MULTIVARIATE APPROACH TO TIME SERIES MODEL IDENTIFICATION

BY

AGADA, JOSEPH OCHE

(PG/M.Sc/13/65067)

BEING PROJECT REPORT SUBMITTED IN PATIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE (M.Sc) DEGREE IN STATISTICS OF THE UNIVERSITY OF NIGERIA, NSUKKA

CERTIFICATION

The work embodied in this project report is original and has not been submitted for any other degree in this or other universities.

Supervisor:
Professor F.I. Ugwuowo
Department of Statistics
University of Nigeria
Nsukka

Head of Department:
Professor F.I. Ugwuowo
Department of Statistics
University of Nigeria
Nsukka

External Examiner

DEDICATED TO

My little daughter

MERCY OCHIJE AGADA

With whom God blessed my family while this work was still in progress

ACKNOWLEDGEMENT

The idea that materialized to this write-up is strictly divine, besides, this work is only a success because I still have breath in me; therefore, my profound gratitude goes to almighty God who made it all possible.

No success of this nature can be attributed to individual effort therefore I must seize this opportunity to appreciate some individuals who have contributed greatly to the success of this work. I am indebted to my supervisor, Professor F.I. Ugwuowo for his kind attention and assistance through the period of this work. He always listens deeply to me each time I sit down to share ideas with him and his response and contributions have always been a great source of motivation to me. It was great working with you sir, thank you. I also want to express my profound gratitude to entire Lecturers in the department of Statistics for they have all contributed directly or indirectly to the success of this work. I want to specifically mention Mr. M.S. Madukaife, Mr. U.C. Nduka and Mr. S.S. Sanni for their technical support.

My profound gratitude also goes to Assoc. Prof. W.I.E. Chukwu and Dr. A.V. Oladugba for their moral support. I am most greatful to my family members especially my late father, Elder Job Agada, my lovely mother, Mrs Grace Agada and my brother, Pastor Monday Agada for they have always been behind me from the very beginning till date at all cost.

My very special thanks is reserved for my lovely wife, Joy Agada for her encouragement especially towards the academic vocation and my little angel, Mercy, they are great sources of inspiration. I love you all.

ABSTRACT

This work suggests an exact and systematic model identification approach which is entirely new and addresses most of the challenges of existing methods. We developed quadratic discriminant functions for various orders of autoregressive moving average (ARMA) models. An Algorithm that is to be used alongside our functions was also developed. In achieving this, three hundred sets of time series data were simulated for the development of our functions. Another twenty five sets of simulated time series data were used in testing out the classifiers which correctly classified twenty three out of the twenty five sets. The two cases of misclassification merely imply that our Algorithm will require a second iteration to correctly identify the model in question. The Algorithm was also applied to some real life time series data and it correctly classified it in two iterations.

TABLE OF CONTENT

CHAPTER ONE: INTRODUCTION

CHAPTER TWO: LITERATURE REVIEW

CHAPTER THREE: METHODOLOGY

CHAPTER FOUR: RESULTS

CHAPTER FIVE: SUMMARY AND CONCLUSION

CHAPTER ONE

INTRODUCTION

1.1 INTRODUCTION

Model identification is a crucial part of Time Series model development. The main task of Time Series Modeling is to first examine the series at hand so as to establish the theoretical model that generates the Series. This task seems to be the most challenging and most ambiguous in Time Series Modeling. It has been approached from different perspectives over time. One of the most popular approach is the Box and Jenkins approach presented in Box and Jenkins (1976). Their method involves going through some iterative steps before a final model is selected. The initial step involves calculating the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of the series at various lags and comparing their behaviour with known behaviour of some theoretical model and the model that best approximates the sample behavior is tentatively selected. There are two serious problems with their method. First is the fact that one will need to fit several models or do several adjustments to arrive at the final model. This makes the method computationally expensive. Another serious problem is the inability of the method to accurately differentiate between some classes of models. For example, it is not easy to determine the values of p and q in fitting ARMA (p, q) when $p, q \neq 0$ since both the ACF and PACF tail off. However, Tsay and Tiao (1984, 1985) addressed this problem by proposing the use of the extended autocorrelation (EACF) and smallest canonical correlation SCAN respectively. Tsay and Tiao methods are mere extension of the Box and Jenkins approach as they involve comparing behaviour of sample EACF and smallest canonical correlation with the theoretical behaviour. The difficulties in matching these behaviour is even more in these approach because the clear cut off in theoretical EACF table for example is hardly observed in sample EACF, Cryer and Chan (2008).

Far away from the Box and Jenkins approach, Akaike (1969) and lots of other scientists have done several works on various forms of information criteria. Their approach is based on values calculated from residual of already fitted models. The statistic calculated from residual of these fitted models is perceived as information loss as a result of fitting the model. The order of the model that minimizes the information loss is finally adopted. Model identification stage of time series modeling has suffered severe deficiency over time. All the available methods are deficient in terms of accurately selecting the model fit. There is no well defined procedure that gives the exact model or at least with a known error margin before going ahead to fit the model. The approach presented here is well spelt out rules that guides model development. Some of the model identification approaches especially the Box and Jenkins approach is more of art than science as it is highly judgmental.

The information criteria approach is not an exemption in this deficiency. Fitting models with virtually all order before selecting a particular order is almost the same as fitting all the possible models and selecting the one that passes the goodness of fit test. Model selection is a stage that should be put to rest before considering estimation, if one must come back to this stage then the initial approach is supposed to have pre-specified the next model/order to consider. One other important shortfall of the information criteria approach is that its results are comparative meaning that only selected tentative models will have their information criteria calculated compared and then the model with minimum information criteria chosen as the best model.

In this work, we are proposing an exact model identification method which will address most of the issues with available model identification methods. Our method does not require fitting any model into series at hand before selecting the right model. It will be capable of selecting the exact model with very high level of certainty and in event where the selected model is not the appropriate model (since there may be a small error margin); the method also predetermines the next model (from all models under consideration) to be considered. With further transformation as we have in this work, The behaviour of sample ACF and PACF is actually enough information for model identification. The choice of the lag is informed by the fact that ACF and PACF of stationary ARMA models usually cut off with higher lags. The ACF at lag 1 to 4 and PACF at lag 2 to 5 are the information used in this proposed method. Our method utilized the variations inherent in selected ACF and PACF in the classifiers and used them to classify models as ARMA (p, q) with $p+q \leq 2$(with ARMA(22) added to demonstrate the usefulness of our methods in cases of mixed models)

1.2 STATEMENT OF PROBLEMS

In Time Series analysis, there is need to establish the theoretical model to be fitted into a particular time Series at hand before proceeding with subsequent steps, this work particularly, seeks to develop a new method of achieving this.

In this work, we want to build discriminant functions for each of the classes of theoretical ARMA models considered (i.e. AR(1), AR(2), MA(1), MA(2), ARMA(1,1) and ARMA(2,2)) and also develop an algorithm which is a well organized iterative steps to be followed in applying the functions. We equally hope to test our new method out using both simulated and real life time series and comment on its performance.

1.3 SIGNIFICANCE OF THE STUDY

Series of work have been done on time series model identification but all existing methods are characterized with lack of precise problem formulation, presence of heavy individual judgments, and associated high computational cost due to several adjustments needed to arrive at the final model. Existing model identification methods only give merely tentative and inexact perception of the appropriate theoretical model before fitting hence several models are usually fitted at the initial stage. This work is aimed at developing an exact model identification approach which will address the challenges indicated above.

Our method is hoped to provide a systematic and well organized problem formulation which will be capable of reducing computational cost and rigour associated with time series modeling.

1.4 OBJECTIVES OF THE STUDY

This work is aimed at developing an entirely new model identification method which addresses most of the challenges of existing methods. In achieving this, our specific objectives are outlined below.

- To develop quadratic discriminant functions for each of the ARMA models considered and define the iterative steps (algorithm) to be followed in application of the functions.
- To apply the method to both simulated and real time series data.
- To briefly compare the proposed method with existing methods

1.5 SCOPE AND LIMITATION OF THE STUDY

This work is limited to six classes of ARMA models which are AR(1), AR(2), MA(1), MA(2), ARMA(1,1) and ARMA(2,2). This method can only be applied to those classes of ARMA model meaning that other classes of models like Autoregressive integrated moving average model (ARIMA), seasonal ARIMA, Autoregressive conditional heteroscedasticity (ARCH) models, Generalized Conditional Heteroscedasticity Model (GARCH) etc. are not covered by our method.

CHAPTER TWO

LITERATUR REVIEW

2.1 INTRODUCTION

Though, no work has ever been done on application of Multivariate analytical methods in Time Series ARMA model identification, a lot has been done on model identification. In this section, we are going to briefly look at the extent of work that has been done on model identification, from use of ACF, PACF, EACF and Canonical correlation to test of hypothesis and later, use of information criteria.

2.2 REVIEW OF LITERATURE

Box and Jenkins (1976) have the credit for the most popular judgmental model identification method. Their approach adopts theoretical models to be fitted into Time Series based on careful examination of the sample ACF and PACF. Behaviour of sample ACF and PACF calculated from the Series to be fitted are compared with that of available theoretical model and the model with similar behavior is suspected to have generated the Series to be fitted and therefore considered. Theoretically, the ACF and PACF of an AR(p) tail off and cut off at lag p respectively. ACF and PACF of MA(q) cut off at lag q and tails off respectively while both ACF and PACF of ARMA(p,q) tail off. This approach has come under serious criticism based on a couple of arguments. First is that the method is highly judgmental because the behaviour of the sample ACF and PACF cannot be perfectly matched with the theoretical behaviour in most cases, therefore deciding whether ACF or PACF cut off at certain lag depends on individual Judgment. Box and Jenkins approach is merely tentative, the model identified at this stage is subject to series of modifications and the models selected at the end are usually different from the initial model. This will obviously lead to fitting and re-fitting of several models and series of goodness-of-fit tests. It is based on the arguments above and other similar arguments that Anderson (1975) concluded that Box and Jenkins approach, though attractive but in practice both time consuming, difficult and relatively expensive computationally.

One of the shortfalls of the Box and Jenkins approach is specifying the value of p and q in ARMA(p,q) model where $p,q \neq 0$ since both ACF and PACF tail off in this case. Tsay and Tiao (1984) proposed the extended autocorrelation approach to address this problem. The EACF method uses the fact that if the AR part of a mixed ARMA model is known, ñfiltering outò the autoregression from the observed series results in a pure MA process which now enjoy the cut off property. Chan (1999) in a comparative simulation study concluded that the EACF approach has a good sampling property for moderately large sample size. In formalizing the procedure, EACF is usually calculated and compared with the theoretical behaviour of EACF for available models and the values of p and q are consequently considered. Cryer and Chan (2008) specifically pointed out that there will never be a clear cut off in the sample EACF as we have in the theoretical EACF. This makes the method even more difficult than the Box and Jenkins approach.

Bhansali (1993) worked on a hypothesis testing approach to model identification. His approach involve pre-fitting all the suspected models and testing hypothesis concerning the order until the right model is established. Consider an AR(p), the procedure is to test the hypothesis that the model is an AR(p) against an alternative that it is an AR(p+1). We continue increasing the value of value of p by 1 and repeating the hypothesis until the null hypothesis is accepted.

Whittle (1963) introduced an order selection procedure which is based on residual variance plot. Take an AR(p) for example, the lowest possible value of p is chosen and the model fitted, the error variance σ_e^2 is estimated from the fitted model as $\hat{\sigma}_e^2$, if the order fitted is lower than the actual order then $\hat{\sigma}_e^2$ is greater than σ_e^2. p is increased and the model fitted with $\hat{\sigma}_e^2$ calculated for each p. p's are plotted against $\hat{\sigma}_e^2$'s and the value of p that corresponds to the point where $\hat{\sigma}_e^2$ stopped reducing is adopted as the order of the process. Jenkins and Watts (1968) also used this method and further suggested that the method can be used for order selection of MA and ARMA models.

Whittleôs (1963) procedure motivated Akaike into research on order selection for which he is so popular today.

Akaike (1969) improved on Whittle (1963) procedure for order selection of AR processes. He suggested that AR(p) be fitted for p=0,1,2,é ,m when m is the upper bound placed on p. Value of order selection criteria which is the standardized error term calculated for each of the models and p that minimizes the calculated order selection criterion is selected as the order of the process.

Akaike (1969) proposed the use of the final prediction error, FPE. For a particular value of p, the FPE is defined as

$$FPE(p) = \frac{N+p}{N-p}\hat{\sigma}_p^2$$

é ..(1)

where N is the series length, $\hat{\sigma}_p^2$ is the MLE of the variance of the residual based on the p^{th} order model. The p for which FPE is lowest is adopted as the order of the AR model.

Akaike (1974) introduced the popular Akaikeôs Information Criterion (AIC), AIC broadened the scope of usage of the information criteria. It can be used in Time Series model identification in the wide range of situations. Its usage is not even restricted to Time Series analysis Zazli (2002). For an AR(p) process, the AIC is given as $AIC(p) = N\log\left(\hat{\sigma}_p^2\right) + 2p$

é ..(2)

the p that minimizes the AIC is selected as the order of the AR(p) process. As stated earlier the AIC can be used in selecting order of general ARMA(p,q) process. For ARMA(p,q), the AIC is given as

$$AIC(p,q) = (N-p)\log\left(\hat{\sigma}_{p,q}^2\right) + 2(P+q+1)$$

é é é é é é é é é é é é é é é é .(3)

the combination of p and q that minimizes the AIC is the order of the ARMA (p,q) process.

Akaike (1979) developed a Bayesian expansion of the AIC called the Bayesian Information Criterion BIC. The BIC can complement AIC as Sibata (1979) has shown that AIC criterion tends to overestimate the true order of an AR model while the BIC under estimates it. The true order value should lie between the AIC and BIC estimates. The BIC is defined as

$$BIC(p) = N\log\hat{\sigma}_e^2 - (N-p)\log\left(1-\frac{p}{N}\right) + p\log N + p\log\left\{p^{-1}\left(\frac{\hat{\sigma}_X^2}{\hat{\sigma}_p^2}\right) - 1\right\}$$

é é é é é .(4)

where $\hat{\sigma}_X^2$ is estimated variance of process that generated the series at hand. Other literatures on information criteria are presented in: Akaike (1974, 1980); Anderson (2008); Burnham and Anderson (2002, 2004); Schwartz (1978); Hannan and Quinn (1979); Parzen (1974).

Zazli (2002) carried out a comparative simulation study on the following order selection criteria:

- FPE
- AIC
- BIC
- Schwartzös selection criterion
- Hannan and Quinn criterion (HQ)

and concluded that BIC and Schwartzös criteria gave almost identical results and FPE and AIC also gave almost identical results. In general , it was concluded that both BIC and Schwartzös criteria gave the best results for AR order selection followed by Hannan-Quinn criterion, FPE and AIC. Zazi (2002) advised that different order selection criteria be used for estimating the order of the fitted short time series model.

Clark and Troskie (2008) pioneered the use of the Bozdogan information criterion (ICOMP) in time series model selection. ICOMP was initially used in building regression models. They concluded that results obtained using the criterion is comparable with results of AIC, BIC and other criterion compared. He emphasized the conclusion of Zazli (2002) which has it that a single information criterion should not be used in model identification.

Pukkila et al (1990) and Pukkila (1993) moved away from use of information criterion and proposed another method of ARIMA model identification. The procedure was based on the fact that a white noise process can be viewed as an AR(0), they fitted various order into ARMA(p,q) and checked if it transfers the residual into an AR(0). The order of the process that transfers the residual into an AR(0) is selected.

CHAPTER THREE

METHODOLOGY

3.1 THE BAYESIAN AND FISHER'S CLASSIFICATION RULE

The problem of Discriminant Analysis is the problem of classifying a given observation into one of the available k-classes $(k \geq 2)$. Consider the population

$X/G = k \sim N_p(\mu_k, \Sigma_k)$;

$$f_k(x) = \frac{1}{(2\pi)^{\frac{p}{2}} \left|\Sigma_k\right|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(x-\mu_k\right)' \Sigma^{-1}\left(x-\mu_k\right)\right\}$$ é é é é é é .é é é é é ..

(5)

Note that $X / G = k \sim N_p\left(\mu_k, \Sigma_k\right)$ implying that the population X whose group is k is normal with parameters μ_k and Σ_k is the basic assumption of Discriminant Analysis whether it is linear or quadratic. When μ_k differ with k and Σ_k is the same for all k then the linear Discriminant function is adopted. Quadratic Discriminant function is adopted when both μ_k and Σ_k are assumed to vary with k. Let π_k be the prior probability that an arbitrary unit belongs to group k defined mathematically $\pi_k = P\left(G = k\right)$, we are interested in $P\left(G = k / X = x\right)$ which is defined as

$$P\left(G = k / X = x\right) = \frac{P\left(G = k, X = x\right)}{P\left(X = x\right)} = \frac{P\left(X = x / G = k\right) P\left(G = k\right)}{P\left(X = x\right)} = \frac{f_k(x)\pi_k}{\sum_{j=}^{k} f_j(x)\pi_j}$$ --

(6)

$$P(G = k / X = x) = \frac{\pi_k f_k(x)}{\sum_{j=1}^{k} \pi_j f_j(x)}$$ é é é é é é é é é é é é é é é .é é é .. (7)

Equation (7) is the Bayes theorem hence the name Bayes classification rule which is being derived. We will compute $P(G = k / X = x)$ for each class k given the observation vector x the class with the highest value of $P(G = k / X = x)$ is the most likely class for the unit being considered. $P(G = k / X = x)$ is the class posterior probability while

$\pi_k = P(G = k)$
é .(8)

is the prior probability. With the knowledge of the prior probability, the posterior probability can be calculated and used to predict class membership of arbitrary units.

Consider $P(G = k_1 / X = x)$ and $P(G = k_2 / X = x)$; class k_1 is more likely than k_2 if

$$P(G = k_1 / X = x) > P(G = k_2 / X = x)$$

Hence,

$$\frac{P(G = k_1 / X = x)}{P(G = k_2 / X = x)} > 1$$ é é é é é é é é é é é é é é é é ..é é é é é é

(9)

Taking log of both sides and substituting appropriately gives

$$\log\left\{\frac{\dfrac{\pi_{k_1}}{(2\pi)^{\frac{p}{2}}\left|\Sigma_{k_1}\right|^{\frac{1}{2}}}\exp\left\{-\dfrac{1}{2}\left(x-\mu_{k_1}\right)'\Sigma^{-1}\left(x-\mu_{k_1}\right)\right\}}{\dfrac{\pi_{k_2}}{(2\pi)^{\frac{p}{2}}\left|\Sigma_{k_2}\right|^{\frac{1}{2}}}\exp\left\{-\dfrac{1}{2}\left(x-\mu_{k_2}\right)'\Sigma^{-1}\left(x-\mu_{k_2}\right)\right\}}\right\}>0 \text{é é é é é é é é é é .}$$

(10)

Therefore, for the observation vector x, group k_1 is more likely if

$$\Rightarrow -\frac{1}{2}\log\left|\Sigma_{k_1}\right|-\frac{1}{2}\left(x-\mu_{k_1}\right)'\Sigma_{k_1}^{-1}\left(x-\mu_{k_1}\right)+\log\pi_{k_1} > -\frac{1}{2}\log\left|\Sigma_{k_2}\right|-\frac{1}{2}\left(x-\mu_{k_2}\right)'\Sigma_{k_2}^{-1}\left(x-\mu_{k_2}\right)+\log\pi_{k_2}$$

---------------------------------------(11)

Generalizing the quadratic discriminant function, given an observation vector x, group k is more likely if

$$QDF_k(x)=-\frac{1}{2}\log\left|\Sigma_k\right|-\frac{1}{2}\left(x-\mu_k\right)'\Sigma_k^{-1}\left(x-\mu_k\right)+\log\pi_k \quad \text{é é é é é é é é é é é .}$$

(12)

is highest for all groups.

where k=1,2,3,é ,n; n is number of groups.

The expression in equation (12) is the Quadratic Discriminant function.

Consider the expression in equation (11), let $\Sigma_{k_1}=\Sigma_{k_1}=\Sigma$, then k_1 is more likely if

$$\Rightarrow -\frac{1}{2}\left(x-\mu_{k_1}\right)^{'}\Sigma^{-1}\left(x-\mu_{k_1}\right)+\log\pi_{k_1}-\frac{1}{2}\left(x-\mu_{k_2}\right)^{'}\Sigma^{-1}\left(x-\mu_{k_2}\right)+\log\pi_{k_2}>0$$

é é é é é ..(13)

Expanding the expression above and re-arranging gives that class k_1 is more likely if

$$\mu_{k_1}^{'}\Sigma^{-1}x-\frac{1}{2}\mu_{k_1}^{'}\Sigma^{-1}\mu_{k_1}+\log\pi_{k_1}>\mu_{k_2}^{'}\Sigma^{-1}x-\frac{1}{2}\mu_{k_2}^{'}\Sigma^{-1}\mu_{k_2}+\log\pi_{k_2}$$

é é é é é é é .(14)

Generalizing equation (14) as well; given an observation vector x, group k is more likely if

$$LDF_k(x)=\mu_k^{'}\Sigma^{-1}x-\frac{1}{2}\mu_k^{'}\Sigma^{-1}\mu_k+\log\pi_k$$ --

(15)

is highest for all groups

where k=1,2,3,é ,n; n is number of groups.

Equation (9) is the Linear Discriminant function . Obviously, when $X/G=k \sim N_p\left(\mu_k,\Sigma_k\right)$ and $\Sigma_1=\Sigma_2=...=\Sigma_k=\Sigma$, we utilize the Linear Discriminant function and when $X/G=k \sim N_p\left(\mu_k,\Sigma_k\right)$ and Σ_k vary with k then we utilize the Quadratic Discriminant function. Classification rules are the same for both functions and it states as follows: given an observation vector x, the class

k for which the discriminant function of x is highest is the most likely group class of the unit on which x is calculated.

The derivations of the linear and quadratic discriminant functions is based on the Bayes theorem and that is why the classification rules are termed Bayes classification rules, Bickel and Levina (2004). The parameters of the two discriminant functions considered, namely Σ_k, Σ, μ_k and π_k can be estimated and when the estimated parameters are plugged directly into the Bayes classification rules, we have Fisherõs classification rules. See Bickel and Levina (2004) for details.

The Bayesian Quadratic Discriminant Analysis QDA was first proposed by Geisser (1964) and Keehn (1965) and their work has attracted series of subsequent works on the Bayesian QDA as can be seen in Srivastava et al (2007). The function adopted in this work is the Fisherõs Quadratic Discriminant Function which resulted from substitution of estimates of parameters of the Bayes Quadratic Discriminant Function.

3.2 DISTRIBUTIONAL ASSUMPTIONS

Multivariate Normality: Multivariate normality of the eight dimensional random variable formed by the extracted ACF and PACF with mean vector varying with groups is a basic assumption for application of discriminant functions be it linear or quadratic. Though, univariate normality of individual variables that form a multivariate random variable do not imply normality of the multivariate variable but in most cases, when all the univariate normal variables are normally distributed then the resulting multivariate variable is expected to be normally distributed. The correlation between univariate random variables that form the eight dimensional multivariate variable can as well be removed via principal component analysis so that the individual principal components are now uncorrelated and captures most of the variations of the original multivariate random variable from which the principal components were extracted. Hence, normality of the individual random variables and normality of the extracted principal components are overwhelming argument to confirm multivariate normality. We test the same variables for each group and both the principal components and the individual variables were normally distributed (except very few which are negligible), implying that the multivariate variable is normal with mean varying from group to group. The null hypothesis is that test distribution is normal. The decision rule is to reject the null hypothesis if p-value is less than the level of significance. The p-values are presented in the table below:

Original variables

Table 1: Table of p-values for normality tests.

Groups	1	2	3	4	5	6	7	8
1	.001	0.872	0.176	0.348	0.999	0.935	0.871	0.989
2	.648	0.04	0.739	0.443	0.127	0.413	0.905	0.744
3	.000	0.484	0.989	0.826	0.543	0.092	0.939	0.978
4	.398	0.02	0.968	0.817	0.013	0.306	0.560	0.621
5	.083	0.065	0.684	0.228	0.072	0.913	0.994	0.876
6	.000	0.031	0.022	0.133	0.695	0.649	0.192	0.890

Table 2: Table of p-values for principal components

	Principal components			
Groups	1	2	3	4
1	0.959	0.598	0.959	
2	.012	0.353	0.840	
3	0.074	0.989	0.657	0.885
4	.404	0.528	1.000	
5	0.036	0.010		
6	0.171	0.003	0.618	

Most of the values lead to acceptance of the null hypothesis at significance level of 0.01.

Linearity: the choice of quadratic discriminant function was as a result of the failure of the linearity assumption (i.e. assumption of equality of variance-covariance matrix). The test was done using the Box M statistic given as

$$M = (n-g)\log|S| - \sum_{i=1}^{g}(n_i)\log|S_i| \quad \text{é é é é é é é é é é é é é é é é} \quad (16)$$

where S and S_i are pooled and covariance matrices respectively and n, n_i and g are observation size for all groups combined, observation size for group i and number of groups respectively. Hypothesis of equal covariance matrices ($H_0 : \Sigma_1 = \Sigma_2 = ... = \Sigma_k$) was tested. The p-value for the test was 0.00 implying rejection of the null hypothesis hence the choice of quadratic discriminant function.

3.3 DEVELOPMENT OF THE PROPOSED CLASSIFIER

To develop the classifiers, we simulated 300 sets of Time Series data each of length 500 with different combinations of parameters. The following classes of linear stationary models AR(1), AR(2), MA(1), MA(2), ARMA(1,1) and ARMA(2,2) were used and 50 sets of data each employed. Limiting the models to six classes is in line with the statement of Box and Jenkins(1970) which has it that ARMA(p,q) with $p+q \le 2$ are adequate in most cases. Also, to ensure parsimony generally, $p,q \le 2$. For each data set, we calculated the ACF at lag 1-4 and PACF at lag 2-5 since PACF at lag 1 is the same as ACF at lag 1. The ACFs and PACFs calculated now form the multivariate variable used in separating the groups. As was pointed out earlier, equality of covariance matrix assumption failed and so we adopted the Quadratic Discriminant rule. For each group say k, the covariance matrix Σ_k is calculated; the mean vector μ_k for the variables is also calculated. The discriminant function which is developed for each group is now given as follow for group k:

$$Q_k(x) = -\frac{1}{2}\log\left|\Sigma_k\right| - \frac{1}{2}\left(x-\mu_k\right)^T \Sigma^{-1}\left(x-\mu_k\right) + \log\pi_k \text{ é é é é é é é é é é .(17)}$$

k=1,2,3,4,5,6.

where π_k is the prior probability that the individual or item considered belong to group k. The quadratic function is a function of the x vector.

Now, given an arbitrary series which you may not know the class it belongs, calculate the ACF at lag 1-4 and represent the values with x_1, x_2, x_3, x_4, represent the values of PACF at lag 2-5 with x_5, x_6, x_7, x_8 so that the x-vector will now be defined as $x^T = \left(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\right)$ é é é é é é é é é é é é é é é é .(18)

The x-vector is now substituted into each of the quadratic discriminant function and the series from which the x-vector was generated belongs to any group that maximizes the discriminant function. The chosen model is fitted into the series from where the x-was generated. There is the tendency for the right model to be correctly identified in this case due to the high accuracy but if it turns out otherwise, we consider the group with discriminant function value that is closest to the highest and fit the model to that series. The earlier lags where chosen because ACF and PACF in most cases become zero or insignificant for higher lags. Most of the important variations of interest are captured by the earlier lags.

3.4 THE PROPOSED ALGORITHM

Given a Series whose theoretical model is to be identified, summary of the proposed model identification algorithm is presented below:

Step 1: Calculate the ACF at lag 1 to 4 and PACF at lag 2 to 5 and denote the observation vector x as

$$x^T = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) \quad \text{é é é é é é é é é é é é é é é é é é} \quad ..(19)$$

Step 2: input (19) into each of the classifiers and calculate the values for each group.

Step 3: the model/class with the highest value of classifiers is fitted into the series at hand

Step 4: carry out model diagnostic check on the fitted model and stop iteration if the model is adequate, else, go back to step 3 and consider the model with the next highest value of the classifier to the one considered earlier and then repeat step 3 and 4. The iteration is continued until an adequate model is identified.

CHAPTER FOUR

RESULTS

4.1 THE PROPOSED FUNCTIONS

Following the procedure described above, the following classifiers were developed for various models considered (note that prior probability is assumed equal for all groups and as such is 1/6)

FOR:

AR (1): $CL_{AR(1)}(x) = -\frac{1}{2}\log\left|\Sigma_{AR(1)}\right| - \frac{1}{2}\left(\underline{x} - \underline{\mu}_{AR(1)}\right)^T \Sigma_{AR(1)}^{-1}\left(\underline{x} - \underline{\mu}_{AR(1)}\right) + Log\pi_{AR(1)}$

$$\Sigma_{AR(1)} = \begin{pmatrix} 0.6610 & -0.0120 & 0.4490 & -0.0120 & 0.0060 & -0.0140 & 0.0030 & -0.0060 \\ -0.0120 & 0.0280 & -0.0080 & 0.0370 & 0 & 0.0030 & 0.0010 & 0 \\ 0.4490 & -0.0080 & 0.3240 & -0.0080 & 0.0020 & -0.0080 & 0.0020 & -0.0040 \\ -0.0120 & 0.0370 & -0.0080 & 0.0530 & 0 & 0.0040 & 0.0020 & 0 \\ 0.0060 & 0 & 0.0020 & 0 & 0.0020 & 0 & 0 & 0 \\ -0.0140 & 0.0030 & -0.0080 & 0.0040 & 0 & 0.0030 & 0 & 0 \\ 0.0030 & 0.0010 & 0.0020 & 0.0020 & 0 & 0 & 0.0020 & 0 \\ -0.0060 & 0 & -0.0040 & 0 & 0 & 0 & 0 & 0.0020 \end{pmatrix}$$

$$\Sigma_{AR(1)}^{-1} = \begin{pmatrix} 33.4184 & 1.5769 & -44.9030 & -3.1327 & -55.3523 & 38.8114 & -2.8804 & 10.4493 \\ 1.5769 & 483.34 & 0.3206 & -338.83 & -5.0513 & -23.3637 & 94.4670 & 5.3719 \\ -44.9030 & 0.3206 & 63.7871 & 2.4495 & 70.9218 & -43.0349 & 0.9576 & -7.1347 \\ -3.1327 & -338.83 & 2.4495 & 259.54 & 6.9486 & -15.3134 & -87.8764 & -4.4991 \\ -55.3523 & -5.0513 & 70.9218 & 6.9486 & 595.135 & -73.3993 & 7.6837 & -24.2132 \\ 38.8114 & -23.37 & -43.0349 & -15.3134 & -73.399 & 443.4751 & 11.8131 & 30.3644 \\ -2.8804 & 94.4670 & 0.9576 & -87.876 & 7.6837 & 11.8131 & 544.01 & -6.7261 \\ 10.4493 & 5.3719 & -7.1347 & -4.4991 & -24.2132 & 30.3644 & -6.7261 & 517.0785 \end{pmatrix}$$

$$\left|\Sigma_{AR(1)}\right| = 1.9725 \times 10^{-17}$$

$$Log\left|\Sigma_{AR(1)}\right| = -16.7050$$

$$\underline{\mu}_{AR(1)} = \begin{pmatrix} -.012281190 \\ .646701146 \\ -.016467286 \\ .435036074 \\ -.007941516 \\ .008384162 \\ .005130720 \\ -.018072549 \end{pmatrix}$$

AR(2): $CL_{AR(2)}(x) = -\frac{1}{2}\log\left|\Sigma_{AR(2)}\right| - \frac{1}{2}\left(\underline{x} - \underline{\mu}_{AR(2)}\right)^T \Sigma_{AR(2)}^{-1}\left(\underline{x} - \underline{\mu}_{AR(2)}\right) + Log\pi_{AR(2)}$

$$\sum_{AR(2)} = \begin{pmatrix} 0.4250 & 0.0180 & 0.3080 & 0.0030 & 0.0070 & -0.0140 & -0.0020 & 0 \\ 0.0180 & 0.2890 & 0.0050 & 0.1660 & 0.2090 & -0.0020 & 0 & 0.0010 \\ 0.3080 & 0.0050 & 0.3180 & 0 & -0.0040 & -0.0140 & 0.0030 & 0.0010 \\ 0.0030 & 0.1660 & 0 & 0.1340 & 0.1160 & 0.0010 & 0.0010 & -0.0010 \\ 0.0070 & 0.2090 & -0.0040 & 0.1160 & 0.1720 & -0.0030 & 0.0010 & 0.0010 \\ -0.0140 & -0.0020 & -0.0140 & 0.0010 & -0.0030 & 0.0030 & 0 & 0 \\ -0.0020 & 0 & 0.0030 & 0.0010 & 0.0010 & 0 & 0.0030 & 0 \\ 0 & 0.0010 & 0.0010 & -0.0010 & 0.0010 & 0 & 0 & 0.0020 \end{pmatrix}$$

$$\sum_{AR(2)}^{-1} = \begin{pmatrix} 8.5738 & 0.1271 & -8.5355 & 0.7674 & -1.3576 & -1.3502 & 14.4481 & 5.2667 \\ 0.1271 & 44.4594 & -2.1644 & -19.5068 & -41.2491 & -14.6145 & 22.5011 & -10.2764 \\ -8.5355 & -2.1644 & 2.7446 & -1.3095 & 4.7265 & 23.3622 & -19.5739 & -8.3081 \\ 0.7674 & -19.5068 & -1.3095 & 28.6684 & 3.8563 & -21.2344 & -9.0205 & 22.8142 \\ -1.3576 & -41.2491 & 4.7265 & 3.8563 & 54.4084 & 41.3451 & -25.0531 & -7.0148 \\ -1.3502 & -14.6145 & 23.3622 & -21.2344 & 41.3451 & 474.7363 & -30.9659 & -35.6636 \\ 14.4481 & 22.5011 & -19.5739 & -9.0205 & -25.0531 & -30.9659 & 373.8972 & 6.5527 \\ 5.2667 & -10.2764 & -8.3081 & 22.8142 & -7.0148 & -35.6636 & 6.5527 & 524.2067 \end{pmatrix}$$

$$\left|\sum_{AR(2)}\right| = 9.7888\text{e-}014$$

$$Log\left|\sum_{AR(2)}\right| = -13.0093$$

$$\underline{\mu}_{AR(1)} = \begin{pmatrix} .031768954 \\ .348790205 \\ .033505763 \\ .389282118 \\ .055346368 \\ -.006051303 \\ -.002800483 \\ -.007268034 \end{pmatrix}$$

MA (1): $CL_{MA(1)}(x) = -\frac{1}{2}\log\left|\Sigma_{MA(1)}\right| - \frac{1}{2}\left(\underline{x} - \underline{\mu}_{MA(1)}\right)^T \Sigma_{MA(1)}^{-1}\left(\underline{x} - \underline{\mu}_{MA(1)}\right) + Log\pi_{MA(1)}$

$$\Sigma_{MA(1)} = \begin{pmatrix} 0.2370 & -0.0030 & 0.0020 & -0.0040 & -0.0020 & 0.0920 & -0.0030 & 0.0420 \\ -0.0030 & 0.0040 & -0.0000 & 0 & 0.0020 & 0 & 0.0010 & 0 \\ 0.0020 & -0.0000 & 0.0020 & 0 & 0 & 0.0010 & 0 & 0 \\ -0.0040 & 0 & 0 & 0.0020 & 0 & -0.0020 & 0.0010 & -0.0010 \\ -0.0020 & 0.0020 & 0 & 0 & 0.0040 & -0.0010 & 0.0010 & 0 \\ 0.0920 & 0 & 0.0010 & -0.0020 & -0.0010 & 0.0380 & -0.0010 & 0.0180 \\ -0.0030 & 0.0010 & 0 & 0.0010 & 0.0010 & -0.0010 & 0.0030 & 0 \\ 0.0420 & 0 & 0 & -0.0010 & 0 & 0.0180 & 0 & 0.0100 \end{pmatrix}$$

$$\Sigma_{MA(1)}^{-1} = \begin{pmatrix} 139.7 & 170.0 & 88.3 & -59.4 & -126.5 & -450.4 & -5.2 & 218.2 \\ 170.0 & 565.0 & 138.9 & -27.5 & -321.9 & -599.3 & -101.6 & 362.0 \\ 88.3 & 138.9 & 612.5 & -28.5 & -113.9 & -397.1 & -42.9 & 340.9 \\ -59.4 & -27.5 & -28.5 & 682.2 & 90.1 & 174.9 & -249.4 & 2.7 \\ -126.5 & -321.9 & -113.9 & 90.1 & 478.2 & 470.1 & -51.9 & -305.9 \\ -450.4 & -599.3 & -397.1 & 174.9 & 470.1 & 1675.4 & 92.8 & -1106.4 \\ -5.2 & -101.6 & -42.9 & -249.4 & -51.9 & 92.8 & 493.4 & -170.1 \\ 218.2 & 362.0 & 340.9 & 2.7 & -305.9 & -1106.4 & -170.1 & 1175.1 \end{pmatrix}$$

$$\left|\Sigma_{MA(1)}\right| = 3.1373 \times 10^{-20}$$

$$Log\left|\Sigma_{MA(1)}\right| = -19.5034$$

$$\underline{\mu}_{MA(1)} = \begin{pmatrix} -.021299086 \\ .015464797 \\ -.006779423 \\ -.007782054 \\ -.283970606 \\ -.009652009 \\ -.143177206 \\ -.014282182 \end{pmatrix}$$

MA (2): $CL_{MA(2)}(x) = -\frac{1}{2}\log\left|\Sigma_{MA(2)}\right| - \frac{1}{2}\left(\underline{x} - \underline{\mu}_{MA(2)}\right)^T \Sigma_{MA(2)}^{-1}\left(\underline{x} - \underline{\mu}_{MA(2)}\right) + Log\pi_{MA(2)}$

$$\Sigma_{MA(2)} = \begin{pmatrix} 0.0960 & 0.0120 & 0 & -0.0030 & 0.0070 & -0.0040 & 0.0050 & 0.0240 \\ 0.0120 & 0.1020 & -0.0060 & 0.0020 & 0.0830 & -0.0020 & 0.0260 & 0.0070 \\ 0 & -0.0060 & 0.0040 & -0.0001 & -0.0050 & 0.0010 & -0.0010 & 0 \\ -0.0030 & 0.0020 & -0.0001 & 0.0020 & 0.0020 & -0.0000 & 0.0020 & 0.0001 \\ 0.0070 & 0.0830 & -0.0050 & 0.0020 & 0.0760 & -0.0020 & 0.0160 & 0.0060 \\ -0.0040 & -0.0020 & 0.0010 & -0.0000 & -0.0020 & 0.0320 & 0 & 0.0030 \\ 0.0050 & 0.0260 & -0.0010 & 0.0020 & 0.0160 & 0 & 0.0150 & 0.0030 \\ 0.0240 & 0.0070 & 0 & 0.0001 & 0.0060 & 0.0030 & 0.0030 & 0.0140 \end{pmatrix}$$

$$\Sigma_{MA(2)}^{-1} = \begin{pmatrix} 21.6843 & -4.2595 & -1.1318 & 35.9120 & 5.4291 & 6.4560 & -2.8256 & -38.3707 \\ -4.2595 & 256.8376 & 36.8477 & 193.8485 & -239.2432 & -2.9424 & -217.4509 & 27.4414 \\ -1.1318 & 36.8477 & 280.094 & 19.7933 & -15.9532 & -7.4040 & -30.0948 & -1.7338 \\ 35.9120 & 193.8485 & 19.7933 & 796.728 & -177.7498 & 8.9662 & -256.4422 & -34.2177 \\ 5.4291 & -239.2432 & -15.9532 & -177.750 & 241.3898 & 4.1586 & 184.5262 & -32.4687 \\ 6.4560 & -2.9424 & -7.4040 & 8.9662 & 4.1586 & 34.1442 & 0.5991 & -18.8791 \\ -2.8256 & -217.4509 & -30.0948 & -256.442 & 184.5262 & 0.5991 & 284.9018 & -25.1024 \\ -38.371 & 27.4414 & -1.7338 & -34.2177 & -32.4687 & -18.8791 & -25.1024 & 147.0381 \end{pmatrix}$$

$$\left|\Sigma_{MA(2)}\right| = 4.1413\times10^{-16}$$

$$Log\left|\Sigma_{MA(2)}\right| = -15.3829$$

$$\underline{\mu}_{MA(2)} = \begin{pmatrix} -.017751319 \\ -.097241212 \\ .003128591 \\ .015433840 \\ -.199210898 \\ -.031544491 \\ -.118628287 \\ -.030528789 \end{pmatrix}$$

ARMA (11):

$$CL_{ARMA(11)}(x) = -\frac{1}{2}\log\left|\Sigma_{ARMA(11)}\right| - \frac{1}{2}\left(\underline{x} - \underline{\mu}_{ARMA(11)}\right)^T \Sigma^{-1}_{ARMA(11)}\left(\underline{x} - \underline{\mu}_{ARMA(11)}\right) + Log\pi_{ARMA(11)}$$

$$\Sigma_{ARMA(11)} = \begin{pmatrix} 0.2520 & 0.0130 & 0.1110 & 0.0110 & 0.0030 & 0.0650 & 0.0060 & 0.0240 \\ 0.0130 & 0.0690 & 0.0100 & 0.0280 & -0.0370 & 0.0030 & -0.0150 & 0.0030 \\ 0.1110 & 0.0100 & 0.0530 & 0.0090 & 0.0010 & 0.0310 & 0.0030 & 0.0110 \\ 0.0110 & 0.0280 & 0.0090 & 0.0160 & -0.0130 & 0.0030 & -0.0030 & 0.0020 \\ 0.0030 & -0.0370 & 0.0010 & -0.0130 & 0.0490 & -0.0030 & 0.0200 & -0.0020 \\ 0.0650 & 0.0030 & 0.0310 & 0.0030 & -0.0030 & 0.0240 & 0 & 0.0090 \\ 0.0060 & -0.0150 & 0.0030 & -0.0030 & 0.0200 & 0 & 0.0110 & 0 \\ 0.0240 & 0.0030 & 0.0110 & 0.0020 & -0.0020 & 0.0090 & 0 & 0.0060 \end{pmatrix}$$

$$\Sigma^{-1}_{ARMA(11)} \begin{pmatrix} 99.0589 & -42.3671 & -252.6208 & 164.1043 & 46.0370 & 69.2136 & -81.8569 & -55.0901 \\ -42.3671 & 116.8153 & 103.2202 & -234.7880 & -42.6411 & -8.8402 & 167.7482 & -0.8668 \\ -252.621 & 103.2202 & 757.2781 & -471.3182 & -143.8348 & -333.3047 & 204.9940 & 179.648 \\ 164.1043 & -234.7880 & -471.3182 & 673.8969 & 183.8204 & 164.0863 & -431.5644 & -84.4282 \\ 46.0370 & -42.6411 & -143.8348 & 183.8204 & 148.2754 & 65.3874 & -263.4892 & -9.0601 \\ 69.2136 & -8.8402 & -333.3047 & 164.0863 & 65.3874 & 363.0916 & -33.0419 & -238.9129 \\ -81.857 & 167.748 & 204.9940 & -431.5644 & -263.4892 & -33.0419 & 669.7703 & -26.6808 \\ -55.0901 & -0.8668 & 179.6481 & -84.4282 & -9.0601 & -238.9129 & -26.6808 & 441.5976 \end{pmatrix}$$

$$\left|\Sigma_{ARMA(11)}\right| = 7.9440\text{e-}017$$

$$Log\left|\Sigma_{ARIMA(11)}\right| = 16.0999$$

$$\underline{\mu}_{ARMA(11)} = \begin{pmatrix} .032453312 \\ .211720759 \\ .017656330 \\ .106578730 \\ -.117280748 \\ .002223160 \\ -.035846598 \\ -.015933867 \end{pmatrix}$$

ARMA (22):

$$CL_{ARMA(22)}(x) = -\frac{1}{2}\log\left|\Sigma_{ARMA(22)}\right| - \frac{1}{2}\left(\underline{x} - \underline{\mu}_{ARMA(22)}\right)^T \Sigma^{-1}_{ARMA(22)}\left(\underline{x} - \underline{\mu}_{ARMA(22)}\right) + Log\pi_{ARMA(22)}$$

$$\Sigma_{ARMA(22)} = \begin{pmatrix} 0.6920 & -0.0250 & 0.5860 & -0.0320 & 0.1410 & -0.0590 & -0.0640 & 0.0060 \\ -0.0250 & 0.1080 & 0.0190 & 0.0900 & 0.0760 & 0.0130 & -0.0010 & 0 \\ 0.5860 & 0.0190 & 0.5610 & 0.0020 & 0.1660 & -0.0340 & -0.0630 & -0.0070 \\ -0.0320 & 0.0900 & 0.0020 & 0.1170 & 0.0620 & 0.0330 & -0.0020 & -0.0060 \\ 0.1410 & 0.0760 & 0.1660 & 0.0620 & 0.1750 & 0.0140 & -0.0280 & -0.0130 \\ -0.0590 & 0.0130 & -0.0340 & 0.0330 & 0.0140 & 0.0740 & 0.0140 & -0.0130 \\ -0.0640 & -0.0010 & -0.0630 & -0.0020 & -0.0280 & 0.0140 & 0.0260 & 0.0000 \\ 0.0060 & 0 & -0.0070 & -0.0060 & -0.0130 & -0.0130 & 0.0000 & 0.0170 \end{pmatrix}$$

$$\Sigma^{-1}_{ARIMA(22)} = \begin{pmatrix} 23.4655 & 19.4542 & -24.9055 & -9.1122 & -4.3313 & 7.4507 & -11.2149 & -19.3670 \\ 19.4542 & 55.3299 & -19.5453 & -34.3781 & -16.4179 & 15.2307 & -25.87 & -27.9537 \\ -24.9055 & -19.5453 & 29.5919 & 10.2612 & 1.4580 & -7.1137 & 15.8349 & 20.2706 \\ -9.1122 & -34.3781 & 10.2612 & 34.4633 & 5.0366 & -13.6820 & 16.5531 & 12.9925 \\ -4.3313 & -16.4179 & 1.4580 & 5.0366 & 16.7419 & -5.7522 & 13.7536 & 12.3096 \\ 7.4507 & 15.2307 & -7.1137 & -13.6820 & -5.7522 & 25.0773 & -19.0615 & 4.3915 \\ -11.2149 & -25.8696 & 15.8349 & 16.5531 & 13.7536 & -19.0615 & 74.5780 & 12.2565 \\ -19.3670 & -27.9537 & 20.2706 & 12.9925 & 12.3096 & 4.3915 & 12.2565 & 91.3618 \end{pmatrix}$$

$$\left|\Sigma_{ARIMA(22)}\right| = 9.3001 \times 10^{-11}$$

$$Log\left|\Sigma_{ARIMA(22)}\right| - 10.0315$$

$$\underline{\mu}_{ARIMA(22)} = \begin{pmatrix} .120436701 \\ .757347586 \\ .079494417 \\ .646596927 \\ .191202580 \\ .096024274 \\ .000749487 \\ -.039129856 \end{pmatrix}$$

4.2 APPLICATION OF THE PROPOSED FUNCTIONS TO SIMULATED TIME SERIES

25 Time Series were simulated and the model was applied in classifying them, summary of the results are presented in the following table.

Table 3: Table classification results

Class of series	$CL_{AR(1)}(x)$	$CL_{AR(2)}(x)$	$CL_{MA(1)}(x$	$CL_{MA(2)}(x)$	$CL_{ARMA(11)}(x)$	$CL_{ARMA(22)}(x)$	Predicted class
AR(1)	5.0087	0.1922	-607.28	-425.32	-27.224	2.5231	AR(1)
AR(2)	-106.8441	2.9465	-907.18	-565.400	-106.99	2.1999	AR(2)
MA(1)	-80.2985	-20.8149	5.5705	-3.8411	-24.0097	-7.2519	MA(1)
MA(2)	-159.2052	-9.0239	-12.2927	5.8699	-23.8468	-10.2926	MA(2)
ARMA(11)	-42.2467	-0.3277	-26.0036	-0.5259	4.3128	-1.4799	ARMA(11)
ARMA(22)	-205.9732	-41.3638	-698.226	-526.475	-191.1478	2.8252	ARMA(22)
AR(1)	6.4989	2.9275	-203.215	-120.68	0.9590	3.095	AR(1)
AR(2)	-39.725	2.7820	-876.205	-575.276	-81.9364	2.5875	AR(2)
MA(1)	-71.3875	-2.9054	5.6523	1.7510	-6.7183	0.7148	MA(1)
MA(2)	-68.9480	2.2212	-3.7370	4.7256	-1.5273	0.1549	MA(2)
ARMA(11)	-25.2539	0.5932	-33.8313	1.5088	5.6942	-0.5869	ARMA(11)
ARMA(22)	-186.4571	-30.7272	-782.687	-590.226	-226.2335	2.1433*	ARMA(22)
AR(1)	3.8598	1.4694	-258.450	-156.5740	-1.1626	2.9018	AR(1)
AR(1)	5.9085	0.9330	-574.995	-423.7169	-24.8434	2.7109	AR(1)
AR(1)	-2.6969	-4.7041	-1119.80	-837.8608	-79.6732	0.5977	ARMA(11)*
AR(1)	0.6320	-4.8443	-1039.50	-803.6329	-69.9680	0.3464	AR(1)
AR(2)	-224.9539	-19.5230	-266.207	-97.8245	-116.0428	-21.4678	AR(2)
AR(2)	-271.4803	-25.9775	-335.736	-125.3184	-158.6506	-21.3591	ARMA(22)*
AR(2)	-235.2235	-17.6035	-239.314	-100.2809	-107.6462	-20.8871	AR(2)
MA(1)	-37.1714	-3.3674	6.0970	1.6311	-8.4713	-5.3951	MA(1)
MA(1)	-36.4434	-3.3674	6.0970	3.7186	-26.8132	-9.0093	MA(1)
MA(1)	-773.8833	-7.6637	5.5542	0.3233	-9.1909	-5.4984	MA(1)
MA(1)	-63.1097	-8.8352	6.9824	1.9684	-17.8298	-6.5374	MA(1)
ARMA(11)	-24.9408	3.8419	-16.5037	1.5999	6.6926	2.9488	ARMA(11)
ARMA(22)	-17.1743	-21.0214	-1338.90	-897.5448	-98.2088	-2.1964	ARM(22)

* Cases of Misclassification

4.3 APPLICATION OF OUR METHOD TO REAL LIFE SERIES

Consider series A (Chemical Process concentration readings: every two hours) which was modeled as ARMA (1, 1) by Box et al (2008). We now apply the model identification Algorithm to this series as follows:

Step 1: Calculate the observation vector as

$$x = (0.483,\ 0.418,\ 0.323,\ 0.296,\ 0.241,\ 0.073,\ 0.081,\ 0.066)$$

Step 2: Input the observation vector x into each of the classifiers to give the resulting values as:

-17.5773, 0.1016, -99.5515, -43.5628, 0.3762, 1.6954 for AR(1), AR(2), MA(1), MA(2), ARMA(1,1) and ARMA(2,2) respectively. From the above result, ARMA (2, 2) gave the highest value of the classifier.

Step 3: ARMA (2, 2) is fitted into the series.

Step 4: the goodness of fit test of course fails see Box et al (2008). We therefore go back to step 2

Step 2: we now consider ARMA (1, 1) whose classifier value is closest to that of ARMA (2, 2)

Step 3: ARMA (1, 1) is fitted into the series.

Step 4: model diagnostic check is carried out which is now positive, see Box et al (2008). We now adopt ARMA (1, 1) and stop iteration.

The Algorithm in this case terminated at the second iteration. The choice of this series (series A) in this illustration is aimed at exposing a case with more than one iteration in the application of this Algorithm and also to see how our Algorithm addressed the problem of ARMA (p, q) model identification when $p, q \neq 0$.

4.4 BRIEF COMPARISON OF OUR METHOD WITH OTHER METHODS

The discriminant values are highly comparable with the conventional information criteria in terms of behaviour, while the information criteria choose the order that minimizes the error term; our discriminant values choose the order that maximizes the probability that the order is the rightful order. According to the comparative work done by Zazli (2002), the following information criteria were compared and their percentage of correct classification are given on the average as: FPE 73%, 73%, BIC 95%, Schwartz 95%, HQ 90%. Our functions classified 23 correctly out of 25 which implies 92% correct classification. While the information criteria select the best model after fitting several models, our function select the rightful model before fitting, this is it major strength over the use of the conventional information criteria.

CHAPTER FIVE

SUMMARY AND CONCLUSION

5.1 SUMMARY

In this work, we simulated 300 sets of Time Series data, 50 each for AR(1), AR(2), MA(1), MA(2), ARMA(1,1) and ARMA(2,2). From each of the 300 series, we calculated ACF for lag 1-4 and PACF for lag 2-5. We combined the calculated ACF and PACF which now form an eight dimensional multivariate random variable with each group now having 50*8 data points. We now used the eight dimensional multivariate random variables to build quadratic discriminant function for each of the groups; we chose to call those discriminant functions ñclassifiersò. We simulated another 25 time series for which we calculated ACF for lag 1-4 and PACF for lag 2-5, the values calculated constitute elements of observation vectors. We then impute the observation vector calculated say from the first of the 25 simulated Time Series into each of the classifiers and the classification rule is to classify the series from which the vector was calculated into the group that maximize the value of the classifiers. This was done for each of the 25 Time Series and the classification result is presented in the table of classification results. We also developed an algorithm to be used alongside our classifiers so that with our algorithm, misclassification at the initial stage merely imply that one will need more than a single iteration to correctly identify an appropriate model. We also went ahead to apply the method to a real life Time Series that has already been fitted by Box et al (2008) and the right model was identify in two iterations.

5.2 DISCUSSION OF RESULTS

Box and Jenkins (1976) have the credit for the most popular judgmental model identification method. Their approach adopts theoretical models to be fitted into Time Series based on careful examination of the sample ACF and PACF.

Behaviour of sample ACF and PACF calculated from the Series to be fitted are compared with that of available theoretical models and the model with similar behavior is suspected to have generated the Series to be fitted and therefore considered. Theoretically, the ACF and PACF of an AR (p) tail off and cut off at lag p respectively. ACF and PACF of MA (q) cut off at lag q and tails off respectively while both ACF and PACF of ARMA (p, q) tail off. This approach has come under serious criticism based on a couple of arguments. First is that the method is highly judgmental because the behaviour of the sample ACF and PACF cannot be perfectly matched with the theoretical behaviour in most cases, therefore deciding whether ACF or PACF cut off at certain lag depend on individual Judgment. Box and Jenkins approach is merely tentative, the model identified at this stage is subject to series of modifications and the models selected at the end are usually different from the initial model. This will obviously lead to fitting and re-fitting of several models and series of goodness of fit tests. It is based on this argument and others that Anderson (1975) concluded that Box and Jenkins approach, though attractive but in practice both time consuming, difficult and relatively expensive computationally. One of the shortfalls of the Box and Jenkins approach is specifying the value of p and q in ARMA (p, q) model where $p,q \neq 0$ since both ACF and PACF tail off in this case.

This new approach has used 300 simulated ARMA models to build the classifiers, one for each class of classes of models. The validation of the classifiers was done using 25 simulated ARMA models. We extracted the ACF for lag 1- 4, PACF for lag 2-5 to form the multivariate variables used in building the quadratic discriminant function (the classifiers). The usefulness of our classifiers does not stop at misclassification, in cases of misclassification as we observed, the model with value of classifier closest to the model wrongly suggested by the classifiers is adopted, and in the two cases we observed, such models were actually the right models. The algorithm for the utilization of this new method is given along-side our classifiers. The two cases of misclassifications observed in our table merely imply that you will need a second iteration to correctly identify such models; only one iteration will do justice to model identification problem in the remaining 23 simulated series. We also went ahead to apply our algorithm to real life Time Series which was correctly identified after two iterations. Our approach addressed the challenge posed by uncertainties associated with behaviour of ACF and PACF in Box and Jenkins approach and it also helps us to identify the right model with high level of uncertainties even before fitting the model unlike information criteria approach that identify the right model after fitting several models.

5.3 CONTRIBUTION

Box et al (2008) specifically said this:

> *…it should also be explained that identification is necessary* ***inexact*** *because the question of what type of model exist in practice in what circumstances, is a property of the behaviour of the physical world and therefore cannot be decided by purely mathematical argument. Furthermore, because at the identification stage,* ***no precise formulation of the problem is available****, statistically* ***"inefficient****" methods must necessarily be used. It is a stage at which graphical methods are particularly useful and j****udgments*** *must be exercised. However it should be borne in mind that preliminary identification commits us to nothing except* ***tentative*** *consideration of a class of models that will later be efficiently fitted and checked*

The nature of challenges of model identification methods presented in the above statement of Box et al (2008) motivated this work. This is the first work ever that considered application of the powerful multivariate analytical tool-discriminant function to ARMA model identification. This is as well the first exact model identification approach which is able to specifically suggest a particular model at the initial stage and also provides the next available course of action in cases where the initially selected model turn out to be the wrong model. Our approach answers most of the research questions in the above paragraph.

Contributions of this work to Time Series model identification are summarised below:

- Our approach has reduced the negative impact of individual judgments from Time Series ARMA model identification.
- There is now a precise formulation of problem in Time Series ARMA model identification.
- Computational cost associated with Time Series modeling due to inexact model identification approach has been greatly reduced

REFERNCES

Akaike, H. (1969). fitting Autoregressive Models for Prediction. *Annals Institute of Statistical Mathematics*, 21: 243-247.

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

Akaike, H. (1979). A Bayesian Extension of the Minimum AIC procedure of

Autoregressive Model fitting. *Biometrica*, 66:237-242

Akaike, H. (1980). Likelihood and the Bayes procedure. *in Bernardo, J. M.; et al., Bayesian Statistics,* Valencia: University Press, pp. 143ï 166

Anderson, D. R. (2008). *Model Based Inference in the Life Sciences*, Springer

Anderson, O.D.(1975). Distinguishing between simple Box and Jenkins models. *International Journal of math. Edu. In Sc. and Tech*, 6(4): 461-465

Bhansali, R.J. (1993). Order Selection for Time Series Models: A Review, in development in Time Series Analysis by Rao, T.S. *Chapman and Hall, London*.

Bickel, P.J. and Levina, E. (2004). some theory for Fisherôs Linear Discriminant Function ñNaïve Bayesò and some alternatives when there are many more variables than observations. *Bernoulli,* 10(6): 989-1010

Box, G.E.P and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day

Box, G.E.P. and Jenkins, G.M.(1970). *Time Series Analysis: Forecasting and Control.* Holden-Day

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C.(2008). *Time Series Analysis: forecasting and control*. New Jersey: Wiley.

Burnham, K. P.; Anderson, D. R. (2002), *Model Selection and Multimodel Inference: A.Practical Information-Theoretic Approach* (2nd ed.), *Springer-Verlag, ISBN 0-387-95364-7*

Burnham, K. P.; Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in Model Selection. *Sociological Methods and Research,* 33: 261ï 304

Chan, W. S. (1999). A comparison of some pattern identification methods for order Determination of mixed ARMA models. *Statistics and Probability Letters*, 42: 69ï 79.

Clark, A.E. and Troskie, C.G. (2008). Time Series Model Selection. *Communication in Statistics-Simulation and Computation,* 37:766-771.

Cryer, J.D. and Kung-Sik, C. (2008) *Time Series Analysis with application in R.* Springer. Science+Business Media, LLC, 2nd edition

Geisser, S. (1964) Posterior odds for multivariate Normal Distribution. *Journal of the Royal Statistical Society Series B Methodological*, 26:69-76

Hannan, E.J. and Quinn, B.G. (1979). The Determination of the order of an Autoregressive. *Journal of the Royal Statistical Society, Series B*, 41:190-195.

Jenkins, G.M. and Watts, D.G.(1968). *Spectral Analysis and its Application*, Holden-Day: San Francisco

Keehn, D.G. (1965). A Note on Learning for Gaussian Properties. *IEEE Trans. On Information Theory,* 11:126-132

Koreisha, S. and Pukkila, T. (1993). Using the Residual White Noise Autoregressive Order-Determination criteria to Identify Unit Roots in ARIMA Models. Working paper, University of Oregon Graduate School of Management.

Parzen, E. (1974). Some Recent Advances in Time Series Modeling. *IEEE Transaction on Automatic Control*, AC-19: 723-730.

Pukkila, T., Koreisha, S., and Kallinen, A. (1990). The Identification of ARIMA Models. *Biometrika,* 77, 537-548.

Schwartz, G. (1978). Estimating the Dimension of a Model. *Annal of Statistics*, 6:461-464

Shibata, R. (1979). Selection of the order of an Autoregressive Model by Akaikeôs Information Criterionö. *Biometrica*, 63: 117-126.

Srivastava, S., Gupta, M.R. and Frigyik, B.A (2007). Bayesian Quadratic Discriminant Analysis. *Journal of Machine Learning Research* 8:1277-1305.

Tsay, R. and Tiao, G. (1985). Use of canonical analysis in time series model identification. *Biometrika*, 72, 299ï315.

Tsay, R. S. and Tiao, G. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMAModels. *Journal of the American Statistical Association*, 79, 385, 84ï96.

Whittle, P. (1963*). Prediction and Regulation.* London University Press

Zazli, C. (2002). Performance of order Selection Criterion for Short Time Series. *PakistanJournal of Applied Sciences, 2(7): 783-788.*