KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

COLLEGE OF SCIENCE

Estimating the Impact of the Cocoa Hi-tech and Mass Spraying Programmes on Cocoa Production in Ghana: An application of Intervention Analysis

By

Omane - Adjepong Maurice (B.Ed Mathematics)

A Thesis submitted to the Department of Mathematics in partial fulfillment of the requirements for the award of the degree of

MASTER OF PHILOSOPHY
(APPLIED MATHEMATICS)

JUNE, 2012

DECLARATION

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

Omane-Adjepong Maurice	…..…………………….	……..……..…………
(Index No.: PG5071410)	Signature	Date

Certified by:

Dr. F. T. Oduro	…………………………	……...……….………....
(Supervisor)	Signature	Date

Certified by:

Mr. F.K Darkwah	…..………………………	..……………………..
(Head of Dept.)	Signature	Date

ACKNOWLEDGMENT

I warmly express my profound appreciation to my supervisor Dr. F.T Oduro for his objective analysis and valuable contributions in making this work a success.

Again, I specially wish to thank Professor (Asst.) James Monogan, a Political Science lecturer at the University of Chicago for assisting me with software codes to run my data. I am equally indebted to Dr. Z.K.M. Batsi, Dean of the Faculty of Applied Sciences at the Kumasi Polytechnic, for proofreading my scripts.

I cannot forget the motivation and support I had from Messrs Samuel Oduro of NHS, Scotland and Kofi Ababio Kicupson, a lecturer at the Garden City University College. Again, I thank all the Statistics lecturers at Kumasi Polytechnic for making their offices available for me to undertake this work.

Another thump up goes to the entire staff at the Monitoring, Research and Evaluation Department of Ghana COCOBOD, especially to the Deputy Director of the Department for providing me with annual Ghana's cocoa production figures. This thesis would not have been complete without the able assistant of Mr. K. Obeng Adjinah, the national coordinator of the Cocoa Diseases and Pests Control (CODAPEC) Unit. I would also extend a hand of appreciation to Dr. Ofori-Frimpong of the Cocoa Hi-technology programme for assisting me with adequate information about the programme.

Finally, to all the 2012 M.Phil Applied Mathematics colleagues, I say "ayekoo" to every one for a journey worth traveling.

DEDICATION

I wish to gratefully dedicate this thesis to the almighty God who by His grace and mercies granted me with knowledge and strength to undertake this academic exercise. I again express my sincere gratitude to my mum Mad. Christiana Tetteh, for her immense contributions towards my formal education and my general up-bring. Another appreciation also goes to a gallant grandmum, Mad. Elizabeth Abena Antwiwaah for her advice and support.

Finally, I dedicate this piece to my very good friend Miss Sonita Sarpong Senya, who has been with me throughout this difficult moment.

I say a big thank you to all of you.

ABSTRACT

Using the intervention time series analysis of Box and Tiao, this thesis presents an estimate of the level and nature of impact of the national mass spraying and the cocoa Hi-technology Government intervention programmes, implemented in 2002 and 2003 respectively, on cocoa production in Ghana. Annual time series cocoa production data covering the period from 1948 to 2011 was obtained from the Monitoring, Research and Evaluation Department of Ghana's COCOBOD. Results from the study indicate that the preintevention period could best be modeled with an AR(1) process. The effects of both intervention programmes were found to be abrupt and permanent. The impact of the mass spraying programme was estimated (at 95% confidence level) to have had a significant increase of 182,398.2 metric tonnes annually. It was also found that the cocoa Hi-technology programme significantly (at 95% confidence level) increased production by 266,515.1 metric tonnes per annum.

TABLE OF CONTENT

CHAPTER 3: METHODOLOGY

CHAPTER 4: ANALYSIS AND DISCUSSION OF RESULTS

CHAPTER 5: SUMMARY, CONCLUSION AND RECOMMENDATIONS

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

1.0 Background of the Study

Since independence, Agriculture has been the key player in Ghana's economic development and growth. Averagely, it accounts for about 40 percent of the country's gross domestic product (GDP) and generates 55 percent of the foreign exchange earnings. The agriculture sector today employs about 51 percent of the labour force in Ghana, and is the main source of income and sustainable employment for nearly 70 percent of Ghanaian rural folks. The cocoa industry alone employs close to about 60 percent of the national agricultural workforce in the country (Asuming-Brempong et al., 2006).

Undoubtedly, cocoa production is the predominant activity in the Ghanaian agricultural sector. During the 2002 season, cocoa made up for 22.4 percent of the total foreign exchange earnings for Ghana. This contributed sixty-three (63) percent of the entire foreign export earnings accrued from the agricultural sector. Also, the "impressive growth performance" of the country's economy in 2003 was mainly attributed to an exceptionally strong output growth in the Agricultural sector, which not only made up for the services sector but moved the overall growth rate up from a projected 4.7 percent to 5.2 percent. In turn, an examination of the sector's performance was exclusively on account of cocoa production and marketing, with output rising from a slight decline of 0.5 percent in 2002 to a sturdy performance of 16.4 percent growth, making a turnaround of nearly 17 percent. Currently, Ghana is rated as the world's leader in premium quality cocoa beans production, ahead of major competitor countries such as Cote d'lvoire and Indonesia (Centre for Policy Analysis, 2004).

Ghana's cocoa production has over the years faced major challenges which have adversely contributed to the country losing her position as the leading producer of cocoa beans in the world. The key challenges include; diseases and pests infestation, climate and soil quality, and the setting of cocoa producer price (Anim-Kwapong and Frimpong, 2005).

In order to boost annual production and the quality of yields, successive Governments since independence have introduced policies and programmes which sought to tackle these major challenges bedeviling the industry. For instance, the yearly setting of cocoa producer price has considerably seen a facelift for the past few decades. During this period, Government yearly reviews the producer price and also gives prompt bonuses to farmers in order to sustain their interest in the cultivation of the cash crop. Again, a nationwide spraying exercise and a fertilizer subsidy programme have, in recent times, been embarked upon to help curb the diseases and pests plaque, as well as the quality of soil in various farmlands. Reportedly, the two recent control measures have been successfully and widely embraced by the cocoa farmers as compared to an earlier policy of cutting out diseased trees.

This study, therefore, examines or evaluates the two recent control measures introduced by Government to boost up cocoa production in Ghana using a statistical technique called intervention time series analysis.

This chapter then describes the origin and development of Ghana's cocoa industry. It further states the statement of the problem and the objectives of the study. Also included in this chapter are: research methodology, significance of the study and organization of the thesis.

1.0.1 Origin and Development of Ghana's Cocoa Industry

Cocoa originated from the headwaters of the Amazon in South America. Large scale cultivation of cocoa was started by the Spanish in the 16th century in Central America. It later spread to British, French and Dutch West Indies in the 17th century, and continued to Brazil in the 18th century. The crop was brought from Brazil to Fernando Po and Sao Tome in 1840. History has it that cocoa was first planted in the coastal areas of the then Gold Coast by the Dutch Missionaries. In 1857, the Basel Missionaries were also recorded to have planted cocoa at Aburi in the Eastern Region. However, cocoa cultivation could not spread throughout the country until Tetteh Quarshie, a native of Osu in Accra, who had traveled to Fernando Po and worked there as a blacksmith, returned with Amelonado cocoa pods in 1879. He established a farm at Akwapim Mampong in the Eastern Region from where enthusiastic farmers bought pods to plant. This resulted in the spread of cocoa farming to other parts of the region. In order to supplement the pods being supplied from the farms of Tetteh Quarshie, Sir William Brandford Griffith, the then Governor of the Gold Coast, also arranged for cocoa pods to be brought from Sao Tome in 1886, from which seedlings were raised at Aburi Botanical Garden and distributed to farmers (Asuming-Brempong et al., 2006).

Commercial growing of cocoa in Ghana began as early as 1879 when the crop was reintroduced into the country by Tetteh Quarshie. Local farmers in the Gold Coast widely embraced the cultivation of the crop and worked assiduously to commercialize it. This is affirmed in the words of A.W.Knapp (1920), that "the seriousness with which the people of the Gold coast took to cocoa farming was said to be phenomenal as it is alleged to have shattered the stereotype image of the "indolent" native, and showed the world that the "natives" were capable of building a strong economy by their own initiative and industry".

The first cocoa beans were recorded to have left the shores of Ghana in 1891. Since then, the crop has remained the major export crop and a readily source of domestic revenue and foreign exchange to the nation. Through the dint of hard work, Ghana became the world's leading producer of cocoa. From 1910 to 1977, its market shares ranged from 30 to 40 percent. Production then increased from 36.3 metric tonnes in 1891 to a peak of 591,031 metric tonnes in the 1964/65 season, giving the country a global productivity share of about 33 percent (Adjinah and Opoku, 2010). In recognition of the immense contribution of cocoa to the economy, the government of Ghana in 1947 established the Ghana Cocoa Marketing Board (CMB), now called COCOBOD, as the main governmental agency responsible for the development of the cocoa industry in Ghana. Presently, Ghana has six major cocoa growing areas namely: Eastern, Central, Brong-Ahafo, Volta, Ashanti and Western regions, with the latter divided into two zones due to its vast nature of production.

Notwithstanding the successes chalked, cocoa production has, over the years, been weighed down with several challenges. The major challenges that have been short listed include diseases and pests attacks, climate and soil quality, and annual setting of producer price. To rescue the industry, the Government of Ghana in collaboration with Cocoa Board instituted a number of policy interventions to control these challenges. Some of these interventions could not hold for long whereas others were embraced by the farmers. For instance, the Akuafo checque payment system which was introduced in 1986 to pay farmers' produce collapsed a few years after its inception. The policy of cutting out diseased cocoa trees enacted by the colonial Government through recommendations by the Research Station at Tafo, had its share of hitches at the beginning when most farmers refused to cooperate with the Agriculture workers who were mandated to cut out all affected cocoa trees.

Recently, the Government through COCOBOD has initiated two intervention programmes geared towards reviving and booming the cocoa industry in terms of production. A nationwide cocoa diseases and pests control programme, commonly known as "Cocoa Mass Spraying" was started in 2001/2 cocoa season aimed at assisting farmers to fight the Black pod and Mirid/Capsid diseases. In all, seventy-two (72) cocoa growing districts are now benefiting from the Mass Spraying programme; thirty-five (35) districts were sprayed against the Mirid/Capsid disease only, twenty-one (21) districts against the Black pod disease only, and sixteen (16) districts benefited from both programmes. This has effectively put the threat from Black pod and Capsid under some level of control. The effects of soil nutrient deficiency and pests' attacks were also reduced by the timely introduction of the Cocoa Hi-technology programme in 2002/2003 cocoa season. Under this programme, farmers were supplied with packages of fungicides, pesticides and fertilizers to help increase their yields per a hectare of farm land. The fertilizers under the policy were supplied on credit basis by COCOBOD to beneficiary cocoa farmers in the initial stages. Payments were later made by the beneficiary farmers during the ensuing harvesting season in installments (COCOBOD, 2007).

1.1 Statement of the Problem

By 1930, after Ghana had been the leading producer of cocoa for about 20 years, cocoa production in the Eastern region, then the biggest cocoa producing region in Ghana was plaqued with pests and diseases, which caused production to decline fast. The situation called for policy interventions that could control the problems and arrest the declining production trend. In line with this, a research station was set up in 1935 at Tafo. The station was mandated to investigate the disease and pest problems of cocoa in the country and was also tasked to recommend pragmatic measures to maintain production levels.

In the early 1960's, the Government of Ghana introduced a mass spraying exercise to help curb the problem of diseases and pests infestation on cocoa farms in Ghana. Though the programme did not last long, it was said to have recorded high increases in cocoa yields. The then Ghana's highest production of 591,031 MT in 1964/65 was mainly attributed to this spraying exercise. Nevertheless, the incidence of diseases and pests infestation was reported to be very high in the Ashanti and Brong Ahafo regions in the early part of 1980's. This resulted in an estimated loss of crop yields of about 50 to 100 percent.

The Research Station at Tafo, now Cocoa Research Institute of Ghana (CRIG), was said to have identified the causal agent to be Phytophthora megakarya and recommended the use of fungicides in spraying the cocoa farms as a means of control.

Figure 1.1: Cocoa tree attacked by black pod disease

In a committed attempt to capture the decline in cocoa production levels, the Ghana Government through its COCOBOD instigated a nationwide cocoa disease and pests control programme, dubbed 'mass spraying' and the cocoa Hi-technology programme in 2001/2 and 2002/3 cocoa seasons respectively. These two programmes were purposely to assist all cocoa farmers to fight against diseases and pests plague, and also to improve soil quality. Since the inception of these two programmes, there have been varied assertions made about the impact of these policies on cocoa production in Ghana. Ghana Cocoa Board through its CODAPEC and Hi-tech programmes claims there has been dramatic turn of event towards increased production levels since the programmes were introduced about ten years ago. According to production figures reported by Asuming-Brempong et al., (2006), the mass spraying programme has had positive impact on national cocoa production, resulting in excess production of about 700,000 metric tonnes during the 2003/04 and 2005/06 cocoa seasons. Again, in the 2012 budget statement and economic policy, the Finance Minister mainly attributed the cocoa sub-sector's remarkable growth of 14 percent to policy interventions and the hard work of cocoa farmers.

However, it appears that no rigorous statistical study has been undertaken to assess the validity of these assertions. This study is therefore being embarked upon to primarily test the impact of these intervention programmes on Ghana's annual cocoa production, using intervention time series analysis originally developed by Box and Tiao (1975).

1.2 Objectives of the Study

The objectives of the study are to:

1. fit an intervention model to the annual cocoa production series from 1948 to 2011 using the Box-Jenkin-Tiao intervention modeling strategy.
2. estimate and assess the nature of impact of the Government's mass spraying and the Hi-technology intervention events on Ghana's annual cocoa production time series.
3. interprete the results of the fitted intervention model in relation to the effectiveness of such Government programmes.

1.3 Research Methodology

The secondary source data used for this study was acquired from the Monitoring, Research and Evaluation Department of Ghana COCOBOD. The data is made up of sixty-four (64) consecutive annual levels (tonnes) of cocoa production in Ghana from 1948 to 2011. The Box-Jenkins ARIMA modeling procedure was used to model the preintervention data. The Intervention time series analysis technique was later applied to obtain a model in the form $Y_t = f(I_t) + N_t$, where N_t is said to be the noise part of the model which is in the form of an ARIMA model obtained from the preintervention series, and is assumed to be stationary.

For a permanent change, the intervention variables I_t were introduced as step functions which were defined in the form of indicator variables where one indicates occurrence and zero for nonexistence of the intervention. In this study, the nature of impact (abrupt or gradual) of the two major intervention events on the cocoa production series was assessed. A broad-spectrum of assessing the goodness-of-fit based on the Ljung-Box Q statistic, residuals analyses and residual plots were applied to determine the adequacies of the models chosen. Models with adequate selection criteria were considered. R and Minitab statistical software were used to run the entire data.

1.4 Significance of the Study

Cocoa has been the bedrock of the Ghanaian economy throughout the century. It continues to play major roles such as creating sustainable employment, foreign exchange earnings to the country, and providing a chunk source of government revenue. As a result of its vital contributions to the nation, policies or programmes initiated to revive the industry need to be assessed in order for stakeholders to make informed decisions. This study is of no exception as it helps the Government and Ghana COCOBOD with a means to evaluate the effectiveness of intervention programmes geared towards boosting cocoa production in Ghana.

Moreover, this paper will also aid as a guide for Government and other cooperate institutions to assess the impact of control measures or events on any series of interest. The study would again add to the body of knowledge for other researchers to possibly build upon.

1.5 Organization of the Thesis

The next four chapters of this thesis are organized as follows: Chapter 2 reviews literature on previous intervention time series analyses that have been used to evaluate or assess programmes and policies. Chapter 3 of the thesis fits a statistical model using the Box-Jenkins ARIMA process to the annual cocoa production series from 1948 to 2001. It also assesses the effect of the intervention programmes on the cocoa series using the intervention time series analysis propounded by Box and Tiao (1975).

Chapter 4 discusses and analyses the annual cocoa production pattern in Ghana from 1948 to 2011. The chapter also discusses the results of the pre- and post-intervention models obtained.

Chapter 5 presents a summary of findings of the study together with recommendations and a conclusion.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

This chapter primarily reviews previous applications of intervention analysis published by various authors or researchers. The works reviewed touched on both univariate and multivariate time series intervention models relevant to this study (ie. as applied to social, health, economic and environmental time series data).

2.1 Previous Applications of Intervention Time Series Analysis

Most time series data which have been well-collected and organized may be specifically analyzed to assess the effect of some input events or interventions upon the series. Campbell and Stanley (1966) called this type of design as the "interrupted time series quasi-experiment". The methods of analysis originally suggested by Campbell and Stanley were found to be invalid due to the bias nature of the variance estimates, which were produced mainly by serial dependence. The autoregressive integrated moving average (ARIMA) model was later tested to be statistically valid for the assessment of the significance of these effects. This work was originally put forward by Box and Tiao (1975); Glass, Wilson and Gottman (1975); and McCleary and Hay (1980). After the development of this ARIMA Intervention model, many analysts have used it in a wide variety of applications. For example, Bhattacharyya and Layton (1979) and Harvey and Durbin (1986) have both applied the intervention analysis technique to analyze the effects of seatbelt laws on road fatality rates in Australia and Britain, respectively.

The subsequent piece generally reviews the intervention analysis technique as being applied in selected areas of motor/road traffic accidents, financial sectors, health, terrorism control and disasters. Others include; environmental challenges, transportation, and the agriculture sector.

2.1.1 Applications to Road Traffic Accident Data

Ledolter and Chan (1996) examined whether a significant change in fatal and major-injury accident rate can be detected following the implementation of a higher speed limit in the state of Iowa. Findings from their study revealed a 20% increase in the number of state-wide fatal accidents as a result of the speed limit change. The impact was largest on rural interstates, where the number of fatal accidents increased by 57%, implying two additional fatal accidents each quarter. However, their paper failed to find an impact on the number of major-injury accidents.

In another work, Chang, et al. (1993) proposed a multiplicative time-series model to capture the stochastic fatality pattern based on a long-term nationwide fatality data in the U.S.A, 'before' a speed limit change (January 1975 - March 1987). This long-term fatality model, along with the 2-year 'after' information (April 1987 - December 1989), was then used to detect the possible impact pattern. The results of the intervention analyses indicate that the increased speed limit had significant initial impacts on highway fatalities as the nationwide level. Such impacts, however, decayed after about a 1-year ‘learning period’. Additional study from their work showed that some large states, such as Texas, California, and Illinois, were insensitive to the speed limit increase.

Furthermore, Zambon et al. (2007) assessed the effects of a demerit points system introduced in Italy in July 2003 on the prevalence of seat belt use (intermediate outcome) and the number of road traffic deaths and injuries (health outcomes) using intervention analysis. A pre and post-intervention regional observation study for seat belt investigation from April 2003 to October 2004, and a national time-series analysis of road traffic deaths and injuries between 1999 and 2004 for health outcomes were considered. In all, 19,551 drivers, 19,057 front passengers and 8,123 rear passengers were included in the investigation into seat belt use, whereas 38,154 fatalities and 1,938,550 injured subjects were examined for the time-series analysis. Findings from their study revealed an increase in observed seat belt use of 51.8% among drivers, 42.3% among front passengers and 120.7% among rear passengers. They later estimated that 1,545 deaths and 91,772 injuries were prevented in 18 months after the introduction of the legislation.

More recently, Yaacob et al. (2011) conducted a study using intervention analysis to investigate the effects of OPS Sikap on road accidents in Malaysia. The aim was to assess the intervention effect in comparison with the standard ARIMA model, and hence to obtain the best model for forecasting purposes. The results of the paper suggested that there was a drop in the number of road accidents during OPS Sikap II, VI, VIII, XII and XIV. Nonetheless, the significant reduction could only be seen after the implementation of OPS Sikap VII and XIV with an expected number of reductions by about 1,227 and 1,484 accidents associated with respective interventions.

2.1.2 Financial Sector Applications

In the context of the financial sector, intervention analysis has mostly been used to estimate the effects of events, policies and regulation changes. For example, Chung, et al. (2009) carried out a study to model and analyze the impact of financial crisis on the manufacturing industry in China using data collected from March 2005 to November 2008 by the China Statistical Databases of the National Bureau of Statistics in China. The intervention effect of the global financial crisis that began in September 2008 on China's manufacturing industry, as measured in this study, was temporary and abrupt. The results indicate that China's manufacturing industry may have to tolerate a significant negative effect caused by the global financial crisis over a period of time, with its gross industry output value declining throughout 2008 and 2009 before reaching a state of equilibrium. The study further compared the results of the ARIMA and ARIMA Intervention models, confirming that the application of intervention analysis is appropriate for explaining the dynamics and impact of interruptions and changes of time-series in a more detailed and precise manner.

In another study, Rock-Antoine and Hannarong (2002) investigated the impact of the North American Free Trade Agreement (NAFTA) on both bilateral trade and income of each member country- US, Canada, and Mexico. The time series data covered before and after NAFTA was a formed span from 1980 to 1999. In the study, NAFTA was considered as a prolonged impulse function in international trade activities among the three trading partners by employing an intervention-function model. Findings from their study revealed that NAFTA increases bilateral trade between US-Canada and US-Mexico, and in terms of income, NAFTA benefits Canada the most "certainly".

Also, Shittu (2009) used the intervention analysis approach to model exchange rate in Nigeria in the presence of finance and political instability. The analyst applied the intervention analysis technique on the monthly exchange rate of Naira vis-à-vis US Dollar from 1970 to 2004 on some identified intervention variables. The result showed that most of the interventions are pulse function with gradual and linear but significant impact in the naira-dollar exchange rates.

In a research, Bianchi et al. (1998) analyzed existing and improved methods for forecasting incoming calls to telemarketing centers for the purposes of planning and budgeting. Their work considered the use of additive and multiplicative versions of Holt–Winters (HW) exponentially weighted moving average models and compared it to Box–Jenkins (ARIMA) modeling with intervention analysis. They also determined the forecasting accuracy of HW and ARIMA models for samples of telemarketing data. Although there was much evidence in recent literature that "simple models" such as Holt–Winters perform equally or better than complex models, they found out that the ARIMA models with intervention analysis perform better for the time series data they studied.

In a paper by Abdalla (2006), the intervention modeling technique was used to measure the impact of Yazegi Company's decision to deliver new kinds of soft drink, and the intervention impact of the Israeli constraints on sales delivery in Gaza. It was found that there was a significant impact on the company's sales due to the company's intervention (positive impact), and the Israeli intervention (negative impact). The paper then made a comparison between the non-intervention model and the intervention model, and concluded that the forecasts of the intervention model were better than that of the non-intervention model, because its values were closer to the actual data and had smaller standard errors.

The control of inflation is said to be central to good monetary policy. In line with the negative influences of inflation on the world's economy, many analysts have assessed monetary and fiscal policies on the impact of inflation. Valadhkhani and Layton (2004) assessed the magnitude and duration of the Goods and Service Tax (GST) effect on inflation in Australia's eight capital cities using Box and Tiao intervention analysis and quarterly data spanning from 1948:4 to 2003:1. They found that GST had a significant but transitory impact on inflation only in the September quarter of 2000 when this new tax system was implemented.

Bonham and Gangnes (1996) explored the effect on hotel revenues of the Hawaii room tax using time series intervention analysis. They specified a time series model of revenue behaviour that captured the long run co-integrating relationships among revenues and important income and relative price variables, as well as other short run dynamic influences. The paper estimated the effect on Hawaii hotel room revenues of the 5% Hawaii hotel room tax introduced in January 1987. It concluded with no evidence of statistically significant tax impacts.

Keum-Rok (1998) also analyzed the impact of elections on tax policy in Japan using an ARIMA (autoregressive integrated moving average)-intervention analysis from 1953 through 1992. The author used discretionary tax revenues, which mean the changes in tax receipts attributable to changes in the tax code, rather than automatic tax revenues due to business cycles in the economy. The result of his study shows that there is a political tax cycle in Japan. That is, discretionary tax revenues decrease with a statistically significant amount in a year immediately before elections for the House of Representatives.

2.1.3 Application to Disaster Cases

A wealth of literature exists regarding the use of intervention analysis to asses or estimates the impact of natural and man-made disasters on time series data. These include the 1986 nuclear disaster in Chernobyl on tourism in Sweden (Hultkrantz and Olsson, 1997); the September 21, earthquake in 1999 and the severe acute respiratory syndrome outbreak in 2003 on the inbound tourism demand from Japan to Taiwan (Min, 2008a); Hurricane Hugo in September 21, 1989 on the business of a public hospital at US mainland in South Carolina (Fox, 1996); and the effect of the September 11 terrorist attack in USA on the performance of the airline industry (Guzhva,2008).

Min et al. (2010) used Autoregressive Integrated Moving Average (ARIMA) with integration model (also known as integration analysis) to evaluate the impact of different local, regional and global incidents of a man-made, natural and health character, in Taiwan over the last decade. The incidents used in this study are the Asian financial crisis starting in mid–1997, the September 21st earthquake in 1999, the September 11th terrorist attacks in 2001, and the outbreak of severe acute respiratory syndrome (SARS) in 2003.

Empirical results revealed that the SARS illness had a significant impact, whereas the Asian economic crisis, the September 21st earthquake and the September 11th terrorist attacks showed no significant effect on air movements.

In a comparable study, Min (2008b) assessed whether two events, the 9-21 earthquake in 1999 and the severe acute respiratory syndrome outbreak in 2003, had a temporary or long-term impact on the inbound tourism demand for Japan. Moreover, the study assessed whether intervention analysis produces better forecasts as compared with forecasts without intervention analysis. Experimental results confirm that the effect of both disasters on Japanese inbound tourism presented only temporarily, and the forecasting efficiency of ARIMA with intervention is superior to that of a model without intervention.

Elsewhere, a general formulation of an intervention model was applied to water-quality data for two streams in the north-eastern Victoria city of Australia, measuring the effect of drought on the electrical conductivity of one stream, and the effect of bushfires on the flow and turbidity of the other. The nature of the intervention is revealed using exploratory data-analysis techniques, such as smoothing and box plots, on the time-series data. Intervention analysis is then used to confirm the identified changes and estimate their magnitude. The increased level of electrical conductivity due to drought was determined by three techniques of estimation and the results compared. The best of these techniques was then used to model changes in stream flow and turbidity following bushfires in the catchments (Welsh and Stewart, 1999).

Collier et al. (2010) illustrates that natural disaster such as those created by extreme El Niño can significantly threaten financial institutions serving the poor. The effects of the 1997-98 El Niño on problem loans (restructured loans and those in arrears) is estimated using intervention analysis for a microfinance institution (MFI) in Piura, a region in northern Peru severely affected by El Niño. Extreme El Niño events like those in 1982-83 and 1997-98 create catastrophic flooding that destroys transportation infrastructure, disturbs the livelihoods of households engaged in a wide range of activities, and destroys productive assets, crops, and private homes. The purpose of their paper was to assess the exposure of a Piura MFI to the consequences associated with an extreme El Niño. Portfolio-level, monthly data from January 1994 to October 2008 were examined using an intervention analysis. While restructured loans averaged 0.5 percent of the total loan portfolio before the 1997-98 El Niño, the estimated cumulative effect of the El Niño indicates that an additional 3.8 percent of the total portfolio value was restructured in a short time period due to this event. No significant effect was found for changes in the proportion of late loans. The analyses demonstrate 1) that the correlated risk exposure of many small borrowers can significantly affect the lender when the catastrophe occurs; 2) the importance of considering bank management in assessing disaster risk to a loan portfolio; and 3) lender strategies to minimize losses may require long-term restructuring that perpetuates the effects of the disaster in the community.

2.1.4 Health Sector Applications

Intervention analysis has most recently been extended to estimate the impact of events and programmes in the health sectors. Examples of such works are the Australian Heroin shortage of 2001 (Stuart, et al., 2006); U.S. Medicaid expenditures on antidepressant agents (Ferrand et al., 2011) and hospital antibiotic prescribing (Ansari et al., 2010).

Many claims have been made in regard to the level of success and efficacy of breast and cervical screening programs within Australia and internationally. Various techniques have been utilised to assess the full ramifications of introducing each screening program. Many techniques have not been adapted to this area of study. One such example that has previously been neglected is intervention analysis. Intervention analysis enables the magnitude of effects resulting from an interruption in a time series to be directly quantified. The technique draws heavily on Autoregressive Integrated Moving Average (ARIMA) modeling and application of the Box-Jenkins methodology. An extension of this technique allows a transfer function to be estimated to ascertain whether other factors are contributing to the effects obtained in the intervention model. The results produced outline reductions in mortality rates of between 5% and 28% within Australia as a result of the national breast and cervical screening programs enacted in 1991. The extents of these reductions in mortality indicate that the programs have been successful in reducing mortality to a certain degree. The transfer function analysis also indicates that these reductions can be directly attributed to the introduction of the screening programs (Quirey, 2001).

Cristina et al. (1998) presented a study that describes the use of time series analysis in the evaluation of the incidence of nosocomial infection at the Guadalajara general hospital, Spain. A monthly data comprising time series incidence of nosocomial infections was sourced from a nosocomial infection surveillance system of a primary-care general hospital. The data was analysed by curve fitting, autoregressive integrated moving average (ARIMA) modeling (Box-Jenkins), and intervention and dynamic regression analysis. Their result indicated that, the imposed control and training of personnel by the surveillance system was associated with a 3.63% decrease in accumulated monthly incidence of nosocomial. There was also a strong indication that an increase of infection incidence of 4.34% corresponded to a medical strike. An increase of 0.18% was associated with each new nursing contract. Evidence was then obtained from the possible relationship between incidence of nosocomial infection and vacation periods.

A University-based hospital in Bogotá, Colombia, developed and implemented an educational intervention to complement a new structured antibiotic order form. This intervention was performed after assessing the appropriateness of the observed antibiotic prescribing practices using a quasi-experimental study. An application of interrupted time series intervention analysis was conducted in three antibiotic groups (aminoglycosides, cephradine/cephalothin, and ceftazidime/cefotaxime) and their hospital weekly rate of incorrect prescriptions before and after the intervention. A fourth time series was defined on prophylactic antibiotic use in elective surgery. Preintervention models were used in the postintervention series to test for pre–post series level differences. An abrupt constant change was significant in the first, third, and fourth time series indicating a 47, 7.3, and 20% reduction of incorrect prescriptions after the intervention. The study concluded that a structured antibiotic order form, coupled with graphic and educational interventions can improve antibiotic use in a university hospital (Adriana et al., 2003).

Medical experts and health administrators have for several years embarked on vaccination exercises to eradicate various forms of diseases. Girard (2000) used an ARIMA model with intervention analysis technique to analyse and assess the epidemiology situation of whooping-cough in England and Wales for the period 1940 to 1990. The ARIMA modeling of this illness contains intervention variables, such as the introduction of widespread vaccination in 1957 and the fall in the level of vaccination down to 30% in 1978. The results of the study confirmed the role of the intervention variables on the evolution of the morbidity due to whooping-cough, by quantifying their impact on the level of the morbidity, as well as the delay needed before they have an influence on the increase of recorded cases of whooping-cough.

2.1.5 Terrorism Control

Terrorist activities have, over the years, been quantified and analysed by many studies all around the world. In most of these studies, anti-terrorist policies and laws are evaluated to ascertain their effectiveness. A recent approach in this direction is the use of intervention analysis to measure the impact or outcome of terrorism on economic activities.

Ismail et al. (2009), dissected the impact of Bali's first terrorist bombing event that occurred on October 12th, 2002, to the tourism industry in Indonesia. An intervention model, particularly a pulse function of intervention was carried out focusing on the differential statistics that can be used to determine the order of intervention model. The study was focused on the derivation of some effect shapes, i.e. temporary, gradually or permanent on the arrival of tourists into Bali. A new model building procedure with three main iterative steps for determining an intervention model was used for the data with extreme change in mean. The result from the study showed a decreasing trend in tourist arrivals in Bali.

Also, Muhammad et al. (2010) focus on the development of a model that could be used to explain the magnitude and periodic impacts of the Asian financial crisis since July 1997 and terrorist attacks referring to the Bali bombings on October 12th 2002 and October 1st 2005, respectively. Monthly data comprising the number of tourist arrivals in Indonesia via Soekarno-Hatta airport are used as the data for this case study. The results show that the Asian financial crisis and Bali bombings yield negative impacts on the number of tourist arrivals to Indonesia via Soekarno-Hatta airport. Generally, the Asian financial crisis gives a negative permanent impact after seven month delay. The first and second Bali bombings also yield negative impacts which were temporary effect after six and twelve months delay respectively.

Enders et al. (1990) employed interrupted time series analysis to assess the effectiveness of four specific terrorist-thwarting policies undertaken between January 5, 1973 and April 15, 1986. These policies include the following: (1) installation of metal detectors in airports, (2) enhanced security for U.S embassies and personnel, (3) the legislation of the Reagan "get-tough" laws on terrorism, and (4) the U.S retaliatory strike against Libya. Both short - , medium - , and long- run effects were ascertained. The most successful policy involved metal detectors. Expenditures to secure U.S embassies had the intended effect, but it also had the unintended effect of putting non- U.S diplomats at somewhat greater risk. The Reagan get-tough laws were ineffective. Unfortunately, the Libya raid had the unintended effect of increasing U.S and U.K attacks temporarily.

Following a line of investigation, Carlos (2003) enquired into the terrorism policies that have worked best in handling the Spanish Euskadi Ta Askatasuna (ETA) terrorism using time series data from 1968 to 2000. The effects of political, deterrence and economic policies were reviewed, and concluded that their influence on terrorism incidents is mixed.

Lloyd et al. (1998) also delves into the impact of anti-dumping and anti-cartel actions using intervention analysis. The researchers adopted the intervention analysis technique to primarily distinguish the effects of the anti-dumping action from those of changes in cartel behaviour.

2.1.6 Application to Environmental Challenges

It is widely accepted that the continued existence of the entire human race is mainly supported by a well-sustainable environment. However, due to the issue of over-population and its associated economic pressures, the immediate environment is more at times exploited in attempting to meet the insatiable needs of humans. Many techniques including Intervention analysis have helped to weigh up the consequences of some of these natural and artificial activities on the environment. For example, Stanley (1984) probes into the effects of spring break-up on Microscale air temperatures in the Mackenzie River Delta

In another development, an Intervention Analysis Model (Box and Tiao, 1975) was used to study the impact of the 'intervention' brought in by the Government of India (GoI), to control the Carbon Monoxide (CO) pollution caused by the vehicular exhaust emissions, by the enforcement of the emission standards for the vehicles, on the mean level of the time-series of CO concentration. The study was conducted for an Air Quality Control Region (AQCR) comprising of an urban road intersection in Delhi, India, where almost 100% CO is contributed by vehicular traffic. Application of the model suggests that the `intervention' has not been effective in bringing down the desired change; some likely causes of which have also been mentioned (Sharma and Khare, 1999).

Intervention analysis is a rigorous statistical method for analyzing the effects of man-induced or natural changes on the environment. For instance, it may be necessary to determine whether a newly installed pollution control device significantly reduces the former mean level of a pollutant. By using intervention analysis, the actual change in the pollutant levels can be statistically determined. Previously, no comprehensive method was available to assess changes in the environment. Intervention analysis is an advanced type of Box-Jenkins model. A general description of Box-Jenkins models and their extensions is given in the work of Hipel et al. (2001). Also, the importance of adhering to sound modeling principles when fitting a stochastic model to a time series is emphasized. Following a discussion of intervention models, three applications of intervention analysis to environmental problems are given. Two applications deal with the environmental effects of man-made projects, while the third example demonstrates how a forest fire can affect the flow regime of a river.

Kuo and Sun (1993) designed an intervention model with Box-Jenkins time series model to deal with the extraordinary phenomena caused by typhoons and other serious abnormalities of the weather on an average 10 day stream-flow of the Tanshui River basin in Taiwan. The model included five types of typhoons and one serious drought as the exogenous factors and was capable of describing the special hydrological time series. The intervention model was applied to stream-flow forecasting and synthesis at Sanhsia gagging station in the basin. They then compared the traditional autoregressive moving average (ARMA) model with the intervention model. It was realized that the intervention model greatly improves the forecasting technique and successfully captures the special pattern of the stream-flow when typhoons invade. In addition, the intervention model was effectively used by the researchers to synthesize stream-flow.

2.1.7 Transportation Sector Applications

Ming-Chan, et al. (2003) discusses the technique with which the impact of policy guidance on transportation demand is examined. Through developing a multivariate ARIMA model encompassing transfer function and intervention analysis capability, the impact of the intervention policy could be examined by simultaneously considering the dynamic regression effects of other explanatory series on response series. After empirical data analysis, it shows that the policy launched by Taiwan government in 1987 to lift the ban on family visits to Mainland China has a statistically significant and positive effect on the demand for air transport service on Taipei-Hong Kong route, while simultaneously taking the possible effect from the change in national income per capita into consideration. Moreover, model estimation also shows a pattern with impact that is gradually rising at initial part, and lasting long afterwards.

Narayan and Considine (1989) presents a study to show how intervention analysis can be used to obtain accurate estimates of the impact of two price changes on ridership in a transit system. A method is proposed in which the components of a time series together with the intervention components are explicitly modeled in a multiple input transfer function model. The method thus combines techniques from regression analysis with those from the ARIMA methodology. It is shown that both price increases are accompanied by significant reductions in ridership. The case study compares the results of this method with those from the usual intervention analysis of Box and Tiao.

Chen (2006) investigated the intervention impacts on tourist flows, and later in the same study, the researcher then evaluated the accuracy of various forecasting techniques to predict travel demand before and after the inclusion of intervention events. Data on air transport passengers of the US from January 1990 to June 2003 was used for the study. In all, five forecasting methods were compared for accuracy. The study focused on two main objectives: 1. Forecasting accuracy of allowing for intervention events in the modeling process. 2. Examining the impacts on tourism demand of the major crises that occurred during the period 2001–2003. The revealed that the total impacts of intervention events on US domestic tourist flows since September 11th 2001 showed a decrease in demand by approximately 61,488,520 in terms of the number airline seats, and international passenger demand also decreased by approximately 5,890,595 in terms of airline seats. In the case of the monthly international arrivals, the SARIMA forecasted values was the best among all the other forecasting techniques when the intervention events were estimated.

2.1.8 Agriculture Sector Applications

In recent years, the contribution from outboard sector has been a major component in the total marine fish production from the states of Kerala and Karnataka. This is as a result of the introduction of crafts fitted with outboard engines for propulsion in the mid eighties, which intensified and developed into a major sector. The impact of this intervention is examined by adopting two popular time series methods used for intervention analysis. The first method is based on seasonal ARIMA modeling and the second is based on regression modeling with ARMA type errors. Quarter-wise total marine fish landings in the two states during 1960-2000 were used for the impact study. The analysis revealed that for Kerala the model found suitable is seasonal ARIMA type model and for Karnataka the feasible model was regression model with ARMA errors. Based on the final estimated intervention models, the effect of the interventions was estimated at 2.26 lakh tonnes and 88 thousand tonnes per annum respectively for Kerala and Karnataka (Sathianandan et al., 2006).

Morales et al. (1988) uses a time series based methodology, intervention analysis, to investigate changes occurring in California's brucellosis program during the 1970's. Monthly records from the Bureau of Animal Health, State of California on herds quarantined for Brucellosis over a ten year period--January 1970 to December 1979 was use for the study. These records were combined with a chronology of relevant changes in the eradication campaign policies, two of which particularly appeared to have paralleled large shifts in the number of infected herds. The first event was the return of Bureau of Animal Health personnel to the brucellosis program in August, 1973 after the eradication of exotic disease outbreaks in California. The second event was a change in regulations effective January, 1976 mandating 30-day retesting of dairy females entering California and special movement procedures for the state's two most infected counties located in Southern California. These events were incorporated in the model building as dummy variables (ie, 0 for all time periods before the event or intervention, and 1 thereafter). A final ARIMA model was then estimated on the differenced data using the method of maximum likelihood and a moving-average model of order 1 and 4 [MA (1, 4)] was fitted. They indicated that brucellosis infected herds are likely to remain infected over several monthly observations as captured by the first-order moving-average [MA (1)] term. The presence of a fourth-order moving-average term (ε_{t-4}) indicates the existence of a four-month cycle which they could not associate with the program procedures.

Remenyi, et al. (2010) attempt at using time-series analysis to describe changes in the heart rate (HR) of free-ranging cows receiving programmed audio cues from directional virtual fencing (DVF™) devices designed to control the animal's location on the landscape as well as non-programmed environmental/physiological cues. Polar Accurex® devices were used to capture HR every minute from March 19-24, 2003 when three mature free-ranging beef cows previously habituated to DVF™ control were confined to a brush-infested area of an arid rangeland paddock. Global positioning system (GPS) electronics were used to locate each cow's location approximately every minute while it was in a 58 ha virtual paddock (VP™) and every second when it penetrated a virtual boundary (VB™). The cows never escaped through the VB™ though they penetrated it a total of 26 times in 11 different events at which time they received programmed audio cues lasting from 1 to 56 s. These data reveal that HR spikes from programmed audio cues all fell within the textbook range for cow HR (40-186 beats per minute, bpm). For both audio and selected environmental /physiological events, HR spikes returned to pre-cuing "baseline" values in about one minute. However, the longest return time to baseline lasted (about 4 minutes) and this was for an environmental/physiological event. HR, animal location, weather and other electronic data should be measured at equally-spaced time intervals using a single time stamp so as to accurately associate HR changes with possible causes.

2.2 Conclusion

The above discussion has mainly highlighted the use of intervention analysis as an analytical and a forecast tool in a wide variety of applications. It reviews a range of previous applications of intervention analysis that have been used by various researchers since its original development by Box and Tiao (1975). From the review, we could affirm that the intervention analysis, as being used by many analysts is a powerful statistical technique in evaluating or assessing the impact of events, policies or shocks on time series data. However, to my best of knowledge, no published study has being extended to assess policies or shocks in the cocoa sectors using intervention analysis.

The next chapter elaborates further on this statistical technique and provides a detailed account of its procedure.

CHAPTER 3

METHODOLOGY

3.0 Introduction

This chapter essentially underlines the elementary theory of Time Series with regards to its definition, component and objectives. It also examines the concept of Univariate Time Series process with key emphasis on the popularly known Box-Jenkins ARIMA process. The chapter further elaborates on the Box and Tiao Intervention Analysis procedure which have been used extensively to model the impacts of the Intervention events under consideration in the study.

3.1 Basic Definition

Time series is a collection of observations on variables, made sequentially in time. These observations are normally taken at equal time period. For example, daily stock prices, weekly petroleum sales, monthly electricity consumption, and the annual cocoa production figures used in this study are all classified as time series.

In notation, it is donated by; $\{Y_1, Y_2, Y_3, \ldots Y_T\}$ or $\{Y_t\}$, where $t = 1,2,3,\ldots T$

The mean and variance functions of such observations at time t are given respectively by;

$$\mu_t = E[Y_t] \tag{3.1}$$

$$\sigma_t^{\,2} = E[(Y_t - \mu_t)^2 \tag{3.2}$$

3.2 Component of Time Series

A vital step in choosing appropriate modeling and forecasting procedure is to consider the type of data patterns exhibited from the time series graphs of the time plots. The sources of variation in terms of patterns in time series data are mostly classified into four main components. These components include seasonal variation; trend variation; cyclic changes; and the remaining "irregular" fluctuations.

3.2.1 Seasonal Effect

A seasonal (*S*) pattern exists when a series is influenced by seasonal factors. Time series data which are collected over the quarter of the year, monthly, or day of the week are basic examples of seasonal factors. Many time series, such as sales of products, temperature readings and household electricity consumption usually show signs of these seasonal patterns. For example, in many western nations, unemployment is typically 'high' in winter but lower in summer. This means unemployment only reaches its peaks during the winter periods. Seasonal series are sometimes also called "periodic" although, they do not exactly repeat themselves over each period.

3.2.2 Other Cyclic Changes

A cyclic (*C*) variation exists when the data exhibit rises and falls that are not of a fixed period. For economic series, these are usually due to economic fluctuations such as those associated with the business cycle. The sales of products such as automobiles, steel and major appliances occasionally shows this type of pattern. The major distinction between a seasonal and a cyclical pattern is that the former is of a constant length and recurs on a regular periodic basis, while the latter varies in length. Moreover, the average length of a cycle is generally longer than that of seasonality and the magnitude of a cycle is usually more variable than that of seasonality.

3.2.3 Trend

Trend may be loosely defined as a long term increase or decrease in the mean level. In other words, a trend pattern exists when there is a long term increase or decrease in a particular data. The difficulty with this definition is deciding what is meant by 'long term'. Granger was reported by Chatfield (1996) to have defined 'trend in mean' as comprising all cyclic components whose wavelength exceeds the length of the observed time series.

3.2.4 Other Irregular Fluctuations

After trend and cyclic variations have been removed from a set of data, we are left with a series of residuals, which may or may not be 'random'. Series which shows this irregular variation may be explained in terms of probability models, such as moving average models, autoregressive models or a mixed process (Chatfield, 1996, pp. 9–10) and (Makridakis et al., 1998, pp.25).

3.3 Objectives of Time Series

There are several possible objectives in analyzing a time series. These objectives may be broadly classified as description, explanation, prediction and control.

3.3.1 Description

The first most important step in time series analysis is usually to plot the data against time, and to report some descriptive measures of the characteristics engulfed by the time series under consideration. In most cases, a plot of such data normally shows the type of variations being depicted by the series. These plots mainly attempt to describe seasonal, cyclic and trend variations in series. In other situations, the time series graph may not only show variation patterns, but it also enables the analyst to look for 'wild observations' or outliers which do not appear to be more consistent with the rest of the data. Another key feature to observe in the graph of time series is the possible presence of turning points, where, for example, an increasing trend has suddenly changed to be a downward trend. Analysts are therefore cautioned to always plot their data to verify these checks before embarking on full analysis of any time series data.

3.3.2 Explanation

Most often, when observations are taken on two or more variables, for example, the extent at which sea level is affected by temperature and pressure, we may make use of the variation in a particular series, say temperature or pressure to explain the variation in another time series, say the sea level. In a similar case, a researcher may want to analyze how Gross Domestic Product is affected by a countries interest and inflation rates.

3.3.3 Prediction

One of the most fundamental practices in time series analysis is to predict future values for the observed series. Many economic and financial time series Analysts see prediction as a core component in making accurate projections.

3.3.4 Control

When an observed time series measures the 'quality' of a manufacturing process, for instance wastewater treatment process, the main aim of the analyst may be to control the process. In practice, control procedures are of several different kinds. For instance, in statistical quality control, the observations are plotted on control charts and the controller takes action, as a result of studying the charts. Nonetheless, many other contributions to control theory have been made by control engineers and mathematicians rather than statisticians (Chatfield, 1996, pp. 5–7).

3.4 Univariate Time Series Models

Under this section, I discussed time series analysis models with only one series of observations such as the cocoa production series used for the study. These kinds of models are usually term as Univariate Time Series models. Such models consider their series, for instance the cocoa production series as a function of its own past, random shocks and time. The rest of this part briefly stressed on some basic Univariate models and Univariate Time Series processes.

3.4.1 Autoregressive (AR) Models

Autoregressive models are based on the idea that the current value of the series Y_t can be explained as a function of p past values, $Y_{t-1}, Y_{t-2}, ..., Y_{t-p}$, where p determines the number of steps into the past needed to forecast the current value. In other words, when the value of the present observation Y_t depends solely on past values of Y_t and a certain random term (shock), the resulting model $Y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots.. + \phi_p y_{t-p} + \varepsilon_t$ (3.3), is known as the p^{th} order Autoregressive model (Shumway and Stoffer, 2006, pp. 85).

3.4.2 Moving Average (MA) Models

When the current output Y_t depends solely on the current input and q prior inputs, the model $Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots. - \theta_q \varepsilon_{t-q}$ (3.4), is known as a moving average model of order q. The phrase moving average in this study is not the normal moving average smoother which consists of arithmetic mean of past historical data. The moving average in (3.4) is rather defined as a moving average of the error series ε_t. It is most at times unclear to nonstatisticians as to how an observation could be modeled as a linear weighted sum of random numbers. However, it is important to realize that a series composed of such linear sums of white noise elements is not itself white noise; but rather has a definite autocorrelation structure (Reagan, 1984).

3.4.3 Autoregressive Moving Average (ARMA) Models

A mixed stochastic process which combines autoregressive (AR) and moving average (MA) models together in the form;

$Y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots. + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots. - \theta_q \varepsilon_{t-q}$ (3.5), is called an autoregressive moving average model of order (p, q).

3.4.4 Continuous vs. Discrete Time Series

A time series is said to be continuous when observations are made continuously in time. An example of a continuous time series can be seen in the readings from a seawater level as is being measured by an automated sensor.

On the other side, a series is said to be discrete when observations are taken only at specific times, usually equally spaced. The term ‘discrete’ is used for series which are made sequentially in time even when the measured variable is a continuous variable. In this study, I mainly concentrated on a discrete time series, where the cocoa production figures were taken at annual time interval (Chatfield, 1996, pp. 4 - 5).

3.4.5 Deterministic vs. Stochastic Models

Deterministic models are time-dependent quantities which may be prescribed by certain known laws of physics or accounting. They may indicate the presence or absence of an event. Such models always assume that an outcome of an event is certain, and easily does predictions exactly. This means, a deterministic model gives forecast values with no error margins.

On the contrary, a stochastic time series model sees each data value of its series as a random variable which may be viewed as a sample mean of a probability distribution from an underlying population at each point in time. Stochastic models such as the Box-Jenkins models usually assume that outcomes from events are uncertain. These models give forecast values with non-zero error margins. The error margins are randomly distributed. The main underlying difference between the two models is seen from the probabilistic nature of stochastic models as compared to the known laws of physics associated with deterministic models. In notation, stochastic models can be expressed in terms of discrete and continuous time series processes. These processes are shown in equations (3.6) and (3.7) respectively:

$$\{Y_t : t \in \{0,1,2,\ldots\}\} = \{Y_0, Y_1, Y_2, \ldots\} \tag{3.6}$$

$$\{Y_t :\ t \in (-\infty, \infty)\} = \{Y_t :\ -\infty < t < \infty\} \tag{3.7}$$

However, a discrete time series data (Ghana's annual cocoa production figures) was entirely used to obtain the pre and post-intervention stochastic models reported in this study (Yaffee and McGee, 2000, pp.5).

3.4.6 Stationary and NonStationary Stochastic Process

Time series may be stationary or nonstationary. According to Box and Jenkins (1976), stationary series are characterized by a kind of statistical equilibrium around a constant mean level, as well as a constant dispersion around that mean level. There are several kinds of stationarity. A series is said to be stationary in the wide sense, weak sense, or second order if it has a fixed mean and a constant variance. A series is believed to be strictly stationary if it has, in addition to a fixed mean and constant variance, a constant autocovariance structure. When a process is climax as weak stationary, both the mean and variance remains constant over time, and the autocovariance depends only on the number of time lags between temporal reference points in the series. Weak stationarity is also called covariance stationarity or stationarity in the second sense. The condition of covariance stationary implies a stable regime, in which the parameters of a model remain constant. If the distributions of the observations are normally distributed, the series is said to possess strict stationarity. Series must be stationary in mean, variance and autocovariance. If a series is stationary, the magnitude of the autocorrelation attenuates fairly rapidly, whereas that of a nonstationary series diminishes gradually over time.

Many economic series are integrated or nonstationary in nature. Nonstationary series may follow from the presence of one or several of five conditions: outliers, random walk, drift, trend, or changing variance. Time series data must be examined in order to ascertain whether any of these nonstationary phenomena inhere within such series. Nonstationary series that lack mean stationarity have no mean attractor toward which the level tends over time. For example, if each realization from a stochastic process appears to be a random fluctuation, as in the haphazard step of a drunken sailor, bereft of his bearings, zapped with random shocks, the series of his movement is a random walk. If the series exhibits such sporadic movement around a level before the end of the time horizon under consideration, it shows signs of random walk plus drift. Drift in other words, is random variation around a nonzero mean. A series with trend normally manifests an average change in mean level over time. The trend may be stochastic or deterministic. In general, transformations to stationary are performed by differencing for stochastic trends and by regression for deterministic trends. If the series has a stochastic trend, then there is change of level in the series that is not entirely predictable from its history. Other nonstationary series have growing or shrinking variances. Changes in variance may come from seasonal effects or the influence of other variables on the series under consideration. Series afflicted with significantly changing variance have homogenous nonstationarity. To prepare such series for statistical modeling, the series are transformed to stationarity either by taking the natural log, by taking a difference, or by taking residuals from a regression. If the series can be transformed to stationarity by differencing, one calls the series difference-stationary. If one can transform the series to stationarity by detrending it in a regression and using the residuals, then we say the series is trend-stationary (Yaffee and McGee, 2000).

3.4.7 The Lag and Differencing Operators

The Lag operator L is defined for a time series $\{Y_t\}$ by;

$$LY_t = Y_{t-1} \tag{3.8}$$

$$L^2Y_t = Y_{t-2} \tag{3.9}$$

The two equations, (3.8) and (3.9) define lag operators for past values of $\{Y_t\}$ which are lagged by one and two time periods respectively. This means, the lag operator L which operates on any time series $\{Y_t\}$ has the effect of shifting the series back by one, two or more periods. We can therefore define powers of L for $\{Y_t\}$ by;

$$
\begin{aligned}
L^2Y_t &= LLY_t = LY_{t-1} = Y_{t-2} \\
L^3Y_t &= LL^2Y_t = Y_{t-3} \\
&. \\
&. \\
&. \\
L^kY_t &= Y_{t-k}
\end{aligned}
$$

Box and Jenkins (1976) initially used a Backshift operator B for the same Lag operator L. But for convenience seek, the Lag operator L was adopted throughout this study.

Other operators can also be defined in terms of L. Example of such operators is the differencing operator. This operator is of fundamental importance when dealing with models for nonstationary time series. The first and second-order differencing operators for $\{Y_t\}$ are given in (3.10) and (3.11) respectively:

$$\nabla Y_t = Y_t - Y_{t-1} = Y_t - LY_{t-1}$$

$$\nabla Y_t = (1-L)Y_t \tag{3.10}$$

$$\nabla^2 Y_t = \nabla(\nabla Y_t) = \nabla(Y_t - Y_{t-1})$$

$$= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$$

$$= Y_t - 2Y_{t-1} + Y_{t-2} = (1 - 2L + L^2)Y_t$$

$$\therefore \nabla^2 Y_t = (1-L)^2 Y_t \tag{3.11}$$

In all, the d^th^-order differencing operator may be written as $(1-L)^d Y_t$, where L is the lag operator and d denotes the order of differencing (Ihaka, 2005).

3.4.8 Unit Root and Stationarity Tests

Most time series data could be nonstationary due to the presence of random walk, drift, or trend as discussed in the earlier sections of this chapter. One best way to test these is to evaluate a regression that nests a mean, a lagged term which checks for difference stationarity, and a term for deterministic trend which also looks for trend stationary in one particular model:

$$Y_t = \alpha + y_{t-1} + \beta t + \varepsilon_t \tag{3.12}$$

By taking the first difference of (3.12), we get; $\nabla Y_t = \alpha + (\rho - 1)y_{t-1} + \beta t + \varepsilon_t$.

This model forms the basis of the Dickey-Fuller unit root test. The application of the Dickey-Fuller test mainly depends on the regression context in which the lagged dependent variable is tested. The three identified model contexts are those of (1) a pure random walk, (2) random walk plus drift, and (3) the combination of deterministic trend, random walk, and drift. In line with this, three different regression equations are considered;

$$Y_t = \rho_1 y_{t-1} + \varepsilon_t\,,\ \varepsilon_t \sim i.i.d(0,\sigma^2) \tag{3.13}$$

$$Y_t = \alpha_0 + \rho_1 y_{t-1} + \varepsilon_t\,,\ \ \varepsilon_t \sim i.i.d(0,\sigma^2) \tag{3.14}$$

$$Y_t = \alpha_0 + \rho_1 y_{t-1} + bt + \varepsilon_t\,,\ \varepsilon_t \sim i.i.d(0,\sigma^2) \tag{3.15}$$

Equation (3.13) specifies a regression model without a constant term. This regression model is used to test for pure random walk process without drift. Here, the null hypothesis of nonstationary random walk is tested against a stationary series. If $\rho_1 = 1$, then the null hypothesis cannot be rejected and the data generating process is inferred to have a unit root.

The second Dickey-Fuller case in (3.14) involves a context of random walk plus drift around a nonzero mean. The null hypothesis is that the series under consideration is integrated at the first order, that is, I (1). In other words, the null hypothesis of whether $\rho_1 = 1$ is tested against a stationary series around a constant mean of $\frac{\alpha_0}{(1-\rho)}$.

The third Dickey-Fuller case is one with a context of random walk plus drift in addition to a deterministic linear trend shown in (3.15). As in the earlier cases, the null hypothesis is that $\rho_1 = 1$ $(\rho_1 - 1 = 0)$ and the alternate hypothesis is that the series is stationary.

Nonetheless, not all Dickey-Fuller regression models have white noise residuals. This means, in a situation where the error term (ε_t) in (3.13), (3.14) and (3.15) are autocorrelated, the Dickey-Fuller distribution might not be applicable. However, if there is autocorrelation in the series, it has to be removed from the residuals (ε_t) of the regression before Dickey-Fuller tests are executed. Under the conditions of residual serial correlation, where the Dickey-Fuller regression models are not valid for the unit root test, a new test called the Augmented Dickey-Fuller (ADF) test in (3.16) may be applied. This new regression model addresses the issue of serial correlation.

$$Y_t = \alpha_0 + \rho_1 y_{t-1} + \sum_{i=2}^{p-1} \beta_i \nabla y_{t-i} + \varepsilon_t \tag{3.16}$$

In a situation where the process is an ARMA(p, q) process, Said and Dickey were reported by Yaffee and McGee (2000) to have discovered that the MA(q) portion of the ARMA(p, q) process under conditions of MA(q) parameter invertibility can be represented by an AR(p) process of the kind in (3.16) when p gets large enough. The Augmented Dickey-Fuller equation is identical to the three Dickey-Fuller equations discussed earlier, except that the ADF equation contains higher order lags of the differenced dependent variable which take care of serial correlation before testing for nonstationarity. If the series has a higher order serial correlation which result to an AR unit root, higher order differencing will be required in order to transform the residuals into white noise disturbances. Moreover, utmost care should be taken since over-differenced series might also result to an MA unit root.

For a series which has random walk plus drift patterns, the lagged order of the differenced dependent variables necessary to remove the autocorrelation from the residuals is usually set to be about three (3). For this situation, equation (3.17) might be used to test for nonstationarity.

$$Y_t = \alpha_0 + \rho_1 y_{t-1} + \sum_{i=2}^{3} \beta_i \nabla y_{t-i} + \varepsilon_t \tag{3.17}$$

In a case where the series under study exhibit patterns of random walk plus drift around a stochastic trend, the Dickey-Fuller test can be reconstructed with the addition of a time trend variable as shown in (3.18)

$$Y_t = \alpha_0 + \rho_1 y_{t-1} + \sum_{i=2}^{p} \beta_i \nabla y_{t-i} + bt + \varepsilon_t \text{, where } \varepsilon_t \sim i.i.d(0, \sigma^2) \tag{3.18}$$

where $\rho_1 y_{t-1} + \sum_{i=2}^{p} \beta_i \nabla y_{t-i}$ is the Augmented part, y_{t-1} is the lagged term of Y_t, ∇y_{t-i} shows the lagged change, t represent the deterministic trend, α is the drift component, ε_t represents a well-behaved error term (unobserved series) and b, ρ_1, β are coefficients to be estimated.

Generally an ADF test with hypothesis $H_0 : \rho_1 = 0$

$H_1 : \rho_1 < 0$ can be tested in the regression model in (3.18).

Kwiatkowski, Phillips, Schmidt and Shin (KPSS, 1992) proposed a procedure for testing stationarity time series data as reported by Kirchgassner and Wolters (2007). The null hypothesis of their test does not depend on the existence of a unit root but instead a stationary series. The starting point of the KPSS test is given as $Y_t = \alpha_t + \beta_t + \mu_t$. Instead of the commonly used constant term, a random walk, $\alpha_t = \alpha_{t-1} + \varepsilon_t$ is allowed, where ε_t are assumed to be independent and identically normally distributed. The hypothesis of their test is given below:

H_0: Y_t is level or trend stationary, I(0)

H_1: Y_t has a unit root

A p-value greater than 0.05 from the result of the KPSS test is enough to accept the null hypothesis at 5% level of significance. On the contrary, in an ADF unit root test where p-value > 0.05, one can not reject the presence of a unit root and should therefore difference the series once or more (∇Y_t) to achieve stationarity, and later display the correlogram of such differenced series (Yaffee and McGee, 2000, pp.81-86) and (Frain, 1992).

3.4.9 Bounds of Stationarity and Invertibility

According to Wold's decomposition theorem, a series may be explained in terms of an infinite linear combination of weighted random shocks. Such series may be interpreted as an infinite moving average of innovations or random shocks. More often than not, moving average models are conceived of as finite, rather than infinite order of weighted past shocks.

Moreover, higher moving average (MA) models tends to be rarer as compared to that of first- or second-order moving average models. In a situation where the observed series converges, a first-order moving average process is said to be equivalent to an infinite-order autoregressive process. Similarly, an infinite-order AR process of a series is also equivalent to the first-order MA process if only the series converges.

For example, in an MA(1) model where $Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$, the parameter θ_1 must conform to certain bounds of invertibility. The bounds of invertibility of the MA(1) model are defined by the inequality of $|\theta_1| < 1$.

The MA(1) process, $Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$ can be written as:

$$Y_t - \mu = \varepsilon_t - \theta_1 \varepsilon_{t-1} = (1-\theta_1 L)\varepsilon_t \quad \Rightarrow \quad \frac{Y_t}{(1-\theta_1 L)} = \varepsilon_t$$

$$\frac{Y_t}{(1-\theta_1 L)} = Y_t + \theta_1 L + \theta_1{}^2 L^2 + \ldots + \theta_1{}^q L^q = \varepsilon_t$$

In this direction, so long as the invertibility condition is obtain for the MA(1) process, whenever $\theta_1 = \phi_1$, the moving average process is said to be another expression of an infinite autoregressive process. This condition exhibits the duality of autoregression and moving average process as discussed earlier. On a more general case, let us consider an MA(1) process: $Y_t = \varphi_1 y_{t-1} + \varepsilon_t$

$$Y_t - \varphi_1 Y_{t-1} = \varepsilon_t \quad \Rightarrow \quad (1-\varphi_1 L) Y_t = \varepsilon_t$$

$$Y_t = \frac{\varepsilon_t}{(1-\varphi_1 L)} \quad \Rightarrow \quad Y_t = (1-\varphi_1 L)^{-1} \varepsilon_t$$

$$Y_t = (1 + \varphi_1 L + \varphi_1{}^2 L^2 + \ldots + \varphi_1{}^p L^p)\varepsilon_t \quad \Rightarrow \quad Y_t = \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_1{}^2 \varepsilon_{t-2} + \ldots + \varphi_1{}^p \varepsilon_{t-p} \qquad (3.19)$$

By extension, if $Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$ and $\varepsilon_{t-1} = Y_{t-1} - \theta_1 \varepsilon_{t-2}$, and $Y_t = \varepsilon_t - \theta_1 (Y_{t-1} - \theta_1 \varepsilon_{t-2})$

$$Y_t = \varepsilon_t - \theta_1 Y_{t-1} + \theta_1{}^2 \varepsilon_{t-2} \qquad (3.20)$$

By further extension: $\varepsilon_{t-2} = Y_{t-2} - \theta_1 \varepsilon_{t-3}$, for which reason $Y_t = \varepsilon_t - \theta_1 Y_{t-1} - \theta_1{}^2 \varepsilon_{t-2}$

$$\Rightarrow \quad Y_t = \varepsilon_t - \theta_1 Y_{t-1} - \theta_1{}^2 Y_{t-2} - \theta_1{}^3 \varepsilon_{t-3} = \varepsilon_t - \sum_{i=1}^{\infty} \theta_1^{\,i} y_{t-i}$$

$$Y_t + \sum_{i=1}^{\infty} \theta_1{}^i y_{t-i} = \varepsilon_t \quad \Rightarrow \quad Y_t + \theta_1 Y_{t-1} + \theta_1{}^2 Y_{t-2} + \ldots = \varepsilon_t \qquad (3.21)$$

The MA process in (3.20) has being expressed as a function of the sum of a current and an infinite series of past observations. Therefore, equation (3.21) can be conceived of as a lag function of Y_t.

For an AR(1) process, if $|\phi_1|<1$, then the series in (3.19) and (3.20) converges to a solution. If $|\phi_1|=1$, there is a unit root and the process becomes a nonstationary random walk that does not stabilize. Also, the series becomes nonstationary and goes out of control if $|\phi_1|>1$. These parameter limits are called the bounds of stationarity for the AR(1) process. Moreover, higher order AR models also have these bounds of stationarity. For example, the AR(2) model, $Y_t = \mu+\phi_1 y_{t-1}+\phi_2 y_{t-2}+\varepsilon_t$, has three sets of statioanarity boundaries:

1. $|\phi_2|<1$ or $-1<\phi_2<1$
2. $\phi_1+\phi_2<1$
3. $\phi_2-\phi_1<1$

Similarly, higher order MA models have limits within which the process remains stable. For example, the MA(2) process $Y_t=\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}$ has the following bounds of invertibility:

1. $|\theta_2|<1$ or $-1<\theta_2<1$
2. $\theta_1+\theta_2<1$
3. $\theta_2-\theta_1<1$

In the case of nonlinear models with two roots, the roots of the AR equation $(1-\phi_1 L-\phi_2{}^2 L^2)Y_t=\varepsilon_t$ or that of the MA equation $(1+\theta_1 L+\theta_2{}^2 L^2)Y_t=\varepsilon_t$ must lie outside the unit circle. By using the quadratic approach, the requirement can be written as; $\left|\frac{\phi_1\pm\sqrt{\phi_1{}^2+4\phi_2}}{-2\phi_2}\right|>1$. This condition also affects higher order polynomial AR and MA models (Yaffee and McGee, 2000).

3.4.10 Normal (Gaussian) White Noise and Linear Time Series

A time series Y_t is called white noise if $\{Y_t\}$ is a sequence of independent and identically distributed random variables with finite mean and variance. In a particular situation, if Y_t is normally distributed with a known variance σ^2 and a mean of zero, then the series is said to be Gaussian white noise. It is denoted by; $\varepsilon_t \sim i.i.d..N(0, \sigma^2{}_{\varepsilon})$ (3.22)

Theoretically, all the Autocorrelation functions of a white noise series must be zero. However, in practice, if all sample ACFs are closer to zero, then the observed series is classified as a white noise series.

A time series Y_t is said to be linear if it can be written as $Y_t = \mu + \sum_{i=0}^{\infty} \varphi_i \varepsilon_{t-i}$ or $Y_t = \mu + \varphi_0 \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \ldots$ (3.23), where μ is the mean of Y_t, φ is a constant with $\varphi_0 = 1$, and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with mean zero and a well-defined distribution (ie. $\{\varepsilon_t\}$ is a white noise series). The general linear process depends on both past and future values of ε_t. However, a linear process which depends only on the past and present values of ε_t is said to be causal (Ruey, 2002, pp.26 – 28).

3.4.11 The Principle of Parsimony vs Simplicity

The principle of parsimony is a statistical concept originally credited to an earlier work by Tukey (1961). It states that the best chosen model for a given set of data is the very simplest models which can account for the observed properties of the data. By applying the principle to univariate models, a parsimonious univariate model would be a model which contains the minimum number of parameters necessary to yield white noise residuals. Parsimony in a wider sense requires statisticians to use the smallest number of parameters that will adequately represent a time series data. The parsimonious principle is always noted for guiding analyst to choose the simplest models, but care should be taken since too simple models might not consider some effects of the time series, especially in the area of intervention analysis (Reagan, 1984) and (Tukey, 1961).

3.5 Box-Jenkins ARIMA Process

The Box-Jenkins ARIMA process actually builds upon an earlier work done by Yule-Walker. Yule had already studied the AR and MA processes and was highly credited for these works. However, Box and Jenkins later developed a new model which combines the AR and MA processes with an integrated term. Their model was simply denoted as ARIMA (*p*, *d*, *q*), where *p* shows the AR process, *d* is the integrated term which takes care of nonsationary processes, and *q* represents the MA process. For example, in the fitted ARIMA(1,1,1) model, both the AR and MA processes shows a significant spike at the first lag and the process is integrated at the first order, I(1) to make it stationary. Theoretically, the integrated term I(*d*) needed to make the process stationary should not exceed the second difference or should not be integrated above the second order, I(2).

Moreover, the Box-Jenkins process of fitting and analyzing univariate time series models can mainly be categorized into three iterative stages namely: The Identification stage; Estimation stage; and the Diagnostic stage. The subsequent sections therefore explain each of these stages with detailed illustrations.

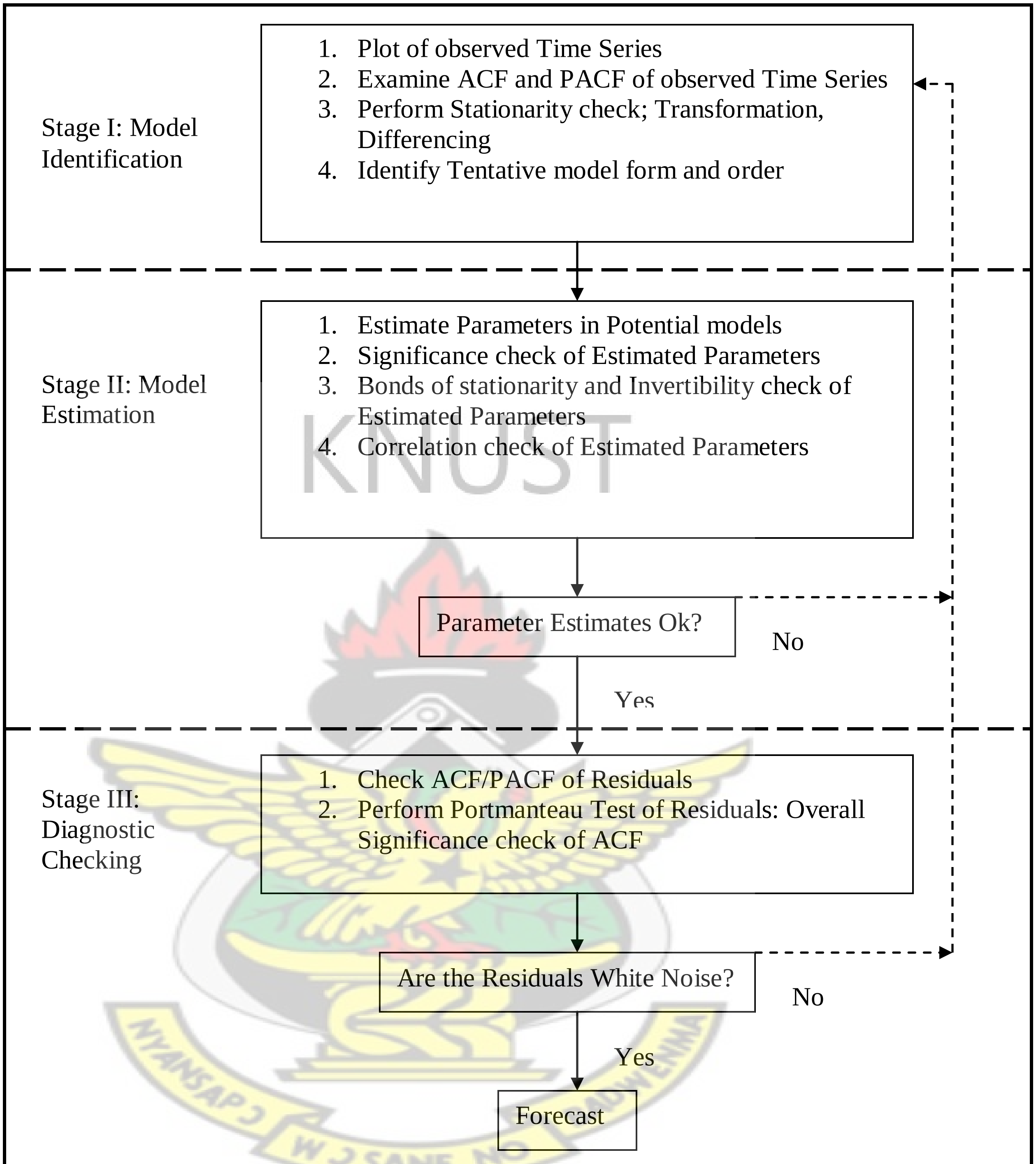

Figure 3.1: Diagrammatic representation of Box-Jenkins Time Series modeling procedure

3.5.1 Model Identification

The first step in fitting and analyzing a model for an observed time series is to plot the series against time, and verify the various data patterns shown from the graph of such plots. These data pattern might give the analyst clues as to the appropriate time series modeling procedure to be used.

However, the most important tools under the identification stage are the autocorrelation function (ACF) and the partial autocorrelation (PACF) of the observed series. The ACF of the observed series measures the degree of correlation between neighbouring observations in the series (for example, Y_t and Y_{t-1}), whereas the PACF measures the degree of association between current values of a series (Y_t) with earlier values of that same series (Y_{t-k}) when the effects of other time lags on the series are held constant.

For instance, assuming there is an autocorrelation between Y_t and Y_{t-1}, there will also be a significant correlation between Y_{t-1} and Y_{t-2}. Since Y_t and Y_{t-2} are related to Y_{t-1}, there will again be a correlation between Y_t and Y_{t-2}. In measuring the real correlation between Y_t and Y_{t-2}, there is the need to hold the effect of the intervening Y_{t-1} constant. In these circumstances, the partial autocorrelation is used in measuring such correlations. Basically, at the first lag, the autocorrelation and the partial autocorrelation are of the same magnitude. Table 3.1 presents a simple calculation of the autocorrelation (r_1) at lag 1 for Ghana's annual cocoa production series.

Table 3.1: Autocorrelation at lag 1 (Y_{t-1}) for the Cocoa Production series

t	Y_t	Y_{t-1}	$(Y_t - \bar{Y})$	$(Y_{t-1} - \bar{Y})$	$(Y_t - \bar{Y})^2$	$(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})$
1	207559	-	-161838	-	26191416870	-
2	278372	207559	-91024.6	-161838	8285482356	14731209130
3	247834	278372	-121563	-91024.6	14777471800	11065192350
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
61	680781	614532	311384.4	245135.4	96960384690	76331325530
62	710642	680781	341245.4	311384.4	116448406000	106258477800
63	632037	710642	262640.4	341245.4	68979966580	89624813260
64	1024533	632037	655156.4	262640.4	429229875700	172070516000
sum	23641384				1739476980000	1265291155000

$$\bar{Y} = \frac{\sum_{t=1}^{64} Y_t}{64} = \frac{23641384}{64} = 369396.625$$

$$r_1 = \alpha_1 = \frac{\sum_{t=1+1=2}^{64} (Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=1}^{64} (Y_t - \bar{Y})^2} = \frac{1265291155000}{1739476980000} = 0.73$$

From Table 3.1, the entire mean $(\bar{Y})$ of the data and the autocorrelation at lag 1 (r_1) is manually calculated to be 369396.625 and 0.73 respectively. Moreover, the partial autocorrelation at lag 1 (α_1) is also taken as 0.73 since the autocorrelation and the partial autocorrelation at lag 1 are of the same magnitude $(r_1 = \alpha_1)$. In all, the autocorrelation and the partial autocorrelation at lag 1, 2, 3, 4 and so on make up the ACF and PACF.

The ACF and PACF are mainly used by analysts to perform two major checks at the identification stage. A check of stationarity may be verified from the ACF and PACF plot of the observed time series. If the time plot shows that the data points oscillate around a finite mean or equivalently, the lags of the ACF and PACF plot quickly drop to zero or get nearer to the zero mark at a faster rate. Usually, this indicates the presence of stationarity in the time plot of the series. Also, when the data appears non-stationary, the lags of the ACF decay exponentially or shows a damped sine-wave pattern and often do not die out to zero. In such cases, the PACF also normally show a high significant spike close to 1 at the first lag, and the subsequent lags quickly decline closely to zero but do not die out to zero.

Most often, when a non-seasonal data appears to be non-stationary, a first difference is taken to restore stationarity. For most practical situations, a maximum of two consecutive differences will transform the data into a stationary series. After achieving stationarity in the time series, the analyst then re-examine the ACF and the PACF patterns to identify the form and order of tentative models. A pure AR or MA model or a mixture of the two may be revealed. If there are no significant autocorrelations after lag q, an MA(q) model may be considered. A pure AR(p) model may also be appropriate if there are no significant partial autocorrelations after lag p. In some situations, a mixture model comprising the AR(p) and MA(q) or ARMA(p, q) model may be necessary. If the series was differenced to achieve stationarity, the MA and AR models may be written in the form IMA(d, q) and ARI(p, d) respectively. Under such series, an integrated term may also be incorporated into the ARMA(p, q) model to obtain a new class of model known as the ARIMA(p, d, q) model. In summary, the behavior of the ACF and PACF for identifying simple tentative models are presented in Table 3.2

Table 3.2: Characteristics of the ACF and PACF for identifying pure AR and MA models

Proces	ACF	PACF
AR(p)	Tails off as exponential decay	Spikes at lag 1 to *p*, then cuts off to zero
MA(q)	Spike at lag 1 to *q*, then cuts off to zero	Tails off as exponential decay
AR(1)	Quickly tails off as exponential decay	Spike at 1, then cuts off to zero
MA(1)	Spike at lag 1, then cuts off to zero	Quickly tails off:damped sine wave/decay

In most cases, analysts may identify more than one plausible model. All these identified models are to be considered for the estimation and diagnostic checking stages before the best model that fits the observed series quite well is chosen. Analysts are advice not to choose the best model by their own preferences, but to subject candidate models to checks under the estimation and diagnostic stages in order to select a suitable model among the lot.

3.5.2 Model Estimation

After identifying a tentative model, the next stage would be to estimate the parameters in the model. The most popular methods for estimation are the least square estimates, the method-of-moments estimates and the maximum likelihood estimates. Preferably, the maximum likelihood estimation method was used in this study for estimating the AR and MA parameters, as well as all other parameters reported in the study.

The Akaike information criteria (AIC) and the Schwarz Bayesian information criteria (BIC) are the two main penalty function statistics which penalizes fitted models based on the principle of parsimony. These statistics were one of the various checks used to verify the adequacy of the chosen models. Comparatively, models with the smallest AIC and BIC are deemed to have residuals which resembles a white noise process. Twice the number of estimated parameters minus two times the log likelihood gives the AIC value of a model. The BIC is computed as $-2\ln(L)+\ln(n)k,$ where L is the likelihood, n denotes the number of residuals and k is the number of free parameters

For each parameter estimate, there will be a reported standard error for that particular parameter. From the parameter estimate and its standard error, a test for statistical significance can be conducted. Over here, a *t*-test, which is a test of whether a parameter is significantly different from zero, is used. The *t*-value of the test is computed from the ratio of each parameter estimate to its standard error. For statistically significant parameters, the absolute values of the *t*-ratios are expected to be greater than 1.96. From the test, significant parameters are to be maintained whereas parameters which are not significant must be trimmed from the model. The *t*-test statistics for the AR and MA parameter estimates are given in (3.24) and (3.25) respectively.

$t=\dfrac{\theta_p}{s.e.(\theta_p)}$ (3.24) ; $t=\dfrac{\phi_p}{s.e.(\phi_p)}$ (3.25), where θ_p and ϕ_q are the estimated p^{th} and q^{th} coefficients, and $s.e.(\theta_p)$ and $s.e.(\phi_q)$ denote the standard error of θ_p and ϕ_q respectively.

Again, the estimated AR and MA parameters must also conform to certain boundary conditions as already discussed in section 3.4.9. If the AR and MA parameters does not lie within those bounds of stationarity and invertibility, then the parameters of the model in question needs to be re-estimated or better still a different candidate model is alternatively considered for estimation.

In conclusion, if the parameters of a candid model are subjected to all the checks discussed above and the result proof to be reliable, such model would then be considered for the diagnostic checking stage else, the analyst has no better option than to re-start the process from the identification stage in section 3.5.1.

3.5.3 Diagnostic Checking

The main objective under the diagnostic stage of the Box-Jenkins ARIMA process is to examine whether the fitted model follows a white noise process. One way of achieving this objective is to study the autocorrelation values (r_k) one at a time, and to develop a standard error formula to test whether a particular r_k value is significantly different from zero. Theoretically, it is envisage that all autocorrelation coefficients for a series of random numbers must be zero. However, because of the presence of finite samples, each sample autocorrelations might not be exactly zero. The ACF coefficients of white noise data is said to have a sampling distribution that can be approximated by a normal curve with mean zero and standard error of $\frac{1}{\sqrt{n}}$, where n gives the number of data points in the observed series. For a white noise process, 95% of all sample autocorrelation values (r_k) must lie within a range specified by the mean plus or minus 1.96 standard errors. In this case, since the mean of the process is zero and the standard error is $\frac{1}{\sqrt{n}}$, one should expect about 95% of all sample autocorrelation values (r_k) to be within the range of

$\pm 1.96\sqrt{n}$ or $(-1.96\sqrt{n} < r_k < 1.96\sqrt{n})$. If this condition does not hold, then the model fitted does not follow a white noise process, or the residuals are not white noise. The correlogram of the ACF would therefore show lines at the critical values of $\pm 1.96\sqrt{n}$ for easily verification.

Rather than studying the autocorrelation values (r_k) one at a time, an alternative approach is to consider a whole set of r_k values say, the first 12 to 24 values and develop a test to verify whether the set of r_k values is significantly different from zero using a portmanteau test. The Ljung-Box test is a modified version of the portmanteau test statistic developed by Ljung and Box (1978). Their test-statistic has a distribution closer to the chi-square distribution rather than an earlier portmanteau test statistic proposed by Box and Pierce. In this study, the modified Ljung- Box Q statistic was used to test the adequacy of the model selected. It tests whether the model's residuals have a mean of zero, constant variance and serially uncorrelated r_k values (a white noise check). The test statistic is given by;

$$Q = n(n+2)\sum_{k=1}^{h}\frac{r_k^{\ 2}}{(n-k)} \qquad (3.26)$$

where n denote the number of data points in the series, $r^2{}_k$ is the square of the autocorrelation at lag k, and h is the maximum lag being considered. The hypothesis to be tested is formulated in the form;

H_0: The set of autocorrelations for residual is white noise (model fit data quite well)

H_1: The set of autocorrelations for residual is different from white noise

The test statistic (Q) is compared with a chi-square distribution written as $\chi^2_{\alpha,(h-p-q)}$, where α is taken to be 5% (0.05), h is the maximum lag being considered, and p and q are the order of the AR and MA processes respectively. The decision is to accept the null hypothesis (H_0) if $Q < \chi^2_{\alpha,(h-p-q)}$, and to reject the alternative hypothesis if $Q > \chi^2_{\alpha,(h-p-q)}$. In other words, the residuals are not white if the test statistic Q lies in the extreme 5% $(\alpha = 0.05)$ of the right-hand tail of the chi-square distribution.

It is important to check whether the residuals of the model adopted are normally distributed with constant variance and a zero mean. This can be done through a residual plot against time, and an objective test called the Shapiro-Wilk Normality test. A residual plot of a histogram and a normal probability graph can also be used to confer normality on the series under consideration. The hypothesis of the Shapiro-Wilk test is given as:

H_0: The error terms in the model are normally distributed

H_1: The error terms in the model are different from a normal distribution

The test is deemed significant if the reported p-value is less than 0.05, else it is not significant. In all, if a candidate model violates these assumptions at the diagnostic stage, the analyst is advice to restart the Box-Jenkins process of identification, estimation and diagnostic checking, and continues in that cycle until a plausible or an appropriate model is obtain. Afterwards, the analyst may then decide to forecast using such model (Reagan, 1984) and (Makridakis et al., 1998).

3.6 Intervention Time Series Analysis

Intervention analysis or impact analysis is a special case of dynamic regression models which is mainly use by analysts to assess the impact or effect of special events or circumstances such as policy changes, strikes, advertising promotions, environmental regulations and many similar events which are normally referred to as intervention events. It permits the study of inputs and output phenomena in the time domain. It can be applied in the modeling of regime changes, contingencies, or even outliers in time series analysis. The impact response model is formulated as a regression function. The regression model contains independent variables consisting of an ARIMA noise model and an intervention function, whereas the independent variable represents the response series Y_t. More specifically, the response variable Y_t is a function of the preintervention ARIMA noise model plus the inputs function of the deterministic intervention indicator for each of the interventions being modeled: $Y_t = \sum_t \sum f(I_t) + N_t$ (3.27), where $f(I_t)$ is the intervention function of a discrete deterministic intervention indicator at time *t*, and N_t is the ARIMA preintervention model or the noise model.

3.6.1 The Assumptions underlying Intervention Analysis Models

Intervention analysis or impact analysis models are characterized by certain assumptions which guide analysts to practically fit these models. The accuracy of impact analysis is contingent upon the fulfillment of these assumptions.

1. The system in which the input event and the impact response take place is usually assumed to be closed. Apart from the noise model of the series, the only exogenous impact on the series is presumed to be that of the intervention event. In

this case, all other factors are presumed to be the same whiles the intervention event alone precipitates the impact.

2. The temporal delimitations of the input event are presumed to be known. The time of onset, the duration, and the time of termination of the input event have to be identifiable. Moreover, the noise model that describes the preintervention series has to be stable with a mean of zero.

3. Another assumption is that there should be enough observations in the series before and after the onset of the event for the analyst to separately model the preintervention and post intervention series by whichever parameter estimation process the analyst chooses to use. In the conventional approach where there is enough dataset prior to the intervention event, the preintervention series is modeled first, and the impact is modeled afterward.

3.6.2 Intervention Indicator Variables

The intervention indicator is an exogenous variable whose discrete coding represents the presence or absence of an input event. The intervention function $f(I_t)$ can be coded as either a step function or a pulse function depending on the onset and duration of the input event. If $f(I_t)$ is a step function, then the intervention indicator is coded 0 prior to the beginning of the event and as 1 at both the onset (T) and for the entire duration of the presence of the event.

$$f(I_t) = s_t^{(T)}, \text{ when } s_t^{(T)} = \begin{cases} 1, & if \quad t \geq T \\ 0, & if \quad t < T \end{cases} \tag{3.28}$$

where $s_t^{(T)}$ denote a step function and T is the onset time of the input event.

In another situation $f(I_t)$ can also be modeled as a conventional pulse function. In such circumstances, the intervention indicator is coded as 0 prior to the event and immediately after the event, and as 1 at the onset of the event.

$$f(I_t) = p_t^{(T)}, \text{ when } \; p_t^{(T)} = \begin{cases} 1, & if \quad t = T \\ 0, & if \quad t \neq T \end{cases} \tag{3.29}$$

In (3.29), $p_t^{(T)}$ represents a pulse function, whereas T is the time the input event occurred.

Example of such coding for the indicator variables representing the step and pulse input functions are presented in Table 3.3. The coding indicates the presence or absence of a particular input event; say a sales promotion exercise which was started in the year 2000. In Table 3.3, the onset of the intervention input begins in 2000 ($t=T$) where it is coded as 1, and remains for just a period in the case of the pulse function, but remains as 1 for the entire presence of the intervention exercise in the case of the step function.

Table 3.3: Intervention Indicator coding for the pulse and step input functions

Pulse function		Step function	
Time(t)	$p_t^{(T)}$	Time(t)	$s_t^{(T)}$
1980	0	1980	0
1981	0	1981	0
1982	0	1982	0
.	.	.	.
.	.	.	.
2000	1	2000	1
2001	0	2001	1
2002	0	2002	1
.	.	.	.
.	.	.	.
2011	0	2011	1

The step and the conventional pulse functions are input variables which are sometimes interrelated. A transformed step function usually reverts to a pulse function or in other words, a differenced step function gives a pulse function. For example, let us consider a step function as follows;

$$I_t = \{\ldots 0, 0, 0, 1, 1, 1 \ldots\}$$

By differencing the step function, we get;

$$= \{\ldots (0-0),\ (0-0),\ (0-0),\ (1-0),\ (1-1),\ (1-1) \ldots\}$$

$$= \{\ldots 0, 0, 0, 1, 0, 0, 0 \ldots\}$$

It can be noted that $(1-L)s_t^{(T)} = P_t^{(T)}$. This process then suggest that step function models $S^{(T)}$ could equally be well represented in terms of pulse transfer models $P^{(T)}$.

However, we now consider the two intervention events for this study as step intervention functions due to their occurrences. It is then hypothesize that each programme produces a positive impact in terms of production levels. In line with this, the cocoa mass spraying and the Hi-tech intervention events were postulated as;

$$Y_t = c + w_1 I_{1t} + w_2 I_{2t} + \frac{\theta(L)}{\phi(L)}\varepsilon_t, \tag{3.30}$$

where

$$I_{1t} = S_t^{(2002)} = \begin{cases} 1, & t \geq 2002 \\ 0, & otherwise \end{cases}$$

$$I_{2t} = S_t^{(2003)} = \begin{cases} 1, & t \geq 2003 \\ 0, & otherwise \end{cases}$$

c is a constant and Y_t is the level of change with respect to gains or losses made in the volume of production. The intervention variable I_{1t} is a step function which corresponds to the mass spraying exercise, whereas I_{2t} represents another step function for the cocoa Hi-technology programme.

3.6.3 Dimensions of Impact Assessment or Intervention Models

Generally, an impact assessment model may be formulated as; $Y_t = \frac{w_0}{(1-\delta L)} I_{t-b} + N_t$ (3.31), where I_{t-b} is the intervention indicator variable normally known as the change agent, scored 0 or 1 for the absence or presence of the intervention event and the subscript b is a possible time delay for the impact to take off. w_0 is the impact parameter which indicates the magnitude of the impact, and δ represent the decay parameter, whereas N_t is the noise model. At certain situations, the response series Y_t may not quickly observed the impact of the intervention event. The b index in I_{t-b} gives the number of periods delayed between the onset of a known intervention and the actual time it's impacted on the response series (Y_t). If b is assign a value of 2, there would be exactly two time periods of delay between the intervention event I_t and the time it takes for its impact to be fully realized on the response series Y_t.

Moreover, there are two major dimensions characterized by impact assessment. These are usually observed by the duration and nature of the impacts. Some interventions could give temporary or permanent effects with respect to the duration. The nature of impacts can also be seen as abrupt or gradual processes. Sudden and constant changes (abrupt permanent) are normally attributed to step functions; sudden and instantaneous changes (abrupt temporary) are modeled with pulse function; gradual and permanent effects are mainly modeled with step function with first-order decay rate; gradual and decaying changes (pulse decay) can also be modeled with pulse function with first-order decay rate.

The abrupt onset and permanent duration effects are popularly called a simple step function. Step functions are mainly used to model permanent changes in the response series (Y_t). A step function with a first-order decay rate may be written as;

$$Y_t = \frac{w_0}{(1-\delta_1 L)} I_{t-b} + N_t \tag{3.32}$$

If after fitting the model in (3.32) the denominator reduces to unity, the model will then be called a simple step function with a zero-order decay, where $f(I_t) = s_t^{(T)} = w_0 I_{t-b}$. Also, if there is no time delays $(b=0)$, then $f(I_t) = s_t^{(T)} = w_0 I_t$, and the full model will now be of the form;

$$Y_t = w_0 I_t + N_t \tag{3.33}$$

However, the gradual permanent effects are characterized by slow changes in the level of the series that usually result in a new permanent level. It is usually modeled with step functions with first-order decays as shown in (3.32). If the index $b=0$ in (3.32), the gradual permanent effect model then becomes; $Y_t = \frac{w_0}{(1-\delta_1 L)} I_t + N_t$ (3.34), where $-1 < \delta_1 < 1.$

Again, if the noise model (N_t) is subtracted from the response series (Y_t), then

$$Y_t - N_t = \frac{w_0}{(1-\delta L)} I_t \text{ or } Y_t^* = \frac{w_0}{(1-\delta L)} I_t .$$

We then expand and simplify to obtain;

$$Y_t^*(1-\delta L) = w_0 I_t$$

$$Y_t^* - \delta L Y_t^* = w_0 I_t$$

$$\Rightarrow \quad Y_t^* - \delta Y_{t-1} = w_0 I_t .$$

Since the impact at t is w_0, then the impact at $t+1$ is given by;

$$Y^*_{t+1} = \delta Y_t^* + w_0 I_{t+1} = w_0(1+\delta) .$$

The impact over time or the change level obtained from the gradual permanent effect is given by;

$$w_0(1+\delta+\delta^2+\delta^3+\delta^4+\delta^5+\delta^6 \ldots) I_t$$

Therefore, $Y^*_{i+n} = \sum_{k=0}^{n} \delta_1{}^k w_0$. Again, since $\delta < 1$, and δ^{10} or δ^{100} is infinitesimal, then the effect as $t \rightarrow \infty$ gets very smaller. In all, the asymptotic or the long term change given by a gradual permanent effect is of the form; $\frac{w_0}{1-\delta}$

Temporary effects are often modeled with pulse functions. The abrupt onset and temporary duration effects are often called the "pulse effect", simply modeled as; $Y_t = w_0 I_{t-b} + N_t$, where I_{t-b} is the intervention indicator coded 0 prior to the event and 1 at the onset, and b may indicate a possible time delay between Y_t and I_t. The gradual temporary effects are often modeled with pulse functions having first-order decay rates. It is also formulated as;

$$Y_t = \frac{w_0}{(1-\delta_1 L)} I_t(1-L) + N_t \qquad (3.34)$$

In summary, Figure 3.2 shows graphical outputs for some intervention events that were modeled with the step and pulse intervention functions discussed in this subsection. For example, the $Y_t = (wL)S_t^{(T)}$ in Figure 3.2(a) may be used to indicate the presence of a permanent step change with unknown level of magnitude w after time T. The model $Y_t = \left(\frac{wL}{1-\delta L}\right)S_t^{(T)}$ in Figure 3.2(b) also corresponds to a gradual permanent change with decay rate δ, which later results in a long-term change in level given by $\frac{w}{1-\delta}$. Again, $Y_t = \left(\frac{w_1 L}{1-\delta L}\right)P_t^{(T)}$ represents a gradual temporary change after time T of unknown magnitude w_1 and decay rate δ. More complex intervention responses are usually modeled with various combinations of any of the simpler forms as shown in Figure 3.2(f).

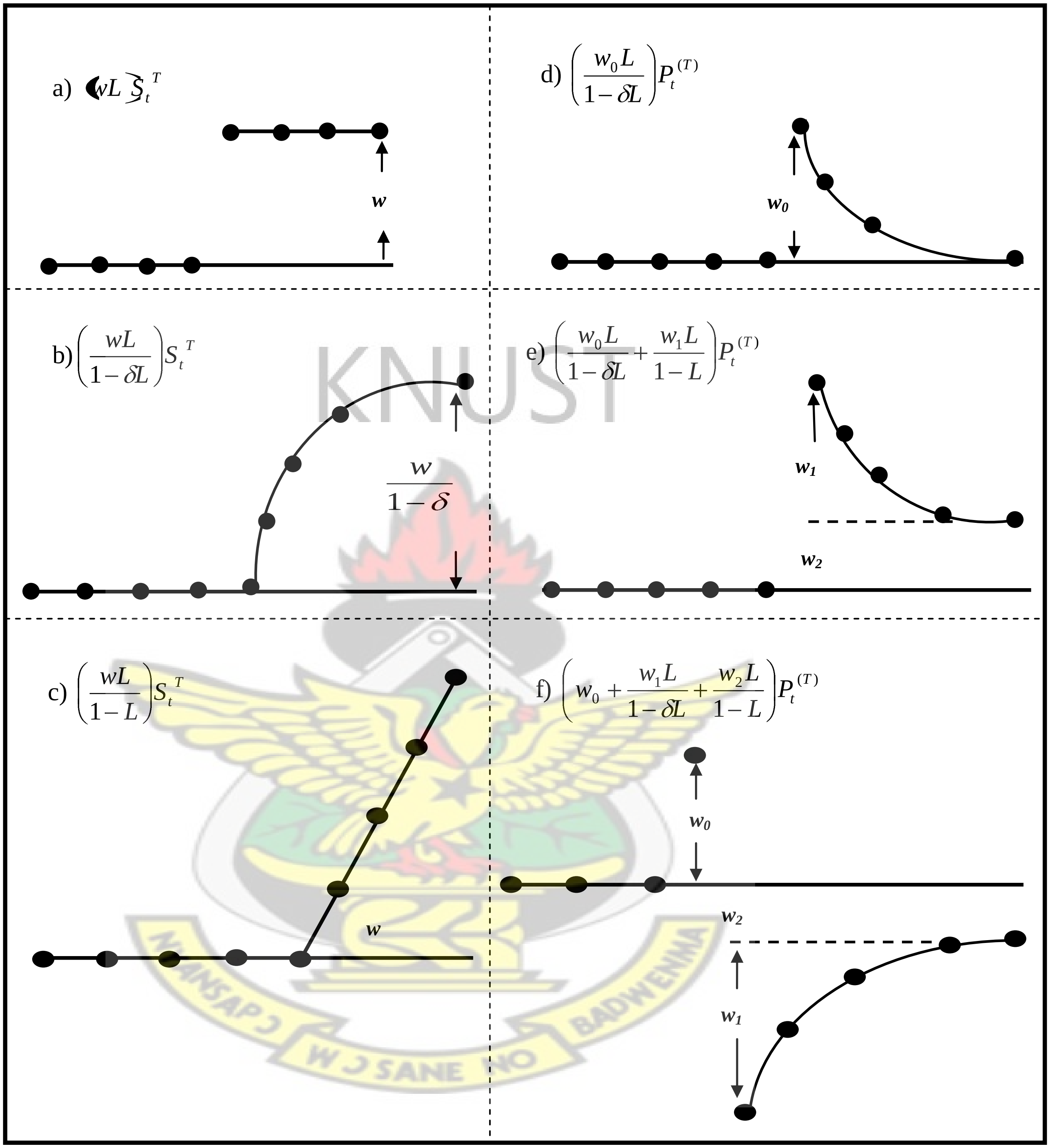

Figure 3.2: Responses to step and pulse intervention inputs: (a), (b), (c) response to a step input; (d), (e), (f) response to a pulse input

3.6.4 Estimation and Diagnostic Checking for the Impact Assessment Models

By considering the general impact assessment model in (3.31), the main parameters to be estimated are the impact (w_0) and the decay (δ) parameters using any chosen estimation method. The time delay, indexed *b* in (3.31) is usually deduced from the spikes of the time plot. For example, if the size of the spikes in the time plot becomes pronounced at the third lag after impact, then the delay time should be set to three time periods. Preferably, the maximum likelihood estimation method was used to estimate the intervention parameters as well as that of the noise, N_t.

An estimated value for w_0 computed as $\frac{\sum_T Y_t I_t}{\sum_T {Y_t}^2}$, would apparently indicate the size of rise or drop in the level of the response series. Also, the delta (δ) parameter associated with the gradual temporary and the gradual permanent effects estimates the adjustments subsequent to the change (pulse decay) or the overall rate of increase (gradual permanent), and is expected to be within the bounds of system stability $(-1 < \delta < 1)$. Moreover, estimation of the impact parameters should reflect what is known and observed about the impact of the intervention event.

Once a tentative impact model has been estimated, the next is to conduct diagnostic checks on the significance of the hypothesized parameters and the behaviour of the residuals. A *t*-test statistic is used to check the significance of all parameters reported from the estimated impact model. Any parameter whose *t*-value is less than 1.96 is deemed to have a nonsignificant effect and must be quickly trimmed from the model. The residuals should also be examined to verify whether it follows a white noise process or otherwise. When the residuals are tested to be white noise, then the adequacy of the impact or intervention model will be fully established. Thereafter, the analyst then interprets the effect of the intervention event as reported by its estimated hypothesized parameters (Yaffee and McGee, 2000) and (Box et al., 1994).

CHAPTER 4

ANALYSIS AND DISCUSSION OF RESULTS

4.0 Introduction

The chapter basically discusses the general pattern of Ghana's annual cocoa production time plot from the year 1948 to 2011. The discussion of the time plot was alienated into four main production periods. The data pattern of the preintervention period (1948-2001) was also analyzed and compared to that of the postintervention period (2002-2011). The chapter then analyses the various results and the models obtained from the Box-Jenkins process of fitting a Noise model for the preintervention data. Thereafter, a full-fitted impact model or an intervention time series model for the entire annual cocoa production series (1948-2011) was also discussed and thoroughly analysed.

4.1 Preliminary Analysis of Ghana's Cocoa Production Series

Figure 4.1 shows the time plot of Ghana's annual cocoa production figures from 1948 to 2011. Table 1 in Appendix C displays the dataset used for this time plot. The plot indicates a steady period of growth in production from 1948 to 1959. Production significantly shot up from 1960 to 1961. There was almost a stagnant growth in production outputs as recorded throughout 1962 to 1964. In 1965, there was a swift increase in production as observed from the time plot, but production output quickly dropped drastically in 1966. Production then continued to decline gradually from 1967 to 1984, where Ghana's lowest production output was recorded since 1948. The decline in production at this period was mainly linked to the incidence of diseases and pests infestation which were reported to be very high in the Ashanti and Brong Ahafo regions around the early part of 1980's. There were slight increases in production characterized by sharp decreases observed from 1985 to around 2001.

Ghana's production level rapidly increased from the year 2002 to 2011. These increases were engulfed with a few sharp decreases in 2005, 2007 and 2010 with an all time peak of production output recorded in the year 2011. Perhaps, the massive turnaround in production from 2002 to 2011 could be attributed to the Government's mass spraying and cocoa Hi-technology programmes initiated in early part of 2000's. Relatively, production periods from 2002 throughout 2011 witnessed more swift increases as compared to periods ranging from 1948 to 2001, which only showed few increases with more decreases, and at certain period a steady form of growth.

In all, Ghana's annual cocoa production time plot can be put in four main levels of production periods: a steady growth period from 1948 to 1959; an apparently increasing period from 1960 to around 1976 with a peak at 1965; a fast decreasing period from 1977 to 1984; and a gradual increasing period from 1985 to 2001, with rapid increases recorded around 2002 throughout to the year 2011.

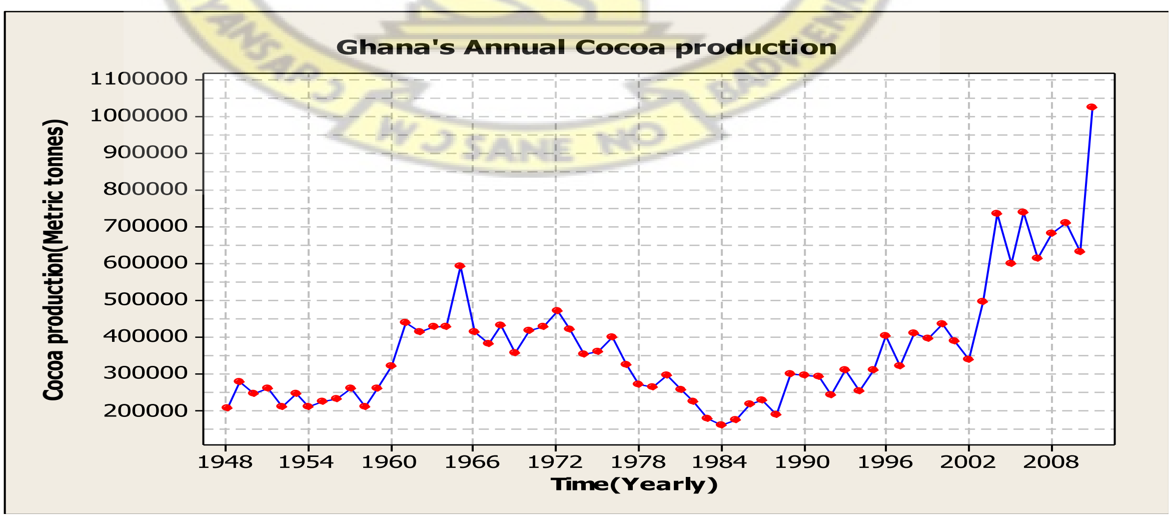

Figure 4.1: Ghana's annual cocoa production time plot from 1948 to 2011

4.2 Model Identification Process

The model identification process is where the form and order of tentative models are basically selected. The form and order of these models are picked from the sample autocorrelation function and the partial autocorrelation function of the observed series. However, such observed series must be stationary before tentative models are selected. From Figure 4.1, the pattern of the preintervention part of the cocoa production series (1948–2001) does not show much evidence of any irregularities in the time plot. Moreover, we shall use the ACF plot and two objective test statistics to check whether the preintervention part of the series is stationary or otherwise.

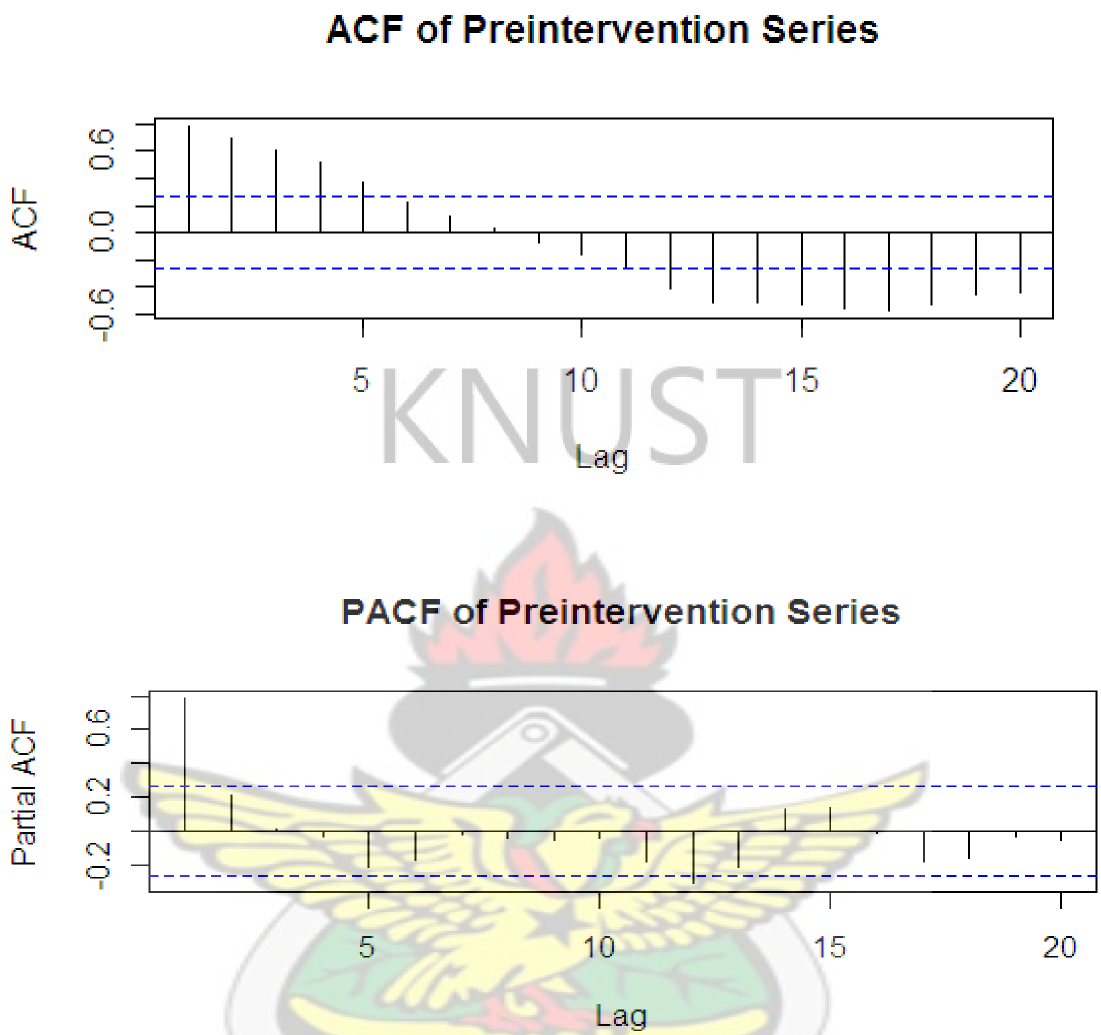

Figure 4.2: Empirical ACF and PACF of the Preintervention (1948–2001) part of the Cocoa Series

In a process of assessing the degree of dependence in any series and selecting a model for such series, one of the most important tools time series analysts normally use is the sample ACF of that particular series. The ACF can be used to detect stationary and nonstationary series. If a series is stationary, its ACF rapidly die out to or near zero. A nonstationary series mainly have high positive lags which slowly decay to or near zero.

Critical observation from Figure 4.2 shows that the autocorrelation function of the preintervention series attenuates at a quicker rate to nearly zero, whiles the partial autocorrelation function shows a significant non-zero spike at the first lag and thereafter geometrically decays to zero. This depicts a clear situation of a series which is stationary in the mean and does not require any form of differencing or transformation. Based on Figure 4.2, one could see quick decay in the ACF and a single significant spike around 0.78 in the PACF. 0.78 seems enough less than 1 for us to reasonably accept that, an AR(1) process might bestow a good fit for the preintervention part of the cocoa production series. In order to verify the adequacy of the AR(1) process, we shall slightly over-fit this model with an AR(2) process and later perform diagnostic checks to obtain the best fitted model.

Table 4.1 Unit Root and Stationarity Tests for the cocoa preintervention series

	Summary of Test Statistic		
	Test Statistic	Lag order	*p*-value
Test Type			
ADF	-4.164	12	0.01
KPSS	0.2853	1	0.1

Table 4.1 presents the Augmented Dickey-Fuller test (ADF) for the null hypothesis of a unit root against the alternative of a stationary series. It also gives the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test for the null hypothesis of a level stationary against an alternative of unit root.

From Table 4.1, the lag order of the ADF test was set to 12. This is supported by the Monte Carlo experiments which suggest that, it is better to error on the side of including too many lags than to bias the test by choosing a very small lag length. However, the Dickey-Fuller test statistic and its *p*-value reported in Table 4.1 rejects the null hypothesis of a unit root at 5% significance level. The alternative hypothesis of a stationary series is therefore accepted. This implies that the preintervention series can be perceived as an integrated series of order zero, I(0).

On the other hand, the KPSS test statistic was recorded as 0.2853 with a *p*-value of 0.10 as presented in Table 4.1. The test failed to reject the null hypothesis of having a level stationary series. Altogether, the two objective tests and the ACF of the observed series confer stationarity in the preintervention cocoa production series.

4.3 Estimation and Significance Checks of Tentative Models Parameters

The parameters of the selected ARIMA(1, 0, 0) and the slightly over-fitted ARIMA(2, 0, 0) models were estimated using the maximum likelihood estimation method, with the R statistical software package. The results from the two estimated models are presented in Table 4.2 and Table 4.3.

Table 4.2: Parameter Estimates for ARIMA(1, 0, 0) model

		Model Fit Statistics		
AIC	AICc	BIC		
1339.92	1340.4	1345.89		
Coefficients		Estimate	STD Error	*t*-value
ar1		0.7966	0.0804	9.90796
Intercept		313689.72	35313.86	8.882907

Table 4.3: Parameter Estimates for ARIMA(2, 0, 0) model

		Model Fit Statistics		
AIC	AICc	BIC		
1338.91	1339.73	1346.87		
		Estimate	STD Error	*t*-value
Coefficients				
ar1		0.6071	0.1325	4.58189
ar2		0.2371	0.1344	1.76414
Intercept		315106.1	42767.09	7.36796

From Table 4.2, the *t*-value of the ar1 coefficient (ϕ_1) indicates a statistically significant coefficient which needs to be maintained in the model. Again, the estimated ar1 coefficient of 0.7966 (0.0804) reported in Table 4.2 strictly conforms to the bounds of parameter stationarity.

The *t*-test conducted on the ARIMA (2, 0, 0) coefficients was not statistically significant for one of the coefficients as presented in Table 4.3. The ar2 coefficient (ϕ_2) is not significantly different from zero, and as a matter of urgency needs to be trimmed from the model. Nonetheless, all the coefficients of ARIMA (2, 0, 0) do not lie outside the bounds of parameter stationarity.

Comparatively, the estimated AIC and BIC does not clearly favour any of the two models under consideration, but the *t*-test of whether a parameter is significantly different from zero, faulted ARIMA (2, 0, 0) as one of its estimated parameters was not statistically significant.

4.4 Diagnostic Checking Process for the Estimated Models

After estimating the parameters of tentative models, the next stage would be to check the adequacies of the estimated models. Models which exhibit adequate diagnostic checking criteria are usually deemed to be of good fit. Under this subsection, the residuals of ARIMA(1, 0, 0) and ARIMA(2, 0, 0) models were checked for adequacies or otherwise, using the ACF and PACF of the residuals and other objective diagnostic checking tests.

Table 4.4: Ljung-Box Test for ARIMA(1, 0, 0) model

	Summary of Test Statistic		
	X-squared	df	*p*-value
Test Type			
Ljung-Box	20.3258	24	0.6781

Table 4.5: Shapiro-Wilk Normality Test for ARIMA(1, 0, 0)

	Summary of Test Statistic	
	W	*P* -Value
Test Type		
Shapiro-Wilk	0.9582	0.05728

From Table 4.4, the Ljung-Box test does not reject randomness of the error terms based on the first twenty-four autocorrelations of the residuals. This indicates that the residuals of the fitted AR (1) model are white noise, and for that matter the model fits the series quite well. The histogram of the residual plot shown in Figure 1 of Appendix A for the fitted AR (1) model exhibit a mound shape distribution. However, the evidence is not sufficient to reject normality, as presented in the normality probability plot at the bottom display of Figure 1 in Appendix A, and the Shapiro-Wilk normality test result in Table 4.5.

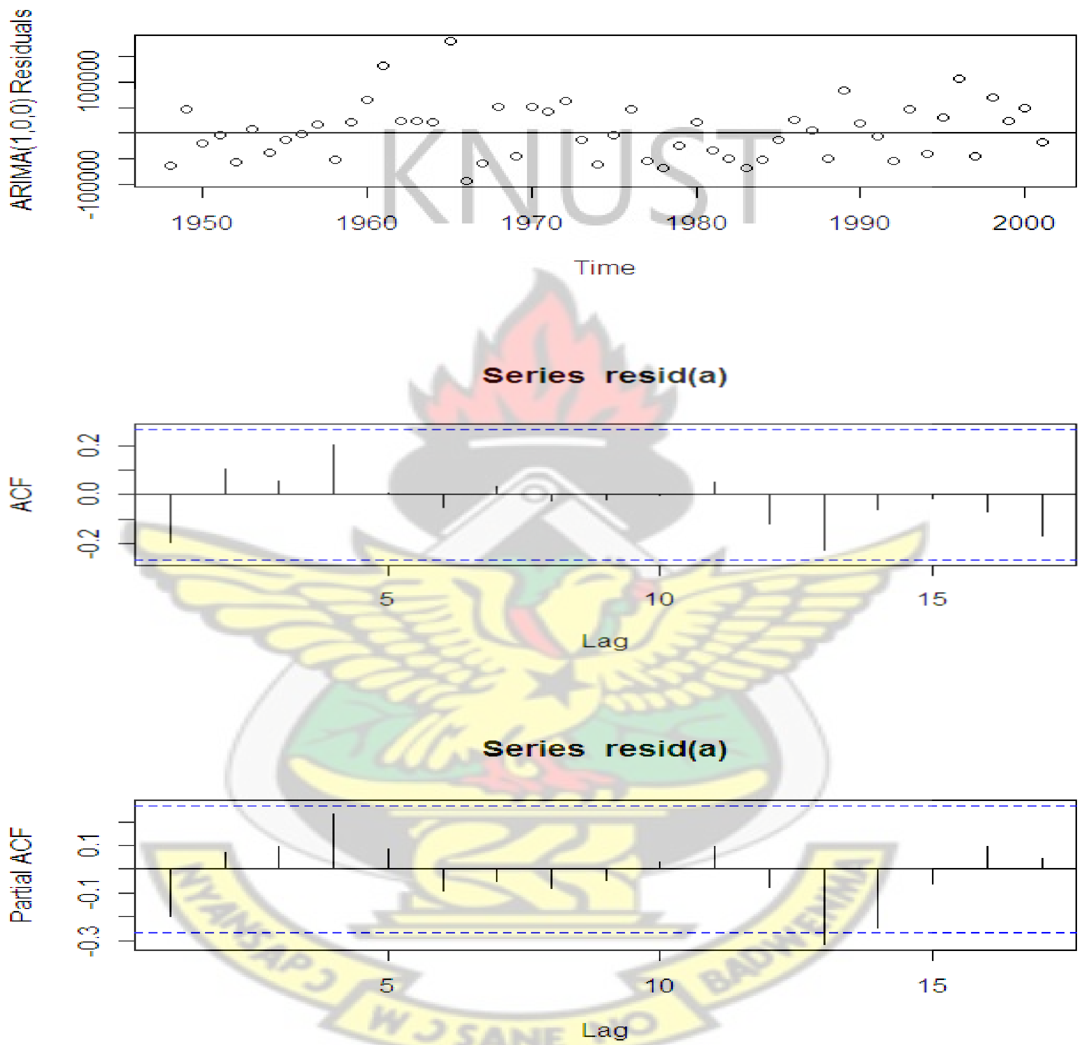

Figure 4.3: Diagnostic Residual plots of ARIMA (1, 0, 0) model

The top display of Figure 4.3 shows a time plot of the residuals for ARIMA (1, 0, 0) model. The only possible outlier seems to occur in 1965. The plot generally shows no clear pattern, and may be conceived of as an i.i.d sequence with a constant variance and a zero mean. The middle part of Figure 4.3 gives the ACF plot of the residuals. There is no evidence of a significant spike in the ACF plot. This shows that the residuals of ARIMA (1, 0, 0) follow a white noise process. Also, the PACF plot in the bottom display of Figure 4.3 confirms white noise residuals for the fitted model.

Table 4.6: Ljung-Box Test for ARIMA(2, 0, 0) model

	Summary of Test Statistic		
	X-squared	df	*p*-value
Test Type			
Ljung-Box	21.7382	24	0.5949

Table 4.7 Shapiro-Wilk Normality Test for ARIMA(2, 0, 0)

	Summary of Test Statistic	
	W	*p* -value
Test Type		
Shapiro-Wilk	0.951	0.02742

The Ljung-box objective test result in Table 4.6 confirms white noise residuals for the over-fitted ARIMA (2, 0, 0) model with a *p*-value of 0.5949. From the bottom display of Figure 1 in Appendix B, although the normal probability plot looks like a normal distribution, the top display shows a distribution which deviates from normality. The Shapiro-Wilk test result in Table 4.7 bears out the normality deviation of the model with a *p*-value of 0.02742.

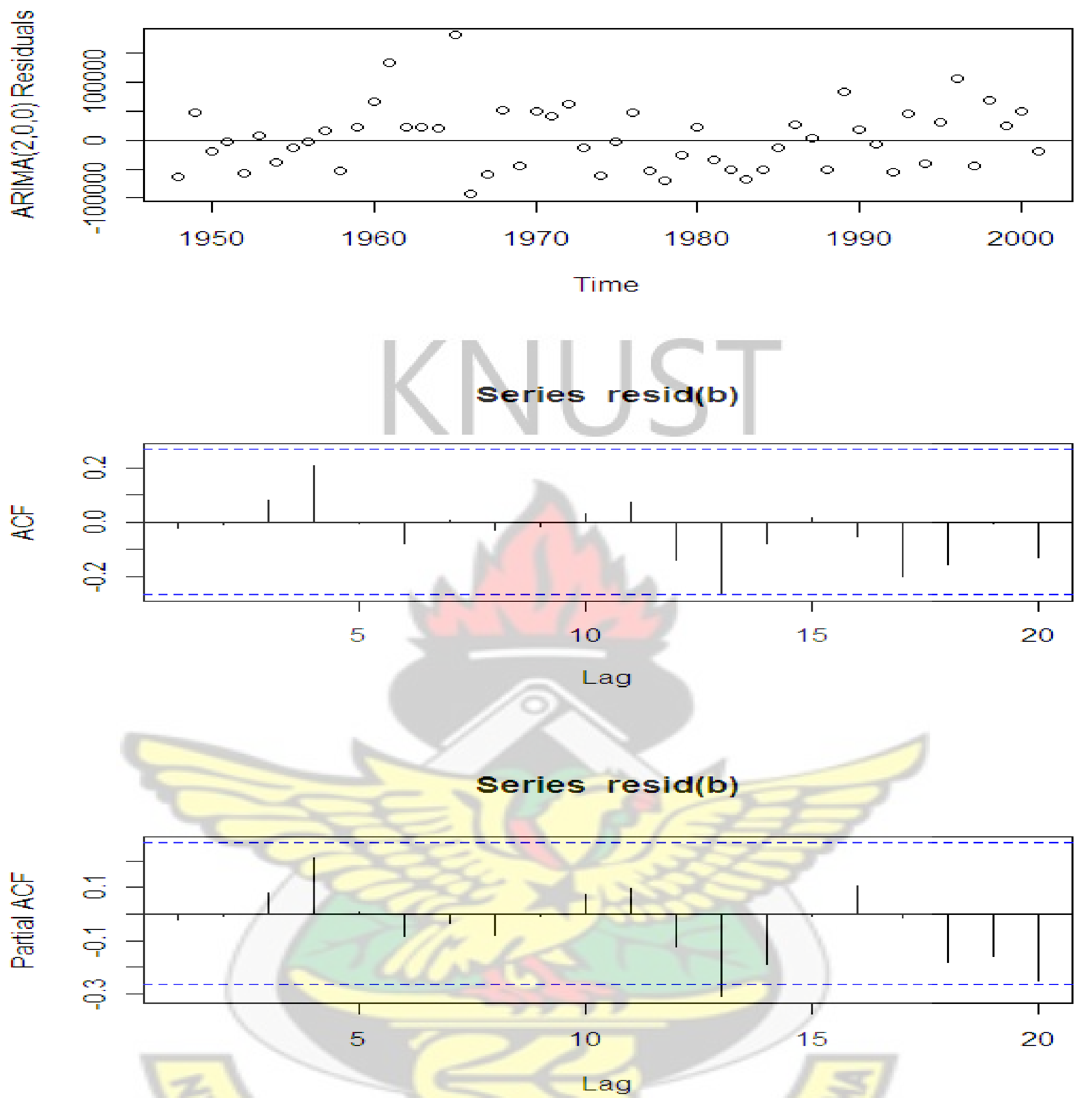

Figure 4.4: Diagnostic Residual plots of the over-fitted ARIMA(2, 0, 0) model

From Figure 4.4, none of the three diagnostic residual plots exhibits difficulties with the fitted model. There is only an outlier recorded in 1965 in the residual plot. The ACF and PACF of the residuals indicate a white noise process for the fitted ARIMA(2, 0, 0) model.

In all, each of the two candidate models has presented a good case of being the best fitted model to the preintervention series. There were not enough flaws for the models at the diagnostic checking stage, except for the normality of ARIMA(2, 0, 0) residuals, and an outlier in 1965 for both models. The two models each follow a white process. The reported penalty function statistics (AIC and BIC) does not clearly favour any of these two candidate models. However, the ar2 coefficient (ϕ_2) of ARIMA(2, 0, 0) was not significant at the 5% level of significance, and could have negative effect on the performance of the model. This therefore, deny ARIMA(2, 0, 0) of being the best fitted model for the series. Besides, the best fitted model with significant parameter estimates and adequate diagnostic checking criteria was that of ARIMA(1, 0, 0) model.

4.5 Estimation and Diagnostic Checks for the Full Intervention Model

After fitting a noise model for the preintervention series, the model and the hypothesized intervention model's parameters would be estimated and later checked for adequacies. From here, the noise ARIMA(1, 0, 0) model of the preintervention cocoa series was estimated together with a dichotomous intervention function using the R software package. The results of the estimated parameters are presented in Table 4.9.

Table 4.8: Za Test for possible break point position in the cocoa series

	Summary of Test Statistics		
	p-value	Test Statistic	Critical Values
	2.20E-16	-3.6517	0.01 = -5.57
Test Type			0.05 = -5.08
Za			0.1 = -4.82
potential break in data at position 55			

The Zivot and Andrews test of no data break point as its null hypothesis was used to ascertain a possible break in the entire cocoa production series from 1948 to 2011. Table 4.8 presents a summarized result from the Za test. The test reported a potential break point at position 55 of the dataset which corresponds to the year 2002, where, the mass spraying intervention event took off. This obviously indicates that the onset of the mass spraying intervention was characterized by an immediate impact, hence the break at the year of onset. Also, there were no time delays for the impact of the Hi-technology intervention event since the size of the spike at the onset of the event is more pronounced, as observed in Figure 4.1

Table 4.9: Parameter Estimates for the hypothesized Impact Assessment model

	Model Fit Statistics			
AIC	AICc	BIC		
1596.33	1597.02	1604.9		
		Estimate	STD. Error	*t*-value
Coefficients				
ar1		0.6777	0.1007	6.72989
I1t-MA0		182398.2	70061.35	2.60341
I2t-MA0		266515.1	71671.58	3.71856
Intercept		315435.3	30113.68	10.4748

From Table 4.9, all the coefficients of the estimated parameters are significantly different from zero. There is stationarity in the ar1 process as its estimate is far below 1. Generally, the full-fitted impact assessment model exhibits a stationary process. I1t-MA0 denotes the mass spraying intervention event, whereas I2t-MA0 represents the cocoa Hi-technology intervention input event. Apparently, no obvious flaws are reported by Table

4.9 except the penalty function statistics which penalized the fitted model based on the principle of parsimony.

Table 4.10: Ljung-Box Test for the Impact or Intervention model

	Summary of Test Statistic		
	X-squared	df	*p*-value
Test Type			
Ljung-Box	13.4239	24	0.9586

Based on the results from Table 4.10, we failed to reject the null hypothesis of white noise residuals at 5% significant level. Thus, the calculated p-value of 0.9586 is far greater than the pre-chosen significant level of 0.05. This follows that the fitted intervention model provides a good fit for the entire cocoa production series.

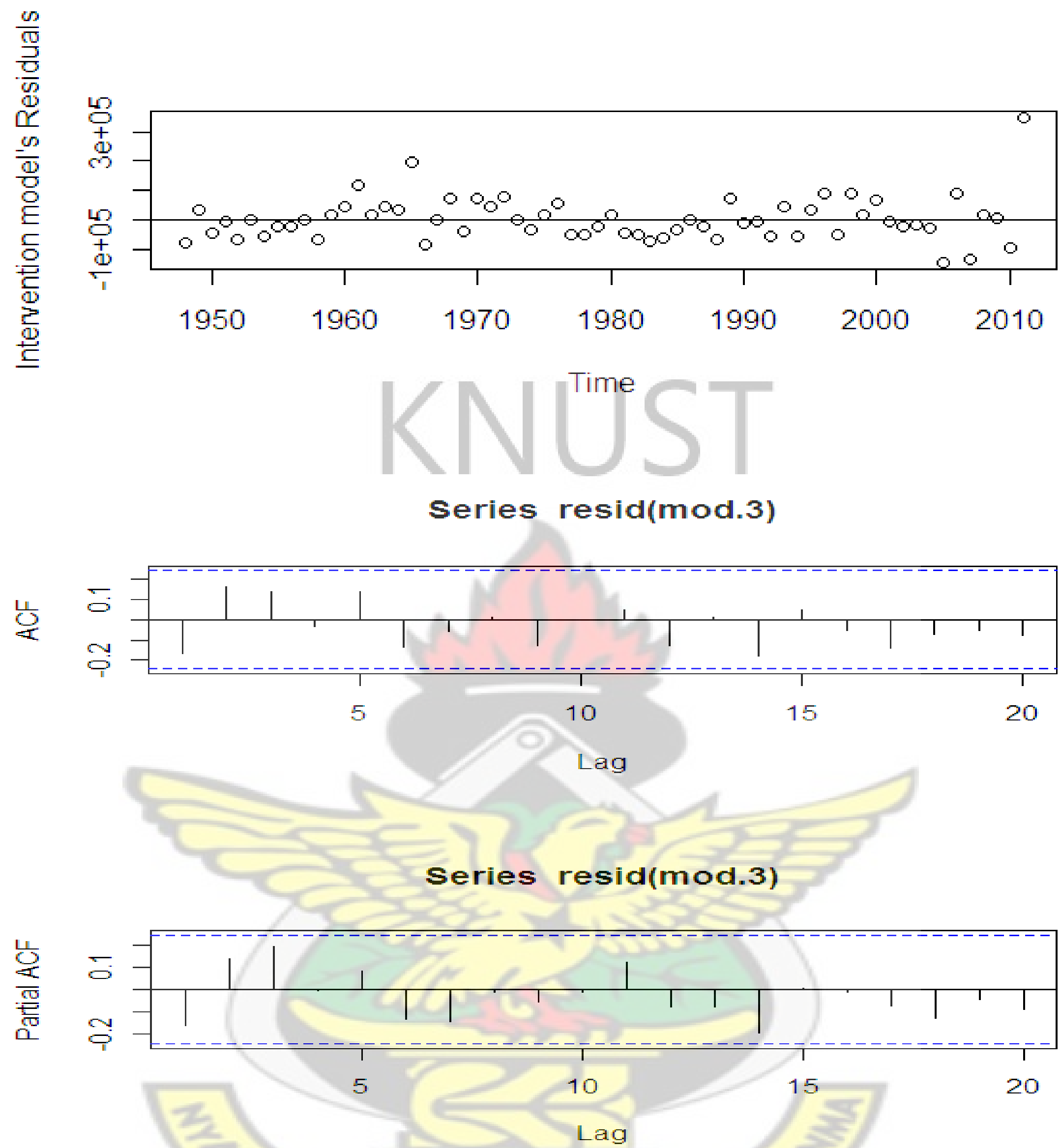

Figure 4.5: Diagnostic Residual plots of the Impact Assessment or Intervention model

The three (3) diagnostic residual plots in Figure 4.5 do not show any anomalies for the fitted Intervention model. There is only one outlier clearly spotted in 2011. Obviously, there are no significant spikes in the ACF and PACF plots of the residuals. These observations attest to the fact that the residuals left after fitting the intervention model are reasonably white noise.

4.6 Graph of the Estimated Intervention Model

The last stage of fitting a full impact assessment model is to graph the output of the intervention input events. Figure 4.6 therefore shows the output graph of the two intervention events. The thick black line indicates the original cocoa series. The red dash line represents the effect of the cocoa Hi-tech programme, whereas the mass spraying effect is shown by the blue dash line. Both programmes exhibit a clear situation of a simple step function with no decays, as shown in Figure 3.2(a) under subsection 3.6.3

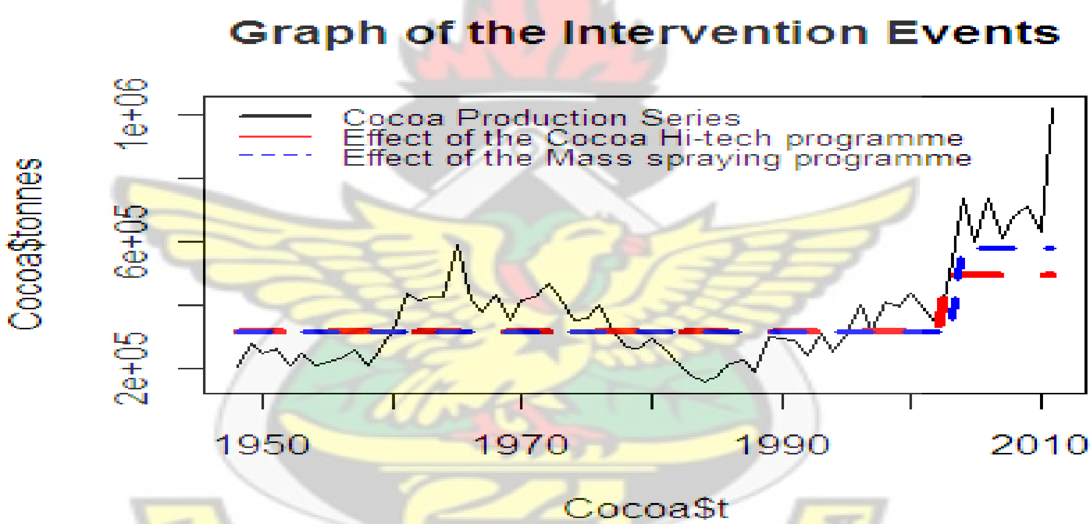

Figure 4.6: Graph of the two Intervention Events

CHAPTER 5

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.0 Introduction

This last chapter presents summary of findings from the study, and recommends pragmatic measures for stakeholders. It again recommends further study possibility areas for future researchers. The chapter also gives a conclusion statement based on the findings.

5.1 Summary of Findings from the Study

The main purpose of the study was to estimate and asses the nature of impact of the cocoa mass spraying and the cocoa Hi-technology intervention events on Ghana's cocoa production output. Annual production figures from 1948 to 2011 were used to obtain a model each for the pre- and postintervention series. It was revealed that the preintervention series could best be fitted with an autoregressive model of the first-order:

$$Y_t = 0.7966 y_{t-1} + 313689.72 + \varepsilon_t \qquad (4.1)$$

with the residuals, ε_t being white noise.

Thereafter, the AR(1) error process plus a deterministic intervention input events were estimated together to obtain parameter estimates for the hypothesized intervention model. In notation, the result of the fully-fitted intervention model as shown in Table 4.9 could also be written as:

$$Y_t = 315435.3 + 182{,}398.2 I_{1t} + 266{,}515.1 I_{2t} + (-0.6777L)\varepsilon_t \qquad (4.2)$$

where ε_t were found to be white noise.

The effect of the two intervention events were both discovered to be abrupt and permanent. Also, empirical results indicate that the intervention events could best be modeled with a permanent change with a zero-order transfer function. From Table 4.9, the cocoa mass spraying programme recorded significant increases in production outputs of 182,398.2 metric tonnes, whereas the cocoa Hi-technology intervention event triggered another significant increase of 266,515.1 metric tonnes annually. These findings reject the null hypothesis of no positive impacts for the two intervention events formulated earlier at the beginning of the study. An alternative hypothesis of positive impacts is therefore accepted for the intervention programmes.

5.2 Conclusion

It was found that the intervention effect of the mass spraying and the cocoa Hi-tech programmes could best be fitted to a simple step function with zero-order decays. However, the two intervention events showed abrupt and permanent nature of impact on the response cocoa production series. These findings successfully help to achieve the set objectives for this study, and do not contradict other findings from the various works reviewed under the literature subsection of chapter 2.

It can then be generally concluded that the mass cocoa spraying and the cocoa Hi-technology intervention programmes have had significantly positive impact on Ghana's cocoa production levels since their inception.

5.3 Recommendations for Ghana's Cocoa Sector

The following are recommended for stakeholders in Ghana's cocoa sector:

1. In line with the positive impacts of the intervention programmes on production outputs, we recommend Government to intensify support for these cocoa mass spraying and Hi-technology programmes by providing the needed resources and funds to possibly sustain the existence of the programmes.
2. There should be strict supervision by the officials of Ghana COCOBOD in the process of allocating and disbursing materials purposely meant for the two intervention programmes. In other words, tighter measures should be taken by Government and COCOBOD to ensure that monies and materials provided for the programmes are not perhaps channeled to achieve parochial interests of persons or group of people.
3. Farmers who receive credit facilities under the cocoa Hi-technology programme should do well to pay on time. This will enable them or new farmers to benefit from such credit facilities at any of the crop seasons.

5.4 Recommendations for Future Researchers

1. It is recommended that future studies should be conducted on the forecast performance of ARIMA-Intervention models as compared to other forecasting techniques.
2. It is further recommended that a study should be carried out to assess the impact of the smuggling of cocoa from Cote d'ivoire, during the political struggle there in 2011 on Ghana's cocoa production levels.

REFERENCES

Abdalla El-Habil (2006). Measuring the Impact of a New Policy on Industry in Gaza. An - Najah Univ. J. Res. (H. Sc.) Vol. 21(2)

Adjinah K.O., and Opoku I.Y., (2010). The National Cocoa Diseases and Pests Control (CODAPEC): Achievements and Challenges, Accra

Adriana P., Rodolfo J. D., Benigno R., Amparo Y. C., Victor D., Juan L.M., Maria C.C., (2003). An interrupted time series analysis of parenteral antibiotic use in Colombia. Journal of Clinical Epidemiology, Volume 56, Issue 10, Pages 1013-1020. htpp://sciencedirect.com/science/journal/08954356

Anim-Kwapong G.J., and Frimpong E.B., (2005). Vulnerability of agriculture to climate change – impact of climate change on cocoa production: Cocoa Research Institute of Ghana, New-Tafo Akim

Ansari F., Kirsteen G., Dilip N., Gabby P., Simon O., Craig R., and Peter D., (2010). Outcomes of an intervention to improve hospital antibiotic prescribing: interrupted time series with segmented regression analysis. Journal of Antimicrobial Chemotherapy, Volume52, Issue5, pp. 842-848. http://jac.oxfordjournals.org

Asuming-Brempong S., Sarpong D.B., Asenso-Okyere K., and Amoo P., (2006). Labour practices in cocoa production in Ghana (pilot survey)

Bhattacharyya M.N., and Layton, A.P., (1979). ‘Effectiveness of seat belt legislation on the Queensland road toll: an Australian case study in intervention analysis’, Journal of the American Statistical Association, vol. 74, pp. 596-603. http://scialert.net/fulltext/?doi=jas.2011.1105.1112&org=11

Bianchi L., Jeffrey J., and Hanumara R.C., (1998). Improving forecasting for telemarketing centers by ARIMA modeling with intervention. International Journal of Forecasting, Volume 14, Issue 4, Pages 497-504. http://www.sciencedirect.com/science?

Bonham S.C., and Gangnes B., (1996). Intervention analysis with cointegrated time series: the case of the Hawaii hotel room tax. Applied Economics, 28, 1281 – 1293

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

Box G.E.P., Jenkins G.M., and Reinsel, G.C., (1994). Time Series Analysis: Forecasting and Control, 3rd ed. Prentice Hall, Upper Saddle River, New Jersey.

Box G.E.P., and Tiao G.C., (1975). Intervention analysis with applications to economic and environmental problems, J. Amer. Stat. Assoc. 70, 70–79. http://scialert.net/fulltext/?doi=jas.2011.1105.1112&org=11

Budget Statement and Economic Policy of the Government of Ghana for the 2012 FINANCIAL YEAR, Theme: “Infrastructure Development for Accelerated Growth and Job Creation”

Carlos P.B., (2003). An Intervention Analysis of Terrorism: The Spanish ETA Case. Defence and Peace Economics,Vol. 14(6), pp. 401–412

Centre for Policy Analysis, Ghana (2004). Selected Economic Issues: Review and Analysis of Budget Statement and Economic Policy, No.7, pp.39

Chang G.L., Chen C.H., and Carter E.C., (1993). Intervention analysis for the impacts of the 65 mph speed limit on rural interstate highway fatalities. J Safety Res 1993; 241:33–53.

Chatfield C., (1996). The analysis of time series: an introduction, 5th ed., London: Chapman and Hall.

Chen J.C.R., (2006). Before and after the Inclusion of Intervention Events: An Evaluation of Alternative Forecasting Methods for Tourist Flows. Tourism Analysis, Vol. 10, pp. 269–276

Chung C.P.R., Ip W.H., and Chan S.L., (2009). An ARIMA-Intervention Analysis Model for the Financial Crisis in China's Manufacturing Industry. International Journal of Engineering Business Management, Vol. 1, No. 1, pp. 15-18

COCOBOD NEWS, 2007

Collier B., Ani L.K., and Skees J.R., (2010). Loan Portfolio Performance and El_iño, an Intervention Analysis. Selected Paper prepared for presentation at the Southern Agricultural Economics Association Annual Meeting, Orlando, FL

Enders W., Sandler T., and Cauley J., (1990). Assessing the impact of terrorist-thwarting policies: An intervention time series approach. Defense Economics 2(1), 1–18.

Fernandez-Perez C., Juan T., and Miguel C., (1998). Multivariate time series analysis in nosocomial infection surveillance: a case study. International Journal of Epidemiology, 27-282-288

Ferrand Y., Kelton C.M., Guo J.J., Levy M.S., and Yu Y., (2011). Using time-series intervention analysis to understand U.S. Medicaid expenditures on antidepressant agents. Res Social Adm Pharm.7(1):64-80. http://www.ncbi.nlm.nih.gov/pubmed

Fox R.T., (1996). "Using Intervention Analysis to Assess Catastrophic Events on Business Environment." International Advances in Economic Research 2: 341-349.

Frain J., (1992). Lecture Notes on Univariate Time Series Analysis and Box-Jenkins Forecasting. Economic Analysis, Research and Publications (reprinted with revisions)

Girard D.Z., (2000). Intervention times series analysis of pertussis vaccination in England and Wales. Health Policy (2000) 54: 13-25.

Guzhva S.V., (2008). Applying intervention analysis to financial performance data: The case of US airlines and September 11th. Journal of Economics and Finance, Volume 32, Number 3, 243-259

Harvey A.C., and Durbin J., (1986). 'The effects of seat belts legislation on British road casualties: a case study in structural time series modelling', Journal of the Royal Statistical Society, Series A. vol. 149, pp. 187-227.

Hipel K.W., Lettenmaier D.P., and McLeod A.I., (2001). Assessment of environmental impacts part one: Intervention analysis. Environmental Management, Volume 2, Issue 6, pp.529-535. http://adsabs.havard.edu/cgi-bin/author

Hultkrantz L., and Olsson K., (1997). “Chernobyl Effects on Domestic and Inbound Tourism in Sweden– A Time-series Analysis.” Environmental and Resource Economics 9, 239-258.

Ihaka R., (2005).Time Series Analysis, Lecture Notes. University of Auckland.

Ismail Z., Suhartono Y.A., and Efendi R., (2009). Intervention Model for Analyzing the Impact of Terrorism to Tourism Industry. Journal of Mathematics and Statistics 5 (4): 322-329, ISSN: 1549-3644

Keum-Rok Y., (1998). Intervention Analysis of Electoral Tax Cycle: The Case of Japan. Public Choice, vol. 96, issue 3-4, pages 241-58. http://enconpapers.org/artic

Kirchgassner G., and Wolters J., (2007). Introduction to Modern Time Series Analysis. Springer Berlin Heidelberg, New York, pp.178

Knapp A.W., (1920). Cocoa and Chocolate: Their History from Plantation to Consumer. London. U.K

Kuo J.T., and Sun Y.H., (1993). An intervention model for average 10 day stream flow forecast and synthesis. J. Hydrol., 151: 35-56.

Ledolter J., and Chan K.S., (1996). Evaluating the impact of the 65mph maximum speed limit on Iowa rural interstates. American Statistician, 50, 79 – 85

Lloyd T., Morrissey O., and Reed G., (1998), Estimating the impact of anti-dumping and anti-cartel actions using intervention analysis, Economic Journal, vol.108, pp. 458-76.

Makridakis S., Wheelwright S. C., and Hyndman R. J., (1998). Forecasting: Methods and Applications, 3rd ed., Wiley.

Min C.H.J., (2008a). Intervention analysis of inbound tourism: A case study of Taiwan, in Joseph S. Chen (ed.) 4 (Advances in Hospitality and Leisure, Volume 4), Emerald Group Publishing Limited, pp.53-74

Min C.H.J., (2008b). "Forecasting Japanese tourism demand in Taiwan using an intervention analysis", International Journal of Culture, Tourism and Hospitality Research, Vol. 2 Iss: 3, pp.197 – 216

Min C.H.J., Hsien-Hung K., and Hsiang L.H., (2010). Interventions affecting air transport passenger demand in Taiwan. African Journal of Business Management Vol. 4(10), pp. 2121-2131, ISSN: 1993-8233

Ming-Chang C., Kee-Kuo C., and Alex P., (2003). Establishing an intervention model to examine the impact of policy guidance on transportation demand. Journal of the Eastern Asia Society for Transportation Studies, Vol.5, October, 2003, 1565 – 1575

Morales R. A., Cowen P., and Van P.P.N., (1988). The use of Intervention Analysis in Disease Control and Eradication Program Evaluation: California's Brucellosis Eradication Campaign. Proceedings of the 5th International Symposium on Veterinary Epidemiology and Economics. Acta Veterinaria Scandinavia, Suppletum 84. www.sciquest.org.nz

Muhammad H., Lee S., and Bahrom S., (2010). Multi Input Intervention Model for Evaluating the Impact of the Asian Crisis and Terrorist Attacks on Tourist Arrivals. MATEMATIKA, Volume 26, Number 1, 83–106

Narayan J., and Considine J., (1989). Intervention analysis to assess the impact of price changes on ridership in a transit system: a case study. Logist. Transport. Rev., 25(1): 53-62.

Quirey T.L., (2001). The impact of breast and cervical cancer screening on mortality in Australia: an application of intervention analysis. Honours Thesis, School of Economics, The University of Queensland.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.or

Reagan K.M., (1984). Unpublished MS Thesis. University of California, Los Angeles.

Remenyi N., Anderson D.M., and Murray L.W., (2010). Using time-series intervention analysis to model cow heart rate affected by programmed audio and environmental/physiological cues. In: Applied Statistics in Agriculture Conference Proceedings, pp.107–136, Kansas State University Publishers, Manhattan Kansas. http://jornada.nmsu.edu/biblio/author/119

Robert A.Yaffee., and McGee M., (2000). Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS. Academic Press Inc., San Dieg.

Rock-Antoine M., and Hannarong S., (2002). Who is benefiting the most from nafta? An Intervention time series analysis. Journal of economic development 69 volume 27, number 2

Ruey S. T., (2002). Analysis of Financial Time Series. A Wiley-Interscience Publication

Sathianandan T.V., Somy K., Mini K. G., and Joji T.V., (2006). Impact of introduction of crafts with outboard engines on marine fish production in Kerala and Karnataka – a study through Intervention analysis. Indian Journal of Fisheries, 53 (3): pp. 271-282

Sharma P., and Khare M., (1999). Application of intervention analysis for assessing the effectiveness of CO pollution control legislation in India. Transport Res D-Tr E, 4(6): 427-432. http://scialert.net/fulltext/?doi=jas.2011.1105.1112&org=11

Shittu O.I., (2009). Modeling Exchange Rate in Nigeria in the Presence of Financial and Political Instability: An Intervention Analysis Approach. Middle Eastern Finance and Economics, ISSN: 1450 – 2889, Issue 5.

Shumway R.H., and Stoffer D.S., (2006). Time Series Analysis and Its Applications (with R Examples), 2nd ed. New York: Springer.

Stanley M.H., (1984). Effects of Spring Breakup on Microscale Air Temperatures in the Mackenzie River Delta. Arctic, Vol. 37, No. 3, pp. 263-269

Stuart G., Louisa D., Wayne H., and Carolyn D., (2006). Using intervention time series analyses to assess the effects of imperfectly identifiable natural events: a general method and example. BMC Medical Research Methodology 2006, **6**:16

Tukey J.W., (1961), "Discussion, Emphasizing the Connection Between Analysis of Variance and Spectrum Analysis," Technometrics, Vol.3, No.2, pp.191-219

Valadkhani A., and Layton A.P., (2004). Quantifying the Effect of GST on Inflation in Australia's Capital Cities: An Intervention Analysis, Australian Economic Review, 37(2), 125-38.

Welsh D., and Stewart D.B., (1999). Applications of intervention analysis to model the impact of drought and bushfires of Water Quality. Australian Journal of Marine and Freshwater Research 40 (3) 241 – 257

Yaacob W.F., Husin W.Z., Abd.Aziz N., and Nordin N. I., (2011). An Intervention Model of Road Accidents: The Case of OPS Sikap Intervention. Journal of Applied Sciences, 11: 1105-1112. http://scialert.net/fulltext/?doi=jas.2011.1105.1112&org=11

Zambon F., Fedeli U., Cristiana V., Marchesan M., Avossa F., Brocco S., and Spolaore P., (2007). Evidence-based policy on road safety: the effect of the demerit points system on seat belt use and health outcomes. Journal of Epidemiology and Community Health, 61:877-881. http://jech.bmj.com/content/61/10/877.full

APPENDIX A

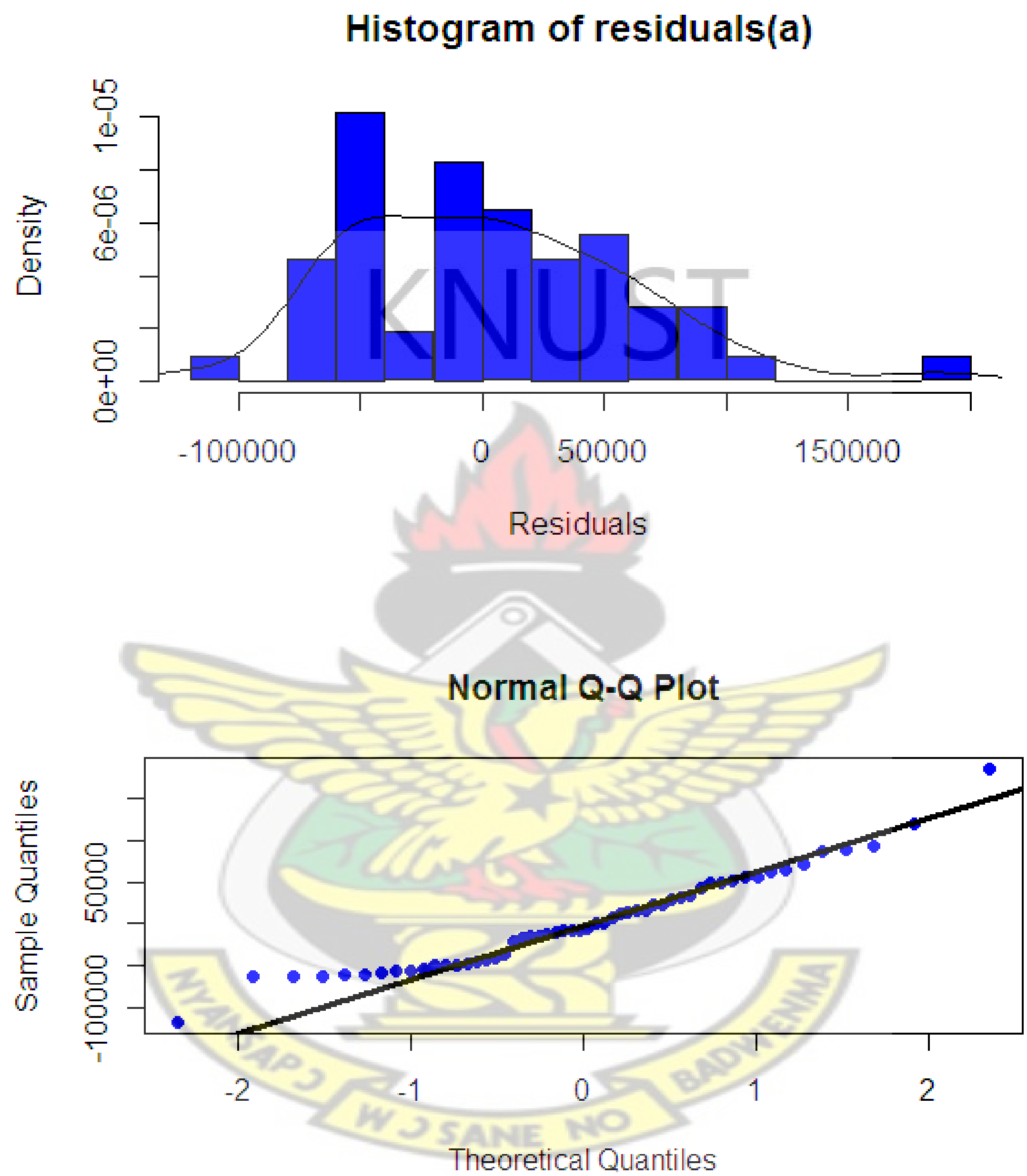

Figure 1: Histogram and Normal Probability plots of Residuals of ARIMA(1, 0, 0)

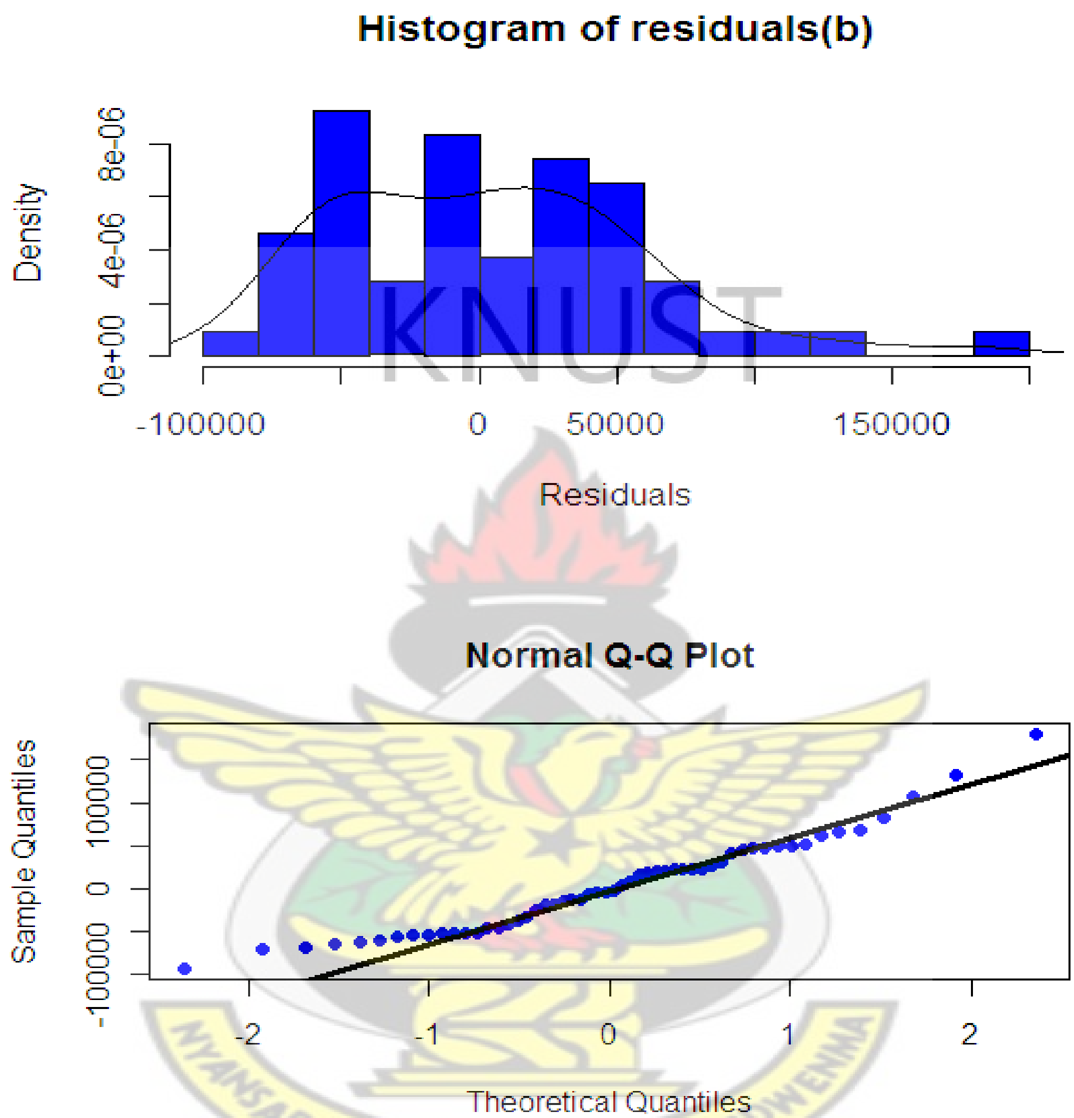

Figure 1: Histogram and Normal Probability plots of residuals for ARIMA(2, 0, 0)

APPENDIX C

Table 1: Ghana's Annual Cocoa Production Data

	Production				***Production***		
Crop Year	Main	Mid	Total	Crop Year	Main	Mid	Total
1947/48	203,808	3,751	207,559	1979/80	281,723	14,696	296,419
1948/49	269,051	9,321	278,372	1980/81	254,961	3,013	257,974
1949/50	246,443	1,391	247,834	1981/82	215,705	9,177	224,882
1950/51	258,282	3,941	262,223	1982/83	176,758	1,868	178,626
1951/52	206,904	3,759	210,663	1983/84	153,737	5,219	158,956
1952/53	243,385	3,597	246,982	1984/85	165,377	9,432	174,809
1953/54	206,038	4,978	211,016	1985/86	205,327	13,707	219,034
1954/55	209,748	13,569	223,317	1986/87	218,255	9,510	227,765
1955/56	221,505	10,943	232,448	1987/88	173,171	15,000	188,171
1956/57	257,112	2,676	259,788	1988/89	290,308	9,793	300,101
1957/58	197,751	12,014	209,765	1989/90	282,578	12,473	295,051
1958/59	229,470	30,102	259,572	1990/91	261,219	32,133	293,352
1959/60	296,338	25,885	322,223	1991/92	229,122	13,695	242,817
1960/61	427,377	11,782	439,159	1992/93	262,431	49,692	312,123
1961/62	402,691	12,495	415,186	1993/94	221,302	33,351	254,653
1962/63	387,712	40,306	428,018	1994/95	287,095	22,359	309,454
1963/64	376,616	51,166	427,782	1995/96	349,305	54,567	403,872
1964/65	575,710	15,321	591,031	1996/97	297,702	24,786	322,488
1965/66	402,481	13,272	415,753	1997/98	365,483	43,900	409,383
1966/67	369,213	12,140	381,353	1998/99	340,033	57,642	397,675
1967/68	404,394	26,271	430,665	1999/2000	406,399	30,548	436,947
1968/69	319,560	36,028	355,588	2000/01	350,359	39,413	389,772
1969/70	385,030	32,427	417,457	2001/02	321,321	19,241	340,562
1970/71	401,591	26,303	427,894	2002/03	444,135	52,711	496,846
1971/72	457,814	12,050	469,864	2003/04	668,787	68,188	736,975
1972/73	415,740	6,103	421,843	2004/05	526,828	72,490	599,318
1973/74	343,328	11,543	354,871	2005/06	649,672	90,786	740,458
1974/75	353,018	8,265	361,283	2006/07	587,502	27,030	614,532
1975/76	388,573	11,748	400,321	2007/08	663,954	16,827	680,781
1976/77	309,384	14,727	324,111	2008/09	634,256	76,386	710,642
1977/78	263,214	8,125	271,339	2009/10	587,179	44,858	632,037
1978/79	249,984	15,092	265,076	2010/11	916,810	107,744	1,024,553

Source: Monitoring, Research and Evaluation Department of Ghana COCOBOD