	0	0	0	0	20	60

Table 4.8 also gives the percentage performance of AIC overestimating the models as order and sample size increases. It can be seen from the tables that AR (2), AR (3), MA (3), ARMA (3, 3) and ARMA (2, 2) performs better than that of the other models. AR (1) and ARMA (1, 1) performs poorly as sample size increases from 25,35,50,100,1000.

Table 4.8: Summary of percentage of overestimating the true model using the AIC varying across sample sizes

	Sample Sizes					
Models	15	25	35	50	100	1000
AR (1)	40	100	100	100	100	100
AR (2)	0	20	20	20	20	0
AR (3)	0	0	0	0	100	0
MA (1)	0	0	60	60	80	100
MA (2)	0	0	0	0	40	100
MA (3)	0	0	0	0	100	0
ARMA (1,1)	40	80	100	80	80	100
ARMA (2,2)	0	0	40	0	0	0
ARMA (3,3)	0	0	0	0	0	0

4.2.6 Overall Performance of the Graphical and Information Criterion overestimating the true order

For sake of clarity of interpretations, five performance capability categories have been defined.

The last performance rating is defined with a weight of 0 and the assigned rate on the category or class is [90% – 100%] which is "unacceptable". The next lowest rating is with a weight of 1 and the alloted rate of category is [70% - 80%] which is "poor". The third performance rating is also with a weight of 2 and the category on the alloted rate is [50% – 60%] and is "acceptable". The second rate is again defined with a weight of 3 with the doled out rate as [30% - 40%] and is termed "good". The first rate, which is "very good", is defined with a weight of 4 and is therefore assigned percentage category as [0% – 20%]. That is, a decreasing weight is alloted to a decreasing performance class.

From table 4.9, it is clear that Akaike Information Criterion performs better than that of the Graphical method since it has the lowest score of 162. It is concluded that the graphical method under -fits the model more that of the information criterion.

Table 4.9: Overall performance rating summary of graphical method and information criterion overestimating the true models

Weight (w_i)	0	1	2	3	4		
Performance Capabilities category						Score	Rank
Methods	Unacceptable	Poor	Acceptable	Good	Very Good		
Graphical	4	2	10	2	36	172	1
AIC	9	4	2	4	35	160	2

It can then be concluded that, the overall relative performance of the two methods, AIC is ranked as the highest performed method with respect to the probability of selecting the true model as compared to the other methods and it is also ranked best in all the three instances of selecting the true model, under-fitting and over-fitting. Therefore, in terms of correctly selecting the true model, the information criterion is highly recommended since the Graphical method seems to have highest probability of under-fits or over-fits models as compared to that of Information Criterion. Although the graphical method is not performing well in general, it did performed in selecting the AR(2) model and ARMA (2,2) so it must be considered in few cases when it is doing well. Therefore the graphical and the AIC methods performs better in the selection of individual models AR(2) and ARMA(2,2) so where they are performing in a particular instance, the two methods can be combined to select the true model rather than using a single method.

4.3 Decision on single or combine model

This aspect of the study is an extension of [Yu et al., 2005]. Thus we will start with the same number of evaluation sets and evaluation interval in [Yu et al., 2005] before varying the number and interval of the evaluation sets to observed whether decision on single or combining methods will change or not. Two different datasets are used to validate the hypothesis. Three forecast methods were fitted namely ARIMA, ETS, NNTAR. As indicated, for each method, the best model was selected based on having the minimum information criteria. The best model is used for forecasting the length of the test data and the forecast accuracy measure are computed.

4.3.1 Conventional Number of Evaluation set and interval based on [Yu et al., 2005]

Table 4.10 gives the accuracy measures on three evaluation sets. From table 4.10, the model with the least performance indicator of the root mean square error(RMSE) is ARIMA with a value of 278454.2 and that of MAPE is 30.8126. This means that if we are to select a single method, ARIMA is the best performing model and must be selected for prediction.

Table 4.10: The Forecast Accuracy Measure on three different models with three evaluation sets having a difference of 10 on Cocoa Production Data

ARIMA			ES			ANN		
Eval Sets	RMSE	MAPE	Eval Sets	RMSE	MAPE	Eval Sets	RMSE	MAPE
10	139355.5	12.7950	10	137698.14	12.6711	10	232791.15	23.3695
20	374484.33	42.9144	20	324484.33	36.0373	20	289854.36	32.4129
30	321522.66	36.7285	30	383595.04	49.8147	30	365318.84	45.1057
Mean	278454.1633	30.8126	Mean	281925.8367	32.8410	Mean	295988.1167	33.6294
σ	123339.2202	15.9073	σ	128353.9652	18.7770	σ	66476.4198	10.9191
S_i				0.9729	0.9029		0.1972	1.5900

A further analysis is done to confirm the decision by calculating the relative prediction performance stability (S_i values). From section 3.3.1 the use of the RMSE, its S_i values are calculated as

$$S_1 = \frac{281925.8/128353.9652}{278454.1633/123339.2202}$$

$$= 0.9729$$

$$S_2 = \frac{295988.1/66476.422}{278454.1633/123339.2202}$$

$$= 0.1972$$

Using MAPE, the S_i values are calculated as

$$S_1 = \frac{32.8410/18.7770}{30.8126/15.9073}$$

$$= 0.9029$$

$$S_2 = \frac{33.6294/10.9191}{30.8126/15.9073}$$

$$= 1.5900$$

Three evaluation sets (Eval set) of size (10, 20, 30) was looked at in order to check for the performance stability for the cocoa production data for the three models ARIMA, ANN and ES .The same calculations are done on the MAPE and the values shown from the table are 0.9327 and 1.5966 respectively. These values all are less than the threshold value ie $S_i < S_o$ so the three models are to be combined. Since this approach is based on [Yu et al., 2005], it can be concluded that the models should be combined.

We again varies [Yu et al., 2005] the interval of evaluation set (Eval set) on the same data in order to check if the decision will change or not.

From table 4.11, the interval on the evaluation sets was increased with 15 units (ie 30, 45, 60). In this experiment too, pertaining to the same forecast accuracy measure, the RMSE and the MAPE, and the best performing model is still ARIMA.

Table 4.11: The Forecast Accuracy Measure on three different Models with three evaluation sets having a difference of 15 on Cocoa Production Data

ARIMA			ES			ANN		
Eval Sets	RMSE	MAPE	Eval Sets	RMSE	MAPE	Eval Sets	RMSE	MAPE
30	321522.66	36.72846	30	383595.04	49.81469	30	365146.92	45.06945
45	227310.74	44.0375	45	227129.96	44.2054	45	223982.71	49.527674
60	272433.61	37.1416	60	272432.8	37.14143	60	271458.3	36.9858
Mean	273755.67	39.30252	Mean	294385.9333	43.7205	Mean	286862.6433	43.8610
σ	47119.8721	4.1058	σ	80509.5359	6.3505	σ	71831.7749	6.3577
S_i				0.6294	0.7192		0.0689	0.7207

Using the RMSE, the S_i values can be formulated as

$$S_1 = \frac{294385.9333/80509.5359}{273755.67/47119.8721}$$

$$= 0.6294$$

$$S_2 = \frac{286862.6433/71831.7749}{273755.67/47119.8721}$$

$$= 0.0689$$

Using MAPE, the S_i values are calculated as

$$S_1 = \frac{43.7205/6.3505}{39.30252/4.1058}$$

$$= 0.0689$$

$$S_2 = \frac{43.8610/6.3577}{39.30252/4.1058}$$

$$= 0.7207$$

All the values for the performance prediction stability were less than the threshold value on the two measures of accuracy that was used and the decision is that combine the models even with an increased in interval on the evaluation sets of same

This analogy was again tested on another set of data (ie Rainfall data) to still verify if the decision still holds with respect to the data at hand.

From table 4.12 it can be realized the ANN model with respect to the RMSE performs better than the ARIMA and the ES models since it had the minimum value of mean and standard deviation. However, pertaining to the MAPE forecast accuracy measure, the ES stands out as compared to the other two.

Table 4.12: The Forecast Accuracy Measure on three different Models with three evaluation sets having a difference of 10 on Rainfall Data

ARIMA			ES			ANN		
Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE
10	203.8851	16.5780	10	221.1879	16.45342	10	192.378	16.55198
20	197.1541	16.5612	20	210.3299	15.7177	20	194.7724	17.2458
30	187.9571	16.5195	30	189.164	15.8661	30	181.1799	16..21514
mean	196.3321	16.5529	Mean	206.8939	16.0124	Mean	189.4434	16.8989
σ	7.9958	0.03011	σ	16.2861	0.3891	σ	7.2559	0.4906
S_i	0.9405	13.3539		0.4866				0.8369

Using the RMSE of Rainfall data with an interval of 10 on the evaluation set, the S_i values can be formulated as

$$S_1 = \frac{196.3321/7.9958}{189.4434/7.2559}$$

$$= 0.9405$$

$$S_2 = \frac{206.8939/16.2861}{189.4434/7.2559}$$
$$= 0.4866$$

Using MAPE, the S_i values are calculated as

$$S_1 = \frac{16.5529/0.0311}{16.0124/0.3891}$$

$$= 13.3539$$

$$S_2 = \frac{16.8989/0.4906}{16.0124/0.3891}$$

$$= 0.8369$$

From table 4.12, the except in the MAPE for ARIMA model where $S_i > S_o$ with the value of 13.3539 where the selecting a single model is possible, all other cases show that the values are consistent with RMSE and MAPE in [Yu et al., 2005] so the combining method is recommended.

Table 4.13 shows the Forecast Accuracy Measure on the three conventional models with a difference of 15 on three evaluation sets ([Yu et al., 2005]) using the rainfall data. In this subsection it was realized that the ANN model had the forecasting accuracy measure for both the RMSE and MAPE. This was then used in calculating for the S_i values.

Table 4.13: The Forecast Accuracy Measure on three different Models with three evaluation sets having a difference of 15 on Rainfall Data

ARIMA			ES			ANN		
Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE
30	187.9571	16.5195	30	189.164	15.8661	30	181.1799	16.21514
45	236.0765	23.7310	45	317.406	32.2396	45	204.6659	20.16308
60	208.0172	16.5734	60	286.6984	16.5734	60	204.9911	17.7980
mean	210.6836	18.9413	Mean	264.4228	21.5597	Mean	196.9456	18.0587
σ	24.1703	4.1481	σ	66.9601	9.2558	σ	13.6545	1.9868
S_i	0.6043	0.5024		0.2738	0.2563			

Using the RMSE of Rainfall data with an interval of 15 on the evaluation set, the S_i values can be formulated as

$$S_1 = \frac{210.6836/24.1703}{196.9456/13.6545}$$

$$= 0.6043$$

$$S_2 = \frac{264.4228/66.9601}{196.9456/13.6545}$$

$$= 0.2738$$

Using MAPE, the S_i values are calculated as

$$S_1 = \frac{18.9413/4.1481}{18.0587/1.9868}$$

$$= 0.5024$$

$$S_2 = \frac{21.5597/9.2558}{18.0587/1.9868}$$

$$= 0.2563$$

From the S_i values calculated, the results show lesser relative prediction performance stability values than that of the threshold value of 4 with three evaluation sets on an increased interval. Per that, the decision is to combine the models.

4.3.2 Varying the number of evaluation sets and interval

This subsection presents different evaluation sets from [Yu et al., 2005] and intervals being varied on the same data as used above. We propose that varying the number of evaluation sets and also increased interval may affect the decision on whether to use single or combined models.

Table 4.14 presents the forecasting performance results on four evaluation sets (10, 12, 14, 16) with an increasing interval of 2 using data from the cocoa production. Here we have reduced the evaluation set interval to 2 and have increase the number of evaluation sets to 4. The minimum value (PI_i) of all the three models of the RMSE measure of accuracy is 251,745.9 and its corresponding standard deviation is 131,616.1 and that of MAPE are 28.47338 and its standard deviation is 18.48352 suggesting that the exponential smoothing model is promising for forecasting . Using the same formulation in calculating for the S_i values as shown for the three evaluation sets, the S_i values for this section as shown in the table are less than that of the threshold. Hence the decision is to combine the models for forecasting or predicting.

Table 4.14: The Forecast Accuracy Measure on four evaluation sets in an increasing interval on Cocoa Production Data

ARIMA			ES			ANN		
Eval Sets	RMSE	MAPE	Eval Sets	RMSE	MAPE	Eval Sets	RMSE	MAPE
10	139355.5	12.79503	10	137698.14	12.67107	10	232791.15	23.36946
12	293957.14	30.1481	12	145388.07	13.4195	12	339402.75	39.5408
14	444973.48	55.6634	14	402937.67	49.9237	14	458415.91	56.1135
16	387205.58	45.6309	16	320959.55	37.8793	16	304086.83	35
Mean	316372.925	36.0594	Mean	251745.8575	28.4734	Mean	333674.16	38.5060
σ	133407.9007	18.7270	σ	131616.0581	18.4835	σ	94244.5626	13.5708
S_i	1.2398	1.2500					1.8510	1.8419

Using the RMSE of Cocoa production data with an interval of 2 on the evaluation set, the S_i values can be formulated as

$$S_1 = \frac{316372.925/133407.9007}{251745.8575/131616.0581}$$

$$= 1.2398$$

$$S_2 = \frac{333674.16/94244.5626}{251745.8575/131616.0581}$$

$$= 1.8510$$

Using MAPE, the S_i values are calculated as

$$S_1 = \frac{36.0594/18.7270}{28.4734/18.4835}$$

$$= 1.2500$$

$$S_2 = \frac{38.5060/13.5708}{28.4734/18.4835}$$

$$= 1.8419$$

The table 4.15 shows the evaluation sets with intervals (20, 30, 40, 50) with 10 intervals each on the three models used in the experiment and the performance indication factors are also calculated with same procedure. In this table too, ANN is selected as the "best" model due to its minimum forecasting accuracy measure in both the RMSE and the MAPE approach.

Also, in case since all cases $S_i < S_o$, hence we combine the models.

Table 4.15: The Forecast Accuracy Measure on four evaluation sets in an increasing interval of 10 on Cocoa Production Data

ARIMA			ES			ANN		
Eval Sets	RMSE	MAPE	Eval Sets	RMSE	MAPE	Eval Sets	RMSE	MAPE
20	374484.33	42.9144	20	324484.33	36.0373	20	289854.36	32.4129
30	321522.66	36.7285	30	383595.04	49.8147	30	365318.84	45.1057
40	245781.13	45.16371	40	246163.01	44.9785	40	238048.63	51.9020
50	255930.62	75.4285	50	238191.62	67.9368	50	214368.96	50.7366
Mean	299429.685	50.0588	Mean	298108.5	49.6918	Mean	276897.6975	45.0393
σ	60254.0051	17.2851	σ	69021.7342	13.4357	σ	66846.2579	8.9254
S_i	1.1997	0.5739		1.0427	0.7329			

Using the RMSE of Cocoa production data with an interval of 2 on the evaluation set, the S_i values can be formulated as

$$S_1 = \frac{299429.685/60254.0051}{276897.6975/66846.2579}$$

$$= 1.1997$$

$$S_2 = \frac{298108.5/69021.7342}{276897.6975/66846.2579}$$

$$= 1.0427$$

Using MAPE, the S_i values are calculated as

$$S_1 = \frac{50.0588/17.2851}{45.0393/8.9254}$$

$$= 0.5739$$

$$S_2 = \frac{49.6918/13.4357}{45.0393/8.9254}$$

$$= 0.7329$$

From table 4.16 the same calculations are done. The S_i values have been given as 0.8369, 0.8957 for RMSE and 1.1817 and 1.0794 for MAPE which also indicates that $S_i < S_o$ so the combination strategy is needed.

Table 4.16: The Forecast Accuracy Measure on four evaluation sets in an increasing interval of 15 on Cocoa Production Data

ARIMA			ES			ANN		
Eval Sets	RMSE	MAPE	Eval Sets	RMSE	MAPE	Eval Sets	RMSE	MAPE
15	405784.5	49.0995	15	352638.99	42.4746	15	326076.62	39.48785
30	321522.66	36.72846	30	383595.04	49.81469	30	365146.92	45.06945
45	227310.74	44.0375	45	227129.96	44.2054	45	223982.71	49.5277
60	272433.61	37.1416	60	272432.8	37.1414	60	271458.3	36.9858
Mean	306762.8775	41.7518	Mean	308949.1975	43.4090	Mean	296666.1375	42.7677
σ	76407.4029	5.9358	σ	71899.5473	5.2225	σ	61840.9512	5.6327
S_i	0.8369		S_i	0.8957	1.1817			1.0794

Using the RMSE of Cocoa production data with an interval of 15 on the evaluation set, the S_i values can be formulated as

$$S_1 = \frac{306762.8775/76407.4029}{296666.1375/61840.9512}$$

$$= 0.8369$$

$$S_2 = \frac{308949.1975/71899.5473}{296666.1375/61840.9512}$$

$$= 0.8957$$

Using MAPE, the S_i values are calculated as

$$S_1 = \frac{43.4090/5.2225}{41.7518/5.9358}$$

$$= 1.1817$$

$$S_2 = \frac{42.7677/5.6327}{41.7518/5.9358}$$

$$= 1.0795$$

Table 4.17 again shows the the prediction accuracy measure on the three conventional models with four different evaluation set using an interval of 2. It can be seen that, with regards to RMSE measure, the ANN model was promising whiles with the MAPE measure the ES model performed well.

With the same assumed threshold value $S_o = 4.0$, the S_i's were calculated from different evaluation sets on the rainfall data and from tables 4.17, the S_i values for RMSE and MAPE are also given as 0.9026 and 0.8470 for both ARIMA respectively, whiles 0.9791 for the RMSE on ES and 0.7710 for MAPE on ANN. They are consistent on the evaluation sets (10, 12, 14, and 16) and since all $S_i < S_o$, then combining strategy of the models is recommended.

Table 4.17: The Forecast Accuracy Measure on four evaluation sets in an increasing interval of 2 on Rainfall Data

ARIMA			ES			ANN		
Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE
10	203.8851	16.57797	10	221.1879	16.45342	10	192.378	16.55198
12	191.16	16.0263	12	209.7344	15.7195	12	181.1799	16.21514
14	183.1688	15.7386	14	199.8959	15.7274	14	175.5889	15.8516
16	172.1152	14.3275	16	189.6525	14.50305	16	164.7866	14.1742
mean	187.5823	15.6676	mean	205.1177	15.6008	mean	178.4834	15.6982
σ	13.3827	0.9589	σ	13.4908	0.8087	σ	11.4936	1.0555
S_i	0.9026	0.8470		0.9791		S_i		0.7710

Using the RMSE of Rainfall data with an interval of 2 on the evaluation set, the S_i values can be formulated as

$$S_1 = \frac{187.5823/13.3827}{178.4834/11.4938}$$

$$= 0.9026$$

$$S_2 = \frac{205.1177/13.4908}{178.4834/11.4938}$$

$$= 0.9791$$

Using MAPE, the S_i values are calculated as

$$S_1 = \frac{15.6676/0.9589}{15.6008/0.8087}$$

$$= 0.8470$$

$$S_2 = \frac{15.6982/1.0555}{15.6008/0.8087}$$

$$= 0.7710$$

It can also be inferred from table 4.18 that ANN is promising to be the best model, and further analysis show that the models must be combined for prediction purposes since all $S_i < S_o$.

Table 4.18: The Forecast Accuracy Measure on four evaluation sets in an increasing interval of 10 on Rainfall Data

ARIMA			ES			ANN		
Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE
20	197.1541	16.5612	20	210.3299	15.7177	20	194.7724	17.2458
30	187.9571	16.51948	30	189.164	15.8661	30	181.1799	16.2151
40	221.0233	22.2215	40	311.4989	31.9834	40	206.7776	20.4538
50	220.3126	21.0333	50	318.0294	31.4886	50	206.5481	19.1980
mean	206.6118	19.0839	mean	257.2556	23.7639	mean	197.3195	18.9659
σ	16.6618	2.9768	σ	67.0181	9.2077	σ	12.1326	1.6165
S_i	0.7625	0.5464	S_i	0.2360	0.2200			

Using the RMSE of Rainfall data with an interval of 10 on the evaluation set, the S_i values can be formulated as

$$S_1 = \frac{206.6118/16.6618}{197.3195/12.1326}$$

$$= 0.7625$$

$$S_2 = \frac{257.556/67.0181}{197.3195/12.1326}$$

$$= 0.2360$$

Using MAPE, the S_i values are calculated as

$$S_1 = \frac{19.0839/2.9768}{18.9659/1.6165}$$

$$= 0.5464$$

$$S_2 = \frac{23.7639/9.2077}{18.9659/1.6165}$$

$$= 0.2200$$

From table 4.19, ANN has the least performance indicators values of 190.2138 for RMSE and 17.4804 for MAPE. Also further analysis with the relative prediction performance stability values are all less than the threshold values so combine the models for forecasting.

Table 4.19: The Forecast Accuracy Measure on four evaluation sets in an increasing interval of 15 on Rainfall Data

ARIMA			ES			ANN		
Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE
15	177.5604	14.7031	15	197.1309	15.1385	15	170.0184	14.9241
30	187.9571	16.5195	30	189.1640	15.8661	30	181.1799	16.2151
45	236.0765	23.7310	45	317.4060	32.2396	45	204.6659	20.1631
60	208.0172	16.5734	60	286.6984	16.5734	60	204.9911	17.7980
mean	202.4028	17.8817	mean	247.5998	19.9544	mean	190.2138	17.2751
σ	25.7634	3.9952	σ	64.1962	8.2110	σ	17.4804	2.2557
S_i	0.7220	0.5844		0.35445	0.31733			

Using the RMSE of Rainfall data with an interval of 15 on the evaluation set, the S_i values can be formulated as

$$S_1 = \frac{202.4028/25.7634}{190.2138/17.4804}$$

$$= 0.7220$$

$$S_2 = \frac{247.5998/64.19622}{190.2138/17.4804}$$

$$= 0.35445$$

Using MAPE, the S_i values are calculated as

$$S_1 = \frac{17.8817/3.9952}{17.2751/2.2557}$$

$$= 0.5844$$

$$S_2 = \frac{19.9544/8.2110}{17.2751/2.2557}$$

$$= 0.31733$$

4.3.3 Discussion on varying the number of evaluation sets and interval

It is observed from the S_i values that the relative prediction performance stability of the models are consistent with the two measures of accuracy of the performance of the prediction or forecasting on cocoa production data. Researchers should consider combining models when using data on cocoa production for a better prediction rather than selecting a single model no matter the size of the evaluation sets and the interval on the evaluation sets.

It again follows that, to predict the rainfall patterns in the future, combined models must be considered as a priority than selecting a single model.

The impact of the evaluation sets on the performance stability of models in forecasting is highly seen from the results. It shows a consistent performance in the decision process of model selection. It can also be said, the size of the evaluation sets does not affect the forecasting decision of a model whether to select a single model or combined individual competing models no matter the sample size of the entire data for different forecasting models.$S_i < S_o$ and the conclusion is that the combining method is allowed. In all, the performance indicator values on the evaluation sets as the number increases from three to four on different intervals was explained by 97% of the entire results and this shows that using different time series models for forecasting, the size of the evaluation sets does not influence the decision to select a single model or combining the models for forecasting no matter the sample size of the entire data. Also whether one measure of accuracy is used or more than one are used, there is no relationship between the measures of accuracy (performance indicators RMSE and MAPE) with the size of the evaluation sets and the interval on the evaluation sets. In conclusion, the size of evaluation sets has no influence or impact on model selection.

4.3.4 Modified Relative Prediction Performance Stability

The performance indicators from [Yu et al., 2005] was based on mean. Hence a further analysis was conducted on the decision rule used by [Yu et al., 2005] in the calculation for the relative prediction stability values by using the median since the mean is affected by outliers and could not be a good measure for accuracy. The median and median deviations as in section 3.4 were calculated and were used by the decision rule to check if the decision will still be the selection of a single model for forecasting or combining the methods as size of evaluation sets are varied across intervals from the rainfall data.

From the varying interval of 10 with three size of the evaluation sets as used in [Yu et al., 2005], it was deduce that there is a slight change in the relative prediction performance stability values (S_i) when the median was used and could been seen from table 4.20 as 1.2258, 17. 5876, 0.6265, 0.6628 as compared with the other values from table 4.12 as 0.9405, 13.3539, 0.4866, 0.8369. In both cases only one value was greater than the threshold value in the case of the ARIMA model for MAPE and the decision rule could be used to select this single model for forecasting but the other three values are less than the threshold value and therefore the conclusion is that the combining strategy is recommended

Table 4.20: The Forecast Accuracy Measure on three evaluation sets on the decision using median with Rainfall Data

ARIMA			ES			ANN		
Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE
10	203.8851	16.57797	10	221.1879	16.45342	10	192.378	16.55198
20	197.1541	16.5612	20	210.3299	15.7177	20	194.7724	17.2458
30	187.9571	16.51948	30	189.164	15.86609	30	181.1799	16.21514
Median	197.1541	16.5612	Median	210.3299	15.7177	Median	194.7724	17.2458
σ	8.0588	0.03180	σ	16.8210	0.5307	σ	9.7593	0.8785
S_i	1.2258	17.5876		0.6265				0.6628

Using the RMSE of Rainfall data with an interval of 10 on the evaluation set, the S_i values which is calculated with equation 3.5 is

$$S_1 = \frac{197.1541/8.0588}{194.7724/9.7593}$$

$$= 1.2258$$

$$S_2 = \frac{210.3299/16.8210}{194.7724/9.7593}$$

$$= 0.6265$$

Using MAPE, the S_i values are calculated as

$$S_1 = \frac{16.5612/0.03180}{15.7177/0.5307}$$

$$= 17.5876$$

$$S_2 = \frac{17.2458/0.8785}{15.7177/0.5307}$$

$$= 0.6628$$

The tables 4.21, 4.22 and 4.23 present the forecast accuracy measure on different evaluation sets on the decision using median with Rainfall Data. The S_i values from these table are all less than the threshold values. Hence the combination of models are required here.

Table 4.21: The Forecast Accuracy Measure on three evaluation sets on the decision using median with Rainfall Data

ARIMA			ES			ANN		
Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE
30	187.9571	16.5195	30	189.164	15.8661	30	181.1799	16.2151
45	236.0765	23.7310	45	317.406	32.2396	45	204.6659	20.1631
60	208.0172	16.5734	60	286.6984	16.5734	60	204.9911	17.7980
median	208.0172	16.5734	Median	286.6984	16.5734	Median	204.6659	17.7980
σ	24.3899	5.0613	σ	72.3046	11.0889	σ	16.6087	2.0123
S_i	0.6921		S_i	0.3218	0.4564	S_i		2.7010

Using the RMSE of Rainfall data with an interval of 10 on the evaluation set, the S_i values which is calculated with equation 3.5 is

$$S_1 = \frac{208.0172/24.3899}{204.6659/16.6087}$$

$$= 0.6921$$

$$S_2 = \frac{286.6984/72.3046}{204.6659/16.6087}$$

$$= 0.3218$$

Using MAPE, the S_i values are calculated as

$$S_1 = \frac{16.5734/11.0889}{16.5734/5.0613}$$

$$= 0.4564$$

$$S_2 = \frac{17.7980/2.0123}{16.5734/5.0613}$$

$$= 2.7010$$

Table 4.22: The Forecast Accuracy Measure on four evaluation sets in an increasing interval of 2 on Rainfall Data

ARIMA			ES			ANN		
Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE
10	203.8851	16.57797	10	221.1879	16.45342	10	192.378	16.55198
12	191.16	16.02629	12	209.7344	15.7195	12	181.1799	16.21514
14	183.1688	15.73862	14	199.8959	15.72739	14	175.5889	15.85159
16	172.1152	14.32749	16	189.6525	14.50305	16	164.7866	14.17415
median	187.1644	15.8825	median	204.8152	15.7235	median	178.3844	16.0334
σ	16.4016	1.2131	σ	16.5283	1.0055	σ	14.0775	1.3769
S_i	0.9006	0.8372	S_i	0.9779				0.7447

With an interval of 2 on four evaluation sets, the S_i values using the RMSE forecast accuracy measure are calculated as follows

$$S_1 = \frac{187.1644/16.4016}{178.3844/14.0775}$$

$$= 0.9006$$

$$S_2 = \frac{204.8152/16.5283}{178.3844/14.0775}$$

$$= 0.9779$$

Using MAPE, the S_i values are calculated as

$$S_1 = \frac{15.8825/1.2131}{15.7235/1.005}$$

$$= 0.8372$$

$$S_2 = \frac{16.0334/1.3769}{15.7235/1.005}$$

$$= 0.7447$$

Table 4.23: The Forecast Accuracy Measure on four evaluation sets in an increasing interval of 10 on Rainfall Data

ARIMA			ES			ANN		
Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE
20	197.1541	16.5612	20	210.3299	15.7177	20	194.7724	17.2458
30	187.9571	16.51948	30	189.164	15.86609	30	181.1799	16..21514
40	221.0233	22.22148	40	311.4989	31.98337	40	206.7776	20.45377
50	220.3126	21.03331	50	318.0294	31.48861	50	206.5481	19.19802
median	208.7335	18.7973	median	260.9144	23.6774	median	200.6603	18.2219
σ	20.6259	3.6683	σ	82.2430	11.2778	σ	15.5923	2.3360
S_i	0.7864	0.6569	S_i	0.2465	0.2691			

With an interval of 10 on four evaluation sets in our case, the S_i values using the RMSE forecast accuracy measure are calculated as follows

$$S_1 = \frac{208.7335/20.6259}{200.6603/15.5923}$$

$$= 0.7864$$

$$S_2 = \frac{260.9144/82.2430}{200.6603/15.5923}$$

$$= 0.2465$$

Using MAPE, the S_i values are calculated as

$$S_1 = \frac{18.7973/3.6683}{18.2219/2.3360}$$

$$= 0.6569$$

$$S_2 = \frac{23.6774/11.2778}{18.2219/2.3360}$$

$$= 0.2691$$

From table 4.24 as interval increased with same size of evaluation sets, $S_i < S_o$, the decision is combine the models for forecasting. The evaluation sets were increased to four sizes across a varying interval to check the decision. It was observed from the table values that the relative prediction performance stability values were not consistent with the other values calculated across the same size of the evaluation sets across the same varying interval in the case of the means. The stability value for ANN is 4.2709 and could be selected for prediction but since all other values are less than the threshold value per the decision rule by [Yu et al., 2005]), the combining method is appropriate for forecasting. From the discussions above it can be concluded that the median causes a slight change in the prediction performance stability values but does not affect the decision rule since majority of these values are less than the assumed threshold values used by [Yu et al., 2005].

Table 4.24: The Forecast Accuracy Measure on four evaluation sets in an increasing interval of 15 using median on Rainfall Data

ARIMA			ES			ANN		
Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE	Eval sets	RMSE	MAPE
15	177.5604	14.7031	15	197.1309	15.13853	15	170.0184	14.92406
30	187.9571	16.51948	30	189.164	15.86609	30	181.1799	16.21514
45	236.0765	23.73097	45	317.406	32.23955	45	204.6659	20.16308
60	208.0172	16.57342	60	286.6984	16.57342	60	204.9911	17.79801
median	197.9872	16.5465	median	241.9147	16.2198	median	192.9229	17.0066
σ	32.1656	5.2448	σ	125.3984	11.3590	σ	21.7492	2.7886
S_i	0.6939	0.2643	S_i	0.2175		S_i		4.2709

With an interval of 15 on four evaluation sets in our case, the S_i values using the RMSE forecast accuracy measure are calculated as follows

$$S_1 = \frac{197.9872/32.1656}{192.9229/21.7492}$$

$$= 0.6939$$

$$S_2 = \frac{241.9147/125.3984}{192.9229/21.7492}$$

$$= 0.2175$$

Using MAPE, the S_i values are calculated as

$$S_1 = \frac{16.5465/5.2448}{16.2198/11.3590}$$

$$= 0.2643$$

$$S_2 = \frac{17.0066/2.7886}{16.2198/11.3590}$$

$$= 4.2709$$

CHAPTER 5

Summary and Conclusion

5.1 Introduction

This chapter presents conclusions drawn from the study and some recommendations to be considered when selecting a model in times series analysis for prediction into future unseen data.

5.2 Summary

Several mathematical models have been used in time series data analysis for prediction or forecasting in recent years. Many researchers have often encountered problems in the model selection and how the selected model can be used to forecast or predict accurately into future data values from series of competing models as the true or best model .In time series forecasting applications, there is the practice of selecting a single model from a series of multiple competing models in terms of some selection procedures for final estimation, interpretation, and projections to be made based on the selected model. The study looks at model selection in time series using autoregressive integrated moving average (ARIMA) models by comparing the performance of graphical method and Akaike Information Criteria (AIC) to select the true model and to evaluate the impact of the size of the evaluation sets on the relative prediction performance stability of root mean square error(RMSE) and mean absolute percentage error (MAPE) of the selected model or models for prediction by using different linear and non-linear models (ARIMA, ES, ANN) were used. To verify the efficiency and the reliability of the proposed method and the modeling technique, simulations were conducted and two real time series data of yearly cocoa production in Ghana and monthly rainfall data in Accra were used. The results of the study reveal that the Akaike information criterion performs better than the graphical method but in exceptional cases both methods perform better in AR(2) and ARMA (2,2). Also there is no significance impact on the size of the evaluation sets on whether to select a single model or combine the models for forecasting into future unseen data. Also the interval on the evaluation sets did not influence the decision no matter the performance indicators used.

5.3 Conclusion

It was observed from the study that, in comparison to the performance of the graphical method and the Akaike information criterion (AIC) in selecting the ARIMA models, the information criteria performs better than the graphical method. It also showed that no single method performs better than the other since both methods can under-fit or over–fit models. The information criteria has minimum error in under-fitting and over-fitting a model as compared to the graphical method. The models performing well from the two methods are AR(2) and ARMA (2,2). This means that though one method may be superior over the other methods of model selection procedures or methods, in instances where the two methods are combined in forecasting, the combined models might be performing better than using only a single method.

Also, in verifying for the size of the evaluation sets in forecasting, whether to select a single model or combine the models of different models, our findings showed that the size of the evaluation sets may not influence the decision of selecting or combining since 97% of the decisions were to combine the models. Also no matter the measures of accuracy of the performance stability used,the decision will not change the outcome. It was also observed that the intervals on the evaluation sets do not also affect the decision of selecting or combining but the decision is highly base on the performance stability indicators which are highly influence by the root mean square values (RMSE) and the mean absolute percentage error (MAPE). From the research, it can be concluded that the combining strategy performs better in most cases in forecasting in time series analysis than selecting a single model.

In addition to that, though there was a modification on the computations of the relative prediction performance stability formerly utilized by [Yu et al., 2005], the decision rule still remains the same. Hence, whether the use of the mean or median on different the size of evaluation sets and interval, the combining strategy still outperforms.

5.4 RECOMMENDATIONS

On the bases of the research findings on the study, the following recommendations are made for considerations in further research.

- Per our findings, it is recommended that, the use of information criteria should be employed by researchers in selecting approximating models.
- It is again recommended that the combined approach should be employed in selecting the true model.
- In the computations of the relative prediction performance stability the median approach should be used since it is not affected by extreme values.

REFERENCES

[Adhikari and Agrawal, 2012] Adhikari, R. and Agrawal, R. (2012). Combining multiple time series models through a robust weighted mechanism. In *Recent Advances in Information Technology (RAIT), 2012 1st International Conference on*, pages 455–460. IEEE.

[Akaike, 1974] Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.

[Ankrah et al., 2015] Ankrah, S., Peiris, B., and Thattil, R. (2015). Weighted ranking procedure for combining univariate time series models. *Tropical Agricultural Research*, 26(3).

[Armstrong, 2001] Armstrong, J. S. (2001). *Principles of forecasting: a handbook for researchers and practitioners*, volume 30. Springer Science & Business Media.

[Bates and Granger, 1969] Bates, J. M. and Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468.

[Box and Jenkins, 1970] Box, G. and Jenkins, G. (1970). Times series analysis forecasting and control. holden-day san francisco.

[Breiman, 1996] Breiman, L. (1996). Stacked regressions. *Machine learning*, 24(1):49–64.

[Buckland et al., 1997] Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997). Model selection: an integral part of inference. *Biometrics*, pages 603–618.

[Burnham and Anderson, 2002] Burnham, K. P. and Anderson, D. (2002). Model selection and multi-model inference. 2002.

[Chatfield,] Chatfield, C. The analysis of time series: An introduction. *Chapman and Hall/ CRC.*

[Clemen, 1989] Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International journal of forecasting*, 5(4):559–583.

[Draper, 1995] Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 45–97.

[George and McCulloch, 1997] George, E. I. and McCulloch, R. E. (1997). Approaches for bayesian variable selection. *Statistica sinica*, pages 339–373.

[Hamzaçebi, 2008] Hamzaçebi, C. (2008). Improving artificial neural networks' performance in seasonal time series forecasting. *Information Sciences*, 178(23):4550–4559.

[Hipel and McLeod, 1994] Hipel, K. W. and McLeod, A. I. (1994). *Time series modelling of water resources and environmental systems*, volume 45. Elsevier.

[Hoeting et al., 1999] Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical science*, pages 382–401.

[Juditsky and Nemirovski, 2000] Juditsky, A. and Nemirovski, A. (2000). Functional aggregation for nonparametric regression. *Annals of Statistics*, pages 681–712.

[Madigan et al., 1995] Madigan, D., York, J., and Allard, D. (1995). Bayesian graphical models for discrete data. *International Statistical Review/Revue Internationale de Statistique*, pages 215–232.

[Raftery, 1995] Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological methodology*, pages 111–163.

[Raicharoen et al., 2003] Raicharoen, T., Lursinsap, C., and Sanguanbhokai, P. (2003). Application of critical support vector machine to time series prediction. In *Circuits and Systems, 2003. ISCAS'03. Proceedings of the 2003 International Symposium on*, volume 5, pages V–V. IEEE.

[Reid, 1968] Reid, D. J. (1968). Combining three estimates of gross domestic product. *Economica*, 35(140):431–444.

[Reid, 1969] Reid, D. J. (1969). *A comparative study of time series prediction techniques on economic data*. University of Nottingham, Library Photographic Unit.

[Schwarz et al., 1978] Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.

[Terui and Van Dijk, 2002] Terui, N. and Van Dijk, H. K. (2002). Combined forecasts from linear and nonlinear time series models. *International Journal of Forecasting*, 18(3):421–438.

[Tseng et al., 2002] Tseng, F.-M., Yu, H.-C., and Tzeng, G.-H. (2002). Combining neural network model with seasonal time series arima model. *Technological Forecasting and Social Change*, 69(1):71–87.

[Van Der Voort et al., 1996] Van Der Voort, M., Dougherty, M., and Watson, S. (1996). Combining kohonen maps with arima time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies*, 4(5):307–318.

[Wolpert, 1992] Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, 5(2):241–259.

[Yang, 2003] Yang, Y. (2003). Regression with multiple candidate models: selecting or mixing? *Statistica Sinica*, pages 783–809.

[Yu et al., 2005] Yu, L., Wang, S., Lai, K., and Nakamori, Y. (2005). Time series forecasting with multiple candidate models: selecting or combining. *Journal of Systems Science and Complexity*, 18(1):1–18.

[Zhang, 2003] Zhang, G. P. (2003). Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175.

[Zou and Yang, 2001] Zou, H. and Yang, Y. (2001). Combining time series models for forecasting, working paper. *Department of Statistics, Iowa State University.*

[Zou and Yang, 2004] Zou, H. and Yang, Y. (2004). Combining time series models for forecasting. *International journal of Forecasting*, 20(1):69–84.

Appendix A

```
set.seed(10)
sim.arma=arima.sim(list(ar=c(0.4,0.4), ma=c(0.7,-0.1)), n=15)
sim.arma
par(mfrow=c(2,2))
acf(sim.arma,main="ACF ofARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
library(forecast)
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order =c( 3,0,3))
m3
set.seed(11)
sim.arma=arima.sim(list(ar=c(0.4,-0.6),ma=c(0.6,-0.2)), n=15)
sim.arma
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order =c( 3,0,3))
m3
set.seed(12)
sim.arma=arima.sim(list(ar=c(0.5,0.3), ma=c(0.5,0.3)),n=15)
```

```
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order =c( 3,0,3))
m3
set.seed(13)
sim.arma=arima.sim(list(ar=c(0.9,-0.2), ma=c(0.9,-0.2)),n=15)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order=c(3,0,3))
m3
set.seed(14)
sim.arma=arima.sim(list(ar=c(0.5,-0.3), ma=c(0.5,-0.3)) ,n=15)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order=c(3,0,3))
m3
set.seed(10)
```

```
sim.arma=arima.sim(list(ar=c(0.4,0.4),ma=c(0.7,-0.1)),n=25)
sim.arma
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(11)
sim.arma=arima.sim(list(ar=c(0.4,-0.6),ma=c(0.6,-0.2)),n=25)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(12)
sim.arma=arima.sim(list(ar=c(0.5,0.3), ma=c(0.5,0.3)),n=25)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.ar,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
```

```
m3
set.seed(13)
sim.arma=arima.sim(list(ar=c(0.9,-0.2), ma=c(0.9,-0.2)),n=25)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,2))
m3
set.seed(14)
sim.arma=arima.sim(list(ar=c(0.3,-0.5), ma=c(0.3,-0.5)) ,n=25)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.ar,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(10)
sim.arma=arima.sim(list(ar=c(0.4,0.4),ma=c(0.7,-0.1)),n=35)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
```

```
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(11)
sim.arma=arima.sim(list(ar=c(0.4,-0.6),ma=c(0.6,-0.2)),n=35)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(12)
sim.arma=arima.sim(list(ar=c(0.5,0.3), ma=c(0.5,0.3)),n=35)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.ar,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(13)
sim.arma=arima.sim(list(ar=c(0.9,-0.2), ma=c(0.9,-0.2)),n=35)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
```

```
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(14)
sim.arma=arima.sim(list(ar=c(0.5,-0.3), ma=c(0.5,-0.3)) ,n=35)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order=c(3,0,3))
m3
set.seed(10)
sim.arma=arima.sim(list(ar=c(0.4,0.4),ma=c(0.7,-0.1)),n=50)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(11)
sim.arma=arima.sim(list(ar=c(0.4,-0.6),ma=c(0.6,-0.2)),n=50)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
```

```
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(12)
sim.arma=arima.sim(list(ar=c(0.5,0.3), ma=c(0.5,0.3)),n=50)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.ar,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(13)
sim.arma=arima.sim(list(ar=c(0.9,-0.2), ma=c(0.9,-0.2)),n=50)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(14)
sim.arma=arima.sim(list(ar=c(0.3,-0.5), ma=c(0.3,-0.5)) ,n=50)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.ar,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
```

```
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(10)
sim.arma=arima.sim(list(ar=c(0.4,0.4),ma=c(0.7,-0.1)),n=100)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1 m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(11)
sim.arma=arima.sim(list(ar=c(0.4,-0.6),ma=c(0.6,-0.2)),n=100)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(12)
sim.arma=arima.sim(list(ar=c(0.5,0.3), ma=c(0.5,0.3)),n=100)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.ar,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
```

```
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(13)
sim.arma=arima.sim(list(ar=c(0.9,-0.2), ma=c(0.9,-0.2)),n=100)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(14)
sim.arma=arima.sim(list(ar=c(0.3,-0.5), ma=c(0.3,-0.5)) ,n=100)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.ar,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(10)
sim.arma=arima.sim(list(ar=c(0.4,0.4),ma=c(0.7,-0.1)),n=1000)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
```

```
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(11)
sim.arma=arima.sim(list(ar=c(0.4,-0.6),ma=c(0.6,-0.2)),n=1000)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(12)
sim.arma=arima.sim(list(ar=c(0.5,0.3), ma=c(0.5,0.3)),n=1000)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.ar,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(13)
sim.arma=arima.sim(list(ar=c(0.9,-0.2), ma=c(0.9,-0.2)),n=1000)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
```

```
pacf(sim.arma,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3
set.seed(14)
sim.arma=arima.sim(list(ar=c(0.3,-0.5), ma=c(0.3,-0.5)) ,n=1000)
acf(sim.arma,main="ACF of ARMA(3,0,3) process")
pacf(sim.ar,main="PACF of ARMA(3,0,3) process")
m1=Arima(sim.arma, order =c( 1,0,1))
m1
m2=Arima(sim.arma, order =c( 2,0,2))
m2
m3=Arima(sim.arma, order = c(3,0,3))
m3

    a<-read.csv('C:/Users/user/Documents/cocoa.csv',header = T)
a
library(forecast)
a1=ts(a,start=1948, frequency = 1)
summary(a1)
train1=window(a1,start=1948, end=2000)
test1=window(a1,start=2001 ,end=2015)
m1=auto.arima(train1,stepwise = FALSE,approximation = FALSE)
m1
f1=forecast(m1,h=15)
f1
```

```
test1
accuracy(f1,test1)
train2=window(a1,start=1948,end=1985)
test2=window(a1,start=1986,end=2015)
m2=auto.arima(train2,stepwise = FALSE,approximation = FALSE)
m2
f2=forecast(m2,h=30)
accuracy(f2,test2)
train3=window(a1,start=1948,end=1970)
test3=window(a1,start=1971,end=2015)
m3=auto.arima(train3,stepwise = FALSE,approximation = FALSE)
m3
f3=forecast(m3,h=45)
accuracy(f3,test3)
train4=window(a1,start=1948,end=1955)
test4=window(a1,start=1956,end=2015)
m4=auto.arima(train4,stepwise = FALSE,approximation = FALSE)
m4
f4=forecast(m4,h=60)
accuracy(f4,test4)
m11=ets(train1)
m11
f11=forecast(m11,h=15)
accuracy(f11,test1)
m12=ets(train2)
m12
f12=forecast(m12,h=30)
accuracy(f12,test2)
m13=ets(train3)
```

```
m13
f13=forecast(m13,h=45)
accuracy(f13,test3)
m14=ets(train4)
m14
f14=forecast(m14,h=60)
accuracy(f14,test4)
m21=nnetar(train1)
m21
f21=forecast(m21,h=15)
accuracy(f21,test1)
m22=nnetar(train2)
m22
f22=forecast(m22,h=30)
accuracy(f22,test2)
m23=nnetar(train3)
m23
f23=forecast(m23,h=45)
accuracy(f23,test3)
m24=nnetar(train4)
m24
f24=forecast(m24,h=60)
accuracy(f24,test4)
```