KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

KUMASI

COLLEGE OF SCIENCE

FACULTY OF PHYSICAL SCIENCE

INTERRUPTED TIME SERIES ANALYSIS OF THE RATE OF INFLATION IN GHANA

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS IN PARTIAL FULFILLMENT OF THE REQUIRMENTS FOR THE AWARD OF MASTER OF SCIENCE DEGREE IN MATHEMATICS

BY

MUHAMMED HUDU

May, 2009

DECLARATION

I hereby declare that this submission is my own work towards the award of the MSc. (Mathematics) 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.

……………………………   ………………………………..

MUHAMMED HUDU   Date

……………………………   …………………………………

Dr. F. T. ODURO   Date

(SUPERVISOR)

……………………………   …………………………………

Dr. S. K. Amponsah   Date

(HEAD OF DEPARTMENT)

DEDICATION

I dedicate this work to the most high God who in his own diverse ways has brought me this far and also to my family, friends and loved ones.

ACKNOWLEDGEMENT

My sincere appreciations first and foremost go to Almighty God who has seen us through this thesis.

To my supervisor Mr. Dominic Nsiah for his useful guidance and constructive criticisms, which enabled me, produce this thesis.

My appreciations to my external examiner Prof. N.N.N. Nsowah-Nuamah (Deputy Government Statistician-Ghana) for his patience and guidance.

I will like to thank Dr. F. T. Oduro for supervising me after my supervisor passed away. Not forgetting Dr. S. K. Amponsah and Mr. S. K. Appiah, the entire department of Mathematics, for their encouragements and advice. Also to the Bank of Ghana for offering me with the relevant data for my Analysis.

Also to Mr. Kumapley Joachim Cupid and Mr. Michael M. Ofori all formally of the Department of Mathematics-KNUST for making themselves available throughout the writing of this thesis and making it a success, to them I say thank you.

This thesis wouldn’t be complete if not for some few hands, i.e. Mr. David Kombat(Census Co-ordinator), Mr. Stephen Amoah, Mr. Michael Ofori, Mr. Samuel Donkoh all of Ghana Statistical Service, to them I say God bless you for your effort.

Finally to my families and all others who have contributed in one way or the other to make the writing of this thesis possible.

ABSTRACT

In 2001, the government of Ghana instituted a set of economic policies aimed at reducing the rate of inflation.

In this study, I attempt to assess the effectiveness of the intervention. Using the statistical hypothesis testing model, this compares the difference in the mean inflationary rates from January **1996** to January **2001** and February **2001** to December **2006**.

Using Box-Jenkins method, we obtained Autoregressive (*AR(1)*) models of the pre-intervention data, post intervention data and the entire interrupted time series data.

Further, using interrupted time series analysis, we are able to show that the parameters before and after the interventions were significantly different at the **5%** level of significance.

In particular, the intercepts of linear trend lines were **1.230** and **3.097** before and after intervention respectively. The slopes were **-0.004** and **-0.036** before and after intervention respectively.

The method used in this study can be used to investigate the effectiveness of any intervention at controlling some time dependent natural or socio-economic phenomena.

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

1.0 BACKGROUND TO THE STUDY

Inflation is Ghana's demon. It is the evil that has dogged all aspects of national life for long. And like the devil, it is a tempting lot. Falling for its bait brings ephemeral gains. But like all fiends, it can wreck havoc.

Inflation causes global concerns because it can distort economic patterns. Inflation can result in the redistribution of wealth when not anticipated. For instance, inflation tends to benefit borrowers at the expense of lenders whenever rates are underestimated over the life of a loan. Inflation has a corrosive effect on savings. As prices surge, the value of savings will decline if the rate of inflation exceeds the rate of interest. People on fixed incomes are hurt by inflation. Workers who retired on fixed incomes are often hurt because of the declining value of their monthly cheques as a result of rising inflation. Thus inflation is a worry for the health of the world's economy and sustenance of the global financial system.

While inflationary pressures or the rise in the general cost of goods affects all nations, it is even more invidious in developing nations such as Ghana. It is an economic fact of Ghanaian life. But more volatile are its political dimensions.

Evidently, inflation in Ghana is caused by both fiscal and non-monetary issues. In the past, Ghana's balance of payments position has been in severe difficulties due to inappropriate trade, fiscal and monetary policies. Excessive money supply is the single most pervasive cause of inflation in Ghana. For instance between 1996 and 1997 inflation was at 25% and 8% respectively, meaning the then PNDC government adopted some prudent economic management tools to tame the demon. But it lost the race when the rate jumped to 40% at some point, again reflecting fiscal mismanagement.

Monetary factors play an important role in the failure to control inflation in Ghana. The devaluation of the cedi at diverse times especially during the Economic Reforms Programme (ERP) led to price raises. The ERP's attempt to promote production in the short term resulted in higher debts and inflation. Ghana's debt profile increased from 4 billion US dollars in 1992 to 4.3 billion dollars by the next year, because of heavy borrowing and excessive printing of money to service domestic and multilateral debts.

A non-monetary source of high inflation is attributable to the poor performance of the agriculture sector. Between 1995 and 1999, Ghana's agriculture sector grew by 4.4 %, but it dropped again to 1.1 % by 2000. This resulted in high food prices in the country. Food prices alone accounts for over half of the average household expenditure in Ghana. The other dimension to inflationary trends in Ghana is the international connection. Inflation in Ghana, like the rest of the world, is partly internationally transmitted. An example is crude oil hikes. Increases in petroleum products often have direct consequences on food prices in Ghana.

Whether internally induced or otherwise, inflation often results in severe socio-economic consequences for the average Ghanaian. This is due to the fact that high inflation reduces purchasing power. Again, the higher the level of inflation, the less amount of goods and services a worker can obtain from his earnings, even if nominal earnings remain constant. In effect, inflation can reduce output and employment.

It is a major cause of civil strife in Ghana. It belies almost every military overthrow in Ghana. As an example, inflation topped 100 % in 1977 moderating to 54 % in 1981. These high rates underlined popular resentments against the Limann government, which led to his removal from office by Jerry John Rawlings in a military putsch in 1981.

As can be gleaned from a survey above, inflation is a worldwide economic fact, but it is an evil in Ghanaian life. And it mirrors the political evolution of the country. Governments who want to stay in power and enjoy the support of the people must resist this devil. They must stick to prudent economic policies and avoid excessive money printing. At a point, many Ghanaians, particularly financial watchers felt the vampire is being leashed.

1.1 PROBLEM STATEMENT

The first democratic change of government through the ballot box in the history of Ghana (since her independence in 1957) was in 2001 when the National Democratic Congress (NDC) government handed power to the New Patriotic Party (NPP) government.

In a budget statement and economic policy presented to parliament on Friday, 9th March 2001, by honourable Minister of Finance, Yaw Osafo-Marfo had described the Economy as an agile one, with external and internal problems culminating into high domestic inflation.

For the year 2001, one of the macro-economic targets set by the NPP government was to achieve an end-of-period rate of inflation of 25%.

In order to attain the monetary policy object for 2001, the government introduced a tight monetary policy stance and actively used Open Market Operations (OMO), Repurchase Agreements (Repos) and prime rate (interest Rate) to influence monetary aggregates in the desired direction.

There after (in 2004) the government claimed that its policies had been successful in reducing the inflation rate.

This study aims at testing the validity of that assertion.

1.2 OBJECTIVES OF THE STUDY

The specific objectives of the study are as follows:

1 To model (using Box-Jenkins ARIMA method) the inflation rate time series of the period 1996 to 2004.

2 To test the effectiveness of the economic policy intervention in the year 2001 on the inflation rate time series for the period 2001 to 2006 (using the interrupted time series experiment)

1.3 METHODOLOGY

The data was obtained from the Ghana Statistical Service and Bank of Ghana Revenue Department. The data on monthly basis consists of the combined data of both food and non-food inflation rate for Ghana, for the period from Jan. 1996 to Dec. 2006.

Box-Jenkins ARIMA time series modeling procedure was used for modeling the data.

Interrupted time series experiment method was used in assessing the impact of the intervention on the series.

Test of hypothesis was used as a preliminary method to compare the mean of the data before and after the intervention.

SPSS was used for the computation and analysis of the data in the series.

1.4 SIGNIFICANCE OF STUDY

This study will provide a method for assessing the effectiveness of government policy intervention on inflation.

1.5 STRUCTURE OF THE THESIS

Chapter one contains the background, problem statement, objectives and significance of the study. Chapter two deals with the review of relevant literature and the theory of time series analysis with special explanation on the Box-Jenkins method.

Chapter three deals with modeling and data analysis; here we apply the Box-Jenkins method of modeling time series. Using the Interrupted time series experiment the effectiveness of the control measure is tested.

Chapter Four deals with the discussion of the results obtained from the Box-Jenkins approach and the interrupted time series experiment.

Finally, from the results it is concluded that the policy to control the inflation rates in Ghana has been effective.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

In this chapter relevant literature, that is other works in different fields where interrupted time series has been applied, has been reviewed and other theoretical works relevant to this study have also been touched on.

2.1 Applications of Interrupted Time Series Experiments

Bloom et al, (2003), also introduced a new approach for measuring the impacts of whole-school reforms. His approach is based on "short" interrupted time-series analysis, which has been used to evaluate programs in many fields. Bloom's approach was used to measure impacts on three facets of student performance: (a) average (mean) test scores, which summarize impacts on total performance; (b) the distribution of scores across specific ranges, which helps to identify where in the distribution of student performance impacts were experienced; and (c) the variation (standard deviation) of scores, which indicates how the disparity in student performance was affected. To help researchers use the approach, his research lays out its conceptual rationale, describes its statistical procedures, explains how to interpret its findings, indicates its strengths and limitations, and illustrates how it was used to evaluate a major whole-school reform—Accelerated Schools.

The use of fluoroquinolones has been linked to increasing bacterial resistance and infection and/or colonization with already resistant pathogens both as a risk factor and based on volume of use. Changes in individual fluoroquinolones used in an institution may also be related to these clinical problems. Interrupted time series analysis, which allows for assessment of the associations of an outcome attributable to a specific event in time, was used to study the effect of changes in a hospital's fluoroquinolone formulary on fluoroquinolone susceptibility rates in select gram-negative pathogens and the methicillin-resistant *Staphylococusaureus* (MRSA) isolation rate.(Bosso, Mauldin et al, 2005) Susceptibility rates to ciprofloxacin were considered for the period of 1993 through 2004, while the MRSA isolation rate was assessed from 1995 through 2004. Levofloxacin was added to the formulary in 1999, and gatifloxacin was substituted for levofloxacin in 2001.

Segmented regression analysis for interrupted time series was used to determine the significance of the differences in levels and slopes over time due to two interventions;

i. The addition of levofloxacin to the formulary in 1999 and

ii. A subsequent switch from levofloxacin to gatifloxacin in 2001

Segmented regression analysis of interrupted time series data allows for the assessment of long-term effects on an outcome attributable to a specific event in time, i.e. the implementation of and intervention.

Total volume of fluoroquinolone use was measured and controlled for in the analysis, allowing for examination of the effect on resistance observed for both levofloxacin and gatifloxacin independent of changes in overall fluoroquinolone use.

Estimations were made of the change in isolation rates of MRSA and susceptibility rates of the gram-negative organisms immediately following the intervention, the difference between the pre- and post-intervention slopes of the outcome, and the periodic average intervention effect after the intervention Proper estimations of standard errors and significance were made through the detection of and correction for autocorrelations. Significance was determined at the 0.05 level and SAS 9.0 was used for the statistical programming.

It was observed that statistically significant changes in the already declining rates of susceptibility of *Pseudomonas aeruginosa* and *Escherichia coli* to ciprofloxacin and in the already rising MRSA isolation rate were associated with the addition of levofloxacin to the formulary. Substitution of gatifloxacin for levofloxacin on the formulary was associated with reversals in the downward trend in *Escherichia coli*susceptibility to ciprofloxacinand the upward trend in MRSA isolation rate. No associations were detected on susceptibility of *Klebsiellapheumoniae* or *Proteus mirabilis* to ciprofloxacin.

These findings suggest that potential changes in susceptibility to fluoroquinolones and isolation of MRSA may vary by both drug and organism.

In the paper by Fleming N.S, Gibson E., Fleming D.G. et al, the researchers' presents a time-series methodology by Box and Tiao (1981) for testing the impact of the American Academy of Pediatrics (APP) recommendation that healthy infants be put to sleep on their side or back to reduce the risk of Sudden Infant Death Syndrome (SIDS).

Quarterly data (for white and non-white kids) that provide more exact parameter estimation i.e. better quantification of effects, than from aggregated (annual) information, was used to determine a “before” baseline level and “after” effect, a design described by Campbell and Stanley(1966). The paper also examined the existence of a seasonal effect, i.e. if the first quarter of each year in Philadelphia has a higher number of deaths than the other three quarters. The evaluation models also considered that data collected at equal intervals were autocorrelated, i.e. correlated with previous data points.

To quantify the impact of both the intervention and seasonality in percentage terms specifically, the natural logarithms of the dependent variable was computed from the natural logarithm of the quarterly death rate per 10,000 annual live births. In instances where no death occurred in the quarter, one death was assumed since there is no natural log of 0.

After modeling, computations were performed using the SAS/ETS (PROC ARIMA) statistical computer software package as described in the manual (SAS Institute Inc., 1988). The program use a method based on maximum likelihood to estimate parameters after obtaining initial values from conditional least squares. Chi-square statistics are computed for each model that test the model’s goodness –of-fit, i.e. the lack of autocorrelation in the residuals and remaining white noise process

The results revealed that, the intervention caused a significant reduction in the SIDS rate for both whites and non-whites. Winter SIDS rates were higher than those in other seasons. A supplementary analysis revealed that no statistical interaction existed between the intervention and seasonality, i.e. the effect of winter did not change after the intervention.

In a paper by Bloom et al., (1999), interrupted time series analysis was employed for estimating the impact of school restructuring programs designed to increase student achievement in primary and secondary school.

With the development of various school restructuring programs, foremost among these being the Henry Levin's Accelerated School Project, Ted Sizer's Coalition of Essential Schools, James Comer's School Development Program etc, all designed to affect all students in a school, it appeared possible to implement them for some students but not for others.

Under these and other conditions, however, it is possible to estimate the impacts of such programs by measuring the extent to which student achievement increased relative to the pre-program trend.

In the paper Bloom first illustrates the approach and considers its strengths and weaknesses. He then describes how to estimate program impacts and their standard errors from a simple regression model. Next the focuses on the statistical precision of these impact estimates and the research design considerations that affect this precision. The paper then concludes by briefly outlining several important issues related to the approach that could be addressed in future research.

Huang et al.(2008) considered forecasting the latent rate profiles of a time series of inhomogeneous Poisson processes. Their work was motivated by operations management of queueing systems, in particular, telephone call centers, where accurate forecasting of call arrival rates was a crucial primitive for efficient staffing of such centers. Their forecasting approach utilizes dimension reduction through a factor analysis of Poisson variables, followed by time series modeling of factor score series. Time series forecasts of factor scores were combined with factor loadings to yield forecasts of future Poisson rate profiles. Penalized Poisson regressions on factor loadings guided by time series forecasts of factor scores were used to generate dynamic within-process rate updating. Methods were also developed to obtain distributional forecasts. Their methods were illustrated using simulation and real data. Their empirical results demonstrate how forecasting and dynamic updating of call arrival rates can affect the accuracy of call center staffing.

Generally, the result of clustering cannot reflect the similarities of time series properly because of the disturbance of noises and details in time series. Liu et al.(2008), proposed a new approach to this problem based on wavelet decomposition and denoising is proposed. Their approach has been tested and analyzed by Synthetic Control Chart Time Series from University of California, Irvine (UCI) knowledge discovery in databases (KDD)

Interrupted time series designs have been frequently employed to evaluate program impact. Analysis strategies to determine if shifts have occurred are not well known. The case where statistical fluctuations (errors) may be assumed independent is considered, and a segmented regression methodology presented. Gillings et al. (1981) assessed the changes in local and state perinatal post neonatal mortality to identify historical trends and which they used to evaluate the impact of the North Carolina Regionalized Perinatal Care Program when seven years of post-program mortality data become available. The perinatal program region was contrasted with a control region to provide a basis for interpretation of differences noted. Relevant segmented regression models provided good fits to the data and highlighted mortality trends over the last 30 years. Considerable racial differences in these trends were identified, particularly for post neonatal mortality. Gillings, et al., later considered segmented regression which was relevant for the analysis of interrupted time series designs in applications when errors can be taken to be independent. Thus, their methodology may be regarded as a general statistical tool for evaluation purposes.

Despite recent considerable advances in structural health monitoring (SHM) of civil infrastructure, converting large amounts of data from SHM systems into usable information and knowledge remains a great challenge. Omenzetter et al. (2006), addresses the problem through their analysis of time histories of static strain data recorded by an SHM system installed in a major bridge structure and operating continuously for a long time. They formulate a vector seasonal autoregressive integrated moving average (ARIMA) model for the recorded strain signals. The coefficients of the ARIMA model were allowed to vary with time and were identified using an adaptive Kalman filter. Their proposed method has been used for analysis of the signals recorded during the construction and service life of the bridge. By observing various changes in the ARIMA model coefficients, unusual events as well as structural change or damage sustained by the structure could be revealed.

Hay et al., (2001) presented a Bayesian analysis of a time series of counts to assess its dependence on an explanatory variable. The time series represented was the incidence of the infectious disease ESBL-producing Klebsiellapneum**on**iae in an Australian hospital and the explanatory variable was the number of grams of antibiotic (third generation) cephalosporin used during that time. They demonstrated that there is a statistically significant relationship between disease occurrence and use of the antibiotic, lagged by three months. The model used is a parameter-driven in the form of a generalized linear mixed model. Comparison of models was made in terms of mean square error.

Correlation techniques are important tools for investigating relationships between crop growth and environment. However when applied to a time series the presence of autocorrelation affects the estimates of correlations between two observed series, an effect which is sometimes overlooked in studies reported in the literature. Appropriate statistical techniques were needed to ensure that proper inferences can be made from these observations. Empirical time series modeling has the potential for removing autocorrelations in many of these cases. To test the feasibility of this technique Kuehl et al., (1976), investigated the relationship between boll retention in cotton (Gossypiumhirsutum L.) and 5-day average minimum temperature during the growing season was investigated with 3 years of data. Their best fitting autoregressive moving average models were selected for the observed time series for each of the 3 years. Residual series, observed minus predicted values, were computed from the estimated time series models for boll retention and minimum temperature and were free of autocorrelation. Negative cross-correlations were found between the residual series for boll retention and the average minimum temperature for the 5-day periods beginning with 10 days and 1 day prior to anthesis, respectively. The negative relationship Kuehl et al. had suggests that high night time temperatures accumulated over 5-day periods prior to and during anthesis increase boll shedding. They conclude that time series modeling was an effective technique to aid the identification of relationships between agronomic variables measured in time sequences.

The need to characterize and forecast time series recurs throughout the sciences, but the complexity of the real world is poorly described by the traditional techniques of linear time-series analysis. Although newer methods can provide remarkable insights into particular domains, they still make restrictive assumptions about the data, the analyst, or the application. Here Gershenfeld et al., showed that signals that are nonlinear, non-stationary, non-gaussian, and discontinuous can be described by expanding the probabilistic dependence of the future on the past around local models of their relationship. The predictors they derived from this general framework have the form of the global combinations of local functions that were used in statistics, machine learning and studies of nonlinear dynamics. Their method offers forecasts of errors in prediction and model estimation, provides a transparent architecture with meaningful parameters, and had straight forward implementations for offline and online applications. They demonstrated their approach by applying it to data obtained from a pseudo-random dynamical system, from a fluctuating laser, and from a bowed violin.

Kei-Mu et al.(1996), in their research compared an endogenous growth model to an exogenous model using time series analysis to determine the effect of government policy changes to the growth rate of GNP per capita. Their results indicated that for endogenous models, changes in fiscal policies lead to permanent changes in GNP. On the other hypothesized that, policy shifts in exogenous growth economies only show temporary changes in the GNP. They established that Non-military equipment capital and non-military structural capital are the only factors that affect long-term GNP levels.

Analysis of spatial patterns in images can provide valuable information in many application domains, such as in geography, meteorology and medicine. Kontos et al. (2004), proposed to apply techniques from the time series domain to analyze the spatial patterns extracted from 3D images. After traversing an image using a space-filling curve, they discovered discriminative patterns by analyzing the spatial sequence in the transformed domain. Because of the similarity of the sequences with time series, they proposed the use of existing time series similarity analysis techniques, including Euclidean distance, and dimensionality reduction techniques, such as singular value decomposition and piecewise aggregate approximation, for further analysis of the spatial patterns. As a case study, they analyzed an fMRI dataset. Their experimental results verified that the discovered spatial patterns have strong discriminative power among different classes and the overall accuracy for clustering and similarity retrieval is above 90% and as high as 100% for certain experimental settings.

The adjusted interrupted time-series (AITS) approach was introduced by Walker et al. (2004) to develop empirical methods for measuring the impacts of place-based local development strategies. The AITS approach was applied to three community development initiatives using single-family home prices as the outcome indicator and it was found that it could measure impacts on both the base level of prices and the rate of price appreciation.

Marquéset el. was focused on the application of neural network based models to the analysis of total ozone (TO) time series. Processes that affect total ozone are extremely nonlinear, especially at the considered European mid-latitudes. Artificial neural networks (ANNs) are intrinsically non-linear systems; hence they are expected to cope with TO series better than classical statistics do. Moreover, neural networks do not assume the stationarity of the data series so they are also able to follow time-changing situations among the implicated variables. These two features turn NNs into a promising tool to catch the interactions between atmospheric variables, and therefore to extract as much information as possible from the available data in order to make, for example, time series reconstructions or future predictions. Models based on NNs have also proved to be very suitable for the treatment of missing values within the data series. In their paper, they present several models based on neural networks to fill the missing periods of data within a total ozone time series, and models able to reconstruct the data series. The results released by the

ANNs were compared with those obtained by using classical statistics methods, and better accuracy was achieved with the nonlinear ANNs techniques. Different network structures and training strategies were tested depending on the specific task to be accomplished.

Focusing on the widely-used Box-Jenkins "airline" model, Aston et al. showed how the class of seasonal ARIMA models with a seasonal moving average factor can be parsimoniously generalized to model time series with heteroskedastic seasonal frequency components. Their frequency-specific models decompose this factor by associating one moving average coefficient with a proper subset of the seasonal frequencies 1, 2, 3, 4, 5 and 6 cycles per year and a second coefficient with the complementary subset. A generalization of Akaike's AIC was presented to determine these subsets. Properties of seasonal adjustment filters and adjustments obtained from the new models were examined as are forecasts.

Research describing slack-adjusted data envelopment analysis was presented by Sueyoshi et al in 1994. In particular their problem of multiple solutions on return to scale was addressed with reference to the Japanese power industry.

A study conducted by Xong et al. (1999) to analyze a novel algorithm for determining the predictability index. The relationship between the predictability index and the position of the poles and lag of a time series were characterized as an AR(p) model. Numerical examples were then presented to determine the effectiveness of the algorithm. Their results indicated that the estimator is an effective tool in predictability ranking applications and time series analysis.

Gil-Alana et al. also proposed a general time series model, whose components are modelled in terms of fractionally integrated processes. This specification allows them to consider the trend, the seasonal and the cyclical components as stochastic processes, including the unit root models as particular cases. A very general version of the tests of Robinson (1994) was used to test the order of integration of each component. Finite-sample critical values of the tests were evaluated and, an empirical application was also carried out at the end of the article.

In the next chapter we discuss the theory of time series and interrupted time series analysis.

CHAPTER THREE

METHODOLOGY

3.0 INTRODUCTION

In this chapter we discuss the theory of time series in terms of its definition, types of time series as well as the theory of time series ARIMA models.

3.1.0 Definition

Time series is a time dependent sequence denoted $Y_1, Y_2, \ldots Y_t$ or Y_t where $t \in N$ where $1, 2, \ldots, n$ denote time steps.

3.1.1 Deterministic Time Series

If from past knowledge, the future of a time series can be exactly predicted, it is a deterministic series and requires no further investigation. It can be expressed as a known function. That is $Y_t = f(t)$.

3.1.2 Stochastic Time Series

If a time series can be expressed as $Y_t = X(t)$, where X is a random variable, then Y_t is a stochastic time series.

3.2 OBJECTIVES OF TIME SERIES ANALYSIS

The main objectives of analyzing a time series are classified as description, explanation, prediction and control.

3.2.1 Description

When presented with time series data, the first step in the analysis is usually to plot the data to obtain simple descriptive measure of the main properties of the series as seasonal effect, trend etc.

Apart from trend and seasonal variations, the outliers to look for in the graph of the time series are the possible presence of turning points, where for example, an upward trend has suddenly changed to a downward trend.

3.2.2 Explanation

When observations are taken on two or more variables, it may be possible to use the variable in one time series variable to explain the variation in the other time series variable. This may give a deeper understanding of the mechanism which generated a given time series. For example, sales are affected by price and economic condition.

3.2.3 Prediction

Given an observed time series one may want to predict the future values of the series. This is an important task in sales forecasting and in the analysis of economic and industrial time series. Prediction is closely related to control problems in many situations. For example if we can predict that manufacturing process is going to move off target, then appropriate corrective action can be taken.

3.2.4 Control

When a time series is generated which measures the quality of a manufacturing process, the aim of the analysis may be to control the process. In statistical quality control, the observations are plotted on control charts and the controller takes action as a result of studying the charts. Box and Jenkins have described a more sophisticated control strategy which is based on fitting a stochastic model to the series from which future values of the series are predicted. The values of the process variables predicted by the model are taken as target values and the variables conform to the target values.

3.3 COMPONENTS OF TIME SERIES

Traditional methods of time series analysis are mainly concerned with decomposing the variation in series into the various components of trend, periodic and stochastic.

3.3.1 Periodic Component

If $Y_t = Y_{t+T} + e_t$ for all $t \in N$, then the time series has a periodic component of period T.

3.3.2 Trend Component

If $Y_t = y + \beta_t + e_t$, then there exist a linear trend with the slope being β.

3.4.0 STATIONARY TIME SERIES

A time series is said to be stationary if the joint distribution of X_{t_1} , ... , X_{t_n} is the same as the joint distribution of X_{t_1+T} , ... , X_{t_n+T} for allt_{1+T} , ... , t_{n+T}. In other words shifting the time origin by an amount T has no effect on the joint distribution which must therefore depend only on the intervals between t_1 , ... , t_n.

3.4.1 Autocorrelation Function (ACF)

The autocorrelation function measures the degree of correlation between neighboring observations in a time series. The autocorrelation function at lag k is defined as

$$\rho_k = \frac{E[(Y_t+\mu_y)(Y_{t+k}-\mu_y)]}{[E(Y_t+\mu_y)^2 E(Y_{t+k}-\mu_y)^{2}]}$$

$$\rho_k = \frac{co(Y_t, Y_{t+k})}{\sigma_{t+k}}$$

The autocorrelation coefficient is estimated from sample observation using the formula;

$$r_k = \frac{\sum_{t=2}^{n}(Y_t - \mu_y)(Y_{t+k} - \mu_y)}{\sum_{t=1}^{n}(Y_t - \mu_t)^2}$$

(Hamilton J.D, 1994)

3.4.2 Sampling Distribution of Autocorrelation Coefficient

The autocorrelation coefficient of a random data are approximately normal with $\mu_{pk} = 0$ and $\sigma_{pk}^2 = \frac{1}{n}$. Where n is the size of the sample. Thus for a random sample of size 40 we expect $-2\sigma_p < r_k < 2\sigma_p$ for significant limits of two standards errors which is

$\frac{-2}{\sqrt{40}} \leq r_k \leq \frac{2}{\sqrt{40}}$ which is $-0.316 \leq r_k \leq 0.316$. Hence any value of r_k lying outsides this interval is said to be significantly different from zero. *(Hamilton J.D, 1994)*

3.4.3 Partial Autocorrelation Coefficient

Partial autocorrelation function measures the degree of association between Y_t andY_{t+k} when the effects of other time lags on Y are held constant. The partial autocorrelation function PACF denoted by $\{\emptyset_{kk}: k = 1,2 \ldots \ldots..\}$ The set of partial autocorrelation at various lags k are defined by $\emptyset_{kk} = |P_k^*|/|P_k|$ where P_k is the KxK autocorrelation matrix and P_k^* is P_k with the last column replaced by $[P_1, P_2, \ldots \ldots, P_k]^T$ and an example is $\emptyset_{11} = \emptyset_1 = P_1$

and $$\emptyset_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

And estimates of $\emptyset_{kk}$ can be obtained by replacing the ρ_i by r_i.

3.4.4 Sampling Distribution of the Partial Autocorrelation Coefficients

The partial autocorrelation coefficients of random data are approximately normal with $\mu_{\emptyset kk} = 0$ and $\sigma_{\emptyset kk} = \frac{1}{\sqrt{n}}$ and n is the size of the sample. Thus for a random sample of size 40 we expect $-2\sigma_{\emptyset kk} \leq \emptyset_{kk} \leq 2\sigma_{\emptyset kk}$ for significant limits of two standard errors which is $\frac{-2}{\sqrt{40}} \leq \emptyset_{kk} \leq \frac{2}{\sqrt{40}}$

Which is equal to $-0.316 \leq \emptyset_{kk} \leq 0.316$. Hence any value of $\emptyset_{kk}$ lying outside this interval is said to be significantly different from Zero. *(Hamilton J.D, 1994)*

3.4.5 An Autoregressive Model of Order p [AR(p)]

An autoregressive model of order p denoted by AR(p) is a special kind of regressive in which the explanatory variables are past values of the process. An autoregressive model of order p is given by $Y_t = \sum_{k=1}^{p} \alpha_k Y_{t-k} + \mu + e_t$ Where μ is the mean of the time series data and e_t is the white noise.

The order of an AR(p) process is determined by the partial autocorrelation function (PACF). An AR(p) process has its PACF cutting off after lag p and the ACF decays. For example the PACF of an AR(1) process cuts off after lag one (1). *(Hamilton J.D, 1994)*

3.4.6 Autoregressive Process of Order one (1) AR(1)

The AR(1) process is

$$Y_t = \alpha_1 Y_{t-1} + \mu + e_t$$

Putting $\mu = 0$ we have

$$Y_t = \alpha_1 Y_{t-1} + e_t$$

Multiplying through by Y_{t-k} we have

$$Y_{t-k} Y_t = \alpha_1 Y_{t-k} Y_{t-1} + e_t Y_{t-k}$$

$$cov(Y_{t-k}, e_t) = \alpha_1 cov(Y_{t-k}, Y_{t-1}) + cov(Y_{t-k}, e_t)$$

But $cov(Y_{t-k}, e_t) = 0$ since Y_{t-k} depends only on $e_{t-k}, e_{t-k-1}, \dots$ .... which are not correlated with e_t for k>0. Hence

$$\gamma_k = \alpha_1 \gamma_{k-1}$$

Dividing through by γ_0 we have $\frac{\gamma_k}{\gamma_0} = \alpha_1 \frac{\gamma_{k-1}}{\gamma_0}$

$$\rho_k = \alpha_1 \rho_{k-1} \ where \ \rho_0 = 1$$

We have $\rho_1 = \alpha_1 \rho_0 = \alpha_1$ since $(\rho_0 = 1)$

$$\rho_1 = \alpha_1$$

For k=2,

$$\rho_2 = \alpha_1 \rho_1 = \alpha_1(\alpha_1) = \alpha_1^2$$

For k=3,

$$\rho_3 = \alpha_1\rho_2 = \alpha_1\alpha_1^2 = \alpha_1^3$$

And in general *(Box and Jenkins, 1971)*

$$\rho_k = \alpha_1^k$$

3.4.7 Estimating AR(p) Parameters using the method of ordinary Least Squares

The method of ordinary least squares can be employed to estimate the parameters of the AR(p) process. In multiple regression (5) we have,

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + e_t$ and $\beta = (X^T X)^{-1} X^T Y$

Where $\beta = [\beta_0, \beta_1, \dots, \beta_k]^T$ $X = [X_1, X_2, \dots, X_n]^T$

And $Y = [Y_1, Y_2, \dots, Y_n]^T$

Similarly with AR process the X vector is formed using the past values of Y. for example consider the AR(1) process :

$$Y_t = \alpha_1 Y_{t-1} + e_t + \mu$$

Hence $Y_2 = \alpha_1 Y_1 + \mu$

$$Y_3 = \alpha_1 Y_2 + \mu$$

$$Y_t = \alpha_1 Y_{t-1} + \mu$$

This equation is over determined and it is solved using the ordinary least squares method.

The X vector is $(Y_1, Y_2, \dots.., Y_{t-1})^T$ and the Y vector is $(Y_2 - \mu, Y_3 - \mu, \dots, Y_t - \mu)^T$

Then *(Hamilton J.D, 1994)*

$\alpha_1 = (X^T X)^{-1} X^T Y$

3.4.8 Moving Average of Order q MA(q)

MA models provide predictions of Y_t based on a linear combination of past forecast errors. In particular the MA model of order q is given by *(Hamilton J.D, 1994)*

$$Y_t = \sum_{k=1}^{q} \theta_k e_{t-k} + \mu + e_t$$

3.4.9 Autocorrelation Function (ACF) of MA(q)

$$\gamma(k) = cov(Y_t, Y\ \ t+k)$$

$$= cov(\theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}, \theta_1 e_{t+k-1} + \quad \theta_2 e_{t+k2} + \cdots + \theta_q e_{t+k-q})$$

$$= \begin{cases} 0 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots k > q \\ \sigma_e^2 \sum_{t-1}^{q-k} \theta_i \theta_{i+k} \dots\dots\dots\dots\dots\dots\dots\dots k = 0,1, \dots q \\ \gamma(-k) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots.. k < 0 \end{cases}$$

Since

$$cov(e_s, e_t) = \begin{cases} \sigma_e^2 \dots\dots\dots\dots\dots\dots\dots\dots\dots s = t \\ 0 \dots\dots\dots\dots\dots\dots\dots\dots\dots s \neq t \end{cases}$$

Hence the autocorrelation function (ACF) of MA(q) process is given by

$$\rho(k) = \begin{cases} 1 \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots k = 0 \\ \sum_{i=1}^{q-k} \theta_i \theta_{i+u} \Big/ \sum_{t=1}^{q} \theta_i^2 \ldots k = 1,2, \ldots q \\ \rho(-k) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots . . k < 0 \end{cases}$$

The order of the MA(q) is given by the autocorrelation function. The ACF cuts after lag q and the partial autocorrelation function decays to zero. Thus an MA(1) process cuts off after lag one. In other words the ACF after lag one will not be significantly different from zero.

3.4.10 Moving Average process of Order one MA(1)

The MA(1) process is given by

$$Y_t = \theta_1 e_{t-1} + \mu + e_t$$

And its autocorrelation is given by

$$\rho(k) = \begin{cases} 1 \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots . k = 0 \\ \theta_1 \big/ (1 + \theta^2) \ldots \ldots \ldots \ldots \ldots \ldots k = +1 \\ 0 \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots otherwise \end{cases}$$

Thus

$$\rho_1 = \frac{\theta_1}{1+\theta_1^2}$$

$$\rho_1 + \rho_1\theta_1^2 - \theta_1 = 0$$

$$\Rightarrow \quad \rho_1\theta_1^2 - \theta_1 = 0$$

The parameters are thus roots of a quadratic. This means that we can find two MA(1) processes that corresponds to the same ACF. To establish a one-to-one correspondence between the ACF and the model and obtain a converging autoregressive representation, we restrict the moving average parameter such that$|\theta| < 1$. This restriction is known as the invertibility implies that the process can be written in terms of an autoregressive model. *(Hamilton J.D, 1994)*

3.4.11 Estimation of the model parameters of the MA(q) process.

For an MA(1) process an iterative method is used since the ordinary least squares cannot be used as the residual sum of squares is not a quadratic function. The approach suggested by box and Jenkins is used. Given the MA(1) model.

$Y_t = \theta_1 e_{t-1} + \mu + e_t$ Where μ and $\theta_1$are constants and $r_1 = \frac{\theta_1}{1+\theta_1^2}$

Then the residual sum of squares is calculated using $Y_t = \mu + e_t + \theta_1 e_{t-1}$ recursively in the form

$$e_t = Y_t - \mu - \theta_1 e_{t-1}$$

With $e_0 = 0$ we have

$e_1 = Y_1 - \mu$

$e_2 = Y_2 - \mu - \theta_1 e_1$

$e_3 = Y_3 - \mu - \theta_1 e_2$

$e_N = Y_N - \mu - \theta_1 e_{N-1}$

Then $\sum_{t-1}^{N} e_t^2$ is calculated.

This procedure is then repeated for other values of μ and θ_1 and the sum of squares $\sum_{t-1}^{N} e_t^2$ computed for a grid of points in the (μ, θ_1) plane. We then determine by inspection the least squares estimates of μ and θ_1 which minimizes$\sum_{t-1}^{N} e_t^2$. *(Box and Jenkins, 1971)*

3.4.12 The Duality of AR and MA processes

We show that the Random Walk process given by

$$Y_t = Y_{t-1} + e_t$$

Can be rewritten as an infinite moving average. Indeed, consider the following moving average,

$$Y_t = e_t + e_{t-1} + e_{t-2} + \cdots = \sum_{t=0}^{\infty} e_{t-i}$$

$$= (1 + B + B^2 + B^3 + \cdots)e_t$$

$$= \left(\sum_{t=0}^{\infty} B^i\right)e_t$$

We recall that$\sum_{t=0}^{\infty} y^t = {}^{1}/_{(1-y)}$. Is valid when $|y| < 1$

Hence $Y_t = \left(\frac{1}{1-B}\right)e_t$

So that

$$(1 - B)Y_t = e_t$$

$$Y_t - BY_t = e_t$$

$$Y_t - Y_{t-1} = e_t$$

$$Y_t = Y_{t-1} + e_t$$

This is the random walk process. This means that a finite autoregressive process. For example, we show that an MA(1) process is an infinite autoregressive process.

For such a process,

$$Y_t = e_t - \theta_1 e_{t-1}$$

Using the B operator notation we have

$$Y_t = (1 - \theta_1 B)e_t$$

$$\frac{Y_t}{1-\theta_1 B} = e_t$$

$$(1 + \theta_1 B + \theta_2^2 B^2 + \cdots)Y_t = e_t$$

$$Y_t + \theta_1 Y_{t-1} + \theta_2^2 Y_{t-2} + \cdots = e_t$$

$$Y_t + \theta_1 Y_{t-1} + \theta_2^2 Y_{t-2} + \cdots = e_t$$

This is an infinite autoregressive process.

3.4.13 ARMA or "Mixed" Process

Consider the process given by;

$$Y_t = \alpha_1 Y_{t-1} + \theta_1 e_{t-1} + e_1$$

This can be rewritten as

$Y_t - \alpha_1 Y_{t-1} = e_t + \theta_1 e_{t-1}$ Or

$(1 - \alpha B)Y_t = (1 + \theta B)e_t$…………………………….. (1)

$AR(B)Y_t = MA(B)e_t$

This is called a mixed or autoregressive moving average (ARMA) process of order (1,1).

Since equation (1) is ARMA(1,1) if $|\theta| < 1$ it can be rewritten as

$$(1 - \alpha B)\left(\frac{1}{1 + \theta B}\right)Y_t = e_t$$

$$(1 - \alpha B)(1 - \theta B + \theta^2 B^2 - \theta^3 B^3 + \cdots)Y_t = e_t$$

$$[(1 - \alpha + \theta)B + (\alpha\theta + \theta^2)B^2 + \cdots]Y_t = e_t$$

This is an infinite order AR process. This is true if $|\alpha| < 1$ and $|\theta| < 1$ i.e. if the AR is stationary and MA is invertible. If we have two polynomial in B, MA(B) and AR(B), and an ARMA model,

$$AR(B)Y_t = MA(B)e_t$$

It is possible to write the model as an infinite AR process:

$$\left(\frac{AR(B)}{MA(B)}\right) Y_t = e_t$$

Or an infinite MA process

$$Y_t = \left(\frac{MA(B)}{AR(B)}\right) e_t$$

And approximate either by finite processes

ARMA processes are parsimonious however identifying those using ACF and PACF may be difficult. The condition necessary for dividing by AR(B) is that the AR process be stationary and by MA(B) is that the MA process be invertible.

3.4.14 Autoregressive Moving Average Model (ARMA)

A more general model is a mixture of the AR(p) and MA(q) models and is called an autoregressive moving average model (ARMA) of order (p,q).

The ARMA(p,q) is given by

$$Y_t = \sum_{i=1}^{p} \alpha_i Y_{i-1} + \sum_{i=1}^{q} \theta_i e_{i-1} + \mu + e_t$$

An example of an ARMA(1,1)

$$Y_t = \alpha_1 Y_{t-1} + \theta_1 e_{t-1} + \mu + e_t$$

An important characteristic of ARMA models is that both the ACF and PACF do not cut off as in AR and MA models.

(Box and Jenkins, 1971)

3.4.15 ARMA(1,1) Model

An example of an ARMA(p,q) model is the ARMA(1,1) model given by

$$Y_t = \alpha_1 Y_{t-1} + \theta_1 e_{t-1} + \mu + e_t$$

The ARMA(1,1) model is stationary if $-1 < \alpha_1 < 1$ and invertible if $-1 < \theta_1 < 1$. Its theoretical autocorrelation function (ACF) and partial autocorrelation function (PACF) trail off to zero in a damped exponential fashion. In an ARMA(1,1) model both ACF and the PACF trail off to zero. *(Hamilton J.D, 1994)*

3.4.16 Estimating the Parameters of an ARMA Model

The procedure for estimating the parameters of the ARMA model is like the one for the MA model it is an iterative method. Like the MA the residual sum of squares is calculated at every point on a suitable grid of the parameter values, and the values, and the values give the minimum sum of squares are the estimates.

For an ARMA(1,1) the model is given by

$$Y_t - \mu = \alpha_1 (Y_{t-1} - \mu) e_t + \theta_1 e_{t-1}$$

Given N observation $Y_1, Y_2, \dots, Y_N$, we guess values for μ, α_1, θ_1, set $e_0 = 0$ and $Y_0 = 0$ and then calculate the residuals recursively by

$$e_1 = Y_1 - \mu$$

$$e_2 = Y_2 - \mu - \alpha_1 (Y_1 - \mu) - \theta_1 e_1$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$e_N = Y_N - \mu - \alpha_1(Y_1 - \mu) - \theta_1 e_{N-1}$$

The residual sum of squares $\sum_{t=1}^{N} e_t^2$ is calculated. Then other values ofμ, α_1, θ_1, are tried until the minimum residual of squares is found.

Note: It has been found that most of the stationary time series occurring in practices can be fitted by AR(1), AR(2), MA(1), MA(2), ARMA(1,1) or white noise models that are customarily needed in practice. *(Hamilton J.D, 1994)*

3.4.17 The Autoregressive Integrated Moving Average Model (ARIMA)

If a non-stationary time series which has variation in the mean is differenced to remove the variation the resulting time series is called an integrated time series. It is called an integrated model because the stationary model which is fitted to the differenced data has to be summed or integrated to provide a model for the non-stationary data. Notationally, all AR(p) and MA(q) models can be represented as ARIMA(1,0,0) that is no differencing and no MA part.

The general model is ARIMA(p,d,q) where p is the order of the AR part, d is the degree of differencing and q is the order of the MA part.

Writing $$W_t = \nabla^d Y_t = (1 - B)^d Y_t$$

The general ARIMA process is of the form

$$W_t = \sum_{t=1}^{p} \alpha_i W_{t-i} + \sum_{t=1}^{q} \theta_i e_{t-i} + \mu + e_t$$

3.4.18 ARIMA(1,1,1) Process

An example of ARIMA(p,d,q) is the ARIMA(1,1,1) which has one autoregressive parameter, one level of differencing and one MA parameter is given by

$$W_t = \alpha_1 W_{t-1} + \theta_1 e_{t-1} + \mu + e_t$$

$$(1 - B)Y_t = \alpha_1(1 - B)Y_{t-1} + \theta_1 e_{t-1} + \mu + e_t$$

Which can be simplified further as

$$Y_t - Y_{t-1} = \alpha_1 Y_{t-1} - \alpha_1 Y_{t-2} + \theta_1 e_{t-1} + \mu + e_t$$

$$Y_t - Y_{t-1} = \alpha_1(Y_{t-1} - Y_{t-2}) + \theta_1 e_{t-1} + \mu + e_t$$

3.4.19 Estimating the parameters of an ARIMA Model

In practice most time series are non-stationary and the series is differenced until the series becomes stationary. An AR, MA or ARMA model is fitted to the differenced series and estimation procedures are as described for the AR, MA, and ARMA above.

3.4.20 Stationarity and Invertibility Conditions of Specific Time Series model

In the table below we display the stationarity and invertibility conditions of specific time series models and the behavior of their theoretical ACF and PACF functions.

Table 3.12.3 Specific Time Series Models

ARIMA MODEL	STATIONARITY CONDITIONS	INVERTIBILITY CONDITION	ACF COEFFICIENTS	PACF COEFFICIENTS
(1,d,0)	$-1 < \alpha_1 < 1$	NONE	Dies down	Cuts off after lag one
(2,d,0)	$\alpha_1 + \alpha_2 < 1$ $\alpha_1 - \alpha_2 < 1$ $-1 < \alpha_2 < 1$	NONE	Dies down	Cuts off after lag two
(0,d,1)	NONE	$-1 < \theta_1 < 1$	Cuts off after lag one	Dies down
(0,d,2)	NONE	$\theta_1 + \theta_2 < 1$ $\theta_1 - \theta_2 < 1$ $\theta_2 < 1$	Cuts off after lag two	Dies down
(1,d,1)	$-1 < \alpha_1 < 1$	$-1 < \theta_1 < 1$	Dies down	Dies down

3.4.21 The Box-Jenkins Method of Modeling time Series

The Box-Jenkins methodology is a statistical sophisticated way of analyzing and building a forecasting model which best represents a time series. The first stage is the identification of the appropriate ARIMA models through the study of the autocorrelation and partial autocorrelation functions. For example if the partial autocorrelation cuts off after lag one and the autocorrelation function decays then ARIMA(1,0,0) is identified. The next stage is to estimates the parameters of the ARIMA model chosen.

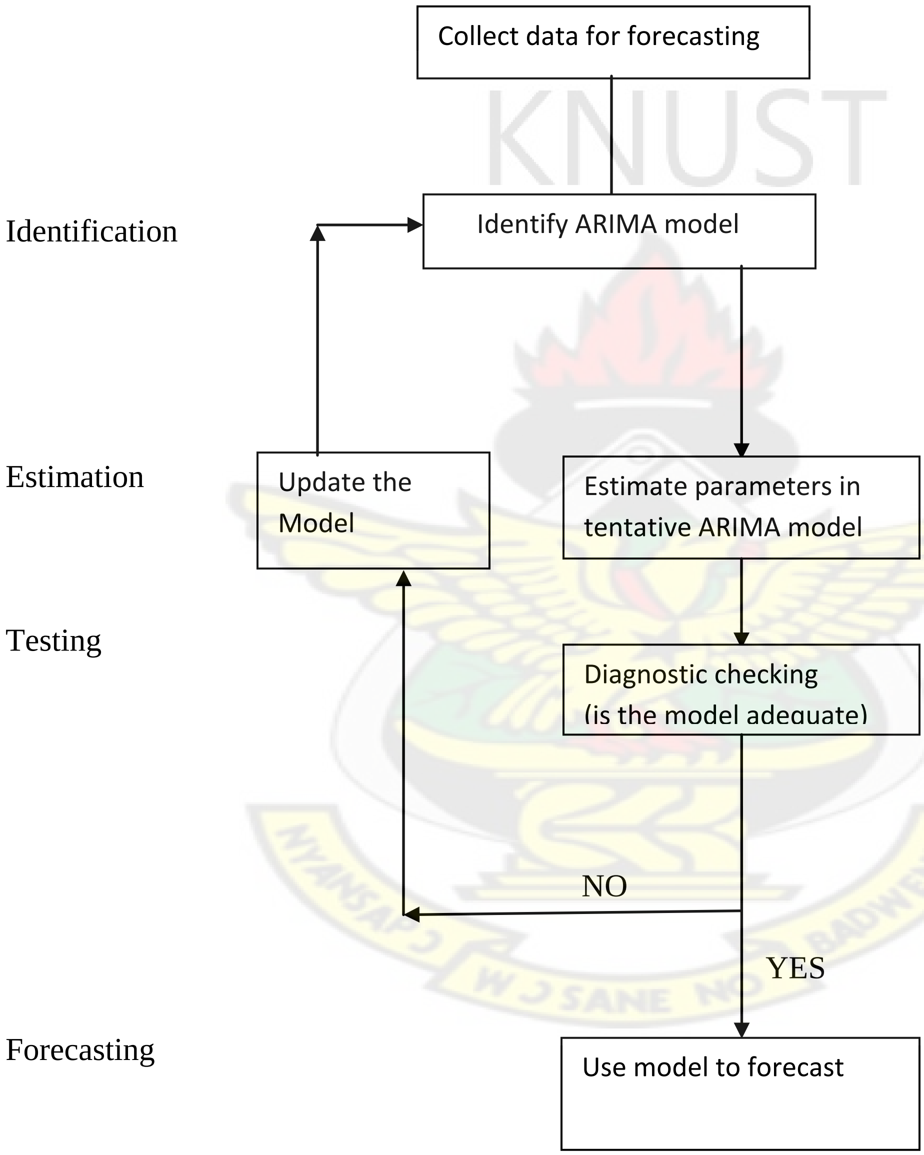
The third stage is the diagnostic checking of the model. The Q-statistic is used for the model adequacy check.

If the model is not adequate then the forecaster goes to stage one to identify an alternative model and it is tested for adequacy and if adequacy then the forecaster goes to the final stage of the process.

The fourth stage is where the analysis uses the model chosen to forecast and the process ends.

Below is a schematic representation of the box-Jenkins process.

Figure 3.13 The Box-Jenkins Process



3.4.22 Identification techniques

Identification methods are rough procedures applied to a set of data indicate the kind of representational model that will be further investigated. The aim here is to obtain some idea of the values p, d and q needed in the general linear ARIMA model and to obtain initial estimates for the parameters.

The task here is to identify an appropriate subclass of models from the general ARIMA family $\alpha(B)\nabla^d Y_t = \theta(B)e_t$ which may be used to represent a given time series. The approach will be as follows;

(a) To difference Y_t as many times as is needed to produce stationarity, reducing the process under study to the mixed autoregressive moving average process $\alpha(B)W_t = \theta_0 + \theta(B)e_t$ where $W_t = (1 - B)^d Y_t = \nabla^d Y_t$

(b) To identify the resulting ARMA process

The principle tools for putting (a) and (b) into effect is the sample autocorrelation function and the sample partial autocorrelation function. Apart from helping to guess the form of the model, they are used to obtain approximate estimates of the parameters of the model. These approximations are useful at the estimates stage to provide starting values for iterative procedures employed at that stage.

3.4.23 Use of the autocorrelation and Partial Autocorrelation functions in Identification

A stationary mixed autoregressive moving average process of order (p,0,q), $\alpha(B)Y_t = \theta(B)e_t$, the autocorrelation function satisfies the difference equation

$$\alpha(B)\rho_k = 0 \qquad\qquad k > q$$

Also, if

$$\alpha(B) = \prod_{i=1}^{p}(1 - G_i B)$$

The solution of this difference equation for the kth autocorrelation is, assuming distinct roots, of the form

$$\rho_k = A_1 G_1^k + A_2 G_2^k + \cdots + A_p G_p^k \qquad\qquad k > q - p$$

The stationarity requirement that the zeros of $\alpha(B)$ lie outside the unit circle implies that the roots $G_1, G_2, G_3, \dots, G_k$ lie inside the unit circle. Inspection of the equation

$$\rho_k = A_1 G_1^k + A_2 G_2^k + \cdots + A_p G_p^k \qquad\qquad k > q - p$$

Shows that in the case of a stationary model in which none of the roots lie close to the boundary of the unit circle, the autocorrelation function will quickly "die out" or decay for moderate and large k.

Suppose that a single real root, say G_1 approaches unity, so that $G_1 = 1 - \delta$ where δ is a small positive quantity. Then, since for k large, $\rho_k = A_1(1 - k\delta)$ the autocorrelation function will not die out quickly and will fall off slowly and very nearly linearly. Similarly if more than one root approaches unity the autocorrelation function will decay slowly. Therefore if the autocorrelation function dies out slowly it implies there is at least a root which approaches unity. As a result failure of the estimated autocorrelation function to die out rapidly might logically suggest that the underlying stochastic process is non-stationary in Y_t but possible stationary in∇Y_t, or in some higher difference.

It is therefore assumed that the degree of differencing d_1 necessary to achieve stationarity has been reached when the autocorrelation function of $W_t = \nabla^d Y_t$ die out fairly quickly.

3.4.24 Identifying the Resulted Stationary ARMA process

The autocorrelation function of an autoregressive process of order p tails off, its partial autocorrelation function has a cut off after lag p. the autocorrelation function of a moving average process of order q cuts off after lag q and its partial autocorrelation tails off.

Furthermore the autocorrelation function for a mixed process, containing a pth order autoregressive component and qth order moving average components, is a mixture of exponentials and damped sine waves after the first q-p lags conversely, the partial autocorrelation function for a mixed process is dominated by a mixture of exponentials and damped sine waves after the first p-q lags.

3.4.25 Akaike's Information Criteria (AIC)

The AIC which was proposed by Akaike uses the maximum likelihood method [7]. In the implementation of the approach, a range of potential ARMA models are estimated by maximum likelihood method, and for each the AIC is calculated, given by

$$AIC(p,q) = \frac{-2\ln(maximum\ likelihood)+2r}{N}$$

$$AIC(p.q) = \ln(\sigma_e^2) + r\frac{2}{n} + constant$$

Where n is the sample size or the number of observation in the historical time series data σ_e^2 is the maximum likelihood estimate of σ_e^2, and it is the residual or shock variance,$r = p + q + 1$, denotes the number of parameters estimated in the model.

Given two or more competing models the one with the smaller AIC value will be selected.

3.4.26 Schwarz's Bayesian Information Criterion (BIC)

Schwarz's BIC like the AIC uses the maximum likelihood method. It is given by

$$BIC(p,q) = \ln(\hat{\sigma}_e^2) + r\frac{\ln(n)}{n},$$

Where $\hat{\sigma}_e^2$ is the maximum likelihood estimate of σ_e^2 , $r = p + q + 1$, denotes the number of parameters estimated in the model, including a constant term and n is the sample size or the number of observations in the time series data. The BIC imposes a greater penalty for the number of estimated model parameters than does AIC.

Use of minimum BIC for model selection results in a chosen model whose number of parameters is less than that chosen under AIC.

One disadvantage of the information criteria approach is the enormous work involved in computing the maximum likelihood estimates of several models which is time consuming and expensive.

However this problem gas been overcome by the introduction of computers since there are software which compute several of these information criteria values. Information criteria are useful tools in model selection. They should not, however, be substituted for the careful examination of the autocorrelation and partial autocorrelation functions.

3.4.27 Estimation of the Parameters of the Model Identified

Once a model is identified the next stage of the Box-Jenkins approach is to estimate the parameters[9]. In this study the estimation of the parameters was done using a statistical package called the Statistical Package for Social Scientists (SPSS).

3.4.28 Testing the Model for Adequacy

After identification an appropriate model for a time series data, it is very important to check that the model is adequate. The error terms e_t are examined and for the model to be adequate the errors should be random. If the error terms are statistically different from zero, the model is not adequate.

The test statistic is the Q –statistic.

$$Q = n(n+2)\sum_{i=1}^{k}\frac{r_i^2}{n-i},$$

Which is approximately distributed as a χ^2 with $k-p-q$ degrees of freedom, where n is the length of the times series , k is the first k autocorrelations being checked , p is the order of the AR process and q is the order of the MA process, and r is the estimated autocorrelation coefficient of the ith residual term.

If the calculated value of Q is greater than χ^2 for $k-p-q$ degrees of freedom, then the model is considered inadequate and adequate if Q is less than χ^2 for $k-p-q$ degrees of freedom.

If the model is tested inadequate then the forecaster should select an alternative model and test for the adequacy of the model.

3.4.29 Forecasting

The fourth stage of the Box-Jenkins approach is to forecast [9] with model selected. Suppose the model chosen to fit a hypothetical data is

$$Y_t = Y_{t-1} + \alpha_1(Y_{t-1} - Y_{t-2}) + e_t$$

And suppose further that the data is of length 60, $\alpha = 0.2178$

$$Y_{60} = 131.2, \qquad Y_{59} = 134.8$$

Then $Y_{61} = Y_{60} + 0.2178(Y_{60} - Y_{59})$

$$Y_{61} = 131.2 + 0.2178(131.2 - 134.8)$$

$$Y_{61} = 130.097$$

Hence, a forecast value for period 61 is 130.097.

3.5 INTERRUPTED TIME SERIES

3.5.1 Introduction

A common research question in time series analysis is whether an outside event affected subsequent observations. For example did the implementation of a new economic policy improve economic performance; did a new anti-crime law affect subsequent crime rates; and so on. In general, we would like to evaluate the impact of one or more discrete events on the values in the time series.

3.5.2 Definition

Given a time series data $Y_1, Y_2, \dots, Y_N, Y_{N+1}, Y_{N+2}, \dots, Y_{N+k}$.

If a control measure or an intervention was introduced at time N+1 then the given time series data $Y_1, Y_2, \dots, Y_{N+k}$ is an interrupted time series.

For example in 1985 the government of Ghana with the help of the World Bank started vaccinating children against measles. Hence a time series data on the incidence of measles from 1980 to date may be considered as an interrupted time series data.

3.5.3 The Interrupted Time Series Experiment

The interrupted time series experiment seeks to measure the effectiveness of a control measure on the intervention. The experiment could be one which seeks to control a system or eradicate a phenomenon like that of the measles programme started in 1985 or for example increase the National Output by increasing wages. In case of the measles the measure is the vaccination of the National Output is the increase in wages.

(Box and Tiao, 1966)

3.5.4 Measurement of the intervention Effect

The procedure used in this work is to fit an AR(p) an autoregressive model of order p to the interrupted time series data using the Box-Jenkins methods of fitting a model to a time series data. The next step is to use the least squares method to estimate the parameters and statistical methods to assess the effectiveness of the intervention. (McDowall, et al)

Let us consider an imaginary interrupted time series data which can be fitted with a stationary AR(1) model which has a zero mean.

$Y_t = \alpha_1 Y_{t-1} + e_t$ where α_1 is the AR(1) parameter and e_t is the white noise. Let us assume further that we have n_1 data points before intervention and n_2 points after intervention and $n_1 + n_2 = N$. Suppose we assume that the effect is to add γ to the mean level.

The data can be expressed as follows

$$Y_2 = \alpha_1 Y_1 + e_2$$

$$Y_3 = \alpha_1 Y_2 + e_t$$

$$\vdots$$

$$Y_n = \alpha_1 Y_{n-1} + e_{n-1}$$

$$Y_{n+1} = \alpha_1 Y_n + e_n + \gamma$$

$$\vdots$$

$$Y_N = \alpha_1 Y_{N-1} + e_{N-1} + \gamma$$

This model can be written in matrix notation as

$$Y = X\beta + E$$

Where

$$X = \begin{bmatrix} 0 & Y_1 \\ 0 & Y_2 \\ . & Y_{n-1} \\ . & Y_n \\ 1 & . \\ 1 & . \\ 1 & Y_{N-1} \end{bmatrix} \qquad Y = \begin{bmatrix} Y_2 \\ Y_3 \\ . \\ . \\ . \\ Y_n \end{bmatrix} \qquad \beta = \begin{bmatrix} \gamma \\ \alpha \end{bmatrix} \quad \text{and} \quad E = \begin{bmatrix} e_2 \\ e_3 \\ . \\ . \\ . \\ . \\ e_N \end{bmatrix}$$

This has the least squares solution as

$\beta = \begin{bmatrix} \gamma \\ \alpha_1 \end{bmatrix} = (X^T X)^{-1} X^T X.$ Where X^T is the transpose of$X$.

In this case, it is easy to show that

$.X^T X = \begin{bmatrix} n_2 & \sum_{n_1}^{N-1} Y_t \\ \sum_{n_1}^{N} Y_t & \sum_1^{N-1} Y_t^2 \end{bmatrix}$ $\qquad X^T Y = \begin{bmatrix} \sum_1^N Y_t \\ \sum_2^N Y_t Y_{t-1} \end{bmatrix}$

If the first element (first row, first column) of $(X^T X)^{-1}$ is denoted by C, it can also be shown that an asymptotic standard normal $[N(0,1)]$ tests can be derived for δ under the null hypothesis that $\delta = 0$ for small samples.

The following is a statistic with an approximate t distribution

$t_{N-3} = \frac{\gamma}{(Se)\sqrt{c}}$ where Se is the square root of the residual variance, computed as

$$S_e^2 = \frac{1}{N-3} (Y - X\beta)^T (Y - X\beta).$$

3.5.5 Extension of the Procedure to AR(p) Models

Assume that n_1 data points before intervention consist of a stationary stochastic component, which is fitted with an autoregressive model, plus a linear trend. Thus before intervention $(t = p+1, p+2, \dots, n_1)$ the data can be represented as

$$Y_t = m_1 t + b_1 + \sum_{i=1}^{p} \alpha_i Y_{t-1} + e_t$$

After intervention $t = n_{t+1}, \dots, N$ and assume $n_2 = N - n$ data points move to a new asymptotic trend line. It is further assumed that the autoregressive parameters have not changed as a result of the intervention. Thus after the intervention the data can be represented as

$$Y_t = m_2 t + b_2 + \sum_{i=1}^{p} \alpha_i Y_{t-1} + e_t$$

The next stage is to estimate the parameters m_1, b_1, α_1, m_2, and b_2 one proceeds to find whether there has been significant changes in the values of m_1 and b_1 as reflected in the values of m_2 and b_2 which will be used to test whether the intervention was successful.

$$\beta = \begin{pmatrix} b_1 \\ m_1 \\ b_2 \\ m_2 \\ \alpha_1 \\ \alpha_2 \\ . \\ . \\ . \\ \alpha_p \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ . \\ . \\ . \\ \beta_{p+4} \end{pmatrix} \qquad E = \begin{pmatrix} e_{p+1} \\ e_{p+2} \\ . \\ . \\ . \\ . \\ e_N \end{pmatrix}$$

As usual the least squares estimates are

$$\beta = (X^T X)^{-1} X^T Y$$

If the estimate of σ_e^2 is denoted by S_e^2 which $\left(\frac{1}{v}\right)(Y - X\beta)^T(Y - X\beta)$

Where v, the degrees of freedom for errors is $N - 2p - 4$ and denote the C as the diagonal of $(X^T X)^{-1}$, then each of the parameters in $\beta_i = (b_1, m_1, b_2, m_2, \alpha_1, \ \dots, \alpha_p)$ can be referred to a t-distribution with v degrees of freedom, where

$$T = \frac{\beta_1}{S_e\sqrt{C}}$$

The data points do not actually lie on the lines $m_1 t + b_1$ before and $m_2 t + b_2$ after intervention. Rather, before intervention the data follow a steady-state trend line of the form $B_1 + M_1 t$ and approach $B_2 + M_2 t$ after intervention.

3.5.6 Steady-state solutions

Suppose that $Y_t = \sum \alpha Y_{t-1} + mt + e_t$

Then the expected value of Y_t is $E(Y_t) = \sum \alpha E(Y_{t-1}) + mt + b$

To find the steady-state solution of this difference equation in $E(Y_t)$, we assume

$$M_t + B = \sum \alpha(M(Y_{t-1}) + B) + mt + b$$

Equating coefficients of t and constant terms we have

$Mt = \left(\sum \alpha_i\right) Mt + mt$

$B = \sum \alpha_i(-i)M + (\sum \alpha_i) + mt + b$

So that $M = \frac{m}{1-\sum \alpha_i}$ and $M = \frac{b-M(\sum \alpha_i)}{1-\sum \alpha_i}$

For example for an AR(p) model with n_1 data points before intervention and n_2 data points after intervention where $n_1 + n_2 = N$, the model becomes

$$Y_t = m_1 t + b_1 + \sum_{i=1}^{p} \alpha_i Y_{t-1} + e_t$$

Before intervention and

$$Y_t = m_2 t + b_2 + \sum_{i=1}^{p} \alpha_i Y_{t-1} + e_t$$

After intervention.

The matrix notation is follows

$$Y = \begin{pmatrix} Y_3 \\ Y_4 \\ Y_{N1-2} \\ . \\ . \\ Y_N \end{pmatrix} \qquad X = \begin{pmatrix} 1 & 1 & 0 & 0 & Y_2 & Y_1 \\ 1 & 2 & 0 & 0 & Y_3 & Y_2 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & N1 & 0 & 0 & Y_{N1-1} & Y_{n1-2} \\ 0 & 0 & 1 & 1 & Y_{N1} & Y_{N1-1} \\ 0 & 0 & 2 & 2 & . & Y_{n1} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 1 & n_2 & Y_{N-1} & Y_{N-2} \end{pmatrix}$$

$$\beta = \begin{bmatrix} M_1 \\ B_1 \\ M_2 \\ B_2 \\ \alpha_1 \\ \alpha \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \end{bmatrix} \qquad E = \begin{bmatrix} e_3 \\ e_4 \\ . \\ . \\ . \\ e_N \end{bmatrix}$$

And

$\beta = (X^T X)^{-1} X^T Y$ and $T = \frac{\beta_1}{S_e \sqrt{C}}$

3.5.7 Sample size Requirements

It has been found that sample size requirements for the t and f distributions in the case of the interrupted time series analysis put no bounds on the sample size requirements, even cases where $n_1 = n_2 = 5$ tend to be a bit below the nominal levels ($\alpha = 0.05$ or $\alpha = 0.01$) and as sample size increases beyond $n_1 = n_2 = 20$. The significance level slowly converges to the nominal value.

However if the significance levels are considerably larger than the nominal values for $n_1 = n_2 = 10$ or $n_1 = n_2 = 20$, sample sizes of $n_1 = 60$ or even $n_1 = n_2 = 100$ may not yield satisfactory results.

The factor that plays an important role in determining the validity of the t and f distributions is the amount of positive autocorrelation present. For AR(1) models the significance test becomes questionable when α_1 is too large, with the dividing line somewhere between $\alpha_1 = 0.5$ and $\alpha_1 = 0.075$.

For the AR(2) models, near the edge of the stationarity reign defined by $\alpha_1 + \alpha_2 = 1$ the behavior is very bad. The worst cases are those for which $\alpha_1 + \alpha_2 = 9$, with $\alpha_1 + \alpha_2 = 0.6$. Following a study of the four cases for which $\alpha_1 + \alpha_2 = 0.6$ gives the worst when

$\alpha_1 = 0$ and $\alpha_2 = 0.6$, followed in order of increasing accuracy by $\alpha_1 = 0.3$, $\alpha_2 = 0.3$, $\alpha_1 = 0.6$, $\alpha_2 = 0$ and $\alpha_1 = 1.2$ and $\alpha_2 = -0.6$. The last case gives very reasonable results.

The negative correlation two steps in the past can thus counter balance a large positive correlation one step back.

The AR(3) models yield the same general results, $\alpha_1 = 0.3$, $\alpha_2 = 0.2$, $\alpha_3 = 0.2$ the case $(\alpha_1 + \alpha_2 + \alpha_3) = 0.7$ is the worst as would be expected with three positive autocorrelations.

The , $\alpha_1 = -0.3$, $\alpha_2 = 0.2$, $\alpha_3 = 2.$ case is worse than the , $\alpha_1 = 0.3$, $\alpha_2 = 0.2$, $\alpha_3 = -0.2$ case because in the former the positive autocorrelation extends further into the past. The factor that controls the accuracy of the proposed interrupted Time series experiment is the design; over fitting and the desired α level have a relatively small influence. But one has to view with caution the cases for which the sum of the autoregressive coefficients exceeds 0.6. Interrupted time series experiment analysis with autoregressive models can be used with confidence even in small sample sizes.

3.5.8 Testing for the significance of the Intervention

Here we test:

$H_0: \; m_1 = m_2 \,, \quad b_1 = b_2$ (Intervention ineffective)

Against:

$H_0: \; m_1 \neq m_2 \,, \quad b_1 = b_2$ (Intervention effective)

Let SS_0 denote the residual error sum of squares in the reduced model.

$$Y_t = m_1 + b + \sum_{i=1}^{p} \alpha_i Y_{t-1} + e_t \qquad for\ all\ t$$

And let SS_0 denote the residual sum of squares in the full model

$$Y_t = m_1 t + b_1 + \sum_{i=1}^{p} \alpha_i Y_{t-1} + e_t \quad t = p+1, \qquad \dots, n_1$$

And

$$Y_t = m_1 t + b_1 + \sum_{i=1}^{p} \alpha_i Y_{t-1} + e_t \quad t = n_1, \qquad \dots, N$$

Then under the null hypothesis,

$$F = \frac{(SS_0 - SS_1)/_2}{SS_1/_v}$$

Has an F(1,v) distribution. Here v, the error degrees of freedom is equal to the number of observations minus the number of "start up" observations (2 in AR(2) model or 3 in AR(3) minus the number of parameters fit, that is 6 in AR(2) and 4 in AR(1).

3.6 INFLATION

Inflation is a general rise in prices across the economy. This is distinct from a rise in the prices of a particular good or services. Individual prices rise and fall all the time in a market economy, reflecting consumer choices and preferences, and changing costs.

If the prices of item increase because demand for them is high, we do not think of this as inflation. Inflation occurs when most prices are rising by some degree across the whole economy.

3.6.1 Measurement of inflation

Inflation is the change in the price level from one year to the next. The change in inflation can be calculated by using whatever price index is most applicable to the given situation. The two most common price indices used in calculating inflation are CPI (Consumer price index) and the GDP (Gross Domestic Product) deflator. The inflation rates derived from different price indices will themselves be different.

3.6.2 Computing a Price Index

A measure of inflation in use today is the Consumer Prices Index (CPI) which forms the basis of monetary policy. It measures the prices of products and services that consumers buy.

When dealing with a large number of goods, some of whose prices have gone up faster than others and some may have even fallen, we pick a representative selection called a "basket" of goods and services and compare the costs of that "basket" over time. When we do this, we obtain a price index, which is defined as the cost of representing basket of goods today, expressed as a percentage of the cost of the same basket of goods of starting year, or base year. In other words,

$$Consumer\ price\ index = \frac{Cost\ of\ basket\ today}{Cost\ of\ basket\ in\ base\ year} * 100$$

3.6.3 Calculating Inflation Using GDP Deflator

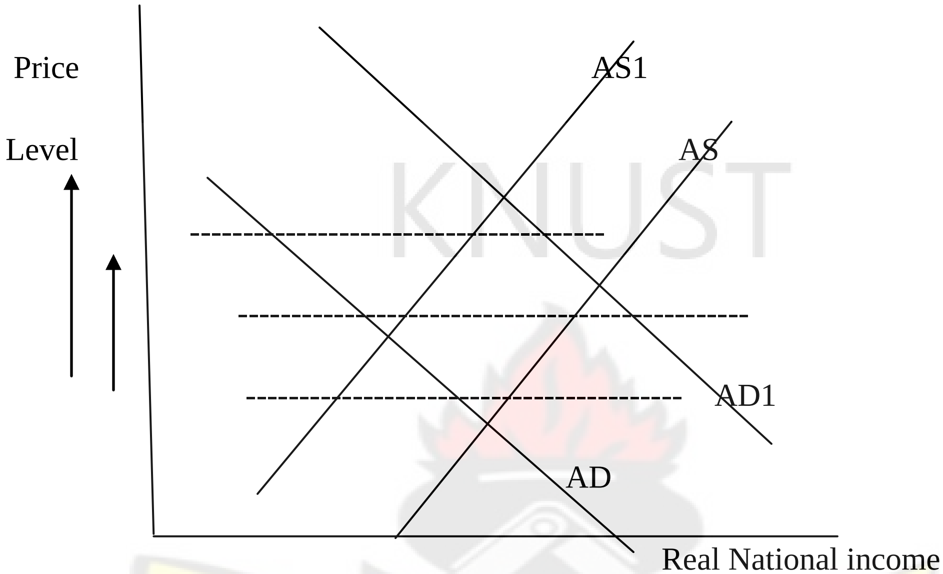
The other major price index used to determine the price level is the GDP deflator, a price index that shows how much of the change in the GDP from a base year is reliant on change in the price level. The GDP deflator is calculated by dividing the nominal GDP by the real GDP minus 1. That is ,

$$GDP\ deflator = \frac{Norminal\ GDP}{real\ GDP} - 1$$

Where real GDP is real gross domestic product is the sum value of goods and services produced in a country and valued at constant prices, calibrated from some base year and nominal GDP is nominal gross domestic product is the sum value of goods and services produced in a country and valued at current prices.

3.6.4 Causes of inflation

The causes of inflation are easiest explained by looking at aggregate supply and demand as shown by the diagram below, any change in either AS or AD will cause a change in the price level.



If aggregate demand increases to AD1 or aggregate supply decreases to AS2, the price level increases- this is inflation. If both increase together the inflation is even worse. If the inflation is cause by an increase in demand, then is it known as demand-pull inflation. The growth in demand literally pulls up prices. However, if the inflation is caused by a change in aggregate supply, then is it usually known as cost-pull inflation.

In practice, the two are often linked together as increases in demand may cause labor shortages, which in turns push up wages. Firms, who have to pay the higher wages, are then forced to put their prices up to maintain their margins. It also important to look at the role of the amount of money in the economy. The quantity theory of money shows how increased growth in the money can cause inflation. This happens because the extra money boosts the level of demand, and so causes demand-pull inflation.

when you consider that people build their expectation of inflation into their wage claim, you can see that this in itself can be a cause of inflation. If you expect inflation to be 5%, you may reasonably expect a wage rise in excess of this. If you manage to get that wage increase, then that may cause further cost-push inflation as firms are then facing higher costs. The higher inflation may then raise people's expectations further. a vicious circle. Expectations can be a bit like a self-fulfilling prophecy! Higher expectations can actually cause higher inflation.

3.6.5 Effects of Inflation

There are two general categories of effects due to inflation. The first groups of effects are caused by expected inflation. That is, these effects are a result of the inflation that economists and consumers plan on year to year. The second of effects are caused by unexpected inflation. These effects are a result of inflation above and beyond what was expected by economists and consumers. In general, the effects of unexpected inflation are much more harmful that the effects of expected inflation.

3.6.6 Cure of Inflation

The most appropriate cure for inflation depends on what the cause is. A considerable amount of research has gone into the causes of inflation, and a much more sophisticated understanding has emerged. So what cures are there? As we have said the cure has to relate to the cause. If inflation is caused by demand growing faster than the economy can cope with (demand-pull inflation). We try to control the level of demand. Inflation is caused by a lack of capacity or by costs rising in the economy, and then supply side solutions may be required. If inflation is caused by excessive monetary growth, then it will be most appropriate to put in place policies to control the level of money supply growth. In practice, all of these will be appropriate to a greater or lesser extent, indeed they are inter-linked and cannot be looked at in isolation.

In practice the most emphasis generally goes into looking at the level of demand. Is it growing faster than supply? If so, do we need to slow the economy down a little? What causes changes in demand? How do people react to interest rates changes? Other factors like the capacity of the economy are very important. The Government may have to deal with these also. In addition, supply side policies tend to be rather long term in nature whereas demand can be influenced more quickly-thus in economic terms.

CHAPTER FOUR

DATA ANALYSIS AND MODELLING

4.0 INTRODUCTION

In this chapter, I systematically go through the interrupted time series analysis described in chapter two to find out whether the intervention has been effective or not effective.

4.1 PRELIMINARY ANALYSIS OF DATA

The figure (Fig. 4.1.1) below show the trajectory of monthly inflation rates in Ghana. From the figure the rate of inflation drops drastically from January 1996 to November 2006 and then gradually to its minimum (about 9.4%) occurring in August 2006. The rate of inflation then rises steadily to 41.9% in June 2002. From July 2002 to December 2003, there was a decline again in the rate of inflation. From April 2003 to July 2003 there has been a quick rise in the inflationary value to relative maximum of 30% and again declines to 14.8 in Dec. 2006.

In general, there has been a downward trend in the rates of inflation over the period under consideration.

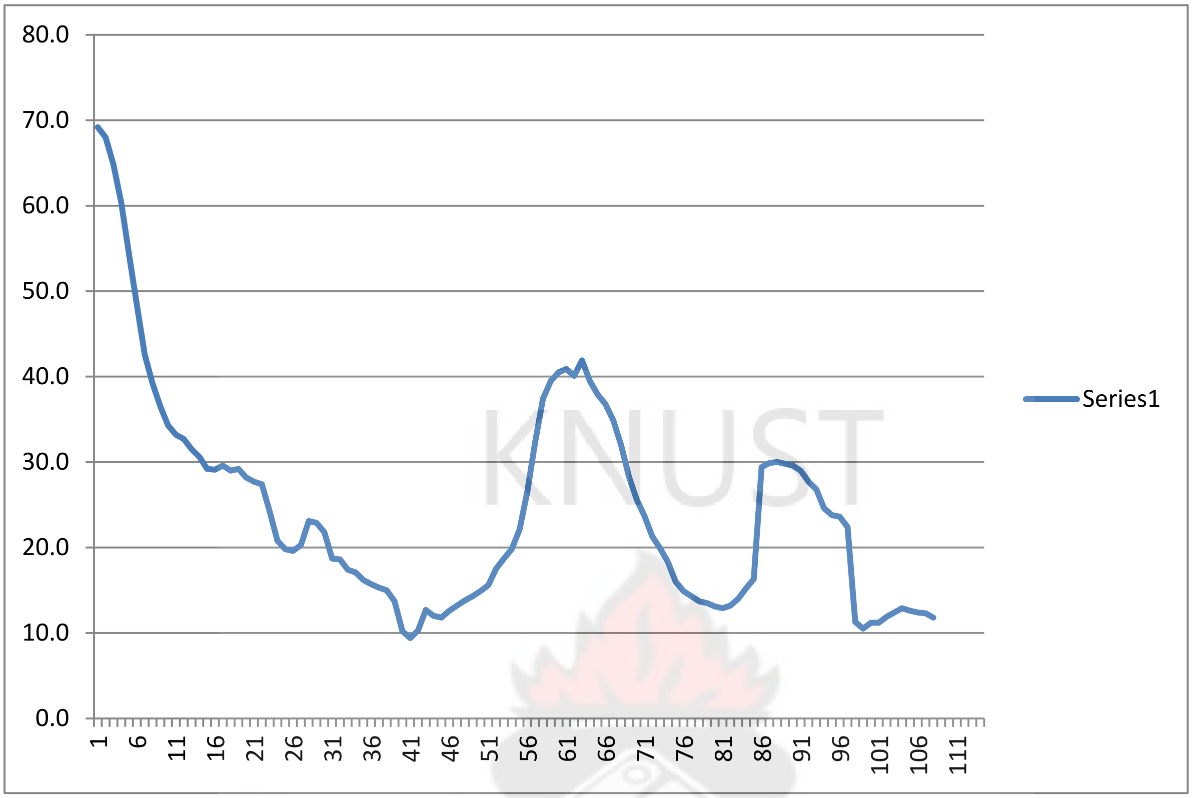


Fig. 4.1.1 Monthly inflation rates in Ghana from Jan. 1996 to Dec. 2006

Table 4.1.2 below displays the means and standard deviations of the pre-intervention data, post-intervention data, and the entire data.

Table 4.1.2 Descriptive statistics

TYPE OF DATA	MEAN	STANDARD DEVIATION
PRE-INTERVENTION	27.1063	15.4752
POST-INTERVENTION	21.3250	9.3326
ENTIRE DATA	23.6375	12.4354

4.2 MODEL IDENTIFICATION AND DIAGNOSTIC TESTING

We implement the Box-Jenkins method of identification in modeling the time series. This was done using Statistical Package for Social Scientist (SPSS).

4.2.1 Adequacy Test for an ARIMA(p,d,q) Model

For any ARIMA(p,d,q) model, the χ^2-distribution can be used to test for the adequacy of the model. The Q-statistics is distributed as χ^2_{k-p-q} where k=24 (maximum lag) used for Q, p is the order of the AR process and q is the order of the MA process. For example, ARIMA(1,0,0) is distributed as $\chi^2_{24-1-0} = \chi^2_{23}$.

In the table below, we display the critical values for some ARIMA models.

Table 4.3.1 χ^2 −DISTRIBUTION

DISTRIBUTION	SIGNIFICANCE LEVEL	CRITICAL VALUE
χ^2_{23}	0.05	35.172
χ^2_{22}	0.05	33.924
χ^2_{21}	0.05	32.671
χ^2_{20}	0.05	31.410

4.2.2 Identification of the Model

Observations from Fig.'s 4.2.1 and 4.2.2 show that the autocorrelation function trails down whiles the partial autocorrelation function truncates after the first lag. These gives an indication of an AR(1) process which is theoretically represented as

$$Y_t = \alpha Y_{t-1} + e_t$$

Fig 3.2.1 and 3.2.2 shows the graph of the autocorrelation and partial autocorrelation functions respectively.

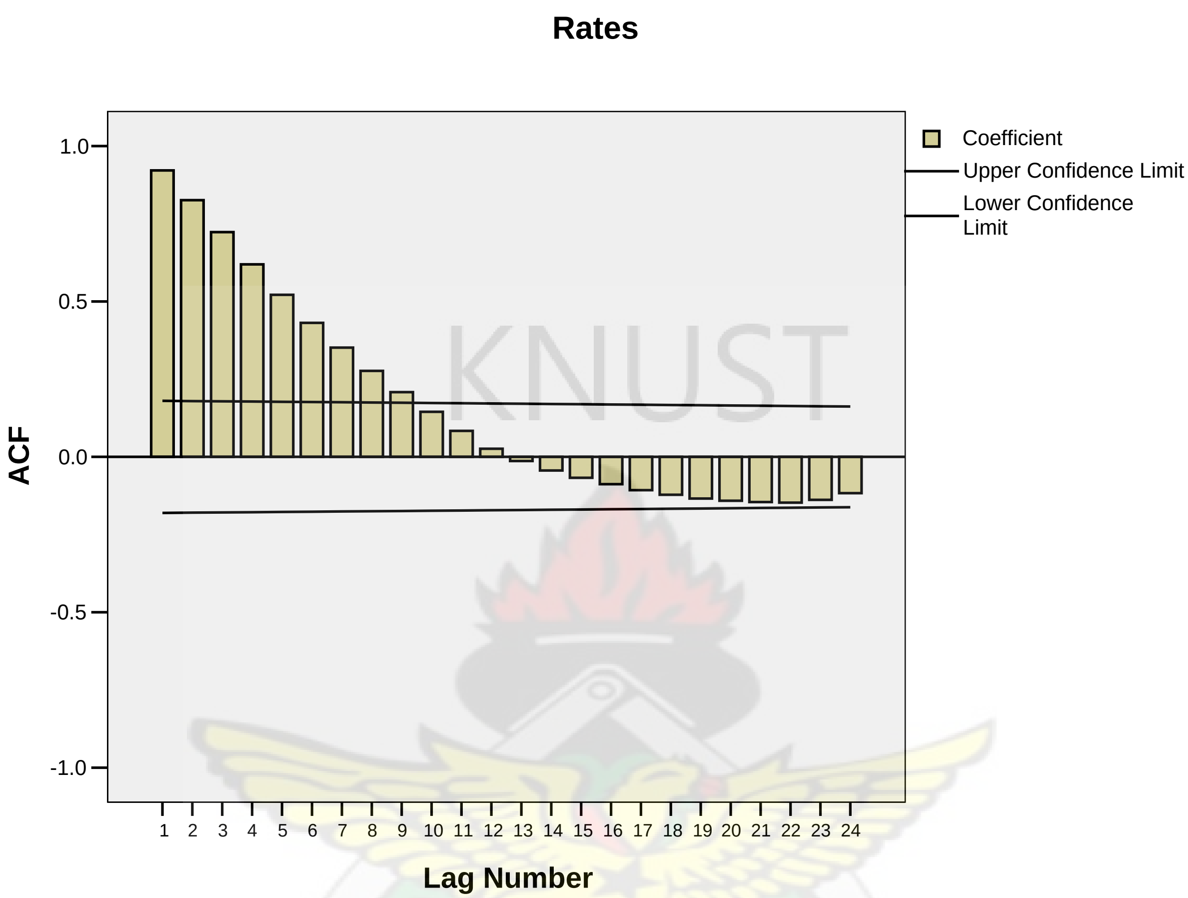


Fig 4.2.1 ACF of monthly inflation rate in Ghana from Jan. 96 to Dec. 06

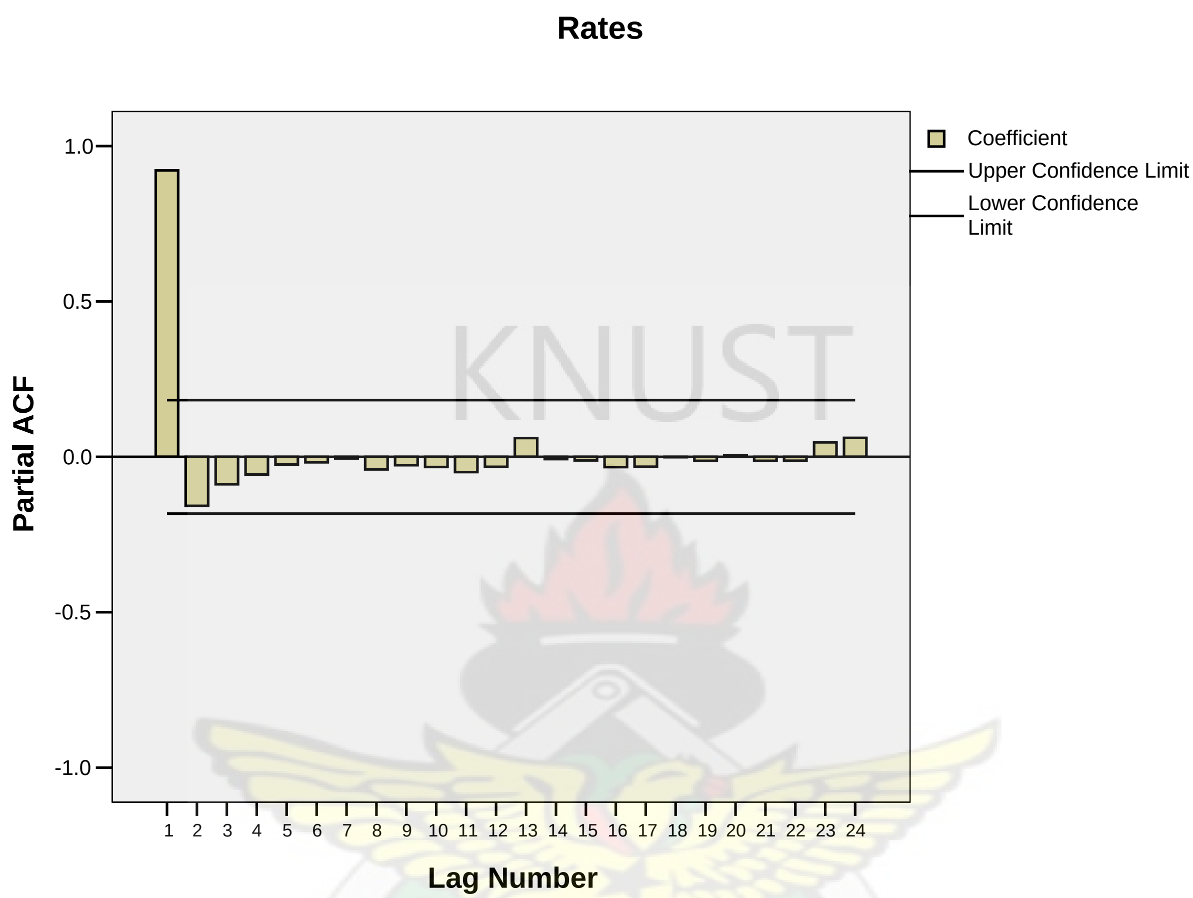


Fig 4.2.2 PACF of monthly inflation rate in Ghana from Jan. 96 to Dec. 06

Table 4.2.1 Analysis of the time series data

ARIMA MODEL	RESIDUAL VARIANCE	AIC	Q-VALUE
(1,0,0)	6.197	566.019	16.780

Since the interrupted time series analysis look out for an ARIMA(p.0.0), the ARIMA(1,0,0) is the best model for the time series data and it is also adequate since the Q- value is less than the critical value. It has the form

$$Y_t = 0.995Y_{t-1} + 37.657$$

4.3.1 TEST FOR SIGNIFICANCE OF DIFFERENCE BETWEEN THE MEANS OF THE PRE-INTERVENTION DATA AND THE POST-INTERVENTION DATA

Let μ_1 *be the sample mean of the pre − intervention data*

μ_2 *be the sample mean of the post − intervention data*

${\sigma_1}^2$ *be the sample variance of the pre − intervention data*

${\sigma_2}^2$ *be the sample variance of the post − intervention data*

$$\mu_1 = 27.1063 \qquad \mu_2 = 21.3250$$

$$n_1 = 48 \qquad n_2 = 72$$

$${\sigma_1}^2 = 239.482 \qquad {\sigma_2}^2 = 87.086$$

$$\sigma_p{}^2 = 147.786 \qquad\qquad \sigma_p = 12.158$$

Where σ_p^2 is the pooled variance of the population which is given by

$$\sigma_p^2 = \frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}$$

HYPOTHESIS TESTING

$H_0 : \mu_1 = \mu_2$ (Intervention not effective)

$H_1 : \mu_1 \neq \mu_2$ (Intervention effective)

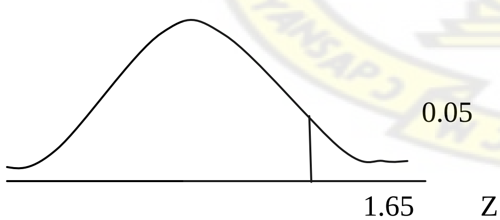
LEVEL OF SIGNIFICANCE

$\alpha = 0.05$

TEST STATISTICS

$$Z = \frac{(\mu_1 - \mu_2)}{\sigma_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$$

CRITICAL REGION (CR)



DECISION RULE

$$If\ Z \geq 1.65,\ reject\ H_0\ \ and\ accept\ H_1$$

$$If\ Z < 1.65, accept\ H_0\ and\ reject\ H_1$$

CALCULATION

$$Z = \frac{(27.1063 - 21.3250)}{12.158\sqrt{\frac{1}{48} + \frac{1}{72}}} = 2.552$$

CONCLUSION

Since Z=2.552 > 1.65 we reject H_0 and accept H_1 and conclude that there is enough evidence at the 5% level of significance of a decrease in the mean level of the inflation in Ghana after the intervention policy in 2001.

4.3.2 USE OF REGRESSION ANALYSIS TO MODEL TREND AND AUTOREGRESSIVE COMPONENTS

We attempt to apply the interrupted time series analysis described in chapter two to estimates b_1, b_2, m_1, m_2, α. Here b_1 and b_2 are the intercepts and m_1 and m_2 are the slopes before and after intervention while α is the AR parameters. This is done by use of SPSS for the regression analysis.

The results for the full and reduced models are displayed in the tables below.

TABLE 4.3.2.1 Variables in the equation of full model

VARIABLE	ESTIMATES	STD. ERROR	95% CI LOWER	95%CI UPPER	T-VALUE	SIGNIFICANCE
B_1	1.230	1.529	-1.800	4.260	0.804	0.423
M_1	-.004	0.035	-0.074	0.066	-0.117	0.907
B_2	3.097	0.961	1.193	5.001	3.222	0.002
M_2	-.036	0.014	-0.064	-0.008	-2.532	0.013
α	.916	0.027	0.000	0.970	34.002	0.000

Table 4.3.2.2 Model efficiency of full model

R	R-SQUARE	ADJUSTED R	STD. ERROR
0.997	0.993	0.993	2.19774

Table 4.3.2.3 ANOVA of full model

MODEL	DF	SSS	MSE	F	SIG. F
REGRESSION	5	80110.663	16022.133	3317.167	0.000
RESIDUAL	114	550.627	4.830		

Table 4.3.2.4 Variables in the equation of reduced model

VARIABLE	ESTIMATES	STD. ERROR	95% CI LOWER	95% CI UPPER	T-VALUE	SIG OF T
B	1.040	0.797	-0.539	2.619	1.304	0.195
M	0.003	0.007	-0.011	0.017	0.395	0.694
α	0.930	0.020	0.891	0.968	47.641	0.000

Table 4.3.2.5 Model efficiency of reduced model

R	R-SQUARE	ADJUSTED R-SQUARE	STD ERROR
0.996	0.993	0.992	2.27386

Table 4.3.2.6 ANOVA of reduced model

MODEL	DF	SSS	MSE	F	SIG. OF F
REGRESSION	3	80061.519	26687.173	5161.486	0.000
RESIDUAL	116	599.771	5.170		

Now from the table of both the full and the reduced models, we have

$$SS_1 = 550.627 \ and \ SS_0 = 599.771$$

$$F = \frac{(SS_0 - SS_1)/2}{SS_1/v}$$

$$F = \frac{(599.771 - 550.627)/2}{550.627/114}$$

$$F = 5.087$$

F (1, 114) at 0.05 level of significance that is $F_{1,114,0.05} = 3.92$.

The conclusion is that since the value of the F statistics is 5.087 which show a significant intervention effect that the intervention has been effective at 5% significance level.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1 SUMMARY OF FINDINGS

Comparing the mean values from Table 3.1.1, the mean value of the post-intervention data (21.325%) is less than that of the pre-intervention data. This means the average rate of inflation over the post-intervention period (21.325%) is less than the average rate of inflation of pre-intervention data (27.1063%). Similarly, the standard deviation of the post-intervention data (9.3326) is less than that of the pre-intervention data (15.4752).

From the analysis, AR(1) was observed to be the most preferred adequate model among other AR(p) models. The model is $Y_t = 0.995Y_{t-1} + 37.657$

5.2 SIGNIFICANCE TESTS

in this section we discuss the tests of Significance of the difference between the means of the pre and post intervention data.

5.3 TESTING THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN THE MEANS OF THE PRE AND POST INTERVENTION DATA

The significance test of the difference between the means of the pre and post intervention data was significant. There was enough evidence at the 5% level of significance that the mean before the intervention was greater than the mean after the intervention.

5.4 CONCLUSION

It was found out that the rate of inflation in Ghana can be fitted with an autoregressive model of order one, i.e. AR(1) model. From the results of the tests of the difference between the means before and after intervention, as well as the interrupted time series experiment, we conclude that the intervention has successfully reduced the rate of inflation in the nation.

5.5 RECOMMENDATION

1. Interrupted Time Series can be used as a tool for analyzing the effectiveness of government on inflation and other economic indicators.

2. We recommend that the Government continues with the tight monetary policy, Open Market Operations (OMO), Repurchase Agreements (Repos) and prime rate (interest Rate) policies that has been used since Jan 2001 to Dec 2006 in trying to reduce the rate of inflation since it was effective.

3. We also recommend a study be carried out to research the effectiveness of some other government interventions e.g. exchange rate depreciation, wages, exogenous shocks in the domestic food supply, petroleum prices etc.

REFERENCES

1. **Abraham, Bovas and Ledolter, Johannes,** (1983), Statistical Methods for Forecasting, John Wiley and Sons Press, N.Y.
2. **Bloomfield, Peter,** (1976), Fourier Analysis of Time Series: An Introduction. 1976. John Wiley and Sons.
3. **Box, George E.P., Jenkins, Gwlym M. and Reinsel**, **Gregory C,** (1994), Time Series Analysis: Forecasting and Control. Third Edition , Prentice Hall, Englewood Cliffs, New Jersey.
4. **Chatfield, C**, (1988), The Analysis of Time Series. An Introduction, Fourth Edition, Chapman and Hall. London.
5. **Dickey, David A. and Wayne A. Fuller,** (1979), "Distribution of the Estimators for Autoregressive Time Series with a Root." Journal of the American Statistical Association
6. **Engle Robert F.,** (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." Econometrica. 50,
7. **Gottman, M John**, (1981), Time Series Analysis, A Comprehensive Introduction for Social Scientists. Cambridge University Press.
8. **Greene, William H.,** (1990), Econometric Analysis (New York: Macmillan)
9. **Hardwick Philip, Khan Bahadur, Langmead John**, (1990), An Introduction to Modern Economics, Longman Group UK Ltd, 3rd Edition.
10. **Harvey Andrew C.,** (1990), The Econometric Analysis of Time Series, 2nd ed.(Oxford: Phillip Allan)
11. **Hoel, G Paul**, (1991), Introduction to mathematical Statistics, John Wiley and Sons, 4th Edition

12. **Jarrett, Jeffrey**, (1991), Business Forecasting methods, Second Edition, Basic Blackwell Ltd. Oxford.

13. **Kennedy, Peter E.,** (1992), A Guide to Econometrics, 3rd Edition (Cambridge, Massachusetts: The MIT Press)

14. **Kreyszig, Erwin**, (1993), Advance Engineering Mathematics. Seventh Edition, John Wiley and Sons Press, N.Y.

15. **Nsowaah Nuamah, N.N.N**, (2002), Lecture Notes on Regression Analysis.

16. **Maddala, G.S.,** (1988), Introduction to Econometrics (New York: Macmillan)

17. **Spiegel Murray R.**, (1974), Fourier Analysis with applications to Boundary Value Problems. (Schaum's Outline Series) McGraw-Hill Book Company.

18. **Akinori Baba, and Tamiki Komatsuzaki et al.,** (1995), Construction of effective free energy landscape from single-molecule time series, **(**PNAS**)** Proceedings of the National Academy of Sciences of the United States of America**.**

19. **Aston, John A.D., Findley, David F., McElroy, Tucker S., Wills, Kellie C., Martin, Donald E. K. et al.,** (2004), New ARIMA Models for Seasonal Time Series and Their Application to Seasonal Adjustment and Forecasting,

20. **Bloom Howard S. et al.,** (2003), Using "Short" Interrupted Time-Series Analysis To Measure The Impacts of Whole-School Reforms With Applications to a Study of Accelerated Schools Manpower Demonstration Research Corporation, Evaluation Review, Vol. 27, No. 1, 3-49

21. **Gillings D., Makuc D. and Siegel E. et al.,** (1981), Analysis of interrupted time series mortality trends: an example to evaluate regionalized perinatal care**,** American Journal of Public Health, Vol. 71, Issue 1 38-46,

22. **Gil-Alana, Luis A., Luis Alberiko Gil-Alana et al.,** Generalized Fractional Time Series Modelling of the Relationship between Consumption and Income in the UK,

23. **Gershenfeld, N. , Schoner, B. & Metois R. E. et al.,** (1999)**,** Cluster-weighted modelling for time-series analysis, Letters to Nature, *Nature*397, 329-332|doi:10.1038/16873

24. **Haipeng Shen and Jianhua Z. Huang et al.,** (2008), Forecasting time series of inhomogeneous Poisson processes with application to call center workforce management, Ann. Appl. Stat. Volume 2, Number 2

25. **Hay J. L. and Pettitt A. N. et al.,** (2001), Bayesian analysis of a time series of counts with covariates: an application to the control of an infectious disease, *Biostatistics* 2:433-444, Oxford University Press.

26. **Hongwei Guo, Yanchi Liu, Helan Liang, Xuedong Gao et al,** (2008), An Application on Time Series Clustering Based on Wavelet Decomposition and Denoising, Fourth International Conference Publication Date: 18-20 Oct. 2008, Volume: 5, on page(s): 419-422

27. **Kuehl, R. O., Buxton, D. R. and Briggs, R. E. et al.,** (1976), Application of Time Series Analysis to Investigate Crop and Environment Relationships, Published in Agron J 68:491-495 (1976), © 1976 American Society of Agronomy

28. **Leiderman P. Herbert and David Shapiro et al.,** (1962), Application of a Time Series Statistic to Physiology and Psychology, *Science* 12 October 1962, Vol. 138. no. 3537, pp. 141 – 142, DOI: 10.1126/science.138.3537.141

29. **Miika Ahdesmäki**, **Harri Lähdesmäki**, **Ron Pearson**, **Heikki Huttunen** and **Olli Yli-Harja et al.,** (2005), Robust detection of periodic time series measured from biological systems, *BMC Bioinformatics* 2005, **6:**117doi:10.1186/1471-2105-6-117

30. **Monge Sanz, B. M. and Medrano Marqués, N. J. et al.,** (1992), Total ozone time series analysis: a neural network model approach, , Electronics Design Group – GDE–, Faculty of Sciences, University of Zaragoza, E-50009 Zaragoza, Spain.

31. **Piotr Omenzetter and James Mark William Brownjohn et al.,** (2006), Application of time series analysis for bridge monitoring, PiotrOmenzetter*et al* 2006 *Smart Mater.Struct.***15** 129-138 doi: 10.1088/0964-1726/15/1/041

32. **Qiang Wang Kontos, D. Guo Li Megalooikonomou, V et al.,** Application of time series techniques to data mining and analysis of spatial patterns in 3D images,

33. **Rogers, Andrei, McNown, Robert, Knudsen, Christin et al.,** (1993), Forecasting fertility: an application of time series methods to parameterized model schedules, Publisher: Academic Press, Inc., Social Science Research, ISSN: 0049-089X,

34. **Russell E. Warren et al.,** (1993), Adaptive Kalman-Bucy filter for differential absorption lidar time series data, Applied Optics, Vol. 26, Issue 22, pp. 4755-4760 doi:10.1364/AO.26.004755

35. **Walker Chris, Galster George, Temkin Kenneth, Sawyer Noah et al.,** (2004), Measuring the impacts of community development initiatives: A new application of the adjusted interrupted time-series method, Publisher: Sage Publications, Inc., Evaluation Review, ISSN: 0193-841X,

36. **Yi, Kei-Mu, Kocherlakota, Narayana R. et al.,** (1996), A simple time series test of endogenous vs. exogenous growth models: an application to the United States, Publisher: MIT Press Journals, Review of Economics and Statistics, ISSN: 0034-6535.

APPENDIX

COMBINED DATA FOR THE ANALYIS

ENTIRE DURATION	PRE-INTERVENTION	POST-INTERVENTION
69.2	69.2	14.3
68	68	14.9
64.8	64.8	15.6
60.3	60.3	17.5
54.2	54.2	18.7
48.4	48.4	19.8
42.6	42.6	22.1
39.2	39.2	26.6
36.5	36.5	32.3
34.3	34.3	37.4
33.2	33.2	39.5
32.7	32.7	40.5
31.5	31.5	40.9
30.6	30.6	40.1
29.2	29.2	41.9
29.1	29.1	39.5
29.6	29.6	37.9
29	29	36.8
29.2	29.2	34.9
28.2	28.2	32
27.7	27.7	28.3
27.4	27.4	25.6
24.2	24.2	23.7
20.8	20.8	21.3
19.8	19.8	19.9
19.6	19.6	18.3
20.3	20.3	16
23.1	23.1	14.9
22.9	22.9	14.3
21.8	21.8	13.7
18.7	18.7	13.5
18.6	18.6	13.1
17.4	17.4	12.9
17.1	17.1	13.2
16.2	16.2	14
15.7	15.7	15.2
15.3	15.3	16.3
15	15	29.4
13.7	13.7	29.9
10.2	10.2	30
9.4	9.4	29.8
10.3	10.3	29.6

12.7	12.7	29
12	12	27.7
11.8	11.8	26.8
12.6	12.6	24.6
13.2	13.2	23.8
13.8	13.8	23.6
14.3		22.4
14.9		11.3
15.6		10.5
17.5		11.2
18.7		11.2
19.8		11.9
22.1		12.4
26.6		12.9
32.3		12.6
37.4		12.4
39.5		12.3
40.5		11.8
40.9		11.6
40.1		14
41.9		16.7
39.5		16.6
37.9		16.3
36.8		15.7
34.9		14.9
32		14.7
28.3		14.9
25.6		15.4
23.7		15.3
21.3		14.8
19.9		
18.3		
16		
14.9		
14.3		
13.7		
13.5		
13.1		
12.9		
13.2		
14		
15.2		
16.3		
29.4		
29.9		
30		
29.8		

29.6
29
27.7
26.8
24.6
23.8
23.6
22.4
11.3
10.5
11.2
11.2
11.9
12.4
12.9
12.6
12.4
12.3
11.8
11.6
14
16.7
16.6
16.3
15.7
14.9
14.7
14.9
15.4
15.3
14.8