CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND OF THE STUDY

Investment is an integral part of life. People undertake investments in various forms. A student invests in education spending time and money to acquire knowledge to use in future. A cocoa farmer who spends years, money and energy cultivating cocoa is undertaking an investment. In the same vein a businessman who invests his money and time in the business is also undertaking an investment. Some people invest thousands of cedis in treatment and cosmetics to make them look more attractive.

One thing that is common to all these people mentioned above is that they all invest with the hope of reaping a reward for their investments. The type of reward and the time (period) it takes to get the reward (return) from your investment will depend on the type of investment undertaken. For instance, where as it may take at least ten years to receive a reward from education it may take between three to four years for a farmer to start enjoying the fruits of his toils.

Investment is putting money into something with expectation of profit. More specifically investment is the commitment of money or capital to the purchase of financial instrument or other assets so as to gain profitable returns in the form of interest, income (dividend or appreciation. Capital gains of the value of the investment. It is related to savings or deferring consumption. Investment is involved in many areas of the economy such as business management finance, firms, government etc.

Investment in the field of finance involves investing in marketable, public traded securities or investing in real assets. A financial asset represents a financial claim on an asset that is usually documented by some form of legal representation. Examples of financial assets are bonds and stocks. Real assets on the other hand represent tangible assets that can be seen, felt, held or collected. Examples are gold, real estates, cars etc. Real assets of an economy include land, buildings, machines, capital and knowledge that are used to produce goods and services.

Investors in securities include individual investors and institutional investors such as investment banks, pension funds, mutual funds and insurance firms.

Investment has been given many definitions by different people in the field of finance.

To the economist, investment involves expenditure on capital goods or on inventory of goods or raw materials that are used to produce other goods and services causing future production and income to rise.

"To the financial economist, investment loosely described involves foregoing current consumption in exchange for future opportunities".

It is therefore expected that the investor would be compensated for the time value of money, loss of purchasing power and risk associated with owning the investment.

Hirt and Block (1998) in their book "Fundamentals of investment management, define investment as the commitment of current funds in anticipation of receiving a larger future flow of funds.

Investment involves the commitment of funds, time, energy and expertise in a venture with the view of getting a reward (return) from such a venture in the future. The investment decision is essentially how much not to consume in the present in order that more can be consumed in the future.

The optimal investment decision maximizes the expected satisfaction (utility) gained from consumption over the planning horizon of the decision maker. It should be noted that idle cash is not an investment. Therefore a person who keeps millions of cedis at home is not investing that fund. Idle cash is not investment because its value reduces with time due to inflation. Secondary, idle cash does not generate any return.

1.2 PROBLEM STATEMENT

In 2007 the Executive Chairman of Databank Financial Services, Ken Ofori-Atta, officially announced the performance of the EPACK fund at an Annual General Meeting. He said the fund's performance in 2006 was over six times better than the second best performing mutual fund in Ghana. He said EPACK continues to be the most sought after mutual fund not only by local residents but also by Ghanaians abroad and that the fund has been doing very well over the years. The fund's market capitalization rose by ¢105.6 billion, representing 40%, to close at ¢382 billion by the end of December 2006; compared to a 19% appreciation in the market capitalization of the Ghanaian bourse.

This study seeks to ascertain the truthfulness of the Executive Chairman and to access the performance of EPACK over the years by time series approach.

1.3 OBJECTIVES

The research work seeks to:

i. Apply a mathematical model to suit the share price of EPACK Mutual Fund.

ii. Analyse the trend of the share prices of EPACK Mutual Fund.

iii. Analyse the behaviour of the EPACK share prices with time.

iv. Ascertain whether or not the assertions being made about the fund is true.

1.4 METHODOLOGY

The data was obtained from the DataBank Asset Management Services Limited, Kumasi branch. The data comprised averaged monthly EPACK share prices from January 1999 to December 2010. Trend analysis was conducted followed by Box-Jenkins Autoregressive Integrated Moving Average (ARIMA). Time series modeling procedure was used in modeling the data. Statistical Package for Social Scientist (SPSS) was used for the computation and analysis of the data in the series after which Matrix Laboratory (MatLab) was used to code for the forecasted values of EPACK share price.

1.5 JUSTIFICATION / SIGNIFICANCE

Since there is no published mathematical work on the performance of EPACK share prices, this thesis will serve as a source of knowledge both in the field of time series and that of EPACK share price.

It will also serve as basis for further study in the field.

1.6 SCOPE/ LIMITATIONS

Aside the stress and financial constraints that one may encounter in putting this piece of work together, I must say that laying hands on a project topic was really a difficult task. Moreover, the course was designed specifically for the weekends but hardly can we get our supervisors on weekends because they might be handling our colleagues in the other levels, hence we have to travel distances on week days to see our supervisors meaning not going for work for three days, this to me was my greatest limitations.

1.7 ORGANIZATION OF THE THESIS

The study will be structured into five chapters as follows:

Chapter One covers the introduction of the study, which discusses the background of the study, problem statement, methodology, and significance of the study, Delimitations/Scope and the organization of study.

In chapter two, relevant literature (publications) in the field of time series analysis, and Box-Jenkins Methodology was reviewed. Detailed description of the methodology employed was done in chapter three. The thesis thoroughly discussed Time series analysis, Box Jenkins Methodology and Forecasting techniques. Chapter four was devoted to the analysis of averaged monthly EPACK share prices from January 1999 to December 2010 by employing the methods discussed in chapter three. Chapter five concludes the study with its findings, conclusions and recommendations.

CHAPTER TWO

2.0 LITERATURE REVIEW

Time series analysis and forecasting has become a major tool in different applications in hydrology and environmental management fields. Among the most effective approaches for analyzing time series data is the model introduced by Box and Jenkins, Autoregressive Integrated Moving Average (ARIMA). In a study Naill, P.E. *et al*, [2009] used Box-Jenkins methodology to build ARIMA model for monthly rainfall data taken for Amman airport station for the period from 1922-1999 with a total of 936 readings. In their research, $ARIMA\ (1, 0, 0)$ [(0, 1, 1).sup.12] model was developed. This model was used to forecasting the monthly rainfall for the upcoming 10 years to help decision makers establish priorities in terms of water demand management. They then recommended an intervention time series analysis to be used to forecast the peak values of rainfall data.

Following the examples of other countries, in April 2005 Japan launched wholesale electric power exchange operations as a primary item of system reform in line with electric liberalization. Only two years have passed since the initiation of these operations. However, in the summer of 2005, the surge in market prices was evident, which suggested that certain measures should be taken to confront potential market risks. Establishing a useful system for forecasting market prices through the modeling of price fluctuations in the wholesale electric market became essential. Until then, various price models were being proposed. Taking both the limited amount of data and the model's purpose into consideration, Hiroshi *et al.*, [2007] adopted the univariate time series model. They conducted a time series analysis on the open price indexes in the JEPX spot market with the Box-Jenkins method. Since a seven-day cycle was observed in the data, they adopted the seasonal ARIMA model. In accordance with the procedures of the Box-Jenkins method, they determined the degree of the model's polynomial using the autocorrelation and partial autocorrelation of the data and estimated the parameters of the model with the maximum likelihood method. They then conducted a forecast on next day JEPX spot market prices with this time series model and examined its validity and utility as a forecasting tool. They hypothesized that, price forecasts made with this model require only a small amount of data and will save substantial analysis work. Consequently, their method is expected to be widely used by market participants as the reference data for their bid pricing.

Box-Jenkins modeling has some advantages over other techniques for the analysis of time series of climatological variables. Not only does it provide more information than other methods of analysis, in a more elegant way, but it is also perfectly acceptable from the mathematical point of view. Other methods may not be immediately applicable because of the problem of autocorrelation in time series. The method of Box-Jenkins univariate modeling was briefly discussed by Davies *et al.*, [2006] in their research. As an example of its application to climatological time series analysis, and as an illustration of its usefulness, they examined the monthly activity of temperature inversions over Hemsby (Eastern England) over a 14-year period. Their results showed that the monthly activity series, for both surface and elevated inversions, are stationary. However, the series for surface midnight inversions had a seasonal non-stationarity of lag 12. There was 12 month seasonality for surface inversions and weaker 6-month seasonality for elevated inversions. The monthly activity of surface inversions exhibits less variation than the monthly activity of elevated inversions. This simply reflected the fact that the physical processes responsible for the formation of surface midnight inversions had more regular evolution overtime than those responsible for the formation of elevated inversions. Their results were in accordance with those obtained by using standard statistical techniques.

In order to improve the control level of district-heating systems, it was necessary for the energy companies to have reliable optimization routines, implemented in their organizations. However, before a plan of heat production, a prediction of the heat demand first needs to be determined. Forecast of this heat demand course is significant for short-term and long-term planning of heat production. This forecast is most important for technical and economic consideration. In a paper Bronislav *et al.*, proposed the forecast model of heat demand based on the Box-Jenkins methodology. Their model was based on the assumption that the course of DDHD can be described sufficiently well as a function of the outdoor temperature and the weather independent component (social components). Time of the day affects the social components. The time dependence of the load reflected the existence of a daily heat demand pattern, which may vary for different week days and seasons. Forecast of social component was realized by means of Box-Jenkins methodology. Their model was used for prediction of heat demand in different locality.

The use of the Box-Jenkins approach for forecasting the population of the United States up to the year 2080 was discussed by Peter Pflaumer in 2002. It was shown that the Box-Jenkins approach is equivalent to a simple trend model when making long-range predictions for the United States. An investigation of forecasting accuracy indicated that the Box-Jenkins method produces population forecasts that were at least as reliable as those done with more traditional demographic methods.

Nihan *et al*., [2005] explored the use of time series techniques for short term traffic volume forecasts. A data set containing monthly volumes on a freeway segment for the years 1968 through 1976 was used to fit a time series model. Their resulting model was used to forecast volumes for the year 1977. The forecast volumes were then compared to actual volumes in1977. The results of their study indicated that time series techniques can be used to develop highly accurate and inexpensive short term forecasts. A discussion of the ways in which such models can be used to evaluate the effects of policy changes or other outside impacts was included.

Dobre *et al*., [2008] published a paper on modeling the evolution of unemployment rate using the Box-Jenkins methodology during the period $1998 - 2007$ monthly data. Their empirical study revealed that the most adequate model for the unemployment rate was $ARIMA\ (2, 1, 2)$. Using the model, they forecasted the values of unemployment rate for January and February2008. Therefore, their forecasted unemployment rate of Romania for January 2008 was 4.06%.

Th.D Popescu, [2003] presented a paper on "Experiences with a computer aided procedure for time series analysis and forecasting using Box-Jenkins philosophy". His paper presented some experiences with model building, parameter estimation and forecasting of different time series from industrial processes, biology, transport, using a program package for time series analysis and forecasting. AUTOB&J program was then designed to meet all software needs pertaining to Box-Jenkins or ARIMA model building.

In 2003 Vähäkyla *et al.*, worked on short-term forecasting of grid load using Box-Jenkins techniques in their paper, they demonstrated the use of Box-Jenkins time series analysis in short-term load forecasting, and a forecasting system developed at the Imatra Power Company was described. Their forecasting algorithm was simple, fast and accurate, which made it suitable for online forecasting. Their transfer function model was used to introduce temperature effects, thus improving accuracy further. They hypothesized that their method gave good results in other forecasting problems of electrical energy systems.

The sociologist or historian who wants to analyze time series data is often confronted with the fact that the data do not meet the requirements of the statistical models that he or she would like to apply. One of the problems commonly encountered is "outliers" that, if not treated properly, may distort model identification and parameter estimation. Statisticians working within the Box-Jenkins approach to time series analysis have recently developed a detection and estimation procedure that promises to be quite effective in modeling outliers. This procedure is introduced here in the version given by Chung Chen and Lon-Mu Liu. Several examples with simulated and real world data are presented.

Chan [2002] in his paper adopted the multiple time-series modeling approach suggested by Tiao and Box (1981) to construct a stochastic investment model for price inflation, share dividends, share dividend yields and long-term interest rates in the United

Kingdom. His method had the advantage of being direct and transparent. The sequential and iterative steps of tentative specification, estimation and diagnostic checking parallel those of the well-known Box-Jenkins method in the univariate time-series analysis. He then concluded that it was not required to specify any *a priori* causality as compared to some other stochastic asset models in his literature.

Fishwick *et al.*, [1991] discussed the results of a comparative study of the performance of Neural networks and conventional methods in forecasting time series. Their work was initially inspired by previously published works that yielded inconsistent results about comparative performance. They had experimented with three time series of different complexity using different feed forward, back propagation neural network models and the standard Box-Jenkins model. Their experiments demonstrated that for time series with long memory, both methods produced comparable results. However, for series with short memory, neural networks outperformed the Box-Jenkins model. They noted that some of the comparable results arose since the neural network and time series model appeared to be functionally similar models. They found that for time series of different complexities there were optimal neural network topologies and parameters that enabled them to learn more efficiently. Their initial conclusions were that neural networks were robust and provided good long-term forecasting. They were also parsimonious in their data requirements. Neural networks represented a promising alternative for forecasting, but there were problems determining the optimal topology and parameters for efficient learning.

CHAPTER THREE

3.0 METHODOLOGY

In this chapter, we shall first review techniques used to identify patterns in time series data (such as smoothing and curve fitting techniques and autocorrelations), then we shall introduce a general class of models that can be used to represent time series data and generate predictions (autoregressive and moving average models). Finally, we shall review some simple but commonly used modeling and forecasting techniques based on linear regression.

3.1 TIME SERIES ANALYSIS

Under this topic, we shall review techniques that are useful for analyzing time series data, that is, sequences of measurements that follow non-random orders. Unlike the analyses of random samples of observations that are discussed in the context of most other statistics, the analysis of time series is based on the assumption that successive values in the data file represent consecutive measurements taken at equally spaced time intervals. Detailed discussions of the methods described in this chapter can be found in Anderson (1976), Box and Jenkins (1976), Kendall (1984), Kendall and Ord (1990), Montgomery, Johnson, and Gardiner (1990), Pankratz (1983), Shumway (1988), Vandaele (1983), Walker (1991), and Wei (1989).

3.1.1 Two Main Goals

There are two main goals of time series analysis:

i. Identifying the nature of the phenomenon represented by the sequence of observations, and

ii. Forecasting (predicting future values of the time series variable).

Both of these goals require that the pattern of observed time series data is identified and more or less formally described. Once the pattern is established, we can interpret and integrate it with other data (i.e., use it in our theory of the investigated phenomenon, e.g., seasonal commodity prices). Regardless of the depth of our understanding and the validity of our interpretation (theory) of the phenomenon, we can extrapolate the identified pattern to predict future events.

3.1.2 Systematic Pattern and Random Noise

As in most other analyses, in time series analysis it is assumed that the data consist of a systematic pattern (usually a set of identifiable components) and random noise (error) which usually makes the pattern difficult to identify. Most time series analysis techniques involve some form of filtering out noise in order to make the pattern more salient.

3.1.3 Two General Aspects of Time Series Patterns

Most time series patterns can be described in terms of two basic classes of components: trend and seasonality. The former represents a general systematic linear or (most often) nonlinear component that changes over time and does not repeat or at least does not repeat within the time range captured by our data (e.g., a plateau followed by a period of exponential growth). The latter may have a formally similar nature (e.g., a plateau followed by a period of exponential growth), however, it repeats itself in systematic intervals over time. Those two general classes of time series components may coexist in real-life data. For example, sales of a company can rapidly grow over years but they still follow consistent seasonal patterns (e.g., as much as 25% of yearly sales each year are made in December, whereas only 4% in August).

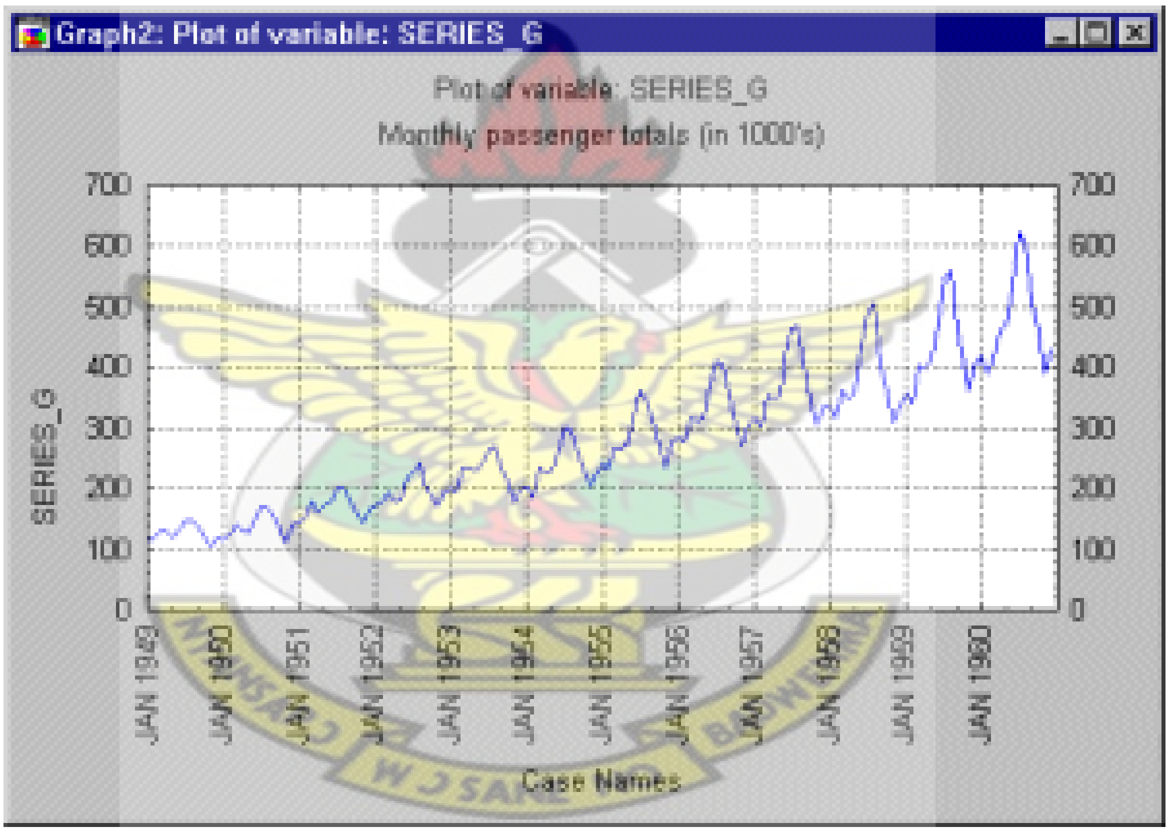


Figure 3.1: Plot of variable: Series G

This general pattern is well illustrated in a "classic" *Series G* data set (Box and Jenkins, 1976, p. 531) representing monthly international airline passenger totals (measured in thousands) in twelve consecutive years from 1949 to 1960 (see example data file *G.sta* and graph above). If you plot the successive observations (months) of airline passenger totals, a clear, almost linear trend emerges, indicating that the airline industry enjoyed a steady growth over the years (approximately 4 times more passengers traveled in 1960 than in 1949). At the same time, the monthly figures will follow an almost identical pattern each year (e.g., more people travel during holidays than during any other time of the year). This example data file also illustrates a very common general type of pattern in time series data, where the amplitude of the seasonal changes increases with the overall trend (i.e., the variance is correlated with the mean over the segments of the series). This pattern which is called *multiplicative seasonality* indicates that the relative amplitude of seasonal changes is constant over time, thus it is related to the trend.

3.1.4 Trend Analysis

There are no proven "automatic" techniques to identify trend components in the time series data; however, as long as the trend is monotonous (consistently increasing or decreasing) that part of data analysis is typically not very difficult. If the time series data contain considerable error, then the first step in the process of trend identification is smoothing.

Smoothing: Smoothing always involves some form of local averaging of data such that the nonsystematic components of individual observations cancel each other out. The most common technique is moving average smoothing which replaces each element of the series by either the simple or weighted average of n surrounding elements, where n is the width of the smoothing "window" (see Box and Jenkins, 1976; Velleman and Hoaglin, 1981). Medians can be used instead of means. The main advantage of median as compared to moving average smoothing is that its results are less biased by outliers (within the smoothing window). Thus, if there are outliers in the data (e.g., due to measurement errors), median smoothing typically produces smoother or at least more "reliable" curves than moving average based on the same window width. The main disadvantage of median smoothing is that in the absence of clear outliers it may produce more "jagged" curves than moving average and it does not allow for weighting.

In the relatively less common cases (in time series data), when the measurement error is very large, the distance weighted least squares smoothing or negative exponentially weighted smoothing techniques can be used. All those methods will filter out the noise and convert the data into a smooth curve that is relatively unbiased by outliers.. Series with relatively few and systematically distributed points can be smoothed with bicubic splines.

Fitting a function: Many monotonous time series data can be adequately approximated by a linear function; if there is a clear monotonous nonlinear component, the data first need to be transformed to remove the nonlinearity. Usually a logarithmic, exponential, or (less often) polynomial function can be used.

3.1.5 Analysis of Seasonality

Seasonal dependency (seasonality) is another general component of the time series pattern. The concept was illustrated in the example of the airline passengers data above. It is formally defined as correlational dependency of order k between each i^{th} element of the series and the $(i-k)^{th}$element (Kendall, 1976) and measured by autocorrelation (i.e., a correlation between the two terms); k is usually called the *lag*. If the measurement error is not too large, seasonality can be visually identified in the series as a pattern that repeats every k element.

Autocorrelation correlogram: Seasonal patterns of time series can be examined via correlograms. The correlogram (autocorrelogram) displays graphically and numerically the autocorrelation function(ACF), that is, serial correlation coefficients (and their standard errors) for consecutive lags in a specified range of lags (e.g., 1 through 30). Ranges of two standard errors for each lag are usually marked in correlograms but typically the size of auto correlation is of more interest than its reliability because we are usually interested only in very strong (and thus highly significant) autocorrelations.

Examining correlograms: While examining correlograms, you should keep in mind that autocorrelations for consecutive lags are formally dependent. Consider the following example. If the first element is closely related to the second, and the second to the third, then the first element must also be somewhat related to the third one, etc. This implies that the pattern of serial dependencies can change considerably after removing the first order auto correlation (i.e., after differencing the series with a lag of 1).

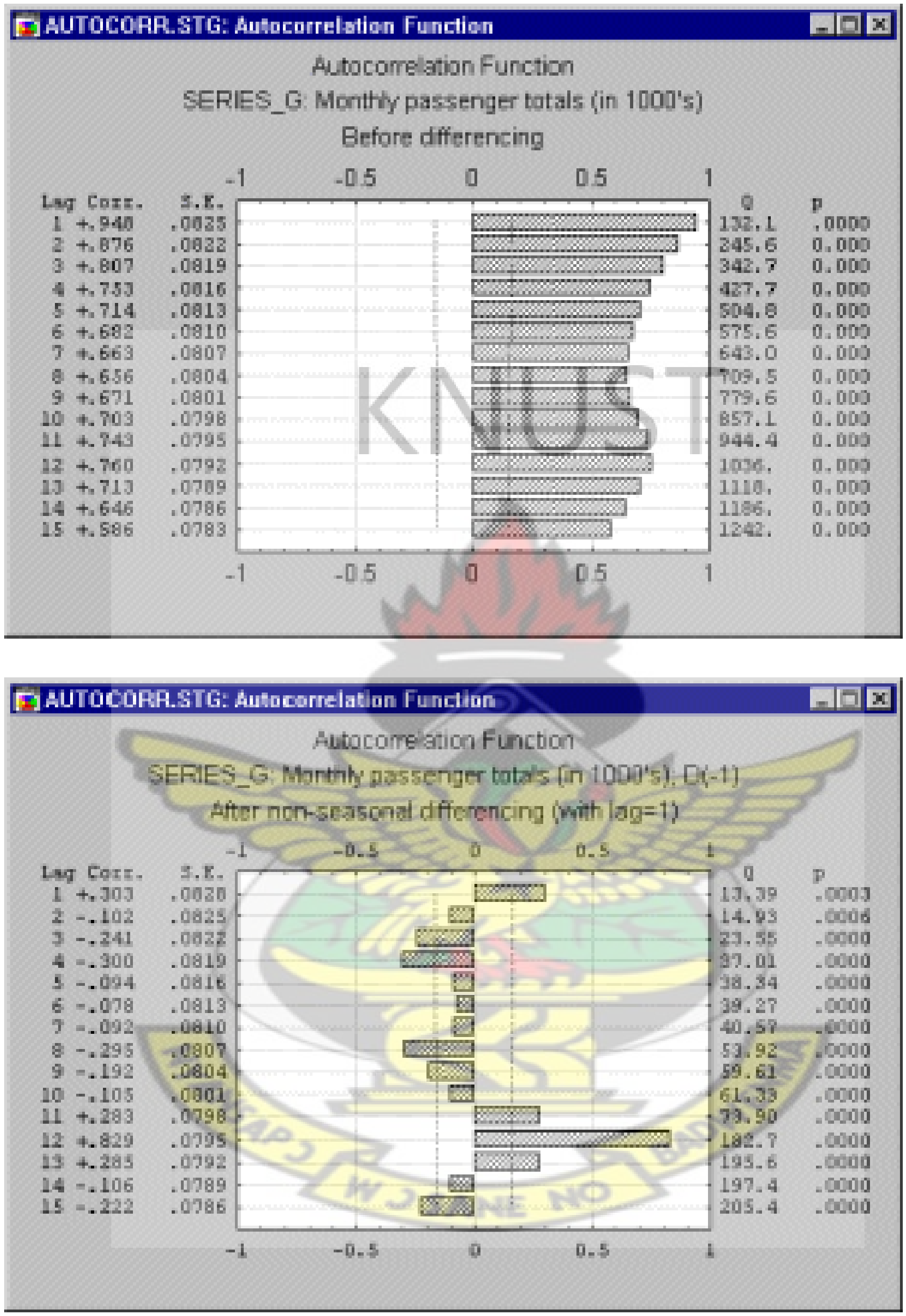


Figure 3.2: Graph of Autocorrelation Function

Partial autocorrelations Function (PACF): Another useful method to examine serial dependencies is to examine the partial autocorrelation function (PACF) - an extension of autocorrelation, where the dependence on the intermediate elements (those within the lag) is removed. In other words the partial autocorrelation is similar to autocorrelation, except that when calculating it, the (auto) correlations with all the elements within the lag are partialled out (Box and Jenkins, 1976; McDowall, McCleary, Meidinger, and Hay, 1980). If a lag of 1 is specified (i.e., there are no intermediate elements within the lag), then the partial autocorrelation is equivalent to auto correlation. In a sense, the partial autocorrelation provides a "clearer" picture of serial dependencies for individual lags (not confounded by other serial dependencies).

Removing serial dependency: Serial dependency for a particular lag of k can be removed by differencing the series, that is converting each i^{th} element of the series into its difference from the $(i\text{-}k)''^{th}$ element. There are two major reasons for such transformations.

First, we can identify the hidden nature of seasonal dependencies in the series. Remember that, as mentioned in the previous paragraph, autocorrelations for consecutive lags are interdependent. Therefore, removing some of the autocorrelations will change other auto correlations, that is, it may eliminate them or it may make some other seasonalities more apparent.

The other reason for removing seasonal dependencies is to make the series *stationary* which is necessary for ARIMA and other techniques.

3.1.6 Autoregressive Integrated Moving Average (ARIMA)

The modeling and forecasting procedures discussed in *Identifying Patterns in Time Series Data* involved knowledge about the mathematical model of the process. However, in real-life research and practice, patterns of the data are unclear, individual observations involve considerable error, and we still need not only to uncover the hidden patterns in the data but also generate forecasts. The ARIMA methodology developed by Box and Jenkins (1976) allows us to do just that; it has gained enormous popularity in many areas and research practice confirms its power and flexibility (Hoff, 1983; Pankratz, 1983; Vandaele, 1983). However, because of its power and flexibility, ARIMA is a complex technique; it is not easy to use, it requires a great deal of experience, and although it often produces satisfactory results, those results depend on the researcher's level of expertise (Bails and Peppers, 1982). The following sections will introduce the basic ideas of this methodology.

3.1.7 Two Common Processes

3.1.7.1 Autoregressive process.

Most time series consist of elements that are serially dependent in the sense that you can estimate a coefficient or a set of coefficients that describe consecutive elements of the series from specific, time-lagged (previous) elements. This can be summarized in the equation:

$$X_t = \xi + \phi_1 X_{(t-1)} + \phi_2 X_{(t-2)} + \phi_3 X_{(t-3)} + \cdots + \epsilon$$

where

ξ is a constant (intercept), and

ϕ_1, ϕ_2, ϕ_3 are the autoregressive model parameters.

Put into words, each observation is made up of a random error component (random shock, ϵ) and a linear combination of prior observations.

Stationarity requirement: An autoregressive process happens to be stable if the parameters are within a certain range; for example, if there is only one autoregressive parameter then it must fall within the interval of -1 < ϕ < 1. Otherwise, past effects would accumulate and the values of successive x_t' s would move towards infinity, that is, the series would not be stationary. If there is more than one autoregressive parameter, similar (general) restrictions on the parameter values can be defined (Box and Jenkins, 1976; Montgomery, 1990).

3.1.7.2 Moving average process.

Independent from the autoregressive process, each element in the series can also be affected by the past error (or random shock) that cannot be accounted for by the autoregressive component, that is:

$$X_t = \mu + \varepsilon_t - \phi_1 \varepsilon_{(t-1)} - \phi_2 \varepsilon_{(t-2)} - \phi_3 \varepsilon_{(t-3)} - \cdots$$

Where

μ is a constant, and

θ_1, θ_2, θ_3 are the moving average model parameters.

Put into words, each observation is made up of a random error component (random shock, ε) and a linear combination of prior random shocks.

Invertibility requirement: Without going into too much detail, there is a "duality" between the moving average process and the autoregressive process (Box and Jenkins, 1976; Montgomery, Johnson and Gardiner, 1990), that is, the moving average equation above can be rewritten (inverted) into an autoregressive form (of infinite order). However, analogous to the stationarity condition described above, this can only be done if the moving average parameters follow certain conditions, that is, if the model is invertible. Otherwise, the series will not be stationary.

3.1.8 AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) Methodology

Autoregressive moving average model: The general model introduced by Box and Jenkins (1976) includes autoregressive as well as moving average parameters, and explicitly includes differencing in the formulation of the model. Specifically, the three types of parameters in the model are: the autoregressive parameters (*p*), the number of differencing passes (*d*), and moving average parameters (*q*). In the notation introduced by Box and Jenkins, models are summarized as ARIMA (*p, d, q*); so, for example, a model described as (0, 1, 2) means that it contains 0 (zero) autoregressive (*p*) parameters and 2 moving average (*q*) parameters which were computed for the series after it was differenced once.

Identification: As mentioned earlier, the input series for ARIMA needs to be stationary, that is, it should have a constant mean, variance, and autocorrelation through time. Therefore, usually the series first needs to be differenced until it is stationary (this also often requires log transforming the data to stabilize the variance). The number of times the series needs to be differenced to achieve stationarity is reflected in the *d* parameter (see the previous paragraph). In order to determine the necessary level of differencing, you should examine the plot of the data and autocorrelogram. Significant changes in level (strong upward or downward changes) usually require first order non seasonal (lag =1) differencing; strong changes of slope usually require second order non seasonal differencing. Seasonal patterns require respective seasonal differencing. If the estimated autocorrelation coefficients decline slowly at longer lags, first order differencing is usually needed. However, some time series may require little or no differencing, and that *over differenced* series produce less stable coefficient estimates.

At this stage (which is usually called *Identification* phase) we also need to decide how many autoregressive (*p*) and moving average (*q*) parameters are necessary to yield an effective but still *parsimonious* model of the process (*parsimonious* means that it has the fewest parameters and greatest number of degrees of freedom among all models that fit the data). In practice, the numbers of the *p* or *q* parameters very rarely need to be greater than 2.

Estimation and Forecasting: At the next step (*Estimation*), the parameters are estimated (using function minimization procedures, so that the sum of squared residuals is minimized. The estimates of the parameters are used in the last stage (*Forecasting*) to calculate new values of the series (beyond those included in the input data set) and confidence intervals for those predicted values. The estimation process is performed on transformed (differenced) data; before the forecasts are generated, the series needs to be *integrated* (integration is the inverse of differencing) so that the forecasts are expressed in values compatible with the input data. This automatic integration feature is represented by the letter I in the name of the methodology (ARIMA = Auto-Regressive Integrated Moving Average).

The constant in ARIMA models: In addition to the standard autoregressive and moving average parameters, ARIMA models may also include a constant, as described above. The interpretation of a (statistically significant) constant depends on the model that is fit.

Specifically, (1) if there are no autoregressive parameters in the model, then the expected value of the constant is μ, the mean of the series; (2) if there are autoregressive parameters in the series, then the constant represents the intercept. If the series is differenced, then the constant represents the mean or intercept of the differenced series; For example, if the series is differenced once, and there are no autoregressive parameters in the model, then the constant represents the mean of the differenced series, and therefore the *linear trend slope* of the un-differenced series.

3.1.9 Identification Phase

Number of parameters to be estimated: Before the estimation can begin, we need to decide on (identify) the specific number and type of ARIMA parameters to be estimated. The major tools used in the identification phase are plots of the series, correlograms of Autocorrelation Function (ACF), and Partial Autocorrelation Function (PACF). The decision is not straightforward and in less typical cases requires not only experience but also a good deal of experimentation with alternative models (as well as the technical parameters of ARIMA). However, a majority of empirical time series patterns can be sufficiently approximated using one of the 5 basic models that can be identified based on the shape of the Autocorrelogram Function (ACF) and Partial Auto correlogram Function (PACF). The following brief summary is based on practical recommendations of Pankratz (1983); Hoff (1983), McCleary and Hay (1980), McDowall, McCleary, Meidinger, and Hay (1980), and Vandaele (1983). Also, since the number of parameters (to be estimated) of each kind is almost never greater than 2, it is often practical to try alternative models on the same data.

1. *One autoregressive (p) parameter*: ACF - exponential decay; PACF - spike at lag 1, no correlation for other lags.
2. *Two autoregressive (p) parameters*: ACF - a sine-wave shape pattern or a set of exponential decays; PACF - spikes at lags 1 and 2, no correlation for other lags.
3. *One moving average (q) parameter*: ACF - spike at lag 1, no correlation for other lags; PACF - damps out exponentially.
4. *Two moving average (q) parameters*: ACF - spikes at lags 1 and 2, no correlation for other lags; PACF - a sine-wave shape pattern or a set of exponential decays.
5. *One autoregressive (p) and one moving average (q) parameter*: ACF - exponential decay starting at lag 1; PACF - exponential decay starting at lag 1.

Seasonal models: Multiplicative seasonal ARIMA is a generalization and extension of the method introduced in the previous paragraphs to series in which a pattern repeats seasonally over time. In addition to the non-seasonal parameters, seasonal parameters for a specified lag (established in the identification phase) need to be estimated. Analogous to the simple ARIMA parameters, these are: seasonal autoregressive (*ps*), seasonal differencing (*ds*), and seasonal moving average parameters (*qs*). For example, the model (*0,1,2*)(*0,1,1*) describes a model that includes no autoregressive parameters, 2 regular moving average parameters and 1 seasonal moving average parameter, and these parameters were computed for the series after it was differenced once with lag 1, and once seasonally differenced. The seasonal lag used for the seasonal parameters is usually determined during the identification phase and must be explicitly specified.

The general recommendations concerning the selection of parameters to be estimated (based on ACF and PACF) also apply to seasonal models. The main difference is that in seasonal series, ACF and PACF will show sizable coefficients at multiples of the seasonal lag (in addition to their overall patterns reflecting the non seasonal components of the series).

3.1.10 Parameter Estimation

There are several different methods for estimating the parameters. All of them should produce very similar estimates, but may be more or less efficient for any given model. In general, during the parameter estimation phase a function minimization algorithm is used (the so-called *quasi-Newton* method; refer to the description of the *Nonlinear Estimation* method) to maximize the likelihood (probability) of the observed series, given the parameter values. In practice, this requires the calculation of the (conditional) sums of squares (SS) of the residuals, given the respective parameters. Different methods have been proposed to compute the SS for the residuals: (1) the approximate maximum likelihood method according to McLeod and Sales (1983), (2) the approximate maximum likelihood method with backcasting, and (3) the exact maximum likelihood method according to Melard (1984).

Comparison of methods: In general, all methods should yield very similar parameter estimates. Also, all methods are about equally efficient in most real-world time series applications. However, method *1* above, (approximate maximum likelihood, no backcasts) is the fastest, and should be used in particular for very long time series (e.g., with more than 30,000 observations). Melard's exact maximum likelihood method (number *3* above) may also become inefficient when used to estimate parameters for seasonal models with long seasonal lags (e.g., with yearly lags of 365 days). On the other hand, you should always use the approximate maximum likelihood method first in order to establish initial parameter estimates that are very close to the actual final values; thus, usually only a few iterations with the exact maximum likelihood method (*3*, above) are necessary to finalize the parameter estimates.

Parameter standard errors: For all parameter estimates, you will compute so-called *asymptotic standard errors*. These are computed from the matrix of second-order partial derivatives that is approximated via finite differencing.

Penalty value: As mentioned above, the estimation procedure requires that the (conditional) sums of squares of the ARIMA residuals be minimized. If the model is inappropriate, it may happen during the iterative estimation process that the parameter estimates become very large, and, in fact, invalid. In that case, it will assign a very large value (a so-called *penalty value*) to the SS. This usually "entices" the iteration process to move the parameters away from invalid ranges. However, in some cases even this strategy fails, and you may see on the screen (during the *Estimation procedure*) very large values for the SS in consecutive iterations. In that case, carefully evaluate the appropriateness of your model. If your model contains many parameters, and perhaps an intervention component, you may try again with different parameter start values.

3.1.11 Evaluation of the Model

Parameter estimates: we shall report approximate t values, computed from the parameter standard errors. If not significant, the respective parameter can in most cases be dropped from the model without affecting substantially the overall fit of the model.

Other quality criteria: Another straightforward and common measure of the reliability of the model is the accuracy of its forecasts generated based on partial data so that the forecasts can be compared with known (original) observations.

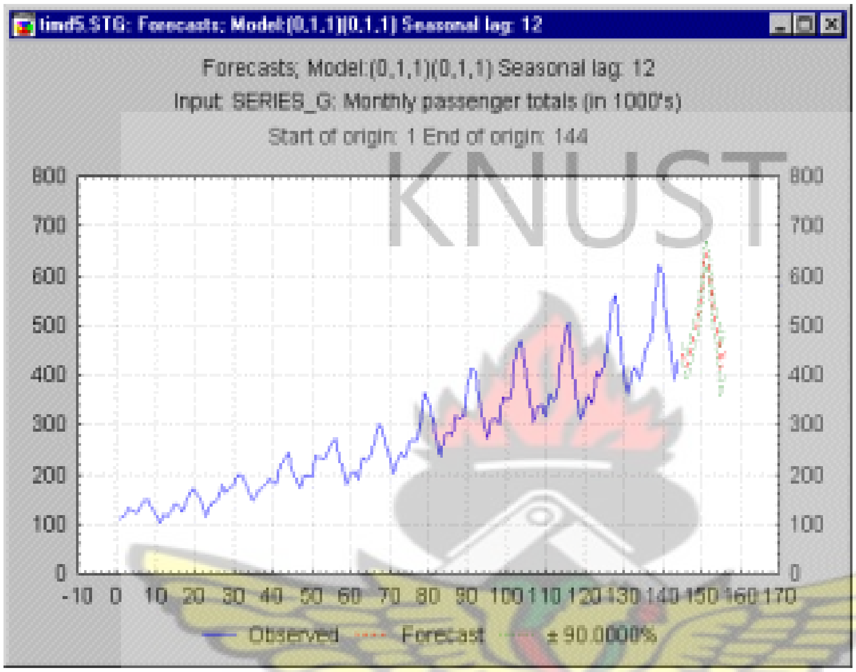


Figure 3.3: Graph of Forecast, Model (0,1,1) (0,1,1) Seasonal lag 12

However, a good model should not only provide sufficiently accurate forecasts, it should also be parsimonious and produce statistically independent residuals that contain only noise and no systematic components (e.g., the correlogram of residuals should not reveal any serial dependencies). A good test of the model is (a) to plot the residuals and inspect them for any systematic trends, and (b) to examine the autocorrelogram of residuals (there should be no serial dependency between residuals)

Analysis of residuals: The major concern here is that the residuals are systematically distributed across the series (e.g., they could be negative in the first part of the series and approach zero in the second part) or that they contain some serial dependency which may suggest that the ARIMA model is inadequate. The analysis of ARIMA residuals constitutes an important test of the model. The estimation procedure assumes that the residual are not (auto-) correlated and that they are normally distributed.

Limitations: The ARIMA method is appropriate only for a time series that is stationary (that is, its mean, variance, and autocorrelation should be approximately constant through time) and it is recommended that there are at least 50 observations in the input data. It is also assumed that the values of the estimated parameters are constant throughout the series.

3.1.12 Estimating the parameters of an ARIMA Model

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

3.1.13 Stationarity and Invertibility Conditions of Specific Time Series model

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

Table 3.1 Specific Time Series Models

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

3.2 THE BOX-JENKINS METHOD OF MODELING TIME SERIES

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

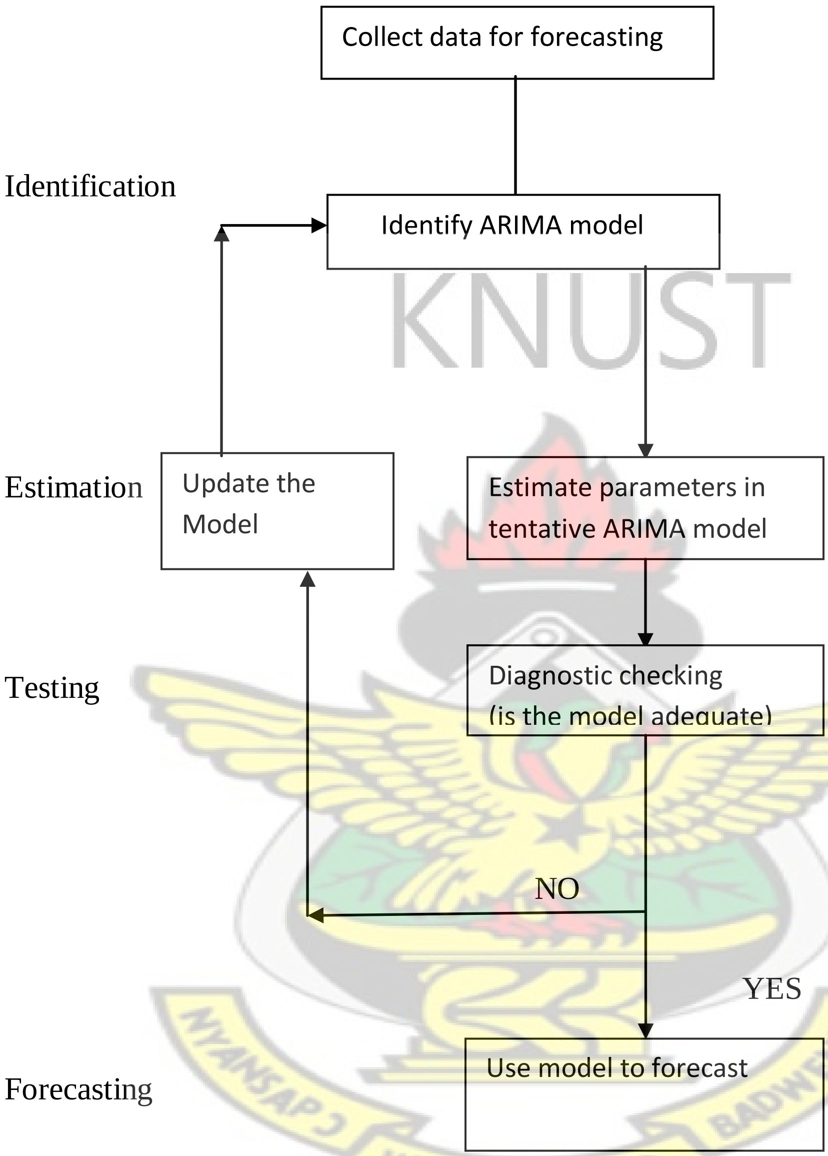
The third stage is the diagnostic checking of the model. The Q-statistic is used for the model adequacy check.

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

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

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

Figure 3.4 The Box-Jenkins Process



3.2.1 Identification techniques

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

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

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

(b) To identify the resulting ARMA process

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

3.2.2 Use of the autocorrelation and Partial Autocorrelation functions in Identification

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

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

Also, if

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

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

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

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

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

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

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

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

3.2.3 Identifying the Resulted Stationary ARMA process

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

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

3.2.4 Akaike's Information Criteria (AIC)

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

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

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

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

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

3.2.5 Schwarz's Bayesian Information Criterion (BIC)

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

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

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

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

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

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

3.2.6 Estimation of the Parameters of the Model Identified

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

3.2.7 Testing the Model for Adequacy

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

The test statistic is the Q –statistic.

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

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

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

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

3.3 FORECASTING

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

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

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

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

Then
$$Y_{61} = Y_{60} + 0.2178(Y_{60} - Y_{59})$$
$$Y_{61} = 131.2 + 0.2178(131.2 - 134.8)$$
$$Y_{61} = 130.097$$

Hence, a forecast value for period *61* is *130.097*.

3.4 EXPONENTIAL SMOOTHING

Exponential smoothing has become very popular as a forecasting method for a wide variety of time series data. Historically, the method was independently developed by Brown and Holt. Brown worked for the US Navy during World War II, where his assignment was to design a tracking system for fire-control information to compute the location of submarines. Later, he applied this technique to the forecasting of demand for spare parts (an inventory control problem). He described those ideas in his 1959 book on inventory control. Holt's research was sponsored by the Office of Naval Research; independently, he developed exponential smoothing models for constant processes, processes with linear trends, and for seasonal data.

Gardner (1985) proposed a "unified" classification of exponential smoothing methods. Excellent introductions can also be found in Makridakis, Wheelwright, and McGee (1983), Makridakis and Wheelwright (1989), Montgomery, Johnson and Gardiner (1990).

3.4.1 Simple Exponential Smoothing

A simple and pragmatic model for a time series would be to consider each observation as consisting of a constant (*b*) and an error component ε(epsilon), that is: $X_t = b + \varepsilon_t$. The constant *b* is relatively stable in each segment of the series, but may change slowly over time. If appropriate, then one way to isolate the true value of *b*, and thus the systematic or predictable part of the series, is to compute a kind of moving average, where the current and immediately preceding ("younger") observations are assigned greater weight than the respective older observations. Simple exponential smoothing accomplishes exactly such weighting, where exponentially smaller weights are assigned to older observations. The specific formula for simple exponential smoothing is:

$$S_t = \alpha * X_t + (1- \alpha) * S_{t-1}$$

When applied recursively to each successive observation in the series, each new smoothed value (forecast) is computed as the weighted average of the current observation and the previous smoothed observation; the previous smoothed observation was computed in turn from the previous observed value and the smoothed value before the previous observation, and so on. Thus, in effect, each smoothed value is the weighted average of the previous observations, where the weights decrease exponentially depending on the value of parameter α (alpha). If α is equal to 1 (one) then the previous observations are ignored entirely; if α is equal to 0 (zero), then the current observation is ignored entirely, and the smoothed value consists entirely of the previous smoothed value (which in turn is computed from the smoothed observation before it, and so on; thus all smoothed values will be equal to the initial smoothed value S_0). Values of α in-between will produce intermediate results.

Even though significant work has been done to study the theoretical properties of (simple and complex) exponential smoothing (Gardner, 1985; Muth, 1960; McKenzie, 1984, 1985), the method has gained popularity mostly because of its usefulness as a forecasting tool. For example, empirical research by Makridakis *et al;* (1982, Makridakis, 1983), has shown simple exponential smoothing to be the best choice for one-period-ahead forecasting, from among 24 other time series methods and using a variety of accuracy measures. Thus, regardless of the theoretical model for the process underlying the observed time series, simple exponential smoothing will often produce quite accurate forecasts

3.4.2 Choosing the Best Value for Parameter α (alpha)

Gardner (1985) discusses various theoretical and empirical arguments for selecting an appropriate smoothing parameter. Obviously, looking at the formula presented above, α should fall into the interval between 0 (zero) and 1 (although, Brenner *et al.*, 1968, for an ARIMA perspective, implying $0 < \alpha < 2$). Gardner (1985) reports that among practitioners, an α smaller than 0.30 is usually recommended. However, in the study by Makridakis *et a*; (1982), α values above .30 frequently yielded the best forecasts. After reviewing the literature on this topic, Gardner (1985) concludes that it is best to estimate an optimum α from the data, rather than to "guess" and set an artificially low value.

Estimating the best α value from the data: In practice, the smoothing parameter is often chosen by a *grid search* of the parameter space; that is, different solutions for α are tried starting, for example, with $\alpha = 0.1$ to $\alpha = 0.9$, with increments of 0.1. Then α is chosen so as to produce the smallest sums of squares (or mean squares) for the residuals (i.e., observed values minus one-step-ahead forecasts; this mean squared error is also referred to as *ex post* mean squared error, *ex post* MSE for short).

3.4.3 Indices of Lack of Fit (Error)

The most straightforward way of evaluating the accuracy of the forecasts based on a particular α value is to simply plot the observed values and the one-step-ahead forecasts. This plot can also include the residuals (scaled against the right *Y*-axis), so that regions of better or worst fit can also easily be identified.

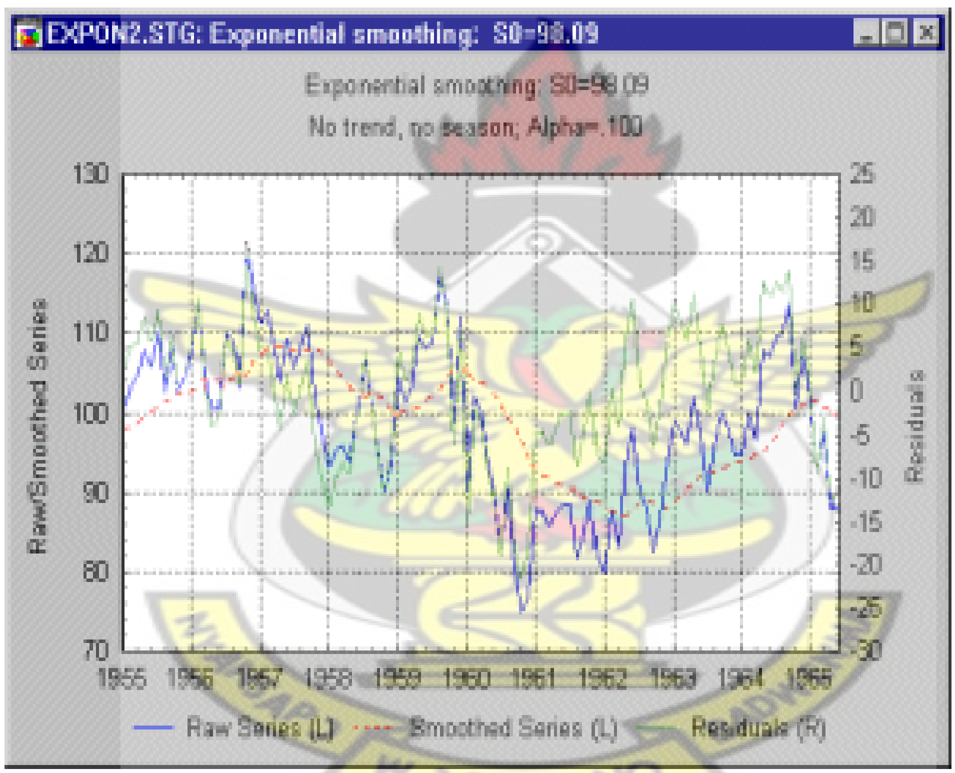


Figure 3.5: Graph of Exponential smoothing

This visual check of the accuracy of forecasts is often the most powerful method for determining whether or not the current exponential smoothing model fits the data. In addition, besides the *ex post* MSE criterion, there are other statistical measures of error that can be used to determine the optimum α parameter (Makridakis, Wheelwright, and McGee, 1983):

Mean error: The mean error (ME) value is simply computed as the average error value (average of observed minus one-step-ahead forecast). Obviously, a drawback of this measure is that positive and negative error values can cancel each other out, so this measure is not a very good indicator of overall fit.

Mean absolute error: The mean absolute error (MAE) value is computed as the average *absolute* error value. If this value is 0 (zero), the fit (forecast) is perfect. As compared to the mean *squared* error value, this measure of fit will "de-emphasize" outliers, that is, unique or rare large error values will affect the MAE less than the MSE value.

Sum of squared error (SSE), Mean squared error: These values are computed as the sum (or average) of the squared error values. This is the most commonly used lack-of-fit indicator in statistical fitting procedures.

Percentage error (PE): All the above measures rely on the actual error value. It may seem reasonable to rather express the lack of fit in terms of the *relative* deviation of the one-step-ahead forecasts from the observed values, that is, relative to the magnitude of the observed values. For example, when trying to predict monthly sales that may fluctuate widely (e.g., seasonally) from month to month, we may be satisfied if our prediction "hits the target" with about ±10% accuracy. In other words, the absolute errors may be not so much of interest as are the relative errors in the forecasts. To assess the relative error, various indices have been proposed (Makridakis, Wheelwright, and McGee, 1983). The first one, the percentage error value, is computed as:

$PE_t = 100*(X_t - F_t)/X_t$

where X_t is the observed value at time t, and F_t is the forecasts (smoothed values).

Mean percentage error (MPE): This value is computed as the average of the PE values.

Mean absolute percentage error (MAPE). As is the case with the mean error value (ME, see above), a mean percentage error near 0 (zero) can be produced by large positive and negative percentage errors that cancel each other out. Thus, a better measure of relative overall fit is the mean *absolute* percentage error. Also, this measure is usually more meaningful than the mean squared error. For example, knowing that the average forecast is "off" by ±5% is a useful result in and of itself, whereas a mean squared error of 30.8 is not immediately interpretable.

Automatic search for best parameter: A quasi-Newton function minimization procedure (the same as in ARIMA is used to minimize either the mean squared error, mean absolute error, or mean absolute percentage error. In most cases, this procedure is more efficient than the grid search (particularly when more than one parameter must be determined), and the optimum α parameter can quickly be identified..

CHAPTER FOUR

4.0 DATA ANALYSIS

In this chapter we analyzed the share prices of EPACK by employing Box-Jenkins method of analyzing time series data.

Data comprising of the average monthly EPACK share prices from January 1999 to December 2010 was obtained from the DATABANK Ghana Ltd, Kumasi office and used for the analysis.

4.1 PRELIMINARY ANALYSIS

This preliminary analysis consists of the computation of the descriptive statistics in the relation to the data. The results are displayed in the table below.

The table below shows the descriptive statistics of the data.

Table 4.1 Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
EPACK SHARE PRICES	144	0.04	0.97	0.42	.32155	.103

Table 4.1 above displays some of the basic descriptive statistics of the share price of EPACK. It can be observed that over the period under consideration EPACK had a minimum share price of GH¢ 0.04 from the beginning (i.e. January 1999) and a maximum of GH¢ 0.97 .

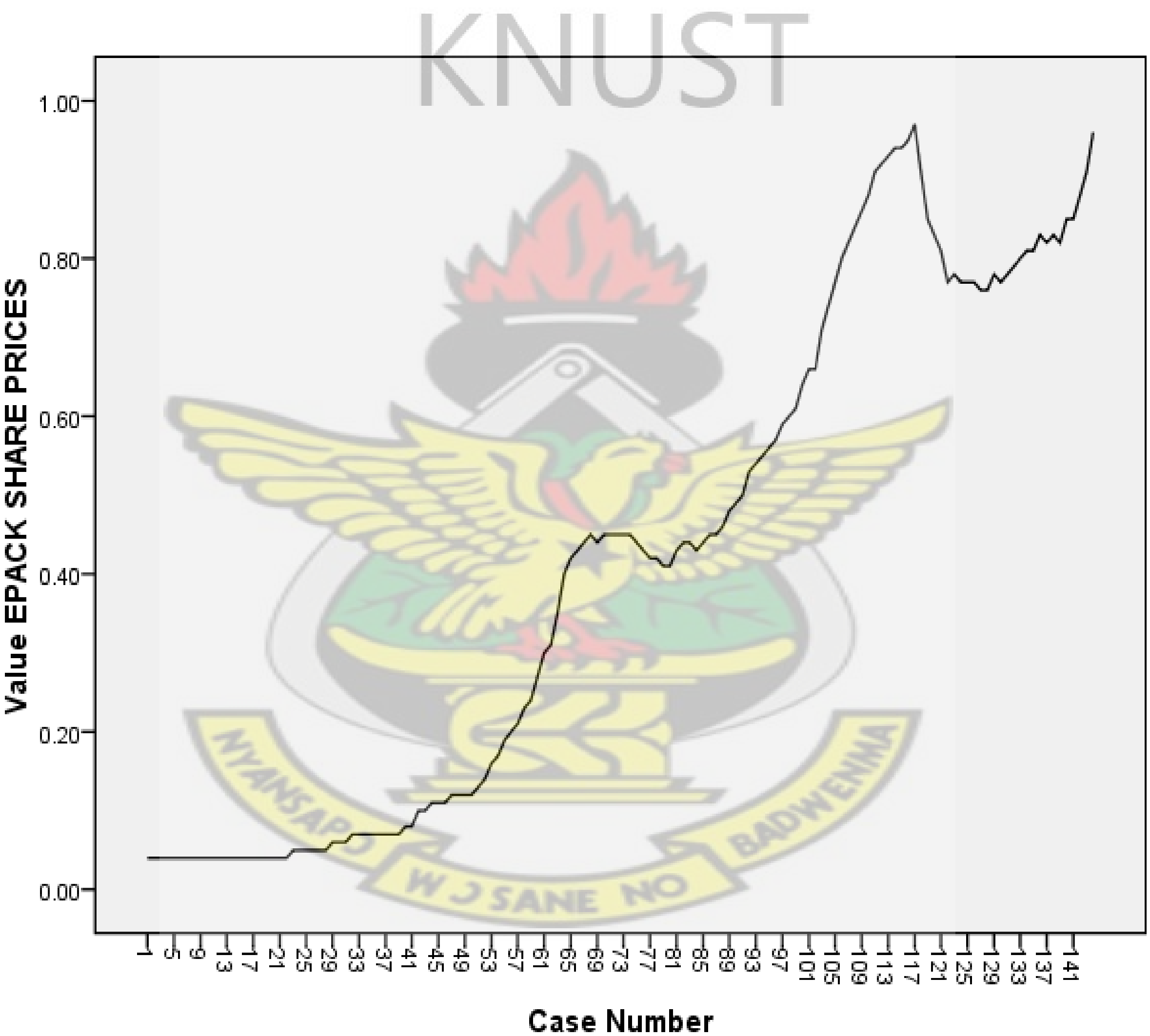


Fig. 4.1 Trajectory of EPACK share prices from January 1995 to February 2011

From Figure 4.1 above the EPACK share price was approximately constant from the first data point to about the 21st data point before having a rise. It can be observed the in general the periodogram of the share prices of EPACK over the years exhibits an upwardly moving quadratic trend in the mean. This shows that the data is not stationary in the mean hence there is the need to difference twice in order to attain stationarity.

Below is the graph obtained after the first differencing.

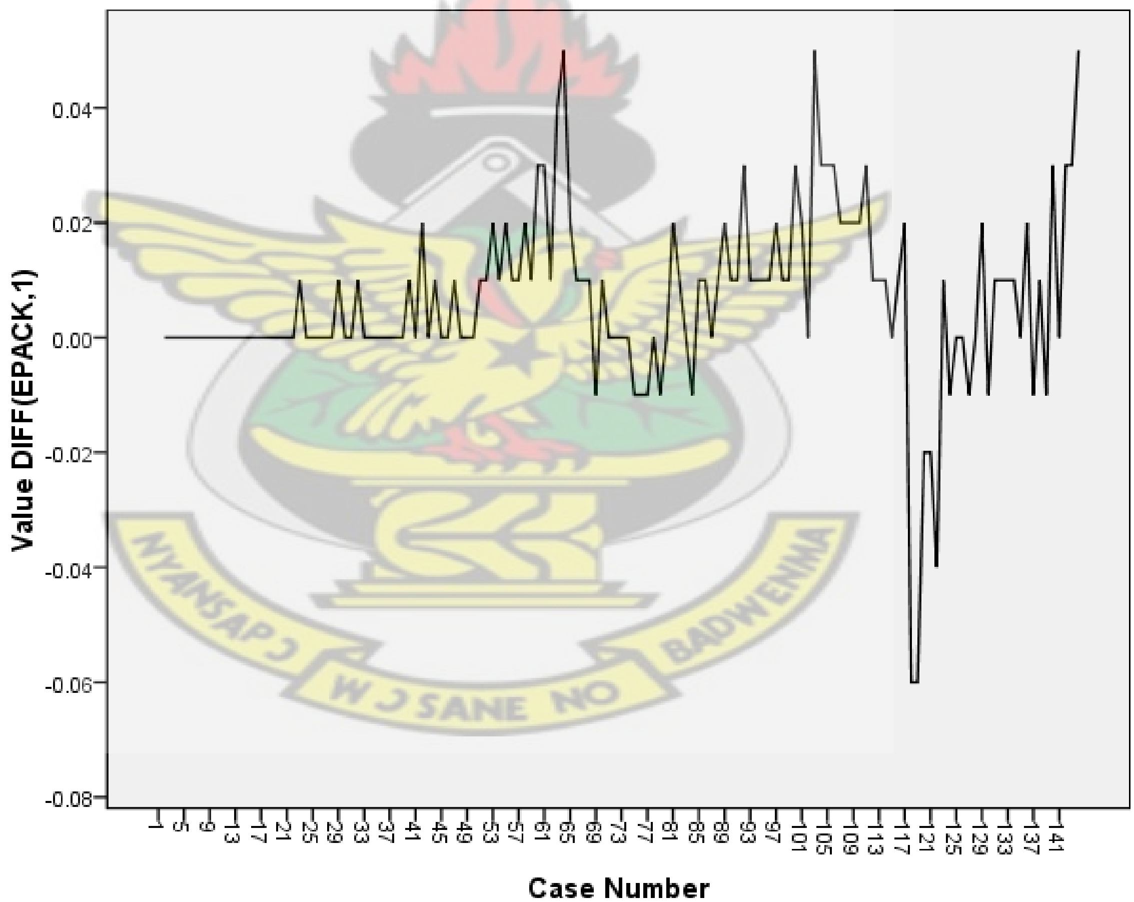


Fig. 4.2 Graph of First order differeced EPACK share prices from January 1999 to February 2011

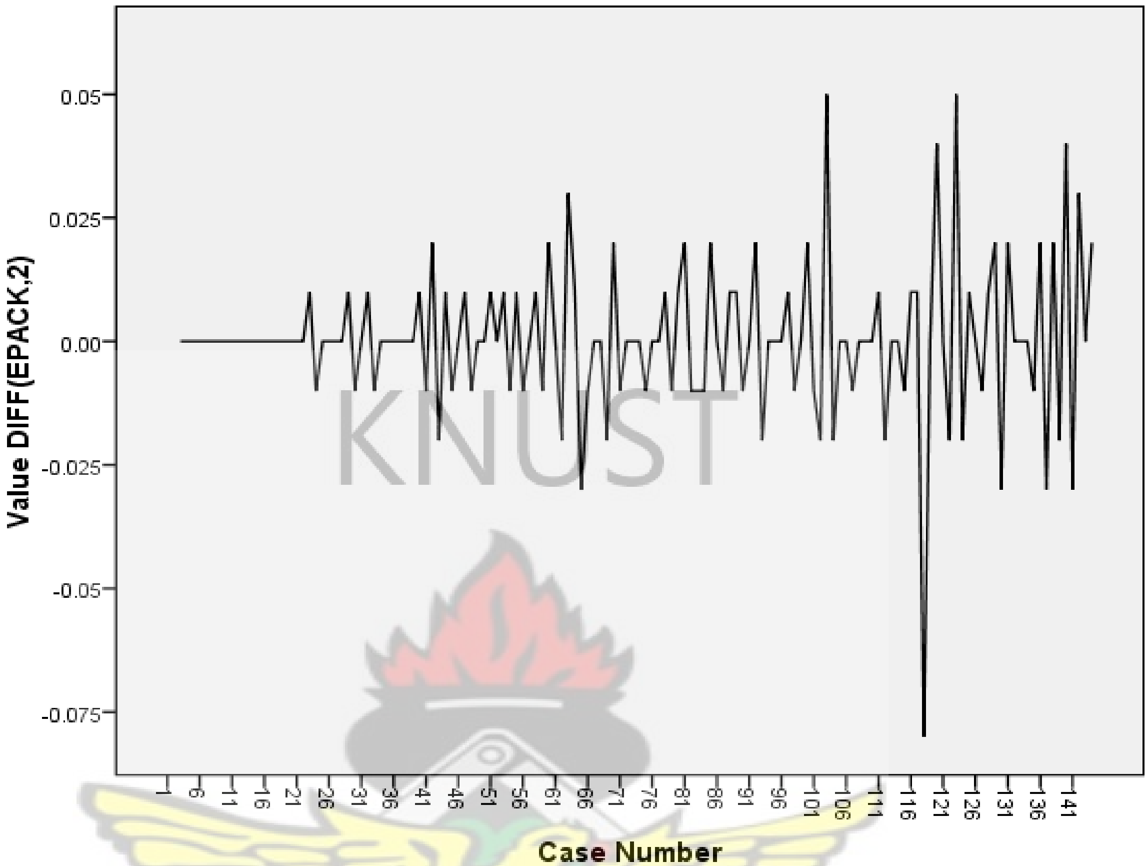


Fig. 4.3 Graph of Second order differeced EPACK share prices from January 1995 to February 2011

From the two graphs above the graph after the first difference exhibits non-stationarity whilst that after second order is stationary. Since the second order differenced data is stationary it is then used to model the share prices.

4.2 MODEL IDENTIFICATION

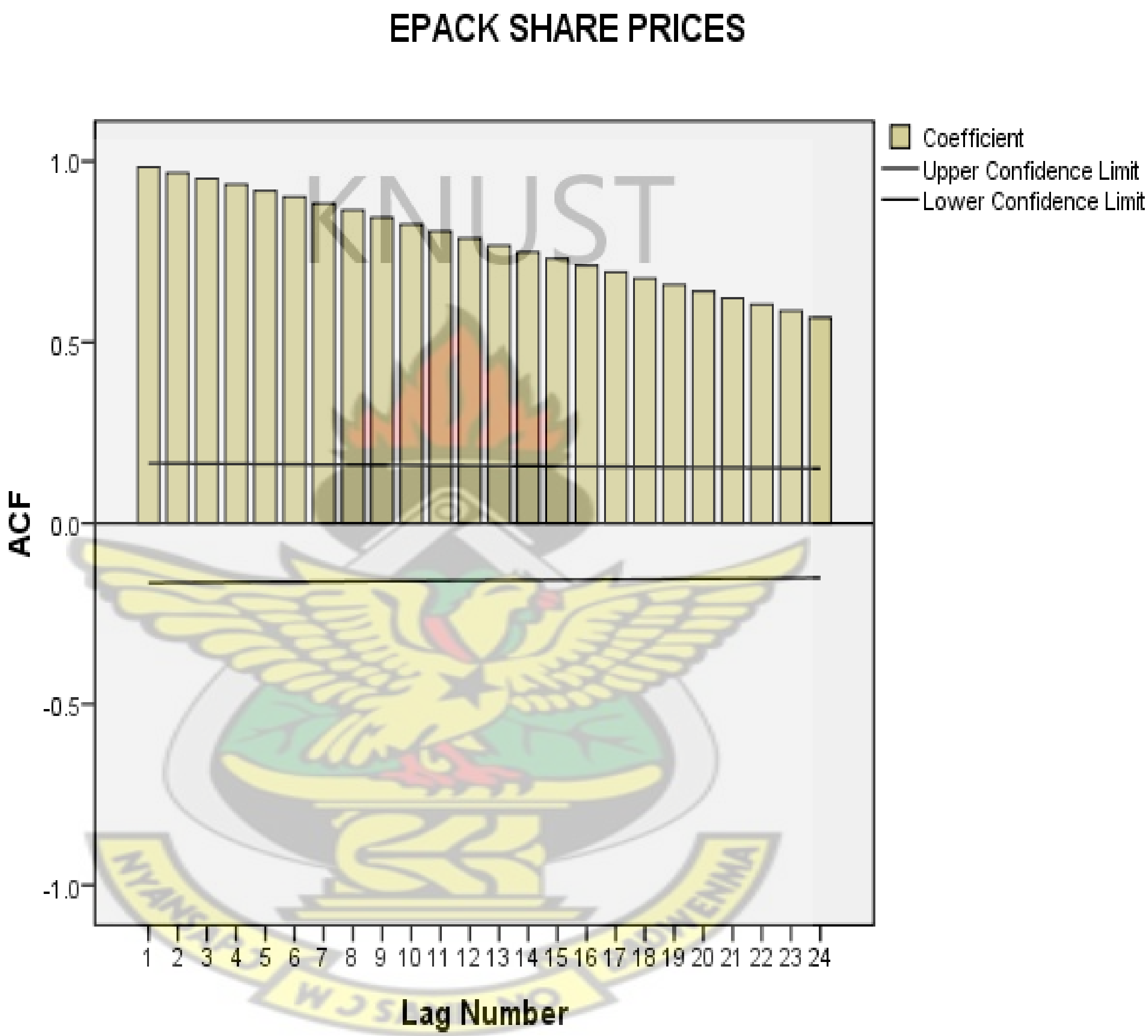


Fig. 4.4 Autocorrelation Function of PACK share prices

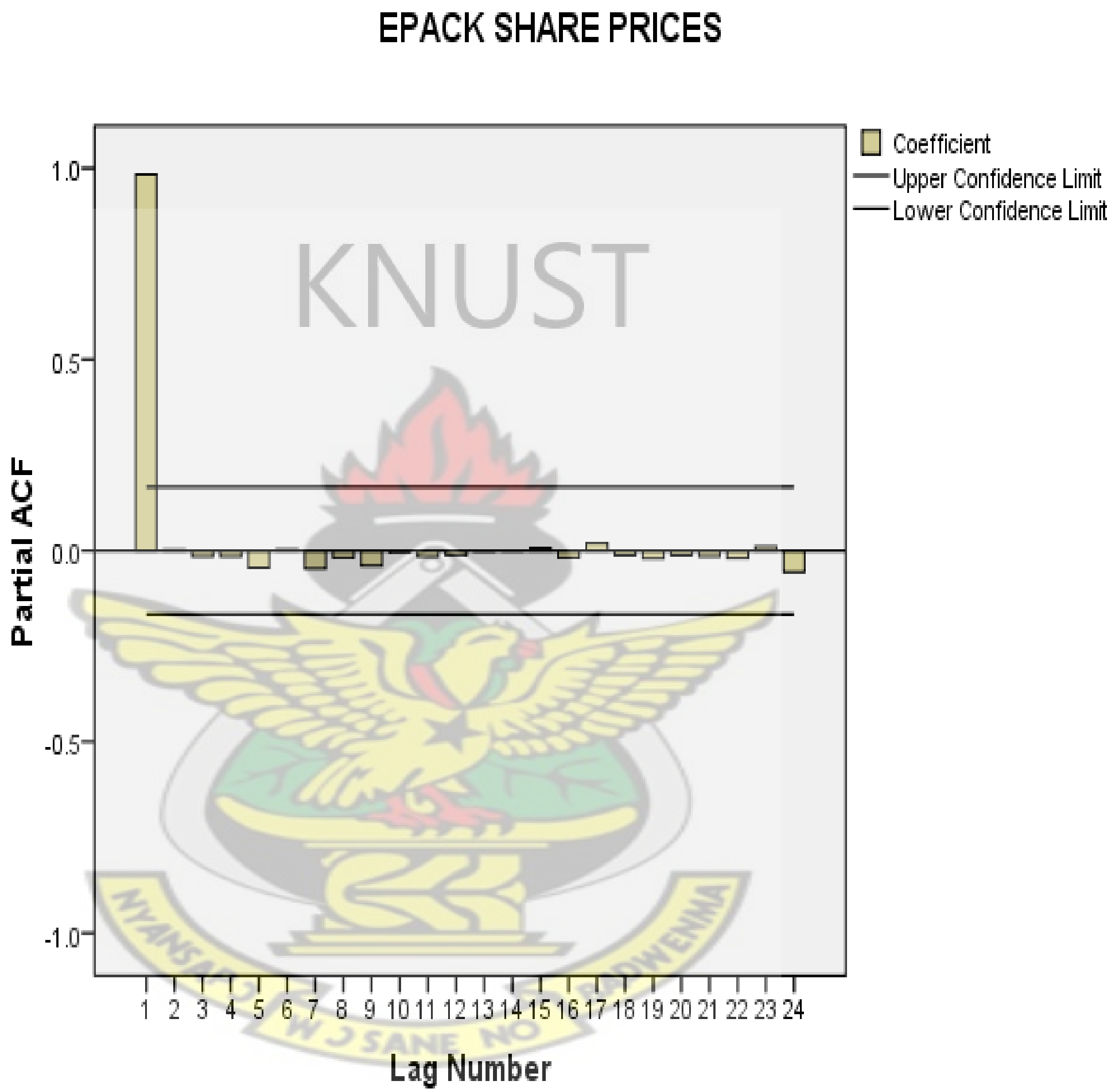


Fig. 4.5 Partial Autocorrelation Function of PACK share prices

The figures 4.4 and 4.3 above shows the autocorrelation and partial autocorrelation functions of the EPACK share prices for the period under consideration. The autocorrelation function dies down at a very slow rate confirm the existence of a trend in the mean whilst the partial autocorrelation function truncates after the first lag.

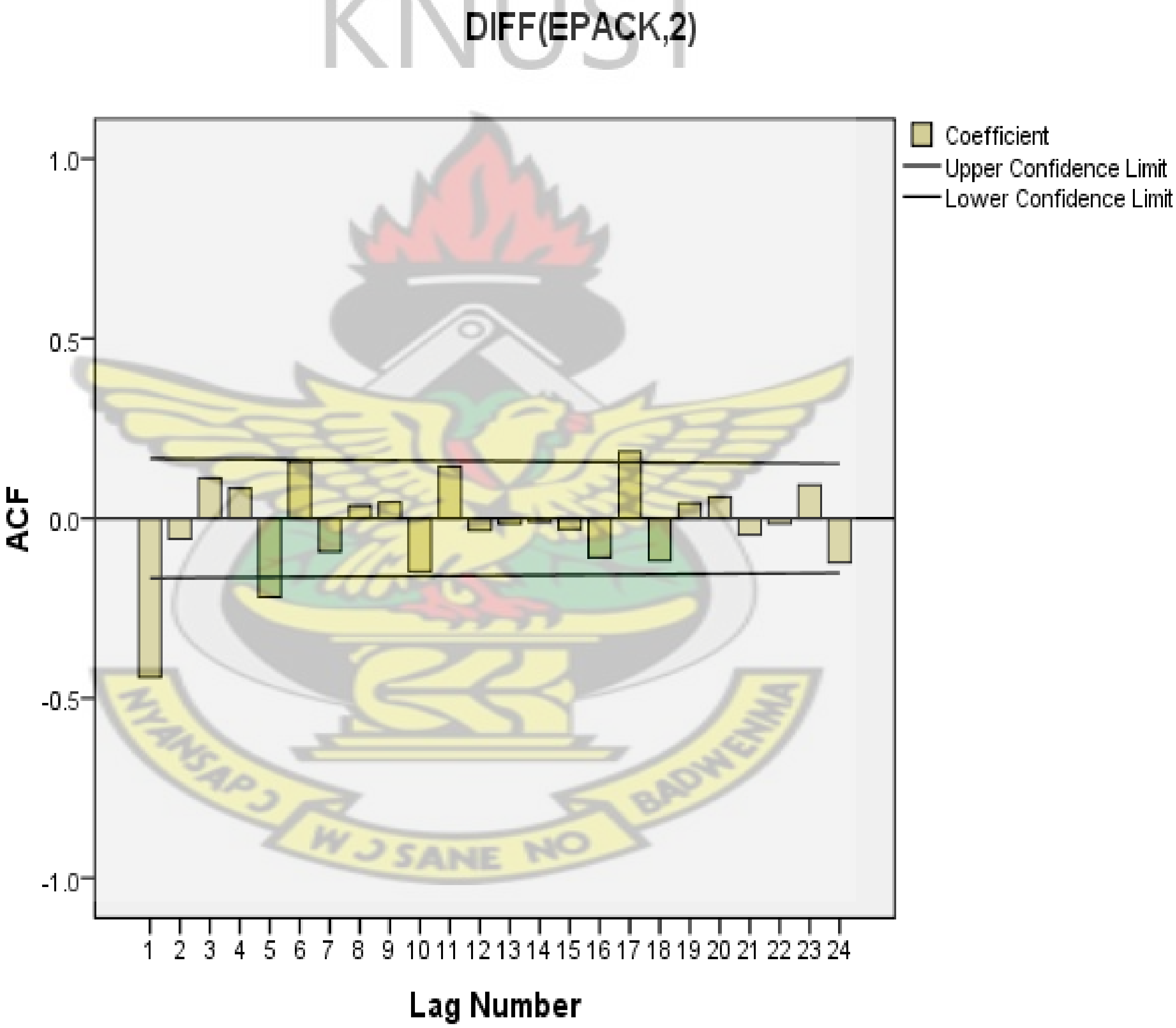


Fig. 4.6 ACF of Second order differeced EPACK share prices

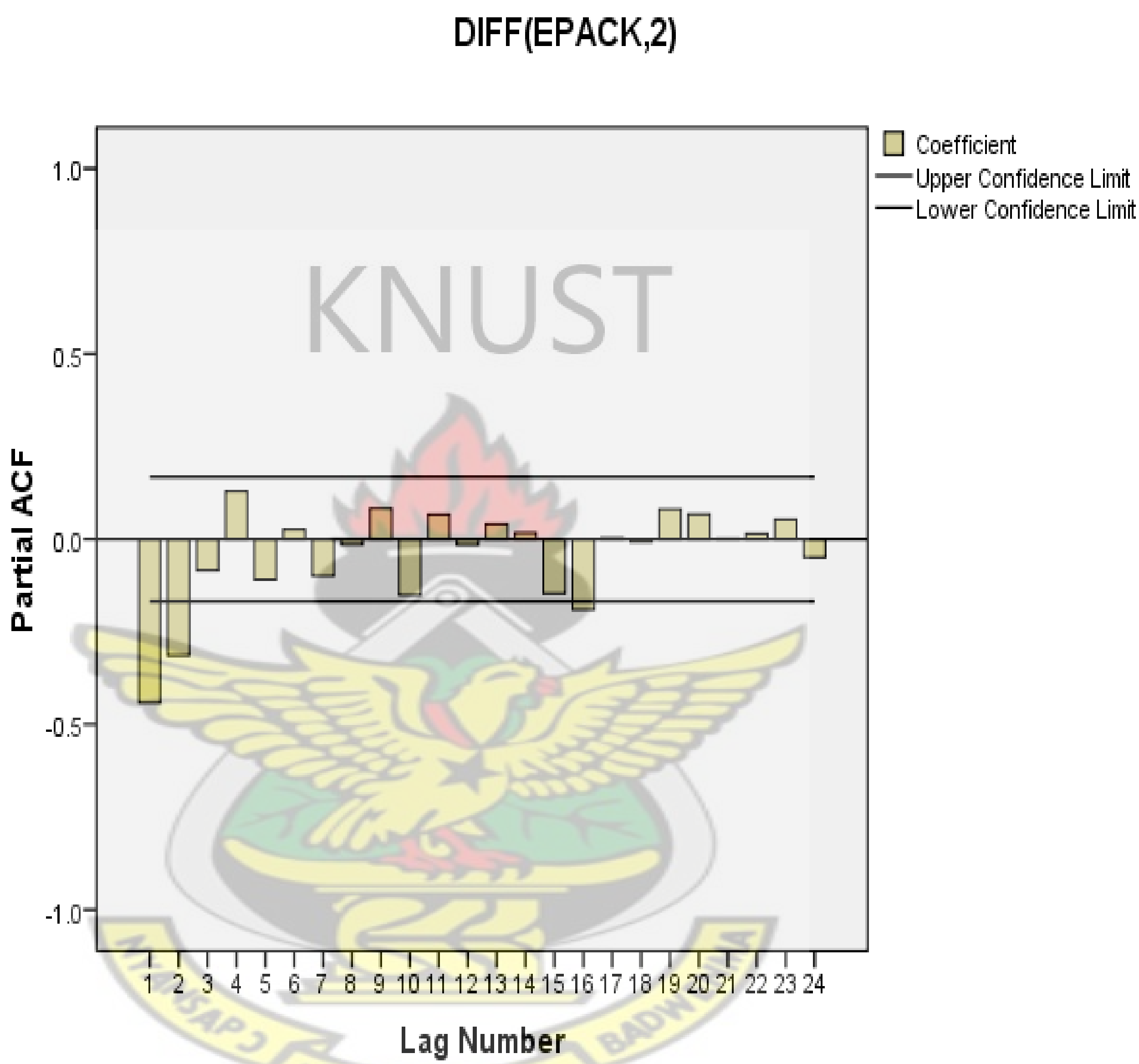


Fig. 4.7 PACF of Second order differeced EPACK share prices

From the correllograms above the autocorrelation functions tails off quickly to zero whilst the partial autocorrelation function truncates after lag 2. This gives an indication of an $AR(2)$ model.

Because the data was first differenced to attain stationarity, the actual model identified is $ARIMA(2,2,0)$. The identified model might not be the best model for the data, therefore it is compared with other probable models and the best is chosen.

4.3 MODEL SELECTION

The identified model $ARIMA(2,2,0)$ is compared with other stationary second order differenced models, that is $ARIMA(2,2,0)$, $ARIMA(0,2,1)$, $ARIMA(0,2,2)$, $ARIMA(1,2,1)$, $ARIMA(2,2,1)$, $ARIMA(1,2,2)$ and $ARIMA(2,2,2)$, and the model of the least residuals is selected.

Table 4.2 Model Selection

MODEL	**RMSE**	**MAPE**	**NORMALIZED BIC**
$ARIMA(1,2,0)$	0.014	80.011	−8.495
$ARIMA(2,2,0)$	0.013	75.006	−8.661
$ARIMA(0,2,1)$	0.013	76.054	−8.591
$ARIMA(0,2,2)$	0.013	75.168	−8.553
$ARIMA(1,2,1)$	0.013	75.442	−8.551
$ARIMA(2,2,1)$	0.013	75.636	−8.526
$ARIMA(1,2,2)$	0.013	75.132	−8.821
$ARIMA(2,2,2)$	0.013	75.110	−8.488

From table 4.2 above, $ARIMA(2,2,2)$ is ignored because of the principle of parsimony, which says that given any two models the model with the least number of parameters is more preferred.

Comparing the rest of the models based on the RMSE(Root Mean Square Error), MAPE (Mean Absolute Percentage Error) and BIC(Bayesians Information Criterion) $ARIMA(2,2,0)$ is has the least error value. Hence our statistically most preferred model is $ARIMA(2,2,0)$.

The theoretical form of the model is given by

$$Y_t = \alpha_1 Y_{t-1} + (1 + \alpha_2) Y_{t-2} - \alpha_1 Y_{t-3} - \alpha_2 Y_{t-4} + e_t$$

4.4 MODEL ADEQUACY

The selected model is now diagnostically tested. This is done by the fact that the Q-statistics is X^2 distributed with $(k - p - q)$ degrees of freedom. Where k is the maximum time lag $(k = 24)$, $p = 2$ is the number of AR parameters and $q = 0$ is the number of Moving Average parameters

Table 4.3 **Model Statistic**

Model	Ljung-Box Q(18)	
	Statistics	Sig.
DIFF(EPACK,2)-Model_1	20.699	0.002

The estimate Box-Ljung Q-statistics for the model is 20.00.

Since $Q - statistics < {X^2}_{22} = 33.924$ the fitted model, that is $ARIMA(2,2,0)$ is of best fit and can be used to forecast.

4.5 PARAMETER ESTIMATE

Table 4.4 ARIMA Model Parameters

			Estimate	SE	Sig.
DIFF(EPACK,2)-Model_1		Constant	0.000	.001	.001
	AR	Lag 1	-0.587	.081	.000
		Lag 2	-0.318	.081	.000

The Table above displays the parameter estimates of the selected model. After the parameters are estimated the model for the share price of EPACK over the period under consideration is given by

$$Y_t = -0.587Y_{t-1} + 0.682Y_{t-2} + 0.587Y_{t-3} + 0.318Y_{t-4}$$

4.6 TIME PATH

The developed model for the EPACK share prices (that is

$Y_t = -0.587 + 0.682Y_{t-2} + 0.587Y_{t-3} + 0.318Y_{t-4}$) is a fourth order homogeneous difference equation with constant coefficients. The characteristic equation corresponding to this model is given by

$$\lambda^4 + 0.587\lambda^3 - 0.682\lambda^2 - 0.587\lambda - 0.318 = 0$$

with characteristic roots $\lambda_1 = 0.6754$, $\lambda_2 = 0.4734$, $\lambda_{3,4} = -0.8689 \pm 0.4895$.

The solution of the model equation is

$$Y_t = A(0.6754)^t + B(0.4734)^t + C(-0.8689 + 0.4895)^t + D(-0.8689 - 0.4895)^t$$

where A, B, C and D are constants.

The characteristic root with the largest absolute value is called the dominant root because it dominates the time path. For convergence, the absolute value of the dominant root must be less than $one(1)$ [that is $|\lambda_*| < 1$].

Since the dominant real characteristic root is $|0.06754| < 1$, the model will converge as $t \rightarrow \infty$.

CHAPTER FIVE

5.0 FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

This chapter summarizes the thesis by enumerating the finding, conclusions and recommendations.

5.1 FORECASTING

The obtained model is used to forecast for some future values of EPACK share prices. This is done by writing a MatLab code to forecast some future values.

Table 5.1 The forecasted values

DATE	FORECAST (GH¢)
January 2011	0.84
February 2011	0.97
March 2011	0.85
April 2011	0.96
May 2011	0.85
June 2011	0.96

5.2 FINDINGS

The following findings were made;

1. The minimum share price of EPACK share price over the period under consideration is ¢ 0.04 and it occurred at the start of the fund, that is January 1999 through to October 2000, and its Maximum share price over the period is ¢ 0.97 which also occurred on September 2008.
2. The average share price over the period is ¢ 0.42 with a standard deviation of ¢ 0.32
3. EPACK share prices over the years exhibited upwardly moving trend with a significant fall in the year 2009, but continued rising from January 2010 to December 2010.
4. EPACK share prices are so autocorrelated that there had to be a second order differencing to attain stationarity.
5. The ARIMA model equation for EPACK share price over the period under consideration is

$$Y_t = -0.587Y_{t-1} + 0.682Y_{t-2} + 0.587Y_{t-3} + 0.318Y_{t-4}$$

6. The EPACK share price ARIMA model is a fourth order difference equation with constant coefficients and since its dominant characteristic root is less than 1 (in absolute terms) the share price will eventually converge sinusoidally to a value.

5.3 CONCLUSIONS

The following conclusions were made after the research.

1. EPACK Fund share prices can be modeled using Box-Jenkins methodology of time series analysis.

2. With a general upwardly moving quadratic trend in the EPACK Fund share prices over time shows confirms the assertion that EPACK Fund is a good investment package.

3. The derived model is practicable because of its convergence and from the model EPACK share prices depends more on its past two observations that is. lag 2.

4. The 2009 fall in the EPACK share prices can be attributed to the global economic meltdown in 2008.

5. EPACK is not only being talked about but its share prices have experienced tremendous increase in value since its commencement.

6. All things being equal the EPACK share will in the long run converge slowly.

5.4 RECOMMENDATIONS

Upon a successful research the following recommendation are made;

1. With the exhibition of the upwardly moving trend in the EPACK share prices investor are advised to invest in the fund.

2. Fund Managers of EPACK should keep their current policy, since its working, but should research into other workable policy so as to maintain their shareholders when conditions change.
3. Further research should be done to ascertain the actual cause for the significant fall in the EPACK share price in 2009.

REFERENCE

1. Barham Catherine and Begum Nasima (2007) ‘‘Time series analysis of the Labour Force Survey longitudinal data sets”, Economic &Labour Market Review **1 (1)**.
2. Bernhard Pfaff (2005). Analysis of Integrated and cointegrated Time Series with R. Springer.
3. Box G.E.P. and Jenkins G.M.. (1976). Time Series Analysis: Forecasting and Control. Holden-Day, revised edition.
4. Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control. San Francisco: Holden-Day.
5. Box, G.E.; Jenkins, G. M.; Reinsel, G. C. (1994). Time series analysis: Forecasting andcontrol. 3rd.Edition. New Jersey: Prentice Hall
6. Box, G.E.P. – Jenkins, G.M. (1970, 1976) Time Series Analysis, Forecasting and Control. Revised Edition 1976, Holden-Day, San Francisco.
7. Brockwell, Peter J. and Davis, Richard A. Time Series: Theory and Methods. New York: Springer-Veriag Inc., 1987.
8. Bronislav Chramcov (2010), Forecast model of heat demand based on the Box-Jenkins Methodology. World Scientific and Engineering Academy and Society (WSEAS). Stevens Point Wisconsin, USA, 252-256.
9. Chan, Wai-Sum (1989) Some Robust Methods for Time Series Modeling. Ph.D. dissertation, Temple University, USA
10. Dobre Ion, Alexandru Adriana AnaMaria (2008), Modelling Unemployment Rate using Box-Jenkins Procedure, *Journal of Applied Quantitative Methods,* **3(2),** 156-166;

11. Engle R.F. and Granger C.W.J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55:251–176.

12. Hirt Geoffrey and Block B. Stanley (1998), Fundamentals of Investment Management. McGraw Hill College.

13. Hoff, J. C. (1983). A practical guide to Box-Jenkins forecasting. London: Lifetime Learning Publications.

14. Katarina Juselius (2006). The Cointegrated VAR Model: Methodology and Applications. Springer,

15. Kendall, M.G. – Ord, J.K. (1990) Time Series. 3rd edition, Sevenoaks: Edward Arnold.

16. McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. A. (1980). Interrupted time series analysis. Beverly Hills, CA: Sage Publications.

17. Mélard, G. – Roy, R. (1987) Confidence intervals and tests for autocorrelations. *Computational Statistics & Data Analysis* 5, 31–44.

18. Melard, G. (1984). A fast algorithm for the exact likelihood of autoregressive-moving average models. *Applied Statistics*, **33,** 104-119.

19. NauR. F., "Introduction to arima: nonseasonal models", 2005. http://www.duke.edu/rnau/411arim.htm.

20. Pankratz, A. (1983). Forecasting with univariate Box-Jenkins models: Concepts and cases. New York: Wiley.

21. Pfaff, B.: "Analysis of Integrated and cointegrated Time Series with R". Springer, September 2005.

22. Shumway, Robert H. and Stoffer, David S. (2005). The Time Series Analysis and Its Applications with R Examples, Second Edition, Springer,

23. Tang Z, De Almeida C, Fishwick P. A (1991). Time series forecasting using neural networks vs. Box- Jenkins methodology. *Simulation*, **57:** 303-310.

24. Tiao, G.C. (1985) Autoregressive moving average models, intervention problems and outlier detection in time series. In: Handbook of Statistics, Vol. 5, E.J. Hannan, P.R. Krishnaiah and M.M. Rao, eds, 85–118.

25. Vähäkyla P., Hakonen E. and Léman P. (1980), *International Journal of Electrical Power & Energy Systems* **2(1)**, 29-34.

26. Vandaele, W. (1983). Applied time series and Box-Jenkins models. New York: Academic Press.

27. Wang, Shoujun (2007). Time Series Analysis of Air Pollution in the City of Bakersfield, California.

28. Witt, Thomas J.(2007), Using the autocorrelation function to characterize time series of voltage measurements.