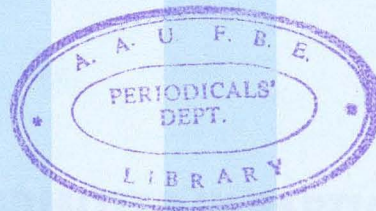
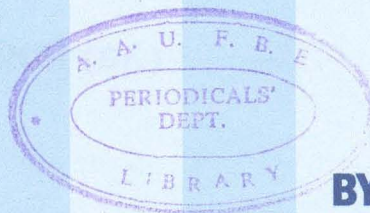


FORECASTING SALES USING THE BOX-JENKINS APPROACH: The Case of Gulele Soap Factory

24



**A PROJECT REPORT SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF
BUSINESS ADMINISTRATION**



BY

ARAYA MULUGETA MULUNEH

**ADDIS ABABA UNIVERSITY
FACULTY OF BUSINESS AND ECONOMICS
MBA PROGRAM**



FEBRUARY 2003

24

**MBA.
ARA**

FORECASTING SALES USING THE BOX-JENKINS APPROACH: The Case of Gulele Soap Factory

**A PROJECT REPORT SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF
BUSINESS ADMINISTRATION**

Advisor
GURMEET SINGH (Dr.)

Researcher
ARAYA MULUGETA

**ADDIS ABABA UNIVERSITY
FACULTY OF BUSINESS AND ECONOMICS
MBA PROGRAM**

**FEBRUARY 2003
ADDIS ABABA**

Gurmeet Singh (Dr.)

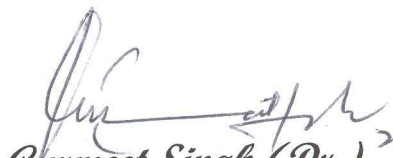
Faculty of Business and Economics

Addis Ababa University,

Ethiopia

APPROVAL

This is to approve that Ato Araya Mulugeta has completed a research project entitled "Forecasting Sales using the Box-Jenkins Approach: The Case of Gulele Soap Factory" with my advice and follow up. I also approve that his work is appropriate enough to be submitted as a partial fulfillment of the degree in Masters of Business Administration (MBA).


Gurmeet Singh (Dr.) 21/02/2003

Research Advisor

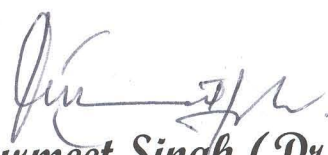
Declaration

I Araya Mulugeta Muluneh assert that this study entitled "Forecasting Sales using the Box-Jenkins Approach: The Case of Gulele Soap Factory" is my own original work that has not been presented for a degree in any other university and that all sources of material used for the study have been duly acknowledged.

Araya Mulugeta



Researcher



Gurmeet Singh (Dr.) 21/02/2003

Research Advisor

ACKNOWLEDGMENT

Nothing can compensate for the priceless contributions that have been made by several individuals throughout the preparation of this paper. In fact, a word of thanks can hardly say how much their contribution meant to the successful completion of the work.

First and for most I would like to put on record my enormous gratitude and appreciation to my advisor Dr. Gurmeet Singh whose unfailing help and tireless devotion to the work has been by far above and beyond a call of duty.

I would like to extend my indebtedness to my friend Abel Tilahun for his unreserved and intensive support, both in terms of material and constructive idea, throughout the lifetime of the study.

I cannot close, however, without forwarding my appreciation and words of thanks to the management of Gulele Soap Factory for their unlimited cooperation and positive sprit without which the objective of the study could have been a dream.

TABLE OF CONTENTS

TITILE	PAGE
List of Tables and Figures	i
Abstract.....	..ii

CHAPTER ONE

Introduction	1
1.1 Background.....	1
1.2 Statement of the Problem	4
1.3 Objective of the Research.....	5
1.4 Importance of the Research	5
1.5 Delimitation of the Research.....	6
1.6 Data and Methodology.....	8

CHAPTER TWO

Review of Literatures	10
2.1 Conceptual Framework.....	10
2.2 Survey of previous works in sales forecasting.....	14

CHAPTER THREE

Data and Methodology	21
3.1 Data.....	21
3.2 Methodology	22
3.2.1 Definition of some important terms	22
3.2.2 Autoregressive and Moving average processes.....	23

3.2.2.1 Autoregressive models.....	23
3.2.2.2 Moving average models	24
3.2.2.3 Mixed Autoregressive-Moving average models.....	25
3.2.3 Stages of the model building process	27
3.3 Forecasting.....	34

CHAPTER FOUR

Data Analysis and Discussion of Results.....36

4.1 Test of Randomness	36
4.2 Model Identification	40
4.3 Model Estimation	47
4.4 Model Diagnostic Checking.....	48
4.5 Interpretation of the final Model.....	52

CHAPTER FIVE

Forecasting.....61

CHAPTER SIX

Summary, Conclusion and Recommendation.....67

6.1 Summary	67
6.2 Conclusion	71
6.3 Recommendation.....	72

List of Figures and Tables

Title	Page
Fig 1. Steps in model building. (Source: Box and Jenkins (1976:19))	26
Fig 2. Plot of Monthly Sales for Laundry Soap.	38
Fig 3. Plot of the Autocorrelation Function for the indifference series.	39
Fig 4. Autocorrelation Function (ACF) for the first-degree Differenced Series	42
Fig 5. Partial Autocorrelation Function (PACF) Plot for the first-degree	42
Fig 6. Autocorrelation and Partial Autocorrelation plots of the residuals	48
Fig 7. P-Values of the Chi-Squared Statistics in the Portmanteau test.....	49
Table 1. Autocorrelation Values for the first 20 lags.....	53
Table 2. Forecasted monthly Sales of Laundry Soap measured in terms of Cartoons.	61
Table 3. Updated Forecasted Monthly Sales of Laundry Soap	65

Abstract

This study is conducted to facilitate the decision making process of Gulele Soap Factory by developing an appropriate forecasting model. It aims at providing accurate sales forecasts for future sales, which is a vital input in decision-making. Particularly, the focus is given for forecasting monthly laundry soaps sold by the factory.

Towards achieving its objective the study considered the Box-Jenkins approach to time series analysis. A total of 68 monthly sales data has been taken for analysis or model building purpose. Moreover, additional 5 months sales data has been used for forecasting purpose.

The analysis of the data, which is carried out using S-Plus 2000 package, suggested that ARIMA (3,1,0) model represents the pattern of monthly laundry soap sales data. According to this model, forecasting current sales essentially requires the inclusion or consideration of the previous four consecutive sales data occurring at the four successive lags. Moreover, it is found that the sales data recorded in the first lag has greater influence or contribution in forecasting current sales volume.

On the other hand, it is observed that the sales data involves a seasonal component that turns out to affect the sales volume approximately in 3.5 months. In other words, the analysis indicated that there is a seasonal component that occurs with 3 or 4 months periodicity. This in turn resulted in attaching high importance to the third and fourth lag coefficients as compared to the coefficient of the third lag.

The results obtained led to the conclusion that the time factor is the major but not the only relevant factor in forecasting sales. Other considerations in relation to promotional activities, competitors action, seasonal factors, etc should be kept in mind.

Chapter One

Introduction

1.1 Background

In today's highly competitive business environment, where there are a number of competitors, executives (managers) are expected to make effective and appropriate decisions about future conditions more than ever before. This is basically because; a wrong decision taken by an organization will give room for its competitors to win.

Among the different conditions of making decisions the one made under the condition of uncertainty is much difficult since it requires forecasting the most probable event likely to occur in the future. Forecasting may be considered as both an art and science concerning primarily on predicting future events based on past (historical) data or through human judgment. Of course, one of the difficulties associated to forecasting is that the future is not known certainly. That is, managers make decisions to do with the future in the present with little to help except past experience. [Lancaster and Reynolds1999: 85-86] Therefore, in making decisions managers need clues as to what event likely to happen in the future. These clues can be obtained through forecasting. Good forecasts are critical importance in all aspects of business. Forecasting forms an

essential part of the planning process. That is, effective business planning in both the short and long run demands sales forecasting. Moreover forecasting provides guidelines and indicators of future trends. [Parsons 1987:66]

Coming to a specific case, marketing (sales) managers need sales forecasting data in order to plan and make wise decisions regarding future conditions. The level of future sales forms the basis for many other resource allocation decisions that managers are called upon to make. Sales forecasts drive a company's production, capacity and scheduling systems and serve as inputs to financial, marketing and personnel planning. [Heizer and Render 1999:143]

Despite the critical role of sales forecasting in effective business planning, its practice seems to be ignored specially in business organizations acting in developing countries. As stated by Kracmar [1971:5-6], marketing and marketing research have been the most underdeveloped part of the economy in developing countries. The case is even worse when it comes to the availability and growth of researches in forecasting in these countries. Moreover, among the few globally available marketing researches referring to forecasting most of them confined to the common regression and correlation analysis even in situation where there is serial or auto correlation, which violates one of the assumptions of these econometric methods. In other words the usual forecasting

techniques like regression and correlation analysis ignore the dynamics or memory of the process involved and treat the data, which are mostly in time series, as being independent. [Kappor and others 1981:94]

In Ethiopia as well it is hardly possible to get business organizations that make use of quantitative forecasting techniques so as to predict the demand for their products in the future. In other words, the different quantitative methods of forecasting sales does not seem applied in most of the business organizations. Most business managers in Ethiopia do not consider forecasting as a systematic approach that needs its own input and analysis so as to arrive at more accurate forecast values.

Gulele Soap Factory, which is the company under consideration, was established in 1969 G.C with a total capital of 1,200,000 Birr. It is found in Addis Ababa located approximately at 9 kms from the center of the city. For a number of years the factory took the lion's share of the market. Till 1972 the factory produced only one product, which is soap. Currently, the factory involves four product lines namely, Laundry soaps, Sodium Silicate, Sodium Alkaline, and Sodium Neutral. Starting from the outset the government has been the owner of the factory.

1.2 Statement of the Problem

Obviously future sales are not precisely known and this induces a problem in making decisions and setting future plans. One possible solution to reduce the severity of the problem is to make accurate sales forecasts. However, obtaining sales forecasts for the selection of proper marketing strategy by itself is becoming a problem of growing concern than ever before in the history of business at large and in the field of marketing in specific. This is mainly due to the ever-growing dynamism of the business environment. Rapid change is becoming the unique nature of today's business environment.

Depending on the prevailing conditions and nature of the data at hand one forecasting approach may outperform the other. Therefore, business organizations need to identify the most appropriate forecasting technique that best fits the sales data under consideration.

In Gulele soap factory, there is no formal way or means of sales forecasting developed. This results in the absence of satisfactorily accurate sales forecasts that are invaluable in devising marketing strategy. Consequently, the decision making process of the management is affected to a marked extent. The management usually encounters problems in making decisions about future actions in connection with raw material requirement, operations financing, human resource, etc.

1.3 Objective of the Research

The core objective of this research is to develop (fit) an appropriate time series model that can be used for forecasting monthly sales of laundry soaps produced by Gulele soap factory. More specifically, the model development process involves model identification, parameter estimation and model diagnostic.

Apart from its core objective, the paper addresses the question that to what an extent univariate time series models are powerful tools (approaches) towards forecasting sales.

1.4 Importance of the Research

The overall significance of the research can be traced into several issues. Some of them are listed below.

- It provides the management of Gulele soap factory with a more reliable and appropriate forecasts that obviously improve, or at least facilitates, the decision making task.
- It serves as a model that greatly helps other similar business organizations in Ethiopia in developing their own forecasting models through time series analysis. In other words the paper

clearly illustrates how time series analysis can be applied in sales forecasting provided all the required assumptions are satisfied.

- It serves as a reference material for future researches, that is, it adds something to the available literatures especially in developing countries like Ethiopia.

1.5 Delimitation of the Research

The scope of the research is limited in terms of two aspects. First, it is only the time factor that is considered as an influential factor that determines future sales. As a result, the study is limited to be univariate in its nature that involves only sales data at different points in time (monthly sales data). In other words, although there are other factors like advertising and other promotional activities, price of other substitute products, volume of soap imports, etc that are believed to affect sales, in this paper any of these factors is not taken into consideration. This is due to lack of required data on the stated variables and of course, some of the factors are uncontrollable. Another reason is that, it is assumed that future sales can be adequately forecasted depending only on historical data of sales (time factor). This assumption is viable because in the short run most of the stated factors are likely to be stable. Furthermore, as reported by Leon [1983:294-5], it has always been

accepted that when the goal is solely forecasting, the univariate time series approach outperforms the econometric models in most cases and it is also more parsimonious.

Second, the scope of the research is also restricted with regard to the type of product taken into consideration. Gulele soap factory basically produces (sales) four products namely, Laundry soaps, Sodium Silicate, Sodium Alkaline, and Sodium Neutral. In this paper, however, it is only sales of laundry soap that is taken into consideration. Of course, a number of literatures stipulate that whenever there exist more than one product sold by an organization each one of them should be treated separately. Consideration of aggregate sales often is difficult for interpretation and also provides less deep information.

1.6 Data and Methodology

✓ *Data*

The research is to be based on secondary sales data collected from Gulele soap factory internal monthly reports. Laundry soap sales data are recorded for 73 consecutive months ranging from January 1997 to January 2003. The first 68 observations will be included in the analysis while the remaining 5 observations are reserved for forecasting purpose.

✓ *Methodology*

A great deal of researches done in the area of sales forecasting use the famous Box-Jenkins model approach, named after the inventors George E. P. Box and Gwilym M. Jenkins. In this research as well this powerful time series modeling approach will be used for the purpose of forecasting sales. A number of literatures stated that the Box-Jenkins approach is suggested as an effective tool for short term forecasting. Moreover, it is reported that the Box-Jenkins approach outperforms all other models investigated and that it should be considered seriously as a powerful forecasting technique. [Naylor (1972), Nelson (1970), Geurts and Ibrahim (1975), Newbold and Granger (1973)]

As pointed out by Geurts and Ibrahim [1975:183] two main characteristics distinguish Box-Jenkins method from other forecasting

techniques. The first is that, unlike econometric methods model is fitted based on the nature of available data. That is it is not data that is fitted to a predetermined theoretical model, rather one searches for the appropriate model that best describes the behavior of the data set. The second is that a systematic process is utilized to build the model, which includes identification, estimation, and diagnostic checking.

Chapter Two

Review of Literatures

2.1 Conceptual Framework

The very nature of managerial decision-making involves forecasting future conditions. Kotler, as quoted by Lancaster and Reynolds [1998], stated that forecasting is the key to success in business environment. Moreover, Murdick and Schaefer [1961: 6] suggested *"there are few, if any, functions of business which are fundamental and far-reaching in their effect on the health of the enterprise than forecasting or prediction of future sales. ... It is the basis for financial planning, production setting, inventory levels directing sales promotion, expanding or contracting production and allied facilities, calculating profit or loss, and many other phases of business activity."*

Sales forecasting is an attempt of predicting future sales volume for the sake of facilitating decision-making. Sales forecast is the primary factor in estimating the future activities of an organization. [Green and Tull 1970: 1] Its purpose is to estimate the net effect on demand of all internal and external factors in an organizational setup. The accuracy of the forecasts depends on the quality of data used and the

appropriateness of the methodology employed. [Keay 1972:42] The unstable the demand for a product the more critical is forecast accuracy. Poor forecasts can lead to overly large inventories and other associated substantial costs. [Kotler 1967: 239]

Broadly speaking forecasts can be identified, according to time horizons, as Short-range, Medium-range, and Long-range. Moreover, forecasts may also be classified as Economic, Technological and Demand forecasts based on the area of their application. [Heizer and Render 1999:142-3]

There are different approaches to forecasting sales. These are categorized in to two major groups, namely qualitative and quantitative approaches.

✓ **Sales forecasting through qualitative approach**

Also called subjective approach, qualitative forecasting approaches are characterized by personal intuition, emotions, and experience. [Heizer and Render 1999:145] Jury of executive opinion, sales force composite, Delphi method, and consumer market survey are the different options that are available in the qualitative forecasting approach.

✓ **Sales forecasting through quantitative approach**

Time series and econometric approaches constitute the overall quantitative approach to forecasting. In both time series and econometric approaches one uses mathematical models to forecast future sales.

Nevertheless, there are two significant differences between these two quantitative approaches. First, a time series analyst admits allowing the data to specify a relationship between variables by trying a variety of alternative models before selecting the best model. That is, in time series analysis an appropriate model is searched that suits the data on hand. An econometrician, on the other hand, tries to fit the available data to a predetermined model suggested from some theoretical perspective.

Second, the two approaches are also different in the treatment of the residuals. In econometric methods error terms are considered to be results of missing variables or model misspecification and usually assumed uncontrolled. In time series approaches, however, attempts are made to control the residuals by including them as vital part of the model building process, which is not the case in econometric methods. It is also in this aspect the time series gains advantage over econometric methods. [Leone 1983:291-5]

Time series methods are based on the assumption that the past is a good indicator of the future and that future demand is simply a function of time. The whole analysis uses the historical series of only one variable to develop a model for predicting future values. [Lancaster and Reynolds 1999:103] Involved in time series approach are naive, moving averages, exponential smoothing, trend projection, and Box-Jenkins approaches. A drawback of time series analysis is that it takes only the time factor as

influential variable affecting sales data. It ignores the significant effect of other variables like sales promotion and price.

In casual or econometric methods sales forecasts are obtained from the analysis of other variable that are believed to be influential. In other words, a model is built that expresses sales as a function of other variables like advertising expenses, price of the product and substitutes, sales incentives and other relevant auxiliary variables. Here the variable "sales" is considered as the dependent variable. The problem with such an approach is that it ignores the dynamics or memory of the process involved and treats the data, which are in the time series, as being independent. [Kapoor and others 1981: 93] Moreover, it is not possible to exhaust all the variables affecting sales in our model, as some of them are immeasurable.

On top of all, it has to be recognized that there is no single best way of making forecasting. The selection of the approach entirely depends on the situation under consideration. In some cases one may be required to consider some combination of the above approaches in order to get more accurate results. Specifically, many researches are emerging that combine the concept of time series and econometric approaches.

2.2 Survey of previous works in sales forecasting

The increase in the interest of sales forecasting is evident from the marketing literatures. Many writers have forwarded their invaluable approaches and the resulting findings in the field of sales forecasting. Honored researchers have suggested different methods or techniques of sales forecasting. Moreover, most of the marketing researches made in the area of sales forecasting have one thing in common; almost all of them use some kind of time-series or econometric analysis approach as their basic methodology. In the subsequent paragraphs of this section review of some of the most common works in forecasting is made. The review includes data used, methodology applied, results obtained and suggestions made in the different publications referring to sales forecasting.

Geurts and Ibrahim [1975] conducted a research to compare the performance of Box-Jenkins approach in forecasting Hawaii's tourists with the exponential smoothing forecasting technique developed by Brown. Box-Jenkins time series analysis, as stated by Geurt and Ibrahim, is suggested as an effective tool for short-term forecasting. For the research purpose, data used was the monthly number of tourists visiting Hawaii for the period 1952-1971 inclusive. Moreover, both the Box-Jenkins and exponential smoothing time series analyses were considered as methodological tools for forecasting the monthly number of

tourists that visit Hawaii. In the analysis process the original data was logarithmically transformed because it was observed that these seasonal data fluctuated in a percentage manner. The analysis and forecasting process indicated that the exponentially smoothed model patterned on Brown's model and the Box-Jenkins approach seem to have equal performance. [Geurts and Ibrahim 1975:182-187]

Helmer and Johansson [May 1977:227-239] stated that experimentation and time series data analysis of past sales data are the two alternative paths towards assessing marketing effectiveness of a business firm. They have presented a fairly detailed analysis of Box-Jenkins transfer function applied to the well-known annual advertising-sales data of Lydia Pinkham vegetable compound for the period 1907 to 1946. The purpose of the study was to see the comparative forecasting efficiency of the Box-Jenkins transfer function models as compared to other different models including Carrying effect model by Houston, Weiss and the five models of cumulative effects that Palda analyzed, and the Box-Jenkins univariate analysis models. The data analysis showed that all the Box-Jenkins models considered, including the univariate ones, forecast better than Palda's five econometric models displaying cumulative effect. Furthermore, Helmer and Johansson concluded that advertising does have an important impact on sales.

Moriarty and Adams [1979: 221-32] through their research contributed to the development of time-series analysis applied to sales forecasting. The objective of the study was to see whether the same model form is applicable for forecasting sales in cases where there is more than one sales territory. In addition, they have tested whether all territories having a common underlying process also have common parameter values. The authors illustrated the application of the Box-Jenkins approach for forecasting sales where there are different sales territories. They developed distinct Box-Jenkins forecasting models for ten sales territories and compare them with the forecasting performance (accuracy) of a single first order moving average model. To that end results suggested that the territory Box-Jenkins models and the first order moving average model have no significant forecasting differences. The authors also concluded that the moving average model is a reasonable representation of a parsimonious form for all territories. In other words, one can employ the same forecasting model for all territories available without losing significant forecasting accuracy.

Leone [1979], who is one of the most commonly sighted author in comparing econometric and time series approaches to sales forecasting, declared that when the goal of a given marketing analysis is mainly of forecasting then it is better to use time series approach than the econometric counter parts. According to him, when the goal is

forecasting the time-series approach outperforms the econometric model in most cases.

Moriarty and Salamon [1980] through their study attempted to demonstrate the improvement in forecast accuracy that can result from a joint analysis of several individual series. They applied a special case of the general multivariate autoregressive moving average (ARIMA) model to the unit sales data of an anonymous firm. They named this special model as SURAMA (Seemingly Unrelated ARMA). The basic objective was to improve the accuracy of the univariate Box-Jenkins procedure to forecasting sales. Based on their results obtained they concluded that the proposed model seems to be economically justified in most situations in which all the required assumptions are satisfied. They observed that the sales forecasts are substantially improved using the SURAMA model, which requires pre-implementation of univariate Box-Jenkins model. It has to be noted, however, that the SURAMA is not a substitute for the Box- Jenkins approach.

Kappor et al [1981] came up with a different time-series analysis of a non-stationary sales data. They introduced a step-by-step approach for forecasting sales that is an alternative method to Box-Jenkins approach. Their overall attempt was to dichotomize a given non-stationary time series into its deterministic and stochastic components. Meaning a deterministic function of time trend in conjunction with a stochastic

model is used to fit the series in which the seasonality is inhomogeneous and of deterministic nature. The initial values of the parameters of the deterministic function are obtained by the method of least squares while the parameters of the stochastic component are estimated through the Box-Jenkins approach. The authors suggested that time series analysis is widely applicable in marketing research area with its advantage to predict and explain both short term and long term effects.

Umashankar and Ledolter [1983] presented the diagonal time series models approach for forecasting sales. This approach is viewed as an intermediate step between the univariate and multiple time series analyses. The analysis performed on set of five competitive sales series each following different ARIMA models revealed that the diagonal time series analysis is preferred as a sales forecasting tool provided correlations of the residuals in the univariate ARIMA models are large at lag zero and small at all other lags.

In his multivariate analysis of monthly totals of aggregate sales and advertising of 10 brands from 1952 to 1966, Leone [1983] verified that time series approach for forecasting short term of sales outperforms the econometric model in most cases. The author also testified that the advertising of a particular brand tends to create a selective demand for that brand, as well as a primary demand for the category of which that brand is a member. [Leone 1983:291-295]

Smith et al [1994], on their part, emphasized that one of the challenges facing managers in existing markets is that they must make numerous decisions regarding price, advertising, and product selection on the basis of noisy sales data obtained in a shifting market environment. They clearly stipulated that a manager's decisions regarding buying merchandise and planning inventory levels require long term sales forecasts. Moreover, the existence of mechanisms for updating forecasts is identified to be essential. Accordingly, the authors use a two stage sales forecasting approach wherein the first stage is the standard regression analysis to develop an econometric model for sales response. In the second stage they applied the discounted least square technique that simultaneously smooth certain key parameters to obtain updated forecasts. Their approach was illustrated using over 1100 weekly sales observations for two seasons of weekly sales. The result obtained confirmed that updating forecasting models significantly improve forecasts' accuracy.

To sum up what has been obtained from literatures, sales forecasting is identified to be the critical input for making decision in the different aspects of a business organization. Different authors have suggested different methods of sales forecasting. Particularly, most of the literatures seems to indicate that time series analysis is most appropriate for forecasting sales data. With this regard Guerts and Ibrahim [1975] performed univariate Box-Jenkins analysis to forecast the monthly

number of tourists visiting Hawaii. Helmer and Johansson [1977] showed how a bivariate transfer function analysis might be applied in forecasting sales. Moriarty and Adams [1979] performed cross-sectional sales behavior analyses and came up with the result that showed similarity of forecasting models over the different sales territories considered. On the other hand Leone [1979] indicted that short-term sales forecasts obtained from a time series analysis are more accurate and appropriate than those that could be obtained through the econometric approach. Moriarty and Salamon [1980] introduced the SURARMA model as a means of improving the accuracy of sales forecasts obtained through the Box-Jenkins approach. Smith and his colleagues [1983] considered a two stage forecasting approach for showing updating some key parameters of a forecasting model lead to more precise forecast values.

Chapter Three

Data and Methodology

3.1 Data

The research is to be based on secondary sales data collected from Gulele soap factory internal monthly reports. Laundry soap sales data are recorded for 73 consecutive months ranging from January 1997 to January 2003.

The first 68 observations will be included in the analysis while the remaining 5 observations are reserved for forecasting purpose. The size of the data, which is 68, is believed to be sufficient, as the Box-Jenkins approach requires it to be at least 50.

The data is tested for randomness using the turning point test and found to be non-random (time dependent) which is the basic requirement of time series analysis. Moreover, the data is identified to fulfill the normality and constant variance assumptions.

Therefore, the data is justified to be suitable for time series analysis both in its size and meeting the assumptions of time series analysis. The full set of the data is presented in the annex.

3.2 Methodology

3.2.1 Definition of some important terms

Time series: A set of observations generated sequentially in time.

Deterministic time series models: These are models wherein future values of a time series are exactly determined by some mathematical function.

Stochastic process: A statistical phenomenon that evolves in time according to probabilistic laws.

Stationary time series: A stochastic process that remains on equilibrium about a constant mean level. A strictly stationary process properties are unaffected by a change of time origin.

Backward shift operator (B): Let X_t be an observed variate value at time t , then BX_t is given by;

$$BX_t = X_{t-1}$$

More generally, the n^{th} order backward shift operator is given by:

$$B^n X_t = X_{t-n}$$

Backward difference operator (∇): Suppose X_t and X_{t-1} are two consecutive observed values. Thus, the backward difference operator applied to X_t is defined as follows.

$$\nabla X_t = X_t - X_{t-1}$$

It can be shown that

$$\nabla X_t = (1 - B) X_t \Leftrightarrow \nabla = 1 - B$$

3.2.2 Autoregressive and Moving average processes

3.2.2.1 Autoregressive models

In an autoregressive model, the current value of the process is expected as a finite, linear aggregate of previous values of the process and a shock e_t .

That is, let $X_t, X_{t-1}, X_{t-2}, \dots, X_{t-p}$ denote the observed values corresponding to equally spaced points $t, t-1, t-2, \dots, t-p$ in time. Moreover, let μ be the mean level of the process (assuming the time series is stationary). Thus,

$$X_t - \mu = \phi_1 (X_{t-1} - \mu) + \phi_2 (X_{t-2} - \mu) + \dots + \phi_p (X_{t-p} - \mu) + e_t$$

let us denote $X_{t-i} - \mu$ by $\tilde{X}_{t-i}$; $i=0,1,2,\dots, p$

$$\tilde{X}_t = \phi_1 (\tilde{X}_{t-1}) + \phi_2 (\tilde{X}_{t-2}) + \dots + \phi_p (\tilde{X}_{t-p}) + e_t$$

$$\tilde{X}_t = \sum_{i=1}^p \phi_i \tilde{X}_{t-i} + e_t$$

Where e_t is the shock or error term distributed according to normal with mean zero and constant variance. It is also called white noise process.

The above presentation is called an autoregressive process of order p , i.e., AR (p). In its more compact appearance the above expression may be written as

$$\phi(B)\tilde{X}_t = e_t$$

$$\text{where } \phi(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - \dots - \phi_p B^p$$

3.2.2.2 Moving average models

A moving average model is a linear model that expresses the deviation $X_t - \mu$ as a linear combination of the previous shocks (errors).

Symbolically,

$$\begin{aligned}\tilde{X}_t &= e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \\ &= e_t - \sum_{i=1}^q \theta_i e_{t-i}\end{aligned}$$

The expression given above is called moving average of order q or MA(q). One can write this expression in a more compact manner using the moving average operator $\theta(B)$.

$$\tilde{X}_t = \theta(B) e_t$$

$$\text{where } \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

3.2.2.3 Mixed Autoregressive-Moving average models

Sometimes a time series data may exhibit the feature of both autoregressive and moving average processes. In such a case we employ a model called mixed autoregressive-moving average, which is given by:

$$\phi(B)\tilde{X}_t = \theta(B)e_t$$

This equation is referred to as mixed autoregressive moving average model of order p and q, ARMA (p,q).

Box and Jenkins suggested that the adequate representation of actually occurring stationary time series can be obtained with autoregressive, moving average, or mixed models, in which p and q are not greater than 2, often less than 2.

Nevertheless, it is not always possible to obtain a stationary time series with constant mean level. As stated by Box and Jenkins many time series referring to business and economics are non-stationary. Such series may nevertheless exhibit homogenous behavior of a kind.

Homogenous non-stationary behavior can be represented by a model that calls for the d^{th} differencing of a process to be stationary. In practice, d is usually 0,1 or at most 2. It can be noted that the whole purpose of differencing is to convert a non-stationary time series to that of stationary. A general model, which can represent homogenous non-stationary behavior, is the difference equation form given by:

$$\phi(B)(1-B)^d X_t = \theta(B)e_t$$

where d is the level of differencing.

The above equation is called an autoregressive integrated moving average (ARIMA) process of order (p,d,q) or $ARIMA(p,d,q)$.

3.2.3 Stages of the model building process

The overall Box-Jenkins approach to model building consists of three well-defined stages, namely

- ✓ Model identification
- ✓ Parameter estimation
- ✓ Model diagnostic checking

The following flow chart best represents the model building process.

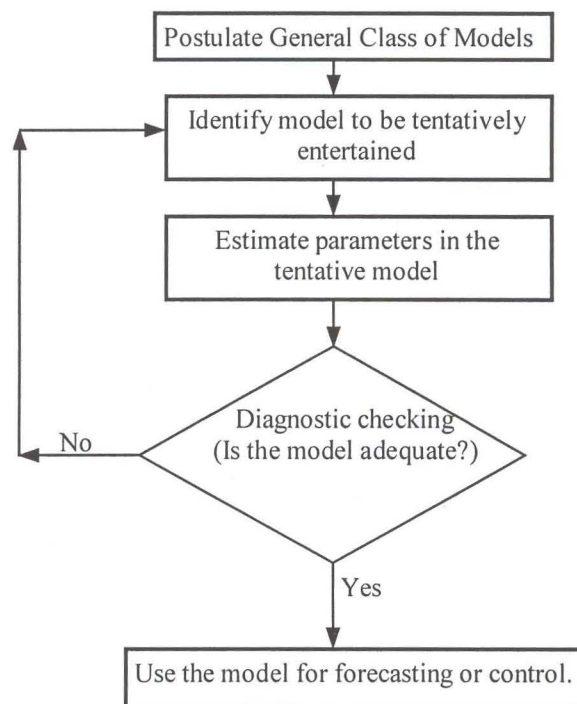


Fig 1. Steps in model building. (Source: Box and Jenkins (1976:19))

i. Model Identification

The whole objective in model identification is to obtain some idea of the values of p, d , and q needed in the general ARIMA(p, d, q) model and to obtain initial guesses for the parameters.

The identification of the values of p, d , and q is usually carried out using the plots of autocorrelation and partial autocorrelation functions.

Autocorrelation function

Clearly, the covariance between X_t and X_{t+k} , separated by k intervals of time, is given by

$$\text{Cov}[X_t, X_{t+k}] = E((X_t - \mu)(X_{t+k} - \mu)) = E(\tilde{X}_t \tilde{X}_{t+k})$$

The above covariance is usually referred to as the auto covariance at lag k and denoted by γ_k . Moreover, it is known that for any two variables X and Y , the correlation between them is given by:

$$\rho_{x,y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E((X - \bar{X})(Y - \bar{Y}))}{\sqrt{E(X - \bar{X})^2 E(Y - \bar{Y})^2}}$$

In time series analysis, we substitute X by X_t and Y by X_{t+k} and call the resulting expression the autocorrelation of X at lag k .

$$\rho_k = \frac{E((X_t - \mu)(X_{t+k} - \mu))}{\sqrt{E(X_t - \mu)^2 E(X_{t+k} - \mu)^2}}$$

Note that the word auto in autocorrelation refers to the fact that we are trying to find the correlation between any two observed values of a single variable separated by k periods. In short, X is correlated to itself at two

different time points. For a stationary process, the autocorrelation at lag k is

$$\rho_k = \frac{\gamma_k}{\gamma_0}, \quad \text{where } \gamma_0 = \sigma_x^2 = E(X_t - \mu)^2 = \text{Var}(X)$$

The plot of the autocorrelation coefficient as a function of the lag k is called the autocorrelation function of the process. Due to the fact that the autocorrelation function is necessarily symmetric about zero (as $\rho_k = \rho_{-k}$) in practice we plot only the positive half of the function. Sometimes, the autocorrelation function is also called correlogram.

Partial autocorrelation function

As discussed by Box and Jenkins, partial autocorrelation function is a device that exploits the fact that whereas an $\text{AR}(p)$ process has an autocorrelation function which is infinite in extent, it can by itself be described in terms of p non-zero functions of the autocorrelations.

For an $\text{AR}(p)$, the partial autocorrelation function will be non-zero for k less than or equal to p and zero for k greater than p . That is, it has a cutoff after lag p . The use of the autocorrelation and partial autocorrelation functions for identifying an appropriate model involves two stages.

Stage 1. Identifying the degree of differencing (d)

It is failure of the estimated autocorrelation function to die out rapidly that suggests non-stationary. In order to achieve stationary a given non-stationary time series must be differenced d times. The value of d is reached, when the autocorrelation function dies out fairly quickly. The order of differencing should be the minimum possible number for which the plot of the autocorrelation function dies out fairly quickly. That means, over differencing is not appropriate. Therefore, by observing the pattern of the autocorrelation function we identify the degree of differencing d . In practice, as Box and Jenkins indicated, d is normally either 0, 1, or 2 and it is usually sufficient to inspect the first 20 or so estimated autocorrelations of the original series and of its first and second differences.

Stage 2. Identification of resulting stationary ARMA process

The aim of this stage of the identification process is to provide tentative values of p and q based on the tentatively determined value of d and the appearance of the autocorrelation and partial autocorrelation functions of the appropriately differenced series.

To that end, whereas the autocorrelation function of an autoregressive process of order p tails off, its partial autocorrelation has a cutoff after lag p . Conversely, the autocorrelation function of a moving average

process of order q has cutoff after lag q , while its partial autocorrelation tails off. Finally, if both the autocorrelation and partial autocorrelation tail off, a mixed process, containing a p^{th} order autoregressive and a q^{th} order moving average components. 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.

It can be clearly observed that the behavior of autoregressive process as measured by the autocorrelation function is the same as the behavior of the moving average process as measured by the partial autocorrelation function, and vice versa.

ii. Model Estimation

After identifying the type of the model, the next step is to estimate the parameters of the model. Nowadays this model estimation process is completely performed using statistical packages like S-Plus.

iii. Model Diagnostic Checking

In model diagnostic checking stage the adequacy of the tentative model is evaluated as a representative of the data on hand. If diagnostic checks, which have been thoughtfully devised, are applied to a model and fail to show serious discrepancies then we shall rightly feel more comfortable about using that model.

Box and Jenkins presented four methods for checking model adequacy. These are,

- ✓ Over fitting
- ✓ Use of residuals
- ✓ Portmanteau test
- ✓ Cumulative periodogram checks

For the purpose of this paper, however, we consider the second and third approaches, as they are more convenient to deal with. Of course, the use of residuals method is used in most of the literatures sited in the previous chapter. Moreover, Box and Jenkins concluded that there is no one best way of checking a models adequacy.

a) Diagnostic checks applied to residuals

In this approach the overall model diagnostic check is done through visual inspection of a plot of the residual autocorrelations. The correlogram of the residuals is plotted and we can then see how many

coefficients are significantly different from zero and whether any further terms are indicated for ARIMA model. [Chatfield 1996:73] The basic premise is that the autocorrelation of the residuals can yield valuable evidence concerning lack of fit and the possible nature of the model adequacy.

b) Portmanteau lack of fit test

As compared to the previous approach the portmanteau approach is somewhat objective.

Suppose that we have the first m autocorrelations $\gamma_k^2(\hat{e})$ ($k=1,2,3,\dots,m$) from any ARIMA (p,d,q) model. Then it can be shown that the expression

$$Q = n \sum_{k=1}^m \gamma_k^2(\hat{e}) \quad \text{where } n = N - d$$

is approximately distributed as $\chi^2(m-p-q)$ if the fitted model is adequate

On the other hand, the value of Q will be inflated if the fitted model is inadequate.

Putting differently, in portmanteau model diagnostic check we test the null hypothesis,

H_0 : *The model is adequate*

Versus its alternative

H_A : *The model is not adequate*

The test statistics used is

$$Q = n \sum_{k=1}^m \gamma_k^2(\hat{e}) \quad \text{where } n = N - d$$

The rejection criterion is $Q > \chi_{\alpha}^2(m - p - q)$ for a selected level of significance. That means, reject the null hypothesis if $Q > \chi_{\alpha}^2(m - p - q)$ for a selected level of significance α .

3.3 Forecasting

The task of forecasting future values, which is the basic aim of time series analysis, is usually done by minimizing the mean square error (MSE). Consequently, the resulting forecast values are called minimum mean squared error forecasts.

The minimum mean square error forecast is defined in terms of the conditional expectation

$$E(X_{t+l}) = E(X_{t+l} / X_t, X_{t-1}, \dots)$$

That is, the minimum mean squared error forecast at origin t , for lead-time l , is the conditional expectation of ... at time t . When $\tilde{X}_t(l)$ is regarded as a function of l for fixed t , it will be called the forecast function for origin t .

Accordingly, the general forecasting equation for ARIMA (p,d,0) model, which is the interest of this paper, is derived from the difference equation as:

$$\phi B(1 - B)^d X_t = e_t \quad \text{Setting } e_t \text{ to zero, we get}$$

$$\phi B(1 - B)^d X_t = 0$$

$$\text{where } \phi B = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

The specific form of the above forecasting equation depends on the value of d. For d=1, the equation will have the form

$$\hat{X}_t = (1 + \phi_1)X_{t-1} + \sum_{i=2}^p (\phi_i - \phi_{i-1})X_{t-i} - \phi_p X_{t-p-1}$$

It can also be written as

$$\nabla X_t = \sum_{i=1}^p \phi_i \nabla X_{t-i}$$

$$\Leftrightarrow \hat{X}_t = X_{t-1} + \sum_{i=1}^p \phi_i \nabla X_{t-i}$$

For example, the forecasting equation of ARIMA (3,1,0) is:

$$\hat{X}_t = (1 + \phi_1)X_{t-1} + (\phi_2 - \phi_1)X_{t-2} + (\phi_3 - \phi_2)X_{t-3} - \phi_3 X_{t-4}$$

Chapter Four

Data Analysis and Discussion of Results

4.1 Test of Randomness

As it is stated in the review of literatures part any time series analysis is necessarily based on the assumption that the data is not random. In other words it is assumed that the data is dependent on the time factor and hence data recorded or observed at time period t will depend on the previous lags namely $t-1$, $t-2$, etc.

It is, therefore, necessary to check whether the sales data on hand meets the non-randomness assumption. Of course, there are a number of techniques that can be used to perform test of randomness. For the purpose of this study the Turning Point test is preferred as it is the most commonly used and also the simplest test.

In the turning point test of randomness we test the null hypothesis

H_0 : The data is random. Versus the alternative

H_A : The data is not random

The test statistic to be used is,

$$Z = \frac{p - E(p)}{S.E(p)}$$

Where

p is the number of turning points in the plot of the data,

$E(p) = \frac{2(n-2)}{3}$ is the expected value of p , and

$S.E(p) = \sqrt{Var(p)} = \sqrt{\frac{16n-29}{90}}$ is the standard error of p .

Moreover, the distribution of the random variable p tends to normality as n , which is the number of observations approaches to infinity. That is, for large value of n ($n > 50$) one can safely assume the distribution of p to be normal.

We reject the null hypothesis if and only if absolute value of the calculated value of Z is greater than its tabulated value at a give level of significance usually taken to be 5% or 10%.

The plot of the sales data for laundry soaps produced and sold by Gulele Soap factory is displayed in the next page. From the plot it can be observed that the graph involves 36 turning points for the available 68 monthly sales. Thus for our data we get

$$E(p) = \frac{2(n-2)}{3} = \frac{2(68-2)}{3} = 44 \text{ and } S.E(p) = \sqrt{\frac{16n-29}{90}} = \sqrt{\frac{16(68)-29}{90}} = \sqrt{11.77} = 3.43$$

Accordingly the calculated Z value will be

$$Z = \frac{p - E(p)}{S.E(p)} = \frac{36 - 44}{3.43} = -2.33$$

$$\Leftrightarrow |Z| = 2.33$$

From the standardized normal distribution we refer that the tabulated values of Z for 5% and 10% levels of significant are respectively 1.96 and 1.644. Hence, since the calculated value is greater than the tabulated values at both 5% and 10% levels of significance, we reject the null hypothesis and accept the alternative, which asserts that the data is not random. Therefore, the data is identified to be time dependent and invites time series analysis.

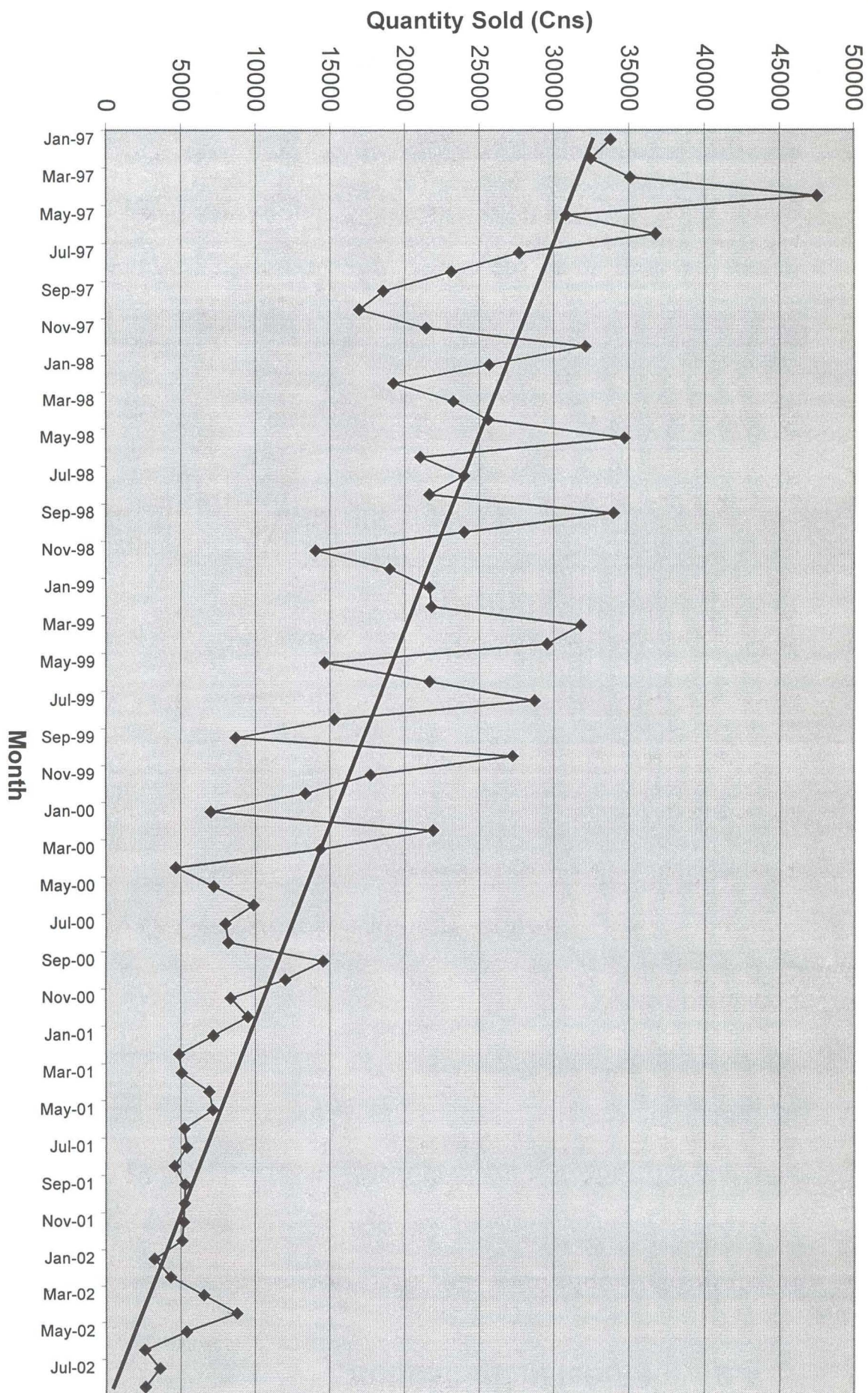


Fig 2. Plot of Monthly Sales for Laundry Soap.

4.2 Model Identification

It is already emphasized that the Box-Jenkins approach, which is the procedure that is followed in this study, requires the data to be stationary. Although it involves some sort of subjectivity, the stationarity of data is usually testified by observing the behavior of the autocorrelation plots. The plot of the autocorrelation function is presented below.

Series : monthlySales\$SalesData

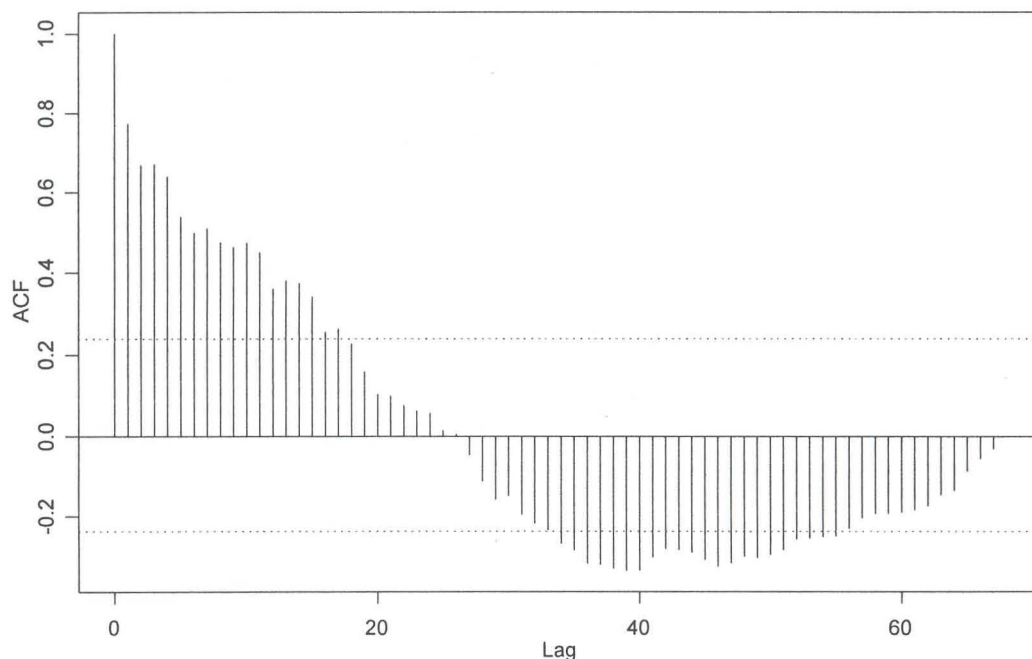


Fig 3. Plot of the Autocorrelation Function for the indifference series.

It can be seen from the above autocorrelation function plot that the graph does not die fairly quickly. Even for the lags around 17 the autocorrelation function is above the 2σ -limit lines. This shows that the

autocorrelation function assumes large values even at higher lags. It is known that one of the indicators of the non-stationary behavior of a given time series data is that the autocorrelation function does not die fairly quickly. Furthermore, as it can be clearly seen from the graph in Fig 2 the sales data has a decreasing or negative trend. Nevertheless, for stationary data there should be no trend in the data. In other words, the data should have a distribution with constant mean. Of course existence of significant increasing or decreasing trend confirms the non-stationary behavior of the data. Therefore, there is evidence to say that the data under consideration has some sort of non-stationarity. If it had been for the autocorrelation function to die quickly after attaining large values in the first few lags, then it would have been said that the data depicts stationary behavior.

If the data is identified to have non-stationary behavior, then our next task will be to apply differencing so as to convert it in to a stationary data. We recall that a stationary data may be obtained by differencing the non-stationary data to some order of differencing, d . The value of d is determined by trial and error, but it usually does not exceed 2 or 3. With regard to the data under consideration the first degree differencing ($d=1$) is found to be appropriate. This first differencing will reduce the total number of observations by one and hence we will deal with 67 observed values rather than 68. The plot of the autocorrelation function for the differenced series (look Fig 4) is observed to die out fairly quickly, which

is the required condition for stationarity. Moreover, there is a constant trend (horizontal) line in the plot of the differenced series, which is also an indication of the stationary behavior of the data.

After the data is once made stationary we proceed with the model identification process. Here it has to be noted that finding the degree of differencing is also part of the model identification process. According to the Box-Jenkins approach the remaining part of the model identification is, therefore, determining the values of p and q in the general Autoregressive Integrated Moving average (ARIMA (p,d,q)) model.

In order to determine the respective tentative or likely values of p and q we have to use the autocorrelation and partial autocorrelation functions. To that end the sketches of Autocorrelation and Partial Autocorrelation functions, obtained using S-Plus 2000 package, are presented in Fig 4 and Fig 5 respectively. We present the plots for all the available lags, though Box and Jenkins suggested that about 20 lags are sufficient for the identification purpose.

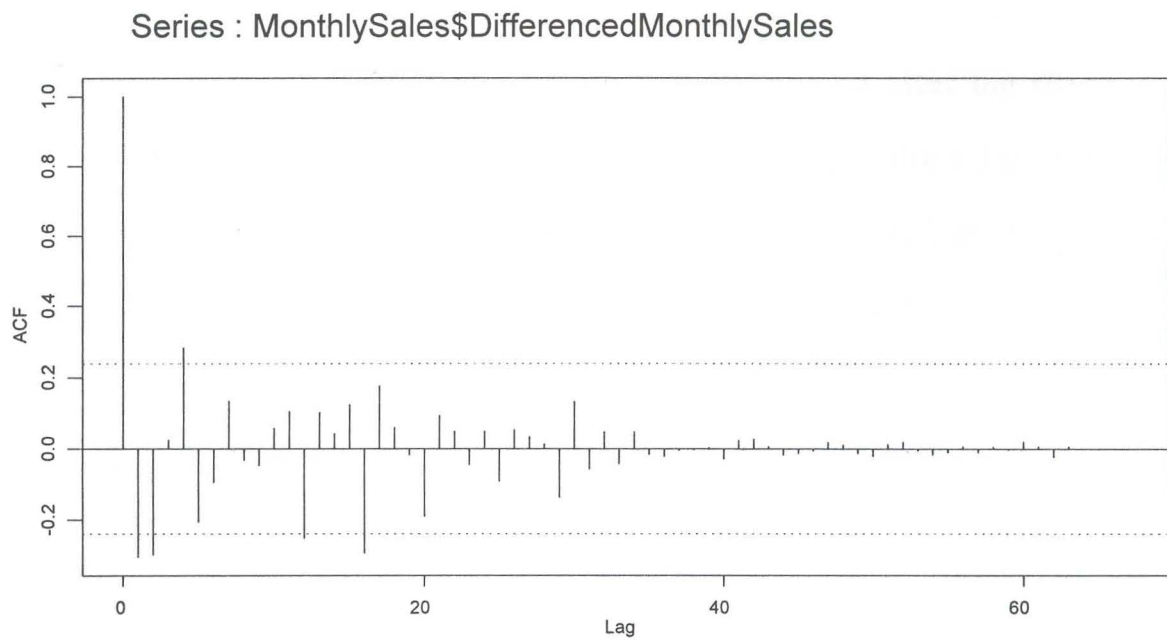


Fig 4. Autocorrelation Function (ACF) for the first-degree Differenced Series

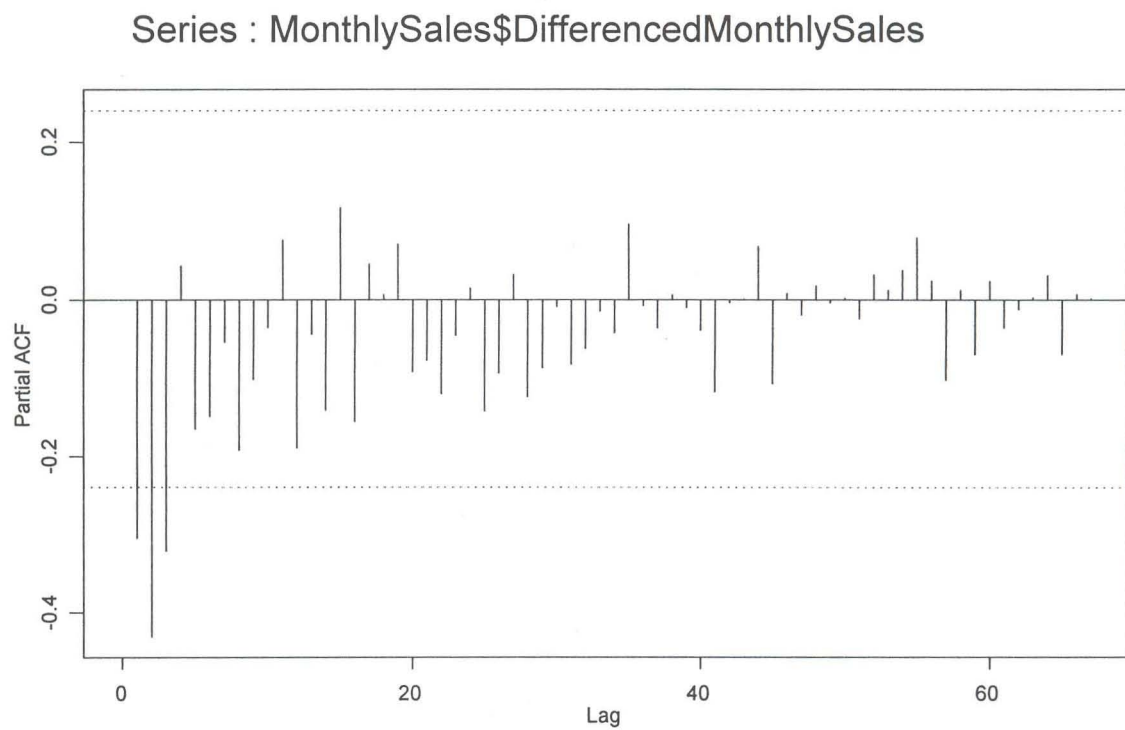


Fig 5. Partial Autocorrelation Function (PACF) Plot for the first-degree

The partial autocorrelation plot in Fig 5 has a cut off after lag three. In other words all the partial autocorrelation function values for the lags exceeding three are not significant as all of them fall within the control limit. Conversely, for all lags below the third the value of each of the partial autocorrelation functions is significant as it is out of the control limit. Moreover, it is known that whenever the partial autocorrelation cuts off after some lag p the implication is that the data under consideration pertains to the Autoregressive process of order p . Hence, the partial autocorrelation plot clearly reveals that the differenced sales data under consideration follows the autoregressive process of order three ($p=3$). Nevertheless, this identified value of p is only temporary or starting value. Its final value will be determined after performing the model diagnostic check.

On the other side, it is natural for the autocorrelation function to tail off given that the partial autocorrelation function showed a cut off after some lag. With this regard the plot displayed in Fig 4 seems to conform the expected behavior of the autocorrelation function. The autocorrelation function tails off rather than cutting off. In fact, the tailing off behavior is not clearly depicted in the plot. It is, however, possible to rely on the partial autocorrelation function analysis and select tentative model(s).

After all the model identification process based on the analysis of autocorrelation and partial autocorrelation functions is highly subjective, but then as far as we are determining the nature of the tentative model it will have no influence on the final results of the model building process provided that the model diagnostic process is to be carried out carefully.

All in all, the analysis of the plots of the autocorrelation and partial autocorrelation functions supports that the most likely model that best represents the sales data on hand is **ARIMA (3,1,0)**. Meaning the differenced data is proposed to behave according to the autoregressive process. In addition, the order of the autoregressive process is identified to be 3. This indicates that the value of an observed sale data at a given time depends on the sales recorded in the previous three consecutive differenced lagged values. That is, in order to forecast the next month differenced sale value we need to take the values of the three most recent differenced monthly sales data in to account. Any sale is assumed to be a function of the previous four consecutive sales observations. Moreover, the value 1 in ARIMA (3,1,0) stands to indicate that the data, as it is, cannot be taken to be stationary. However, it will become stationary after applying first order differencing, which results in reducing the total number of observations by one.

It has to be emphasized, however, that it is possible to consider more than one temporary or candidate model. It is just as a matter of

incidence that we obtained one single tentative models. Other alternative models could not be observed from the autocorrelation and partial correlation functions analysis. That is, the analysis does not suggest any other alternative model other than ARIMA (3,1,0).

4.3 Model Estimation

The purpose in the model estimation process is to determine the parameters or coefficients of the model identified in the previous section namely, ARIMA (3,1,0). In fact the model estimation process is very difficult and sophisticated process. But thanks for computers we can handle it using different statistical packages. Accordingly, the following parameter estimates are obtained using S-Plus 2000 Software Package.

$$\tilde{X}_t = -.55675 \tilde{X}_{t-1} - 0.5408 \tilde{X}_{t-2} - 0.2943 \tilde{X}_{t-3}$$

Where $\tilde{X}_i$ represents the sale at the i^{th} period .

This result tells that the values of ϕ_1, ϕ_2 , and ϕ_3 are -0.55675 , -0.5408 and -0.2943 respectively. This model is a tentative model that relates the sale at time t with the values at periods $t-1$, $t-2$, and $t-3$. Nevertheless, this model cannot be directly used for forecasting purpose at least for two reasons. First the model is only tentative. It needs further investigation for its adequacy. This is what we call model diagnostic check. Second, even then the model was adequate it is presented in terms of the differenced series and hence we need to rearrange terms and express it in terms of the original series in order to make forecasts.

4.4 Model Diagnostic Checking

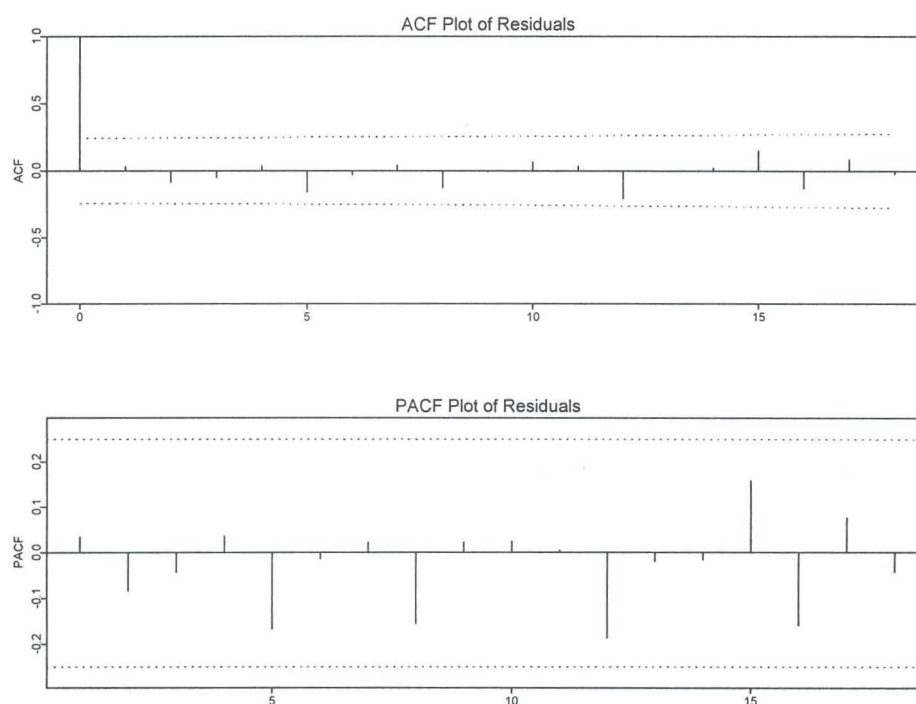
In model diagnostic checking we testify the adequacy of the proposed ARIMA model as a representative of the data under consideration. As it is stated in the methodology, there are different techniques that can be employed to perform the adequacy of a tentative model. For the purpose of this study we use the Autocorrelation check and the Portmanteau lack of fit test.

The plot of the autocorrelations of the residuals, presented in fig 6 shows that all the autocorrelations are insignificant as they all are within the reference limit. Residuals are the difference or deviation of the observations from the fitted value. A satisfactory model is the one that minimizes the residuals. The absence of significant autocorrelation among the residuals indicates that there is no implied pattern in the error terms. The error terms are random in their nature and do not exhibit any common pattern that could result from the inadequacy of the model. Putting differently, whenever the model is inadequate the influence of other influential factors or variables that are not considered in the proposed model would be included in the error terms and tends to introduce a recognizable pattern in the error terms. This recognizable pattern would definitely result in high autocorrelation of residuals. This is due to the fact that high autocorrelation will occur if and only if the residuals have some common behavior. Meaning all the error terms

would be influenced by a common essential variable not considered in the model and hence they would be correlated to one another.

Therefore, since the autocorrelations of the residuals do not exceed the significance limit we have no evidence to argue for the inadequacy of the model. That is, according to the autocorrelation of the residuals test we can conclude that the proposed tentative model can be regarded to be adequate.

ARIMA Model Diagnostics: MonthlySales\$SalesData



ARIMA(3,1,0) Model with Mean 0

Fig 6. Autocorrelation and Partial Autocorrelation plots of the residuals

On the other hand, the portmanteau test also confirms that the tentative ARIMA (3,1,0) model is adequate. The detailed hypothesis testing for the portmanteau test is presented in the annex. Here we simply present the plot of the P-Values for different number of observation considered in the test.

ARIMA Model Diagnostics: MonthlySales\$SalesData

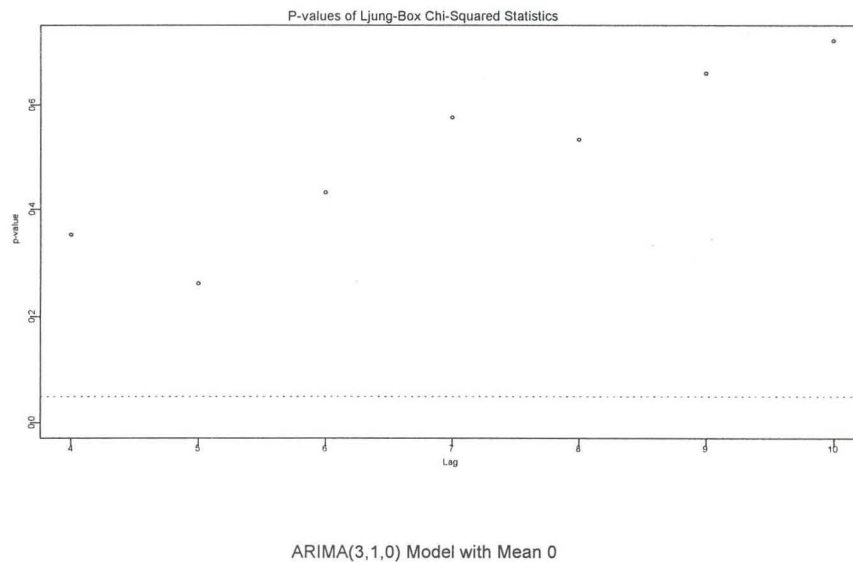


Fig 7. P-Values of the Chi-Squared Statistics in the Portmanteau test

The horizontal dotted line represents the 5% level of significance ($\alpha=0.05$). In addition, each point in the plot corresponds to the P-Value of the calculated Chi-squared value. Obviously, the higher the P-value is the smaller the calculated Chi-Square value. In other words, in terms of the P-Value we reject the null hypothesis if and only if the P-Value of the calculated Chi-Square is less than $\alpha=0.05$. As it can be seen from the above plot, however, all the P-Values of the calculated Chi-squares are

above the significance level ($\alpha=0.05$) suggesting that the null hypotheses stating “the model is adequate” is acceptable. If at least one point in the plot were to fall below the limit then we would say there is some evidence as to the inadequacy of the model.

For the sake of flexibility other similar models were tested (over fitting) for adequacy and all of them turns out to have less precise result than the ARIMA (3,1,0) model. This leads to the conclusion that the proposed ARIMA (3,1,0) is the most appropriate time series model that best fits the laundry soaps sales data collected from Gulele Soap Factory.

4.5 Interpretation of the final Model

From the analysis performed in the previous sections of this chapter we have determined that the appropriate time series model that represents the laundry soap sales data is ARIMA (3,1,0).

In its explicit form the model may be written as:

$$\tilde{X}_t = -0.55675 \tilde{X}_{t-1} - 0.5408 \tilde{X}_{t-2} - 0.2943 \tilde{X}_{t-3}$$

Where $\tilde{X}_i = X_i - X_{i-1}$

Though the model seems to refer to only the previous three consecutive lags, it actually refers to four consecutive lags. This is due to the fact that the model is presented in the form of the differenced series. If we try to convert the above model to the one that involves the original monthly sales rather than the differenced ones, we get the following model representation.

$$\hat{X}_t = 0.44325 X_{t-1} + 0.01595 X_{t-2} + 0.2465 X_{t-3} + 0.2943 X_{t-4}$$

Where $\hat{X}_t$ is the estimated or forecasted value of the actual value X_t .

As it can be observed from the above representation the forecasted or the estimated value of current sales depends on the previous four lags. Meaning, in order to make sales forecasts of the monthly sales of Gulele

laundry soap we need to take the previous four consecutive sales data in to consideration.

According to the model, the four most recent sales data determines the current sales data. In other words, it is found that the sales data for a given month is a function of the sales data for the proceeding four consecutive months.

The above result, forecasting sales based on the previous four lags, seems to be in line with the implication of the autocorrelation values presented in Table1. It can be seen that the degree of correlation between current sale and each one of the past sales data for the different lags is significant only for the first four lags. Except for the fifth lag all the remaining autocorrelation magnitudes are less than 0.5. This shows that the correlation is insignificant. In addition, the correlation between the fifth lag and the current sale is 0.53, which can also be regarded not significant as compared to the remaining lower lags. These significant autocorrelation values for the first four lags imply that current sales can be adequately forecasted from the previous four monthly sales data.

Lag	1	2	3	4	5	6	7	8	9	10
Auto-correlation	0.7732	0.6682	<u>0.6706</u>	0.6391	0.5374	0.4976	<u>0.5088</u>	0.4737	0.4628	<u>0.4725</u>
Lag	11	12	13	14	15	16	17	18	19	20
Auto-correlation	0.4487	0.3592	<u>0.3790</u>	0.3726	0.3392	0.2543	<u>0.2623</u>	0.2258	0.1574	0.1034

Table 1. Autocorrelation Values for the first 20 lags

It is also observed that all the coefficients in the model turn out to be positive. The implication being the sales volume at any given month is directly related to the previous sales records. In other words, if there is a general decreasing trend in sales volume of soaps in the past four months then the next month sales volume is also expected to decrease from its the current value, and vice versa. Clearly, this should be the case as we have observed positive autocorrelation values for different lags in Table1. Of course, this interpretation is based on the assumption that the influence of other external factors remains to be constant. Otherwise, if there is an extraordinary event occurring in any single month that affects sales volume and that has not been observed in the other months, then it is possible to come up with a reversed trend than the past. Such an event can be, for example, new contracts, unusual advertisement campaigns, and other promotional activities.

Fortunately, however, the impact of such new actions or factors may not extend for more than four months. This is because the model will take the action made in to consideration through sales observed after the happening of the new event. Meaning the influence of the new happening will be reflected in the sales data recorded for four or more consecutive months so that the model will take it in to consideration. Hence, after that point the model may serve as a forecasting tool with no or little modification. Nevertheless, if the change is going to happen frequently then it will be necessary to involve the changing variable into the model, which leads us to formulating a multivariate model that involves both time series and econometric data.

In addition, the sum of all the coefficients adds up to the value one. This result can be interpreted in such a way that if the sales for the preceding four months are identical, then the sales for the current period will remain to be the same. In this case we will have a constant sales forecast each month. As a result, the model will generate the same result as that of the naïve approach.

Another interesting finding is that the sale at time t is influenced by or dependent much on the previous month, i.e. $t-1$, sale than the other months. This is observed from the relative comparison of magnitude of the different coefficients. The value 0.44325, which is the coefficient of the most recent sales data relative to the current period, is the greatest

as compared to other coefficients. Through this result the model tries to suggest that the most influential factor in determining current sale is the sale recorded for the previous month. The model attaches the highest weight to the most recent data, which is the observed sale at the period $t-1$.

This finding is logical because for any two consecutive months the probability of having same pattern is high due to the lesser tendency of change to occur in such short period. It is to mean that recent sales data usually have more similar behavior with current or future sale as compared to other relatively older sales data. After all, in time series analysis the basic assumption is that current sales can be forecasted from previous sales. In other words, it is accepted that the major factor that affects sales is the time factor. Moreover, it is natural to expect that recent sales data will give more information than old data in forecasting future sales. This is for the reason that if time is assumed to be the determinant variable in forecasting sales then the shorter the time between an observed sale and current sale the higher will be their correlation. High correlation means that current sales can be forecasted well by using recent sales data that occur at low lags. With this regard it can be seen from table1 that the highest correlation occurs for the first lag. Thus, it is convincing for the coefficient of the first lag to be highest as compared to the others.

Further more, it is natural that the environmental and other influential factors tend to remain stable in the short run. In the case the only considerable change comes from the time factor. Hence, the impact of other influential factors will be more or less similar for monthly sales that are consecutive to one another as compared to those that are apart from one another.

On the other hand, it is observed from the model that the coefficients for the fourth and third lags, i.e. X_{t-3} and X_{t-4} , are relatively greater than the coefficient for the second lag X_{t-2} . The analysis indicates that the fourth and third lags seem to have greater relevance or contribution in forecasting the current sales volume than the sales value recorded in the second lag. This may leave an impression that the result is in contrast with the above discussion that as far as the time factor is concerned more recent data has greater relevance in the forecasting current sales. Differently put, we expect that the second coefficient to have a greater coefficient than the third and the fourth coefficients because the second lag is more near to the current period than the fourth and the third lags.

Nevertheless, the above result is not a contradiction to the general assumption that greater importance is usually attached to more recent data than relatively old sales data. The justification is that we come up with such a result not due to the time factor but due to some implied seasonal factor that affects the monthly sales. It can be observed from

table 1 that there is a general fall in the value of the autocorrelation. An exception, however, is that the autocorrelation shows increment or rise, rather than keeping the decreasing trend, for the fourth, seventh, tenth, thirteenth, and seventeenth lags. The above high autocorrelations occur at fairly constant interval, which is approximately 3.5 months. This is a clear indication that there is some special correlation between current sales and sales at approximately the third or fourth lags. The implication is that there is some seasonal factor that occurs at approximately 3.5 months time, which is its periodicity. This seasonal factor, therefore, makes the third and fourth lags to be more relevant than the second lag in forecasting current sales and its impact is more than that of the impact induced by the time factor. Had it not been for the existence of this seasonal factor the third and fourth lags might not been considered in the model. There are several seasonal or cyclical factors that affect sales among which periodical consumer purchase, periodical advertisement, and other periodical actions taken by competitors are only few to mention.

Another very important point is that it is not as such possible to interpret the coefficients of the developed time series model in the same way as in the case of regression analysis. One cannot consider the coefficients as the amount by which current sales will change for a unit change in the value of previous sales value that corresponds to the coefficient under consideration. The restrictive factor is that in time

series analysis we are considering the explanatory variables to be past sales that are of course not under our control. They are not as such inputs that can be controlled in order to achieve a required current or future sales. Had it been for the model to include other explanatory variables like advertisement expense, there would have been the possibility to control current or future sales by controlling or adjusting the advertisement expense.

In fact, there is no direct or explicit interpretation associated to the coefficients of a time series analysis unlike their counter parts in regression analysis. After all the major purpose of time series analysis is to serve as a forecasting tool rather than explaining the variation or degree of responsiveness of the dependent variable to the explanatory variables.

The first coefficient, 0.44325, that corresponds to the sales at the period $t-1$ indicates that approximately 44.3 % of the sales volume in the first lag will be carried over to the current sales. In other words, the volume of soap sold at any two successive months is found to have significant interrelationship. The amount of sales that occurred at a given month will also significantly determine or affect the sales volume that is going to be observed in the next month. This high percentage value indicates that the sales volume observed at lag 1 has significant impact of the value of

current sales. The greater the sales volume at lag1 (previous sales volume) the greater will be the expected sale at the current period.

On the other hand, the coefficient of the fourth lag, which is 0.2943, indicates that about 30% of the sales volume recorded at the fourth lag will be carried over the current sales volume. That is, the fourth lag also has significant contribution in determining or forecasting the value of current sales volume. According to the model, there seems to be high interrelationship between the current sales and the sales data at the fourth lag than the third and second lags. As compared to the weight associated to the second and third lags, 1.6% and 25% respectively, the weight for the fourth lag is more significant. This could be an indication of the existence of an external variable that influence the sales data once in four months and resulted in introducing common behavior in the pattern for the current and fourth lag sales data. In fact, this is also the reason why the model extends to the fourth lag. The usual number of lags that are considered in a forecasting model is one or two.

One cannot conclude, however, that since the percentage contribution of the second lag is small its significance is also small. It has to be recalled that smaller percentage may result in greater actual or absolute figure if the sales in the second lag turns out to be greater as compared to the other lag sales records.

Chapter Five

Forecasting

The purpose of this chapter is to generate forecasts for future sales and examine the forecasting capability of the model. In the previous chapter we have developed an adequate ARIMA (3,1,0) model that can be used for forecasting purpose. The adequacy of a model, however, is not a guarantee for its capability of forecasting.

It is, therefore, essential to examine the forecasting power of the identified model. Strictly speaking there is no hard and fast rule for evaluating the forecasting power of a model. Usually we refer to the standard error of forecasts and comment on the forecasting power of the model, which is quite subjective and case dependent. To that end we recall that the identified model have the form:

$$\hat{X}_t = 0.44325X_{t-1} + 0.01595X_{t-2} + 0.2465X_{t-3} + 0.2943X_{t-4}$$

Moreover, this model is obtained using the monthly sales data for the 68 months ranging from January 1997 to August 2002. The remaining monthly sales data for the five months that range from September 2002 to January 2003 are used here for forecasting purpose.

To generate the forecasted sales for the month September, which is the first lead-time, we substitute the actual values of the previous four months (May, June, July and August) in to the model. On the other hand, when it comes to the sales forecast of October we do not have the actual value for the previous month, i.e. September, and hence we use its forecasted value in place of the actual value. The forecasting process will then continue by the same manner. Meaning, whenever there is actual value we use it while forecasting the sales of any given month from a fixed forecasting point. Otherwise, if the actual value is not yet achieved or recorded (reached) then we use the corresponding forecasted value.

The following table summarizes the forecasts for the five lead times.

Month	Actual Sales	Forecasted Sales	Absolute Forecast Error
September 2002	3441	3449	8
October 2002	3150	3215	65
November 2002	3544	3187	357
December 2002	3196	3090	106
January 2003	3780	3228	552

Table 2. Forecasted monthly Sales of Laundry Soap measured in terms of Cartoons.

An overall look at the above result bears an impression that whenever the lead-time increases the accuracy of the forecasts tends to decline. For example, the forecast error for the first lead-time, which is 8, is smaller when it is compared to the error for the forecast of October, which is 65.

This implies that the forecast done for the first month or the first lead-time is more precise as compared to that of the second, and so on. In other words it is observed that whenever the lead-time increases the error also increases. The justification for this pattern is that whenever the lead-time gets larger and larger there will be increased probability for different changes to take place. It is evident that a number of factors may bring changes in the pattern of sales data whenever the forecast interval gets wider.

It is generally agreed that the time factor can be a determinant variable for forecasting sales data but not the only. There are other variables that readily affect the sales in a given month. With this regard the decline in the accuracy of the forecasts for the sales of the laundry soap produced by Gulele Soap Factory can be attributed to the change that has been introduced in terms of advertisements or other promotional activities. It is true that the factory carry out advertisements in media like television and radio. Nonetheless, since the pattern of the advertisement carried out by the factory lacks consistency and planning it is hardly possible to take this factor into account. Obviously, unless there is consistent pattern of advertisement and other promotional activities, it would be very difficult to detect the impact of such activities on the sales volume.

This behavior of the model is of course in line with the previous findings that Box-Jenkins time series analysis is suggested as an effective tool for short-term forecasting. [Geurts and Ibrahim 1975:182]

The higher accuracy of the forecasts for the first and second lead times indicates that the sales data for soaps in the factory can be forecasted more satisfactorily in the short run. In other words, the sales of the factory are dynamic that it is not possible to accurately forecast the sales for the long run. Another possible explanation is in relation to the methodology used. As it is stated previously the Box-Jenkins approach is best suited for short term forecasting than for long term forecasting.

The smaller magnitude of the absolute error for the forecast of the month can be attributed to the consequence of the interdependence between the current period (the base period) and the fourth lead-time (X_{t+4}). We recall that from the analysis of the autocorrelation, in the previous chapter, it has been found that there is some seasonal factor or component that occurs with 4 months of periodicity. Therefore, due to the existence of such common seasonal component the forecast for the fourth lead-time resulted in higher accuracy- lower absolute error.

Another consideration of the forecasting process is that of error analysis. Till now we were dealing with the absolute deviations of the forecasted values from the actual values. This approach, however, is more crude measure for evaluating the forecasting power of a model. Though there

are a number of other related error analysis techniques the most commonly used technique for evaluating the forecasting accuracy of a model is the U-Statistics. The interpretation of the U-statistics can be considered for three distinct cases. Although the interpretation is adapted for the purpose of this specific study, it can also be extended to the evaluation of any forecasting technique.

Case 1: $U=1$ implies that the naïve forecasting method is as good as the Box Jenkins approach.

Case 2: $U<1$ confirms that the Box-Jenkins forecasting method has outperformed the naïve forecasting.

Case 3: $U>1$ suggests that the naïve forecasting approach outperforms the Box-Jenkins approach.

For the soaps sales data on hand the U-statistics is computed to be 0.88. The interpretation of this figure is, therefore, that the derived Box-Jenkins time series model outperforms the naïve approach. In other words, the failure of the naïve approach to outperform the Box-Jenkins approach leads to a conclusion that the latter should be preferred. Nevertheless, this conclusion applies only to the forecasts of the monthly sales of laundry soaps by Gulele soap factory; it cannot be in general extended to other time series data.

Therefore, it can be said that the developed model can be used as a satisfactory tool for forecasting the monthly sales of soaps produced by Gullele Garment factory. Particularly the forecasting model is found to be more powerful for short forecasting horizons than for long forecasting intervals.

The final point with regard to forecasting is the process of updating forecasts. The idea of updating is that whenever new sales data comes to existence then it should be used in forecasting other future values. For instance when we forecast the sales volume for the month October we have used the forecasted sales volume for September. Now in updating forecasts we use the actual value of the sales for September instead of the forecasted value. It is obvious that the updating process usually improves the forecasting process. Accordingly, the following updated forecasts are obtained for the five lead times considered previously.

Month	Actual Sales	Forecasted Sales	Absolute Forecast Error
September 2002	3441	3448.6	7.6
October 2002	3150	3211.2	61.2
November 2002	3544	3157.7	386.3
December 2002	3196	3245.1	49.1
January 2003	3780	3262.3	517.7

Table 3. Updated Forecasted monthly Sales of Laundry Soap

Chapter Six

Summary, Conclusion and Recommendation

6.1 Summary

Business managers of the day are involved in making complex and critical decision-making process. Few of these decisions by and large determine the success or failure of the organization. In the process of decision-making and strategy formulation managers use different indicators among which sales forecasting is the interest of this study. Particularly, this study concentrates on forecasting monthly sales of laundry soap produced and sold by Gulele Soap Factory.

Despite the indispensable role played by sales forecasting in decision-making and strategy formulation, there has been no formal means of sales forecasting developed or used in Gulele Soap factory that enables management forecast monthly sales of laundry soaps. This situation has brought problem in the decision making process of the management of the factory. The management usually encounters problems in making decisions about future actions in connection with raw material requirement planning, operations planning and financing, human resource planning, marketing activities, etc.

The purpose of the study is, therefore, to provide a possible solution to the aforementioned problem(s). To that end the objective of the research is to develop a sound sales forecasting model that is capable of forecasting monthly sales with satisfactory accuracy level. In addition, the paper addresses the question that to what extent univariate time series models are powerful tools (approaches) towards forecasting sales.

Moreover the importance of the study is stated to be:

- It provides the management of Gulele soap factory with a more reliable and appropriate forecasts that obviously improve, or at least facilitate, the decision making task.
- It serves as a model that greatly helps other similar business organizations in Ethiopia in developing their own forecasting models through time series analysis. In other words the paper clearly illustrates how time series analysis can be applied in sales forecasting provided all the required assumptions are satisfied.
- It serves as a reference material for future researches, i.e., it adds something to the available literatures especially in developing countries like Ethiopia.

In order to achieve its basic objective the study applies the Box-Jenkins Time Series Analysis approach on the monthly sales of laundry soaps produced and sold by Gulele Soap Factor. A total of 73 monthly sales

data are collected from which the last five are reserved for forecasting purpose.

Review of related literatures is made in the second chapter of the material. The review is made in relation to the overall views and findings of sales forecasting. Moreover, the different approaches to forecasting that have been suggested by different authors are also incorporated in the review.

The third chapter is dedicated to highlight the cores of Box-Jenkins approach to time series analysis. It is stated that the Box-Jenkins approach involves three model building stages namely, Model Identification, Model Parameter(s) Estimation, and Model Diagnostic Checking. The present study presents a fairly detailed discussion on each one of these three stages.

The analysis and discussion of results, which is the major team of any study, is presented in the fourth chapter. The analysis is performed using S-Plus 2000 statistical package and gives the result that the monthly sales pattern of soaps sold by Gulele Soap Factory can be represented by Autoregressive Integrated Moving Average model of order 3,1,0 (ARIMA (3,1,0)). The general model in its explicit form is found to have the form $\hat{X}_t = 0.44325X_{t-1} + 0.01595X_{t-2} + 0.2465X_{t-3} + 0.2943X_{t-4}$.

Furthermore, the model is testified for its forecasting power or accuracy using the reserved five months sales data. As stipulated in chapter five, the model is found to be satisfactorily powerful for the intended forecasting purpose. Besides, the model is also compared with the naïve forecasting method, which is assumed to be the simplest and also less costly method. With this regard the U-Statistics confirmed that the identified ARIMA (3,1,0) model performs better as compared to the naïve forecasting method.

6.2 Conclusion

The data analysis and resulting findings lead to the following major conclusions.

- The monthly sales of soaps can be forecasted using the ARIMA (3,1,0) model with a satisfactory degree of accuracy. Moreover, the sales data for the previous four lags is vital in making current sales forecast.
- The suggested model seems to serve better for short forecasting horizons. Whenever the forecasting interval gets larger the accuracy of forecasts are observed to reduce significantly.
- There is a seasonal component pertaining to the sales data considered that affects the monthly sales pattern. The periodicity of this seasonal factor is approximated to be between 3 and 4 months.
- The sales data recorded for the first lag (X_{t-1}) is the major determinant factor in forecasting current sales (X_t). That is, the time factor has a significant role in forecasting sales.

6.3 Recommendation

Finally, the following recommendations are forwarded vis-à-vis the problem that has been researched based on the overall findings of the study.

- Management of Gulele Soap Factory may facilitate its decision making task by conducting sales forecasting based on the suggested ARIMA (3,1,0) model.
- The suggested forecasting model is most appropriate to be used for forecasting interval(s) preferably not more than three months.
- The model should be updated with some interval using the new sales record for the new months that has not been included within the model formulation process. The model should also be revised in cases where there is some significant change in the environment that is believed to have some sort of impact on sales of laundry soaps produced by the factory.
- Although more than satisfactory results can be obtained, pure time series may not be sufficient in conducting sales forecasting. In addition to the time factor, it is extremely advantageous to incorporate other independent variables that readily affect sales volume. One such critical variable could be advertisement expense

both by the factory and also by other competitors in the same industry.

- Other companies may also use the same approach or procedure for building their own forecasting model that satisfactorily forecast their future sales, and hence facilitating the overall effectiveness of the companies.

Bibliography

- Box and Jenkins. "*Time Series Analysis: Forecasting and Control*," Revised Edition, California: Holden-Day, 1976.
- Chatfield, "*The Analysis of Time Series: An Introduction*," Fifth Edition, UK: Chapman and Hall, 1996.
- Geurts D. and Ibrahim B., "*Comparing the Box-Jenkins Approach with the Exponential Smoothing Forecasting Model Application to Hawaii Tourists*." *Journal of Marketing Research*, Vol. XII, May 1975(182-187).
- Green and Tull. "*Research for Marketing Decisions*," Second Ed. USA: Prentice-Hall, 1970.
- Heizer and Render. "*Principles of Operations Management*." Third Ed, USA: Prentice-Hall, 1999.
- Helmer M. and Johansson K., "*An Exposition of the Box-Jenkins Transfer Function Analysis with an Application to the Advertising-Sales Relationship*." *Journal of Marketing Research*, Vol. XIV, May 1977(227-239).
- Kappoor S.G., Madhok P., and WU, "*Modeling and Forecasting Sales Data by Time Series Analysis*." *Journal of Marketing Research*, Feb. 1981.
- Keay F. "*Marketing and Sales Forecasting*." UK: Pergamon Press, 1972.
- Kotlet P. "*Marketing Management*," Third Ed., USA: Prentice Hall Inc., 1967.

- Lancaster and Reynolds. *"Introduction to Marketing: A Step-By-Step Guide to all the Tools of Marketing."* UK, 1999.
- Lancaster and Reynolds. *"Marketing."* UK: Reed Educational and Professional Pub., 1998.
- Leone P. *"Modeling Sales-Advertising Relationship: An Integrated Time Series Econometric Approach."* Journal of Marketing Research, Vol. XX, Aug. 1983(291-295)
- Moriarty and Adams, *"Issues in Sales Territory Modeling and Forecasting Using Box-Jenkins Analysis."* Journal of Marketing Research, Vol. XVI, May 1979(221-232)
- Moriarty and Gerald Salamon. *"Estimating and Forecasting Performance of a Multivariate Time Series Model of Sales."* Journal of Marketing Research, Vol. XVII Nov. 1980(558-564)
- Murdick G., and Schaefer E. *"Sales Forecasting for Lower Costs and Higher Profits."* USA: Prentice-Hall Inc., 1967.
- Parsons W. *"Improving Marketing Performance."* UK: Gower Publishing, 1987.
- Smith and Shelby. *"A two Stage Sales Forecasting Procedure Using Discounted Least Squares."* Journal of Marketing Research, Vol. XXXI, Feb. 1994(44-56)
- Umashanhar and Johannes. *"Forecasting with Diagonal Multiple time Series Models: an Extension of Univariate Models."* Journal of Marketing Research, Vol. XX, Feb 1983(58-63).

ANNEX

Portmanteau Lack Of Fit Test

r	r ²	lag	r	r ²	lag	r	r ²	lag	r	r ²
0.033	0.001	6	-0.029	0.001	11	0.036	0.001	16	-0.130	0.017
-0.083	0.007	7	0.043	0.002	12	-0.205	0.042	17	0.087	0.008
-0.049	0.002	8	-0.124	0.015	13	-0.001	0.000	18	-0.021	0.000
0.038	0.001	9	-0.006	0.000	14	0.023	0.001			
-0.155	0.024	10	0.066	0.004	15	0.154	0.024			

Autocorrelation of the residuals for 18 successive lags

As it is stated in chapter three the portmanteau approach we are interested to test the null hypothesis

H₀: *The model is adequate*

Versus its alternative

H_A: *The model is not adequate*

From the above table it can be calculated that

$$\sum_{k=1}^{18} r^2 = 0.151 \text{ where } m=18 \text{ is the number of lags taken.}$$

Therefore,

$$Q = n \sum_{k=1}^m \gamma_k^2(\hat{e}) = 67(0.151) = 10.117 \text{ where } n = N - d = 68 - 1 = 67$$

The tabulated 5% and 10% Chi-square values with $m-p-q=18-3-0=15$ degrees of freedom are 25 and 22.3 respectively.

The rejection criterion or rule is to reject the null hypothesis if

$$Q > \chi_{5,10}^2(15)$$

But then the calculated value of Q is less than that of the tabulated chi square value both at 5% and 10% levels of significance.

Therefore, we accept the null hypothesis that the model is adequate.

Monthly sales data of Laudry Soap(Measured in Cartoons)

Jan-97	33772		Sep-98	33976		May-00	7212		Jan-02	3237
Feb-97	32469		Oct-98	24008		Jun-00	9878		Feb-02	4321
Mar-97	35074		Nov-98	14039		Jul-00	7996		Mar-02	6558
Apr-97	47494		Dec-98	19023		Aug-00	8184		Apr-02	8795
May-97	30801		Jan-99	21663		Sep-00	14543		May-02	5394
Jun-97	36760		Feb-99	21789		Oct-00	12022		Jun-02	2578
Jul-97	27672		Mar-99	31802		Nov-00	8342		Jul-02	3591
Aug-97	23127		Apr-99	29532		Dec-00	9501		Aug-02	2636
Sep-97	18583		May-99	14641		Jan-01	7183		Sep-02	3441
Oct-97	16973		Jun-99	21671		Feb-01	4865		Oct-02	3150
Nov-97	21444		Jul-99	28700		Mar-01	5057		Nov-02	3544
Dec-97	32103		Aug-99	15302		Apr-01	6894		Dec-02	3196
Jan-98	25682		Sep-99	8694		May-01	7140		Jan-03	3780
Feb-98	19261		Oct-99	27254		Jun-01	5232			
Mar-98	23275		Nov-99	17740		Jul-01	5397			
Apr-98	25610		Dec-99	13368		Aug-01	4587			
May-98	34688		Jan-00	7008		Sep-01	5299			
Jun-98	21064		Feb-00	21912		Oct-01	5269			
Jul-98	23985		Mar-00	14371		Nov-01	5203			
Aug-98	21669		Apr-00	4663		Dec-01	5077			